Introducing A New Model of Class Identification: A Mixed Method of Mathematical Modeling and Bayesian Statistical Modeling*

Atsushi Ishida

(Osaka University of Economics)

Abstract

This paper aims to introduce a new analytical framework for class identification by applying a mixed method of simple mathematical modeling and Bayesian statistical modeling. First, I constructed a simple mathematical model which can explain the middle concentration tendency of class identification where the majority of people tend to regard themselves as middle, assuming that the succession of the same Bernoulli m-trials with success probability p determines one's subjective class identification. Second, I estimated parameters of the model from SSM survey data by applying a Bayesian statistical model. The distribution of latent success probability p and number of trials m was estimated by the Markov Chain Monte Carlo method. I also analyzed differences in distributions of p and m among age cohorts and educational levels by hierarchical models.

From the analysis, I found the following point: (1) approximately five trials of fifty-fifty games with around 0.5 success probability describes well the observed class identification distribution in 2015 data; (2) the Japanese postwar period can be divided in two based on people's subjective evaluations—the period of expanding opportunity (1955 to 1975) and the period of high and constant

* The study was supported by JSPS KAKENHI Grant Number JP25000001 and 15K13080. The author thanks the Social Stratification and Social Mobility (SSM) Research Committee of 2015 for the permission to use the SSM 1955-2015 data.

success probability, but less chance of trials (from 1985); and (3) the different games model on educational levels is always better in terms of goodness of predictions than the common game model in each survey period. However, these models were closest in terms of goodness of prediction in 1975 and 1985, possibly indicating that during the “all middle-class society” period, people evaluated their society almost the same as the common game where all players have the same opportunities, but dierent luck.
Keywords: class identification, random walk, Bayesian statistical modeling

1 Introduction

Social psychological aspects of social inequality and social stratification, such as people’s cognitions, attitudes, and emotions in an unequal or stratified society, can be regarded as antecedent conditions of people’s rational choices or actions that, if aggregated, would lead to macro (un)change in the society. Thus, it is crucially important for micro-macro linkage approaches on social inequality to shed light on the social psychological aspects as with other aspects of inequality.

Among social psychological features, this paper focuses on class identification. Class identification is the extent to which people identify themselves as members of a certain social class or stratum, and it has been one of the main subjects in social psychological studies of social stratification. In this paper, I will introduce a new analytical framework for class identification by applying a mixed method of simple mathematical modeling and Bayesian statistical modeling, rather than the conventional frequentist statistical analysis that typically applies regression analysis blindly.

First, I construct a simple mathematical model that explains one of the major tendencies of class identification, that is the middle concentration tendency in which the majority of people tend to regard themselves as middle. This is the life-is-like-a-random-walk model where it is assumed that succession of the same Bernoulli m-trials with success probability p determines one’s subjective class identification.

Second, I will estimate parameters of the model from empirical data by applying a Bayesian statistical model. The Bayesian modeling enables us to construct more flexible model which directly reflects the mathematical model and is able to explain generative mechanism of observed distribution. The distribution of latent success probability p and number of trial m are estimated by MCMC estimation and differences in distributions of p and m among different social categories are analyzed by hierarchical models.

The new analytical framework is based on an intermediate modeling strategy. In general, formal modeling of a social phenomenon has three features:

(1) a more realistic mechanism description of the phenomenon;
(2) more constraints on parameters; and
(3) lower fitness of empirical data.

On the contrary, conventional statistical modeling, which tries to summarize empirical data structures as typically linear equations, has opposite features:

(1) less or no realistic mechanism description of the phenomenon;
(2) less or no constraints on parameters; and
(3) higher fitness of empirical data.

The intermediate modeling strategy, composed of simple mathematical modeling and Bayesian statistical modeling, aims to have not the highest but the most reasonable qualities of all features:

(1) a reasonable realistic mechanism description of the phenomenon;
(2) reasonable constraints on parameters; and
(3) reasonable fitness of empirical data.

A series of Japanese cross-sectional survey (SSM survey) data from 1955 to 2015, collected every 10 years, will be analyzed using the Bayesian model. The results of this analysis will be interpreted and discussed. Section 2 will discuss preceding studies on class identification; section 3 will introduce the life-is-like-a-random-walk model; section 4 will provide a methodological description of Bayesian modeling; section 5 will offer results of analyses of SSM survey data; and finally, section 5 will provide a conclusion.

2 Class Identification

Class identification (or self-location in the class system or status identification) is a kind of subjective evaluation of one's own social status in society. More specifically, class identification is the extent to which people identifies themselves as members of a certain social class or stratum.

In a series of studies across many countries, a tendency, referred to as the "middle concentration tendency," in which the majority of people tend to regard themselves as middle regardless of their actual objective status, was common. For example, Kelley and Evans (1995) and Evans and Kelley (2004) found, using cross-national survey datasets, that in many countries people tend to select the middle category for subjective status location. These researchers argued that this tendency existed because of a comparison mechanism in homogeneous reference groups, that is, an assumption that people tend to assess their social status by comparing themselves with others whose status are similar with them. Even in Japan, the middle concentration tendency has consistently been observed since the 1970s when Japan enjoyed affluent economics as a result of the postwar economic growth (Kikkawa, 2016).

The formal model by Fararo and Kosaka (2003) further pursued the idea of a mechanism of class identification formation via comparison with others. In their model, it is assumed that actors evaluate their class position by a sequential comparison process with others in order of importance of several social economic dimensions. Under the model assumption, the distribution of class identification approximates to a normal distribution under the condition that all dimensions are independent of each other (Ishida, 2012; Hamada, 2012).

Following these preceding studies, in this paper, I introduce an alternative and simpler mathematical model explaining the middle concentration tendency.

3 Life-Is-like-a-Random-Walk Model

An ancient Chinese saying states "good luck and bad luck alternate like the strands of a rope." This saying tells us that good events and bad events in life are like two sides of a coin flipped by Fate. Many other sayings share the same message relevant to fortune and life, such as "everything for men is like a Saioh's horse." In line with these wisdoms, I insist here "life is like a random walk."

The random walk, as well known, is a simple stochastic process in which actors flip a coin to decide to move left with a probability p or right with $1 - p$ in every step in m times sequential steps. As long as the flipping-a-coin trials are independent from one another, distribution at a sufficiently large m-th step approximates a normal distribution. Figure 1 is an example of the random walk process in which the left graph shows trajectories of 100 different random walks with the probability 0.5 and the right graph shows the histogram of their positions at 500th step.

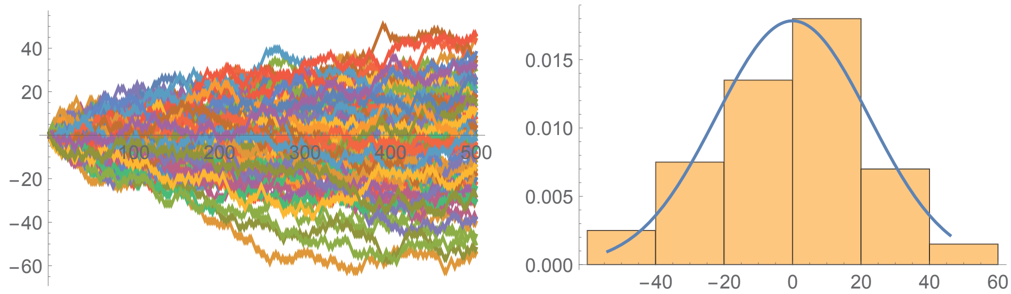

Fig. 1 An example of the random walk process ($p = 0.5, m = 500$)

Based on the random walk process, I construct the life-is-like-a-random-walk model (LLRW model) for class identification distribution. Here are two assumptions:

Assumption 1 Life is like a series of Bernoulli trials that result in success with a probability p or failure with $1 - p$.

Assumption 2 People evaluate their relative status in society by the identical function of their success rate (z) of Bernoulli trials at the end of m-th trial.

