Extremes

Robert L. Wolpert
Department of Statistical Science
Duke University, Durham, NC, USA

1 Extreme Values

Most probability books (including our text) do a fine job of covering the approximate probability distribution of *sums* (or averages) of independent random variables. If $\{X_j\}$ are independent and identically distributed (i.i.d.) with any distribution having a finite mean μ and variance σ^2, the sum and average

$$S_n := \sum_{j=1}^{n} \qquad \bar{X}_n := \frac{1}{n} S_n$$

are each asymptotically normally distributed in the sense that their standardized version

$$Z_n := \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

satisfies

$$\lim_{n\to\infty} \Pr[a < Z_n \le b] = \Phi(b) - \Phi(a)$$

uniformly in $-\infty < a < b < \infty$, where

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-z^2/2}\, dz$$

denotes the standard Normal CDF function. Some texts go further and discuss limits for sums of random variables X_j that do *not* have finite means or variances— in that case the α-Stable distribution emerges as another (in fact, the only other) possible limiting distribution for normalized sums of the form

$$\frac{S_n - a_n}{b_n}$$

for suitable non-random sequences $\{a_n\}$, $\{b_n\}$.

In light of recent concerns about economic crises and climate changes leading to catastrophes in storm and drought severity, temperature, hurricane intensity, and such, there is a new interest in looking not at the probability distributions of *averages* (like $\bar{X}_n$) but at those of *extremes*, like:

$$X_n^* := \max_{1\le j\le n} X_j.$$

1.1 Example 1: Exponential Distribution

Let $\{X_j\}$ have independent Exponential distributions $X_j \overset{\text{iid}}{\sim} \mathsf{Ex}(\lambda)$, and let X_n^* be the largest of the first n. Can we find non-random sequences $\{a_n\}$, $\{b_n\}$ and a limiting CDF $G(z)$ for which

$$\lim_{n\to\infty} \Pr\left[\frac{X_n^* - a_n}{b_n} \le z\right] = G(z)?$$

For any sequences $\{a_n\}$, $\{b_n\}$ the exact probabilities are

$$\begin{aligned}
\Pr\left[\frac{X_n^* - a_n}{b_n} \le z\right] &= \Pr[X_n^* \le a_n + b_n z] \\
&= \Pr\left\{\cap_{j=1}^n [X_j \le a_n + b_n z]\right\} \\
&= \{\Pr[X_1 \le a_n + b_n z]\}^n \\
&= \left\{1 - e^{-\lambda(a_n + b_n z)}\right\}^n
\end{aligned}$$

The goal is to find $\{a_n, b_n\}$ for which this converges as $n \to \infty$ to a DF. If we now choose $a_n = (\log n)/\lambda$ and $b_n = 1/\lambda$,

$$\begin{aligned}
&= \left\{1 - \frac{1}{n} e^{-z}\right\}^n \\
&\to G(z) := \exp\big(-e^{-z}\big), \qquad (1)
\end{aligned}$$

the standard Gumbel Distribution. Evidently its median is $-\log\log 2$ (since $G(-\log\log 2) = \exp(-\log 2) = 1/2$), so the median m_n^* for X_n^* defined by $1/2 = \Pr[X_n^* \le m_n^*]$ is

$$m_n^* = \frac{1}{\lambda}\log n - \frac{1}{\lambda}\log\log 2,$$

which grows with n at a logarithmic rate.

For example, if we imagine that sprinters' speed in m/s are given by the $\mathsf{Ex}(1)$ distribution, then the fastest speed of n independently-drawn sprinters would have approximately the rescaled Gumbel Distribution with median $m_n^* = \log n - \log\log 2$; this has even odds of exceeding Usain Bolt's 2009 world-record 9.69s 100m pace if

$$\begin{aligned} \log n - \log\log 2 &\geq \frac{100\text{m}}{9.69\text{s}} \\ &= 10.32\text{m/s} \\ \log n &\geq \log\log 2 + 10.32 \\ n &\geq \exp(-0.37 + 10.32 = 9.95) \\ &= 21\,023.73, \end{aligned}$$

i.e., there's about an even chance that one of 21,024 independent $\mathsf{Ex}(1)$ random variables would exceed Bolt's pace.

