行動科学演習・数理行動科学研究演習『社会科学のためのベイズ統計モデリング』第 10 章

Hiroshi Hamada
Tohoku University

Jul, 2020
at Tohoku University

8 章より再掲

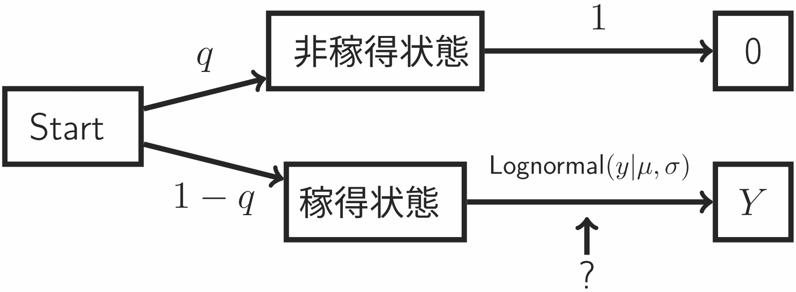
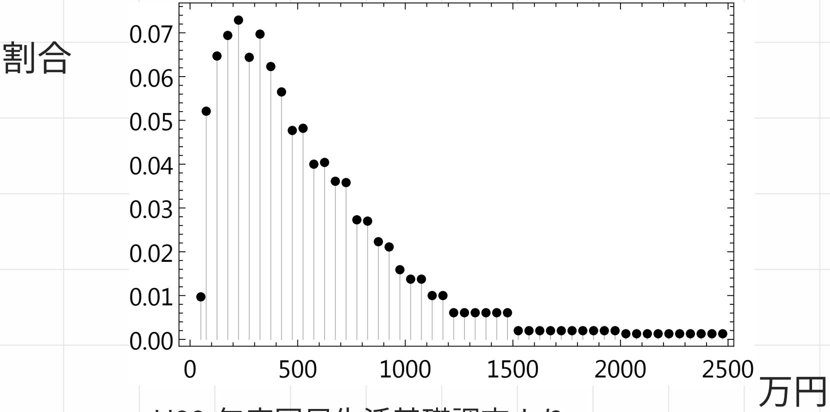


Figure: 分布の合成プロセスを表した樹形図

A Generative Model of Income Distribution(Hamada 1999;2003;2016; 2018)

Q:所得分布は，なぜ対数正規分布にしたがうのか？



H28 年度国民生活基礎調査より

Previous Studies in Economics

Mathematical Model

- McAlister (1879) presented a possible model
- Kapteyn (1903) firstly applied the lognormal to income distribution
- Gibrat (1931) illustrated the law of proportionate effect with extensive data
- Champernowne (1953) specified additional condition for Pareto distribution
- Rutherford (1955) applied random shock model to age cohorts

Human Capital Approach

- It explains differences in earnings of workers (Mincer 1958; 1974, Becker 1964, Becker & Chiswick1966, Ben-Porath 1967)
- Mincer's earnings function $\log y = y_0 + rS + \beta_1 X + \beta_2 X^2$ is one of the most widely used model in empirical economics (Lemieux 2006).
- 《対数収入を説明変数に回帰》は，いまや定番中の定番

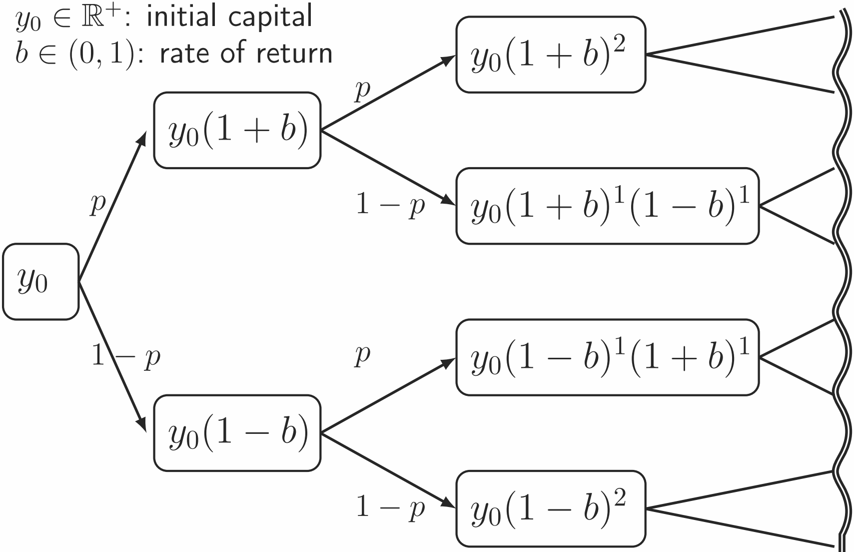
分布生成モデル (Hamada 1999;2003;2016; 2018)

所得分布生成プロセスの仮定（Hamada2016, Hamada2019）.

