A Higher Order Expansion for the Joint Density of the Sum and the Maximum with Applications to the Estimation of Climatological Trends

by
Amy Marie Grady

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics.

Chapel Hill
2000

Approved by

Advisor: Professor Richard L. Smith

Reader: Professor Amarjit Budhiraja

Reader: Professor Edward Carlstein

Reader: Professor Peter Robinson

Reader: Professor Gordon Simons

ABSTRACT

Amy Marie Grady

A Higher Order Expansion for the Joint Density of the Sum and the Maximum with Applications to the Estimation of Climatological Trends

(Under the direction of Richard L. Smith)

The higher order expansion for the joint density of the sum and maximum of an *iid* sequence answers two questions: the theoretical question of what is the rate of the asymptotic independence between these two which was established by Chow and Teugels (1978) and Anderson and Turkman (1991), and the practical question of how to describe and model the dependence when the asymptotic result is not yet realized. Developing such an expansion under the three different domains of attraction for the maximum and modeling the annual total and maximum precipitation across the contiguous US, using a combined generalized extreme value (GEV) version of the expansion, are the key elements of this thesis.

The three important developments necessary for the derivation of the expansion of the joint density are as follows. First we develop an Edgeworth expansion for the density of the sum given the maximum. Second we establish expansions for the density of the maximum under the Gumbel, the Fréchet, and the Weibull domains of attraction. Finally, we calculate the first two conditional moments present in the above Edgeworth expansion. The key in the last step is reformulating the moments in terms of exceedances over a threshold.

Finally in the data analysis section we give a heuristic development of the expansion of the joint density for the combined GEV version. We then model the annual total and maximum rainfall across the continental US. The relevancy of such an expansion for the joint density of the sum and the maximum in studying US precipitation and to larger climate change questions lies in the emergence of climate variability and/or extreme events as having an important impact on the overall climate. In particular, at the US level, we address issues raised in Karl *et al.* (1995, 1996) whose premise is that the climate in the US is becoming more extreme. With respect to the precipitation in the contiguous US, Karl and Knight (1998) found that the increase seen in the annual total rainfall is being driven by an increase in the upper 10% of rainfall events, suggesting the annual total rainfall and extreme rainfalls are dependent. Using the above model which incorporates this first order approximation to the dependence between the annual total and annual maximum rainfall, we find substantial evidence for a positive trend in both with the trend in the annual total rainfall being more dominant both in terms of significant evidence and magnitude. Since the beginning of the 20th century, we estimate a 7.4% (s.e. 0.1%) increase in the national average of annual total rainfall and a 3.0% (s.e. 0.1%) increase in the national average of annual maximum rainfall.

To those who

truly believed in me,

and encouraged me,

and supported me

down this long path.

CONTENTS

1 INTRODUCTION AND LITERARY REVIEW **1**

1.1 Introduction 1

1.2 Theoretical Literary Review 2

1.2.1 Limit Laws for Sums 2

1.2.2 Limit Laws for Maxima 5

1.2.3 Limit Laws for the Joint Distribution of the Sum and Maximum . 27

1.3 Climate Literary Review 30

2 EXPANSION OF THE JOINT DENSITY UNDER THE GUMBEL DOMAIN OF ATTRACTION **43**

2.1 Introduction 43

2.2 Main Theorem 48

2.3 Lemmas 51

2.3.1 Lemmas for Proposition 50 52

2.3.2 Lemmas for Proposition 51 61

2.3.3 Lemmas for Proposition 53 62

2.4 Propositions 64

2.4.1 Expansion of the Conditional Density of Sum Given the Maximum 64

2.4.2 Expansion of the Density of the Maximum 68

2.4.3 Expansions Involving the Conditional Mean and Variance 77

2.5 Proof of Main Theorem and Its Corollary 82

2.6 Simulation Project 92

2.6.1 Objective 92

2.6.2 Form 93

2.6.3 Procedure 94
2.6.4 Issues to Resolve in the Simulation Project 95
2.6.5 Results 98
2.6.6 Conclusions 99

3 EXPANSION OF THE JOINT DENSITY UNDER THE FRÉCHET DOMAIN OF ATTRACTION 106

3.1 Introduction 106
3.2 Main Theorem 110
3.3 Propositions 111
3.3.1 Expansion of the Conditional Density of the Sum Given the Maximum 111
3.3.2 Expansion of the Density of the Maximum 111
3.3.3 Expansions of Conditional Mean and Variance 122
3.4 Proof of Main Theorem and Its Corollary 125

4 EXPANSION OF THE JOINT DENSITY UNDER THE WEIBULL DOMAIN OF ATTRACTION 136

4.1 Introduction 136
4.2 Main Theorem 140
4.3 Propositions 141
4.3.1 Expansion of Conditional Density of the Sum given the Maximum 141
4.3.2 Expansion of Density of the Maximum 141
4.3.3 Expansions of Conditional Mean and Variance 147
4.4 Proof of Main Theorem and Its Corollary 149

5 DATA ANALYSIS OF CONTINENTAL US RAINFALL 160

5.1 Introduction 160
5.2 Review of Precipitation Studies 161
5.3 The Data 164

5.4 Models . . . 166
5.4.1 Notation . . . 166
5.4.2 Independent Case . . . 167
5.4.3 Expansion Case . . . 169
5.5 Statistical Procedure . . . 175
5.5.1 Level of Analysis . . . 175
5.5.2 Method . . . 175
5.5.3 Form of the Log Likelihood Functions . . . 176
5.5.4 Maximum Likelihood in Action . . . 181
5.6 Data Analysis . . . 183
5.6.1 Descriptive Statistics . . . 183
5.6.2 Introductory Model Specification . . . 188
5.6.3 Model Analysis . . . 196
5.6.4 Applying the Results of the Data Analysis . . . 227
5.7 Comparison of Results to Other Findings . . . 236

BIBLIOGRAPHY **245**

LIST OF TABLES

5.1 Results of the AR models for annual maximum rainfall series. 190

5.2 Results of the AR models for annual total rainfall series. 190

5.3 Results on comparing Gumbel distribution to GEV distribution. 191

5.4 Results on the Kolmogorov-Smirnov test for the sum. 193

5.5 Models run in the trend analysis. 198

5.6 Standard error comparison of parameter estimates in the independent base model and expansion base model. 200

5.7 Likelihood ratio tests to be performed. 201

5.8 Results of the likelihood ratio tests in the independent case. 203

5.9 Results of the likelihood ratio test in the expansion case. 204

5.10 Results of t-tests of trend parameters. 208

5.11 Spatial model results for γ in expansion case. 216

5.12 Spatial model results for β in expansion case. 217

5.13 Spatial model results for τ in expansion case. 219

5.14 Spatial model results for γ in independent case. 220

5.15 Spatial model results for β in independent case. 220

5.16 Spatial model results for τ in independent case. 221

5.17 National averages of the trend estimates produced by spatial model with polynomial of order 0. 222

5.18 National averages of the trend estimates produced by spatial model with polynomial of order 4. 223

5.19 Some individual station results for the trend estimates and their associated likelihood ratio tests. 234

LIST OF FIGURES

2.1 Finite bin width consideration in the simulation project. 96

2.2 Effect of increasing standard deviation within underlying distribution. 101

2.3 Effects of sample size on expansion of joint density. Normal distribution with $\mu = 0$ and $\sigma = 1$ as sample size increases from: (a) 30, (b) 60, (c) 90, to (d)120. 102

2.4 Effects of sample size on expansion of joint density. Lognormal distribution with standard deviation 0.50 as sample size increases from: (a) 30, (b) 60, (c) 90, to (d)120. 103

2.5 Effects of sample size on expansion of joint density. Weibull distribution with $\alpha = 1$ and $\beta = 1/2$ as sample size increases from: (a) 30, (b) 60, (c) 90, to (d)120. 104

2.6 Effects of sample size on expansion of joint density. Weibull distribution with $\alpha = \beta = 1$ [Exponential distribution] as sample size increases from: (a) 30, (b) 60, (c) 90, to (d)120. 105

5.1 Map of the contiguous United States of America showing the five regions defined for this analysis and the 187 stations in the data set. 165

5.2 Box-plot for the annual maximum rainfall in the five regions. The units are in hundredths of an inch. 184

5.3 Box-plot for the annual total rainfall in the five regions. The units are in hundredths of an inch. 184

5.4 The average annual maximum rainfall across the US each year from 1901 to 1997. 185

5.5 The average annual total rainfall across the US each year from 1901 to 1997. 186

5.6 The annual total rainfall to annual maximum rainfall for stations: (a) Berkeley, CA (b) Caldwell, ID (c) Le Mars, IA (d) Eastport, ME (e) Savannah, GE. 238

5.7 The annual maximum rainfall for the years of record in stations: (a) Berkeley, CA (b) Caldwell, ID (c) Le Mars, IA (d) Eastport, ME (e) Savannah, GE..239

5.8 The annual total rainfall for the years of record in stations: (a) Berkeley, CA (b) Caldwell, ID (c) Le Mars, IA (d) Eastport, ME (e) Savannah, GE. 240
5.9 Gumbel plots for stations: (a) Albany, Texas (b) Big Timber,Montana (c) Princeton, Indiana (d) Fredonia, New York (e) St. Leo, Florida and (f) Goldsboro, North Carolina. 241
5.10 Quantile plots for stations: (a) Goldsboro, North Carolina (b) Ojai, California (c) Fort Collins, Colorado (d) Dillion, Minnesota (e) Fredonia, New York (f) St. Leo, Florida. 242
5.11 The parameter estimates in model I1 plotted against the parameter estimates in model E1: (a) η (b) ψ (c) μ (d) σ. 243
5.12 Mapped estimates for (a.) β and (b.) γ in model E5. 243

Chapter 1

INTRODUCTION AND LITERARY REVIEW

1.1 Introduction

To study the behavior of the sum and the maximum of a series, the literature shows that these two are asymptotically independent – see Chow and Teugels (1978) and Anderson and Turkman (1991). In other words, we should – arguably – model them separately. There are problems both theoretical and practical with that approach. Theoretically, we know that the convergence rate of the distribution of the maximum to a limiting extreme value distribution is relatively slow and hence this should impede the rate to this asymptotic independence. In many applications, the dependence between these two can be shown; that is, the asymptotic independence is yet to be realized. For example, Karl and Knight (1998) show that, in the precipitation series across the US, an increase in the extremes is driving the increase in the total annual rainfall. Moreover, often the important question to study is what is this dependence structure between the sum and maximum? For example, one of the predictions by climate models about a warmer world due to global warming is for an increase in the global mean precipitation with more intense storms. So we can ask: Given an increase in the largest daily rainfalls, how will that effect total amount of annual rainfall? In theoretical terms, the problem is to calculate a higher order term in this asymptotic joint distribution of the sum and the maximum. Specifically since we wish to model the sum and the maximum via maximum likelihood estimation, we really need to establish a *higher order term* associated with the asymptotic joint *density* of the sum and the maximum.

This thesis deals with this higher order expansion of the joint density of the sum and maximum – both establishing it and then modeling precipitation across the contiguous US using it. In fact, this thesis breaks down into two parts. In Chapters 2, 3, and 4, we derive the higher order expansion for the joint density of the sum and the maximum for the three domains of attraction of the maximum. In Chapter 5, we look at a data analysis of the annual precipitation series across the contiguous US. Here, in Chapter 1, we begin by reviewing the pertinent literature on both the theoretical and application parts. Section 2 reviews the theoretical publications. This includes: Subsection 1 which reviews limit laws for sums; Subsection 2 which reviews limit laws for maximum including introducing the "three types" of limiting distributions for both the *iid* and stationary sequence cases, the expansions of these asymptotic distributions, local laws, the limiting distribution of exceedances, and the penultimate approximation; and Subsection 3 which reviews limit laws for the joint distribution of the sum and the maximum. Finally, Section 3 reviews the relevant results on global warming – in particular, results concerning the precipitation across the contiguous US.

1.2 Theoretical Literary Review

1.2.1 Limit Laws for Sums

Limit laws for sums – central limit theory results – are fundamental to statistical science both in the depth of research and breadth of application. In 1718, de Moivre established the first central limit theory results for the sum of binary random variables with $p = 1/2$. Laplace proved the result for general proportions in 1812 but it would be 1887 until Chebyshev solved for the arbitrary sum results. Liapunov is credited with the first modern discussion of the central limit theorem which included the first rates of convergence ideas. His proofs entailed characteristic functions. This work dates to 1900 and 1901. For further background, see Johnson and Kotz (1982), Vol 2, p. 651-655.

Given the volume of research on limit laws for sums during the 20th century, here we limit our review to results pertinent to the work of this thesis. The focus of this work is to establish a higher order expansion for the joint density of the sum and the maximum. In particular, the derivation of the expansion of the joint density proceeds by multiplying an expansion of the conditional density of the sum given the maximum by an expansion for the density of the maximum. Specifically, we need to make use of central limit theorems for the non-identical case since our result must apply to the conditional density of the sum given the maximum where the value of the maximum can vary as the limit is taken. We also need to focus on results for densities as opposed to distributions. In particular we need to focus on expansions for the densities of the sum (given the maximum).

We begin with the triangular array form of the Lindeberg-Feller central limit theorem, taken from Hall (1982). This is the most general form for the non-identical case. Suppose that for each n the variables $X_{1n}, \ldots, X_{nn}$ are independent with zero means and are normalized so that their variances add up to one; i.e.,

$$\sum_{j=1}^{n} E(X_{nj}^2) = 1 \quad n \geq 1.$$

Let $S_n = \sum_{j=1}^{n} X_{nj}$ and define the uniformly asymptotically negligible (UAN) condition on the variances as

$$max_{1 \leq j \leq n} E(X_{nj})^2 \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty. \tag{1.1}$$

Theorem 1 (Thm 2.1, Hall, 1982) *(1.1) holds and S_n is asymptotically normal if and only if*

$$\forall \epsilon > 0 \quad \sum_{j=1}^{n} E\{X_{nj}^2 I(|X_{nj}| > \epsilon)\} \rightarrow 0 \quad as \quad n \rightarrow \infty. \tag{1.2}$$

Note the sufficiency part was first solved by Lindeberg in 1922 with the necessary part solved by Feller in 1935. See Johnson and Kotz (1982) Vol. 4, p. 651 for further details.

Now to present the necessary central limit result for densities, which is the starting point for the results of this paper, we introduce some new definitions. Here we let $X_1, \ldots, X_n$ be mutually independent random variables with common distribution function F and characteristic function ϕ. We suppose $EX_j = 0$ and $EX_j^2 = 1$ and put $S_n = X_1 + \ldots + X_n$. Also let $\mu_k = \int_{-\infty}^{\infty} x^k dF(x)$ so that $\mu_1 = 0$ and $\mu_2 = \sigma^2$. Theorem 2 (Feller, 1971) tells when the asymptotic density of $S_n/\sqrt{n}$ exists and what that asymptotic density is. Theorem 3 (Feller, 1971) gives the expansion of this density.

Theorem 2 (Thm XV.5.2, Feller, 1971) *If $|\phi|$ is integrable, then $S_n/\sqrt{n}$ has a density f_n which tends uniformly to the normal density $\mathcal{N}'$ where $\mathcal{N}' = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.*

Theorem 3 (Thm XVI.2.1, Feller, 1971) *Suppose μ_3 exists and that $|\phi|^\nu$ is integrable for some $\nu \geq 1$. Then f_n exists for $n \geq \nu$ and as $n \to \infty$*

$$f_n(x) - \mathcal{N}'(x) - \frac{\mu_3}{6\sigma^3\sqrt{n}}(x^3 - 3x)\mathcal{N}'(x) = o(1/\sqrt{n})$$

uniformly in x.

Note this (first order) expansion is called (or was called) the Edgeworth expansion for f_n. For further results see Feller (1971), Chapter XVI, Section 2.

In the development of the higher order expansion for the joint density, it will also be necessary to bound the density of the sum multiplied by a polynomial. Under the same conditions of Theorem 3, Petrov (1975) gives the needed result. Here we present the "$k = 3$" version of Petrov (1975).

Theorem 4 (Thm 17, Petrov, 1975) *Let $\{X_n\}$ be a sequence of independent random variables having a common distribution with zero mean, non-zero variance, and $E|X|^3 < \infty$. Let the random variable $\frac{1}{\sqrt{n\sigma^2}}S_n$ have for some $n > N$ a bounded density $f_n(x)$. Then*

$$(1 + |x|^3)\{f_n(x) - \mathcal{N}'(x) - \frac{\mu_3}{6\sigma^3\sqrt{n}}(x^3 - 3x)\mathcal{N}'(x)\} = o(1/\sqrt{n}) \tag{1.3}$$

uniformly in x.

This theorem will be important in establishing the uniformity of the results.

The results of Chapters 2, 3, and 4 will utilize this expansion for the density of the sum with the non-identical case of the Lindeberg-Feller theorem to establish an Edgeworth expansion for the conditional density of the sum given the maximum.

1.2.2 Limit Laws for Maxima

Although the limit laws for the maximum – fundamental in extreme value theory – do not have the same prominence in traditional statistical literature as those for sums, extreme value theory is an important area of research and application. The first formal look at the asymptotic distributions of the extremes began with the works of Dodd (1923), Fréchet (1927), and Fisher and Tippett (1928) with Fisher and Tippett establishing the "three types" of extreme value distributions. The first rigorous derivation of the extreme value distributions comes from Gnedenko (1943) but it was de Haan (1970) who completed the domain of attraction problem. At that time, he solved for the explicit representation of the auxillary function connected with the Gumbel domain of attraction. Some of the earliest work in applying extreme value analysis is found in Weibull (1939, 1951) and Gumbel (1958). Prescott and Walden (1980, 1983) established a rigorous theory for the maximum likelihood estimation – a modern bridge between the theoretical results and the numerous applications – for this family. This work is the cornerstone of much of the proliferation of extreme value application in the last 20 years. Like the review for the limit laws for sums, a review of extreme value results in the 20th century would be extensive. One of the best references in the field of extreme value theory is Leadbetter, Lindgren, and Rootzén (1983). Here we focus on the results applicable to the discussions in this work. In particular, we focus on the results for the expansions and densities since we want to establish the expansion of the density of the maximum in subsequent chapters.

Overview of the "three types"

The primary focus of classical extreme value theory involves the distribution of the maximum – in particular, the limiting distribution of the maximum (as $n \to \infty$). We begin with an *iid* sequence of random variables $X_1, \dots, X_n$ with common distribution function F. Define $M_n = max(X_1, \dots, X_n)$. From this we obtain

$$P(M_n \le x) = F^n(x).$$

We see without proper normalizing constants that M_n converges in probability to the upper endpoint of the distribution $x_o = sup\{x : F(x) < 1\}$. In other words,

$$\forall x < x_o \quad P(M_n \le x) \to 0 \;\text{ as }\; n \to \infty.$$

Thus establishing a (asymptotic) non-degenerate distribution function for the maximum, say H, is essentially finding normalizing constants $a_n > 0$ and b_n real such that

$$P(\frac{M_n - b_n}{a_n} \le x) = F^n(a_n x + b_n) \to H(x) \;\text{ as }\; n \to \infty.$$

When this limit holds, the non-degenerate distribution function H is said to be max-stable and F is said to be in the domain of attraction of H – i.e. $F \in \mathcal{D}(H)$.

Now the fundamental theorem of extreme value theory – from Theorem 1.4.2 of Leadbetter *et al.* (1983) – is as follows.

Theorem 5 (Thm 1.4.2, Leadbetter *et al.*, 1983) *Let* $M_n = max(X_1, \dots, X_n)$ *where* X_i *are iid random variables. If for some constants* $a_n > 0, b_n$ *we have*

$$P(\frac{M_n - b_n}{a_n} \le x) \stackrel{w}{\to} H(x) \tag{1.4}$$

for some non-degenerate d.f. H, then H is one of the three extreme value types listed below:

$$\begin{array}{llll} I & \textit{Gumbel type} & (\Lambda): & H(x) = \exp(-e^{-x}), \quad -\infty < x < \infty \\ II & \textit{Fréchet type} & (\Phi_\alpha): & H(x) = \begin{cases} 0 & x \le 0 \\ \exp(-x^{-\alpha}) \ \textit{for some} \ \alpha > 0 & x > 0 \end{cases} \\ III & \textit{Weibull type} & (\Psi_\alpha): & H(x) = \begin{cases} \exp(-(-x)^{\alpha}) \ \textit{for some} \ \alpha > 0 & x \le 0. \\ 0 & x > 0 \end{cases} \end{array}$$

Conversely each d.f. H of extreme value type may appear as a limit in (1.4) and, in fact, appears when H is the distribution function of each X_i.

For statistical purposes, it is better to have a single family. This unified form was first discovered by von Mises (1936) and can be written as

$$H(x;\eta,\psi,k) = \exp\{-[1-\frac{k(x-\eta)}{\psi}]^{1/k}\} \quad \text{where} \quad 1-\frac{k(x-\eta)}{\psi} > 0, \psi > 0 \qquad (1.5)$$

where the case $k = 0$ is interpreted as the limit $k \to 0$; that is,

$$H(x) = \exp(-e^{-\frac{x-\eta}{\psi}}) \quad -\infty < x < \infty.$$

This latter case corresponds to the Gumbel extremal type(I). The Fréchet and Weibull type (II and III, respectively) distributions are associated with $k < 0$ and $k > 0$, respectively. The unified form of the extreme value distribution (1.5) is commonly referred to as the Generalized Extreme Value (GEV) distribution.

Although we have only considered the *iid* case – in fact, restrict our attention to this case in this thesis – much of the prominent research in extreme value theory involves dependent sequences.

Dependent Sequences

Often in application the *iid* assumption is not realistic. Although there are numerous ways to extend the *iid* case, much of the literature focuses on stationary sequences. In fact, the focus lies on establishing conditions for which the limiting distribution of the maximum of such sequence still exists. More specifically, constructing conditions on the stationary sequence such that limiting distribution of the maximum is that same type as that for the associated independent sequence is of particular interest and can be shown to be surprisingly general. Again, the literature pertaining to extreme value theory for stationary sequences is extensive and we make no attempt to present a thorough overview of the topic. One of the best references, containing a comprehensive study of the topic, is Leadbetter *et al.* (1983). Here we only present associated with some topics, primarily from Leadbetter *et al.* (1983), used in some of the literature for the joint distribution of the sum and the maximum.

In general, there are two types of conditions that limit the dependency of the stationary sequences. The first type limits the amount of long range dependence. The second limits local dependence or what is known as the amount of clustering associated with exceedances over a threshold. The first set is general referred to as mixing conditions and we begin with strong mixing, introduced by Rosenblatt (1956).

Definition 6 (Strong Mixing, from Leadbetter *et al.*, 1983) *A sequence $\{X_i\}$ is strong mixing if there exists a function $g(k)$ tending to zero as k tends to infinity and such that*

$$|P(A \cap B) - P(A)P(B)| < g(k)$$

when $A \in \mathcal{F}(X_1, \ldots, X_p)$ and $B \in \mathcal{F}(X_{p+k+1}, X_{p+k+2}, \ldots)$ for any p and k where $\mathcal{F}()$ denotes the σ-field generated by the indicated random variables.

This essentially tells us that as we separate the past and the future – i.e. let $k \to \infty$ – the two behave independently. Now since the events of interest in extreme value theory are primarily of the form $\{X \leq u_n\}$ or $\{X > u_n\}$, we can in fact weaken the mixing condition in definition 6 and ultimately still obtain the results pertaining to the limiting distribution of M_n. The condition D is of this form. Define $F_{k_1,\ldots,k_n}(u) = P\{X_{k_1} \leq u, \ldots, X_{k_n} \leq u\}$. We have

Definition 7 (Condition D, Leadbetter *et al.*, 1983) *The condition D will be said to hold if for any integers*

$$i_1 < \ldots < i_p \quad and \quad j_1 < \ldots < j_{p'}$$

for which $j_1 - i_p \geq l$, and any real u,

$$|F_{i_1,\ldots,i_p,j_1,\ldots,j_{p'}}(u) - F_{i_1,\ldots,i_p}(u)F_{j_1,\ldots,j_{p'}}(u)| \leq g(l),$$

where $g(l) \to 0$ as $l \to \infty$.

Now we can weaken this mixing condition even further by requiring the condition D hold for only certain sequences of values $\{u_n\}$. This condition is defined as $D(u_n)$.

Definition 8 (Condition $D(u_n)$, Leadbetter *et al.*, 1983) *The condition $D(u_n)$ will be said to hold if for any integers*

$$1 \leq i_1 < \ldots < i_p < j_1 < \ldots < j_{p'} \leq n$$

for which $j_1 - i_p \geq l$, we have

$$|F_{i_1,\ldots,i_p,j_1,\ldots,j_{p'}}(u_n) - F_{i_1,\ldots,i_p}(u_n)F_{j_1,\ldots,j_{p'}}(u_n)| \leq \alpha_{n,l},$$

where $\alpha_{n,l} \to 0$ as $n \to \infty$ for some sequence $l_n = o(n)$.

Again this mixing condition restricts the amount and type of long range dependence. Note strong mixing implies condition D which implies condition $D(u_n)$.

Now it has been shown by Leadbetter (1983) that condition $D(u_n)$ is sufficient to guarantee the result concerning the three types of possible extreme value distributions for the limiting distribution M_n.

Theorem 9 (Thm 1.1, Leadbetter, 1983) *Let* $\{X_n\}$ *be a stationary sequence such that* $M_n = \max\{X_1, \ldots, X_n\}$ *has a non-degenerate limiting distribution* H *as in 1.4 for some constants* $a_n > 0, b_n$*. Suppose that* $D(u_n)$ *holds for all sequences* u_n *given by* $u_n = a_n x + b_n, -\infty < x < \infty$*. Then* H *is one of the three classical types given above.*

Now condition $D(u_n)$ gives the possible types for the limiting distribution of the maximum of a stationary sequence but it does not define the existence of such a limit distribution. To do that, conditions that define the clustering of the exceedances are necessary. We begin with condition $D'(u_n)$ of Leadbetter *et al.* (1983).

Definition 10 (Condition $D'(u_n)$**, Leadbetter *et al.*, 1983)** *The condition* $D'(u_n)$ *will be said to hold for the stationary sequence* $\{X_n\}$ *and sequence* $\{u_n\}$ *of constants if*

$$\limsup_{n\to\infty} n \sum_{j=2}^{[n/k]} P\{X_1 > u_n, X_j > u_n\} \to 0, \quad \text{as} \quad k \to \infty$$

(where [] denotes the integer part.)

Together conditions $D(u_n)$ and $D'(u_n)$ provide the classical result for stationary sequence. Here we define $\tilde{M}_n$ as $\max\{\tilde{X}_1, \ldots, \tilde{X}_n\}$ where $\tilde{X}_1, \ldots, \tilde{X}_n$ is the independent sequence associated with $X_1, \ldots, X_n$.

Theorem 11 (Thm 3.5.2, Leadbetter *et al.*, 1983) *Suppose that* $D(u_n)$ *and* $D'(u_n)$ *are satisfied for the stationary sequence* $\{X_n\}$ *when* $u_n = a_n x + b_n$ *for each* x *(*$\{a_n > 0\}, \{b_n\}$ *being given sequences of constants). Then* $P\{\frac{M_n - b_n}{a_n} \leq x\} \to H(x)$ *for some non-degenerate* H *if and only if* $P\{\frac{\tilde{M}_n - b_n}{a_n} \leq x\} \to H(x)$*.*

In fact, Leadbetter (1983) shows that in most cases of practical interest, the condition $D(u_n)$ is sufficient to guarantee that the limiting distribution of M_n is of precisely the same type as $\tilde{M}_n$. Note when dropping the condition $D'(u_n)$, we need to be able to describe the amount of clustering associated with exceedances of a sequence over a threshold $\{u_n\}$. We do so by defining the extremal index of the sequence. Note we define $\{u_n(\tau)\}$ as a sequence of thresholds which satisfies $n\{1-F(u_n(\tau))\} \to \tau$ as $n \to \infty$ and $D(u_n(\tau))$ as the extenstion of the condition $D(u_n)$ with u_n replaced by $u_n(\tau)$ for each $\tau > 0$.

Definition 12 (Extremal index, Leadbetter *et al.*, 1983) *A process $\{X_i\}$ has extremal index θ, where $0 \leq \theta \leq 1$, if for each $\tau > 0$*

(i.) There exists a sequence $u_n(\tau)$ such that $n[1-F(u_n(\tau))] \to \tau$ as $n \to \infty$.

$$(ii.) \quad P(M_n \leq u_n(\tau)) = F^{n\theta}(u_n(\tau)) + o(1) \to e^{-\theta\tau} \quad as \quad n \to \infty.$$

Note that $\theta = 1$ indicates a process which behaves like an independent process.

Note if (i) holds and condition $D(u_n(\tau))$ holds for each τ then (ii) holds and the sequence $\{X_n\}$ has extremal index θ.

Now given the definition of the extremal index θ of a sequence $\{X_n\}$ we can redefine the conditions on a stationary sequence so that the classical asymptotic distributions still hold.

Theorem 13 (Thm 2.5, Leadbetter, 1983) *Let the stationary sequence $\{X_n\}$ have extremal index $\theta > 0$. Then M_n has a non-degenerate limiting distribution if and only if $\tilde{M}_n$ does, and these are then of the same type based on the same normalizing constants. In the case $\theta = 1$ the limiting distributions for M_n and $\tilde{M}_n$ are identical.*

Note the extremal index may be interpreted as the inverse of the mean cluster size. The extremal index thus plays a key role in the study of extremes of stationary sequences including the work of Anderson and Turkman(1991) which is reviewed in Section 1.2.3. The rest of the work in this thesis involves the *iid* case.

Expansions

Regularly Varying Functions Before we can look at the literature pertaining to the expansions for the distribution for the maximum, we need to briefly go over the concept of regular variation and the Karamata representation for regularly varying functions. Given these concepts we can refashion the "three types" theorem and see the connection to the expansions.

Definition 14 (Regularly Varying, Resnick, 1987) *A measurable function* $U\colon \mathbb{R}_+ \to \mathbb{R}_+$ *is regularly varying at* ∞ *with index* ρ *[written* $U \in \mathcal{R}_\rho$*] if* $\forall x > 0$

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\rho.$$

Note this limit is locally uniform on $(0, \infty)$, see Resnick (1987), Proposition 0.5, p. 17. The parameter ρ is called the exponent of variation. If $\rho = 0$ the function U is said to be slowly varying. The general notation for slowly varying functions is $\mathcal{L}(x)$. Using this notation we see that we can always rewrite a regularly varying function U in terms of a slowly varying function $\mathcal{L}$; i.e., $U(x) = x^\rho \mathcal{L}(x)$.

Next we give Karamata's Theorem [Theorem 0.6 in Resnick (1987), p. 17] which details the integral properties of regularly varying functions and provides the connection between regularly varying functions and the extreme "three types" of limiting distributions. Finally from Karamata's Theorem we get the Karamata Representation of slowly varying functions. This is another important piece in the development of the expansion of the distribution function for the maximum, particularly, in the Gumbel case.

Theorem 15 (Karamata's Theorem 0.6 of Resnick, 1987) *(a) If* $\rho \geq -1$ *then* $U \in \mathcal{R}_\rho$ *implies* $\int_0^x U(t)dt \in \mathcal{R}_{\rho+1}$ *and*

$$\lim_{x\to\infty} \frac{xU(x)}{\int_0^x U(t)dt} = \rho + 1$$

If $\rho < -1$ (or if $\rho = -1$ and $\int_x^\infty U(s)ds < \infty$) then $U \in \mathcal{R}_\rho$ implies $\int_x^\infty U(t)dt$ is finite $\int_x^\infty U(t)dt \in \mathcal{R}_{\rho+1}$ and

$$\lim_{x\to\infty} \frac{xU(x)}{\int_x^\infty U(t)dt} = -\rho - 1$$

(b) If U satisfies

$$\lim_{x\to\infty} \frac{xU(x)}{\int_0^x U(t)dt} = \lambda \in (0,\infty)$$

then $U \in \mathcal{R}_{\lambda-1}$. If $\int_x^\infty U(t)dt < \infty$ and

$$\lim_{x\to\infty} \frac{xU(x)}{\int_x^\infty U(t)dt} = \lambda \in (0,\infty)$$

then $U \in \mathcal{R}_{-\lambda-1}$.

Corollary 16 (The Karamata Representation, Resnick, 1987) *$\mathcal{L}$ is slowly varying iff $\mathcal{L}$ can be represented as*

$$\mathcal{L}(x) = c(x)\exp\left\{\int_1^x t^{-1}\epsilon(t)dt\right\}$$

for $x > 0$ where $c : \mathbb{R}_+ \to \mathbb{R}_+, \epsilon : \mathbb{R}_+ \to \mathbb{R}_+$ and

$$\begin{aligned} \lim_{x\to\infty} c(x) &= c \in (0,\infty) \\ \lim_{t\to\infty} \epsilon(t) &= 0. \end{aligned}$$

From the theory of regular variation we can rework the conditions for the three possible types of asymptotic distributions for the maximum. Here we present Theorem 1.6.2 of Leadbetter *et al.* (1983) slightly reworded in the notation of this thesis.

Theorem 17 (Thm 1.6.2, Leadbetter *et al.*, 1983) *Necessary and sufficient conditions for the distribution function F of the random variables of an iid sequence $\{X_n\}$ to belong to each of the three types are:*
(a.) Type II (Φ): Let $U(t) = 1 - F(t), x_o = \infty$ where $x_o = \sup\{x : F(x) < 1\}$ and $U(t)$ is regularly varying with index α, $\alpha > 0$. [Theorem 2.3.1 in De Haan (1970)]

(b.) Type III (Ψ)*: Let* $U(t) = 1 - F(x_o - \frac{1}{t})$ *where* $x_o < \infty$ *and* $U(t)$ *is regularly varying with index* α, $\alpha > 0$. *[Theorem 2.3.2 in De Haan (1970)]*

(c.) Type I (Λ)*: There exists some strictly positive (auxiliary) function* ϕ *such that*

$$\lim_{t \to x_o} \frac{1 - F(t + x\phi(t))}{1 - F(t)} = e^{-x} \quad \forall x \in \mathcal{R}$$

(Theorem 2.5.1 in De Haan(1970)].

Note condition (c) was originally due to Gnedenko (1943) but de Haan (1970) provided an explicit characterization of ϕ.

Now we present the important results concerning the rates of convergence and, in more detail, the expansion for the distribution of the maximum for the three domains of attractions. We break the results up into the Fréchet and Weibull type – the Weibull result being just a transformation of the Fréchet type – and, finally, the Gumbel results.

Fréchet Case First we extend our notions concerning slowly varying functions. In particular, to calculate rates of convergence for the distribution of the maximum to the limiting distributions, we need to establish "remainders" to the definition of slowly varying functions. We distinguish between two forms of slow variation of remainder. The first gives enough detail to establish rates of convergence for the asymptotic distribution of M_n. The second allows us to calculate higher order terms for the expansions of the distribution of M_n.

Definition 18 (Slow variation with remainder, Smith, 1982) *Let* $\mathcal{L}$ *and* g *be two functions defined on* $(0, \infty)$. *Assume* $g(t) \to 0$ *as* $t \to \infty$ *and* $\mathcal{L}$ *measurable. Define the two forms of slow variation with remainder as*

$$\begin{aligned} \textit{SR1:} \quad \frac{\mathcal{L}(tx)}{\mathcal{L}(t)} - 1 &= O(g(t)), \quad x > 0, t \to \infty \\ \textit{SR2:} \quad \frac{\mathcal{L}(tx)}{\mathcal{L}(t)} - 1 &\sim g(t)\nu(x), \quad x > 0, t \to \infty \end{aligned}$$

If SR2 holds and ν *satisfies the condition: There exists an* x *such that* $\nu(x) \neq 0$ *and* $\forall y$ $\nu(xy) - \nu(y) \neq 0$, *then* g *is regularly varying with index* ρ, *i.e.*

$$\lim_{t \to \infty} \frac{g(tx)}{g(t)} = x^{\rho} \quad x > 0 \ \textit{ for some } \ \rho \leq 0.$$

and $\nu(x) = ch_\rho(x)$ *for some constant* c *and* $h_\rho(x) = \int_1^u x^{\rho-1}dx$.

Theorem 1 in Smith (1982) gives the rate of convergence for the distribution of M_n under the Fréchet domain of attraction.

Theorem 19 (Thm 1, Smith, 1982) *Suppose that* $F(x) < 1, \forall x < \infty$, *that* $-\log F(x)$ *is regularly varying with index* α *for some* $\alpha > 0$ *and that* $\mathcal{L}(t) = -t^{-\alpha}\log F(t)$, $t > 0$ *satisfies SR1 for some positive function* g *satisfying* $\frac{g(tx)}{g(t)} < C$ *for* $x > 1$, $t \geq t_o$ *and*

$$\forall t \geq t_o, \; x \leq 1, \; \frac{g(tx)}{g(t)} \geq Bx^{-\theta} \quad (\theta > 0, B > 0).$$

Let the normalizing constants for maximum defined in equation (1.4) be $b_n = 0$ *and* a_n *such that* $-\log F(a_n) \leq n^{-1} \leq -\log F(a_n^-)$ *where* $F(x^-) = sup_{y<x}F(y)$. *Then as* $n \to \infty$

$$\sup_x |F^n(a_n x) - \Phi_\alpha(x)| = O(g(a_n)).$$

Thus we see that the rate of convergence to the limiting distribution when $F \in \mathcal{D}(\Phi_\alpha)$ is $g(a_n)$. To see the first order term in this expansion the SR2 condition is necessary. Theorem 2 in Smith (1982) gives this expansion.

Theorem 20 (Thm 2, Smith, 1982) *Let* $F(\cdot), \mathcal{L}(\cdot), a_n$ *and* b_n *be as in Theorem 19. Suppose* $\mathcal{L}$ *satisfies SR2 where* $g \in RV_\rho$, $\nu(x) = ch_\rho(x)$ *for some* $\rho \leq 0$, *then*

$$F^n(a_n x) - \Phi_\alpha(x) = -cg(a_n)h_\rho(x)x^{-\alpha}\Phi_\alpha(x) + o(g(a_n))$$

uniformly in $0 < x < \infty$.

Weibull results The results for the Fréchet domain of attraction immediately provide the results for the Weibull domain of attraction, through the simple transformation that exists between the two distributions.

Given the distribution function F, a necessary and sufficient for the existence of $a_n > 0$ and b_n real such that $\lim_{n\to\infty} F^n(a_n x + b_n) = \Psi_\alpha(x)$, $\alpha > 0$ is that F has a finite endpoint – i.e., $x_o = \sup\{x : F(x) < 1\} < \infty$ – and that $F_1(x) = F(x_o - \frac{1}{x})$ be in the domain of attraction of Φ_α. We see this connection by looking at Theorem 17, part (a) and (b). Using Theorem 19, Smith (1982) obtains the rate of convergence when $F \in \mathcal{D}(\Psi_{\alpha)}$, see Theorem 5 of Smith (1982).

Theorem 21 (Thm 5, Smith, 1982) *Suppose F_1 satisfies the hypothesis of Theorem 19. Define $b_n = x_o$ and a_n such that $-\log F(x_o - a_n) \leq n^{-1} \leq -\log F(x_o - a_n^-)$, then*

$$\sup_x |F^n(a_n x + b_n) - \Psi_\alpha(x)| = O(g(\frac{1}{a_n})).$$

Using Theorem 20 Smith (1982) obtains the first order term in the expansion, see Theorem 6 in Smith (1982).

Theorem 22 (Thm 6, Smith, 1982) *Suppose F_1 satisfies the hypotheses in Theorem 20. Then uniformly in $-\infty \leq x < 0$*

$$F^n(a_n x + b_n) - \Psi_\alpha(x) = cg(\frac{1}{a_n})h_{-\rho}(-x)^\alpha \Psi_\alpha(x) + o(g(\frac{1}{a_n}))$$

.

Note we use the relationship $h_\rho(1/y) = -h_{-\rho}(y)$.

Gumbel case Unlike the Fréchet and Weibull case, there does not exist a unified result for the uniform rates of convergence and/or expansion of the distribution of M_n for the underlying distributions which lie in the Gumbel domain of attraction. Although work first began with specific distributions within the Gumbel domain – exponential (Hall and Wellner, 1979), normal (Hall, 1979), and powers of the normal distribution (Hall, 1980) – results are typically produced for classes of distributions within the Gumbel domain of attraction. Anderson (1971) produced results for two classes of distributions in $\mathcal{D}(\Lambda)$ which Cohen (1982b) denotes as $A1$ and $A2$. Cohen (1982b) also defined two classes of distributions which are referred to as N and E. Cohen (1982b) produced uniform rates of convergence and expansions for the distribution of M_n for these latter two classes. Here we only present his results for class N since these are the results that will be used in the derivation of the joint density of the sum and the maximum.

First we present Cohen's (1982b) characteristic theorem for $-\log F$ when $F \in \mathcal{D}(\Lambda)$. Note although there are many ways to quantify the conditions necessary and/or sufficient so that $F \in \mathcal{D}(\Lambda)$, essentially we need $-\log F(x)$ or its asymptotic equivalent $1 - F(x)$ to be slowly varying. [In particular, if x_n is such that $1 - F(x_n) = O(\frac{1}{n})$, then $1 - F(x_n) + \log F(x_n) = O(\frac{1}{n^2})$. The relative difference being of smaller order than the other terms we ultimately consider.] Hence if $-\log F$ is slowly varying, we may rewrite $-\log F$ using a Karamata representation. Based on this representation we define of Cohen's (1982b) class N. Finally we present his uniform expansions.

Theorem 23 (Thm 1, Cohen, 1982b) $F \in \mathcal{D}(\Lambda) \leftrightarrow \exists$ *functions* a, b, c *and* f *defined on* $[X, x_o)$, *for some constants* $X < x_o \leq \infty$, *such that* a/f *and* b *are Lebesque-integrable over finite sub-intervals of* $[X, x_o)$ *and, for some constants* c_1 *and* c_2,

$$-\log F(x) = c(x) \exp\left\{-\int_X^x \frac{a(t)}{f(t)} dt\right\}, \quad X \leq x < x_o \tag{1.6}$$

$$f(x) = \begin{cases} c_2 + \int_X^x b(t)dt & x = \infty \quad X \leq x < x_o \\ -\int_x^{x_o} b(t)dt & x < \infty \quad X \leq x < x_o \end{cases} \tag{1.7}$$

$$f(x) > 0, \quad X \le x < x_o; \tag{1.8}$$

$$c(x) \to c_1 \quad as \quad x \to x_o; \tag{1.9}$$

$$a(x) \to 1 \quad as \quad x \to x_o; \tag{1.10}$$

$$b(x) \to 0 \quad as \quad x \to x_o. \tag{1.11}$$

Further, if $F \in \mathcal{D}(\Lambda)$, we may take

$$b(x) = -1 + \frac{\{-\log F(x)\}\{\int_x^{x_o}\int_y^{x_o} -\log F(s)dsdy\}}{\{\int_x^{x_o} -\log F(s)ds\}^2} \tag{1.12}$$

$$a(x) = 1 + 2b(x); \tag{1.13}$$

$$c(x) = c_1\{1 + b(x)\}; \tag{1.14}$$

for $X \le x < x_o$ and some constant c_1.

Now we can define Cohen's class N – a family of distributions within the Gumbel domain of attraction.

Definition 24 (Cohen's class N, Cohen, 1982b) *We shall say F is in class N if there exists a characterization a, b, c, f such that (1.6) - (1.11) holds and in addition there exists $K > 1$ and r, s such that*

$$Either\ \underline{(a)}\ \ b(x) > 0 \quad or\ \underline{(b)}\ \ b(x) < 0, \quad X \le x < x_o; \tag{1.15}$$

$$\{a(x) - 1\}/b(x) \to r \quad as \quad x \to x_o; \tag{1.16}$$

$$\{c(x) - c_1\}/b(x) \to s \quad as \quad x \to x_o; \tag{1.17}$$

$$f'(x) \quad exists\ and \quad f'(x) = b(x) \quad for \quad X \le x < x_o; \tag{1.18}$$

$$xf(x) + y \to x_o \quad uniformly\ in \quad |x| \le Klog|(b(y)|, \quad as \quad y \to x_o; \tag{1.19}$$

$$\frac{b(xf(x) + y)}{b(y)} \to 1 \quad uniformly\ in \quad |x| \le Klog|(b(y)|, \quad as \quad y \to x_o; \tag{1.20}$$

Note the distributions which belong to class N include the normal, log-normal, gamma (shape parameter $\neq 1$) and Weibull (shape parameter $\neq 1$). Distributions which do not belong to class N include the exponential.

Embedded in Theorem 9 of Cohen(1982b) is an alternative definition of class N.

Definition 25 (from Thm 9, Cohen, 1982b) *Let F satisfy (1.6) - (1.11) and (1.19) with $K > 2$. Assume $b'(x)$ exists and has constant sign for $X \leq x < x_o$ (hence (1.15) holds) and*

$$\{a(x) - 1\}/b(x) \to r \quad as \quad X \leq x < x_o; \tag{1.21}$$

$$\{c(x) - c_1\}/b(x) \to s \quad as \quad X \leq x < x_o; \tag{1.22}$$

$$\frac{b'(xf(x) + y)}{b'(x)} \to 1 \quad uniformly\ in \quad |x| \leq Klog|(b(y)|, \quad as \quad y \to x_o; \tag{1.23}$$

$$\frac{f(x)}{b(x)} b'(x) \log|b(x)| \to 0 \quad as \quad x \to x_o \tag{1.24}$$

Then F is in class N.

Finally, we present the uniform expansion of the distribution of M_n presented by Cohen (1982b).

Theorem 26 (Thm 2, Cohen, 1982b) *Let F be in class N with the appropriate functions a, b, c, and f satisfying (1.6) – (1.11) and (1.15) – (1.20). Define the normalizing constants a_n and b_n by*

$$-\log F(b_n) \leq n^{-1} \leq -\log F(b_n -);$$

$$a_n = f(b_n).$$

Then, uniformly on $-\infty < x < \infty$,

$$F^n(a_n x + b_n) = \Lambda(x) + O(b(b_n));$$

$$F^n(a_n x + b_n) = \Lambda(x) - (\frac{1}{2}x^2 - rx)\Lambda'(x)b(b_n) + o(b(b_n)).$$

Threshold case In the derivation of the joint density of the sum and the maximum, we need expansions associated with the limiting distribution of exceedances above a threshold. In particular, we need expansions for the first two moments of this distribution. In the section on limit laws of exceedances, we see the relationship between the GEV distribution and the Generalized Pareto distribution via threshold methods. Here we present the results of Smith (1987) in which he establishes both the rate of convergence for F_u – the tail of the conditional density of $X-u$ given $X>u$ where u is an arbitrary threshold – and then the first order term in this expansion. Here we only present the results associated with the Gumbel domain of attraction. First we present a refinement of the Karamata's representation by Balkema and de Haan (1972).

Definition 27 (Ext. to Karamata rep., Balkema and de Haan, 1972) *If F is in the domain of attraction of Λ, then there exists a representation*

$$1-F(x)=c(x)\exp\left\{-\int_{-\infty}^{x}\frac{x}{\phi(t)}\right\},\quad x<x_o, \tag{1.25}$$

where $c(x)\to 1$ as $x\to x_o\leq\infty$, ϕ is a positive differentiable function and $\phi'(x)\to 0$ as $x\to x_o\leq\infty$.

Proposition 28 (Prop. 9.1, Smith, 1987) *Suppose (1.25) holds and*

$$\phi'(u+y\phi(u))/\phi'(u)\to 1,\quad \textit{as}\ \ u\to x_o,\quad \textit{uniformly over}$$

$$0\leq y\leq -K\log|\phi'(u)|,\quad \textit{for some}\ \ K>1,$$

$$c(u)-1\sim s\phi'(u),\quad \textit{as}\ \ u\to x_o\ \ \textit{for finite}\ \ s.$$

Then, for each $\delta>0$, there exist $u_\delta\leq x_o$ and a function ϵ_u tending to 0 as $u\to x_o$, such that for $u\geq u_\delta, 0\leq y<x_o-u$,

$$|1-F_u(y\phi(u))-e^{-y}\{1+\phi'(u)y^2/2\}|<\epsilon_u\phi'(u)\min(1,y^{-\delta}).$$

Proposition 29 (Prop. 9.2, Smith, 1987) *Suppose (1.25) holds and*

$$\phi''(u+y\phi(u))/\phi''(u) \to 1, \quad as \quad u \to x_o, \quad uniformly\ over$$
$$0 \le y \le -K \log|\phi'(u)|, \quad for\ some \quad K > 2,$$
$$c(u) - 1 \sim s\{(\phi'(u))^2 + |\phi(u)\phi''(u)|\}, \quad as \quad u \to x_o \quad for\ finite \quad s.$$
$$\phi(u)\phi''(u)\log|\phi'(u)|/\phi'(u) \to 0, \quad as \quad u \to x_o.$$

Then, for each $\delta > 0$, there exist $u_\delta < 0$ and a function ϵ_u tending to 0 as $u \to x_o$, such that for $u \ge u_\delta, 0 \le y < x_o - u$,

$$|1 - F_u(y\phi(u)) - e^{-y}[1 + \phi'(u)y^2/2 - y^3\{2\phi'(u))^2 - \phi(u)\phi''(u)\}/6 + y^4(\phi^4(\phi')(u))^2/8]|$$
$$< \epsilon_u[(\phi'(u))^2 + |\phi(u)\phi''(u)|]\min(1, y^{-\delta}).$$

Note from Proposition 9.2 of Smith (1987), he calculates the first two moments of $Y = X - u$,

$$\begin{aligned} E(\frac{Y}{\phi(u)}) &= 1 + \phi'(u) + ((\phi'(u))^2 + \phi(u)\phi''(u)) + o(\phi'(u)^2 + |\phi(u)\phi''(u)|) \\ Var(\frac{Y}{\phi^2(u)}) &= 2 + 6\phi'(u) + 14(\phi'(u))^2 + 8\phi(u)\phi''(u) + o(\phi'(u)^2 + |\phi(u)\phi''(u)|). \end{aligned}$$

Limit Laws involving densities: The domain of local attraction

The result concerning the convergence of the density of M_n which we wish to emphasize here is that the so-called von Mises conditions are equivalent to the necessary and sufficient conditions needed in establishing the local uniform convergence of the density of the maximum. Although these conditions were first studied by von Mises (1936) with respect to the general domain of attraction problem, the local domain of attraction problem was first looked at by Pickands (1968) and Anderson (1971). Results most closely connected to this thesis – involving the uniform results and rates of convergence – are de Haan and Resnick (1982) and Sweeting (1985). To proceed we first outline the required notation for this section. Next we present the adaptive von

Mises conditions. Note the original von Mises conditions for the Gumbel case assumed a second derivative to F. Finally, we tie these von Mises conditions to the conditions for the density convergence. In fact, we reintroduce a function ϕ, which is found in the Karamata representation of a slowly varying function. Letting this function take the form $\phi(t) = \frac{1-F(t)}{f(t)}$ where $f(t) = dF(t)/dt$, we show that the behaviour of $\phi(t)$ governs the von Mises conditions and hence the local limit laws for M_n. In the next section we exploit this function to establish the Generalized Pareto distribution. In the last section of the theoretical review, we see that how we deal with the limit of the derivation of this function can lead to either the ultimate or penultimate approximation to the asymptotic distribution of M_n.

Traditionally the von Mises conditions are introduced primarily as sufficient conditions for the general domain of attraction problem. In many statistical applications these conditions are more easily verified and lead to direct calculation of the normalizing constants a_n and b_n. For the specific conditions in the following theorem, parts (a) and (b) which deal with the Fréchet and Weibull cases are due primarily to von Mises (1936) although the proofs usually accompanying these results [see Resnick(1987), p. 85 or 63] are due to de Haan and Resnick(1982). Part (c) which deals with the Gumbel case comes from de Haan(1970).

Suppose $P(\frac{M_n-b_n}{a_n} \le x) = F^n(a_n x + b_n) \to H(x)$ for some extreme value distribution H. We suppose F is absolutely continuous with density $f(x)$. Define f_n as the density of $\frac{M_n-b_n}{a_n}$; that is,

$$f_n(x) = n a_n F^{n-1}(a_n x + b_n) f(a_n x + b_n). \tag{1.26}$$

The local domain of attraction problem centers on conditions such that

$$f_n(x) \to H'(x) \tag{1.27}$$

where H' is the derivative of H.

Theorem 30 (Von Mises c., Prop. 1.15(a), 1.16(a), 1.17(a), Resnick, 1987) *(a) Fréchet case: Suppose F is absolutely continuous with positive density f in some neighborhood of ∞. If for some $\alpha > 0$,*

$$\lim_{x\to\infty} \frac{x}{\phi(x)} = \lim_{x\to\infty} \frac{xf(x)}{1-F(x)} = \alpha \tag{1.28}$$

then $F \in \mathcal{D}(\Phi_\alpha)$. We may choose a_n to satisfy $a_n f(a_n) \sim \alpha/n$.

(b) Weibull case: Suppose F has a finite right endpoint x_o and is absolutely continuous in a left neighborhood of x_o with positive density f. If for some $\alpha > 0$,

$$\lim_{x\to x_o} \frac{x_F - x}{\phi(x)} = \lim_{x\to x_o} \frac{(x_F - x)f(x)}{1-F(x)} = \alpha \tag{1.29}$$

then $F \in \mathcal{D}(\Psi_\alpha)$.

(c) Gumbel case: Let F be absolutely continuous in a left neighborhood of x_o with density f. If

$$\lim_{x\to x_o} \frac{1}{\phi(x)} \frac{\int_x^{x_o}(1-F(t))dt}{1-F(x)} = \lim_{x\to x_o} \frac{f(x)\int_x^{x_o}(1-F(t))dt}{[1-F(x)]^2} = 1 \tag{1.30}$$

then $F \in \mathcal{D}(\Lambda)$. In this case, we may choose $a_n = \phi(b_n)$ and b_n such that $1-F(b_n) = \frac{1}{n}$.

Note it can be shown that the key in solving each of these cases involves assuming $\lim_{x\to\infty} \phi'(x) = k$ where $k \in \mathbb{R}$ is some constant. Also note if we take $\phi'(b_n) = -k_n$ where k_n is not a constant, this leads to the penultimate approximation. For more details on the penultimate approximation, see the last section in the theoretical review.

Now de Haan and Resnick (1982) showed that conditions (1.28), (1.29), and (1.30) are the necessary and sufficient conditions for $f_n(x) \to H'(x)$ uniformly over compact subsets. Specifically we have the following:

Proposition 31 (Prop 2.5, Resnick, 1987) *Suppose F is absolutely continuous with a density f and right endpoint x_o. If $F \in \mathcal{D}(H)$ and*

(a) $H = \Phi_\alpha$ then (1.27) is true locally uniformly on $(0,\infty)$ iff (1.28).

(b) $H = \Psi_\alpha$ then (1.27) is true locally uniformly on $(-\infty,0)$ iff (1.29).

(c) $H = \Lambda$ then (1.27) is true locally uniformly on $(-\infty,\infty)$ iff (1.30).

Sweeting (1985) gave alternative necessary and sufficient conditions which he showed were equivalent to (1.28), (1.29), and (1.30). He also established uniformity of convergence over the full range of H under the additional additional condition

$$f(x) \leq C\{F(x)\}^{-B} \quad \forall x \in (-\infty, \infty) \quad \text{for some} \quad B > 0, C > 0.$$

Sweeting (1985) also extended these results by establishing the kth times differentiable domain of attraction – i.e. when

$$\frac{d}{dx^{(k)}} F^n(a_n x + b_n) \to H^{(k)}(x) \quad \text{as} \quad n \to \infty.$$

This is an important result for this paper because it gives not only rates of convergence for densities but also higher order expansion terms.

Limit Laws of Exceedances: Generalized Pareto Distribution

There is a connection between the GEV distribution and the Generalized Pareto (GP) distribution. We let $Y = X - u$ where u is a threshold which we may take to equal $a_n v + b_n$ where a_n and b_n are defined as the normalizing constants of M_n. Recall $F_u(y) = P(Y \leq y | X > u) = \frac{F(y+u)}{1-F(u)}$ which is the conditional distribution of an exceedance over the threshold.

Pickands (1975) showed that the GP distribution is the limiting distribution for the exceedance over a threshold under precisely the same conditions (on F) as the ordinary GEV distribution is the limiting distribution for the maximum. A quick look at the reasoning behind this connection is as follows. Using our definitions of a_n, b_n, ϕ and the Karamata representation of a slowly varying function, we can write

$$\frac{1 - F(u + x\phi(u))}{1 - F(u)} = \exp\left\{-\int_0^x \frac{\phi(u)}{\phi(u + s\phi(u))} ds\right\} \approx \{1 + x\phi'(y)\}^{-1/\phi'(y)} \qquad (1.31)$$

for some $y \in (u, u + x\phi(u))$.

Using the last step in (1.31) – a Taylor's expansion for its integrand – and assuming $\phi'(y) \to k$ then

$$\frac{1 - F(u + x\phi(u))}{1 - F(u)} \approx \begin{cases} (1 + kx)^{-1/k} & k \neq 0 \\ e^{-x} & k = 0. \end{cases}$$

If we let $y = x\phi(u)$ and $\sigma = \phi(u)$ then

$$1 - F_u(y) = \frac{1 - F(y + u)}{1 - F(u)} \approx \begin{cases} (1 + \frac{ky}{\sigma})^{-1/k} & k \neq 0 \\ e^{-y/\sigma} & k = 0 \end{cases} \tag{1.32}$$

where $\sigma > 0, -\infty < k < \infty$ and for $k \leq 0$, $0 \leq y < \infty$ and for $k \geq 0, 0 \leq y \leq -\sigma/k$.

The righthand side of (1.32) is the tail function for the GP distribution. The GP distribution is important to the statistical methodology which utilizes the POT(peaks over threshold) point process approach in fitting data. For more details, see Davison and Smith (1990). For this thesis it is a key to the derivation of the expansion of the joint density of the sum and the maximum since to establish the higher order term we rewrite the mean of the underlying distribution in terms of the mean of an exceedance over a threshold.

Penultimate Approximation

Finally, we look briefly at the construction of the penultimate approximation to the extreme value distributions. Recall the function introduced in Theorem 30 – $\phi(x)$. In Theorem 30, we see that the $\lim_{x\to\infty} \phi(x)$ characterizes the *ultimate* approximation of the distribution of the maximum – i.e., which domain of attraction it belongs to. In fact, the ultimate approximation involves substituting $\lim_{x\to\infty} \phi(x) = k$ (say) into the appropriate formula in Theorem 30. If instead we substitute a sequence $\{k_n\}$ into the formula, we obtain a sequence of approximate distributions to the distribution of the maximum which is referred to as the *penultimate* approximation. We revisit (1.31) but replace $1 - F(x)$ in the Karamata's representation by its asymptotic equivalent $-\log F$, u by b_n, and $\phi(u)$ by a_n. In doing so, we may rewrite the righthand side as

$$-\log F^n(a_n x + b_n) \approx \{1 + x\phi'(y)\}^{-1/\phi'(y)}$$

for some $y \in (b_n, a_n x + b_n)$ which gives

$$F^n(a_n x + b_n) \approx \exp[-\{1 + x\phi'(y)\}^{-1/\phi'(y)}].$$

Now assuming $\phi'(y) \to k$, then the ultimate approximation is

$$F^n(a_n x + b_n) \approx \exp[-\{1 + kx\}^{-1/k}].$$

Note in the case where $k = 0$ we should interpret the above formula as

$$F^n(a_n x + b_n) \approx \exp[-\exp\{-x\}].$$

which is the Gumbel case.

Now if we replace $\phi'(y)$ not by k but by $\phi'(b_n)$ – i.e. let $k_n = -\phi'(b_n)$ – then

$$F^n(a_n x + b_n) \approx \exp[-\{1 + x\phi'(b_n)\}^{-1/\phi'(b_n)}] \approx \exp[-\{1 + k_n x\}^{-1/k_n}].$$

which is the penultimate approximation. For more details, see Smith(1990).

This penultimate approximation has received more attention in the case where $k = 0$. In this case, the ultimate approximation is of the Gumbel type while the penultimate approximation is either the Fréchet or Weibull type depending on the sign of $\phi'(b_n)$ for large n. The penultimate approximation was first suggested by Fisher and Tippett (1928). In their specific case for normal extremes, they showed numerically that a better approximation to F^n was in the Weibull family. Both Cohen (1982a) and Gomes (1984) have shown that in many cases that this approximation improves upon the rate of convergence. The statistical implication is that it is better to estimate F^n by fitting the penultimate approximation rather than the ultimate approximation. Note for the normal extremes, Cohen (1982b) gives the rates for the ultimate and penultimate approximation respectively at $O((\log n)^{-1}))$ and $O(\log n)^{-2})$ which shows that the penultimate approximation improves on the ultimate approximation as sample size increases. The case where $k \neq 0$ – either the Fréchet or Weibull case – is more recent. Gomes and Pestana (1986) give examples of the penultimate's efficacy.

1.2.3 Limit Laws for the Joint Distribution of the Sum and Maximum

Chow and Teugels (1978), who solved the asymptotic joint distribution of the sum and the maximum for the *iid* case, and Anderson and Turkman (1991), who extended the results to the dependent case, are the two primary sources for this joint distribution problem although others working on this problem precede them – for example, D.A. Darling (1952) and Arov and Bobrov (1960) – and continue to work on this problem – for example, Mori (1981) and Vangel (1999).

Here we first introduce some definitions necessary to understanding the Chow and Teugels (1978) result and then give this *iid* result. Next, we define conditions which limit the dependency in the Anderson and Turkman (1991) paper. Finally, we present the Anderson and Turkman (1991) result.

Let $X, X_1, \ldots, X_n$ be an *iid* sequence of random variables with common distribution F.

Definition 32 (Stable distribution, Feller, 1971) *The distribution F is stable if for each n there exist constants $c_n > 0, \gamma_n$ such that*

$$S_n \stackrel{d}{=} c_n X + \gamma_n$$

and F is not concentrated at one point. F is stable in the strict sense if the above formula holds with $\gamma_n = 0$.

Theorem 33 (Thm VI.1.1, Vol. II, Feller, 1971) *The norming constants are of the order* $c_n = n^{1/\alpha}$ *with* $0 < \alpha \leq 2$.

The constant α is called the characteristic exponent of F.

Note in their derivation Chow and Teugels (1978) utilize the characteristic function – $\omega_\alpha(t,p)$ – of a stable law. In particular, they define

$$\log \omega_\alpha(t,p) = \begin{cases} |t|^\alpha C \frac{\Gamma(3-\alpha)}{\alpha(\alpha-1)}[\cos\frac{\alpha\pi}{2} - i\frac{t}{|t|}(p-q)\sin\frac{\alpha\pi}{2}] & \text{if } \alpha \neq 1 \\ -t|C[\frac{\pi}{2} + i\frac{t}{|t|}(p-q)\log|t|] & \text{if } \alpha = 1 \end{cases}$$

In the above expression, $0 \leq \alpha \leq 2, 0 \leq \rho \leq 1, q = 1-p$ and $C > 0$.

Definition 34 (Domain of a stable law, by Anderson and Turkman, 1991) F *is said to belong in the domain of attraction of the stable law with characteristic function* $\omega_\alpha(t,p)$ – $F \in \mathcal{D}(\alpha,p)$ – *iff either (1)* $0 < \alpha < 2$ *and for some slowly varying function* $\mathcal{L}$

$$1 - F(x) \sim p\frac{2-\alpha}{\alpha}x^{-\alpha}\mathcal{L}(x), \quad x > 0$$

or (2) $\alpha = 2$ *and for some slowly varying function* $\mathcal{L}$

$$\mu(x) = \int_{-x}^{+x} y^2 dF(y) \sim \mathcal{L}(x), \quad x > 0.$$

Using the above concepts, we can now present Chow and Teugels (1978) result.

Theorem 35 (Thm 1, Chow and Teugels, 1978) *Theorem: Let* $X_1, \ldots, X_n$ *be an iid sequence of random variables with non-degenerate common distribution function* F*. Define* $S_n = \sum_{i=1}^n X_i$ *and* $M_n = \max(X_1, \ldots, X_n)$*. Let* $a_n > 0$ *and* $c_n > 0$*. Then* $(U_n, V_n) = (\frac{S_n}{a_n} - nb_n, \frac{M_n - d_n}{c_n})$ *converges in distribution to a limit* (U, V) *where neither* U *nor* V *is degenerate iff* F *lies in both the domain of attraction of an extreme value distribution and a stable law; i.e.,* $F \in \mathcal{D}(\alpha,p) \cap \mathcal{D}(H)$*. The random variables* U *and* V *are independent unless* $0 < \alpha < 2, p > 0, H = \Psi_\alpha$.

Thus it follows from Chow and Teugels (1978) that if F has a finite variance and the limits U and V are non-degenerate then U must be normally distributed and if $F(x) < 1$ for all x, V must follow either a Gumbel distribution Λ or a Fréchet distribution Φ_α with $\alpha \geq 2$. In this case U and V are independent and F must belong to the domain of attraction of V.

Anderson and Turkman (1991) extended the results of Chow and Teugels (1978) to stationary sequences. They also extended the results to hold when F is in the domain of attraction of the Weibull (Ψ) extreme value distribution. Thus they showed even under certain dependency conditions that S_n and M_n are asymptotically independent. Recall the definitions of strong mixing – definition 6 — and the extremal index for a stationary sequence – definition 12. Anderson and Turkman (1991) define the following condition to restrict the amount of clustering.

Definition 36 *The condition $D'(a_n, u_n)$ of Anderson and Turkman(1991) is said to hold if for each $\tau > 0$*

$$\lim_{k\to\infty} k \limsup_{n\to\infty} \sum_{j=1}^{[n/k]} E\{|\exp(ita_n^{-1} \sum_{\substack{l=1 \\ l\neq j}} X_l) - 1|\chi(X_j > u_n)\} = 0$$

where [] denotes the integer part and χ is the indicator function.

Note the definition of $D'(a_n, u_n)$ is a variant of condition $D'(u_n)$ – see definition 10 from Leadbetter *et al.*(1983). Recall the condition $D'(u_n)$ limits the amount of clustering. In particular it limits the probability of one exceedance following another. Also note in the derivation, Anderson and Turkman (1991) have taken $E(X) = 0$ so that the normalizing constant b_n can be taken to be 0. Specifically their main result is as follows.

Theorem 37 (Thm 1, Anderson and Turkman, 1991) *Suppose the sequence $\{X_i\}$ is stationary, has zero mean and finite variance, and satisfies,* *(i.)* $\{X_i\}$ *is strong mixing and has positive extremal index;*

(ii.) For some constants $a_n > 0$ *with* $\lim_{n\to\infty} a_n = \infty, c_n > 0,$ *and* d_n *that marginally,*

$$S_n^* = S_n/a_n \xrightarrow{d} \mathcal{N}(0,1),$$
$$M_n^* = (M_n - d_n)/c_n \xrightarrow{d} H$$

where H *is one of the extreme value distributions* Λ, Φ_α *for some* $\alpha > 2$, Ψ_α *for some* $\alpha > 0$.

(iii.) $\{X_i\}$ *satisfies* $D'(a_n, u_n)$ *condition.*

Then

$$\lim_{n\to\infty} E[\exp(itS_n^*)\chi(M_n^* \leq x)] = \exp(-t^2/2)H(x)$$

so that S_n^* *and* M_n^* *are asymptotically independent.*

At this point we conclude the literary review of the theoretical results around which this thesis is developed. Chapters 2 through 4 will utilize these results to develop a higher order expansion for the joint density of M_n and S_n.

1.3 Climate Literary Review

The impetus for such an expansion of the joint density of the mean and the maximum involves a climatological question which has and continues to draw much attention. In the broadest sense, this question – the application we look at in Chapter 5 – focuses on how the extremes in a weather series affect the average weather – for example, how a change in the hottest day of the year or the storm with the most rainfall might affect the average annual temperature or total rainfall? This discussion was instigated shortly after the publicity of global warming – a now well established increase in the average surface temperature of between 0.3^{o}C to 0.6^{o}C since the latter half of the nineteenth century (Karl *et al.*, 1991). Again we do not want to review all the literature concerning global warming and its related topics. Our main focus will be studies which tie the extreme climate events to the overall weather patterns.