For this example we can compute exactly the median for X_n^* or, if we prefer, the probability that X_n^* exceeds 9.95 for $n = 21024$; the latter, for example, is

$$\Pr[X_{21024}^* > 10.32] = \big[1 - \exp(-10.32)\big]^{21024} = 0.5000176,$$

just as expected.

1.2 Example 2: Normal Distribution

Now let $\{X_j\}$ have independent standard Normal distributions $X_j \overset{\text{iid}}{\sim} \mathsf{No}(0,1)$, set $X_n^* := \max_{1 \leq j \leq n} X_j$, and seek non-random $\{a_n\}$, $\{b_n\}$ and a limiting CDF $G(z)$ for $b_n^{-1}(X_n^* - a_n)$. First we need to note that, for $x > 0$,

$$\begin{aligned} \Phi(-x) &= \int_x^\infty \phi(z)\,dz \\ &\leq \int_x^\infty \frac{z}{x}\phi(z)\,dz = \frac{1}{x\sqrt{2\pi}} \int_x^\infty z e^{-z^2/2}\,dz = \frac{1}{x}\phi(x); \end{aligned}$$

Gordon's Inequality improves this to the two-sided bound

$$1 \leq \frac{\phi(x)}{x\Phi(-x)} \leq 1 + \frac{1}{x^2}$$

for every $x > 0$. Now let $a_n = -\Phi^{-1}(1/n)$ be the $(1 - 1/n)$'th quantile and set $b_n = 1/a_n$; note that $a_n \asymp \sqrt{2\log n}$ grows as $n \to \infty$, while $b_n \to 0$. By

Taylor's theorem and the evenness of $\phi(z)$, for fixed $z \in \mathbb{R}$,

$$\begin{aligned}
\log \Phi(-a_n - b_n z) &= \log \Phi(-a_n) - b_n z \frac{\phi(-a_n)}{\Phi(-a_n)} + o(b_n z) \\
&= \log \frac{1}{n} - z \frac{\phi(a_n)}{a_n \, \Phi(-a_n)} + o(b_n z) \\
&= \log \frac{1}{n} - z + o(b_n z)
\end{aligned}$$

so

$$\begin{aligned}
\Pr[X_1 \le a_n + b_n z] &= \Phi(a_n + b_n z) \\
&= 1 - \frac{1}{n} e^{-z + o(1/\sqrt{\log n})}, \text{ and} \\
\Pr[X_n^* \le a_n + b_n z] &\approx \left[1 - n^{-1} e^{-z}\right]^n \\
&\approx \exp(-e^{-z}) =: G(z),
\end{aligned}$$

again the Gumbel distribution. Similarly, if $\{X_i\} \overset{\text{iid}}{\sim} \mathsf{No}(\mu, \sigma^2)$ (now with arbitrary mean and variance) then we simply change the location and scale to find that with $a_n = \mu - \sigma \Phi^{-1}(1/n)$ and $b_n = -\sigma/\Phi^{-1}(1/n)$ we have

$$\Pr\left[\frac{X_n^* - a_n}{b_n} \le z\right] \to G(z),$$

with median

$$m_n^* = \mu - \sigma \Phi^{-1}(1/n) + (\log \log 2)/\Phi^{-1}(1/n)$$

growing like $\sigma\sqrt{2 \log n}$ as $n \to \infty$.

Typically unbounded distributions like the Exponential and Normal (as well as the Gamma, Lognormal, Weibull, *etc.*) whose tails fall off exponentially or faster will have this same Gumbel limiting distribution for the maxima, and will have medians (and other quantiles) that grow as $n \to \infty$ at the rate of (some power of) $\log n$.

