Springer Series in
Operations Research
and Financial Engineering

Sidney I. Resnick

Extreme Values, Regular Variation, and Point Processes

极值正则变差和点过程

Springer

世界图书出版公司
www.wpcbj.com.cn

ISBN 978-7-5100-3746-7

9 787510 037467 >

上架建议：数学/影印版

ISBN 978-7-5100-3746-7/O·886

定价：39.00元

数学
经典教材
（影印版）

极值正则变差和点过程

Extreme Values, Regular Variation, and Point Processes

世界图书出版公司

Sidney I. Resnick

Extreme Values, Regular Variation, and Point Processes

Springer-Verlag
New York Berlin Heidelberg
London Paris Tokyo

Sidney I. Resnick
Operations Research and
Industrial Engineering
Cornell University
Upson Hall
Ithaca, NY 14853-7501
USA

Series Editors

J. Gani
Statistics Program
Department of Mathematics
University of California
Santa Barbara, CA 93106
USA

C.C. Heyde
Department of Statistics
Institute of Advanced Studies
The Australian National University
GPO Box 4, Canberra ACT 2601
Australia

AMS Classification: 60G55

Library of Congress Cataloging-in-Publication Data
Resnick, Sid.
Extreme values, regular variation, and point processes.
(Applied probability)
Bibliography: p.
Includes index.
1. Point processes. 2. Extreme value theory.
3. Distribution (Probability theory) I. Title.
II. Series.
QA274.42.R47 1987 519.2'32 87-318

ISBN 0-387-96481-9 Springer-Verlag New York Berlin Heidelberg
ISBN 3-540-96481-9 Springer-Verlag Berlin Heidelberg New York

Preface

Extreme value theory is an elegant and mathematically fascinating theory as well as a subject which pervades an enormous variety of applications. Consider the following circumstances:

Air pollution monitoring stations are located at various sites about a city. Government regulations mandate that pollution concentrations measured at each site be below certain specified levels.

A skyscraper is to be built near Lake Michigan and thus will be subject to wind stresses from several directions. Design strength must be sufficient to withstand these winds. Similarly, a mechanical component such as an airplane wing must be designed to withstand stresses from several sources.

Dams or dikes at locations along a body of water such as a river or sea must be built high enough to exceed the maximum water height.

A mining company drills core samples at points of a grid in a given region. Continued drilling will take place in the direction of maximum ore concentration.

Athletic records are frequently broken.

A common feature of these situations is that observational data has been or can be collected and the features of the observations of most interest depend on largest or smallest values; i.e., on the extremes. The data must be modeled and decisions made on the basis of how one believes the extreme values will behave.

This book is primarily concerned with the behavior of extreme values of independent, identically distributed (iid) observations. Within the iid framework there are surprising depth, beauty, and applicability. The treatment in this book is organized around two themes. The first is that the central analytic tool of extreme value theory is the theory of regularly varying functions, and the second is that the central probabilistic tool is point process theory and in particular the Poisson process. Accordingly we have presented a careful exposition of those aspects of regular variation and point processes which are essential for a proper understanding of extreme value theory.

Chapter 0 contains some mathematical preliminaries. Some authors might relegate these to appendices, but I believe these should be read first, before plunging into the following sections, in order that readers can get used to my way of doing things. Chapter 0 also contains a derivation of the three families of classical, Gnedenko limit distributions for extremes of iid variables and an account of regular variation and its extensions.

Chapter 1 discusses thoroughly questions of domains of attraction. If iid random variables have common distribution F, what criteria on F or its density F' guarantee that suitably scaled and centered extremes have limit distributions as the sample size gets large? What are suitable scaling and centering constants? These results provide a theoretical underpinning to statistical practice as discussed, for example, in Gumbel (1958): Suppose data are obtained such that each observation is an extreme. For example, our data may consist of maximal yearly flow rates at a particular site on the Colorado River over the last 50 years. Suppose by stretching the imagination, one believes the data to be modeled adequately by the iid assumption. The underlying distribution of the model is unknown, so the parametric assumption is made that the data comes from a limiting extreme value distribution. Usually it is the Gumbel, also called *double exponential*, distribution $\Lambda(x) = \exp\{-e^{-x}\}$ that is chosen, and the estimation problem reduces to choosing location and scale constants; this is sometimes done graphically using loglog paper. The underlying distribution may not be extreme value, but we robustly hope it is at least in a domain of attraction, so that the distribution of extremes is close to a limiting extreme value distribution.

