5-13-2019

Bounding the Rates of Convergence Towards the Extreme Value Distributions

James Palmer
Bridgewater State University

Bounding the Rates of Convergence Towards the Extreme Value Distributions

James Palmer

Submitted in Partial Completion of the Requirements for
Departmental Honors in Mathematics

Bridgewater State University

May 13, 2019

Dr. John Pike, Thesis Advisor
Dr. Stephen Flood, Committee Member
Dr. Laura Gross, Committee Member

Bounding the Rates of Convergence Towards the Extreme Value Distributions

James Palmer

1 Introduction

Extreme value theory is a branch of probability which examines the extreme outliers of probability distributions. Three extreme value distributions arise as the limits of the maxima of sequences of random variables with certain properties. In this paper, we will first give information about these three distributions and prove that they are the only limit distributions of maxima. After that, we switch to a discussion about Stein's method.

Stein's method is commonly used to prove central limit theorems. Stein's method also develops bounds on the distance between probability distributions with regards to a probability metric. There are three essential steps to Stein's method: finding a characterizing operator, solving the Stein equation, and then using the solution to that equation to generate bounds on the distance to the target distribution. We will give a general overview of the method with some basic examples, and then go over various ways to find operators for any distribution. We outline the generalized density method, a recent technique for finding operators, and apply it to the extreme value distributions to find a particularly good operator. Next, we work with two operators simultaneously in an attempt to bound the distance between maxima and the extreme value distributions. Specifically, we apply this idea to the convergence of the exponential distribution to the Gumbel.

2 Extreme Value Distributions

Our first goal is to develop a basic understanding of the maxima of a sequence of independent, identically distributed random variables. We will start by proving that the only three limit distributions of maxima are the Gumbel, Fréchet, and Weibull distributions, and to do this, we first need a specific result called the convergence to types theorem.

Convergence to Types Theorem: Suppose $a_n \geq 0, A_n \geq 0, b_n \in \mathbb{R}, B_n \in \mathbb{R}$, and that for each $n \geq 1, F_n$ is a cumulative distribution function. If for every $x \in \mathbb{R}$,

$$F_n(a_n x + b_n) \to U(x) \quad \text{and} \quad F_n(A_n x + B_n) \to V(x) \tag{1}$$

as $n \to \infty$, then

$$\frac{A_n}{a_n} \to \alpha > 0, \qquad \frac{B_n - b_n}{a_n} \to \beta \in \mathbb{R}, \quad \text{and} \quad V(x) = U(\alpha x + \beta). \tag{2}$$

The quantities α and β can be found by inverting the two relations in Equation (1). That $V(x) = U(\alpha x + \beta)$ follows from Skorokhod's representation theorem. For a proof of this result, see chapter 0 of Resnick (1987). Now, we can move on to the desired proof involving the extreme value distributions.

2.1 Fisher-Tippett-Gnedenko Theorem:

Let $X_n, n \geq 1$ be a sequence of independent and identically distributed random variables with distribution function $F(x)$, and let $M_n = \max(X_1, X_2, ..., X_n)$. We note that M_n has distribution $F(x)^n$ by the simple calculation

$$P(M_n \leq x) = P\left\{\bigcap_{i=1}^{n}(X_i \leq x)\right\} = \prod_{i=1}^{n} P(X_i \leq x) = F(x)^n.$$

Now, suppose there exist $a_n > 0$, $b_n \in \mathbb{R}$, $n \geq 1$ such that

$$\lim_{n\to\infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = \lim_{n\to\infty} F(a_n x + b_n)^n \to G(x)$$

weakly and G is not degenerate. Then G has one of the following cumulative distribution functions:

$$\text{Fréchet} : \Phi_\alpha(x) = \begin{cases} 0, & x < 0 \\ e^{-x^{-\alpha}}, & x \geq 0 \end{cases} \qquad \text{for some } \alpha > 0;$$

$$\text{Weibull} : \Psi_\alpha(x) = \begin{cases} e^{-(-x)^{\alpha}}, & x < 0 \\ 0, & x \geq 0 \end{cases} \qquad \text{for some } \alpha > 0;$$

$$\text{Gumbel} : \Lambda(x) = e^{-e^{-x}}, x \in \mathbb{R}.$$

Here, we define that X_n converges weakly to X if and only if $P(X_n \leq x) \to P(X \leq x)$ for all continuity points of $F(x) = P(X \leq x)$. Additionally, we define that a distribution is degenerate if and only if it focuses at a single point.

Proof: Let $t \in \mathbb{R}$. Then for $t > 0$,

$$F(a_{\lfloor nt \rfloor} x + b_{\lfloor nt \rfloor})^{\lfloor nt \rfloor} \to G(x)$$

and

$$F(a_n x + b_n)^{\lfloor nt \rfloor} = (F(a_n x + b_n)^n)^{\frac{\lfloor nt \rfloor}{n}} \to G^t(x),$$

because $\frac{\lfloor nt \rfloor}{n} \to t$. As $G(x)$ and $G^t(x)$ are the same type, by the convergence to types theorem there exists $\alpha(t) > 0$ and $\beta(t) \in \mathbb{R}$ such that

$$\lim_{n\to\infty} \frac{a_n}{a_{\lfloor nt \rfloor}} = \alpha(t)$$

and

$$\lim_{n\to\infty} \frac{b_n - b_{\lfloor nt \rfloor}}{a_{\lfloor nt \rfloor}} = \beta(t)$$

with

$$G\left(\alpha(t)x + \beta(t)\right) = G^t(x). \tag{3}$$

Here, we can replace the variable t with the product of two variables st and see that for $s > 0$ and $t > 0$,

$$G\left(\alpha(st)x + \beta(st)\right) = G^{st}(x). \tag{4}$$

By evaluating one exponent at a time according to Equation (3) we see that

$$\begin{aligned} G^{st}(x) &= G^{s}\left(\alpha(t)x+\beta(t)\right) \\ &= G\left(\alpha(s)(\alpha(t)x+\beta(t))+\beta(s)\right) \\ &= G\left(\alpha(s)\alpha(t)+\alpha(s)\beta(t)+\beta(s)\right). \end{aligned} \tag{5}$$