The study of climate change which refers to the shift of weather features over relatively large areas (global or continental) and over long periods of time (decades, centuries, or millennia) has represented and continues to represent an enormous undertaking at all levels of its investigations: from the global perspective down to the regional level, from improving and understanding the historical climate records to predicting events over the next millennium, and from understanding what influences humans have had on the process to what humans can do to alleviate and/or adapt to these changes. Thousands of scientists have worked and are working on issues relating to this problem in international agencies, national agencies, academia, and private industry. Some of the notable groups include the Intergovernmental Panel on Climate Change (IPCC) which was established in 1988 by the United Nations; in the United States, the National Climate Data Center (NCDC) which is a branch of the National Oceanic and Atmospheric Agency (NOAA) and was established in 1951; the Goddard Institute in the National Aeronautical and Space Administration (NASA); and the Pew Center for Climate Change at the National Center of Atmospheric Research (NCAR). The private industries which have people working with this topic are as diverse as Dow Chemical Inc., Glaxo Welcome, Inc., Meryll Lynch, and Ford Industries. To see a more complete look at this diversity, see IPCC (1996).

Obviously there is too much literature to review – maybe too much to expect any consensus given the differences between the groups and their motives. Here we give the broadest of overviews – some of the basic definitions and notions, some of the basic methods people are using, some major accomplishments that have led to an increased confidence in climate change results, and some of these major results. Throughout we highlight what we specifically will be studying in Chapter 5 – the data analysis of the total and maximum annual precipitation in the contiguous United States. We give the more general information about the study of climate change to give a perspective to our data analysis – both to understand the impetus for our analysis and to better interpret its results.

The climate – the "average" weather – is obviously a very complicated and interactive system which is influenced by both internal and external forces. The internal components include the atmosphere, the oceans, biosphere, cryosphere (ice and snow), and the lithosphere (Earth's crust). The external forces, those which influence but are not influenced by the climate, include the Sun and its output, the Earth's rotation, the Sun-Earth geometry, and volcanic eruptions.

These components are linked by the flows of energy and mass. Mass flows are in fact cycles. Since mass cannot be destroyed, mass – in one form or the other – flows from one reservoir to another in the climate system. Mass cycles include the water cycle, the carbon cycle, sulphur cycle, and nutrients (i.e. phosphorus and nitrate) cycle. Energy flows include the transfer of momentum between the atmosphere and the ocean, sensible heat, latent heat, and solar and infared radiation. Note this last piece has been the focus of much research and popular speculation. In particular, the greenhouse effect refers to the reduction in the loss of infared radiation to space by the atmosphere which tends to make the climate warmer. The gases which play a major role in the greenhouse effect by either trapping or reflecting the radiation are water vapor, carbon dioxide CO_2, methane CH_4, nitrous oxide N_2O, and ozone O_3. Note clouds also play an intricate part of this process. These greenhouse gases – GHG – are natural and necessary for the climate but it is the "enhanced" greenhouse effect that has been the focus of much research and debate. It refers to the *increased* levels of the GHG above their natural existing levels and their impact on the climate including the average net radiation. A change in the average net radiation – the average amount of solar radiation that is absorbed into the Earth's atmosphere and how much long wave radiation is released – is referred to as a radiative forcing of the system. This increase in concentration of these GHG will trap a significant amount of heat. This heat will almost certainly lead to changes in the climate that are significant from a human and ecological point of view (Harvey, 2000). For example, a simplified version of this effect shows that this greater concentration of carbon-based gases and water vapor better insulate the earth which in turn leads to an increase in temperature which leads to more evaporation and hence more precipitation. The actual process can be quite intricate and highly interdependent.

To be complete there are both natural and human causes for possible climate change. Natural causes on a global scale involve changes in topography, land-sea geography, bathymetry (ocean basins), and internal variability of the atmosphere-ocean system. On a shorter time scale, the natural causes include solar luminosity, the Earth's orbit, volcanic eruptions, and even El Niño – the warming of the equatorial Pacific ocean which occurs as irregular intervals of 3 to 5 years. Human causes include the increases in concentration of well mixed GHG, changes in ozone, levels of aerosols, and changes in land surfaces such as deforestation. Note the last three have more impact at a regional or continental level but which ultimately effects the global climate. Note it is the opinion of many – including Harvey (2000) – that the natural forces will almost certainly be overwhelmed by the heating effect of the increasing GHGs concentration during the 21st century. For more information on the climate system and changes in it, see Chapter 2 of Harvey (2000).

Note often the terms climate change and climate variation are used to differentiate, respectively, between changes in the climate which are directly or indirectly due to human activity versus due to natural variation. Throughout this thesis we will not distinguish between these two. More specifically climate change will refer to a change in the climate regardless of the cause. For further information on the climate system, a good reference is K.E. Trenberth *et al.* (1996) or Harvey (2000).

Although climate change refers to change in the average weather, the weather itself is made up of many variables, what are referred to as climate indicators. The two indicators most studied and most recognizable are temperature and precipitation.

Another feature of the climate that has recently been emphasized is extreme events and/or climate variability – both in temperatures and precipitation. In fact, the relationship between extremes and the average climate has garnished attention from Karl *et al.* (1995, 1996, 1998) at the NCDC and Wigley(1999) of the Pew Center for Climate Change at NCAR, among others. Other climate indicators include the atmospheric circulation and storms – hurricanes, tornadoes, cyclones, etc.

There are two broad areas of investigations related to climate change. The first involves the detection of trends. Our data analysis falls into this category. In general, detection is more straightforward and usually – but not always – pertains to just the analysis of the historical record. The detection questions can be further specified – although this is not the objective of this paper – to discern significant difference between natural variability in the climate system and anthropogenically forced variability. The second area of interest in climate change which draws the most attention and the most debate is the attribution of the cause for climate change. Establishing cause and effect is a challenging endeavor. Since we only have one Earth, "experiments" run to investigate causation must involve simulation projects which is where the General Circulation Models (GCMs) play their prominent role. For more information, see IPCC (1996). As stated previously, at no step in our analysis do we imply any causation to the trends we estimate so this attribution question is not addressed.

There are two basic vehicles to study climate change – analysis of historical records (observational data) and the use of the GCMs (modeled data). A third approach involves studying the socioeconomic factors associated with climate change – for example, population growth and/or shift, necessary insurance reserves, and educational programs. This third factor is of interest mainly to policymakers but given the interactive nature man has with his environment these factors can both influence climate change and be influenced by it.

One of the major developments in the past 25 years in this field has been the collection and maintenance of the historical climate records and a more uniform structure for providing quality control for these large data sets. In the United States, the NCDC was established for just this purpose. The United States Historical Climate Network (USHCN) is the culmination and now ongoing task of the NCDC and the Carbon Dioxide Information and Analysis Center (CDIAC) of Oak Ridge National Laboratory. The USHCN was constructed to help address issues concerning climate change. It is a high quality, moderate size data set that includes minimum and maximum temperatures along with precipitation. There are a total of 1221 stations in this particular network within the contiguous US, although roughly between 180 and 190 stations are considered primary. In general, the records run from 1901 to 1997. The criteria for a station's inclusion include length of record, percentage of missing data, number of station moves, and spatial coverage. These data have gone through extensive quality control to correct for human error, time of observation bias, equipment adjustment, and urban warming. They do have a procedure to estimate missing data. For precipitation, it involves generating gamma random variables. For more information concerning the USHCN, please see NOAA website (http://www.ncdc.noaa.gov/ol/climate/research/ushcn/ushcn.html).

Other data sets which have been meticulously groomed since their initial collection exist. In fact, there are three well established global scale data sets of mainly land surface-air temperatures: (1) produced at the University of East Anglia (UK) by Professor Philip Jones, (2) produced by Doctor James Hansen and colleagues at the Goddard Institute of Space Studies (GISS) in the United States of America, and (3) produced by Doctor Konstantin Vinnokov and colleagues at the State Hydrological Institute in St. Petersburg, Russia. See Section 5.1 of Harvey (2000) for more details on these data sets. In general, data sets that are considered very reliable have roughly the same length of record as the USHCN data set – usually starting in the late 19th century or early 20th century. There are some studies, see Lamb (1965, 1988), which discuss data going back to 1400 AD but spatial coverage is quite limited. There are obviously other sources of climate data – ice core, tree ring, lake level, and coral data, see Cook (1995) – but the information that can be extracted from these data sets are not of the same time scale or accuracy that we need in our analysis of the annual precipitation across the contiguous US.

Studies which involve only the analysis of historical data sets are limited in their scope. Although they may help to determine if a trend in a climate indicator has occurred, they cannot tell us what caused this trend. We need to keep this in mind when interpreting our results. Given this limitation, data analysis projects are often teamed with climate simulation models to aid in the interpretation and causation of results of the data analysis part.

The second biggest development in the past 25 years in this field is the improvement of the GCMs. Although these models are not used in our analysis in Chapter 5, they are important to know since some of the work that instigated this work, see Karl *et al.* (1995, 1996), involves climate simulation data. The climate models' primary role in studying climate change is to simulate the climate as it now operates and then to perturb certain features to see the effects. It is necessary to calculate the effects of all key processes in the climate system. This means representing the processes as mathematical formulae. Climate models are those which contain enough of these processes to be considered useful in representing the entire system.

Obviously the climate system, even just the parts that are currently understood, is too complex to be thoroughly modeled. There are simplifications which of a necessity must be made. Although invariant principles such as Newton's laws of motion are ideal in formulating these models, in some cases empirically derived relationships are included. Because they were not derived rigorously, it is possible these relationships are not applicable under different circumstances – i.e. when the climate changes. Another simplification is the discretization of the continuous climate process, in particular, selecting a fixed set of points to represent a region or even globe. The spacing between the points on the grid used is the "spatial resolution" which is typically hundreds of kilometers in the horizontal on the global scale. It is important to realize that many key elements of the climate system have scales much smaller than this – note, in particular, many extreme events. A formulation of the effects of a small-scale process on a larger scale is called "parameterization." A third simplification involves averaging over a complete spatial dimension; for example, simulating the three dimensional climate in two or even one dimension.

There are a variety of complexity to the climate models. Some of the more simple models are one dimensional atmosphere models or one dimensional energy-balance models. The most complex are the three dimensional atmospheric and oceanic general circulation models (AGCMs and OGCMs). In fact, only when these two are coupled do we have what is considered to be a realistic climate model. These models include simulation of winds; ocean currents, temperature, and salinity; clouds; precipitation evaporation; soil moisture; and many other features. Due to the complexity of the model, a majority of these processes are parameterized. The more simple models are used to investigate basic relationships between components and exploring global-scale equations. The more complex models can give insights on regional climate change, at shorter time scales, or on processes that need a finer resolution. The more complex models are also costly, can be difficult to understand, and with the higher resolution can produce "noise" in the system. Note the simpler models can be calibrated – i.e. have variables identified and parameters estimated – by the more complex models. For more information on climate models, see Chapter 5 of Harvey (2000). Note successfully modeling the major impact that the oceans have on the climate is considered one of the significant improvements in these GCMs along with including the radiative forcing especially due to the sulfate aerosols and the secondary solar effects. For more details, see Wigley *et al.*(1997).

Kattenberg *et al.*(1996) is a very good resource for the results using the GCM models. Research shows that increasing CO_2 in all models produces an annual mean warming, particularly in the high northern latitudes and especially in the autumn and winter months. How big this increase is depends on the CO_2 forced into the system although the best estimate for this increase is 0.3^oC/decade in the early 21st century.

When aerosol forcing is included in the model this increase drops to 0.2^oC/decade. For precipitation, an increase in CO_2 leads to an increase in the global mean precipitation for all models. This increase is notable in the high latitudes in the winter and in most cases extends well into the mid-latitudes. Again, when including aerosol forcing, this predicted increase in global mean precipitation is reduced. With respect to changes in extreme events and/or variability, predictions using GCMs are more difficult since these events are "small scale"; that is, the resolution at which these GCMs are modeled is not fine enough to accurately portray these rare events. Essentially, predicting changes in extremes means predicting changes in probability distributions. Current beliefs are that (1) a change in the mean temperature will have a substantial impact on exceedance probabilities (due to the short upper tail of the temperature distribution) and (2) a change in temperature variability also affects the occurrence of extreme events. There is some debate which has more impact on extreme events – a change in the mean or in the variability. Katz and Brown (1992) found a change in variability has a greater influence on monthly maximum temperatures. On the other hand, Cao *et al.* (1992) and Hennessey and Pittock (1995) found using enhanced greenhouse simulations that changes in mean temperatures have a greater affect than climate variability. Some of the predictions of the GCMs for extreme events and/or variability are: (1) general warming tends to lead to an increase in extremely high temperatures and a decrease in winter days with extremely low temperatures, (2) a decrease of daily temperature variability (most notably a decrease in the diurnal temperature) for some regions and an increased daily precipitation variability in a few regions, and (3) an increase in precipitation intensity – implying more extreme rainfall events – and also more frequent and/or severe drought periods in a warmer climate. For further details, see Kattenberg *et al.* (1996).

Again we could not begin to summarize the results from the data analysis studies, from GCM studies, and studies that used both methods that continue to be calculated. Overall we see that some of these climate indicators have significant trends while others do not. Even those with significant trends are not necessarily uniform with respect to time or space. To see a comprehensive survey of the results, see IPCC (1996). At this point it becomes necessary to limit our review to those studies directly applicable to our work. In particular we focus on extreme events and their relationship to the mean climate of the contiguous US – in particular, the precipitation series. Studying precipitation is considered more difficult than temperature because in general there is greater variability from one point to another. There are two global scale data sets: one by Professor Hulmes and the other is the Global Historical Climate Network. See Harvey (2000) for more details on these data sets. On a global scale, the findings are consistent with a warmer climate. The mean precipitation is generally increasing outside the tropics and decreasing in Sahel. As for the precipitation extremes, there is growing evidence that they are taking on more importance. (Harvey, 2000) For our particular analysis of rainfall in the contiguous US, there are three main papers which give the necessary background information. In fact, they were the impetus of the analysis. They are: Karl, Knight, and Plummer (1995), Karl, Knight, Easterling, and Quayle (1996), and Karl and Knight (1998). They are important because they give the best look at how climatologists model the connection between means and extremes.

Karl *et al.* (1996) provides the best general overview of the US climate with respect to the climate extremes. The two instruments used in this paper to study climate extremes were the CEI, the Climate Extremes Index which is an aggregate set of conventional climate extreme indicators, and the GCRI, the US Greenhouse Climate Response Index which includes indicators that measure the changes in the climate of the US that have been projected to occur as a result of increased emissions of greenhouse gases. The method used was to fit ARMA models to the time series of each indicator, using the BIC criterion to select the most appropriate order. Trends were removed prior to the fitting. Then the trend of the observed series is compared to 1000 Monte Carlo simulations from the generated time series. The fraction of the time the observed trend exceeds those calculated from the simulated series is used as a measure of the statistical significance of the observed trend. The general conclusions were: (1) The changes in the CEI support the notion that the climate of the US has become more extreme in the recent decades – yet the magnitude and persistence of the changes are not now large enough to conclude that the increases in the extremes could not have arisen from the quasi-stationary climate and (2) changes in the GCRI are consistent with the expected significance of change due to the enhanced greenhouse effect but the increases are not large enough to unequivocally reject the possibility that the increases in the GCRI may have resulted from other factors including natural climate variability. In other words, they found evidence – but not overwhelming evidence – the climate of the contiguous US is becoming more extreme.

The premise of Karl *et al.* (1995) is that understanding climate change requires attention to changes in climate variability and extremes but that knowledge of recent behaviour of these variables had been limited by the unavailability to data. Specifically, they focused on climate indicators of temperature and precipitation. They analyzed four data sets: 187 stations from the USHCN, 223 stations from the former USSR, 197 from the People's Republic of China, and 40 stations from Australia. For temperature data, they calculated anomalies in the daily maxima, minima, average, and the diurnal range series where the daily anomalies were calculated using the first three harmonics of the period of record mean annual daily temperature series being analyzed. Then they defined temperature variation as the mean of the series defined by absolute difference in the mean temperature anomaly (maximum, minimum, average, or diurnal range) from time frame i to $i + 1$. Values were arithmetically averaged within regions and then area weighted across the countries. Daily rainfalls were aggregated into 5 categories from very light to extremely heavy and the proportion in each category calculated. A national mean was calculated by area weighting. For both temperature and precipitation, significance of the trends was established via the Monte Carlo method used and described in the above paper of Karl *et al.* (1996). With respect to their temperature analysis, they found that the interseasonal temperature variability has generally decreased in the Northern Hemisphere. Specifically for the contiguous US, the day-to-day temperature variability for all elements in the US is significantly decreasing. These decreases in interseasonal variability in the US are primarily due to decreases in the spring and summer. With respect to precipitation for the US, they found a clear signal that there is an increase in the proportion of precipitation derived from the extreme precipitation category (> 2 inches). This increase is a result of increases during all seasons although more readily observed in the summer and then spring. This increase is seen throughout the country except the far west and the southeast. The trend in the total precipitation in the summer and spring is near zero. Note their finding are consistent with model projections of a warmer world.

Finally, Karl and Knight (1998) gives the most comprehensive look at precipitation across the contiguous US. The data used in the analysis consists of 142 stations from the USHCN plus an additional 48 stations which used a standard 8 inch gauge in the collecting of the precipitation data. Essentially their method is via summary statistics – certain weighted spatial averages. More specifically, the precipitation data was arithmetically averaged into $1^o \times 1^o$ grid cells. These cells were area weighted to calculate changes in precipitation for the nine regions they used. A national average was calculated by area weighting the nine regions. Finally, they used a nonparametric Kendall τ test ($\alpha = 0.05$) to detect significant trends. They found that since 1910 precipitation has increased in the continental US by about 10% but that one statement, which is often quoted, is an oversimplification of their results. By focusing on different quantiles of the precipitation data, they maintain the precipitation distribution itself has changed, making this precipitation increase fairly complex. They found: this increase is affected by both the frequency and intensity of precipitation; in all categories, the probability of precipitation on any given day has increased; precipitation intensity has increased only in the extremes; and in fact, the increase in total precipitation derived from the extreme events is higher relative to the moderate and low events. The last finding – that the increase in precipitation is primarily due to heavy/extreme daily precipitation events – is most interesting. In fact, they found that 53% of the rise in the total increase is due to a positive trend in the upper 10% of the probability distribution despite the fact these upper tail events only constitute about 35-40% of the total annual precipitation. This is seen predominantly in the summer and then spring and in general holds for all regions across the US except for the far West and the Southeast. In summary, since

1910 they are seeing a positive trend in total precipitation and in the number and intensity of extreme events. Moreover the increase in the upper percentiles is driving the increase in the total precipitation.

Their findings are consistent with other contributors to the 1996 IPCC report. There is evidence of a small positive (1%) global trend in precipitation over land during the 20th century although precipitation has been relatively low since the 1980s. This trend is non-uniform – positive trends exist in some regions and not others (Groisman and Legates 1995, Nicholls *et al.* 1996). Globally the data on extremes in precipitation are inadequate to say anything about a global change (IPCC 1996, p137). On the US scale, the increase in annual precipitation is most apparent after 1950 and is in large part due to increases during the autumn (September to November). The increases average out to about 5% across the contiguous US. Also the increases are more prevalent in the eastern two-thirds of North America.(Findlay *et al.*, 1994, Lettenmaier *et al.* 1994). Finally for extremes in precipitation in the continental US, Iwashima and Yamamoto (1993) also found positive trends in the higher frequency of extreme 24 hour rainfall totals. Finally, prevalent in all precipitation analysis is the marked year to year variability (Wigley, 1999).

In conclusion, in the past decade there has been increasing focus on the extreme events of some of the climate indicators, in particular, precipitation and temperature. This discussion has lead to inquires about the relationship between these extremes and the mean climate. Although globally there are issues concerning the availability of reliable data which has the necessary information, the USHCN affords the study of US precipitation a better opportunity as can be demonstrated with the above studies. Another importance of these three studies, especially Karl and Knight (1998), is that they show how the climatologists connect the mean and the extremes.

Chapter 2

EXPANSION OF THE JOINT DENSITY UNDER THE GUMBEL DOMAIN OF ATTRACTION

2.1 Introduction

This chapter studies the relationship between the sum and maximum of an *iid* sequence of random variables. The work of Chow and Teugels (1978) – followed by Anderson and Turkman (1991) – established that asymptotically the sum and the maximum are independent. Specifically they showed that, under appropriate conditions, the joint distribution of the sum and the maximum converges to the product of their individual asymptotic distributions – namely, the normal distribution and one of the three extreme value distributions. The chief theoretical question that remains is to determine the rate of convergence of this asymptotic independence. The associated statistical methodology question is, in moderate sample size where the asymptotic result is not yet realized, how does one model the dependence structure between the sum and the maximum? Given the importance of both the rate of convergence and the statistical modeling questions, an important development in this area is a higher order *expansion* for the joint *density* of the sum and the maximum. This is the goal of this chapter, along with the two subsequent chapters.

Specifically, we look at the following structure. Let $X_1, \ldots, X_n$ be an *iid* sequence of random variables with common distribution function F which has density f and characteristic function φ where the support of F lies on (x_l, x_o) where $-\infty \leq x_l$, $x_o \leq \infty$. We assume the existence of the third moment μ_3 from which follows the existence of the mean μ, variance σ^2, and third cumulant $\mathcal{K}_3$.

Define $S_n = \sum_{i=1}^{n} X_i$ with the normalized version as

$$S_n^* = \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \tag{2.1}$$

and $M_n = \max_{1 \leq i \leq n} X_i$ with the normalized version being $M_n^* = \frac{M_n - b_n}{a_n}$ where $a_n > 0, b_n$ real.

We define the distribution function of S_n^* as $F_{S_n^*}$ with density $f_{S_n^*}(w) = dF_{S_n^*}(w)/dw$. We also define the distribution function of M_n^* as $F_{M_n^*}(v) = F^n(a_n v + b_n)$ with density $f_{M_n^*}(v) = dF_{M_n^*}(v)/dv = nF^{n-1}(a_n v + b_n) f(a_n v + b_n) a_n$.

If we let h_n denote the higher order term in the expansion of the joint density of S_n^* and M_n^* with R_n as the remaining error term associated with this expansion, then one of our specific goals is to establish $h_n(v, w)$ and R_n such that

$$|f_{S_n^*, M_n^*}(w, v) - f_{S_n^*}(w) f_{M_n^*}(v)\{1 + h_n(v, w)\}| = o(R_n) \tag{2.2}$$

uniformly in some interval of w and v. The term $h_n(v, w)$ is important because as a first order approximation it describes the dependence structure between S_n^* and M_n^*. The term R_n is important since it gives the error associated with this approximation.

Since we are assuming a finite variance, we have that F lies in the domain of attraction of a stable law with index equal to 2; that is, $F_{S_n^*}(w)$ converges to $\mathcal{N}(w)$ where $\mathcal{N}(w)$ denotes the normal distribution function. Note the normal density is denoted by $\mathcal{N}'(w)$. As seen in Chapter 1, there is a rich literature detailing not only when the distribution of S_n converges but also when the density and its higher order expansions converge. We will rely on Feller (1971), Chapter XVI, Section 2, Theorem 1 which gives the uniform expansion of the density of S_n^* – the unconditional version. Given the conditions of Feller's expansion for the density, we also have the result due to Petrov (1975) for this expansion which says that we can bound $x^m\{f_{S_n^*}(x) - \mathcal{N}'(x)\}$ uniformly $\forall x$ when $m \le 3$.

We also assume F lies in the domain of attraction of one of three extreme value distributions. Specifically, in Chapter 2 we assume that F lies in the Gumbel domain of attraction. We let Λ denote the Gumbel distribution function and Λ' denote its density where

$$\Lambda(x) = \exp -e^{-x},\ -\infty < x < \infty,$$

and

$$\Lambda'(x) = e^{-x}\exp\{-e^{-x}\},\ -\infty < x < \infty.$$

Under the Gumbel domain of attraction, Balkema and DeHaan (1972) refined the Karamata representation from Chapter 1. This representation is as follows. If F is in the domain of attraction of Λ than $F(x)$ may be written in the form

$$-\log\ F(x) = c(x)\exp\ \{-\int_{-\infty}^{x}\frac{dt}{\phi(t)}\}\quad \forall x < x_o \tag{2.3}$$

where ϕ is a positive twice differentiable function, $\phi'(x) \to 0$ and $c(x) \to 1$ as $x \to x_o$, and $x_o = \sup\ \{x : 1 - F(x) > 0\}$. We can then take a_n and b_n – the normalizing constants for M_n under the Gumbel distribution – to be

$$a_n = \phi(b_n)\ \text{ where } \phi \text{ is defined in (2.3)} \tag{2.4}$$

and

$$b_n = \inf\ \{x : -\log\ F(x) > 1/n\}. \tag{2.5}$$

With these assumptions we can specify another form of the result – useful in statistical modeling

$$|f_{S_n^*,M_n^*}(w,v) - \mathcal{N}'(w)\Lambda'(v)\{1 + h_n'(v,w)\}| = o(R_n') \tag{2.6}$$

uniformly in some interval of w and v. Note that h'_n may be be different from h_n and R'_n may be different from R_n. The idea in statistical modeling is to replace $f_{S_n^*,M_n^*}(w,v)$ by $\mathcal{N}'(w)\Lambda'(v)\{1+h'_n(v,w)\}$. The term $h'_n(w,v)$ should lead to a better modeling of S_n^* and M_n^* than simply modeling them separately, i.e., by $\mathcal{N}'(w)\Lambda'(v)$. The R'_n term gives the scale of error associated with the above substitution.

Our approach in solving either (2.2) or (2.6) is to rewrite $f_{S_n^*,M_n^*}(w,v) = f_{S_n^*|M_n^*}(w|v) \times f_{M_n^*}(v)$ where $f_{S_n^*|M_n^*}(w|v)$ is the conditional density of S_n^* given M_n^*. Then we need to establish three key expansions. These expansions form the three main propositions of this chapter which in turn combine to give the main results. First we need to establish the expansion for the conditional density of $S_n^*|M_n^*$. Then we need to derive the expansion for the density of M_n^*. Finally, we need the expansions for the conditional mean and variance of $S_n^*|M_n^*$.

A key in the derivation is connecting the two random variables – S_n and M_n. This is done by conditioning one on the other, here S_n given $M_n = u_n$ where $u_n = a_n v + b_n$ with a_n and b_n defined in (2.4) and (2.5) and with v fixed.

The conditional distribution of S_n given M_n may be written in the form

$$P[S_n \le x | M_n = u_n] = P[(\sum_{i=1}^{n-1} X_i^* + u_n) \le x | M_n = u_n]$$

where X_i^*s are *iid* conditional random variables with

$$P[X_i^* \le x] = P[X \le x | X \le u_n].$$

Hereforth we write X^* for the random variable with the distribution $F(x)/F(u_n), \forall x \le u_n$, and write $\overline{S_n} = X_i^* + u_n$ where $X_1^*, \ldots, X_{n-1}^*$ are *iid* with the same distribution as X^*. The dependence on a given sequence $\{u_n\}$ is implicit in the notation. Thus the conditional distribution of S_n given $M_n = u_n$ is the same as the unconditional distribution of $\overline{S_n}$, and we shall use this equivalence in the following discussion.

We may write

$$E(\overline{S_n}) = (n-1)\mu(u_n) + u_n$$

where $\mu(u_n) = E(X^*) = E(X|X \le u_n)$ and

$$Var(\overline{S_n}) = (n-1)\sigma^2(u_n)$$

where $\sigma^2(u_n) = Var(X^*) = Var(X|X \le u_n)$. Note we let $\kappa_3(u_n)$ denote the third cumulant of X^*. Now we also let $\tilde{S_n} = \{\overline{S_n} - E(\overline{S_n})\}/\sqrt{var(\overline{S_n})}$ as the normalized version of $\overline{S_n}$. Then

$$\begin{aligned}
P[\tilde{S_n} \le x] &= P[\frac{\overline{S_n} - \{(n-1)\mu(u_n) + u_n\}}{\sqrt{(n-1)\sigma^2(u_n)}} \le x] \\
&= P[\frac{S_n - \{(n-1)\mu(u_n) + u_n\}}{\sqrt{(n-1)\sigma^2(u_n)}} \le x|M_n = u_n] \\
&= P[\frac{S_n - \{(n-1)\mu(u_n) + u_n\}}{\sqrt{(n-1)\sigma^2(u_n)}} \le x|M_n^* = v] \\
&= P[\frac{S_n^*\sqrt{n\sigma^2} + n\mu - \{(n-1)\mu(u_n) + u_n\}}{\sqrt{(n-1)\sigma^2(u_n)}} \le x|M_n^* = v] \qquad (2.7)
\end{aligned}$$

where the last step comes from substituting (2.1) into the formula. Thus (2.7) gives the form of the transformation we need. Note the Jacobian of this transformation is $\frac{\sqrt{n\sigma^2}}{\sqrt{(n-1)\sigma^2(u_n)}}$.

The final key in the derivation is exploiting the relationship between the generalized extreme value (GEV) distribution and the generalized Pareto (GP) distribution. Let us define Y as the exceedance over the threshold u_n; that is, $Y = X - u_n|X > u_n$. The conditional distribution function of Y given $X > u_n$ is

$$F_{u_n}(y) = P(X \le u_n + y|X > u_n) = \frac{F(u_n + y) - F(u_n)}{1 - F(u_n)}.$$

Note that the tail distribution function is $1 - F_{u_n}(y) = \frac{1-F(u_n+y)}{1-F(u_n)}$.

Define $m(u_n)$ and $s^2(u_n)$ as the conditional mean and variance of Y given $X > u_n$. Note we have $E(X|X > u_n) = m(u_n) + u_n$ and $Var(X|X > u_n) = s^2(u_n) = Var(Y|X > u_n)$.

This construction allows us to write the mean and variance of $\tilde{S_n}$ in terms of $F(u_n)$, $m(u_n)$ and u_n. An explicit expansion for $F(u_n)$ and then $m(u_n)$ depend obviously on assumption made on F.

Broadly, we need conditions on F to obtain the expansion of the conditional density of $S_n^*|M_n^*$ and for the expansion of the density of M_n^*. The conditions necessary for the expansion of the conditional moments are included in those needed for the two density expansions.

The plan of this chapter is as follows. The main theorem and its corollary are presented in Section 2.2. Section 2.3 contains the lemmas and their corollaries necessary for proving the three main propositions. Once these three propositions are established in Section 2.4, Section 2.5 sets out the proofs of the main theorem and its corollary. Finally Section 2.6 presents a simulation project conducted to study the behavior of the higher order expansion term of the joint density of the sum and the maximum and its improvement over the asymptotic result.

2.2 Main Theorem

Here we present the main theorem, followed by its corollary. The main theorem provides a useful result for statistical application. It delineates a first order correction for the dependence structure that exists between S_n^* and M_n^*; that is, it gives the approximation of the joint density as the product of the individual densities plus this higher order term.

The corollary provides an explicit algebraic approximation for the joint density of the S_n^* and M_n^*; that is, we can approximate the joint density as the product of the normal density and the Gumbel density plus the following higher order terms. The error associated with this approximation can be viewed as the maximum of two parts – one associated with the expansion of the density of the maximum and the expansion of the density of the sum and the other part a combination of the expansion of the conditional density of the sum given the maximum and the expansions of the conditional moments written in terms of the parameters of the maximum.

Theorem 38 *Let $X_1, \ldots, X_n$ be an iid sequence of random variables with distribution function F, density function f, characteristic function φ, mean μ, and variance σ^2. Let u_n be a threshold level and φ_{u_n} be the characteristic function of $X|X < u_n$.*

Given the following two sets of assumptions
Set A: Assume f' is integrable, μ_3 exists, φ''' exists and is continuous in a neighborhood of 0, and $|\varphi_{u_n}(t)|^n$ is integrable for $n \geq$ some $n^* > 1$.

Set B: Assume F is in the domain of attraction of Λ so that the representation in (2.3) holds. Use the same form of a_n and b_n as defined in (2.4) and (2.5). Also assume that for a constant $K > 2$,

$$a_n v + b_n \to x_o \quad \textit{uniformly in } |v| \leq -K \log\ |\phi'(b_n)| \ \textit{as } n \to \infty \tag{2.8}$$

$$\phi''(a_n v + b_n)/\phi''(b_n) \to 1 \quad \textit{uniformly in } |v| \leq -K \log\ |\phi'(b_n)| \ \textit{as } n \to \infty \tag{2.9}$$

$$\phi(b_n)\phi''(b_n) \log\ |\phi'(b_n)|/\phi'(b_n) \to 0 \quad \textit{as } n \to \infty. \tag{2.10}$$

Also for any $v^* < x_o$

$$\inf_{v \leq v^*} \frac{f'(v)F(v)}{f^2(v)} > -\infty \tag{2.11}$$

and when $x_o = \infty$,

$$\frac{\phi'(t) \log |\phi'(t)|}{t} \to 0, \quad \textit{as } t \to \infty,. \tag{2.12}$$

Defining

$$r_n = \begin{cases} \frac{b_n}{\sqrt{n\sigma^2}}, & x_o = \infty \\ \frac{x_o - \mu}{\sqrt{n\sigma^2}} & x_o < \infty, \end{cases} \tag{2.13}$$

then

$$|f_{S_n^*, M_n^*}(w, v) - f_{S_n^*}(w) f_{M_n^*}(v)\{1 - r_n(e^{-v} - 1)w\}| = o(r_n) \tag{2.14}$$

uniformly $\forall w$ and $\forall |v| \leq -K \log |\phi'(b_n)|$.

Corollary 39 *Given the conditions in Theorem 38,*

if $x_o = \infty$

$$|f_{S_n^*,M_n^*}(w,v) - \mathcal{N}'(w)\Lambda'(v)\{1 + (\frac{v^2}{2} - v - \frac{v^2 e^{-v}}{2})\phi'(b_n)\}\{1 - r_n(e^{-v} - 1)w\}| = o\{\max(r_n, |\phi'(b_n)|)\} \quad (2.15)$$

and if $x_o < \infty$

$$|f_{S_n^*,M_n^*}(w,v) - \mathcal{N}'(w)\Lambda'(v)\{1 + \frac{\kappa_3}{6\sigma^3\sqrt{n}}(w^3 - 3w)\} \times\{1 + (\frac{v^2}{2} - v - \frac{v^2 e^{-v}}{2})\phi'(b_n)\}\{1 - r_n(e^{-v} - 1)w\}| = o\{\max(r_n, |\phi'(b_n)|)\} \quad (2.16)$$

uniformly $\forall w$ *and* $|v| \leq -K \log|\phi'(b_n)|$.

Corollary 40 *Given the conditions in Theorem 38 and defining*

$$H(x;\eta,\psi,k) = \exp\left[-\{1 - \frac{k(x-\eta)}{\psi}\}^{1/k}\right]$$

Let $k_n = -\phi'(b_n)$ *and replace*

$$\Lambda'(v)\{1 + (\frac{v^2}{2} - v - \frac{v^2 e^{-v}}{2})\phi'(b_n)\} \quad (2.17)$$

in (2.16) and (2.15) by

$$H'(v;0,1,k_n) \quad \text{where} \quad H'(x;\eta,\psi,k) = \frac{d}{dx}H(x;\eta,\psi,k),$$

then we get the same result as Theorem 38.

Remark: This gives us the penultimate version of the approximation.

2.3 Lemmas

Here we present the lemmas necessary for the propositions of Section 2.4. This section has three subsections – lemmas necessary for Proposition 50, 51, and then 53. Proposition 50 gives the uniform expansion for the conditional density of S_n^* given M_n^*. The lemmas needed for Proposition 50 are as follows: Lemma 41 is a retooled Riemann-Lebesgue theorem, similar to Feller (1971) , Chapter XV, Section 4, Theorem 4. It gives conditions necessary so that the conditional characteristic function goes to zero uniformly at infinity. Lemma 42 provides a rate of the convergence in Lemma 41. Its first corollary, Corollary 43 gives conditions when the nth power of the conditional characteristic function is integrable. Its second corollary, Corollary 44, provides the Fourier norm of the conditional density of S_n^* given M_n^*. Lemma 45 shows that we can bound the conditional characteristic function by a number less than 1, uniformly. Finally, Lemma 46 is a technical result that shows the difference between a function of the conditional characteristic function and its third derivative tends to zero uniformly in a neighborhood of zero.

Proposition 51 gives the expansion of the density of the maximum when the maximum is in the Gumbel domain of attraction. Given $F \in \mathcal{D}(\Lambda)$, Balkema and DeHaan (1972) provide the representation for $-\log F(x)$, see (2.3). Using this representation, in Lemma 47, we are able to write an expansion for the ratio $\frac{-\log F(u_n)}{-\log F(b_n)}$. Given the definition of b_n, this allows us to write an expansion for $-\log F(u_n)$ and via a Taylor's expansion one for $1-F(u_n)$. In Lemma 48, we bound the function $e^{-mv}e^{-e^{-v}}$ uniformly $\forall v$. This is important to Proposition 51 in establishing the convergence rates uniformly.

Finally, Proposition 53 gives expansions for the conditional mean and variance of X given $X \leq u_n$. Its lemma, Lemma 49, deals with rates of convergence. In particular, it deals with how quickly the auxilary function defined in the Balkema and DeHaan (1972) representation tends to zero. This is important in interpreting which term in the expansion is of the highest order.

2.3.1 Lemmas for Proposition 50

For Lemma 41, we need the following notation. For a function $f(x)$ defined on the interval from (x_l, x_o), let

$$F(x) = \int_{x_l}^{x} f(t)dt, \quad \varphi(t) = \int_{x_l}^{x_o} e^{itx} f(x)dx,$$

and

$$\tilde{f_{u_n}}(x) = f(x)/F(u_n) \text{ for } x_l < x < u_n, \quad \varphi_{u_n}(t) = \int_{x_l}^{u_n} e^{itx} f_{u_n}(x)dx.$$

Lemma 41 *Let f be any integrable function, then $|\varphi_{u_n}(t)| \to 0$ uniformly in n as $|t| \to \infty$.*

PROOF:

$$|\varphi_{u_n}(t)| = |\int_{x_l}^{x_o} e^{itx} \tilde{f_{u_n}}(x)dx| < |\frac{1}{F(u_n)} \int_{x_l}^{u_n} \cos(tx)f(x)dx| + |\frac{1}{F(u_n)} \int_{x_l}^{u_n} \sin(tx)f(x)dx| \tag{2.18}$$

By the mean approximation theorem of Feller (1971), Chapter IV, Section 2, for any arbitrary integrable function f and $\epsilon > 0$ there exists a step function h such that $\int_{-\infty}^{\infty} |f(x) - h(x)|dx < \epsilon$.

The key to the uniformity of the result is that the same h can be shown to hold for all f_{u_n}.

Before we start, let us define $M = \sup_n \frac{1}{F(u_n)}$. Note $M < \infty$ since u_n is a threshold; that is, $u_n > x_l + \delta \ \forall n$ and for some $\delta > 0$.

It suffices to prove $|\frac{1}{F(u_n)} \int_{x_l}^{u_n} \cos(tx)f(x)dx| \to 0$ uniformly in n as $|t| \to \infty$ in (2.18) since the sine term will follow by the same argument.

First the assertion is easily verifiable for finite step function h; that is, letting

$$h(x) = c_i \text{ for } \zeta_{i-1} < x \leq \zeta_i, i = 1, \ldots, k \text{ with } \zeta_k = u_n$$

$$\begin{aligned}
\left|\frac{1}{F(u_n)}\int_{x_l}^{u_n}\cos(tx)h(x)dx\right| &\leq \left|M\sum_{i=1}^{k}c_i\int_{\zeta_{i-1}}^{\zeta_i}\cos(tx)dx\right| \\
&\leq M\ \frac{1}{|t|}\ \sum_{i=1}^{k}|c_i|\ |\sin(t\zeta_i)-\sin(t\zeta_{i-1})| \\
&\leq \frac{2M\sum_{i=1}^{k}|c_i|}{|t|}.
\end{aligned}$$

Thus

$$\left|\frac{1}{F(u_n)}\int_{x_l}^{u_n}\cos(tx)h(x)dx\right| \to 0 \text{ as } |t|\to\infty \text{ uniformly in } n.$$

Finally, we have

$$\begin{aligned}
\left|\frac{1}{F(u_n)}\int_{x_l}^{u_n}\cos(tx)f(x)dx\right| &= \left|\frac{1}{F(u_n)}\int_{x_l}^{u_n}\cos(tx)[f(x)-h(x)+h(x)]dx\right| \\
&\leq M\int_{-\infty}^{u_n}|\cos(tx)||f(x)-h(x)|dx \\
&\qquad +\frac{1}{F(u_n)}\left|\int_{-\infty}^{u_n}\cos(tx)h(x)dx\right| \\
&\leq M\int_{-\infty}^{\infty}|f(x)-h(x)|dx+\frac{1}{F(u_n)}\left|\int_{-\infty}^{u_n}\cos(tx)h(x)|dx\right| \\
&\leq M\epsilon+\frac{2M\sum_{i=1}^{k}|c_i|}{|t|}.
\end{aligned}$$

Now let $|t|\to\infty$ and note that ϵ is arbitrary. Thus adding the corresponding sine term,

$$|\varphi_{u_n}(t)|\to 0 \text{ uniformly in } n \text{ as } |t|\to\infty.$$

□

Lemma 42 *Let f' exist and be integrable, then*

$$\limsup_{n\to\infty} \sup_t |t|\ |\varphi_{u_n}(t)| < \infty$$

PROOF: From (2.18) in lemma 41 we see that it suffices to show

$$\limsup_{n\to\infty} \sup_t |t|\ |\int_{-\infty}^{u_n} \cos(tx)\frac{f(x)}{F(u_n)}dx| < \infty$$

since again the corresponding sine term would follow via the same argument. Now

$$\begin{aligned}
|t|\ |\int_{-\infty}^{u_n} \cos(tx)\frac{f(x)}{F(u_n)}dx| &= |t|\ |\frac{1}{t}\frac{1}{F(u_n)}\{\sin(tu_n)f(u_n) \\
&\qquad - \lim_{x\to-\infty}\sin(tx)f(x) - \int_{-\infty}^{u_n}\sin(tx)f'(x)dx\}| \\
&\qquad\qquad \text{by integration by parts} \\
&\le \frac{1}{F(u_n)}\{|\sin(tu_n)f(u_n)| \\
&\qquad +|\lim_{x\to-\infty}\sin(tx)f(x)| + |\int_{-\infty}^{u_n}\sin(tx)f'(x)dx|\} \\
&\le \frac{1}{F(u_n)}\{2\int_{-\infty}^{\infty}|f'(x)|dx + \lim_{x\to-\infty}|f(x)|\} \\
&< \infty
\end{aligned}$$

since $\frac{1}{F(u_n)} \le M < \infty$ where M is defined in Lemma 41, f' is assumed integrable, and f is a proper density.

So we can say

$$\limsup_{n\to\infty} \sup_t |t|\ |\int_{-\infty}^{u_n} \cos(tx)\frac{f(x)}{F(u_n)}dx| < \infty$$

and thus with the corresponding sine term

$$\limsup_{n\to\infty} \sup_t |t|\ |\varphi_{u_n}(t)| < \infty.$$

□

Corollary 43 *Given*

$$\limsup_{n\to\infty} \sup_t |t| \; |\varphi_{u_n}(t)| < \infty \tag{2.19}$$

then there exists an n^* *such that* $|\varphi_{u_n}(t)|^n$ *is integrable for* $n \geq n^* > 1$.

PROOF: Clearly if (2.19) holds, then there exists an n^* and a constant $c < \infty$ such that for all $n \geq n^*$,

$$\sup_t |t| \; |\varphi_{u_n}(t)| \leq c < \infty.$$

We also have

$$|\varphi_{u_n}(t)| \leq 1, everywhere$$

so

$$|\varphi_{u_n}(t)| \leq \begin{cases} \frac{c}{|t|} & \text{if } |t| \geq c \quad \forall n \geq n^* \\ 1 & \text{if } |t| \leq c. \end{cases}$$

So

$$\int_{-\infty}^{\infty} |\varphi_{u_n}(t)|^n dt \leq 2c + 2c^n \int_c^{\infty} \frac{dt}{t^n} = 2c + \frac{2c}{n-1} < \infty \quad \forall n \geq n^*.$$

Thus $|\varphi_{u_n}(t)|^n$ is integrable $\forall n \geq n^*$. □

Corollary 44 *If* $|\varphi_{u_n}(t)|^n$ *is integrable for some* $n \geq n^* > 1$, *then* $f_{\tilde{S_n}}$ *exists and* $\forall n \geq n^* > 1$ *has Fourier norm*

$$N_n = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \varphi_{u_n}^{n-1}\left(\frac{t}{\sqrt{(n-1)\sigma^2(u_n)}}\right) \exp\left(-\frac{it\mu(u_n)\sqrt{n-1}}{\sigma(u_n)}\right) \right| dt. \tag{2.20}$$

PROOF: This falls from an application of the Fourier inversion formula of Feller (1971), Chapter XV, Section 3, Theorem 3.

We see that the characteristic function of $\tilde{S_n}$ is

$$\varphi_{\tilde{S_n}}(t) = E\exp\left(it\left[\frac{\sum_{i=1}^{n-1} X_i^* - (n-1)\mu(u_n)}{\sqrt{(n-1)\sigma^2(u_n)}}\right]\right)$$

where $X_i^* = X_i | X_i \leq u_n$.

Thus

$$\varphi_{\tilde{S_n}}(t) = \exp(-\frac{it\mu(u_n)\sqrt{n-1}}{\sigma(u_n)}) \quad \varphi_{u_n}(\frac{t}{\sqrt{(n-1)\sigma^2(u_n)}})^{n-1}$$

From corollary 43, we see

$$\int_{-\infty}^{\infty} |\varphi_{u_n}(t)|^n dt < \infty \quad \forall n \geq n^* > 1.$$

so the assumptions for the Fourier inversion theorem holds and the result follows. □

Lemma 45 *For a continuous underlying distribution, given a* $\delta > 0$ *there exists a number* $q_\delta < 1$ *such that*

$$|\varphi_{u_n}(t)| < q_\delta \quad \forall |t| > \delta \text{ and } \forall n \geq n^*.$$

Proof: We begin with

$$|\varphi_{u_n}(t) - \varphi(t)| = |\frac{1}{F(u_n)}\{\varphi(t) - \int_{u_n}^{\infty} e^{itx} f(x)dx\} - \varphi(t)| \leq \frac{1-F(u_n)}{F(u_n)} + M\int_{u_n}^{\infty} f(x)dx \tag{2.21}$$

where M is defined in Lemma 41. Now (2.21) tends to 0 as $n \to \infty$. The bound is obviously uniform over all t.

From Feller(1971), Chapter XV, Section 1, Lemma 4, we have

$$|\varphi(t)| < 1 \quad \text{whenever} \quad |t| \neq 0. \tag{2.22}$$

From Feller(1971), Chapter XV, Section 1, Lemma 1(a), we have

$$\varphi(t) \text{ is continuous.} \tag{2.23}$$

From the Riemann-Lebesgue theorem Feller (1971), Chapter XV, Section 4, Lemma 3, we have

$$\exists\ \delta^* \text{ such that } |\varphi(t)| < \frac{1}{2}\ \ \forall |t| > \delta^*. \tag{2.24}$$

Now fix $\delta > 0$. Using (2.22) and (2.23), we have

$$\exists\ q_{\delta^*} \in (\frac{1}{2}, 1) \text{ such that } |\varphi(t)| < q_\delta^* \text{ on } \delta \le |t| \le \delta^*. \tag{2.25}$$

Using (2.24) and (2.25), we have

$$|\varphi(t)| < q_\delta^* \quad |t| \ge \delta.$$

Now, let $q_\delta = (q_\delta^* + 1)/2$. Then by (2.21), we can choose n^* so that

$$|\varphi_{u_n}(t) - \varphi(t)| \le q_\delta - q_\delta^* \text{ on } \forall t \quad \forall n \ge n^*.$$

Hence,

$$|\varphi_{u_n}(t)| < q_\delta \quad \forall |t| \ge \delta, \quad \forall n \ge n^*.$$

□

For the following lemma, we need to define the following notation

$$\psi_{u_n}(t) = \log \varphi_{u_n}(t) - it\mu(u_n) + \frac{t^2}{2}\sigma^2(u_n)$$

$$\psi(t) = \log \varphi(t) - it\mu + \frac{t^2}{2}\sigma^2$$

$$\mathcal{K}_3(u_n) = \mu_3(u_n) - 3E(X^2|X < u_n)\mu(u_n) + 2\mu(u_n)^3$$

$$\mathcal{K}_3 = \mu_3 - 3E(X^2)\mu + 2\mu^3$$

Lemma 46 *Assume μ_3 exists, there exists a $\delta > 0$ such that φ''' exists and is continuous for some $|t| < \delta$, and $|\varphi_{u_n}(t)|^{n^*}$ is integrable for some $n^* > 1$, then*

$$|\psi_{u_n}(t) - \frac{(it)^3 \mathcal{K}_3(u_n)}{6}| < \epsilon |t|^3, \quad \forall |t| < \delta, \forall n \geq n^*. \tag{2.26}$$

Proof: First let us look more closely at the moments, the characteristic functions, the function ψ_{u_n}, and their derivatives.

Since μ_3 exists this implies EX^2 and EX exist, and thus $\mathcal{K}_3$ also exists.

By the dominated convergence theorem, $\mu_3(u_n)$, $E(X^2|X < u_n)$, $\mu(u_n)$, and $\mathcal{K}_3(u_n)$ exist and as $n \to \infty$

$$\mu_3(u_n) \to \mu_3, \quad E(X^2|X < u_n) \to E(X^2), \quad \mu(u_n) \to \mu \text{ and } \mathcal{K}_3(u_n) \to \mathcal{K}_3.$$

Next since φ''' exists and is continuous, then φ'' and φ' exist and are continuous.

Since $|\frac{(iX)^m e^{itx} 1_{(-\infty, u_n)}}{F(u_n)}| < M|X|^m$ for each m and $\int_{-\infty}^{\infty} |x|^m f(x)dx < \infty$ for each $m = 0, 1, 2, 3$ again by the dominated convergence theorem we have ${\varphi_{u_n}}'''$, ${\varphi_{u_n}}''$, and ${\varphi_{u_n}}'$ exist and as $n \to \infty$

$${\varphi_{u_n}}''' \to \varphi''', \quad {\varphi_{u_n}}'' \to \varphi'', \text{and } {\varphi_{u_n}}' \to \varphi'.$$

By usual characteristic function properties,

$$\varphi_{u_n}(0) = 1, \;\; {\varphi_{u_n}}'(0) = i\mu(u_n), \; {\varphi_{u_n}}''(0) = i^2 E(X^2|X < u_n), \; {\varphi_{u_n}}'''(0) = i^3 E(X^3|X < u_n)$$

Using the definition of ψ_{u_n} we have

$${\psi_{u_n}}'''(t) = \frac{{\varphi_{u_n}}'''(t)}{\varphi_{u_n}(t)} - 3\frac{{\varphi_{u_n}}''(t){\varphi_{u_n}}'(t)}{{\varphi_{u_n}}^2(t)} + 2(\frac{{\varphi_{u_n}}'(t)}{\varphi_{u_n}(t)})^3 \tag{2.27}$$

We also have

$$\psi_{u_n}(0) = {\psi_{u_n}}'(0) = {\psi_{u_n}}''(0) = 0 \text{ with } {\psi_{u_n}}'''(0) = i^3\mathcal{K}_3(u_n). \tag{2.28}$$

Now equation 2.26 is solved by looking at the three term Taylor expansion of ψ_{u_n} and putting a bound on the remainder term.

We have by the Cauchy form of Taylor's expansion (see Johnson and Kotz (1982), Vol. 9, p. 187) of $\psi_{u_n}(t)$ about 0, for each t and n

$$\psi_{u_n}(t) = \sum_{j=0}^{2} \frac{{\psi_{u_n}}^{(j)}(0)t^j}{j!} + \frac{t^3}{3!}{\psi_{u_n}}'''(\theta t) \;\text{ for some } \theta \in (0,1)$$

where ${\psi_{u_n}}^{(i)}$ is the ith derivative of ψ_{u_n}.

Using (2.28), we see

$$|\psi_{u_n}(t) - \frac{(it)^3\mathcal{K}_3(u_n)}{6}| = |\frac{t^3}{6}\{{\psi_{u_n}}'''(\theta) - {\psi_{u_n}}'''(0)\}|.$$

So proving equation (2.26) is equivalent to proving there exists a $\delta > 0$ and $n^* > 1$ such that

$$|{\psi_{u_n}}'''(t) - {\psi_{u_n}}'''(0)| < \epsilon \quad \forall |t| < \delta \text{ and } \forall n \geq n^*. \tag{2.29}$$

To prove (2.29) we look at (2.27). Since $\lim_{t\to 0}\varphi_{u_n}(t) \neq 0$, it suffices to prove for each $m = 0, 1, 2, 3$ that given $\epsilon > 0$ there exists a $\delta > 0$ (and $n^* > 1$) such that

$$|{\varphi_{u_n}}^{(m)}(t) - {\varphi_{u_n}}^{(m)}(0)| < \epsilon \;\; \forall |t| < \delta \text{ and } \forall n \geq n^*. \tag{2.30}$$

We break (2.30) up into the following parts:

$$|{\varphi_{u_n}}^{(m)}(t) - {\varphi_{u_n}}^{(m)}(0)| \leq |{\varphi_{u_n}}^{(m)}(t) - \varphi^{(m)}(t)| + |\varphi^{(m)}(t) - \varphi^{(m)}(0)| + |{\varphi_{u_n}}^{(m)}(0) - \varphi^{(m)}(0)|. \tag{2.31}$$

For the first term in (2.31), for each fixed t

$$\begin{aligned} |\varphi_{u_n}{}^{(m)}(t) - \varphi^{(m)}(t)| &= \left|\frac{\varphi^{(m)}(t) - \int_{u_n}^{\infty}(ix)^m e^{itx} f(x)dx}{F(u_n)} - \varphi^{(m)}(t)\right| \\ &\leq |\varphi^{(m)}(t)| \left(\frac{1 - F(u_n)}{F(u_n)}\right) + \frac{1}{F(u_n)} \left|\int_{u_n}^{\infty}(ix)^m e^{itx} f(x)dx\right| \\ &\leq M\{1 - F(u_n)\} + M \int_{-\infty}^{\infty} 1_{(u_n,\infty)}|x|^m f(x)dx. \end{aligned}$$

Since $1-F(u_n) \to 0$ as $n \to \infty$, there exists a n' such that $\forall n \geq n'$, $M\{1-F(u_n)\} < \epsilon/6$.

Since $\int_{-\infty}^{\infty} |x|^m dF < \infty$ for $m = 0, 1, 2, 3$ and $u_n \to \infty$, there exists (for each m) a n'' such that $\forall n \geq n''$ $M \int_{-\infty}^{\infty} 1_{(u_n,\infty)}|x|^m f(x)dx < \epsilon/6$.

Thus there exists a $n^* = max(n', n'')$ such that $\forall n \geq n^*$

$$|\varphi_{u_n}{}^{(m)}(t) - \varphi^{(m)}(t)| < \epsilon/3 \quad \forall t. \tag{2.32}$$

For the second term in (2.31), recall $\varphi^{(m)}$ is continuous in a neighborhood of 0 for each m. Thus for each m, given an $\epsilon > 0$ there exists a $\delta > 0$ such that

$$|\varphi^{(m)}(t) - \varphi^{(m)}(0)| < \epsilon/3 \;\; \forall |t| < \delta. \tag{2.33}$$

Finally, for the third term in (2.31), recall $\varphi_{u_n}{}^{(m)}(0) \to \varphi^{(m)}(0)$ as $n \to \infty$ by dominated convergence theorem. Thus for each m, there exists a n^* such that

$$|\varphi_{u_n}{}^{(m)}(0) - \varphi^{(m)}(0)| < \epsilon/3 \;\; \forall n \geq n^* \;\; \forall t. \tag{2.34}$$

Putting (2.32), (2.33), and (2.34) together, define δ as smallest necessary in (2.33) and n^* as large as necessary in (2.32) and (2.34). Then (2.31) holds, thus we have (2.30) which is sufficient to prove (2.29). □

2.3.2 Lemmas for Proposition 51

Lemma 47 *Let F be in the domain of attraction of Λ so that the representation in (2.3) holds with $c(x) = 1$. Recall a_n and b_n are the appropriate normalizing constants for the Gumbel distribution defined in (2.4) and (2.5). Here $u_n = a_n v + b_n$. Also assume that*

$$\frac{\phi'(a_n v + b_n)}{\phi'(b_n)} \to 1, \quad \textit{uniformly on } |v| \leq e_n = -K \log|\phi'(b_n)|$$

then

$$\frac{-\log F(u_n)}{-\log F(b_n)} = e^{-v}[1 + O(v^2|\phi'(b_n)|)], \quad \textit{uniformly on } |v| \leq e_n \tag{2.35}$$

Remarks We assume that $c(x) \equiv 1$ in (2.3) – see comments in Proposition 51 section. The conditions of Lemma 47 are not as strong as those in Proposition 51. In other words, the results for Lemma 47 immediately apply in Proposition 51.

Proof: Here we mimic the proof for Proposition 9.2 of Smith (1987). In fact, Proposition 9.2 of Smith (1987) gives the results for the threshold case.

From (2.3),

$$\frac{-\log F(u_n)}{-\log F(b_n)} = \exp\{-\int_0^v \frac{\phi(b_n)}{\phi(a_n t + b_n)} dt\}. \tag{2.36}$$

But

$$\int_0^v \frac{\phi(b_n)}{\phi(a_n t + b_n)} dt = \int_0^v \frac{\phi(b_n)}{\phi(b_n) + a_n t\theta} dt$$

where $\theta = \phi'(a_n s + b_n)$ for some s between 0 and v. Recall that $\phi(b_n) = a_n$. Thus (2.36) is equal to

$$\exp\{-\frac{1}{\theta}\log(1 + \theta v)\}. \tag{2.37}$$

which using a Taylor's expansion of $\log(1 + x)$ (2.37) is equal to

$$\exp\{-(v + O(v^2\theta)\}. \tag{2.38}$$

From (2.61) we have $\frac{\phi'(a_n s+b_n)}{\phi'(b_n)} \to 1$ uniformly on $|v| \le e_n = -K\log|\phi'(b_n)|$ so that we may substitute $\phi'(b_n)$ for θ in (2.38). Now using $e^{-v+r} = e^{-v}e^{r}$ and a Taylor expansion for $e^r = 1 + r + o(r)$ when $r \to 0$, we conclude

$$\frac{-\log F(u_n)}{-\log F(b_n)} = e^{-v}[1 + O(v^2|\phi'(b_n)|)], \quad \text{uniformly on } |v| \le e_n. \tag{2.39}$$

□

Lemma 48 *Let $m \ge 1$ be a finite constant, $\delta > 0$ be an arbitrary finite constant, and κ be a finite constant. The function $h(v) = e^{-e^{-v}}e^{-mv}$ is uniformly bounded $\forall v$. In fact,*

$$h(v) \le \min(e^{-m}m^m, \kappa|v|^{-\delta}), \quad \forall v \tag{2.40}$$

Proof: Now, $h(v)$ is a continuous function on $-\infty < v < \infty$. We have $\sup_v h(v) = e^{-m}m^m$ at $v = -\log m$. Its inflection points are at $v = LIP = -\log\{(2m+1) + \sqrt{4m+1}\}$ and $v = UIP = -\log\{(2m+1)1\sqrt{4m+1}\}$. Note these inflection points (LIP and UIP) are finite for finite m.

Now we have for any $\delta > 0$

For $LIP < v < UIP$	$\lvert h(v)\rvert < e^{-m}m^m$	since h reaches its max. in this interval.
For $v > UIP$	$\lvert h(v)\rvert < \kappa\lvert v\rvert^{-\delta}$	since exp. term in $h(v)$ will dominate.
For $v < LIP$	$\lvert h(v)\rvert < \kappa\lvert v\rvert^{-\delta}$	since double exp. in $h(v)$ will dominate.

where κ is a constant.

Thus we have our result. □

2.3.3 Lemmas for Proposition 53

Lemma 49 *Let F be in the domain of attraction of Λ so that representation in (2.3) holds with $c(x) = 1$, then as $u \to x_o$*

$$\phi(u)/u \to 0, \quad x_o = \infty \tag{2.41}$$

or

$$\phi(u) \to 0, \quad x_o < \infty \tag{2.42}$$

PROOF: If $x_o < \infty$, then we can write $-\log F(x_o) = 0$ so in the Balkema and DeHaan (1972) representation [(2.3) with $c(x) = 1$], we have

$$\exp\{-\int_0^{x_o} \frac{dt}{\phi(t)}\} = 0$$

Assume $\phi(t) \not\to 0$. Since ϕ is continuous, if $\phi(x_o) \neq 0$, then $\phi(t) \to c$ where $c \neq 0$ is a constant (as $t \to x_o$). In other words, given an $\epsilon > 0$ there exists a t_c such that for some $t > t_c$

$$c - \epsilon < |\phi(t)| < c + \epsilon.$$

Thus we would have

$$\exp\{-\int_0^{x_o} \frac{dt}{\phi(t)}\} > \exp\{-\int_0^{t_c} \frac{dt}{\phi(t)}\} + \exp\{-\int_{t_c}^{x_o} \frac{dt}{c-\epsilon}\} > \exp\{-\frac{x_o - t_c}{c - \epsilon}\}.$$

Now this last term is a positive constant; i.e., it does not equal to 0. Thus we have a contradiction so

$$\phi(u) \to 0, \quad x_o < \infty.$$

Now for the case when $x_o = \infty$. In the Balkema and DeHaan (1972) representation (2.3) with $c(x) = 1$, we have $\phi'(u) \to 0$ as $u \to 0$. This implies that given a $\delta > 0$ there exists a u_δ such that

$$|\phi'(u)| < \frac{\delta}{2}, \ \forall u > u_\delta \tag{2.43}$$

Now assume $\frac{\phi(u)}{u} \not\to 0$. This implies that exists an infinite sequence of u such that along this sequence

$$\phi(u) > \kappa_1 u \quad \text{for some } \kappa_1 > 0. \tag{2.44}$$

Note by the Fundamental Theorem of Calculus and using (2.43) we have

$$\begin{aligned}\phi(u) &= \phi(u_\delta) + \int_{u_\delta}^{u} \phi'(t)dt \\ &< \phi(u_\delta) + \frac{\delta}{2}\{u - u_\delta\} \\ &\leq \kappa_2 + \frac{\delta}{2}u \end{aligned} \tag{2.45}$$

Note in the last line we are using that both u_δ and $\phi(u_\delta)$ are some finite constants.

Now combining (2.44) and (2.45) we have that

$$u < \frac{\kappa_2}{\kappa_1 - \frac{\delta}{2}} = \kappa_3$$

where κ_3 is some finite constant. But we have $u \to x_o = \infty$. Thus we have a contradiction so we must have

$$\frac{\phi(u)}{u} \to 0, \quad x_o = \infty$$

□

2.4 Propositions

Here we present the three main propositions of the chapter. They contain the fundamental pieces necessary for the formulation of the main theorem. They are as follows: Proposition 50 gives the expansion of $f_{\tilde{S}_n}$; Proposition 51 gives the expansion of $f_{M_n^*}$; and Proposition 53 gives the expansions of the conditional mean and variance to substitute into Proposition 50.

2.4.1 Expansion of the Conditional Density of Sum Given the Maximum

In establishing the expansion for the density of $\tilde{S}_n$, we base the following proof on the argument in Feller (1971), Chapter XVI, Section 2, see p. 533. Deriving the expansion for $f_{\tilde{S_n}}$ involves conditioning on $X \leq u_n$. Recall the random variable X^* which has distribution function $\tilde{F_{u_n}}(x) = F(x)/F(u_n)$, density $\tilde{f_{u_n}}(x) = f(x)/F(u_n)$, and characteristic function $\varphi_{u_n}(t) = \int_{-\infty}^{u_n} e^{itx}\tilde{f_{u_n}}(x)dx$. Its mean and variance are $\mu(u_n)$ and $\sigma^2(u_n)$, with $\mu_3(u_n)$ as its third moment and $\mathcal{K}_3(u_n)$ as third cumulant. Recall $\mathcal{N}'$ denotes the standard normal density.

Proposition 50 *Assume f' is integrable, μ_3 exists, φ''' exists and is continuous in a neighborhood of 0, and $|\varphi_{u_n}(t)|^n$ is integrable for some $n \geq n^* > 1$, then $f_{\tilde{S_n}}$ exists for $n \geq n^*$ and as $n \to \infty$*

$$f_{\tilde{S_n}}(x) - \mathcal{N}'(x) - \frac{\mathcal{K}_3(u_n)}{6\sigma^3(u_n)\sqrt{n}}(x^3 - 3x)\mathcal{N}'(x) = o(\frac{1}{\sqrt{n}}) \quad \textit{uniformly in } x \textit{ and } v. \tag{2.46}$$

Proof: By Corollary 44, the left hand side of (2.46) exists for $n \geq n^*$ and has Fourier norm

$$\begin{aligned} N_n &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\Big|\exp(-\frac{it\mu(u_n)\sqrt{n-1}}{\sigma(u_n)}\varphi_{u_n}(\frac{t}{\sqrt{n-1}\sigma(u_n)})^{n-1} \\ &\quad - \exp(-\frac{t^2}{2}) - \frac{\mathcal{K}_3(u_n)}{6\sigma^3(u_n)\sqrt{n-1}}(it)^3\exp(-\frac{t^2}{2})\Big|dt. \end{aligned} \tag{2.47}$$

Let N_L be N_n where the integral is restricted to the interval $|t| \leq \delta\sigma(u_n)\sqrt{n-1}$ and N_G be N_n where the integral is restricted to the intervals $|t| > \delta\sigma(u_n)\sqrt{n-1}$.

Now, choose $\delta > 0$ arbitrary but fixed. By Lemma 45 there exists a number $q_\delta < 1$ and a n^{**} such that

$$|\varphi_{u_n}(t)| < q_\delta \quad \forall|t| > \delta \text{ and } \forall n \geq n^{**}. \tag{2.48}$$

If we replace t by $\frac{t}{\sqrt{n-1}\sigma(u_n)}$ in (2.48), we have that

$$|\varphi_{u_n}(\frac{t}{\sqrt{n-1}\sigma(u_n)})| < q_\delta \quad \forall|t| > \delta\sigma(u_n)\sqrt{n-1} \text{ and } \forall n \geq n^{**}.$$

Thus the contribution of the intervals $|t| > \delta\sigma(u_n)\sqrt{n-1} \ \ \forall n \geq n' = max(n^*, n^{**})$ in (2.47) is then

$$N_G < \tag{2.49}$$
$$q_\delta^{n-1-n'} \int_{-\infty}^{\infty} |\varphi_{u_n}(\frac{t}{\sqrt{n-1}\sigma(u_n)})|^{n'} dt + \int_{|t|>\delta\sigma(u_n)\sqrt{n-1}} \exp(-\frac{t^2}{2})(1 + |\frac{\mathcal{K}_3(u_n)t^3}{6\sigma^3(u_n)\sqrt{n-1}}|)dt.$$

Since $\int_{-\infty}^{\infty} |\varphi_{u_n}(\frac{t}{\sqrt{n-1}\sigma(u_n)})|^n dt < \infty$ for all $n \geq n' = max(n^*, n^{**})$ and q_δ does not dependent on n, we have the first term of (2.50) tends to zero more rapidly than any power of $1/n$. The same holds for the second term and can be seen by substituting into the inequality $\underline{\sigma(u_n)} = \inf \sigma(u_n)$ for $\sigma(u_n)$ – i.e. $N_G = o(1/\sqrt{n})$ uniformly in n.