- 成功確率 $p \in (0,1)$ で人的資本の投資を n 回反復
- 初期資本: $y_0 \in \mathbb{R}^+$，利益率: $b \in (0,1)$.
- 投資コスト: $y_t b$
- 成功: $y_t = y_{t-1} + y_{t-1}b = y_{t-1}(1+b)$
- 失敗: $y_t = y_{t-1} - y_{t-1}b = y_{t-1}(1-b)$

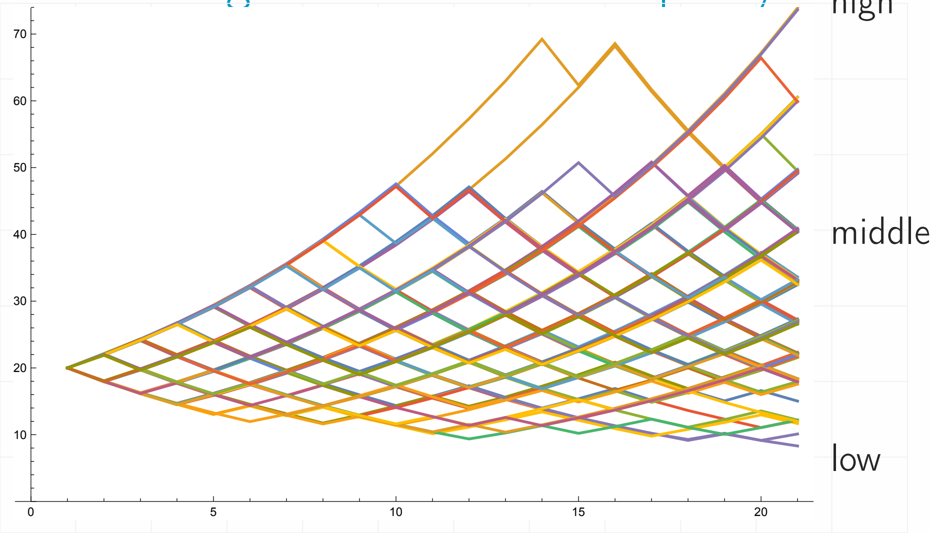
A Generative Model (Hamada 1999;2003;2016; 2018)

$y_0 \in \mathbb{R}^+$: initial capital
$b \in (0, 1)$: rate of return



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

Emergence of Income Disparity



Concentration in middle-low range

Probabilistic Human Capital Theory

$$
\begin{tikzcd}
\boxed{X_i \sim \text{Bernoulli}(p)} \arrow[d, "\text{chance}"] & & \boxed{Y \sim \Lambda(\mu, \sigma)} \\
\boxed{W \sim \text{Binomial}(n, p)} & & \boxed{\log Y \sim \text{N}(\mu, \sigma)} \arrow[u, "\text{capital dist.}"'] \\
W = \sum_{i=1}^{n} X_i \arrow[d, "\text{total success}"] & & \\
\underset{\text{capital at } n}{Y = y_0(1+b)^W(1-b)^{n-W}} \arrow[rr] & & \log Y_n \arrow[u, "n \text{ is large}"']
\end{tikzcd}
$$

Proof of Ideal Type

n 回試行後の資本 Y_n

$$Y_n = y_0(1+b)^X(1-b)^{n-X}$$
$$\log Y_n = \alpha X + \beta$$

n 回の試行で x 回成功する確率

$$P(X=x) = {}_nC_x p^x(1-p)^{n-x}$$

中心極限定理（厳密には分布収束は使えない）

$$Z = \frac{X-np}{\sqrt{np(1-p)}}, n \to \infty \Rightarrow Z \sim N(0,1)$$
$$X \sim N(np, npq) \Rightarrow \alpha X + \beta \sim N(\alpha np + \beta, \alpha^2 npq)$$
$$\log Y_n \sim N(\alpha np + \beta, \alpha^2 npq)$$

