July 27th, 2012

Law of large random numbers

Takaaki SHIMURA

(The Institute of Statistical Mathematics)

1 Introduction

1.1 Motivation and Problem

In EVT, the data we are interested in are

“large (or small) and random numbers”.

We consider a numerical characteristic of large random numbers.

For example,

【Problem 1】 Which is the most suitable as a large random number?

A. 70081 . . .
B. 62320 . . .
C. 19808 . . .

The above express the first figures (digths) of larger numbers in the decimal system.

【Problem 2】 What kind of distribution appears as the distribution of the second figure (digit)? The second figure in A, B and C is 0, 2, 9, respectively.

A. 70081 . . .
B. 62320 . . .
C. 19808 . . .

The distribution of the second figure is a distribution on $\{0, 1, 2, \ldots, 9\}$.

Of course, those things depend on
"the randomness that produces numbers".

Precisely, the tail behavior :

$$P(X > x)\ (x \to \infty)$$

is essential.

The tail of a distribution F means $1 - F(x)$.

1.2 Mathematical setting

We consider a transformation from a large number to a number in $[0, 1)$, which moves the decimal point and excludes the first figure.

【Transformation on $[1, \infty)$ to $[0, 1)$】

$$d_1 d_2 d_3 \ldots d_n . d_{n+1} \ldots \text{ in } [10^{n-1}, 10^n)$$
$$\to 0.d_2 d_3 \ldots \text{ in } [0, 1),$$

where n is a natural number.

Let X be a random variable with infinite end point.
If $X = d_1 d_2 d_3 \dots d_n . d_{n+1} \dots$ on $[1, \infty)$, then $Y = 0.d_2 d_3 \dots d_n d_{n+1} \dots$ is a random variable on $[0, 1)$.
We are interested in the distribution of Y for large X, which implies the behavior of the large random number except the first figure.

F is a distribution on real line with infinite end point : $\sup\{x : F(x) < 1\}$.
Let X be a random variable with distribution F.

Define N and K as
$N : 10^{N-1} \leq X < 10^N$,
$K : K10^{N-1} \leq X < (K+1)10^{N-1}$

Then previous transformation is written as

$$Y = (X - K10^{N-1})/10^{N-1}$$

Let us consider the conditional distribution.

$$F^{k,n}(x) = P(Y \leq x | K = k, N = n),$$
$$for\ k = 1, 2, \ldots, 9.$$

Our main interest is in the behavior of $F^{k,n}$ for each k as $n \to \infty$.

1.3 Agenda

1. Introduction
2. Preliminaries
3. Main result
4. Secondary results

2 Preliminaries

We introduce regularly varying function and its related notion.
They are used to characterize distribution classes defined by tail behaviors.

【Regularly varying function】

A positive measurable function $f(x)$ is said to be regularly varying with exponent (index) $\rho (\in \mathbf{R})$ ($f \in \mathbf{R}_\rho$) if for each $k > 0$

$$\lim_{x \to \infty} f(kx)/f(x) = k^\rho .$$

In the case of $\rho = 0$, it is called slowly varying. $f(x) \in \mathbf{R}_\rho$ is written as $f(x) = x^\rho l(x)$ with a slowly varying $l(x)$. e.g. $f(x) = x^2 \log x$

【Distributions with regularly varying tails】

The stable distributions except the normal distribution, the Pareto distribution, F-distribution and the Ziph distribution have regularly varying tail.

【Π-varying function】

A positive measurable function $f(x)$ on $(0, \infty)$ is Π-varying if there exists a positive function $a(x)$ on $(0, \infty)$ such that for $\lambda > 0$,

$$\lim_{x \to \infty} \frac{f(\lambda x) - f(x)}{a(x)} = \log \lambda.$$

We write $f \in \Pi$ or $f \in \Pi(a)$. $a(x)$ is called an auxiliary function of $f(x)$.

For example,
$f(x) = \log x$ is Π-varying with $a(x) = 1$.

