mathematics

MDPI

Article

A Discrete GEV Distribution

Adamu Abubakar Umar [1,2] and Saralees Nadarajah [1,*]

[1] Department of Mathematics, University of Manchester, Manchester M13 9PL, UK; adamu.umar@postgrad.manchester.ac.uk
[2] Department of Statistics, Ahmadu Bello University, Zaria 810107, Nigeria
* Correspondence: mbbsssn2@manchester.ac.uk

Abstract

Annual maximum data are frequently recorded as integers, yet they are routinely modeled with the continuous generalized extreme value (GEV) distribution. We study the discrete GEV distribution of Dkengne et al. (2016), which is the distribution of the integer part of a continuous GEV random variable, and we provide a systematic treatment of its distributional, inferential and finite-sample properties. The model retains the location, scale and shape parameters of its continuous counterpart and contains the discrete Gumbel, discrete Fréchet and discrete (reversed) Weibull distributions as particular cases. We give the support (including the endpoint cases), the probability mass and cumulative distribution functions, their tail behavior, moments, the quantile function, and quantile-based measures of skewness and kurtosis. For maximum likelihood estimation, we identify the subsets of the parameter space on which the support is locally fixed, show that the model is differentiable in quadratic mean on a larger set that also contains integer endpoints when $\beta > -\frac{1}{2}$, and state the resulting asymptotic normality. We also show that flooring, ceiling and rounding of a GEV variable all lead to the same discrete family up to a shift in location. Simulation studies examine the finite-sample performance of the estimators and compare the discrete GEV with the continuous GEV fitted to the integer values under several observation mechanisms, including block maxima of integer counts. Applications to annual maximum wind speeds at Lisbon and annual maximum temperatures at Oxford are analyzed with discrete versions of the Kolmogorov–Smirnov, Cramér–von Mises and Anderson–Darling statistics calibrated by a parametric bootstrap.

Keywords: annual maximum data; generalized extreme value distribution; maximum likelihood estimates; discrete extreme value distributions; goodness of fit for discrete data

MSC: 62E99

Check for updates

Academic Editor: Yuri A. Iriarte

Received: 6 September 2026
Revised: 3 November 2026
Accepted: 6 November 2026
Published: 9 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

The Generalized Extreme Value (GEV) distribution is a family of continuous probability distributions that arises in the modeling of extreme values, such as the maximum or minimum of a large set of independent random variables. It is grounded in extreme value theory, a branch of statistics that focuses on understanding the behavior of the tails of distributions. The GEV distribution unifies three distinct types of extreme value distributions—the Gumbel, Fréchet and (reversed) Weibull—into a single parametric framework. Its flexibility makes it a popular choice in fields such as hydrology, finance, environmental science, and engineering, where the prediction of rare events, such as floods, market crashes, or structural failures, is critical.

The GEV distribution can be specified by the cumulative distribution function (Fisher and Tippett [1]; Gnedenko [2]; Jenkinson [3])

$$F(x) = \exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right] \tag{1}$$

for $\mu - \frac{\sigma}{\beta} \leq x < \infty$ if $\beta > 0$, $-\infty < x < \infty$ if $\beta = 0$ and $-\infty < x \leq \mu - \frac{\sigma}{\beta}$ if $\beta < 0$, where $-\infty < \mu < \infty$ denotes a location parameter, $\sigma > 0$ denotes a scale parameter and $-\infty < \beta < \infty$ denotes a shape parameter. Note that if X has distribution function (1) and $\beta > 0$, then X has a heavy tail bounded below by $\mu - \frac{\sigma}{\beta}$. If $\beta < 0$, then X has a short tail bounded above by $\mu - \frac{\sigma}{\beta}$.

The GEV distribution has been extensively used for modeling block maxima. Recent applications include analysis of a stationary climatic covariate in a Mediterranean city (Cherif [4]), assessing the risk of extreme precipitation in Japan (Jiao et al. [5]), estimating extreme design rainfall in Tasmania (Hossain et al. [6]), extreme value capture in day-ahead photovoltaic scenario generation (Zhang et al. [7]) and SAR image change detection (Zhang et al. [8]). Estimation typically proceeds by maximum likelihood, the method of moments or the method of probability-weighted moments (Hosking et al. [9]), and the model's properties have been studied extensively—see, for example, Leadbetter et al. [10], Galambos [11], Resnick [12], Embrechts et al. [13], Kotz and Nadarajah [14], Coles [15], Beirlant et al. [16], Gumbel [17], Castillo et al. [18], Malevergne and Sornette [19], de Haan and Ferreira [20], Reiss and Thomas [21], Falk et al. [22], Novak [23], Ahsanullah [24] and references therein for comprehensive treatments.

Alongside these continuous developments, a growing body of work has sought discrete counterparts of classical extreme value laws, motivated in part by integer-valued data such as the block maxima considered here. Nakagawa and Osaki [25] obtained the discrete Weibull distribution by discretizing the survival function at the integers; Chakraborty et al. [26] used the same survival-function discretization to construct a discrete analogue of the Gumbel distribution and studied its reliability properties, and Hadi and Khudhair [27] used it to obtain a discrete Fréchet distribution. Das and Das [28] applied the same technique to obtain a discretized Fréchet–Weibull distribution and studied its properties and applications. Akkanphudit [29] extended this line of work to a five-parameter discrete Gompertz–Weibull–Fréchet family nesting several discrete sub-models, and Das et al. [30] further generalized the discretized Fréchet–Weibull construction to the bivariate setting by a maximization construction in which each component is the maximum of its own variable and a shared one. Also relevant, although from a peaks-over-threshold rather than a block-maxima perspective, is the work of Hitz et al. [31], who revisited extreme value analysis for discrete-valued data, proposing a discrete generalized Pareto distribution and a generalized Zipf distribution as threshold-exceedance models for integer data such as word frequencies and tornado outbreak counts. The discrete GEV distribution considered in this paper is not new: its probability mass function was given by Dkengne et al. [32] as the distribution of the integer part of a continuous GEV random variable within a framework of extreme value theory for integer-valued random variables. What has not been provided, to our knowledge, is a systematic treatment of this model. The contributions of the present paper are therefore not the model itself but the following results for it:

(a) an explicit description of the support, including the cases where an endpoint of the continuous GEV is an integer, together with the tail behavior, moments and their existence, the quantile function and quantile-based skewness and kurtosis;
(b) the observation that flooring, ceiling and rounding of a continuous GEV variable all lead to the same discrete family, up to a shift of the location parameter, so that the model does not depend on which of these deterministic recording conventions was used;

(c) a description of the subsets of the parameter space on which the support is locally fixed, a proof that the model is differentiable in quadratic mean on a set that also contains integer endpoints when $\beta > -\frac{1}{2}$, and the resulting asymptotic normality of the maximum likelihood estimators;
(d) simulation evidence on finite-sample performance and on the comparison with the continuous GEV fitted to integer values, under several observation mechanisms including block maxima of integer counts;
(e) discrete versions of the Kolmogorov–Smirnov, Cramér–von Mises and Anderson–Darling goodness-of-fit statistics calibrated by a parametric bootstrap that respects the discretization.

The distribution (1) is continuous, yet annual maximum data are frequently integer-valued. Two representative examples are:

- annual maximum wind speeds, in kilometers per hour, from 1941 to 1970, at Lisbon, Portugal (Tiago de Oliveira [33]): 129, 117, 100, 100, 132, 94, 108, 113, 96, 113, 96, 72, 98, 85, 124, 108, 102, 102, 112, 107, 86, 91, 96, 89, 90, 89, 89, 84, 107, 111. The unique values in these data are 129, 117, 100, 132, 94, 108, 113, 96, 72, 98, 85, 124, 102, 112, 107, 86, 91, 89, 90, 84, 111. Although Tiago de Oliveira [33] does not say how the wind speed measurements were taken, wind speed is usually measured by instruments whose output is naturally truncated. A cup or propeller anemometer, for example, produces electrical pulses as it rotates, and the logging system counts the completed pulses (or revolutions) within a sampling interval. Any partial rotation at the end of the interval is simply not counted, so the recorded value is effectively the floor of the true speed. Similarly, many digital data loggers and analogue-to-digital converters truncate fractional parts rather than round them, because truncation is computationally simpler and was the default in early integer-based hardware and storage formats, where saving whole numbers in knots, m/s, or km/h reduced memory and transmission costs. Truncation also has a mild conservative appeal, since it never overstates the observed speed.
- annual maximum temperatures, in degrees Fahrenheit, from 1901 to 1980 at Oxford, England (Tabony [34]): 89, 84, 84, 85, 81, 92, 79, 84, 85, 79, 95, 87, 83, 87, 82, 83, 89, 84, 86, 79, 89, 86, 93, 86, 85, 85, 80, 87, 87, 89, 75, 95, 89, 86, 86, 83, 87, 84, 84, 85, 89, 88, 92, 87, 84, 84, 90, 90, 89, 86, 82, 89, 90, 83, 87, 82, 86, 80, 91, 82, 87, 77, 81, 87, 81, 80, 83, 89, 88, 88, 84, 77, 85, 77, 91, 94, 80, 80, 85, 83. The unique values in these data are 89, 84, 85, 81, 92, 79, 95, 87, 83, 82, 86, 93, 80, 75, 88, 90, 91, 77, 94. Although Tabony [34] does not say how the temperature measurements were taken, flooring is a natural and widely used recording convention. With traditional liquid-in-glass thermometers, an observer typically reads the highest graduation the column has actually reached rather than guessing where it sits between marks, which amounts to truncation. Many digital sensors and data loggers likewise truncate the fractional part in firmware or at the storage stage because it is computationally trivial and saves space, and archives built from such devices, as well as older paper registers kept in whole degrees, inherit that convention. Flooring also has a conservative, reproducible appeal—the recorded value never overstates what was observed, and two observers reading the same instrument will agree, whereas rounding to the nearest integer requires judging whether the reading is above or below the half-degree point.

Both datasets contain only integer values, yet they have previously been analyzed with continuous extreme value models. The Oxford temperatures were analyzed with continuous extreme value distributions by Tabony [34] and were fitted with the continuous GEV model (1) by Chandler and Bate [35] (Section 5.2, jointly with the annual maxima at Worthing); the Lisbon wind speeds were analyzed with continuous extreme value models

by Tiago de Oliveira [33] and are distributed as an example dataset with the R package evd (Stephenson [36]) for use with its continuous extreme value fitting functions.

We stress that integer values do not, by themselves, imply that the data arise by flooring an underlying continuous GEV variable. Throughout, the discrete GEV is interpreted through the observation model

$$Y = \lfloor X \rfloor, \qquad X \sim \mathrm{GEV}(\mu, \sigma, \beta), \tag{2}$$

in which the continuous annual maximum X is recorded with its fractional part discarded. The arguments above make this mechanism plausible for the two datasets, but the sources do not document it, so it is an assumption. Two remarks limit its consequences. First, if the values were instead obtained by ceiling or by rounding to the nearest integer, then $Y = \lfloor X + c \rfloor$ with $c = 1$ or $c = \frac{1}{2}$, and $X + c$ is again GEV with location $\mu + c$. Hence, ceiling and rounding lead to exactly the same discrete family, with the location parameter shifted by one or one half unit (Section 2.1); only the interpretation of μ depends on which convention was used. Second, when the data are not a discretization of a continuous variable at all (for example, maxima of integer-valued counts), the discrete GEV can only be an approximation; this case is examined via simulation in Section 3.2.

The discrete GEV distribution and its mathematical properties are given in Section 2. Section 3 presents simulation studies assessing finite-sample performance and comparing fits to integer data. The two datasets are revisited in Section 4, where the discrete GEV is compared with the continuous GEV and with other discrete models. Conclusions are given in Section 5. All computations were performed using R software, version 4.5.2 (R Core Team [37]).

2. The Discrete Distribution

The following subsections develop the fundamental mathematical properties of the discrete GEV distribution, providing the foundation for model fitting, inference, and risk quantification in practical extreme-event analysis. Further mathematical properties are given in Appendix A.

The discrete GEV distribution studied here was introduced by Dkengne et al. [32], who gave its probability mass function (4) as the law of $\lfloor X \rfloor$ for a continuous GEV variable X. The remaining results in this section are, to our knowledge, not given there.

2.1. Probability Mass and Cumulative Distribution Functions

The survival function corresponding to (1) is

$$S(x) = 1 - F(x) = 1 - \exp\left[-\left(1 + \beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]. \tag{3}$$

The discretization of a continuous distribution can be achieved using several techniques in the literature. One technique is to take the probability mass function as $p(x) = S(x) - S(x+1)$. Using (3), the probability mass function of a discrete generalized extreme value distribution can be taken to be

$$\begin{aligned} p(x) &= S(x) - S(x+1) \\ &= \left\{1 - \exp\left[-\left(1 + \beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]\right\} - \left\{1 - \exp\left[-\left(1 + \beta\frac{x+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]\right\} \\ &= \exp\left[-\left(1 + \beta\frac{x+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right] - \exp\left[-\left(1 + \beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]. \end{aligned} \tag{4}$$

The cumulative distribution function corresponding to (4) is

$$P(x) = \exp\left[-\left(1+\beta\frac{x+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]. \tag{5}$$

In (3)–(5), F is understood as the GEV distribution function extended to the whole real line, that is, $F(x) = 0$ for $x \le \mu - \frac{\sigma}{\beta}$ when $\beta > 0$ and $F(x) = 1$ for $x \ge \mu - \frac{\sigma}{\beta}$ when $\beta < 0$. With this convention $p(x) = F(x+1) - F(x) = \Pr(x \le X < x+1) = \Pr(\lfloor X \rfloor = x)$ and $P(x) = F(x+1)$ for every integer x, so (4) is the law of $Y = \lfloor X \rfloor$ in (2). The support follows from the endpoints of X. If $\beta > 0$, let $a = \mu - \frac{\sigma}{\beta}$ be the lower endpoint of X. Then, $p(x) > 0$ if and only if $F(x+1) > 0$, that is, $x + 1 > a$, and the smallest admissible integer is $\lfloor a \rfloor$ (if a is an integer then $\lfloor a \rfloor = a$ and $p(a) = F(a+1) > 0$, while $p(a-1) = F(a) - F(a-1) = 0$). If $\beta < 0$, let $b = \mu - \frac{\sigma}{\beta}$ be the upper endpoint of X. Then, $p(x) > 0$ if and only if $F(x) < 1$, that is, $x < b$, and the largest admissible integer is $\lceil b \rceil - 1$ (which equals $b - 1$ if b is an integer and $\lfloor b \rfloor$ otherwise). If $\beta > 0$ in (4), then $x \ge \left\lfloor \mu - \frac{\sigma}{\beta} \right\rfloor$ and the probability mass function corresponds to a general form of the discrete Fréchet distribution due to Hadi and Khudhair [27]. If $\beta < 0$ in (4), then $x \le \left\lceil \mu - \frac{\sigma}{\beta} \right\rceil - 1$ and the probability mass function corresponds to a general form of the reversed discrete Weibull distribution: if $b = \mu - \frac{\sigma}{\beta}$ is an integer, then $\Pr(b - 1 - Y \ge k) = q^{k^{1/|\beta|}}$, $k = 0, 1, \ldots$, with $q = \exp\left[-\left(\frac{|\beta|}{\sigma}\right)^{1/|\beta|}\right]$, so that $b - 1 - Y$ has the discrete Weibull distribution due to Nakagawa and Osaki [25]. If $\beta = 0$, then (4) reduces to

$$p(x) = \exp\left[-\exp\left(-\frac{x+1-\mu}{\sigma}\right)\right] - \exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right],$$

which is the probability mass function of the discrete Gumbel distribution due to Chakraborty *et al.* [26]. Hence, the domain of (4) can be summarized as follows:

$$\text{Domain of (4)} = \begin{cases} \left\lfloor \mu - \frac{\sigma}{\beta} \right\rfloor, \left\lfloor \mu - \frac{\sigma}{\beta} \right\rfloor + 1, \ldots, & \text{if } \beta > 0, \\ \ldots, \left\lceil \mu - \frac{\sigma}{\beta} \right\rceil - 2, \left\lceil \mu - \frac{\sigma}{\beta} \right\rceil - 1, & \text{if } \beta < 0, \\ \ldots, -1, 0, 1, \ldots, & \text{if } \beta = 0. \end{cases} \tag{6}$$

The support therefore depends on the parameters through the endpoint $\mu - \frac{\sigma}{\beta}$ when $\beta \neq 0$, and it changes exactly when this endpoint crosses an integer. The consequences for likelihood inference are discussed in Section 2.4.

Remark 1 (**other rounding conventions**). *Let c be a constant and $Y_c = \lfloor X + c \rfloor$ with $X \sim GEV(\mu, \sigma, \beta)$. Since $X + c \sim GEV(\mu + c, \sigma, \beta)$, Y_c has probability mass function (4) with μ replaced by $\mu + c$. Ceiling ($\lceil X \rceil = \lfloor X \rfloor + 1$ almost surely, i.e., $c = 1$) and rounding to the nearest integer ($c = \frac{1}{2}$) are special cases. Thus, the family (4) is the same under all three conventions; they differ only in how μ relates to the location of the underlying continuous variable. In particular, if the recording convention is unknown, the fitted discrete GEV is unaffected, and only the reconstruction of the continuous location is uncertain by at most one unit.*

The discretization method employed to construct (4) is particularly well-suited to the GEV framework for several compelling reasons. First, this survival-function-based approach directly inherits the cumulative distribution function of the continuous GEV, yielding the elegant closed form

$$P(x) = \exp\left[-\left(1+\beta\frac{x+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right], \tag{7}$$

which preserves the parametric interpretability of the original location μ, scale σ and shape β parameters without reparametrization. Second, and crucially, this formulation guarantees that the resulting discrete distribution unifies the discrete Fréchet ($\beta > 0$), discrete Weibull ($\beta < 0$), and discrete Gumbel ($\beta = 0$) distributions as special cases, exactly mirroring the unification property of the continuous GEV—a theoretical coherence that alternative discretization schemes based on, for example, sampling the continuous density at the integers and renormalizing, or matching moments would generally destroy. Third, unlike density-based discretization (which requires a normalizing constant given by an infinite sum and does not yield a closed-form CDF) or hazard-rate-based schemes (which do not naturally preserve the GEV tail behavior), the survival-function method concentrates probability mass precisely at integer values by taking differences of the survival function at consecutive integers, so that, under the observation model (2), the likelihood of integer-valued annual maxima—such as wind speeds in km/h or temperatures in degrees Fahrenheit—is the exact grouped-data likelihood of the continuous GEV, rather than an approximation obtained by evaluating a density at the integers. Indeed, under (2), the interval (grouped-data) likelihood of the continuous GEV, $\prod_i [F(x_i + 1) - F(x_i)]$, coincides with the likelihood of the discrete GEV; the two are the same model. The practical difference between the discrete GEV and the continuous GEV fitted by its density to integer values is examined in Sections 3 and 4.

If $\beta > 0$, then $p(x)$ and $P(x)$ behave as

$$p(x) \sim \frac{1}{\sigma}\left(1 + \beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}-1}$$

and

$$P(x) \sim 1 - \left(1 + \beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}$$

as $x \to \infty$. Furthermore, since $F(\lfloor a \rfloor) = 0$ with $a = \mu - \frac{\sigma}{\beta}$,

$$p\left(\left\lfloor \mu - \frac{\sigma}{\beta} \right\rfloor\right) = P\left(\left\lfloor \mu - \frac{\sigma}{\beta} \right\rfloor\right) = \exp\left[-\left(1 + \beta\frac{\left\lfloor \mu - \frac{\sigma}{\beta} \right\rfloor + 1 - \mu}{\sigma}\right)^{-\frac{1}{\beta}}\right].$$

If $\beta < 0$, then, since $F(\lceil b \rceil) = 1$ with $b = \mu - \frac{\sigma}{\beta}$,

$$p\left(\left\lceil \mu - \frac{\sigma}{\beta} \right\rceil - 1\right) = 1 - \exp\left[-\left(1 + \beta\frac{\left\lceil \mu - \frac{\sigma}{\beta} \right\rceil - 1 - \mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]$$

and

$$P\left(\left\lceil \mu - \frac{\sigma}{\beta} \right\rceil - 1\right) = 1.$$

If $\beta = 0$, then $p(x)$ and $P(x)$ behave as

$$p(x) \sim \exp\left(-\frac{x-\mu}{\sigma}\right) - \exp\left(-\frac{x+1-\mu}{\sigma}\right)$$

and

$$P(x) \sim 1 - \exp\left(-\frac{x+1-\mu}{\sigma}\right)$$

as $x \to \infty$.

2.2. Mean, Variance and Characteristic Function

The mean and variance corresponding to (4), that is, of $Y = \lfloor X \rfloor$ in (2), are

$$E(Y) = \sum_{j=-\infty}^{\infty} j\left\{\exp\left[-\left(1+\beta\frac{j+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right] - \exp\left[-\left(1+\beta\frac{j-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]\right\} \tag{8}$$

and

$$Var(Y) = \sum_{j=-\infty}^{\infty} j^2\left\{\exp\left[-\left(1+\beta\frac{j+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right] - \exp\left[-\left(1+\beta\frac{j-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]\right\} - [E(Y)]^2, \tag{9}$$

where the sums run over the support (6). For $\beta < 0$, the support is bounded above, and the lower tail of (4) decays faster than any power of $|j|$; for $\beta = 0$, both tails decay at least exponentially; for $\beta > 0$, the support is bounded below and $p(j) \sim cj^{-\frac{1}{\beta}-1}$ as $j \to \infty$ for some $c > 0$. Hence, $E(|Y|^k) < \infty$ if and only if $k\beta < 1$, so (8) is finite if and only if $\beta < 1$ and (9) is finite if and only if $\beta < \frac{1}{2}$. (For $\beta > 0$, a ratio test on consecutive terms of the sums cannot decide existence since the ratio of consecutive probabilities tends to one.)

The characteristic function corresponding to (4) is

$$E[\exp(itY)] = \sum_{j=-\infty}^{\infty} \exp(itj)\left\{\exp\left[-\left(1+\beta\frac{j+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right] - \exp\left[-\left(1+\beta\frac{j-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]\right\},$$

where $\mathrm{i} = \sqrt{-1}$.