Bayes Model

$$
\begin{aligned}
\log Y_i &\sim \mathsf{Normal}(\mu, \sigma) \quad i = 1, 2, \ldots N \\
\mu &= \log y_0 + n \log(1-b) + \log \frac{1+b}{1-b} np \\
\sigma &= \sqrt{npq} \log \frac{1+b}{1-b} \\
p &\sim \mathrm{Beta}(1,1), \qquad b \sim \mathrm{Beta}(1,1)
\end{aligned}
$$

Stan code

```
parameters {
real <lower=0, upper=1> p; //成功確率
real <lower=0, upper=1> b; //投資利益率
}

transformed parameters{//y0=10, n=10を仮定
real m; real s;
real y0;// initial capital
real n;// number of chance
y0=10;n=10;
m = log(y0)+n*log(1-b)+log((1+b)/(1-b))*n*p;
s = sqrt(n*p*(1-p))*log( (1 + b )/(1 - b));
}

model {
for (i in 1:N) y[i] ~ normal(m, s);
}//データ上,対数所得に変換済み
```

予測分布と MCMC 結果

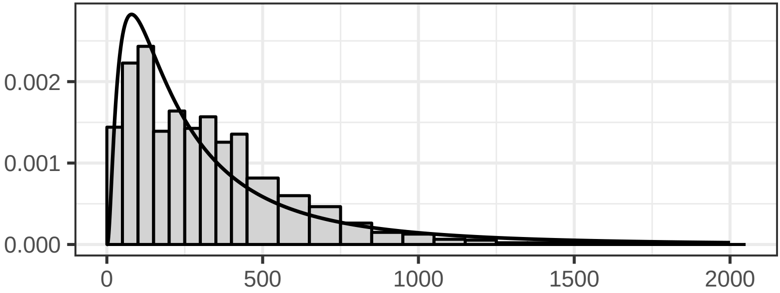


Figure: 予測分布とデータ（SSP2015 個人年収）の比較

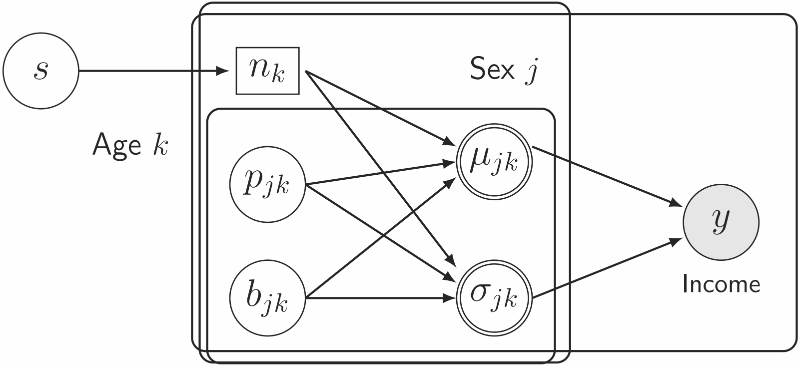
	mean	2.5%	97.5%	n_eff	Rhat
m	5.4845	5.4507	5.5184	4000	0.9993
s	0.9541	0.9297	0.9799	2132	1.0009
p	0.9192	0.9165	0.9219	2319	1.0008
b	0.5032	0.4968	0.5102	2913	1.0000