Based on the assumptions, number of success denoted by s through m trials comes from the binomial distribution parameterized by m and p, that is,

$$s \sim \text{Binomial}(m, p). \tag{1}$$

And then, according to the central limit theorem, the distribution of success rate ($z = s/m$) will approximate a normal distribution parameterized by p as mean and $\sqrt{p(1-p)/m}$ as standard deviation, as m gets sufficiently larger, that is,

$$z \approx \text{Normal}(p, \sqrt{p(1-p)/m}). \tag{2}$$

Figure 2 shows a simple outline illustration of the life-is-like-a-random-walk model.

From the model, a simple explanation for the middle concentration tendency of class identification is derived: if the lives of most people are like a series of fifty-fifty games and they evaluate their status based on their wins, then distribution of class

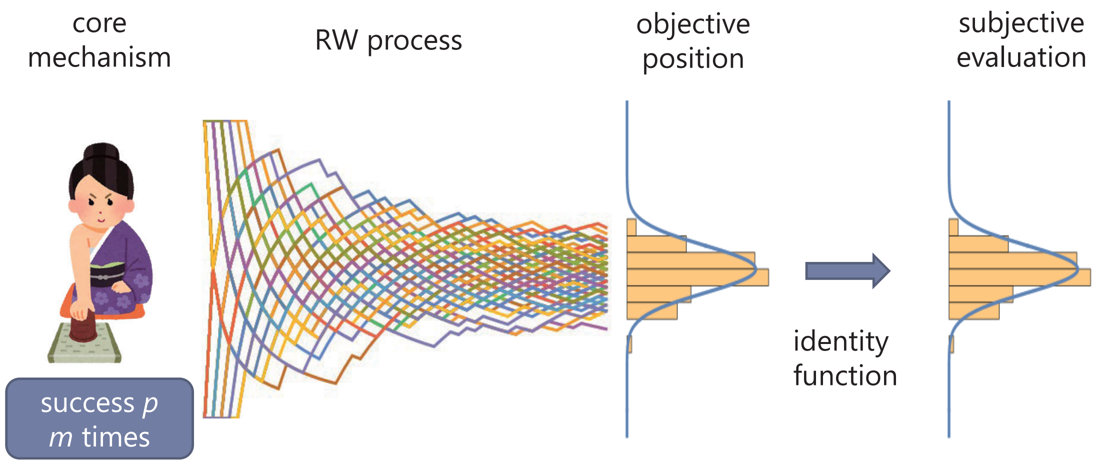

Fig. 2 Outline of the life-is-like-a-random-walk model

identity eventually shows middle concentration. In addition, two implications with respect to parameters can be derived immediately from the model.

Implication 1 The higher success probability (p) in society, the higher the social average subjective status.

Implication 2 The larger the number of trials (m) in society, the smaller the variance of subjective status.

It isnoteworthy that the LLRW model has the same mathematical structure as the specific version of Fararo-Kosaka model of class identification with the assumption of two ranks and identical distributions (Fararo and Kosaka, 2003). Hence, it can be said that the LLRW model and the Fararo-Kosaka model have different allegorical mechanisms, but share the same fundamental mechanism.

4 Parameter Estimation by the Bayesian Statistical Model

I will apply the Bayesian statistical modeling approach for parameter estimation of the life-is-like-a-random-walk model where there are two parameters to be estimated: latent success probability p and latent number of trials m.

In contrast to the conventional frequentists approach, the Bayesian modeling approach enables us to construct a flexible model by means of a hierarchical model and to express the generative process of observed distribution. These features help us conduct direct estimations of the mathematical model and interpret an outcome of quantitative analysis based on the concrete theoretical framework.

4.1 Data and Variables

Data to be analyzed comes from the National Survey of Social Stratification and Social Mobility (SSM) surveys, which are national level random sampling cross-sectional surveys conducted every 10 years from 1955 to 2015. Accordingly, we have 7 waves of data[*1]. SSM surveys focus mainly on the current situation and changes in social stratification structure as well as inter- and intra-generational class mobility in Japan. Moreover, SSM surveys also focus on social psychological aspects of social stratification, such as people's cognition and attitude towards stratification, inequality, unfairness, and related policies. By using this longitudinal data, I examine changes on latent mechanisms of generating class identification distribution in postwar Japan. It should be noted that since female samples are only available from the 1985 survey, in this study, I decide to analyze only male samples aged 20 to 69 for examination of longitudinal trends of class identification[*2].

The outcome variable is relative class identification score z ranging from 1 as the highest to 0 as the lowest. The variable is recoded from 5-scaled class identification. In the SSM survey, class identification has almost invariably been asked using the following question[*3]: "suppose we were to divide the people living in Japanese society today into the following five strata (Upper, Upper Middle, Lower Middle, Upper Lower, Lower Lower), to which group do you think you would belong?" I assign $4/4 = 1$ to "Upper," 3/4 to "Upper Middle," 2/4 to "Lower Middle," 1/4 to "Upper Lower," $0/4 = 0$ to "Lower Lower," respectively[*4].

[*1] The published data and detailed information on surveys from 1955 to 2005 can be found in the SSJ Data Archive (`http://csrda.iss.u-tokyo.ac.jp`). The short summary of 2015 survey can be found in the following URL (`http://www.l.u-tokyo.ac.jp/2015SSM-PJ/2015ssmjisshigaiyo.pdf`) [in Japanese].

[*2] Analyses of female samples are tasks to be completed in the future.

[*3] SSM surveys have been simultaneously conducted by two modes of administration: interviewer-administered and self-administered. The question of class identification has been investigated by interviewer-administered questionnaire in every wave except in 2005.

[*4] As for the 2005 data, I employ a 10-scaled class identification variable (from 1 as the highest to 10 as the lowest) investigated by interviewer-administered questionnaire instead of 5-scaled variable to avoid bias by mode of administration. According to suggestions by Kobayashi (2008), I recoded the 10-scaled variable to the 5-scaled variable in the following manner: 1 and 2 to "Upper," 3 and 4 to "Upper Middle," 5 and 6 to "Lower Middle," 7 and 8 to "Upper

I try to examine how a difference in social category results in a difference in class identification in the different games model, which I will describe in detail in the next section. In this study, I focus on age cohorts and educational levels as factors that would differentiate social economic opportunities among people in Japan (Kikkawa, 2016). I categorize five age groups according to respondents' age in each wave: 20–29, 30–39, 40–49, 50–59, and 60–69. Then, 11 age cohorts are obtained from age groups in each wave according to year of birth: 1886–1895, 1896–1905, 1906–1915, 1916–1925, 1926–1935, 1946–1955, 1956–1965, 1966–1975, 1976–1985, and 1986–1995. Three educational levels are categorized by years of schooling: lower-secondary (under 9 years), upper-secondary (11–13 years) and tertiary (over 14 years).

4.2 Comparison of Two Models

In the following analysis, I test two models, the common game model and the different games model. The common game model assumes that every player of the game has the same opportunities in terms of success probability p and number of trials m; thus, the dispersion of success rate and identified class is a matter of probabilistic luck. On the other hand, the different games model assumes that players in different categories play different games with different opportunities determined by p_j and m_j; therefore, the dispersion of success rate depends on different opportunities as well as probabilistic luck.

Figure 3 shows a simple outline illustration of the two models.

The preference of the models in terms of model prediction of new data will be evaluated by the WAIC (the Watanabe-Akaike or widely applicable information criterion) in every wave (Watanabe, 2010; Vehtari and Gelman, 2014).

5 Results of Analyses

5.1 Common Game Model

First, I examine the common game model using 2015 data with the assumption that all members in society have experienced the same kind of trials with common p and

Lower," 9 and 10 to "Lower Lower," and then assigned relative class identification scores in the same way as the other waves.

