The Bernoulli Core Approach and Bayesian Modeling: An Analysis of Income Distribution in Japan *

Hiroshi Hamada
(Tohoku University)

Abstract

The purpose of this study is to demonstrate the usefulness of the Bernoulli core approach and its affinity to the Bayesian statistical analysis. The Bernoulli core approach is a mathematically organized system of probabilistic models that consists of self-contained sub-models (Hamada 2017). Each sub-model is expressed as a probabilistic model and explains the process of genesis of distribution for specific outcome variables. We test the empirical validity of one of the sub-models, the generative model for income distribution (Hamada 2004; 2016). Our toy model, which is different from the black-box generalized linear model, formally represents a sociological theory and can explain the generative process of social action in terms of a rigorous micro-macro linkage. Our theory can be tested empirically by the Bayesian statistical analysis, since it is expressed as a stochastic model. To demonstrate the linkage between the toy model and statistical analysis, we estimate posterior distributions of the parameters of the probabilistic toy model by Markov chain Monte Carlo estimation. Using nationwide survey data in Japan, SSM2015, we compare a non-theoretical model with a theoretical one that includes a hierarchical model, by the widely appreciable information criteria and the leave-one-out cross validation. We find that predictive accuracy of the theory-based hierarchical model is fine and provides interesting information about latent parameters.

Keywords: income distribution, human capital, Bayesian statistical analysis

1. Introduction

1.1. The genesis of income distribution

In the field of social science, lognormal distribution[1] has often been used for the mathematical description of income distribution. McAlister (1879) was the first to present a possible model of genesis of the lognormal distribution. Kapteyn (1903) established more clearly the genesis of the distribution and was the first to apply the

* This research is supported by JSPS Grant-in-Aid for Specially Promoted Research (Grant number 25000001 and 16K13406), and I thank the 2015 SSM Survey Management Committee for allowing me to use the SSM data.

[1] The lognormal distribution is intuitively defined as the distribution of a random variable whose logarithm is normally distributed (Aitchison & Brown 1957; Crow & Shimizu 1988).

lognormal distribution to income distribution. Gibrat (1931) illustrated the law of proportionate effect with extensive income data from many countries and for a longer time period[2]. In Gibrat's model, the random chance factor plays an important role. If the random shock is applied to a proportional rather than absolute income change, the process converges to a lognormal distribution (Gibrat 1931; Mincer 1970). Champernowne (1953) elaborated Gibrat's model and showed that when certain assumptions about random shock are introduced, the income distribution converges to a Pareto distribution, which is used for approximation of the upper tail of income distribution. Aitchison and Brown (1957) showed the condition of lognormality for a society that consists of infinite subgroups; namely, if variances in the component distribution are of the same size and the means of the components are lognormally distributed, the aggregate distribution remains lognormal. Even though the assumptions are not realistic, their model provided a theoretical framework for the decomposition of income distribution of the entire society into subpopulations (Hamada 2005). Rutherford (1955) applied random shock to age cohorts and showed that the income variance increases with age for each cohort, but that aggregate variance does not change much with a relatively stable age distribution (Rutherford 1955; Mincer 1970).

1.2. The Bernoulli core approach

These previous studies suggest that explaining the genesis of income distribution itself can be an interesting and important research topic[3]. In the field of mathematical sociology, Hamada (2003, 2004) formalized a repeated investment game to explain the genesis of income distribution[4]. The generative model of income distribution is a sub-model of the general systematic theory, which is called the Bernoulli core approach.

[2] Strictly speaking, the lognormal distribution approximates incomes in the middle range, but fails in the upper tail, where the Pareto distribution is more appropriate (Crow & Shimizu 1988). For simplicity, we use only the lognormal distribution as an ideal type of income distribution.

[3] Friedman (1953) pointed out that the absence of a satisfactory theory of the personal distribution of income and of a theoretical bridge connecting the functional distribution of income with the personal distribution is a major gap in the modern economic theory.

[4] The generative model of income distribution was first proposed by Hamada (2003; 2004). Although the model successfully proved the lognormality of distribution of profit in the repeated game, assumptions are slightly complicated and the difference between the concepts of income (flow) and capital (stock) is not sufficiently clear. Therefore, Hamada (2016) attempted to simplify the model without loss of generality and extract more useful implications. In this study, we add an analysis of the model by the Bayesian statistical method to support the empirical validity of the model.

The Bernoulli core approach is an implementation of the general theoretical sociology, which is proposed by Fararo (1989). The Bernoulli core approach is a theory described by a system of random variables, in which each random variable represents specific distribution of resources such as income, education, or well-being. Traditionally, mathematical sociology has been used to build original models for explaining various social phenomena such as group process, individual action, resource distribution, and social institution. The Bernoulli core approach preserves the systematic relation of middle-range models in mathematical sociology, since each logical relation can be expressed as a transformation of random variables.

