Introduction to Extreme Value Theory

MATH 561 - Theory of Probability I
Professor Partha Dey
Michael Wieck-Sosa

May 4, 2021

Abstract

In this exposition, we will give a brief introduction of Extreme Value Theory for the maxima of i.i.d. random variables. Afterwards, two examples of deriving the limiting distributions for the maximum of Normal and Exponential i.i.d r.v.’s will be given. Lastly, we conclude by giving several reading recommendations to give the reader an impression of the wider landscape of Extreme Value Theory. The proofs and theorems for this exposition are based on Leadbetter (2012), Gumbel (1958), and de Haan (1976), with clarifications and more details included in the proofs when we believed them to be helpful.

1 Proof of the Extreme Types Theorem

We will be concerned with deriving distribution of the maximum of \( n \) i.i.d. r.v.’s \( X_1, X_2, ..., X_n \) denoted by

\[
M_n = \max(X_1, X_2, ..., X_n)
\]
as \( n \to \infty \).

The main result we will build up to is called the Extremal Types Theorem (Theorem 10), which states that if a sequence of normalizing constants \( a_n > 0, b_n \) then \( a_n(M_n - b_n) \) has a non-degenerate limiting distribution \( G(x) \), then that limiting distribution must be one of the three extreme value distributions: the Gumbel or Type I extreme value distribution, the Fréchet or Type II extreme value distribution, or the Weibull or Type III extreme value distribution. First, it will be useful to derive some preliminary results which will be used in the proof of the Extremal Types Theorem.

Now, we will prove several results, which will be used for the proof of the Extremal Types Theorem for i.i.d. r.v.s. Note that there are generalizations of the Extremal Value Theory, including the Extremal Types Theorem, to cases when the r.v.’s are dependent and also on the process level, which we discuss in the last section.

**Definition 1**

If \( \psi(x) \) is a non-decreasing right continuous function, define the inverse function \( \psi^{-1} \) on the interval \( (\inf\{\psi(x)\}, \sup\{\psi(x)\}) \) by

\[
\psi^{-1}(y) = \inf\{x \mid \psi(x) \geq y\}
\]

**Lemma 2**

(i) If \( a > 0, b, c \) are constants and \( H(x) = \psi(ax + b) - c \) then \( H^{-1}(y) = a^{-1}(\psi^{-1}(y + c) - b) \)
(ii) If \( \psi^{-1} \) is continuous, then \( \psi^{-1}(\psi(x)) = x \).

(iii) If \( G \) is a nondegenerate distribution function, then \( \exists y_1, y_2 \in \mathbb{R} \) where \( y_1 < y_2 \) s.t. \( G^{-1}(y_1) < G^{-1}(y_2) \) are well defined and finite.

Proof:

(i) Using the definition of the inverse defined above, observe that

\[
H^{-1}(y) = \inf\{ x \mid \psi(ax + b) - c \geq y \} = a^{-1}(\inf\{ (ax + b) \mid \psi(ax + b) \geq y + c \} - b) = a^{-1}(\psi^{-1}(y + c) - b)
\]

(ii) By definition of \( \psi^{-1} \) and inf, we have

\[
\psi^{-1}(\psi(x)) = \inf\{ x \mid \psi(x) = \psi(x) \} \leq x
\]

Next, suppose for the purposes of contradiction that \( \psi^{-1}(\psi(x)) < x \), then by the definition of \( \psi^{-1} \), \( \exists z < x \) s.t. \( \psi(z) \geq \psi(x) \), but since \( \psi \) is non-decreasing we must have \( \psi(z) = \psi(x) \). If we let \( y = \psi(z) = \psi(x) \), then we have \( \psi^{-1}(y) \geq x \) which clearly contradicts the assumption that \( \psi^{-1} \) is continuous. Therefore, we must have \( \psi^{-1}(\psi(x)) = x \).

(iii) If \( G \) is non-degenerate distribution function, then \( \exists z_1, z_2, z_1 < z_2 \) s.t. \( 0 < G(z_1) = y_1 < G(z_2) = y_2 \leq 1 \). Then \( x_1 = G^{-1}(y_1) \) and \( x_2 = G^{-1}(y_2) \) are well defined. We have \( G^{-1}(y_2) \geq z_1 \), and if \( G^{-1}(y_2) = z_1 \) then \( G(a) \geq z_2 \) \( \forall a > x_1 \), so \( G(z_1) = \lim_{\epsilon \to 0} G(z_1 + \epsilon) \geq y_2 \), which contradicts the fact that \( G(z_1) = y_1 \). Therefore, we must have \( G^{-1}(y_2) > z_1 \geq x_1 = G^{-1}(y_1) \).