By the end of Chapter 1 much analytic technique has been developed and this is exercised in the specialist Chapter 2. If normalized extremes of iid random variables have a limit distribution, when do moments and densities of normalized extremes have limits? We also discuss rates of convergence to the limit extreme value distributions and large deviation questions which emphasize sensitivity to the quality of the approximation of the right tail of the distribution of the maximum of n iid random variables by the tail of the limit extreme value distribution.

Chapter 3 shifts the focus from the analytic to the probabilistic, and is a thorough discussion of those aspects of point processes (and in particular the Poisson process) which are essential, in my view, for a proper understanding of the structural behavior of extremes. The core of the probabilistic results is in Chapter 4, which views records and maxima of iid random variables as stochastic processes. In a sequence of iid observations from a continuous distribution, the records (i.e., those observations bigger than all previous ones) form a Poisson process, and the indices when records occur are approximately a Poisson process. Several extensions to these ideas are discussed.

Also in Chapter 4 is an account of extremal processes. If maxima of n iid random variables are viewed as a stochastic process indexed by n, there is a continuous parameter process called an *extremal process*, which is a useful approximation. The structural properties of such processes are studied and the uses for weak convergence problems are detailed. We also give an account in Sections 3.5, 4.4, and 4.5 of a weak convergence technique called the *point process method*, which has proved invaluable for weak convergence problems involving heavy tailed phenomena. If it is necessary to prove some functional (say the maximum) of n heavy tailed random variables converges weakly as $n \to \infty$, it is often simplest to first prove a point process based on the n variables converges ($n \to \infty$) and then to get the desired result by continuous mapping theorems. The power of this technique is illustrated in Section 4.5, where it is applied to extremes of moving averages.

The last chapter examines some multivariate extreme value problems. In one dimension notions such as maximum and record have unambiguous meanings. In higher dimensions this is no longer the case. The maximum of n multivariate observations could be the convex hull, or it could be the vector of componentwise maxima depending on the application. We concentrate on the latter definition, which seems most natural for the applications mentioned at the beginning of the introduction. We discuss characterizations of the limiting multivariate extreme value distributions and give domains of attraction criteria. A theory of multivariate regular variation is needed, and this is developed. Criteria for asymptotic independence are given, and it is proved that a concept of positive dependence called *association* applies to limiting extreme value distributions.

Notation will ideally seem clear and simple. One quirk that needs to be mentioned is that if a distribution F has a density, it is denoted by F' (even in the multivariate case) and never by f. The symbol f is reserved for the auxiliary function of a class Γ monotone function. Sometimes, in the completely separate context of point processes, f denotes a bounded, continuous real function, but the important point to remember is that f is not the density of F; rather F' is the density of F.

Extreme value results are always phrased for maxima. One can convert results about maxima to apply to minima by using the rule

$$-\max - = \min.$$

For example, $2 = \min\{2, 3\} = -\max\{-2, -3\} = -(-2) = 2$. We denote $\max\{x_i: 1 \le i \le n\}$ by $\bigvee_{i=1}^{n} x_i$ and similarly min is denoted by $\bigwedge$. Also it is usually clear how to adapt weak convergence results for maxima so that they apply to the kth largest of a sample of size n (k fixed, $n \to \infty$). The point process method usually makes this adaptation transparent. See, for instance, Section 4.5.