Because G is non-degenerate, we can conclude from Equations (4) and (5) that

$$\alpha(st)=\alpha(s)\alpha(t) \tag{6}$$

and

$$\begin{aligned} \beta(st) &= \alpha(s)\beta(t)+\beta(s) \\ &= \alpha(t)\beta(s)+\beta(t). \end{aligned} \tag{7}$$

Equation (6) is a functional equation, and with our constraints (finite, measurable, nonnegative), the only possible solution is that $\alpha(t)=t^{-\theta}$, for some $\theta \in \mathbb{R}$. Our value for θ determines which of the three families G belongs to.

Case 1: $\theta = 0$. (G belongs to Gumbel family)
If $\theta = 0$, then $\alpha(t)=1$ and we can simplify Equation (7) to

$$\beta(st)=\beta(s)+\beta(t). \tag{8}$$

The solutions to this functional equation are logarithmic functions. We can say the solution is of the form $\beta(t)=-c\log(t)$ with $t>0$ and $c\in\mathbb{R}$. Now we simplify Equation (3) to

$$G^{t}(x)=G(x-c\log t). \tag{9}$$

Next, we can say that c is nonzero, because $c=0$ implies G is degenerate, violating our earlier assumption. We can also rule out c being negative by noting that $G^t(x)$ must be non-increasing in t because $|G(t)|\leq 1$. Therefore, $c>0$. Additionally, we know that G never takes on the value 0 or 1. If for some $x_0 \in \mathcal{R}$ we had that $G(x_0)=0$ then $0=G(x_0-c\log t)$ would arise directly from Equation (9). That result would hold for every value of t, which is a contradiction. Clearly the same argument works for $G(x_0)=1$. Now, let $x=0$ to see that

$$G^{t}(0)=G(-c\log t)$$

for every positive t. Because we proved $0\leq G(x)\leq 1$, we can set p such that $G(0)=e^{-e^{-p}}$. Now letting $u=-c\log t$ and substituting into the above expression we find that

$$G(u)=e^{(-e^{-p})t}=\exp(-e^{-\frac{u}{c}-p})$$

and thus G belongs to the Gumbel family.

Case 2: $\theta>0$. (G belongs to Fréchet family)
We can utilize both parts of Equation (7) to obtain

$$\frac{\beta(s)}{1-\alpha(s)}=\frac{\beta(t)}{1-\alpha(t)}$$

for $s \neq 1$ and $t \neq 1$, with $\beta(x)(1-\alpha(x))^{-1} = c$ for some constant c. Then we have that

$$\beta(t) = c(1 - t^{-\theta}).$$

Now we can substitute back into Equation (3) to find

$$\begin{aligned} G^t(x) &= G(t^{-\theta}x + c(1-t^{-\theta})) \\ &= G(t^{-\theta}(x-c) + c). \end{aligned}$$

We can now simplify by changing variables. Let $H(x) = G(x+c)$. Then

$$H^t(x) = G^t(x+c) = G(t^{-\theta}x + c),$$

and as G and H are of the same type, we can now work with H. We know that $H^t(x) = H(t^{-\theta x})$ and that H is nondegenerate. Now we want to show that $H(0) = 0$. Letting $x = 0$, we see that $\log H(0) = t \log(H(0))$ for any $t > 0$. As a result, we can conclude that either $\log H(0) = 0$ or $\log H(0) = -\infty$. Now the only possible values for $H(0)$ are 0 and 1. If $H(0) = 1$, then there would exist an $x_0 < 0$ such that $H^t(x)$ is decreasing in t, while $H(t^{-\theta x})$ is increasing in t. As the two expressions are equivalent, we know that $H(0) \neq 1$, and furthermore that $H(0) = 0$. Now we examine $H^t(1) = H(t^{-\theta})$. If $H(1) = 0$ or $H(1) = 1$ then H would take on those values everywhere and thus be degenerate, violating our earlier assumption. As $0 < H(1) < 1$, we can make the substitution $H(1) = \exp(-p^{-\alpha})$. We now set $\alpha = \theta^{-1}$, and make the change of variable $u = t^{-\theta}$ so that $t = u^{-\alpha^{-1}} = u^{-\alpha}$. We can compute

$$H(u) = H^t(1) = e^{-tp^{-\alpha}} = e^{-(pu)^{-\alpha}} = \Phi_\alpha(pu)$$

and see that H belongs to the Fréchet family.

Case 3: $\theta < 0$. (G belongs to Weibull family)

This case is almost identical to the last one. We can follow the same steps to immediately find that

$$H^t(x) = H(t^{-\theta}x),$$

keeping in mind that here θ is taking on negative values. By a similar argument to case 2, we can show that $H(0) = 1$. Now we plug in $x = -1$ and get that $H^t(-1) = H(-t^{-\theta})$. We know that $H(-1)$ does not equal 0 or 1, because if it did H would hold that value everywhere. Considering that, we can make the substitution $\alpha = -\theta^{-1}$ and let $p > 0$ be such that $H(-1) = e^{-p^\alpha}$. Then we make the change of variables $u = -t^{-\theta}$ so that $t = -u^{-\alpha}$. Then we have that

$$H(u) = H^t(-1) = e^{(-tp^\alpha)} = e^{-tp^\alpha} = e^{-(pu)^\alpha} = \Psi_\alpha(pu),$$

which means H belongs to the Weibull family.