\( \square \)

**Corollary 3** If \( G \) is a non-degenerate distribution function and \( a > 0, \alpha > 0, b, \beta \) are constants s.t. \( G(ax + b) = G(ax + \beta) \), \( \forall x \), then \( a = \alpha \) and \( b = \beta \).

Proof: By part (iii) of the previous lemma, if \( y_1 < y_2 \) and \( -\infty < x_1 < x_2 < \infty \), then \( x_1 = G^{-1}(y_1), x_2 = G^{-1}(y_2) \). By part (i) of the previous lemma, the inverses are

\[
a^{-1}G^{-1}(y) - b = a^{-1}(G^{-1}(y) - \beta)
\]

Then we have the desired result

\[
a^{-1}(x_1 - b) = a^{-1}(x_1 - \beta)
\]

\[
a^{-1}(x_2 - b) = a^{-1}(x_2 - \beta)
\]

Therefore \( a = \alpha \) and \( b = \beta \).

\( \square \)

A quick note on the notation. In the proofs below, we denote the convergence at continuity points of the limiting function by \( \rightarrow_{\mathrm{w}} \).

**Theorem 4** The theorem of Khintchine states the following. Let \( \{ F_n \} \) be a sequence of distribution functions and \( G \) a non-degenerate distribution function. Let \( a_n > 0 \) and \( b_n \) be constants s.t.

\[
F_n(a_n x + b_n) \rightarrow_{\mathrm{w}} G(x)
\]

Then for some non-degenerate distribution function \( G_* \) and constants \( a_n > 0, \beta_n \) we have

\[
F_n(a_n x + \beta_n) \rightarrow_{\mathrm{w}} G_*(x)
\]
Theorem 6

1. A non-degenerate distribution function for i.i.d. r.v.'s belongs to the i.i.d. domain of attraction, for maxima, if we have

\[ a_n^{-1} \alpha_n \to a \]
\[ a_n^{-1}(\beta_n - b_n) \to b \]

then we have

\[ G_*(x) = G(ax + b) \]

Proof: Let \( \alpha' = a_n^{-1} \alpha_n \) and \( \beta' = a_n^{-1}(\beta_n - b_n) \) and \( F'_n(x) = F_n(a_n x + b_n) \). Then we can equivalently prove that

\[ F'_n(x) \overset{w}{\to} G(x) \]
\[ F'_n(\alpha' x + \beta') \overset{w}{\to} G_*(x) \]
\[ \alpha' \to a \]
\[ \beta' \to b \]

Since \( G_* \) is non-degenerate by assumption, \( \exists x', x'' \), such that \( 0 < G_*(x') < 1 \) and \( 0 < G_*(x'') < 1 \).

Next, suppose for the purposes of contradiction that \( \{ \alpha'n x' + \beta'_n \} \) is not bounded. Then there is a subsequence \( \{ n_k \} \) s.t. \( \alpha'_{n_k} x' + \beta'_{n_k} \to \pm \infty \). Since \( G \) is a distribution function, this implies that

\[ F'_n(\alpha'_{n_k} x' + \beta'_{n_k}) = G_*(x) \]

for \( x = x' \). Therefore, \( \{ \alpha'n x' + \beta'_n \} \) and \( \{ \alpha'n x'' + \beta'_n \} \) are both bounded, so \( \{ \alpha'_n \} \) and both \( \{ \beta'_n \} \) are bounded. Therefore, there is a sequence \( a, b, \{ m_k \} \) s.t. \( \alpha'_{m_k} \to a \) and \( \beta'_{m_k} \to b \), so we have

\[ F'_n(\alpha'_{m_k} x + \beta'_{m_k}) = G(ax + b) \]

by Corollary 3. Therefore, \( a' = a \) and \( b' = b \) by Corollary 3. Therefore \( \alpha' \to a \) and \( \beta' \to b \), which implies that

\[ F'_n(\alpha' x + \beta'_n) \overset{w}{\to} G_*(x) \] with \( G_*(x) = G(ax + b) \).

Now that we have proven these preliminary results, we are ready to discuss the class of max-stable distributions for maxima of i.i.d. sequences.

**Definition 5** If \( F^n(a_n^{-1} x + b_n) \overset{w}{\to} G(x) \) holds for some sequences \( \{ a_n \}, \{ b_n \} \), we say that \( F \) belongs to the i.i.d. domain of attraction, for maxima, of \( G \), and we denote it by \( F \in D(G) \).