1.3 Example 3: Pareto Distribution

Distributions with "fatter tails" (*i.e.*, those for which $\Pr[X > x]$ falls off no faster than a *power* of x) will have a different limit. For example, let $\{U_j\}$ be i.i.d. Uniform random variables and set $X_j = 1/U_j$; then X_j has the "unit Pareto distribution" determined by

$$\Pr[X_j > x] = 1/x, \qquad x \ge 1$$

and the maximum X_n^* of n i.i.d. unit Paretos will satisfy

$$\Pr[X_n^* \le a_n + b_n z] = \big(1 - [a_n + b_n z]^{-1}\big)^n \qquad a_n + b_n z \ge 1.$$

With $a_n = 0$ and $b_n = n$,

$$= \big(1 - \frac{1}{nz}\big)^n \to e^{-1/z} =: F(z), \qquad z > 0, \tag{2}$$

the "unit Fréchet Distribution". Similarly for $X_j = \epsilon U_j^{-1/\alpha}$ with the $\mathsf{Pa}(\alpha, \epsilon)$ distribution satisfying

$$\Pr[X_j > x] = \epsilon^\alpha / x^\alpha, \qquad x \ge \epsilon,$$

set $a_n = 0$ and $b_n = n^{1/\alpha}\epsilon$, then

$$\Pr[X_n^* \le a_n + b_n z] = \big(1 - \frac{1}{n} z^{-\alpha}\big)^n \to e^{-z^{-\alpha}} =: F(z \mid \alpha), \qquad z > 0,$$

the Fréchet distribution with shape parameter $\alpha > 0$. The Fréchet median is $(\log 2)^{-1/\alpha}$, so X_n^* has median

$$m_n^* = n^{1/\alpha}\epsilon(\log 2)^{-1/\alpha}$$

that grows like a power of n. This is typical for heavy-tailed distributions.

1.4 Example 4: Uniform Distribution

The *minimum* of n i.i.d. $\mathsf{Ex}(\lambda)$ random variables has the $\mathsf{Ex}(n\lambda)$ distribution, so it converges to zero at rate $1/n$; similarly the minimum of n $\mathsf{We}(\alpha, \lambda)$ (Weibull) random variables has the $\mathsf{We}(\alpha, n\lambda)$ distribution and converges to zero at rate $n^{-1/\alpha}$. It follows immediately that for the maximum X_n^* of n *reversed* (or *negative*) Weibull random variables with cdf and pdf

$$\begin{aligned} F(x \mid \alpha, \lambda) &= e^{-\lambda(-x)^\alpha} \qquad x < 0 \\ f(x \mid \alpha, \lambda) &= \lambda\,(-x)^{\alpha-1} e^{-\lambda(-x)^\alpha} \mathbf{1}_{\{x<0\}}, \end{aligned} \tag{3}$$

the limiting distribution of $b_n^{-1}[X_n^* - a_n]$ is the reversed $\mathsf{We}(\alpha, 1)$ Weibull if $a_n = 0$ and $b_n = (\lambda n)^{-1/\alpha}$, with median $m_n^* = -(n\lambda/\log 2)^{-1/\alpha}$ increasing to zero as $n \to \infty$.

Similarly the maximum X_n^* of n i.i.d. uniform random variables $X_j \sim \mathsf{Un}(L,R)$ has limiting distribution:

$$\begin{aligned}
\Pr[b_n^{-1}[X_n^* - a_n] \le z] &= \Pr[X_n^* \le a_n + b_n z] \\
&= \left[1 - \frac{R - a_n - b_n z}{R - L}\right]^n \qquad \text{if } L \le a_n + b_n z \le R \\
&= (1 + z/n)^n \to e^z \qquad \text{if } z \le 0 \text{ and } n > |z|
\end{aligned}$$

for $a_n = R$ and $b_n = (R-L)/n$, the unit Reversed $\mathsf{We}(1,1)$ Weibull with median for X_n^* of

$$m_n^* = R - (R-L)(\log 2)/n$$

increasing at rate $1/n$ to an upper bound of R. The suitably standardized minimum and maximum of n independent $\mathsf{Be}(\alpha,\beta)$ random variables have asymptotic $\mathsf{We}(\alpha,1)$ and reverse $\mathsf{We}(\beta,1)$ distributions, respectively. These are typical of the maximal behavior for bounded random variables with continuous distributions.