2.3. Quantiles and Associated Measures

The quantile function corresponding to (4) is

$$Q(q) = \left\lceil \mu + \frac{\sigma}{\beta}\left\{[-\log(q)]^{-\beta} - 1\right\} - 1 \right\rceil \tag{10}$$

provided $\beta \neq 0$. In particular,

$$Q\left(\frac{1}{8}\right) = \left\lceil \mu + \frac{\sigma}{\beta}\left\{[\log 8]^{-\beta} - 1\right\} - 1 \right\rceil,$$

$$Q\left(\frac{1}{4}\right) = \left\lceil \mu + \frac{\sigma}{\beta}\left\{[\log 4]^{-\beta} - 1\right\} - 1 \right\rceil,$$

$$Q\left(\frac{3}{8}\right) = \left\lceil \mu + \frac{\sigma}{\beta}\left\{\left[-\log\left(\frac{3}{8}\right)\right]^{-\beta} - 1\right\} - 1 \right\rceil,$$

$$Q\left(\frac{1}{2}\right) = \left\lceil \mu + \frac{\sigma}{\beta}\left\{[\log 2]^{-\beta} - 1\right\} - 1 \right\rceil,$$

$$Q\left(\frac{5}{8}\right) = \left\lceil \mu + \frac{\sigma}{\beta}\left\{\left[-\log\left(\frac{5}{8}\right)\right]^{-\beta} - 1\right\} - 1 \right\rceil,$$

$$Q\left(\frac{3}{4}\right) = \left\lceil \mu + \frac{\sigma}{\beta}\left\{\left[-\log\left(\frac{3}{4}\right)\right]^{-\beta} - 1\right\} - 1 \right\rceil$$

and

$$Q\left(\frac{7}{8}\right) = \left\lceil \mu + \frac{\sigma}{\beta}\left\{\left[-\log\left(\frac{7}{8}\right)\right]^{-\beta} - 1\right\} - 1 \right\rceil.$$

If $\beta = 0$, then

$$Q(q) = \lceil \mu - \sigma \log(-\log q) - 1 \rceil.$$

In particular,

$$Q\left(\frac{1}{8}\right) = \lceil \mu - \sigma \log(\log 8) - 1 \rceil,$$

$$Q\left(\frac{1}{4}\right) = \lceil \mu - \sigma \log(\log 4) - 1 \rceil,$$

$$Q\left(\frac{3}{8}\right) = \left\lceil \mu - \sigma \log\left(-\log \frac{3}{8}\right) - 1 \right\rceil,$$

$$Q\left(\frac{1}{2}\right) = \lceil \mu - \sigma \log(\log 2) - 1 \rceil,$$

$$Q\left(\frac{5}{8}\right) = \left\lceil \mu - \sigma \log\left(-\log \frac{5}{8}\right) - 1 \right\rceil,$$

$$Q\left(\frac{3}{4}\right) = \left\lceil \mu - \sigma \log\left(-\log \frac{3}{4}\right) - 1 \right\rceil$$

and

$$Q\left(\frac{7}{8}\right) = \left\lceil \mu - \sigma \log\left(-\log \frac{7}{8}\right) - 1 \right\rceil.$$

Bowley's measure of skewness and Moors' measure of kurtosis can be computed using

$$\text{Skewness} = \frac{Q\left(\frac{3}{4}\right) - 2Q\left(\frac{1}{2}\right) + Q\left(\frac{1}{4}\right)}{Q\left(\frac{3}{4}\right) - Q\left(\frac{1}{4}\right)} \tag{11}$$

and

$$\text{Kurtosis} = \frac{Q\left(\frac{3}{8}\right) - Q\left(\frac{1}{8}\right) + Q\left(\frac{7}{8}\right) - Q\left(\frac{5}{8}\right)}{Q\left(\frac{3}{4}\right) - Q\left(\frac{1}{4}\right)}, \tag{12}$$

respectively.

Figures 1 and 2 display the Bowley skewness (11) and Moors kurtosis (12) of the discrete GEV distribution as functions of the shape parameter β over the range $-1 \le \beta \le 1$. The skewness plot shows a clear step-like relationship with β: the skewness equals -1 (maximal negative skewness) for $\beta \le -0.37$, corresponding to the short-tailed discrete Weibull regime; it equals 0 for $-0.36 \le \beta \le 0.70$, an interval containing the Gumbel case $\beta = 0$; and it equals $\frac{1}{3}$ for $\beta \ge 0.71$ (on a grid of spacing 0.01), corresponding to the heavy-tailed discrete Fréchet regime. This pattern confirms the expected ordering of skewness across the three sub-families unified by the discrete GEV, with the Weibull-type left tail producing pronounced negative asymmetry and the Fréchet-type right tail producing mild positive asymmetry. The kurtosis plot in Figure 2 is more irregular, taking only the values 1, $\frac{4}{3}$, $\frac{3}{2}$, $\frac{5}{3}$, 2 and $\frac{5}{2}$ rather than varying smoothly with β; this behavior reflects the fact that the quantile function $Q(q)$ in (12) is defined via a ceiling operator, so that small changes in β can leave the relevant quantiles unchanged over intervals before jumping discontinuously, producing the observed piecewise-constant, staircase appearance. Taken

together, the two figures illustrate that the shape parameter β governs both the direction and magnitude of asymmetry in the discrete GEV distribution while also revealing that measures based on discrete quantiles inherit a step-function character not present in their continuous GEV counterparts.

We have fixed $\mu = 0$ and $\sigma = 1$ in Figures 1 and 2. Since $Q(q)$ increases by one when μ increases by one, both measures are periodic in μ with period one; for $\sigma = 1$ and $\beta \in \{-0.4, -0.2, 0, 0.2, 0.4\}$, as μ varied over $[0, 1)$, the skewness took at most three distinct values (-1, 0 and 1), and the kurtosis took at most four. The effect of σ is not monotone. As σ increases, the step character disappears and both measures approach those of the continuous GEV; for example, for $\beta = 0$, the skewness is 0, 0.333, 0.250, 0.125 and 0.114 for $\sigma = 1, 2, 5, 10, 50$, against the continuous GEV value 0.118, and the kurtosis is 1.500, 1.667, 1.375, 1.312 and 1.266, against 1.278.

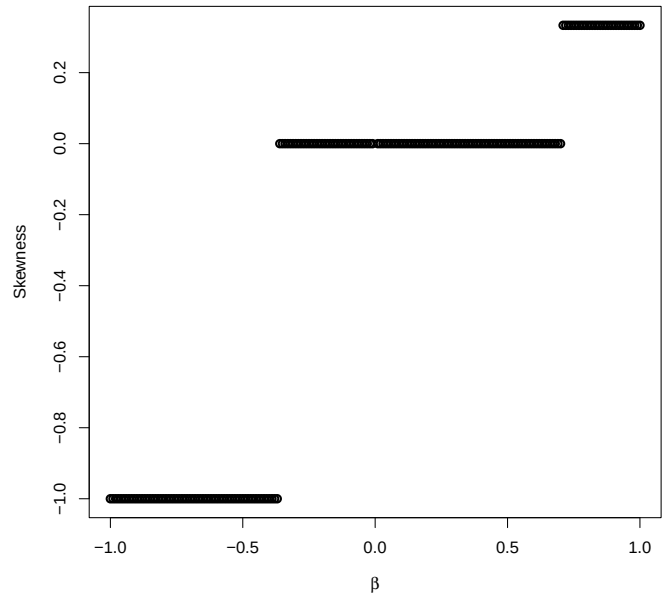


Figure 1. Plot of (11) versus β for $\mu = 0$ and $\sigma = 1$.

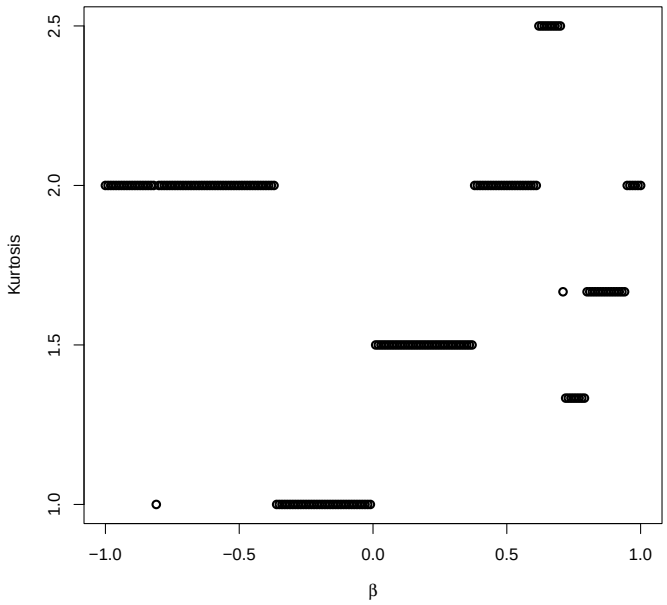


Figure 2. Plot of (12) versus β for $\mu = 0$ and $\sigma = 1$.

2.4. Maximum Likelihood Estimation

The estimation of parameters in (4) can be performed using the method of maximum likelihood. Suppose $x_1, x_2, \ldots, x_n$ is a random sample from (4). The log likelihood function is

$$\log L(\mu, \sigma, \beta) = \sum_{i=1}^{n} \log\left\{\exp\left[-\left(1+\beta\frac{x_i+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right] - \exp\left[-\left(1+\beta\frac{x_i-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]\right\}. \tag{13}$$

The partial derivatives of (13) with respect to the parameters μ, σ and β are

$$\frac{\partial \log L}{\partial \mu} = \sum_{i=1}^{n} \frac{a_1(x_i;\mu,\sigma,\beta) - a_1(x_i+1;\mu,\sigma,\beta)}{\exp\left[-\left(1+\beta\frac{x_i+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right] - \exp\left[-\left(1+\beta\frac{x_i-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]},$$

$$\frac{\partial \log L}{\partial \sigma} = \sum_{i=1}^{n} \frac{a_2(x_i;\mu,\sigma,\beta) - a_2(x_i+1;\mu,\sigma,\beta)}{\exp\left[-\left(1+\beta\frac{x_i+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right] - \exp\left[-\left(1+\beta\frac{x_i-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]}$$

and

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^{n} \frac{a_3(x_i;\mu,\sigma,\beta) - a_3(x_i+1;\mu,\sigma,\beta)}{\exp\left[-\left(1+\beta\frac{x_i+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right] - \exp\left[-\left(1+\beta\frac{x_i-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]},$$

where

$$a_1(x;\mu,\sigma,\beta) = \frac{1}{\sigma}\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}-1}\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right],$$

$$a_2(x;\mu,\sigma,\beta) = \frac{x-\mu}{\sigma^2}\left[\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}-1}\right]\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]$$

and

$$a_3(x;\mu,\sigma,\beta) = \frac{1}{\beta^2}\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\log\left(1+\beta\frac{x-\mu}{\sigma}\right)\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]$$
$$-\frac{x-\mu}{\beta\sigma}\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}-1}\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right].$$

The maximum likelihood estimates of (μ, σ, β), say $\left(\hat{\mu}, \hat{\sigma}, \hat{\beta}\right)$, can be obtained as simultaneous solutions of $\frac{\partial \log L}{\partial \mu} = 0$, $\frac{\partial \log L}{\partial \sigma} = 0$ and $\frac{\partial \log L}{\partial \beta} = 0$. An alternative approach is to maximize (13) directly to obtain the maximum likelihood estimates. We pursued the alternative approach in Sections 3 and 4. R code for this maximization, using the function optim, and R code computing other properties of the discrete GEV distribution, are given in Appendix B.

When maximizing the likelihood surface directly using the optim function in R with the Nelder–Mead method, the optimization was initialized at the starting values $\mu^{(0)} = \bar{x} - 0.45s$, $\sigma^{(0)} = 0.78s$ and $\beta^{(0)} = 0$, where $\bar{x}$ and s are the sample mean and standard deviation. These are the moment estimates for the Gumbel special case ($\sigma = \frac{\sqrt{6}}{\pi}s$, $\mu = \bar{x} - 0.5772\sigma$), so the starting point lies in the region of the data, and the shape parameter starts at the Gumbel value as a neutral default. Data-independent starting values such as $(0, 1, 0)$ cannot be used: for data of the size of the annual maxima in Section 4, every probability in (13) is then zero in floating point, so the log-likelihood cannot be evaluated. To guard against the possibility of convergence to a local rather than global maximum, the optimization was repeated from additional sets of perturbed starting values obtained by adding independent Uniform$(1, 20)$ and Uniform$(-0.2, 0.2)$ random perturbations to the initial values of $\sigma^{(0)}$

and $\beta^{(0)}$, respectively; in all cases, the optimizer converged to the same parameter estimates (perturbations that made the log-likelihood infinite were discarded). Agreement across starting values, and a positive definite observed information matrix at the solution, show only that a local maximum has been found; they do not prove that it is the global maximum. For the two datasets in Section 4, we therefore also computed the profile log-likelihood of β, maximizing over (μ, σ) from several starting values on a grid of β values in $[-0.7, 0.6]$ with spacing 0.02. In both cases, the profile was unimodal on this grid, with its maximum at the grid point nearest to $\hat{\beta}$, which supports, but does not prove, that the reported estimates are global maxima. The convergence criterion was the default tolerance of $\sqrt{\epsilon} \approx 1.5\times10^{-8}$, where ϵ is the machine precision, on the relative change in the objective function value between successive iterations, with a maximum of 10,000 iterations permitted via control = list(maxit = 10,000). Apart from the discarded starting values mentioned above, no computational or statistical issues were encountered during optimization: there were no warnings or errors related to non-finite values, flat likelihood regions, or parameter boundary constraints; the optimization was carried out over $(\mu, \log\sigma, \beta)$ (Appendix B) so that the constraint $\sigma > 0$ is satisfied automatically. At the reported maximum likelihood estimates, the observed information matrix J (whose elements are given in Appendix C) was computed numerically and verified to be positive definite for both datasets, with all eigenvalues strictly positive, confirming that the stationary point identified by optim is a local maximum of the log-likelihood.

Regularity problems with maximum likelihood estimation of the continuous GEV distribution have been well investigated in the literature; see, for example, Smith [38] and Davison and Smith [39]. In the parametrization (1), the usual asymptotic properties of efficiency and normality of maximum likelihood estimators hold if $\beta > -\frac{1}{2}$, but not for $\beta \le -\frac{1}{2}$; if $\beta < -1$, the continuous likelihood is unbounded because the density tends to infinity at the upper endpoint, and maximum likelihood estimators do not exist. These problems arise from the behavior of the continuous density near a parameter-dependent endpoint. They do not carry over to the discrete GEV automatically because $p(x)$ is a probability bounded by one, but the support (6) still depends on the parameters, so standard likelihood theory cannot be applied without further argument. We now make this precise.

Let $\Theta = \{(\mu, \sigma, \beta) : -\infty < \mu < \infty, \sigma > 0, -\infty < \beta < \infty\}$, write $\theta = (\mu, \sigma, \beta)$, $e(\theta) = \mu - \frac{\sigma}{\beta}$ for $\beta \neq 0$, $S(\theta) = \{x \in \mathbb{Z} : p(x;\theta) > 0\}$ for the support, and define

$$\Theta_F = \{\theta \in \Theta : \beta \neq 0,\ e(\theta) \notin \mathbb{Z}\}, \qquad \Theta_D = \Theta \setminus \left\{\theta \in \Theta : \beta \le -\frac{1}{2},\ e(\theta) \in \mathbb{Z}\right\}.$$

Proposition 1. *(i) If $\theta_0 \in \Theta_F$, there is an open neighborhood N of θ_0 on which $S(\theta) = S(\theta_0)$, and, for every $x \in S(\theta_0)$, $\theta \mapsto p(x;\theta)$ is infinitely differentiable and positive on N. (ii) For every $\theta_0 \in \Theta_D$ and every integer x, $\theta \mapsto \sqrt{p(x;\theta)}$ is continuously differentiable in a neighborhood of θ_0. (iii) For a sample $x_1, \ldots, x_n$ with $\beta < 0$ (respectively $\beta > 0$), the log-likelihood (13) is finite if and only if $e(\theta) > x_{(n)}$ (respectively $e(\theta) < x_{(1)} + 1$), and it tends to $-\infty$ as $e(\theta) \downarrow x_{(n)}$ (respectively $e(\theta) \uparrow x_{(1)} + 1$).*

Proof. (i) If $\beta_0 < 0$ and $b_0 = e(\theta_0) \notin \mathbb{Z}$, then $\lceil e(\theta) \rceil = \lceil b_0 \rceil$ for θ close to θ_0 by continuity of e, so the support $\{\ldots, \lceil b_0 \rceil - 1\}$ is fixed. For $x \le \lceil b_0 \rceil - 2$, both x and $x+1$ lie strictly inside the support of X, and for $x = \lceil b_0 \rceil - 1$, we have $p(x;\theta) = 1 - F(x;\theta)$ with x strictly inside; in either case, $p(x;\theta)$ is a smooth positive function of θ. The case $\beta_0 > 0$ is identical, with the lower endpoint. (ii) For $\theta_0 \in \Theta_F$, this follows from (i), and for $\beta_0 = 0$, it follows because, for fixed x, $F(x;\theta)$ is an analytic function of θ near $\beta = 0$. The remaining cases are those in which the endpoint is an integer. If $\beta_0 > 0$ and $e(\theta_0) = a_0 \in \mathbb{Z}$, the only probability that changes from

zero to positive values near θ_0 is that of $x = a_0 - 1$, namely $F(a_0;\theta) = \exp\left\{-\left[\beta\frac{a_0-e(\theta)}{\sigma}\right]^{-\frac{1}{\beta}}\right\}$ when $e(\theta) < a_0$ and zero otherwise; this function and all its derivatives tend to zero as $e(\theta) \uparrow a_0$, so its square root is continuously differentiable. If $-\frac{1}{2} < \beta_0 < 0$ and $e(\theta_0) = b_0 \in \mathbb{Z}$, the probability that changes is that of $x = b_0$, namely $1 - F(b_0;\theta) \sim \left[|\beta|\frac{e(\theta)-b_0}{\sigma}\right]^{\frac{1}{|\beta|}}$ as $e(\theta) \downarrow b_0$, and zero otherwise. Its square root behaves like $(e(\theta) - b_0)_+^{\frac{1}{2|\beta|}}$, and $\frac{1}{2|\beta|} > 1$ when $|\beta| < \frac{1}{2}$, so it is continuously differentiable with zero derivative at $e(\theta) = b_0$. The probability of $x = b_0 - 1$ is $1 - F(b_0 - 1;\theta)$ or $F(b_0;\theta) - F(b_0 - 1;\theta)$ on the two sides, and is continuously differentiable for the same reason. (iii) If $\beta < 0$, $p\big(x_{(n)};\theta\big) > 0$ if and only if $x_{(n)} < e(\theta)$, and as $e(\theta) \downarrow x_{(n)}$, we have $p\big(x_{(n)};\theta\big) \le 1 - F\big(x_{(n)};\theta\big) \to 0$ while the other terms of (13) are bounded above by zero. The case $\beta > 0$ is analogous with $p\big(x_{(1)};\theta\big) \le F\big(x_{(1)} + 1;\theta\big) \to 0$. □

Part (iii) contrasts with the continuous GEV with $\beta < -1$, where the likelihood tends to $+\infty$ as the endpoint approaches the largest observation: in the discrete model, the boundary of the region of finite likelihood repels the maximizer instead of attracting it. Boundedness of p prevents divergence of the likelihood, but it does not by itself guarantee that a maximizer exists or that it is asymptotically normal; these require the conditions below.

Let $I(\theta) = \sum_{x\in S(\theta)} p(x;\theta)\,\nabla\log p(x;\theta)\,\nabla\log p(x;\theta)^\top$ denote the Fisher information matrix of one observation. For $\beta < 0$, the lower tail of (4) decays faster than any power, and for $\beta > 0$, the upper tail decays like a power while the score grows only logarithmically, so $I(\theta)$ is finite on Θ_D. We computed $I(\theta)$ numerically, summing over the support, at $\mu = 0$, $\sigma = 1$ and the five values of β used in Section 3.1, namely $\beta = 0, 0.2, -0.2, -0.5$ and -1; in every case, it was positive definite, with the smallest eigenvalue between 0.31 and 0.69. All four non-zero values give an integer endpoint ($e(\theta) = -5, 5, 2$ and 1, respectively). For $\beta = \pm 0.2$, these are covered by Proposition 1(ii), but $\beta = -0.5$ and $\beta = -1$ lie in the case $\beta \le -\frac{1}{2}$ with an integer endpoint that is excluded from Θ_D, so the simulations at these two values examine the estimators outside the scope of Theorem 1.

Theorem 1. *Let θ_0 be an interior point of Θ_D, and suppose that (A1) $I(\theta)$ is continuous at θ_0 and $I(\theta_0)$ is nonsingular; (A2) there is a function $\dot{m}$ with $\sum_x \dot{m}(x)^2 p(x;\theta_0) < \infty$ such that $|\log p(x;\theta_1) - \log p(x;\theta_2)| \le \dot{m}(x)\|\theta_1 - \theta_2\|$ for all $x \in S(\theta_0)$ and all θ_1, θ_2 in a neighborhood of θ_0; and (A3) the maximum likelihood estimator $\hat{\theta}_n$ is consistent. Then,*

$$\sqrt{n}\big(\hat{\theta}_n - \theta_0\big) \to N_3\big(0, I(\theta_0)^{-1}\big)$$

in distribution as $n \to \infty$.

Proof. Based on Proposition 1(ii) and (A1), the model is differentiable in quadratic mean at θ_0 (van der Vaart [40], Lemma 7.6). The result then follows from Theorem 5.39 of van der Vaart [40]. □

Three comments on Theorem 1 are in order. First, on Θ_F, the support is locally fixed (Proposition 1(i)), so there, the conditions reduce to the classical Cramér-type conditions (see, for example, Cox and Hinkley [41]) applied to a family with a fixed countable support. Second, Proposition 1(ii) shows that crossing of an integer by the endpoint is harmless when $\beta > -\frac{1}{2}$ because the probability that enters or leaves the support is then of smaller order than the squared parameter change. For $\beta \le -\frac{1}{2}$ with an integer endpoint, this argument fails and the asymptotics may be non-standard; we do not study this case theoretically, but two of the simulation settings of Section 3.1 lie in it. Third, although Theorem 1 formally allows $\beta_0 \le -\frac{1}{2}$ when the endpoint is not an integer, the neighborhood in Proposition 1(i) shrinks as the endpoint approaches an integer, so the normal approximation can be expected

to be poor in finite samples in that region. We have not proved consistency (A3) in general; it holds, for example, under Wald's conditions when Θ is replaced by a compact subset of Θ_D on which the model is identifiable, and identifiability is examined numerically in Section 3.1.

Under the conditions of Theorem 1, $\left(\hat{\mu}, \widehat{\sigma}, \hat{\beta}\right)$ is approximately trivariate normal with mean (μ, σ, β) and variance-covariance matrix $\frac{1}{n} I(\mu, \sigma, \beta)^{-1}$, which can be estimated by J^{-1}, the inverse of the observed information matrix

$$J = \begin{pmatrix} J_{1,1} & J_{1,2} & J_{1,3} \\ J_{1,2} & J_{2,2} & J_{2,3} \\ J_{1,3} & J_{2,3} & J_{3,3} \end{pmatrix}$$

evaluated at $\left(\hat{\mu}, \widehat{\sigma}, \hat{\beta}\right)$; expressions for its elements are given in Appendix C. Wald-type tests and confidence intervals follow in the usual way.

3. Simulation Studies

In this section, we perform two simulation studies. The first is to check the finite sample performances of the maximum likelihood estimates in Section 2.4. The second is to compare the fits of (4) and (1) to integer data under several observation mechanisms.

3.1. Finite Sample Performance of Maximum Likelihood Estimates

In this section, we assess finite sample performances of the maximum likelihood estimates of the parameters and of the extreme quantile $Q(0.999)$ in terms of biases, mean squared errors, coverage lengths (average lengths of the confidence intervals) and coverage probabilities, and we check if the parameters in the model are identifiable. We used the scheme described in Appendix D.