**Theorem 6**

(i) A non-degenerate distribution function \( G \) is max-stable iff there is a sequence \( \{ F_n \} \) of distributions functions and constants \( a_n > 0 \) and \( b_n \) such that for each \( k \in \mathbb{N} \) we have

\[ F_n(a_n^{-1} x + b_n) \overset{w}{\to} G^{1/k}(x) \] as \( n \to \infty \)

(ii) If \( G \) is non-degenerate, and \( D(G) \) is nonempty iff \( G \) is max-stable, meaning we have \( G \in D(G) \). The class of non-degenerate distribution functions \( G \) which are the limit laws

\[ P(a_n(M_n - b_n) \leq x) = P(M_n \leq a_n^{-1} x + b_n) = \prod_{i=1}^{n} P(M_i \leq a_i^{-1} x + b_i) = F^n(a_n^{-1} x + b_n) \overset{w}{\to} G(x) \]

for i.i.d. r.v.'s \( X_1, X_2, ... \) coincides with the class of max-stable distribution functions.
Proof:
(i) If $G$ is non-degenerate, then $G^{1/k} \forall k \in \mathbb{N}$, and if $F_n(a_n^{-1}x + b_n) \xrightarrow{w} G^{1/k}(x)$ as $n \to \infty$ holds $\forall k \in \mathbb{N}$. Khintchine’s Theorem implies that $G^{1/k}(x) = G(\alpha_k x + \beta_k)$ for some $\alpha_k > 0$ and $\beta_k$. Thus, $G$ is max-stable by definition. On the other hand, if $G$ is max-stable and $F_n = G^n$, we have $G^n(a_n^{-1}x + b_n) = G(x)$ for some $a_n > 0$ and $b_n$, then we have

$$F_n(a_n^{-1}x + b_n) = (G^{nk}(a_n^{-1}x + b_n))^{1/k} = (G(x))^{1/k}$$

therefore $F_n(a_n^{-1}x + b_n) \xrightarrow{w} G^{1/k}(x)$ as $n \to \infty$.

(ii) If $G$ is max-stable, then $G^n(a_nx + b_n) = G(x)$ for some $a_n > 0$ and $b_n$, so $G \in D(G)$. On the other hand, if $D(G)$ is nonempty, then $F \in D(G)$ with $F^n(a_n^{-1}x + b_n) \xrightarrow{w} G(x)$. Therefore, $F^n(a_n^{-1}x + b_n) \xrightarrow{w} G(x)$ or $F^n(a_n^{-1}x + b_n) \xrightarrow{w} G^{1/k}(x)$. Therefore $F_n(a_n^{-1}x + b_n) \xrightarrow{w} G^{1/k}(x)$ as $n \to \infty$ holds with $F_n = F^n$ so by part (i) proven above, $G$ is max-stable.

\[\square\]

**Corollary 7** If $G$ is max-stable, there exist function $a(s) > 0$ and $b(s)$ defined for $s > 0$ s.t.

$$G'(a(s)x + b(s)) = G(x)$$

for all real $x$, $s > 0$.

Proof: By assumption, $G$ is max-stable, and there exist $a_n > 0$, $b_n$ such that

$$G^n(a_nx + b_n) = G(x)$$

so we have

$$G^{[ns]}(a_{[ns]}x + b_{[ns]}) = G(x)$$

where $[ns]$ is the integer part of $ns$. Therefore, we have

$$G^n(a_{[ns]} + b_{[ns]} \xrightarrow{w} G^{1/s}(x)$$

Therefore by the previous limit, and $G^n(a_nx + b_n) = G(x)$, and since $G^{1/s}$ is non-degenerate, then Khintchine’s Theorem with $a_n = a_{[ns]}$ and $\beta_n = b_{[ns]}$ implies that $G(a(s)x + b(s)) = G^{1/s}(x)$ for some $a(s) > 0$ and $b(s)$.

\[\square\]

**Definition 8** We say that two distribution functions $G_1, G_2$ are of the same type if

$$G_2(x) = G_1(ax + b)$$

for some constants $a > 0$, $b$.

Moreover, note that by the definition of $D(G)$ (Definition 5), we have that $D(G_1) = D(G_2)$ if $G_1$ and $G_2$ are of the same type, otherwise if $G_1$ and $G_2$ are not of the same type then $D(G_1) \cap D(G_2) = \emptyset$.