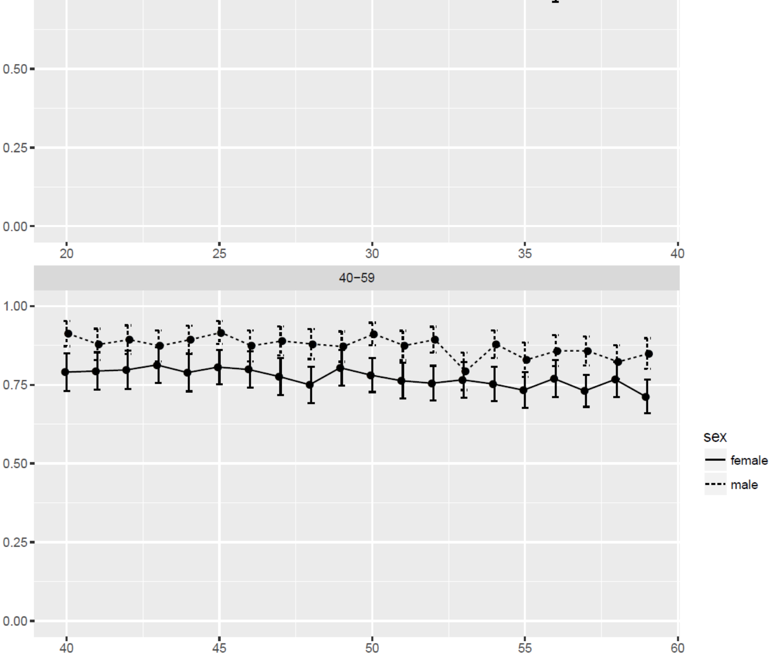
Hierarchical Bayes Model

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

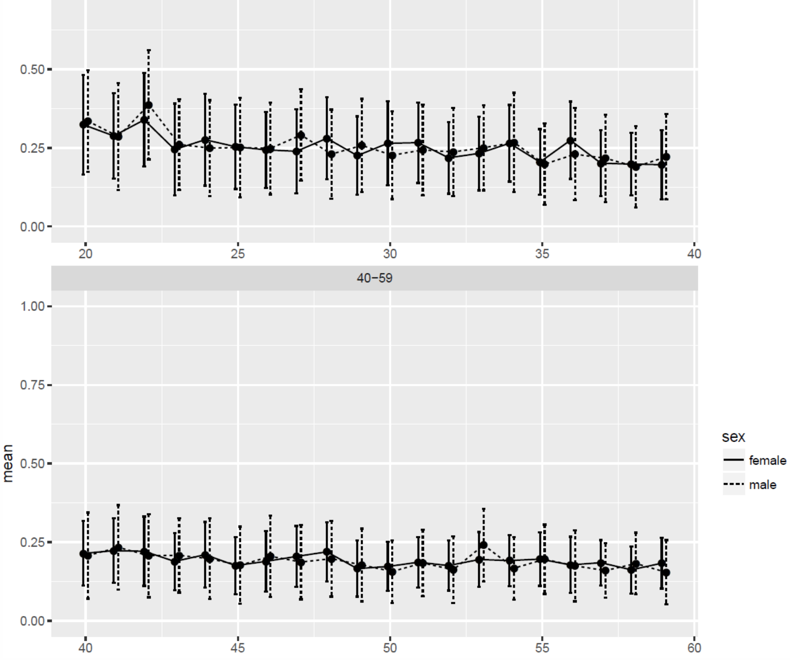
$$p_{jk}, b_{jk} \sim \mathrm{Beta}(1,1) \quad \mu_{jk} = \log y_0 (1-b_{jk})^{n_k} \frac{1+b_{jk}}{1-b_{jk}} n_k p_{jk}$$

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





Posterior distribution of success probability p, model 2.



Posterior distribution of rate of return b, model 2.

ベイズモデルのメリット

- Toy model + Bayesian modelling により理論の検証とインプリケーションの考察が可能.
- GLM 的な変数の追加投入は，ベイズモデルと両立可能（方法論的な上位互換性）
- WAIC により予測精度の比較が可能（ただし GLM とのモデル比較には注意）

数理モデルのメリット

- 応答変数の分布のパラメータが理論的な概念の関数として明示的に特定できる

$$
\begin{aligned}
\mathbb{E}[Y] &= \exp\left\{\mu + \frac{\sigma^2}{2}\right\} \\
\sigma &= \sqrt{np(1-p)}\log\left(\frac{1+b}{1-b}\right) \\
G &= 2\int_{-\infty}^{\sigma/\sqrt{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx - 1 \\
G &= f(\sigma)
\end{aligned}
$$