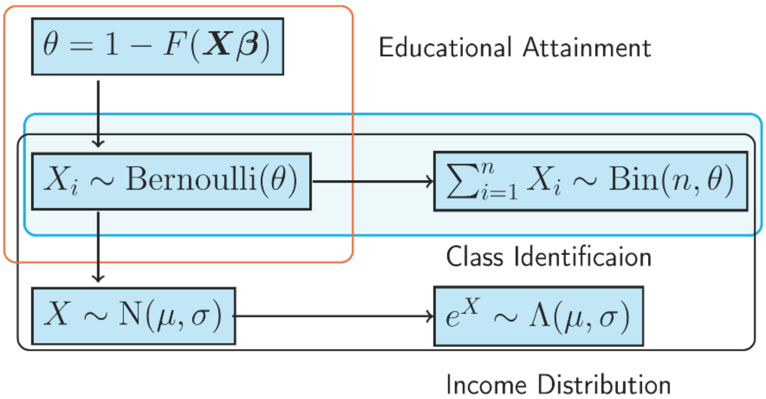

Figure 1: Illustration of the Bernoulli core approach

Figure 1 illustrates the systematic relation of random variables that correspond to various sociological models. The core of this systematic model is Bernoulli distribution since it is one of the simplest probability distributions. In this study, we focus on the model of genesis of distribution; however, the model is just a sub-model of a more general and systematic framework. Similar to Lego blocks or piano variations, we can build different models from combinations of transformation of random variables. For example, the model of middle-class identification, also known as the Fararo-Kosaka model of class identification (Kosaka & Fararo 2003), can be expressed as a combination of Bernoulli, binomial, and normal distribution. Ishida (2018) proposed a new model for class identification based on the random walk process (LLRW model). The LLRW model and the Farao-Kosaka model have the same mathematical structure, binomial distribution. Therefore these models also can be viewed as sub-models of the Bernoulli core approach since the random walk process is fully described by Bernoulli and binomial distribution. A model of class differentials on educational attainment, known as the relative risk aversion hypothesis can be viewed as the stochastic model whose outcome is Bernoulli distribution, in other words, to stay around or to leave advanced level of education (Breen & Goldthorpe 1997). Additionally, the generative model of income distribution can be expressed as a combination of Bernoulli, binomial, normal, and lognormal distribution as we will show in next section.

Thus, the Bernoulli core approach provides general, systematic, clear, and rigorous framework for sociological theory. Our model is different from previous studies in Economics in terms of orientation for general theory. Each sub-model of the Bernoulli core approach attempts to integrate interpretative sociology and analytical action theory (Fararo 1989).

1.3. Organization of this paper

We will organize this paper as follows. In Section 1, we propose the Bernoulli core approach as a general sociological theory framework. In Section 2, we briefly summarize the generative income distribution model proposed by Hamada (2016). The model focuses on the accumulation process of human capital by random chances and describes income as a gain from the capital. The main results of theoretical analysis suggest that capital and income distribution are asymptotically subject to a lognormal distribution. In Section 3, we construct a Bayesian model in order to test our toy model empirically using SSM2015 data. In Section 4, we estimate posterior distributions of parameters by Markov chain Monte Carlo (MCMC) method and compare the models by the widely applicable information criterion (WAIC) and the leave-one-out cross-validation (LOO). In terms of those criteria, we find that our toy model may have better predictive accuracy compared to models that are not based on the theory for genesis of income distribution.

2. Mathematical model of genesis of income distribution

2.1. Basic assumptions of the model

The assumptions of our simplified model are as follows (Hamada 2016). Hereafter, we use the symbols Y and W for capital and the number of success outcome, respectively, to emphasize that they are random variables.

1. People in a society experience random chance n times with success and failure probabilities of p and $1-p$, respectively, where $p \in (0,1)$. The probability p is

common to all members in a society and fixed through time[5].

2. $y_0 \in R^+$ and $b \in (0, 1)$ denote "an initial capital" and "an interest rate," respectively. b is a constant. Y_t indicates the amount of capital at time t.
3. At each chance, people invest a constant proportion b of capital Y_t. In other words, the investment cost is $Y_t b$ at time t.
4. On one hand, people earn a profit of $Y_{t-1}b$ when they succeed at time t. If they succeed, then the capital 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. If they fail, then the capital at time t is defined as $Y_t = Y_{t-1} - Y_{t-1}b$.

Figure 2 illustrates the process of capital accumulation under the assumptions. Each bifurcation indicates success or failure by random chances. As the diagram suggests, the capital Y_t may differ among people in a society depending on the result of random chances. Inequality of capital will emerge as the random chance is repeated.

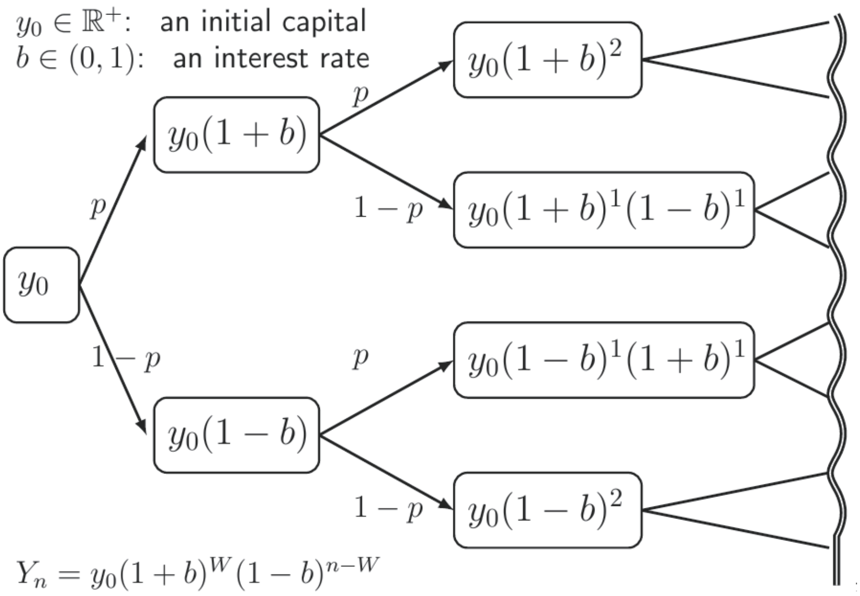

Figure 2: Tree diagram of the model.