Now, we are ready to discuss the distributions of the maxima of i.i.d. r.v.’s in the sense of

$$\mathbb{P}(a_n(M_n - b_n) \leq x) \xrightarrow{w} G(x)$$

**Theorem 9** Every max-stable distribution is of extreme value type, meaning that it is equal to $G(ax + b)$ for some $a > 0$ and $b$, and each distribution of extreme value type is max-stable. The three extreme value types are:
Type I: \( G(x) = \exp(-e^{-x}) \) for \( -\infty < x < \infty \)

Type II: \( G(x) = \begin{cases} 
0 & \text{for } x \leq 0 \\
\exp(-x^{-\alpha}) & \text{for some } \alpha > 0 \text{ and for } x > 0
\end{cases} \)

Type III: \( G(x) = \begin{cases} 
\exp(-(-x)^\alpha) & \text{for some } \alpha > 0 \text{ and for } x \leq 0 \\
1 & \text{for } x > 0
\end{cases} \)

Proof: First, we will prove the converse that each distribution of extreme value type is max-stable. For Type I we have

\[
\exp\left\{ -\alpha - \frac{\alpha}{\lambda}\right\} = \exp\left\{ -\alpha - \log\left(\frac{\alpha}{\lambda}\right)\right\}
\]

with similar expressions for the other types using the same approach. Now, let us prove the direct statement.

Following de Haan (1976), if \( G \) is max-stable, then we have

\[
-s\log G(a(s)x + b(s)) = -\log G(x)
\]

such that

\[
-\log(-\log G(a(s)x + b(s))) = -\log(-\log G(x))
\]

By the max-stable property, \( G \) has no jump at any finite endpoint. The function \( \psi(x) = -\log(-\log G(x)) \) has inf \( \psi(x) = -\infty \) and sup \( \psi(x) = \infty \), therefore it has an inverse function \( U(y), \forall y \in \mathbb{R} \). Then we have

\[
\psi(a(s)x + b(s)) - \log s = \psi(x)
\]

so by Lemma 2 part (i) we have

\[
\frac{U(y + \log s) - b(s)}{a(s)} = U(y)
\]

Subtracting for \( y = 0 \) yields

\[
\frac{U(y + \log s) - U(\log s)}{a(s)} = U(y) - U(0)
\]

and letting \( z = \log s, \hat{a}(z) = a(e^z) \) and \( \hat{U}(y) = U(y) - U(0) \) so that for all \( y, z \in \mathbb{R} \) we have

\[
\hat{U}(y + z) - \hat{U}(y) = \hat{U}(y)\hat{a}(z)
\]

Then substituting \( y \) and \( z \) and subtracting, we have

\[
\hat{U}(y)(1 - \hat{a}(z)) = \hat{U}(z)(1 - \hat{a}(y))
\]

Two cases are possible.

Case 1: \( \hat{a}(z) = 1, \forall z \), then we have

\[
\hat{U}(y + z) = \hat{U}(y) + \hat{U}(z)
\]
The monotone increasing solution is $\hat{U}(y) = \rho y$ for some $\rho > 0$ s.t. $U(y) - U(0) = \rho y$ and
\[
\psi^{-1}(y) = U(y) = \rho y + U(0),
\]
so by Lemma 2 part (ii) we have
\[
x = \psi^{-1}(\psi(x)) = \rho \psi(x) + U(0)
\]
\[
\psi(x) = (x - U(0))/\rho
\]
thus $G(x) = \exp(-e^{-(x-U(0))/\rho}$ when $G(x) \in (0,1)$.

Therefore, $G$ has no jump any any finite endpoint, thus this limit law is of Type I.

Case 2: $\hat{a}(z) \neq 1$ for some $z$ when we have
\[
\hat{U}(y)(1 - \hat{a}(z)) = \hat{U}(z)(1 - \hat{a}(y))
\]
so we have
\[
\hat{U}(y) = \frac{\hat{U}(z)}{1 - \hat{a}(z)}(1 - \hat{a}(y)) = c(1 - \hat{a}(y))
\]
where $c = \frac{\hat{U}(z)}{1 - \hat{a}(z)} \neq 0$ since $\hat{U}(z) = 0$ would imply $\hat{U}(y) = 0$ for all $y$, so we would have $U(y) = U(0)$, therefore $c = \frac{\hat{U}(z)}{1 - \hat{a}(z)} \neq 0$. Then from the equation
\[
\hat{U}(y + z) - \hat{U}(y) = \hat{U}(y)\hat{a}(z)
\]
we have
\[
c(1 - \hat{a}(y + z)) - c(1 - \hat{a}(z)) = c(1 - \hat{a}(y))\hat{a}(z)
\]
so we have $\hat{a}(y+z) = \hat{a}(y)\hat{a}(z)$. However, we since $\hat{a}$ is monotone, the only solution is $\hat{a}(y) = e^{\rho y}$ for $\rho \neq 0$ so we have
\[
\psi^{-1}(y) = U(y) = U(0) + c(1 - e^{\rho y})
\]
Since $-\log(-\log G(x))$ is increasing, $U$ is increasing, so we have $c < 0$ if $\rho > 0$ and $c > 0$ if $\rho < 0$.