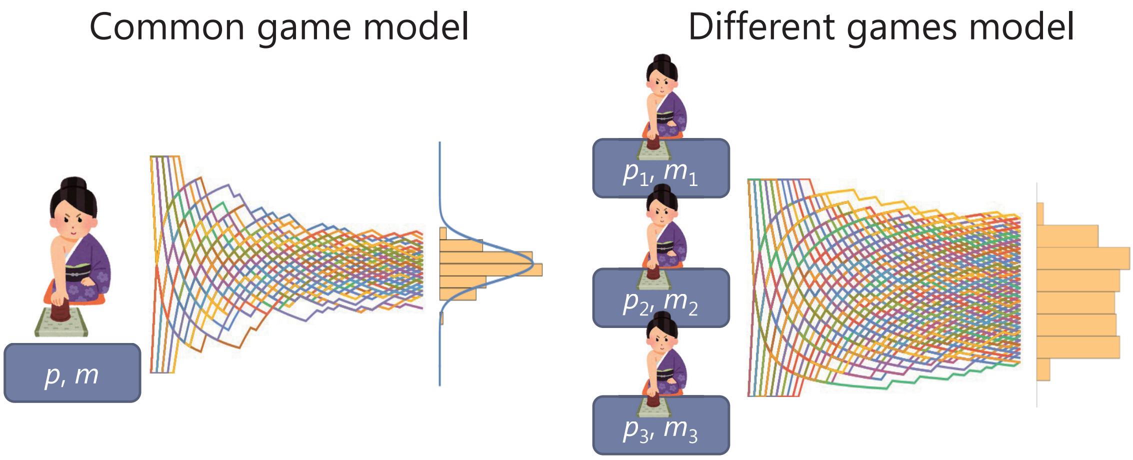

Fig. 3　Outline of the two models

m. From the life-is-like-a-random-walk model, it is assumed that observed relative status z_i is approximately normally distributed with parameters p and m.

$$z_i \sim \mathrm{Normal}(p, \sqrt{p(1-p)/m}) \tag{3}$$

Here, p is assumed to be a logistic function of beta for future model extension.

$$p = \frac{1}{1+\exp(-\beta)} \tag{4}$$

It is assumed that prior distribution of β is a normal distribution with a relatively larger standard deviation, that is,

$$\beta \sim \mathrm{Normal}(0, 10^2). \tag{5}$$

It is also assumed that m is a continuous variable for simplification of estimation and prior distribution of m is a a truncated normal distribution, such that

$$m \sim \mathrm{Normal}_+(0, 10^2). \tag{6}$$

Figure 4 is the graphical model of the common game model, where grey circle nodes indicate observed continuous variables (z_i in this model), double circle nodes indicate generative continuous variables (parameter p), and single circle nodes indicate latent continuous variables with prior distribution (parameters β and m). Figure 5 is the outline illustration of the model.

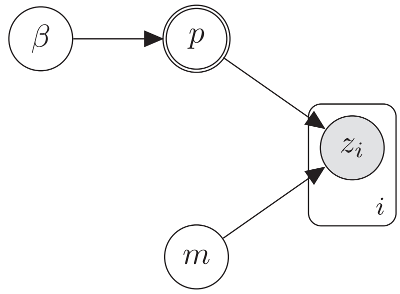

Fig. 4 Graphical model of the common game model

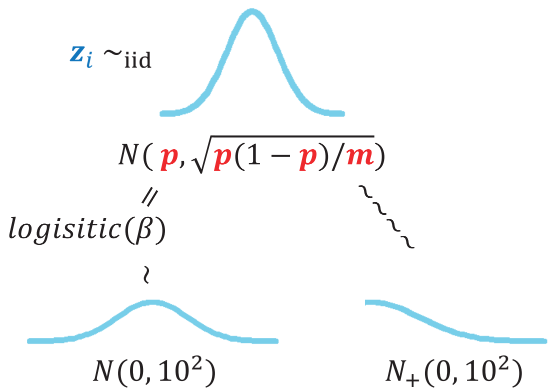

Fig. 5 Outline of the common game model

Posterior distributions for the parameters were estimated by the Markov Chain Monte Carlo (MCMC) method. In this study, I employed Stan 2.16.0 and RStan 2.16.2 (Stan Development Team, 2017a,b) for the MCMC programming[*5]. I conducted four chains of sampling for 5,000 iterations each, which includes 1,000 initial iterations as burn-in samples. The thin interval was set as one, to generate 16,000 sampled points of posterior distributions. The sample size (n) used for this analysis was 2895.

Table 1 shows the summary of the MCMC estimation of posterior distributions for the common game model for SSM 2015 data. The Gelman-Rubin MCMC convergence statistic ($\hat{R}$) of each parameter is around 1.000; hence, we can safely conclude that the MCMC sampling converged (Gelman et al., 2013, 284–286).

[*5] I referred Kruschke (2015) and Matsuura (2016) for building Stan codes for the models. Stan codes are available in Appendix 2.

Table 1 Result of the MCMC estimation of the common game model in 2015 data ($n = 2895$)

	mean	SE	SD	2.50%	25%	50%	75%	97.50%	$\hat{R}$
β	0.033	0.000	0.016	0.001	0.023	0.034	0.044	0.065	1.000
p	0.508	0.000	0.004	0.500	0.506	0.508	0.511	0.516	1.000
m	5.417	0.001	0.141	5.146	5.319	5.416	5.510	5.694	1.000

SE: standard error of posterior mean, SD: standard deviance of posterior distribution, x%: x-percentile of posterior distribution, $\hat{R}$: Gelman-Rubin MCMC convergence statistic.

The success probability of a trial p is estimated around 0.51, meaning that the game is almost a fifty-fifty game. The number of trials m is estimated to be five times, approximately. Therefore, if everyone played a fifty-fifty game approximately five times, then we obtain nearly the same distribution as observed relative status distribution. Figure 6 shows the histogram of observed relative status overlapped by predicted normal distribution determined by the medians of posterior distributions of p and m.

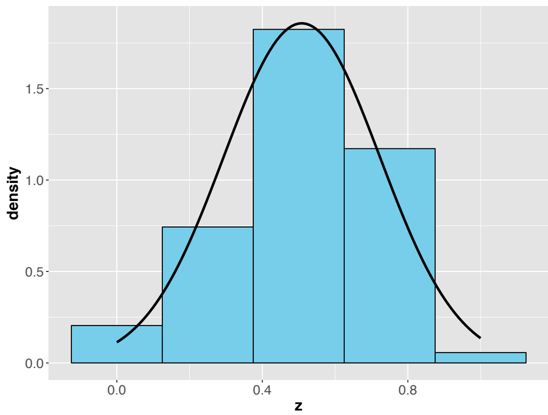

Fig. 6 Histogram of observed relative status and predicted distribution in 2015 data

5.2 Different Games Model on Age Cohorts

Next, I examine the different games model with the assumption that each category in the society has experienced different games with different opportunities. I apply the framework of the Bayesian hierarchical modeling to construct the different games model.

First, I examine the different games model on age cohorts in 2015 data. In this model, it is assumed that the parameters p and m differ by age cohort. A person i's relative status z_i is assumed to be normally distributed, and the distribution to be parameterized by $p_{j(i)}$ and $m_{j(i)}$, which indicates success probability and number of trials commonly held in the cohort j to which person i belongs, that is,

$$z_i \sim \text{Normal}(p_{j(i)}, \sqrt{p_{j(i)}(1 - p_{j(i)})/m_{j(i)}}). \tag{7}$$

p_j is assumed to be a logistic function of a liner equation composed of β_0 as the intercept and β_j as the slope of each category. β_j is assumed to be under the zero-sum constraint and has common normal distribution around zero with σ_β as the standard deviation. β_0 and σ_β as hyper-parameters have a normal distribution and a uniform distribution from 0 to ∞ as prior distributions, respectively. These assumptions can be described in the following notations,

$$p_j = \frac{1}{1 + \exp(-\beta_0 - \beta_j)}, \tag{8}$$

$$\beta_0 \sim \text{Normal}(0, 10^2), \tag{9}$$

$$\beta_j \sim \text{Normal}(0, \sigma_\beta), \sum \beta_j = 0, \tag{10}$$

$$\sigma_\beta \sim \text{Uniform}(0, \infty). \tag{11}$$

m_j is assumed to have a common truncated normal distribution at zero with hyper-parameters μ_m and σ_m, that is,

$$m_j \sim \text{Normal}_+(\mu_m, \sigma_m), \tag{12}$$

$$\mu_m \sim \text{Uniform}(0, \infty), \tag{13}$$

$$\sigma_m \sim \text{Uniform}(0, \infty). \tag{14}$$

Figures 7 and 8 are the graphical model and the outline illustration of the different games model, respectively. The different games model can be regarded as a Bayesian version of analysis of variance (ANOVA) model (Kruschke, 2015, Ch.19).