The random chance p represents uncertainty of the return from an investment of capital. Interest rate b indicates the magnitude of profit from capital. As b increases, the

[5] This strong assumption can be generalized (Hamada 2016). Even though we assume the probability p differs among individuals in a society, the main implication of the model does not change. Hamada (2016) showed the capital and gained interest follow lognormal distribution even when success probability p has a probability distribution. This generalization is obtained by the application of Lyapunov's central limit theorem.

expected increment of capital increases. In our model, we define capital as an accumulated individual resource, such as human capital (i.e., knowledge or skills). We assume that human capital is the main resource of individual income or equivalently, individual labor supply (Hamada 2016).

Figure 3 shows the cumulative process of human capital graphically. Each line graph represents individual history of repeated investment. The line graph corresponds to the trajectory of a random walk with cumulative effect[6]. As the number of Bernoulli trials (time steps) increases, the variance of capital increases. On the one hand, the gap between individuals in the low range diminishes, and, on the other hand, the gap in the high range expands as the time steps increase.

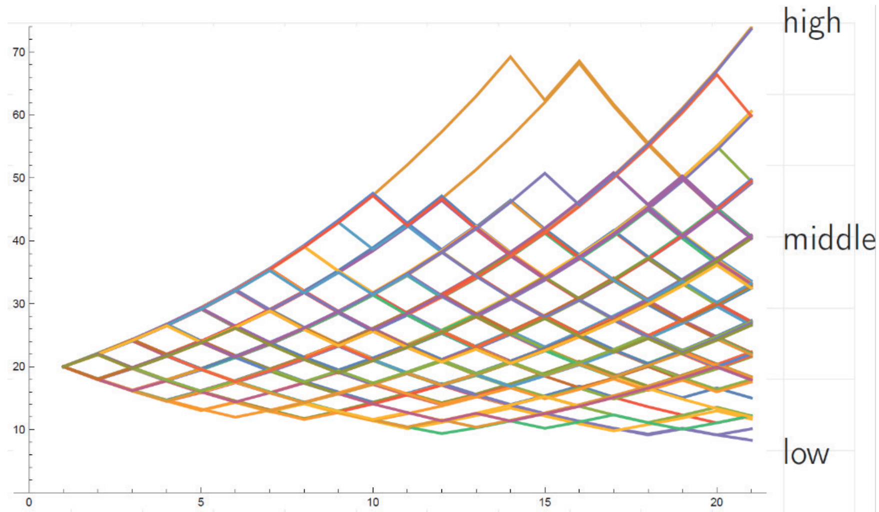

Figure 3: Illustration of the capital accumulation process. The horizontal and vertical axes indicate the number of Bernoulli trials and the amount of capital, respectively.

2.2. Basic properties and propositions of the model

One of the main results of our theoretical analysis is that the capital Y follows a lognormal distribution. The model specifies the probability density function of the capital distribution as a function of exogenous parameters of our model, namely, probability p, investment rate b, and time steps n. Additionally, by the virtues of the mathematical toy model, the average and inequality of capital distribution can be

[6] In terms of the Bernoulli core approach, our model can be viewed as an extended version of the LLRW model (Ishida 2018), since both model contain the random walk process, sum of Bernoulli random variables, and have Bayesian model expression.

expressed as functions of exogenous parameters. Our parsimonious model endogenously derives the average income and Gini coefficient of the distributions from exogenous parameters. Namely, the change of average income and inequality can be analyzed systematically by basic parameters p and b (Hamada 2016).

After n times random chance, capital 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 outcomes, respectively, in repeated games (Hamada 2016).

Proposition 1 (lognormal distribution of capital and income). If n is sufficiently large, the distribution of capital Y_n follows approximately a lognormal distribution. Namely, $Y_n \sim \Lambda(B + Anp,\, np(1-p)A^2)$ where

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

Moreover, the gained interest bY_n follows a lognormal distribution by the nature of probability density function of lognormal distribution. The probability distribution of income bY_n is

$$bY_n \sim \Lambda(\log b + B + Anp,\, np(1-p)A^2).$$

The probability density function of a capital distribution that is derived from a repeated random chance is

$$\frac{1}{\sqrt{2\pi\sigma^2}}\frac{1}{y}\exp\left\{-\frac{(\log y-\mu)^2}{2\sigma^2}\right\} \text{ where}$$

$$\mu = \log y_0 + n\log(1-b) + np\log\left(\frac{1+b}{1-b}\right),\quad \sigma^2 = np(1-p)\left(\log\left(\frac{1+b}{1-b}\right)\right)^2.$$

Proof. See Hamada (2016). □

Proposition 1 shows that stock (capital) and flow (interest of capital, income) are lognormally distributed, respectively, and this implication plays an important role in Bayesian modeling. Figure 4 indicates the relation of random variables that correspond to the derivation of lognormal distribution.

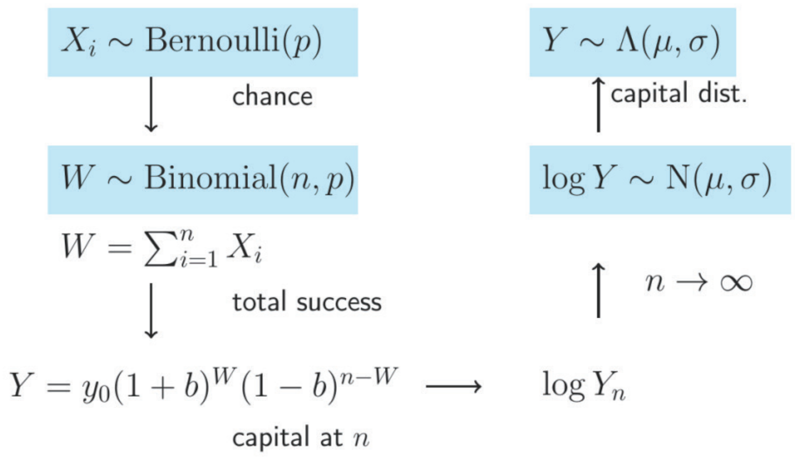

Figure 4: Transition process of lognormal distribution (capital distribution).