By Lemma 2 part (ii) we have
\[
x = \psi^{-1}(\psi(x)) = U(0) + c(1 - e^{\rho \psi(x)}) = U(0) + c(1 - (-\log G(x))^{-\rho})
\]
where $G(x) \in (0,1)$ so we have
\[
G(x) = \exp(-(1 - \frac{x - U(0)}{c})^{-1/\rho})
\]
By the continuity of $G$ at any finite endpoints, we have that $G$ is of Type II or Type III with $\alpha = 1/\rho$ or $\alpha = -1/\rho$ according as $\rho > 0$ and $c < 0$ or $\rho < 0$ and $c > 0$.

Finally, we are ready to prove the main result: the Extreme Types Theorem. The following proof for i.i.d. r.v.'s is due to de Haan (1976), although the general case for dependent r.v.'s was first rigorously proved by Gnedenko (1943).

**Theorem 10** The Extremal Types Theorem states the following. Let $M_n = \max\{X_1, X_2, \ldots, X_n\}$, where $X_i$ are i.i.d. r.v.'s. If for some constants $a_n > 0$ and $b_n$ we have
\[
P(a_n(M_n - b_n) \leq x) \rightarrow G(x)
\]
for some non-degenerate distribution function $G$, then $G$ is one of the three extreme value types. On the other hand, each distribution function $G$ of extreme value type may be a limit of
\(\mathbb{P}(a_n(M_n - b_n) \leq x) \xrightarrow{w} G(x).\) When \(G\) is the distribution function of each of the \(X_i\)’s, \(G\) is the limit of \(\mathbb{P}(a_n(M_n - b_n) \leq x) \xrightarrow{w} G(x).\)

Proof: Suppose \(\mathbb{P}(a_n(M_n - b_n) \leq x) \xrightarrow{w} G(x)\) for some constants \(a_n > 0\) and \(b_n\). Then by Theorem 6, \(G\) is max-stable. Therefore, by Theorem 9, \(G\) is of extreme value type.

On the other hand, if \(G\) is of extreme value type, then it is max-stable by Theorem 9. Moreover, by Theorem 6 part (ii), we have \(G \in D(G).\)

\(\square\)

## 2 Examples

In this section, we will introduce a theorem and corollary that will allow us to obtain the limit law for i.i.d r.v.’s. Afterwards, we will give two examples of deriving the limiting distributions of the maximum of i.i.d r.v.’s with Normal and Exponential distributions.

**Theorem 11** Let \(X_1, X_2, \ldots, X_n\) be i.i.d. r.v.’s. Let \(0 \leq \tau \leq \infty\) and suppose \(\{u_n\}\) is a sequence of real numbers s.t.

\[ n(1 - F(u_n)) \rightarrow \tau \text{ as } n \rightarrow \infty \]

Then we have

\[ \mathbb{P}(M_n \leq u_n) \rightarrow e^{-\tau} \text{ as } n \rightarrow \infty \]

On the other hand, if

\[ \mathbb{P}(M_n \leq u_n) \rightarrow e^{-\tau} \text{ as } n \rightarrow \infty \]

holds for some \(0 \leq \tau \leq \infty\), then we have

\[ n(1 - F(u_n)) \rightarrow \tau \text{ as } n \rightarrow \infty \]

Proof:

Suppose \(\tau \in [0, \infty).\) If \(n(1 - F(u_n)) \rightarrow \tau \text{ as } n \rightarrow \infty\), then we have

\[ \mathbb{P}(M_n \leq u_n) = F^n(u_n) = (1 - (1 - F(u_n)))^n \]

which may be rewritten as

\[ (1 - \tau/n + o(1/n))^n \]

so we must have

\[ \mathbb{P}(M_n \leq u_n) \rightarrow e^{-\tau} \text{ as } n \rightarrow \infty \]