All simulations used $\mu = 0$ and $\sigma = 1$ and sample sizes $n = 40, 41, \ldots, 200$. Plots of the biases, mean squared errors, coverage lengths and coverage probabilities are shown in Figures A1–A4 in Appendix E when $\sigma = 1$ and $\beta = 0$. Plots of the biases, mean squared errors, coverage lengths and coverage probabilities are shown in Figures A5–A8 in Appendix E when $\sigma = 1$ and $\beta = 0.2$. Plots of the biases, mean squared errors, coverage lengths and coverage probabilities are shown in Figures A9–A12 in Appendix E when $\sigma = 1$ and $\beta = -0.2$. Plots of the biases, mean squared errors, coverage lengths and coverage probabilities are shown in Figures A13–A16 in Appendix E when $\sigma = 1$ and $\beta = -0.5$. Plots of the biases, mean squared errors, coverage lengths and coverage probabilities are shown in Figures A17–A20 in Appendix E when $\sigma = 1$ and $\beta = -1$. The horizontal reference lines in the figures correspond to the biases being zero or the coverage probabilities being equal to 0.95.

Before describing the figures in detail, we highlight their most important limitation. The settings $\beta = -0.5$ and $\beta = -1$ have integer endpoints and lie outside the scope of Theorem 1 (see Section 2.4). For $\beta = -1$, the bootstrap was unreliable: the average lengths of the bootstrap intervals in Figure A19 could be computed only for a minority of sample sizes, so that the curves in Figure A19 are interrupted; this indicates that for most values of n, the bootstrap did not give a finite interval in at least one replication. In the same setting, the estimators of σ and β have a clear negative bias (Figure A17): the bias of $\widehat{\sigma}$ is about -0.16 at $n = 40$ and still about -0.05 at $n = 200$, and the bias of $\hat{\beta}$ is about -0.2 at $n = 40$ and approaches zero only for n above about 150, while the bias of $\widehat{Q}(0.999)$ does not decrease with n: it increases from about -0.1 at $n = 40$ to about 0.2 at $n = 200$, with peaks above 0.4. The bootstrap also failed occasionally for $\beta = -0.5$: the curves of the average interval lengths for μ, σ and β in Figure A15 have a few short gaps. Users should therefore treat bootstrap intervals and point estimates with caution when $\hat{\beta}$ is close to or below -1, and they

should report such fits with this caveat. The simulations also do not establish consistency of the estimators: they show the finite-sample behavior for the settings considered, which is consistent with, but does not prove, the asymptotic results of Section 2.4.

For $\beta = 0$ and $\beta = \pm 0.2$, the bias curves fluctuate around zero with an amplitude that contracts as n increases. For the Gumbel case $\beta = 0$ (Figure A1), the bias of $\hat{\mu}$ is small and positive at small n, the biases of $\widehat{\sigma}$ and $\hat{\beta}$ are negative (about -0.05 at $n = 40$), and all three are close to zero for $n \gtrsim 100$; the bias of $\widehat{Q}(0.999)$ is about 1 at $n = 40$ and settles close to zero. For $\beta = 0.2$ (Figure A5), the heavy-tailed case, the bias of $\widehat{Q}(0.999)$ is larger, about 5 near $n = 45$, and decreases slowly towards zero; for $\beta = -0.2$ (Figure A9), it stays within about ± 0.2. In both cases, the bias in $\widehat{\sigma}$ is negative for small to moderate n, indicating a downward finite-sample bias in the scale estimate that diminishes only slowly. For $\beta = -0.5$ (Figure A13), the biases of $\hat{\beta}$ and $\widehat{Q}(0.999)$ are negative throughout, about -0.04 and -0.08 at $n = 200$, and small in absolute terms; the case $\beta = -1$ was described above.

The mean squared errors of $\hat{\mu}$, $\widehat{\sigma}$ and $\hat{\beta}$ decrease, apart from sampling noise, as n increases in all five parameter configurations, which is in line with, though it does not by itself establish, the $\sqrt{n}$-consistency implied by Theorem 1 where that theorem applies. The mean squared error of $\widehat{Q}(0.999)$ also decreases with n, but its size depends strongly on the tail: it falls from about 15 to about 2 for $\beta = 0$ (Figure A2) and from about 1.5 to about 0.2 for $\beta = -0.2$ (Figure A10), whereas for $\beta = 0.2$ (Figure A6), it reaches about 400 near $n = 50$ before falling to about 15 at $n = 200$. For the short-tailed settings $\beta = -0.5$ and $\beta = -1$ (Figures A14 and A18), it is at most about 0.2 throughout, because $Q(0.999)$ is then close to the upper endpoint of the support. Thus, estimation of the extreme quantile is most variable in the heavy-tailed case.

The average lengths of the 95% bootstrap confidence intervals for μ, σ and β decrease smoothly with n for $\beta = 0, \pm 0.2$ and -0.5 (Figures A3, A7, A11 and A15), apart from the few gaps for $\beta = -0.5$ noted above. The average length for $Q(0.999)$ decreases from about 14 to about 5 for $\beta = 0$, from about 55 to about 15 for $\beta = 0.2$, from about 4.5 to about 1.8 for $\beta = -0.2$ and from about 1.4 to about 0.7 for $\beta = -0.5$ as n increases from 40 to 200. For $\beta = -1$, see the limitation described above.

Across all five values of β, the empirical coverage probabilities for μ, σ, β and $Q(0.999)$ fluctuate around the nominal 95% level, mostly between 0.90 and 1.00, without a systematic departure from it, and the amplitude of the fluctuations decreases slightly as n increases. This holds also for $\beta = -0.5$ and $\beta = -1$ (Figures A16 and A20). For $\beta = -1$, however, approximately correct coverage should not be read as evidence that inference is reliable in this region, given the bootstrap failures and the biases described above.

The values of (A1) in Appendix D, giving the proportion of observed information matrices that are positive definite, for the five parameter choices were 0.9999, 0.9993, 0.9991, 0.9998 and 0.9996. The proportions of cases when Kreutz [42]'s test for identifiability was not rejected were 0.9993, 0.9999, 0.9992, 0.9993 and 0.9991. Hence, there is no evidence that the parameters of the discrete GEV distribution are not identifiable.

3.2. Comparison of (4) and (1) for Integer Data

The comparison of (4) and (1) depends on how the integer data arise. If the data are generated as $Y = \lfloor X \rfloor$ with X following (1), the discrete GEV is the exact model for Y, and a simulation under this mechanism alone necessarily favors it. We therefore considered several observation mechanisms, and we compared the two models under each using the following scheme:

(i) generate integer data $y_1, y_2, \ldots, y_{1000}$ by one of the mechanisms

- *floor* : $y_i = \lfloor x_i \rfloor$, *ceiling*: $y_i = \lceil x_i \rceil$, or *rounding*: y_i equal to x_i rounded to the nearest integer, where $x_1, \ldots, x_{1000}$ is a random sample from (1) with $\mu = 0$, $\sigma \in \{1, 5\}$ and $\beta \in \{-0.4, -0.2, 0, 0.2, 0.4\}$;
- *integer block maxima*: y_i is the maximum of 365 independent Poisson variables with mean 10, or of 365 independent negative binomial variables with mean 5 and size parameter 2; these data are integer-valued by nature and are not a discretization of a GEV variable, so neither model is exact;

(ii) fit (4) by maximizing (13), and fit (1) by maximizing the likelihood built from its density at $y_1, \ldots, y_{1000}$;

(iii) for each distinct value y, compute the fitted probability of y under each model: $\hat{p}(y)$ for (4), and for (1), the fitted probability of the interval that the integer y represents under the mechanism used, namely $\hat{F}(y+1) - \hat{F}(y)$ for floor, $\hat{F}(y) - \hat{F}(y-1)$ for ceiling and $\hat{F}\left(y + \frac{1}{2}\right) - \hat{F}\left(y - \frac{1}{2}\right)$ for rounding; for the floor, ceiling and rounding mechanisms, this is the most favorable choice for the continuous model, since it uses the true mechanism, which the discrete GEV is not told; for integer block maxima, there is no natural discretization mechanism and the choice of interval for the continuous GEV is a convention: we used the midpoint-binning convention $\hat{F}\left(y + \frac{1}{2}\right) - \hat{F}\left(y - \frac{1}{2}\right)$ as the primary choice, since it is the standard convention implicitly assumed by the density-based fit, and as a sensitivity check, we also compute the SSD under the floor convention $\hat{F}(y+1) - \hat{F}(y)$ (results are reported in the paragraph following Table 1);

(iv) compute, for each model, the sum of squared differences (SSD) between these fitted probabilities and the empirical probabilities, where the sum runs only over the observed distinct values of y (integers with zero empirical frequency are excluded); as a sensitivity check, also compute an alternative SSD summing over the full integer range—from $\min_i y_i$ to $\max_i y_i$— (results are reported in the paragraph following Table 1);

(v) repeat steps (i) to (iv) 1000 times.

With 1000 replications,and the Monte Carlo standard error of a reported proportion is at most 0.016. Table 1 gives, for each setting, the proportion of replications in which the discrete GEV had the smaller SSD, and the median over replications of the ratio of the SSD of the continuous GEV to that of the discrete GEV.

Table 1. Comparison of the discrete GEV (4) and the continuous GEV (1) fitted to integer data ($n = 1000$, 1000 replications): proportion of replications in which (4) has the smaller sum of squared differences (Prop.), and median ratio of the sums of squared differences of (1) and (4) (Ratio).

		Floor		Ceiling		Rounding	
σ	β	Prop.	Ratio	Prop.	Ratio	Prop.	Ratio
1	−0.4	1.000	140.0	1.000	122.9	0.932	3.48
1	−0.2	1.000	87.5	1.000	89.0	0.785	1.80
1	0	1.000	75.6	1.000	68.6	0.812	2.11
1	0.2	1.000	82.0	1.000	77.4	0.822	2.43
1	0.4	1.000	88.6	1.000	93.2	0.997	5.47
5	−0.4	1.000	1.37	1.000	1.39	0.555	1.00
5	−0.2	1.000	1.27	0.999	1.29	0.516	1.00
5	0	0.997	1.28	0.998	1.29	0.489	1.00
5	0.2	1.000	1.36	0.998	1.36	0.492	1.00
5	0.4	1.000	1.51	0.999	1.51	0.557	1.00

Integer block maxima	Prop.	Ratio
Poisson (mean 10)	0.803	1.28
Negative binomial (mean 5, size 2)	0.522	1.00

Three conclusions emerge. First, under flooring or ceiling, the discrete GEV gives the smaller SSD in essentially every replication. The advantage is very large when the scale is comparable to the recording unit ($\sigma = 1$, median ratios 69–140) because the density-based

fit implicitly treats each integer as the midpoint of its interval and so misplaces the location by half a unit; it is modest when the scale is large relative to the unit ($\sigma = 5$, median ratios 1.27–1.51). Second, under rounding, which is the convention implicitly assumed by the density-based fit, the discrete GEV is still better for $\sigma = 1$ (in 79–100% of replications, median ratios 1.8–5.5), because the density at the midpoint is a poor approximation to the interval probability when the scale is small, but the two models are indistinguishable for $\sigma = 5$ (proportions near one half and median ratios of 1.00). Since the family (4) is the same under all three conventions (Section 2.1), the discrete GEV does not need to know which one was used. Third, for integer block maxima, which are not discretized GEV variables, the discrete GEV was better in 80% of replications for the Poisson maxima, whose values are concentrated on a few integers, and indistinguishable from the continuous GEV for the more dispersed negative binomial maxima. Thus, the conclusion that (4) fits integer data better than (1) holds under discretization mechanisms that are not rounding, or when the scale of the data is small relative to the recording unit; when the scale is large and the data are rounded, or more generally dispersed over many integers, the two models give practically the same fit. In no replication did the continuous GEV fit have $\hat{\beta} < -1$.

The two sensitivity checks described in steps (iii) and (iv) did not materially change any of the conclusions. For the integer block maxima settings, replacing the midpoint-binning convention used for the continuous GEV with the floor convention $\hat{F}(y+1) - \hat{F}(y)$ altered the proportions by at most 0.01 and the median SSD ratios by at most 0.03; the qualitative conclusions—that the discrete GEV is better in 80% of replications for Poisson maxima and indistinguishable from the continuous GEV for negative binomial maxima—were unchanged. Computing the SSD over the full integer range (from $\min_i y_i$ to $\max_i y_i$) rather than only over observed distinct values also left the relative comparisons essentially unchanged: at integers with zero empirical frequency, the fitted probabilities of both models are small, so their contributions to the SSD are negligible, and the proportions and median ratios agreed to within Monte Carlo error with those in Table 1.

4. Data Applications

We return to the two datasets mentioned in Section 1 and compare the fits of (4) and (1). We also fitted the discrete Gumbel distribution, the particular case of (4) for $\beta = 0$; the discrete Fréchet distribution, the particular case of (4) for $\beta > 0$; the discrete Weibull distribution, the particular case of (4) for $\beta < 0$; the discrete inverse Weibull distribution given by the probability mass function $p(x) = q^{(x+1)^{-a}} - q^{x^{-a}}$, $x = 0, 1, 2, \ldots$, for $0 < q < 1$ and $a > 0$; the discrete Lomax distribution given by the probability mass function $p(x) = \left(1 + \frac{x}{b}\right)^{-a} - \left(1 + \frac{x+1}{b}\right)^{-a}$, $x = 0, 1, 2, \ldots$, for $a > 0$ and $b > 0$. The discrete Gumbel, discrete Fréchet and discrete Weibull distributions are not independent competitors of the discrete GEV: they are nested in it, corresponding to $\beta = 0$, $\beta > 0$ and $\beta < 0$, respectively. Comparing them with the discrete GEV is therefore a comparison of tail regimes within one family. The discrete inverse Weibull distribution is also nested in the discrete GEV: its cumulative distribution function $q^{(x+1)^{-a}}$ is (5) with $\beta = \frac{1}{a}$, $\sigma = \frac{\lambda}{a}$ and $\mu = \lambda$, where $\lambda = (-\log q)^{\frac{1}{a}}$, that is, it is the discrete Fréchet distribution with lower endpoint $\mu - \frac{\sigma}{\beta} = 0$. Only the continuous GEV and the discrete Lomax distributions are genuinely different models, so the comparison below involves the discrete GEV, its four nested sub-models and two other models, rather than seven unrelated distributions.

Fitting the continuous GEV to integer data. The continuous GEV (1) was fitted, as is customary in applications, by maximizing the likelihood $\prod_{i=1}^{n} f(x_i)$ built from its density f evaluated at the observed integer values. We did not use the interval (grouped-data) likelihood $\prod_{i=1}^{n} [F(x_i + 1) - F(x_i)]$: under the observation model (2); this likelihood is identical to the discrete GEV likelihood (13), so its maximizer is the discrete GEV fit itself.

The comparison between (1) and (4) in this section is therefore a comparison between treating the integers as exact continuous observations and treating them as the integer parts of continuous observations. To compare the two on the same probability scale, all fitted probabilities for the continuous GEV were computed under (2), that is, $\hat{F}(x+1) - \hat{F}(x)$ for the probability of the integer x and $\hat{F}(x+1)$ for its cumulative probability.

The parameter estimates and standard errors computed via parametric bootstrapping (1000 samples simulated from the fitted model, each refitted by the same method) are shown in Table 2 for the wind speed data. For the continuous GEV, bootstrap samples were simulated from the fitted continuous GEV and floored, in accordance with (2). For the discrete inverse Weibull, we report the equivalent parametrization $P(x) = \exp\left\{-\left[\frac{x+1}{\lambda}\right]^{-a}\right\}$, $\lambda = (-\log q)^{\frac{1}{a}}$, because the maximum likelihood estimate of q is too close to zero to be represented in floating point. The data intervals and associated probabilities, computed as $\hat{F}(x+1) - \hat{F}(x)$, where $\hat{F}$ denotes the fitted distribution function of the continuous counterpart of each model, are given in Table 3. Also given in Table 3 are the sum of differences between the empirical and fitted probabilities and the sum of squares of the differences between the empirical and fitted probabilities. The probability plots comparing the fits are shown in Figure 3. The expected probabilities were computed as $\frac{\text{rank}(x_i)}{n+1}$ for $i = 1, 2, \ldots, n$, where $x_1, x_2, \ldots, x_n$ denote an enumeration of the data values and n the number of observations, with tied values receiving their average rank. The fitted cumulative probabilities for the fits of (1) and (4) were both computed by

$$\exp\left[-\left(1 + \hat{\beta}\frac{x_i + 1 - \hat{\mu}}{\hat{\sigma}}\right)^{-\frac{1}{\hat{\beta}}}\right],$$

with the respective estimates, so that both are evaluated on the same scale $\hat{F}(x_i + 1)$ implied by (2).

Table 2. Parameter estimates and parametric bootstrap standard errors for the wind speed data.

Model	Parameter Estimates (Standard Errors)
(1)	$\hat{\mu} = 96.032$ (2.800), $\hat{\sigma} = 12.852$ (1.885), $\hat{\beta} = -0.199$ (0.160)
(4)	$\hat{\mu} = 96.533$ (2.781), $\hat{\sigma} = 12.849$ (1.871), $\hat{\beta} = -0.199$ (0.156)
discrete Gumbel	$\hat{\mu} = 95.212$ (2.502), $\hat{\sigma} = 12.486$ (1.752)
discrete Fréchet	$\hat{\mu} = 95.212$ (2.481), $\hat{\sigma} = 12.486$ (1.713), $\hat{\beta} = 7.5 \times 10^{-15}$ (0.037)
discrete Weibull	$\hat{\mu} = 96.533$ (2.757), $\hat{\sigma} = 12.849$ (1.845), $\hat{\beta} = -0.199$ (0.146)
discrete inverse Weibull	$\hat{\lambda} = 94.363$ (2.556), $\hat{a} = 7.424$ (1.132)
discrete Lomax	$\hat{a}, \hat{b} \to \infty$, $\frac{\hat{a}}{\hat{b}} \to \hat{\lambda} = 0.00982$ (0.00198)

Table 2 shows that the continuous GEV fitted by its density and the discrete GEV give almost the same estimates of σ and β (12.852 versus 12.849, and -0.199 in both cases), while the location estimate of the continuous GEV is about half a unit lower (96.032 versus 96.533). This is what one expects: evaluating the density at an integer x approximates the probability of the interval $\left[x - \frac{1}{2}, x + \frac{1}{2}\right)$, whereas under (2), the integer x represents the interval $[x, x+1)$, so the density-based fit effectively assumes rounding rather than flooring (see the Remark in Section 2.1). When the scale is large relative to the unit of recording, as here ($\hat{\sigma} \approx 13$ km/h), the two fits are nearly equivalent; the continuous GEV fit is regular, with $\hat{\beta}$ well inside the region $\beta > -\frac{1}{2}$. The discrete GEV is nonetheless the coherent model for these data under (2): its likelihood is the exact probability of the recorded values, and its location parameter is the location of the continuous variable before flooring.

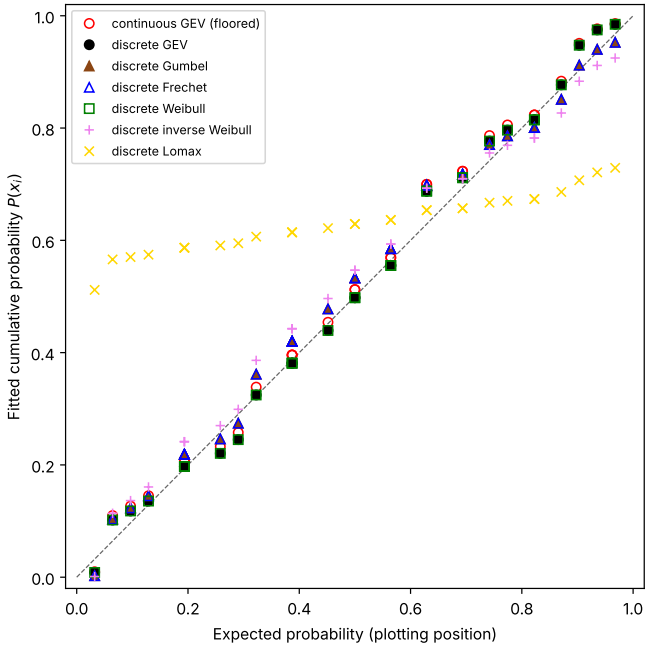


Figure 3. Probability plots comparing the fits of the continuous GEV (red), discrete GEV (black), discrete Gumbel (brown), discrete Fréchet (blue), discrete Weibull (green), discrete inverse Weibull (violet) and discrete Lomax (yellow) distributions for the wind speed data. All fitted cumulative probabilities are evaluated as $\widehat{P}(x_i)$, which equals $\hat{F}(x_i + 1)$ for the continuous GEV, the discrete GEV and its four nested sub-models.

Table 3. Data intervals and associated probabilities for the wind speed data.

Interval	Empirical	Continuous GEV	Discrete GEV	Discrete Gumbel	Discrete Fréchet	Discrete Weibull	Discrete Inverse Weibull	Discrete Lomax
[129, 130)	0.0333	0.0040	0.0043	0.0048	0.0048	0.0043	0.0050	0.0028
[117, 118)	0.0333	0.0134	0.0139	0.0114	0.0114	0.0139	0.0102	0.0031
[100, 101)	0.0667	0.0290	0.0291	0.0272	0.0272	0.0291	0.0247	0.0037
[132, 133)	0.0333	0.0027	0.0029	0.0038	0.0038	0.0029	0.0042	0.0027
[94, 95)	0.0333	0.0278	0.0274	0.0294	0.0294	0.0274	0.0289	0.0039
[108, 109)	0.0667	0.0233	0.0239	0.0196	0.0196	0.0239	0.0170	0.0034
[113, 114)	0.0667	0.0178	0.0184	0.0147	0.0147	0.0184	0.0129	0.0032
[96, 97)	0.1000	0.0288	0.0286	0.0293	0.0293	0.0286	0.0279	0.0038
[72, 73)	0.0333	0.0023	0.0021	0.0010	0.0010	0.0021	0.0006	0.0048
[98, 99)	0.0333	0.0292	0.0292	0.0285	0.0285	0.0292	0.0265	0.0037
[85, 86)	0.0333	0.0169	0.0161	0.0198	0.0198	0.0161	0.0226	0.0042
[124, 125)	0.0333	0.0071	0.0075	0.0070	0.0070	0.0075	0.0067	0.0029
[102, 103)	0.0667	0.0282	0.0285	0.0256	0.0256	0.0285	0.0228	0.0036
[112, 113)	0.0333	0.0189	0.0195	0.0156	0.0156	0.0195	0.0136	0.0033
[107, 108)	0.0667	0.0244	0.0248	0.0206	0.0206	0.0248	0.0179	0.0034
[86, 87)	0.0333	0.0184	0.0176	0.0216	0.0216	0.0176	0.0243	0.0042
[91, 92)	0.0333	0.0251	0.0245	0.0280	0.0280	0.0245	0.0290	0.0040
[89, 90)	0.1000	0.0226	0.0220	0.0261	0.0261	0.0220	0.0279	0.0041
[90, 91)	0.0333	0.0239	0.0233	0.0272	0.0272	0.0233	0.0286	0.0040
[84, 85)	0.0333	0.0154	0.0146	0.0179	0.0179	0.0146	0.0206	0.0043
[111, 112)	0.0333	0.0201	0.0206	0.0166	0.0166	0.0206	0.0144	0.0033
sum of differences with respect to empirical probabilities		0.601	0.601	0.604	0.604	0.601	0.614	0.924
sum of squared differences with respect to empirical probabilities		0.025	0.025	0.026	0.026	0.025	0.027	0.051

The discrete GEV gives $\hat{\beta} = -0.199$, placing the data in the Weibull-type (bounded above) domain, with implied upper endpoint

$$\hat{\mu} - \frac{\widehat{\sigma}}{\hat{\beta}} = 96.533 + \frac{12.849}{0.1987} \approx 161.2,$$

about 29 km/h above the sample maximum of 132 km/h. The endpoint is thus estimated from the shape of the whole sample rather than being pinned to the largest observation. However, with only $n = 30$ observations, the shape parameter is poorly determined: its bootstrap standard error is 0.156, so $\frac{|\hat{\beta}|}{\mathrm{se}(\hat{\beta})} \approx 1.3$, and the likelihood ratio statistic for the discrete Gumbel sub-model $\beta = 0$ is $2 \times (121.657 - 120.623) = 2.07$ ($p \approx 0.15$ on one degree of freedom). The Gumbel special case is therefore not rejected at the 5% level for the wind speed data, and the evidence for a bounded upper tail is weak.