- 微分で分析し放題

Implication

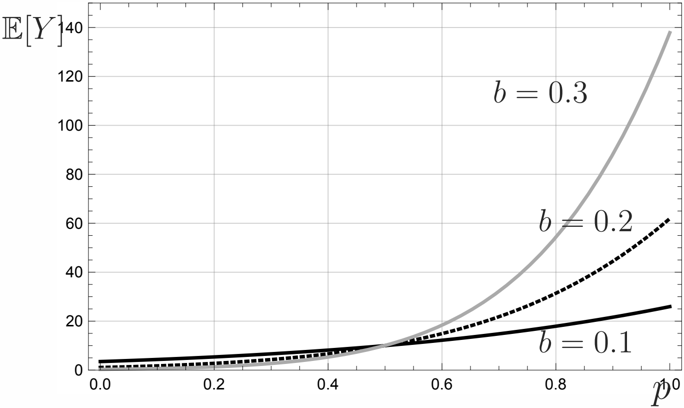


Figure: 資本獲得確率 p，利益率 b と所得平均 $\mathbb{E}[Y]$ の関係．
$n = 10, y_0 = 10$

Implication

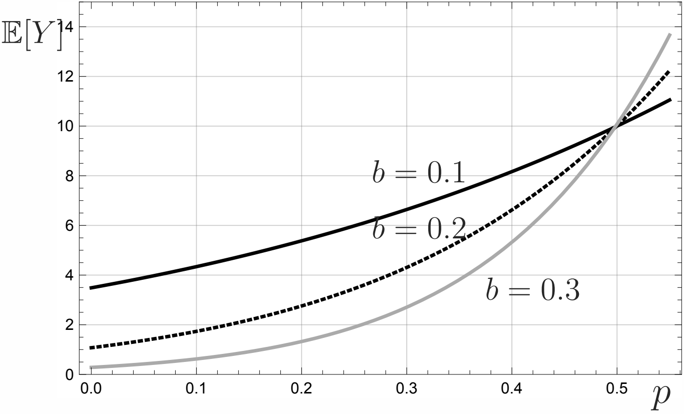


Figure: 資本獲得確率 p，利益率 b と所得平均 $\mathbb{E}[Y]$ の関係．
$n = 10, y_0 = 10$

Implication

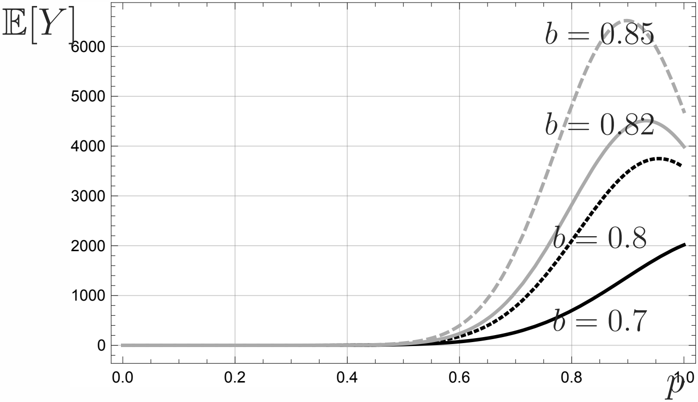


Figure: 資本獲得確率 p，利益率 b と所得平均 $\mathbb{E}[Y]$ の関係．
$n = 10, y_0 = 10$

Implication

命題

利益率 $b < (e^2 - 1)/(e^2 + 1)$ のとき，人的資本の平均 $\mathbb{E}[Y]$ は成功確率 p の増加関数である（*Hamada 2016*）

命題

人的資本 Y_n のジニ係数は，利益率に関して増加である．また p に関しては $p = 0.5$ で極大値をとる．

$$G = \frac{1}{2\mu} \int_0^\infty \int_0^\infty |x - y| f(x) f(y) dx dy$$

Implication

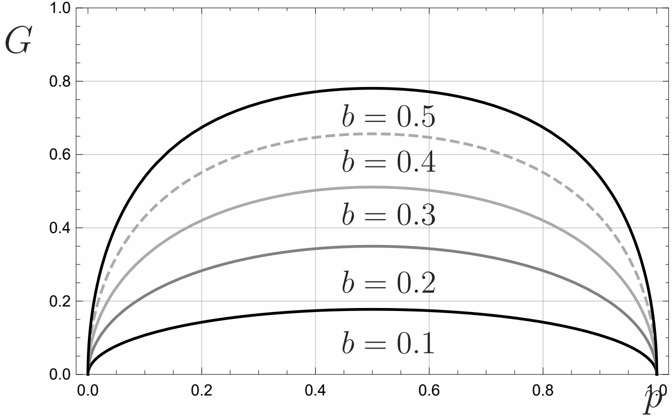


Figure: 資本獲得確率 p，利益率 b とジニ係数 G の関係．
$n = 10, y_0 = 10$

数理モデル + ベイズ統計の含意

- 数理モデルからの理論的インプリケーションとベイズ統計分析の結果を組み合わせて，新しいインプリケーションを導出できる
- p の事後分布平均 0.92，b の事後分布平均は 0.48: 資本獲得確率 p が増加すれば，不平等度は減少し，逆に p が減少すれば，不平等度は増加
- 人的資本獲得確率 p だけを上昇させることができれば，経済的発展と不平等改善が同時に成立