Now, for the converse. If

\[ \mathbb{P}(M_n \leq u_n) \rightarrow e^{-\tau} \text{ as } n \rightarrow \infty \]

for \(\tau \in [0, \infty)\) then we must have \(1 - F(u_n) \rightarrow 0\), otherwise \(1 - F(u_n)\) would be bounded away from 0 for some subsequence \(\{n_k\}\) so by

\[ \mathbb{P}(M_n \leq u_n) = F^n(u_n) = (1 - (1 - F(u_n)))^n \]

we would have \(\mathbb{P}(M_n \leq u_n) \rightarrow 0\). By taking the log of both sides, we have

\[ n \cdot \log(1 - (1 - F(u_n))) \rightarrow -\tau \]

\[ 7 \]
so that \( n(1 - F(u_n))(1 + o(1)) \to T \) yields the result. Suppose, for the purposes of contradiction, that \( T = \infty \) and \( n(1 - F(u_n)) \to T \) but it is not true that \( P(M_n \leq u_n) \to e^{-T} \) as \( n \to \infty \), then there must be a subsequence \( \{ n_k \} \) such that \( P(M_{n_k} \leq u_{n_k} \to e^{-T'} \) as \( k \to \infty \) for some \( T' < \infty \). However, \( P(M_n \leq u_n) \to e^{-T} \) as \( n \to \infty \) implies \( n(1 - F(u_n)) \to T \) as \( n \to \infty \) with \( n_k \) replacing \( n \) s.t. \( n_k(1 - F(u_{n_k})) \to T' < \infty \) which contradicts \( n(1 - F(u_n)) \to T \) with \( T = \infty \). Similarly, we have that \( P(M_n \leq u_n) \to e^{-T} \) as \( n \to \infty \) implies \( n(1 - F(u_n)) \to T \) when \( T = \infty \). □

Next, we will prove a corollary that will allow us to easily find the limit laws of i.i.d. r.v.’s.

**Corollary 12**

(i) \( M_n \to x_F \) a.s., that is, \( P(\lim_{n \to \infty} M_n = x_F) = 1 \)

(ii) If \( x_F < \infty \) and \( F(x_F) < 1 \), meaning that the distribution function \( F \) has a jump at the right endpoint, and if there exists a sequence \( \{ u_n \} \) s.t. \( P(M_n \leq u_n) \to \rho \) as \( n \to \infty \). Then \( \rho = 0 \) or \( \rho = 1 \).

*Proof: Let \( x_F = \sup \{ x \mid F(x) < 1 \} \).

If \( \lambda < x_F \) then \( 1 - F(\lambda) > 0 \) s.t. \( n(1 - F(u_n)) \to T \) as \( n \to \infty \) holds with \( u_n = \lambda \) and \( T = \infty \), therefore by Theorem 11 we have \( P(M_n \leq u_n) \to e^{-T} \) as \( n \to \infty \) which gives \( P(M_n \leq \lambda) = 0 \). Since \( P(M_n > x_F) = 0, \forall n \in \mathbb{N} \), we must have \( M_n \to x_F \) in probability. Since \( \{ M_n \} \) is monotone, we must have \( M_n \to x_F \) a.s., therefore \( P(\lim_{n \to \infty} M_n = x_F) = 1 \).

To prove part (ii), suppose \( x_F < \infty \) and \( F(x_F) < 1 \), and let \( \{ u_n \} \) be a sequence s.t. \( P(M_n \leq u_n) \to \rho \) where \( \rho = e^{-T} \) with \( 0 \leq T \leq \infty \) which is valid since \( \rho \in [0, 1] \). Then by Theorem 11 we have \( n(1 - F(u_n)) \to T \). If \( u_n < x_F \) for infinitely many \( n \) s.t. \( 1 - F(u_n) \geq 1 - F(x_F) > 0 \) we must have \( T = \infty \). Otherwise, \( u_n \geq x_F \) for large enough \( n \), which means \( n(1 - F(u_n)) = 0 \) so \( T = 0 \). Therefore we must have \( T = 0 \) or \( T = \infty \), so we must have \( \rho = 0 \) or \( \rho = 1 \), which proves part (ii). □

Now, we prove that the asymptotic distribution for the maximum of i.i.d standard normal r.v.’s is Gumbel or Type I extreme value distribution.

**Example 13** Let \( X_1, X_2, ..., X_n \sim N(0, 1) \) be i.i.d. r.v.’s, and let \( M_n = \max(X_1, X_2, ..., X_n) \). Find the asymptotic distribution of \( M_n \).

*Proof:

Recall the standard result that \( 1 - \Phi(u) - \frac{\phi(u)}{u} \) as \( u \to \infty \).

