Math 606, Summer 2024[1]: Extreme Values and Rare Events.

Notations.

(1) $1_{\{\cdot\}}$ — the indicator function for the event $\{\cdot\}$.
(2) $\mathcal{N}(0,1)$ — standard Gaussian random variable.
(3) $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ — the ordered sample $X_1, \dots, X_n$ (order statistics).

Abbreviations

(1) iid: independent and identically distributed.

Ideas for Homework

General Exercises

(1) Recall that $F = F(x)$, $x \in \mathbb{R}$, is a `cumulative distribution function` if F has the following properties:
   - for every x, F is right-continuous and has a limit from the left ($F(x) = F(x+)$, $F(x-)$ exists);
   - F is non-decreasing (for every $x < y$, $F(x) \le F(y)$);
   - $\lim_{x\to-\infty} F(x) = 0,\ \lim_{x\to+\infty} F(x) = 1$.

   Let $f_n = f_n(x)$, $x \in \mathbb{R}$, $n \ge 1$, and $f = f(x)$ be cumulative distribution functions such that $\lim_{n\to\infty} f_n(x) = f(x)$ for every $x \in \mathbb{R}$, and assume that the limit function f is continuous. Explain why the convergence of f_n to f is uniform on $\mathbb{R}$. [Make an effort to provide all the details how, given $\varepsilon > 0$, there exists an N_ε such that $|f_n(x) - f(x)| < \varepsilon$ for all $n > N_\varepsilon$ and all $x \in \mathbb{R}$. You can start by noticing that common limits at infinity effectively reduce the problem to a compact interval, and then uniform continuity of f and monotonicity of everything will finish the job. The main challenge is to keep track of all the epsilons while ensuring that the inequalities are in the right direction].

(2) Let $f_n, n \ge 1$, and f be mappings from a metric space X to a metric space Y. Define the following two types of convergence:
   - Uniform $f_n \rightrightarrows f$ if $\lim_{n\to\infty} \sup_{x\in X} d_Y\big(f_n(x), f(x)\big) = 0$.
   - Continuous $f_n \Rrightarrow f$ if $\lim_{n\to\infty} d_X(x_n, x_0) = 0$ implies $\lim_{n\to\infty} d_Y\big(f_n(x_n), f(x_0)\big) = 0$.

   (a) Explain why uniform convergence is equivalent to continuous convergence if X and Y are complete and separable, X is compact, and the limit mapping f is continuous.

   (b) Is is possible to converge uniformly but not continuously?

   (c) Is it possible to converge continuously but not uniformly?

(3) Let $F = F(x)$ be a cumulative distribution function on $\mathbb{R}$. Define the `left-continuous inverse` of F by

$$F^{\leftarrow}(u) = \min\{x \in \mathbb{R} : F(x) \ge u\},\ u \in (0,1).$$

   Confirm the following claims:
   - The function $F^{\leftarrow}$ is well defined, non-decreasing, and is continuous from the left: $F^{\leftarrow}(u) = F^{\leftarrow}(u-)$;
   - $u \le F(x)$ if and only if $F^{\leftarrow}(u) \le x$;
   - If U is a uniform random variable on $(0,1)$, then the random variable $F^{\leftarrow}(U)$ has cumulative distribution function F;
   - If the function F is continuous and X is a random variable with cumulative distribution function F, then $F(X)$ is a uniform random variable on $(0,1)$;
   - If $F_n \rightrightarrows F$ and F is continuous, then $F_n^{\leftarrow} \rightrightarrows F^{\leftarrow}$.

[1]Sergey Lototsky, USC; version of May 14, 2024

(4) Let $X_1, X_2, \ldots$ be iid random variables with cdf $F = F(x)$: $F(x) = \mathbb{P}(X_1 \le x)$. Define $M_n = \max\{X_1, \ldots, X_n\}$ and $x_0 = \sup\{x : F(x) < 1\}$. Explain why

$$\lim_{n\to\infty} M_n = x_0$$

with probability one. Note that $x_0 = +\infty$ is a possibility.

(5) Identify the extreme value distribution (for the max) in each of the following cases:
- Normal, Gamma, Beta, Cauchy;
- Poisson, Geometric, Negative Binomial;
- Your favorite discrete and/or continuous distribution not mentioned above.