Now for the other interval $|t| < \delta\sigma(u_n)\sqrt{n-1}$ we use the formula

$$\psi_{u_n} = \log \varphi_{u_n}(t) - it\mu(u_n) + \frac{t^2}{2}\sigma^2(u_n).$$

We have

$$N_L = \frac{1}{2\pi} \int_{|t|<\delta\sigma(u_n)\sqrt{n-1}} e^{-t^2/2} \left| \exp[(n-1)\psi_{u_n}(\frac{t}{\sqrt{n-1}\sigma(u_n)})] - 1 - \frac{\mathcal{K}_3(u_n)}{6\sigma^3(u_n)\sqrt{n-1}}(it)^3 \right| dt. \tag{2.50}$$

The integral will be evaluated using the following inequality from Feller (1971), equation 2.9 on p. 534.

$$|e^\alpha - 1 - \beta| \leq (|\alpha - \beta| + \frac{1}{2}\beta^2)e^\gamma \text{ where } \ \gamma = \max(|\alpha|, |\beta|) \tag{2.51}$$

Let

$$\alpha = (n-1)\psi_{u_n}(\frac{t}{\sqrt{n-1}\sigma(u_n)})$$

and

$$\beta = \frac{\mathcal{K}_3(u_n)}{6\sigma^3(u_n)\sqrt{n-1}}(it)^3 = (n-1)\frac{\mathcal{K}_3(u_n)}{6\sigma^3(u_n)(\sqrt{n-1})^3}(it)^3.$$

Thus

$$N_L = \frac{1}{2\pi}\int_{|t|<\delta\sigma(u_n)\sqrt{n-1}} e^{-t^2/2}|(e^{\alpha}-1-\beta)|dt$$
$$\leq \frac{1}{2\pi}\int_{|t|<\delta\sigma(u_n)\sqrt{n-1}} e^{-t^2/2}|e^{\gamma}(|\alpha-\beta|+\frac{\beta^2}{2})|dt. \quad (2.52)$$

Now to solve for $|\alpha-\beta|$ we utilize Lemma 46. By substituting $\frac{t}{\sqrt{n-1}\sigma(u_n)}$ for t, we have

$$|\alpha-\beta| = (n-1)|\psi_{u_n}(\frac{t}{\sqrt{n-1}\sigma(u_n)}) - \frac{1}{6}\mathcal{K}_3(u_n)(\frac{it}{\sigma(u_n)(\sqrt{n-1})})^3|$$
$$\leq (n-1)|\frac{t}{\sigma(u_n)\sqrt{n-1}}|^3\epsilon$$
$$\leq \frac{\epsilon|t|^3}{\sqrt{n-1}\sigma(u_n)}. \quad (2.53)$$

For $\frac{1}{2}\beta^2$ we have

$$\frac{1}{2}\beta^2 = \frac{1}{2}[\frac{\mathcal{K}_3(u_n)}{6\sigma^3(u_n)\sqrt{n-1}}(it)^3]^2$$
$$= \frac{t^6}{72(n-1)}\frac{\mathcal{K}_3^2(u_n)}{\sigma^6(u_n)}. \quad (2.54)$$

Finally we note that $\varphi_{u_n}(t)\to\varphi(t)$, $\mu(u_n)\to\mu$, $\sigma^2(u_n)\to\sigma^2$, $\mu_3(u_n)\to\mu_3$, and $\mathcal{K}_3(u_n)\to\mathcal{K}_3$ by dominated convergence theorem. Thus $\psi_{u_n}(t)\to\psi(t)$ as $n\to\infty$ uniformly in t. Then we have $\lim_{t\to 0}\psi_{u_n}(t)=0$ uniformly in n by the dominated convergence theorem.

Now let K_1 be a constant, then we can write

$$|\beta| \leq \frac{K_1|t|^3}{\sqrt{n-1}\sigma^3(u_n)} \leq \frac{K_1|t|^2}{\sqrt{n-1}\sigma^3(u_n)}\delta\sigma(u_n)\sqrt{n-1}$$
$$\leq \frac{K_1\delta|t|^2}{\sigma^2(u_n)}$$
$$< \frac{|t|^2}{4} \text{ if we choose } \delta \text{ so that } \frac{K_1\delta}{\sigma(u_n)} < \frac{1}{4}\ \ \forall n > n^*. \quad (2.55)$$

Similarly by Lemma 46 and defining K_2 as another constant,

$$\begin{aligned}|\alpha| &< K_2(n-1)|\frac{t}{\sqrt{n-1}\sigma(u_n)}|^3 \leq \frac{K_2 t^2}{\sqrt{n-1}\sigma^3(u_n)}\delta\sigma(u_n)\sqrt{n-1} \\ &\leq \frac{K_2\delta|t|^2}{\sigma^2(u_n)} \\ &< \frac{|t|^2}{4} \text{ if we choose } \delta \text{ so that } \tfrac{K_2\delta}{\sigma(u_n)} < \tfrac{1}{4} \ \ \forall n > n^*. \end{aligned} \tag{2.56}$$

Using (2.55) and (2.56), we have

$$\gamma < \frac{t^2}{4}. \tag{2.57}$$

By substituting in (2.53), (2.54), and (2.57), we have the integrand in (2.52),

$$< e^{\frac{-t^2}{4}}[\epsilon\frac{|t|^3}{\sqrt{n-1}\sigma(u_n)} + \frac{t^6}{72(n-1)}\frac{\mathcal{K}_3^2(u_n)}{\sigma^6(u_n)}]$$

Since ϵ is arbitrary and independent of x, $\sigma^2(u_n) \to \sigma^2$ and $\mathcal{K}_3(u_n) \to \mathcal{K}_3$ where σ^2 and $\mathcal{K}_3$ are assumed finite, and $\int_{-\infty}^{\infty} t^6 e^{-t^2} dt < \infty$, then $N_L = o(1/\sqrt{n})$ uniformly in n. Thus (2.46) holds. □

2.4.2 Expansion of the Density of the Maximum

Proposition 51 involves deriving a higher order expansion for the density of M_n, the maximum of a sequence, when the underlying distribution of the observations, F, lies in the domain of attraction of the Gumbel distribution, Λ. Recall the density of M_n^* as $f_{M_n^*}(v) = dF_{M_n^*}(v)/dv$ where $F_{M_n^*}(v) = F^n(a_n v + b_n)$ and $\Lambda'(v)$as the Gumbel density. The constants a_n and b_n are the appropriate normalizing constants for the Gumbel distribution and can be taken to be $a_n = \phi(b_n)$ where ϕ is defined in (2.3) and $b_n = \inf\{x : -\log\ F(x) > 1/n\}$.

Note that the Balkema and DeHaan (1972) representation in (2.3) does not uniquely determine the functions c and ϕ. A change of order $o(1)$ to ϕ would affect c but not the overall representation. Any assumptions about the function c would presuppose smoothing conditions on F.

REMARK For this chapter, we assume that $c(x) \equiv 1$. This, of course, presupposes smoothing conditions on F. In fact, $c \equiv 1$ is equivalent to the twice differentiable domain of attraction of Pickands (1986) which is itself equivalent to the von Mises conditions – conditions known to be sufficient yet not necessary for the general domain of attraction problem. In Theorem 5.2 of Pickands (1986), the conditions for F to be in the twice differentiable domain of attraction of an extreme value distribution are that F is twice differentiable and

$$d\{(1 - F(t))/f(t)\}/dt \to c^* \quad \text{as } t \to x_o \text{ for some constant } c^*. \tag{2.58}$$

Now, Pickands (1986) shows that for the Gumbel case, c^* in (2.58) is equal to 0. Thus it is easily shown that the condition (2.58), in the Gumbel case, is equivalent to

$$\frac{f'(t)\{1 - F(t)\}}{f^2(t)} \to -1 \quad \text{as } t \to x_o \tag{2.59}$$

which is just, again, the von Mises condition, see Leadbetter *et al.* (1983), Thm 1.6.1. Now most well-behaved distributions in the Gumbel domain of attraction which have differentiable densities also satisfy the von Mises conditions so in this case the assumption $c(x) \equiv 1$ is justified.

Proposition 51 *Suppose $F \in \mathcal{D}(\Lambda)$ such that that the representation in (2.3) holds with the definitions of a_n and b_n from (2.4) and (2.5) and,*

Assume for a constant $K > 2$,

$$a_n v + b_n \to x_o \quad \textit{uniformly in } |v| \le -K \log\ |\phi'(b_n)| \textit{ as } n \to \infty. \tag{2.60}$$

$$\phi''(a_n v + b_n)/\phi''(b_n) \to 1 \quad \textit{uniformly in } |v| \le -K \log\ |\phi'(b_n)| \textit{ as } n \to \infty. \tag{2.61}$$

$$\phi(b_n)\phi''(b_n) \log\ |\phi'(b_n)|/\phi'(b_n) \to 0 \quad \textit{as } n \to \infty. \tag{2.62}$$

Suppose also that for any $v^* < x_o$

$$\inf_{v \le v^*} \frac{f'(v)F(v)}{f^2(v)} > -\infty. \tag{2.63}$$

Then for each $\delta > 0$ *there exists an* n^* *and a function* ϵ_{n^*} *tending to 0 as* $n \to \infty$ *such that* $\forall n > n^*$ *and* $\forall v$

$$\begin{aligned} |f_{M_n^*}(v) - \Lambda'(v)\{1 + (\frac{v^2}{2} - v - \frac{v^2 e^{-v}}{2})\phi'(b_n) \\ +(\frac{v^2}{2} + \frac{v^3}{6}[e^{-v} - 1])(2\phi'(b_n)^2 - \phi(b_n)\phi''(b_n)) \\ +(\frac{v^3}{2}[e^{-v} - 2] + \frac{1}{8}v^4[e^{-2v} - 3e^{-v} + 1])\phi'(b_n)^2\}| \\ < \epsilon_{n^*}[\phi'(b_n)^2 + |\phi(b_n)\phi''(b_n)| \min(1, |v|^{-\delta})]. \end{aligned} \tag{2.64}$$

REMARK Condition (2.60) is part of the definition of Cohen's (1982b) class N, see equation 1.24 of Cohen (1982b). Conditions (2.61) and (2.62) are equations (1.59) and (1.60) of Cohen's (1982b) Theorem 9 which list sufficient conditions for his class N. In Cohen (1982b), he lists in his Table 1 many distributions which belong to this class, for example, the normal and lognormal. In other words, there exist a deep pool of distributions satisfying the conditions in Proposition 51.

PROOF: Since F is assumed continuous, there exists a b_n such that $-\log\ F(b_n) = 1/n$, so

$$F_{M_n^*}(v) = F^n(a_n v + b_n) = \exp\{-\frac{-\log\ F(a_n v + b_n)}{-\log\ F(b_n)}\}.$$

Then using (2.3) with $c(v) = 1$,

$$\begin{aligned} F_{M_n^*}(v) &= \exp\{-\exp(-\int_{b_n}^{a_n v + b_n} \frac{dt}{\phi(t)})\} \\ &= \exp\{-\exp(-\int_0^v \frac{\phi(b_n)}{\phi(a_n t + b_n)} dt)\} \end{aligned}$$

$\forall a_n v + b_n < x_o$ where a_n and b_n are defined as above.

Thus

$$f_{M_n^*}(v) = \frac{dF_{M_n^*}(v)}{dv} = \exp\{-\exp(-\int_0^v \frac{\phi(b_n)}{\phi(a_n t + b_n)}dt)\}$$
$$\times \exp(-\int_0^v \frac{\phi(b_n)}{\phi(a_n t + b_n)}dt) \times \frac{\phi(b_n)}{\phi(a_n v + b_n)}$$

or, say,

$$f_{M_n^*}(v) = f_1(v) \times f_2(v) \times f_3(v) \quad \forall v < x_o$$

First, we restrict our attention to the interval $|v| \leq -K \log\ |\phi'(b_n)|$.

For the expansion of $f_2(v)$, in a similar argument that led to equation (9.20) of Smith (1987), we have

$$\begin{aligned} f_2(v) &= \exp\{-\int_0^v \frac{\phi(b_n)}{\phi(a_n t + b_n)}dt\} \\ &= e^{-v}(1 + \frac{v^2}{2}\phi'(b_n) - \frac{v^3}{6}\{2\phi'(b_n)^2 - \phi(b_n)\phi''(b_n)\} + \frac{v^4}{8}\phi'(b_n)^2 \\ &\qquad + o((1+v^4)R(b_n))) \end{aligned} \tag{2.65}$$

where $R(b_n) = \{\phi'(b_n)\}^2 + |\phi(b_n)\phi''(b_n)|$, uniformly on $0 < v < -K \log\ |\phi'(b_n)|$.

Note that the argument in Smith (1987) primarily relies on a Taylor expansion of $\phi(a_n t + b_n)$. Although here we use the Balkema and deHaan (1972) representation for $-\log F(x)$ as opposed to $1 - F(x)$ which Smith (1987) used, the Taylor expansion argument is the same; that is, the exact form of ϕ may be different but not the form of the Taylor expansion. Now all the steps of Smith's (1987) derivations of equation (9.20) apply also in the case $K \log|\varphi'(b_n)| < v < 0$ under the stronger assumptions (2.60) - (2.61). So (2.65) holds uniformly for $|v| < -K \log\ |\phi'(b_n)|$.

The expansion of f_3, embedded into the expansion of f_2, falls from the condition (2.61) which implies

$$\phi'(a_n v + b_n)/\phi'(b_n) \to 1 \quad \text{and} \tag{2.66}$$

$$\phi(a_n v + b_n)/\phi(b_n) \to 1 \quad \text{as} \quad n \to \infty \tag{2.67}$$

each uniformly over $|v| \leq -K \log\ |\phi'(b_n)|$.

In fact, using (2.60),(2.61), (2.66), and (2.67), completing the Taylor's expansion

$$\begin{aligned} f_3(v) &= \frac{\phi(b_n)}{\phi(a_n v + b_n)} \\ &= 1 - v\phi'(b_n) + \frac{v^2}{2}\{2\phi'(b_n)^2 - \phi(b_n)\phi''(b_n)\} - \frac{v^3}{2}\phi'(b_n)^2 + o((1+v^4)R(b_n)) \end{aligned}$$

uniformly over $|v| \leq -K \log\ |\phi'(b_n)|$.

The expansion of $f_1(v) = \exp\ \{-\exp\ (\ -\int_0^v \frac{\phi(b_n)}{\phi(a_n t + b_n)} dt)\}$ is as follows. First using (2.61), (2.66), and (2.67) in the argument which led to the Smith (1987) equation above equation 9.19, we have

$$\int_0^v \frac{\phi(b_n)}{\phi(a_n t + b_n)} dt = v - \frac{v^2}{2}\phi'(b_n) + \frac{v^3}{6}\{2\phi'(b_n)^2 - \phi(b_n)\phi''(b_n)\} + o((1+|v|^3)R(b_n)))$$

uniformly over $|v| \leq -K \log\ |\phi'(b_n)|$.

Now writing,

$$\int_0^v \frac{\phi(b_n)}{\phi(a_n t + b_n)} dt = v + T_n$$

where

$$T_n = T_n(v) = -\frac{v^2}{2}\phi'(b_n) + \frac{v^3}{6}\{2\phi'(b_n)^2 - \phi(b_n)\phi''(b_n)\} + o((1+|v|^3)R(b_n)).$$

Then $f_1(v) = \exp(\ -\exp(\ -v + T_n)))$.

Expanding f_1 about v,

$$\begin{aligned} f_1(v) &= e^{-e^{-v}} + T_n e^{-v} e^{-e^{-v}} + \frac{T_n^2}{2} e^{-e^{-v}}(e^{-2v} - e^{-v}) \\ &\quad + \frac{T_n^3}{6}(1 + o(1)) e^{-e^{-v}}(e^{-3v} - 3e^{-2v} + e^{-v}) \\ &= e^{-e^{-v}}\{1 - e^{-v}\frac{v^2}{2}\phi'(b_n) + \frac{v^3 e^{-v}}{6}[2\phi'(b_n)^2 - \phi(b_n)\phi''(b_n)] \\ &\quad + \frac{v^4(e^{-2v} - e^{-v})}{8}\phi'(b_n)^2 + o((1+e^{-2v})(1+v^4)R(b_n))\}. \end{aligned}$$

Thus multiplying across $f_{M_n^*}(v) = f_1(v) \times f_2(v) \times f_3(v)$ we have

$$f_{M_n^*}(v) = e^{-v}e^{-e^{-v}}\{1 + (\frac{v^2}{2} - v - \frac{v^2e^{-v}}{2})\phi'(b_n) + (\frac{v^2}{2}$$
$$+\frac{v^3}{6}[e^{-v} - 1])(2\phi'(b_n)^2 - \phi(b_n)\phi''(b_n)) + (\frac{v^3}{2}[e^{-v} - 2]$$
$$\frac{v^4}{8}[e^{-2v} - 3e^{-v} + 1])\phi'(b_n)^2 + o((1 + e^{-2v})(1 + v^4)R(b_n))\} \tag{2.68}$$
$$\text{for } |v| \leq -K \log\ |\phi'(b_n)|\}$$

Now to write this in the following form we need to bound the last term. To rewrite (2.69) we use Lemma 48, specifically (2.40) with $m = 3$. If we absorb the constants $e^{-3}3^3$ and κ from (2.40) into the following ϵ_{n^*} function, we get for $|v| \leq -K \log\ |\phi'(b_n)|$

$$|f_{M_n^*}(v) - \Lambda'(v)\{1 + (\frac{v^2}{2} - v - \frac{v^2e^{-v}}{2})\phi'(b_n)$$
$$(\frac{v^2}{2} + \frac{v^3}{6}[e^{-v} - 1])(2\phi'(b_n)^2 - \phi(b_n)\phi''(b_n))$$
$$+(\frac{v^3}{2}[e^{-v} - 2] + \frac{v^4}{8}[e^{-2v} - 3e^{-v} + 1])\phi'(b_n)^2\}|$$
$$< \epsilon_{n^*}R(b_n)\min(1, |v|^{-\delta})$$

for some $\delta > 0$ and where the function ϵ_{n^*} tends to 0 as $n^* \to \infty$.

To extend this expansion to the intervals $|v| > -K \log\ |\phi(b_n)|$, we need to show that all the terms in (2.64) are $o(|v|^{-\delta}\phi'(b_n)^2)$ for $|v| \geq e_n$ where $e_n = -K \log\ |\phi(b_n)|$.

Specifically, we will first show

$$|\Lambda'(v)| = o(|v|^{-\delta}\phi'(b_n)^2), \quad |v| \geq e_n. \tag{2.69}$$

Note the higher order terms associated with Λ' in (2.64) are of smaller order and thus will follow.

Second, we need to show

$$f_{M_n^*}(v) = o(|v|^{-\delta}\phi'(b_n)^2), \quad |v| \geq e_n. \tag{2.70}$$

We begin with Λ'. Note that on the interval $|v| \geq e_n$ $\Lambda'(v)$ is maximized for sufficiently large n at $-e_n$ and e_n. Thus we only need to verify (2.69) at these two points.

First, for $v = e_n$, since $K > 2$ we can fix K' such that $1 < K' < K/2$. Thus (for $v > 0$)

$$|\Lambda'(v)| \leq e^{-v} = e^{-\frac{v}{K'}} e^{-v(1-\frac{1}{K'})}$$

But since exponential rates dominate polynomial, we have

$$e^{-v(1-\frac{1}{K'})} \leq \kappa_1 |v|^{-\delta} \text{ for some constant } \kappa_1. \tag{2.71}$$

Now substitute in $v = e_n = -K \log |\phi'(b_n)|$,

$$e^{-\frac{v}{K'}} \leq e^{\frac{K}{K'} \log |\phi'(b_n)|} = |\phi'|^{\frac{K}{K'}} < |\phi'(b_n)|^2 \text{ since } \frac{K}{K'} > 2. \tag{2.72}$$

Together, we have $|\Lambda'(v)| = o(|v|^{-\delta} \phi'(b_n)|$ for $v = e_n$.

Second, for $v = -e_n$, again define K' as before

$$|\Lambda'(v)| < \frac{\kappa_2}{e^{-v}} = \kappa_2 (e^{-v})^{-\frac{1}{K'}} (e^{-v})^{-1+\frac{1}{K'}} \text{ for some constant } \kappa_2. \tag{2.73}$$

Note in this case $v < 0$. Now, using the same argument in (2.71) and (2.72) we have $|\Lambda'(v)| = o(|v|^{-\delta} \phi'(b_n))$ for $v = -e_n$.

Now (2.64) and (2.69) implies that (2.70) holds for $v = \pm e_n$. Thus to prove (2.70) for $|v| > e_n$ it suffices to show that

(a.) $f_{M_n}(x)$ is increasing for $x \leq -a_n e_n + b_n$,

(b.) $f_{M_n}(x)$ is decreasing for $x \geq a_n e_n + b_n$.

Now we may write

$$f_{M_n}(x) = n f(x) F^{n-1}(x).$$

So that

$$\frac{d}{dx} f_{M_n}(x) = n F^{n-2}(x) \{ f'(x) F(x) + (n-1) f^2(x). \} \tag{2.74}$$

We see that

$$(2.74)\begin{cases} \geq 0 & \text{if } \frac{f'(x)F(x)}{f^2(x)} \geq -(n-1) \\ \leq 0 & \text{if } \frac{f'(x)F(x)}{f^2(x)} \leq -(n-1). \end{cases} \tag{2.75}$$

So proving (2.70) is equivalent to showing (2.75).

Using (2.3) with $c(x) \equiv 1$

$$F(x) = \exp\{-\exp\{-\int_{-\infty}^{x} \frac{dt}{\phi(t)}\}\}.$$

$$f(x) = \exp\{-\exp\{-\int_{-\infty}^{x} \frac{dt}{\phi(t)}\}\} \exp\{-\int_{-\infty}^{x} \frac{dt}{\phi(t)}\} \frac{1}{\phi(x)}$$

$$\begin{aligned} f'(x) &= \frac{1}{\phi^2(x)} \exp\{-\exp\{-\int_{-\infty}^{x} \frac{dt}{\phi(t)}\}\} \exp\{-\int_{-\infty}^{x} \frac{dt}{\phi(t)}\} \\ &\quad \times [\exp\{-\int_{-\infty}^{x} \frac{dt}{\phi(t)}\} + 1 + \phi'(x)]. \end{aligned}$$

Thus

$$\frac{-f'(x)F(x)}{f^2(x)} = 1 + \{1 + \phi'(x)\} \frac{1}{-\log F(x)}.$$

Now $\phi' \to 0$ as $x \to x_o$ by (2.3), so

$$\frac{-f'(x)F(x)}{f^2(x)} \sim \frac{1}{-\log F(x)}, \quad x \to x_o.$$

Hence, there exists some x^* such that, $\forall x \geq x^*$,

$$\frac{1}{2} \frac{1}{-\log F(x)} < \frac{-f'(x)F(x)}{f^2(x)} < \frac{2}{-\log F(x)}. \tag{2.76}$$

Now, by condition (2.60), we have $a_n e_n + b_n \geq x^*$ for all sufficiently large n, say $n > n_1$.

By condition (2.63), there exists a n_2 such that, whenever $n \geq n_2$

$$\inf_{x \leq x^*} \frac{f'(x)F(x)}{f^2(x)} > -(n-1). \tag{2.77}$$

Let $n^* = \max(n_1, n_2)$. For $n \geq n^*, x^* \leq x \leq -a_n e_n + b_n$,

$$\frac{-f'(x)F(x)}{f^2(x)} < \frac{2}{-\log F(-a_n e_n + b_n)} \sim 2ne^{-e_n} < n-1. \tag{2.78}$$

Note in (2.78) we use

$$-\log F(-a_n e_n + b_n) = -\log F(b_n)\frac{-\log F(-a_n e_n + b_n)}{-\log F(b_n)} \sim \frac{1}{n}e^{e_n} \tag{2.79}$$

which we get by substituting $-e_n$ in for v in (2.35) and using the definition of b_n – namely, $-\log F(b_n) = \frac{1}{n}$.

To continue, putting (2.77) and (2.78) together, we have the $\frac{f'(x)F(x)}{f^2(x)} > -(n-1)$ for the range $x \leq -a_n e_n + b_n$.

For $x \geq a_n e_n + b_n$,

$$\frac{-f'(x)F(x)}{f^2(x)} > \frac{1}{2}\frac{1}{-\log F(a_n e_n + b_n)} \sim 2ne^{e_n} > n-1. \tag{2.80}$$

Thus we have for sufficiently large n,

$$\frac{-f'(x)F(x)}{f^2(x)} \begin{cases} \leq n-1 & \text{if } x \leq -a_n e_n + b_n \\ \geq n-1 & \text{if } x \geq a_n e_n + b_n \end{cases} \tag{2.81}$$

This is equivalent to (2.75). Thus the result (2.70) holds. □

Corollary 52 *Given the notation and conditions in Proposition 51, then for each $\delta > 0$ there exists an n^* and a function ϵ_{n^*} tending to 0 as $n^* \to \infty$ such that $\forall n > n^*$ and $\forall |v| \leq e_n = -K \log |\phi'(b_n)|$, and constant $j > 0$ finite,*

$$\begin{aligned} |e^{-jv} f_{M_n^*}(v) &- e^{-jv}\Lambda'(v)\{1 + (\frac{v^2}{2} - v - \frac{v^2 e^{-v}}{2})\phi'(b_n) \\ &+ (\frac{v^2}{2} + \frac{v^3}{6}[e^{-v} - 1])(2\phi'(b_n)^2 - \phi(b_n)\phi''(b_n)) \\ &+ (v^2 + \frac{5}{6}v^3[e^{-v} - 1] + \frac{1}{8}v^4[e^{-2v} - 3e^{-v} + 1])\phi'(b_n)^2\}| \\ &< \epsilon_{n^*}[\phi'(b_n)^2 + |\phi(b_n)\phi''(b_n)| \min(1, |v|^{-\delta})]. \end{aligned} \tag{2.82}$$

PROOF: This follows immediately from the proof of Proposition 51. For the interval $|v| \le e_n$, the proof is the same as in Proposition 51 except that $f_2(v)$ is replaced by $e^{-jv} f_2(v)$ to absorb the extra e^{-jv}. The changes are in (2.65) where

$$\begin{aligned} e^{-jv} f_2(v) &= e^{-(j+1)v}(1 + \frac{v^2}{2}\phi'(b_n) - \frac{v^3}{6}\{2\phi'(b_n)^2 - \phi(b_n)\phi''(b_n)\} + \frac{v^4}{8}\phi'(b_n)^2 \\ &\quad + o((1+e^{-2v})(1+v^4)R(b_n))). \end{aligned}$$

The only effect this would have on the rate of convergence would be in (2.40) where m would now be $(j+3)$ but again this would be absorbed into the ϵ_{n^*} function.

REMARK This gives us that $e^{-jv} f_{M_n^*}(v)$ is uniformly bounded on $|v| \le e_n$ for any $j > 0$ finite. □

2.4.3 Expansions Involving the Conditional Mean and Variance

Recall the definitions: Let $Y = (X - u)_+$ where u is an arbitrary threshold; that is, Y is an exceedance. The random variable X has mean μ and variance σ^2. The random variable $X|X \le u$ has mean $\mu(u_n)$ and variance $\sigma^2(u_n)$. Finally, the random variable $Y|X > u$ has mean $m(u)$ and variance $s^2(u)$. Let $u = u_n = a_n v + b_n$ where a_n and b_n are the normalizing constants for M_n defined in (2.4) and (2.5).

Proposition 53 *Under conditions of Proposition 51, we have*

$$\mu - \mu(u_n) \sim \begin{cases} \frac{u_n e^{-v}}{n} & x_o = \infty \\ \frac{(x_o - \mu)e^{-v}}{n} & x_o < \infty \end{cases} \tag{2.83}$$

and

$$\sigma^2 - \sigma^2(u_n) \sim \begin{cases} \frac{u_n^2 e^{-v}}{n} & x_o = \infty \\ \{(x_o - \mu)^2 - \sigma^2\}\frac{e^{-v}}{n} & x_o < \infty \end{cases} \tag{2.84}$$

uniformly on $|v| \leq e_n = -K \log|\phi'(b_n)|$.

If $x_o = \infty$ *and*

$$\frac{\phi'(t)\log|\phi'(t)|}{t} \to 0, \quad \textit{as } t \to \infty, \tag{2.85}$$

then

$$\mu - \mu(u_n) \sim \frac{b_n e^{-v}}{n} \tag{2.86}$$

$$\sigma^2 - \sigma^2(u_n) \sim \frac{b_n^2 e^{-v}}{n} \tag{2.87}$$

uniformly in $|v| \leq e_n = -K log|\phi'(b_n)|$.

Proof Now we may write $\mu = \mu(u_n)F(u_n) + (m(u_n) + u_n)\{1 - F(u_n)\}$ from which we may solve

$$\mu - \mu(u_n) = \{\frac{1 - F(u_n)}{F(u_n)}\}\{u_n + m(u_n) - \mu\}. \tag{2.88}$$

We then need to compare the expansions in each term of the above formula.

Under conditions similar to Proposition 51, in his section 9.2, Smith (1987) established expansions for the mean and variance of Y – the exceedance. He showed

$$E\{Y|X > u_n\} = m(u_n) = \phi(u_n) + \phi(u_n)\phi'(u_n) + o(|\phi(u_n)\phi'(u_n)|) \tag{2.89}$$

and

$$E\{Y^2|X > u_n\} = 2\phi^2(u_n) + 6\phi^2(u_n)\phi'(u_n) + o(\phi^2(u_n)(\phi'(u_n))^2).$$

Note that Smith (1987) assumes the Balkema and deHaan (1972) representation for $1 - F(x)$ as opposed to $-\log F(x)$ used in Proposition 51. He also did not assume the function $c(x) \equiv 1$. This does not have an effect on the above formulae in our application. Although the exact form of ϕ may differ from Smith (1987) to here, how it is plugged into the above formulae does not change in the derivation. Specifically, we can solve for $\phi(x) = \frac{\{-\log F(x)\}F(x)}{f(x)}$ using $-\log F(x)$ in the Balkema and deHaan (1972) representation. The equivalent when using $1 - F(x)$ is $\phi(x) = \frac{1-F(x)}{f(x)}$. Asymptotically these are equivalent. Specifically, in our application, the difference between $-\log F$ and $1 - F$ is $o(\frac{1}{n})$ which is smaller than any of the terms derived in the approximations in this thesis. As for the function c, Smith (1987) assumes in his Proposition 9.2, that $c(u) - 1 \sim s\{(\phi'(u))^2 + |\phi(u)\phi''(u)|\} \to 0$ as $u \to x_o$ for finite s. Therefore the remainder term for c is again of smaller order than is being considered in this thesis and therefore does not impact the validity of the above formula for this application.

Now we solve for,

$$s^2(u_n) = E\{Y^2|X > u_n\} - E^2\{Y|X > u_n\} = \phi^2(u_n) + 4\phi^2(u_n)\phi'(u_n) + o(\phi^2(u_n)\phi'(u_n)). \tag{2.90}$$

Next, we look at the expansion of $1 - F(u_n)$. We have

$$\begin{aligned} 1 - F(u_n) &= 1 - e^{\log F(u_n)} \\ &= 1 - \{1 + \log F(u_n) + (1/2 + o(1)) \log^2 F(u_n)\} \\ &= -\log F(u_n)\{1 + (1/2 + o(1)) \log F(u_n)\}. \end{aligned}$$

From (2.35) in Lemma 47 and the fact that $u_n = a_n v + b_n$ and $-\log F(b_n) = 1/n$, we have

$$1 - F(u_n) = \frac{e^{-v}}{n}(1 + o(1)). \tag{2.91}$$

Now recall equations (2.41) and (2.42) from Lemma 49; i.e.,

$$\frac{\phi(u_n)}{u_n} \to 0, \quad x_o = \infty$$

$$\phi(u_n) \to 0, \quad x_o < \infty.$$

Using (2.89), (2.91), (2.41), (2.42) and (2.88), we have our result:

$$\mu - \mu(u_n) \sim \begin{cases} \frac{u_n e^{-v}}{n} & x_o = \infty \\ \frac{(x_o - \mu)e^{-v}}{n} & x_o < \infty \end{cases}$$

uniformly on $|v| \leq e_n = -K \log|\phi'(b_n)|$.

From (2.83), we can solve for

$$\mu^2 - \mu^2(u_n) \sim \begin{cases} 2\mu u_n \frac{e^{-v}}{n} & x_o = \infty \\ 2(x_o - \mu)\mu \frac{e^{-v}}{n} & x_o < \infty \end{cases} \tag{2.92}$$

uniformly on $|v| \leq e_n = -K \log|\phi'(b_n)|$.

Now, we use

$$\begin{aligned} \sigma^2 &= \{EX^2 - (EX)^2\} \\ &= [E(X^2|X \leq u_n)F(u_n) + E(X^2|X > u_n)(1 - F(u_n))] \\ &\quad -[E(X|X \leq u_n)F(u_n) + E(X|X > u_n)(1 - F(u_n))]^2 \end{aligned}$$

and solve for

$$\begin{aligned} \sigma^2 - \sigma^2(u) &= \sigma^2 + \mu^2(u) - \{\frac{\mu^2 + \sigma^2}{F(u)}\} + \{\frac{1}{F(u)} - 1\}\{(u + m(u))^2 + s^2(u)\} \\ &= \mu^2(u) - \mu^2 + \{\frac{1 - F(u)}{F(u)}\}\{(u + m(u))^2 + s^2(u) - \mu^2 - \sigma^2\}. \end{aligned}$$

Now substituting the expansions (2.89), (2.90),(2.92), (2.91),(2.41) and (2.42) into this formula we have,

$$\sigma^2 - \sigma^2(u_n) \sim \begin{cases} \frac{u_n^2 e^{-v}}{n} & x_o = \infty \\ \{(x_o - \mu)^2 - \sigma^2\}\frac{e^{-v}}{n} & x_o < \infty \end{cases}$$

uniformly on $|v| \leq e_n = -K \log|\phi'(b_n)|$.

REMARK Now in case of $x_o = \infty$, we use assumption (2.85). Condition (2.85) guarantees that a_n increases sufficiently slower than b_n does. Note this condition results from assumptions in Lemma 1 of Cohen (1982b). These assumptions are (1) if either $\phi'(u) > 0$ for all sufficiently large u or (2) if either $\phi'(u) < 0$ for all sufficiently large u and $\phi'(u)$ is regularly varying for $u \to \infty$, then (2.85) falls by formula (a) at the bottom of p. 846 of Cohen (1982b).

Now this gives us on the interval $|v| < e_n = -K \log |\phi'(b_n)|$ where we use the definition $\phi(b_n) = a_n$,

$$\begin{aligned} \frac{u_n}{b_n} = \frac{a_n v + b_n}{b_n} &\leq \frac{\phi(b_n)\{-K \log |\phi'(b_n)|\} + b_n}{b_n} \\ &= 1 + -K\frac{\phi(b_n) \log |\phi'(b_n)|}{b_n} \\ &\to 1 \text{ as } n \to \infty, \text{ uniformly in } |v| < e_n. \end{aligned}$$

Again the last line falls from condition (2.85). Thus $u_n = b_n(1 + o(1)), |v| \leq e_n$. In this case, we may replace u_n by b_n in (2.83) and (2.84) without changing the result. This leads to (2.86) and (2.87).

Note in (2.86) and (2.87) we use the fact that

$$\frac{b_n^2}{n} \to 0. \tag{2.93}$$

We see this by looking at

$$\frac{b_n^2}{n} = b_n^2\{-\log F(b_n)\} = b_n^2\{1 - F(b_n) + o(\frac{1}{n})\}.$$

Thus we see (2.93) is equivalent to show $b_n^2\{1 - F(b_n)\} \to 0$. Now

$$b_n^2\{1 - F(b_n)\} \leq \int_{b_n}^{x_o} x^2 dF(x) = \int_{-\infty}^{\infty} 1_{(b_n, x_o)} x^2 dF(x).$$

Now we have assumed that variance of X is finite. So by dominated convergence theorem we have

$$b_n^2\{1 - F(b_n)\} \to 0 \text{ as } n \to \infty.$$

Note this will also give us that

$$\frac{b_n}{\sqrt{n}} \to 0 \text{ as } n \to \infty. \tag{2.94}$$

□

2.5 Proof of Main Theorem and Its Corollary

Proof of Theorem 38 We can write the joint density of S_n^* and M_n^* as

$$\begin{aligned} f_{S_n^*,M_n^*}(w,v) &= f_{M_n^*}(v) f_{S_n^*|M_n^*}(w|v) \\ &= f_{M_n^*}(v)\sqrt{\frac{n\sigma^2}{(n-1)\sigma^2(u_n)}} f_{\tilde{S_n}}(z) \end{aligned} \tag{2.95}$$

where

$$z = \frac{n\mu + \sqrt{n\sigma^2}w - (n-1)\mu(u_n) - u_n}{\sqrt{(n-1)\sigma^2(u_n)}}. \tag{2.96}$$

.

Here we let $u_n = a_n v + b_n$. To enable the uniformity results we will allow v, w and hence z to be dependent on n. In general, we suppress this so as to make the notation easier to read.

Note the transformation from S_n^* to $\tilde{S_n}$ and the form of (2.96) comes from (2.7).

Now to establish (2.14) we break it up as follows

$$\begin{aligned} & f_{M_n^*}(v)\sqrt{\frac{n\sigma^2}{(n-1)\sigma^2(u_n)}} f_{\tilde{S_n}}(z) - f_{S_n^*}(w) f_{M_n^*}(v)\{1 - r_n(e^{-v}-1)w\} \\ &\qquad = E_1 + E_2 + E_3 + E_4 + E_5 + E_{5b} \end{aligned} \tag{2.97}$$

where

$$E_1 = f_{M_n^*}(v) f_{\tilde{S_n}}(z)\left[\sqrt{\frac{n\sigma^2}{(n-1)\sigma^2(u_n)}} - 1\right] \tag{2.98}$$

$$E_2 = f_{M_n^*}(v)\left[f_{\tilde{S_n}}(z) - \mathcal{N}'(z)\{1 + \frac{\kappa_3(u_n)}{6\sigma^3(u_n)\sqrt{n}}(z^3 - 3z)\}\right] \tag{2.99}$$

$$E_3 = f_{M_n^*}(v)\left[\mathcal{N}'(z) - \mathcal{N}'(w)\{1 - r_n(e^{-v}-1)w\}\right] \tag{2.100}$$

$$E_4 = f_{M_n^*}(v)\left[\mathcal{N}'(w)\{1 + \frac{\kappa_3}{6\sigma^3\sqrt{n}}(w^3 - 3w)\} - f_{S_n^*}(w)\right]\{1 - r_n(e^{-v}-1)w\} \tag{2.101}$$

$$E_5 = f_{M_n^*}(v)\left[\frac{\mathcal{N}'(z)\kappa_3(u_n)}{6\sigma^3(u_n)\sqrt{n}}(z^3 - 3z) - \frac{\mathcal{N}'(w)\kappa_3}{6\sigma^3\sqrt{n}}(w^3 - 3w)\right] \tag{2.102}$$

$$E_{5b} = f_{M_n^*}(v)\mathcal{N}'(w)\frac{\kappa_3}{6\sigma^3\sqrt{n}}(w^3 - 3w)[r_n(e^{-v}-1)w]. \tag{2.103}$$

Now to prove (2.14) we show that (2.98) – (2.103) are $o(r_n)$ uniformly $\forall w$ and $|v| \leq -K \log|\phi'(b_n)|$. And to prove (2.98) – (2.103) are $o(r_n)$, it suffices to prove that for any $\epsilon > 0$, $\sum_{j=1}^{6} |E_j| = \epsilon r_n$ for all sufficiently large n. It is necessary to consider two cases where the dependence on n for w, v, and z need to be explicitly expressed. These two cases are: Case (a) $|z_n - w_n| \leq \delta$ for a $\delta > 0$ and Case (b) $|z_n - w_n| > \delta$.

Proof for E_1 The following argument for E_1 holds irrespective of δ and so holds for both Case (a) and Case (b).

We start with $E_1 = f_{M_n^*}(v) f_{\tilde{S}_n}(z) \left[\sqrt{\frac{n\sigma^2}{(n-1)\sigma^2(u_n)}} - 1\right]$. Specifically we begin with its third term.

From (2.84) we have

$$\frac{\sigma^2(u_n)}{\sigma^2} = \begin{cases} 1 - \frac{b_n^2 e^{-v}}{n} + o(\frac{b_n^2 e^{-v}}{n}) & x_o = \infty \\ 1 - \{(x_o - \mu)^2 - \sigma^2\}\frac{e^{-v}}{n} + o[\{(x_o - \mu)^2 - \sigma^2\}\frac{e^{-v}}{n}] & x_o < \infty. \end{cases}$$

From this we get

$$\frac{\sigma^2}{\sigma^2(u_n)} = \begin{cases} 1 + \frac{b_n^2 e^{-v}}{n} + o(\frac{b_n^2 e^{-v}}{n}) & x_o = \infty \\ 1 + \{(x_o - \mu)^2 - \sigma^2\}\frac{e^{-v}}{n} + o[\{(x_o - \mu)^2 - \sigma^2\}\frac{e^{-v}}{n}] & x_o < \infty \end{cases} \tag{2.104}$$

as long as $\frac{b_n^2 e^{-v}}{n} \to 0$ uniformly in v. To see this we note that $\frac{b_n^2 e^{-v}}{n} \sim b_n^2\{1 - F(u_n)\} \sim u_n^2\{1 - F(u_n)\} \leq \int_{u_n}^{\infty} x^2 dF(x) \to 0$ since the variance is assumed finite.

Now we establish a bound for the third term of E_1. Note we have $\sqrt{\frac{n}{n-1}} = 1 + O(\frac{1}{n})$ and that $\frac{\sigma^2}{\sigma^2(u_n)}$ is greater than 1 and bounded.

$$|\sqrt{\frac{n\sigma^2}{(n-1)\sigma^2(u_n)}} - 1| \leq \frac{\sigma^2}{\sigma^2(u_n)} - 1 + O(\frac{1}{n}). \tag{2.105}$$

Now to complete the inequality (2.105), we substitute in (2.104) to get for some κ

$$|\sqrt{\frac{n\sigma^2}{(n-1)\sigma^2(u_n)}} - 1| \leq \begin{cases} \kappa(\frac{b_n^2 e^{-v}}{n} + \frac{1}{n}) & x_o = \infty \\ \kappa(\frac{e^{-v}}{n} + \frac{1}{n}) & x_o < \infty. \end{cases} \tag{2.106}$$

Let us go back to E_1 which we now have is $O(e^{-v} f_{M_n^*}(v) f_{\tilde{S_n}}(z) r_n^2)$. We have that $f_{\tilde{S_n}}(z)$ is bounded $\forall z$ by Proposition 50. Now for the interval $|v| \leq e_n$, we have $e^{-v} f_{M_n^*}(v)$ is bounded by Corollary 52. Therefore

$$E_1 = o(r_n) \quad \text{for } |v| \leq e_n \text{ and } \forall w.$$

or we may write for sufficiently large n ($\epsilon > 0$)

$$|E_1| = \frac{\epsilon}{6}\, r_n \quad \text{for } |v| \leq e_n \text{ and } \forall w. \tag{2.107}$$

Proof for E_2 Like the proof for E_1, the following argument holds for both Case (a) and Case (b).

Next we look at $E_2 = f_{M_n^*}(v) \left[f_{\tilde{S_n}}(z) - \mathcal{N}'(z)\{1 + \frac{\kappa_3(u_n)}{6\sigma^3(u_n)\sqrt{n}}(z^3 - 3z)\} \right]$.

Here, $\left[f_{\tilde{S_n}}(z) - \mathcal{N}'(z)\{1 + \frac{\kappa_3(u_n)}{6\sigma^3(u_n)\sqrt{n}}(z^3 - 3z)\} \right] = o(\frac{1}{\sqrt{n}})$ uniformly in z by Proposition 50. We also have that $f_{M_n^*}(v)$ is bounded $\forall v$ by Proposition 51. Hence

$$E_2 = o(\frac{1}{\sqrt{n}}) = o(r_n), \quad \forall w, \forall v$$

or we may write for sufficiently large n ($\epsilon > 0$)

$$|E_2| = \frac{\epsilon}{6}\, r_n, \quad \forall w, \forall v. \tag{2.108}$$

Proof for E_3 Here we will proceed simultaneously with both cases until *Step 3* where at that point we will need to divide the proof.

Recall $E_3 = f_{M_n^*}(v) \left[\mathcal{N}'(z) - \mathcal{N}'(w)\{1 - r_n(e^{-v} - 1)w\}\right]$. Establishing $E_3 = o(r_n)$ involves a longer argument than needed for E_1 or E_2. We thus break this argument into the following steps.

Step 1: From (2.96), we see that z is a function of w. We now get an explicit form of the difference between z and w.

$$z - w = \frac{n(\mu - \mu(u_n))}{\sqrt{(n-1)\sigma^2(u_n)}} + \left\{ \sqrt{\frac{n\sigma^2}{(n-1)\sigma^2(u_n)}} - 1 \right\} w + \frac{\mu(u_n) - u_n}{\sqrt{(n-1)\sigma^2(u_n)}}. \quad (2.109)$$

Now for $x_o = \infty$ we use (2.86), (2.87), and (2.106) and simplify to

$$z - w = \frac{b_n e^{-v}}{\sqrt{n\sigma^2}}\{1 + o(1)\} + O(we^{-v}(\frac{b_n}{\sqrt{n\sigma^2}})^2) - \frac{b_n}{\sqrt{n\sigma^2}}\{1 + o(1)\}. \quad (2.110)$$

For $x_o < \infty$, we use (2.83), (2.84), and (2.106) and simplify to

$$z - w = \frac{(x_o - \mu)e^{-v}}{\sqrt{n\sigma^2}}\{1 + o(1)\} + O(w(1 + e^{-v})(\frac{b_n}{\sqrt{n\sigma^2}})^2) + \frac{(\mu - x_o)(1 + o(1))}{\sqrt{n\sigma^2}}. \quad (2.111)$$

Using the definition of r_n from (2.13), we can combine (2.110) and (2.111):

$$z - w = r_n(e^{-v} - 1) + o\{r_n(e^{-v} + 1)\} + O(w(e^{-v} + 1)r_n^2). \quad (2.112)$$

Step 2: We take a Taylor expansion for $\mathcal{N}'(z)$ about w.

Using $z = w + t_n$ where t_n is seen in (2.112), we can write

$$\mathcal{N}'(z) = \mathcal{N}'(w + t_n) = \mathcal{N}'(w) + t_n\mathcal{N}''(z^*) \quad \text{where} \quad z^* \text{ between } w \text{ and } z.$$

Using the identity $\mathcal{N}''(x) = -x\mathcal{N}'(x)$, we have

$$\mathcal{N}'(z) = \mathcal{N}'(w) - t_n z^* \mathcal{N}'(z^*)$$

for z^* between w and z.

Substituting this into E_3 we have

$$\begin{aligned}
E_3 &= f_{M_n^*}(v)(z - w)\left[w\mathcal{N}'(w) - z^*\mathcal{N}'(z^*)\right] + f_{M_n^*}(v)\mathcal{N}'(w)r_n(e^{-v} - 1)w \\
&\quad - f_{M_n^*}(v)(z - w)w\mathcal{N}'(w) \\
&= f_{M_n^*}(v)(z - w)\left[w\mathcal{N}'(w) - z^*\mathcal{N}'(z^*)\right] + o(r_n(e^{-v} + 1)w\mathcal{N}'(w)f_{M_n^*}(v)) \\
&\quad + O(w(e^{-v} + 1)r_n^2 w\mathcal{N}'(w)f_{M_n^*}(v)) \\
&= E_6 + E_7 + E_8.
\end{aligned}$$

Now, of course, we have $w^2\mathcal{N}'(w)$ is bounded and $(1+e^{-v})f_{M_n^*}(v)$ is bounded as seen in Corollary 52. Thus

$$E_7 = o(r_n) \ \text{ uniformly } \forall w \text{ and } |v| \le e_n,$$

or for sufficiently large n ($\epsilon > 0$)

$$|E_7| = \frac{\epsilon}{18}\, r_n \ \text{ uniformly } \forall w \text{ and } |v| \le e_n. \tag{2.113}$$

Also,

$$E_8 = O(r_n^2) \ \text{ uniformly } \forall w \text{ and } |v| \le e_n$$

or for sufficiently large n ($\epsilon > 0$)

$$|E_8| = \frac{\epsilon}{18}\, r_n \ \text{ uniformly } \forall w \text{ and } |v| \le e_n. \tag{2.114}$$

Step 3: Now we focus on E_6. The important details in this formula concern $z - w$ since the other terms are bounded. Recall v, w, z, and z^* actually depend on n. So fix the notation by writing $v = v_n$, $z = z_n$, $w = w_n$, and $z^* = z_n^*$ so the dependence on n is explicit.

Now substituting (2.112) into E_6 we obtain

$$\begin{aligned} E_6 &= f_{M_n^*}(v_n)\left\{r_n(e^{-v_n}-1) + o\{r_n(e^{-v_n}+1)\} + O(we^{-v_n}r_n^2)\right\}\left[w_n\mathcal{N}'(w_n) - z_n^*\mathcal{N}'(z_n^*)\right] \\ &= E_9 + E_{10} + E_{11}. \end{aligned} \tag{2.115}$$

At this point, it is necessary to separate the argument into the two cases.

Case (a) If $|z_n - w_n| \le \delta$ then $|z_n^* - w_n| \le \delta$. By uniform continuity of $w_n^k\mathcal{N}'(w_n)$ for $k = 0, 1, 2$ given $\epsilon > 0$ we can find a $\delta > 0$ such that

$$|z_n - w_n| < \delta \Rightarrow |z_n^k\mathcal{N}'(z_n) - w_n^k\mathcal{N}'(w_n)| < \frac{\epsilon}{C}$$

for $k = 0, 1, 2$ and any given constant $C > 0$.

Now since $(e^{-v_n}+1)f_{M_n^*}(v_n)$ is bounded on $|v_n| \le e_n$, we have that each E_9, E_{10}, and E_{11} is bounded by some

$$constant \times |z_n^{*k}\mathcal{N}'(z_n^*) - w_n^k\mathcal{N}'(w_n)|$$

for $k = 0, 1, 2$.

In other words, we can choose a δ so that $\forall w_n$ and $\forall |v_n| \le e_n$

$$|z_n - w_n| < \delta \Rightarrow |E_6| < \frac{\epsilon}{18}\, r_n \;\text{ for all sufficiently large } n$$

which with (2.113) and (2.114) gives

$$|E_3| < \frac{\epsilon}{6}\, r_n \;\text{ for all sufficiently large } n \tag{2.116}$$

for $|v_n| \le e_n$ and $\forall w_n$ and when $|z_n - w_n| \le \delta$. [End Case (a)]

Case (b) Here we show if $|z_n - w_n| > \delta$ then the entire left-hand side of (2.97) is $o(r_n)$.

Part 1 Suppose $|z_n - w_n| > \delta$ and $r_n e^{-v_n} < \delta^2$.

From (2.112) we deduce $|w_n r_n| >$ some $\delta_1 > 0$ for all sufficiently large n. So $|w_n| > \frac{\delta_1}{r_n}$. We also have from (2.112),

$$\begin{aligned} z_n &= w_n + r_n(e^{-v_n} - 1) + o(r_n(e^{-v_n}+1)) + O(w_n(e^{-v_n}+1)r_n^2) \\ &= w_n + o(1) + o(1) + O(w_n o(1)) \\ &= w_n(1 + o(1)) \\ &> \frac{\delta_1}{2r_n}, \quad \text{say,} \end{aligned}$$

for all sufficiently large n.

Thus we have [with (2.106)]

1. $|f_{\tilde{S_n}}(z_n)| = o(r_n)$ by Proposition 50.

2. $|w_n|^k |f_{S_n^*}(w_n)| = o(r_n)$ for k=0,1 by Petrov's central limit theorem.

3. $|(e^{-v_n}+1)f_{M_n^*}(v_n)|$ is bounded on $|v_n| \le e_n$. by Proposition 51 and Corollary 52.

Hence the left-hand side of (2.97) is $o(r_n)$ on $|v_n| \le e_n$. and $\forall w_n$ when $|z_n - w_n| > \delta$ and $r_n e^{-v_n} < \delta^2$.

<u>*Part 2*</u> Suppose $|z_n - w_n| > \delta$ and $r_n e^{-v} \ge \delta^2$. Corollary 52 allows us to say that, for finite $m' > 2$, $e^{-m' v_n} e^{-v_n} f_{M_n^*}(v_n)$ is uniformly bounded on $|v_n| \le e_n$. This gives us that $e^{-v_n} f_{M_n^*}(v_n) = O(e^{(m'-1)v_n})$ but if $r_n e^{-v_n} \ge \delta^2 > 0$ then $r_n e^{-v_n} \not\to 0$ so $|v_n| < \log r_n$ for sufficiently large n. Thus

$$e^{-v_n} f_{M_n^*}(v_n) = O(e^{(m'-1)v_n}) \le O(r_n^{(m'-1)}) = o(r_n) \text{ since } m' \text{ can be taken} > 2. \quad (2.117)$$

Now looking at the parts of (2.97), we have that $f_{\tilde{S_n}}(z_n)$ is bounded – see Proposition 50 – and $w_n f_{S_n^*}(w_n)$ is bounded – again, see Petrov's (1975) central limit theorem. From (2.117) we have $e^{-v} f_{M_n^*}(v_n) = o(r_n)$ which also gives $f_{M_n^*}(v_n) = o(r_n)$ [End of Case(b)]

Altogether we see that the left-hand side of (2.97) is $o(r_n)$ when $|z_n - w_n| > \delta$ and $r_n e^{-v} \ge \delta^2$.

Thus for E_3 we have either for sufficiently large n ($\epsilon > 0$)

$$|E_3| = \frac{\epsilon}{6}\, r_n \quad \text{for } |v| \le e_n \text{ and } \forall w. \quad (2.118)$$

or left-hand side of (2.97) is $o(r_n)$.

Proof of E_4 Like the proofs for E_1 and E_2 this argument holds irrespective of δ and hence is the same for both Case (a) and Case (b).

Recall we have

$$E_4 = f_{M_n^*}(v)\left[\mathcal{N}'(w)\{1 + \frac{\kappa_3}{6\sigma^3\sqrt{n}}(w^3 - 3w)\} - f_{S_n^*}(w)\right]\{1 - r_n(e^{-v}-1)w\}.$$

Again we have $f_{M_n^*}(v)$ is bounded and also by Feller (1971), Chapter XVI, Section 2, Theorem 1 the term inside $[\ldots]$ is $o(\frac{1}{\sqrt{n}})$ uniformly in w. To handle the term $r_n(e^{-v}-1)w$, we need

(a) $\sup_v (e^{-v} - 1) f_{M_n^*}(v)$ to be bounded on $|v| \le e_n$ which we again have by Corollary 52

(b) $\sup_w |w\{\mathcal{N}'(w) - f_{S_n^*}(w)\}| \to 0$ as $n \to \infty$ which follows from Petrov's (1975) central limit theorem under the same assumptions as Proposition 50.

(c.) $w^4 \mathcal{N}'(w)$ to be bounded so $\mathcal{N}'(w)\frac{\kappa_3}{6\sigma^3\sqrt{n}}(w^3 - 3w)w \to 0$ uniformly in w which we have by properties of the normal density.

Thus

$$E_4 = o(r_n) \quad \text{for } |v| \le e_n \text{ and } \forall w.$$

or we may write for sufficiently large n ($\epsilon > 0$)

$$|E_4| < \frac{\epsilon}{6}\, r_n \quad \text{for } |v| \le e_n \text{ and } \forall w. \tag{2.119}$$

Proof of E_5 Finally we look at the term

$$E_5 = f_{M_n^*}(v) \left[\frac{\mathcal{N}'(z)\kappa_3(u_n)}{6\sigma^3(u_n)\sqrt{n}}(z^3 - 3z) - \frac{\mathcal{N}'(w)\kappa_3}{6\sigma^3\sqrt{n}}(w^3 - 3w) \right].$$

Now, again we have $f_{M_n^*}(v)$ is uniformly bounded on $|v_n| \le e_n$. We also have that the function $\mathcal{N}'(z)\{z^3 - 3z\}$ is uniformly continuous so by similar argument to proof of E_3, particularly *Step 3*, we can conclude

$$|E_5| < \frac{\epsilon}{6} r_n \quad \text{or (2.97) is } o(r_n), \forall |v_n| \le e_n, \forall w. \tag{2.120}$$

Proof of E_{5b} Like the proofs for E_1 and E_2 this argument holds irrespective of δ and hence is the same for both Case (a) and Case (b).

Recall we have $E_{5b} = f_{M_n^*}(v)\mathcal{N}'(w)\frac{\kappa_3}{6\sigma^3\sqrt{n}}(w^3 - 3w)[r_n(e^{-v} - 1)w\}]$.

Now, we have this immediately since,

(a) $\sup_v (e^{-v} - 1) f_{M_n^*}(v)$ is bounded on $|v| \le e_n$ by Corollary 52.

(b.) $w^4 \mathcal{N}'(w)$ is bounded uniformly in w so $\mathcal{N}'(w) \frac{\kappa_3}{6\sigma^3}(w^3 - 3w)w$ is bounded uniformly in w which we have by properties of the normal density.

Thus we have

$$E_{5b} = O(\frac{r_n}{\sqrt{n}}) = o(r_n) \quad \text{for } |v| \le e_n \text{ and } \forall w.$$

or we may write for sufficiently large n ($\epsilon > 0$)

$$|E_{5b}| < \frac{\epsilon}{6} r_n \quad \text{for } |v| \le e_n \text{ and } \forall w. \tag{2.121}$$

In conclusion, using (2.107),(2.108), (2.118), (2.119), (2.120),and (2.121), we have shown that $\sum_{j=1}^{6} |E_j| \le \epsilon r_n$ uniformly $\forall w$ and $|v| \le e_n = -K \log |\phi'(b_n)|$. Hence our result. □

Proof of Corollary 39

First we will show the result (2.16) – i.e., for the case when $x_o < \infty$.

Using the result of Theorem 38 – (2.14) – to show (2.16) we need to prove that

$$\Big| \Big[f_{S_n^*}(w) f_{M_n^*}(v) - \mathcal{N}'(w)\{1 + \frac{\kappa_3}{6\sigma^3\sqrt{n}}(w^3 - 3w)\}\Lambda'(v) \\ \times \{1 + (\frac{v^2}{2} - v - \frac{v^2 e^{-v}}{2})\phi'(b_n)\}\Big] \{1 - r_n(e^{-v} - 1)w\} \Big| \\ = o(\max\{r_n, |\phi'(b_n)|\}). \tag{2.122}$$

for $\forall w$ and $\forall |v| \le e_n = -K \log |\phi'(b_n)|$.

First we show (2.122) without the $r_n(e^{-v} - 1)w$ term.

Let

$$A_n = f_{S_n^*}(w),$$

$$A'_n = \mathcal{N}'(w)\{1 + \frac{\kappa_3}{6\sigma^3\sqrt{n}}(w^3 - 3w)\},$$

$$B_n = f_{M_n^*}(v),$$

and

$$B'_n = \Lambda'(v)\{1 + (\frac{v^2}{2} - v - \frac{v^2 e^{-v}}{2})\phi'(b_n)\}.$$

Then we can write the left-hand side of (2.122) without the $r_n(e^{-v} - 1)w$ term as

$$\begin{aligned} |A_n B_n - A'_n B'_n| &= |A_n B_n - A_n B'_n + A_n B'_n - A'_n B'_n| \\ &\leq |A_n|\,|B_n - B'_n| + |B'_n|\,|A_n - A'_n| \qquad (2.123) \end{aligned}$$

Now $|A_n|$ is bounded $\forall w$ by Petrov's result and $|B_n - B'_n| = o(|\phi'(b_n)|)$ $\forall v$ by Proposition 51. Thus the first term on the right-hand side of the inequality in (2.123) is $o(|\phi'(b_n)|)$, $\forall v$ and $\forall w$. For the second term in the inequality in (2.123), $|B'_n|$ is bounded $\forall v$ by Proposition 51 and $|A_n - A'_n| = o(r_n)$ by Feller (1971), Chapter XVI, Section 2, Theorem 1 uniformly in w. Thus this second term is $o(r_n)$, $\forall v$ and $\forall w$. Thus the right-hand side of (2.123) is $o(\max\{r_n, |\phi'(b_n)|\})$, $\forall v$ and $\forall w$.

When we add in the $r_n(e^{-v} - 1)w$ term, we need to strength this to

(a.) $|wA_n|$ bounded – which we again have $\forall w$ by Petrov's result.

(b.) $|e^{-v} + 1|\,|B_n - B'_n| = o(|\phi'(b_n)|)$ – which we have by Corollary 52, note now on $|v| \leq -K \log|\phi'(b_n)|$.

(c.) $|e^{-v} + 1|\,|B'_n|$ bounded – which we again have by Corollary 52, note now on $|v| \leq -K \log|\phi'(b_n)|$.

(d.) $|w|\,|A_n - A'_n| = o(r_n)$ – which we again have $\forall w$ by Petrov's result.

Hence, we have result (2.16) for $|v| \leq -K \log|\phi'(b_n)|$ and $\forall w$.

For the case when $x_o = \infty$ – (2.15) – the term $\frac{\kappa_3}{6\sigma^3\sqrt{n}}(w^3 - 3w)$ need not be included with the higher order terms since in this case it is $o(r_n)$ by the definition of r_n when $x_o = \infty$, see (2.13). □

PROOF OF COROLLARY 40: We have $H(v;0,1,k) = \exp[-(1-kv)^{1/k}]$ so that

$$H'(v;0,1,k) = (1-kv)^{1/k}\exp[-(1-kv)^{1/k}].$$

If we look at $-\log H'$, we have

$$(1-\frac{1}{k})\log(1-kv) + (1-kv)^{1/k}. \tag{2.124}$$

Expanding $(1-kv)^{1/k}$ by $e^{-v}(1-\frac{kv^2}{2}) + O(k^2)$ and $(1-\frac{1}{k})\log(1-kv)$ by $v - kv + \frac{kv^2}{2} + O(k^2)$, we have (2.124) is equal to

$$v + e^{-v} + k(\frac{v^2}{2} - v - \frac{e^{-v}v^2}{2}) + O(k^2). \tag{2.125}$$

Now if we look at the log of (2.17), we have

$$-\log\left[\Lambda'(v)\{1+(\frac{v^2}{2} - v - \frac{v^2e^{-v}}{2})\phi'(b_n)\}\right] =$$
$$v + e^{-v} - (\frac{v^2}{2} - v - \frac{e^{-v}v^2}{2})\phi'(b_n) + O(\phi(b_n)^2). \tag{2.126}$$

With the definition $k_n = -\phi'(b_n)$ we have that (2.125) and (2.126) match up to $O(\phi'(b_n)^2)$. □

2.6 Simulation Project

2.6.1 Objective

The overall objective of the simulation project is to study the performance of the higher order term under different underlying distributions and as n, the sample size, increases. To do that we will evaluate the dependence between the maximum and sum of simulated data, looking to see the extent of dependence that exists between these two random variables at different levels of n and different underlying distributions. Then we can check the fit of the asymptotic independence case; that is, we will see how much error we incur by applying the limiting version in each case. Finally, using the higher order expansion term, we will be able to determine how much this term improves the fit of the maximum and sum.

2.6.2 Form

In the previous section, we have established the higher order term for the joint density of the sum and the maximum $f_{S_n^*,M_n^*}(w,y)$. The joint density gives the widest range of application. It can be used to model S_n^* and M_n^* simultaneously or can be manipulated to model the conditional random variables $S_n^*|M_n^*$ or $M_n^*|S_n^*$. In fact, the conditional density of the maximum given the sum, $f_{M_n^*|S_n^*}(y|w)$, is the form of the density that will be used in this simulation project. Although there is not necessarily a best form to study this new higher order term, this particular conditional density allows us to see the effect changing the sum has on the maximum. Now as we try to compare the simulated data of the maximum given the sum to the limiting Gumbel density and to the higher order expansion of $f_{M_n^*|S_n^*}(y|w)$, we need to choose a particular functional of these competing densities to compare. Traditionally, in extreme value theory, one computes percentiles (or n th year return levels) of the appropriate distribution. Here, we will choose to compare the 80th percentile under these three cases:

A.) "True" conditional density: represented by the simulated data of $M_n^*|S_n^*$.

B.) Limiting independent case: $f_{M_n^*|S_n^*}(y|w) = f_{M_n^*}(y) = \Lambda'(y)$ where Λ' is the Gumbel density.

C.) The higher order expansion term case: $f_{M_n^*|S_n^*}(y|w) = \Lambda'(y) \times \{1 - \frac{w(e^{-y}-1)b_n}{\sqrt{n\sigma^2}}\}$.

The 80th percentile has no intrinsic value. The 80th is selected since traditionally we look at percentiles in the upper tail but in this case we did not want to go very high into the tail so that the emphasis stays with the behavior of the data as opposed to the extreme behavior of the tail.

A comparision of the simulated data, the Gumbel limiting density, and the expansion depends not only on the sample size but also on the underlying distribution of the process. The underlying distribution defines the b_n term in the expansion term and thus affects the rate of convergence of the conditional density. To see this effect three different underlying distributions are considered: the Normal distribution, the Lognormal distribution, and the Weibull distribution.

2.6.3 Procedure

For each of the above underlying distributions, we apply the following steps for n = 30,60,90, and 120.

Step 1: Simulate 10,000 independent samples of size n.

Step 2: Calculate the sum and the maximum for each sample.

Step 3: Order the pairs (S_n, M_n) with respect to S_n and divide into 50 bins which contain 10,000/50 or 200 pairs apiece.

Step 4: In each bin, order the maxima and find the 80th percentile.

Step 5: Plot the simulated data by graphing the bin midpoint versus that bin's 80th percentile.

Step 6: Plot the 80th percentile of the limiting Gumbel density. Note this will be constant with respect to the sum.

Step 7: Using the bin midpoints upon which to condition, calculate and plot the 80th percentile of the expansion of the conditional density.

Step 8: Overlay steps 5, 6, and 7.

2.6.4 Issues to Resolve in the Simulation Project

"Numerical Recipes in C" by Press *et al.* (1990) provide many standard subroutines for the program including those for sorting, providing the gamma function, providing uniform deviates, providing normal deviates, and integrating.

As with all computer results, the level of round-off error has to be established. The C+ program defines all necessary variables at double precision; that is, at 10^{-16}. The only exception to this is in the Weibull case when evaluating the gamma function to calculate the mean and variance. The subroutine **gammaln()** is significant only to 8 digits. For more detail, see Press *et al.* (1990), p. 168.

The subroutine **ran0()** is used to produce uniform deviates within (0,1). The routine in based on the algorithm of Bays and Durham as described in Knuth (1981, Sections 3.2-3.3). Essentially, it provides an additional random shuffling of the random numbers generated by the basic random number generator of the computer. This should free the random numbers from sequential correlation. For more details, see Press *et al.* (1990) p. 207.

The subroutine **gesdev()** is used to produce standard normal deviates. The routine is an adaptation of the Box-Muller method for generating random deviates with a normal (Gaussian) distribution. The adaptation is to pick a point randomly inside a unit circle versus a unit square so as to avoid calls to trig functions. Again for further details, see Press *et al.* (1990) p. 217.

The integration subroutine used for going from the density to the cumulative distribution function was a combination of **trapzd()** and **qsimp()**. The qsimp subroutine which utilizes the trapzd subroutine is based on the extended trapezoidal rule and has error $O(1/N^4)$ where N is the number of points the interval to be integrated over is broken. Together this integration subroutine is considered reliable for uncomplicated work. Again, for further details, see Press *et al.* (1990), pp. 120-123.

As with any simulation of a continuous random variable, the effect of finite bin widths has to be considered. The issue concerning finite bin widths in this project is: In theory when conditioning upon the sum, we would let S_n vary along the appropriate interval, continuously. Obviously, we cannot condition upon all possible values of S_n. Instead, we condition on a S_n in a particular bin, one in each of the 50 bins. In fact, we use the bin midpoint. Having a finite bin width, versus an "infinitesimal" one, would present a problem if using different values in the bin affects the overall appearance of the graph. To judge the effect of finite bin width on the results, the graphs in the simulation project are refit using the left and the right endpoints of the bins. Figure 2.1 shows this graph for the Normal distribution with sample size of 30. Other permutations of the simulation project, under different underlying distributions and different sample sizes, show similar results. Except for a few bins at either end, there is no discernible finite bin width effect on the results.