I would like to discuss the interpretation of difference of parameters. What are the actual meanings of parameters p and m? How can we interpret the outcome of empirical data analysis? p can be regarded as a subjective evaluated advantage in

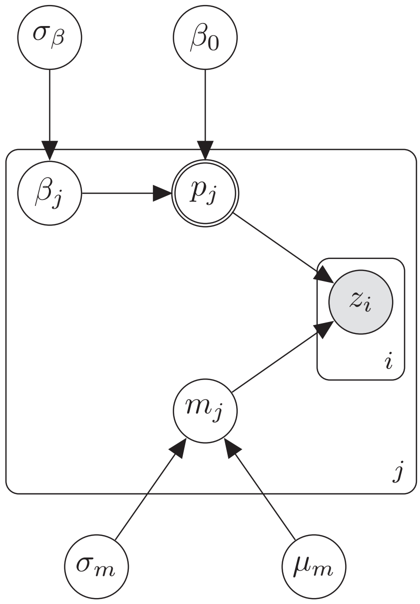

Fig. 7　Graphical model of the different games model

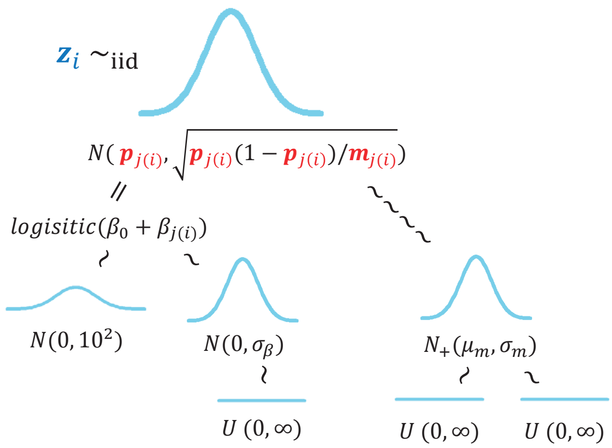

Fig. 8　Outline of the different games model

a life event, such as enrollment in a university or getting a job or promotion shared in a social category. So, a category that has a higher value of p can be assumed to be a category that achieves higher advantage. Meanwhile, m can be regarded as a subjective evaluated number of important life events that members of a social category have experienced. If you played a few games, you could all win from beginner's luck or could all lose. But, if you played many games, a winning rate becomes more reflective of their actual advantages, so that, the variance within a category becomes smaller.

However, the value of m should be regarded not as an absolute but rather as a relative value, because it partly depends on measuring scales.

Let us examine the result of the analysis. Table 2 shows the summary of the MCMC estimation of posterior distributions for the different games model on age cohorts in SSM 2015 data. The MCMC setting for parameter estimation is the same as that in the common game model. MCMC sampling for all parameters can be judged as being converged from the values of $\hat{R}$. The index of β_j, p_j, and m_j indicates age cohorts: 1 1986-95 (20s in 2015), 2 1976-85 (30s), 3 1966-1975 (40s), 1956-1965 (50s), and 1946-1955 (60s).

Table 2 Result of the MCMC estimation of the different games model on age cohorts in 2015 data ($n = 2895$)

	mean	se_mean	sd	2.50%	25%	50%	75%	97.50%	$\hat{R}$
β_0	0.028	0.000	0.017	−0.005	0.017	0.028	0.040	0.062	1.001
β_1	−0.030	0.000	0.038	−0.106	−0.055	−0.030	−0.004	0.039	1.003
β_2	−0.099	0.000	0.033	−0.164	−0.120	−0.099	−0.076	−0.033	1.000
β_3	0.029	0.000	0.030	−0.029	0.009	0.029	0.049	0.089	1.000
β_4	0.096	0.000	0.032	0.032	0.074	0.096	0.118	0.159	1.000
β_5	0.004	0.000	0.027	−0.046	−0.015	0.004	0.022	0.056	1.003
σ_β	0.121	0.002	0.084	0.038	0.071	0.099	0.143	0.339	1.003
μ_m	5.427	0.006	0.317	4.784	5.288	5.438	5.574	6.028	1.001
σ_m	0.433	0.019	0.454	0.024	0.162	0.317	0.552	1.543	1.005
p_1	0.500	0.000	0.011	0.477	0.492	0.500	0.507	0.521	1.003
p_2	0.482	0.000	0.009	0.464	0.476	0.483	0.489	0.501	1.000
p_3	0.514	0.000	0.008	0.498	0.509	0.514	0.520	0.531	1.000
p_4	0.531	0.000	0.009	0.513	0.525	0.531	0.537	0.548	1.001
p_5	0.508	0.000	0.007	0.495	0.503	0.508	0.513	0.522	1.001
m_1	5.143	0.016	0.349	4.369	4.921	5.187	5.403	5.696	1.007
m_2	5.456	0.003	0.250	4.964	5.293	5.457	5.606	5.968	1.001
m_3	5.611	0.004	0.263	5.155	5.432	5.583	5.769	6.194	1.001
m_4	5.579	0.003	0.262	5.115	5.402	5.557	5.736	6.158	1.000
m_5	5.379	0.004	0.213	4.947	5.241	5.383	5.524	5.789	1.004

The index of β_j, p_j, m_j indicates age cohorts: 1 1986-95, 2 1976-85, 3 1966-1975, 4 1956-1965, 5 1946-1955.

Figures 9 and 10 show the difference of parameters p_j and m_j depending on age cohorts where each dot indicates the median and the error bar indicates 95% high density interval of posterior distribution, respectively. Figure 11 is the path plot of the medians of p_j and m_j by age cohorts. Roughly saying, opportunities of games in terms of both the success probability and number of trials increases with age up to respondents in their 50s, except those in their 30s, who, in terms of success probability, suffered the negative effects of severe recession when entering graduate job markets. Figure 12 shows the histograms of observed relative status overlapped by predicted normal distributions according to age cohorts.

5.3 Different Games Model on Educational Levels

Next, I examine the different games model on educational levels. Table 3 shows the summary of the MCMC estimation of posterior distributions for the different games model on educational levels in SSM 2015 data. The MCMC setting for parameter estimation is the same as that in the common game model. MCMC sampling for all parameters can be judged as being converged from the values of $\hat{R}$. The index of β_j, p_j, and m_j indicates educational levels: 1 lower secondary, 2 upper secondary, and 3 tertiary.

Table 3 Result of the MCMC estimation of the different games model on educational levels in 2015 data ($n = 2895$)

	mean	se_mean	sd	2.50%	25%	50%	75%	97.50%	$\hat{R}$
β_0	−0.048	0.000	0.023	−0.093	−0.063	−0.048	−0.032	−0.002	1.000
β_1	−0.280	0.000	0.042	−0.364	−0.307	−0.280	−0.252	−0.196	1.000
β_2	−0.054	0.000	0.026	−0.105	−0.071	−0.054	−0.037	−0.001	1.000
β_3	0.334	0.000	0.027	0.282	0.316	0.334	0.351	0.385	1.000
σ_β	1.345	0.104	3.525	0.192	0.370	0.609	1.146	6.853	1.003
μ_m	6.449	0.396	11.553	2.175	4.983	5.559	6.159	13.160	1.007
σ_m	4.329	0.690	23.783	0.313	0.848	1.509	3.076	20.581	1.006
p_1	0.419	0.000	0.015	0.389	0.409	0.419	0.429	0.449	1.000
p_2	0.475	0.000	0.005	0.464	0.471	0.475	0.478	0.485	1.000
p_3	0.571	0.000	0.006	0.560	0.567	0.571	0.575	0.582	1.000
m_1	4.687	0.006	0.451	3.833	4.377	4.673	4.988	5.611	1.001
m_2	5.562	0.002	0.200	5.184	5.427	5.557	5.694	5.960	1.000
m_3	6.179	0.003	0.260	5.685	6.002	6.172	6.352	6.700	1.000

The index of β_j, p_j, m_j indicates educational levels: 1 lower secondary, 2 upper secondary, and 3 tertiary.