2.2 Generalized Extreme Value Distribution

Later in this paper, these three distributions may be compressed into the generalized extreme value distribution. The cumulative distribution function is then given by

$$F_\xi(s) = \begin{cases} e^{-(1+\xi s)^{\frac{-1}{\xi}}}, & \xi \neq 0 \\ e^{-e^{-s}}, & \xi = 0 \end{cases}.$$

The support is understood here and henceforth to be (a, b) where $a = -1/\xi$, $b = \infty$ for $\xi > 0$; $a = -\infty$, $b = -1/\xi$ for $\xi < 0$; and $a = -\infty$, $b = \infty$ for $\xi = 0$. Observe that the second expression is the limit of the first expression as $\xi \to 0$. In that case the generalized extreme value distribution is the Gumbel. The Fréchet arises if $\xi > 0$, in which case it can be retrieved through the change of variables $\xi = \alpha^{-1}$ and $y = 1 + \xi x$. If $\xi < 0$ then the generalized distribution becomes the reversed Weibull, the appropriate change of variables being $\xi = -\alpha^{-1}$ and $y = -1 - \xi x$. The density is given by

$$f_\xi(s) = \begin{cases} (1+\xi s)^{-\frac{1}{\xi}-1} e^{-(1+\xi s)^{\frac{-1}{\xi}}}, & \xi \neq 0 \\ e^{-s} e^{-e^{-s}}, & \xi = 0. \end{cases}$$

Once the aforementioned change of variables has been made, the $-1/\xi$ terms become 0 in the definitions of a, b.

3 Stein's Method

Stein's method is a collection of techniques to get bounds on the distance between probability distributions. We measure the distance between distributions with respect to certain probability metrics. Given measures μ and ν, we measure distance using an appropriate class of test functions $\mathcal{H}$ by the formula

$$d_{\mathcal{H}}(\mu, \nu) = \sup_{h \in \mathcal{H}} \left| \int h(x) d\mu(x) - \int h(x) d\nu(x) \right|.$$

To improve clarity, we abuse notation and write these metrics as

$$d_{\mathcal{H}}(W, Z) = \sup_{h \in \mathcal{H}} |E[h(W)] - E[h(Z)]|$$

where the random variables W and Z have laws μ and ν respectively. The family $\mathcal{H}$ dictates which particular metric is being evaluated. When studying continuous distributions, we normally look at the Wasserstein metric, which is found by letting $\mathcal{H}$ be the set of all Lipschitz continuous functions with Lipschitz constant 1. The first step of Stein's method is to rewrite the right side of the above equation into something that is easier to work with.

To accomplish this, we find a characterizing operator $\mathcal{A}$ for our target distribution $\mathscr{L}(Z)$ such that

$$E[\mathcal{A}f(W)] = 0 \Longleftrightarrow W =_d Z$$

for all f in a distribution determining class of functions $\mathcal{F}$. Here, $W =_d Z$ means that W is equal in distribution to Z. Normally there are infinitely many operators for a distribution, and the choice of which to work with has impact later in the method. Once we have defined our characterizing operator, we set up and solve the Stein equation. For each $h \in \mathcal{H}$, the Stein equation is

$$\mathcal{A}f(x) = h(x) - E[h(Z)],$$

and if we've chosen a suitable operator, we can find a solution f_h that belongs to the class of functions $\mathcal{F}$ attached to the operator. If we plug in our random variable W and take expectations, we get that

$$E[\mathcal{A}f_h(W)] = E[h(W)] - E[h(Z)],$$

and then by taking suprema and absolute values we find that

$$\sup_{h\in\mathcal{H}}|E[\mathcal{A}f_h(W)]| = \sup_{h\in\mathcal{H}}|E[h(W)] - E[h(Z)]| = d_{\mathcal{H}}(W,Z). \tag{10}$$

Here we've redefined the desired distance as being dependent on W by incorporating Z into the choice of operator, and in turn f_h. When employing Stein's method, the hope is that if W is close to Z, then given that $E[\mathcal{A}f(Z)] = 0$, $E[\mathcal{A}f_h(W)]$ should be close to 0.

Given that Stein's method was originally conceived with the standard normal distribution in mind, it's natural to introduce the classic characterizing operator for the normal distribution as an example. Stein's Lemma, given in Chen et al. (2010), defines the most notable operator, stating that

$$E[f'(W) - Wf(W)] = 0 \iff W \sim \mathcal{N}(0,1), \text{ the standard normal distribution,}$$

for all absolutely continuous f with $E|f(Z)|, E|f'(Z)| < \infty$, $Z \sim \mathcal{N}(0,1)$. Writing $\Phi(x) = E[h(Z)]$, we can set up the Stein equation, finding that

$$f'(w) - wf(w) = h(w) - \Phi(h)$$

has the bounded solution

$$f_h(w) = e^{\frac{w^2}{2}} \int_w^\infty e^{-\frac{t^2}{2}} (\Phi(h) - h(t))dt,$$

where f_h satisfies

$$||f_h||_\infty \leq 2||h'||_\infty, \qquad ||f_h'||_\infty \leq \frac{2}{\pi}||h'||_\infty, \quad \text{and} \quad ||f_h''||_\infty \leq 2||h'||_\infty$$

if h is absolutely continuous. Following the general process given in Equation (10), we get that

$$d_{\mathcal{H}}(W,Z) = \sup_{h\in\mathcal{H}}|E[f_h'(W) - Wf_h(W)]|.$$

After setting up and solving the Stein equation, the next step in Stein's method is to bound the right side of Equation (10). At this point, the method splinters into several techniques for generating bounds.

4 General Methods of Finding Operators

Given a particular distribution, the first step in Stein's method is to define a characterizing operator. We start off by going over some basic ideas and methods.