(6) Let $X_1, X_2, \ldots$ be iid random variables. Determine the sequence of (non-random) numbers a_n, $n \ge 1$, such that

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{X_n}{a_n} = 1\right) = 1 \qquad (1.1)$$

if the distribution of X_k is
- Normal, Gamma, Cauchy;
- Poisson, Geometric, Negative Binomial;
- Some other discrete and/or continuous distribution that is not bounded from above.

(7) Do the previous problem with

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{\max\{X_1, \ldots, X_n\}}{a_n} = 1\right) = 1 \qquad (1.2)$$

instead of (1.1). What distributions require different sequences a_n in (1.1) and (1.2)?

(8) Let $X_1, X_2, \ldots$ be iid random variables with characteristic function $\varphi(t) = e^{-|t|^\alpha}$, and assume that $\alpha \in (0, 2)$. Define $S_n = X_1 + \ldots + X_n$.
- Show that

$$\mathbb{P}\left(\limsup_{n\to\infty} \left|\frac{S_n}{n^{1/\alpha}}\right|^{1/(\ln\ln n)} = e^{1/\alpha}\right) = 1. \qquad (1.3)$$

  This is Problem 3 from §4 of Chapter IV in the second edition of Shiryaev's *Probability*.
- What happens if $\alpha = 2$?
- What happens if $\alpha > 2$?

(9) Let $U_1, \ldots, U_n$ be iid random variables, uniform on the interval $[a, a+\theta]$. Denote by $U_{(1)} < \ldots < U_{(n)}$ the corresponding order statistics and define $V_1 = U_{(1)}$, $V_k = U_{(k)} - U_{(k-1)}$, $k = 2, \ldots, n$, $V_{n+1} = a + \theta - U_{(n)}$.
- Identify the distribution of each $U_{(k)}$.
- Confirm that the random variables $V_1, \ldots, V_{n+1}$ are exchangeable and compute $\mathbb{E}V_k$.
- Determine the non-random numbers α_n so that the limit $\lim_{n\to\infty} \alpha_n \max_{1\le k\le n+1} V_k$ exists in distribution, and identify the limit.
- Determine the non-random numbers β_n so that the limit $\lim_{n\to\infty} \beta_n \min_{1\le k\le n+1} V_k$ exists in distribution, and identify the limit.

(10) Let Z be a standard normal random variable.
- Prove that, for every $x > 1$,

$$\frac{1}{\sqrt{2\pi}}\, \frac{x}{1+x^2} e^{-x^2/2} \le \mathbb{P}(Z \ge x) \le \frac{1}{\sqrt{2\pi}\, x}\, e^{-x^2/2}. \qquad (1.4)$$

  [This is all about integration by parts. For the lower bound, start by computing the derivative of $x^{-1}e^{-x^2/2}$; then note that the function $f(x) = x^2/(1+x^2)$ is increasing for $x > 0$.]
- Derive an analog of (1.4) for $\mathbb{P}(|Z| \ge x)$ that would hold for all $x \ge 0$.
- Confirm, numerically or otherwise, that the function

$$F(x) = 2^{-22^{1-41^{x/10}}}, \ x > 0, \qquad (1.5)$$

can be a good approximation of $\mathbb{P}(Z \le x)$. In what range of values x would you use such an approximation? The reference for (1.5): A. Soranzo and E. Epure, Very Simply Explicitly Invertible Approximations of Normal Cumulative and Normal Quantile Function, Applied Mathematical Sciences, volume 8, number 87, 4323–4341, 2014, http://dx.doi.org/10.12988/ams.2014.45338.

(11) Let $X_k,\ k \ge 1$ be iid with zero mean, unit variance, and finite third moment. Denote by Φ the cdf of the standard normal random variable.
- Show that if $\varepsilon \in (0,1)$ and $x_n = (1-\varepsilon)\sqrt{\ln n}$, then

$$\lim_{n\to\infty} \frac{1-\mathbb{P}\left(\frac{X_1+\ldots+X_n}{\sqrt{n}} \le x_n\right)}{1-\Phi(x_n)} = 1. \tag{1.6}$$

- Will (1.6) continue to hold with $\varepsilon = 0$?

(12) Confirm that

$$\ln(n+1) < \sum_{k=1}^{n} \frac{1}{k} < 1 + \ln n,\ \ n = 2,3,4,\ldots$$

and

$$\lim_{n\to\infty} \left(\sum_{k=1}^{n} \frac{1}{k} - \ln n\right)$$

exists (the limit is the Euler-Marscheroni constant $\gamma = 0.577\ldots$).