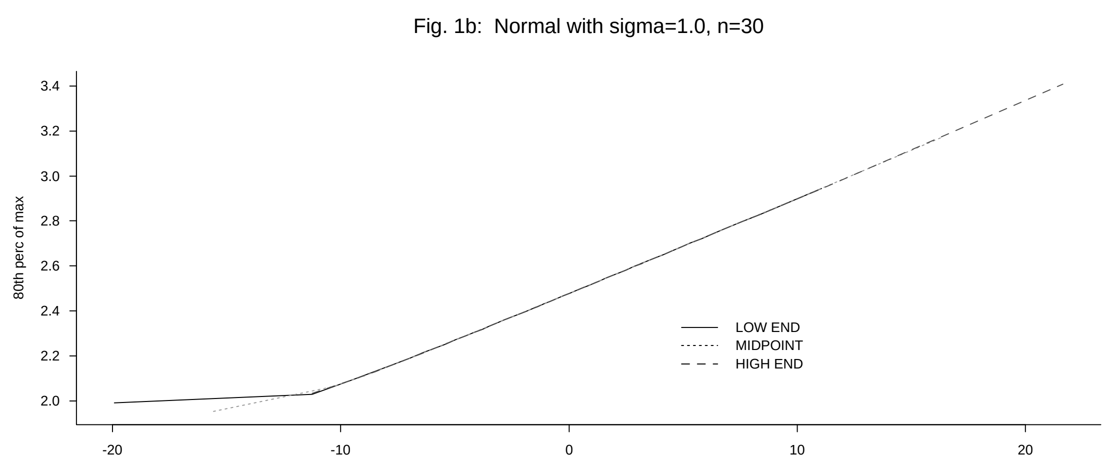


Figure 2.1: Finite bin width consideration in the simulation project. Normal distribution with $\sigma = 1$. Graphs simulated data plotted against the low (solid line), midpoint (short dashed line), and high (long dashed line) end of the bin.

Other issues to resolve are specific to this simulation: (1) How to deal with the fact that the higher order expansion of the conditional density is not a proper density; that is, for some values it can be negative and does not necessarily integrate to 1. (2) What form of a_n and b_n to use in the conditional density.

If we look at the form of the expansion of the conditional density

$$f_{M_n^*|S_n^*}(y|w) = \Lambda'(y) \times \{1 - \frac{w(e^{-y}-1)b_n}{\sqrt{n\sigma^2}}\},$$

we see that for some values of y, this expansion of the conditional density can be negative. To rectify this problem, we actually program a different form of the expansion. Since the higher order term is contained in the conditional mean, i.e. in $E(S_n|M_n = u_n)$, this is the term on which we wish to focus our attention. Instead of pulling the higher order term out in the final Taylor expansion in the derivation, we leave it on the inside. Specifically, we use the higher order expansion of the conditional mean and variance in the density of $S_n^*|M_n^*$ and hence, since we use the normal density for $S_n^*|M_n^*$, a proper density, we get a non-negative function, although this form does not necessarily integrate to one. Note we can rescale to compensate when the expansion does not integrate to one. Thus for the expansion of the conditional density of $M_n^*|S_n^*$, we program

$$f_{M_n^*|S_n^*}(y|w) \approx \frac{\Lambda'(y) \times \mathcal{N}'(z)}{\mathcal{N}'(w)}$$

where $\mathcal{N}'$ is the normal density and

$$z = \frac{n\mu + \sqrt{n\sigma^2}w - \{(n-1)\mu(u_n) + u_n\}}{\sqrt{(n-1)\sigma(u_n)}}.$$

Note from our notation in the previous chapter we have $\mu = EX$, $\sigma^2 = Var(X)$, $\mu(u_n) = E(X|X < u_n)$, and $\sigma^2(u_n) = Var(X|X < u_n)$.

Thus it is the 80th percentile of this form of the conditional density that is evaluated in the simulation project.

Finally, both a_n and b_n contribute to the expansion of the conditional mean and variance through u_n, $\mu(u_n)$ and $\sigma^2(u_n)$, as can be seen through formulae (2.86) and (2.87) in the Section 2.4. The following forms of these two normalizing constants are used in the program:

$$b_n \text{ such that } -\log(F(b_n)) = \frac{1}{n}$$

and

$$a_n = \phi(b_n) = \frac{-\log(F(b_n))}{f(b_n)}$$

where F corresponds to the underlying distribution being simulated and f is the corresponding density. In particular, for each of the underlying distributions, we first invert

$$-\log(F(b_n)) = \frac{1}{n} \tag{2.127}$$

to solve for b_n.

Then the formula for a_n simplifies to

$$a_n = \frac{1}{nf(b_n)} \tag{2.128}$$

Now for each distribution, we use its density and substitute the b_n we solved for in (2.127) into (2.128) to solve for a_n. Thus for each distribution, we can explicitly solve for both b_n and a_n.

2.6.5 Results

Before we investigate the main findings of the simulation project, there is one more issue to raise.

In each of the three underlying distributions, its variance/shape parameter is also varied in the simulation. The graphs show the same overall features except as the variance/shape parameter increases there are more extreme events. This has a tendency to affect greatly the upper (and lower) bins in the simulation. Then what dominates the graph is not the behavior of the higher order expansion or its asymptotic counterpart at the center of the support but rather the quirky behaviour at the outside bins.

How severely this shows up in the graphs depends on the underlying distribution. Changing the variance for the Normal family does not alter the behaviour of the simulation, only the vertical scale of results. The Weibull distribution is most affected with the last two bins dominating the results. The Lognormal family shows moderate effects to the increase. Figure 2.2 (a.), (b.), (c.), and (d.) show the Lognormal distribution for X where $\log(X) = Z \sim \mathcal{N}(\mu, \sigma^2)$. In all cases, $\mu = 0$ and σ varies through 0.25, 0.50, 1.0, and 2.0.

Note in cases where this change in variance/shape parameter affects the results, the higher order term gives a poorer fit as this parameter increases. This results when the variance/shape parameter changes in the Lognormal and Weibull distribution so that the distributions become"farther" away from the Normal distribution.

So as not to run into this large deviation in scale in the outer bins, we will try to use parameters in each family which focus the results to the center of the graphs. Note we also present results for the Weibull distribution with $\alpha = \beta = 1$, the Exponential distribution. This particular distribution is not a member of Cohen's (1982) class N; that is, does not fulfill the assumptions in this chapter. We include this example so that we may compare it to the Weibull $(\alpha = 1, \beta = 1/2)$ case.

The results are:

1.) Figure 2.3 (a.), (b.), (c.), and (d.) shows results for the Normal distribution with $\mu = 0$ and $\sigma = 1$ for n=30,60,90,120.

2.) Figure 2.4 (a.), (b.), (c.), and (d.) shows results for the Lognormal distribution with normal parameters $\mu = 0$ and and $\sigma = 0.50$ for n=30,60,90,120.

3.) Figure 2.5 (a.), (b.), (c.), and (d.) shows results for the Weibull distribution with $\alpha = 1$ and $\beta = 1/2$ for n=30,60,90,120.

4.) Figure 2.6 (a.), (b.), (c.), and (d.) shows results for the Weibull distribution with $\alpha = \beta = 1$ [Exponential case] for n=30,60,90,120.

2.6.6 Conclusions

In all cases, the sum and the maximum do not behave independently. Looking at the vertical scale under all three underlying distributions, the behavior moves toward independence; that is, the slope of the simulated data decreases but even at n=120, we can clearly see the dependence – a positive slope. See Figures 9,13, and 17. Hence we can conclude that under these conditions using the asymptotic result – the limiting Gumbel 80th percentile –would not be appropriate.

This we see in all cases: The limiting Gumbel does not appear to be a good fit. As n increases more of the simulated data crowds toward the Gumbel limit in the center. The vertical scale has been fixed for each distribution to show this shrinkage as n increases. What we do see for all n is that the Gumbel is only reliable at the center

Finally for all underlying distributions and for all n considered here, the higher order expansion provides a better fit. It is always closer to the simulated data. In fact except for a few bins at the lower and the upper ends, the higher order expansion follows the behaviour of the simulated data very closely in the Normal and Lognormal cases. Although the approximation does not fit the simulated data with Weibull ($\alpha = 1, \beta = 1/2$) as well as in the first two cases, it still is an improvement over the limiting Gumbel case. It is interesting in the Exponential case, Figure 2.6, that the approximation fits the simulated data better than in Figure 2.5 which actually belongs to Cohen's (1982) class N. Note the Exponential distribution belongs to Cohen's (1982) class E. When the underlying distribution belongs to class E, the maximum has a faster rate of convergence to the extreme value distribution. The implication for this simulation project is that this rate of convergence to the asymptotic extreme value distribution has an impact on the effectiveness of the higher order term. Obviously in all cases, the fit between the simulated and approximation improves as the sample size becomes larger.

We conclude that adding the higher order expansion significantly improves the fit for all sample sizes and for all underlying distributions. Thus if one needs to model the sum or maximum and has information on the other random variable or wants to model both variables simultaneously, one should introduce the higher order term. The new model will give a closer fit to the data.

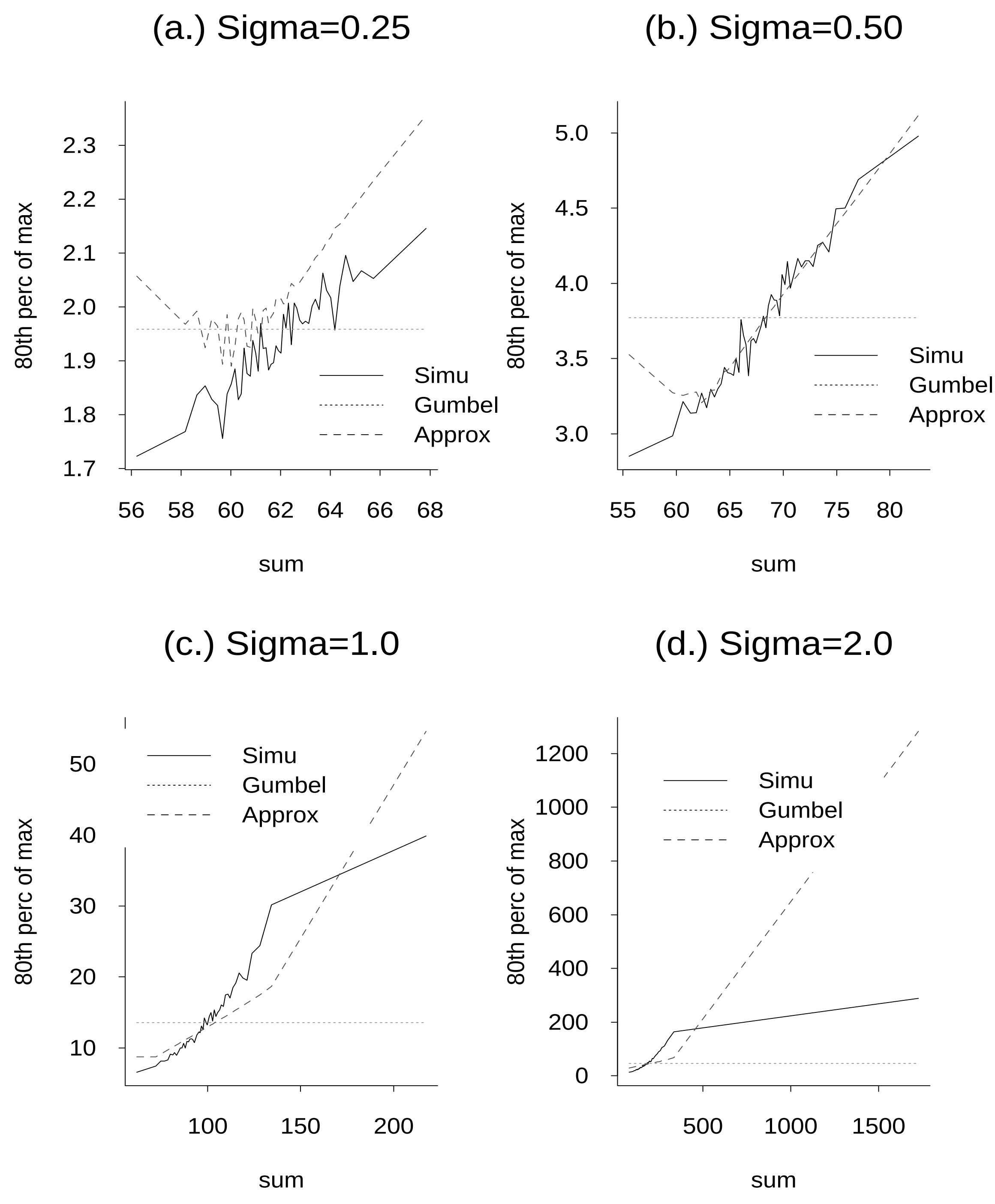


Figure 2.2: Effect of increasing standard deviation within underlying distribution. Log-normal distribution with standard deviation: (a) 0.25, (b) 0.50, (c) 1.0, (d) 2.0. The solid line represents the simulated data; the short dashed line represents the Gumbel limit; and the long dashed line represents the higher order approximation of the conditonal denstiy.

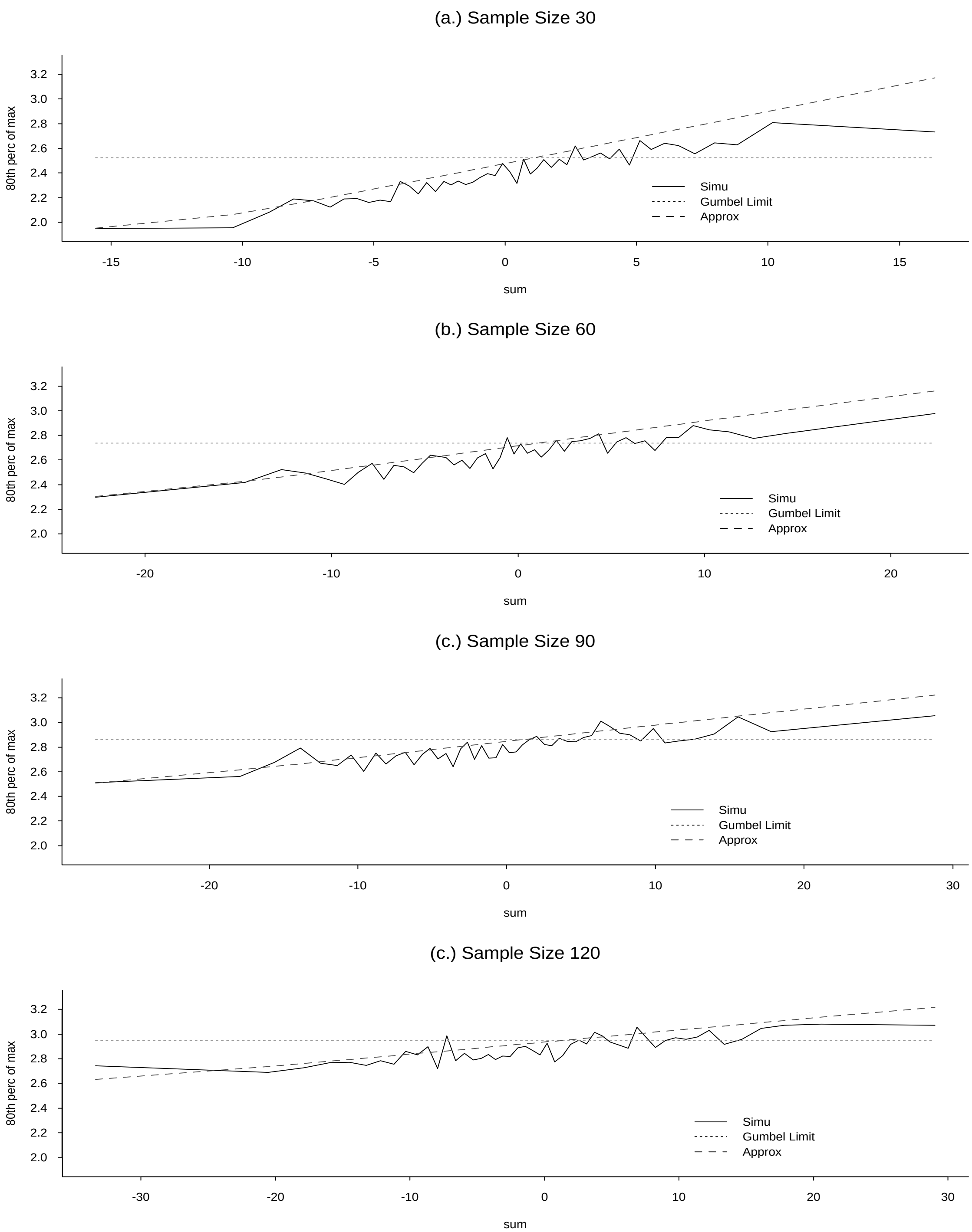


Figure 2.3: Effects of sample size on expansion of joint density. Normal distribution with $\mu = 0$ and $\sigma = 1$ as sample size increases from: (a) 30, (b) 60, (c) 90, to (d)120. The solid line represents the simulated data; the short dashed line represents the Gumbel limit; and the long dashed line represents the higher order approximation of the conditional density.

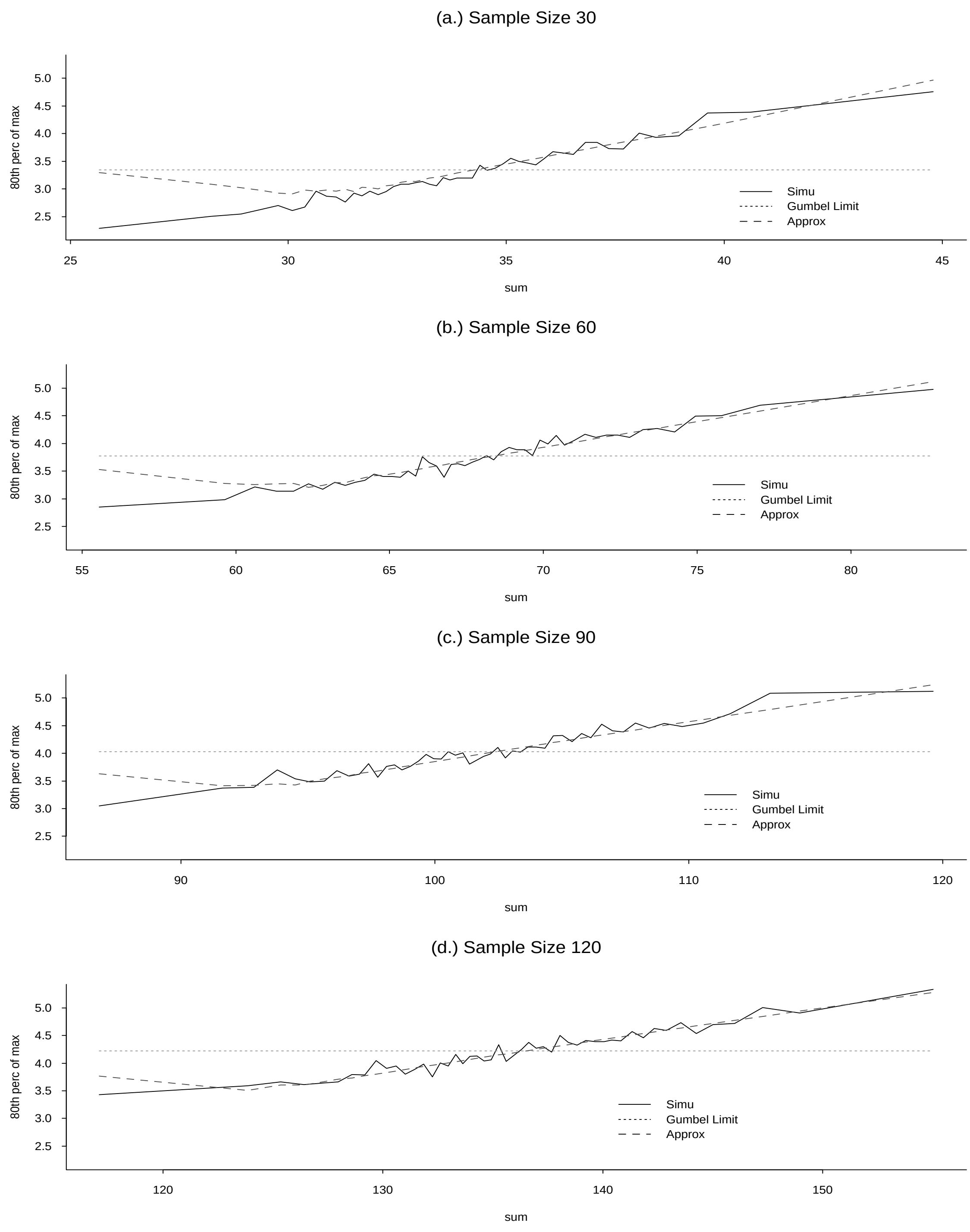


Figure 2.4: Effects of sample size on expansion of joint density. Lognormal distribution with standard deviation 0.50 as sample size increases from: (a) 30, (b) 60, (c) 90, to (d)120. The solid line represents the simulated data; the short dashed line represents the Gumbel limit; and the long dashed line represents the higher order approximation of the conditional density.

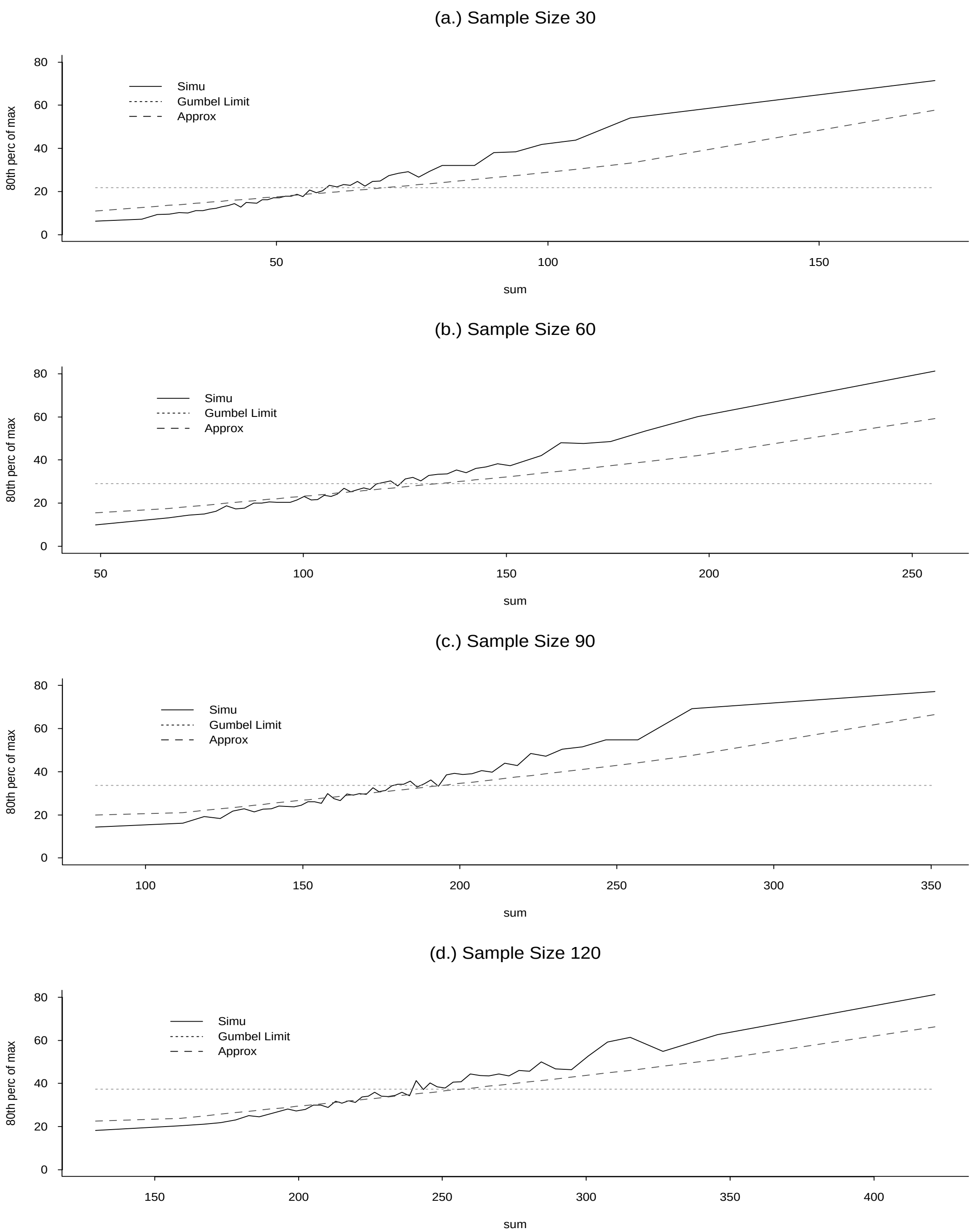


Figure 2.5: Effects of sample size on expansion of joint density. Weibull distribution with $\alpha = 1$ and $\beta = 1/2$ as sample size increases from: (a) 30, (b) 60, (c) 90, to (d) 120. The solid line represents the simulated data; the short dashed line represents the Gumbel limit; and the long dashed line represents the higher order approximation of the conditional density.

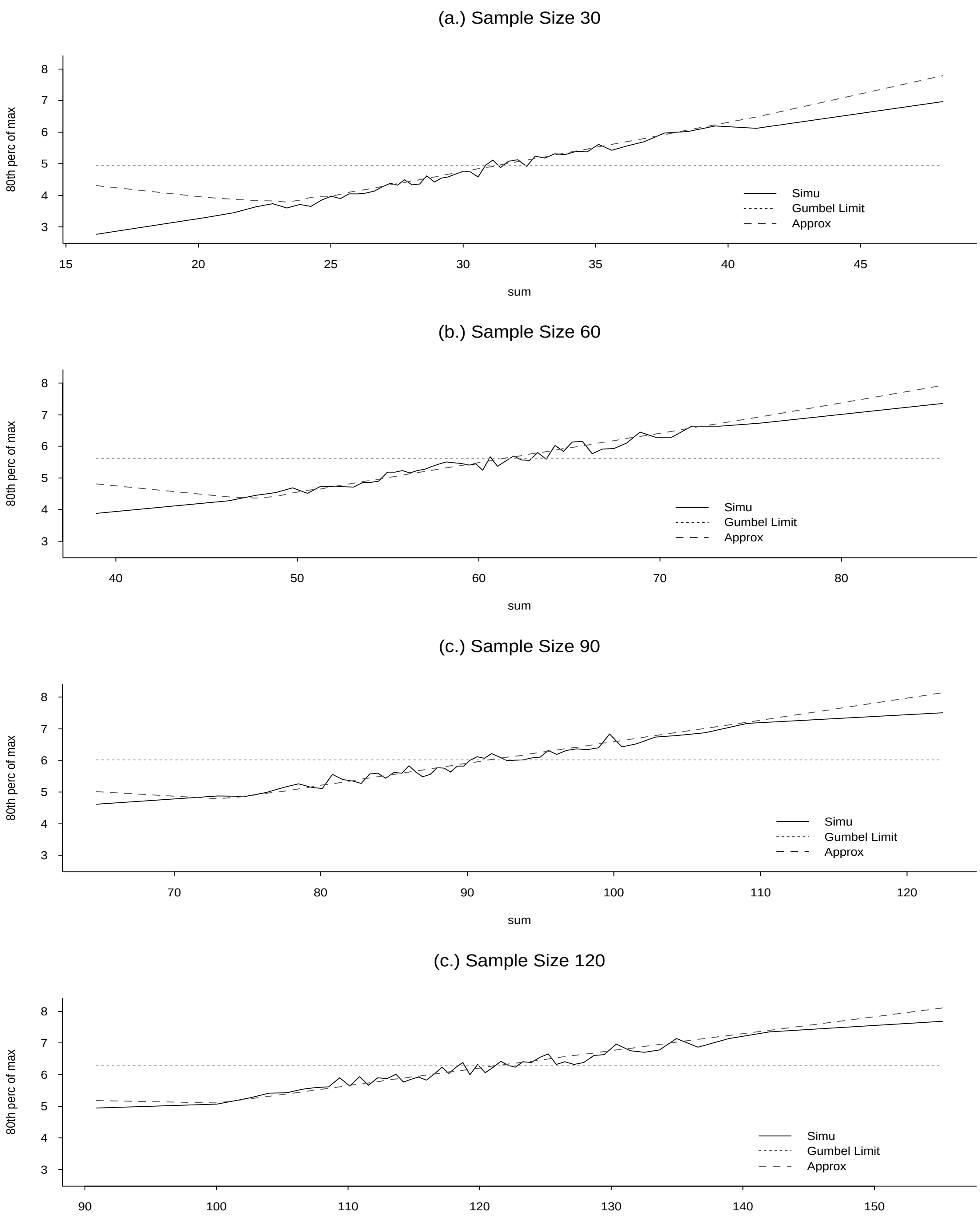


Figure 2.6: Effects of sample size on expansion of joint density. Weibull distribution with $\alpha = \beta = 1$ [Exponential distribution] as sample size increases from: (a) 30, (b) 60, (c) 90, to (d) 120. The solid line represents the simulated data; the short dashed line represents the Gumbel limit; and the long dashed line represents the higher order approximation of the conditional density.

Chapter 3

EXPANSION OF THE JOINT DENSITY UNDER THE FRÉCHET DOMAIN OF ATTRACTION

3.1 Introduction

In this chapter we develop the joint density of the sum and maximum of an *iid* sequence of random variables when the underlying distribution lies in the Fréchet domain of attraction.

The structure on which the development is based is similar to that of the Gumbel case. Let $X_1, \ldots, X_n$ be an *iid* sequence of random variables with common distribution function F which has density f. We again assume the existence of the mean μ, variance σ^2, and also the third moment μ^3 and third cumulant κ_3.

We again define $S_n = \sum_{i=1}^n X_i$ with the normalized version as

$$S_n^* = \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \tag{3.1}$$

and $M_n = \max_{1 \leq i \leq n} X_i$ with the normalized version being $M_n^* = \frac{M_n - b_n}{a_n}$ where $a_n > 0, b_n$ real.

Recall the notation for the distribution function of S_n^* is $F_{S_n^*}(w)$ with density $f_{S_n^*}(w) = dF_{S_n^*}(w)/dw$. Also the definition of the distribution function of M_n^* is $F_{M_n^*}(v) = F^n(a_n v + b_n)$ with density

$$f_{M_n^*}(v) = dF_{M_n^*}(v)/dv = nF^{n-1}(a_n v + b_n)f(a_n v + b_n)a_n.$$

Like Chapter 2, throughout this chapter since we are assuming a finite variance, we have that F lies in the domain of attraction of a stable law with index equal to 2 so that $F_{S_n^*}(w)$ converges to $\mathcal{N}(w)$ where $\mathcal{N}(w)$ denotes the normal distribution function. Recall the normal density is denoted by $\mathcal{N}'(w)$. In fact, we utilize Proposition 50 developed in Chapter 2 directly in this chapter. Again we use the result from Petrov (1975) which says that we can bound $|x|^m\{f_{S_n^*}(x) - \mathcal{N}'(x)\}$ uniformly $\forall x$ when $m \leq 3$.

For this chapter, we also assume that F lies in the Fréchet domain of attraction. We let Φ_α denote the Fréchet distribution and Φ'_α denotes its density where

$$\Phi_\alpha(x) = \begin{cases} e^{-x^{-\alpha}} & x > 0, \quad \alpha > 0 \\ 0 & \text{otherwise} \end{cases}$$

and

$$\Phi'_\alpha(x) = \begin{cases} \alpha x^{-\alpha-1}e^{-x^{-\alpha}} & x > 0, \quad \alpha > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Note since we are assuming that the underlying distribution has a finite third moment, we are in fact assuming that $\alpha \geq 3$ in the Fréchet formulae.

Recall from Chapter 1, F lies in the Fréchet domain of attraction when $1 - F(x)$ is regularly varying with index α where $\alpha > 0$. Thus we may write $1 - F(x)$ as $x^{-\alpha}\mathcal{L}(x)$ where $\mathcal{L}(x)$ is a slowly varying function. In fact, to establish the higher order terms associated with the expansion of the density of the maximum we assume that $\mathcal{L}(x)$ satisfies a slow variation with remainder condition. Moreover, since we want to develop an expansion for the density, we write $f(x) = x^{-\alpha-1}\mathcal{L}(x)$ where we assume $\mathcal{L}(x)$ satisfies the SR2 condition defined by Smith (1982). Recall the function $\mathcal{L}$ is said to satisfy the SR2 condition if $\mathcal{L}$ and g are two functions defined on $(0, \infty)$ where we assume $g(t) \to 0$

as $t \to \infty$, $\mathcal{L}$ is measurable, and

$$\frac{\mathcal{L}(\lambda x)}{\mathcal{L}(x)} - 1 \sim g(x)\nu(\lambda), \quad \lambda > 0, x \to \infty$$

If SR2 holds and ν satisfies the condition: There exists a λ such that $\nu(\lambda) \neq 0$ and $\forall y$ $\nu(\lambda y) - \nu(y) \neq 0$, then g is regularly varying with index ρ, i.e.

$$\lim_{t\to\infty} \frac{g(t\lambda)}{g(t)} = \lambda^{\rho} \quad \lambda > 0 \text{ for some } \rho \leq 0.$$

and $\nu(\lambda) = ch_{\rho}(\lambda)$ for some constant c and $h_{\rho}(\lambda) = \int_1^{\lambda} u^{\rho-1} du$.

Without loss of generality, we take the constant c to be equal to 1 both here and in Chapter 4. In practice, this constant is absorbed into the function g.

We can take the normalizing constants of M_n under the Fréchet domain of attraction to be

$$b_n \equiv 0 \tag{3.2}$$

and

$$a_n \text{ such that } 1 - F(a_n) = \tfrac{1}{n}. \tag{3.3}$$

We again solve for two forms of the expansion for the joint density. One is of the form of (2.2) from Chapter 2. The other is similar to (2.6) of Chapter 2 with Λ' replaced by Φ'. Like Chapter 2, the derivation starts by rewriting $f_{S_n^*,M_n^*}(w,v) = f_{S_n^*|M_n^*}(w|v) f_{M_n^*}(v)$ where $f_{S_n^*|M_n^*}(w|v)$ is the conditional density of S_n^* given M_n^*. Again we need the three key expansions. The first is the expansion for the conditional density of $S_n^*|M_n^*$ which we have already established in Chapter 2. Now under the Fréchet domain of attraction, we need to derive the expansion for the density of M_n^*. Finally, we need the expansions for the conditional mean and variance of $S_n|M_n$ under the Fréchet domain of attraction. Thus the main propositions of this chapter are the expansions for the density of M_n^* and for the conditional mean and variance of $S_n|M_n$ under the Fréchet domain of attraction.

Again which value M_n is conditioned upon is important to this derivation. Set $M_n = u_n = a_n v + b_n$ with v is fixed and where u_n is defined as a threshold with a_n and b_n defined in (3.3) and (3.2). Thus $M_n = u_n = a_n v$ or $M_n^* = M_n/a_n$.

Recall we can rewrite the distribution of $S_n|M_n$ in terms of $\sum_{i=1}^{n-1} X_i^* + u_n$ where the X^*s are *iid* random variables which have distribution function $\tilde{F_{u_n}}(x) = F(x)/F(u_n)$, and density $\tilde{f_{u_n}}(x) = f(x)/F(u_n)$. We also define $\mu(u_n)$ and $\sigma^2(u_n)$ as its mean and variance.

Recall the distributional relationship between $\tilde{S_n}$ and $S_n^*|M_n^*$

$$P[\tilde{S_n} \leq x] = P[\frac{n\mu + S_n^*\sqrt{n\sigma^2} - [(n-1)\mu(u_n) + u_n]}{\sqrt{(n-1)\sigma^2(u_n)}} \leq x|M_n^* = v] \tag{3.4}$$

with the Jacobian of the transformation as $\frac{\sqrt{n\sigma^2}}{\sqrt{(n-1)\sigma^2(u_n)}}$.

Recall from Chapter 2 that in deriving the expansions for the conditional mean and variance, we first need to solve the mean and variance of an exceedance above a threshold. Let us define Y as the exceedance over the threshold u_n; that is, $Y = X - u_n$. The conditional distribution function of Y given $X > u_n$ is

$$F_{u_n}(y) = P(X \leq u_n + y|X > u_n) = \frac{F(u_n + y) - F(u_n)}{1 - F(u_n)}.$$

Note that the tail distribution function is $1 - F_{u_n}(y) = \frac{1-F(u_n+y)}{1-F(u_n)}$.

Define $m(u_n)$ and $s^2(u_n)$ as the conditional mean and variance of Y given $X > u_n$. Note we have $E(X|X > u_n) = m(u_n) + u_n$ and $Var(X|X > u_n) = s^2(u_n) = Var(Y|X > u_n)$.

The outline of this chapter is as follows. The main theorem and corollary are presented in Section 3.2. Section 3.3 contains the propositions necessary in establishing the main result – (1) the expansion of the condition density of $\tilde{S_n}$, (2) the expansion of the density of M_n^*, and (3) the expansions for the mean and variance of $S_n|M_n = u_n$. Since the first expansion has already been derived in Chapter 2, in this chapter, only the latter two propositions and their corollaries are presented. Finally, Section 3.4 contains the proofs of the main theorem and its corollary.

3.2 Main Theorem

Here we present the main theorem and its corollary in the same format as in Chapter 2.

Theorem 54 *Let $X_1, \ldots, X_n$ be an iid sequence of random variables with distribution function F, density function f, characteristic function φ, mean μ, and variance σ^2. Let u_n be a threshold and φ_{u_n} be the characteristic function of $X|X < u_n$.*

Given the following two sets of assumptions

Set A: Assume f' is integrable, μ_3 exists, φ''' exists and is continuous in a neighborhood of 0, and $|\varphi_{u_n}(t)|^n$ is integrable for some $n \geq n^* > 1$.

Set B: Assume

$$f(x) = x^{-\alpha-1}\mathcal{L}(x), \quad \alpha > 1 \quad \textit{where} \tag{3.5}$$

$$\frac{\mathcal{L}(\lambda x)}{\mathcal{L}(x)} = 1 + h_\rho(\lambda)g(x) + o(g(x)), \quad \lambda \geq 1, \quad \textit{and where}$$

$$g \in \mathcal{R}_\rho \quad \textit{for some} \quad \rho < 0 \quad \textit{and} \quad h_\rho(\lambda) = \frac{\lambda^\rho - 1}{\rho}. \tag{3.6}$$

Also assume that

$$\frac{|f'(y)|F(y)}{f^2(y)} \leq n - 1, \quad \forall y \leq a_n e_n. \tag{3.7}$$

where $e_n = \{-\gamma \log g(a_n)\}^{-1/\alpha}$ for some $\gamma > 1$.
Define the normalizing constants of M_n as $b_n \equiv 0$ and a_n such that $1 - F(a_n) = \frac{1}{n}$ so that $u_n = a_n v$ and that

$$ng(a_n) \to \infty, \quad \textit{as } n \to \infty. \tag{3.8}$$

Define

$$r_n = \frac{a_n}{\sqrt{n\sigma^2}}. \tag{3.9}$$

Then,

$$|f_{S_n^*, M_n^*}(w, v) - f_{S_n^*}(w) f_{M_n^*}(v)\{1 - r_n v(\frac{\alpha}{\alpha - 1} v^{-\alpha} - 1)w\}| = o(r_n) \tag{3.10}$$

uniformly $\forall w$ and $\forall v \geq e_n$ where $e_n = \{-\gamma \log g(a_n)\}^{-1/\alpha}$ for some $\gamma > 1$.

Corollary 55 *Given the conditions of Theorem 54, then uniformly* $\forall w$ *and* $\forall v \geq e_n$ *where* $e_n = \{-\gamma \log g(a_n)\}^{-1/\alpha}$ *for some* $\gamma > 1$

$$
\begin{aligned}
&|f_{S_n^*,M_n^*}(w,v) - \\
&\mathcal{N}'(w)\Phi'_\alpha(v)\{1 + [h_\rho(v)(1 - \frac{\alpha}{\alpha-\rho}v^{-\alpha}) - \frac{1}{\alpha-\rho}]g(a_n)\}\{1 - r_n v(\frac{\alpha}{\alpha-1}v^{-\alpha} - 1)w\}| \\
&\qquad = o(\max\{r_n, g(a_n)\}). \qquad (3.11)
\end{aligned}
$$

3.3 Propositions

Here we present the main propositions and their corollaries of this chapter. Along with Proposition 50 of Chapter 2, they contain the fundamental parts necessary for establishing the expansion of the joint density of S_n^* and M_n^* for the Fréchet case.

3.3.1 Expansion of the Conditional Density of the Sum Given the Maximum

We do not need to repeat the result because all the conditions needed for Proposition 50 in Chapter 2 also apply in this chapter.

3.3.2 Expansion of the Density of the Maximum

The basis of this proposition is Smith (1982) which established an expansion for the distribution function of M_n^* under the Fréchet and Weibull domains of attractions. We also rely on Goldie and Smith (1987), particularly their Proposition 2.5.1 which gives integral forms of the SR2 conditions.

Proposition 56 *Suppose*

$$f(x) = x^{-\alpha-1}\mathcal{L}(x), \quad \alpha > 1 \quad \textit{and}$$

$$\frac{\mathcal{L}(\lambda x)}{\mathcal{L}(x)} = 1 + h_\rho(\lambda)g(x) + o(g(x)), \quad \lambda \geq 1, \quad \textit{where}$$
$$g \in \mathcal{R}_\rho \quad \textit{for some} \quad \rho < 0 \quad \textit{and} \quad h_\rho(\lambda) = \frac{\lambda^\rho - 1}{\rho}.$$

Also assume that

$$\frac{|f'(y)|F(y)}{f^2(y)} \leq n-1, \quad \forall y \leq a_n e_n.$$

where $e_n = \{-\gamma \log g(a_n)\}^{-1/\alpha}$ *for some* $\gamma > 1$.

Define the normalizing constants of M_n *as* $b_n \equiv 0$ *and* a_n *such that* $1 - F(a_n) = \frac{1}{n}$ *and that*

$$ng(a_n) \to \infty, \quad \textit{as } n \to \infty.$$

Then

$$|f_{M_n^*}(v) - \alpha v^{-\alpha-1}e^{-v^{-\alpha}}\{1 + [h_\rho(v)(1 - \frac{\alpha}{\alpha-\rho}v^{-\alpha}) - \frac{1}{\alpha-\rho}]g(a_n)\}| = o(g(a_n)) \quad (3.12)$$

uniformly on $v > 0$.

REMARK: Assumption (3.7) allows us to use $-\log F(x)$ and $1 - F(x)$ interchangeably; that is, the difference between them will be of smaller magnitude than the other error terms that are being considered.

PROOF: Let $M_n^* = \frac{M_n}{a_n}$ with density

$$f_{M_n^*}(v) = nF^{n-1}(a_n v)f(a_n v)a_n. \quad (3.13)$$

We begin by defining $\mathcal{L}_1(x) = x^\alpha\{1 - F(x)\}$ and showing

$$\frac{\mathcal{L}_1(\lambda x)}{\mathcal{L}_1(x)} = 1 + \frac{\alpha}{\alpha-\rho}h_\rho(\lambda)g(x) + o(g(x)), \quad \lambda \geq 1, \quad (3.14)$$

where the functions – h_ρ and g – and the constants – α and ρ – are defined in (3.5) and (3.6). Note (3.14) is of the same form as (3.6) where $g(x)$ has been replaced by

$\frac{\alpha}{\alpha-\rho}g(x)$; that is, $\mathcal{L}_1$ is also SR2. We can see in (3.14) why there is no loss of generality in setting $c = 1$ in the definition of SR2. This has no impact on the results. To see the relationship of $\frac{\mathcal{L}_1(\lambda x)}{\mathcal{L}_1(x)}$ we use Proposition 2.5.1 of Goldie and Smith (1987).

Since $\mathcal{L}$ satisfies (3.6) and $\int_1^\infty |\lambda^{-\alpha-1}| d\lambda < \infty$ for $\alpha > 1$, then by Proposition 2.5.1 of Goldie and Smith (1987) we have

$$\int_1^\infty \lambda^{-\alpha-1} \frac{\mathcal{L}(\lambda x)}{\mathcal{L}(x)} d\lambda = \frac{1}{\alpha} + \frac{g(x)}{\alpha(\alpha-\rho)} + o(g(x)). \tag{3.15}$$

From (3.15), we have

$$\begin{aligned} \frac{\mathcal{L}_1(x)}{\mathcal{L}(x)} &= \frac{x^\alpha}{\mathcal{L}(x)} \int_x^\infty f(t)dt = \frac{x^\alpha}{\mathcal{L}(x)} \int_x^\infty t^{-\alpha-1}\mathcal{L}(t)dt \\ &= \int_1^\infty \lambda^{-\alpha-1} \frac{\mathcal{L}(\lambda t)}{\mathcal{L}(t)} d\lambda \\ &= \frac{1}{\alpha} + \frac{g(x)}{\alpha(\alpha-\rho)} + o(g(x)), \quad \alpha > 1. \end{aligned} \tag{3.16}$$

Now using (3.16) and the fact that $g \in \mathcal{R}_\rho$, then

$$\frac{\mathcal{L}_1(\lambda x)}{\mathcal{L}(\lambda x)} = \frac{1}{\alpha} + \frac{\lambda^\rho g(x)}{\alpha(\alpha-\rho)} + o(g(x)) \quad \lambda \geq 1, \quad \alpha > 1. \tag{3.17}$$

Also by inverting (3.16),

$$\frac{\mathcal{L}(x)}{\mathcal{L}_1(x)} = \alpha[1 - \frac{g(x)}{(\alpha-\rho)}] + o(g(x)), \quad \alpha > 1. \tag{3.18}$$

Thus with (3.16), (3.17), and (3.18) we have

$$\begin{aligned} \frac{\mathcal{L}_1(\lambda x)}{\mathcal{L}_1(x)} &= \frac{\mathcal{L}_1(\lambda x)}{\mathcal{L}(\lambda x)} \frac{\mathcal{L}(x)}{\mathcal{L}_1(x)} \frac{\mathcal{L}(\lambda x)}{\mathcal{L}(x)} \\ &= \{1 + \frac{\lambda^\rho g(x)}{(\alpha-\rho)}\}\{1 - \frac{g(x)}{(\alpha-\rho)}\}\{1 + h_\rho(\lambda)g(x)\} + o(g(x)) \\ &= 1 + \frac{\alpha}{\alpha-\rho} h_\rho(\lambda)g(x) + o(g(x)), \quad \lambda \geq 1, \quad \alpha > 1. \end{aligned} \tag{3.19}$$

This proves (3.14).

Now we break up the proof into three sections. First we derive functions E_1, E_2, E_3 and E_4 so that we may write

$$f_{M_n^*}(v) = \alpha v^{-\alpha-1} e^{-v^{-\alpha}} \left\{1 + \left[h_\rho(v)(1 - \frac{\alpha}{\alpha-\rho} v^{-\alpha}) - \frac{1}{\alpha-\rho}\right] g(a_n)\right\} + \sum_{j=1}^{4} E_j. \quad (3.20)$$

Second we show (3.12) holds by showing that E_1, E_2, E_3 and E_4 are $o(g(a_n))$ uniformly on $v \geq e_n$ where $e_n = \{-\gamma \log g(a_n)\}^{-1/\alpha}$ for some $\gamma > 1$. Finally we show on the interval $v < e_n$ that each term in (3.12) is uniformly $o(g(a_n))$.

Section 1 We approximate (3.13) in two parts. First we find the approximation for $F^{n-1}(a_n v)$ and then for $a_n n f(a_n v)$.

If (3.5) and (3.6) hold then by Lemma 3 of Smith (1982), given an $\epsilon > 0$ there exists a x_ϵ such that if $x \geq x_\epsilon$, $\lambda x \geq x_\epsilon$

$$|\log\{\frac{\mathcal{L}(\lambda x)}{\mathcal{L}(x)}\} - h_\rho(\lambda) g(x)| \leq \epsilon g(x) \lambda^\beta \quad (3.21)$$

where $\beta = \rho - \epsilon < 0$ since we are assuming that $\rho < 0$.

Let $\theta_i, i = 1, 2, \ldots, 12$ be generic constants between -1 and 1. Using the Taylor expansion

$$e^y = 1 + y e^{\theta y}, \quad \text{where } \theta \in (-1, 1) \quad (3.22)$$

and (3.21), we can write

$$\begin{aligned}
&\left(\exp\{-\epsilon\lambda^\beta g(x)\} - 1\right) + h_\rho(\lambda) g(x) \left(\exp\{\theta_1 h_\rho(\lambda) g(x) - \epsilon\lambda^\beta g(x)\} - 1\right) \\
&\quad \leq \frac{\mathcal{L}(\lambda x)}{\mathcal{L}(x)} - 1 - h_\rho(\lambda) g(x) \leq \\
&\qquad \left(\exp\{\epsilon\lambda^\beta g(x)\} - 1\right) + h_\rho(\lambda) g(x) \left(\exp\{\theta_1 h_\rho(\lambda) g(x) + \epsilon\lambda^\beta g(x)\} - 1\right).
\end{aligned}$$

Thus if $x \geq x_\epsilon$, $\lambda x \leq x_\epsilon$

$$\begin{aligned}
&\frac{\mathcal{L}(\lambda x)}{\mathcal{L}(x)} - 1 - h_\rho(\lambda) g(x) = \\
&\left(\exp\{\epsilon\theta_2\lambda^\beta g(x)\} - 1\right) + h_\rho(\lambda) g(x) \left(\exp\{\theta_1 h_\rho(\lambda) g(x) + \theta_3\epsilon\lambda^\beta g(x)\} - 1\right). \quad (3.23)
\end{aligned}$$

Note we can get a similar equation for $\mathcal{L}_1(\lambda x)/\mathcal{L}_1(x)$ by using the definition of $\mathcal{L}_1$ in (3.17) and applying Lemma 3 of Smith (1982). We get (3.23) with the same x_ϵ but now $g(x)$ is replaced by $g_1(x) = \frac{\alpha g(x)}{\alpha-\rho}$; that is,

$$\frac{\mathcal{L}_1(\lambda x)}{\mathcal{L}_1(x)} - 1 - h_\rho(\lambda) g_1(x) = \Big(\exp\{\epsilon\theta_{10}\lambda^\beta g_1(x)\} - 1\Big) + h_\rho(\lambda) g_1(x) \Big(\exp\{\theta_9 h_\rho(\lambda) g_1(x) + \theta_{11}\epsilon\lambda^\beta g_1(x)\} - 1\Big) \tag{3.24}$$

Another Taylor expansion gives $\log(1-y) + y = \frac{-\theta_4 y^2}{1-\theta_4 y}$, $\forall y < 1$.

Apply $y = 1 - F(a_n v)$,

$$-\log F(a_n v) = 1 - F(a_n v) + \frac{\theta_4(1-F(a_n v))^2}{1-\theta_4(1-F(a_n v))}.$$

Thus

$$-(n-1)\log F(a_n v) = \left(1 - \frac{1}{n}\right)\frac{v^{-\alpha}\mathcal{L}_1(a_n v)}{\mathcal{L}_1(a_n)} + (n-1)\frac{\theta_4(1-F(a_n v))^2}{1-\theta_4(1-F(a_n v))}.$$

Now substituting in (3.24) with $\lambda = v$ and $x = a_n$ (and labeling x_ϵ as v_ϵ),

$$-(n-1)\log F(a_n v) = v^{-\alpha} + v^{-\alpha} h_\rho(v) g_1(a_n) + R_n, \tag{3.25}$$

where

$$\begin{aligned} R_n &= -\frac{v^{-\alpha}}{n} - \frac{v^{-\alpha} h_\rho(v) g_1(a_n)}{n} + (1-\frac{1}{n})v^{-\alpha}\Big(\exp\{\epsilon\theta_{10} v^\beta g_1(a_n)\} - 1\Big) \\ &+ (1-\frac{1}{n})v^{-\alpha} h_\rho(v) g_1(a_n)\Big(\exp\{\theta_9 h_\rho(\lambda) g_1(a_n) + \theta_{11}\epsilon\lambda^\beta g_1(a_n)\} - 1\Big) \\ &+ (n-1)\frac{\theta_4(1-F(a_n v))^2}{1-\theta_4(1-F(a_n v))} \end{aligned} \tag{3.26}$$

on $a_n v \geq v_\epsilon$.

Another application of (3.22) yields

$$\exp\{-v^{-\alpha} h_\rho(v) g_1(a_n)\} = 1 - v^{-\alpha} h_\rho(v) g_1(a_n) \exp\{\theta_5 v^{-\alpha} h_\rho(v) g_1(a_n)\}. \tag{3.27}$$

With (3.25) and (3.27), we have on $a_n v \geq v_\epsilon$

$$F^{n-1}(a_n v) = e^{-v^{-\alpha}}\left(1 - v^{-\alpha} h_\rho(v) g_1(a_n) e^{\theta_5 v^{-\alpha} h_\rho(v) g_1(a_n)}\right) e^{-R_n}. \tag{3.28}$$

Define

$$U_n = \left(e^{\theta_5 v^{-\alpha} h_\rho(v) g_1(a_n)} - 1\right) h_\rho(v) g_1(a_n). \tag{3.29}$$

Thus on $a_n v \geq v_\epsilon$

$$F^{n-1}(a_n v) = e^{-v^{-\alpha}} \left(1 - v^{-\alpha} h_\rho(v) g_1(a_n) + U_n\right) e^{-R_n}. \tag{3.30}$$

Now for the second term in $f_{M_n^*}(v)$

$$a_n n f(a_n v) = \frac{a_n f(a_n v)}{1 - F(a_n)} = \alpha v^{-\alpha-1} \frac{\mathcal{L}(a_n)}{\alpha \mathcal{L}_1(a_n)} \frac{\mathcal{L}(a_n v)}{\mathcal{L}(a_n)}.$$

For $\frac{\mathcal{L}(a_n)}{\alpha \mathcal{L}_1(a_n)}$ we substitute in a_n for x in (3.18). For the last term we use (3.23) with $\lambda = v$ and $x = a_n$. Thus for $a_n v \geq v_\epsilon$

$$\begin{aligned} a_n n f(a_n v) &= \alpha v^{-\alpha-1}[1 - \frac{g(a_n)}{(\alpha-\rho)} + o(g(a_n))] \\ &\times \left(1 + h_\rho(v) g(a_n) + \{e^{\epsilon\theta_7 v^\beta g(a_n)} - 1\} + h_\rho(v) g(a_n)\{e^{\theta_6 h_\rho(v) g(a_n) + \theta_8 \epsilon v^\beta g(a_n)} - 1\}\right) \\ &= \alpha v^{-\alpha-1}\{1 - \frac{g(a_n)}{(\alpha-\rho)} + S_n\}\{1 + h_\rho(v) g(a_n) + T_n\} \end{aligned} \tag{3.31}$$

where

$$S_n = o(g(a_n))) \tag{3.32}$$

and

$$T_n = \{e^{\epsilon\theta_7 v^\beta g(a_n)} - 1\} + h_\rho(v) g(a_n)\{e^{\theta_6 h_\rho(v) g(a_n) + \theta_8 \epsilon v^\beta g(a_n)} - 1\}. \tag{3.33}$$

By multiplying (3.30) and (3.31) and rearranging the terms, we have on $a_n v \geq v_\epsilon$

$$f_{M_n^*}(v) = \alpha v^{-\alpha-1} e^{-v^{-\alpha}} \left\{1 - v^{-\alpha} h_\rho(v) g_1(a_n) - \frac{g(a_n)}{\alpha-\rho} + h_\rho(v) g(a_n)\right\} + \sum_{j=1}^{4} E_j \tag{3.34}$$

where

$$\begin{aligned} E_1 = \alpha v^{-\alpha-1} e^{-v^{-\alpha}} &\{1 - v^{-\alpha} h_\rho(v) g_1(a_n) + U_n\}\{1 - \frac{g(a_n)}{\alpha-\rho} + S_n\} \\ &\times\{1 + h_\rho(v) g(a_n) + T_n\}\{e^{-R_n} - 1\}, \end{aligned} \tag{3.35}$$

$$E_2 = \alpha v^{-\alpha-1} e^{-v^{-\alpha}} U_n \{1 - \frac{g(a_n)}{\alpha - \rho} + S_n\}\{1 + h_\rho(v) g(a_n) + T_n\}, \tag{3.36}$$

$$E_3 = -\alpha v^{-\alpha-1} e^{-v^{-\alpha}} v^{-\alpha} h_\rho(v) g_1(a_n) \{-\frac{g(a_n)}{\alpha - \rho} + S_n + (h_\rho(v) g(a_n) + T_n)(1 - \frac{g(a_n)}{\alpha - \rho} + S_n)\}, \tag{3.37}$$

$$E_4 = \alpha v^{-\alpha-1} e^{-v^{-\alpha}} \{(-\frac{g(a_n)}{\alpha - \rho} + S_n)(h_\rho(v) g(a_n) + T_n) + S_n + T_n\}. \tag{3.38}$$

Now recall $e_n = \{-\gamma \log g(a_n)\}^{-1/\alpha}$ for some $\gamma > 1$. Note as $n \to \infty$, we have $e_n \to 0$. We also have $g(a_n) > \frac{1}{n}$ by (3.8). Using the definition of a_n and the Fréchet domain of attraction – i.e., $1 - F(x)$ is regularly varying with index $-\alpha$ – we have

$$a_n = O(n^{\frac{1}{\alpha}}). \tag{3.39}$$

Thus

$$a_n e_n > O(n^{\frac{1}{\alpha}} (\gamma \log n)^{-\frac{1}{\alpha}}) \to \infty$$

so the condition $a_n v \geq v_\epsilon$ is still satisfied. In other words,

$$v \geq e_n \Rightarrow a_n v \geq v_\epsilon$$

so we have completed Section 1; that is, we have (3.20) for $v \geq e_n$ (i.e. $a_n v \geq v_\epsilon$).

Section 2 We need to show that (3.12) holds uniformly on $v \geq e_n$. We first look at the error terms R_n, S_n, T_n, and U_n.

Note in what follows we use

$$e^y - 1 \sim y \quad \text{when } y \to 0. \tag{3.40}$$

In other words, we need in each application of (3.40) that the exponent goes to zero – i.e., $y \to 0$ – for the following relations for R_n, S_n, T_n, and U_n to hold.

For R_n, we need

$$\{1 - F(a_n v)\}^2 \quad = \quad \{1 - F(a_n)\}^2 \left\{\frac{1 - F(a_n v)}{1 - F(a_n)}\right\}^2 = \frac{1}{n^2}\left\{v^{-\alpha} \frac{\mathcal{L}_1(a_n v)}{\mathcal{L}_1(a_n)}\right\}^2$$

$$
\begin{aligned}
&= \frac{v^{-2\alpha}}{n^2}\{1 + h_\rho(v)g_1(a_n) + (\exp\{\epsilon\theta_{10}v^\beta g_1(a_n)\} - 1) \\
&\quad +h_\rho(v)g_1(a_n)(\exp\{\theta_9 h_\rho(v)g_1(a_n) + \theta_{11}\epsilon v^\beta g_1(a_n)\} - 1)\}^2 \\
&\sim \frac{v^{-2\alpha}}{n^2}\{1 + h_\rho(v)g_1(a_n) + \{\epsilon\theta_9 v^\beta g_1(a_n)\} \\
&\quad +h_\rho(v)g_1(a_n)\{\theta_{10}h_\rho(v)g_1(a_n) + \theta_{11}\epsilon v^\beta g_1(a_n)\}\}^2 \\
&\sim \frac{v^{-2\alpha}}{n^2}\{1 + o(1)\}.
\end{aligned}
$$

This gives us for $v \geq e_n$, (when the exponents in the exponentials above go to 0 – i.e. in (3.40), $y \to 0$)

$$
\begin{aligned}
R_n &\sim v^{-\alpha}\left(\epsilon\theta_{10}v^\beta g_1(a_n)\right) + v^{-\alpha}h_\rho(v)g_1(a_n)\left\{\theta_9 h_\rho(v)g_1(a_n) + \theta_{11}\epsilon v^\beta g_1(a_n)\right\} \\
&+ \frac{v^{-\alpha}}{n}\left[-1 - h_\rho(v)g_1(a_n)\right] \\
&+ \frac{v^{-\alpha}}{n}\left(\epsilon\theta_{10}v^\beta g_1(a_n) + h_\rho(v)g_1(a_n)\{\theta_9 h_\rho(v)g_1(a_n) + \theta_{11}\epsilon v^\beta g_1(a_n)\}\right) \\
&+ \frac{-\theta_4 v^{-2\alpha}}{n^2}\frac{(n-1)(1+o(1))}{1 - \theta_4\frac{v^{-\alpha}}{n}\{1+o(1)\}} \qquad (3.41)
\end{aligned}
$$

and

$$S_n \sim o(g(a_n)) \qquad (3.42)$$

and

$$T_n \sim \epsilon\theta_7 v^\beta g(a_n) + h_\rho(v)g(a_n)[\theta_6 h_\rho(v)g(a_n) + \theta_8\epsilon v^\beta g(a_n)] \qquad (3.43)$$

and, finally,

$$U_n \sim \theta_5 v^{-\alpha}h_\rho^2(v)g_1^2(a_n). \qquad (3.44)$$

Recall $g_1(a_n) = \frac{\alpha}{\alpha-\rho}g(a_n)$ and $g(a_n) \geq \frac{1}{n}$. We also have ϵ is arbitrary. On any finite range of v, in all applications of (3.40) in R_n, T_n, and U_n, we have the exponent $y \to 0$. Thus using (3.41), (3.42), (3.43), and (3.44), we see $|R_n|$, $|S_n|$, $|T_n|$, and $|U_n|$ are $o(g(a_n))$ for any finite range of v. This also gives us $(e^{-R_n} - 1) = o(g(a_n))$ for any finite range of v. Looking at (3.35), (3.36), (3.37), and (3.38), we have the leading term $(\alpha v^{-\alpha-1}e^{-v^{-\alpha}})$ is uniformly bound for any finite range of v and thus this gives us that E_1, E_2, E_3, and E_4 are each $o(g(a_n))$ for any finite range of v. In other words, (3.12) holds for compact sets of $(0, \infty)$.

Therefore we are left to check the limits as (a.) $v \to \infty$ and (b.) $v \to 0$ via e_n.

Case $v \to \infty$: Since we assume $\rho < 0$ we have $\beta < 0$ which gives us v^β and h_ρ are uniformly bounded (away from 0). We can apply (3.40). Looking at (3.41), (3.42), (3.43), and (3.44), we thus have $|R_n|, |S_n|, |T_n|$, and $|U_n|$ are $o(g(a_n))$. The same is then true for E_1, E_2, E_3, and E_4 since $\alpha v^{-\alpha-1}e^{-v^{-\alpha}}$ is bounded on $1 \le v < \infty$. Hence for some $K > 0$, independent of ϵ,

$$|f_{M_n^*}(v) - \alpha v^{-\alpha-1}e^{-v^{-\alpha}}\{1+[h_\rho(v)(1-\frac{\alpha}{\alpha-\rho}v^{-\alpha}) - \frac{1}{\alpha-\rho}]g(a_n)\}| \le K\epsilon g(a_n), \quad 1 < v < \infty \tag{3.45}$$

since ϵ is arbitrary, this establishes (3.12) on $1 < v < \infty$.

Case $v \to 0$ *via* e_n: Recall $v \ge e_n$ implies $a_n v \ge v_\epsilon$, so that we have (3.34) with (3.35) – (3.38). Recall $e_n^{-\alpha} = -\gamma \log g(a_n)$. Now for this interval ($v \ge e_n$), we have for any $\delta > 0$

$$sup_v v^{-\delta} g(a_n) = (-\gamma \log g(a_n))^{\delta/\alpha} g(a_n) \to 0 \text{ as } n \to \infty. \tag{3.46}$$

Again, we can use (3.40) to evaluate R_n, T_n, and U_n Thus we again can show that $|R_n|, |S_n|, |T_n|$, and $|U_n|$ are $o(g(a_n))$. For example to evaluate $|U_n|$, we have (recall $\rho < 0, \alpha > 2$)

$$\theta_5 v^{-\alpha} h_\rho(v) g_1(a_n) \to 0 \text{ uniformly on } v \ge e_n.$$

Hence using (3.46) we have

$$U_n \sim [\theta_5 v^{-\alpha} h_\rho(v) g_1(a_n)] h_\rho(v) g_1(a_n) = o(g(a_n)) \text{ uniformly on } v \ge e_n.$$

For $|S_n|$, we have already have uniformly in $v \ge e_n$.

For $|T_n|$ and $|R_n|$ we look at (3.43) and (3.41). Again we have ϵ is arbitrary and $\rho < 0$. Like the argument for $|U_n|$ at each application of (3.40), $y \to 0$, so

$$|T_n| = o(g(a_n)) \text{ and } |R_n| = o(g(a_n))$$

uniformly $v \geq e_n$. Thus with (3.46) ensuring $\alpha v^{-\alpha} e^{-v^{-\alpha}}$ is bounded on $v \geq e_n$, this again gives E_1, E_2, E_3, and E_4 are each $o(g(a_n))$ uniformly on $v \geq e_n$ and thus (3.12) on $v \geq e_n$.

Finally to complete the proof we need to extend the argument to $v < e_n$. To do so we show all terms in (3.12) are $o(g(a_n))$ for $v = e_n$ and then all terms are monotonically decreasing in the interval $v < e_n$.

Now (3.12) holds for $v = e_n$ and by definition of e_n we have

$$\alpha v^{-\alpha-1} e^{-v^{-\alpha}} \{1 + [h_\rho(v)(1 - \frac{\alpha}{\alpha - \rho} v^{-\alpha}) - \frac{1}{\alpha - \rho}] g(a_n)\} \tag{3.47}$$

is $o(g(a_n))$ at $v = e_n$. Since (3.47) is monotonic for sufficiently small v, it is $o(g(a_n))$ uniformly on $v < e_n$. From (3.12) and (3.47), we have $f_{M_n^*}(e_n) = o(g(a_n))$.

We have seen in the Gumbel chapter that

$$\frac{d}{dx} f_{M_n^*}(v) = n a_n^2 F^{n-2}(a_n v)\{(n-1) f^2(a_n v) + f'(a_n v) F(a_n v)\}. \tag{3.48}$$

Now (3.48) ≥ 0 for $v \geq e_n$ by (3.7). Thus $f_{M_n^*}(v)$ is monotonic increasing on $v < e_n$ so $f_{M_n^*}(v) = o(g(a_n)), \forall v < e_n$. □

REMARKS:

1. The uniformity result does not appear to hold when $\rho = 0$. For example, in T_n if $v > 1$ when $\rho = 0$, then $\beta > 0$ so

$$T_n \sim \epsilon \theta_7 v^\beta g(a_n) \to \infty, \quad \text{as } v \to \infty.$$

2. If assumption (3.8) fails because $|f'(y)|/f^2(y)$ becomes infinite at some finite y^*, we can redefine f on some interval $(-\infty, y^*)$. This will have no impact on the conclusions since $P[M_n \leq y^*] = F^n(y^*) = o(g(a_n))$ for any fixed y^*.

Corollary 57 *Given the conditions and set-up of Proposition 56, for any $m < 5$ and $\alpha > 2$*

$$|v^{-\alpha+m}\left\{f_{M_n^*}(v) - \alpha x^{-\alpha-1}e^{-v^{-\alpha}}\{1+[h_\rho(v)(1-\frac{\alpha}{\alpha-\rho}v^{-\alpha}) - \frac{1}{\alpha-\rho}]g(a_n)\}\right\}| = o(g(a_n)) \tag{3.49}$$

uniformly on $v \geq e_n$ where $e_n = \{-\gamma \log g(a_n)\}^{-1/\alpha}$ for some $\gamma > 1$. And

$$|v\left\{f_{M_n^*}(v) - \alpha x^{-\alpha-1}e^{-v^{-\alpha}}\{1+[h_\rho(v)(1-\frac{\alpha}{\alpha-\rho}v^{-\alpha}) - \frac{1}{\alpha-\rho}]g(a_n)\}\right\}| = o(g(a_n)) \tag{3.50}$$

uniformly on $v \geq e_n$ where $e_n = \{-\gamma \log g(a_n)\}^{-1/\alpha}$ for some $\gamma > 1$.