4.1 Density Method

The most common method to find an operator for any given distribution is the density method, which notably generates Stein's operator for the normal distribution. Suppose that Z is absolutely continuous with continuously differentiable density $p(x)$, which is supported on $[a, b]$ and strictly positive on (a,b) for some $-\infty \le a < b \le \infty$. Let $\mathcal{F}$ be a class of test functions for which the expectations below exist. Given $f \in \mathcal{F}$, the density method gives us the operator

$$(\mathcal{A}f)(x) = f'(x) + \frac{p'(x)}{p(x)} f(x).$$

We can prove that $E[(\mathcal{A}f)(Z)] = 0$ by taking expectations and finding that

$$\begin{aligned} E[\mathcal{A}f(Z)] &= E\left(f'(Z)\right) + E\left(\frac{p'(Z)}{p(Z)} f(Z)\right) \\ &= \int_a^b f'(x)p(x)dx + \int_a^b f(x)p'(x)dx \\ &= f(x)p(x)|_a^b - \int_a^b f(x)p'(x)dx + \int_a^b f(x)p'(x)dx \\ &= 0. \end{aligned}$$

Here we are assuming that $\lim_{x \searrow a} p(x)f(x) = 0 = \lim_{x \nearrow a} p(x)f(x)$. This is a consistent way of finding operators, but oftentimes using the operator from the density method leads to problems later on. As a result, it is still worthwhile to explore other methods.

4.2 Generalized Density Method

Under the same initial assumptions as the density method, also let $\varphi(x)$ be a continuously differentiable function which is nonzero on (a,b) and satisfies $\lim_{x \searrow a} \varphi(x)p(x)f(x) = 0 = \lim_{x \nearrow b} \varphi(x)p(x)f(x)$ for all f in a distribution determining class $\mathcal{F}$. Then we can compute

$$\begin{aligned} E[\varphi'(Z)f(Z)] &= \int_a^b \varphi'(x)f(x)p(x)dx \\ &= \varphi(x)f(x)p(x)|_a^b - \int_a^b \varphi(x)[f'(x)p(x) + f(x)p'(x)]dx \\ &= -\int_a^b \left[f'(x) + \frac{p'(x)}{p(x)}f(x)\right] \varphi(x)p(x)dx \\ &= -E[G_f(Z)\varphi(Z)] \end{aligned} \tag{11}$$

where $G_f(x) = f'(x) + \frac{p'(x)}{p(x)} f(x)$. If we define an operator $(\mathcal{A}f)(x) = \varphi'(x)f(x) + G_f(x)\varphi(x)$, then by the above computations $E[(\mathcal{A}f)(Z)] = 0$. The density method arises from this method by setting $\varphi(x) = 1$, so this approach might be termed the *generalized density method*. This method is outlined in Ley et al. (2017). For $\mathcal{H}$ a family of functions determining a probability metric $d_{\mathcal{H}}(X,Y) = \sup_{h \in \mathcal{H}} |E[h(X)] - E[h(Y)]|$, the associated Stein equation is

$$\begin{aligned} f'(x)\varphi(x) + \left[\varphi'(x) + \frac{p'(x)}{p(x)}\varphi(x)\right] f(x) &= \varphi'(x)f(x) + \left[f'(x) + \frac{p'(x)}{p(x)}f(x)\right]\varphi(x) \\ &= (\mathcal{A}f)(x) \\ &= h(x) - \mu_h \end{aligned}$$

where $\mu_h = E[h(Z)]$. If we define $P(x) = \frac{\varphi'(x)}{\varphi(x)} + \frac{p'(x)}{p(x)}$ and $Q(x) = \frac{h(x)-\mu_h}{\varphi(x)}$, we can rewrite the above equality as

$$f'(x) + P(x)f(x) = Q(x).$$

Seeing that it's a linear ODE with integrating factor $e^{\int P(x)dx} = e^{\log(\varphi(x)p(x))} = \varphi(x)p(x)$, we reduce the equation to

$$\frac{d}{dx}(f(x)p(x)\varphi(x)) = Q(x)\varphi(x)p(x) = (h(x) - \mu_h)p(x).$$

The general solution to this equation is thus

$$f_h(x) = \frac{1}{\varphi(x)p(x)} \int_a^x (h(t) - \mu_h)p(t)dt. \tag{12}$$

4.3 Generalized Bounds

We typically want to choose our operator so that f_h has desirable properties, as Stein did in the normal case discussed earlier. This will enable us to determine $\mathcal{F}$ such that if $E[\mathcal{A}f(W)] = 0$ for all $f \in \mathcal{F}$, then $W =_d Z$. In this section, we rewrite the general solution given in Equation (12) into a form that is oftentimes easier to bound. Working in the Wasserstein metric, we can compute that

$$\begin{aligned}
h(t) - \mu_h &= \int_a^b [h(t) - h(x)]p(x)dx \\
&= \int_a^t \int_x^t h'(u)p(x)dudx - \int_t^b \int_t^x h'(u)p(x)dudx \\
&= \int_a^t \int_a^u h'(u)p(x)dxdu - \int_t^b \int_u^b h'(u)p(x)dxdu \\
&= \int_a^t h'(u)F(u)du - \int_t^b h'(u)(1 - F(u))du,
\end{aligned}$$

where $F(u) = \int_a^u p(x)dx$ is the distribution function for Z. We can plug this into our integral in the general solution to get

$$\begin{aligned}
\int_a^x (h(t)-\mu_h)p(t)dt &= \int_a^x \int_a^t h'(u)F(u)p(t)dudt - \int_a^x \int_t^b h'(u)(1-F(u))p(t)dudt \\
&= \int_a^x \int_u^x h'(u)F(u)p(t)dtdu - \int_a^x \int_a^u h'(u)(1-F(u))p(t)dtdu \\
&\qquad - \int_b^x \int_a^x h'(u)(1-F(u))p(t)dtdu \\
&= \int_a^x h'(u)F(u)(F(x)-F(u))du - \int_a^x h'(u)(1-F(u))F(u)du \\
&\qquad - \int_x^b h'(u)(1-F(u))F(x)du \\
&= \int_a^x h'(u)F(u)F(x)du - \int_a^x h'(u)F(u)du \\
&\qquad - \int_x^b h'(u)F(x)du + \int_x^b h'(u)F(u)F(x)du \\
&= (F(x)-1)\int_a^x h'(u)F(u)du + F(x)\int_x^b h'(u)(F(u)-1)du.
\end{aligned}$$