Figures 13 and 14 show the difference of parameters p_j and m_j depending on educational levels where each dot indicates the median and the error bar indicates 95% high density interval of posterior distribution, respectively. Figure 15 is the path

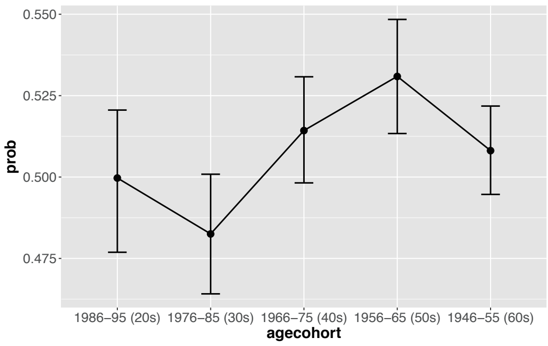

Fig. 9 p_j by age cohorts (median and 95% HDI)

Fig. 10 m_j by age cohorts (median and 95% HDI)

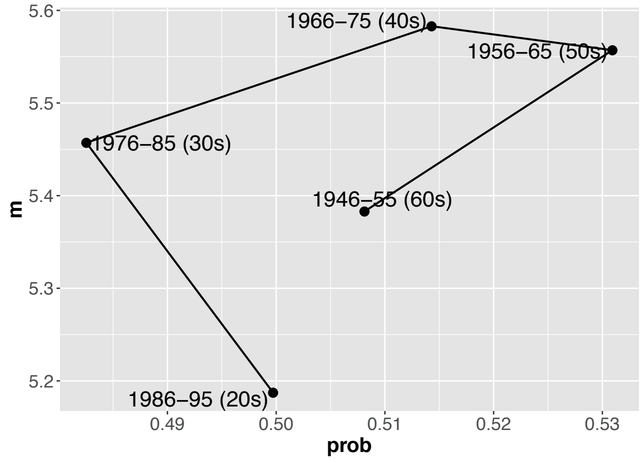

Fig. 11 Path plot of medians of p_j and m_j by age cohorts

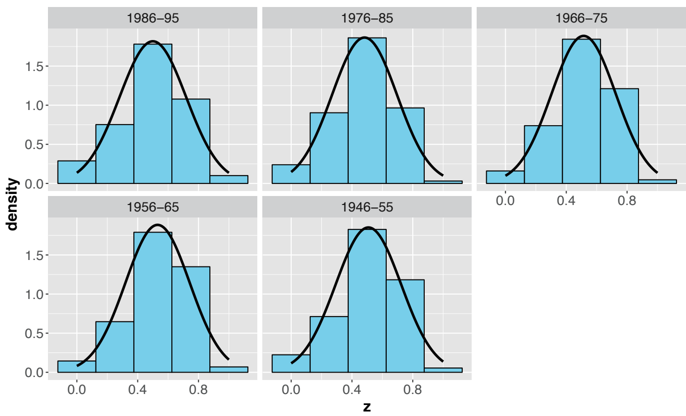

Fig. 12 Histograms of observed relative status and predicted distributions by age cohorts.

plot of medians of p_j and m_j by educational levels. We can see from these figures that both the success probability p_j and number of trials m_j increase as educational level gets higher, indicating that higher educational levels provide better opportunities in terms of both success probability and number of trials in a random walk game. Figure 16 shows the histograms of observed relative status overlapped by predicted normal distributions according to educational levels.

5.4 Trends of the Parameters of the Common Game Model and the Different Games Model on Educational Levels

Finally, I examine the trends of the model parameters in both the common game model and the different games model on educational levels in postwar Japan from 1955 to 2015[*6]. The trends of the parameters of the different games model on age cohorts will be shown in Appendix 1.

Figures 17 and 18 are the trends of parameters of the common game model, and Figure 19 is the path plot of medians of p and m by survey years. As for success probability p, it increased sharply from 0.33 in 1955 to around 0.5 in 1975 following the high economic growth in Japan, and then it became almost stable around 0.5 from 1975 to 2015, except for a slight decrease in 2005. As for number of trials m, it increased from 4.7 in 1955 to 6.5 in 1975 in the same way with success probability, then it decreased by about 1 point in 1985, thereafter it became almost stable. We can see clearly from the path plot (Figure 19) that the postwar period can be divided into two periods in terms of people's subjective evaluation: the period of expanding opportunity in terms of both the success probability and possible chance of trials from 1955 to 1975 and the period of high and constant opportunity, but less chance of trials from 1985.

Figures 20 and 21 are the trends of parameters of the different game model on educational levels, and Figure 22 is the path plot of the medians of p_j and m_j by educational levels by survey years. Both success probability p_j and number of trials m_j increase as educational level increases in every survey period. As for p_j, the difference of success probabilities among educational levels became smaller from 1955 to 1975,

[*6] The sample size of the data in each survey wave was 1982 in 1955 survey, 1989 in 1965, 2665 in 1975, 2391 in 1985, 2372 in 1995, 2567 in 2005, and 2895 in 2015.

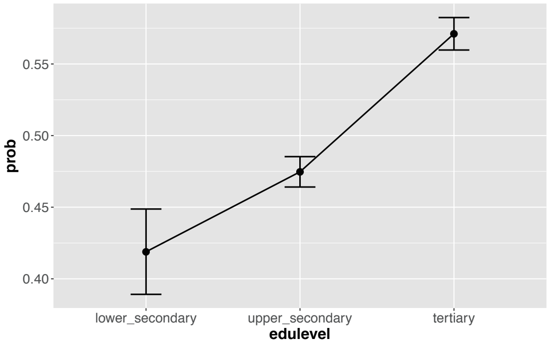

Fig. 13 p_j by educational levels (median and 95% HDI)

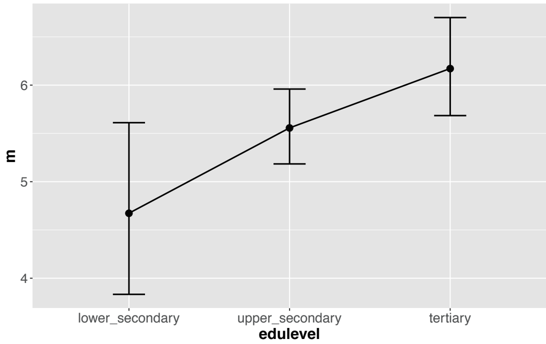

Fig. 14 m_j by educational levels (median and 95% HDI)

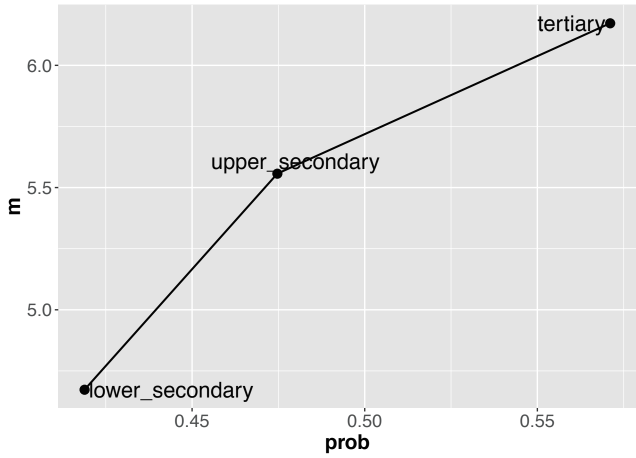

Fig. 15 Path plot of medians of p_j and m_j by educational levels

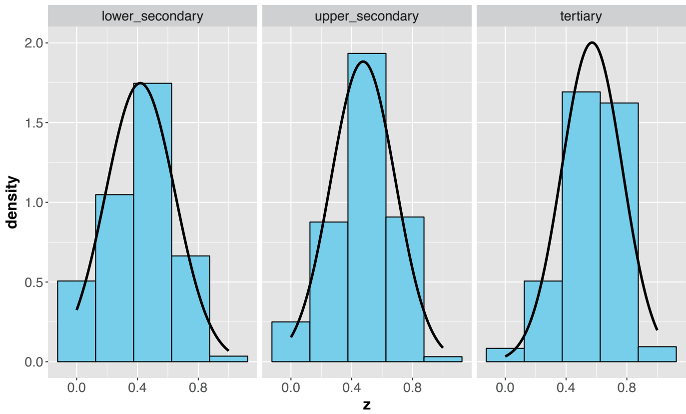

Fig. 16 Histograms of observed relative status and predicted distributions by educational levels

