Stochastic Processes and their Statistics in Finance

A numerical characteristic of extreme values

Takaaki SHIMURA

(The Institute of Statistical Mathematics)

0.1 Plan of Presentation

1. Introduction
2. Main result
3. Limit distributions
4. Summary

1 Introduction

1.1 Motivation and Problem

We consider a numerical characteristic of random numbers. Especially,

“ Extremely large random numbers or small random numbers”.

Classification of normal random numbers by the first figure

$$[0, 1) : 0.59922,\ \ 0.39319,\ \ 0.11950,\ \ 0.01336$$

$$[1, 2) : 1.04194,\ \ 1.43943,\ \ 1.38955,\ \ 1.66662$$

$$[2, 3) : 2.19377,\ \ 2.40794,\ \ 2.14139,\ \ 2.32582$$

$$[3, 4) : 3.06956,\ \ 3.86446,\ \ 3.20402,\ \ 3.04337,$$
$$3.07787,\ \ 3.16713,\ \ 3.45392,\ \ 3.04813$$

1.2 Mathematical setting

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

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

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

$$\to 0.d_2 d_3 \ldots \text{ in } [0, 1),$$

where n is a natural number.

We call d_m the mth figure.

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

If $X = d_1 d_2 d_3 \ldots d_n . d_{n+1} \ldots$ on $[1, \infty)$, then $Y = 0.d_2 d_3 \ldots d_n d_{n+1} \ldots$ is a random variable on $[0, 1)$.

We consider the distribution of Y for large X, which implies the behavior of the large random number except the first figure.

Define N and K as

N : the number of figures before the decimal point of X : $10^{N-1} \leq X < 10^N$,

K : the first figure of X :

$$K10^{N-1} \leq X < (K+1)10^{N-1}$$

Then previous transformation is written as

$$Y = X/10^{N-1} - K.$$

Let us consider the conditional distribution.

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

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

1.3 Classification of distributions

Denote the tail of a distribution F by

$$\bar{F}(x) = 1 - F(x).$$

F is said to have regularly varying tail with index $\alpha > 0$ if

$$\lim_{x \to \infty} \bar{F}(\lambda x)/\bar{F}(x) = \lambda^{-\alpha} \text{ for each } \lambda > 0.$$

For example, the Cauchy, the Pareto, the F and the Ziph distributions have regularly varying tail.

【Regularly varying function】

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

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

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

【Rapidly varying function】

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

$$\lim_{x \to \infty} f(\lambda x)/f(x) = \infty.$$

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

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

Rapidly varying tail distributions are various.

· Very rapid tail decay : the normal distribution and the Rayleigh distribution.

· Middle tail decay: the exponential type, i.e. the exponential distribution, the Gamma distribution, the Chi-square distribution, the generalized inverse Gaussian distribution.

· Little bit heavy tail : the log-normal distribution.

【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 $a(x) = 1$.

$f(x) = \log x + 2^{-1}\sin(\log x)$ is NOT Π-varying.

Roughly speaking, a Π-varying function is nondecreasing slowly varying with good (local) property.

Distribution with $1/\Pi$-varying tail is so to speak "super heavy" and not so familiar.
The log Cauchy distribution is the case.

Thus we have three kind of tail behaviors :

- $\bar{F}(x) \in \mathbf{RV}_{-\alpha}(\alpha > 0)$,
- $\bar{F}(x) \in \mathbf{RV}_{-\infty}$,
- $1/\bar{F}(x) \in \Pi \ (\subset \mathbf{RV_0})$.

These classes cover most well-known distributions with infinite endpoint.

Distributions with finite endpoint are classified in the same way.

Let F be a distribution on $(-\infty, 0)$ with the endpoint 0 : $\sup\{x : F(x) < 1\} = 0$.

We say that F has regularly varying tail at 0 if

$$\bar{F}(-1/x) \in \mathbf{RV}_\alpha (\alpha < 0).$$

Rapidly varying and $1/\Pi$ varying tail at 0 are defined in a similar way.
The definition for a general finite endpoint is also done.

【Examples】

(i) Regularly varying tail at their finite endpoint: The Beta distribution and the Pareto distribution.

(ii) Rapidly varying tail at their finite endpoint : The exponential distribution and the log-normal distribution.

