A Model of Genesis of Income Distribution (minimal version)

Hiroshi Hamada
Tohoku University

2016/01/28

The lognormal distribution is intuitively defined as the distribution of a random variable whose logarithm is normally distributed (Aitchison and Brown 1957; Crow and Shimizu 1988). In the field of social science, lognormal distribution has been often used for mathematical description of income distribution.

McAlister is the first person to describe a possible model of genesis of the lognormal distribution. Kapteyn established more clearly the genesis of the distribution (Kapteyn 1903). The lognormal distribution was first applied to income distribution by Kapteyn. Gibrat illustrated the law of proportionate effect with extensive income data from many countries and over many years (Gibrat 1931)[1].

In this paper, we show alternative process of generation of lognormal distribution. Namely, we prove that a simple repeated investment game can generates a lognormal distribution.

Basic Assumptions

- People in a society repeat an investment game n times with success probability $p \in (0, 1)$ and failure probability $1 - p$. The success probability p is common to all members in a society and p is fixed through the time.
- $y_0 \in \mathbb{R}^+$ and $b \in (0, 1)$ denote an initial income and an investment rate respectively. b is a constant. y_t indicates the income at time t.
- People invest at each stage game with constant proportion b of income y_t, in other words an investment cost is $y_t b$ at time t.
- y_t may differ among people in a society depending on the result of the investment game. Inequality will emerge as time passes.
- On the one hand, people gain the profit $y_{t-1}b$ when they succeed at time t. In a success situation, the income at time t is defined as $y_t = y_{t-1} + y_{t-1}b$. On the other hand, people lose $y_{t-1}b$ when they fail at time t. In a failure situation, the income at t is defined as $y_t = y_{t-1} - y_{t-1}b$.

[1] It is known that the lognormal distribution approximates incomes in the middle range but fails in the upper tale where the Pareto distribution is more appropriate (Crow and Shimizu 1988).

Lemma 1 (An income at time n)**.** After repeating an investment game n times, income y_n can be written as $y_n = y_0(1+b)^w(1-b)^{n-w}$ where w and $n-w$ are the numbers of success and failure in n times repeated game.

Proof. This lemma is proved by mathematical induction. Consider the case $n = 1$. People obtain an income $y_1 = y_0 + y_0 b = y_0(1+b)$ with success, and $y_1 = y_0 - y_0 b = y_0(1-b)$ with failure. Thus

$$y_n = y_0(1+b)^w(1-b)^{n-w}$$

is true for the case $n = 1$.

Consider the case $n = k$. Suppose that $y_k = y_0(1+b)^w(1-b)^{k-w}$ is hold with w times success and $k-w$ times failure at time k.

Next, consider the case $k+1$. People obtain an income

$$y_{k+1} = y_k + y_k b = y_k(1+b)$$

if the player succeeds at time $k+1$. By the assumption $y_k = y_0(1+b)^w(1-b)^{k-w}$, we have

$$\begin{aligned} y_{k+1} &= y_k(1+b) = (y_0(1+b)^w(1-b)^{k-w})(1+b) \\ &= y_0(1+b)^{w+1}(1-b)^{k+1-(w+1)}. \end{aligned}$$

If the player fails at $k+1$. By the assumption $y_k = y_0(1+b)^w(1-b)^{k-w}$, we have

$$\begin{aligned} y_{k+1} &= y_k(1-b) = (y_0(1+b)^w(1-b)^{k-w})(1-b) \\ &= y_0(1+b)^w(1-b)^{(k+1)-w}. \end{aligned}$$

This implies that y_{k+1} holds if y_k holds. Thus

$$y_n = y_0(1+b)^w(1-b)^{n-w}$$

holds for every n by mathematical induction. □

Proposition 1 (Genesis of a lognormal distribution)**.** If n is sufficiently large, the income y_n obeys a lognormal distribution.