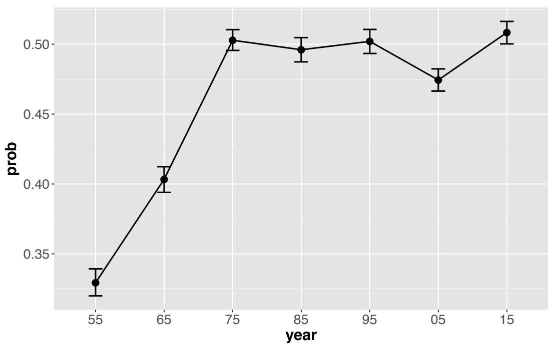

Fig. 17 p by survey years (median and 95% HDI)

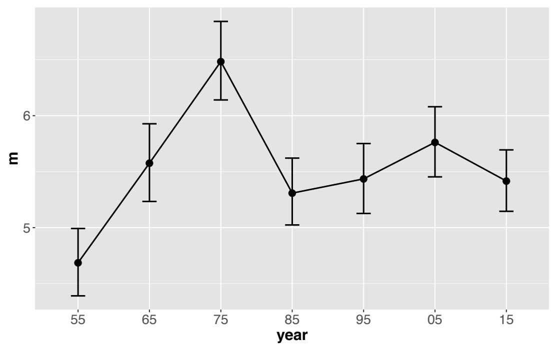

Fig. 18 m by survey years (median and 95% HDI)

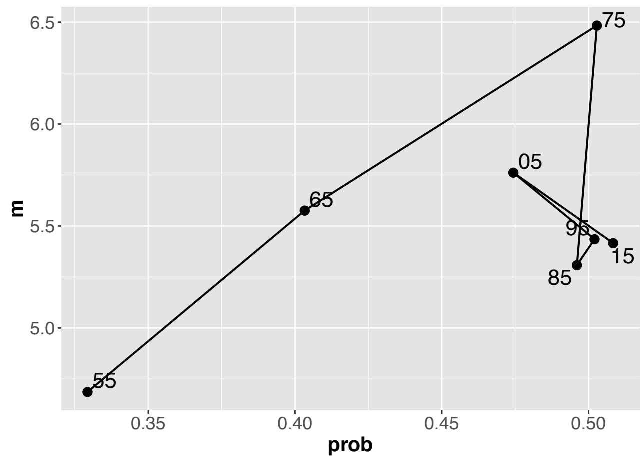

Fig. 19 Path plot of medians of p and m by survey years

and then become larger after 1975. The shapes of trajectories for each educational level in the path plot looks mostly homogeneous, but the location moves to the right-top corner as educational level increases, indicating that higher educational levels provide better opportunities in terms of both the success probability and number of trials in a random walk game in every survey period.

Figure 23 shows the values of WAIC for model comparison[*7]. In terms of WAIC, the different games model is adopted as a better model than the common game model in every survey period. Figure 24 shows differences of WAIC between the two models[*8].

[*7] I employed ‘loo’ R package (Vehtari et al., 2016) for calculation of WAIC.

[*8] More precisely, Figure 24 shows differences of estimated expected log pointwise predictive density for a new data (eldp), which is related to WAIC as WAIC $= -2$elpd. See Vehtari and Gelman (2014) for more information.

We can see from Figure 24 that with respect to educational level, the common game model and the different games model were closest in 1975 and 1985. In those days, it was commonly discussed in mass media as well as academia that Japan had become an "all middle-class society" in which almost of all the people regarded themselves as "middle." From the outcome of our analysis, it can be claimed that in the time of the "all middle-class society," people evaluated their society almost the same as the common game in which all players have the same opportunities, but different luck.

6 Conclusion

Thus far, I have introduced a new analytical model for class identification by applying a mixed method of simple mathematical modeling and Bayesian statistical modeling. In the first part, I introduced the life-is-like-a-random-walk model for class identification distribution, which may be the simplest model for explaining middle concentration. In the second part, I attempted a parameter estimation of latent success probability and number of trials by the Bayesian (hierarchical) modeling method.

From the analysis, I determined that:

(1) assuming approximately five trials of fifty-fifty games with around 0.5 success probability can well describe the observed class identification distribution in 2015 data;

(2) the Japanese postwar period can be divided into two periods in terms of people's subjective evaluation: the period of expanding opportunity from 1955 to 1975 and the period of high and constant success probability, but less chance of trials after 1985; and

(3) the different games model assuming differences in parameters among educational levels is always better in terms of goodness of prediction than the common game model assuming common parameters in the society in every survey period. However, these models were closest in terms of goodness of prediction in 1975 and 1985, possibly indicating that in the time of "all middle-class society," people evaluated their society almost the same as the common game in which all player has the same opportunities, but different luck.

Thus, although some future tasks, which includes constructing and testing more

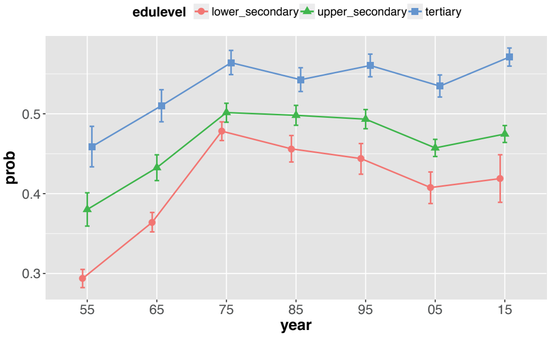

Fig. 20 p_j by educational levels by survey years (median and 95% HDI)

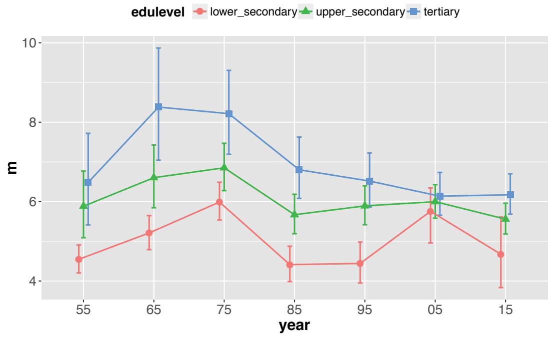

Fig. 21 m_j by educational levels by survey years (median and 95% HDI)

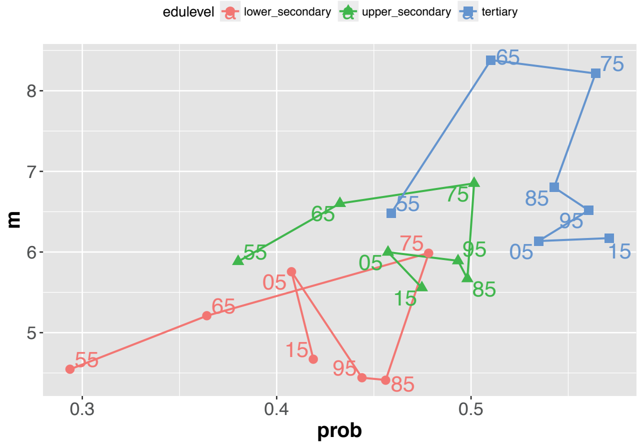

Fig. 22 Path plot of medians of p_j and m_j by educational levels by survey years