So far, the parameters of income distribution have been identified. Before we estimate the posterior distribution of parameters, we need to confirm the property of average capital. Differentiating the mean of capital distribution with respect to p, we obtain the following proposition.

Proposition 2 (distribution mean and success probability (Hamada 2016)). If interest rate b is smaller than $(e^2 - 1)/(e^2 + 1) \approx 0.761594$, then the mean of the capital distribution is an increasing function of the success probability p of random chance.

Proof. See Hamada (2016). □

In general, sociologists who solely rely on generalized linear models are likely to assume that explanatory variables linearly affect outcome variables such as income. However, we cannot know whether a generalized linear model is a true probability model for true distribution since we can never know the true distribution. If we consider the human capital theory as a verbal model or generalized linear model, we can never make a non-linear prediction like Proposition 2. The hypothesis that simply income is an increasing function of p and b may not be true. Intuitively, an interest rate b enhances the impact of success probability on economic growth. However, Proposition 2 claimed that there is a range in which the average of capital (income) distribution becomes a decreasing function of the probability p.

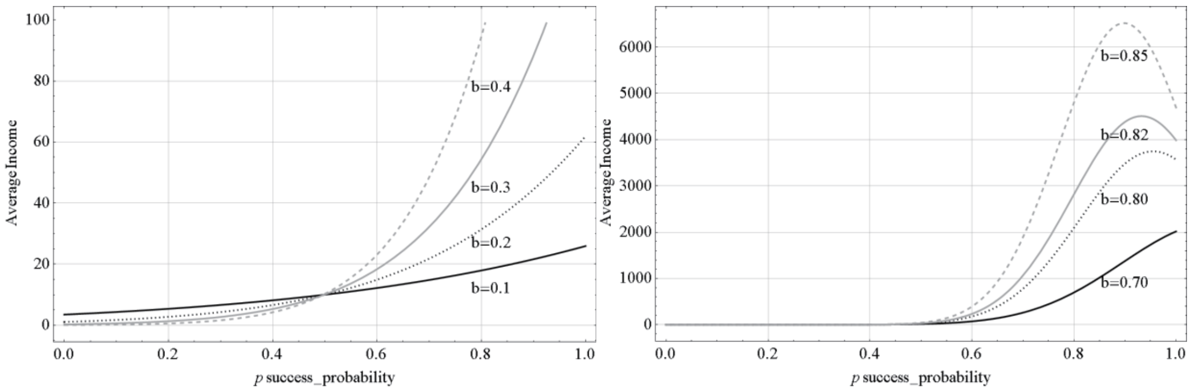

Figure 5: Average of capital distribution with $n = 10, b \in [0.1, 0.85], y_0 = 10$

(reprinted from author's previous work; Hamada 2016).

Figure 5 illustrates the nonlinear relation between success probability p and average of capital distribution under a specific constant b. As the right panel of Figure 5 shows, the average of capital distribution is not a monotone function anymore when $b > (e^2 - 1)/(e^2 + 1)$. This suggests that when we attempt to explain outcome variables theoretically, a generalized linear model is not always the best choice because Proposition **2** implies that the parameter of the outcome variable is not a monotone function of explanatory variables.

Many quantitative researches in sociology assume a generalized linear model without theoretical reasoning because it is easy to estimate the parameters of outcome variables. Certainly, it was very difficult for us to estimate the parameters for theoretical models such as the generative model of income distribution. However, by development of probabilistic programing language such as BUGS, JAGS, or Stan, Bayesian modeling with MCMC method allows us to estimate the parameters for complicated models that cannot be expressed by the generalized linear model. We will show that the stochastic toy model has a strong affinity to the Bayesian statistical analysis in next section.

3. Bayesian statistical analysis

3.1. Statistical model based on the toy model

We construct a statistical model based on the mathematical toy model described in the previous section. To facilitate comparison, we also define model 0 and model 3 as non-theoretical models. Hereafter, for computational convenience, the outcome variable Y is defined as a logarithmic form in the data and models. Therefore, in our statistical model, we assume that the logarithm of capital Y is subject to normal distribution without loss of generality.

$$\begin{aligned}
&Y[i] \sim N(\mu,\sigma) && i=1,2,\ldots,N(\text{individual})\\
&\mu \sim \text{Uniform}(-2,2)\\
&\sigma \sim \text{Uniform}(0,2)
\end{aligned}$$

Next, we define the baseline model (model 1) as follows.

$$\begin{aligned}
&Y[i] \sim N(\mu,\sigma) \quad i=1,2,\ldots N(\text{individual})\\
&\quad \mu = \log y_0 + n\log(1-b) + \log\frac{1+b}{1-b}np\\
&\quad \sigma = \sqrt{npq}\log\frac{1+b}{1-b}\\
&\quad p \sim \text{Beta}(1,1), \qquad b \sim \text{Beta}(1,1)
\end{aligned}$$

Note that $q=1-p$. It is extremely important that model 1 has theoretically defined parameters μ and σ that are given by Proposition 1, the functions of endogenous parameters of our toy model. In this sense, model 1 represents theoretical model. Since the parameters p and b are probabilities, we assume their prior distributions are both subject to beta distribution. Note that the parameters p and b are not observable, and that posterior distributions of p and b are estimated by the MCMC method. In model 1, we assume y_0 and n are constant ($y_0=10,\ n=10$)