The discrete Weibull fit is identical to the discrete GEV fit. This is expected: because the unconstrained maximizer has $\hat{\beta} < 0$, restricting β to be negative leaves the maximum unchanged. Conversely, the discrete Fréchet fit returns $\hat{\beta} = 7.5 \times 10^{-15}$, i.e., the estimate sits on the boundary $\beta \to 0^+$ of its constrained parameter space, so that the discrete Fréchet collapses onto the discrete Gumbel ($\hat{\mu} = 95.212$ and $\widehat{\sigma} = 12.486$ in both cases). Since the Fréchet shape estimate is on the boundary, its standard error is not informative. Among the four extreme-value-type discrete models, there are therefore only two distinct fits: the bounded discrete GEV/Weibull fit and the unbounded discrete Gumbel/Fréchet fit. The practical advantage of the discrete GEV is that the data determine the tail regime rather than the analyst having to choose between the Fréchet, Gumbel and Weibull forms in advance.

The discrete inverse Weibull fit, $\widehat{P}(x) = \exp\left\{-\left[\frac{x+1}{94.363}\right]^{-7.424}\right\}$, is a Fréchet-type distribution whose support starts at 0 (it is the law of the integer part of a Fréchet variable with lower endpoint 0); with a large shape parameter a, it can mimic a unimodal distribution concentrated around 90–100 km/h, and its negative log-likelihood (123.28) is only slightly larger than that of the discrete Gumbel (121.66). Since it is the discrete Fréchet distribution with its lower endpoint fixed at 0, it cannot fit better than the discrete Fréchet; the likelihood ratio statistic against the discrete GEV is $2 \times (123.277 - 120.623) = 5.31$ ($p \approx 0.021$ on one degree of freedom). For the discrete Lomax, the likelihood increases along a ridge on which a and b both tend to infinity with $\frac{a}{b}$ tending to a limit λ. Since $\left(1 + \frac{x}{b}\right)^{-a} \to \exp(-\lambda x)$ as $b \to \infty$ with $\frac{a}{b} = \lambda$ fixed, the maximum likelihood fit is the limiting geometric distribution, with $\hat{\lambda} = 0.00982$. Table 2 therefore reports $\hat{\lambda}$ and its bootstrap standard error rather than the arbitrarily large values of $\hat{a}$ and $\hat{b}$ at which an optimizer stops. The fitted geometric distribution has mean $\frac{1}{e^{\hat{\lambda}}-1} \approx 101.3$, which reproduces the sample mean (101.33) almost exactly. In other words, the discrete Lomax matches the first moment but, having its mode at zero and a standard deviation of about 102 (against a sample standard deviation of 13.9), cannot describe the location or spread of the data. The failure has a structural cause: the model has no location parameter and, unlike the discrete inverse Weibull, no shape parameter that can concentrate its mass, so it cannot place its probability in a narrow band far from the origin, which is exactly where annual maxima lie.

Two features of the data must be borne in mind when reading Table 3. First, with $n = 30$, each observation contributes $\frac{1}{30} \approx 0.0333$; five values occur twice (0.0667) and two values, 89 and 96, occur three times (0.1000). Second, the observations are spread over the 61 integers from 72 to 132, so any reasonable unimodal model must assign modest probability to each individual integer; the largest single-integer probability under the discrete GEV fit is only about 0.029 (near 98–100 km/h). Consequently, every fitted probability in Table 3 is smaller than the corresponding empirical frequency, all differences are positive, and the reported "sum of differences" equals one minus the total fitted probability placed on the 21 observed values.

Read this way, the row of sums is informative. The discrete GEV (and discrete Weibull) places $1 - 0.601 = 39.9\%$ of its mass on the 21 observed values, as does the continuous GEV, while the discrete Gumbel (and discrete Fréchet) places 39.6%, the discrete inverse Weibull 38.6% and the discrete Lomax only 7.6%. The sums of squared differences are 0.0252 for the discrete GEV, 0.0253 for the continuous GEV, 0.0259 for the discrete Gumbel, 0.0270 for the discrete inverse Weibull and 0.0508 for the discrete Lomax. For this short series, therefore, the discrete GEV, the continuous GEV, the discrete Gumbel and the discrete inverse Weibull describe the data about equally well, and only the discrete Lomax is clearly inadequate.

The difference between the discrete GEV and the discrete Gumbel is small in aggregate for this short series, but the cell-by-cell comparison shows a pattern. Relative to the discrete Gumbel, the discrete GEV

- assigns more probability to the upper-middle range 107–117 km/h (for example, 0.0239 versus 0.0196 at 108, 0.0184 versus 0.0147 at 113, and 0.0139 versus 0.0114 at 117), where the data contain several repeated values;
- assigns less probability to the two largest values (0.0043 versus 0.0048 at 129 and 0.0029 versus 0.0038 at 132);
- assigns less probability to the range 84–91 km/h but about twice as much to the isolated low value 72 km/h (0.0021 versus 0.0010).

This is precisely the pattern expected from a negative shape parameter: a shorter right tail and a longer left tail, i.e., negative skewness. Under the fitted models, the probability that all 30 annual maxima fall at or below the observed maximum of 132 km/h is about 0.63 for the discrete GEV but only about 0.23 for the discrete Gumbel, so the observed sample maximum is considerably more typical under the discrete GEV. Given the likelihood ratio test above, this difference should be regarded as suggestive rather than conclusive for $n = 30$.

In Figure 3, each point plots the fitted cumulative probability at an observation against its plotting position; a well-fitting model produces points close to the diagonal. Because the discrete GEV and discrete Weibull fits are identical, the black symbols are hidden beneath the green ones; likewise, the brown (discrete Gumbel) symbols are hidden beneath the blue (discrete Fréchet) ones. Once the continuous GEV is evaluated on the same scale $\hat{F}(x_i + 1)$, its points (red) lie on top of, or very close to, those of the discrete GEV. The discrete Gumbel/Fréchet and discrete inverse Weibull points also scatter closely about the diagonal; the maximum vertical deviations from the diagonal are 0.059 for the discrete GEV, 0.069 for the discrete Gumbel, 0.064 for the discrete inverse Weibull and 0.071 for the continuous GEV. Only the discrete Lomax points (yellow), which run from about 0.51 to 0.73, depart clearly from the diagonal: the whole observed range 72–132 km/h covers only about a fifth of its probability mass. The small steps in the sequences arise from tied observations. For $n = 30$, the probability plot does not separate the discrete GEV from the continuous GEV, the discrete Gumbel or the discrete inverse Weibull.

Tables 4 and 5 and Figure 4 repeat the same model comparison for the 80 years (1901–1980) of annual maximum Oxford temperature data. The pattern of Table 2 recurs in Table 4. The continuous GEV fitted by its density and the discrete GEV have almost identical scale and shape estimates (4.260 versus 4.250 and −0.287 in both cases), with the continuous location again about half a unit lower (83.839 versus 84.341).

The discrete GEV fit gives $\hat{\mu} = 84.341$, $\hat{\sigma} = 4.250$ and $\hat{\beta} = -0.287$ with a bootstrap standard error of 0.078, so that $\frac{|\hat{\beta}|}{\mathrm{se}(\hat{\beta})} \approx 3.7$, and the likelihood ratio statistic for the discrete Gumbel sub-model is $2 \times (234.818 - 228.900) = 11.84$ ($p \approx 0.0006$). Unlike the wind speed data, the temperature data therefore provide clear evidence for a bounded upper tail. The implied upper endpoint is approximately 99.2 °F, four degrees above the largest value recorded in 1901–1980. This endpoint is a feature of a stationary model fitted to that

period and should not be interpreted as a permanent physical ceiling, particularly in view of warming since 1980.

Table 4. Parameter estimates and parametric bootstrap standard errors for the temperature data.

Model	Parameter Estimates (Standard Errors)
(1)	$\hat{\mu} = 83.839\ (0.529)$, $\hat{\sigma} = 4.260\ (0.368)$, $\hat{\beta} = -0.287\ (0.077)$
(4)	$\hat{\mu} = 84.341\ (0.530)$, $\hat{\sigma} = 4.250\ (0.369)$, $\hat{\beta} = -0.287\ (0.078)$
discrete Gumbel	$\hat{\mu} = 83.707\ (0.490)$, $\hat{\sigma} = 4.136\ (0.358)$
discrete Fréchet	$\hat{\mu} = 83.706\ (0.499)$, $\hat{\sigma} = 4.135\ (0.368)$, $\hat{\beta} = 3.529 \times 10^{-10}\ (0.051)$
discrete Weibull	$\hat{\mu} = 84.341\ (0.530)$, $\hat{\sigma} = 4.250\ (0.369)$, $\hat{\beta} = -0.287\ (0.078)$
discrete inverse Weibull	$\hat{\lambda} = 83.601\ (0.478)$, $\hat{a} = 20.104\ (1.817)$
discrete Lomax	$\hat{a}, \hat{b} \to \infty$, $\frac{\hat{a}}{\hat{b}} \to \hat{\lambda} = 0.01165\ (0.00133)$

Table 5. Data intervals and associated probabilities for the temperature data.

Interval	Empirical	Continuous GEV	Discrete GEV	Discrete Gumbel	Discrete Fréchet	Discrete Weibull	Discrete Inverse Weibull	Discrete Lomax
$[89, 90)$	0.1125	0.0590	0.0648	0.0466	0.0466	0.0648	0.0443	0.0041
$[84, 85)$	0.1125	0.0891	0.0873	0.0873	0.0873	0.0873	0.0855	0.0044
$[85, 86)$	0.0875	0.0901	0.0903	0.0819	0.0819	0.0903	0.0791	0.0043
$[81, 82)$	0.0375	0.0638	0.0576	0.0747	0.0747	0.0576	0.0773	0.0045
$[92, 93)$	0.0250	0.0255	0.0303	0.0256	0.0256	0.0303	0.0250	0.0040
$[79, 80)$	0.0375	0.0386	0.0327	0.0421	0.0421	0.0327	0.0445	0.0046
$[95, 96)$	0.0250	0.0051	0.0073	0.0132	0.0132	0.0073	0.0136	0.0038
$[87, 88)$	0.1125	0.0799	0.0839	0.0648	0.0648	0.0839	0.0616	0.0042
$[83, 84)$	0.0750	0.0840	0.0802	0.0886	0.0886	0.0802	0.0883	0.0044
$[82, 83)$	0.0500	0.0752	0.0699	0.0846	0.0846	0.0699	0.0860	0.0045
$[86, 87)$	0.0875	0.0868	0.0891	0.0739	0.0739	0.0891	0.0707	0.0043
$[93, 94)$	0.0125	0.0168	0.0208	0.0207	0.0207	0.0208	0.0204	0.0039
$[80, 81)$	0.0625	0.0512	0.0447	0.0597	0.0598	0.0447	0.0628	0.0046
$[75, 76)$	0.0125	0.0064	0.0046	0.0013	0.0013	0.0046	0.0010	0.0048
$[88, 89)$	0.0375	0.0703	0.0755	0.0555	0.0555	0.0755	0.0526	0.0042
$[90, 91)$	0.0375	0.0472	0.0531	0.0386	0.0386	0.0531	0.0369	0.0041
$[91, 92)$	0.0250	0.0357	0.0413	0.0316	0.0316	0.0413	0.0304	0.0040
$[77, 78)$	0.0375	0.0181	0.0142	0.0125	0.0125	0.0142	0.0124	0.0047
$[94, 95)$	0.0125	0.0100	0.0131	0.0166	0.0165	0.0131	0.0167	0.0039
sum of differences with respect to empirical probabilities		0.047	0.039	0.080	0.080	0.039	0.091	0.919
sum of squared differences with respect to empirical probabilities		0.008	0.008	0.012	0.012	0.008	0.013	0.066

As before, the discrete Weibull fit coincides with the discrete GEV fit, and the discrete Fréchet fit collapses to the discrete Gumbel fit ($\hat{\beta} = 3.529 \times 10^{-10}$; $\hat{\mu} = 83.706$ versus 83.707; $\hat{\sigma} = 4.135$ versus 4.136). The discrete Gumbel location is about 0.6 °F lower than that of the discrete GEV, so its fitted mode is shifted to the left. The discrete inverse Weibull is again a Fréchet-type fit, with $\hat{\lambda} = 83.601$ and a very large shape $\hat{a} = 20.104$, and its negative log-likelihood (236.59) is slightly larger than that of the discrete Gumbel (234.82); the likelihood ratio statistic against the discrete GEV is $2 \times (236.593 - 228.900) = 15.39$ ($p < 0.001$). The discrete Lomax is again degenerate: it again collapses to a geometric law ($\hat{\lambda} = 0.01165$) whose mean, $\frac{1}{e^{0.01165}-1} \approx 85.33$, reproduces the sample mean (85.33 °F) while being unable to describe the concentration of the data between 75 and 95 °F.

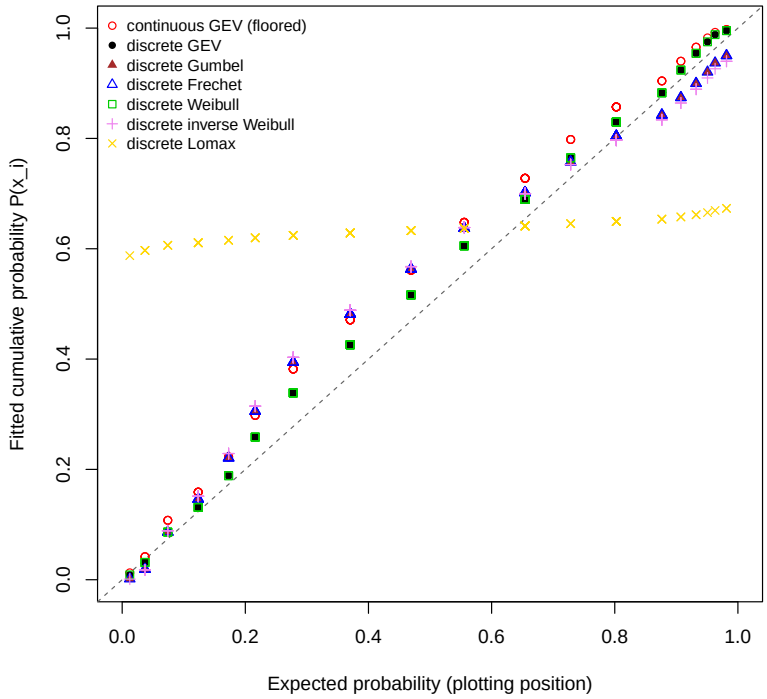


Figure 4. Probability plots comparing the fits of the continuous GEV (red), discrete GEV (black), discrete Gumbel (brown), discrete Fréchet (blue), discrete Weibull (green), discrete inverse Weibull (violet) and discrete Lomax (yellow) distributions for the temperature data. All fitted cumulative probabilities are evaluated as $\widehat{P}(x_i)$, which equals $\hat{F}(x_i + 1)$ for the continuous GEV, the discrete GEV and its four nested sub-models.

With $n = 80$, each observation contributes 0.0125, the 19 distinct values span only the 21 integers from 75 to 95 (only 76 and 78 are unobserved), and the fitted probabilities of the adequate models are comparable in size to the empirical frequencies. Hence, unlike in Table 3, the differences in Table 5 take both signs, and the "sum of differences" is a signed sum equal to one minus the total fitted probability on the observed values. The sum of squared differences is the more informative summary.

The discrete GEV places 96.1% of its mass on the 19 observed values, compared with 95.3% for the continuous GEV, 92.0% for the discrete Gumbel, 90.9% for the discrete inverse Weibull and 8.1% for the discrete Lomax. The sum of squared differences for the discrete GEV (0.0079) is about one third lower than that of the discrete Gumbel (0.012) and of the discrete inverse Weibull (0.013), and about one eighth of that of the discrete Lomax (0.066). The continuous GEV has a sum of squared differences of 0.0081, close to that of the discrete GEV. The improvement of the discrete GEV over the discrete Gumbel is therefore clearer here than for the wind speed data.

The cell-by-cell comparison explains where the improvement over the discrete Gumbel comes from:

- The empirical frequencies are largest at 84, 87 and 89 °F (each 0.1125). The discrete GEV assigns its largest probabilities to 84–87 °F (0.0873, 0.0903, 0.0891, 0.0839), whereas the discrete Gumbel peaks at 82–84 °F (0.0846, 0.0886, 0.0873) and substantially over-predicts 81 and 82 °F (0.0747 and 0.0846 against empirical 0.0375 and 0.0500). The discrete GEV reduces these over-predictions (0.0576 and 0.0699).
- At the coolest values, the discrete GEV is closer to the data: 0.0046 versus 0.0013 at 75 °F (empirical 0.0125) and 0.0142 versus 0.0125 at 77 °F (empirical 0.0375), consistent with the longer left tail implied by $\hat{\beta} < 0$.

- At 94 °F, the discrete GEV is almost exact (0.0131 against 0.0125). At 95 °F, observed twice, it predicts 0.0073 against the Gumbel's 0.0132 and the empirical 0.0250. This is the one cell where the bounded tail visibly costs the discrete GEV, and it is the natural price of a finite endpoint near 99 °F.
- The largest single residuals of the discrete GEV occur at 89 °F (under-prediction by 0.048) and 88 °F (over-prediction by 0.038). These largely cancel: the combined empirical frequency for 88–89 °F is 0.150 against a fitted 0.140. Such paired residuals of opposite sign in neighboring cells are typical of sampling variation and could also reflect a mild reading preference for particular values in the historical record.

Summarizing the upper tail in a single number: under the fitted models, the probability that none of the 80 annual maxima exceeds the observed maximum of 95 °F is about 0.70 for the discrete GEV but only about 0.017 for the discrete Gumbel. The observed record is thus entirely typical under the discrete GEV but would be surprising under the discrete Gumbel, whose unbounded upper tail places too much probability above 95 °F.

Figure 4 confirms these conclusions graphically and more sharply than Figure 3. With 80 observations but only 19 distinct values, the adequate fits appear as staircases: each tied value produces a horizontal run of points sharing a common fitted probability. As before, black is hidden beneath green and brown beneath blue.

The continuous GEV points (red), now evaluated on the same scale as the discrete models, lie close to the discrete GEV staircase, slightly above it because of the half-unit difference in location (maximum vertical deviation from the diagonal 0.104 against 0.061 for the discrete GEV). The discrete Lomax points (yellow) are confined to about 0.59–0.67 so that the observed range 75–95 °F corresponds to less than a tenth of its probability mass.

The discrete GEV/Weibull (green) and discrete Gumbel/Fréchet (blue) staircases both follow the diagonal, but they differ in the upper tail. For plotting positions above about 0.85, the blue points (and the violet discrete inverse Weibull points) fall systematically below the diagonal, reaching only about 0.93–0.95 at the largest observations, whereas the green points remain on the diagonal up to about 0.99. This is the graphical counterpart of the upper-tail calculation above: the discrete Gumbel assigns too much probability to temperatures beyond those observed, while the bounded discrete GEV does not. In the central part of the plot, the two staircases are interleaved about the diagonal, with neither dominating, in agreement with the mixed-sign residuals in Table 5.

Finally, formal tests of goodness of fit were performed. The Kolmogorov–Smirnov (Kolmogorov [43]; Smirnov [44]), Cramér–von Mises and Anderson–Darling statistics were designed for continuous distributions, and their classical null distributions are not valid for discrete data: with a discrete null distribution, the classical tests are conservative, and when parameters are estimated, the null distributions change further. We therefore modified both the statistics and their calibration.

Discrete statistics. Let $\widehat{P}$ be the fitted cumulative distribution function on the integers and $\hat{p}(j) = \widehat{P}(j) - \widehat{P}(j-1)$, let F_n be the empirical distribution function, and let $Z_j = F_n(j) - \widehat{P}(j)$. Following Choulakian et al. [45], we used

$$KS = \sqrt{n}\max_j \left|Z_j\right|, \qquad W^2 = n\sum_j Z_j^2\,\hat{p}(j), \qquad A^2 = n\sum_j \frac{Z_j^2\,\hat{p}(j)}{\widehat{P}(j)\left[1-\widehat{P}(j)\right]},$$

where the sums and the maximum run over all integers j in the support of the fitted distribution, not only over the observed values, and terms with $\widehat{P}(j) \in \{0, 1\}$ are omitted from A^2. In practice, an unbounded support was truncated where the remaining tail probability was below 10^{-12} (or at 2×10^5 units beyond the data for the two models with very heavy upper tails, where the omitted contribution is negligible). These are the discrete analogues of the Kolmogorov–Smirnov, Cramér–von Mises and Anderson–Darling statistics;

for a discrete distribution, $\sup_x$ of the continuous Kolmogorov–Smirnov statistic is attained at the integers, so KS coincides with it. We also used the Pearson chi-square statistic with cells $(-\infty, x_{(1)}]$, $\{x_{(1)}+1\}, \ldots, \{x_{(n)}-1\}$ and $[x_{(n)}, \infty)$, fixed from the observed data.

Calibration. Because the parameters are estimated, p values were obtained by a parametric bootstrap that respects both the estimation and the discretization:

1. compute the statistic T for the data $x_1, \ldots, x_n$ and the fitted model;
2. simulate $B = 1000$ samples of size n from the fitted model, denoted $\left\{x_1^{(j)}, \ldots, x_n^{(j)}\right\}$ for $j = 1, \ldots, B$; for the continuous GEV, the samples are simulated from the fitted continuous GEV and then floored, as in (2), so that they are integer-valued like the data;
3. refit the same model by the same method to each sample, obtaining $\hat{\theta}^{(j)}$ (for the continuous GEV, by maximizing its density-based likelihood at the floored values);
4. recompute the statistic, T_j, for each sample and its own fitted model;
5. compute the p value as $\frac{1+\sum_{j=1}^{B} I\{T_j \geq T\}}{B+1}$.

Since the bootstrap samples are drawn from the same discrete law as that assumed for the data, the resulting p values account for both the discreteness of the data and the estimation of the parameters; the classical tables for continuous distributions are not used at any stage. The smallest attainable p value is $\frac{1}{1001} \approx 0.001$.

The p values are given in Table 6. For the wind speed data, no model except the discrete Lomax is rejected at the 5% level by the three empirical distribution function tests; the discrete inverse Weibull is rejected only by the chi-square test ($p = 0.035$), and the discrete Lomax by all four tests. The discrete GEV and the continuous GEV have similar, large p values. For the temperature data, the discrete GEV, discrete Weibull and continuous GEV are not rejected by any test ($p \geq 0.56$), whereas the discrete Gumbel and discrete Fréchet are rejected at the 5% level by the Cramér–von Mises, Anderson–Darling and chi-square tests (and by the Kolmogorov–Smirnov test for the discrete Fréchet), and the discrete inverse Weibull and discrete Lomax are rejected by all four tests. These results agree with the likelihood ratio tests above: the data do not discriminate between the discrete GEV and the continuous GEV fitted by its density, they favor the bounded discrete GEV over the discrete Gumbel clearly for the temperature data but not for the shorter wind speed series, and they reject the discrete Lomax for both datasets.