The best plan for reading this book is to start from the beginning and read each page lovingly until the end. There is only one section that is tedious. The second best plan is to start from the beginning and go through, passing lightly over certain material depending on background, taste, and interests but slowing down for the important results. Chapter 2, parts of 3, 4.4.1, and part of 4.6 may be skimmed, but the motto to be kept in mind is "skim; don't skip." This includes the exercises, which contain complementary material and alternative approaches. The extent to which readers will actively attempt the exercises will determine the extent of their progress from observer to practitioner. If plans 1 and 2 seem too ambitious, readers could consider making a module of Chapters 0, 1, and 5 and another module of a skimmed Chapter 3 and a heavily studied Chapter 4.

There are a number of things this book is not. It is not an encyclopedia and it is not a history book. Using a literary analogy, think of this as a novel. There is a story to be told, and readers should pay attention to matters of style and exposition and to how cosmic themes and characters relate. This book provides excellent coverage of problems arising from iid observations and offers good grounding in the subject, but does not pretend to offer comprehensive coverage of the whole subject of extreme values. This is now so broad and vast that it is doubtful that one book would do it justice. Consequently, a reader needing a rounded view of the whole subject is encouraged to consult other books and sources, as well as this one. For instance, with the exception of Section 4.5 on extremes of moving averages, I do not give attention to the important case of extremes of dependent variables. Fortunately, there is already a superb book on this subject by M.R. Leadbetter, G. Lindgren, and H. Rootzen, entitled *Extremes and Related Properties of Random Sequences and Processes.* It is very well written and elegant and is highly recommended.

Chapters 0, 1, and 2 bear the intellectual influence of my colleague and friend Laurens de Haan, with whom I have had the privilege and pleasure of working and learning since 1972. Professor de Haan has had enormous influence on the subject, and his 1970 monograph remains, despite the huge quantity of research it stimulated, an excellent place to learn about the relationship of extreme value theory and regular variation.

Now the acknowledgments. It is customary at this point for authors to make a maudlin statement thanking their families for all the sacrifices which made the completion of the book possible. This may be rather out of tune in these pseudo-quasi-semiliberated eighties. I will merely thank Minna, Nathan, and Rachel Resnick for a cheery, happy family life. Minna and Rachel bought me the mechanical pencil that made this project possible, and Rachel generously shared her erasers with me as well as providing a back-up mechanical pencil from her stockpile when the original died after 400 manuscript pages. I appreciate the fact that Nathan was only moderately aggressive about attacking my Springer-Verlag correspondence with a hole puncher.

Robyn Kelley (with timely assistance from Waydene Casey and Pat Key) provided excellent typing worthy of a raise, and cheerfully coped with substandard penmanship, transatlantic mailings, the differences between $8\frac{1}{2} \times 11$ inch and A4 paper, European hole punchers that punch in the wrong places for U.S. binders, and a professor who though on leave of absence kept sending her work.

The U.S. National Science Foundation has been magnificent in its continuing support. As a young pup I was allowed to schnor from other people's grants, and then at a crucial stage in my career NSF made me co-principal investigator of a series of grants that continue to the present. Other institutions which have generously provided support which has generated time away from teaching for thinking and learning are Erasmus University, Rotterdam; University of Amsterdam; Mathematics Center, Holland; Lady Davis Fellowship Trust; Technion, Haifa, Israel; United Kindgom Science and Engineering Research Council (SERC). The first two-thirds of the book were written while I was on sabbatical support—gratefully acknowledged—from Colorado State University. Duane Boes has been an enlightened and supportive chairman at CSU. Sussex University, Brighton, United Kingdom, provided a hospitable and pleasant environment, where the last third was finished while I was being supported by the SERC Fellowship.

As they have been so often in the past, Joe Gani and Chris Heyde were very helpful and encouraging during the preparation of this book.