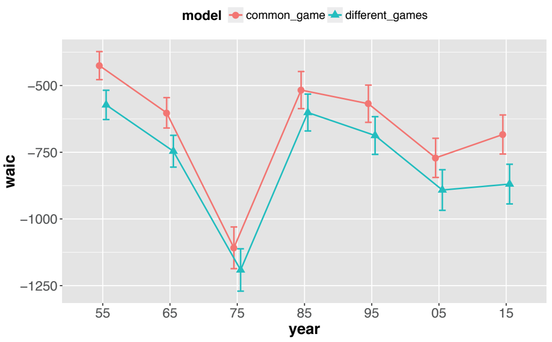

Fig. 23 Trends of WAIC (error bar: standard error)

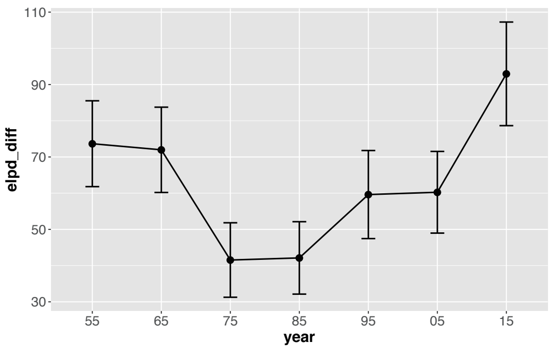

Fig. 24 Trends of WAIC difference (error bar: standard error)

empirically realistic and complex models, remain, at least in this study, we could shed new light on the well-known Japanese class identification trends using a new analytical framework.

References

Evans, M. D. R. and J. Kelley (2004). Subjective Social Location: Data From 21 Nations. *International Journal of Public Opinion Research 16*(1), 3–36.

Fararo, T. J. and K. Kosaka (2003). *Generating Images of Stratification: A Formal Theory.* Dordrecht, Netherlands: Kluwer Academic Publisher.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. David B. Dunson, A. Vehtari, and D. B. Rubin (2013). *Bayesian Data Analysis, 3rd Edition.* Boca Raton: Chapman & Hall/CRC.

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*, 23–35.

Ishida, A. (2012). A Detailed Derivation of the Distribution of Class Identification in a 'Chance Society': A Note on the Fararo-Kosaka Model. *Kwansei Gakuin University School of Sociology Journal 114*, 257–65.

Kelley, J. and M. D. R. Evans (1995). Class and Class Conflict in Six Western Nations. *American Sociological Review 60*(2), 157–178.

Kikkawa, T. (2016). *Social Mentality in Contemporary Japan: Quantitative Social Consciousness Studies.* Osaka: Osaka University Press.

Kobayashi, D. (2008). Kaisou Kizoku Ishiki ni Tsuite no Kiso Bunseki: Jiten Hikaku no Tame no Chuiten [A Basic Analysis of Status Identification: Points of Attention for Time Comparisons]. In S. Miwa and D. Kobayashi (Eds.), *2005-Nen SSM Nihon Chousa no Kiso Bunseki: Kouzou, Suusei, Houhou: 2005-Nen SSM Chosa Shiriizu 1 [Basic Analysis of the SSM 2005 Survey: Structure, Trend and Method]*, pp. 111–126. SSM 2005 Survey Management Committee.

Kruschke, J. K. (2015). *Doing Bayesian Data Analysis, Second Edition.* London: Academic Press.

Matsuura, K. (2016). *Stan to R de Bayes Toukei Modeling [Bayesian Statistical Modeling Using Stan and R].* Tokyo: Kyoritsu Shuppan.

Stan Development Team (2017a). RStan: the R interface to Stan. R package version

2.16.2. http://mc-stan.org.

Stan Development Team (2017b). Stan Modeling Language Users Guide and Reference Manual, Version 2.16.0. http://mc-stan.org.

Vehtari, A. and A. Gelman (2014). WAIC and Cross-validation in Stan. http://www.stat.columbia.edu/~gelman/research/unpublished/waic_stan.pdf.

Vehtari, A., A. Gelman, and J. Gabry (2016). loo: Efficient Leave-One-Out Cross-Validation and WAIC for Bayesian Models, R Package Version 1.1.0. https://CRAN.R-project.org/package=loo.

Watanabe, S. (2010). Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *Journal of Machine Learning Research 11*, 3571–3594.

7 Appendix 1: The Trends of the Parameters of the Different Games Model on Age Cohorts

This appendix shows the trends of the model parameters in the different games model on age cohorts in the postwar Japan from 1955 to 2015, and the results of model comparison between the common game model and the different games model in each survey wave by WAIC.

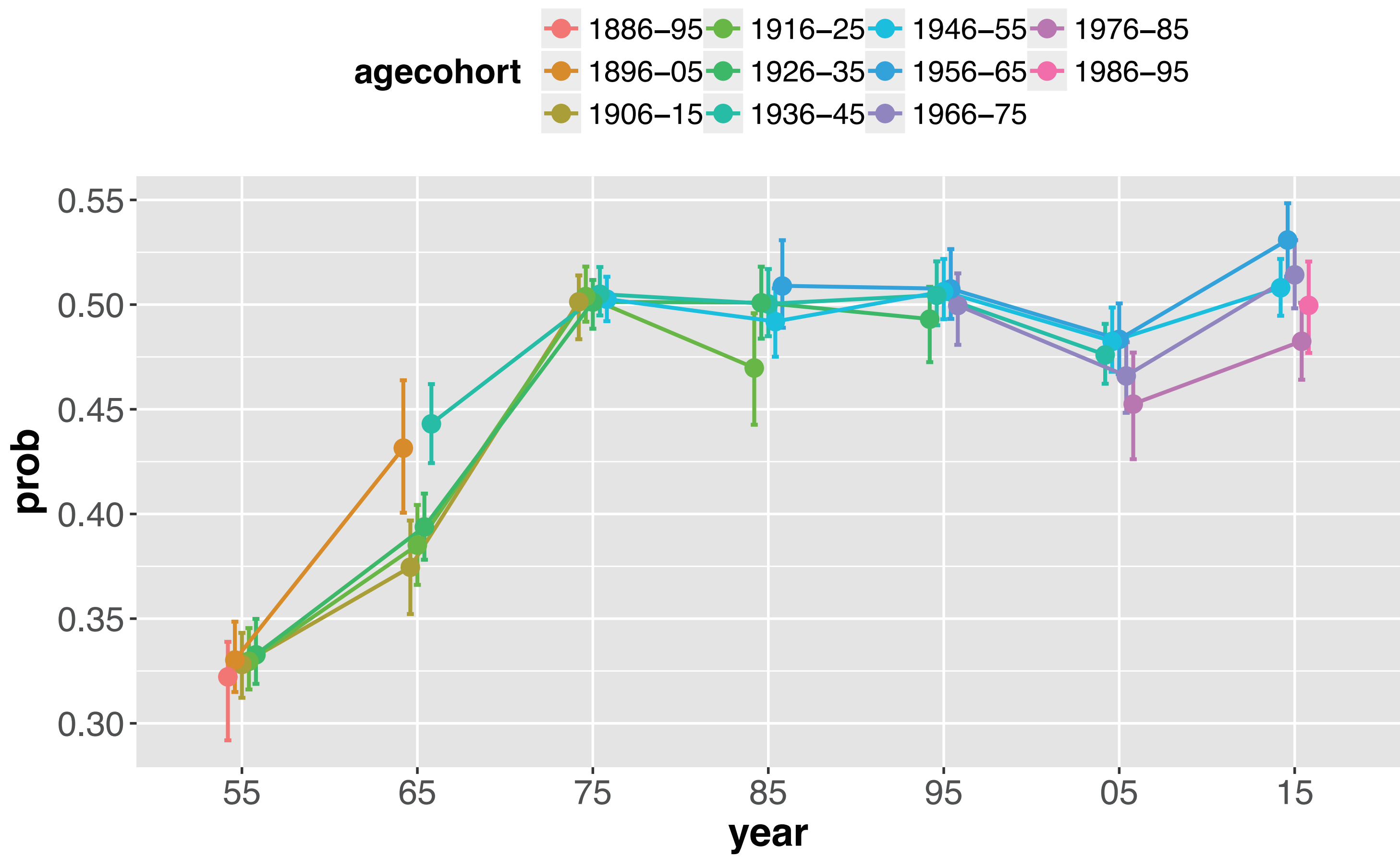

Fig. 25 p_j by age cohorts by survey years (median and 95% HDI)