Fisher and Tippett (1928) first proved that location-scale families of these three distributions— Gumbel (Equation (1)), Fréchet (Equation (2)), and reversed Weibull (Equation (3))— are the *only* possible limits for maxima of independent random variables. Much later McFadden (1978) discovered that all three of these limiting distributions could be expressed in the same functional form as special cases of a single three-parameter "Generalized Extreme Value" (GEV) distribution, with CDF

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

In some ways I feel this was unfortunate, because now it is common for people to model and fit the GEV without thinking very clearly about the specific form of their data and distributions.

2 Threshold Exceedances

As before let $\{X_j\}$ be i.i.d. for $1 \le j \le n$ and set $T_j = \frac{j-1/2}{n} \in (0,1)$. Let a_n and b_n be real numbers and set $Y_j = a_n + b_n X_j$. For sufficiently large u, the numbers $N(R_i)$ of points (T_j, Y_j) in disjoint rectangles $R_i = (s_i, t_i] \times (u_i, v_i]$ with $0 \le s_i < t_i \le 1$ and $u \le u_i < v_i \le \infty$ will be approximately independent Poisson random variables, with means

$$\lambda_i = n(t_i - s_i)[F(a_n + b_n v_i) - F(a_n + b_n u_i)]$$

Here we look for choices of a_n and b_n for which this has a simple form, and then exploit it.

2.1 Example 1: Weibull Distribution

If $\Pr[X_j > x] = e^{-\beta x^\alpha}$ for $x > 0$, then for the choice $a_n = [\beta^{-1} \log n]^{1/\alpha}$ and $b_n = a_n/(\alpha \log n)$ we have for all large enough z,

$$\begin{aligned} n[1 - F(a_n + b_n z)] &= n \exp\big(-\beta(a_n + b_n z)^\alpha \big) \\ &= n \exp\big(-\log n(1 + z/\alpha \log n)^\alpha \big) \\ &= n \exp\big(-\log n(1 + z/\log n + o(1/\log n)) \big) \\ &\approx e^{-z}, \end{aligned}$$

so $\{T_j, Y_j = (X_j - a_n)/b_n\}$ have approximately the Poisson distribution on $[0, 1] \times \mathbb{R}$ with intensity measure $\nu(dt\,dy) = e^{-y}\,dt\,dy$. A similar approach with suitable a_n, b_n works for any other distribution in the Gumbel domain. The maximum $M_t := \max\{Y_j : \ T_j \le t\}$ is a non-decreasing stochastic process on the unit interval $0 < t < 1$, with CDF

$$\begin{aligned} F_t(z) &= \Pr[M_t \le z] \\ &= \Pr[\text{No Poisson points in } (0, t] \times (z, \infty)] \\ &= e^{-te^{-z}}, \end{aligned}$$

the Gumbel distribution. The events $[M_t \le z]$ and $[X^*_{\lfloor nt \rfloor} \le a_n + b_n z]$ are identical.

2.1.1 Related Max-Stable Process

Let $\{(T_j, Y_j)\}$ be the points of a $\mathsf{Po}(e^{-y} dt\,dy)$ random field on all of $\mathbb{R}^d \times \mathbb{R}_+$, and let $f(t)$ be any strictly positive function; define a random process by

$$Z(t) = \sup_j \{Y_j / f(T_j - t)\}.$$

If $f(t) = \sum a_i \mathbf{1}_{A_i}(t)$ is a simple function, then

$$\begin{aligned}
\Pr[Z(t) \le z] &= \prod_i \Pr\big[\sup_j\{Y_j/a_i \le z:\ T_j - t \in A_i\big] \\
&= \prod_i \Pr\big[\text{No Poisson pts in } (A_i + t) \times (a_i z, \infty)\big] \\
&= \prod_i \exp\big(-|A_i| e^{-a_i z}\big) \\
&= \exp\left(-\int e^{-zf(s)} ds\right),
\end{aligned}$$

so $Z(t)$ is a stationary process. For any (not necessarily simple) strictly positive $f(t)$ on $\mathbb{R}^d$, the same identity follows from LDCT.