Fort Collins, Colorado
April 1987

Sidney I. Resnick

Contents

Chapter 0

Preliminaries

Some of the topics discussed here are sometimes relegated to appendices. However, since these topics must be well understood before arriving at the core of the subject, it seems sensible to discuss the preliminary topics first. The reader is advised to skim 0.1, 0.2 according to taste and background, but slow down for 0.3, which discusses the possible limiting distributions for normalized maxima of independent, identically distributed (iid) random variables. Section 0.4 treats the basic facts in the theory of regular variation and some important extensions. Regular variation is the basic analytical theory underpinning extreme value theory, and its importance cannot be overemphasized.

0.1. Uniform Convergence

If f_n, $n \geq 0$ are real valued functions on $\mathbb{R}$ (or any metric space) then f_n converges uniformly on $A \subset \mathbb{R}$ to f_0 if

$$\sup_{x \in A} |f_n(x) - f_0(x)| \to 0$$

as $n \to \infty$.

If U_n, $n \geq 0$ are nondecreasing real valued functions on $\mathbb{R}$ then it is a well known and useful fact that if U_0 is continuous and $U_n(x) \to U_0(x)$ as $n \to \infty$ for all $x \in \mathbb{R}$ then $U_n \to U$ locally uniformly; i.e., for any $a < b$

$$\sup_{x \in [a,b]} |U_n(x) - U_0(x)| \to 0. \tag{0.1}$$

One proof of this fact is outlined as follows: If U_0 is continuous on $[a, b]$, then it is uniformly continuous. For any x there is an interval-neighborhood in $[a, b]$, 0_x containing x, on which U oscillates by less than ε. This gives an open cover of $[a, b]$. Compactness allows us to prune $\{0_x, x \in [a, b]\}$ to obtain a finite subcover. Using this finite collection and the monotonicity of the functions leads straightaway to the desired uniform convergence.

Another proof of (0.1) is obtained by using the concept of continuous convergence (Kuratowski, 1966). Suppose $\mathscr{X}$, $\mathscr{Y}$ are two complete, separable metric spaces and $f_n: \mathscr{X} \to \mathscr{Y}$, $n \geq 0$. Then f_n converges to f continuously if whenever $x_n \in \mathscr{X}$, $n \geq 0$ and $x_n \to x_0$ we have $f_n(x_n) \to f_0(x_0)$.

The connection with uniform convergence is this: If $\mathscr{X}$ is compact and f_0 is continuous then $f_n \to f_0$ continuously iff $f_n \to f_0$ uniformly on $\mathscr{X}$. This equivalence sometimes provides a convenient way of proving uniform convergence since it allows us to prove convergence of a sequence of points in $\mathscr{Y}$ rather than having to deal with functions.

The equivalence of the two concepts is seen readily: Let d be the metric on $\mathscr{Y}$. If $f_n \to f_0$ uniformly on $\mathscr{X}$ and $x_n \to x_0$ then we have

$$\begin{aligned} d(f_n(x_n), f_0(x_0)) &\leq d(f_n(x_n), f_0(x_n)) + d(f_0(x_n), f_0(x_0)) \\ &\leq \sup_{x \in \mathscr{X}} d(f_n(x), f_0(x)) + d(f_0(x_n), f_0(x_0)). \end{aligned}$$

The first term goes to zero as $n \to \infty$ by uniform convergence and the second term vanishes by continuity. Conversely suppose $f_n \to f_0$ continuously but not uniformly. Then there is a subsequence $\{n(k')\}$ and $\varepsilon > 0$ such that for all $n(k')$

$$\sup_{x \in \mathscr{X}} d(f_{n(k')}(x), f_0(x)) > 2\varepsilon.$$

Using the definition of sup we find points $\{x_{k'}\} \subset \mathscr{X}$ such that

$$d(f_{n(k')}(x_{k'}), f_0(x_{k'})) > \varepsilon. \tag{0.2}$$

Since $\mathscr{X}$ is assumed compact there is a limit point x_0 and a subsequence $\{x_k\} \subset \{x_{k'}\}$ with $x_k \to x_0$. Continuous convergence and continuity of f_0 require

$$d(f_{n(k)}(x_k), f_0(x_k)) \leq d(f_{n(k)}(x_k), f_0(x_0)) + d(f_0(x_0), f_0(x_k)) \to 0$$

in violation of (0.2). The contradiction occurs because we supposed f_n did not converge to f_0 uniformly.

