Aio cursus Einmahl / de Haan
3 september 1999

1 Empirical Processes

Suppose $E_1, E_2, \ldots,$ i.i.d. standard exponential

Consider $E_{1,n}, E_{2,n}, \ldots, E_{n,n},$ order statistics from $(E_1, E_2, \ldots, E_n)$

We consider limit distributions of $E_{k,n}, n \to \infty$.

1. $\mathbf{k = 1}$

   $P\{E_{1,n} > x\} = P\{E_i > x,\ i = 1, \ldots, n\} = e^{-nx}$
   i.e. $nE_{1,n}$ is standard exponential.

   $\Rightarrow nE_{1,n} \longrightarrow$ standard exponential distribution
   The same is true for example for the uniform distribution.

2. Limit distribution of $(E_{1,n}, E_{2,n}, \ldots, E_{k,n}), n \to \infty, k$ fixed.
   Well-known: Rényi's representation:

   $$(E_{1,n}E_{2,n}, \ldots, E_{n,n}) \stackrel{d}{=} \left(\frac{E_1^*}{n}, \frac{E_1^*}{n} + \frac{E_2^*}{n-1}, \ldots, \frac{E_1^*}{n} + \cdots + \frac{E_n^*}{1}\right)$$

   with $E_1^*, \ldots, E_n^*$ i.i.d. standard exponential. (This holds for *fixed* n).
   Hence

   $$n(E_{1,n}, E_{2,n}, \ldots E_{k,n}) \stackrel{d}{=} \left(E_1^*, E_1^* + \frac{n}{n-1}E_2^*, \ldots, E_1^* + \cdots + \frac{n}{n-k}E_k^*\right)$$

   $$\xrightarrow{a.s.} (E_1^*, E_1^* + E_2^*, \ldots, E_1^* + \cdots + E_k^*).$$

   Conclusion: If we consider $nE_{1,n}, nE_{2,n}, \ldots, nE_{k,n}$ as random points on the line, then –as $n \to \infty$–, these points converge to the first k points of a homogeneous Poisson process. And, in fact, the point process represented by the points

   $$nE_1, nE_2, nE_3, \ldots$$

converges to a homogeneous Poisson process in the sense of convergence of point processes (aio course last year). So, in particular,
$nE_{k,n} \longrightarrow$ a gamma distribution with k degrees of freedom if $n \to \infty, k$ fixed.
All of this is also true for example for the uniform distribution (exercise).

3. Limit distribution of $E_{k,n},\ n \to \infty,\ k \to \infty,\ n-k \to \infty$ (we need the last condition because what holds for minima, also holds for maxima: $-\min(-a,-b) = \max(a,b)$).

   Here it is easier to start from the uniform distribution (the result will be the same).
   So take $U_1, U_2, \ldots$ i.i.d. uniform $(0,1)$. Let $U_{1,n}, U_{2,n}, \ldots, U_{n,n}$ be the n-th order statistics.
   Limit distribution of $U_{k,n},\ n \to \infty,\ k \to \infty,\ n-k \to \infty$.
   We derive this by simply working out the limit procedure for the density.

Lemma 1.1. *$X_n \to X$ in distribution if for the density we have $f_{X_n}(x) \to f_X(x)$ for all x.*

Now $P\{U_{k,n} < x\} = P\{$ at least k of $U_1, \ldots, U_n$ are smaller than $x\} = \sum_{i=k}^{n} \binom{n}{i} x^i (1-x)^{n-1}$.