Roughly speaking, a Π-varying function is slowly varying with good (local) property. Indeed, the behavior of a slowly varying is various. e.g.
$2 + \sin \log^2 x$.
A distribution with $1/\Pi$-varying tail seems to be rare. The log Cauchy distribution is the case.

【Rapidly varying function】

$f(x)$ is said to be a rapidly varying with exponent ∞ ($f \in \mathbf{R}_\infty$) if for each $\lambda > 1$ $\lim_{x\to\infty} f(\lambda x)/f(x) = \infty$.

$f(x) = \exp x$ is rapidly varying.

In the same way, $f \in \mathbf{R}_{-\infty}$ if for each $\lambda > 1$ $\lim_{x\to\infty} f(\lambda x)/f(x) = 0$.

【Distributions with rapidly varying tails】

Many (Most?) distributions have rapidly varying tail.
As a distribution with very rapid tail decay, the normal distribution and the Rayleigh distribution.
As middle tail decay, the exponential type, i.e. the exponential distribution, the Gamma distribution, the Chi-square distribution, the generalized inverse Gaussian distribution.

As a little bit heavy tail, the lognormal distribution also has a rapidly varying tail.

3 Main result

Remember

$$F^{k,n}(x) = P(Y \leq x | K = k, N = n),$$

for $k = 1, 2, \ldots, 9$.
Our main interest is in the behavior of $F^{k,n}$ for each k as $n \to \infty$.

We show that

· the limit distribution $F^k(x) = \lim_{n\to\infty} F^{k,n}$ exists for most distributions.

· 3 kinds of limit distributions appear depending on the tail behaviors.

【Notation】

$\bar{F}(x) = 1 - F(x)$: the tail of a distribution F.

$$
\begin{aligned}
F^{k,n}(x) &= P(Y \leq x | K = k, N = n) \\
&= \frac{P(k10^{n-1} \leq X \leq (k+x)10^{n-1})}{P(k10^{n-1} \leq X < (k+1)10^{n-1})} \\
&= \frac{\bar{F}(k10^{n-1}) - \bar{F}((k+x)10^{n-1})}{\bar{F}(k10^{n-1}) - \bar{F}((k+1)10^{n-1})} \\
&= \frac{1 - \bar{F}((k+x)10^{n-1})/\bar{F}(k10^{n-1})}{1 - \bar{F}((k+1)10^{n-1})/\bar{F}(k10^{n-1})}
\end{aligned}
$$

The third equality holds for continuous F, but it is not essential.

If $\bar{F}(x) \in \mathbf{R}_{-\infty}$, for $x > 0$

$$\lim_{n\to\infty} \bar{F}((k+x)10^{n-1})/\bar{F}(k10^{n-1}) = 0.$$

If $\bar{F}(x) \in \mathbf{R}_{-\alpha}(\alpha > 0)$,

$$\lim_{n\to\infty} \bar{F}((k+x)10^{n-1})/\bar{F}(k10^{n-1}) = (1+x/k)^{-\alpha}.$$

If $1/\bar{F}(x) \in \Pi$,

$$\bar{F}(k10^{n-1}) - \bar{F}((k+x)10^{n-1}) \sim \log(1+\frac{x}{k})a(k10^{n-1}$$

Theorem 1

(i) If $\bar{F}(x) \in \mathbf{R}_{-\infty}$, for every k,

$$\lim_{n\to\infty} F^{k,n}(x) = 1_{\{x\geq 1\}},$$

where 1_A denotes the indicator function of a set A.
Namely, $F^{k,n}$ converges to δ_0 (a distribution concentrates at $\{0\}$ as $n \to \infty$.

(ii) If $\bar{F}(x) \in \mathbf{R}_{-\alpha}(\alpha > \mathbf{0})$, for $0 \le x \le 1$,

$$F^k(x) = \lim_{n\to\infty} F^{k,n}(x) = \frac{1 - (1 + \frac{x}{k})^{-\alpha}}{1 - (1 + \frac{1}{k})^{-\alpha}}.$$

(iii) If $1/\bar{F}(x) \in \Pi$, for $0 \le x \le 1$,

$$F^k(x) = \lim_{n\to\infty} F^{k,n}(x) = \frac{\log(1 + \frac{x}{k})}{\log(1 + \frac{1}{k})}.$$

4 Secondary results

We give a secondary result for each case.