We now check (0.1) by using continuous convergence: Suppose $\{x_n, n \geq 0\} \subset [a, b]$ and $x_n \to x_0$. We check $U_n(x_n) \to U_0(x_0)$. It suffices to consider two cases: (a) $x_n > x_0$. (b) $x_n < x_0$. (If necessary, partition $\{x_n\}$ into two subsequences.) We consider only (a). The following are evident: There exists $\eta > 0$ such that

$$|U_0(x_0 + \eta) - U(x_0)| < \varepsilon \tag{0.3}$$

because U is continuous. Furthermore there is n_0 such that if $n \geq n_0$

$$|x_n - x_0| < \eta \tag{0.4}$$

and

$$|U_n(x_0 + \eta) - U_0(x_0 + \eta)| \vee |U_n(x_0) - U_0(x_0)| < \varepsilon \tag{0.5}$$

since $U_n \to U_0$ pointwise. We then have for $n \geq n_0$ on the one hand

$$\begin{aligned} U_n(x_n) &\leq U_n(x_0 + \eta) && \text{(from (0.4))} \\ &\leq U_0(x_0 + \eta) + \varepsilon && \text{(from (0.5))} \\ &\leq U_0(x_0) + 2\varepsilon && \text{(from (0.3))} \end{aligned}$$

and on the other hand

$$\begin{aligned} U_n(x_n) &\geq U_n(x_0) && \text{(monotonicity)} \\ &\geq U_0(x_0) - \varepsilon && \text{(from (0.5)).} \end{aligned}$$

Continuous convergence follows.

If F_n, $n \geq 0$ are distribution functions on $\mathbb{R}$ (always understood to be nondefective) then $F_n \to F_0$ pointwise and F_0 continuous imply uniform convergence on $\mathbb{R}$. Local uniform convergence comes from (0.1), and off a large interval $[a, b]$ there is not much possibility of oscillation. Given ε pick b such that $F_0(b) > 1 - \varepsilon$ and there exists n_0 such that for $n \geq n_0$

$$|F_n(b) - F_0(b)| < \varepsilon$$

whence for $x \geq b$

$$|F_n(x) - F_n(b)| \leq 1 - F_n(b) \leq 1 - F_0(b) + |F_0(b) - F_n(b)| \leq 2\varepsilon$$

and therefore for $n \geq n_0$

$$\begin{aligned} \sup_{x>b} |F_n(x) - F_0(x)| &\leq \sup_{x>b} |F_n(x) - F_n(b)| \\ &\quad + |F_n(b) - F_0(b)| + |F_0(b) - F_0(x)| \\ &\leq 2\varepsilon + \varepsilon + \varepsilon. \end{aligned}$$

Similarly for $x < a$. Combined with uniform convergence on $[a, b]$ this gives convergence uniformly on $\mathbb{R}$.

Alternatively since $F_n(\infty) = 1$, $F_n(-\infty) = 0$ for all $n \geq 0$ we may compactify $\mathbb{R}$ and work on $[-\infty, \infty]$. If $F_n \to F_0$ pointwise on $[-\infty, \infty]$ and F_0 is continuous, local uniform convergence coincides with uniform convergence.

EXERCISES

0.1.1. Suppose $U_n^{(i)}$, $n \geq 0$ are real valued functions on $\mathbb{R}$ and as $n \to \infty$

$$U_n^{(i)} \to U_n^{(i)}$$

locally uniformly on $\mathbb{R}$ for $i = 1, 2$. Prove

(a) $U_n^{(1)} + U_n^{(2)} \to U_0^{(1)} + U_0^{(2)}$
locally uniformly.

(b) $U_n^{(1)} \cdot U_n^{(2)} \to U_0^{(1)} \cdot U_0^{(2)}$
locally uniformly.