Since $|h'(u)| \leq 1$ almost everywhere, we have that

$$|f_h(x)| \leq \frac{1}{|\varphi(x)|p(x)}\left[(1-F(x))\int_a^x F(u)du + F(x)\int_x^b (1-F(u))du\right].$$

We also want to rewrite the general form of $f_h'(x)$ and $f_h''(x)$. We differentiate Equation (12) and find that

$$f_h'(x) = \frac{h(x)-\mu_h}{\varphi(x)} - \frac{\varphi'(x)p(x)+\varphi(x)p'(x)}{\varphi(x)^2 p(x)^2}\int_a^x (h(t)-\mu_h)p(t)dt$$

and then substitute f_h back in to get

$$f_h'(x) = \frac{h(x)-\mu_h}{\varphi(x)} - \left(\frac{\varphi'(x)}{\varphi(x)} + \frac{p'(x)}{p(x)}\right)f_h(x).$$

We can differentiate the Stein equation to see that

$$\begin{aligned}
h'(x) = f''(x)\varphi(x) + f_h'(x)\varphi'(x) + \left(\varphi''(x)\frac{p''(x)p(x)-p'(x)^2}{p(x)^2}\varphi(x) + \frac{p'(x)}{p(x)}\varphi'(x)\right)f_h(x) \\
+ \left(\varphi'(x) + \frac{p'(x)}{p(x)}\varphi(x)\right)f_h'(x)
\end{aligned}$$

and isolating the desired $f_h''(x)$ we get

$$f_h''(x) = \frac{h'(x)}{\varphi(x)} - f_h'(x)\left(2\frac{\varphi'(x)}{\varphi(x)} + \frac{p'(x)}{p(x)}\right) - f_h(x)\left(\frac{\varphi''(x)}{\varphi(x)} + \frac{p''(x)p(x)-p'(x)^2}{p(x)^2} + \frac{p'(x)\varphi'(x)}{p(x)\varphi(x)}\right).$$

Now that we've accomplished a reasonable amount in generality, it's time to employ these methods with regards to the distributions we're focusing on.

5 Operator for the Extreme Value Distributions

Bartholmé and Swan (2013) define a particularly useful operator for the Fréchet, and it can be derived using the generalized density method. The Fréchet has density $p(y) = y^{-1-\alpha}e^{-y^{-\alpha}}$, and $p'(y) = y^{-2-2\alpha}e^{-y^{-\alpha}}(\alpha - y^\alpha - \alpha y^\alpha)$. The operator is given by the generalized density method when we let $\varphi(y) = y^{1+\alpha}$, noting that $\varphi'(y) = (\alpha+1)y^\alpha$. We compute that

$$\begin{aligned}\varphi'(y) + \frac{p'(y)}{p(y)}\varphi(y) &= (\alpha+1)y^\alpha + \frac{y^{-2-2\alpha}e^{-y^{-\alpha}}(\alpha - y^\alpha - \alpha y^\alpha)}{y^{-1-\alpha}e^{-y^{-\alpha}}}(y^{1+\alpha}) \\ &= \alpha y^\alpha + y^\alpha + y^{-1-\alpha}y^{1+\alpha}(\alpha - y^\alpha - \alpha y^\alpha) \\ &= \alpha.\end{aligned}$$

Remember that our test functions must satisfy

$$\lim_{y \searrow a} \varphi(y) f(y) p(y) = 0 = \lim_{y \nearrow b} \varphi(y) f(y) p(y).$$

In this instance our restrictions become

$$\lim_{y \searrow 0} e^{-y^{-\alpha}} f(y) = 0$$

and

$$\lim_{y \nearrow \infty} f(y) = \lim_{y \nearrow \infty} e^{-y^{-\alpha}} f(y) = 0.$$

We let $\mathcal{F} = AC_0(a,b)$, the space of absolutely continuous functions on (a,b) which vanish at $a = 0$ and $b = \infty$. The following result is stated in Bartholmé and Swan (2013). We provide a proof for completeness.

Theorem 1. *Define the functional operator $\mathcal{A}$ by*

$$(\mathcal{A}f)(y) = y^{\alpha+1}f'(y) + \alpha f(y).$$

Then $E[(\mathcal{A}f)(Y)] = 0$ for all $f \in \mathcal{F}$ if and only if Y has the Fréchet distribution.

Proof: Letting Z have the Fréchet distribution, we clearly have that $E[(\mathcal{A}f)(Z)] = 0$ for all $f \in \mathcal{F}$ given the calculations in Equation (11).

Letting $E[(\mathcal{A}f)(Y)] = 0$ for all $f \in \mathcal{F}$, we first obtain the unique bounded solution f_x of the differential equation

$$y^{\alpha+1}f_x'(y) + \alpha f_x(y) = \mathbb{1}[y \le x] - \Phi_\alpha(x),$$

where $\Phi_\alpha(x)$ is the c.d.f. of the Fréchet distribution. After dividing by $y^{\alpha+1}$, we find that the integrating factor is $e^{\int \alpha y^{-\alpha-1}dy} = e^{-y^{-\alpha}}$. Next, we see that f_x is given by

$$f_x(y) = e^{y^{-\alpha}} \int_0^y (\mathbb{1}[y \le x] - \Phi_\alpha(x)) t^{-\alpha-1} e^{-t^{-\alpha}} dt.$$

By definition, we have that

$$|P(Y \leq x) - \Phi(x)| = |E[y^{\alpha+1} f_x'(Y) + \alpha f_x(Y)]|.$$

By our assumptions, the right side is 0, and as a result Y has the Fréchet distribution. $\square$

We also want to solve the Stein equation. Similar to above, we find that the solution is

$$f_h(y) = e^{y^{-\alpha}} \int_0^y (h(t) - \mu_h) t^{-\alpha-1} e^{-t^{-\alpha}} dt.$$

This holds for general test functions h, not just $h(y) = 1\{y \leq x\}$, the test functions corresponding to the Kolmogorov metric which gives the sup norm of the difference between cdfs.