2.2 Example 2: Pareto Distribution

If $\Pr[X_j > x] = \epsilon^\alpha x^{-\alpha}$ for $x > \epsilon$, then for the choice $a_n = 0$ and $b_n = \epsilon n^{1/\alpha}$ we have for all large enough z,

$$\begin{aligned}
n[1 - F(a_n + b_n z)] &= n\big(0 + \epsilon n^{1/\alpha} z)^\alpha\big) \\
&= z^{-\alpha},
\end{aligned}$$

so $\{T_j, Y_j = (X_j - a_n)/b_n\}$ have approximately the Poisson distribution on $[0,1] \times \mathbb{R}_+$ with intensity measure $\nu(dt\,dy) = \alpha y^{-\alpha-1}\,dt\,dy$. A similar approach with suitable a_n, b_n works for any other distribution in the Fréchet domain.

The maximum $M_t := \max\{Y_j :\ T_j \le t\}$ is a non-decreasing stochastic process on the unit interval $0 < t < 1$, with CDF

$$\begin{aligned}
F_t(z) &= \Pr[M_t \le z] \\
&= \Pr[\text{No Poisson points in } (0,t] \times (z, \infty)] \\
&= e^{-tz^{-\alpha}},
\end{aligned}$$

the Fréchet distribution. The events $[M_t \le z]$ and $[X^*_{\lfloor nt \rfloor} \le a_n + b_n z]$ are identical.

Note that the *sum* of the $\{Y_j :\ T_j \le t\}$ will be finite almost-surely if $\int_0^\infty (z \wedge 1)\alpha z^{-\alpha-1}\,dz < \infty$, i.e., if $0 < \alpha < 1$; in that case the non-decreasing process

$$S_t := \sum\{Y_j :\ T_j \le t\}$$

is a fully-skewed α-Stable process with distribution

$$\sim \mathsf{St}_0\left(\alpha, \beta = 1, \gamma = \frac{\pi t}{2\Gamma(\alpha)\sin(\pi\alpha/2)}\right)$$

and the $\{Y_j\}$ are the "jumps" of S_t. A similar representation holds for $1 \le \alpha < 2$, but "compensation" is required (sort of like subtracting an infinite drift from S_t). There is no α-Stable process for $\alpha > 2$, although the connection between Fréchet distribution and the Poisson point process remains.

2.2.1 Related Max-Stable Process

Let $\{(T_j, Y_j)\}$ be the points of a $\mathsf{Po}(\alpha y^{-\alpha-1} dt\, dy)$ random field on all of $\mathbb{R}^d \times \mathbb{R}_+$, and let $0 \le f(t) \in L_\alpha(\mathbb{R}^d, dt)$; define a random process by

$$Z(t) = \sup_j \{Y_j f(T_j - t)\}.$$

If $f(t) = \sum a_i \mathbf{1}_{A_i}(t)$ is a simple function, then

$$\begin{aligned}
\Pr[Z(t) \le z] &= \prod_i \Pr\big[\sup_j \{Y_j\, a_i \le z : \; T_j - t \in A_i\big] \\
&= \prod_i \Pr\big[\text{No Poisson pts in } (A_i + t) \times (z/a_i, \infty)\big] \\
&= \prod_i \exp\big(-|A_i|(z/a_i)^{-\alpha}\big) \\
&= \exp\left(-z^{-\alpha}\int f(s)^\alpha ds\right),
\end{aligned}$$

so $Z(t)$ is a stationary process with a Fréchet $\mathsf{Fr}\big(\alpha, \|f\|_\alpha^\alpha\big)$ distribution. For non-simple $0 \le f \in L_\alpha$, the same identity follows from LDCT.