(c) If $g: R \to R$ is bounded and continuous then $g(U_n^{(1)}) \to g(U_n^{(1)})$ locally uniformly.

Use continuous convergence.

0.2. Inverses of Monotone Functions

Suppose H is a nondecreasing function on $\mathbb{R}$. With the convention that the infimum of an empty set is $+\infty$ we define the (left continuous) inverse of H as

$$H^{\leftarrow}(y) = \inf\{s: H(s) \geq y\}.$$

To check $H^{\leftarrow}$ is left continuous at $x \in \mathbb{R}$, suppose $x_n \uparrow x$ but $H^{\leftarrow}(x_n) \uparrow H^{\leftarrow}(x-) < H^{\leftarrow}(x)$. Then there exist $\delta > 0$ and y such that for all n

$$H^{\leftarrow}(x_n) < y < H^{\leftarrow}(x) - \delta.$$

The left inequality and the definition of $H^{\leftarrow}$ yield $H(y) \geq x_n$ for all n, and hence letting $n \to \infty$ we get $H(y) \geq x$ whence again by the definition of $H^{\leftarrow}$ we get $y \geq H^{\leftarrow}(x)$, which coupled with $y < H^{\leftarrow}(x) - \delta$ leads to the desired contradiction.

In case the function H is right continuous we have the following interesting properties:

$$A(y) := \{s: H(s) \geq y\} \qquad \text{is closed} \tag{0.6a}$$

$$H(H^{\leftarrow}(y)) \geq y \tag{0.6b}$$

$$\begin{cases} H^{\leftarrow}(y) \leq t \text{ iff } y \leq H(t) \\ t < H^{\leftarrow}(y) \text{ iff } y > H(t). \end{cases} \tag{0.6c}$$

For (0.6a) observe if $s_n \in A(y)$ and $s_n \downarrow s$ then $y \leq H(s_n) \downarrow H(s)$ so $H(s) \geq y$ and $s \in A(y)$. If $s_n \uparrow s$ and $s_n \in A(y)$ then $y \leq H(s_n) \uparrow H(s-) \leq H(s)$ and $H(s) \geq y$ so $s \in A(y)$ again and $A(y)$ is closed. Since $A(y)$ is closed, $\inf A(y) \in A(y)$, i.e., $H^{\leftarrow}(y) \in A(y)$, which means $H(H^{\leftarrow}(y)) \geq y$. Last, (0.6c) follows from the definition of $H^{\leftarrow}$.

The probability integral transform follows: Let $([0, 1], \mathscr{B}[0, 1], m)$ be the Lebesgue probability space; m is Lebesgue measure. Suppose U is the identity function on $[0, 1]$: i.e., U is a uniformly distributed random variable. If F is a distribution function (df) then $F^{\leftarrow}(U)$ is a random variable on $[0, 1]$ with df F. This is readily checked: For $t \in \mathbb{R}$

$$\begin{aligned} m[F^{\leftarrow}(U) \leq t] &= m[U \leq F(t)] \qquad \text{(from (0.6c))} \\ &= F(t). \end{aligned}$$

A slight variant of this involves an exponential distribution rather than the uniform: Let X be a real random variable with distribution F. Set $R = -\log(1 - F)$. If $P[E > x] = e^{-x}$, $x > 0$ then $R^{\leftarrow}(E)$ and X have the same distribution which we write as

$$R^{\leftarrow}(E) \stackrel{\mathrm{d}}{=} X.$$

To check this is simple: For $x \in \mathbb{R}$

$$\begin{aligned} P[R^{\leftarrow}(E) > x] &= P[E > R(x)] \\ &= \exp\{-R(x)\} = 1 - F(x). \end{aligned}$$

We now discuss convergence of monotone functions. For any function H denote

$$\mathscr{C}(H) = \{x \in \mathbb{R}: H \text{ is finite and continuous at } x\}.$$

A sequence $\{H_n, n \geq 0\}$ of nondecreasing functions on $\mathbb{R}$ converges weakly to