(13) Confirm that, as $x \to +\infty$, the functions $x \mapsto \sum_{k\le x}(1/k)$ and $x \mapsto (\ln x)^{\sin(\ln x)}$ are slowly varying, and the functions $x \mapsto 2 + \sin x$ and $x \mapsto xe^{\sin(\ln x)}$ are not regularly varying.

(14) Let $X_k,\ k \ge 1$, be iid with cdf F, and define $x_0 = \sup\{x : F(x) < 1\}$. Recall that X_n is a record if $X_n > \max\{X_1,\ldots,X_{n-1}\}$.
- Prove that the total number of records is finite if and only if $F(x_0-) < 1$.
- Prove that, for $a < b \le x_0$, the probability to have no records taking values in the interval $(a,b]$ is $\mathbb{P}(X_1 > b)/\mathbb{P}(X_1 > a)$.

(15) Confirm that a two-dimensional cdf $F = F(x,y)$ is max-id if and only if

$$F(a,y)F(b,x) \le F(a,x)F(b,y)$$

for all $a < b$ and $x < y$.

(16) Let $X_{(k)},\ k = 1,2,3$ be the order statistics of an independent sample from the geometric distribution: $\mathbb{P}(X = n) = (1-p)p^n, n = 0,1,2,\ldots$. Compute the probability that $X_{(1)} < X_{(2)} < X_{(3)}$.

An outline: after some combinatorial analysis, the required probability turns out to be $1 - 3\mathbb{P}(X_1 = X_2) + 2\mathbb{P}(X_1 = X_2 = X_3)$; by direct computation, for m iid geometric,

$$\mathbb{P}(X_1 = \ldots = X_m) = \frac{(1-p)^m}{1-p^m}.$$

(17) Let $X_{(k)},\ k = 1,2,3$ be the order statistics of an independent sample from the standard normal distribution. Confirm that $\mathbb{E}X_{(3)} = 3/(2\sqrt{\pi})$.

(18) Let $X_{(k)},\ k = 1,\ldots,n$ be the order statistics of an independent sample from the standard Cauchy distribution. Show that if r is a real number such that $r < k < n-r+1$, then $\mathbb{E}|X_{(k)}|^r < \infty$.

(19) For each of the three extreme value distributions (Frechet, Gumbell, Weibull), compute the mean, the variance, the mode, and the median. Some of them are infinite and some of them involve the Euler Gamma function and/or the Euler-Mascheroni constant.

(20) Let

$$F(x) = 1 - \frac{1}{x},\ \ x \ge 1. \tag{1.7}$$

Determine numbers $a_n > 0$ and $b_n \in \mathbb{R}$ so that $\lim_{n\to\infty}\Big(F(a_n x + b_n)\Big)^n = e^{-x^{-1}},\ x > 0$.

(21) Let $X_1, X_2, \ldots$ be iid with cdf (1.7). Confirm that if $\lim_{n\to\infty}(k/n) = 0$, then $\lim_{n\to\infty} X_{(n-k)} = +\infty$ with probability one. [Assume that $X_{(n-k)} < r$ infinitely often for some $r > 0$ and get a contradiction with $k/n \to 0$; this is Lemma 3.2.1 in de Haan and Ferreira book].
(22) Confirm that, when it comes to extreme values, every Gamma distribution is in the domain of attraction of Gumbel.
(23) True of false: each of the three extreme value distributions is in its own domain of attraction? Explain your conclusion.
(24) Let X and Y be iid and not bounded from above. Compute the value of the limit