Table 6. Parametric bootstrap p values ($B = 1000$) of the discrete goodness-of-fit tests.

	Wind Speed Data				Temperature Data			
Model	KS	W^2	A^2	χ^2	KS	W^2	A^2	χ^2
continuous GEV (floored)	0.814	0.904	0.822	0.476	0.697	0.568	0.856	0.587
discrete GEV	0.905	0.836	0.790	0.483	0.878	0.896	0.872	0.562
discrete Gumbel	0.776	0.828	0.716	0.216	0.051	0.012	0.012	0.044
discrete Fréchet	0.772	0.823	0.709	0.219	0.035	0.011	0.009	0.044
discrete Weibull	0.907	0.839	0.794	0.482	0.878	0.896	0.872	0.562
discrete inverse Weibull	0.656	0.426	0.240	0.035	0.019	0.006	0.001	0.012
discrete Lomax	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001

5. Conclusions

We have studied the discrete analogue of the generalized extreme value distribution introduced by Dkengne et al. [32], and we provided a systematic treatment of its distributional properties, estimation and finite-sample performance. It is constructed by differencing the survival function of the continuous GEV at consecutive integers, and it is the law of the integer part of a continuous GEV variable. The construction has three useful consequences. First, the cumulative distribution function of the discrete model is simply the continuous GEV distribution function evaluated at $x + 1$. This keeps the interpretation of the loca-

tion, scale and shape parameters and yields closed-form expressions for the probability mass function, the cumulative distribution function and the quantile function. Second, the discrete GEV unifies the discrete Gumbel ($\beta = 0$), discrete Fréchet ($\beta > 0$) and discrete (reversed) Weibull ($\beta < 0$) distributions within a single three-parameter family, so the data, rather than the analyst, determine the tail regime. Third, the family is the same whether the continuous variable is floored, ceiled or rounded; only the meaning of the location parameter changes.

We derived the support, including the cases of integer endpoints, the tail behavior, moments and the conditions for their existence, the quantile function, and Bowley skewness and Moors kurtosis. The quantile-based measures inherit a step-function character from the ceiling operator in the quantile function. For maximum likelihood estimation, we showed that the support is locally fixed away from integer endpoints, that the model is differentiable in quadratic mean except when $\beta \leq -\frac{1}{2}$ and the endpoint is an integer, and that the maximum likelihood estimators are then asymptotically normal under the conditions of Theorem 1, which include consistency. Unlike the continuous GEV with $\beta < -1$, the discrete likelihood cannot diverge, and it tends to $-\infty$ at the boundary of the region where it is finite; this does not by itself guarantee existence of a maximizer, which we checked for the two datasets based on profile likelihood.

The simulation studies showed that the mean squared errors of the maximum likelihood estimators decrease as the sample size increases, and that the biases decrease towards zero for $\beta = 0$ and $\beta = \pm 0.2$. Coverage probabilities of the bootstrap confidence intervals were close to the nominal level for moderate sample sizes, and the observed information matrix was positive definite in almost all replications. For $\beta = -1$, which lies outside the scope of Theorem 1, the bootstrap intervals could not be computed in some replications for most sample sizes, and the estimators of σ and β were noticeably biased downwards even for moderate n; occasional bootstrap failures also occurred for $\beta = -0.5$; inference should therefore be treated with caution when $\hat{\beta}$ is near or below -1. Estimation of the extreme quantile $Q(0.999)$ was most variable in the heavy-tailed case $\beta = 0.2$. In a simulation with 1000 replications under several observation mechanisms, the discrete GEV gave smaller squared deviations from the empirical probabilities than the continuous GEV fitted by its density in essentially all replications when the data were floored or ceiled, with a large advantage when the scale was comparable to the recording unit and a modest one otherwise; under rounding, and for dispersed integer block maxima, the two models were indistinguishable when the scale was large relative to the recording unit.

For both the Lisbon wind speed and the Oxford temperature data, the continuous GEV fitted by its density is regular and almost identical to the discrete GEV, apart from a shift of about half a unit in location, because the scale of the data is large relative to the recording unit. Under the flooring model, the discrete GEV is nevertheless the coherent choice since its likelihood is the exact probability of the recorded values. Within the discrete family, the bounded discrete GEV was clearly preferred to the discrete Gumbel for the temperature data (likelihood ratio $p \approx 0.0006$, rejection of the discrete Gumbel by three of four goodness-of-fit tests, and a more plausible sample maximum) but not for the shorter wind speed series, where the Gumbel sub-model was not rejected. The discrete Lomax, which has neither a location parameter nor a shape parameter able to concentrate its mass, was unable to describe annual maxima concentrated far from the origin. Goodness of fit was assessed with discrete versions of the Kolmogorov–Smirnov, Cramér–von Mises and Anderson–Darling statistics, calibrated by a parametric bootstrap that respects both the discreteness and the estimation of the parameters.

Several extensions deserve further study. These include non-stationary versions of the discrete GEV with covariate-dependent parameters, for example, to accommodate trends in

temperature extremes since 1980. Other directions are Bayesian and penalized likelihood estimation for small samples or shape parameters near $\beta = -\frac{1}{2}$ and discrete versions of the r-largest order statistics and point process models. Further work could also establish consistency of the maximum likelihood estimators in general, study the case $\beta \leq -\frac{1}{2}$ with an integer endpoint, and develop multivariate discrete extreme value models for jointly recorded integer-valued maxima.

Author Contributions: Conceptualization, A.A.U. and S.N.; Methodology, A.A.U. and S.N.; Software, A.A.U. and S.N.; Investigation, A.A.U. and S.N.; Writing—Original Draft Preparation, A.A.U. and S.N.; Writing—Review and Editing, A.A.U. and S.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: All relevant data are within the manuscript; both datasets are also distributed with the R package evd (Stephenson [36]).

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Further Mathematical Properties

Appendix A.1. Mode

Let x_0 be the real root of the following equation, which states that the GEV density takes equal values at x and $x+1$ (we assume, as is the case for the GEV density, that the difference of the two sides changes sign once, from positive to negative, and take $\beta \neq 0$); this root exists for $\beta > -1$. For $\beta \leq -1$, the GEV density increases up to the upper endpoint $b = \mu - \frac{\sigma}{\beta}$, the difference of the two sides changes sign at $x = b - 1$, and we take $x_0 = b - 1$

$$\left(1+\beta\frac{x+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}-1}\exp\left[-\left(1+\beta\frac{x+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right] = \left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}-1}\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right].$$

If $p(\lfloor x_0 \rfloor) > p(1+\lfloor x_0 \rfloor)$, then the mode of (4) is $\lfloor x_0 \rfloor$. If $p(\lfloor x_0 \rfloor) < p(1+\lfloor x_0 \rfloor)$, then the mode is $1+\lfloor x_0 \rfloor$; if the two probabilities are equal, both values are modes.

Appendix A.2. Hazard Rate Functions

The hazard rate function is defined by $h(x) = \frac{p(x)}{S(x)}$, where $S(x) = 1 - F(x) = \Pr(Y \geq x)$ for integer x. Therefore, it follows from (3) and (4) that

$$h(x) = \frac{\exp\left[-\left(1+\beta\frac{x+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right] - \exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]}{1-\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]}.$$

The reversed hazard rate function is defined by $r(x) = \frac{p(x)}{P(x)}$. Therefore, it follows from (4) and (5) that

$$r(x) = 1 - \exp\left[\left(1+\beta\frac{x+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}} - \left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right].$$

Appendix A.3. Further Moment Properties

The moment generating function corresponding to (4) is

$$E[\exp(tY)] = \sum_{j=-\infty}^{\infty}\exp(tj)\left\{\exp\left[-\left(1+\beta\frac{j+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right] - \exp\left[-\left(1+\beta\frac{j-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]\right\}.$$

Furthermore, the factorial moment corresponding to (4) is

$$E[Y(Y-1)\cdots(Y-r+1)]$$
$$= \sum_{j=-\infty}^{\infty} j(j-1)\cdots(j-r+1)\left\{\exp\left[-\left(1+\beta\frac{j+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right] - \exp\left[-\left(1+\beta\frac{j-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]\right\}.$$

The sums run over the support (6). The moment generating function is finite for $t \le 0$ if $\beta > 0$, for $t < \frac{1}{\sigma}$ if $\beta = 0$, for all real t if $-1 < \beta < 0$, for $t > -\frac{1}{\sigma}$ if $\beta = -1$, and for $t \ge 0$ if $\beta < -1$. These ranges follow from the tails of (4): for $\beta > 0$, $p(j) \sim cj^{-\frac{1}{\beta}-1}$ as $j \to \infty$; for $\beta = 0$, $p(j) \sim \left(1-\mathrm{e}^{-\frac{1}{\sigma}}\right)\exp\left(-\frac{j-\mu}{\sigma}\right)$ as $j \to \infty$; and for $\beta < 0$, $\log p(j)$ behaves like $-c|j|^{\frac{1}{|\beta|}}$ as $j \to -\infty$, for some constants $c > 0$. For the same reason, the factorial moment of order r is finite if and only if $r\beta < 1$ when $\beta > 0$, and is finite for every r when $\beta \le 0$.

Appendix A.4. Sum of Two Independent Variables

Suppose X_1 and X_2 are independent random variables with the probability mass function (4) specified by the parameters $(\mu_1, \sigma_1, \beta_1)$ and $(\mu_2, \sigma_2, \beta_2)$, respectively. Let $S = X_1 + X_2$. The probability mass function of S is

$$\begin{aligned} p_S(s) = &\sum_{j=-\infty}^{\infty} \exp\left[-\left(1+\beta_1\frac{s-j+1-\mu_1}{\sigma_1}\right)^{-\frac{1}{\beta_1}}\right]\exp\left[-\left(1+\beta_2\frac{j+1-\mu_2}{\sigma_2}\right)^{-\frac{1}{\beta_2}}\right] \\ &- \sum_{j=-\infty}^{\infty} \exp\left[-\left(1+\beta_1\frac{s-j+1-\mu_1}{\sigma_1}\right)^{-\frac{1}{\beta_1}}\right]\exp\left[-\left(1+\beta_2\frac{j-\mu_2}{\sigma_2}\right)^{-\frac{1}{\beta_2}}\right] \\ &- \sum_{j=-\infty}^{\infty} \exp\left[-\left(1+\beta_1\frac{s-j-\mu_1}{\sigma_1}\right)^{-\frac{1}{\beta_1}}\right]\exp\left[-\left(1+\beta_2\frac{j+1-\mu_2}{\sigma_2}\right)^{-\frac{1}{\beta_2}}\right] \\ &+ \sum_{j=-\infty}^{\infty} \exp\left[-\left(1+\beta_1\frac{s-j-\mu_1}{\sigma_1}\right)^{-\frac{1}{\beta_1}}\right]\exp\left[-\left(1+\beta_2\frac{j-\mu_2}{\sigma_2}\right)^{-\frac{1}{\beta_2}}\right]. \end{aligned}$$

The corresponding cumulative distribution function of S is

$$\begin{aligned} P_S(s) = &\sum_{k=-\infty}^{s}\sum_{j=-\infty}^{\infty} \exp\left[-\left(1+\beta_1\frac{k-j+1-\mu_1}{\sigma_1}\right)^{-\frac{1}{\beta_1}}\right]\exp\left[-\left(1+\beta_2\frac{j+1-\mu_2}{\sigma_2}\right)^{-\frac{1}{\beta_2}}\right] \\ &- \sum_{k=-\infty}^{s}\sum_{j=-\infty}^{\infty} \exp\left[-\left(1+\beta_1\frac{k-j+1-\mu_1}{\sigma_1}\right)^{-\frac{1}{\beta_1}}\right]\exp\left[-\left(1+\beta_2\frac{j-\mu_2}{\sigma_2}\right)^{-\frac{1}{\beta_2}}\right] \\ &- \sum_{k=-\infty}^{s}\sum_{j=-\infty}^{\infty} \exp\left[-\left(1+\beta_1\frac{k-j-\mu_1}{\sigma_1}\right)^{-\frac{1}{\beta_1}}\right]\exp\left[-\left(1+\beta_2\frac{j+1-\mu_2}{\sigma_2}\right)^{-\frac{1}{\beta_2}}\right] \\ &+ \sum_{k=-\infty}^{s}\sum_{j=-\infty}^{\infty} \exp\left[-\left(1+\beta_1\frac{k-j-\mu_1}{\sigma_1}\right)^{-\frac{1}{\beta_1}}\right]\exp\left[-\left(1+\beta_2\frac{j-\mu_2}{\sigma_2}\right)^{-\frac{1}{\beta_2}}\right]. \end{aligned}$$

The corresponding mean and variance of S are

$$E(S) = E(X_1) + E(X_2)$$

and

$$Var(S) = Var(X_1) + Var(X_2),$$

respectively, where $E(X_1)$, $E(X_2)$ are given by (8) and $Var(X_1)$, $Var(X_2)$ are given by (9), provided that $\beta_1, \beta_2 < 1$ for the mean and $\beta_1, \beta_2 < \frac{1}{2}$ for the variance. The family (4) is not closed under addition, so S does not in general have a discrete GEV distribution.

Appendix A.5. Shannon Entropy

Given a discrete random variable X with probability mass function $p(\cdot)$, Shannon entropy is defined by $E[-\log p(X)]$ (Shannon [46,47]). For the probability mass function given by (4),

$$E[-\log p(X)] = -\sum\left\{\exp\left[-\left(1+\beta\frac{x+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]-\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]\right\}$$
$$\cdot\log\left\{\exp\left[-\left(1+\beta\frac{x+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]-\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]\right\},$$

where the sum is over the domain given by (6). If $\beta = 0$, then

$$E[-\log p(X)] = -\sum_{x=-\infty}^{\infty}\left\{\exp\left[-\exp\left(-\frac{x+1-\mu}{\sigma}\right)\right]-\exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right]\right\}$$
$$\cdot\log\left\{\exp\left[-\exp\left(-\frac{x+1-\mu}{\sigma}\right)\right]-\exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right]\right\}$$

provided that the infinite sum exists.

Appendix A.6. Kullback–Leibler Divergence

Given two discrete random variables X_1 and X_2 with probability mass functions $p_1(\cdot)$ and $p_2(\cdot)$, respectively, Kullback–Leibler divergence (Kullback and Leibler [48]) is defined by $E_{X_1}\left[\log\frac{p_1(X)}{p_2(X)}\right]$. Suppose $p_1(\cdot)$ and $p_2(\cdot)$ are given by (4) with parameters (μ_1,σ_1,β_1) and (μ_2,σ_2,β_2), respectively. Then,

$$E_{X_1}\left[\log\frac{p_1(X)}{p_2(X)}\right] = \sum\left\{\exp\left[-\left(1+\beta_1\frac{x+1-\mu_1}{\sigma_1}\right)^{-\frac{1}{\beta_1}}\right]-\exp\left[-\left(1+\beta_1\frac{x-\mu_1}{\sigma_1}\right)^{-\frac{1}{\beta_1}}\right]\right\}$$
$$\cdot\log\left\{\frac{\exp\left[-\left(1+\beta_1\frac{x+1-\mu_1}{\sigma_1}\right)^{-\frac{1}{\beta_1}}\right]-\exp\left[-\left(1+\beta_1\frac{x-\mu_1}{\sigma_1}\right)^{-\frac{1}{\beta_1}}\right]}{\exp\left[-\left(1+\beta_2\frac{x+1-\mu_2}{\sigma_2}\right)^{-\frac{1}{\beta_2}}\right]-\exp\left[-\left(1+\beta_2\frac{x-\mu_2}{\sigma_2}\right)^{-\frac{1}{\beta_2}}\right]}\right\}$$

provided that $p_2(x) > 0$ for all x in the support of p_1, where the sum is over the domain given by (6) with (μ,σ,β) replaced by (μ_1,σ_1,β_1). If $\beta_1 = 0$ and $\beta_2 = 0$, then

$$E_{X_1}\left[\log\frac{p_1(X)}{p_2(X)}\right] = \sum_{x=-\infty}^{\infty}\left\{\exp\left[-\exp\left(-\frac{x+1-\mu_1}{\sigma_1}\right)\right]-\exp\left[-\exp\left(-\frac{x-\mu_1}{\sigma_1}\right)\right]\right\}$$
$$\cdot\log\left\{\frac{\exp\left[-\exp\left(-\frac{x+1-\mu_1}{\sigma_1}\right)\right]-\exp\left[-\exp\left(-\frac{x-\mu_1}{\sigma_1}\right)\right]}{\exp\left[-\exp\left(-\frac{x+1-\mu_2}{\sigma_2}\right)\right]-\exp\left[-\exp\left(-\frac{x-\mu_2}{\sigma_2}\right)\right]}\right\}$$

provided that the infinite sum exists.

Appendix A.7. Minimum and Maximum

Suppose $X_1, X_2, \ldots, X_N$ is a random sample from (4) (we write N for the sample size to avoid confusion with the probability mass function p). Let $m = \min(X_1, X_2, \ldots, X_N)$ and $M = \max(X_1, X_2, \ldots, X_N)$. The cumulative distribution functions of m and M are

$$1-\left\{1-\exp\left[-\left(1+\beta\frac{x+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]\right\}^N$$

and

$$\exp\left[-N\left(1+\beta\frac{x+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right],$$

respectively. The corresponding probability mass functions are

$$\left\{1-\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]\right\}^{N}-\left\{1-\exp\left[-\left(1+\beta\frac{x+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]\right\}^{N}$$

and

$$\exp\left[-N\left(1+\beta\frac{x+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]-\exp\left[-N\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right],$$

respectively.

Appendix A.8. Order Statistics

Suppose $X_1, X_2, \ldots, X_N$ is a random sample from (4). Let $X_{(r)}$ denote the rth smallest of $X_1, X_2, \ldots, X_N$. The cumulative distribution function and probability mass function of $X_{(r)}$ are

$$\sum_{k=r}^{N}\binom{N}{k}\exp\left[-k\left(1+\beta\frac{x+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]\left\{1-\exp\left[-\left(1+\beta\frac{x+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]\right\}^{N-k}$$

and

$$\sum_{k=r}^{N}\binom{N}{k}\exp\left[-k\left(1+\beta\frac{x+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]\left\{1-\exp\left[-\left(1+\beta\frac{x+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]\right\}^{N-k}$$
$$-\sum_{k=r}^{N}\binom{N}{k}\exp\left[-k\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]\left\{1-\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]\right\}^{N-k},$$

respectively.

Appendix B. R Code

The following R code (base R only) computes the probability mass function (4), the cumulative distribution function (5), the quantile function (10), random numbers, Bowley skewness (11) and Moors kurtosis (12), fits the model by maximizing (13) with optim, computes the observed information matrix J, and computes the discrete goodness-of-fit statistics (including the chi-square statistic) and their parametric bootstrap p values for the discrete GEV, as in Section 4. For example, `fit_dgev(x)` for the Oxford temperatures returns $\hat{\mu} = 84.34$, $\hat{\sigma} = 4.25$, $\hat{\beta} = -0.287$, as in Table 4.

```
# Discrete GEV distribution: Y = floor(X), X ~ GEV(mu, sigma, beta)

# GEV distribution function, extended to the whole real line
pgev_ext <- function(x, mu, sigma, beta) {
  z <- (x - mu) / sigma
  if (abs(beta) < 1e-12) return(exp(-exp(-z)))
  t <- beta * z
  out <- rep(if (beta > 0) 0 else 1, length(x))
  ok <- t > -1
  out[ok] <- exp(-exp(-log1p(t[ok]) / beta))
  out
}

# cumulative distribution function (5): P(x) = F(x + 1)
pdgev <- function(x, mu, sigma, beta) pgev_ext(floor(x) + 1, mu, sigma, beta)

# probability mass function (4): p(x) = F(x + 1) - F(x)
ddgev <- function(x, mu, sigma, beta)
  pgev_ext(x + 1, mu, sigma, beta) - pgev_ext(x, mu, sigma, beta)
```

```
# quantile function (10)
qdgev <- function(q, mu, sigma, beta) {
  v <- if (abs(beta) < 1e-12) mu - sigma * log(-log(q))
       else mu + sigma / beta * ((-log(q))^(-beta) - 1)
  ceiling(v - 1)
}

# random generation by inversion of the continuous GEV, then flooring
rdgev <- function(n, mu, sigma, beta) {
  u <- runif(n)
  x <- if (abs(beta) < 1e-12) mu - sigma * log(-log(u))
       else mu + sigma / beta * ((-log(u))^(-beta) - 1)
  floor(x)
}

# Bowley skewness (11) and Moors kurtosis (12)
bowley <- function(mu, sigma, beta) {
  Q <- function(q) qdgev(q, mu, sigma, beta)
  (Q(3/4) - 2 * Q(1/2) + Q(1/4)) / (Q(3/4) - Q(1/4))
}
moors <- function(mu, sigma, beta) {
  Q <- function(q) qdgev(q, mu, sigma, beta)
  (Q(3/8) - Q(1/8) + Q(7/8) - Q(5/8)) / (Q(3/4) - Q(1/4))
}

# negative log-likelihood (13); theta = (mu, log sigma, beta)
nll_dgev <- function(theta, x) {
  p <- ddgev(x, theta[1], exp(theta[2]), theta[3])
  if (any(!is.finite(p)) || any(p <= 0)) return(1e10)
  -sum(log(p))
}

# maximum likelihood fit by optim (Nelder-Mead), started at the Gumbel
# moment estimates and restarted from perturbed values of sigma and beta
fit_dgev <- function(x, nstart = 10) {
  s0 <- sd(x) * sqrt(6) / pi
  start <- c(mean(x) - 0.5772 * s0, log(s0), 0)
  best <- optim(start, nll_dgev, x = x, method = "Nelder-Mead",
                control = list(maxit = 10,000))
  for (k in seq_len(nstart)) {
    st <- c(start[1], log(s0 + runif(1, 1, 20)), runif(1, -0.2, 0.2))
    if (nll_dgev(st, x) >= 1e10) next
    r <- optim(st, nll_dgev, x = x, method = "Nelder-Mead",
               control = list(maxit = 10,000))
    if (r$value < best$value) best <- r
  }
  est <- c(mu = best$par[1], sigma = exp(best$par[2]), beta = best$par[3])
  # observed information matrix J in the parametrization (mu, sigma, beta)
  f <- function(p) nll_dgev(c(p[1], log(p[2]), p[3]), x)
  J <- optimHess(est, f)
  list(estimate = est, nll = best$value, J = J,
       se_asymptotic = tryCatch(sqrt(diag(solve(J))),
                                error = function(e) rep(NA, 3)))
}

# discrete Kolmogorov-Smirnov, Cramer-von Mises and Anderson-Darling
# statistics (Choulakian et al., 1994) for a fitted cdf Pfun on the integers,
# and the Pearson chi-square statistic with cells (-Inf, c1], {c1 + 1}, ...,
# {c2 - 1}, [c2, Inf), where (c1, c2) are fixed from the observed data
gof_stats <- function(x, Pfun, lo, hi, cells) {
  j <- lo:hi
  Pj <- Pfun(j)
  pj <- Pj - Pfun(j - 1)
  Fn <- ecdf(x)(j)
  Z <- Fn - Pj
  n <- length(x)
```

```
  ok <- Pj > 0 & Pj < 1
  c(KS = sqrt(n) * max(abs(Z)),
    W2 = n * sum(Z^2 * pj),
    A2 = n * sum(Z[ok]^2 * pj[ok] / (Pj[ok] * (1 - Pj[ok]))),
    chi2 = {
      k <- cells[1]:(cells[2] - 1)
      E <- n * c(Pfun(cells[1]), diff(Pfun(k)), 1 - Pfun(cells[2] - 1))
      O <- c(sum(x <= cells[1]), tabulate(match(x, k[-1]), length(k) - 1),
             sum(x >= cells[2]))
      sum(ifelse(E > 0, (O - E)^2 / E, 0))
    })
}

# integer range carrying all but 1e-12 of the probability
support_range <- function(x, Pfun) {
  lo <- min(x) - 1
  while (Pfun(lo - 1) > 1e-12) lo <- lo - 10
  hi <- max(x) + 1
  while (1 - Pfun(hi) > 1e-12) hi <- hi + 10
  c(lo, hi)
}

# parametric bootstrap p values for the discrete GEV
gof_dgev <- function(x, B = 1000) {
  fit <- fit_dgev(x, nstart = 0)
  e <- fit$estimate
  P0 <- function(j) pdgev(j, e[1], e[2], e[3])
  r0 <- support_range(x, P0)
  cells <- range(x)
  T0 <- gof_stats(x, P0, r0[1], r0[2], cells)
  Tb <- matrix(NA, B, 4)
  for (b in 1:B) {
    xb <- rdgev(length(x), e[1], e[2], e[3])
    fb <- fit_dgev(xb, nstart = 0)
    eb <- fb$estimate
    Pb <- function(j) pdgev(j, eb[1], eb[2], eb[3])
    rb <- support_range(xb, Pb)
    Tb[b, ] <- gof_stats(xb, Pb, rb[1], rb[2], cells)
  }
  (1 + colSums(sweep(Tb, 2, T0, ">="))) / (B + 1)
}
```

