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 11 章

Hiroshi Hamada
Tohoku University

Aug, 2020
at Tohoku University

収入の評価

- 国レベルで幸福と所得には弱い関連しか見られない
- 国レベルの経済発展による所得の増加は平均的な幸福の増加をもたらさない（Easterlin 1974,1975）
- 階層イメージのモデル:　自分がどの階層に属しているかという階層帰属意識が，他者との比較のなかでバイアスを伴って形成される過程（Fararo and Kosaka 2003）

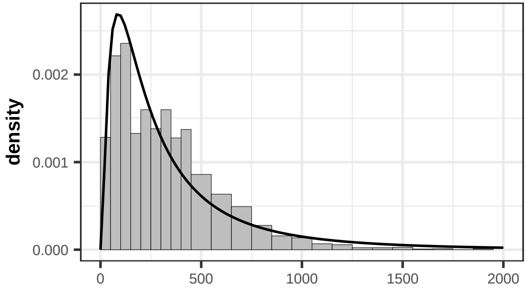

Figure: 個人収入の客観分布（SSP2015 データ，実線は推定された対数正規分布）

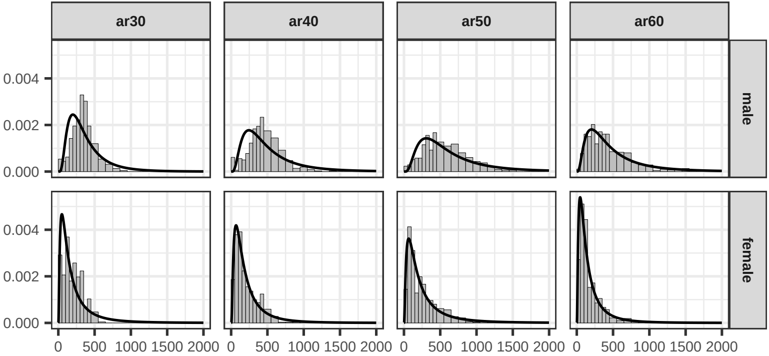

Figure: 性別・年齢階層ごとの個人収入の客観分布（SSP2015 データ，実線は推定された対数正規分布）

「収入について，現代日本社会おける最高水準を 1，最低の水準を 10 とすると，現在のあなたご自身はどの位にあたると思われますか」

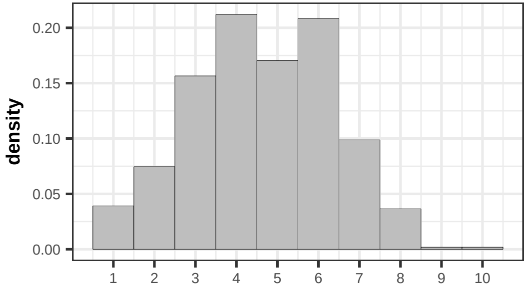

Figure: 収入評価の分布（SSP2015 データ）

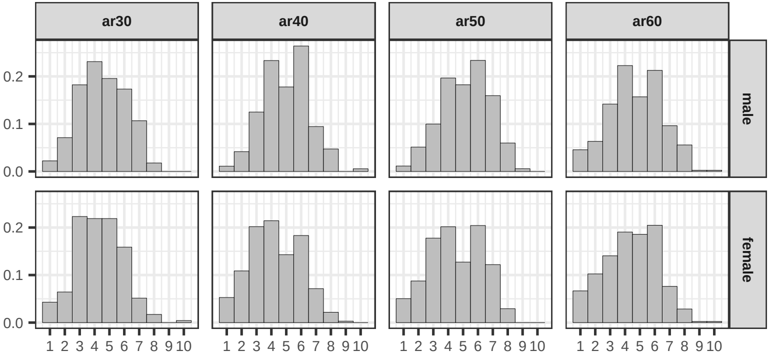

Figure: 性別・年齢階層ごとの収入評価の分布（SSP2015 データ）

ベースラインモデル

1. 人はある次元での自らの相対的地位を評価する際に，ランダムに出会う他者との比較を行う
2. その際，自分が他者よりも，その次元において上か下かを判断する
3. 他者よりも自分が上になる（自分より下となる他者に出会う）回数が多ければ多いほど，自らの相対的地位を高く評価する
4. 具体的な評価の際には，直近の出会いから評価を構成する

ベースラインモデル

1. 人はある次元での自らの相対的地位を評価する際に，ランダムに出会う他者との比較を行う
2. その際，自分が他者よりも，その次元において上か下かを判断する
3. 他者よりも自分が上になる（自分より下となる他者に出会う）回数が多ければ多いほど，自らの相対的地位を高く評価する
4. 具体的な評価の際には，直近の出会いから評価を構成する