Let \( T = e^{-x} \) and \( 1 - \Phi(u_n) = (1/n)e^{-x} \). Since \( 1 - \Phi(u_n) - \phi(u_n)/u_n \), we have the limit

\[
(1/n)e^{-x}u_n/\phi(u_n) \to 1
\]

Taking the log yields \(-\log(n) - x + \log(u_n) - \log(\phi(u_n)) \to 0 \) and also

\[-\log(n) - x + \log(u_n) + \frac{1}{2}\log(2\pi) + \frac{u_n^2}{2} \to 0\]

Therefore, we have \( \frac{u_n^2}{2\log(n)} \to 1 \) and

\[2\log(u_n) - \log(2) - \log(\log(n)) \to 0\]

8
or equivalently
\[ \log(u_n) = \frac{1}{2}(\log(2) + \log(\log(n))) + o(1) \]
where we write \( f(x) = o(g(x)) \) as \( x \to \infty \) if \( \forall \epsilon, \exists N \) s.t. \( |f(x)| \leq \epsilon g(x), \forall x \geq N \).

Now, substituting this back into the limit
\[ -\log(n) - x + \log(u_n) + \frac{1}{2}\log(2\pi) + \frac{u_n^2}{2} \to 0 \]
yields the equation
\[ \frac{u_n^2}{2} = x + \log(n) - \frac{1}{2}\log(4\pi) - \frac{1}{2}(\log(\log(n))) + o(1) \]
or equivalently, factoring out \( \log(n) \), we have
\[ u_n^2 = 2\log(n)(1 + \frac{x - \frac{1}{2}\log(4\pi) - \frac{1}{2}(\log(\log(n)))}{\log(n)}) + o\left(\frac{1}{\log(n)}\right) \]
so taking the square root of both sides yields
\[ u_n = (2\log(n))^{1/2}(1 + \frac{x - \frac{1}{2}\log(4\pi) - \frac{1}{2}(\log(\log(n)))}{\log(n)}) + o\left(\frac{1}{\log(n)}\right) \]

Therefore, distributing \( (2\log(n))^{1/2} \) yields
\[ u_n = \frac{x}{a_n} + b_n + o((\log(n))^{-1/2}) = \frac{x}{a_n} + b_n + o(a_n^{-1}) \]

By Theorem 11, with the result that \( P(M_n \leq u_n) \to e^{-T} \), if we let \( T = \exp(-e^{-x}) \) then we have \( P(M_n \leq u_n) \to e^{-e^{-x}} \). Therefore, we have the desired result that
\[ P(M_n \leq \frac{x}{a_n} + b_n + o(a_n^{-1}) \to \exp(-e^{-x}) \]
and rearranging the equation yields
\[ P(a_n(M_n - b_n) + o(1) \leq x) \to \exp(-e^{-x}) \]

□

Therefore, we have shown that the maximum of \( X_1, X_2, ..., X_n \) i.i.d standard normal r.v.’s has the Gumbel or Type I extreme value distribution.

Next, we will show that the maximum of \( X_1, X_2, ..., X_n \) i.i.d exponential r.v.’s has the Gumbel or Type I extreme value distribution.

**Example 14** Let \( X_1, X_2, ..., X_n \sim \text{Exp}(\lambda) \) be i.i.d. r.v.’s, and let \( M_n = \max(X_1, X_2, ..., X_n) \). Find the asymptotic distribution of \( M_n \).

Observe that
\[ P(a_n(M_n - b_n) \leq x) = P(M_n \leq \frac{x}{a_n} + b_n) = P\left(\bigcap_{j=1}^{n} [M_j \leq \frac{x}{a_n} + b_n]\right) = (P(X_1 \leq \frac{x}{a_n} + b_n))^n \]
Now, since $X_i \sim \text{Exp}(\lambda)$ we have

$$F(a_n^{-1}x + b_n) = (1 - e^{-\lambda(a_n^{-1}x + b_n)})^n$$

Choosing $a_n = \frac{\log(n)}{\lambda}$ and $b_n = \frac{1}{\lambda}$ yields

$$(1 - \frac{1}{n} e^{-x})^n$$

By the definition of $e^x = (1 - \frac{x}{n})^n$ we have

$$\mathbb{P}(a_n(M_n - b_n) \leq x) = (1 - \frac{1}{n} e^{-x})^n \to \exp(-e^{-x})$$

as $n \to \infty$

□

Therefore, we have shown that the maximum of $X_1, X_2, ..., X_n$ i.i.d exponential r.v.'s has the Gumbel or Type I extreme value distribution.