5.1 GEVD Operator

We can use the generalized extreme value distribution to extend the operator to apply to the Weibull and Gumbel distributions. The generalized extreme value distribution has density

$$p(x) = (1 + \xi x)^{(-1-\frac{1}{\xi})} \exp(-(1 + \xi x)^{-\frac{1}{\xi}})$$

with

$$p'(x) = \left[1 - (\xi + 1)(1 + \xi x)^{\frac{1}{\xi}}\right] (1 + \xi x)^{-2-\frac{2}{\xi}} \exp(-(1 + \xi x)^{-\frac{1}{\xi}})$$

and

$$\frac{p'(x)}{p(x)} = \left[1 - (\xi + 1)(1 + \xi x)^{\frac{1}{\xi}}\right] (1 + \xi x)^{-1-\frac{1}{\xi}}.$$

We perform a change of variables and let $\varphi(x) = (1 + \xi x)^{1+\frac{1}{\xi}}$, with derivative $\varphi'(x) = (\xi + 1)(1 + \xi x)^{\frac{1}{\xi}}$. The associated operator is then

$$\begin{aligned}
(\mathcal{A}f)(x) &= \varphi'(x) f(x) + \left[f'(x) + \frac{p'(x)}{p(x)} f(x)\right] \varphi(x) \\
&= (\xi + 1)(1 + \xi x)^{\frac{1}{\xi}} f(x) + \left[(1 + \xi x)^{1+\frac{1}{\xi}} f'(x) + \left[1 - (\xi + 1)(1 + \xi x)^{\frac{1}{\xi}}\right] f(x)\right] \\
&= (1 + \xi x)^{1+\frac{1}{\xi}} f'(x) + f(x).
\end{aligned}$$

We plug into Equation (12) to see that the general solution is

$$f_h(x) = e^{(1+\xi x)^{-1/\xi}} \int_a^x (h(t) - \mu_h)(1 + \xi t)^{-1-1/\xi} e^{(1+\xi t)^{-1/\xi}} dt.$$

Next, we will apply this operator to the Weibull and Gumbel.

5.2 Reversed Weibull

In the reversed Weibull case, if we let $\alpha = -\xi^{-1}$, $y = -(1+\xi x)$ we have that $\varphi(x) = (-y)^{1-\alpha}$ with $\varphi'(x) = (\alpha-1)(-y)^{-\alpha}$, noting that

$$p(y) = \alpha(-y)^{\alpha-1}e^{-(-y)^\alpha} \text{ and } p'(y) = e^{-(-y)^\alpha}(-\alpha(-y)^{\alpha-2}(\alpha-1) + \alpha^2(-y)^{2\alpha-2}).$$

Then we can calculate

$$\begin{aligned}\varphi'(y) + \frac{p'(y)}{p(y)}\varphi(y) &= (\alpha-1)(-y)^{-\alpha} + \left((1-\alpha)(-y)^{-1} + \alpha(-y)^{\alpha-1}\right)(-y)^{1-\alpha} \\ &= (\alpha-1)(-y)^{-\alpha} + (1-\alpha)(-y)^{-\alpha} + \alpha \\ &= \alpha.\end{aligned}$$

Our test functions must satisfy

$$\lim_{y\searrow a}\varphi(y)f(y)p(y) = 0 = \lim_{y\nearrow b}\varphi(y)f(y)p(y).$$

Here, we have that $\varphi(y)p(y) = \alpha e^{-(-y)^\alpha}$. As a result, our restrictions become

$$\lim_{y\searrow-\infty} f(y) = \lim_{y\searrow-\infty} f(y)\alpha e^{-(-y)^\alpha} = 0$$

and

$$\lim_{y\nearrow 0} f(y)\alpha e^{-(-y)^\alpha} = 0.$$

By a slight abuse of notation, we let $\mathcal{F} = AC_0(a,b)$ as before, keeping in mind that a and b vary according to type.

Theorem 2. *Define the functional operator $\mathcal{A}$ by*

$$(\mathcal{A}f)(y) = (-y)^{1-\alpha}f'(y) + \alpha f(y).$$

Then $E[(\mathcal{A}f)(Y)] = 0$ for all $f \in \mathcal{F}$ if and only if Y has the Weibull distribution.

The proof of this theorem is analogous to the proof in the Fréchet case. To solve the Stein equation we divide out by $(-y)^{1-\alpha}$, and find that our integrating factor is $e^{\int \alpha(-y)^{\alpha-1}dy} = e^{-(-y)^\alpha}$, and our solution is

$$f_h(y) = e^{(-y)^\alpha}\int_{-\infty}^{y}(h(t)-\mu_h)e^{-(-t)^\alpha}(-t)^{\alpha-1}dt.$$

5.3 Gumbel

The Gumbel distribution arises from the generalized extreme value distribution when we send ξ to 0. The density of the Gumbel is

$$\begin{aligned}p(x) &= \lim_{\xi\to 0}(1+\xi x)^{-1-\xi^{-1}}e^{-(1+\xi x)^{-\xi^{-1}}} \\ &= e^{-x-e^{-x}}.\end{aligned}$$

We also note that

$$p'(x) = (-1 + e^{-x})e^{-x-e^{-x}}.$$

Next, we take the limit of $\varphi(x)$ to obtain

$$\begin{aligned}
\lim_{\xi\to 0} \varphi(x) &= \lim_{\xi\to 0}(1+\xi x)^{1+\xi^{-1}} \\
&= \exp\left(\lim_{\xi\to 0}(1+\xi^{-1})\log(1+\xi x)\right) \\
&= \exp\left(\lim_{\xi\to 0} \frac{\log(1+\xi x)}{\frac{\xi}{\xi+1}}\right) \\
&= \exp\left(\lim_{\xi\to 0} \frac{\frac{x}{1+\xi x}}{\frac{1}{(\xi+1)^2}}\right) \\
&= e^x.
\end{aligned}$$