Appendix C. Elements of the J Matrix

The elements are

$$J_{1,1} = \sum_{i=1}^{n} \left\{ \frac{a_1(x_i;\mu,\sigma,\beta) - a_1(x_i+1;\mu,\sigma,\beta)}{\exp\left[-\left(1+\beta\frac{x_i+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right] - \exp\left[-\left(1+\beta\frac{x_i-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]} \right\}^2 - \sum_{i=1}^{n} \frac{b_1(x_i;\mu,\sigma,\beta) - b_1(x_i+1;\mu,\sigma,\beta)}{\exp\left[-\left(1+\beta\frac{x_i+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right] - \exp\left[-\left(1+\beta\frac{x_i-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]},$$

$$J_{1,2} = \sum_{i=1}^{n} \frac{[a_1(x_i;\mu,\sigma,\beta) - a_1(x_i+1;\mu,\sigma,\beta)][a_2(x_i;\mu,\sigma,\beta) - a_2(x_i+1;\mu,\sigma,\beta)]}{\left\{\exp\left[-\left(1+\beta\frac{x_i+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right] - \exp\left[-\left(1+\beta\frac{x_i-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]\right\}^2} - \sum_{i=1}^{n} \frac{b_2(x_i;\mu,\sigma,\beta) - b_2(x_i+1;\mu,\sigma,\beta)}{\exp\left[-\left(1+\beta\frac{x_i+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right] - \exp\left[-\left(1+\beta\frac{x_i-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]},$$

$$J_{1,3} = \sum_{i=1}^{n} \frac{[a_1(x_i;\mu,\sigma,\beta) - a_1(x_i+1;\mu,\sigma,\beta)][a_3(x_i;\mu,\sigma,\beta) - a_3(x_i+1;\mu,\sigma,\beta)]}{\left\{\exp\left[-\left(1+\beta\frac{x_i+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right] - \exp\left[-\left(1+\beta\frac{x_i-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]\right\}^2} - \sum_{i=1}^{n} \frac{b_3(x_i;\mu,\sigma,\beta) - b_3(x_i+1;\mu,\sigma,\beta)}{\exp\left[-\left(1+\beta\frac{x_i+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right] - \exp\left[-\left(1+\beta\frac{x_i-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]},$$

$$J_{2,2} = \sum_{i=1}^{n} \left\{\frac{a_2(x_i;\mu,\sigma,\beta) - a_2(x_i+1;\mu,\sigma,\beta)}{\exp\left[-\left(1+\beta\frac{x_i+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right] - \exp\left[-\left(1+\beta\frac{x_i-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]}\right\}^2 - \sum_{i=1}^{n} \frac{b_4(x_i;\mu,\sigma,\beta) - b_4(x_i+1;\mu,\sigma,\beta)}{\exp\left[-\left(1+\beta\frac{x_i+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right] - \exp\left[-\left(1+\beta\frac{x_i-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]},$$

$$J_{2,3} = \sum_{i=1}^{n} \frac{[a_2(x_i;\mu,\sigma,\beta) - a_2(x_i+1;\mu,\sigma,\beta)][a_3(x_i;\mu,\sigma,\beta) - a_3(x_i+1;\mu,\sigma,\beta)]}{\left\{\exp\left[-\left(1+\beta\frac{x_i+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right] - \exp\left[-\left(1+\beta\frac{x_i-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]\right\}^2} - \sum_{i=1}^{n} \frac{b_5(x_i;\mu,\sigma,\beta) - b_5(x_i+1;\mu,\sigma,\beta)}{\exp\left[-\left(1+\beta\frac{x_i+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right] - \exp\left[-\left(1+\beta\frac{x_i-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]}$$

and

$$J_{3,3} = \sum_{i=1}^{n} \left\{\frac{a_3(x_i;\mu,\sigma,\beta) - a_3(x_i+1;\mu,\sigma,\beta)}{\exp\left[-\left(1+\beta\frac{x_i+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right] - \exp\left[-\left(1+\beta\frac{x_i-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]}\right\}^2 - \sum_{i=1}^{n} \frac{b_6(x_i;\mu,\sigma,\beta) - b_6(x_i+1;\mu,\sigma,\beta)}{\exp\left[-\left(1+\beta\frac{x_i+1-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right] - \exp\left[-\left(1+\beta\frac{x_i-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]},$$

where

$$b_1(x;\mu,\sigma,\beta) = \frac{\beta+1}{\sigma^2}\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}-2}\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right] - \frac{1}{\sigma^2}\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{2}{\beta}-2}\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right],$$

$$b_2(x;\mu,\sigma,\beta) = -\frac{1}{\sigma^2}\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}-1}\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right] + \frac{(\beta+1)(x-\mu)}{\sigma^3}\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}-2}\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right] - \frac{x-\mu}{\sigma^3}\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{2}{\beta}-2}\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right],$$

$$b_3(x;\mu,\sigma,\beta) = \frac{1}{\sigma\beta^2}\log\left(1+\beta\frac{x-\mu}{\sigma}\right)\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}-1}\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]$$
$$-\frac{(\beta+1)(x-\mu)}{\beta\sigma^2}\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}-2}\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]$$
$$-\frac{1}{\sigma\beta^2}\log\left(1+\beta\frac{x-\mu}{\sigma}\right)\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{2}{\beta}-1}\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]$$
$$+\frac{x-\mu}{\beta\sigma^2}\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{2}{\beta}-2}\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right],$$

$$b_4(x;\mu,\sigma,\beta) = -\frac{2(x-\mu)}{\sigma^3}\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}-1}\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]$$
$$+\frac{(\beta+1)(x-\mu)^2}{\sigma^4}\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}-2}\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]$$
$$-\frac{(x-\mu)^2}{\sigma^4}\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{2}{\beta}-2}\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right],$$

$$b_5(x;\mu,\sigma,\beta) = \frac{x-\mu}{\sigma^2\beta^2}\log\left(1+\beta\frac{x-\mu}{\sigma}\right)\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}-1}\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]$$
$$-\frac{(\beta+1)(x-\mu)^2}{\beta\sigma^3}\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}-2}\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]$$
$$-\frac{x-\mu}{\sigma^2\beta^2}\log\left(1+\beta\frac{x-\mu}{\sigma}\right)\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{2}{\beta}-1}\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]$$
$$+\frac{(x-\mu)^2}{\beta\sigma^3}\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{2}{\beta}-2}\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]$$

and

$$b_6(x;\mu,\sigma,\beta) = -\frac{2}{\beta^3}\log\left(1+\beta\frac{x-\mu}{\sigma}\right)\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]$$
$$+\frac{1}{\beta^4}\left[\log\left(1+\beta\frac{x-\mu}{\sigma}\right)\right]^2\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]$$
$$+\frac{2(x-\mu)}{\beta^2\sigma}\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}-1}\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]$$
$$-\frac{1}{\beta^4}\left[\log\left(1+\beta\frac{x-\mu}{\sigma}\right)\right]^2\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{2}{\beta}}\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]$$
$$+\frac{2(x-\mu)}{\beta^3\sigma}\log\left(1+\beta\frac{x-\mu}{\sigma}\right)\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{2}{\beta}-1}\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]$$
$$-\frac{2(x-\mu)}{\beta^3\sigma}\log\left(1+\beta\frac{x-\mu}{\sigma}\right)\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}-1}\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]$$
$$+\frac{(\beta+1)(x-\mu)^2}{\beta^2\sigma^2}\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}-2}\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right]$$
$$-\frac{(x-\mu)^2}{\beta^2\sigma^2}\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{2}{\beta}-2}\exp\left[-\left(1+\beta\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\beta}}\right].$$

Appendix D. Simulation Scheme

We used the following scheme in Section 3.1:

(i) simulate a random sample of size n from (4) for given values of μ, σ and β (we took $\mu = 0$, $\sigma = 1$ and $\beta = 0, 0.2, -0.2, -0.5, -1$), using the inversion method;

(ii) compute the maximum likelihood estimates $\hat{\mu}$, $\hat{\sigma}$, $\hat{\beta}$, $\widehat{Q}(0.999)$ by directly maximizing (13) and the associated 95 percent confidence intervals $[L(\hat{\mu}), U(\hat{\mu})]$, $[L(\hat{\sigma}), U(\hat{\sigma})]$, $\left[L\left(\hat{\beta}\right), U\left(\hat{\beta}\right)\right]$, $\left[L\left(\widehat{Q}(0.999)\right), U\left(\widehat{Q}(0.999)\right)\right]$ by parametric bootstrapping;

(iii) repeat steps i) and ii) ten thousand times, resulting in the estimates $\hat{\mu}_i$, $\hat{\sigma}_i$, $\hat{\beta}_i$, $\widehat{Q}_i(0.999)$ and the associated 95 percent confidence intervals $[L(\hat{\mu}_i), U(\hat{\mu}_i)]$, $[L(\hat{\sigma}_i), U(\hat{\sigma}_i)]$, $\left[L\left(\hat{\beta}_i\right), U\left(\hat{\beta}_i\right)\right]$, $\left[L\left(\widehat{Q}_i(0.999)\right), U\left(\widehat{Q}_i(0.999)\right)\right]$ for $i = 1, 2, \ldots, 10{,}000$;

(iv) compute the observed information matrix J of Section 2.4 (with elements given in Appendix C) for each of the ten thousand samples, resulting in $J_{n,i}$ for $i = 1, 2, \ldots, 10{,}000$;

(v) compute the biases as

$$\text{bias}(\hat{\mu}) = \frac{1}{10{,}000} \sum_{i=1}^{10{,}000} (\hat{\mu}_i - \mu),$$

$$\text{bias}(\hat{\sigma}) = \frac{1}{10{,}000} \sum_{i=1}^{10{,}000} (\hat{\sigma}_i - \sigma),$$

$$\text{bias}\left(\hat{\beta}\right) = \frac{1}{10{,}000} \sum_{i=1}^{10{,}000} \left(\hat{\beta}_i - \beta\right)$$

and

$$\text{bias}\left(\widehat{Q}(0.999)\right) = \frac{1}{10{,}000} \sum_{i=1}^{10{,}000} \left(\widehat{Q}_i(0.999) - Q(0.999)\right);$$

(vi) compute the mean squared errors as

$$\text{mse}(\hat{\mu}) = \frac{1}{10{,}000} \sum_{i=1}^{10{,}000} (\hat{\mu}_i - \mu)^2,$$

$$\text{mse}(\hat{\sigma}) = \frac{1}{10{,}000} \sum_{i=1}^{10{,}000} (\hat{\sigma}_i - \sigma)^2,$$

$$\text{mse}\left(\hat{\beta}\right) = \frac{1}{10{,}000} \sum_{i=1}^{10{,}000} \left(\hat{\beta}_i - \beta\right)^2$$

and

$$\text{mse}\left(\widehat{Q}(0.999)\right) = \frac{1}{10{,}000} \sum_{i=1}^{10{,}000} \left(\widehat{Q}_i(0.999) - Q(0.999)\right)^2;$$

(vii) compute the coverage lengths as

$$\text{cov len}(\hat{\mu}) = \frac{1}{10{,}000} \sum_{i=1}^{10{,}000} [U(\hat{\mu}_i) - L(\hat{\mu}_i)],$$

$$\text{cov len}(\hat{\sigma}) = \frac{1}{10{,}000} \sum_{i=1}^{10{,}000} [U(\hat{\sigma}_i) - L(\hat{\sigma}_i)],$$

$$\text{cov len}\left(\hat{\beta}\right) = \frac{1}{10{,}000} \sum_{i=1}^{10{,}000} \left[U\left(\hat{\beta}_i\right) - L\left(\hat{\beta}_i\right)\right]$$

and

$$\text{cov len}\left(\widehat{Q}(0.999)\right) = \frac{1}{10{,}000} \sum_{i=1}^{10{,}000} \left[U\left(\widehat{Q}_i(0.999)\right) - L\left(\widehat{Q}_i(0.999)\right)\right];$$

if the bootstrap did not give a finite interval in some replication, the average length is not defined for that n; this occurred for most values of n when $\beta = -1$, which is why the curves in Figure A19 are interrupted, and for a few values of n when $\beta = -0.5$, which explains the short gaps in Figure A15;

(viii) compute the coverage probabilities as

$$\text{cov prob}(\hat{\mu}) = \frac{1}{10{,}000}\sum_{i=1}^{10{,}000} I\{L(\hat{\mu}_i) < \mu < U(\hat{\mu}_i)\},$$

$$\text{cov prob}(\hat{\sigma}) = \frac{1}{10{,}000}\sum_{i=1}^{10{,}000} I\{L(\hat{\sigma}_i) < \sigma < U(\hat{\sigma}_i)\},$$

$$\text{cov prob}\big(\hat{\beta}\big) = \frac{1}{10{,}000}\sum_{i=1}^{10{,}000} I\big\{L\big(\hat{\beta}_i\big) < \beta < U\big(\hat{\beta}_i\big)\big\}$$

and

$$\text{cov prob}\big(\widehat{Q}(0.999)\big) = \frac{1}{10{,}000}\sum_{i=1}^{10{,}000} I\big\{L\big(\widehat{Q}_i(0.999)\big) < Q(0.999) < U\big(\widehat{Q}_i(0.999)\big)\big\},$$

where $I\{\cdot\}$ denotes the indicator function;

(ix) repeat steps (i) to (viii) for $n = 40, 41, \ldots, 200$;

(x) compute the proportion of observed information matrices that are positive definite as

$$\text{prop} = \frac{1}{1610000}\sum_{n=40}^{200}\sum_{i=1}^{10000} I\big\{\text{all eigenvalues of } J_{n,i} \text{ are positive}\big\}. \tag{A1}$$

The is.positive.definite function in the R package matrixcalc was used;

(xi) compute the proportion of cases when Kreutz [42]'s test for identifiability was not rejected.

Appendix E. Figures for Simulation Studies

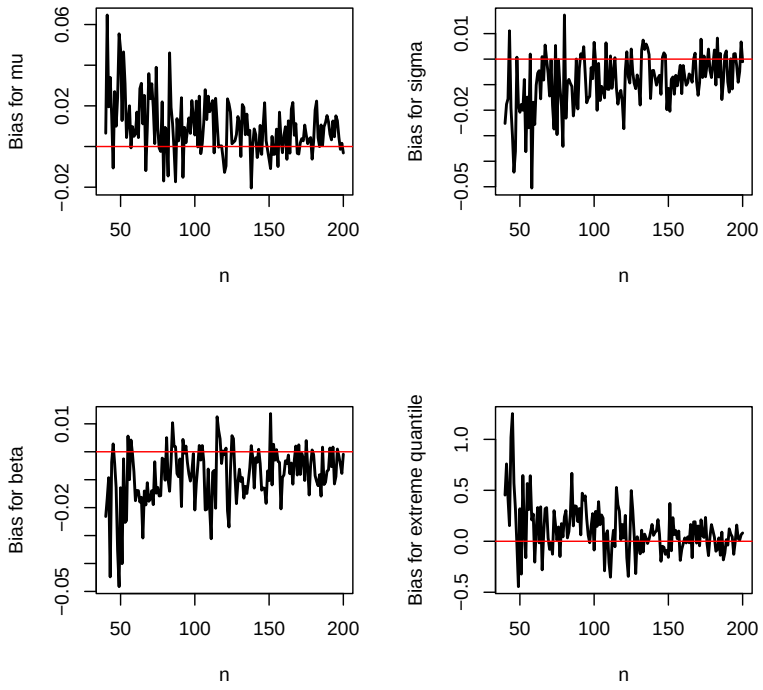


Figure A1. Biases of $\hat{\mu}$, $\hat{\sigma}$, $\hat{\beta}$ and $\widehat{Q}(0.999)$ versus n when $\sigma = 1$ and $\beta = 0$.

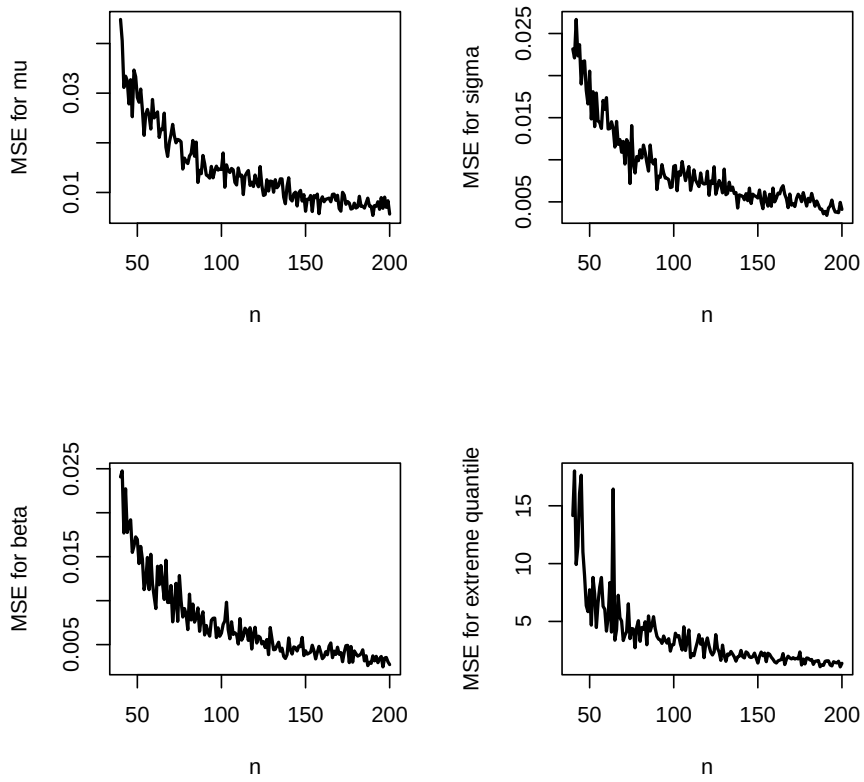


Figure A2. Mean squared errors of $\hat{\mu}$, $\widehat{\sigma}$, $\hat{\beta}$ and $\widehat{Q}(0.999)$ versus n when $\sigma = 1$ and $\beta = 0$.

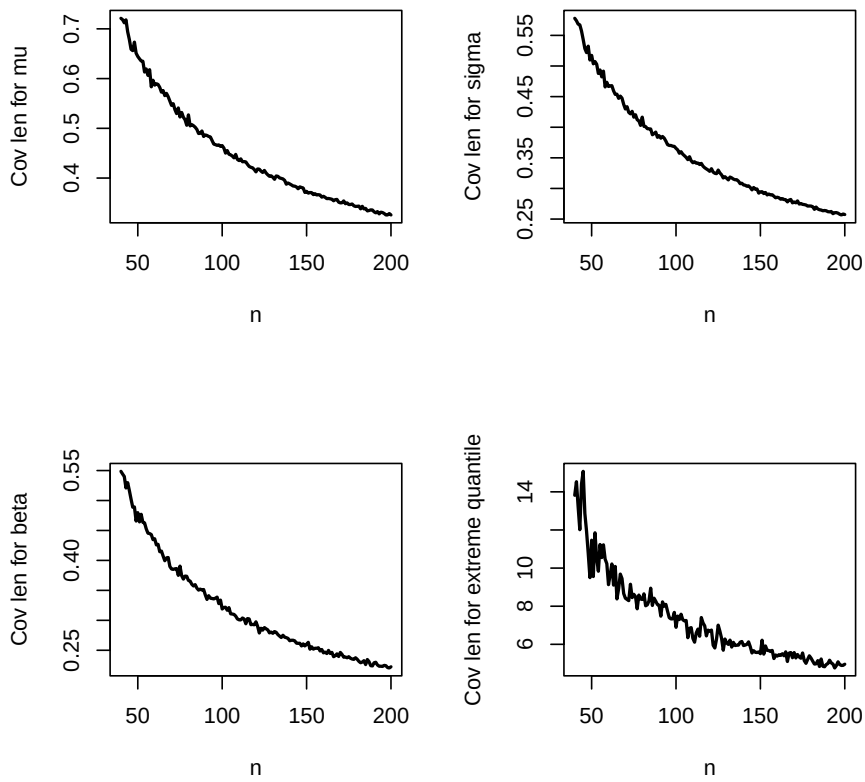


Figure A3. Coverage lengths for $\hat{\mu}$, $\widehat{\sigma}$, $\hat{\beta}$ and $\widehat{Q}(0.999)$ versus n when $\sigma = 1$ and $\beta = 0$.