Hence

$$\begin{aligned}
f(x) &= \sum_{i=k}^{n} \binom{n}{i} \frac{d}{dx}\{x^i(1-x)^{n-i}\} \\
&= \sum_{i=k}^{n} \binom{n}{i} i x^{i-1}(1-x)^{n-i} - \sum_{i=k}^{n} \binom{n}{i} (n-i) x^i (1-x)^{n-i-1} \\
&= \sum_{i=k}^{n} \binom{n}{i} i x^{i-1}(1-x)^{n-i} - \sum_{j=k+1}^{n+1} \binom{n}{j-1} (n-j+1) x^{j-1} (1-x)^{n-j} \\
&= \left(\sum_{i=k+1}^{n} 0\right) + \binom{n}{k} k x^{k-1}(1-x)^{n-k} + o = \frac{n!}{(k-1)!(n-k)!} x^{k-1}(1-x)^{n-k}.
\end{aligned}$$

Write

$$b_n := \frac{k-1}{n-1}$$

$$a_n := \sqrt{b_n(1-b_n)\frac{1}{n-1}}.$$

Then: density of normalized order statistic is

$$\frac{d}{dx}P\left\{\frac{U_{k,n}-b_n}{a_n} \le x\right\} = \frac{d}{dx}P\left\{U_{k,n} \le b_n + xa_n\right\}$$

$$= a_n f_{U_{k,n}}(b_n + xa_n) = a_n \cdot \frac{n!}{(k-1)!(n-k)!}(b_n + xa_n)^{k-1}(1-b_n-xa_n)^{n-k}$$

$$= \left[\frac{n!}{(k-1)!(n-k)!}a_n b_n^{k-1}(1-b_n)^{n-k}\right]\left[\left(1+x\frac{a_n}{b_n}\right)^{k-1}\left(1-x\frac{a_n}{1-b_n}\right)^{n-k}\right].$$

Stirling's formula:

$$n! \sim \sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n}$$

$$(k-1)! \sim \sqrt{2\pi}(k-1)^{k-\frac{1}{2}}e^{-k+1}$$

$$(n-k)! \sim \sqrt{2\pi}(n-k)^{n-k+\frac{1}{2}}e^{-n+k}$$

Further

$$b_n^{k-1} = \frac{(k-1)^{k-1}}{(n-1)^{k-1}}$$

$$(1-b_n)^{n-k} = \frac{(n-k)^{n-k}}{(n-1)^{n-k}}.$$

So the first factor tends to $\dfrac{1}{\sqrt{2\pi}}$.

The second factor: we take logarithms:

$$(k-1)\log(1+x\frac{a_n}{b_n}) + (n-k)\log\left(1-x\frac{a_n}{1-b_n}\right)$$

(note $\frac{a_n}{b_n} \to 0$ and $\frac{a_n}{1-b_n} \to 0$)

$$= (k-1)\left\{x\frac{a_n}{b_n} - \frac{x^2}{2}\left(\frac{a_n}{b_n}\right)^2 + \dots\right\}$$

$$+(n-k)\left\{-x\frac{a_n}{1-b_n}-\frac{x^2}{2}\left(\frac{a_n}{1-b_n}\right)+\dots\right\}$$

(first order term drops out)

$$=-\frac{x^2}{2}\left\{(k-1)\left(\frac{a_n}{b_n}\right)^2+(n-k)\left(\frac{a_n}{1-b_n}\right)^2\right\}+\dots$$

$$=-\frac{x^2}{2}\left\{(k-1)\frac{b_n(1-b_n)\frac{1}{n-1}}{b_n^2}+(n-k)\frac{b_n(1-b_n)\frac{1}{n-1}}{(1-b_n)^2}\right\}$$

$$=-\frac{x^2}{2}\frac{1}{n-1}\left[(k-1)\frac{1-b_n}{b_n}+(n-k)\frac{b_n}{1-b_n}\right]$$

$$=-\frac{x^2}{2}\frac{1}{n-1}\left[\frac{(k-1)(n-k)}{k-1}+\frac{(n-k)(k-1)}{n-k}\right]$$

$$=-\frac{x^2}{2}$$

We have proved:

Lemma 1.2. *(Smirnov 1952) Suppose $n\to\infty,\ k\to\infty,\ n-k\to\infty$. Then*

$$\frac{U_{k,n}-b_n}{a_n}$$

is asymptotically standard normal with

$$b_n=\frac{k-1}{n-1}$$

$$a_n=\sqrt{b_n(1-b_n)\frac{1}{n-1}}.$$

3a **Theorem 1.3.** *Suppose $k=[np]$ with $0<p<1$ (or more generally $\sqrt{n}\left(\frac{k_n}{n}-p\right)\to 0$). Then*

$$\sqrt{n}\cdot\frac{U_{k,n}-p}{\sqrt{p(1-p)}}$$

is asymptotically standard normal.