Next, we define the hierarchical model (model 2) as follows.

$$\begin{aligned}
Y[i] &\sim N(\mu[Sex[i],Age[i]],\sigma[Sex[i],Age[i]])\\
i &= 1,2,\ldots N(\text{individual})\\
\mu[j,k] &= \log y_0 + n[k]\log(1-b[k,j])\\
&\qquad + \log\frac{1+b[j,k]}{1-b[j,k]}n[k]p[j,k]\\
\sigma[j,k] &= \sqrt{n[k]p[j,k]q[j,k]}\log\frac{1+b[j,k]}{1-b[j,k]}\\
p[j,k] &\sim \text{Beta}(1,1), \qquad b[j,k] \sim \text{Beta}(1,1)\\
n[k] &\sim N(k+19,s), \qquad s \sim \text{Unif}(0,1000)\\
j &= 1,2(\text{sex}), \quad k=1,2,\ldots,K(\text{age})
\end{aligned}$$

Model 2, same as model 1, represents theoretical model since it has the parameters given by Proposition 1. In model 2, Index i stands for individual, j for sex, and k for age respectively. μ and σ are functions of latent parameters p and b. Additionally, p and b, whose prior distributions are beta distributions (in this model, it is equivalent to uniform distribution), are clustered by age and sex. As a result, μ and σ are also clustered by age and sex. Model 2 looks complicated, however it is just a clustered version of model

1.

Intuitively, the hierarchical model 2 can be expressed as the following diagram in Figure 6.

$$Y \sim \Lambda(\mu_{\mathsf{Sex}[i],\mathsf{Age}[i]}, \sigma^2_{\mathsf{Sex}[i],\mathsf{Age}[i]})$$

$$p_{jk}, b_{jk} \sim \mathrm{Beta}(1,1) \quad \mu_{jk} = \log y_0(1-b_{jk})^{n_k}\frac{1+b_{jk}}{1-b_{jk}} n_k p_{jk}$$

$$n_k \sim \mathrm{N}(k,s) \qquad \sigma^2_{jk} = n_k p_{jk} q_{jk}\left(\log\frac{1+b_{jk}}{1-b_{jk}}\right)^2$$

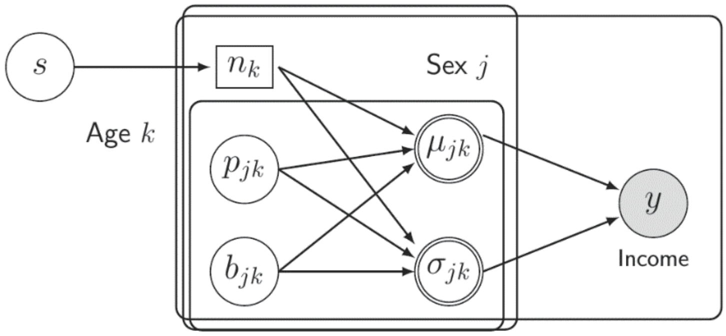

Figure 6: Bayesian model of income distribution.
Index *i* stands for individual and *k* for age.

We assume that μ_{jk} and σ_{jk} have a group-level variance, since individuals in the same age group experience nearly equal times of random chance, and male workers have more advantages in acquiring human capital than female workers empirically.

Finally, we define a linear model (model 3) as follows.

$$\begin{aligned} Y[i] &\sim N(\mu[i], \sigma) \\ \mu[i] &= b_0 + b_1 Male[i] + b_2 Age[i] \\ i &= 1, 2, \ldots N(\text{individual}) \\ \sigma &\sim \mathrm{Uniform}(0, 2) \end{aligned}$$

Model 3 represents a typical simplified linear model.

3.2. Data

We used the following variables from the SSM2015 dataset:

Y: individual income (logarithmic scale)

Sex: male or female (0 for female and 1 for male in model 3)

Age: 20–80.

Sex and Age are used for clustering parameters in model 2.

3.3. Result of MCMC estimation

We estimate posterior distributions of parameters by MCMC method. R (version 3.4.1) and Stan (version) are used for computation. Additionally, Rstan (version 2.16.2) and loo (version 1.1.0) package are used for the implementation of Stan model from R and computation of a WAIC and a leave-one-out cross-validation. The MCMC settings are chains=3, warmup=1000, and sampling=1000.

Table 1: Summary of MCMC samples of parameters.

	mean	2.50%	97.50%	n_eff	$\hat{R}$
			model 0		
mu	5.36	5.339	5.382	3000.000	1.000
sigma	0.899	0.884	0.915	3000.000	1.001
			model 1		
mu	5.36	5.338	5.381	3000.000	0.999
sigma	0.899	0.884	0.914	1369.915	1.001
p	0.919	0.917	0.92	1770.568	1.000
b	0.478	0.474	0.482	1545.447	1.000
			model 3		
b0	5.193	5.144	5.240	1152.432	1.002
b1	0.885	0.847	0.924	938.618	1.002
b2	-0.008	-0.009	-0.006	1591.467	1.000
sigma	0.772	0.758	0.787	1442.755	1.001

Table **1** summarizes the distribution samples of parameters we obtained from MCMC estimation. Since we estimate $61(age)\times 2(sex)\times 2(p,b)=244$ posterior distribution of parameters p and b, we omit information of model 2 from Table 1 and show the graph of mean and standard deviation in **Figure 7**, rather than showing unnecessary long tables.