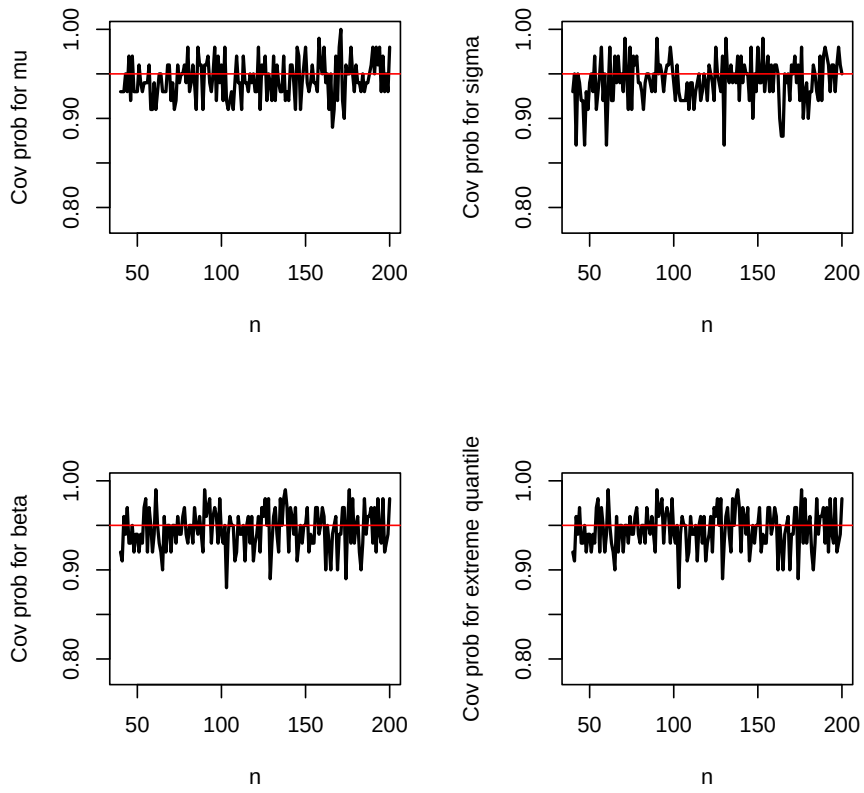


Figure A4. Coverage probabilities for $\hat{\mu}$, $\hat{\sigma}$, $\hat{\beta}$ and $\widehat{Q}(0.999)$ versus n when $\sigma = 1$ and $\beta = 0$.

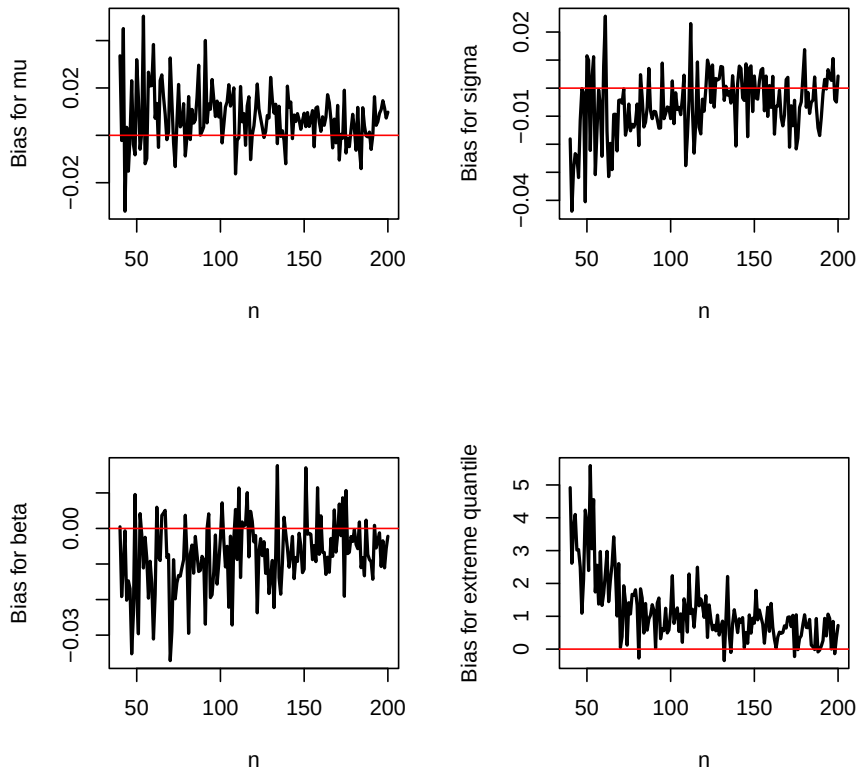


Figure A5. Biases of $\hat{\mu}$, $\hat{\sigma}$, $\hat{\beta}$ and $\widehat{Q}(0.999)$ versus n when $\sigma = 1$ and $\beta = 0.2$.

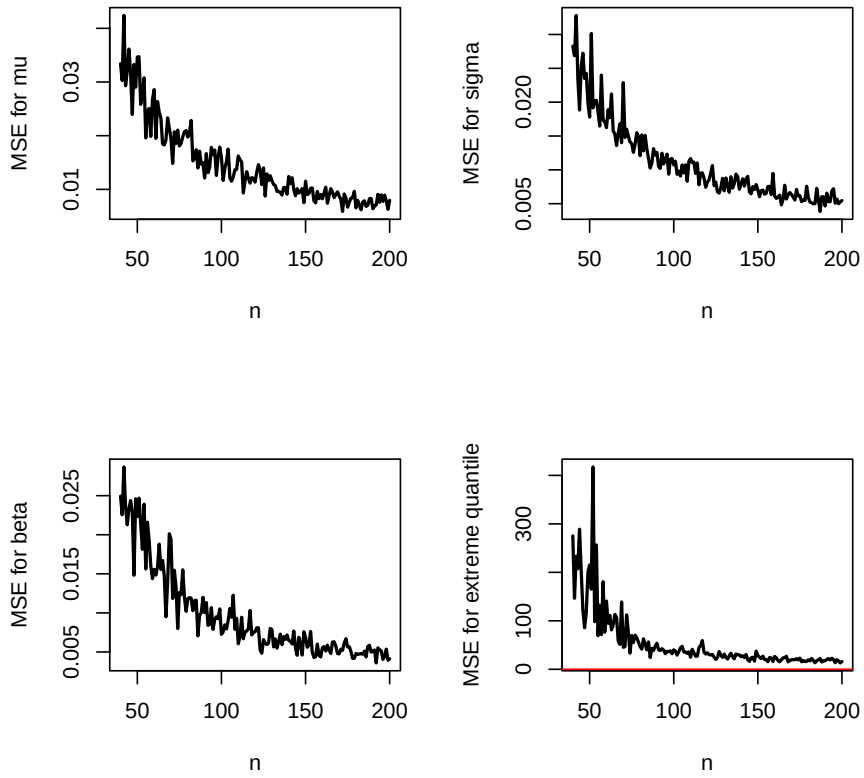


Figure A6. Mean squared errors of $\hat{\mu}$, $\widehat{\sigma}$, $\hat{\beta}$ and $\widehat{Q}(0.999)$ versus n when $\sigma = 1$ and $\beta = 0.2$.

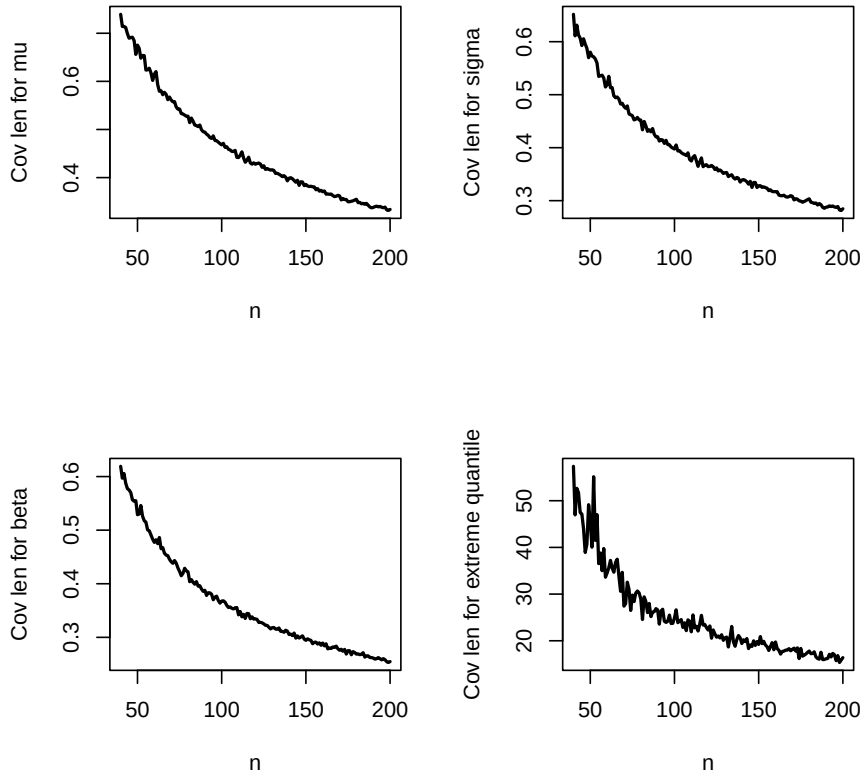


Figure A7. Coverage lengths for $\hat{\mu}$, $\widehat{\sigma}$, $\hat{\beta}$ and $\widehat{Q}(0.999)$ versus n when $\sigma = 1$ and $\beta = 0.2$.

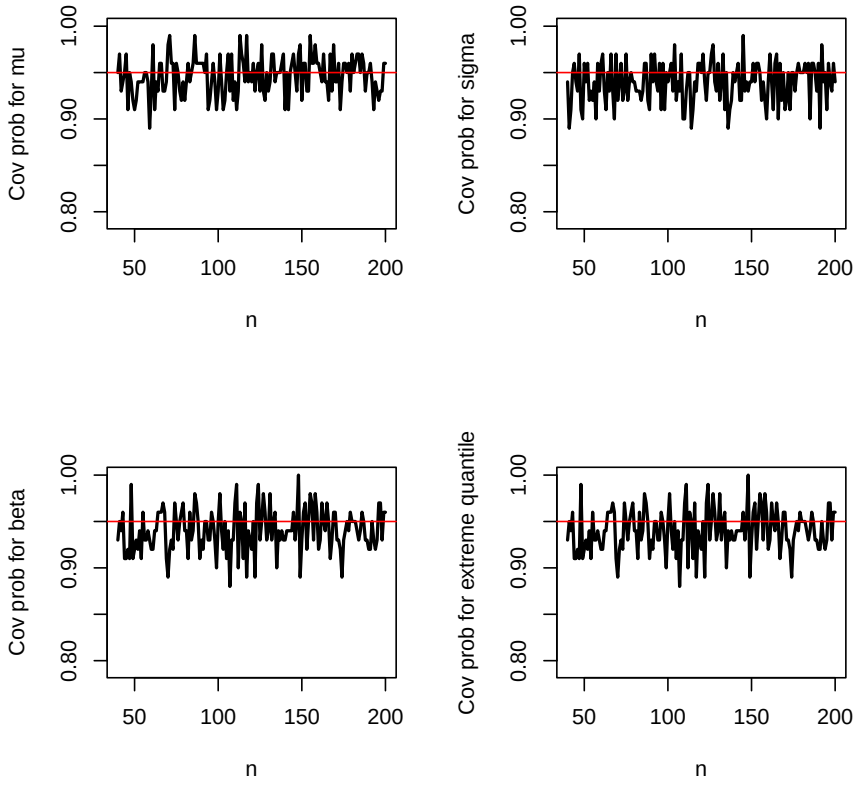


Figure A8. Coverage probabilities for $\hat{\mu}$, $\widehat{\sigma}$, $\hat{\beta}$ and $\widehat{Q}(0.999)$ versus n when $\sigma = 1$ and $\beta = 0.2$.

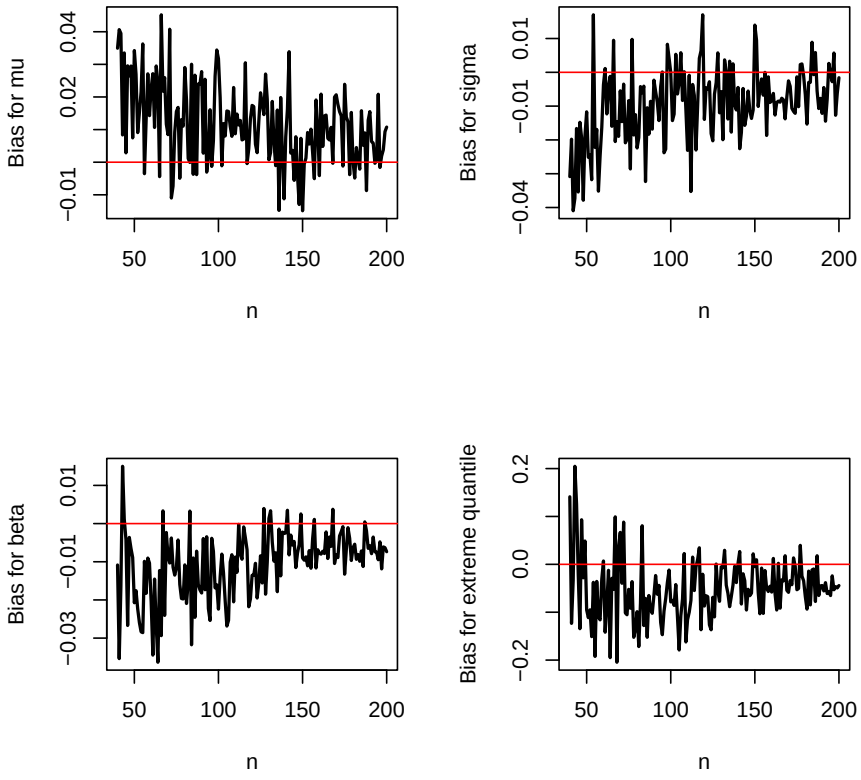


Figure A9. Biases of $\hat{\mu}$, $\widehat{\sigma}$, $\hat{\beta}$ and $\widehat{Q}(0.999)$ versus n when $\sigma = 1$ and $\beta = -0.2$.

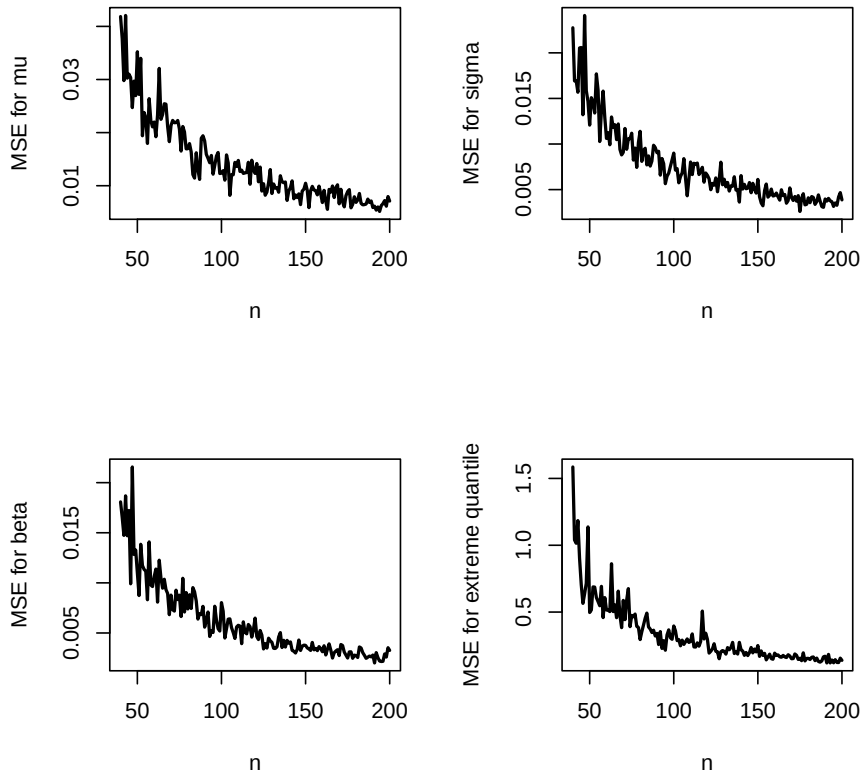


Figure A10. Mean squared errors of $\hat{\mu}$, $\widehat{\sigma}$, $\hat{\beta}$ and $\widehat{Q}(0.999)$ versus n when $\sigma = 1$ and $\beta = -0.2$.

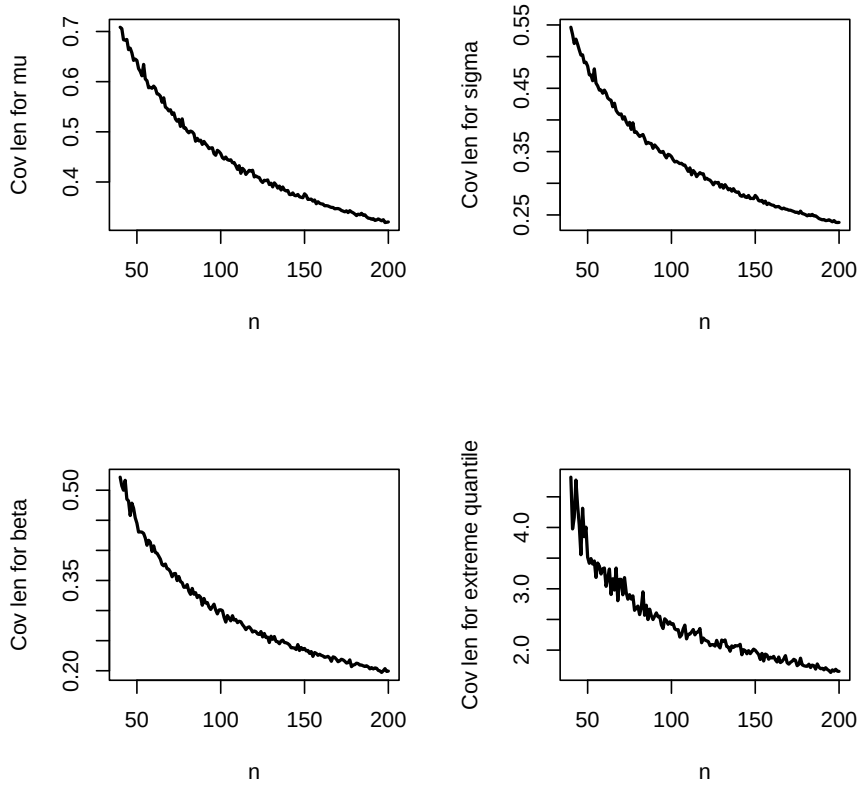


Figure A11. Coverage lengths for $\hat{\mu}$, $\widehat{\sigma}$, $\hat{\beta}$ and $\widehat{Q}(0.999)$ versus n when $\sigma = 1$ and $\beta = -0.2$.

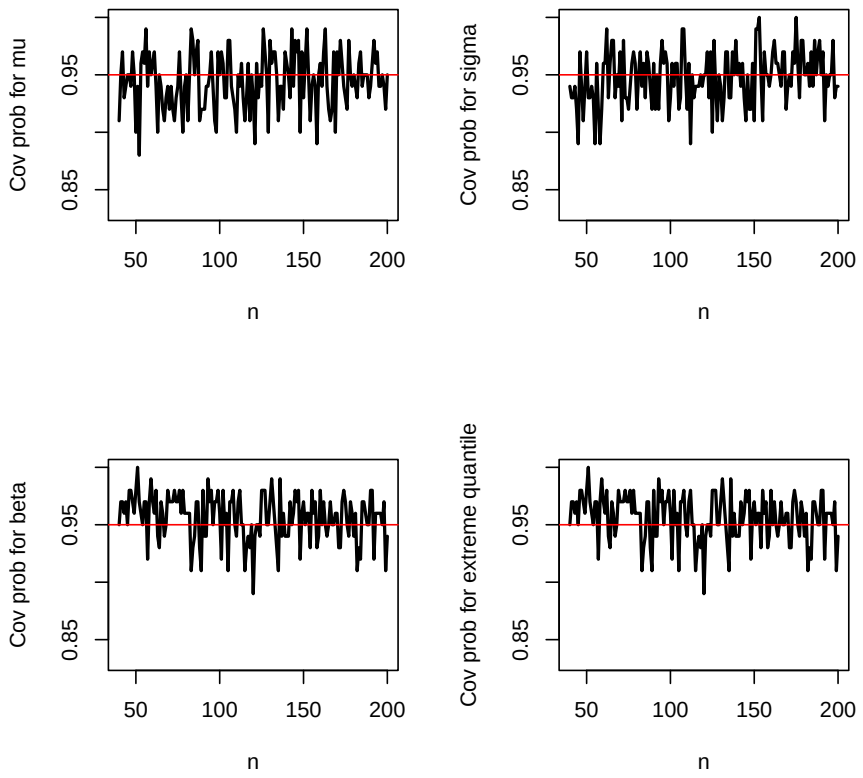


Figure A12. Coverage probabilities for $\hat{\mu}$, $\widehat{\sigma}$, $\hat{\beta}$ and $\widehat{Q}(0.999)$ versus n when $\sigma = 1$ and $\beta = -0.2$.

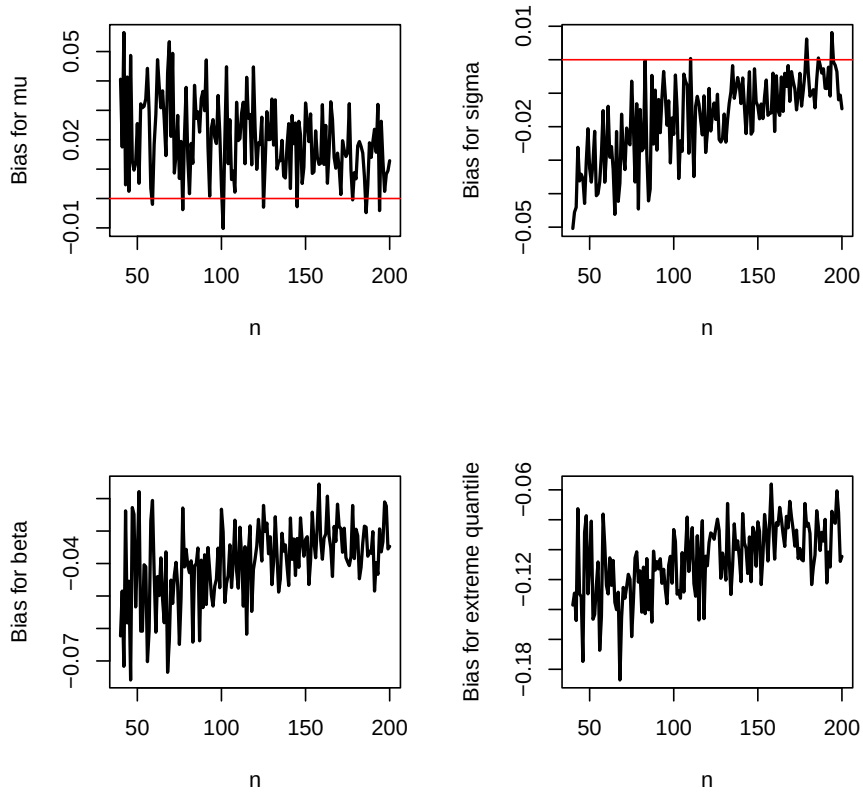


Figure A13. Biases of $\hat{\mu}$, $\widehat{\sigma}$, $\hat{\beta}$ and $\widehat{Q}(0.999)$ versus n when $\sigma = 1$ and $\beta = -0.50$.