This leads to a well-known result:
The stochastic processes

$$\sqrt{n}\left\{U_{[nt],n} - t\right\},\ 0 \le t \le 1$$

converge ($n \to \infty$) to a continuous Gaussian stochastic process $\{X(t)\}$ known as Brownian bridge:

$$EX(t) = 0 \text{ for all } t \text{ and Cov } (X(s),\ X(t)) = s(1-t) \text{ for } 0 < s < t < 1.$$

John Einmahl will elaborate on this. So up to now the order statistics have led us to the Poisson process and the Brownian bridge. But there is more.

3b Smirnov's lemma also gives us:

Theorem 1.4. *Suppose $n \to \infty,\ k \to \infty,\ k/n \to o$. Then*

$$\frac{U_{k,n} - b_n}{a_n} = \sqrt{k}\left(\frac{n+1}{k}U_{k,n} - 1\right)$$

is asymptotically standard normal.
A similar results holds for $n-k \to \infty,\ n-k = o(n),\ n \to \infty$.

In fact we can extend this also to a statement on stochastic processes:

The stochastic processes

$$\sqrt{k}\left(\frac{n+1}{k}U_{[kt],n}^{-t}\right),\ 0 \le t < \infty$$

converge ($n \to \infty$) to Brownian motion.

Again John Einmahl will elaborate on this.

So there are three (or five) options, for observing the process from 3(5) distances.

Note that I have not yet defined the empirical process.
Remember the process $U_{[nt],n}$ of the (central) order statistics. This process is called the (uniform) **inverse empirical process** $(0 \le t \le 1)$.

Now the random function $U_{[nt],n}$ is a **nondecreasing** function, so it has an inverse. The inverse function is

$$\frac{1}{n}\sum_{i=1}^{n} 1\{U_i \le t\}$$

(check!) and this process is called the **(uniform) empirical process**.

In fact this process has very similar asymptotic behaviour as the process $U_{[nt],n}$, so there are 3 (or 5) different possible asymptotic perspectives:

1. Poisson process
2. Brownian bridge
3. Brownian motion.

2 Link with extreme value theory.

We have seen:

$$nU_{1,n} \xrightarrow{d} \quad E \text{ (standard exponential distribution).}$$

I prefer to work with maxima instead of minima:

$$\frac{1}{n}\cdot \max\left(\frac{1}{U_1},\frac{1}{U_2},\ldots,\frac{1}{U_n}\right) = \frac{1}{nU_{1,n}}$$

$\xrightarrow{d}$ the distribution with distribution function $\exp -\frac{1}{x}, x > 0$.

Now in general:

$$X_1, X_2, \ldots \quad \text{i.i.d.} \quad F.$$

Limit distribution of $\max(X_1, X_2, \ldots, X_n)$?
One approach is as follows:
Take $\frac{1}{U_1}, \frac{1}{U_2}, \ldots$ as above
and the function $U := \left(\frac{1}{1-F}\right)^{\leftarrow}$ (inverse function).
Then

$$U\left(\frac{1}{U_1}\right),\ U\left(\frac{1}{U_2}\right),\ldots \quad \text{i.i.d.} \quad F.$$