2.3 Example 3: Beta Distribution

If $X_j \overset{\text{iid}}{\sim} \mathsf{Be}(\alpha, \beta)$ then for small ϵ,

$$\begin{aligned}
\Pr[X_j > 1 - \epsilon] &\approx \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_{1-\epsilon}^1 (1-x)^{\beta-1}\, dx \\
&= \frac{\epsilon^\beta}{\beta B(\alpha,\beta)}, \qquad B(\alpha,\beta) := \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}
\end{aligned}$$

so for $a_n = 1$ and $b_n = (\beta B(\alpha,\beta)/n)^{1/\beta}$, for $z < 0$ we have

$$\begin{aligned} n \Pr[X_j > a_n + b_n z] &\approx \frac{n}{\beta B(\alpha,\beta)}(1 - a_n - b_n z)^\beta \\ &= (-z)^\beta \end{aligned}$$

so $\{T_j, Y_j = (X_j - a_n)/b_n\}$ have approximately the Poisson distribution on $[0,1] \times \mathbb{R}_-$ with intensity measure $\nu(dt\,dy) = \beta(-y)^{\beta-1}\,dt\,dy$. A similar approach with suitable a_n, b_n works for any other distribution in the Reverse Weibull domain.

The maximum $M_t := \max\{Y_j :\ T_j \le t\}$ is a non-decreasing stochastic process on the unit interval $0 < t < 1$, with CDF

$$\begin{aligned} F_t(z) &= \Pr[M_t \le z] \\ &= \Pr[\text{No Poisson points in } (0,t] \times (z,\infty)] \\ &= e^{-t(-z)^\beta}, \end{aligned}$$

the reversed Weibull distribution. The events $[M_t \le z]$ and $[X^*_{\lfloor nt \rfloor} \le a_n + b_n z]$ are identical.

The *minimum* of n i.i.d. $\mathsf{Be}(\alpha,\beta)$ random variables can be studied in the same way; for $a_n = 0$ and $b_n = (\alpha B(\alpha,\beta)/n)^{1/\alpha}$, the points $\{T_j, Y_j = (X_j - a_n)/b_n\}$ have approximately the Poisson distribution on $[0,1] \times \mathbb{R}_+$ with intensity measure $\nu(dt\,dy) = \alpha y^{\alpha-1}\,dt\,dy$, and the cumulative minimum $m_t = \min\{Y_j :\ T_j \le t\}$ is a non-increasing stochastic process satisfying $\Pr[m_t > z] = e^{-tz^\alpha}$, the usual (un-reversed) Weibull.

2.3.1 Related Max-Stable Process

Let $\{(T_j, Y_j)\}$ be the points of a $\mathsf{Po}(\alpha y^{\alpha-1} dt\,dy)$ random field on all of $\mathbb{R}^d \times \mathbb{R}_+$, and let $0 < f(t) \in L_\alpha(\mathbb{R}^d, dt)$. Define a random process by

$$Z(t) = \inf_j \{Y_j / f(T_j - t)\}.$$

If $f(t) = \sum a_i \mathbf{1}_{A_i}(t)$ is a simple function, then

$$\begin{aligned}
\Pr[Z(t) > z] &= \prod_i \Pr\left[\sup_j \{Y_j/a_i > z : \; T_j - t \in A_i\right] \\
&= \prod_i \Pr\left[\text{No Poisson pts in } (A_i + t) \times (0, z\, a_i]\right] \\
&= \prod_i \exp\left(- |A_i| (z\, a_i)^\alpha \right) \\
&= \exp\left(-z^\alpha \int f(s)^\alpha ds \right),
\end{aligned}$$

so $Z(t)$ is a stationary process with a Weibull $\mathsf{We}\big(\alpha, \|f\|_\alpha^\alpha\big)$ distribution. For non-simple $0 \le f \in L_\alpha$, the same identity follows from LDCT.

Leftovers

- Order statistics from PP perspective
- Log likelihood— exponential family?
- Conjugate priors? Fisher information? Jeffreys' rule?