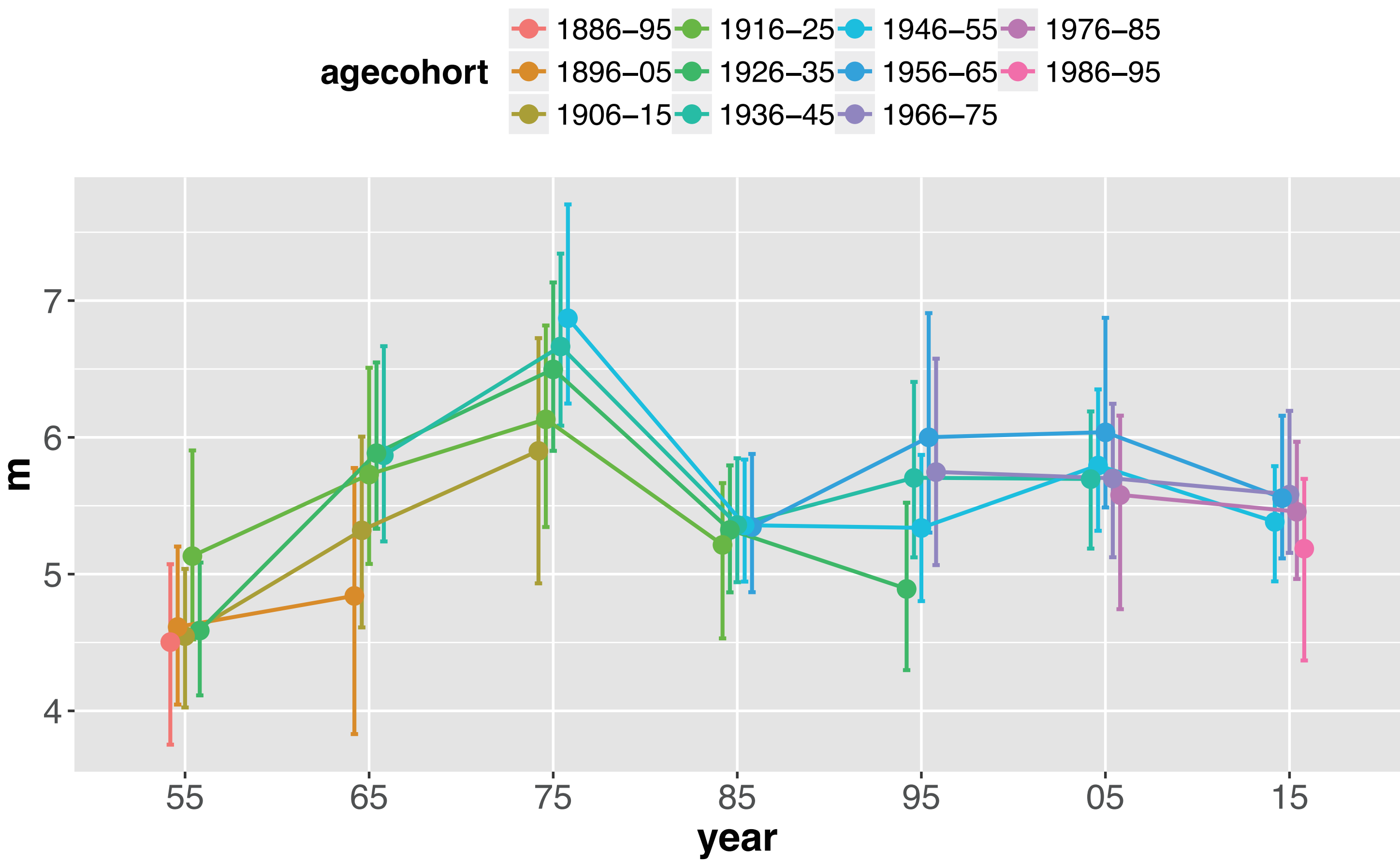

Fig. 26 m_j by age cohorts by survey years (median and 95% HDI)

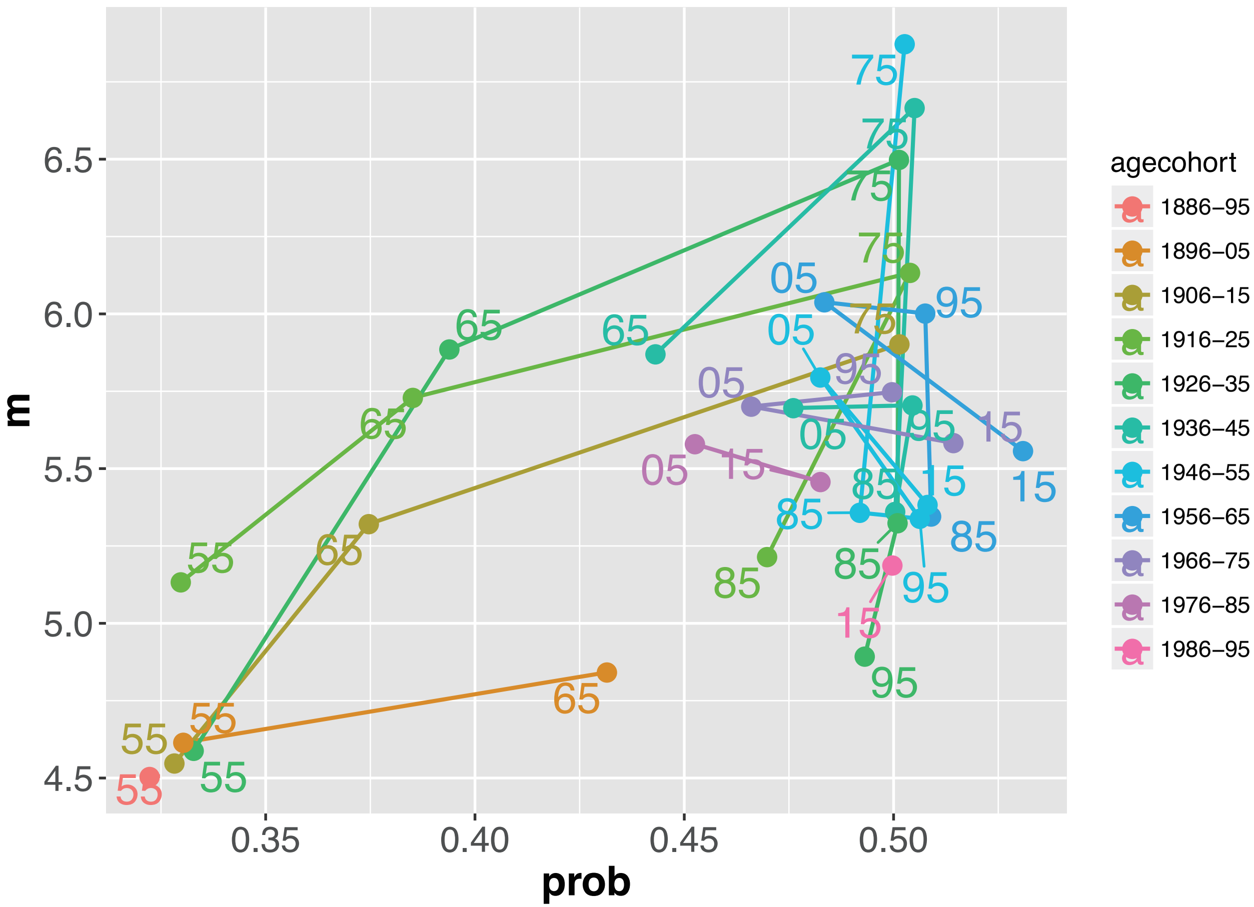

Fig. 27 Path plot of medians of p_j and m_j by age cohorts by survey years