The Fisher–Tippett–Gnedenko Theorem

Below is a proof of the Fisher-Tippett-Gnedenko Theorem, which is also referred to as the extreme value theorem in probability. A version of this proof can be found in section 14 of Billingsley's *Probability and Measure* [1]. This proof is presented below with more detail and some clarifying comments, along with all of the necessary intermediate results.

Prerequisite Definitions and Results

Distribution Types

Distribution functions F and G are of the same type if there exist constants $a > 0$ and b such that $F(ax+b) = G(x)$ for all x. Informally, two distributions are of the same type if they are equivalent up to scaling and shifting.

Extremal Distributions

A distribution function is extremal if it is nondegenerate and if, for some distribution function G and constants $a_n > 0$ and b_n,

$$G^n(a_n x + b_n) \xrightarrow{n} F(x)$$

The motivation for this definition is that if random variables $X_i \overset{iid}{\sim} G$ for $i = 1, ..., n$, then $F_{max(\{X_i\})}(a_n x + b_n) = G^n(a_n x + b_n)$, and so extremal distributions are the limiting distribution types of normalized maxima of i.i.d. random variables. The main result characterizes the possible types of extremal distributions, of which there are only three.

Useful Results

Three intermediate results are required for the below proof of the main result, and these are presented here without proof. The following lemma and theorem are from section 14 of Billingsley.

Lemma 5. If $F(x) = F(ax+b)$ for all x and F is nondegenerate, then $a = 1$ and $b = 0$.

Theorem 14.2 Suppose that $F_n(u_n x + v_n) \to^n F(x)$ and $F_n(a_n x + b_n) \to G(x)$, where $u_n > 0$, $a_n > 0$, and F and G are nondegenerate. Then there exist $a > 0$ and b such that $a_n/u_n \to a$, $(b_n - v_n)/u_n \to b$, and $F(ax + b) = G(x)$.

Theorem 14.2 is an important result on its own, as it says that limiting distributions are unique up to type. In other words, a sequence of distribution functions cannot converge to multiple distribution functions of different types.

The following general results related to Cauchy's equation will also be useful, and are stated and proven in the appendix of Billingsley.

Theorem A20 Let f be a real function on $(0, \infty)$, and suppose that f satisfies Cauchy's equation: $f(x+y) = f(x) + f(y)$ for $x, y > 0$. If there is some interval on which f is bounded above, then $f(x) = xf(1)$ for $x > 0$.

Corollary to Theorem A20 Let U be a real function on $(0, \infty)$ and suppose that $U(x + y) = U(x)U(y)$ for $x, y > 0$. Suppose further that there is some interval on which U is bounded above. Then either $U(x) = 0$ for $x > 0$, or else there is an A such that $U(x) = exp(Ax)$ for $x > 0$.

Theorem Statement

Let $X_i \overset{iid}{\sim} G$ for $i = 1, ..., n$. If there exist real sequences $a_n > 0$ and b_n and a nondegenerate distribution function F such that:

$$lim_{n \to \infty} P\left(\frac{max(\{X_i\}) - a_n}{b_n} \leq x\right) = F(x)$$

Or, equivalently:

$$lim_{n \to \infty} G^n(a_n x + b_n) = F(x)$$

Then F belongs to one of the Gumbel, Frechet, or reversed Weibull distribution types.

Proof

Assume that F is extremal. Then, by definition, for some distribution function G and for $n > 0$:

$$G^n(a_n x + b_n) \to F(x)$$

This implies the following two limiting results:

$$G^{nk}(a_n x + b_n) \to F^k(x)$$
$$G^{nk}(a_{nk} x + b_{nk}) \to F(x)$$

Therefore, by **theorem 14.2** there exists constants $c_k > 0$ and d_k such that:

$$F^k(x) = F(c_k x + d_k) \tag{1}$$

Clearly $c_1 = 1$ and $d_1 = 0$. Note that if there exists some $k_0 > 1$ for which $c_{k_0} \neq 1$, then there exists an x' such that $x' = c_{k_0} x' + d_{k_0}$, and so by (1) we have:

$$\begin{aligned} F^{k_0}(x') &= F(c_{k_0} x' + d_{k_0}) \\ \implies F^{k_0}(x') &= F(x') \\ \implies F(x') &\in \{0, 1\} \end{aligned} \tag{2}$$

And so F attains either 0 or 1. From (1), we derive the following two equalities:

$$\begin{aligned} F(c_{jk} x + d_{jk}) = F^{jk}(x) = F^j(c_k x + d_k) &= F(c_j(c_k x + d_k) + d_j) \\ &= F(c_j c_k x + c_j d_k + d_j) \end{aligned}$$

$$\begin{aligned} F(c_{jk} x + d_{jk}) = F^{jk}(x) = F^k(c_j x + d_j) &= F(c_k(c_j x + d_j) + d_k) \\ &= F(c_j c_k x + c_k d_j + d_k) \end{aligned}$$

Since each of these holds for all x, by **lemma 5** we get the equalities:

$$c_{jk} = c_j c_k \tag{3}$$
$$d_{jk} = c_j d_k + d_j = c_k d_j + d_k \tag{4}$$

We now consider three distinct (exhaustive) cases for the scaling constant sequence c_k, each of which implies a different type for F.