This also gives us $\varphi'(x) = e^x$. We can now start to work towards our operator, first finding that

$$\begin{aligned}
\varphi'(x) + \frac{p'(x)}{p(x)}\varphi(x) &= e^x + \left(\frac{(-1+e^{-x})e^{-x-e^{-x}}}{e^{-x-e^{-x}}}\right)e^x \\
&= e^x + (-1+e^{-x})e^x \\
&= 1.
\end{aligned}$$

Our test functions must satisfy

$$\lim_{x\searrow a}\varphi(x)f(x)p(x) = 0 = \lim_{x\nearrow b}\varphi(x)f(x)p(x).$$

Here, our restrictions simply become

$$\lim_{x\searrow -\infty} e^{-e^{-x}}f(x) = 0 = \lim_{x\nearrow\infty} e^{-e^{-x}}f(x),$$

which works if we let $\mathcal{F} = AC_0(\mathbb{R})$.

Theorem 3. *Define the functional operator $\mathcal{A}$ by*

$$(\mathcal{A}f)(y) = e^x f'(y) + f(x).$$

Then $E[(\mathcal{A}f)(Y)] = 0$ for all $f \in \mathcal{F}$ if and only if Y has the Gumbel distribution.

Again, the proof of this theorem is similar to the proof given in the Fréchet case. We solve the Stein equation by first dividing out by e^x and seeing that the integrating factor is $e^{\int e^{-x}} = e^{-e^{-x}}$. The solution is

$$f_h(x) = e^{-e^{-x}}\int_{-\infty}^{x}(h(t)-\mu_h)e^{-t-e^{-t}}dt.$$

6 Employing Two Operators

In the proof of the Fisher-Tippett-Gnedenko Theorem in Section 2.1, we discussed the convergence of $W_n = (M_n - b_n)/a_n$, where W_n had distribution $G_n(x) = F(a_n x + b_n)^n$ and density $p_n(x) = n a_n F(a_n x + b_n)^{n-1} f(a_n x + b_n)$. Now, we want to construct an operator for G_n and use it in conjunction with an operator for our target distribution.

First, let Z have our target distribution with density $q(x)$. Let $\mathcal{A}$ be the Stein operator for the target distribution, chosen with $\varphi_1(x)$ in the generalized density method, and $\mathcal{B}_n$ be the Stein operator for G_n, chosen with $\varphi_2(x)$ in the generalized density method. Then we have that

$$(\mathcal{A}f)(x) = \varphi_1(x) f'(x) + \left(\varphi_1'(x) + \frac{q'(x)}{q(x)}\right) f(x)$$

and

$$\begin{aligned}(\mathcal{B}_n f)(x) &= \varphi_2(x) f'(x) + \left(\varphi_2'(x) + \frac{p_n'(x)}{p_n(x)}\right) f(x) \\ &= \varphi_2(x) f'(x) + \left(\varphi_2'(x) + \rho_n(x)\right) f(x),\end{aligned}$$

where $\rho_n(x) = \frac{d}{dx}\log(p_n(x))$ is the score function for W_n. When applying this method, we can simplify the general form for $\rho_n(x)$, seeing that

$$\begin{aligned}\rho_n(x) &= \frac{n(n-1)a_n^2 F(a_n x + b_n)^{n-2} f(a_n x + b_n)^2 + n a_n^2 F(a_n x + b_n)^{n-1} f'(a_n x + b_n)}{n a_n F(a_n x + b_n)^{n-1} f(a_n x + b_n)} \\ &= a_n \left(\frac{(n-1) f(a_n x + b_n)}{F(a_n x + b_n)} + \frac{f'(a_n x + b_n)}{f(a_n x + b_n}\right).\end{aligned} \tag{13}$$

Next, we want to combine the two operators into one, getting that

$$E[h(Z) - h(W_n)] = E[(\mathcal{A}f_h)(W_n)] = E[(\mathcal{A}f_h)(W_n)] - E[(\mathcal{B}_n f_h)(W_n)] = E[\mathcal{C}_n f_h(W_n)],$$

where
After reaching this point for a particular W_n, we would then examine $|E[(\mathcal{C}_n f_h)(W_n)]|$. If we wanted to pick $\varphi_1(x)$ and $\varphi_2(x)$ such that the f' term is eliminated, it's simple enough. Just choose $\varphi_1(x) = \varphi_2(x)$. We can then work with any $\varphi(x)$ and say that

$$\mathcal{C}_n f(x) = \varphi(x) \left(\frac{q'(x)}{q(x)} - \rho_n(x)\right) f(x). \tag{14}$$

However, when we compute $|E[(\mathcal{C}_n f_h)(x)]|$, with this method, referencing Equation (12) for the value of f_h, we get

$$|E[(\mathcal{C}_n f_h)(x)]| = \left| E\left[\varphi(x)\left(\frac{q'(x)}{q(x)} - \rho_n(x)\right) \frac{1}{\varphi(x)p(x)} \int_a^x (h(t) - \mu_h)p(t)dt\right]\right|. \tag{15}$$

There is cancellation with the $\varphi(x)$ terms, so when employing this method the choice of $\varphi(x)$ is irrelevant, and the result is the same as when employing the density method for both operators (though the choice still impacts the class of test functions for the two operators $\mathcal{A}$ and $\mathcal{B}_n$, as well as the properties of the solutions to the associated Stein equations). Bartholmé and Swan (2013) used a similar strategy along with the operator given earlier to compute rates of convergence of the maxima of Pareto random variables to the Fréchet distribution.