So

$$\max(X_1, X_2, \ldots, X_n) \stackrel{d}{=} \max\left(U\left(\frac{1}{U_1}\right),\ldots,U\left(\frac{1}{U_n}\right)\right)$$

$$= U\left(\max\left(\frac{1}{U_1}, \dots, \frac{1}{U_n}\right)\right)$$

$$= U\left(1/\min(U_1, U_2, \dots, U_n)\right)$$

$$= U\left(n \cdot \frac{1}{n\min(U_1, U_2, \dots, U_n)}\right)$$

We want

$$\frac{\max(X_1, X_2, \dots, X_n) - b_n}{a_n} \xrightarrow{d} .$$

Now suppose that there is a positive function a such that

$$(1) \qquad \lim_{t\to\infty} \frac{U(tx) - U(t)}{a(t)} = \Psi(x)$$

exists for all $x > 0$ and is not identically zero.
Then

$$\frac{U\left(n\frac{1}{n\min(U_1,\dots,U_n)}\right) - U(n)}{a(n)} \xrightarrow{d} \Psi\left(\frac{1}{E}\right)$$

since $n.\min(U_1, \dots, U_n) \xrightarrow{d} E$.

Lemma 2.1. *The limit function in* (1) *is always of the form*

$$c \cdot \frac{x^\gamma - 1}{\gamma}$$

with $c > 0$ and $\gamma \in \mathbb{R}$.

Theorem 2.2. *There exists constants $a_n > 0$ and b_n such that*

$$\frac{\max(X_1, X_2, \dots, X_n) - b_n}{a_n}$$

converges to a non-degenerate distribution if and only if the function U satisfies (for all $x > 0$)

$$\lim_{t\to\infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}$$

with a a positive function and $\gamma \in \mathbb{R}$.
The limit distribution (with $b_n = U(n)$ and $a_n = a(n)$) is

$$\exp\left\{-(1+\gamma x)^{-1/\gamma}\right\}$$

for values of x for which this is in (0,1).

An important statistical problem is how to estimate γ from a finite sample $X_1, X_2, \ldots, X_n$ from F.

I shall restrict myself here to the case $\gamma > 0$ and in particular to the case $P\{X_i > 0\} = 1$.

Theorem 2.3. *there exist constants $a_n > 0$ such that*

$$\frac{\max(X_1, X_2, \ldots, X_n)}{a_n}$$

converges to a non-degenerate distribution if and only if the function U satisfies

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\gamma$$

for all $x > 0$ with $\gamma > 0$. The limit distribution (with $a_n = U(n)$) is

$$\exp -x^{-1/\gamma}$$

for $x > 0$.

Remark 2.4. *As we have seen, the uniform distribution satisfies the conditions with $\gamma = 1$.*

A well-known estimator for γ is due to B. Hill (1975):

$$\hat{\gamma}_{k,n} := \frac{1}{k} \sum_{i=0}^{k-1} \log X_{n-i,n} - \log X_{n-k,n}.$$

So the estimator involves the k top order statistics of the sample.

Relevant for $\hat{\gamma}_n$ are the Poisson approximation of the extreme order statistics and the Brownian motion approximation of the intermediate order statistics.

Theorem 2.5. *If*

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\gamma \tag{2}$$

for $x > 0$, then

$$\hat{\gamma}_{k,n} \to \gamma \quad \textit{in probability}$$

for sequences $k = k(n) \to \infty$, $k(n)/n \to 0$ ("any intermediate sequence").

Remark 2.6. *Condition (2) is sufficient for the Poisson approximation but not for the Brownian motion approximation.*

For the asymptotic normality of $\hat{\gamma}_{n,k}$ we do need the Brownian motion approximation:

Theorem 2.7. *If for some positive function a*

$$\lim_{t\to\infty} \frac{\frac{U(tx)}{U(t)} - x^\gamma}{a(t)} =: \Phi(x) \tag{3}$$

exists for all $x > 0$ with Φ not constant, then

$$\sqrt{k}\left(\frac{\hat{\gamma}_{n,k}}{\gamma} - 1\right)$$

is asymptotically standard normal for intermediate sequences $k(n)$ that do not grow to fast.

Remark 2.8. *Condition (3) is sufficient for the Brownian motion approximation. The limit random variable can be expressed in terms of the limiting Brownian motion $\{W(x)\}$ as follows*

$$\int_0^1 \frac{W(x)}{x} dx - W(1).$$