Case 1: $c_k = 1$ for all k

Suppose that $c_k = 1$ for all k. Then, by (1) we have that:

$$F^k(x) = F(x + d_k) \tag{5}$$

And therefore:

$$F(x) = F(x - d_k + d_k) = [F(x - d_k)]^k$$
$$\implies F(x - d_k) = F^{1/k}(x)$$

And so $F^{j/k} = F(x + d_j - d_k)$. Note that this agrees with (3) and (4). Let $r = j/k$, and let $\delta_r = d_j - d_k$. Then $F^r(x) = F(x + \delta_r)$. We assume that F is nondegenerate, and so there exists an x^* such that $0 < F(x^*) < 1$. Since $F^k(x^*)$ is decreasing in k, then by (5) $F(x^* + d_k)$ is decreasing in k and by the monotonicity of F it follows that d_k is decreasing in k. Similarly, δ_r is strictly decreasing in r.

We now have a functional equation involving F that is defined for positive rational numbers r. The next construction allows us to extend this equation to positive real numbers t. Define:

$$\varphi(t) = inf_{0<r<t}\delta_r$$

Where r is rational in the infimum. Since δ_r is decreasing in r, then $\varphi(t)$ is decreasing in t. We have:

$$F^t(x) = lim_{r\uparrow t}F^r(x) = F(x + lim_{r\uparrow t}\delta_r) = F(x + \varphi(t)) \tag{6}$$

Where r is rational in the limit. Note that the monotonicity (and boundedness) of the sequence δ_r implies existence of the limit. Applying the argument that yielded (3) and (4) from $F^{jk}(x)$ to $F^{ts}(x)$ yields:

$$\varphi(st) = \varphi(s) + \varphi(t)$$

It follows that $f(x) = \varphi(exp(x))$ satisfies $f(x+y) = f(x)+f(y)$ for positive real x and y. Therefore, by **theorem A20** we have that $f(x) = x * f(1) = x * \varphi(e)$. Setting $t = exp(x)$ yields:

$$\varphi(exp(x)) = x\varphi(e)$$
$$\implies \varphi(t) = -\beta ln(t)$$

For some β that is positive since $\varphi(t)$ is decreasing in t. Setting $t = exp(x/\beta)$ in (6) (which holds for all positive real t) yields:

$$F^{exp(x/\beta)}(x) = F(x + \varphi(exp(x/\beta)))$$
$$\implies F^{exp(x/\beta)}(x) = F(x - x) = F(0)$$
$$\implies log(F^{exp(x/\beta)}(x)) = log(F(0))$$
$$\implies exp(x/\beta)log(F(x)) = log(F(0))$$
$$\implies F(x) = exp(exp(-x/\beta)logF(0))$$

Therefore F has the same type as:

$$F_1(x) = exp(-exp(-x))$$

F_1 is the distribution function of a standard **Gumbel** random variable. Note that this distribution function does not attain the values 0 or 1, and so the Gumbel distribution is supported on the entire real line.

Case 2: $c_k \neq 1$ for some k, F attains 0

Now suppose that there exists a $c_k \neq 1$, such that there exists an x' for which $F(x') = 0$ (see (2)). Let $x_0 = sup\{x : F(x) = 0\}$ and without loss of generality assume that $x_0 = 0$. This is possible since $F(x + x')$ is of the same type as F. Then, if $d_k = 0$ for some value of k, there exists some x in a neighborhood of 0 such that $c_k x + d_k$ has a different sign than x. Then, one side of equation (1) is zero and the other is strictly positive, which is a contradiction. Therefore $d_k = 0$ for all k, and we have:

$$\begin{aligned} F^k(x) &= F(c_k x) \\ \implies F^{1/k}(x) &= F(\frac{x}{c_k}) \\ \implies F^{j/k}(x) &= F(\frac{c_j x}{c_k}) \end{aligned} \tag{7}$$

For positive rational $r = j/k$, set $\gamma_r = c_j/c_k$ such that $F^r(x) = F(\gamma_r x)$. Since F is nondegenerate, there is some x for which $0 < F(x) < 1$ and $F^k(x)$ is decreasing in k. This implies that c_k and is decreasing in k and thus γ_r is strictly decreasing in r.