PROOF: The result follows from the proof of Proposition 56. If we focus on the interval $v \geq e_n$ which implies $a_n v \geq v_\epsilon$, we see that neither multiplying by $v^{-\alpha+m}$ nor by v affects the terms $R_n =$ (3.41), $S_n =$ (3.42), $T_n =$ (3.43), and $U_n =$(3.44) which are central to the uniformity result on this interval. The only detail to check when constructing E_1, E_2, E_3, and E_4 is that each would now be multiplied by $v^{-\alpha+m}$ or v. It then suffices to show that $v^{-\alpha+m}\{v^{-\alpha-1}e^{-v^{-\alpha}}\}$ and $v\{v^{-\alpha-1}e^{-v^{-\alpha}}\}$ remain bounded on the interval $v \geq e_n$.

Now, certainly $v^{-\alpha-2}e^{-v^{-\alpha}}$ is bounded since $\alpha > 2$. For $v^{-\alpha+m}\{v^{-\alpha-1}e^{-v^{-\alpha}}\}$ we need $-2\alpha + m - 1 < 0$. But again $\alpha > 2$ so this is bounded when $m > 5$. □

REMARKS

1. This corollary is important when proving a uniform bound for $|v^{-\alpha+m}f_{M_n^*}(v)|$, $m = 1, 2$, or 3 and for $|vf_{M_n^*}(v)|$ on $v \geq e_n$ in the main theorem.

2. If $\alpha \geq 3$, then $m < 7$ in (3.49) and v can be replaced by v^2 in (3.50).

3. Likewise we can prove the corollary when the leading term is $v^{-2\alpha+1}$. In fact, with the exponent of -2α we can let m be ≤ 6 in remark 1.

3.3.3 Expansions of conditional mean and variance

Given the SR2 conditions we have already established for $1-F(x)$, this proposition is primarily an application of Proposition 2.5.1 of Goldie and Smith (1987) which again solves integrals of SR2 conditions.

Proposition 58 *Suppose*

$$f(x) = x^{-\alpha-1}\mathcal{L}(x), \quad \alpha > 2 \quad \textit{and}$$

$$\frac{\mathcal{L}(\lambda x)}{\mathcal{L}(x)} = 1 + h_\rho(\lambda)g(x) + o(g(x)), \quad \lambda \geq 1, \quad \textit{where}$$
$$g \in \mathcal{R}_\rho \quad \textit{for some} \quad \rho < 0 \quad \textit{and} \quad h_\rho(\lambda) = \frac{\lambda^\rho - 1}{\rho}.$$

Then

$$\mu - \mu(u) \sim \frac{\alpha}{\alpha-1}u\{1-F(u)\}, \quad (\alpha > 1) \tag{3.51}$$

and

$$\sigma^2 - \sigma^2(u) \sim \frac{\alpha}{\alpha-2}u^2\{1-F(u)\}, \quad (\alpha > 2). \tag{3.52}$$

Proof: From the Gumbel chapter, we have

$$\mu - \mu(u) = \frac{1-F(u)}{F(u)}\{u + m(u) - \mu\} \tag{3.53}$$

Here we have

$$\begin{aligned} m(u) &= E(Y|X>u) = \int_0^\infty P[Y>y|X>u]dy = \int_0^\infty \frac{1-F(u+y)}{1-F(u)}dy \\ &= u\int_1^\infty \frac{1-F(ux)}{1-F(u)}dx \\ &= u\int_1^\infty x^{-\alpha}\frac{\mathcal{L}_1(ux)}{\mathcal{L}_1(u)}dx. \end{aligned}$$

Substituting in (3.19) for $\mathcal{L}_1(ux)/\mathcal{L}_1(u)$ and using Proposition 2.5.1 of Goldie and Smith (1987), we have

$$m(u) = u\{\frac{1}{\alpha - 1} + g(u)\frac{1}{(\alpha-1)(\alpha-\rho-1)} + o(g(u))\}, \quad \alpha > 1. \tag{3.54}$$

Substituting in (3.54) into (3.53) we have the result (3.51).

Also from Gumbel chapter

$$\sigma^2 - \sigma^2(u) = \mu^2(u) - \mu^2 + \left(\frac{1-F(u)}{F(u)}\right)\{(u+m(u))^2 + s^2(u) - \mu^2 - \sigma^2\}. \tag{3.55}$$

Now to calculate $s^2(u)$ we begin with $E(Y^2|X > u)$,

$$E(Y^2|X > u) = \int_0^\infty y^2 f_u(y)dy = 2u\int_1^\infty (x-1)x^{-\alpha}\frac{\mathcal{L}_1(ux)}{\mathcal{L}_1(u)}dx$$

Again we use formula (3.19) for $\mathcal{L}_1(ux)/\mathcal{L}_1(u)$ and Proposition 2.5.1 of Goldie and Smith (1987) with $\nu(x) = (x-1)x^{-\alpha}$. If $\alpha > 2$, we have

$$E(Y^2|X > u) = \\ 2u^2\left\{\frac{1}{(\alpha-1)(\alpha-2)} + \frac{(2\alpha-\rho-3)}{(\alpha-1)(\alpha-2)(\alpha-\rho-1)(\alpha-\rho-2)}g(u) + o(g(u))\right\}.$$

Thus

$$\begin{aligned} s^2(u) &= Var(Y|X > u) \qquad (3.56) \\ &= u^2\{\frac{\alpha}{(\alpha-1)^2(\alpha-2)} + \frac{(\alpha^2-\alpha-\rho-3)}{(\alpha-1)^2(\alpha-2)(\alpha-\rho-1)(\alpha-\rho-2)}g(u) + o(g(u))\}. \end{aligned}$$

From (3.51) we have

$$\mu^2 - \mu^2(u) \sim 2\mu u(\frac{\alpha}{\alpha-1})\{1-F(u)\} \tag{3.57}$$

We see the important terms of (3.55) involve $(u+m(u))^2$ and $s^2(u)$ and by using (3.54) and (3.57) in (3.55, we have the result – (3.52). □

Corollary 59 *Assume the condition of Proposition 58. Let the normalizing constants for M_n be defined as $b_n \equiv 0$ and a_n such that $1 - F(a_n) = \frac{1}{n}$ so that we may define $u_n = a_n v$. Also assume $ng(a_n) \to \infty$ as $n \to \infty$. Then for all $v \geq e_n$*

$$\mu - \mu(u) \sim \frac{\alpha}{\alpha - 1} a_n v \{\frac{v^{-\alpha}}{n}\}, \quad \alpha > 1 \tag{3.58}$$

and

$$\sigma^2 - \sigma^2(u) \sim \frac{\alpha}{\alpha - 2}(a_n v)^2 \{\frac{v^{-\alpha}}{n}\}, \quad \alpha > 2. \tag{3.59}$$

PROOF: These results fall from Proposition 58 with $u = a_n v$ when $1 - F(a_n v) = \frac{v^{-\alpha}}{n} + o(1)$, for $v \geq e_n$.

From (3.24) which is based on Lemma 3 of Smith (1982) we have

$$\frac{1 - F(a_n v)}{1 - F(a_n)} = v^{-\alpha}\{1 + h_\rho(v) g_1(a_n) + Q_n\}$$

where

$$Q_n = \left(e^{\epsilon\theta_{10} v^\beta g_1(a_n)} - 1\right) + h_\rho(v) g_1(a_n) \left(e^{\theta_9 h_\rho(v) g_1(a_n) + \theta_{11}\epsilon v^\beta g_1(a_n)} - 1\right).$$

Now as long as we are bounded away from 0, using the argument for establishing the error of R_n in (3.41) we have $Q_n = o(g_1(a_n)) = o(g(a_n))$, for $v \geq e_n$.

Thus

$$\frac{1 - F(a_n v)}{1 - F(a_n)} = v^{-\alpha}\{1 - h_\rho(v) g_1(a_n) + o(g(a_n))\}$$

or using the definition of $1 - F(a_n)$ and the definition of u

$$\begin{aligned} 1 - F(u) &= \frac{v^{-\alpha}}{n}\{1 - h_\rho(v) g_1(a_n) + o(g_1(a_n))\} \\ &= \frac{v^{-\alpha}}{n}\{1 + o(1)\} \end{aligned} \tag{3.60}$$

uniformly on $v \geq e_n$. Note $h_\rho(v) g_1(a_n)$ is uniformly bounded on $v \geq e_n$ by (3.46).

Substituting (3.60) into (3.51) and (3.52), we get (3.58) and (3.59). □

3.4 Proof of Main Theorem and Its Corollary

PROOF OF THEOREM 54: We can write the joint density of S_n^* and M_n^* as

$$\begin{aligned} f_{S_n^*,M_n^*}(w,v) &= f_{M_n^*}(v) f_{S_n^*|M_n^*}(w|v) \\ &= f_{M_n^*}(v)\sqrt{\frac{n\sigma^2}{(n-1)\sigma^2(u_n)}} f_{\tilde{S_n}}(z) \end{aligned} \tag{3.61}$$

where

$$z = \frac{n\mu + \sqrt{n\sigma^2}w - (n-1)\mu(u_n) - u_n}{\sqrt{(n-1)\sigma^2(u_n)}}. \tag{3.62}$$

.

Here we let $u_n = a_n v$. To enable the uniformity results we will allow v, w and hence z to be dependent on n. We suppress this so as to make the notation easier to read.

Note the transformation from S_n^* to $\tilde{S_n}$ and the form of (3.62) comes from (2.7) of Chapter 2.

Now to establish (3.10) we break it up as follows

$$\begin{aligned} &f_{M_n^*}(v)\sqrt{\frac{n\sigma^2}{(n-1)\sigma^2(u_n)}} f_{\tilde{S_n}}(z) - f_{S_n^*}(w) f_{M_n^*}(v)\{1 - r_n v(\frac{\alpha}{\alpha-1}v^{-\alpha} - 1)w\} \\ &= E_1 + E_2 + E_3 + E_4 + E_5 + E_{5b} \end{aligned} \tag{3.63}$$

where

$$E_1 = f_{M_n^*}(v) f_{\tilde{S_n}}(z)\left[\sqrt{\frac{n\sigma^2}{(n-1)\sigma^2(u_n)}} - 1\right] \tag{3.64}$$

$$E_2 = f_{M_n^*}(v)\left[f_{\tilde{S_n}}(z) - \mathcal{N}'(z)\{1 + \frac{\kappa_3(u_n)}{6\sigma^3(u_n)\sqrt{n}}(z^3 - 3z)\}\right] \tag{3.65}$$

$$E_3 = f_{M_n^*}(v)\left[\mathcal{N}'(z) - \mathcal{N}'(w)\{1 - r_n v(\frac{\alpha}{\alpha-1}v^{-\alpha} - 1)w\}\right] \tag{3.66}$$

$$\begin{aligned} E_4 &= f_{M_n^*}(v)\left[\mathcal{N}'(w)\{1 + \frac{\kappa_3}{6\sigma^3\sqrt{n}}(w^3 - 3w)\} - f_{S_n^*}(w)\right] \\ &\quad \{1 - r_n v(\frac{\alpha}{\alpha-1}v^{-\alpha} - 1)w\} \end{aligned} \tag{3.67}$$

$$E_5 = f_{M_n^*}(v)\left[\frac{\mathcal{N}'(z)\kappa_3(u_n)}{6\sigma^3(u_n)\sqrt{n}}(z^3 - 3z) - \frac{\mathcal{N}'(w)\kappa_3}{6\sigma^3\sqrt{n}}(w^3 - 3w)\right] \tag{3.68}$$

$$E_{5b} = f_{M_n^*}(v)\mathcal{N}'(w)\frac{\kappa_3}{6\sigma^3\sqrt{n}}(w^3 - 3w)[r_n v(\frac{\alpha}{\alpha-1}v^{-\alpha} - 1)w]. \tag{3.69}$$

Now to prove that (3.10) we show that (3.64) – (3.69) are $o(r_n)$ uniformly $\forall w$ and $v \geq e_n$. And to prove (3.64) – (3.69) are $o(r_n)$, it suffices to show that for any $\epsilon > 0$, $\sum_{j=1}^{6} |E_j| = \epsilon\, r_n$ for all sufficiently large n. It is necessary to consider two cases where the dependence on n for w, v, and z need to be explicitly expressed. These two cases are: Case (a) $|z_n - w_n| \leq \delta$ for a $\delta > 0$ and Case(b) $|z_n - w_n| > \delta$.

Proof for E_1 The following argument for E_1 holds irrespective of δ and so holds for both Case (a) and Case (b). We start with $E_1 = f_{M_n^*}(v) f_{\tilde{S}_n}(z) \left[\sqrt{\frac{n\sigma^2}{(n-1)\sigma^2(u_n)}} - 1\right]$. Specifically we begin with its third term.

From the Gumbel proof, we have

$$|\sqrt{\frac{n\sigma^2}{(n-1)\sigma^2(u_n)}} - 1| \leq \sqrt{\frac{\sigma^2}{\sigma^2(u_n)}} - 1 + O(\frac{1}{n}). \tag{3.70}$$

Now from (3.59), we have

$$\frac{\sigma^2(u_n)}{\sigma^2} = 1 - \frac{\alpha}{\alpha - 2} r_n^2 v^{-\alpha+2} + o(r_n^2 v^{-\alpha+2}).$$

Now $r_n^2 v^{-\alpha+2} \to 0, \forall v \geq e_n$ since

$$r_n^2 v^{-\alpha+2} \sim \frac{a_n^2 v^{-\alpha+2}}{n} \to 0. \tag{3.71}$$

To see this, we have for $v \geq e_n$ that $\frac{a_n^2 v^{-\alpha+2}}{n} = (a_n v)^2 \frac{v^{-\alpha}}{n} \sim u_n^2\{1 - F(u_n)\} \leq \int_{u_n}^{\infty} x^2 dF(x) \to 0$ since the variance is assumed finite. Since $\frac{a_n^2 v^{-\alpha+2}}{n} \to 0, \forall v \geq e_n$, we can invert (3.71) to get

$$\frac{\sigma^2}{\sigma^2(u_n)} - 1 = \frac{\alpha}{\alpha - 2} r_n^2 v^{-\alpha+2} + o(r_n^2 v^{-\alpha+2}). \tag{3.72}$$

Note that

$$\frac{\sigma^2}{\sigma^2(u_n)} = 1 + o(1) \text{ uniformly on } v \geq e_n \tag{3.73}$$

since $v > 0$ and $\alpha > 3$ since the third moment is assumed to exist.

Substituting (3.72) into (3.70), we establish the inequality

$$|\sqrt{\frac{n\sigma^2}{(n-1)\sigma^2(u_n)}}-1| \le K\left(\frac{\alpha}{\alpha-2}r_n^2 v^{-\alpha+2}+\frac{1}{n}\right), \quad \text{for some constant } K>0. \quad (3.74)$$

Thus

$$E_1 = O(f_{M_n^*}(v) f_{\tilde{S_n}}(z) v^{-\alpha+2} r_n^2).$$

Now by Proposition 50 in Chapter 2, $f_{\tilde{S_n}}(z)$ is uniformly bounded $\forall z$. By Corollary 57, $v^{-\alpha+2} f_{M_n^*}(v)$ is uniformly bounded on $v \ge e_n$.

Thus

$$E_1 = O(r_n^2) = o(r_n), \quad \forall z, v \ge e_n.$$

or we may write for sufficiently large n

$$|E_1| = \frac{\epsilon}{6}\, r_n, \forall w, \forall v \ge e_n. \quad (3.75)$$

Proof for E_2 Like proof for E_1, the following argument holds for both Case (a) and Case (b).

Next we look at $E_2 = f_{M_n^*}(v)\left[f_{\tilde{S_n}}(z) - \mathcal{N}'(z)\{1+\frac{\kappa_3(u_n)}{6\sigma^3(u_n)\sqrt{n}}(z^3-3z)\}\right]$.

Here, $\left[f_{\tilde{S_n}}(z) - \mathcal{N}'(z)\{1+\frac{\kappa_3(u_n)}{6\sigma^3(u_n)\sqrt{n}}(z^3-3z)\}\right] = o(\frac{1}{\sqrt{n}})$ uniformly in z by Proposition 50 of Chapter 2. We also have that $f_{M_n^*}(v)$ is bounded by Proposition 56. Hence

$$E_2 = o(\frac{1}{\sqrt{n}}) = o(r_n), \quad \forall v \text{ and } \forall w.$$

or we may write for sufficiently large n

$$|E_2| = \frac{\epsilon}{6}\, r_n, \forall w, \forall v. \quad (3.76)$$

Proof for E_3 Here we will proceed simultaneously with both until *Step 3* where at that point we will need to divide the proof into the two cases.

Recall

$$E_3 = f_{M_n^*}(v)\left[\mathcal{N}'(z) - \mathcal{N}'(w)\{1 - r_n v(\frac{\alpha}{\alpha-1}v^{-\alpha} - 1)w\}\right].$$

Establishing $E_3 = o(r_n)$ involves a longer argument than needed for E_1 or E_2. We thus break this argument into the following steps.

Step 1: Recall from the Gumbel case,

$$z - w = \frac{n(\mu - \mu(u_n))}{\sqrt{(n-1)\sigma^2(u_n)}} + \{\sqrt{\frac{n\sigma^2}{(n-1)\sigma^2(u_n)}} - 1\}w + \frac{\mu(u_n) - u_n}{\sqrt{(n-1)\sigma^2(u_n)}}. \tag{3.77}$$

Substitute in (3.58) into the first and third term and (3.74) into the second term,

$$\begin{aligned} z - w &= \frac{(\frac{\alpha}{\alpha-1})a_n v^{-\alpha+1}}{\sqrt{n\sigma^2}}\{1 + o(1)\} + O(wr_n^2 v^{-\alpha+2}) - \frac{a_n v}{\sqrt{n\sigma^2}}\{1 + o(1)\} \\ &= r_n v\left[(\frac{\alpha}{\alpha-1})v^{-\alpha} - 1\right] + o\{r_n v\left[(\frac{\alpha}{\alpha-1})v^{-\alpha} - 1\right]\} + O(r_n^2 v^{-\alpha+2}w) \end{aligned} \tag{3.78}$$

Step 2: We again take a Taylor expansion for $\mathcal{N}'(z)$ about w – i.e. write $z = w + t_n$ where t_n can be seen in (3.78). Recall

$$\mathcal{N}'(z) = \mathcal{N}'(w) - t_n z^* \mathcal{N}'(z^*).$$

for z^* between w and z.

Substituting this into E_3 we have

$$\begin{aligned} E_3 &= f_{M_n^*}(v)(z - w)\left[w\mathcal{N}'(w) - z^*\mathcal{N}'(z^*)\right] + o(r_n v(\frac{\alpha}{\alpha-1}v^{-\alpha} - 1)w\mathcal{N}'(w)f_{M_n^*}(v)) \\ &\quad + O(wv^{-\alpha+2}r_n^2 w\mathcal{N}'(w)f_{M_n^*}(v)) \\ &= E_6 + E_7 + E_8. \end{aligned}$$

Now $w\mathcal{N}'(w)$ and $w^2\mathcal{N}'(w)$ are uniformly bounded $\forall w$. From Corollary 57, we have $v^{-\alpha+1}f_{M_n^*}(v)$ and $v^{-\alpha+2}f_{M_n^*}(v)$ are uniformly bounded for $v \geq e_n$. Thus $E_7 = E_8 = o(r_n)$ on $v \geq e_n$ or we may write for sufficiently large n ($\epsilon > 0$)

$$|E_7| < \frac{\epsilon}{18}\, r_n,\ \forall w,\ \forall v \geq e_n, \tag{3.79}$$

$$|E_8| < \frac{\epsilon}{18}\, r_n,\ \forall w,\ \forall v \geq e_n. \tag{3.80}$$

Step 3: Now we focus on E_6. The important details in this formula concern $z - w$ since the other terms are bounded. Recall v, w, z, and z^* actually depend on n. So again fix the notation by writing $v = v_n, w = w_n, z = z_n$, and $z^* = z_n^*$ so the dependence on n is explicit.

By substituting (3.78) into E_6

$$\begin{aligned}
E_6 &= f_{M_n^*}(v_n)\left(r_n v_n\left[(\frac{\alpha}{\alpha-1})v_n^{-\alpha}-1\right] + o\{r_n v_n\left[(\frac{\alpha}{\alpha-1})v_n^{-\alpha}-1\right]\} + O(r_n^2 v_n^{-\alpha+2} w_n)\right) \\
&\quad \times \{w_n \mathcal{N}'(w_n) - z_n^* \mathcal{N}'(z_n^*)\} \\
&= r_n f_{M_n^*}(v_n) v_n \left((\frac{\alpha}{\alpha-1})v_n^{-\alpha}-1\right)\{w_n \mathcal{N}'(w_n) - z_n^* \mathcal{N}'(z_n^*)\} \\
&\quad + f_{M_n^*}(v_n) o\{r_n v_n\left[(\frac{\alpha}{\alpha-1})v_n^{-\alpha}-1\right]\}\{w_n \mathcal{N}'(w_n) - z_n^* \mathcal{N}'(z_n^*)\} \\
&\quad + f_{M_n^*}(v_n) O(r_n^2 v_n^{-\alpha+2} w_n)\{w_n \mathcal{N}'(w_n) - z_n^* \mathcal{N}'(z_n^*)\} \\
&= E_9 + E_{10} + E_{11}.
\end{aligned} \tag{3.81}$$

At this point it is necessary to separate the argument into the two cases.

Case (a) If $|z_n - w_n| \leq \delta$ then $|z_n^* - w_n| \leq \delta$. By uniform continuity of $w^k \mathcal{N}'(w)$ for $k = 0, 1, 2$ given $\epsilon > 0$, we can find a $\delta > 0$ such that

$$|z_n - w_n| < \delta \Rightarrow |z_n^k \mathcal{N}'(z_n) - w_n^k \mathcal{N}'(w_n)| < \frac{\epsilon}{C}$$

for $k = 0, 1, 2$ and any given constant $C > 0$.

Now since $(v_n^{-\alpha+2} + 1) f_{M_n^*}(v_n)$ is bounded on $v_n \geq e_n = \{-\gamma \log g(a_n)\}^{-1/\alpha}$ by Proposition 56 and Corollary 57, we have that each E_9, E_{10}, and E_{11} is bounded by some

$$\mathit{constant} \times |z_n^{*k} \mathcal{N}'(z_n^*) - w_n^k \mathcal{N}'(w_n)|$$

for $k = 0, 1, 2$.

In other words, we can choose a δ so that $\forall w_n$ and $\forall v_n \geq e_n$

$$|z_n - w_n| < \delta \Rightarrow |E_6| < \frac{\epsilon}{18}\, r_n \;\text{ for all sufficiently large } n\ (\epsilon > 0)$$

which with (3.79) and (3.80) gives

$$|E_3| < \frac{\epsilon}{6}\, r_n \;\text{ for all sufficiently large } n \tag{3.82}$$

for $v_n \geq e_n$ and $\forall w$ and when $|z_n - w_n| \leq \delta$. [End Case (a)]

Cases (b) Here we show that if $|z_n - w_n| > \delta$ then the entire left-hand side of (3.63) is $o(r_n)$.

<u>*Part 1*</u> Suppose $|z_n - w_n| > \delta$ and $|w_n r_n| \leq \delta^2$. From (3.78) we deduce either $|r_n v^{-\alpha+1}| >$ some $\delta_1 > 0$ for sufficiently large n ($v_n \downarrow 0$) or $|r_n v_n| >$ some δ_2 for sufficiently large n. ($v_n \uparrow \infty$).

Look at the right-hand side of (3.63) we see $\sqrt{\frac{n\sigma^2}{(n-1)\sigma^2(u_n)}} = 1 + o(1)$ is uniformly bounded on $v_n \geq e_n$ by (3.73). Also $f_{\tilde{S_n}}(z_n) = O(1)$ is uniformly $\forall z_n$ by Proposition 50 of Chapter 2. Finally, $f_{S_n}(w_n)$ and $w_n f_{S_n}(w_n)$ are uniformly bounded $\forall w$ by Petrov's (1975) Central Limit Theory result. So we need only look at terms associated with v_n in the right-hand side of (3.63).

When

$$|r_n v^{-\alpha+1}| > \delta_1 \Rightarrow |v_n| > \left(\frac{\delta_1}{r_n}\right)^{\frac{1}{-\alpha+1}}.$$

We need to show on this interval

(a) $|f_{M_n^*}(v_n)| = o(r_n)$.

(b) $|v_n f_{M_n^*}(v_n)| = o(1)$.

(c) $|v_n^{-\alpha+1} f_{M_n^*}(v_n)| = o(1)$.

As $n \to \infty$, v_n can go $\downarrow 0$ on this interval. Thus it suffices to show (c). From the Remarks following Corollary 57, we have

$$|v_n^{-2\alpha+1} f_{M_n^*}(v_n)| = O(1) \;\text{ or }\; |v_n^{-\alpha+1} f_{M_n^*}(v_n)| = O(v_n^{\alpha})$$

On $|v_n| > \left(\frac{\delta_1}{r_n}\right)^{\frac{1}{-\alpha+1}}$, we have

$$|v_n^{-\alpha+1} f_{M_n^*}(v_n)| = O(\delta_1^{\frac{\alpha}{-\alpha+1}} r_n^{\frac{-\alpha}{-\alpha+1}}) = o(r_n).$$

Thus the right-hand side of (3.63) is $o(r_n)$ on this interval.

When

$$|r_n v_n| > \delta_2 \Rightarrow |v_n| > \frac{\delta_2}{r_n}.$$

We again need to show that (a), (b), and (c) from above. As $n \to \infty$, v_n can go $\uparrow \infty$ on this interval so it suffices to show (a). By the remark following Corollary 57,

$$|f_{M_n^*}(v_n)| = O(v_n^{-2}).$$

On $|v_n| > \frac{\delta_2}{r_n}$, we thus have

$$|f_{M_n^*}(v_n)| = O(\delta_1^2 r_n^2) = o(r_n) = o(1).$$

Thus on this interval the right-hand side of (3.63) is $o(r_n)$.

<u>*Part 2*</u> Suppose $|z_n - w_n| \geq \delta$ and $|w_n r_n| > \delta^2$. (and $|r_n v_n| \leq \delta_1$ and $|r_n v_n^{-\alpha+1}| \leq \delta_2$). From (3.78) we deduce $|r_n w_n| >$ some $\delta_1 > 0$ for sufficiently large n. So $|w_n| > \frac{\delta_1}{r_n}$. We also have from (3.78)

$$\begin{aligned} z_n &= w_n + r_n v_n \left[(\frac{\alpha}{\alpha - 1}) v_n^{-\alpha} - 1 \right] + o\{ r_n v_n \left[(\frac{\alpha}{\alpha - 1}) v_n^{-\alpha} - 1 \right] \} + O(r_n^2 v_n^{-\alpha+2} w_n) \\ &= w_n + o(1) + o(1) + O(w_n o(1)) \\ &= w_n(1 + o(1)) \end{aligned}$$

for sufficiently large n.

Thus we have [with (3.73)]

1. $|f_{\tilde{S_n}}(z_n)| = o(r_n)$ by Proposition 50.

2. $|w_n|^k |f_{S_n^*}(w_n)| = o(r_n)$ for k=0,1 by Petrov's central limit theorem.

3. $|(e^{-v} + 1) f_{M_n^*}(v_n)|$ is bounded on $v_n \geq e_n$. by Proposition 56 and Corollary 57.

Hence the left-hand side of (3.63) is $o(r_n)$ on $v_n \geq e_n$. and $\forall w_n$ when $|z_n - w_n| > \delta$ and $|r_n w_n| > \delta^2$. [End of Case (b)]

Hence $\forall w$ and $\forall v \geq e_n$

$$E_6 = o(r_n) \quad \text{or entire (3.63)} = o(r_n).$$

In other words, for E_3 we have either for sufficiently large n ($\epsilon > 0$)

$$|E_3| < \frac{\epsilon}{56} r_n, \forall v_n \geq e_n, \forall w_n \tag{3.83}$$

of the left-hand side of (3.63) is $o(r_n)$.

Proof for E_4 Like the proofs for E_1 and E_2 this argument holds irrespective of δ and hence is the same for both Case (a) and Case (b).

Recall we have
$E_4 = f_{M_n^*}(v) \left[\mathcal{N}'(w)\{1 + \frac{\kappa_3}{6\sigma^3\sqrt{n}}(w^3 - 3w)\} - f_{S_n^*}(w)\right] \{1 - r_n v(\frac{\alpha}{\alpha-1}v^{-\alpha} - 1)w\}.$

Again we have $f_{M_n^*}(v)$ is bounded and also by Feller (1971), Chapter XVI, Section 2, Theorem 1 the term inside $[\ldots]$ is $o(\frac{1}{\sqrt{n}})$ uniformly in w. To handle the term $r_n v(\frac{\alpha}{\alpha-1}v^{-\alpha} - 1)w$, we need show

(a) $\sup_v v(\frac{\alpha}{\alpha-1}v^{-\alpha} - 1) f_{M_n^*}(v)$ is bounded on $v \geq e_n$ which we again have by Corollary 57.

(b) $\sup_w |w\{\mathcal{N}'(w) - f_{S_n^*}(w)\}| \to 0$ as $n \to \infty$ which follows from Petrov's (1975) central limit theorem under the same assumptions as Proposition 50 of Chapter 2.

(c.) $w^4\mathcal{N}'(w)$ to be bounded so $\mathcal{N}'(w)\frac{\kappa_3}{6\sigma^3\sqrt{n}}(w^3 - 3w)w \to 0$ uniformly in w which we have by properties of the normal density.

Thus

$$E_4 = o(r_n), \quad \forall v \geq e_n, \forall w.$$

or we may write for sufficiently large n ($\epsilon > 0$)

$$|E_4| < \frac{\epsilon}{6}\, r_n, \forall w, \forall v \geq e_n \tag{3.84}$$

Proof for E_5 Finally we look at the term

$$E_5 = f_{M_n^*}(v)\left[\frac{\mathcal{N}'(z)\kappa_3(u_n)}{6\sigma^3(u_n)\sqrt{n}}(z^3 - 3z) - \frac{\mathcal{N}'(w)\kappa_3}{6\sigma^3\sqrt{n}}(w^3 - 3w)\right].$$

Now, again we have $f_{M_n^*}(v)$ is uniformly bounded. We also have that the function $\mathcal{N}'(z)\{z^3 - 3z\}$ is uniformly continuous so by similar argument to the proof of E_3, particularly *Step 3*, we can conclude for sufficiently large n

$$|E_5| < \frac{\epsilon}{6} r_n \;\text{ or (3.63) is } o(r_n), \forall v_n \geq e_n, \forall w. \tag{3.85}$$

Proof of E_{5b} Like the proofs for E_1 and E_2 this argument holds irrespective of δ and hence is the same for both Case (a) and Case (b).

Recall we have $E_{5b} = f_{M_n^*}(v)\mathcal{N}'(w)\frac{\kappa_3}{6\sigma^3\sqrt{n}}(w^3 - 3w)[r_n v(\frac{\alpha}{\alpha-1}v^{-\alpha} - 1)w\}]$.

Now, we have the result necessary immediately since,

(a) $\sup_v v(\frac{\alpha}{\alpha-1}v^{-\alpha} - 1)f_{M_n^*}(v)$ is bounded on $v \geq e_n$ by Corollary 57.

$\sup_v(e^{-v} - 1)f_{M_n^*}(v)$ is bounded on $|v| \leq e_n$ by Corollary 52.

(b.) $w^4\mathcal{N}'(w)$ is bounded uniformly in w so $\mathcal{N}'(w)\frac{\kappa_3}{6\sigma^3}(w^3 - 3w)w$ is bounded uniformly in w which we have by properties of the normal density.

Thus we have

$$E_{5b} = O(\frac{r_n}{\sqrt{n}}) = o(r_n) \;\text{ for } v \geq e_n \text{ and } \forall w.$$

or we may write for sufficiently large n ($\epsilon > 0$)

$$|E_{5b}| < \frac{\epsilon}{6}\, r_n \quad \text{for } v \geq e_n \text{ and } \forall w. \tag{3.86}$$

In conclusion, using (3.75), (3.76), (3.83), (3.84), (3.85), and (3.86), we have shown for sufficiently large n $\sum_{j=1}^{6} \leq \epsilon\, r_n$ $\forall w$ and $\forall v \geq e_n = \{\gamma \log g(a_n)\}^{-1/\alpha}$. □

Proof of Corollary 55

Using the result of Theorem 54 – equation (3.10) – to show (3.11) we need to prove that

$$\begin{aligned}
&\Big|\Big[f_{S_n^*}(w) f_{M_n^*}(v) - \mathcal{N}'(w)\Phi'_\alpha(v) \\
&\times\{1 + [h_\rho(v)(1 - \frac{\alpha}{\alpha-\rho}v^{-\alpha}) - \frac{1}{\alpha-\rho}]g(a_n)\}\Big]\{1 - r_n v(\frac{\alpha}{\alpha-1}v^{-\alpha} - 1)w\}\Big| \\
&\qquad\qquad = o(\max\{r_n, g(a_n)\}).
\end{aligned} \tag{3.87}$$

for $\forall w$ and $\forall v \geq e_n = \{-\gamma \log g(a_n)\}^{-1/\alpha}$.

Let

$$A_n = f_{S_n^*}(w),$$

$$A'_n = \mathcal{N}'(w),$$

$$B_n = f_{M_n^*}(v),$$

and

$$B'_n = \Phi'_\alpha(v)\{1 + [h_\rho(v)(1 - \frac{\alpha}{\alpha-\rho}v^{-\alpha}) - \frac{1}{\alpha-\rho}]g(a_n)\}.$$

Then we can write the left-hand side of (3.87) without the $r_n v(\frac{\alpha}{\alpha-1}v^{-\alpha} - 1)w$ term as

$$\begin{aligned}
|A_n B_n - A'_n B'_n| &= |A_n B_n - A_n B'_n + A_n B'_n - A'_n B'_n| \\
&\leq |A_n|\, |B_n - B'_n| + |B'_n|\, |A_n - A'_n|
\end{aligned} \tag{3.88}$$

Now $|A_n|$ is bounded $\forall w$ by Petrov's result and $|B_n - B'_n| = o(g(a_n))$ $\forall v$ by Proposition 56. Thus the first term on the right-hand side of the inequality in (3.88) is $o(g(a_n))$, $\forall v$ and $\forall w$. For the second term in the inequality in (3.88), $|B'_n|$ is bounded $\forall v$ by Proposition 56 and $|A_n - A'_n| = O(\frac{1}{\sqrt{n}})$ by Feller (1971) Chapter XVI, Section 2, Theorem 1 uniformly in w. Given the definition of r_n, we have $|A_n - A'_n| = O(\frac{1}{\sqrt{n}}) = o(r_n)$ uniformly in w. Thus this second term is $o(r_n)$, $\forall v$ and $\forall w$. Thus the right-hand side of (2.123) is $o(\max\{r_n, g(a_n)\})$, $\forall v$ and $\forall w$.

When we add in the $r_n v(\frac{\alpha}{\alpha-1} v^{-\alpha} - 1)w$ term, we need to strength this to

(a.) $|wA_n|$ bounded – which we again have $\forall w$ by Petrov's result.

(b.) $|v^{-\alpha+1} + 1|\ |B_n - B'_n| = o(g(a_n))$ – which we have by Corollary 57, note now on $v \geq \{-\gamma \log g(a_n)\}^{-1/\alpha}$.

(c.) $|e^{-v} + 1|\ |B'_n|$ bounded – which we again have by Corollary 52, note now on $v \geq \{-\gamma \log g(a_n)\}^{-1/\alpha}$.

(d.) $|w|\ |A_n - A'_n| = o(r_n)$ – which we again have $\forall w$ by Petrov's result.

Hence we have the result (3.10) $\forall w$ and $\forall v \geq e_n = \{-\gamma \log g(a_n)\}^{-1/\alpha}$. □

Chapter 4

EXPANSION OF THE JOINT DENSITY UNDER THE WEIBULL DOMAIN OF ATTRACTION

4.1 Introduction

In this chapter we develop the joint density of the sum and maximum of an *iid* sequence of random variables when the underlying distribution lies in the Weibull domain of attraction.

The structure is similar that of Chapter 2 and 3. Here we will use slightly different notation from the previous chapters so as to distinguish from the Fréchet case. Let $Y_1, \ldots, Y_n$ be a *iid* sequence of random variables with common distribution function F which has density f. We again assume the existence of the mean μ, variance σ^2, and also the third moment μ^3 and third cumulant κ_3.

We again define $S_n = \sum_{i=1}^n Y_i$ with the normalized version as

$$S_n^* = \frac{S_n - n\mu}{\sqrt{n\sigma^2}}. \tag{4.1}$$

Recall the notation for the distribution function of S_n^* is $F_{S_n^*}(w)$ with density $f_{S_n^*}(w) = dF_{S_n^*}(w)/dw$.

Like Chapter 2 and 3, throughout this chapter we assume a finite variance and thus we know F lies in the domain of attraction of a stable law with index equal to 2; that is, $F_{S_n^*}(w)$ converges to $\mathcal{N}(w)$ where $\mathcal{N}(w)$ denotes the normal distribution function. Recall the normal density is denoted by $\mathcal{N}'(w)$. In fact, we again utilize Proposition 50 developed in Chapter 2 directly in this chapter. Also we use the central limit result from Petrov (1975) which says that we can bound $|x|^m\{f_{S_n^*z}(x) - \mathcal{N}'(x)\}$ uniformly $\forall x$ when $m \leq 3$.

We again define $M_n = \max\{Y_1, \ldots, Y_n\}$ but here we denote the normalized version as $M_n^\dagger$ and define it as

$$M_n^\dagger = a_n\{M_n - d_n\} \tag{4.2}$$

where $a_n > 0$, d_n real. With this notation, we define the distribution function of $M_n^\dagger$ as $F_{M_n^\dagger}(y)$ and its density as $f_{M_n^\dagger}(y) = \frac{d}{dy}F_{M_n^\dagger}(y)$.

For this chapter, we also assume that F lies in the Weibull domain of attraction; i.e., $F \in \mathcal{D}(\Psi_\alpha)$. Let Ψ_α denote the Weibull distribution and Ψ'_α denotes its density where

$$\Psi_\alpha(y) = \begin{cases} e^{-(-y)^\alpha} & y < 0, \quad \alpha > 0 \\ 0 & \text{otherwise} \end{cases}$$

and

$$\Psi'_\alpha(y) = \begin{cases} \alpha(-y)^{\alpha-1}e^{-(-y)^\alpha} & y < 0, \quad \alpha > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Note since we are assuming the underlying distribution has a finite third moment, again we are in fact assuming that $\alpha \geq 3$ in the Weibull formulae.

The following definition of the Weibull domain of attraction is taken from Smith (1982), proved in Theorem 2.3.2 of de Haan (1970), originally given by Gnedenko (1943).

Definition 60 (Weibull Domain of Atrraction) *Given F, a necessary and sufficient condition for the existence of $a_n > 0, d_n$ real such that*

$$\lim_{n\to\infty} F^n\left(\frac{y}{a_n} + d_n\right) = \Psi_\alpha(y), \quad \alpha > 0$$

is that F has a finite endpoint; that is,

$$x_o = \sup\{y : F(y) < 1\} < \infty$$

and that $F_1(x) = F(x_o - \frac{1}{x})$ is in the Fréchet domain of attraction, Φ_α.

Thus we see a simple transformation connects the Weibull and Fréchet domain of attraction. In fact, through this transformation we are able to apply the propositions and corollaries of Chapter 3 directly in this chapter. We have defined $M_n^\dagger$ in such a way that the normalizing constant a_n is the same as in the Fréchet case. Specifically we can take the normalizing constants of M_n under the Weibull domain of attraction to be

$$d_n \equiv x_o \tag{4.3}$$

and

$$a_n \text{ such that } 1 - F_1(a_n) = 1 - F(x_o - \tfrac{1}{a_n}) = \tfrac{1}{n}. \tag{4.4}$$

Recall $M_n^* = \frac{\max(X_1,\ldots,X_n)-b_n}{a_n}$ where $X_1, \ldots, X_n$ is an *iid* sequence random variables with underlying distribution function in the *Fréchet* case. Note we defined $b_n = 0$ so that $M_n^* = \frac{\max(X_1,\ldots,X_n)}{a_n}$.

Again we exploit this connection between $M_n^\dagger$ (the normalized maximum under the Weibull domain) and M_n^* (the normalized maximum under the Fréchet domain) in this chapter.

We again solve for two forms of the expansion for the joint density. One is of the form of (2.2) from Chapter 2. The other is similar to (2.6) of Chapter 2 with Λ' replaced by Ψ'_α. Like Chapter 2, the derivation starts by rewriting $f_{S_n^*,M_n^\dagger}(w,y) = f_{S_n^*|M_n^\dagger}(w|y) f_{M_n^\dagger}(y)$ where $f_{S_n^*|M_n^\dagger}(w|y)$ is the conditional density of S_n^* given $M_n^\dagger$. Again we need the three key expansions. The first is the expansion for the conditional density of $S_n^*|M_n^\dagger$. Note the expansion of the (conditional) density derived in Proposition 50 of Chapter 2 still applies in this chapter. Although we have switched notation from M_n^* to $M_n^\dagger$, the only difference in the derivation is in the exact form of the threshold (u_n) where we now divide by a_n as opposed to multiply. The process is still to condition on M_n. Now under the Weibull domain of attraction, we need to derive the expansion for the density of $M_n^{\dagger}$. Finally, we need the expansions for the conditional mean and variance of $S_n^*|M_n^{\dagger}$ under the Weibull domain of attraction. Thus the main propositions of this chapter are the expansions for the density of $M_n^{\dagger}$ and for the conditional mean and variance of $S_n^*|M_n^{\dagger}$ under the Weibull domain of attraction.

Again which value M_n is conditioned upon is important to this derivation. Set $M_n = u_n = \frac{y}{a_n} + d_n$ with y is fixed and where u_n is defined as a threshold with a_n and d_n defined in (4.4) and (4.3). Thus $M_n = u_n = \frac{y}{a_n} + x_o$ or $M_n^{\dagger} = a_n\{M_n - x_o\}$.

Recall we can rewrite the distribution of $\overline{S_n}$ in terms of $\sum_{i=1}^{n-1} Y_i^* + u_n$ where the Y^*s are *iid* random variables which have distribution function $\tilde{F_{u_n}}(y) = F(y)/F(u_n)$, and density $\tilde{f_{u_n}}(y) = f(y)/F(u_n)$. We also define $\mu(u_n)$ and $\sigma^2(u_n)$ as its mean and variance.

Again we have the distributional relationship between $\tilde{S}_n$ and $S_n^*|M_n^{\dagger} = y$ as

$$P[\tilde{S}_n \leq x] = P[\frac{n\mu + S_n^*\sqrt{n\sigma^2} - [(n-1)\mu(u_n) + u_n]}{\sqrt{(n-1)\sigma^2(u_n)}} \leq |M_n^{\dagger} = y] \qquad (4.5)$$

with the Jacobian of the transformation as $\frac{\sqrt{n\sigma^2}}{\sqrt{(n-1)\sigma^2(u_n)}}$.

Recall from Chapter 2 that in deriving the expansions for the conditional mean and variance, we first need to solve the mean and variance of an exceedance over a threshold. Recall the conditional mean and variance of an exceedance, given an exceedance over u_n exists, is $m(u_n)$ and $s^2(u_n)$, respectively.

The outline of this chapter is as follows. The main theorem and corollary are presented in Section 4.2. Section 4.3 contains the propositions necessary in establishing the main result – (1) the expansion of the condition density of $S_n^*|M_n^{\dagger}$, (2) the expansion of the density of $M_n^{\dagger}$, and (3) the expansions for the mean and variance of $S_n|M_n$. Since the first expansion has already been derived in Chapter 2, in this chapter only the latter two propositions and their corollaries are presented. Finally, Section 4.4 contains the proofs of the main theorem and its corollary.

4.2 Main Theorem

Here we present the main theorem and its corollary in the same format as in Chapter 2 and 3.

Theorem 61 *Let $Y_1, \ldots, Y_n$ be an iid sequence of random variables with distribution function F, density function f, characteristic function φ, mean μ, and variance σ^2. Let u_n be a threshold and φ_{u_n} be the characteristic function of $Y|Y < u_n$.*

Given the following two sets of assumptions

Set A: Assume f' is integrable, μ_3 exists, φ''' exists and is continuous in a neighborhood of 0, and $|\varphi_{u_n}(t)|^n$ is integrable for some $n \geq n^* > 1$.

Set B: Assume $x_o = sup_y\{y : F(y) < 1\} < \infty$. and, defining $F_1(x) = F(x_o - \frac{1}{x})$, we suppose F_1 satisfies Set B assumptions in Theorem 54 of Chapter 3. We also take the normalizing constants of $M_n^\dagger$ to be $d_n = x_o$ and a_n such that $1 - F_1(a_n) = \frac{1}{n}$. Define

$$r_n = \frac{x_o - \mu}{\sqrt{n\sigma^2}} \tag{4.6}$$

Then,

$$|f_{S_n^*, M_n^\dagger}(w, y) - f_{S_n^*}(w) f_{M_n^\dagger}(y)\{1 - r_n[(-y)^\alpha - 1]w\}| = o(r_n). \tag{4.7}$$

$\forall w$ and $\forall y \geq e_n^\dagger$ where $e_n^\dagger = -\{-\gamma \log g(a_n)\}^{1/\alpha}$ for some $\gamma > 1$.

Corollary 62 *Given the conditions of Theorem 61, then uniformly* $\forall w$ *and* $\forall y \geq e_n^\dagger$ *where* $e_n^\dagger = -\{-\gamma \log g(a_n)\}^{1/\alpha}$ *for some* $\gamma > 1, \alpha > 1$

$$\Big| f_{S_n^*, M_n^\dagger}(w,y) - \mathcal{N}'(w)\Psi'_\alpha(y)\{1 - r_n[(-y)^\alpha - 1]w\}\{1 + \frac{\kappa_3}{6\sigma^3\sqrt{n}}(w^3 - 3w)\}$$
$$\times\{1 - [h_{-\rho}(-y)(1 - \frac{\alpha}{\alpha-\rho}(-y)^\alpha) + \frac{1}{\alpha+\rho}]g(a_n)\}\Big| = o(\max\{r_n, g(a_n)\}). \quad (4.8)$$

4.3 Propositions

Here we present the main propositions and their corollaries. These propositions are central to the derivation of the expansion of the joint density of the sum and the maximum. They parallel the propositions in the previous two chapters.

4.3.1 Expansion of Conditional Density of the Sum given the Maximum

We do not need to repeat the results because all the conditions needed for Propostion 50 also apply in this chapter.

4.3.2 Expansion of Density of the Maximum

Like in the Fréchet case, this proposition is based on Smith (1982). As in Smith (1982), the results for the Weibull case are derived using the transformation from the Fréchet case.

Proposition 63 *Let* $Y_1, \ldots, Y_n$ *be an iid sequence of random variables with common distribution function* F*. Define* $M_n^\dagger = a_n\{\max(Y_1, \ldots, Y_n) - d_n\}$*. Also define* $F_1(x) =$

$F(x_o - \frac{1}{x})$, $d_n = x_o$ and a_n such that $1 - F_1(a_n) = \frac{1}{n}$. Suppose F_1 satisfies the conditions of Proposition 56 of Chapter 3 and $\alpha > 1$, then

$$|f_{M_n^\dagger}(y) - \Psi'_\alpha(y)\{1 - [h_{-\rho}(-y)(1 - \frac{\alpha}{\alpha - \rho}(-y)^\alpha) + \frac{1}{\alpha - \rho}]g(a_n)\}| = o(g(a_n)) \quad (4.9)$$

uniformly on $y < 0$.

PROOF: Let $X = \frac{1}{x_o - Y}$. Definition 60 states that $F \in \mathcal{D}(\Psi_\alpha)$ if and only if $F_1 \in \mathcal{D}(\Phi_\alpha)$. We use $M_n^\dagger = a_n\{\max(Y_1, \ldots, Y_n) - d_n\} = a_n\{\max(Y_1, \ldots, Y_n) - x_o\}$ and recall in the Fréchet case, $M_n^* = \frac{\max(X_1, \ldots, X_n)}{a_n}$ where the constant a_n is the same for both cases.

Note that

$$\begin{aligned} M_n^\dagger &= a_n\{\max(Y_1, \ldots, Y_n) - x_o\} = a_n\{\max(x_o - \frac{1}{X_1}, \ldots, x_o - \frac{1}{X_n}) - x_o\} \\ &= \frac{-a_n}{\max(X_1, \ldots, X_n)} = -\frac{1}{M_n^*}. \end{aligned}$$

Thus the two densities are related by

$$f_{M_n^\dagger}(y) = \frac{1}{y^2} f_{M_n^*}(-\frac{1}{y}). \quad (4.10)$$

Therefore for each $y \in (-\infty, 0)$, we have by (3.12) of Chapter 3

$$\begin{aligned} f_{M_n^\dagger}(y) &= \frac{1}{y^2}\Big\{\alpha(-\frac{1}{y})^{-\alpha-1} e^{-(-\frac{1}{y})^{-\alpha}} \\ &\quad \times\{1 + [h_\rho(-\frac{1}{y})(1 - \frac{\alpha}{\alpha - \rho}(-\frac{1}{y})^{-\alpha}) - \frac{1}{\alpha - \rho}]g(a_n)\} + o(g(a_n))\Big\} \\ &= \alpha(-y)^{\alpha-1} e^{-(-y)^\alpha}\Big\{\{1 - [h_{-\rho}(-y)(1 - \frac{\alpha}{\alpha - \rho}(-y)^\alpha) + \frac{1}{\alpha - \rho}]g(a_n)\} \\ &\quad + o(g(a_n))\Big\}. \end{aligned} \quad (4.11)$$

Note we need $\alpha > 1$ since (4.11) becomes infinite otherwise.

Define $e_n^\dagger = -\frac{1}{e_n}$ where e_n is defined in Chapter 3. Specifically, we have

$$e_n^\dagger = -\{-\gamma \log g(a_n)\}^{1/\alpha}. \quad (4.12)$$

As $n \to \infty$, we have $e_n^\dagger \to -\infty$.

Note that the proof for Weibull case falls directly from the Fréchet case using the transformation v to $-1/y$. With this transformation we can rewrite (4.10), $\forall y \geq e_n^\dagger$

$$f_{M_n^\dagger}(y) = \alpha(-y)^{\alpha-1}e^{-(-y)^\alpha}\left\{1 - \left[h_{-\rho}(-y)(1 - \frac{\alpha}{\alpha-\rho}(-y)^\alpha) + \frac{1}{\alpha-\rho}\right]g(a_n)\right\} + \sum_{j=1}^{4} E_j. \tag{4.13}$$

where

$$\begin{aligned} E_1^\dagger &= \alpha(-y)^{\alpha-1}e^{-(-y)^\alpha}\{1 + (-y)^\alpha h_{-\rho}(-y)g_1(a_n) + U_n\} \\ &\quad \times\{1 - \frac{g(a_n)}{\alpha-\rho}\}\{1 - h_{-\rho}(-y)g(a_n) + T_n\}\{e^{-R_n} - 1\}, \end{aligned} \tag{4.14}$$

$$E_2^\dagger = \alpha(-y)^{\alpha-1}e^{-(-y)^\alpha}U_n\{1 - \frac{g(a_n)}{\alpha-\rho} + S_n\}\{1 - h_{-\rho}(-y)g(a_n) + T_n\}, \tag{4.15}$$

$$\begin{aligned} E_3^\dagger &= \alpha(-y)^{\alpha-1}e^{-(-y)^\alpha}(-y)^\alpha(-h_{-\rho}(-y))g_1(a_n) \\ &\quad \times\{-\frac{g(a_n)}{\alpha-\rho} + S_n + (-h_{-\rho}(-y))g(a_n) + T_n)(1 - \frac{g(a_n)}{\alpha-\rho} + S_n)\}, \end{aligned} \tag{4.16}$$

$$E_4 = \alpha(-y)^{\alpha-1}e^{-(-y)^\alpha}\{(-\frac{g(a_n)}{\alpha-\rho} + S_n)(-h_{-\rho}(-y))g(a_n) + T_n) + S_n + T_n\}, \tag{4.17}$$

with

$$\begin{aligned} R_n^\dagger &\sim (-y)^\alpha\left(\epsilon\theta_{10}(-y)^{-\beta}g_1(a_n)\right) \\ &+ (-y)^\alpha(-h_{-\rho}(-y))g_1(a_n)\left\{\theta_9(-h_{-\rho}(-y))g_1(a_n) + \theta_{11}\epsilon(-y)^{-\beta}g_1(a_n)\right\} \\ &+ \frac{(-y)^\alpha}{n}[-1 + h_{-\rho}((-y))g_1(a_n)] \\ &+ \frac{(-y)^\alpha}{n}\Big(\epsilon\theta_{10}(-y)^{-\beta}g_1(a_n) \\ &\qquad -h_{-\rho}(-y)g_1(a_n)\{\theta_9(-h_{-\rho}(-y))g_1(a_n) + \theta_{11}\epsilon(-y)^{-\beta}g_1(a_n)\}\Big) \\ &+ \frac{-\theta_4(-y)^{2\alpha}}{n^2}\frac{(n-1)(1+o(1))}{1 - \theta_4\frac{(-y)^\alpha}{n}\{1+o(1)\}} \quad \text{where } \beta = \rho - \epsilon < 0 \end{aligned} \tag{4.18}$$

and

$$S_n^\dagger \sim o(g(a_n)) \tag{4.19}$$

and

$$T_n^\dagger \sim \epsilon\theta_7(-y)^{-\beta}g(a_n) - h_{-\rho}(-y)g(a_n)[\theta_6(-h_{-\rho}(-y))g(a_n) + \theta_8\epsilon(-y)^{-\beta}g(a_n)] \tag{4.20}$$

and, finally,

$$U_n^\dagger \sim \theta_5(-y)^\alpha(h_{-\rho}^2(-y))g_1^2(a_n) \tag{4.21}$$

when R_n, S_n, T_n, and U_n go to 0.

Note we use the relation $h_\rho(-\frac{1}{y}) = -h_{-\rho}(-y)$ where recall $\rho > 0$.

Now the proof – like the Fréchet case – breaks down into three parts: Case 1: $y \to 0$ on $-1 \leq y < 0$; Case 2: $y \to -\infty$ via $e_n^\dagger$; Case 3: $y \to -\infty$ on $y < e_n^\dagger$.

The key is that, using the transformation to the Fréchet case, these three cases are equivalent to: Case A: $v \to \infty$ on $1 \leq v < \infty$; Case B: $v \to 0$ via e_n; Case C: $v \to 0$ on $v < e_n$, respectively.

Case 1 $y \to 0$ is equivalent to proving in the Fréchet case

$$v^2E_1 = v^2E_2 = v^2E_3 = v^2E_4 = v^2E_4 = o(g(a_n)) \tag{4.22}$$

uniformly on $1 \leq v < \infty$.

The leading terms associated with the polynomial of v in each formula in (4.22) is $v^{-\alpha+1}$. Since $\alpha > 1$, then $\lim_{v\to\infty} v^{-\alpha+1} = 0$. In other words, on $1 \leq v < \infty$, (4.22) holds. With the transformation,

$$E_1^\dagger = E_2^\dagger = E_3^\dagger = E_4^\dagger = E_4^\dagger = o(g(a_n)) \tag{4.23}$$

uniformly on $-1 \leq y < 0$.

Case 2 $y \to -\infty$ via $e_n^\dagger$ is equivalent to $v \to 0$ via e_n in the Fréchet case. Recall a_n is defined so that $1 - F_1(a_n) = \frac{1}{n}$. Since the Weibull and the Fréchet case also share the same function g, we again have $g(a_n) > 1/n$ and $a_ne_n \to \infty$ so the condition in Lemma 3 of Smith (1982) still holds and so does the formula for $f_{M_n^*}$ along with E_1, E_2, E_3, and E_4, with R_n, S_n, T_n, and U_n from Chapter 3. With the transformation we have (4.13) – (4.21) on $y \geq e_n^\dagger$.

Again proving the result for $E_1^\dagger, E_2^\dagger, E_3^\dagger, E_4^\dagger$, and $E_4^\dagger$ on $y \geq e_n^\dagger$ is equivalent to proving in the Fréchet case that

$$v^2 E_1 = v^2 E_2 = v^2 E_3 = v^2 E_4 = v^2 E_4 = o(g(a_n)) \tag{4.24}$$

uniformly as $v \to 0$ via e_n as $n \to \infty$; i.e., on $v \geq e_n$.

Once more the leading term in (4.22) is $v^{-\alpha+1}$ where $-\alpha + 1 < 0$. By (3.45) of Chapter 3, we have for any $\delta > 0$,

$$v^{-\delta} g(a_n) \to 0 \quad \text{as } n \to \infty.$$

Thus (4.24) holds $\forall v \geq e_n$. Equivalently

$$E_1^\dagger = E_2^\dagger = E_3^\dagger = E_4^\dagger = E_4^\dagger = o(g(a_n)) \tag{4.25}$$

uniformly on $\forall y \geq e_n^\dagger$.

Case 3 $y \to -\infty$ on $y < e_n^\dagger$. Here we need to show that each term in (4.9) is $o(g(a_n))$ on $y \leq e_n^\dagger$.

First look at the second term of (4.9),

$$\alpha(-y)^{\alpha-1} e^{-(-y)^\alpha} \{1 - [h_{-\rho}(-y)(1 - \frac{\alpha}{\alpha - \rho}(-y)^\alpha) + \frac{1}{\alpha - \rho}] g(a_n)\}. \tag{4.26}$$

Obviously the highest order terms is $\alpha(-y)^{\alpha-1} e^{-(-y)^\alpha}$. As $y \to -\infty$, (4.26) $\sim e^{-(-y)^\alpha}$ which is an increasing function on this interval. In other words, setting $y = e_n^\dagger$, for sufficiently large n,

$$(4.26) \quad \sim e^{-(-e_n^\dagger)^\alpha} = e^{-(-\gamma \log g(a_n))} = g^\gamma(a_n) = o(g(a_n)) \ \text{ since } \lambda > 1. \tag{4.27}$$

Now (4.9) holds for $y = e_n^\dagger$, thus $f_{M_n^\dagger}(e_n^\dagger) = o(g(a_n))$. To complete the proof we need only show that $f_{M_n^\dagger}(y)$ is increasing on $y < e_n^\dagger$.

Now

$$\frac{d}{dy} f_{M_n^\dagger}(y) = \frac{d}{dy}\left\{\frac{1}{y^2} f_{M_n^*}(-\frac{1}{y})\right\} = -\frac{2}{y^3} f_{M_n^*}(-\frac{1}{y}) + \frac{1}{y^4} f'_{M_n^*}(-\frac{1}{y}). \tag{4.28}$$

Since $y < 0$ – or $-2/y^3$ is positive – and $f_{M_n^*}$ is a proper density, the first term of (4.28) is non-negative. Finally, the interval $y < e_n^\dagger$ is equivalent to the interval $v < e_n$ in the Fréchet case. From Chapter 3 equation (3.47), we see on this interval that $f'_{M_n^*} \geq 0$. Hence the second term in (4.28) is non-negative and together we have

$$\frac{d}{dy} f_{M_n^\dagger}(y) \geq 0 \quad \text{on } y \leq e_n^\dagger. \tag{4.29}$$

So $f_{M_n^\dagger}(y)$ is monotone on $y \leq e_n^\dagger$ □

Corollary 64 *Given the conditions of Proposition 63, we have for $j = 0, 1$*

$$|(-y)^{\alpha+j} f_{M_n^\dagger}(y) - (-y)^{\alpha+1}\Psi'_\alpha(y)\{1 - [h_{-\rho}(-y)(1 - \frac{\alpha}{\alpha-\rho}(-y)^\alpha) + \frac{1}{\alpha-\rho}]g(a_n)\}| = o(g(a_n)) \tag{4.30}$$

uniformly on $y < 0$.

PROOF: The proof falls directly from the proof of Proposition 63. The only difference in the derivation is that we replace $E_i^\dagger$ by $(-y)^{\alpha+j}E_i^\dagger$ $i = 1, 2, 3, 4$ where now the leading term is

$$\alpha(-y)^{2\alpha+j-1}e^{-(-y)^\alpha} \sim e^{-(-y)^\alpha}, \quad \text{as } y \to -\infty.$$

Thus the case $y \to 0$ follows immediately from proof of Proposition 63. The case $y \to -\infty$ via $e_n^\dagger$ also follows immediately by noting that transforming this to the Fréchet case, the highest order polynomial term in each $v^{-\alpha+1}E_i$ would be $-2\alpha + j - 1$. Since $\alpha > 1$, we have the argument in (3.46) of Chapter 3 which leads to the result – i.e., with the transformation $E_1^\dagger = E_2^\dagger = E_3^\dagger = E_4^\dagger = o(g(a_n))$.

Since (4.30) holds when $y = e_n^\dagger$, to complete the proof we need only show that $(-y)^{\alpha+j} f_{M_n^\dagger}(y)$ is increasing on $y < e_n^\dagger$.

We have

$$\frac{d}{dy}(-y)^{\alpha+j} f_{M_n^\dagger}(y) = (\alpha + j)(-y)^{\alpha+j-1} f_{M_n^\dagger}(y) + (-y)^{\alpha+j} f'_{M_n^\dagger}(y).$$

For $y \leq e_n^\dagger$, we have $(\alpha + j)(-y)^{\alpha+j-1} f_{M_n^\dagger}(y) \geq 0$ since $\alpha > 1$ and $f_{M_n^\dagger}$ is a proper density. We also have $(-y)^{\alpha+j} > 0$ and by (4.29) $f'_{M_n^\dagger} > 0$ on this interval. Thus $(-y)^{\alpha+j} f_{M_n^\dagger}(y)$, for $j = 0$ or 1 is increasing on $y \leq e_n^\dagger$. □

4.3.3 Expansions of Conditional Mean and Variance

Since we assume that F has a finite endpoint in the Weibull case, this proposition is equivalent to the Gumbel case when the upper endpoint is finite. The only difference is in the corollary where we substitute in for $1 - F$ the appropriate Weibull expansion as opposed to the Gumbel expansion.

Proposition 65 *Suppose that $Y_1, \ldots, Y_n$ are iid with common distribution F which has finite upper endpoint x_o. Let u be a threshold. Then*

$$\mu - \mu(u) \sim (x_o - \mu)\{1 - F(u)\} \tag{4.31}$$

and

$$\sigma^2 - \sigma^2(u) \sim \{(x_o - \mu)^2 - \sigma^2\}\{1 - F(u)\}. \tag{4.32}$$

Proof Recall from Chapter 2

$$\mu - \mu(u) = \frac{1 - F(u)}{F(u)}\{u + m(u) - \mu\}.$$

Note as $n \to \infty$, we have $u \to x_o$, $m(u) \to 0$, and $F(u) \to 1$. So

$$\mu - \mu(u) \sim \{x_o - \mu\}\{1 - F(u)\}. \tag{4.33}$$

From this we get

$$\mu^2 - \mu^2(u) \sim 2\mu\{x_o - \mu\}\{1 - F(u)\}. \tag{4.34}$$

Recall

$$\sigma^2 - \sigma^2(u) = \mu^2(u) - \mu^2 + \left(\frac{1 - F(u)}{F(u)}\right) \{(u + m(u))^2 + s^2(u) + \mu^2 - \sigma^2\}. \quad (4.35)$$

As $n \to \infty$ $(u \to x_o)$, we also have

$$s^2(u) = Var(X - u|X > u) \to 0. \quad (4.36)$$

Thus the leading term in (4.35) involves $\mu^2(u) - \mu^2$, u^2, $\mu^2 - \sigma^2$, and $1 - F(u)$. Hence substituting in (4.33), (4.34), and (4.36) into (4.35), we have the result (4.32). □

Corollary 66 *Suppose that $Y_1, \ldots, Y_n$ are iid with common distribution F which has finite upper endpoint x_o. Let $F_1(x) = F(x_o - \frac{1}{x})$. Define the normalizing constants of $M_n^\dagger = max(Y_1, \ldots, Y_n)$ as $b_n = x_o$ and a_n such that $1 - F_1(a_n) = 1/n$ so that $u_n = \frac{y}{a_n} + x_o$. Assume conditions in Proposition 63, then uniformly on $\forall y \geq e_n^\dagger$*

$$\mu - \mu(u_n) = (x_o - \mu)\frac{(-y)^\alpha}{n} + o(1) \quad (4.37)$$

and

$$\sigma^2 - \sigma^2(u_n) = \{(x_o - \mu)^2 - \sigma^2\}\frac{(-y)^\alpha}{n} + o(1). \quad (4.38)$$

Remark Given the definition of $M_n^\dagger$, we define

$$u_n = \frac{y}{a_n} + x_o. \quad (4.39)$$

Recall the threshold in Fréchet case was ultimately defined as $u_n = a_n v + b_n = a_n v$. Here we use the inverted formula for u_n in (4.39) so that we may use the same a_n in both the thresholds. Recall we define $F \in \mathcal{D}(\Psi_\alpha)$ if and only if $F_1 \in \mathcal{D}(\Phi_\alpha)$ with normalizing constant a_n defined so that $1 - F_1(a_n) = 1/n$. This consistency allows us to use the results in the propositions and corollaries of Chapter 3 here.

PROOF: We need only show that on $y \geq e_n^\dagger$ that $1 - F(u_n) = \frac{(-y)^\alpha}{n}(1 + o(1))$ and substitute this into (4.31) and (4.32) to get the corollary results.

We have

$$\begin{aligned} 1 - F(u_n) &= \{1 - F(x_o - \frac{1}{a_n})\}\frac{1 - F(u_n)}{1 - F(x_o - \frac{1}{a_n})} \\ &= \{1 - F_1(a_n)\}\frac{1 - F_1(\frac{1}{x_o - u_n})}{1 - F_1(a_n)} \\ &= \{1 - F_1(a_n)\}\frac{1 - F_1(a_n(-\frac{1}{y}))}{1 - F_1(a_n)} \\ &= \frac{1}{n}\left\{\left(-\frac{1}{y}\right)^{-\alpha}(1 + o(1))\right\}. \end{aligned} \quad (4.40)$$

The last line follows from (3.60) of Chapter 3, since $y \geq e_n^\dagger \Rightarrow v \geq e_n$. □

4.4 Proof of Main Theorem and Its Corollary

PROOF OF THEOREM 61: We write the joint density of S_n^* and $M_n^\dagger$ as

$$\begin{aligned} f_{S_n^*, M_n^\dagger}(w, y) &= f_{M_n^\dagger}(y) f_{S_n^*|M_n^\dagger}(w|y) \\ &= f_{M_n^\dagger}(y)\sqrt{\frac{n\sigma^2}{(n-1)\sigma^2(u_n)}} f_{\tilde{S}_n}(z) \end{aligned} \quad (4.41)$$

where

$$z = \frac{n\mu + \sqrt{n\sigma^2}w - (n-1)\mu(u_n) - u_n}{\sqrt{(n-1)\sigma^2(u_n)}}. \quad (4.42)$$

.

Here we let $u_n = \frac{y}{a_n} + x_o$. Again to enable the uniformity results we will allow w, y and hence z to be dependent on n. Again we suppress this so as to make the notation easier to read.

Note the transformation from S_n^* to $\tilde{S}_n$ and the form of (4.42) comes from (2.7) of Chapter 2.