2 Main result

2.1 Large random numbers

Remember

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

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

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

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

$$F^{k,n}(y) = \frac{1 - \bar{F}((k+y)10^{n-1})/\bar{F}(k10^{n-1})}{1 - \bar{F}((k+1)10^{n-1})/\bar{F}(k10^{n-1})}.$$

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

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

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

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

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

$$\begin{aligned} \bar{F}(k10^{n-1}) - \bar{F}((k+y)10^{n-1}) \\ \sim \log(1+\frac{y}{k})a(k10^{n-1}). \end{aligned}$$

Theorem 1

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

$$\lim_{n\to\infty} F^{k,n}(y) = 1_{\{y\geq 0\}},$$

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 $\bar{F}(x) \in \mathbf{RV}_{-\alpha}(\alpha > \mathbf{0})$, for $0 \leq y \leq 1$,

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

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

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

Let

$$G^k_\alpha(y) = \frac{1-(1+\frac{y}{k})^{-\alpha}}{1-(1+\frac{1}{k})^{-\alpha}},$$

$$G^k_0(y) = \frac{\log(1+\frac{y}{k})}{\log(1+\frac{1}{k})}.$$

(i) and (iii) are regarded as the limit of (ii) : $G^k_\alpha(y)$ converges to δ_0 and G^k_0 as $\alpha \to \infty$ $\alpha \to 0$, respectively.

We add some secondary results.

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

$$\lim_{x\to\infty} \bar{F}_G(x)/\bar{F}(x) = 1 \text{ and } \; F_G^{k,n} = G.$$

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 $Y = X_1 + 10^{N-1}Z$.

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

The rate of converge to δ_0 in (i) is as follows.

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

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

where

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

$c(\rho, k, y)$ expresses the rate of convergence to δ_0.

【Property of $c(\rho, k, y)$ as a function of k】

(i) If $-1 \le \rho < 0$, $c(\rho, k, y)$ is a decreasing function of k.

(ii) $c(0, k, y) = (\rho + 1)^{-1} y$ does not depend on k. Especially, $F^{k,n}$ does not depend on k if F is an exponential distribution.

(iii) If $\rho > 0$, $c(\rho, k, y)$ is an increasing function of k.

2.2 Small random numbers

From now, we deal with distributions with finite endpoint. We assume F has finite end point : $x_F = \sup\{x : F(x) < 1\} < \infty$.

Let X be a random variable with distribution F and consider the length until the endpoint :

$$x_F - X.$$

Let $x_F = 0$ for simplicity and transform from $(-\infty, 0)$ to $[0, 1)$:

$$X = -0.0 \cdots 0 d_1 d_2 d_3 \ldots \in [-\infty, 0),\ d_1 \neq 0$$

$$\to Y = 0.d_2 d_3 \ldots \in [0, 1).$$

Define a normalized random variable Y as

$$Y = -10^N X - K,$$

where K is the first non-zero figure of X and N is the number of zeros before K :

$$-10^{-N+1} < X \leq -10^{-N}$$
$$-10^N X - 1 < K \leq -10^N X.$$

Y expresses the behavior of X except the first non-zero figure.

The following conditional distribution is considered.

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

As the large case, the behavior of $F^{k,n}$ as $n \to \infty$ for each k is investigated and similar results are given.

Theorem 4

(i) If $\bar{F}(-1/x) \in \mathbf{RV}_{-\infty}$, then for every $k = 1, 2, \ldots, 9$,

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

where 1_A is the indicate function of a set A.

(ii) If $\bar{F}(-1/x) \in \mathbf{RV}_\alpha(\alpha < \mathbf{0})$, then for $0 \le y \le 1$,

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

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

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

These limit distributions for small random numbers have the same form as ones for large case.
But the parameter range is different. While the large case is in $\alpha \geq 0$, the small case moves in $\alpha (\leq 0)$.
Thus we get two parameter distribution class :

$$G_{\alpha}^{k}(\alpha \in (-\infty, \infty), k = 1, 2, \ldots, 9.),$$

Theorem 5 For arbitrary distribution F with slowly varying tail at 0 and arbitrary distribution G on $[0, 1)$, there exists a distribution F_G such that

$$\lim_{x \uparrow 0} \bar{F}_G(x)/\bar{F}(x) = 1 \text{ and } F_G^{k,n} = G.$$

Theorem 6 Assume that $\bar{F}(-1/x) \in \mathbf{RV}_{-\infty}$ is absolutely continuous and its hazard function satisfies $h(-1/t) \in \mathbf{R}_\rho(\rho \geq \mathbf{1})$. For $0 < y \leq 1$,

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

where

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

In this case, $\tilde{c}(\rho, k, y)$ is decreasing on k.

3 Property of limit distributions

3.1 Property variety of α and k

$G_\alpha^k \ (k = 1, 2, \ldots, 9,\ \alpha \in (-\infty, \infty))$.

$$G_\alpha^k(x) = \frac{1 - (1 + \frac{x}{k})^{-\alpha}}{1 - (1 + \frac{1}{k})^{-\alpha}} \quad (\alpha \neq 0),$$

$$G_0^k(x) = \frac{\log(1 + \frac{x}{k})}{\log(1 + \frac{1}{k})} \qquad (\alpha = 0).$$

Note $\alpha \geq 0$: large case. $\alpha \leq 0$: small case.