We proceed with a similar construction as before, the strategy being to extend γ_r to the positive real numbers. Set $\psi(t) = inf_{0<r<t}\gamma_r$ for positive real t, so that (similar to as in (6)):

$$F^t(x) = F(\psi(t)x) \tag{8}$$

$$\implies F^{ts}(x) = F(\psi(t)\psi(s)x) \tag{9}$$

$$\implies \psi(ts) = \psi(t)\psi(s) \tag{10}$$

Letting $U(x) = \psi(exp(x))$, we have that U satisfies $U(x + y) = U(x)U(y)$, and so by the **corollary to theorem A20** it must be that $U(x) = \psi(exp(x)) = exp(Ax)$ for some A, and so setting $t = exp(x)$ we have that:

$$\psi(t) = exp(Alog(t)) = t^{-\xi}$$

For some $\xi > 0$. Then, setting $t = x^{1/\xi}$ in (8) yields:

$$F^{x^{1/\xi}}(x) = F(1)$$
$$\implies F(x) = exp(x^{-1/\xi} log(F(1)))$$

And therefore F is of the same type as:

$$F_2(x) = \begin{cases} 0 & if\ x < 0 \\ exp(-x^{-\alpha}) & if\ x \geq 0 \end{cases}$$

For $\alpha \equiv 1/\xi > 0$. F_2 is the distribution function of a standard **Frechet** random variable.

Case 3: $c_k \neq 1$ for some k, F attains 1

Now suppose that $c_k \neq 1$ for some k and F attains 1 (see (2)). As in the argument for case 2, we set $x_1 = inf\{x : F(x) = 1\}$ and suppose that $x_1 = 0$. Then $F(x) < 1$ for $x < 0$ and $F(x) = 1$ for $x \geq 0$. By the same argument made in case 2, it must be that $d_k = 0$ for all k so as to avoid contradicting (1). Therefore (7) still holds and for $\gamma_{j/k} = c_j/c_k$, we still have the result that $F^r(x) = F(\gamma_r x)$.

Now there exists some x for which $0 < F(x) < 1$, but by assumption $x < 0$ and so c_k is increasing in k. Using an analogous argument as in case 2, we extend γ_r to positive real t such that $F^t(x) = F(t^\xi x)$ for all x and all $t > 0$, and apply the corollary to theorem A20 to derive that $F(x) = exp((-x)^{1/\xi} log(F(-1))$ for $x < 0$ and $\xi < 0$. Then F is of the type:

$$F_3(x) = \begin{cases} exp(-(-x)^\alpha & if\ x \leq 0 \\ 1 & if\ x \geq 0 \end{cases}$$

For $\alpha \equiv -1/\xi > 0$. This distribution function characterizes the **reversed Weibull distribution**.

Comment

The three extremal distribution types can be considered as special cases of the **generalized extreme value distribution**, with distribution function:

$$F(x) = \begin{cases} exp\left([1 + \xi(\frac{x-\mu}{\sigma})]^{-1/\xi}\right) & if\ \xi \neq 0 \\ exp\left(exp(-\frac{x-\mu}{\sigma})\right) & if\ \xi = 0 \end{cases}$$

Where the support is given by:

$$X \in \begin{cases} [\mu - \sigma/\xi, \infty) & if\ \xi > 0 \\ (-\infty, \infty) & if\ \xi = 0 \\ (-\infty, \mu - \sigma/\xi] & if\ \xi < 0 \end{cases}$$

Clearly $\xi = 0$ corresponds to a Gumbel type distribution, $\xi > 0$ corresponds to a Frechet type distribution, and $\xi < 0$ corresponds to a reversed Weibull type distribution.

Sufficient and necessary conditions for convergence of a given G into a particular extremal family are known, and these are often referred to as the **Von Mises conditions**. These conditions provide the normalizing sequences a_n and b_n that lead to convergence. The class of distribution functions G that converge to a given extremal type is known as the **domain of attraction** associated with that type.

The Fisher-Tippett-Gnedenko theorem has an important connection with the **Pickands-Balkema-De Haan theorem**, as a distribution function $F(x)$ has a generalized Pareto upper tail if and only if it lies in the domain of attraction of some extremal distribution type [2].

References

[1] Billingsley, P. (1995). Probability and measure. New York [u.a.]: Wiley. ISBN: 0471007102

[2] Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. The Annals of Statistics, 3(1), 119–131. http://www.jstor.org/stable/2958083