In Figure 7 and Figure 8, the error bar indicates the standard deviation of posterior distribution of the parameters. All $\hat{R}$ of parameters in model 2 are under 1.035; thus, the MCMC sampling can be seen as converged.

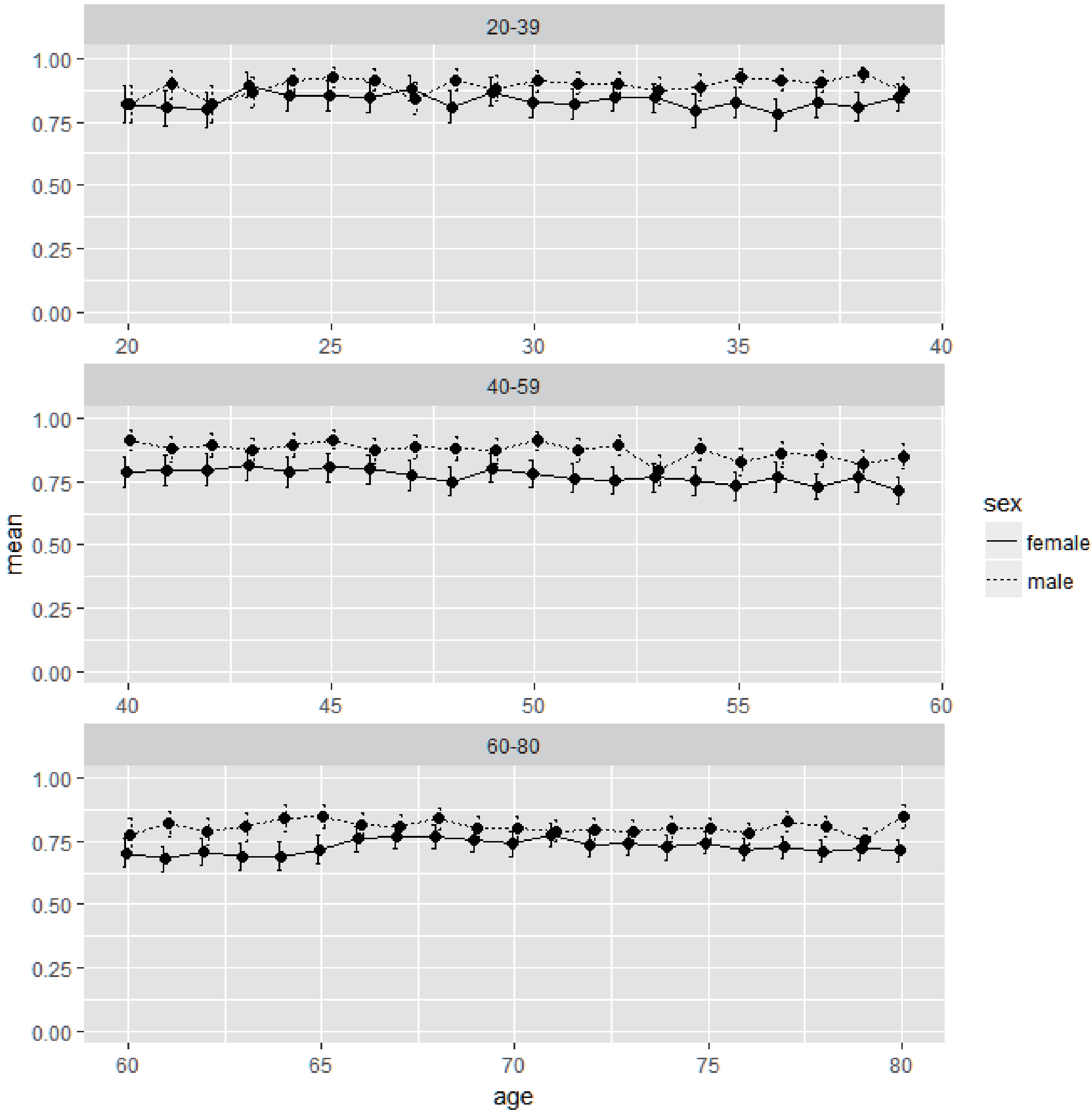

Figure 7: Posterior distribution of *p* (success probability of random chance) computed by MCMC. Age: 20–80, female and male.

Approximately, the success probability p for males is slightly larger than that for females in the area of over 30. The success probability p is almost invariant from age or slightly decreases with age. Meanwhile, the interest rate b is almost the same among both males and females, and decreasing with age in general. With respect to the mean level, the interest rate b decreases from around 0.3 to 0.1 as age increases. Furthermore, the variance of interest rate b is decreasing with age.