$$\lim_{x\to\infty} \frac{\mathbb{P}(X+Y > x)}{\mathbb{P}(X > x)}$$

for as many different distributions as possible (sometimes, the limit is infinite). Can you state and prove a general result?
(25) Consider the function $F(x) = 1 - e^{-x-\sin(x)}$, $x \geq 0$.
   - Confirm that F is a cdf.
   - What can you sam about the limit, as $n \to \infty$, of the maximum of n iid random variables with distribution F? [Apparently, F is not in the domain of attraction of any of the extreme value distributions; looking at a sequence $e^{2\pi n}$ can help.]
(26) Use normal approximation of the Beta distribution to establish normal approximation of k-th uniform order statistic in the limit $k \to \infty$, $n \to \infty$, $n - k \to \infty$. What happens if $n - k$ does not grow to infinity? For related ideas, check out the extreme value book by de Haan and Ferreira: Lemma 2.2.2 and Exercise 2.6.
(27) (Scheffé's lemma) If f_n, $n \geq 1$, is a sequence of pdf-s such that the point-wise limit $\lim_{n\to\infty} f_n(x) = f(x)$ exists and is also a pdf, then $\lim_{n\to\infty} \int_{\mathbb{R}} |f_n(x) - f(x)|\, dx = 0$.
(28) (D-norm in $\mathbb{R}^{\mathrm{d}}$, generated by random vector Z) Let $Z = (Z_, \ldots, Z_{\mathrm{d}})$ be a random vector in $\mathbb{R}^{\mathrm{d}}$ satisfying $Z_k \geq 0$ and $\mathbb{E}Z_k = 1$ for all $k = 1, \ldots, \mathrm{d}$. For $x = (x_1, \ldots, x_{\mathrm{d}})$, define

$$\|x\|_D = \mathbb{E} \max_{1\leq k\leq \mathrm{d}} \Big(|x|_k Z_k\Big). \tag{1.8}$$

Confirm the following claims:
   - (1.8) defines a norm on $\mathbb{R}^{\mathrm{d}}$.
   - $\|x\|_\infty = \max_{1\leq k\leq \mathrm{d}} |x_k| \leq \|x\|_D \leq \|x\|_1 = \sum_{k=1}^{\mathrm{d}} |x|_k$.
   - If $|x_k| \leq |y_k|$ for all k, then $\|x\|_D \leq \|y\|_D$.
   - $\|\cdot\|_\infty = \|\cdot\|_D$ with $Z_k = 1$ (non-random Z);
   - $\|\cdot\|_1 = \|\cdot\|_D$ with *binomial* Z_k satisfying $Z_1 + \ldots + Z_{\mathrm{d}} = \mathrm{d}$ and $\mathbb{P}(Z_k = \mathrm{d}) = 1/\mathrm{d}$, $\mathbb{P}(Z_k = 0) = 1 - (1/\mathrm{d})$.
   - For $p \in (1, +\infty)$, the corresponding norm $\|x\|_p = \Big(\sum_{k=1}^{\mathrm{d}} |x_k|^p\Big)^{1/p}$ is also a D-norm, with iid Z_k such that each $\Big(\Gamma((p-1)/p) Z_k\Big)^p$ has standard Fréchet distribution with cdf $e^{-1/x}$, $x > 0$.

Computer Exercises

(1) Investigate several ways to sample from your favourite extreme value distribution.
(2) Compute the first 100 digits in the decimal representation of γ.

Further Thoughts

(1) Is there a limiting version of (1.3) as $\alpha \to 0+$?
(2) Does it make sense to talk about LLN, CLT, LIL, and LDP in the setting of the extreme value distributions? What kind of results would you get? Are there connections with the corresponding results for the original distributions? How about extreme value for extreme values?

(3) Assume that X stochastically dominates Y, that is, $1 - F_X(x) \geq 1 - F_Y(x)$ (and strict inequality holds for some x). What can we say about the corresponding extreme value distributions? What if we consider some other stochastic ordering?
(4) The word *kernel* has many different meanings, even in mathematics; as a non-negative definite function, it can determine the distribution of a zero-mean Gaussian process and the law of a determinant point process. Are there any interesting/useful connections between the two (very different) processes?
(5) How will Stein's method work to study convergence to extreme value distributions?
(6) What kind of functional limit theorems do we get in the setting of extreme values? One example is (rather scary-looking) Theorem 2.4.2 in the de Haan and Ferreira book; Example A.0.3 in the same book is much better-looking.

Main facts of general interest.

(1) The Cauchy functional equation and variations.
(2) Karamata's theorem about properties of regularly varying functions.
(3) *Convergence of/to types* lemma/theorem.
(4) A simple proof of the one-dimensional Skorokhod representation theorem using the probability integral transform.
(5) The Skorokhod space $\mathbf{D}$, with the standard definition of the metric, is an example of a Polish space that is not a Polish metric space: to make it complete, another metric is required.
(6) The extremal process and its properties.