Now to establish (4.7) we break it up as follows

$$f_{M_n^\dagger}(y)\sqrt{\frac{n\sigma^2}{(n-1)\sigma^2(u_n)}}f_{\tilde{S_n}}(z) - f_{S_n^*}(w)f_{M_n^\dagger}(y)\{1 - r_n((-y)^\alpha - 1)w\}$$
$$= E_1 + E_2 + E_3 + E_4 + E_5 + E_{5b} \qquad (4.43)$$

where

$$E_1 = f_{M_n^\dagger}(y)f_{\tilde{S_n}}(z)\left[\sqrt{\frac{n\sigma^2}{(n-1)\sigma^2(u_n)}} - 1\right], \qquad (4.44)$$

$$E_2 = f_{M_n^\dagger}(y)\left[f_{\tilde{S_n}}(z) - \mathcal{N}'(z)\{1 + \frac{\kappa_3(u_n)}{6\sigma^3(u_n)\sqrt{n}}(z^3 - 3z)\}\right], \qquad (4.45)$$

$$E_3 = f_{M_n^\dagger}(y)\left[\mathcal{N}'(z) - \mathcal{N}'(w)\{1 - r_n((-y)^\alpha - 1)w\}\right], \qquad (4.46)$$

$$E_4 = f_{M_n^\dagger}(y)\left[\mathcal{N}'(w)\{1 + \frac{\kappa_3}{6\sigma^3\sqrt{n}}(w^3 - 3w)\} - f_{S_n^*}(w)\right]$$
$$\{1 - r_n((-y)^\alpha - 1)w\}, \qquad (4.47)$$

$$E_5 = f_{M_n^\dagger}(y)\left[\frac{\mathcal{N}'(z)\kappa_3(u_n)}{6\sigma^3(u_n)\sqrt{n}}(z^3 - 3z) - \frac{\mathcal{N}'(w)\kappa_3}{6\sigma^3\sqrt{n}}(w^3 - 3w)\right]. \qquad (4.48)$$

$$E_{5b} = f_{M_n^\dagger}(y)\mathcal{N}'(w)\frac{\kappa_3}{6\sigma^3\sqrt{n}}(w^3 - 3w)[r_n((-y)^\alpha - 1)w]. \qquad (4.49)$$

Now to prove that (4.7) we show that (4.44) – (4.49) are $o(r_n)$ uniformly $\forall w$ and $y \geq e_n^\dagger$. And to prove (4.44) – (4.49) are $o(r_n)$, it suffices to show that for any $\epsilon > 0$, $\sum_{j=1}^{6}|E_j| = \epsilon\, r_n$ for all sufficiently large n. It is necessary to consider two cases where the dependence on n for w, y, and z need to be explicitly expressed. These two cases are: Case (a) $|z_n - w_n| \leq \delta$ for a $\delta > 0$ and Case(b) $|z_n - w_n| > \delta$.

Proof for E_1 The following argument for E_1 holds irrespective of δ and also for both Case (a) and Case (b).

We start with $E_1 = f_{M_n^\dagger}(y)f_{\tilde{S_n}}(z)\left[\sqrt{\frac{n\sigma^2}{(n-1)\sigma^2(u_n)}} - 1\right]$. Specifically we begin with its third term.

From the Gumbel proof, we have

$$|\sqrt{\frac{n\sigma^2}{(n-1)\sigma^2(u_n)}} - 1| \leq \sqrt{\frac{\sigma^2}{\sigma^2(u_n)}} - 1 + O(\frac{1}{n}). \qquad (4.50)$$

Now from (4.38)

$$\frac{\sigma^2(u_n)}{\sigma^2} = 1 - r_n^2(-y)^\alpha\{1 - \frac{\sigma^2}{(x_o-\mu)^2}\} + o(1)$$

on $-\infty < e_n^\dagger \leq y < 0$.

Note

$$r_n^2(-y)^\alpha = \frac{(x_o-\mu)^2}{\sigma^2}\frac{(-y)^\alpha}{n} \sim 1 - F(u_n) \text{ [by (4.40)] which} \to 0.$$

Thus $r_n^2(-y)^\alpha\{1 - \frac{\sigma^2}{(x_o-\mu)^2}\} \to 0$ on this interval so

$$\frac{\sigma^2}{\sigma^2(u_n)} = 1 + r_n^2(-y)^\alpha\{1 - \frac{\sigma^2}{(x_o-\mu)^2}\} + o(1)$$

on $0 > y \geq e_n^\dagger$.

Thus

$$\sqrt{\frac{n\sigma^2}{(n-1)\sigma^2(u_n)}} - 1| \leq K\left(r_n^2[(-y)^\alpha + 1]\right), \quad \text{for some constant } K > 0. \tag{4.51}$$

Thus

$$E_1 = O(f_{M_n^\dagger}(y) f_{\tilde{S_n}}(z) r_n^2[(-y)^\alpha + 1]).$$

Now by Proposition 50 in Chapter 12, $f_{\tilde{S_n}}(z)$ is uniformly bounded $\forall z$. By Corollary 64, $(-y)^\alpha f_{M_n^\dagger}(y)$ is uniformly bounded on $y \geq e_n^\dagger$ and by Proposition 63 we have $f_{M_n^\dagger}(y)$ is uniformly bounded $\forall y$.

Thus

$$E_1 = O(r_n^2) = o(r_n), \quad \forall z, \forall y \geq e_n^\dagger.$$

or we may write for sufficiently large n ($\epsilon > 0$)

$$|E_1| < \frac{\epsilon}{6}\, r_n, \forall w, \forall y \geq e_n^\dagger. \tag{4.52}$$

Proof for E_2 Like the proof for E_1, the following argument holds for both Case (a) and Case (b). This argument is the same as in the Fréchet case. The only difference is that $f_{M_n^\dagger}$ replaces $f_{M_n^*}$ but this is also bounded, here by Proposition 63. So

$$E_2 = o(\frac{1}{\sqrt{n}}) = o(r_n), \quad \forall y \text{ and } \forall w.$$

or we may write for sufficiently large n ($\epsilon > 0$)

$$|E_2| < \frac{\epsilon}{6}\, r_n, \forall w, \forall y \geq e_n^\dagger. \tag{4.53}$$

Proof for E_3 Here we will proceed simultaneously with both cases until *Step 3* where at that point we will need to divide the proof into the two cases.

Recall $E_3 = f_{M_n^\dagger}(y)\,[\mathcal{N}'(z) - \mathcal{N}'(w)\{1 - r_n((-y)^\alpha - 1)w\}]$. Again establishing $E_3 = o(r_n)$ involves a longer argument than needed for E_1 or E_2. We again break this argument into the following steps.

Step 1: Recall from the Gumbel case,

$$z - w = \frac{n(\mu - \mu(u_n))}{\sqrt{(n-1)\sigma^2(u_n)}} + \{\sqrt{\frac{n\sigma^2}{(n-1)\sigma^2(u_n)}} - 1\}w + \frac{\mu(u_n) - u_n}{\sqrt{(n-1)\sigma^2(u_n)}}. \tag{4.54}$$

Substitute in (4.37) into the first and third term and (4.51) into the second term,

$$\begin{aligned} z - w &= \frac{(x_o - \mu)(-y)^\alpha}{\sqrt{n\sigma^2}}\{1 + o(1)\} + O\left(r_n^2[(-y)^\alpha + 1]w\right) - \frac{(x_o - \mu)}{\sqrt{n\sigma^2}}\{1 + o(1)\} \\ &= r_n((-y)^\alpha - 1) + o\left(r_n((-y)^\alpha - 1)\right) + O\left(r_n^2[(-y)^\alpha + 1]w\right). \end{aligned} \tag{4.55}$$

Step 2: We again take a Taylor expansion for $\mathcal{N}'(z)$ about w – i.e. write $z = w + t_n$ where t_n can be seen in (4.55). Recall

$$\mathcal{N}'(z) = \mathcal{N}'(w) - t_n z^* \mathcal{N}'(z^*)$$

for z^* between w and z.

Substituting this into E_3 we have

$$\begin{aligned}
E_3 &= f_{M_n^\dagger}(y)(z-w)\,[w\mathcal{N}'(w) - z^*\mathcal{N}'(z^*)] + o(r_n((-y)^\alpha - 1)w\mathcal{N}'(w)f_{M_n^\dagger}(y)) \\
&\quad + O(r_n^2[(-y)^\alpha + 1]w^2\mathcal{N}'(w)f_{M_n^\dagger}(y)) \\
&= E_6 + E_7 + E_8.
\end{aligned}$$

Now $w\mathcal{N}'(w)$ and $w^2\mathcal{N}'(w)$ are uniformly bounded $\forall w$. From Proposition 63, $f_{M_n^\dagger}(y)$ is uniformly bounded $\forall y$. From Corollary 64, we also have $(-y)^\alpha f_{M_n^\dagger}(y)$ is uniformly bounded on $y \geq e_n^\dagger$. Thus $E_7 = E_8 = o(r_n)$ on $y \geq e_n^\dagger$ or we may write for sufficiently large n

$$|E_7| < \frac{\epsilon}{6}\, r_n,\ \forall w,\ \forall y \geq e_n^\dagger, \tag{4.56}$$

$$|E_8| < \frac{\epsilon}{6}\, r_n,\ \forall w,\ \forall y \geq e_n^\dagger \tag{4.57}$$

Step 3: Now we focus on E_6. The important details in this formula concern $z - w$ since the other terms are bounded. Recall y, w, z, and z^* are actually dependent on n. So again fix the notation by writing $y = y_n, w = w_n, z = z_n$, and $z^* = z_n^*$ so the dependence on n is explicit.

By substituting (4.55) into E_6 we have

$$\begin{aligned}
E_6 &= f_{M_n^\dagger}(y_n)\left(r_n((-y_n)^\alpha - 1) + o(r_n\{(-y_n)^\alpha - 1\}) + O(r_n^2[(-y_n)^\alpha + 1]w_n)\right) \\
&\quad \times\{w_n\mathcal{N}'(w_n) - z_n^*\mathcal{N}'(z_n^*)\} \\
&= r_n f_{M_n^\dagger}(y_n)[((-y_n)^\alpha - 1)]\{w_n\mathcal{N}'(w_n) - z_n^*\mathcal{N}'(z_n^*)\} \\
&\quad + o\left(r_n f_{M_n^\dagger}(y_n)[((-y_n)^\alpha - 1)]\{w_n\mathcal{N}'(w_n) - z_n^*\mathcal{N}'(z_n^*)\}\right) \\
&\quad + O\left(r_n^2 f_{M_n^\dagger}(y_n)[(-y_n)^\alpha + 1]w_n\{w_n\mathcal{N}'(w_n) - z_n^*\mathcal{N}'(z_n^*)\}\right) \\
&= E_9 + E_{10} + E_{11}.
\end{aligned}$$

At this point, it is necessary to separate the argument into the two cases.

Case (a) If $|z_n - w_n| \leq \delta$ then $|z_n^* - w_n| \leq \delta$. By uniform continuity of $w^k\mathcal{N}'(w)$ for $k = 0, 1, 2$ given $\epsilon > 0$ we can find a $\delta > 0$ such that

$$|z_n - w_n| < \delta \Rightarrow |z_n^k\mathcal{N}'(z_n) - w_n^k\mathcal{N}'(w_n)| < \frac{\epsilon}{C}$$

for $k = 0, 1, 2$ and any given constant $C > 0$.

Now since $((-y_n)^\alpha + 1)f_{M_n^\dagger}(y_n)$ is bounded on $y_n \geq e_n^\dagger$ by Proposition 63 and Corollary 64, we have that each E_9, E_{10}, and E_{11} is bounded by some

$$constant \times |z_n^{*k}\mathcal{N}'(z_n^*) - w_n^k\mathcal{N}'(w_n)|$$

for $k = 0, 1, 2$.

In other words, we can choose a δ so that $\forall w_n$ and $\forall y_n \geq e_n^\dagger$

$$|z_n - w_n| < \delta \Rightarrow |E_6| < \frac{\epsilon}{18}\, r_n \quad \text{for all sufficiently large } n$$

which with (4.56) and (4.57) gives for sufficiently large n ($\epsilon > 0$)

$$|E_3| < \frac{\epsilon}{6}\, r_n \tag{4.58}$$

for $y_n \geq e_n^\dagger$ and $\forall w_n$ and when $|z_n - w_n| \leq \delta$. [End Case (a)]

Case (b) Here we show if $|z_n - w_n| > \delta$ then the entire left-hand side of (4.43) is $o(r_n)$.

<u>*Part 1*</u> Suppose $|z_n - w_n| > \delta$ and $r_n(-y_n)^\alpha < \delta^2$.

From (4.55) we deduce $|w_n r_n| >$ some $\delta_1 > 0$ for all sufficiently large n. So $|w_n| > \frac{\delta_1}{r_n}$. We also have from (4.55),

$$\begin{aligned} z_n &= w_n + r_n((-y_n)^\alpha - 1) + o\left(r_n((-y_n)^\alpha - 1)\right) + O\left(r_n^2[(-y_n)^\alpha + 1]w_n\right) \\ &= w_n + o(1) + o(1) + O(w_n o(1)) \\ &= w_n(1 + o(1)) \\ &> \frac{\delta_1}{2r_n}, \quad \text{say,} \end{aligned}$$

for all sufficiently large n.

Thus we have [with (4.55)]

1. $|f_{\tilde{S_n}}(z_n)| = o(r_n)$ by Proposition 50.

2. $|w_n|^k |f_{S_n^*}(w_n)| = o(r_n)$ for k=0,1 by Petrov's central limit theorem.

3. $|((-y_n)^\alpha + 1) f_{M_n^*}(y_n)|$ is bounded on $y_n \geq e_n^\dagger$ by Proposition 63 and Corollary 64.

Hence the left-hand side of (4.43) is $o(r_n)$ on $y_n \geq e_n^\dagger$. and $\forall w_n$ when $|z_n - w_n| > \delta$ and $r_n(-y_n)^\alpha < \delta^2$.

<u>*Part 2*</u> Suppose $|z_n - w_n| > \delta$ and $r_n(-y_n)^\alpha \geq \delta^2$. Now if $r_n(-y_n)^\alpha \geq \delta^2$ then

$$(-y_n)^\alpha > \frac{\delta^2}{r_n} \quad \text{or } (-y_n) > \delta^{2/\alpha} r_n^{-\frac{1}{\alpha}}.$$

The first term of (4.43) is $f_{M_n^\dagger}(y_n)\sqrt{\frac{n\sigma^2}{(n-1)\sigma^2(u_n)}} f_{\tilde{S_n}}(z_n)$. Now, by Proposition 50 of Chapter 2 we have $f_{\tilde{S_n}}(z_n) = O(1)$ uniformly $\forall z_n$. By (4.51), we have $\sqrt{\frac{n\sigma^2}{(n-1)\sigma^2(u_n)}} = 1 + O(r_n^2[(-y_n)^\alpha + 1])$ uniformly on $y_n \geq e_n^\dagger$. By Corollary 64, we have

$$(-y_n)^{\alpha+1} f_{M_n^\dagger}(y_n) = O(1) \quad \text{on } y_n \geq e_n^\dagger. \tag{4.59}$$

This gives us on $(-y_n) > \delta^{2/\alpha} r_n^{-\frac{1}{\alpha}}$

$$\begin{aligned} f_{M_n^\dagger}(y_n) &= O((-y_n)^{-(\alpha+1)}) \\ &\leq O\left(\delta^{2/\alpha}(r_n^{-\frac{1}{\alpha}})^{-(\alpha+1)}\right) \\ &= O(r_n^{1+1/\alpha}) = o(r_n). \end{aligned} \tag{4.60}$$

So the first term in (4.43) is $o(r_n)$.

The second term in (4.43) is

$$f_{S_n^*}(w_n) f_{M_n^\dagger}(y_n)\{1 - r_n((-y_n)^\alpha - 1)w_n\}.$$

Using the central limit result from Petrov (1975) we have $f_{S_n^*}(w_n)$ and $w_n f_{S_n^*}(w_n)$ are $O(1), \forall w_n$. We also have $f_{M_n^\dagger}(y_n) = o(r_n)$ for $(-y_n) > \delta^{2/\alpha} r_n^{-\frac{1}{\alpha}}$ by (4.60). We need only show that $f_{M_n^\dagger}(y_n)(-y_n)^\alpha = o(1)$ on this interval. But we have seen by (4.59),

$$(-y_n)^\alpha f_{M_n^\dagger}(y_n) = O((-y_n)^{-1}). \tag{4.61}$$

On $(-y_n) > \delta^{2/\alpha} r_n^{-\frac{1}{\alpha}}$,

$$(4.61) = O((\delta^{2/\alpha} r_n^{-\frac{1}{\alpha}})^{-1}) = O(r_n^{1/\alpha}) = o(1).$$

Altogether the second term in (4.43) is $o(r_n)$ when $(-y_n) > \delta^{2/\alpha} r_n^{-\frac{1}{\alpha}}$.

Thus $E_6 = o(r_n)$ or entire (4.43) is $o(r_n)$.

With (4.56) and (4.57), we have either for sufficiently large n ($\epsilon > 0$)

$$|E_3| < \frac{\epsilon}{6}\, r_n, \ \forall w, \ \forall y \geq e_n^\dagger \tag{4.62}$$

or the left-hand side of (4.43) is $o(r_n)$.

Proof for E_4 Like the proofs for E_1 and E_2 this argument holds irrespective of δ and hence is the same for both Case (a) and Case (b).

Recall we have $E_4 = f_{M_n^\dagger}(y)\left[\mathcal{N}'(w)\{1 + \frac{\kappa_3}{6\sigma^3\sqrt{n}}(w^3 - 3w)\} - f_{S_n^*}(w)\right]\{1 - r_n((-y)^\alpha - 1)w\}$.

Again we have $f_{M_n^\dagger}(y)$ bounded and also by Feller (1971), Chapter XVI, Section 2, Theorem 1 the term inside $[\ldots]$ is $o(\frac{1}{\sqrt{n}})$ uniformly in w. To handle the term $r_n((-y)^\alpha - 1)w$, we need show

(a) $\sup_y((-y)^\alpha - 1) f_{M_n^\dagger}(y)$ is bounded on $y \geq e_n^\dagger$ which we again have by Corollary 64.

(b) $\sup_w |w\{\mathcal{N}'(w) - f_{S_n^*}(w)\}| \to 0$ as $n \to \infty$ which follows from Petrov's (1975) central limit theorem under the same assumptions as Proposition 50 of Chapter 2.

(c.) $w^4 \mathcal{N}'(w)$ to be bounded so $\mathcal{N}'(w)\frac{\kappa_3}{6\sigma^3\sqrt{n}}(w^3 - 3w)w \to 0$ uniformly in w which we have by properties of the normal density.

Thus

$$E_4 = o(r_n) \ \ \forall y \leq e_n^\dagger \text{ and } \forall w.$$

or we may write for sufficiently large n ($\epsilon > 0$)

$$|E_4| < \frac{\epsilon}{6}\, r_n \quad \text{for } y \geq e_n^{\dagger} \text{ and } \forall w. \tag{4.63}$$

Proof for E_5 Finally we look at the term

$$E_5 = f_{M_n^{\dagger}}(y)\left[\frac{\mathcal{N}'(z)\kappa_3(u_n)}{6\sigma^3(u_n)\sqrt{n}}(z^3 - 3z) - \frac{\mathcal{N}'(w)\kappa_3}{6\sigma^3\sqrt{n}}(w^3 - 3w)\right].$$

Now the only difference between this Weibull case and the Fréchet case in Chapter 3 is that $f_{M_n^*}$ is now replaced by $f_{M_n^{\dagger}}$. Again $f_{M_n^{\dagger}}$ is uniformly bounded $\forall y$ by Proposition 63. The real argument depends on the term inside $[\ldots]$ being $o(r_n)$ but this term is the same as in the Fréchet case – involving the Normal argument. Thus, like the Fréchet proof, for sufficiently large n ($\epsilon > 0$)

$$|E_5| < \frac{\epsilon}{6} r_n \quad \text{or (4.43) is } o(r_n), \forall y_n \geq e_n^{\dagger}, \forall w. \tag{4.64}$$

Proof of E_{5b} Like the proofs for E_1 and E_2 this argument holds irrespective of δ and hence is the same for both Case (a) and Case (b).

Recall we have $E_{5b} = f_{M_n^{\dagger}}(y)\mathcal{N}'(w)\frac{\kappa_3}{6\sigma^3\sqrt{n}}(w^3 - 3w)[r_n((-y)^{\alpha} - 1)w]$.

Now, we have this result immediately since,

(a) $\sup_y((-y)^{\alpha} - 1)f_{M_n^{\dagger}}(y)$ is bounded on $y \geq e_n^{\dagger}$ which we again have by Corollary 64.

(b.) $w^4\mathcal{N}'(w)$ to be bounded so $\mathcal{N}'(w)\frac{\kappa_3}{6\sigma^3}(w^3 - 3w)w$ is uniformly bounded on w which we have by properties of the normal density.

Thus

$$E_{5b} = O\left(\frac{r_n}{\sqrt{n}}\right) = o(r_n) \quad \forall y \leq e_n^{\dagger} \text{ and } \forall w.$$

or we may write for sufficiently large n ($\epsilon > 0$)

$$|E_{5b}| < \frac{\epsilon}{6}\, r_n \ \text{ for } y \geq e_n^\dagger \text{ and } \forall w. \tag{4.65}$$

In conclusion, using (4.52), (4.53), (4.62), (4.63), (4.64), and (4.65), we have shown for sufficiently large n $\sum_{j=1}^{6} |E_j| \leq \epsilon\, r_n$ uniformly $\forall w$ and $y \geq e_n^\dagger$. Hence our result. □

Proof of Corollary 62

Using the result of Theorem 61 – equation (4.7) – to show (4.8) we need to prove that

$$\begin{aligned}
&\Big|\Big[f_{S_n^*}(w) f_{M_n^*}(y) - \mathcal{N}'(w)\{1 + \frac{\kappa_3}{6\sigma^3\sqrt{n}}(w^3 - 3w)\}\Psi'_\alpha(y) \\
&\times\{1 - [h_{-\rho}(-y)(1 - \frac{\alpha}{\alpha-\rho}(-y)^\alpha) + \frac{1}{\alpha-\rho}]g(a_n)\}\Big]\{1 - r_n((-y)^\alpha - 1)w\}\Big| \\
&\qquad\qquad\qquad\qquad\qquad\qquad = o(\max\{r_n, g(a_n)\}).
\end{aligned} \tag{4.66}$$

for $\forall w$ and $\forall y \geq e_n^\dagger = -\{-\gamma \log g(a_n)\}^{1/\alpha}$.

First we show (4.66) without the $r_n((-y)^\alpha - 1)w$ term.

Let

$$\begin{aligned}
A_n &= f_{S_n^*}(w), \\
A'_n &= \mathcal{N}'(w)\{1 + \frac{\kappa_3}{6\sigma^3\sqrt{n}}(w^3 - 3w)\}, \\
B_n &= f_{M_n^\dagger}(y),
\end{aligned}$$

and

$$B'_n = \Psi'_\alpha\{1 - [h_{-\rho}(-y)(1 - \frac{\alpha}{\alpha-\rho}(-y)^\alpha) + \frac{1}{\alpha-\rho}]g(a_n)\}.$$

Then we can write the left-hand side of (4.66) without the $r_n((-y)^\alpha - 1)w$ term as

$$\begin{aligned}
|A_n B_n - A'_n B'_n| &= |A_n B_n - A_n B'_n + A_n B'_n - A'_n B'_n| \\
&\leq |A_n|\, |B_n - B'_n| + |B'_n|\, |A_n - A'_n|
\end{aligned} \tag{4.67}$$

Now $|A_n|$ is bounded $\forall w$ by Petrov's result and $|B_n - B'_n| = o(g(a_n))$ $\forall y$ by Proposition 63. Thus the first term on the right-hand side of the inequality in (4.67) is $o(g(a_n))$, $\forall y$ and $\forall w$. For the second term in the inequality in (4.67), $|B'_n|$ is bounded $\forall y$ by Proposition 63 and $|A_n - A'_n| = o(r_n)$ by Feller (1971), Chapter XVI, Section 2, Theorem 1 uniformly in w. Thus this second term is $o(r_n)$, $\forall y$ and $\forall w$. Thus the right-hand side of (4.67) is $o(\max\{r_n, g(a_n)\})$, $\forall y$ and $\forall w$.

When we add in the $r_n((-y)^\alpha - 1)w$ term, we need to strength this to

(a.) $|wA_n|$ bounded – which we again have $\forall w$ by Petrov's result.

(b.) $|(-y)^\alpha + 1|\ |B_n - B'_n| = o(|\phi'(b_n)|)$ – which we have by Corollary 64, note now on $y \geq e^\dagger_n$.

(c.) $|(-y)^\alpha + 1|\ |B'_n|$ bounded – which we again have by Corollary 64, note now on $y \geq e^\dagger_n$.

(d.) $|w|\ |A_n - A'_n| = o(r_n)$ – which we again have $\forall w$ by Petrov's result.

Hence, we have result (4.8) for $y \geq e^\dagger_n = -\{-\gamma \log g(a_n)\}^{1/\alpha}$ and $\forall w$.

Chapter 5

DATA ANALYSIS OF CONTINENTAL US RAINFALL

5.1 Introduction

In Chapters 2, 3, and 4, we derived the higher order expansion term for the joint density of the sum and maximum of a series under the three different domains of attraction for the maximum. We also looked at a simulation project that illustrated how the higher order expansion model provided a better fit than the independent model. In Chapters 5, we present a data analysis on U.S. precipitation. The analysis of this series is pertinent to our particular higher order expansion since current work indicates that the extremes in this series have an impact on its mean.

Although weather at the local level and over a relatively short time span has always been relevant to the public interest, climate change – the shift in fundamental aspects of the environment on a continental or global stage with respect to longer time spans such as a century or millennium – has not interested the public until the emergence of greenhouse effect models. Essentially, the greenhouse models imply that man is changing his climate. These greenhouse models predict that increasing amounts of carbon dioxide and aerosols, massive deforestation, and other anthropogenic sources are having a variety of consequences on the global climate: an increase in the mean surface temperature, more precipitation in winter, more severe droughts, an increase in nighttime temperatures, greater proportions of precipitation from heavy events, and a decrease in day to day variability, see Kattenberg *et al.* (1996).

One of the leading groups to study climate change has been Thomas Karl and associates at National Climatic Data Center (NCDC), a branch of the National Oceanic and Atmospheric Association (NOAA). After investigating many indicators of climate change including the rise in the minimum temperature and in precipitation, Karl *et al.* (1996) maintain that the climate is becoming more extreme; that is, many of the changes they have observed in these climate change indicators are influenced by their extreme events. Their results, consistent with the greenhouse models, imply a change in the joint distribution of the means and extremes of these climate indicators.

The purpose of this chapter is to develop and then implement more precise tests of these assertions about the joint distribution of the means and extremes based on the new extreme value theory results developed in the previous chapters. In particular, we focus on the contiguous US rainfall series and the trends in the means and the extremes of that series.

5.2 Review of Precipitation Studies

Others besides Karl's group at NCDC have studied precipitation data – among them the group at NASA's Goddard Institute, see Karl and Knight (1998) for representative list – although throughout the literature there lacks a standard technique on how to do this. To date, the most comprehensive look at US precipitation is presented in Karl and Knight (1998) which followed Karl, Knight, Easterling, and Quayle (1996) and Karl, Knight, and Plummer(1995). Essentially their method is via summary statistics – certain weighted spatial averages. More specifically, in Karl and Knight (1998) the precipitation data was arithmetically averaged into $1^o \times 1^o$ grid cells. These cells were area weighted to calculate changes in precipitation for the nine regions they used. A national average was calculated by area weighting the nine regions. Finally, they used a nonparametric Kendall τ test ($\alpha = 0.05$) to detect significant trends. They found that since 1910 precipitation has increased in the continental US by about 10% but that one statement, which is often quoted, is an oversimplification of their results. By focusing on different quantiles of the precipitation data, they maintain the precipitation distribution itself has changed, making this precipitation increase fairly complex. They found: this increase is affected by both the frequency and intensity of precipitation; in all categories, the probability of precipitation on any given day has increased; precipitation intensity has increased only in the extremes; and in fact, the increase in total precipitation derived from the extreme events is higher relative to the moderate and low events. The last finding – that the increase in precipitation is primarily due to heavy/extreme daily precipitation events – is most interesting. In fact, they found that 53% of the rise in the total increase is due to a positive trend in the upper 10% of the probability distribution despite the fact these upper tail events only constitute about 35-40% of the total annual precipitation. This is seen predominantly in the summer and then spring and in general holds for all regions across the US except for the far West and the Southeast. In summary, since 1910 they are seeing a positive trend in total precipitation and in the number and intensity of extreme events. Moreover the increase in the upper percentiles is driving the increase in the total precipitation.

For the statistician, the tie between the trends of the means and extremes established by Karl and Knight (1998) raises many interesting questions. The most important is, what is the best way to model these seemingly dependent trends? Now Karl and Knight (1998) derive their findings on statistically simple models run on summary statistics of large data sets. What is unclear is exactly how some of the percentages are generated. For example, they say that 53% of the rise in total increase is due to the increase in the upper 10th percentile yet they do not provide a reliability measure for these percentages. Certainly the use of simple models based on summary statistics obscures some of the information in the system. Moreover while their results can be viewed by conditioning on the extremes or means (or any particular quantile) and looking at the change in the other, they do not vary both. They do not model the joint distribution.

Another important question raised about their methodology is can additional insights be gained by incorporating extreme value theory into the analysis? This argument is presented in Smith (1999). He looks at many of the issues in Karl and Knight (1998). He uses the same data set, basing his analysis on daily precipitation and minimum temperature. He begins by analyzing each station individually. For the extremes in these series, he bases the analysis on the peaks over threshold (POT) model in extreme value theory and estimates the trend of these extremes with maximum likelihood estimates. To integrate over the individual stations, he first spatially smoothes the estimates. Using standard kriging theory, he then obtains region averages with standard errors. His results are consistent with Karl and Knight (1998) but his methodology is entirely different. Note his analysis focused just on the extremes in the rainfall series, again, not on modeling the joint distribution of the means and extremes.

In this chapter, we want to investigate some of the ideas in Karl and Knight(1998) and Smith (1999), focusing on the relationship between the means and the extremes in the contiguous US rainfall series. To do so we use their joint distribution. In particular, we want to study the trends in these series. As Smith(1999) utilized the extreme value distributions to study the extremes in this rainfall series, we utilize the current results on their joint distribution to study the mean and maximum in this series. In doing so, we expect our analysis to represent an improvement due to the better methodology. There are two important results on the joint distribution of the means and extremes: Chow and Teugels' (1978) asymptotic independence result and the higher order expansion for the joint density developed in the previous chapters. Due to the empirical evidence of dependence between the mean and the maximum, indicated in Karl and Knight (1998), utilizing the higher order expansion for the joint density developed should prove an even better model for the annual rainfall series.

Overall, the statistical questions that emerge are: In practice, does this new higher order expansion for the joint density give better results than assuming independence? Using the new expansion model, what can we say about the trends in the mean and extremes of contiguous US rainfall? In particular, are there trends in the annual total rainfall, or the annual maximum rainfall? Are these trends increasing (decreasing) at the same rate or at different rates?

5.3 The Data

The data set is calculated from the United States Historical Climatalogical Network (USHCN), the same as used by Karl and Knight (1998) and Smith (1999). The USHCN was consolidated from the US Cooperative Observing Network and is maintained by the NCDC and the Carbon Dioxide Information and Analysis Center (CDIAC) of Oak Ridge National Laboratory. It was constructed to help address issues concerning climate change. It is a high quality, moderate size data set that includes minimum and maximum temperatures along with precipitation. The data set includes a total of 1221 stations within the contiguous US, although roughly between 180 and 190 stations are considered primary. In general, the records run from 1901 to 1997. The criteria for a station's inclusion include length of record, percentage of missing data, number of station moves, and spatial coverage. These data have gone through extensive quality control to correct for human error, time of observation bias, equipment adjustment, and urban warming. They do have a procedure to estimate missing data. For precipitation, it involves generating gamma random variables. For more information concerning the USHNC, please see the NOAA website (http://www.ncdc.noaa.gov/ol/climate/research/ushcn/ushcn.html).

Using the USHCN, a data set was constructed for 187 stations that contains annual total precipitation (units in hundredths of an inch), annual maximum precipitation (units in hundredths of an inch), number of wet days in each year, and number of recorded days in each year. To each station we assign one of five regions which were defined explicitly for this analysis, see Figure 5.1. Region 1 is the West Coast which consists of 16 stations in Washington, Oregon, and California. Region 2 is the Mountains which consisting of 50 stations in Idaho, Montana, Nevada, Utah, Wyoming, Colorado, Arizona, and New Mexico. Region 3 is the Plains which consists of 51 stations in North Dakota, South Dakota, Minnesota, Iowa, Nebraska, Kansas, Missouri, Illinois, Wisconsin, Oklahoma, and Texas. Region 4 is the Northeast which consists of 32 stations in Maine, Vermont, New Hampshire, Massachusetts, Connecticut, Rhode Island, New York, Pennsylvania, New Jersey, Ohio, Indiana, and Michigan. Finally, region 5 is the South which consists of 38 stations in Arkansas, Louisiana, Mississippi, Alabama, Florida, Georgia, South Carolina, North Carolina, Tennessee, Kentucky, Virginia, West Virginia, Maryland, and Delaware. Note these regions do not correspond to the nine regions used in Karl and Knight (1998) nor were they dictated by any other group's work. The regions were chosen so as to contain enough stations to get meaningful comparisons between the groups. We also wanted the regions to have some geographical identity and hopefully within the region some similarity in the climate.

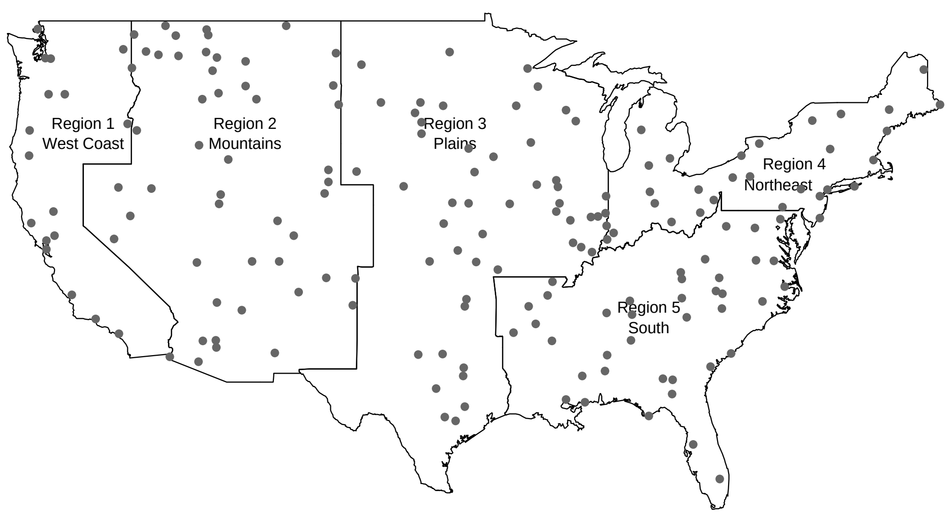


Figure 5.1: Map of the contiguous United States of America showing the five regions defined for this analysis and the 187 stations in the data set.

5.4 Models

Since one of the objectives of this analysis is to characterize the impact the expansion of the joint density has over the simple independent joint density model, we need to run both cases when fitting the annual total and annual maximum rainfall series: (1) The joint density of sum and the maximum assuming independence between them and (2) the expansion of the joint density, derived from the conditional density of the sum given the maximum, presented in the previous chapters.

5.4.1 Notation

Recall some of the notation introduced in Chapter 2. Let $X_1, \ldots, X_n$ be an *iid* sequence of random variables with common distribution function F, density f, mean μ, and variance σ^2.

Define $S_n = \sum_{i=1}^{n} X_i$ with the normalized version as $S_n^* = \frac{S_n - n\mu}{\sqrt{n\sigma^2}}$.

Let $M_n = \max\{X_1, \ldots, X_n\}$ with the normalized version being $M_n^* = \frac{M_n - \eta_n}{\psi_n}$ where $\psi_n > 0, \eta_n$ real. Note η_n and ψ_n are the appropriate normalizing constants needed for the standard extreme value limit. In Chapter 2, 3, and 4, we used a_n and b_n for these constants. We switch notation here for better consistency to the final applied form. Recall the definition $u_n = \psi_n v + \eta_n$ for some fixed v.

The conditional random variable X^* which is defined as X given $X \leq u_n$ has distribution function $\tilde{F_{u_n}}(x) = F(x)/F(u_n)$, density $\tilde{f_{u_n}}(x) = f(x)/F(u_n)$, mean $\mu(u_n)$, and variance $\sigma^2(u_n)$.

The random variable Y is defined as the exceedance over the threshold u_n; that is, $Y = (X - u_n)_+$. The conditional distribution function of Y given $X > u_n$ is $F_{u_n}(y) = \frac{F(u_n+y) - F(u_n)}{1 - F(u_n)}$, with conditional mean $m(u_n)$ and conditional variance $s^2(u_n)$.

Note this means that $E(X|X > u_n) = u_n + m(u_n)$ and $Var(X|X > u_n) = Var(Y|X > u_n) = s^2(u_n)$. Also recall that $\mathcal{N}'$ denotes the normal density.

5.4.2 Independent Case

Recall the result of Chow and Teugels (1978) on the asymptotic joint distribution of S_n^* and M_n^*: In our notation, we have (S_n^*, M_n^*) converge in distribution to a limit (S, M) where neither S nor M is degenerate if and only if F belongs in the domain of attraction of a stable law and an extreme value distribution. The limits, S and M, are independent unless F belongs in the domain of attraction of a stable law with index less than 2. It follows that if F has a finite variance and S and M are non-degenerate then S has a normal distribution and if $F(x) < 1, \forall x$ then M has a Gumbel distribution or a Fréchet distribution – Φ_α – with $\alpha \geq 2$. One of the extensions of Anderson and Turkman (1991) is that the above results hold when F belongs in the domain of attraction of a Weibull distribution. Thus current results show that the asymptotic joint distribution of S_n^* and M_n^* when F has a finite variance is just the product of their marginal asymptotic distributions.

Thus arguably a model for the joint density of S_n^* and M_n^* is

$$f_{S_n^*, M_n^*}(w, v) = f_{S_n^*}(w) \times f_{M_n^*}(v);$$

that is, we assume the asymptotic independence result for the joint density.

Now standard central limit theory (see Feller, 1971, Chapter XVI, Section 2, p. 533) shows that S_n^* has an asymptotically normal density. Recall $\mu_{S_n} = n\mu$ and variance $\sigma_{S_n}^2 = n\sigma^2$. Note these forms for μ_{S_n} and $\sigma_{S_n}^2$ fall from the independent assumption of the underlying random variables. We will comment on the credibility of this assumption for precipitation data in Section 5.5.3. Thus again assuming the asymptotic result, we use

$$f_{S_n^*}(w) = \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}} \quad \infty \leq w \leq \infty. \tag{5.1}$$

Standard extreme value theory (Thm 1.4.2, Leadbetter *et al.*, 1983, p. 11) provides the asymptotic density of M_n^*. Depending on which domain of attraction M_n^* belongs to, we have M_n^* has either an asymptotic Gumbel, Fréchet, or Weibull density. Since we do not know *a priori* to which domain of attraction M_n^* belongs, it is convenient to combine the three classical extreme value densities into the Generalized Extreme Value (GEV) density. So again if we use the asymptotic result for the density of M_n^*, we use:

$$f_{M_n^*}(v) = \exp\{-(1-kv)^{1/k}\} \times \{(1-kv)^{1/k-1}\}, \tag{5.2}$$

for $-\infty < v < \infty$ and $(1-kv) > 0$.

The case when $k = 0$ should be interpreted as the limiting Gumbel density

$$\lim_{k\to 0} f_{M_n^*}(v) = e^{-e^{-v}} \times e^{-v} \tag{5.3}$$

where $-\infty < v < \infty$.

Thus assuming the asymptotic results (independence, the normal density, and the GEV density), a model for a joint density of S_n^* and M_n^* is (5.1)×(5.2),

$$f_{S_n^*,M_n^*}(w,v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}} \times \exp\{-(1-kv)^{1/k}\} \times \{(1-kv)^{1/k-1}\} \tag{5.4}$$

or, when assuming that M_n^* is in the domain of attraction of the Gumbel density, we use (5.1)× (5.3),

$$f_{S_n^*,M_n^*}(w,v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}} \times e^{-e^{-v}} \times e^{-v}. \tag{5.5}$$

Finally in practice, to implement this independent case we fit the following approximation of the joint density to the annual maximum (M_n) and the annual total (S_n):

$$f_{S_n,M_n}(x,y) = \frac{1}{\sqrt{2\pi n\sigma^2}} e^{-\frac{(x-n\mu)^2}{2n\sigma^2}} \times \exp\{-(1-\frac{k(y-\eta_n)}{\psi_n})^{1/k}\} \times \{(1-\frac{k(y-\eta_n)}{\psi_n})^{1/k-1}\} \tag{5.6}$$

or when assuming that M_n is in the domain of the Gumbel density

$$f_{S_n,M_n}(x,y) = \frac{1}{\sqrt{2\pi n\sigma^2}} e^{-\frac{(x-n\mu)^2}{2n\sigma^2}} \times e^{-e^{-\frac{y-\eta_n}{\psi_n}}} \times e^{-\frac{y-\eta_n}{\psi_n}}. \tag{5.7}$$

Now we can estimate the parameters η_n, ψ_n, k, σ, and μ by fitting these approximations. Note we will use the same parametric form for both the independent and dependent case.

5.4.3 Expansion Case

In the previous chapters, we have derived a higher order expansion term for the joint density of the (annual) sum and the (annual) maximum under the 3 cases possible for the domain of attraction of the maximum: the Gumbel (Λ), Fréchet (Φ_α), and Weibull (Ψ_α). Now we wish to combine these into a higher order term for the Generalized Extreme Value (GEV) form and in doing so establish an alternative model to the asymptotic independence model for the joint density of M_n and S_n given in equation (5.6).

We will assume the same general set-up and notation as used in the previous sections.

In the previous chapters we focused on exhibiting explicitly the rate of convergence in the three different forms of the expansion. The actual expansion we now seek for the combined GEV form is slightly different since our purpose is now primarily for data analysis, specifically using the maximum likelihood method. What we want to focus on now is the additional information we can introduce into the sum and variance of the (annual) total by conditioning on the (annual) maximum. The conditional mean is extremely important since it contains information on the higher order term in the expansion.

Recall the derivation of the standardized form of the higher order expansion term in the previous chapters proceeded by

$$f_{S_n^*,M_n^*}(w,v) = f_{S_n^*|M_n^*}(w|v) f_{M_n^*}(v)$$

and consisted of three major steps: establishing the Edgeworth expansion of $f_{S_n^*|M_n^*}$, establishing the expansion for $f_{M_n^*}$, and deriving the approximations of the conditional mean and variance of $S_n^*|M_n*$. The key to the last step is in being able to write the conditional mean and variance as a function of the mean and variance of an exceedance above the threshold.

Here we do not wish to work with the standardized form. Since we are now focusing on the actual form to be used in practice, it is important to write it in the following form

$$f_{S_n,M_n}(x,y) = f_{S_n|M_n}(x|y)f_{M_n}(y) \tag{5.8}$$

Although here we proceed using equation (5.8), we make two important changes in this derivation so as to insure that the approximation to the joint density is positive at all data values. One is that we now look at the combined GEV form for the density of M_n. The second is that we do not establish expansions for the densities of M_n and $S_n|M_n$ but rather focus on finding the approximations to the conditional mean and variance of $S_n|M_n$. Our steps are thus:

Step A Establish using the GEV density for f_{M_n}

Step B Establish using the (conditional) normal approximation for $f_{S_n|M_n}$

Step C Compute the approximations to the conditional mean and variance of $S_n|M_n$.

Step D Plug the approximations to the conditional mean and variance found in step C into step B. Multiply steps A and B together to get an expansion to the joint density of S_n and M_n.

Note: Usually, we condition on $M_n = u_n$; that is, in (5.8) we set $y = u_n$ where, again, $u_n = \psi_n v + \eta_n$.

Step A: Here we assume $F \in \mathcal{D}(GEV)$; that is, there exist constants $\psi_n > 0, \eta_n$ real such that

$$\lim_{n\to\infty} F^n(u_n) = \lim_{n\to\infty} F^n(\psi_n v + \eta_n) = \exp\{-(1-kv)_+^{1/k}\}. \tag{5.9}$$

Thus if we assume the asymptotic result, we may approximate

$$F_{M_n}(u_n) \approx \exp\{-(1-\frac{k(u_n-\eta_n)}{\psi_n})^{\frac{1}{k}}\}$$

or, in terms of the density, we can use the approximation

$$f_{M_n}(u_n) \approx \exp\{-(1-\frac{k(u_n-\eta_n)}{\psi_n})^{1/k}\} \times \{(1-\frac{k(u_n-\eta_n)}{\psi_n})^{1/k-1}\}$$

$$\text{where } -\infty \leq \frac{u_n-\eta_n}{\psi_n} \leq \infty, (1-\frac{k(u_n-\eta_n)}{\psi_n}) > 0.$$

We justify using just the asymptotic GEV density without any expansion terms by looking at the results of the previous chapters: In each of case $(\Lambda, \Phi_\alpha, \Psi_\alpha)$, the higher order terms associated with the expansion of $f_{M_n}(u_n)$ did not play a role in the final higher order term of the joint density in the main theorems. Also in practice, people model annual maxima using the GEV density despite the existence of higher order approximations which theoretically improve on it.

Step B: Here we assume F lies in the domain of attraction of a stable law with index equal to 2 or, in other words, that the variance associated with F is finite. This gives us that $S_n|M_n$ has an asymptotic normal distribution. Thus if we assume the asymptotic result at the density level, we may approximate $f_{S_n|M_n}$ by

$$f_{S_n|M_n} = \mathcal{N}'(x|u_n) = \frac{1}{\sqrt{2\pi Var(S_n|M_n=u_n)}} \exp\left\{-\frac{(x-E(S_n|M_n=u_n))^2}{2\sqrt{Var(S_n|M_n=u_n)}}\right\}$$

where $u_n = \psi_n v + \eta_n$, and $E(S_n|M_n)$ and $Var(S_n|M_n)$ are determined in Step C.

Again we justify using just the asymptotic normal density without any expansion terms by looking at the results of the previous chapters. In each of the previous cases for $f_{S_n|M_n}$ we use the Edgeworth expansion of $\mathcal{N}'$. Now the Edgeworth correction is of order $O(\frac{1}{\sqrt{n}})$. We have seen that this is in general smaller than the other errors being considered and ultimately we drop these terms in the Edgeworth expansion in the final calculations in the previous chapters. Note when transforming from $S_n|M_n$ to $\tilde{S}_n$ we need to introduce the Jacobian. Now, the main contribution to the Jacobian is the ratio of σ^2 to $\sigma^2(u_n)$. With Step C we are in fact incorporating the information from that ratio already. Thus, we need not consider the Jacobian as a separate term.

Step C: Here we wish to establish $E(S_n|M_n)$ and $Var(S_n|M_n)$ for the combined GEV case. Recall the formulae for these expressions from the previous chapters:

$$\begin{aligned} E(S_n|M_n = u_n) &= (n-1)\mu(u_n) + u_n \\ &= (n-1)[\frac{1}{F(u_n)}\{\mu - (1 - F(u_n))(m(u_n) + u_n)\}] + u_n \end{aligned} \tag{5.10}$$

where $E(X|X > u_n) = u_n + m(u_n)$ and $F(u_n)$ is the distribution function of X evaluated at u_n.

$$\begin{aligned} Var(S_n|M_n = u_n) &= (n-1)\sigma^2(u_n) \\ &= (n-1)\Big\{\frac{1}{F(u_n)}\left[E(X^2) - E(X^2|X > u_n)(1 - F(u_n))\right] \\ &\quad - \left[\frac{1}{F(u_n)}\{E(X) - E(X|X > u_n)(1 - F(u_n))\}\right]^2\Big\} \\ &= (n-1)\Big\{\frac{\mu^2 + \sigma^2}{F(u_n)} - \frac{1 - F(u_n)}{F(u_n)}\{(m(u_n) + u_n)^2 + s^2(u_n)\} \\ &\quad + \{\frac{\mu}{F(u_n)} - \frac{((1 - F(u_n))(u_n + m(u_n))}{F(u_n)}\}^2\Big\} \end{aligned} \tag{5.11}$$

where $s^2(u_n) = Var(X|X > u_n) = Var(Y|X > u_n)$.

In the above formulae we need to solve for $m(u_n)$, $s^2(u_n)$, and $F(u_n)$ in the combined GEV form. Again, the key information involves the mean and variance of exceedances above a threshold.

In this step we cannot cut off the higher order terms in the individual cases and immediately get the results. This is because the terms which are important – particularly in $m(u_n)$ – depend on the underlying tail properties of F and hence in which particular domain of attraction F belongs.

To see the general case for $m(u_n)$ and $\sigma^2(u_n)$, we need to exploit the relationship between the GEV distribution and the Generalized Pareto distribution so as to be able to write the mean and variance of the exceedances in terms of the equivalent GEV parameters.

Recall the GP distribution:

$$G(y;k,\gamma) = \begin{cases} 1-(1-ky/\gamma)^{1/k} & \text{if } k \neq 0, \gamma > 0 \\ 1-\exp(-y/\gamma) & \text{if } k = 0, \gamma > 0 \end{cases}$$

where the range of y is $0 < y < \infty (k \leq 0)$ or $0 < y < \gamma/k \quad (k > 0)$ and which has mean $E(Y) = \frac{\gamma}{1+k}$ where $k > -1$ and variance $Var(Y) = \frac{\gamma^2}{(1+k)^2(1+2k)^2}$ where $k > -\frac{1}{2}$.

Since we are assuming that $F \in \mathcal{D}(GEV)$, we have the relationship in equation (5.9). We may define η_n such that $1 - F(\eta_n) = \frac{1}{n}$. As a consequence

$$\lim_{n\to\infty} \frac{1-F(\psi_n v + \eta_n)}{1-F(\eta_n)} = (1-kv)_+^{1/k}. \tag{5.12}$$

So given the above GEV assumptions, we show the limiting distribution of the exceedances is the GP distribution with parameters k and $\gamma(u_n)$ where the function $\gamma(u_n) = \psi_n(1-kv)$:

We have $\lim_{n\to\infty} P(Y > y | X > u_n)$ is

$$\begin{aligned}
\lim_{n\to\infty} \frac{1-F(\gamma(u_n)y + u_n)}{1-F(u_n)} &= \lim_{n\to\infty} \frac{1-F(\gamma(u_n)y + \psi_n v + \eta_n)}{1-F(\eta_n)} \times \frac{1-F(\eta_n)}{1-F(\psi_n v + \eta_n)} \\
&= \lim_{n\to\infty} \frac{1-F(\psi_n(v + (1-k)y) + \eta_n)}{1-F(\eta_n)} \times \frac{1-F(\eta_n)}{1-F(\psi_n v + \eta_n)}.
\end{aligned}$$

Using (5.12),

$$\begin{aligned}
\lim_{n\to\infty} \frac{1-F(\gamma(u_n)y + u_n)}{1-F(u_n)} &= \frac{(1-k(v+(1-kv)y))^{1/k}}{(1-kv)^{1/k}} \\
&= (1-ky)^{1/k} \\
&= 1-G(y;k,\gamma), \text{where } \ \gamma = \gamma(u_n) = \psi_n(1-kv)).
\end{aligned}$$

Hence, we have shown that the limiting distribution of the exceedances is the GP distribution with parameters k and $\gamma(u_n)$. Thus the derivation of the approximate values of $m(u_n)$ and $s^2(u_n)$, the mean and variance of the exceedances, follows directly from the limiting GP distribution. We use the following approximations:

$$m(u_n) = \frac{\gamma(u_n)}{1+k} = \begin{cases} \frac{\psi_n(1-k(u_n-\eta_n)/\psi_n)}{1+k} & \text{if } k > -1 \\ \psi_n & \text{if } k = -1 \end{cases} \quad (5.13)$$

and

$$s^2(u_n) = \frac{\gamma^2(u_n)}{(1+k)^2(1+2k)} = \begin{cases} \frac{(\psi_n(1-k(u_n-\eta_n)/\psi_n))^2}{(1+k)^2(1+2k)} & \text{if } k > -\frac{1}{2} \\ \psi_n^2 & \text{if } k = -\frac{1}{2}. \end{cases} \quad (5.14)$$

Finally, using the limiting GEV distribution from (5.9) we solve for an approximation of the combined form of $F(u_n)$:

$$F(u_n) = \begin{cases} \{\exp(-(1-kv)_+^{1/k})\}^{1/n} & \text{if } k \neq 0 \\ \{\exp(-\exp(-v))\}^{1/n} & \text{if } k = 0\text{, Gumbel case.} \end{cases} \quad (5.15)$$

In conclusion, if we substitute the approximations (5.13), (5.14), and (5.15) into formulae (5.10) and (5.11), then we have the approximations of the conditional mean and variance of $S_n|M_n$.

Step D:

So in practice, to implement this expansion case we fit the following approximation of the joint density to the annual maximum (M_n) and the annual total (S_n):

$$\begin{aligned} f_{S_n,M_n}(x,u_n) \approx & \frac{1}{\sqrt{2\pi Var(S_n|M_n=u_n)}} \exp\left(-\frac{(x-E(S_n|M_n=u_n))^2}{2\sqrt{Var(S_n|M_n=u_n)}}\right) \times \\ & \exp\{-(1-\frac{k(u_n-\eta_n)}{\psi_n})^{1/k}\} \times \{(1-\frac{k(u_n-\eta_n)}{\psi_n})^{1/k-1}\}\frac{1}{\psi_n} \end{aligned} \quad (5.16)$$

or when assuming that M_n is in the domain of the Gumbel density

$$f_{S_n,M_n}(x,u_n) = \exp\left(-\frac{(x-E(S_n|M_n=u_n))^2}{2\sqrt{Var(S_n|M_n=u_n)}}\right) \times e^{-e^{-\frac{u_n-\eta_n}{\psi_n}}} \times e^{-\frac{u_n-\eta_n}{\psi_n}}\frac{1}{\psi_n}. \quad (5.17)$$

where $E(S_n|M_n=u_n)$ and $Var(S_n|M_n=u_n)$ are given in (5.10) and (5.11), respectively. Now we can again estimate the parameters η_n, ψ_n, k, σ, and μ by fitting these expansions.

5.5 Statistical Procedure

5.5.1 Level of Analysis

Recall that the data are collected at 187 stations throughout the US. At each station, there is a single rain gauge. This presents a problem inherent to all precipitation analysis. Specifically, since rain gauges are not at every spot, it is difficult to objectively estimate how much rain has fallen over a large area. This makes inference about precipitation over a large area challenging. (Del Genio *et al.*, 1999, http://www.giss.nasa.gov/research/intro/delgenio.02/index.html)

We begin the analysis by fitting both forms of the joint density of S_n and M_n on an individual station basis, in the same manner Smith (1999) begins his analysis on extremes. This is a marked distinction from Karl and Knight (1998) who began by averaging over a large number of stations. The obvious drawback is the increased variability at the individual station level. This may impact the ability to make inferences about the precipitation trends at the individual station level.

To combine the results from the individual stations, we use the two different methods: (1) relevant tests of hypotheses – likelihood ratio tests to compare models and t-tests to test for significance in individual trend parameters and (2) a spatial smoothing method provided by Smith (2000). This spatial method will lead to both a national average for the estimates of the trends and a map of a smoothed version of the relevant trend estimates.

5.5.2 Method

Based on the two competing forms of the joint density of S_n and M_n found in (5.6) and (5.16), we can easily write down the likelihood functions and hence calculate the maximum likelihood estimates. This is a standard method and is justified for the extreme value parameters when the shape parameter in the extreme value density is less than 0.50; i.e., $k < \frac{1}{2}$. In that case, the maximum likelihood estimators exist and have their classical asymptotic properties, see Smith (1985). Under very general conditions, the maximum likelihood estimates for the normal parameters exist and also have the standard asymptotic properties, see Cox and Hinkley (1974), p. 284.

The likelihood functions are maximized – actually the negative log likelihood functions are minimized – using a non-linear optimization routine, specifically a quasi-Newton algorithm based on an algorithm number 21 of Nash (1979), p. 159-161. Quasi-Newton algorithms are an effective class for optimizations. When they converge, they do so relatively rapidly. Also, a beneficial by-product of using such an algorithm is an estimate of the Hessian matrix from which we may estimate the standard errors of the parameter estimates. For more details, see Nash (1979). The quasi-Newton algorithm in the form implemented here employs an approximation to the gradient of the function to be optimized. Here we use a first order approximation to the gradient:

$$f'(x_o) = \frac{f(x_o + h) - f(x_o)}{h} \quad \text{where we use } h = 10^{-6}.$$

The program runs this particular analysis efficiently when given reasonable starting values. We use method of moment estimators for the starting values except for the trend parameters which are initialized to 0. From the program we can obtain the maximum likelihood estimates with the estimated standard errors and, of course, the value of the log likelihood function evaluated at it maximum.

5.5.3 Form of the Log Likelihood Functions

In practice, we do not maximize the likelihood functions but rather minimize the negative log likelihood function. The following formulae of the negative log likelihoods come directly from equations (5.6) and (5.16) for the independent and expansion case with one adaptation.

Adaptation An adaptation is necessary due to the missing data within each year. Not every year has a complete daily record; i.e.,the number of days in a year – actually the number of recorded days in a year – is not necessarily 365. This presents a problem in the parameters to be estimated. Recall four of the parameters depend on n: μ_{S_n}, $\sigma^2_{S_n}$, η_n, and ψ_n. To compensate, we need to adjust for the different values of n throughout a station's record.

Denote

$$N = \text{number of years on record for the station and}$$

$$n_i = \text{number of days recorded for year } i \text{ at the station.}$$

For the parameters associated with S_n, we simply use n_i in the formulae; i.e., define S_{n_i} as the total in year i and let the mean total rainfall in year i (based on n_i observations) be

$$\mu_{S_{n_i}} = n_i \mu$$

and the variance of the total rainfall in year i be

$$\sigma^2_{S_{n_i}} = n_i \sigma^2.$$

Note in our formulae for $\sigma^2_{S_{n_i}}$ and $\mu_{S_{n_i}}$ we assume that the daily precipitation series is independent. Although we do not try to justify this assumption, we claim that our objectives are not unduly compromised by this. Recall our objectives are primarily comparisons: differences between the fit in the independent case and the expansion cases and between the estimates of the trend parameters for the different models. Since we assume independence between daily rainfall for each model we run, we expect that any effect this assumption may have will cancel out in the comparison; that is, essentially each model will be handicapped in the same way so the comparisons should still be valid. Also for our particular application of precipitation data, current findings indicate that although rainfall events are dependent from day to day, rainfall amounts are not necessarily. For more details, see Stern and Coe (1984) and Smith (1994).

For the parameters associated with M_n, in particular η_n and ψ_n, the adaptation involves the distribution itself. The basic formulation of the distribution of the maximum for *iid* random variables begins with:

$$P(M_n \leq y) = P(X_1 \leq y, \ldots, X_n \leq y) = [F(y)]^n.$$

Now standard extreme value theory (Thm 1.4.2, Leadbetter *et al.* (1983)) tells us $F \in \mathcal{D}(GEV)$; that is, $M_n \xrightarrow{\mathcal{D}} GEV$, if we can find series of norming constants, η_n and ψ_n, such that

$$\lim_{n\to\infty} F^n(\psi_n v + \eta_n) = H(v)$$

where H is the GEV distribution.

In practice, when we approximate $P(M_n \leq y)$ by its asymptotic GEV distribution, we use $y = \psi_n v + \eta_n$ and set

$$[F(y)]^n = H(\frac{y - \eta_n}{\psi_n}). \tag{5.18}$$

Now remember that the problem with the data set is that the annual record for each year is not consistently based on $n = 365$ days; that is, n in equation (5.18) is not fixed at 365 but varies throughout the years due to missing data. If we define M_{n_i} as the maximum in year i (based on n_i days), the parameters associated with M_{n_i} are η_{n_i} and ψ_{n_i} – they vary from year to year. In our application we want to estimate the parameters η_n and ψ_n when $n = 365$; that is, we want to estimate the parameters associated with the *annual* maximum. So what we need is to adapt equation (5.18) for n_i, the actual number of days recorded in year i.

Define the distribution of the maximum rainfall for year i (based on n_i days) as

$$F_{M_{n_i}}(y) = P(M_{n_i} \leq y) = [F(y)]^{n_i}. \tag{5.19}$$

Through our set-up, we can relate $F_{M_{n_i}}(y)$ to $F_{M_{n=365}}(y)$

$$F_{M_{n_i}}(y) = [F(y)]^{n_i} = \{[F(y)]^n\}^{\frac{n_i}{n}} \text{where } n = 365. \tag{5.20}$$

Now we substitute (5.18) into (5.20) to find the asymptotic result we use in this application

$$F_{M_{n_i}}(y) = \{[F(y)]^n\}^{\frac{n_i}{n}} = \{H(\frac{y-\eta_n}{\psi_n})\}^{\frac{n_i}{n}}.$$

In the density form, we have

$$f_{M_{n_i}}(y) = \frac{n_i}{n} \times \{H(\frac{y-\eta_n}{\psi_n})\}^{\frac{n_i}{n}-1} \times H'(\frac{y-\eta_n}{\psi_n}) \times \frac{1}{\psi_n} \text{ where } n = 365. \tag{5.21}$$

So in calculating the negative log likelihood functions from equations (5.6) and (5.16) for the independent and expansion case, we replace the form of f_{M_n} by equation (5.21). This will allow us consistently to be estimating the parameters for the *annual* maximum while compensating for the fact the *annual* record within each station is not consistently based on 365 days. In other words, the adaptation in equation (5.21) allows us to write the density of M_{n_i}, the maximum in year i based on n_i days, as a function of the parameters for the maximum based on $n = 365$ days. We now drop the subscript n in the parameters η_n and ψ_n since with this adaptation we can take n – associated with the parameters – to be 365 consistently. Note we will still need to keep the density of the annual maximum a function of n_i but NOT the parameters.

Set-up Now we may write down the negative log likelihood functions that are used in this analysis. First we define

$$l_{S,M} = \text{likelihood function for the joint density of } S_{n_i} \text{ and } M_{n_i},$$

$$l_M = \text{likelihood function for the density of } M_{n_i},$$

$$l_S = \text{likelihood function for the density of } S_{n_i} \text{ and}$$

$$l_{S|M} = \text{likelihood function for the conditional density of } S_{n_i} \text{ given } M_{n_i}.$$

Independent case For the independent case, the negative log likelihood function of the joint density of S_{n_i} and M_{n_i} is:

$$\begin{aligned} l_{S,M}(\mu_{S_{n_i}}, \sigma^2_{S_{n_i}}, \eta, \psi, k; x, y) &= l_{S_{n_i}}(\mu_{S_{n_i}}, \sigma^2_{S_n}; x) + l_{M_{n_i}}(\eta, \psi, k; y) \\ &= N \log \sqrt{2\pi} + \sum_{i=1}^{N} \log \sqrt{\sigma^2_{S_{n_i}}} + \sum_{i=1}^{N} \frac{(x_i - \mu_{S_{n_i}})^2}{2\sigma^2_{S_{n_i}}} \\ &\quad - \sum_{i=1}^{N} \log(\frac{n_i}{365}) + \sum_{i=1}^{N} \frac{n_i}{365}(1 - \frac{k(y_i - \eta)}{\psi})^{1/k} \\ &+ (1 - \frac{1}{k}) \sum_{i=1}^{N} \log(1 - \frac{k(y_i - \eta)}{\psi}) + N \log \psi \end{aligned} \tag{5.22}$$

where we use $\mu_{S_{n_i}} = n_i \mu$ and $\sigma^2_{S_{n_i}} = n_i \sigma^2$ and the constraints are $\sigma > 0$, $\psi > 0$ and $(1 - \frac{k(y_i - \eta)}{\psi}) > 0$.

In the Gumbel case, the above formula simplifies to

$$\begin{aligned} l_{S,M}(\mu_{S_{n_i}}, \sigma^2_{S_{n_i}}, \eta, \psi, k; x, y) &= N \log \sqrt{2\pi} + \sum_{i=1}^{N} \log \sqrt{\sigma^2_{S_{n_i}}} + \sum_{i=1}^{N} \frac{(x_i - \mu_{S_{n_i}})^2}{2\sigma^2_{S_{n_i}}} \\ &\quad - \sum_{i=1}^{N} \log(\frac{n_i}{365}) + \sum_{i=1}^{N} \frac{n_i}{365} \exp\{-(\frac{y_i - \eta}{\psi})\} + \sum_{i=1}^{N} \frac{y_i - \eta}{\psi} + N \log \psi \end{aligned} \tag{5.23}$$

where the constraints are $\sigma > 0$, $\psi > 0$.

Note in the independence case, the estimates are the same as if M_n and S_n are fit separately.

Expansion case For the expansion case, the negative log likelihood function of the joint density of S_n and M_n is:

$$\begin{aligned} l_{S,M}(\mu_{S_{n_i}}, \sigma^2_{S_{n_i}}, \eta, \psi, k; x, y) &= l_{S_{n_i}|M_{n_i}}(\mu_{S_{n_i}}, \sigma^2_{S_{n_i}}, \eta, \psi, k; x, y) + l_{M_{n_i}}(\eta, \psi, k; y) \\ &= N \log \sqrt{2\pi} + \sum_{i=1}^{N} \log \sqrt{\sigma^2_{S_{n_i}|M_{n_i}}} + \sum_{i=1}^{N} \frac{(x_i - \mu_{S_{n_i}|M_{n_i}})^2}{2\sigma^2_{S_{n_i}|M_{n_i}}} \end{aligned}$$

$$
\begin{aligned}
& -\sum_{i=1}^{N}\log(\frac{n_i}{365})+\sum_{i=1}^{N}\frac{n_i}{365}(1-\frac{k(y_i-\eta)}{\psi})^{1/k} \\
+ \quad & (1-\frac{1}{k})\sum_{i=1}^{N}\log(1-\frac{k(y_i-\eta)}{\psi})+N\log\psi \qquad (5.24)
\end{aligned}
$$

where

$$\mu_{S_{n_i}|M_{n_i}} = E(S_{n_i}|M_{n_i}=u_n)$$

and is given in equation (5.10),

$$\sigma^2_{S_{n_i}|M_{n_i}} = Var(S_{n_i}|M_{n_i}=u_n)$$

and is given in equation (5.11). The constraints on the parameters are the same as in the independent case.

For the Gumbel case, the above formula simplifies to

$$
\begin{aligned}
l_{S,M}(\mu_{S_{n_i}},\sigma^2_{S_{n_i}},\eta,\psi,k;x,y) \quad = \quad & N\log\sqrt{2\pi}+\sum_{i=1}^{N}\log\sqrt{\sigma^2_{S_{n_i}|M_{n_i}}}+\sum_{i=1}^{N}\frac{(x_i-\mu_{S_{n_i}|M_{n_i}})^2}{2\sigma^2_{S_{n_i}|M_{n_i}}} \\
& -\sum_{i=1}^{N}\log(\frac{n_i}{365})+\sum_{i=1}^{N}\frac{n_i}{365}\exp\{-(\frac{y_i-\eta}{\psi})\}+\sum_{i=1}^{N}\frac{y_i-\eta}{\psi}+N\log\psi
\end{aligned}
\qquad (5.25)
$$

with the same constraints as in the independent case.

Note to make the program run more effectively, we rescaled the trend parameters β and γ in the likelihood functions to $\frac{\beta}{1000}$ and $\frac{\gamma}{1000}$.

5.5.4 Maximum Likelihood in Action

There are many advantages to the maximum likelihood method both in modeling and in model selection. One advantage in modeling is the flexibility of adding trends, actually any covariate, to the parameters. A simple modification of the likelihood function is all that is necessary – here, changing any parameter of interest into a function of time. In the above log likelihood functions, we allow μ and η to be fit with a time component.