7 Examples of Operators

In this section, we begin to apply the method discussed in the previous sections to the maxima of random variables with various distributions. Further work needs to be done in terms of bounding various integrals in order to obtain convergence rates.

7.1 Exponential

Let $X_1, X_2, ..., X_n$ be a sequence of random variables with the standard exponential distribution. Then $M_n = \max(X_1, X_2, ..., X_n)$ has distribution $F^n(x) = (1 - e^{-x})^n$. After applying the normalizing constants $a_n = 1$ and $b_n = \log n$ we see that

$$F^n(a_n x + b_n) = (1 - e^{-x-\log n})^n = \left(1 - \frac{e^{-x}}{n}\right)^n \to \Lambda(x)$$

as $n \to \infty$. Note that $f(a_n x + b_n) = e^{-x}/n$ and $f'(a_n x + b_n) = -e^{-x}/n$. Next, we find $\rho_n(x)$ using Equation (13), getting

$$\begin{aligned}\rho_n(x) &= \frac{(n-1)\frac{e^{-x}}{n}}{1 - \frac{e^{-x}}{n}} - 1 \\ &= \frac{n-1}{ne^x - 1} - 1.\end{aligned}$$

We can look back at Equation (15) to generate an operator. First we let $q(x)$ be the p.d.f. of the Gumbel distribution, getting that $q'(x)/q(x) = e^{-x} - 1$. Now we see that

$$(\mathcal{C}_n f)(x) = f'(x)\left(\varphi_1(x) - \varphi_2(x)\right) + f(x)\left(\varphi_1'(x) - \varphi_2'(x) + \varphi_1(x)(e^{-x} - 1) - \varphi_2(x)\left(\frac{n-1}{ne^x - 1} - 1\right)\right).$$

If we let $\varphi_1(x) = \varphi_2(x)$ then we see that

$$(\mathcal{C}_n f)(x) = \varphi(x)\left(e^{-x} - \frac{n-1}{ne^x - 1}\right) f(x),$$

where $\varphi(x)$ can be tuned to change the class of test functions and the solution f_h.

7.2 Pareto

Let $X_1, X_2, ..., X_n$ be a sequence of random variables with the standard Pareto distribution. Then $M_n = \max(X_1, X_2, ..., X_n)$ has distribution $F^n(x) = (1 - x^{-\alpha})$ for $x \geq 1$. The maxima of Pareto random variables converge to the Fréchet distribution by letting $a_n = n^{\alpha^{-1}}$ and $b_n = 0$. From Equation (13) we see that

$$\begin{aligned}\rho_n(x) &= n^{\alpha^{-1}}(n-1)\alpha x^{-\alpha-1} n^{-1-\alpha^{-1}}\left(1 - \frac{x^{-\alpha}}{n}\right)^{-1} + \left(n^{\alpha^{-1}}\right)\frac{\alpha(-\alpha-1)x^{-\alpha-2}n^{-1-2\alpha^{-1}}}{\alpha x^{-\alpha-1}n^{-1-\alpha^{-1}}} \\ &= \frac{n-1}{n}\alpha x^{-1-\alpha}\left(1 - \frac{x^{-\alpha}}{n}\right)^{-1} - \frac{\alpha+1}{x}.\end{aligned}$$

Bartholmé and Swan (2013) developed $O(1/n)$ rates of convergence by taking an equivalent approach with $\varphi(x) = x^{1+\alpha}$.

7.3 Cauchy

Let $X_1, X_2, ..., X_n$ be a sequence of random variables with the standard Cauchy distribution. Then $M_n = \max(X_1, X_2, ..., X_n)$ has distribution $F^n(x) = (\pi^{-1}\arctan(x) + \frac{1}{2})^n$. The maxima of random variables with the Cauchy distribution converge to the Fréchet distribution by letting $a_n = n/\pi$ and $b_n = 0$. We use Equation (13) to compute

$$\begin{aligned}\rho_n(x) &= \frac{n}{\pi}\left(\frac{(n-1)\pi^{-1}\left(1+(\frac{nx}{\pi})^2\right)^{-1}}{\pi^{-1}\arctan(x)+\frac{1}{2}} - \frac{\pi^{-2}2xn\left(1+(\frac{nx}{\pi})^2\right)^{-2}}{\pi^{-1}\left(1+(\frac{nx}{\pi})^2\right)^{-1}}\right)\\ &= \frac{n}{\pi}\left(\frac{\pi^{-1}(n-1)\left(1+(\frac{nx}{\pi})^2\right)^{-1}}{\pi^{-1}\arctan(x)+\frac{1}{2}} - \frac{\pi^{-1}2xn}{1+(\frac{nx}{\pi})^2}\right).\end{aligned}$$

Thus, our operator would be of the form

$$\begin{aligned}(\mathcal{C}_n f)(x) &= f'(x)\left(\varphi_1(x)-\varphi_2(x)\right)\\ &+ f(x)\left(\varphi_1'(x)-\varphi_2'(x)+\varphi_1(x)(e^{-x}-1)-\varphi_2(x)\frac{n}{\pi}\left(\frac{\pi^{-1}(n-1)\left(1+(\frac{nx}{\pi})^2\right)^{-1}}{\pi^{-1}\arctan(x)+\frac{1}{2}} - \frac{\pi^{-1}2xn}{1+(\frac{nx}{\pi})^2}\right)\right).\end{aligned}$$

References

Bartholmé C, Swan Y. 2013. Rates of convergence towards the Fréchet distribution.

Chen LH, Goldstein L, Shao QM. 2010. Normal approximation by Stein's method. Probability and Its Applications. Berlin Heidelberg: Springer.

Ley C, Reinert G, Swan Y. 2017. Stein's method for comparison of univariate distributions. Probability Surveys 14:1–52.

Resnick SI. 1987. Extreme values, regular variation, and point processes. Springer Series in Operations Research and Financial Engineering. New York: Springer.