First, the tail condition in the case (iii) is $1/\Pi$-varying, not general slowly varying. The following shows that this restriction is significant.

Theorem 2 For any distribution F with slowly varying tail and any distribution G on $[0,1)$, there exists a distribution F_G such that $\lim_{x\to\infty} \bar{F}_G(x)/\bar{F}(x) = 1$ and $F_G^{k,n} = G$ (independent of n and k).

Proof.

Let $X_1 = K10^{N-1}$ and $X_2 = X - X_1$.

Since $X_1 \leq X < 2X_1$, we have

$P(X > x) \sim P(X_1 > x)$.

For $Z \sim G$, set $Y = X_1 + 10^{N-1}Z$.

$P(X > x) \sim P(X_1 > x) \sim P(Y > x)$. ■

In the case (i), the following explains the rate of converge to δ_0.

Theorem 3 $\bar{F}(x) \in \mathbf{R}_{-\infty}$ Moreover, assume that F is absolutely continuous and its hazard function $h(t)$ belongs to $\mathbf{R}_\rho(\rho \geq -\mathbf{1})$.
For $0 \leq x < 1$,

$$\lim_{n\to\infty} \frac{1}{10^{n-1}h(10^{n-1})} \log \overline{F^{k,n}}(x) = -c(\rho, k, x),$$

where

$$c(\rho, k, x)$$
$$= \begin{cases} (\rho+1)^{-1}\{(k+x)^{\rho+1} - k^{\rho+1}\} & \rho > -1 \\ \log(1+\frac{x}{k}) & \rho = -1 \end{cases}$$

【The property of limit distribution】

In the cases (ii) and (iii), we investigate the property of limit distribution.
In these cases,

“the limit distribution depends on the first figure k.”

The density of the limit distribution in (ii) is as follows.

$$p_\alpha^k(x) = \frac{\alpha k^{-1}(1+\frac{x}{k})^{-\alpha}}{1-(1+\frac{1}{k})^{-\alpha}} \quad \text{for} \quad 0 \le x \le 1.$$

$p_\alpha^k(x)$ is decreasing for each $k = 1, 2, \ldots, 9$. The degree of decrease tends to be flat as k increases.

The density of the limit distribution in (iii) is as follows.

$$p_0^k(x) = \frac{1}{\log(1+\frac{1}{k})}\frac{1}{k+x} \quad \text{for} \;\; 0 \le x \le 1.$$

$p_0^k(x)$ is decreasing for each $k = 1, 2, \ldots, 9$. The degree of decrease tends to be flat as k increases.

It means that

· smaller figure appears frequently compared with larger figures as the second figure.

This property is common in every figure larger order than the second.

· This tendency decays as k increases.

Proposition M_k, the mean of the limit distributions F^k in (ii) and (iii), are

$$M_k = \begin{cases} \frac{(\alpha+k)(k+1)^{-\alpha} - k^{-\alpha+1}}{(1-\alpha)(k^{-\alpha} - (k+1)^{-\alpha})} & \alpha \neq 1 \\ k(k+1)\log(1+\frac{1}{k}) - k & \alpha = 1 \\ 1 - k\log(1+\frac{1}{k}) & 1/\bar{F}(x) \in \Pi \end{cases}$$

M_k increases for k.

Proposition The distribution of the mth figure after the decimal point of F^k converges to the uniform distribution on $\{0, 1, \ldots, 9\}$ as $m \to \infty$.

4.1 Conclusion

【Problem 1】

Which is the most suitable as a large random number?

A. 70081 . . .

B. 62320 . . .

C. 19808 . . .

【Answer】

“A” seems to be the most suitable.

“C” does not seem to be enough large.

Thank you for your attention!

References

Shimura(2011). Limit distribution of a roundoff error, Statistics and Probability Letters **82**, 713-719.