For each k, G_α^k moves between δ_1 and δ_0.

Proposition 1

(i) G_α^k converges to δ_0 ad $\alpha \to \infty$.
(ii) G_α^k converges to δ_1 as $\alpha \to -\infty$.
(iii) G_{-1}^k is the uniform distribution on $[0, 1]$.

The density functions of $G_\alpha^k (\alpha \geq 0)$ denoted by $p_\alpha^k(y)$ are given as

$$p_\alpha^k(y) = \frac{\alpha k^{-1}(1 + \frac{y}{k})^{-\alpha-1}}{1 - (1 + \frac{1}{k})^{-\alpha}},$$

$$p_0^k(y) = \frac{1}{\log(1 + \frac{1}{k})} \frac{1}{k + y}$$

for $0 \leq y \leq 1$.

Proposition 2

(i) For each k, $p_\alpha^k(y)$ is a decreasing (resp. constant, increasing) function of y and $\alpha > -1$ $(resp.\ \alpha = -1, \alpha < -1)$.

(ii) $p_\alpha^k(0)$ is an increasing function of α for each k. While $p_\alpha^k(1)$ is decreasing function of α for each k.

(iii) $p_\alpha^k(0)$ is a decreasing (resp. constant, increasing) function of k for $\alpha > -1$ $(resp.\ \alpha = -1, \alpha < -1)$.
$p_\alpha^k(1)$ is an increasing (resp. constant, decreasing) function of k for $\alpha > -1$ $(resp.\ \alpha = -1, \alpha < -1)$.

The probability density tends to be flat as k increases.

The distribution function G_α^k has the following monotonicity.

Proposition 3

(i) $G_\alpha^k(y)$ is an increasing function of α for each k and y.

(ii) $G_\alpha^k(y)$ is is a decreasing (resp. constant, increasing) function of k for $\alpha > -1$ $(resp.\ \alpha = -1, \alpha < -1)$.

M_α^k denotes the mean of G_α^k.

Corollary 1

(i) M_α^k is a decreasing function of α for each k.

(ii) M_α^k is an increasing (resp. constant, decreasing) function of k for $\alpha > -1$ $(resp.\ \alpha = -1, \alpha < -1)$.

3.2 The limit distribution of m th figure

H_m^k : the distribution of the $(m-1)$th figure after the decimal point of G_α^k $(m = 2, 3, \ldots)$.
H_m^k is a distribution on $\{0, 1, \cdots, 9\}$
Originally, H_m^k implies the distribution of mth figure of a original random number.

$$X = d_1 d_2 d_3 \ldots d_m \ldots$$

with $d_1 = k$.

Although $H_m^k(j)$ is decreasing for j from Proposition 2, this property disappears as m goes to ∞.

Proposition 4 For each k, H_m^k converges to the uniform distribution on $\{0, 1, \ldots, 9\}$ as $m \to \infty$.

This suggest the distribution of the second figure expresses the original distribution.

【Distribution of the second figure $\alpha = 1$】

$$k = 1 : H_2^1(0) = 2/11 \quad H_2^1(9) = 1/19$$
$$k = 9 : H_2^9(0) = 10/91 \quad H_2^9(9) = 1/11$$

$$H_2^1(0)/H_2^1(9) = 38/11 = 3.45\ldots$$
$$H_2^9(0)/H_2^9(9) = 110/91 = 1.20\ldots$$

The ratio is $2.857\ldots$.

【Distribution of the second figure $\alpha = 0$】

$$k = 1 : H_2^1(0) = \log(11/10)/\log 2$$
$$H_2^1(9) = \log(20/19)/\log 2$$
$$k = 9 : H_2^9(0) = \log(91/90)/\log(10/9)$$
$$H_2^9(9) = \log(100/99)/\log(10/9)$$

$$H_2^1(0)/H_2^1(9) = 1.85\dots$$
$$H_2^9(0)/H_2^9(9) = 1.10\dots$$

The ratio is $1.69\dots$.

4 Summary

· Random numbers have a numerical characteristic. Especially, it is remarkable in extreme values.

· An extreme value (conditioned by the first figure) converges to a limit distribution depends on each tail behavior.

· The limit distribution depends on the tail behavior and the first figure.

References

Bingham, N.H., Goldie, C.M. and Teugels, J.L.(1987). Regular Variation. Cambridge. Cambridge University press.

Shimura, T.(2012). Limit distribution of a roundoff error, Statistics and Probability Letters **82**, 713-719.

Shimura, T.(2013).A numerical characteristic of extreme values, submitted.

Thank you for your attention!