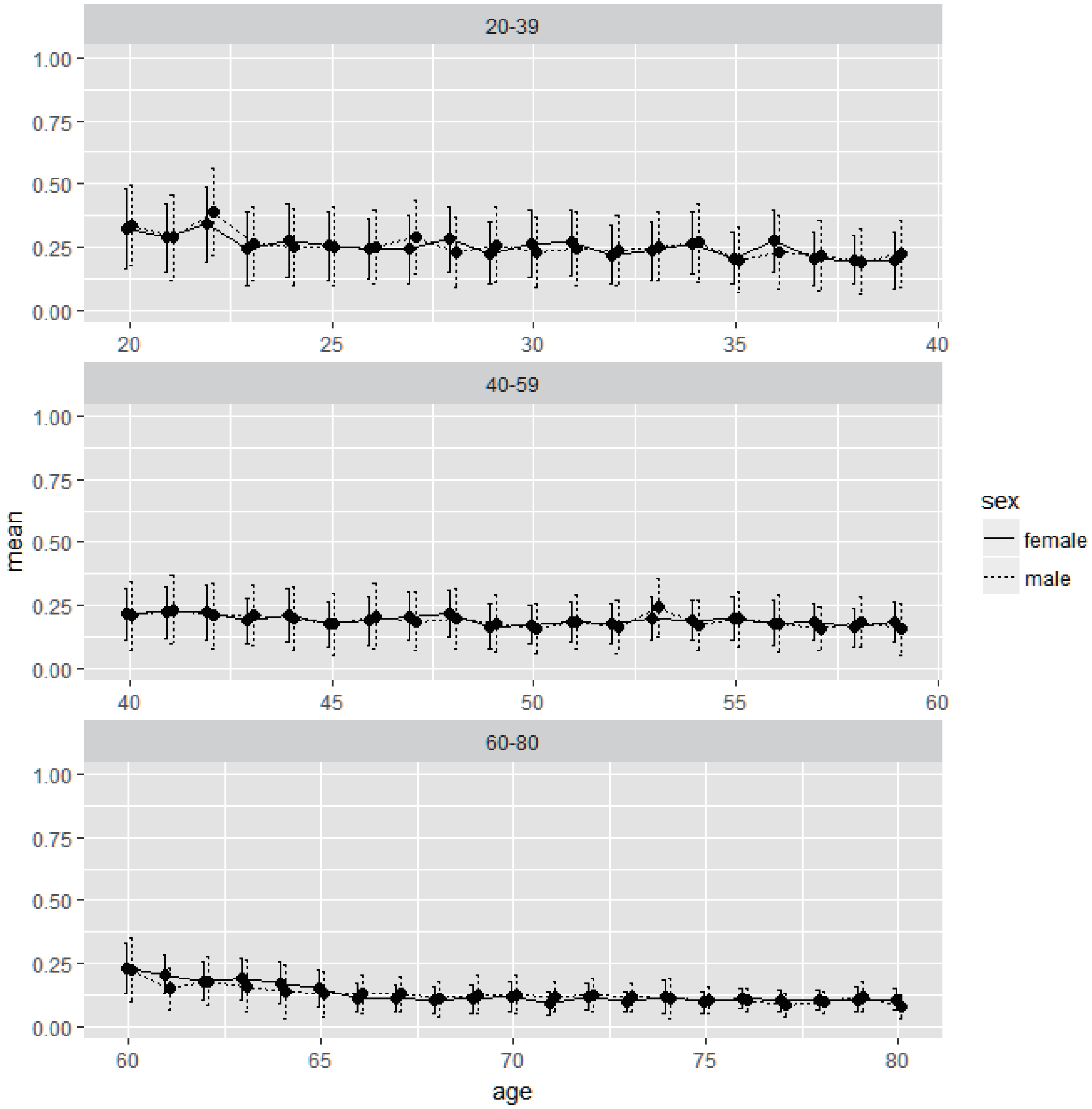

Figure 8: Posterior distribution of *b* (interest rate) computed by MCMC.
Age: 20–80, female and male.

3.4. A comparative analysis of models

We approximately computed the widely applicable information criterion (WAIC; Watanabe 2010) and the leave-one-out cross-validation (LOO) from MCMC samples. WAIC and LOO are methods for estimating pointwise out-of-sample prediction accuracy from a fitted Bayesian model using the log-likelihood evaluated at the posterior simulations of the parameter values (Vehtariy et al 2017).

In **Table 2**, “elpd_waic” and “elpd_loo” are expected log point-wise predictive density, “p_waic” and “p_loo” are estimated effective number of parameters. “waic” is converted to the deviance scale, namely waic= $-2\times$ elpd_waic. Similarly, "looic" is converted to the deviance scale, thus looic= $-2\times$ elpd_loo.

Table 2: Summary of WAIC and leave-one-out cross-validation.

	model_0		model_1		model_2		model_3	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
elpd_waic	-8586.2	68.6	-8586.2	68.6	-7142.9	85.8	-7590	80.3
p_waic	2.4	0.1	2.4	0.1	327.5	19.4	5	0.2
waic	17172.4	137	17172.4	137	14285.8	172	15179.9	160.5
elpd_loo	-8586.2	68.6	-8586.2	68.6	-7154.1	87.1	-7590	80.3
p_loo	2.4	0.1	2.4	0.1	338.7	21.3	5.1	0.2
looic	17172.5	137	17172.4	137	14308.2	174	15179.9	160.5

The estimated effective number of parameters of model 0 and model 1 is both equal to 2.4. An estimated effective number of parameters for WAIC defined as

$$\hat{p}_{waic2}=\sum_{i=1}^{N}V_i[\log f(x_i\mid\theta^{(t)})]$$

which is asymptotically equal to the number of unrestricted parameters (Gelman et al. 2013; Toyoda 2017)[7]. The WAIC of model 2 is smaller than model 0, model 1, and model 3, which implies that the clustered model based on the theory may have better predictive accuracy than other models without theory.

4. Conclusion

In the present paper, we have proposed a general theoretical framework called the Bernoulli core approach. We tested empirical validity of one of sub-models, the generative model of income distribution by constructing Bayesian model. As a result of analysis, we have shown that our model can have better predictive accuracy than black box linear model in terms of WAIC and the leave-one-out cross validation. The mathematical toy model provides not only good predictive accuracy but also interesting implications about latent parameters such as success probability and interest rate.

[7] Readers may wonder why p_waic of model 0 and model 1 are equal as (μ,σ) are the parameters for model 0, while (μ, σ, p, b) are those for model 1. We conjectured that the estimated effective number of parameters are same because μ and σ are deterministic functions of p and b in model 1. The estimated effective number of parameters of model 2 is 327.5 because we used many parameters clustered by age and sex.