Another advantage of the maximum likelihood method is the ability to test significance between models using a likelihood ratio test. Thus for any two models we wish to compare, the ratio of their maximum likelihood functions measures the weight of evidence for model 1 against model 2 provided by the data. Let $L(X;\theta)$ be defined as the likelihood function for $X = (X_1, \ldots, X_n)$ and θ is the parameter space. Under a suitable test of significance we let $\lambda(x)$ be the likelihood ratio test; i.e.,

$$\lambda(X) = \frac{sup_{model1}\{L(X,\theta)\}}{sup_{model2}\{L(X,\theta)\}}.$$

Under suitable regularity conditions (see Cox and Hinkley, 1974, p 313) for large n

$$-2\log\lambda(X) \text{ has an approximate } \chi^2 \text{ distribution.}$$

When testing nested models with the same underlying distribution where model 1 has an r dimensional parameter space and model 2 has a q ($r > q$) dimensional parameter space, the degrees of freedom for the χ^2 test are $r - q$, see (Cox and Hinkley, 1974, p. 323). When choosing between models – i.e. let f_1 and f_2 be two p.d.f.s that are candidates for the p.d.f of *iid* random variables $X_1, \ldots, X_n$ – if the likelihood ratio is suitable, the degrees of freedom are 1 (see Cox and Hinkley, 1974, Example 9.22). Note the regularity conditions necessary for this asymptotic distribution of log $\lambda(X)$ are usually those sufficient for asymptotic normality and asymptotic consistency of the maximum likelihood estimates. So when the standard asymptotic properties for the normal and extreme value parameters hold, it follows that the likelihood ratio test statistic is asympototically χ^2. See Cox and Hinkley, 1974, Section 9.3, p. 311 for further details. Thus using the asymptotic theory we can perform tests of significance between independent models and expansion models. We can also test within both the independent case and the expansion case nested models. Specifically, we can test for adding the trend parameters to the models using this method.

Further, the asymptotic normality of the maximum likelihood estimators (again see Cox and Hinkley, 1974, p. 294, and Smith, 1985) allow us to evaluate the parameter estimates with their estimated standard errors. Asymptotic normality allows us to calculate confidence intervals for the parameters and for the difference between parameters. This will allow us to perform t-tests for testing if individual trend estimates are significantly different from zero. We will also be able to test for a significant difference between those parameters.

5.6 Data Analysis

5.6.1 Descriptive Statistics

Figures (5.2) and (5.3) show box-plots for the annual maximum and annual total rainfall, respectively, for the five regions. Overall, region 2 – the Mountain states – has both the lowest annual maximum and total rainfall. Region 5 – the South – is at the other end: the highest annual maximum and total rainfall. Regions 3 and 4 – the Plains and the Northeast – are very similar. The Northeast has slightly higher annual total rainfall but the Plains have slightly higher annual maximum rainfall. In other words, the Plains are more likely to see a very heavy rainfall but have less overall rainfall than the Northeast. Finally, the West coast is harder to categorize with the stations covering the range from region 2 – the Mountains – up through region 4 – the Northeast.

As to any trend in the annual maximum precipitation series throughout the past century, we first look at the average annual maximum rainfall across the stations for each year from 1901 to 1997, Figure 5.4. We see that the average annual maximum ranges from 2.0 to 2.8 inches. Even after averaging over all the stations, there is still a great amount of variability from year to year. Overall there is a slight decrease from 1901 to 1930, then the process evens out until 1950 where there is a upwards shift. A least squares fit shows a slope of 0.14868 (s.e. 0.04942) with associated p-value 0.0034

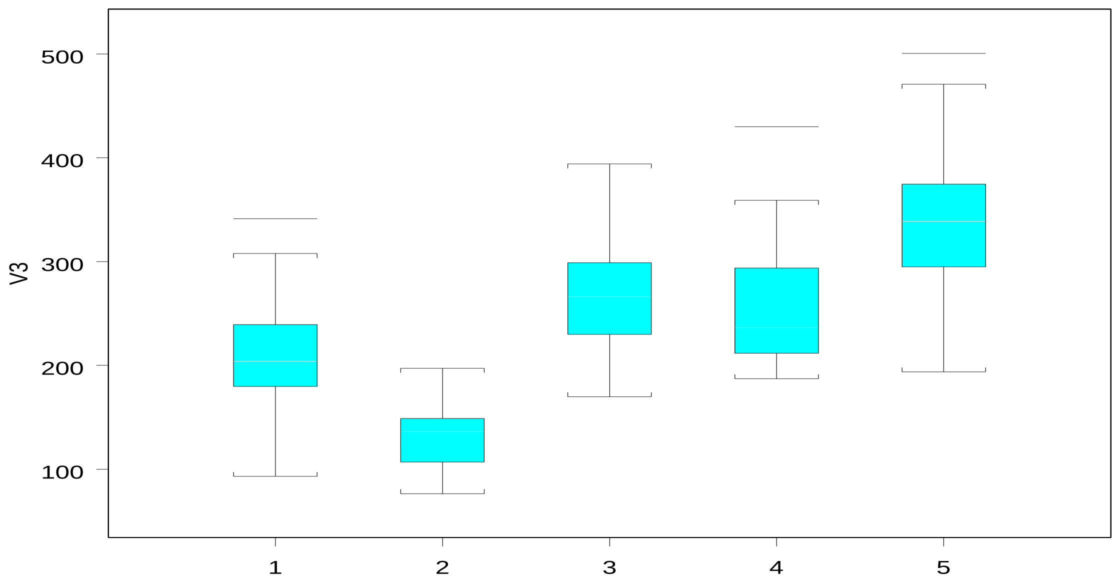


Figure 5.2: Box-plot for the annual maximum rainfall in the five regions. The units are in hundredths of an inch.

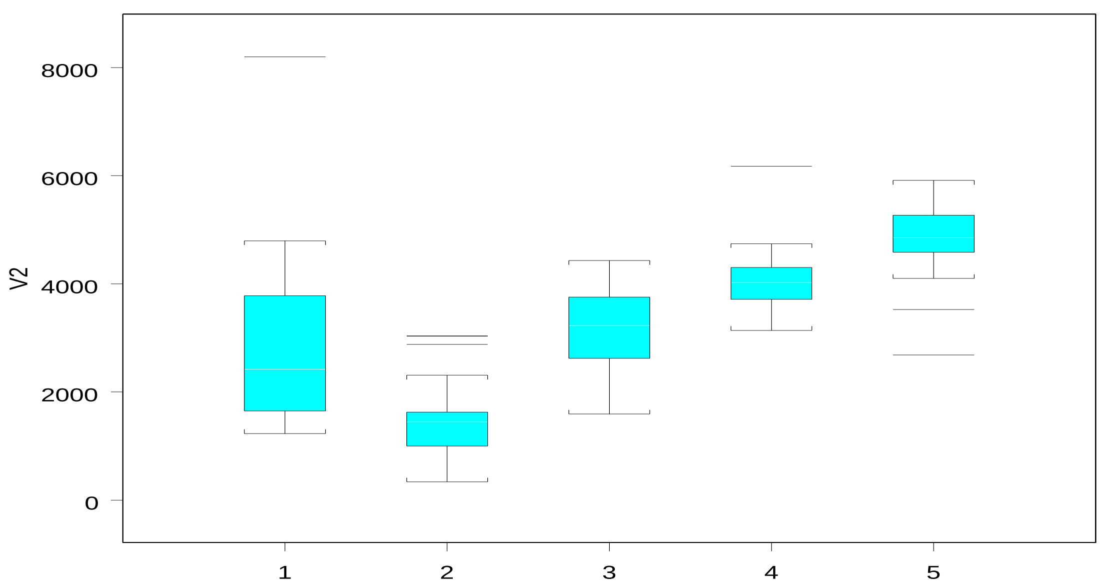


Figure 5.3: Box-plot for the annual total rainfall in the five regions. The units are in hundredths of an inch.

which implies roughly a 6.5% increase in the past century. Note that this standard error calculation assumes independence from year to year.

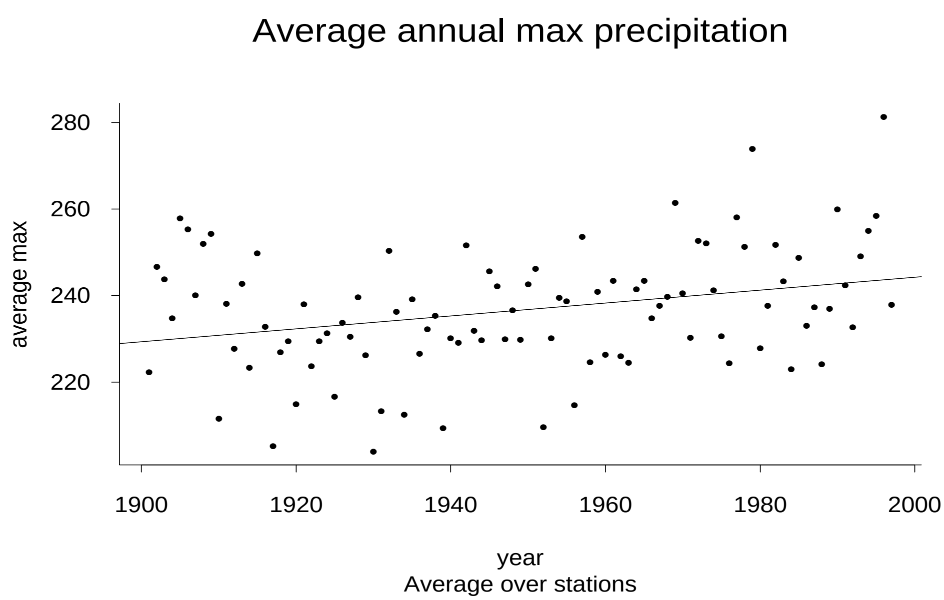


Figure 5.4: The average annual maximum rainfall across the US each year from 1901 to 1997. The units are in hundredth of an inch. The line represents a least square fit to the data.

If we look at the average annual precipitation across these 187 stations from 1901 to 1997, see Figure 5.5, we do not see any obvious trend. For the average annual total series, the data are relatively stable, between 24 and 37 inches. There is more dispersion after 1950 and this may represent an increase after 1980. Overall the graph suggests a possible change in the annual total precipitation from the first half of the century to the latter half. Here a least squares fit shows a slope of 2.566 (s.e. 0.924) with associated p-value .0066. This implies roughly an 8.5% increase in the past century. Again, this standard error calculation assumes year to year independence. To compare it with the average annual maximum series, each are relatively stable from the 1930s to the 1960s and then show possible increases after that. Both have significant least squares slope parameter. It is notable that the annual maximum has a smaller p-value associated with its least squares trend than the annual total trend. Given the increased variability in the maximum series this is an interesting result.

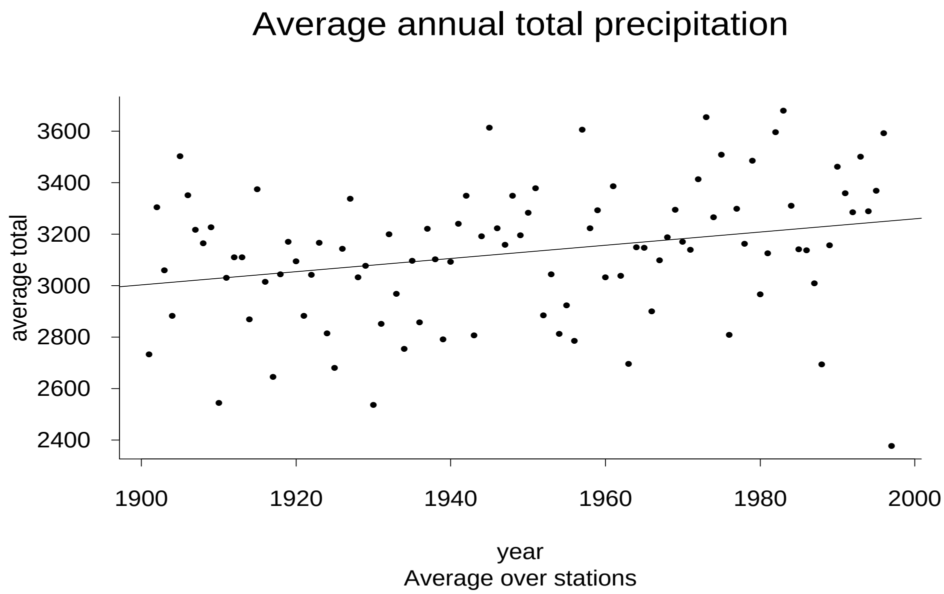


Figure 5.5: The average annual total rainfall across the US each year from 1901 to 1997. The units are in hundredth of an inch. The line represents a least square fit to the data.

Now since we analyze the data on a station level, we look at the above graphs for a few individual stations. In particular we select a station from each region. At the station level, we expect more variability in the graphs.

Figure 5.6 show the plots of the annual total rainfall against the annual maximum rainfall for each of the five stations: Berkeley, CA (region 1); Caldwell, ID (region 2); Le Mars, IA (region 3); Eastport, ME (region 4); and Savannah, GE (region 5). The Pearson correlation coefficient associated with each graphs (with associated p-values in parenthesis) are $0.4724(7.2 \times 10^{-6}), 0.3839(6.7 \times 10^{-5}), 0.3742(8.0 \times 10^{-5}), 0.4757(4.2 \times 10^{-7})$, and $0.4624(1.4 \times 10^{-6})$, respectively. We see on a station level that the annual maximum and annual total are positively correlated. This bolsters the argument for using the expansion model which has the higher order term that models this dependent structure.

Figure 5.7 give the graphs of the annual maximum precipitation series for the five stations in the previous paragraph. Note at present we look only at the scatterplots in Figure 5.7, not the lines that are fit. In Section 5.6.4, we discuss the trends that are fit. Not only is there a greater variation at the station level than the national average (Figure 5.4) but also there is no overall pattern among these stations. The Berkeley annual maximum rainfall series has a possible increase in both the average and in dispersion throughout the century. The Caldwell series shows a slight decrease in the average with less dispersion in the latter half of the century. Le Mars is stable throughout the century except for 3 outliers in the last 15 years. Eastport has the greatest amount of dispersion throughout the century with a possible increase in the average. Finally, Savannah has a relatively stable average with possible decrease in dispersion. Overall no clear patterns in the annual maximum precipitation series emerge from looking at these 5 stations. Clearly at a station level there is a greater amount of variability in the annual maximum rainfall series.

Figure 5.8 shows the graphs of the annual total precipitation series for the 5 stations. Again we see the greater variability in the station series as opposed to the national average series (Figure 5.5). Again there is no overall consensus to a pattern in these series, although these five stations' annual total rainfall series are more similar than their annual maximum rainfall series. In Berkeley (5.8c), Eastport (5.8c), Savannah (5.8c), and even to a lesser extent Caldwell (5.8c), we see a possible increase in the annual total rainfall with a possible increase in dispersion throughout the century in Berkeley, Eastport, and Savannah. Le Mars(5.8d) is the exception with a possible decrease in the average annual total rainfall and stable dispersion. In other words, at the station level there is more visual evidence for a trend in the annual total precipitation series. Due to the significant amount of variability, particularly at the station level, proving that these trends are significant could be problematic.

In conclusion from the descriptive statistics,

1. At the national average level, the annual maximum precipitation series depicts a possible increase towards the end of the second half of the century. Certainly a least squares fit to a linear trend is significant.

2. At the national average level, the annual total rainfall series appears to behave differently from the first half of the century to the latter half. Again a least squares linear fit is significant.

3. At the station level, the annual maximum and annual total rainfall are positively correlated; i.e., evidence of dependence.

4. At the station level, there is no consistency among the 5 stations highlighted here with respect to a trend to their annual maximum rainfall series.

5. At a station level, there is more consistency among the 5 stations with respect to a trend to their annual total rainfall series which may show an overall increase.

6. At the station level, there is a significant amount of variability in both the annual maximum and annual total rainfall series which may obscure any trends.

5.6.2 Introductory Model Specification

Before we begin fitting possible trends and comparing the independent joint density model to the expansion of the joint density model, we need to make some initial decisions about how to proceed with the analysis.

The first is how much of a year's daily precipitation record must exist before we count that year as representing a complete year? The period for which data are available differs from station to station. The number of stations who start their record in 1901 is 105 but the last 3 stations did not start recording until somewhere between 1940 and

1952. All stations suffer from some missing values throughout the time period of the study. To reduce the bias that could result from too much missing data, it was decided to establish a cutoff level for the number of days data required for a particular station to be included in a particular year. Four cutoffs were considered, at 180, 240, 300, and 330 days. There is a big jump between the number of station-year combinations included with a 240 day cutoff compared with a 300 day cutoff. In fact, the percentage of years in the total record which are cut off jumps more than doubles from 1.6% to 3.4% as the cutoff goes from 240 to 300 days. Initially, all four cutoffs were considered and the results compared.

Varying the cutoff level (180, 240, 300, and 330) we fit the Gumbel distribution with its two parameters to the annual maximum rainfall series for all the stations and the normal distribution with its two parameters to the annual total rainfall series for all the stations. We then compare the parameter estimates using a cut-off level of 180 days to the parameter estimates using a cut-off of 240 days, then compare the parameter estimates based on 240 days to parameter estimates based on 300 days, and finally compare the parameter estimates based on 300 days to estimates based on 330 days – ultimately 2244 comparisons. We flagged parameter estimates as being different if the difference between them is larger than the estimated standard error of their difference. Recall an estimate of the standard error of the individual parameter estimates is provided by taking the square root of the estimated Hessian matrix that the computer program produces. Only 3 of these 2244 comparisons were flagged. This means the model fit is essentially the same no matter where the cut-off is set. Our focus on a cut-off level has narrowed to between 240 days and 300 days, the first big jump in missing daily records. Since there is no significant difference in the parameters fit, we choose 240 since it allows us to use more of the data.

Second, often when modeling climatological data, an important issue to address is the amount of serial dependence that exists. To test the serial dependence in the annual maximum and annual total rainfall series, we fit autoregressive time series models

Model Selected	AR(0)	AR(1)	AR(2)	AR(3)	AR(4)
# stations	142 (76%)	24 (13%)	12 (6%)	6 (3%)	3 (1.5%)

Table 5.1: Results of the AR models for annual maximum rainfall series: The table gives the AR model, the AIC criterion selected and the number of stations out of 187 which fell into the 5 models with the percentage of stations in parenthesis.

Model Selected	AR(0)	AR(1)	AR(2)	AR(3)	AR(4)
# stations	120 (64%)	33 (18%)	17 (9%)	12 (6%)	4 (2%)

Table 5.2: Results of the AR models for annual total rainfall series: The table gives the AR model, the AIC criterion selected and the number of stations out of 187 which fell into the 5 models with the percentage of stations in parenthesis.

of orders 0, 1, 2, 3, and 4 by the Yule-Walker equations, as implemented in S-Plus (Venables and Ripley, 1994, p. 361). The best fitting model was selected using the AIC criterion.

For the annual maximum precipitation series, the results of the time series analysis are in Table 5.1. A majority (76%) of the stations selects the best model as AR(0); that is, no significant serial dependence. Based on these findings, we justify an independence assumption for annual maximum rainfall series.

For the annual total precipitation series, the results of the time series analysis are in Table 5.2. Although the percentage of stations which are best modeled with AR(0) is not as high as in the maximum case, the majority (here 64%) of the stations still shows no significant dependence. Thus we conclude that if we take the whole data set into consideration that we do not see any significant serial dependence in either series.

Note as stated in Section 5.5.3, we assume daily precipitation series is *iid*. Again we do not justify this assumption although we reiterate the findings of Stern and Coe (1984) and Smith (1994) that whether it rains or not from day to day is more dependent than the daily amounts of rain.

Region (# st.)	1 (16)	2 (51)	3 (50)	4 (38)	4 (32)	Overall (187)
# sign GEV (%)	2 (12.5)	12 (24)	15 (30)	15 (47)	9 (24)	53 (28)

Table 5.3: Results on comparing Gumbel distribution to GEV distribution. The table provides the five regions in the country with the number of stations in each region in parenthesis and the number of stations which rejected the Gumbel distribution with the % of rejection for each region in parenthesis.

The third step in the data analysis is to determine the appropriate model for the annual maximum series. In the classical set-up, this would be deciding between the Gumbel, Fréchet, or Weibull distribution. Using the GEV distribution, the decision is formalized to testing whether there is significant evidence for the introduction of the shape parameter of the GEV distribution into the model. We can do this using the standard likelihood ratio test theory by calculating the difference between the maximum likelihood under the Gumbel distribution and under the GEV distribution. We reject the null hypothesis that the Gumbel distribution is appropriate when the likelihood test statistic is greater than $\chi^2_1 = 3.84$. In the analysis, we find 55 out of 187 stations (less than 30%) are significantly different from the Gumbel distribution. Thus for the majority of the stations there is no significant reason to reject the Gumbel distribution as the appropriate model for the annual maxima. If we make one decision between modeling with the Gumbel or the GEV distribution for all the stations, then an appropriate distribution to use for the annual maximum precipitation is the Gumbel distribution.

Note the actual percentage of stations for which we reject the Gumbel distribution as the appropriate model varies throughout the different regions, see Table 5.3. Although for all regions, this percentage is less than 50%, there is a big difference between region 1 (West coast) and region 4 (Northeast). In region 1 (West Coast), the Gumbel distribution appears to govern the behavior of the annual maximum. In region 4 (Northeast) that is not as straightforward.

For a few of the stations which rejected the Gumbel distribution as being the better fit, Gumbel plots were produced. A Gumbel plot is a standard graphing device used in extreme value theory which aids in judging the fit of the distribution and in detecting outliers. In it, the ordered observations are plotted against the expected Gumbel series. Obviously, the closer the graph is to a straight line the closer is the Gumbel fit to the data. If the plot curves upward, this usually indicates a Fréchet fit. If it curves downward, this usually indicates a Weibull fit. For more details, see Smith (1990). Figure (5.9) show Gumbel plots for the following stations (the difference in the negative log likelihood functions between the Gumbel and the GEV fit are in parenthesis): Albany, Texas (15.45), Big Timber, Montana (13.90), Princeton, Indiana (7.63), Fredonia, New York (4.01), St. Leo, Florida (11.2), and Goldsboro, North Carolina (3.38). From the plots, we infer a variety of causes for the deviations from the Gumbel distribution. Figure (5.9) (a) and (b) – Albany, Texas and Big Timber, Montana – show the presence of outlier/s which distort the Gumbel fit. Figure (5.9) (c) and (d) show the presence of an upward curve indicating the Fréchet distribution as a more appropriate fit for these stations. Finally, figure (5.9) (e) and (f) are not as readily deciphered – although not a straight line, it does not indicate either the characteristic curve of the Fréchet or Weibull distribution. Our conclusion is to use the Gumbel distribution to model the annual maximum for all stations. Clearly some of the Gumbel plots do not support the Gumbel distribution as the most appropriate model. On the other hand, the majority of the stations do not reject the Gumbel distribution and even for some which do, there is not an obvious better choice.

A final comment of the preliminary fit of the annual maxima concerns the shape parameter of the GEV distribution. For all the 187 stations the estimate of k, the shape parameter, is less than 0.50. In fact, the estimates live between -0.35 and 0.15 with approximately 85% negative. Recall the maximum likelihood estimates exist and have their classical asymptotic properties needed for the likelihood ratio test and the t-tests for the extreme value parameter estimates provided $k < 0.50$.

Region	1	2	3	4	5	Overall
# sign Normal (%)	4 (25%)	7 (14%)	6 (12%)	5 (11%)	8 (25%)	29 (15.5%)

Table 5.4: Results on the Kolmogorov-Smirnov test for the sum. The table provides the five regions in the country with the number of stations in each region in parenthesis and the number of stations which rejected the Normal distribution with the % of rejection for each region in parenthesis.

In the fourth preliminary decision to be made, we look at the asymptotic distribution of the annual total precipitation which is approximately normal by standard central limit theory. Since we model each station individually, the question is how well does the normal distribution fit the 187 annual total precipitation series? We calculate the Kolmogorov-Smirnov test for normality for each of the stations. The Kolmogorov-Smirnov test is a standard non-parametric test with null hypothesis, in this case, that the distribution is normal with alternative hypothesis that the distribution is not normal. We set an $\alpha = 0.05$ level of significance and find 29 out of 187 stations (15.5%) are significantly different than normal. Thus for the majority of the stations there is no significant reason to reject the normal distribution for modeling the annual total rainfall although we recognize some evidence of a discrepancy from the normal fit across the US.

The actual percentage of stations where we reject the normal distribution as the appropriate fit varies from region to region, but less than for the Gumbel distribution in the annual maximum case, see Table 5.4. For all regions the percentage of stations which reject the normal distribution is substantially below 50% but we see the interior of the country (regions 2 and 3) and the Northeast (region 4) have fewer stations that reject the normal distribution. The far West and the South (regions 1 and 5) have the higher percentage of rejections.

Quantile plots were calculated for some of these stations which reject the normal distribution for the annual total precipitation series. The quantile plot, like the Gumbel plot, is a standard graphing technique which aids in judging the fit of the distribution and in detecting outliers. In it, the ordered observations are plotted against the expected normal series. A linear relation reflects a good fit to the normal distribution. See Johnson and Kotz(1982), p. 509 under probability plotting in the Ordered statistics section for more details. Figure (5.10) shows quantile plots for the following stations (the p-value associated with the Kolmogorov-Smirnov test is in parenthesis): (a) Goldsboro, North Carolina (.0231), (b) Ojai, California (.0056), (c) Fort Collins, Colorado (.0137), (d) Dillion, Minnesota (.0074), (e) Fredonia, New York (.0027), and (f) St. Leo, Florida (.002). Figure (5.10) (a) shows 4 outliers which distort the normal fit. Except for Figure (5.10) (a), the other plots show some variation of the S pattern associated with heavy tailed distribution. These particular plots are important because they give evidence that there is a small percentage of annual total rainfall series where the normal distribution is not a convincing fit. Note Fredonia, New York, St. Leo Florida, and Goldsboro, North Carolina show both deviations from normal and deviations from Gumbel. On the other hand, given the relatively quick rate of convergence of the central limit theorem and the fact that the majority of the stations do not reject normality, we conclude that using the normal distribution is the only viable choice.

Finally before we address the trend analysis of the joint density of M_n and S_n, we give a cursory look to what form the trend should take. Specifically, we compare the exponential trend

$$\eta_t = \alpha e^{\beta t} \qquad \mu_t = \delta e^{\gamma t} \tag{5.26}$$

to the simple linear trend

$$\eta_t = \alpha + \beta t \qquad \mu_t = \delta + \gamma t.$$

Note Smith (1999) also fits trends to the scale parameter and likewise we could fit trends to scale parameters, ψ and σ. This was not done primarily because our focus of this analysis is on the change in the location parameters – overall increases – not dispersion.

The linear form is offered primarily due to its simplicity and ease in interpretation. The exponential form has the benefit that we can compare the trends in the annual maximum and annual total rainfall; that is, the exponential trend can be interpreted as a rate of increase as opposed to a magnitude.

We return to the plots of the annual maximum rainfall over the past century (Figure 5.7). At the station level for annual maxima, we find no overall pattern, including the linear or the exponential trend. For the annual total rainfall over the past century at the station level (Figure 5.8), there is more evidence of trends but due to the station level variability there is no visual differentiation between either form.

To compare the two forms of the trend, we fit the Gumbel distribution to the 187 annual maximum rainfall series with the exponential form of the trend and the linear form of the trend in the location parameter. Both the linear and exponential form model significant trends in the exact same stations. In terms of log likelihood functions, each form gives essentially the same value of the log likelihood function at its maximum.

We also fit linear and exponential trends to the location parameter of the annual total rainfall series. We obtain the same type of results. Each form detects significant trends in the same stations. Again there is no difference between the log likelihood functions maximized. Clearly there is no significant difference in modeling these series with either the exponential or linear trend.

A final note is that there are alternative forms of trend which we have not considered. Figure 5.4 – the average annual maximum precipitation for the 187 stations – would support a quadratic trend since the graph appears to dip down into the 1920s and 1930s, then it slowly increases throughout 1940s, 50s and 60s, and finally the slope increases from the 1950s. On the other hand, Figure 5.5 – the average annual total precipitation – appears closer to an exponential trend. No particular form for the trend emerges as the obvious choice for all stations for the annual maximum and annual total precipitation series. We want to select one form for both series and for all stations. Due to lack of any significant evidence in this data set to differentiate between the two forms and the interpretation of exponential form, we stay with the exponential form of the trend as defined in (5.26).

Conclusions from the introductory model specification

1. We use 240 days as the minimum number of days in a year which must be recorded in order to consider the year as not missing.
2. Overall, we find no significant evidence for serial dependence in the annual maximum and annual total precipitation series for the majority of the stations.
3. For the majority of the stations, the Gumbel distribution is an appropriate choice for the limiting distribution for the annual maximum precipitation series and we use it to model all the annual maximum precipitation series.
4. For the majority of the stations, the normal distributions is a good fit to the annual total precipitation series and we use it to model all the annual total precipitation series.
5. We use the exponential form of the trend in the location parameter μ and η in our trend analysis.

5.6.3 Model Analysis

As stated before there are two main themes in this analysis: (1) analyzing annual maximum and annual total rainfall with particular interest in the trends in these series and (2) comparing the independent and expansion models. At this point, we wish to specify the objectives in this analysis, in particular with respect to the trend analysis. The comparison between independent and expansion models fall from the results of the trend analysis. So if we focus on the trend analysis, we can ask the following questions:

Q1 Is there any evidence of a trend in either the means or the maxima of the annual rainfall series?

Q2 Is there any evidence that the trends in Q1 are predominantly in either the means or in the maxima?

Q3 Is there any evidence that the trends in Q1 are common?

Q4 Ultimately what model best estimates the trends in the annual rainfall series?

To study these questions we select the following models in Table 5.5 to run. The independent case refers to models based on the joint density defined in equation (5.7) whose log likelihood function is (5.23), the Gumbel versions. The models assume independence between the maximum and sum. The density has four base parameters: location and scale parameter for M_n – η and ψ – and location and scale parameter for S_n – μ and σ. The expansion case refers to models based on the higher order expansion of the joint density of M_n and S_n defined in (5.17) whose log likelihood function is (5.25), the Gumbel versions. It is developed using the expansion of the conditional density of $S_n|M_n = u_n$. This expansion uses the same four base parameters as the independent case: η, ψ, μ, σ. In models I4 and E4, the parameters β and γ take a common value denoted by τ which is referred to in Q3.

Preliminary Comparison between the Independent and Expansion Models

First we compare the overall fit of the data using the independent models versus using the expansion models. Let us focus on the results obtained in models I1 and E1, the base models (i.e. no trends). We find that in 186 out of 187 stations the expansion model has a larger maximum log likelihood function, indicating an overall better fit for the expansion model. The only exception was station # 308944, Wanakena Ranger Station, New York. In 185 stations, the difference between the maximum log likelihood

Models Run		
Parameters Fit	Independent Case	Expansion Case
η, ψ, μ, σ	I1 (Base Model)	E1 (Base Model)
$\eta = \alpha e^{\beta t}, \psi, \mu, \sigma$	I2 (Max trend)	E2 (Max trend, Total trend =0)
$\eta, \psi, \mu = \delta e^{\gamma t}, \sigma$	I3 (Total trend)	E3 (Total trend, Max trend = 0)
$\eta = \alpha e^{\tau t}, \psi, \mu = \delta e^{\tau t}, \sigma$	I4 (Common trend)	E4 (Common Trend)
$\eta = \alpha e^{\beta t}, \psi, \mu = \delta e^{\gamma t}, \sigma$	I5 (Full Model)	E5 (Full Model)

Table 5.5: Models run in the trend analysis. The table give the independent models and expansion models that are run, labeling each model and specifying the parameters.

functions of I1 and E1 was greater than 2. By analog to the nested model case this difference implies the improvement of the fit using the expansion model is significant. Note if we compare the maximum log likelihood functions of models I5 and E5 (the full models), we get the exact same comparison – the expansion model has a larger maximum likelihood function; that is, the expansion model gives a better fit.

Next we consider the difference between the models which are fit using I1 and E1 – in particular, the parameter estimates and their standard errors. Figure 5.11 shows the plots of the parameter estimates based on I1 versus E1 for η, ψ, μ, and σ, respectively. Given the linear relation in all four plots, we conclude that the independent and expansion models are giving comparable estimates. The estimates for μ are strikingly similar. The estimates for σ – the standard deviation of daily rainfall – are most different. This is readily explained. In the expansion model, σ is estimated conditionally on knowing what the maximum daily rainfall in the year is. In the independent case, the maximum rainfall – the value with the most variability – is treated as unknown and thus adds to the spread of daily rainfall. In fact, if we look more closely at the estimates of σ, the number of stations whose estimate of σ in E1 is smaller than in I1 is 170 out of 187. Further, the number of times the difference between these estimates is greater than the estimate of the standard error of this difference is 56 out of 187.

This decrease in the scale parameter of the total is mirrored to a lesser extent in the scale parameter of maximum, ψ. In 175 out of 187 stations, the estimate of ψ under E1 is smaller than under I1. However these differences are not statistically significant – none of these differences were larger than the estimated standard error of their difference. A final comment on the parameter estimate is that for the locations parameters – μ and η – the expansion model E1 produces estimates that are larger that the independent model I1. In 162 stations, the estimates of η are larger for the E1 model than the I1 model. In 169 stations, the estimates of μ are larger for the E1 model than the I1 model. None of these differences are significant but it does show a pattern. Overall the addition of the higher order term in the expansion model influences more significantly the estimate of σ than the other parameters.

A last note on the difference between the fit of the models from the independent to the expansion cases in I1 and E1 involves how accurately the above parameters are being estimated. We assess this by looking at the standard errors of these parameters. Recall a by-product of the quasi-Newton optimization routine used to calculate the maximum likelihood estimates is an estimate of the standard errors. If the estimates of the standard error in the expansion case are smaller than in the independent case, we claim the expansion model estimates the parameters more precisely. Table 5.6 lists the parameters and the number of stations out of 187 that the estimate of the standard error in the expansion case are less than or equal to the independent case. In over 60% of the stations the expansion model more accurately estimates each of the four parameters. Note it is possible that this decrease in the estimates of the standard error is due to a latent bias in the expansion versus the independent model. There is no basis for assuming this bias and currently it is believed that this decrease truly represents a genuine difference in the precision of the estimates.

In conclusion from the comparison between the independent and expansion models:

1. Except for one station, the expansion model always gives a better fit as measured by the likelihood function.

Parameter	η	ψ	μ	ψ
# stations	111	118	124	124

Table 5.6: Standard error comparison of parameter estimates in the independent base model and expansion base model. The table give the parameters in the base models and the number of stations whose estimated standard error for each parameter is smaller in model E1 than in model I1

2. The expansion and independent model give comparable parameter estimates although the estimates of the the scale parameters, particularly the standard deviation of daily rainfall, are smaller for the expansion model while the estimates of the location parameters are larger.

3. The expansion model has smaller estimated standard errors for the parameter estimates which we believe to represent an increase in the accuracy of the parameter estimates.

Trend Analysis

There are essentially two methods used in the trend analysis which aid in the interpretation of the 187 individual station results: tests of hypotheses and a spatial smoothing method. There are two groups of tests of hypothesis that are used: (1) Asymptotic likelihood ratio tests are used to compare the models set up in the previous section. (2) t-tests for the estimates of the trend parameters to test whether the trends are significantly different from zero. The latter uses the asymptotic normality of the maximum likelihood estimators. The spatial smoothing technique will provide both a national average for the trends of interest and a smoothed version of the estimates of the trends across the entire contiguous US. Using these methods, we will draw conclusions about the estimates of the trends of the annual maxima and totals across the contiguous US at both the national, regional, and local level.

Test	H_o	H_a	Ind Models	Exp Models	χ^2_{df}
T1	$\beta = \gamma = 0$	$\beta \neq 0, \gamma = 0$	I2-I1	E2-E1	$\chi^2_1 = 3.84$
T2	$\beta = \gamma = 0$	$\gamma \neq 0, \beta = 0$	I3-I1	E3-E1	$\chi^2_1 = 3.84$
T3	$\tau = 0$	$\tau \neq 0$	I4-I1	E4-E1	$\chi^2_1 = 3.84$
T4	$\beta = \gamma\ (\tau \neq 0)$	$\gamma \neq \beta$	I5-I4	E5-E4	$\chi^2_1 = 3.84$
T5	$\beta = 0, \gamma = 0$	$\gamma \neq \beta$	I5-I1	E5-E1	$\chi^2_1 = 5.99$
T6	$\beta = 0, \gamma \neq 0$	$\beta \neq 0, \gamma \neq 0$		E5-E3	$\chi^2_1 = 3.84$
T7	$\beta \neq 0, \gamma = 0$	$\beta \neq 0, \gamma \neq 0$		E5-E2	$\chi^2_1 = 3.84$

Table 5.7: Likelihood ratio tests to be performed. The table gives the test number, the null and alternative hypothesis (H_o and H_a) being tested, the two models whose log likelihood functions will form the likelihood ratio test, and the critical value for the likelihood ratio test based on $\alpha = 0.05$ level of significance. Recall β = the trend in the annual maximum rainfall, γ = the trend in the annual total rainfall, and τ = the common trend ($\beta = \gamma$).

Likelihood ratio test of hypothesis The first way to address the questions in Section 5.6.3 is to perform the following hypothesis tests using standard asymptotic likelihood ratio test theory, comparing the log likelihood functions of the above models. Note that the critical values correspond to the limiting χ^2 test with 0.05 level of significance. Table (5.7) lists the tests performed in the analysis concerning the trends in the annual maximum and annual total rainfall series.

Specifically we are testing for the following information:

In the independent case, T1 tests for the presence of (just) a trend in the annual maximum rainfall disregarding any information about the annual total rainfall. This test compares a model (which is based only on annual maximum rainfall) with no trend in the maximum to a model with a trend in the maximum. In the expansion case, T1 tests for the presence of a trend in the annual maximum while the trend in the annual total rainfall is set to 0; i.e., this assumes no trend in the annual total rainfall and tests for a possible trend in the maximum. In the expansion case, T6 tests for the presence of a trend in the annual maximum rainfall given a trend for the annual total rainfall is already present in the model. Note the difference between the independent and the expansion test. In the independent test, no matter what we wish to test about the annual maximum what is happening to the trend in the annual total rainfall is irrelevant. In the expansion cases, that is not the case. Since it models the first order correction to the dependent structure, it contains information on both the annual maximum and annual total rainfall. So whether we include the trend in the annual total or not implies an assumption about this trend in the total rainfall. In the expansion case, we can no longer separate the trends completely.

In the independent case, T2 tests for a presence of (just) a trend in the annual total rainfall disregarding any information about the annual maximum rainfall. This test compares a model (which is based only on annual total rainfall) with no trend in the annual total to a model with a trend in the annual total rainfall. In the expansion case, T2 tests for the presence of a trend in the annual total while the trend in the annual maximum rainfall is set to 0. In the expansion case, T7 tests for the presence of a trend in the annual total rainfall given that a trend for the annual maximum rainfall is already present in the model. Again note the difference in the interpretations between the independent and the expansion cases.

In both the independent and the expansion cases, T3 tests for the significance of a single trend for both the maximum and the total annual rainfall; i.e., this test compares a model with no trends in either the annual maximum or annual total to a model with a common trend to the annual maximum and the annual total rainfall.

In both the independent and expansion cases, T4 tests for the significance of adding a second trend parameter to the model. This test compares a model with a common trend to both the annual maximum and annual total rainfall to a model that has a separate trend for the annual maximum and annual total rainfall.

In both the independent and expansion cases, T5 tests for the significance of adding separate trend parameters for the annual maximum and annual total rainfall to the

#	H_a	Independence Case US(187)	R1(16)	R2(51)	R3(50)	R4(32)	R5(38)
T1	$\beta \neq 0$	19 (10)	2 (12.5)	6 (12)	4 (8)	2 (6)	5 (13)
T2	$\gamma \neq 0$	42 (22)	3 (19)	12 (24)	9 (18)	11 (32)	7 (18)
T3	$\tau \neq 0$	49 (26)	4 (25)	13 (25)	13 (26)	11 (32)	8 (21)
T4	$\beta \neq \gamma \vert \beta = \gamma$	2 (1)	0 (0)	1 (2)	0 (0)	0 (0)	1 (3)
T5	$\beta \neq 0, \gamma \neq 0$	40 (21)	4 (25)	10 (20)	11 (22)	8 (25)	7 (18)

Table 5.8: Results of the likelihood ratio tests in the independent case. The table gives the results of the tests in Table (5.7) for the independent models. It breaks down the number of significant alternative hypotheses throughout the regions and gives the percentage of significant alternative hypothesis in parenthesis.

model simultaneously. This test compares a model which has no trend parameters to one which has a trend for the annual maximum rainfall and another for the annual total rainfall.

The results for these tests are given in Table 5.8 and Table 5.9 for the independent and expansion models, respectively. Since we use a 5% level of significance in our likelihood ratio test, we would expect that in 5% of these stations a trend would be detected by chance alone; that is, the test would claim a trend is significant when in fact no trend exists. Thus the percentages in Table 5.8 and 5.9 take on more meaning. When the percentage goes over 5% (note we do not define how far over 5% we need to be), we claim that there exist evidence for this trend across the US or region. In other words, some underlying change has occurred to explain the percentage of significant trends – they did not occur by chance. Obviously the bigger the percentage the more proof there is behind the significance of the trend.

Results of the likelihood ratio tests Plainly the majority of the stations do not yield statistically significant trends in the annual maximum or annual total rainfall series no matter what model is being fit. This is understandable due to the large variability at the station level. That is not to say that the results do not give insights into the changes in the annual maximum and annual total rainfall over the past century.

#	H_a	Expansion Case US(187)	R1(16)	R2(51)	R3(50)	R4(32)	R5(38)
T1	$\beta \neq 0$	10(5)	2 (12.5)	0 (0)	1 (2)	3 (9)	4 (11)
T2	$\gamma \neq 0$	42 (22)	2 (12.5)	14 (27)	8 (16)	9 (28)	9 (24)
T3	$\tau \neq 0$	40 (21)	4 (25)	11 (22)	9 (18)	9 (28)	7 (18)
T4	$\beta \neq \gamma \vert \beta = \gamma$	15 (8)	1 (8)	6 (12)	2(4)	2 (6)	4 (11)
T5	$\beta \neq 0, \gamma \neq 0$	40 (21)	4 (25)	14 (27)	6 (12)	9 (28)	7 (18)
T6	$\beta \neq 0 \vert \gamma \neq 0$	14 (7)	2 (12.5)	5(10)	2 (4)	2 (6)	3 (8)
T7	$\gamma \neq 0 \vert \beta \neq 0$	42(22)	2 (12.5)	14 (27)	8 (16)	10 (31)	8 (21)

Table 5.9: Results of the likelihood ratio test in the expansion case. The table gives the results of the tests in Table (5.7) for the expansion models. It breaks down the number of significant alternative hypotheses throughout the regions and gives the percentage of significant alternative hypothesis in parenthesis.

Having previously established that the expansion models fit the data better than the independent models, we look for the "best" model among the expansion models although we have also run all the independent models. The results for the independent models will be used both to continue the comparison with the expansion models and to help validate these findings.

First let us concentrate on β – the trend in the annual maximum precipitation. If we look at tables (5.8) and (5.9), we see there is little evidence for significant trends in the maximum no matter which model is fit – expansion T1, expansion T6, or independent T1 which have 7%, 5%, and 10% of the stations with significant βs, respectively. The drop from 10% in independent T1 to 5% in expansion T1 has some interesting interpretations although we caution against too much weight on this particular comparison due to the small percentage of significant βs. This drop is primarily due to region 2. In independent T1, there are 6 stations with significant βs. In expansion T1, none of these stations have significant βs. In fact, in expansion T1, there are no stations with significant βs west of Texas except for the far west coast. Interestingly in expansion T6, five of these six stations in region 2 have significant βs again. Why is this notable? In the independent case, disregarding information on the annual total trend and setting this trend equal to 0 are equivalent. In the expansion models, this is not the case. In expansion T1, H_o assumes no trend in the annual total. In expansion T6, H_o assumes the trend in the annual total is not equal to 0. With respect to region 2, these two expansion models give different results and we need to be mindful of the interpretation of these expansion models throughout the analysis. Overall we conclude there little evidence of a trend – on its own – in the annual maximum precipitation across the continental US.

Looking at γ – the trend in the annual total rainfall series, we see substantial evidence of this trend no matter which model is fit. In all models which test for significance of γ by itself – expansion T2, expansion T7, and independent T2 – we have 22% of the stations with significant trend in the annual total rainfall. Given the amount of station variability this is a high number – more than 4 times the percentage that would be detected by chance alone if $\gamma = 0$. The stations with significant γs are spread throughout the country. We conclude there is substantial evidence of a trend – on its own – in the annual total precipitation in the continental US.

The test for the inclusion of a common trend for the annual maximum and annual total precipitation series into the model turns out to be insightful from the point of view of highlighting the differences between the expansion and independent cases. The comparison between the results for the expansion test T3 and the independent test T3 illustrate the biggest difference between the models in terms of the likelihood ratio test results. First both models show a substantial percentage of stations where including a common trend is significant; that is, there is significant evidence that some trend exists in this data. The expansion test T3 detects a significant common trend in 21% of the stations. For the independent model, this is 26%. Although this change is not big enough to be significant, it does indicate a difference between the expansion and independent models. Further, these stations that drop out do so evenly throughout the country. We conclude that there is strong evidence in the data set to indicate that if we assume a common trend to the annual maximum and total rainfall series, then it is non-zero.

Next we look at tests for adding a second trend parameter above a common trend parameter to the model – expansion T4 and independent T4. These tests are the flip side of tests T3. In both cases, there is little to no evidence for adding a separate trend parameter for both the annual maximum and annual total rainfall series. In the independent T4 test just 1% of the stations detect a significant need for differentiating between the trend in the annual maximum and annual trend in the total. This percentage goes up to 8% for the expansion case. Although neither of these cases lead us to believe differentiating between the trends is significant, this difference in percentages between the tests E4 and I4 reiterates the fact that there is a subtle distinction between how the expansion and independent models estimate these trends. Altogether we see in the independent case more evidence for a common trend for the maximum and the total but once the common trend is in the model essentially no evidence that adding a second parameter is significant. On the other hand in the expansion models we see less evidence for a common trend but more evidence for the need of a separate parameter for differentiating between the trends in the annual maximum and the annual total rainfall. Ultimately we conclude there is no significant evidence to differentiate between the two trends in terms of the likelihood ratio tests.

Finally, the results for test T5 – adding a separate parameter for the maximum and the total, simultaneously, to the model – are identical in the expansion and independent cases. Both show that simultaneously adding β and γ into the model is significant in 21% of the stations. Certainly this is a substantial percentage and indicates that there exist some kind of a trend in the annual maximum and annual total rainfall series. Note of the stations which are significant in expansion T2 ($\gamma \neq 0, \beta = 0$), T3($\tau \neq 0$), and T5($\gamma \neq 0, \beta \neq 0$), 26 are common to all tests.

In conclusion for the likelihood ratio tests:

1. There is significant indication that some trend exists in the annual maximum and annual total rainfall series.

2. Three of the expansion models are arguably an appropriate model: (1) E3, model

with only the trend for the annual total rainfall ($\gamma \neq 0, \beta = 0$); (2) E4, model with a common trend for the annual maximum and annual total rainfall ($\tau \neq 0$); and (3) E5, model with separate trends for the annual maximum and annual total ($\gamma \neq 0, \beta \neq 0$).

3. Due to the substantial evidence of a significant trend in the annual total rainfall in expansion test T2 and T7 and little evidence of a significant trend in the annual maximum rainfall in the expansion test T1 and T6, the likelihood ratio test results lean toward model E3 as the most appropriate model.

4. Finally, in general the expansion and the independent models give similar results. This is particularly true concerning models and tests associated with (just) the trend in the annual total rainfall – expansion T2, expansion T7, and independent T2. This also holds for tests T5. What appears to be the biggest difference is in estimating the trend in the annual maximum and fitting a common trend.

t-tests To get another perspective on the estimates of these trend parameters, we look at t-tests for these parameters based on the maximum likelihood estimates of these trends. Recall we fit each model using the maximum likelihood method and thus obtain maximum likelihood estimates for our trend parameters with estimates of their associated standard errors. Using the asymptotic normality property of maximum likelihood estimates, we can therefore calculate a test statistic for testing if a trend parameter for an individual station is 0 or not. We give the results of the t-tests for the individual stations for all the models that we have run. The models which give estimates for β are E2, E5, and I5 (which is equivalent to I2). The models which give estimates for γ are E3, E5, and I5 (which is equivalent to I3). The models which give estimates for the common trend τ are E4 and I4. We focus primarily on the results of the estimates for the trend parameters in model E5 and E4 although for comparison we give all of the results. Note we do not use the results for the estimates of β in E2

	Trend in max $H_o: \beta = 0$ $H_a: \beta \neq 0$			Trend in total $H_o: \gamma = 0$ $H_a: \gamma \neq 0$			Common trend $H_o: \tau = 0$ $H_a: \tau \neq 0$	
	$\hat{\beta}$ in I5	$\hat{\beta}$ in E5	$\hat{\beta}$ in E2	$\hat{\gamma}$ in I5	$\hat{\gamma}$ in E5	$\hat{\gamma}$ in E3	$\hat{\tau}$ in I4	$\hat{\tau}$ in E4
$t < -2$	7(4)	13(7)	16(9)	6(3)	8(4)	5(3)	14(8)	9(5)
$t < -1$	26(14)	32(17)	44(24)	18(10)	19(10)	19(10)	25(13)	23(12)
$t > 0$	111(59)	109(58)	90(48)	128(68)	131(70)	134(72)	130(70)	132(71)
$t > 1$	57(31)	56(30)	27(14)	77(41)	83(44)	84(45)	86(46)	75(40)
$t > 2$	28(15)	23(12)	10(5)	36(19)	39(21)	37(20)	48(24)	44(24)

Table 5.10: Results of t-tests of trend parameters. The table gives the t-test for the parameter estimates by dividing the results into the three different trend parameters that were estimated. For each the table gives the test which is performed, the model the estimate comes from, and the number of stations(% in parenthesis) which shows significant evidence to reject H_a. Let τ be the common trend $\gamma = \beta$.

$(\beta \neq 0, \gamma = 0)$ since the likelihood ratio tests conclude convincingly that γ does not equal 0 and thus model E2 is not an appropriate model.

Table 5.10 lists the results of these t-tests. It breaks the results down into the estimates for β, γ and τ where recall τ is the common trend parameter. It gives the null and alternative hypothesis that are being tested. Finally it shows the percentage of stations which have a test statistic less than -2, less than -1, greater than 0, greater than 1, and greater than 2. In the spirit we interpreted the results of the likelihood ratio tests, we claim evidence that the associated trend is not zero for these models if the percentage of stations with these test statistics are larger than 2.5%, 15%, 50%, 15%, and 2.5%, respectively.

Results of the t-tests To begin, the trend in the annual total rainfall series is certainly easiest to interpret. First, the results of these t-tests are nearly identical no matter if the tests are based on estimates of γ from model I5, E5 or E3. In other words, these three models give comparable estimates for the trend in the annual total rainfall. Looking at the results in the $t > 0$ row, we see around 70% of the estimates of the trend in the total are positive. If there was no trend in the annual total rainfall series, this should be near 50%. If there was no trend the percentage of stations whose test statistic would be above 1 would be near 15% but the actual percentage is around 3 times that or 44%. The percentage of test statistics above 2 would be near 2.5% if no trend existed but here the actual percent is 8 times that number or 20%. Not only are the estimates for γ positive but they are significantly positive. We also see that these estimates are not significantly negative. Granted 30% of the estimates are negative but by looking at the rows associated with $t < -2$ and $t < -1$, we see the percentage in each row is actually lower than the percentage expected by chance alone if there were no such trend; that is, there are no significantly negative trends in the annual total rainfall. Thus these results concur with the likelihood ratio tests that γ is significantly different from 0, here positive, and the independent and expansion models give the same results.

The results of the t-test for τ are also similar to results both for the t-tests associated with γ and also the likelihood ratio tests for the common trend estimates. Certainly there is substantial evidence that the common trend is not zero. In both the expansion and independent models, 70% of the estimates are positive (compared to 50%) and over 24% of the test statistics are over 2 (compared to 2.5%). Clearly overall the common trend is positive. In fact, there appears to be even a higher percentage of truly significant (test statistic greater than 2) estimates than for any t-test associated with the annual total. Finally, the results for the common trend are nearly identical to the results for the trend in the annual total. This substantiates the idea that the trend in the annual total rainfall dominates the common trend and hence the trend in the annual maximum rainfall.

Finally the results in the t-test concerning the annual maximum rainfall are markedly different from the likelihood ratio tests. First, there is evidence that $\beta \neq 0$. For both the estimates in E5 and I5, we have roughly 60% are positive. More importantly, we see more than five times the percentage of stations whose test statistic is greater than 2 than would happen by chance if $\beta = 0$ – somewhere between 12% and 15%. This indicates $\beta > 0$. Another difference from the likelihood ratio tests is that the expansion and the independent models give comparable results. Currently there is no obvious reason why the likelihood ratio tests and the t-test are giving such different results. Since each of these tests rely on asymptotic results, this may explain the difference. It may also have to due with how well the estimates of the standard errors perform. Note there is a slight difference between the estimates for β in E5 versus I5 with respect to those estimates which are negative. For the estimates of β in E5, we see slightly more significant negative estimates. In particular, there are almost twice as many estimates in E5 whose test statistic is below -2 than in the I5 versions. This 7% for E5 is nearly three times as much as would be expected by chance alone if $\beta = 0$. Thus although the percentage of estimates of β whose absolute value of the test statistic is greater than 2 is the same for E5 (19%) and I5 (19%) and thus both show clearly $\beta \neq 0$, the estimates in E5 are more significantly negative than in I5. Whether this slight difference has an impact on the results is not readily known. So to conclude, the results for the t-test concerning the trend in the annual maximum rainfall indicates that $\beta \neq 0$. More specifically, in the E5 estimates of β there are a significant number of positive estimates and a significant number of negative estimates.

Note although the model E2 is clearly not a viable model since it assumes that $\gamma = 0$ and all evidence is to the contrary, it does show us the effect assuming that the trend in the annual total rainfall has on the estimates of the trend in the annual maximum. First it lowers the estimates for the trend in the annual maximum – only 48% are positive. Second it eliminates a large percentage of the significant positive estimates of the trend in the annual maximum and increases the number of significant negative estimates of the trend in the annual maximum.

Finally, although the results in the t-tests show that both the trend in the annual maximum and annual total rainfall are significantly different from 0, there are important differences between the two trends. The t-tests indicate that more of the estimates for trend in the annual total rainfall are positive ($t > 0$), more are significantly positive ($t > 2$), and less are significantly negative ($t < -2$). So there is less significance in the annual maximum rainfall series and lower estimates.

In conclusion from the t-test results:

1. Some of the results concur with and refine the result of the likelihood ratio tests.

   A There is significant evidence that the trend in the annual total rainfall series is different from 0 ($\gamma \neq 0$). In fact, there is significant evidence that $\gamma > 0$.

   B There is significant evidence that the common trend is different from 0 ($\tau \neq 0$). In fact, there is substantial evidence that $\tau > 0$.

   C The trend in the annual total rainfall series dominates the common trend and hence the trend in the annual maximum rainfall.

2. Some of the results differ from the results in the likelihood ratio tests.

   A Here there is significant evidence that the trend in the annual maximum rainfall series is different from 0 ($\beta \neq 0$). In fact, there is some evidence that overall $\beta > 0$ although in the expansion model there is a higher percentage of stations which are significantly negative.

   B Roughly, the results in the independent and expansion models when estimating β are comparable – again, except for the slight increase in significant negative βs.

3. Based on these t-test, there is evidence that in selecting an appropriate model to estimate the trends in the annual maximum and annual total rainfall series, one must use a model which has a trend in the annual maximum rainfall somehow (E5 or E4). Since the results for the estimates for β are different from the results for τ and hence γ, we would conclude based on the t-test results that the most appropriate model is E5 ($\beta \neq 0, \gamma \neq 0$).

Introduction to Spatial Smoothing Method The other two methods we use to help interpret the 187 station trend estimates are products of a spatial smoothing method established by Smith (2000). From this method we will procure a national average for the estimates of these trends and a smoothed version of these 187 stations results, interpolated to the entire contiguous US. We begin by summarizing the procedure in Smith (2000). Then we outline the steps taken in that method for this application of the trend analysis. Next we organize the national averages obtained in the previous step and explain these results with respect to the results of the previous hypothesis tests and the questions to be answered. After making our recommendation for the most appropriate model in this analysis, we interpret the results of this model's trend estimates on a national level, using the national averages; on a regional level by looking once again at the likelihood ratio table *and* by producing mapped estimates of the trends in the annual maximum and annual total rainfall series; and at a local level by looking at how well these final estimates of the trends fit the five stations we studied in Section 5.6.1.

Spatial smoothing method In this Section we outline the method developed in Smith (2000) for spatially smoothing the trend estimates of the previous sections. To do so, we introduce the following model.

Let $\lambda(s)$ be a parameter of interest. Note Smith (2000) develops the method with a parameter vector. Since in any given instance we are looking at just a trend estimate, we drop reference to the parameter vector. It is assumed $\lambda(s)$ varies smoothly as a function of spatial location s lying in some domain S. For each finite subset of locations $s \in \{s_1, \ldots, s_n\}$ we observe a time series $Y(s,t)$ where t is time and whose distribution depends on $\lambda(s)$. Suppose we have the following heirarchical model:

$$(\theta, \phi) \sim \pi(\theta, \phi) \tag{5.27}$$

$$\lambda|\theta \sim f(\lambda|\theta) \tag{5.28}$$

$$Y(s,.) \sim g(y(s,.)|\lambda(s), \phi). \tag{5.29}$$

where (5.27) gives the prior density of the nuisance parameters θ and ϕ, (5.28) defines the spatial distribution of $\{\lambda(s), s \in S\}$ as a function of θ, and (5.29) gives the distribution of the time series at one site s as a function of $\lambda(s)$ and possible nuisance parameter ϕ.

Instead of implementing a full Bayesian approach, Smith (2000) develops an alternative approximate method for estimating a smoothed estimate of $\lambda(s)$. From (5.29) one can estimate the parameter $\lambda(s)$ from the observed data at each $s \in \{s_1, \ldots, s_n\}$. Let $\hat{\lambda}(s)$ be the estimate and define the error as $\eta(s) = \hat{\lambda}(s) - \lambda(s)$. Now to obtain the distribution of $\hat{\lambda}$, we use the following three pieces. First we exploit the asymptotic normality of $\{\eta(s_1), \ldots, \eta(s_n)\}$ which has mean 0 and covariance matrix W. Then we assume the underlying random field $\{\lambda(s), s \in S\}$ is Gaussian with mean and covariance matrix given by a finite parameter model with respect to θ. In particular, let the mean and covariance matrix be $\mu(\theta)$ and $\Sigma(\theta)$, respectively. Finally, we assume $\lambda(s)$ and $\eta(s)$ are independent, Thus Smith (2000) can then writes the estimate as

$$\hat{\lambda}(s) \sim N(\mu(\theta), \Sigma(\theta) + W) \tag{5.30}$$

Note we may let $\Sigma(\theta)$ take the form

$$\Sigma(\theta) = \alpha V(\theta)$$

where α is an unknown scale parameter and $V(\theta)$ is a vector of standardized covariances determined by an unknown parameter θ.

Once θ has been estimated – using any standard estimation method, such as maximum likelihood – one can estimate $\lambda(s)$ at both the actual stations and across the entire domain S by standard kriging. These estimates are the smoothed estimates $\tilde{\lambda}(s)$.

Now to fit the above model, one must specify the parametric models for $\mu(\theta)$, $\Sigma(\theta)$ and specify the error covariance matrix W.

Finally, from $\tilde{\lambda}(s)$ the smoothed estimates of $\lambda(s)$ for $s \in S$, we can obtain regional averages for any subset of S by integrating these estimates over the region of interest

A where $A \subset S$. Define $\overline{\lambda}(A)$ as the regional average of $\tilde{\lambda}(s)$ over region A. We have:

$$\overline{\lambda}(A) = \frac{\int_A \tilde{\lambda}(s)ds}{|A|} \tag{5.31}$$

where $|A|$ is the area in the grid corresponding to A.

The corresponding standard errors to these regional averages can be calculated through standard kriging techniques, see Smith (2000) for further details.

Smith (2000) has established an algorithm/program which fits the above method. In it, he specifies the following features to be considered and the options that are available:

1. For estimating θ, we can choose either ML, maximum likelihood method, or REML, restricted maximum likelihood method. See Smith(2000) for details between the two.

2. For the error covariance matrix W, one can take W to be a diagonal matrix, implying independence among the stations' errors. Note the obvious input for W when using a quasi-Newton algorithm in the estimation process is the estimate of the Hessian matrix that the algorithm provides. Alternatively, one can input a sample correlation matrix into the program to model dependence between the stations in the error process.

3. For $\Sigma(\theta)$, one can choose from five forms: exponential, Gaussian, wave, spherical, and Matérn. See Smith (2000) for further details on the difference between these.

4. There is an option to include the nugget effect in $\Sigma(\theta)$. The nugget effect arises from a discontinuity point and has the interpretation as either measurement error when observations are being replicated at a site or a *microscale* process which involves discontinuities at distances smaller than the distances between the sites themselves.

5. There is an option to take an orthogonal transformation if the data are suspected of geometric anisotropy. Anisotropy refers to the situation when the dependence between the sites is both in terms of distance and direction.

6. Finally for $\mu(\theta)$, it may take a deterministic form or a polynomial of order 0,1,2,3, or 4.

Application to the Spatial Smoothing Method The above method allows us to take the trend parameters from our analysis of the 187 stations and interpolate to the entire contiguous US. Thus we can get an idea of how the trends in the annual maximum and annual total rainfall behave across the US. Specifically in the above model, let $\lambda(s)$ be the trend estimate (either β, γ, or the the common trend τ), $s = \{s_1, \ldots, s_{187}\}$ be the 187 station sites, and S be the contiguous US grid.

In the Smith (2000) program we use:

1. The maximum likelihood method of estimation for θ.

2. To specify the error covariance matrix W, we assume W to be diagonal with entries determined by the standard errors of the maximum likelihood analysis. Note in doing so, we are assuming no correlations between the stations' estimated trends.

3. No nugget effect in $V(\theta)$.

4. The error covariance matrix $\alpha V(\theta)$ takes an exponential form; that is, where the (i^{th}, j^{th}) entry – without a nugget effect – is

$$v_{i,j} = \alpha e^{-d_{ij}/R}$$

   where d_{ij} is the distance between the i^{th} and the j^{th} sampling points and $R > 0$ is the unknown range parameter. Note the other forms make no overall difference to the mapped estimates.

5. Assuming no orthogonal transformations, i.e., assuming no anisotropy in the data.

6. For $\mu(\theta)$: We fit polynomials of order 0, 1, 2, 3, or 4 to the data and choose the best fit by comparing likelihood ratio test statistics, AIC criterion, and BIC criterion.

Trend in Total (γ): Expansion Model						
order of poly	# of par's	mle of log(α)	mle of R	nllh	AIC/2	BIC/2
0	2	.202	1.249	166.83	168.83	172.06
1	4	.192	1.248	166.61	170.61	177.07
2	7	.171	1.223	165.89	172.89	184.20
3	11	.150	1.154	165.42	176.42	194.20
4	16	-.012	.904	159.12	175.12	200.97

Table 5.11: Spatial model results for γ in expansion case. The table gives the order of the polynomial and number of parameters in the model, mle of parameters, negative log likelihood, and the AIC and BIC criterion.

Note the decision rules for the criterion in the above model selection are as followed:

Likelihood ratio test Let l_q be the log likelihood function of a model based on q parameters and l_p be the log likelihood function of a nested model based on p parameters with $q > p$. We claim there is significant evidence to reject the model based on p parameters in favor of the model based on q parameters if $l_q - l_p > \frac{1}{2}\chi^2_{q-p}$. Note for negative log likelihood functions, we use $l_p - l_q$.

AIC Choose model which minimizes the AIC function: $-2l_p + 2p$.

BIC Choose the model which minimizes the BIC function: $-2l_p + p\log(n)$ where n is sample size.

Note the following tables give the step by step results in this process for the trend in the annual total rainfall – γ – for full expansion model, the trend in the annual maximum rainfall – β – for full expansion model, and the common trend – τ – for expansion model.

Trends for γ in expansion model

The results for the spatial analysis for the trend parameter for the annual total rainfall in the full expansion model are in Table 5.11. We find for $\mu(\theta)$ we should use a polynomial

Trend in Maximum (β): Expansion Model						
order of poly	# of par's	mle of log(α)	mle of R	nllh	AIC/2	BIC/2
0	2	.180	.965	170.48	172.48	175.71
1	4	.134	.844	168.84	172.84	179.30
2	7	.041	.628	165.99	172.99	184.30
3	11	-.002	.565	164.25	175.25	193.02
4	16	-.176	.360	156.97	172.97	198.82

Table 5.12: Spatial model results for β in expansion case. The table gives the order of the polynomial and number of parameters in the model, mle of parameters, negative log likelihood, and the AIC and BIC criterion.

of order 0 because: (a) The drop in the log likelihood function from the model using polynomial of order 0 to order 4 is not significant; i.e., $2 \times (166.83 - 159.12) = 15.42 < \chi^2_{14,.05} = 23.685$. (b) AIC is minimized for polynomial of order 0. (c) BIC is minimized for polynomial of order 0. Thus we maintain that the best model for $\mu(\theta)$ is μ. Using the estimates in the table 5.11 for order 0 polynomial, we can estimate γ at the 187 stations and across the 100×100 grid covering the contiguous US. Converting the integral in (5.31) into the appropriate sum over the grid, we can evaluate this average. Since we sum over the gird encompassing the continetal US, this is a national average. Standard kriging techniques (see Cressie(1993) for more details) provides a similar formula to calculate standard errors. Based on a fit with order 0 we find a national average of .71131 with a standard error of .01265.

Note for comparison we find the national average based on polynomial of order 4 to be .67059 with a standard error of .01304. The difference between the national averages in the model which uses a polynomial of order 0 and one which uses order 4 is relatively small, what will amount to 0.4% difference in the percent increase of annual total rainfall in the past century. Thus we see the spatial smoothing method is relatively stable, although the difference in the national averages are not negligible, with respect to the order of the polynomial in the mean of the Gaussian random field.

Trends for β in expansion model

The results for the spatial analysis for the trend parameter for the annual total rainfall in the full expansion model are in Table 5.6.3. We find the overall the best model for $\mu(\theta)$ appears to be a polynomial of order 0. The BIC criterion clearly signals the polynomial of order 0 as the best fit. For the AIC criterion, the polynomial of order 0 is again selected but the margin is slight compared to the fit with polynomial of order 1 or 4. Finally, the likelihood ratio test is somewhat contradictory. The drop in the log likelihood function from a model with a polynomial of order 0 to order 4 is only just significant at 0.05 level of significance. Here $2 \times (170.48 - 156.97) = 27.02 > \chi^2_{14,.05} = 23.685$ and the biggest jump in log likelihood functions is from order 3 to order 4. This implies that polynomial of order 4 should be considered a competitor for the most appropriate model. There are three reasons as to why a polynomial of order 4 may not be suitable: (a) It is just significant – not overwhelming evidence that this is the correct model. (b) It is difficult to extrapolate a polynomial of order 4 beyond its exact domain. (Recall we are trying to obtain smoothed estimates for the entire contiguous US.) (c) In all other cases, a model using a polynomial of order 0 is the best fit.

When we look at the difference in the national averages based on a model which fit a polynomial of order 0 and of order 4, we find for a model based on a fit with order 0, a national average of .29675 with a standard error of .01304. Based on a polynomial of order 4, we find a national average of .22573 with a standard error of .01265. At this point we do not elect one of these averages as the correct version. Rather we view these as giving an idea of a range in which the true national average fits. In fact, for all of these trends which we find a smoothed version of the estimates we give both the national average for the fit with a polynomial of order 0 and of order 4. We consider the averages calculated as representing this range of possible values for the national average. Ultimately, we will quote the national average based on a polynomial of order 0 fit but here we show both averages to understand the spatial smoothing process.

Common Trend (τ): Expansion Model						
order of poly	# of par's	mle of $\log(\alpha)$	mle of R	nllh	AIC/2	BIC/2
0	2	.277	.580	164.76	166.76	167.00
1	4	.240	.506	163.30	167.30	173.46
2	7	.184	.364	160.40	167.40	176.09
3	11	.142	.285	158.32	169.32	187.09
4	16	.120	.216	157.32	173.32	199.17

Table 5.13: Spatial model results for τ in expansion case. The table gives the order of the polynomial and number of parameters in the model, mle of parameters, negative log likelihood, and the AIC and BIC criterion.