Proof. Let Y_n be a random variable which indicates income at time n. Let W be a random variable which indicates a number of success in n times. by lemma 1, Y_n can be expressed as

$$Y_n = y_0(1+b)^W(1-b)^{n-W}.$$

Apparently Y_n is a function of random variable W. Taking logarithm of Y_n, we obtain

$$\log Y_n = \log \frac{1+b}{1-b} W + \log y_0 + n \log(1-b)$$

By the De Moivre=Laplace theorem, W obeys a normal distribution if $n \to \infty$. Let us define A and B as

$$A = \log \frac{1+b}{1-b} \quad \text{and} \quad B = \log y_0 + n \log(1-b).$$

Note that A and B are not random variables but constants. Thus $\log Y_n$ is simply written as

$$\log Y_n = AW + B.$$

The right hand side of the equation obeys a normal distribution because a linear function of W obeys a normal distribution. Therefore the left hand side of the equation, $\log Y_n$, obeys a normal distribution too. It implies Y_n obeys a lognormal distribution by definition. □

Proposition 2 (The probability density function of an income distribution)**.** The probability density function of an income distribution that is derived from repeated investment games is

$$f(y) = \frac{1}{\sqrt{2\pi npqA^2}} \frac{1}{y} \exp\{-\frac{1}{2}\frac{(\log y - BAnp)^2}{npqA^2}\}$$
$$\text{where} \qquad A = \log\frac{1+b}{1-b}, \quad B = \log y_0 + n\log(1-b).$$

Proof. The probability function of W (a number of success in n times repetition) is

$$P(W = k) =_n C_k p^k (1-p)^{n-k}.$$

By the De Moivre=Laplace theorem, $P(W = k)$ can be approximated by the following probability density function of a normal distribution

$$f(w) = \frac{1}{\sqrt{2\pi npq}} \exp\{-\frac{(w-np)^2}{2npq}\}.$$

Suppose that the random variable Y represents an income of infinitely repeated investment game. Remind that $\log Y$ can be written as the following function,

$$\log Y = AW + B, \text{where} A = \log\frac{1+b}{1-b}, B = \log y_0 + n\log(1-b).$$

Let $\log Y = X$, then $X = AW + B$. The probability density function of X is

$$\begin{aligned} f(x) &= \frac{1}{\sqrt{2\pi npq}} \exp\left\{-\frac{(\frac{x-B}{A} - np)^2}{2npq}\right\} \frac{dw}{dx} \\ &= \frac{1}{\sqrt{2\pi npqA^2}} \exp\left\{-\frac{(w - B - Anp)^2}{2npqA^2}\right\} \end{aligned}$$

by change of variable. By assumption $\log Y_n = X$, we obtain the probability density function of Y_n. Namely,

$$f(y) = \frac{1}{\sqrt{2\pi npqA^2}} \frac{1}{y} \exp\left\{-\frac{1}{2}\frac{(\log y - BAnp)^2}{npqA^2}\right\}$$
$$\text{where} \qquad A = \log\frac{1+b}{1-b}, \quad B = \log y_0 + n\log(1-b).$$

□

Proposition 3 (The Gini index of income distribution derived by repeated games)**.** As time n increases, the inequality of income distribution enlarges. Also as investment rate b increases, the inequality enlarges. The inquality maximizes if $p = 0.5$.

Proof. By the theorem of Aitchison and Brown (1957), the Gini index of the income distribution that obeys a lognormal distribution $\Lambda(\mu, \sigma^2)$ can be written as

$$G = 2 \int_{-\infty}^{\sigma/\sqrt{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx - 1.$$

The Gini index increases as the parameter σ increases, since the upper range of the integral is $\sigma/\sqrt{2}$. In the repeated investment game, the parameter σ can be written as

$$\sigma = np(1-p) \log \frac{1+b}{1-b}.$$

Differentiating σ by n, p, and b, we obtain the proposition. □

References

Aitchison, J. and J. A. C., Brown, 1957, *The Lognormal Distribution with Special Reference to It's Uses in Economics*, Cambridge: The Cambridge University Press.

Gibrat, R.,*1931, Les inegalites economiques*, Paris: Librarie du Recueil Sirey.

Kapteyn, J. C., 1903, *Skew Frequency Curves in Biology and Statistics*, Astronomical Laboratory, Groningen: Noordhoff.