The flexibility of Bayesian modeling may facilitate us to integrate mathematical toy models that represent specific sociological and economic theory and statistical empirical analysis.

References

Aitchison, J., and Brown, J. A. C. 1957. *The Lognormal Distribution: with Special Reference to its uses in Economics*, Cambridge University Press.

Becker, G. S. 1964. *Human Capital*, Columbia University Press.

Breen, R. and Goldthorpe, J. H. 1997. "Explaining educational differentials: Towards a formal rational action theory," *Rationality and Society*, 9(3):275-305.

Champernowne, D. G. 1953. "A model of income distribution," *The Economic Journal*, 63 (250): 318-351.

Crow, E. L., and Shimizu, K. 1988. *Lognormal Distributions: Theory and Applications*, Marcel Dekker, Inc.

Fararo, T.J. 1989. *The Meaning of General Theoretical Sociology: Tradition and Formalization*, Cambridge University Press.

Fararo, T. J., and Kosaka, K. 2003. *Generating Images of Stratification: A Formal Theory*, Springer.

Friedman, M. 1953. "Choice, chance, and the personal distribution of income," *Journal of Political Economy*, 61(4): 277-290.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin D. B. 2013. *Bayesian Data Analysis, third edition*. CRC Press.

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

Hamada, H. 2003. "A generative model of income distribution, " *Journal of Mathematical Sociology*, 27 (4): 279-299.

Hamada, H. 2004. "A generative model of income distribution 2: Inequality of the iterated investment game," *Journal of Mathematical Sociology*, 28 (1): 1-24.

Hamada, H. 2005. "Parametric decomposition of the Gini coefficient: How changes of subgroup affect an overall inequality," *Sociological Theory and Methods*, 20 (2): 241-256.

Hamada, H. 2012. "A model of class identification: Generalization of the Fararo-Kosaka model using Lyapounov's central limit theorem, " *Kwansei Gakuin University School of Sociology Journal*, 114: 21-33.

Hamada, H. 2016. "A generative model for income and capital inequality," *Sociological Theory and Methods*, 31 (2): 241-256.

Hamada, H. 2017. A generative model for action by hierarchical Bayes approach. Proceedings paper of the 1st RC33 Regional Conference on Social Science Methodology: Asia.
(http://140.109.171.200/2017/abstract/cfpa021.pdf)

Ishida, A. 2018. Introducing a new model of class identification: a mixed method of mathematical modeling and bayesian statistical modeling. Retrieved from osf.io/preprints/socarxiv/mv8pk (In printing for SSM 2015 report)

Kalecki, M. 1945. "On the Gibrat distribution," *Econometrica*, 13 (2): 161-170.

Kapteyn, J. C. 1903. *Skew Frequency Curves in Biology and Statistics*, Astronomical

Laboratory, Noordhoff.
Matsuura, K. 2016. *Bayesian Statistical Modeling Using Stan and R*. Kyoritsu Shuppan Co.（松浦健太郎. 2016. 『Stan と R でベイズ統計モデリング』共立出版.）
McAlister, D. 1879. “The law of the geometric mean," *Proceedings of the Royal Society of London*, 29: 367-376.
Merton, R. 1957. *Social Theory and Social Structure*,The Free Press.
Mincer, J. 1958. “Investment in human capital and personal income distribution," *Journal of Political Economy*, 66 (4): 281-302.
Mincer, J. 1970. “The distribution of labor incomes: A survey with special reference to the human capital approach," *Journal of Economic Literature*, 8 (1): 1-26.
Mitzenmacher, M. 2003. “A brief history of generative models for power law and lognormal distributions," *Internet Mathematics*, 1 (2): 226-251.
Rutherford, R. S. G. 1955. “Income distributions: A new model," *Econometrica*, 23 (3): 277-294.
Toyoda, H. 2017. *Practical Bayes Modeling*, Asakura Publishing Co. (豊田秀樹(編). 2017. 『実践ベイズモデリング：解析技法と認知モデル』朝倉書店)
Watanabe, S. 2010. “Asymptotic equivalence of Bayes cross validation and widely application information criterion in singular learning theory," *Journal of Machine Learning Research*, 11: 3571-3594.

Appendix

Stan code for estimation of model 2 is the following[8]:

```
data{
    int N;// sample size
    int K;// range of age
    real y[N]; //log of individual income
    int age[N];
    real y0; // initial income
    int sex[N];
    }

parameters {
    real <lower=0, upper=1> p[2,K]; //clustered by sex and age
    real <lower=0, upper=1> b[2,K]; // clustered by sex and age
    real <lower=0> n[K];
    real <lower=0> s;// variance parameter for n
    }

transformed parameters{
    //two kinds of mu and sigma are defined for male and female
    real mu[2,K];
    real sigma[2,K];
    for (i in 1:2){
            for (j in 1:K){
                    mu[i,j] = log(y0)+n[j]*log(1-b[i,j])+
                   log((1+b[i,j])/(1-b[i,j]))*n[j]*p[i,j];
                    sigma[i,j] = sqrt(n[j]*p[i,j]*(1-p[i,j]))*
```

[8] I referred to excellent examples of Stan code from Matsuura’s textbook (Matsuura 2016).

```
                                   log( (1 + b[i,j] )/(1 - b[i,j]));
                                     }# for loop of index j
                      }# for loop of index i
     }#transformed parameters block ends here

model {
for (i in 1:K){
       n[i] ~ normal(19+i,s);
       }
for (i in 1:N){
      y[i] ~ normal(mu[sex[i]+1,age[i]], sigma[sex[i]+1,age[i]]);
      }
       }# model block ends
```