Trends for τ in expansion model

The results for the spatial analysis for the trend parameter for the annual total rainfall in the full expansion model are in Table 5.6.3. We find for $\mu(\theta)$, we should clearly use a polynomial of order 0. Again the drop in log likelihood functions from a model with polynomial of order 0 and of order 4 is not significant. Here we have $2 \times (164.76 - 157.32) = 14.88 < \chi^2_{14,.05} = 23.685$. Also, both the AIC and BIC are minimized for a fit with polynomial of order 0. Based on a fit with order 0 we find a national average of .56835 with a standard error of .01378.

Note based on a fit with order 4, we find a national average of .47076 with standard error of .01378. Here the difference in the national average is more substantial but since the common trend is not ultimately considered the best overall model, this difference is not considered important.

Spatial analysis for independent model

Note: The above steps in Smith's (2000) program were also performed for the independent version of the estimates for the trend in the annual total rainfall, annual maximum rainfall, and common trend in the annual total and annual maximum rainfall. The following tables give the results:

Trend in Total (γ): Independent Model						
order of poly	# of par's	mle of log(α)	mle of R	nllh	AIC/2	BIC/2
0	2	.295	1.147	175.36	177.36	180.59
1	4	.286	1.144	175.06	179.06	185.52
2	7	.279	1.151	174.72	181.72	193.03
3	11	.252	1.070	173.99	184.99	202.76
4	16	.087	.910	166.94	182.94	208.79

Table 5.14: Spatial model results for γ in independent case. The table gives the order of the polynomial and number of parameters in the model, mle of parameters, negative log likelihood, and the AIC and BIC criterion.

Trend in Maximum (β): Independent Model						
order of poly	# of par's	mle of log(α)	mle of R	nllh	AIC/2	BIC/2
0	2	.135	.668	189.99	191.99	195.22
1	4	.098	.564	188.19	192.19	198.65
2	7	.030	.421	186.42	193.42	204.73
3	11	.001	.458	185.47	196.47	214.24
4	16	-.188	.344	178.95	194.95	220.80

Table 5.15: Spatial model results for β in independent case. The table gives the order of the polynomial and number of parameters in the model, mle of parameters, negative log likelihood, and the AIC and BIC criterion.

Decision for γ : From Table 5.6.3, we find the following. For $\mu(\theta)$ we should use a polynomial of order 0. The drop in the log likelihood function is $2 \times (175.36 - 166.94) = 16.84 < \chi^2_{14,.05} = 23.685$ which is not significant. Also the AIC and BIC are minimized for polynomial of order 0. Based on a fit with order 0 we find a national average of .73837 with a standard error of .01342.

For comparison, based a model with polynomial of order 4 we find a national average of .65125 with a standard error of .01342. Note in all the comparisons this is the only case where the national average of the estimates of the trend in the independent model are smaller than the expansion model.

Common Trend (τ): Independent Model						
order of poly	# of par's	mle of $\log(\alpha)$	mle of R	nllh	AIC/2	BIC/2
0	2	.258	.348	161.64	163.64	166.87
1	4	.236	.319	160.29	164.29	170.75
2	7	.216	.290	158.96	165.96	177.27
3	11	.154	.230	156.27	167.27	185.04
4	17	.152	.194	155.86	171.86	197.71

Table 5.16: Spatial model results for τ in independent case. The table gives the order of the polynomial and number of parameters in the model, mle of parameters, negative log likelihood, and the AIC and BIC criterion.

Decision for β : From Table 5.6.3 we make the following decision. For $\mu(\theta)$ we should use a polynomial of order 0. The drop in the log likelihood function is $2 \times (189.99 - 178.95) = 22.08 < \chi^2_{14,.05} = 23.685$ which is not significant. Again AIC and BIC concur with the likelihood ratio test. Based on a fit with order 0 we find a national average of .45144 with a standard error of .01304.

To compare with a fit using order 4, we find a national average of .3342 with a standard error of .01304.

Decision for τ : Table 5.6.3 contains the results for this case. For $\mu(\theta)$ we should use a polynomial of order 0. The drop in the log likelihood function is $2 \times (161.64 - 155.86) = 11.56 < \chi^2_{14,.05} = 23.685$. Again, the AIC and BIC criterion agree. Based on a fit with order 0 we find a national average of .58763 with a standard error of .01414.

To compare with a fit based on a polynomial of order 4 we find a national average of .51714 with a standard error of .01449.

Not only did the above spatial fit using the Smith (2000) method and program yield the national averages but the program also produced smoothed estimates of the above trends across the contiguous US. Using these smoothed estimates we may plot mapped

Fit with Polynomial of Order 0			
	$\overline{\tau}$	$\overline{\beta}$	$\overline{\gamma}$
Independent	0.588 (0.014)	0.451 (0.013)	0.738 (0.013)
Expansion	0.568 (0.014)	0.297 (0.013)	0.711 (0.013)
Difference: I-E	0.020	0.153	0.027

Table 5.17: National averages of the trend estimates produced by spatial model with polynomial of order 0. The table gives the national averages for parameter estimates of τ in model I4 and E4 and the estimates of β and γ in model I5 and in model E5 with estimated standard errors in parentheses. Table also gives the difference between the independent and expansion models for the estimates of τ, β, and γ.

estimates of these trend estimates. These plots are used in the analysis, see Section (5.6.4).

National Average Results From the spatial smoothing analysis we construct the following tables. First, let us denote the national average for the trend in the annual maximum rainfall series as $\overline{\beta}$; the national average for the trend in the annual total rainfall as $\overline{\gamma}$; and the national average for the common trend in the annual maximum and annual total rainfall series as $\overline{\tau}$. Table 5.17 gives the national averages for the fit using a polynomial of order 0 in $\mu(\sigma)$. Table 5.18 gives the national averages for a fit using a polynomial of order 4 in $\mu(\sigma)$. In each table we present the national average calculated from the estimates of β in model E5 and I5, the estimates of γ in model E5 and I5, and the estimate for the τ in models E4 and I4. Given the results of the likelihood ratio test, we narrow the possible models for the trend analysis to E3, E4, or E5. Given the results of the t-tests, particularly, $\beta \neq 0$, we narrow the results further to model E4 or E5. Here we present the national averages for the estimates in E5 and E4 to distinguish the "best" model. The independent versions are given for a cross reference.

Recall we do not take either the national averages based on a polynomial of order 0 or of order 4 as being correct and the other as being wrong. Again, we view the difference between these national averages as a range in which the national average lie.

Fit with Polynomial of Order 4			
	$\overline{\tau}$	$\overline{\beta}$	$\overline{\gamma}$
Independent	0.517 (0.014)	0.334 (0.013)	0.651 (0.013)
Expansion	0.471 (0.014)	0.226 (0.013)	0.671 (0.013)
Difference: I-E	0.046	0.108	-0.020

Table 5.18: National averages of the trend estimates produced by spatial model with polynomial of order 4. The table gives the national averages for parameter estimates of τ in model I4 and E4 and the estimates of β and γ in model I5 and in model E5 with estimated standard errors in parentheses. Table also gives the difference between the independent and expansion models for the estimates of τ, β, and γ.

First we note that in all the cases the averages based on a polynomial of order 0 are higher than the corresponding averages based on order 4. The spatial model based on order 4 tends to smooth out the estimates, in particular, extending the negative regions. Also the spatial model using order 0 always produces national averages for the expansion models which are lower than the independent models. In the spatial model using order 4 the average common trend in expansion model E4 is less – not significantly less – than the average for the common trend in the independent I4. In further discussion we will quote results based on the spatial model using a polynomial of order 0 but the conclusions are equivalent for the results using a polynomial of order 4.

We conclude the following from the national averages for the estimates in E4 and E5. Clearly by comparing the national averages to their standard errors, at a national level the trend in the annual total rainfall is significantly positive: $\overline{\gamma} = 0.711$ (0.013). This concurs with both the likelihood ratio test and the t-tests. Also concurrent with both types of tests, we have at a national level the common trend is significantly positive: $\overline{\tau} = 0.568$ (0.014). *Now* one of the most important results from these averages is that at a national level the trend in the annual maximum rainfall is significantly positive: $\overline{\beta} = 0.296$ (0.013). This reinforces the t-test results (as opposed to the likelihood ratio tests) that the appropriate model for the annual rainfall must have a parameter that models, in some way, the trend in the annual maximum – either E4 or E5.

There are two more important conclusions that can be drawn from these national averages. The first is that at a national level $\overline{\gamma} \neq \overline{\beta}$. For example, a t-test for $H_o : \overline{\gamma} = \overline{\beta}$ vs $H_a : \overline{\gamma} \neq \overline{\beta}$ gives a test statistic of 22.186 – overwhelmingly significant. In fact, we see the national average for the trend in the annual total is more than twice the national average for the trend in the annual maximum. This again gives evidence that the most appropriate model is E5 which contains a separate parameter for both the trend in the annual maximum rainfall and the trend in the annual total rainfall. Second is that the national average for the common trend $\overline{\tau}$ is "closer" to the national average for the trend in the annual total. The difference between $\overline{\tau}$ and $\overline{\gamma}$ is 0.143. The difference between $\overline{\tau}$ and $\overline{\beta}$ is 0.271. This supports the previous conclusions from the tests that the trend in the annual total rainfall dominate the common trend.

Finally, a quick comparison between the expansion and independent models. We see the national averages for the estimates of the trend in the annual total rainfall $\overline{\gamma}$ and for the common trend $\overline{\tau}$ there is no real difference between the results in the expansion and independent models. This is not the case for the national averages for the trend in the annual maximum rainfall $\overline{\beta}$. The national average based on model E5 is substantially and significantly (test statistic 8.32) less than the national average based on model I5. In fact, the national average for the model E5 is roughly a third smaller. This bears out the results in the likelihood tests that for the trend in the annual maximum rainfall using the expansion model *does* impact the analysis.

In conclusion from the national averages produced by the spatial method

1. The national average of the trend in the annual total rainfall is significantly positive: 0.711 (0.013).

2. The national average of the trend in the annual maximum rainfall is significantly positive: 0.296 (0.013).

3. The national average for the common trend is significantly positive and is dominated by the trend in the annual total rainfall: 0.568 (0.014).

4. The national average of the trend in the annual maximum rainfall does not equal the national average of the trend in the annual total rainfall: $\overline{\beta} \neq \overline{\gamma}$.

5. There is a significant difference between the expansion model and the independent model with respect to the national average for the trend of the annual maximum. The expansion model produces a smaller estimate for this national average. This difference is not seen in the national averages for the annual total rainfall or the common trend. Note this difference between the expansion and the independent model with regard to the trend in the annual maximum rainfall also is seen in the likelihood ratio test results, although to a much lesser extent.

Conclusion of Model Analysis

What emerges from the results in the likelihood ratio tests, the t-tests, and the national averages is a clear yet more subtle view of the relationship between the trend of the annual maximum and annual total rainfall. We look back to the questions raised in Section 5.6.3.

For Q1 (question 1) We find that there is substantial evidence that both the trend in the annual total and annual maximum rainfall are overall significantly positive. This is well established for the trend in the annual total rainfall: the models in the likelihood ratio tests that have the highest percentage of significance are those which include a parameter for the trend in the annual total rainfall; the t-tests for γ that detect significance from 0 are those in the positive range; finally, the national average is significantly positive. This conclusion of significant evidence for the trend in the annual maximum rainfall is more ambiguous: the likelihood ratio tests are non-significant for models with (just) a trend in the annual maximum; the t-test have a considerable percentage of significant positive trends but also a distinct percentage of significant negative trends; the national average is significantly positive but not as high as the national average in the annual total rainfall.

For Q2 The trend in the annual total rainfall is more dominant both in terms of significant evidence and magnitude. In the likelihood ratio tests and the t-tests, more stations are significant for models which contain a parameter for the trend of the annual total than one which contain a parameter for the trend in the annual maximum. The national average for γ in E5 is 0.711 which is much larger than the national average for β in E5 which is 0.297. This does not mean that the trend in the annual maximum rainfall should be disregarded.

For Q3 There is sufficient evidence to conclude that introducing a common trend parameter into the model is significant and that trend is positive: the likelihood ratio tests have a substantial percentage of significant outcomes, the t-tests have a substantial percentage of significant positive test statistics, and the national average is significantly positive. What is less conclusive is whether a model which differentiates between the trend in the annual maximum and the trend in the annual total is a better model? The likelihood ratio test results would indicate that the model should not distinguish between the trend in the annual maximum and the trend in the annual total rainfall. Those results differ from the results of the t-tests and the national averages. The t-test results indicate that the majority of the estimates for the trend in the annual maximum is positive (60%). In fact, the percentage of significant positive trends is substantially higher than expected if there was no trend in the annual maximum. There is also a number of significant negative trends. Overall the national average is significantly positive. In comparing the estimates of the trends in the annual maximum and annual total rainfall, we see see fewer significant positive trends (test statistics over 2) and more significant negative trends (test statistics under -2) for the annual maximum as opposed to the annual total. This is evident in the national averages where the average annual total rainfall [0.711(0.013)] is twice as large as the average for the annual maximum rainfall [0.297 (0.013)]. We conclude the trends in the annual maximum and the annual total are different – both in level of significance and magnitude. The trend in the annual total rainfall is higher for both criteria. We conclude that the trends are not the same and we should model them differently.

For Q4 The expansion model which contains a separate parameter for the trend in the annual maximum and the annual total rainfall is the model E5 ($\beta \neq 0, \gamma \neq 0$). Based on the discussions for the above questions (Q1-Q3), we claim this is the most appropriate model for the annual rainfall series with respect to modeling the trends in the annual maximum and annual total rainfall for the contiguous US.

5.6.4 Applying the Results of the Data Analysis

At the National Level

Obviously the best way to look at the trend in the annual maximum and annual total rainfall at the national level is by looking at the national averages of those trends: $\overline{\beta}$ and $\overline{\gamma}$. Having decided the full expansion model E5 which has a separate parameter for both the trends is the "best" fit, we find that over the past century the annual maximum rainfall has increased by

$$e^{\frac{.297}{1000}(100)} = e^{.0297} = 1.030 \text{ or } 3.0\%.$$

The annual total rainfall has increased by

$$e^{\frac{.711}{1000}(100)} = e^{.0711} = 1.074 \text{ or } 7.4\%.$$

This means the annual total rainfall is increasing more than twice the rate of the annual maximum rainfall.

In comparison, the independent model I5 gives rates of 4.6% and 7.7%, respectively, for estimate of the trend in the annual maximum and annual total rainfall. Recall from the beginning descriptive statistics that the least squares estimates for the trends for the average annual maximum and average annual total rainfall across the stations were 6.5% and 8.5%, respectively.

Again we see that the expansion model and the independent models give very similar results for the trend in the annual total rainfall – around a 7.5% increase. We also see that the least squares estimates results – although an entirely different approach and methodology – is also very close to the other results for the annual total. Clearly the annual total rainfall in the contiguous US is increasing and the rate is somewhere near 7.5%.

As for the trend in the annual maximum rainfall, we see results that are substantially different from the expansion model to the independent model and certainly from the results based on the summary statistics. The independent model gives an estimate of this increase 50% higher than the expansion case. The summary statistics method – least squares fit – gives a result over 100% higher than the expansion model. Although the estimates of this increase differ from model to model, each is significantly positive; thus, we conclude the annual maximum rainfall at the national level is increasing. We can see for all the models that this increase is less than the increase for the annual total rainfall.

At the Regional Level

To see how the 187 station estimate of the trends in the above model vary across the US, we produce mapped estimates of these trends – both β and γ – from the full expansion model E5.

Using the "best" spatial model from the previous section which used a polynomial of order 0 in the mean of the Gaussian random field, a standard kriging procedure interpolates these 187 stations estimates to a 100×100 grid which covers the entire continental US. Due to the fact the US is not a rectangle, there are actually 6319 sites estimated in this procedure.

An important part of this spatial procedure is that the 6319 estimates produced have been smoothed. This smoothing procedure extends to the original 187 individual stations. The diversity of the 187 individual stations sites is dampened due to the spatial model which is now incorporated. These smoothed estimates tend to be more consistent and thus have a smaller variance. For example, the original estimates of β from model E5 for the 187 stations range from -5.10 to 5.11 with a mean of .261 and standard deviation 1.50. For the original 187 estimates of γ in E5, the range is -6.51 to 11.47 with a mean of .701 and standard deviation of 1.64. The 6319 smoothed estimates for β have a range of -1.91 to 2.34 with a mean of .297 and standard deviation of .374. The 6319 smoothed estimates for γ have a range of -2.73 to 2.74 with a mean of .711 and standard deviation of .481. From these summary statistics, we see the estimates for the trend in the annual maximum tend to be more consistent, with a smaller range and less variability, than the smoothed estimates of the trend in the annual total rainfall. We see more variance in the annual total trend estimates.

These smoothed estimates were then entered into an S-Plus program to produce the graphs. The graphs were constructed by sorting the trend estimates into nine colors using a linear transformation. The scales of the graphs for the smoothed estimates of β and γ were made consistent so that we could make direct comparisons between the two graphs. Note due to the large range in the four smallest smoothed estimates of γ, these four estimates were Winsorized (to -2) before being inputted into the S-Plus program. The mapped Winsorized smoothed estimates of β and γ are shown in Figure 5.12 (a) and (b) respectively.

To interpret these graphs, particularly in differentiating between colors on these graphs, we look at the standard errors associated with these 6319 Winsorized smoothed trend estimates of β and γ. For the smoothed estimates of β, the average of these 6319 standard error is 1.05. For the γ case, the average of these 6319 standard errors is 1.02. If we look at the legend associated with these graphs, the range of these nine colors is 4.7445 so that the range between any of the two adjacent colors is 0.527. We see first that the negative estimates are associated with colors purple down to black. The positive estimates are associated with the colors magenta up to white. We also see to declare any significance between colors we need them to be separated by 3 almost 4 shades. For example, we could infer that the black regions are significantly different from the purple, certainly the magenta or lighter colors. The true yellow region is significantly higher than the magenta, certainly the purple region. The hardest to decipher and most important color comparison is the blue to magenta colors. Each is not far enough away from the border of purple/magenta which is roughly the break between the negative and positive estimates to declare significantly different from 0. But the question is are they significantly different from one another? The estimates at the center of the two blue zones are around -.92. The estimates at the center of the two magenta zones are around 0.66. The distance between these centers is thus 1.58 – near significance. *Roughly* we can infer that the colors are far enough away from each other to indicate a difference in these trends (negative/positive) but that this difference is not overwhelming. In other words, each color is not significantly different from the next color, but across the whole range of colors, it is significant. In fact, the lowest level (black) is significantly different from the mid-level (magenta) which is significantly different from the highest estimates (yellow).

We can draw some basic conclusions. First is that for both the estimates of β and γ, the graphs display more positive than negative estimates – strengthening our conclusion that these trends are overall positive. Second is that the estimates for γ are higher than the estimates for β. The estimates for β are negative for a larger percentage of the country. For the estimates of γ, there are only a few patches of negative estimates and more importantly a higher percentage of very positive (yellow) estimates. This also shows the estimates of γ have a larger range than the estimates of β. Clearly the estimates of the trend in the annual total rainfall are higher than in the annual maximum rainfall.

Due to the somewhat inconclusive distinction between the magenta and blue zones, interpreting within the graphs is more problematic, particularly for the estimates of β. Although there are sporadic patches of blue in the mountain region, the upper midwest, the mid-atlantic states and southern Florida, the only concentrated regions of negative estimates of the trend in the annual maximum rainfall (blue) are in the northwest coast and the Missouri/Mississippi area. Also there are no extremely high estimates of β (yellow) except for a handful of isolated sites. Again we reiterate that the estimates of β tend to be more consistent throughout the US. The few significant differences we see tend to be more local as opposed to fitting into any regions – particularly the five regions defined previously.

At first the estimates of γ look more consistent throughout the country than estimates of β – mostly positive (magenta). On closer inspection, we see more levels in this graph than in the one for β. Given that the difference between the yellow and the magenta is roughly 2.00 in the magnitude of the estimates (recall the average standard error is around 1.00), these colors are more significantly different than the blue to magenta – the main comparison in the β graph. The west coast and the mountain region is the most active – certainly a number of areas of very high estimates of γ (yellow) within very close proximity to areas with low to very low estimates of γ (black/blue). East of the Rockies, the estimates of γ tend to be uniform except for a cluster of very high estimates (yellow) in upper Mississippi and a cluster of somewhat negative estimates (blue) in the mid-atlantic states. In other words for the estimates of γ there are numerous pockets of very different trends – significantly negative and significantly positive estimates within 250 miles of each other – which make the estimates of γ less consistent than the estimates of β. This does not imply there are significant regional differences with respect to the estimates of γ. Again these differences tend to be more local than any regional classification.

Finally, we make some observations of the difference between the estimates of β and γ. Again we caution on emphasizing these observations based on the color differences since we lack an exact criterion for differentiating between these colors. Having said that, there are two interesting pieces to note. The first is that every pocket of blue in figure (5.12)(b) for γ corresponds to a blue zone in figure (5.12) (a) for β. In other words, if a region has a decreasing trend in the annual total rainfall, we see a decreasing trend in the annual maximum rainfall. This does not work the other way. There are regions where the trend in the annual maximum is not increasing but the trend in the annual total rainfall is increasing. The biggest difference between the trends in the annual maximum and the annual total rainfall is our second comment. Look at the region around Missouri/Louisiana to Alabama. For the trend in the annual maximum, this area is blue, indicating a region where the trend is decreasing. For the trend in the annual total rainfall, this area is magenta to bright yellow, indicating significantly increasing trends. Over the past century this region is less likely to see a very heavy rainfall but overall to see more rain throughout the year. Due to the lack of overall significantly regional differences in either the estimates of β or γ, depicting any significant difference between the graphs is not necessarily reliable.

Clearly there is no visual confirmation to the five regions previously defined in Section 5.3. Not only are both the estimates for β and γ roughly uniform throughout the country but any exceptions are in local spots throughout the US. To confirm this, χ^2 tests between the five regions were performed on the results of the likelihood ratio tests for the expansion models, see Table 5.9. The χ^2 tests looked for significant regional differences between the results. None of the tests was significant. The smallest p-value was 0.16. This was associated with the test T4 – the significance of adding a second trend parameter above the common trend parameter into the model. Region 2 – the mountains – shows the highest percentage of stations which significantly differentiate between β and γ. Note in Figure 5.12(b) for γ we see many pockets of yellow and even black compared to Figure 5.12(a) which is uniform for the mountain region. Region 3 – the plains – has the smallest percentage of stations which differentiate between the two trends. Note in both Figures 5.12(a) and 5.12(b) that this region is mostly magenta – positive estimates.

In conclusion, at a regional level,

1. The range of Winsorized smoothed estimates mapped in figures 5.12(a) and 5.12(b)

is -2 to 2.74 with an average standard error near 1.00. This implies making any declaration of significant difference based on the different colors in the graphs should be done cautiously. This does not mean that the different colors represent statistical noise. We do need 3 or 4 shades between the colors though, before we infer the estimates as being significantly different.

2. Overall the estimates for γ – the trend in the annual total rainfall – tend to be higher than the estimates for β – the trend in the annual maximum rainfall.

3. The estimates of β tend to be more consistent throughout the country.

4. The estimates of γ show more variability within closer proximity, particularly in the western third of the country.

5. Neither the estimates of β nor γ show any significant regional differences. Each displays pockets where the estimates are different from neighboring spots – more for the estimates of γ – but these pockets tend to be more local than any sweeping regions.

6. This lack of visual differentiation into regions is consistent with the lack of any significant difference in the regions with respect to the χ^2 tests performed on the likelihood ratio test results.

7. The best interpretation of these trends is either (1) to ignore the batches of variability and focus on the national level or (2) to find a better definition for a local average.

At the Local Level

Given the results of the trend analysis, we go back to the five individual stations we have highlighted to see how these results fit at the individual station level. To understand the differences between these stations, we look at the results of the analysis for the individual stations themselves in Table 5.19.

	Ind T1 ($\beta \neq 0$)	Ind T2 ($\gamma \neq 0$)	Exp T3 ($\tau \neq 0$)	Exp T4 ($\beta \neq \gamma$)
Berkeley, CA	5.46 (4.63)	3.77 (2.16)	4.18 (4.75)	5.11, 3.25 (.44)
Caldwell, ID	-.71 (.14)	1.76 (1.66)	1.14 (.87)	-.78,1.87 (2.15)
Le Mars, IA	.19 (.01)	-.28 (.05)	-.32 (.09)	.20, -.41 (.12)
Eastport, ME	3.71 (6.07)	2.62 (6.98)	2.84 (11.00)	3.36, 2.66 (.33)
Savannah, GA	.71 (.12)	1.63 (2.45)	1.63 (2.77)	.48, 1.70 (.55)

Table 5.19: Some individual station results for the trend estimates and their associated likelihood ratio tests. The table gives the station and the parameter estimates of the trends with the difference in the log likelihood functions for the given tests in parenthesis. Recall significance when the difference in log likelihood functions greater than 1.92

In Table 5.19 we see that Berkeley and Eastport are similar: Both have γ and β significant in the independent models, common trend significant for the expansion model but not significant for the additional parameter estimate. Savannah has just γ significant in the independent model and just the common trend significant in the expansion models. Caldwell has just γ almost significant but not a significant common trend but significant separate trends for the expansion models. Le Mars has nothing significant. Note that Le Mars has a negative estimate of γ and Caldwell has a negative estimate of β although neither is significant.

If we go back to Figures 5.7 and 5.8, we now focus on the trends that were fit: the separate trends in model E5 on an individual basis, the separate trends in model E5 on a national average level.

We see that the national average for β and γ in the expansion model E5 is acceptable. Acceptable means that the trend lies within the observations and does not visually contradict the fit except for the estimate of β and γ in Eastport and the estimate of β in Berkeley. Note that these estimates at the individual station level are very high compared to other stations and the national average. Certainly for the stations whose estimates are significantly higher than the national average, the station results obviously fit better.

The benefit of the national average is in the stations whose parameter estimates are not significant at the individual station level. For example, look at the following:

Caldwell The estimate of β has an individual station estimate which is not significant and slightly negative but the national average is not visually too far off and significant. Thus adding the information from the rest of the country helps in this case.

Le Mars The estimate of β which is not significant is essentially the same as the national average which is significant. The estimate of γ which was not significant and negative is not so different from the national average of γ.

Savannah The estimate of β follows the national average closely and with the additional information from the rest of the US is now significant.

In conclusion at the local level,

1. For the pockets of stations where β and γ are very different from the national average, certainly the individual station parameter estimates fit the data more closely and display a larger trend.

2. Except for the stations which are extremely different from the national average, these national averages of γ and β do a credible job with the trends.

3. The national average for β does a better job than for γ. This is because the estimates for β are more compact; that is, there is less fluctuation between the largest and the smallest of the estimates of these trends across the US and thus the national average is closer to the individual station results. Thus the national average has more meaning at the individual station level.

4. Since the estimates of γ have a greater range between the largest and the smallest estimates throughout the country, the national average for γ has less meaning at the individual station level and reveals a need to solve correctly for a more local averaging of these γs.

5.7 Comparison of Results to Other Findings

Smith (1999) found a 4.2% increase in the annual maximum precipitation across the US from 1951 to 1996. We found a 3.0% increase across the US from 1901 to 1997. Note our independent model estimated this increase in the annual maximum at 4.6% – comparable to Smith (1999). Thus we see the influence of adding information on the annual total rainfall is a decrease in the estimate of the trend in the annual maximum. Currently, there is no explanation as to why our result concerning the increase in the annual maximum rainfall is so different from Smith (1999) but recall it is this estimate which has conflicting evidence as to its significance.

Karl and Knight (1998) found a 10% increase in annual total rainfall across the US from 1901 to 1997. We found only a 7.4% increase. In both the expansion and independent cases, our models produce a smaller estimate than Karl and Knight (1998). Since the model used in Karl and Knight (1998) is completely a different model, the difference between the increase in the annual total rainfall is understandable. Recall from Chapter 1, Findlay *et al.* (1994) and Letternaier *et al.* (1994) found a 5% increase in annual precipitation. In other words, our estimate comes out in between the 5% and 10% others are finding.

As to the hypothesis of Karl and Knight (1998) that the increase in the extremes are driving the increase in the totals, we found that there is less evidence of a significant increase in annual maximum rainfall and this increase is significantly smaller than the increase in annual total rainfall. Finally, although the hypothesis of Karl and Knight (1998) suggests that the trend in the annual maximum rainfall should be driving the trend in the annual total rainfall; that is, β should be increasing at least at the same magnitude as γ, we do not see this. These two rates are different, implying the change in the annual rainfall distribution is more complex than a shift in the distribution.

For regional differences, Smith (1999) found no regional differences for the trend in the annual maxima. Karl and Knight (1998) found no significant regional differences in upper percentiles of the rainfall distribution except for the far west and Southeast. Our findings here are consistent with their results. We find no significant regional differences with respect to the percentage of significant trends in either the annual maximum or annual total rainfall series. Further from the mapped estimates, we see that the estimates of β are more uniform across the US except for the Northwest and possibly the Southeast. As to the trends in the annual total rainfall, we see more variability across the US but the different levels of the estimates do not fall into the regions as defined in this chapter, or in Smith (1999), or Karl and Knight (1998).

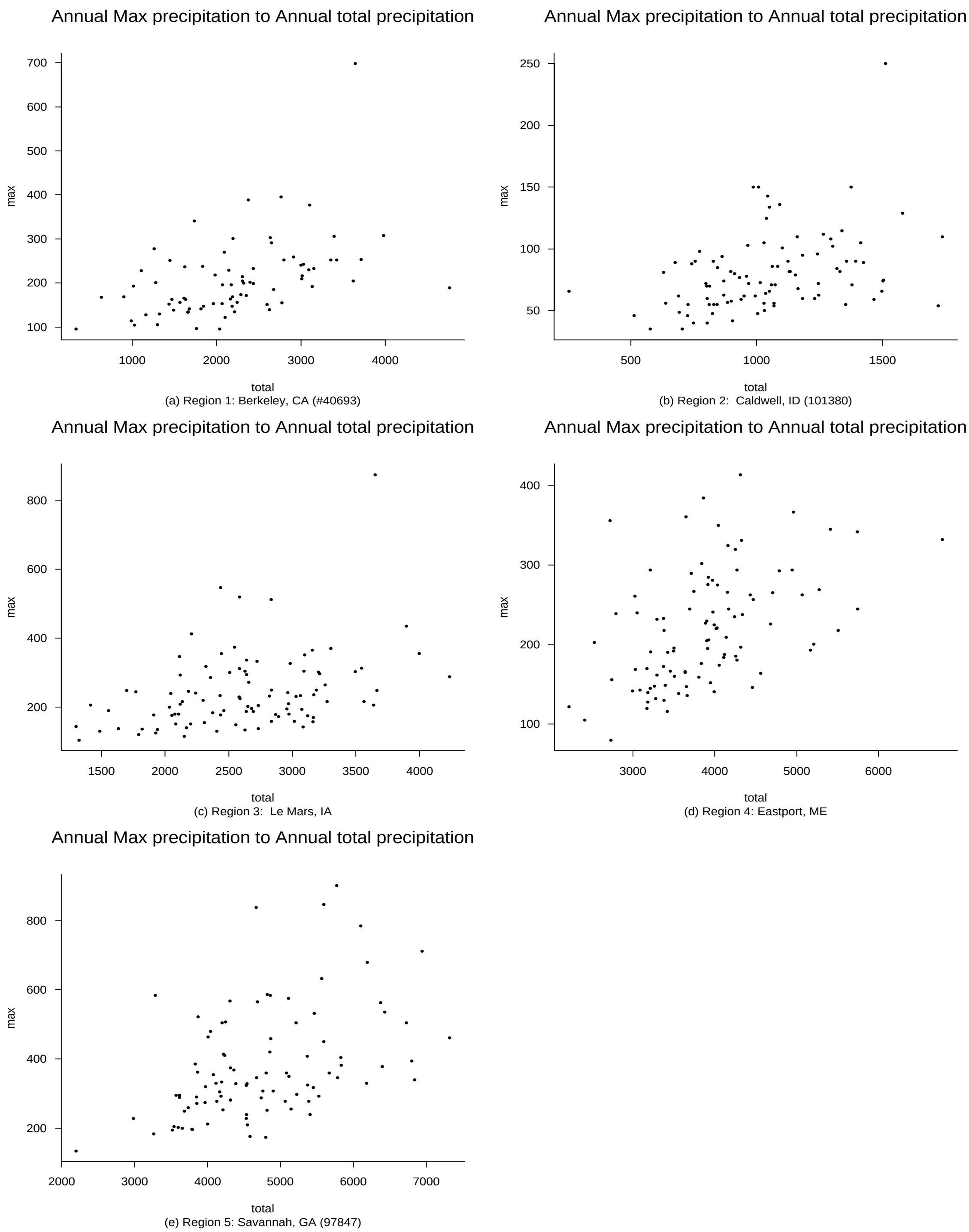
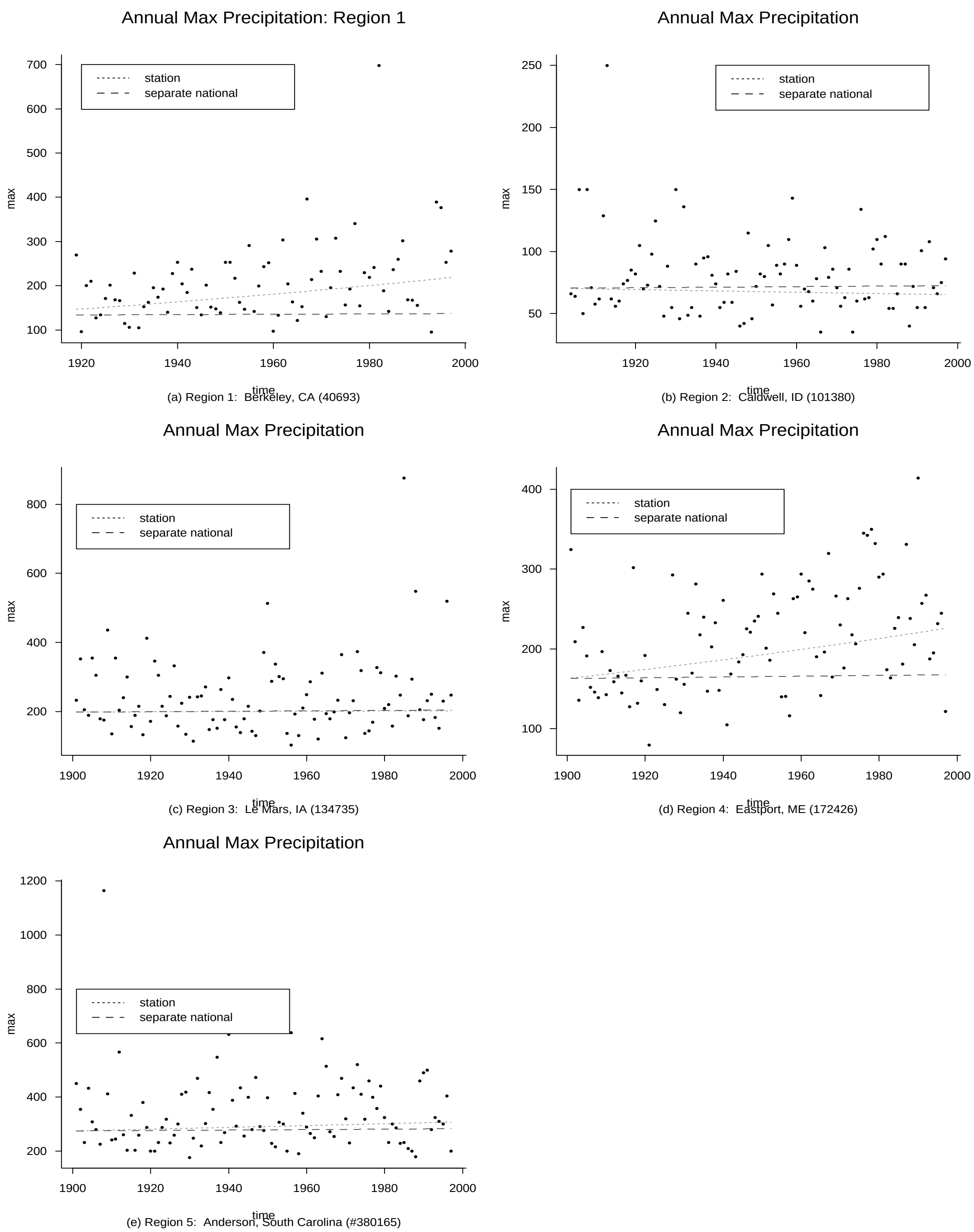


Figure 5.6: The annual total rainfall to annual maximum rainfall for stations: (a) Berkeley, CA (b) Caldwell, ID (c) Le Mars, IA (d) Eastport, ME (e) Savannah, GE. The units are in hundredth of an inch.



(a) Region 1: Berkeley, CA (40693)

(b) Region 2: Caldwell, ID (101380)

(c) Region 3: Le Mars, IA (134735)

(d) Region 4: Eastport, ME (172426)

(e) Region 5: Anderson, South Carolina (#380165)

Figure 5.7: The annual maximum rainfall for the years of record in stations: (a) Berkeley, CA (b) Caldwell, ID (c) Le Mars, IA (d) Eastport, ME (e) Savannah, GE. The units are in hundredth of an inch. The lines fit are (1) the individual station estimate of the trend from the full expansion model and (2) the national average based on the trend estimates from the full expansion model.

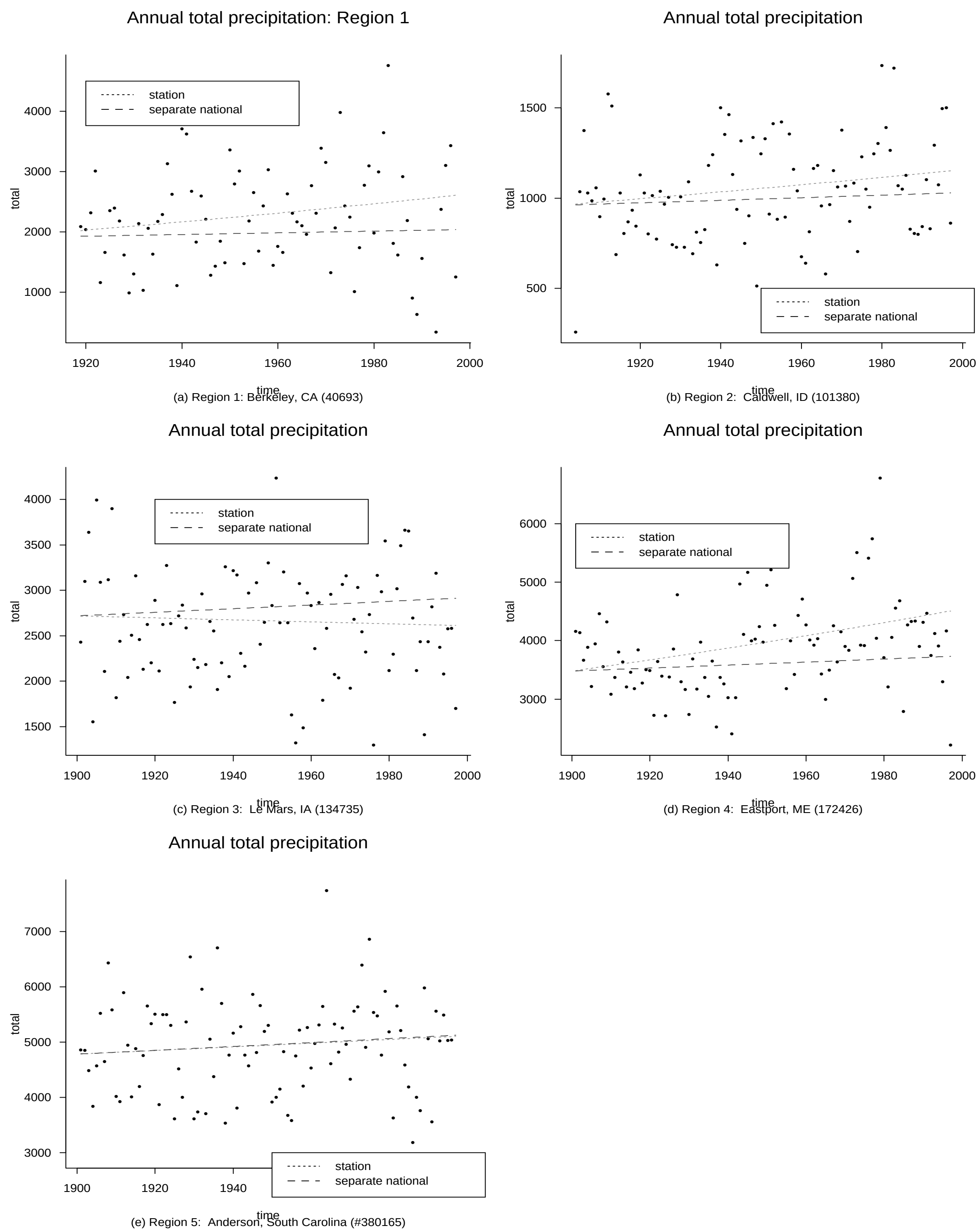


Figure 5.8: The annual total rainfall for the years of record in stations: (a) Berkeley, CA (b) Caldwell, ID (c) Le Mars, IA (d) Eastport, ME (e) Savannah, GE. The units are in hundredth of an inch. The lines fit are (1) the individual station estimate of the trend from the full expansion model and (2) the national average based on the trend estimates from the full expansion model.

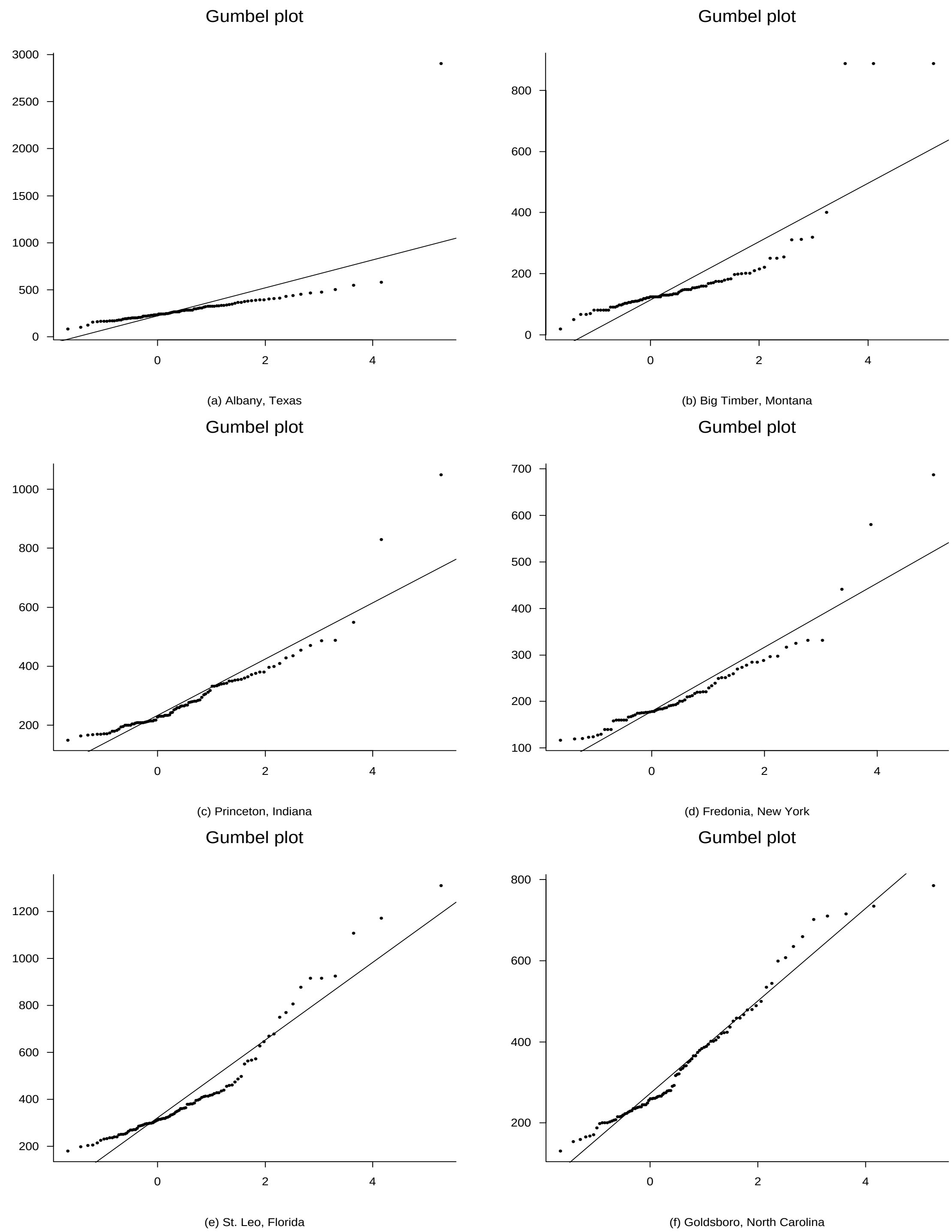


Figure 5.9: Gumbel plots for stations: (a) Albany, Texas (b) Big Timber,Montana (c) Princeton, Indiana (d) Fredonia, New York (e) St. Leo, Florida and (f) Goldsboro, North Carolina.

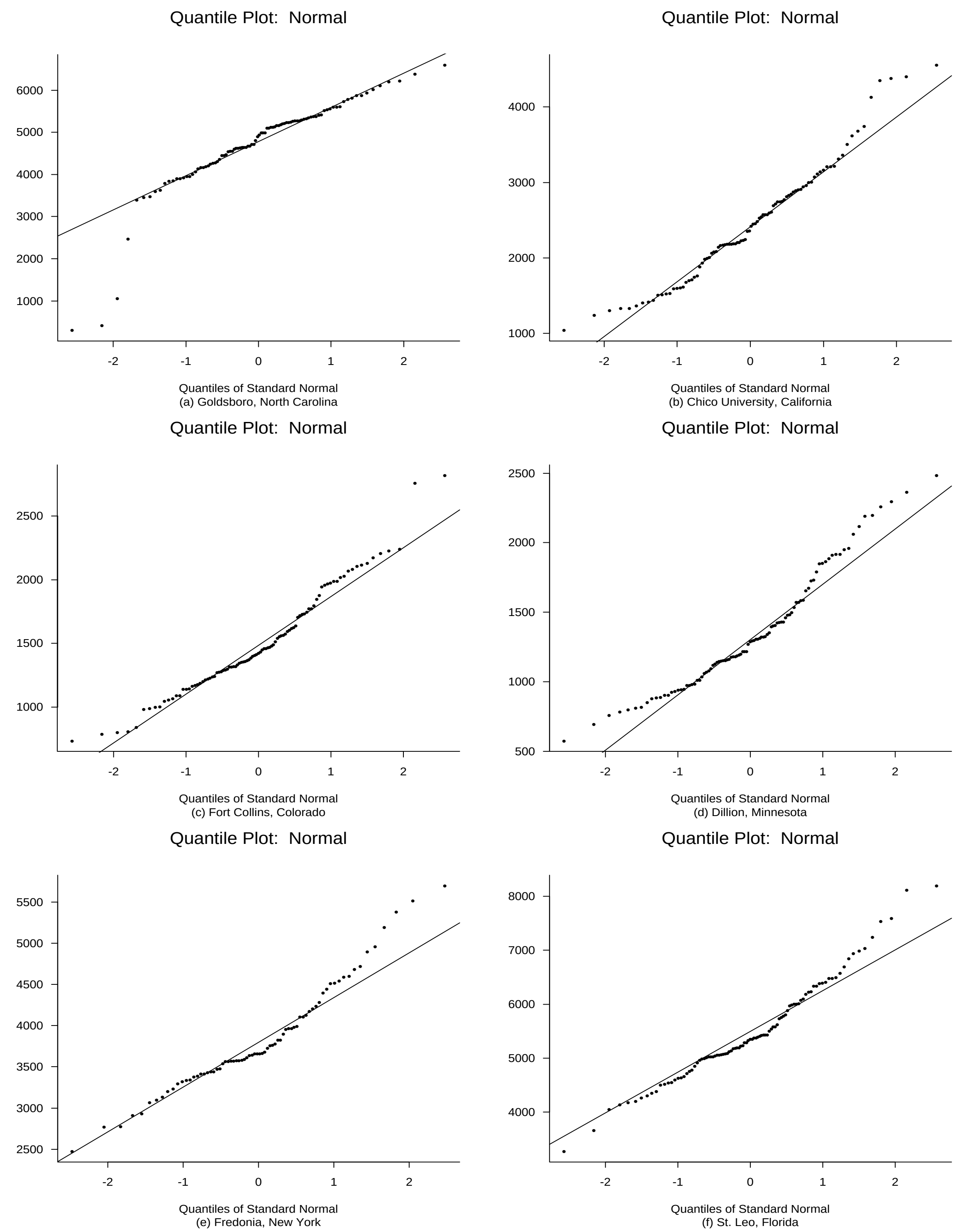


Figure 5.10: Quantile plots for stations: (a) Goldsboro, North Carolina (b) Ojai, California (c) Fort Collins, Colorado (d) Dillion, Minnesota (e) Fredonia, New York (f) St. Leo, Florida.

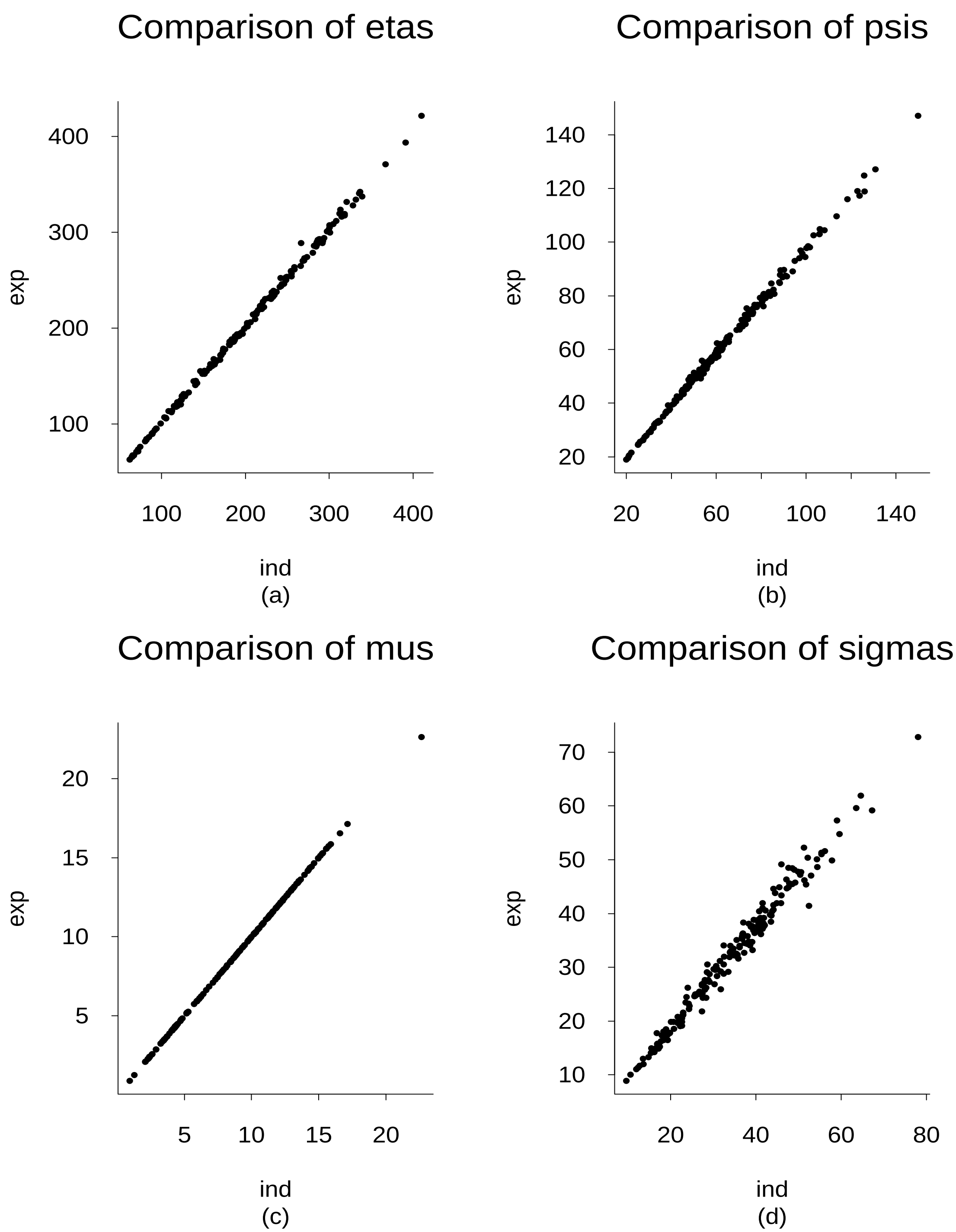
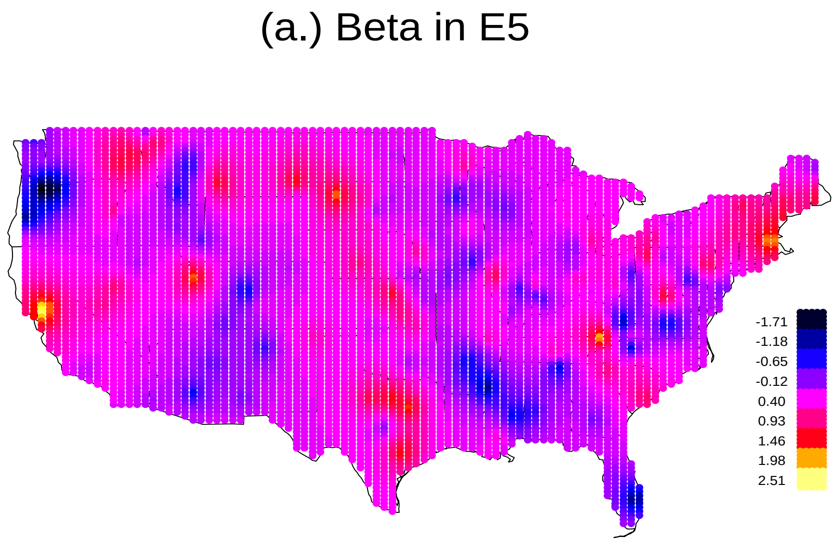


Figure 5.11: The parameter estimates in model I1 plotted against the parameter estimates in model E1: (a) η (b) ψ (c) μ (d) σ.



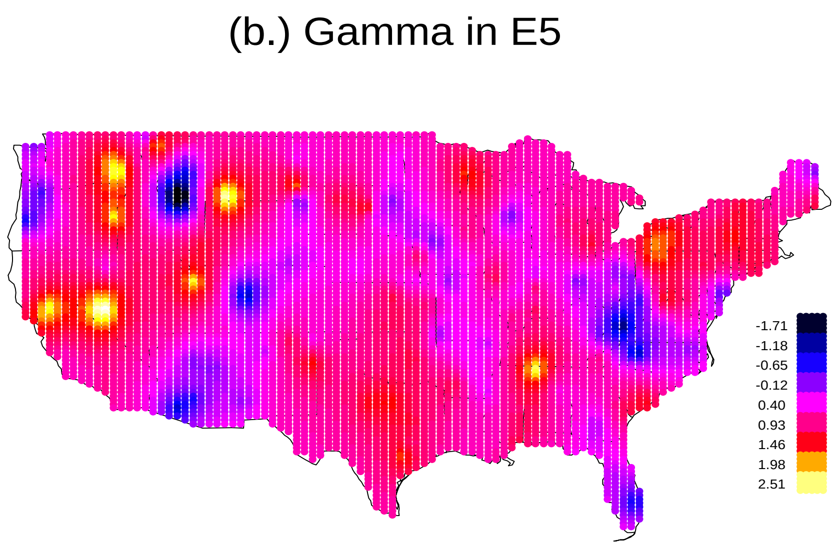


Figure 5.12: Mapped estimates for (a.) β and (b.) γ in model E5. Note color scale is consistent for both β and γ estimates. The four smallest estimates of γ were Winsorized so as to give a more informative graph.

Bibliography

Anderson, C.W. (1971). *Contributions to the Asymptotic Theory of Extreme Values.* Ph.D. Thesis, University of London.

Anderson, C.W. and Turkman, K.F. (1991). The joint limiting distribution of sums and maxima of stationary sequences. *J. Appl. Prob.* **28**, 33-44.

Arov, D.Z. and Bobrov, A.A. (1960). The extreme terms of a sample and their role in the sum of independent variables. *Theor. Probability Appl.* **5** 377-396.

Balkema, A.A. and de Haan, L. (1972). On R. von Mises' condition for the domain of attraction of $\exp(-e^{-x})$. *Ann. Math. Statist.* **43** 1352-1354.

Cao, H.-X., Mitchell, J.F.B. and Lavery, J.R. (1992). Simulated diurnal range and variability of surface temperature in a global model for present and double CO_2 climates. *Journal of Climate* **5** 920-943.

Chow, T.L. and Teugels, J.L. (1978). The sum and the maximum of i.i.d. random variables. In *Proc. Second Prague Symp. Asymptotic Statistics*, 81-92.

Cohen, J.P. (1982a). Convergence rates for the ultimate and penultimate approximations in extreme-value theory. *Adv. Appl. Probab.* **14** 833-854.

Cohen, J.P. (1982b). The penultimate form of approximation to normal extremes. *Adv. Appl. Probab.* **14** 324-339.

Cook, E.R. (1995). Temperature histories from tree rings and corals. *Clim. Dyn.* **11** 211-222.

Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics.* Chapman and Hall, London.

Darling, D.A. (1952). The influence on the maximum term in the addition of independent random variables. *Trans. Amer. Math. Soc.* **73** 95-107.

Davison, A.C. and Smith, R.L. (1990). Models for Exceedances over High Thresholds. *J. R. Statist. Soc. B* **52** 393-442.

Del Genio (1999). http://www.giss.nasa.gov/research/intro/delgenio.02/index.html.

Dodd, E.L. (1923). The greatest and least variate under general laws of error. *Trans. Amer. Math. Soc.* **25** 525-539.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications.* Vol. II, Wiley, New York.

Findlay, B.F., Gullet, D.W., Malone, L., Reycraft, J., Skinner, W.R., Vincent, L., and Whitewood, R. (1994). Canadian national and regional standardized annual precipitation departures. In *Trends '93: A Compendium of Data on Global Change*, Boden, T.A., Kaiser, DP., Sepanski, R.J., and Stoss, F.W. (eds.) ORNL/CDIAC-65, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge USA, 800-828.

Fisher, R.A. and Tippett, L.H.C. (1928). Limiting forms of the frequency distributions of the largest or smallest member of a sample. *Proc. Camb. Phil. Soc.* **24** 180-190.

Fréchet, M. (1927). Sur la loi de probabilité de l'écart maximum. *Ann. Soc. Polonaise Math.* (Cracow) **6** 93.

Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* **44** 423-453.

Goldie, C.M. and Smith, R.L. (1987). Slow variation with remainder: a survey of the theory and its applications. *Quart. J. Math. Oxford Ser. (2)* **38** 45-71.

Gomes, M.I. (1984). Penultimate limiting forms in extreme value theory. *Ann. Inst. Statist. Math.* **36** 71-85.

Gomes, M.I. and Pestana, D.D. (1986). Nonstandard domains of attraction and rates of convergence. In *New Perspectives in Theoretical and Applied Statistics*, Puri, M.L., Vilaplana, J.P., and Wertz, W. (eds.), John Wiley, 467-477.

Groisman, P. Y. and Legates, D.R. (1995). Documenting and detecting long-term precipitation trends: where we are and what should be done. *Climate Change* **31** 601-622.

Gumbel, E.J. (1958). *Statistics of Extremes.* Columbia University Press.

Haan, L. de (1970). *On Regular Variation and Its Application to the Weak Convergence of Sample Extremes.* Mathematical Centre Tracts **32**, Amsterdam.

Haan, l. de and Resnick, S.I. (1982). Local limit theorems for sample extremes. *Ann. Probab.* **10** 396-413.

Hall, P. (1979). On the rate of convergence of normal extremes. *J. Appl. Prob.* **16** 433-439.

Hall, P. (1980). Estimating probabilities for normal extremes. *Adv. Appl. Probab.* **12** 491-500.

Hall, P. (1982a). *Rates of Convergence in the Central Limit Theorem.* Pitman Publishing, Inc., London.

Hall, P. (1982b). On some simple estimates of an exponent of regular variation. *J. R. Statist. Soc. B* **44** 37-42.

Hall, W.J. and Wellner, J.A. (1979). The rate of convergence in law of the maximum of an exponential sample. *Statistica Neerlandica* **33** 151-154.

Harvey, L.D.D. (2000). *Global Warming: The Hard Science.* Prentice Hall.

Hennessey, K.J. and Pittock, A.B. (1995). Greenhouse warming and threshold temperature events in Victoria, Australia. *Int. J. Climatol.* **15** 591-612.

IPCC [Intergovernmental Panel on Climate Change] (1996): *Climate Change 1995: The Science of Climate Change.* Houghton, J.T., Meira Filho, L.G., Callander, B.A., Harris, N., Katternberg, A. and Maskell, K. (eds.), Cambridge University Press, Cambridge, UK.

Iwashima, T. and Yamamoto, R. (1993). A statistical analysis of extreme events: Long term trend of heavy daily precipitation. *J. Met. Soc. Japan* **71** 637-640.

Johnson, N.L. and Kotz, S. (1982). *Encyclopedia of Statistical Sciences.* John Wiley, New York.

Karl, T.R., Kukla, G., Razuvayev, V.N., Changery, M.J., Quayle, R.G., Heim, R.R., Easterling, D.R., and Fu, C.B. (1991). Global warming: evidence for asymmetric diurnal temperature change. *Geophs. Res. Lett.* **18** 2253-2256.

Karl, T.R., Jones, P.D. and Knight, R.W.(1995). Trends in high-frequency climate variability in the twentieth century. *Nature* **377** 217-220.

Karl, T.R., Knight, R.W., Easterling, D.R. and Quayle, R.G. (1996). Indices of climate change for the United States. *Bull. Amer. Meteor. Soc.* **77** 279-292.

Karl, T.R. and Knight, R.W. (1998). Secular trends of precipitation amount, frequency, and intensity in the United States. *Bull. Amer. Meteor. Soc.* **79** 231-241.

Kattenberg, A., Giorgi, F. Grassl, H., Meehl, G.A., Mitchell, J.F.B., Stouffer, R.J., Tokioka, T., Weaver, A.J., and Wigley, T.M.L. (1996). Climate models – projections of future climate. In Houghton, J.T., Filho, L.G.F., Callander, B.A., Harris, N., Kattenberg, A., and Maskell, K. (eds.) *Climate Change 1995: The Science of Climate Change.* Cambridge University Press, Cambridge UK, 285-357.

Katz, R.W. and Brown, B.G. (1992). Extreme events in a changing climate: variability is more important than averages. *Climatic Change* **21** 289-302.

Knuth, D.E. (1981). *Seminumerical Algorithms*, 2nd ed., vol. 2 of *The Art of Computer Programming.* Addison-Wesley, Reading, Massachusetts, §§3.2-3.3.

Lamb, H.H.(1965). The medieval warm epoch and its sequel. *Palaeogeography, Paloeoclimatology, Palaeoecology* **1** 13-37.

Lamb, H.H. (1988). Climate and life during the middle ages studied especially in the mountains of Europe. In *Weather, Climate and Human Affairs.* Lamb, H.H. (ed.), Routledge, London, 40-74.

Leadbetter, M.R. (1983). Extreme and local dependence in stationary sequences. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **65** 291-306.

Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983). *Extreme and Related Properties of Random Sequences and Processes.* Springer-Verlag, New York.

Lettenmaier D., Wood. E.F. and Wallis, J.R. (1994). Hydroclimatological trends in the continental United States, 1948-1988. *J. Climate* **7** 586-607.

Mori, T. (1981). The relation of sums and extremes of random variables. *Proc. 43rd Meeting of ISI, Buenos Aires* 879-894.

Nash, J.C. (1979). *Compact Numerical Algorithms for Computers.* Adam Hilger.

Nicholls, N., Gruza, G.V., Jouzel, J., Karl, T.R., Ogallo, L.A. and Parker, D.E. (1996). Observed climate variability and change. In Houghton, J.T., Filho, L.G.F., Callander, B.A., Harris, N., Kattenberg, A., and Maskell, K. (eds.) *Climate Change 1995: The Science of Climate Change.* Cambridge University Press, Cambridge UK, 133-192.

NOAA [National Oceanic and Atmospheric Administration]

http://www.ncdc.noaa.gov./ol/climate/research/ushcn.html

Petrov, V.V. (1975). *Sum of Independent Random Variables.* Springer-Verlag, Berlin.

Pickands, J. (1968). Moment convergence of sample extremes. *Ann. Math. Statist.* **39** 881-889.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3** 119-131.

Pickands, J. (1986). The continuous and differentiable domains of attraction of the extreme value distributions. *Ann. Probab.* **14** 996-1004.

Prescott, P. and Walden, A.T. (1980). Maximum likelihood estimation of the parameters of the generalized extreme value distribution. *Biometrika* **67** 723-724.

Prescott, P. and Walden, A.T. (1983). Maximum likelihood estimation of the parameters of the three-parameter generalized extreme value distribution from censored samples. *J. Statist. Comput. Simul.* **16** 241-250.

Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T. (1990). *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, New York.

Resnick, S.I. (1987). *Extreme Values, Regular Variation, and Point Processes.* Springer-Verlag, New York.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. USA* **42** 43-47.

Smith, R.L. (1982). Uniform rates of convergence in extreme value theory. *Adv. Appl. Probab.* **14** 600-622.

Smith, R.L. (1985). Maximum likelihood estimation in a class of non-regular cases. *Biometrika* **72** 67-90.

Smith, R.L. (1987). Estimating tails of probability distributions. *Ann. Statist.* **15** 1174-1207.

Smith, R.L. (1990). Extreme value theory. In *Handbook of Applicable Mathematics* (ed. W. Ledermann), vol. 7 Chichester, John Wiley.

Smith, R.L. (1994) Spatial modelling of rainfall data. In *Statistics for the Environment 2: Water Related Issues*, (ed. V. Barnett and F. Turkmen, editors), Chichester, John Wiley, 19-42.

Smith, R.L. (1999). Trends in Rainfall Extremes. Preprint.

http://www.stat.unc.edu/faculty/rs/papers/RLS_Papers.html

Smith, R.L. (2000). Environmental Statistics. Course notes.

http://www.stat.unc.edu/postscript/rs/envstat/env/html

Stern, R.D. and Coe, R. (1984). A modeling fitting analysis of daily rainfall data (with discussion). *J. R. Statist. Soc. A* **147** 1-34.

Sweeting, T.J. (1985). On domains of uniform local attraction in extreme value theory. *Ann. Probab.* **13** 196-205.

Trenberth, K.E., Houghton, J.T., and Meira Filho, L.G.F. (1996). The climate system: on overview. In Houghton, J.T., Filho, L.G.F., Callander, B.A., Harris, N., Kattenberg, A., and Maskell, K. (eds.) *Climate Change 1995: The Science of Climate Change.* Cambridge University Press, Cambridge UK, 51-64.

Vangel, M.G. (1999). Lot acceptance using the sample mean and an extremum. Preprint. Duke University.

Venables, W.N. and Ripley, B.D. (1994). *Modern Applied Statistics with S-Plus.* Springer-Verlag, New York.

Von Mises, R. (1936). La distribution de la plus grande de n valeurs, reprinted in *Selected Papers II*, Amer Math. Soc., Providence, R.I. (1954), 271-294.

Weibull, W. (1939). A statistical theory of strength of materials. *Proc. Ing. Vetenskapsakad* **151**.

Weibull, W. (1951). A statistical distribution of wide applicability. *J. Appl. Mechanics* **18** 293-297.

Wigley, T.M.L. (1999). The science of climate change: global and U.S. perspectives. Prepared for the Pew Center on Global Climate Change, National Center of Atmospheric Research, Boulder, Colorado.

Wigley, T.M.L., Jones, P.D. and Raper, S.C.B. (1997). The observed global warming record: what does it tell us? *Proc. Natl. Acad. Sci. USA* **94** 8314-8320.