References

Fisher, R. A. and Tippett, L. H. C. (1928), "Limiting forms of the frequency distributions of the largest or smallest member of a sample," *Proceedings of the Cambridge Philosophical Society*, 24, 180–190.

McFadden, D. (1978), "Modeling the Choice of Residential Location," in *Spatial Interaction Theory and Planning Models*, eds. A. Karlqvist, L. Lundqvist, F. Snickars, and J. W. Weibull, Amsterdam, NL: North Holland, *Studies in Regional Science and Urban Economics*, volume 3, pp. 75–96, reprinted in ed. J. Quigley, London, UK: Edward Elgar, *The Economics of Housing, Volume I*, 1997, 531–552.

3 Martingales

Now let $\{X_n\}$ be a sequence of random variables with the property that $\mathsf{E}|X_n| < \infty$ for each j and, for $m > n$,

$$\mathsf{E}[X_m \mid X_1, ..., X_n] = X_n. \tag{4}$$

Such a sequence is called a "martingale". They arise as random walks with mean-zero steps and (more interesting) as the fortune at time n of a player in a fair game who uses any (causal) strategy whatsoever for selecting bets, and also as tools in the modern study of Markov chains and Markov processes. A particularly interesting property of martingales is that the bound

$$\Pr[X_n^* > a] \le \frac{E[(X_n)_+]}{a}$$

for every $a > 0$, where $(x)_+ = \max(0, x)$ is the "positive part" of a real number x (and satisfies $(x)_+ \le |x|$). Note Markov's inequality would give the same value as a bound on $\Pr[X_n > a]$; this is much stronger in that it gives a bound on the *maximum* value of $\{X_j\}_{j\le n}$. For example, a simple random walk $X_n = \sum_{j\le n} \zeta_j$ with independent normally-distributed steps $\zeta_j \overset{\text{iid}}{\sim} \mathsf{No}(0, \sigma^2)$ has $X_n \sim \mathsf{No}(0, n\sigma^2)$ with expectation $\mathsf{E}[|X_n|] = \sqrt{n\sigma^2/2\pi}$, so

$$\Pr[\sup_{0\le j\le n} X_j > a] \le \frac{\sigma\sqrt{n}}{a\sqrt{2\pi}}$$

Taking $\sigma^2 = 1/n$ and $n \to \infty$ leads to a bound on the maximum of Brownian Motion over the unit interval.

If Y_n is an asymmetric random walk that steps up and down one unit with probabilities p, q respectively with $p < q$ (for example, Y_n might be the fortune of a gambler at a casino that offers customers only the opportunity to wager \$1 on any of the even-money Roulette bets (red, odd, 1–18, etc.), with $p = 9/19$ and $q = 10/19$). While Y_n itself is not a martingale,

$$X_n = (q/p)^{Y_n}$$

is, and has constant expectation $\mathsf{E}[X_n] = (q/p)^y$ if $Y_0 = y > 0$ is the player's original fortune. If the Casino maintains a reserve of \$R then the probability that the gambler can *ever* "break the bank" by attaining a fortune of $y + R$ is bounded above by

$$\begin{aligned}\Pr[X_n^* > y + R] &= \Pr[Y_n^* > (q/p)^{y+R}] \\ &\le \frac{\mathsf{E}(Y_n)_+}{(q/p)^{y+R}} = \frac{(q/p)^y}{(q/p)^{y+R}} = (p/q)^R;\end{aligned}$$

the Casino can limit the probability that a customer will exhaust its resources to, say, no more than one in a million by maintaining a cash reserve exceeding $\log(10^{-6})/\log(9/10) = \$131.13$. Of course if it allows bets of up to a limit $\$L$ (instead of \$1) the required reserves increase to $R \geq 131L$, and it increases by another factor of 18 of the Casino allows all the usual Roulette bets including single numbers (which pay off at 36:1, instead of 2:1). How (if at all) does the possibility of multiple simultaneous customers change this?