### 3 Conclusion

In this exposition, we have introduced several results that were used to give a proof of one of the main results of Extreme Value Theory. Namely, we gave the proof by de Haan (1976) of the Extreme Types Theorem (Theorem 10), which states that the maximum of i.i.d random variables and a sequence of normalizing constants $a_n > 0, b_n$ then $a_n(M_n - b_n)$ has a non-degenerate limiting distribution $G(x)$, then that limiting distribution must be one of the three extreme value distributions: the Gumbel or Type I extreme value distribution, the Fréchet or Type II extreme value distribution, or the Weibull or Type III extreme value distribution. Afterwards, we gave examples showing that the maximum of the Normal and Exponential distributions follow the Gumbel or Type I extreme value distribution.

While these results just scratch the surface of Extreme Value Theory, being acquainted with the Extreme Types Theorem and the mathematical methods used for the proof will allow you to appreciate much of the literature which we will recommend below. Next, we will give several recommendations to give an impression of the wider landscape of Extreme Value Theory and to discuss some sub-fields that are being actively researched such that an interested reader can explore further on their own.

### 4 Recommendations for Further Reading

First, some basic extensions. In this exposition, we only talked the maximum of i.i.d random variables, which is the simplest case. Of course, everything could be extended to the minimum through the obvious relation that $m_n = \min\{X_1, X_2, ..., X_n\} = -\max\{-X_1, -X_2, ..., -X_n\}$. Moreover, the limiting distributions can be extended to the $k$-th largest maxima of $X_1, X_2, ..., X_n$ denoted by $M_n^{(k)}$ where $k$ can be fixed integer or can tend to infinity as $n \to \infty$. In an introductory text, Leadbetter et. al (2012), the case when $k$ is a fixed integer is discussed in terms of asymptotic Poisson properties of the exceedance of $X_1, X_2, ...$ over a threshold.

Second, some additional extensions and active sub-fields of research. Things get much more complicated in the case when sequences of random variables are dependent and when considering extreme values in continuous time on the process level, which the interested reader will
be able to appreciate more after taking MATH 562. For an introduction to the wider landscape of Extreme Value Theory, interested readers should read *Extremes and Related Properties of Random Sequences and Processes* by Leadbetter (2012), a widely cited modern introduction to Extreme Value Theory. This text discusses the dependent case in chapters 3-6. In these chapters, the authors discuss how with certain restrictions dependent sequences converge have the same limit laws, and in some other non-normal cases where this is not true. In Leadbetter (2012) chapters 7-13, the authors discuss the case of extreme values in continuous time, where a general form of the Extreme Types Theorem (Theorem 10) is obtained for the maximum

\[ M(T) = \sup\{X(t) : 0 \leq t \leq T\} \]

where \( X(t) \) is a stationary stochastic process that satisfies certain regularity and dependence conditions. In this case, extreme value theory can be approached through the properties of upcrossings of a high level. Moreover, this approach shows that the point process of upcrossings of a level take on certain Poisson properties, which allow for asymptotic joint distributions for the locations and heights of any any number \( k \) of the largest local maxima. Additionally, the authors discuss the local behavior of a stationary normal process near a high-level upcrossing using the Slepian model process to detail sample paths at that upcrossing. Using this approach, the limiting distribution for the heights of excursions by these stationary normal processes above a high level can be obtained. The general extremal theory, and in particular the generalization of the Extremal Types Theorem (Theorem 10), for stationary continuous-time processes is proven in chapter 13. The method used for these proofs relies on the method of using the maximum of a continuous parameter process in time \( T = n \) as the maximum of \( n \) submaxima over fixed intervals

\[ M(n) = \max\{X_1, X_2, ..., X_n\} \]

where \( X_i = \sup\{X(t) : i-1 \leq t \leq i\} \). This area has seen rekindled interest in the last decade because of the importance of optimization in machine learning, which deals with understanding the maxima when there is a dependent structure between the random variables.

Third, we give two recommendations for readers interested in Statistics. There is a text called *Upper and Lower Bounds for Stochastic Processes: Modern Methods and Classical Problems* by Michel Talagrand, which is in the references section below. In this text, the authors discuss how to use generic chaining to get bounds for stochastic processes such as Gaussian processes, Bernoulli processes, stable processes, infinitely divisible processes, and the authors also discuss the convergence of random Fourier series, orthogonal series with applications of functional analysis. This text gives considerable attention to lower bounds, which is a much harder and less studied problem than upper bounds. That being said, this text is about deriving inequalities for mathematical statistics, and as such, does not contain much of the rigorous mathematical details of measure theory or sigma algebras that were of primary importance in MATH 561. Lastly, readers interested in Statistics should read the classical text *Statistics of Extremes* by E.J. Gumbel (1958) for a thorough introduction to the derivations of the classical results of Extreme Value Theory, with a plethora motivating examples and applications of Extreme Value Theory.
5 References