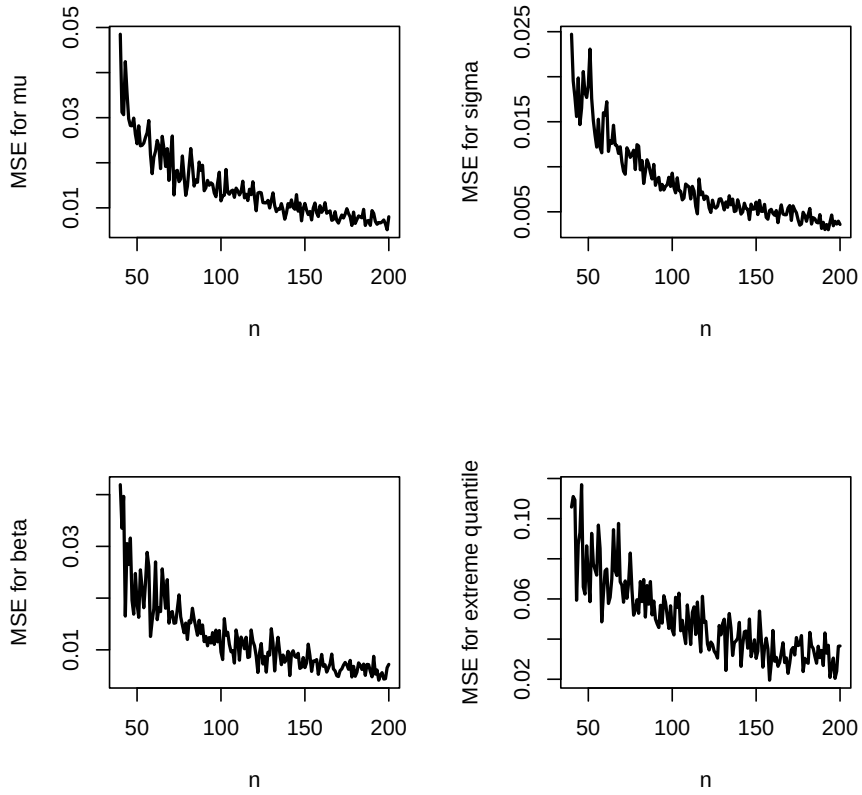


Figure A14. Mean squared errors of $\hat{\mu}$, $\widehat{\sigma}$, $\hat{\beta}$ and $\widehat{Q}(0.999)$ versus n when $\sigma = 1$ and $\beta = -0.50$.

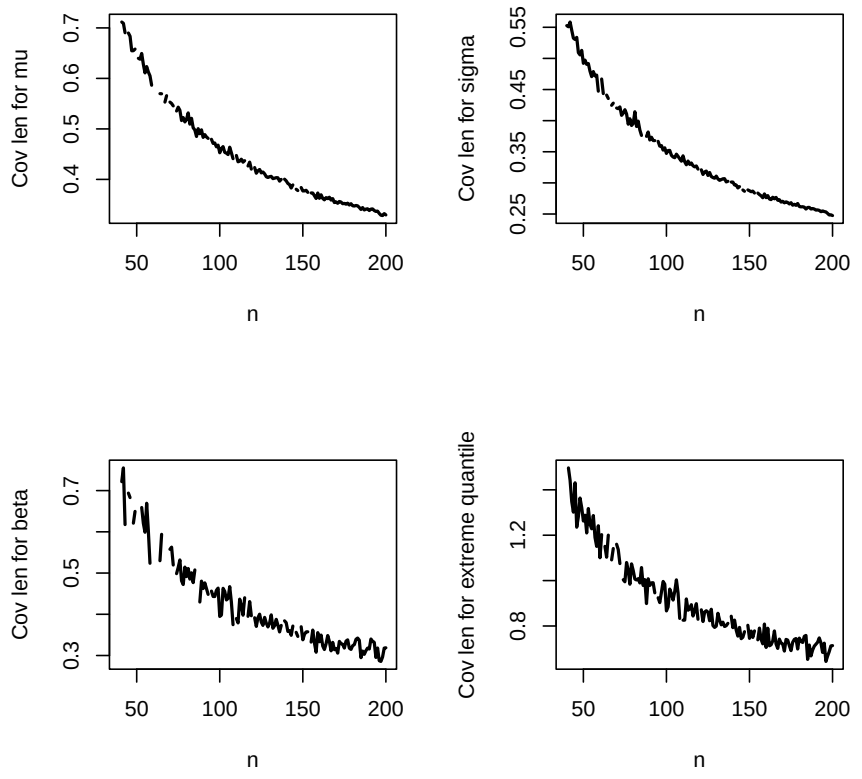


Figure A15. Coverage lengths for $\hat{\mu}$, $\widehat{\sigma}$, $\hat{\beta}$ and $\widehat{Q}(0.999)$ versus n when $\sigma = 1$ and $\beta = -0.50$.

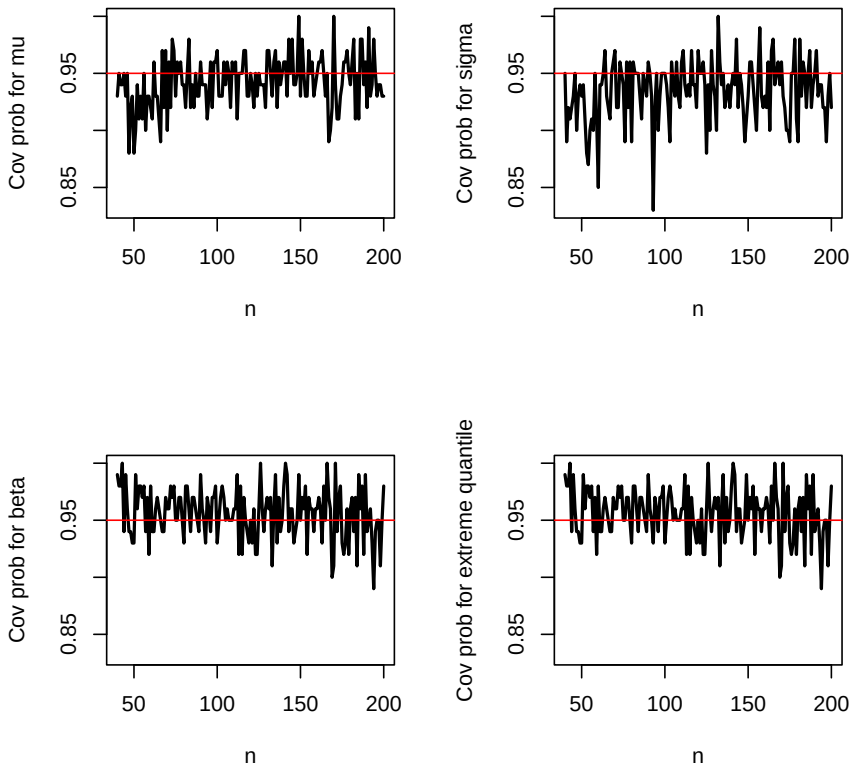


Figure A16. Coverage probabilities for $\hat{\mu}$, $\hat{\sigma}$, $\hat{\beta}$ and $\widehat{Q}(0.999)$ versus n when $\sigma = 1$ and $\beta = -0.50$.

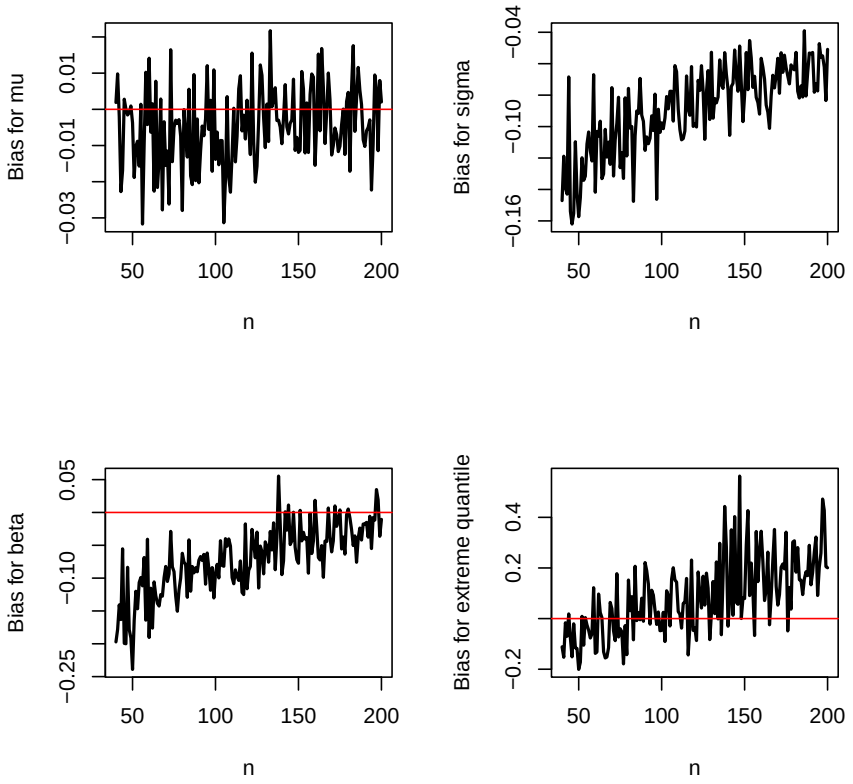


Figure A17. Biases of $\hat{\mu}$, $\hat{\sigma}$, $\hat{\beta}$ and $\widehat{Q}(0.999)$ versus n when $\sigma = 1$ and $\beta = -1$.

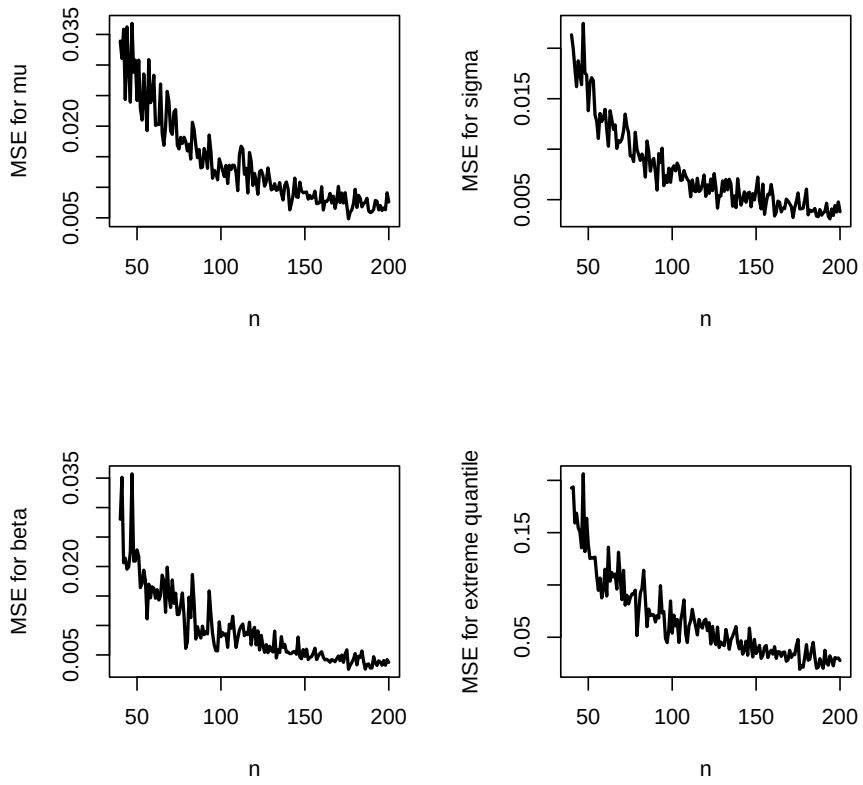


Figure A18. Mean squared errors of $\hat{\mu}$, $\widehat{\sigma}$, $\hat{\beta}$ and $\widehat{Q}(0.999)$ versus n when $\sigma = 1$ and $\beta = -1$.

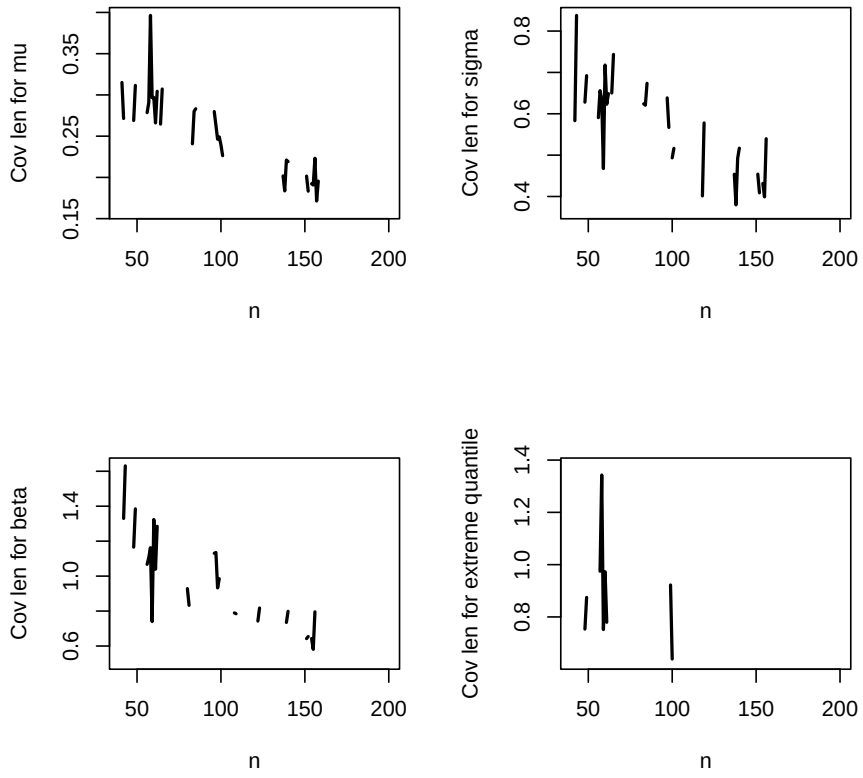


Figure A19. Coverage lengths for $\hat{\mu}$, $\widehat{\sigma}$, $\hat{\beta}$ and $\widehat{Q}(0.999)$ versus n when $\sigma = 1$ and $\beta = -1$.

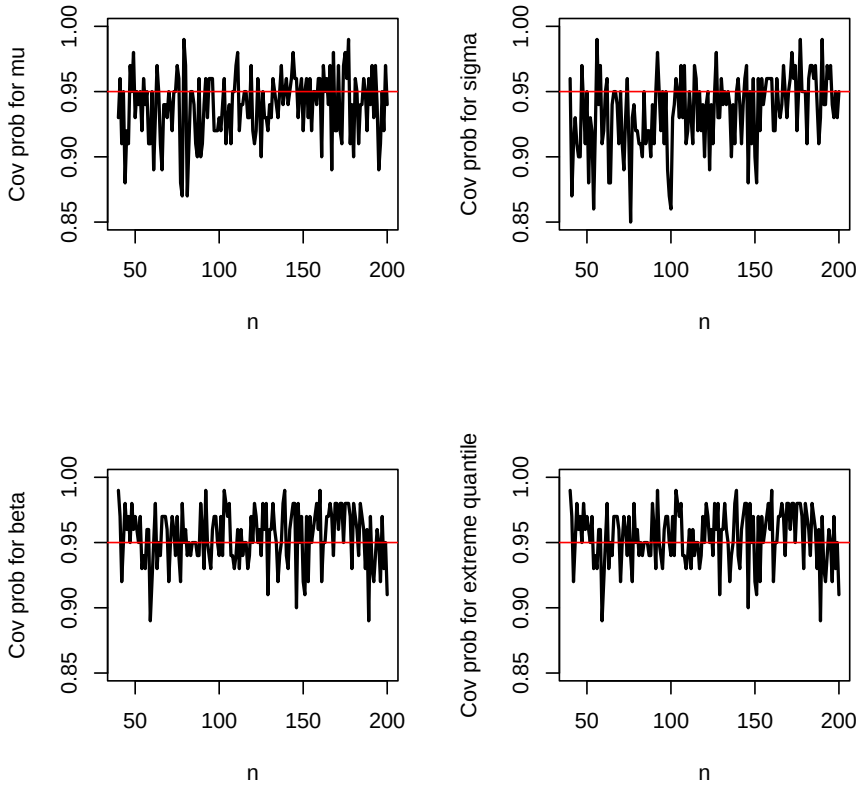


Figure A20. Coverage probabilities for $\hat{\mu}$, $\hat{\sigma}$, $\hat{\beta}$ and $\widehat{Q}(0.999)$ versus n when $\sigma = 1$ and $\beta = -1$.

References

1. Fisher, R.A.; Tippett, L.H.C. Limiting forms of the frequency distribution of the largest and smallest member of a sample. *Proc. Camb. Philos. Soc.* **1928**, *24*, 180–190. [CrossRef]
2. Gnedenko, B.V. Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* **1943**, *44*, 423–453. [CrossRef]
3. Jenkinson, A.F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Q. J. R. Meteorol. Soc.* **1955**, *81*, 158–171. [CrossRef]
4. Cherif, S. A generalized extreme value approach for the analysis of stationary climatic covariate in a Mediterranean city. *Atmos. Sci. Lett.* **2025**, *26*, e1291. [CrossRef]
5. Jiao, Z.; Yuan, J.; Farnham, C.; Emura, K. Assessing the risk of extreme precipitation in Japan through GEV distribution and spatial modeling. *J. Hydrol. Reg. Stud.* **2026**, *63*, 103107. [CrossRef]
6. Hossain, I.; Gato-Trinidad, S.; Imteaz, M.A. Seasonal and spatial variations in General Extreme Value (GEV) distribution shape parameter for estimating extreme design rainfall in Tasmania. *Water* **2026**, *18*, 319. [CrossRef]
7. Zhang, C.; Fu, X.; Yang, D.; Zhang, P.; Zhang, Y. GEV distribution-enhanced Fourier diffusion model for extreme value capture in day-ahead photovoltaic scenario generation. *Appl. Energy* **2026**, *409*, 127472. [CrossRef]
8. Zhang, F.; Zheng, S.; Ma, F.; Yin, Q.; Zhou, Y. SAR image change detection via generalized extreme value (GEV) modeling. *Pattern Recognit.* **2026**, *170*, 112040. [CrossRef]
9. Hosking, J.R.M.; Wallis, J.R.; Wood, E.F. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* **1985**, *27*, 251–261. [CrossRef]
10. Leadbetter, M.R.; Lindgren, G.; Rootzén, H. *Extremes and Related Properties of Random Sequences and Processes*; Springer: New York, NY, USA, 1983.
11. Galambos, J. *The Asymptotic Theory of Extreme Order Statistics*; Robert E. Krieger Publishing Company: Malabar, FL, USA, 1987.
12. Resnick, S.I. *Extreme Values, Regular Variation and Point Processes*; Springer: New York, NY, USA, 1987.
13. Embrechts, P.; Klüppelberg, C.; Mikosch, T. *Modelling Extremal Events for Insurance and Finance*; Springer: Berlin/Heidelberg, Germany, 1997.
14. Kotz, S.; Nadarajah, S. *Extreme Value Distributions: Theory and Applications*; Imperial College Press: London, UK, 2000.
15. Coles, S.G. *An Introduction to Statistical Modeling of Extreme Values*; Springer: London, UK, 2001.
16. Beirlant, J.; Goegebeur, Y.; Segers, J.; Teugels, J. *Statistics of Extremes: Theory and Applications*; John Wiley and Sons: Chichester, UK, 2004.
17. Gumbel, E.J. *Statistics of Extremes*; Dover Publications: Mineola, NY, USA, 2004.

18. Castillo, E.; Hadi, A.S.; Balakrishnan, N.; Sarabia, J.M. *Extreme Value and Related Models with Applications in Engineering and Science*; John Wiley and Sons: Hoboken, NJ, USA, 2005.
19. Malevergne, Y.; Sornette, D. *Extreme Financial Risks: From Dependence to Risk Management*; Springer: Berlin/Heidelberg, Germany, 2006.
20. de Haan, L.; Ferreira, A. *Extreme Value Theory: An Introduction*; Springer: New York, NY, USA, 2006.
21. Reiss, R.-D.; Thomas, M. *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*; Birkhäuser Verlag: Basel, Switzerland, 2007.
22. Falk, M.; Hüsler, J.; Reiss, R.-D. *Laws of Small Numbers: Extremes and Rare Events*; Birkhäuser/Springer Basel AG: Basel, Switzerland, 2011.
23. Novak, S.Y. *Extreme Value Methods with Applications to Finance*; CRC Press: Boca Raton, FL, USA, 2012.
24. Ahsanullah, M. *Extreme Value Distributions*; Atlantis Press: Paris, France, 2016.
25. Nakagawa, T.; Osaki, S. The discrete Weibull distribution. *IEEE Trans. Reliab.* **1975**, *24*, 300–301. [CrossRef]
26. Chakraborty, S.; Chakravarty, D.; Mazucheli, J.; Bertoli, W. A discrete analog of Gumbel distribution: Properties, parameter estimation and applications. *J. Appl. Stat.* **2021**, *48*, 712–737. [CrossRef] [PubMed]
27. Hadi, N.S.; Khudhair, J.K. On discrete Fréchet distribution: Estimation and application. *J. Adm. Econ.* **2024**, *49*, 189–199. [CrossRef]
28. Das, D.; Das, B. Discretized Fréchet–Weibull distribution: Properties and application. *J. Indian Soc. Probab. Stat.* **2023**, *24*, 243–282. [CrossRef]
29. Akkanphudit, T. The discrete Gompertz–Weibull–Fréchet distribution: Properties and applications. *Lobachevskii J. Math.* **2023**, *44*, 3663–3672. [CrossRef]
30. Das, D.; Alshammari, T.S.; Rashedi, K.A.; Das, B.; Hazarika, P.J.; Eliwa, M.S. Discrete joint random variables in Fréchet–Weibull distribution: A comprehensive mathematical framework with simulations, goodness-of-fit analysis, and informed decision-making. *Mathematics* **2024**, *12*, 3401. [CrossRef]
31. Hitz, A.S.; Davis, R.A.; Samorodnitsky, G. Discrete extremes. *J. Data Sci.* **2024**, *22*, 524–536. [CrossRef]
32. Dkengne Sielenou, P.A.; Eckert, N.; Naveau, P. Extreme value theory for integer-valued random variables. In Proceedings of the Conference on Extreme Value Modeling and Water Resources, Lyon, France, 13–24 June 2016. [CrossRef]
33. Tiago de Oliveira, J. *Statistical Analysis of Extremes*; Pendor: Lisbon, Portugal, 1997.
34. Tabony, R.C. Extreme value analysis in meteorology. *Meteorol. Mag.* **1983**, *112*, 77–98.
35. Chandler, R.E.; Bate, S. Inference for clustered data using the independence loglikelihood. *Biometrika* **2007**, *94*, 167–183. [CrossRef]
36. Stephenson, A.G. evd: Extreme value distributions. *R News* **2002**, *2*, 31–32.
37. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2025.
38. Smith, R.L. Maximum likelihood estimation in a class of non-regular cases. *Biometrika* **1985**, *72*, 67–90. [CrossRef]
39. Davison, A.C.; Smith, R.L. Models for exceedances over high thresholds. *J. R. Stat. Soc. Ser. B* **1990**, *52*, 393–442. [CrossRef]
40. van der Vaart, A.W. *Asymptotic Statistics*; Cambridge University Press: Cambridge, UK, 1998.
41. Cox, D.R.; Hinkley, D.V. *Theoretical Statistics*; Chapman and Hall: London, UK, 1974.
42. Kreutz, C. An easy and efficient approach for testing identifiability. *Bioinformatics* **2018**, *34*, 1913–1921. [CrossRef] [PubMed]
43. Kolmogorov, A. Sulla determinazione empirica di una legge di distribuzione. *G. Dell'istituto Ital. Degli Attuari* **1933**, *4*, 83–91.
44. Smirnov, N. Table for estimating the goodness of fit of empirical distributions. *Ann. Math. Stat.* **1948**, *19*, 279–281. [CrossRef]
45. Choulakian, V.; Lockhart, R.A.; Stephens, M.A. Cramér–von Mises statistics for discrete distributions. *Can. J. Stat.* **1994**, *22*, 125–137. [CrossRef]
46. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]
47. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 623–656. [CrossRef]
48. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.