例

- 他者とランダムに 4 回出会い，自分が上かどうか確かめる
- 《自分が上》の回数: $0 \sim 4$
- 《自分が上》の回数に応じて回答選択肢（最低 1，最高 5）を選ぶ

回数		回答
0	$\rightarrow$	1
1	$\rightarrow$	2
2	$\rightarrow$	3
3	$\rightarrow$	4
4	$\rightarrow$	5

例

- 他者とランダムに n 回出会い，自分が上かどうか確かめる
- 《自分が上》の回数: $0 \sim n$
- 《自分が上》の回数（パーセンタイル）に応じて回答選択肢（最低 1，最高 5）を選ぶ

回数	回答
20	$\rightarrow$ 1
40	$\rightarrow$ 2
60	$\rightarrow$ 3
80	$\rightarrow$ 4
100	$\rightarrow$ 5

収入評価のトイモデル

対数正規分布に従う収入の分布関数 $F_\Lambda(x|\mu,\sigma)$.
x の値をもつ個人が自分より下の値 Z をもつ他者とランダムに出会う確率

$$\Pr(Z \leq x) = F_\Lambda(x|\mu,\sigma) = F_\Phi\left(\frac{\log x - \mu}{\sigma}\right) \tag{1}$$

ただし，F_Φ は標準正規分布 $\mathrm{Normal}(0,1)$ の分布関数.
ランダムな出会いを m 回繰り返したとき，他者よりも自分が上になる回数（確率変数）Y は試行回数 m，確率 $F_\Lambda(x|\mu,\sigma)$ の二項分布に従う

$$Y \sim \mathrm{Binomial}(m, F_\Lambda(x|\mu,\sigma)). \tag{2}$$

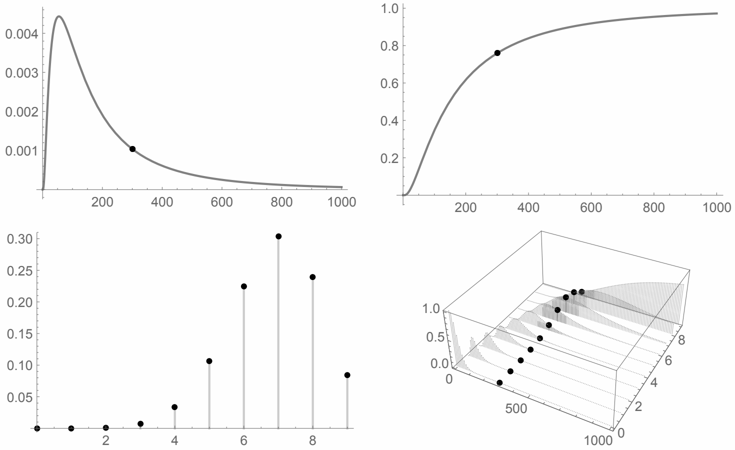

Figure: 収入評価のトイモデルにおける対数正規分布 $\mathrm{Lognormal}(5,1)$ の密度関数（左上），分布関数 $F_\Lambda(x|5,1)$（右上），評価の二項分布 $\mathrm{Binomial}(9, F_\Lambda(300|5,1))$（左下），収入を 0 から 1000 まで変化させたときの評価の二項分布の変化，黒点の列は $x=300$ の評価の二項分布．その断面は $\mathrm{Binomial}(9, F_\Lambda(300|5,1))$ に一致（右下）

注意

収入評価（自分が上になる回数）の 2 項分布は，収入毎に異なる．$Y \sim \mathrm{Binomial}(9, F_\Lambda(100|5,1))$.

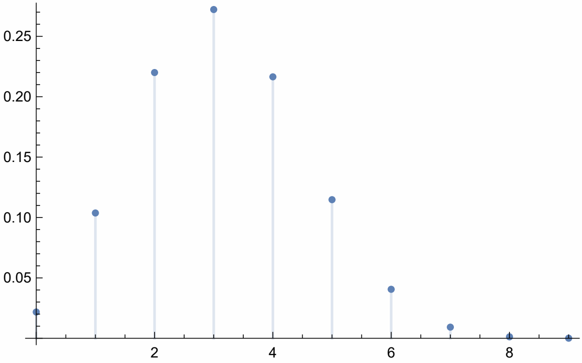

Figure: 収入評価の分布（100 万円の人）

注意

収入評価（自分が上になる回数）の 2 項分布は，収入毎に異なる．$Y \sim \text{Binomial}(9, F_\Lambda(100|5,1))$.

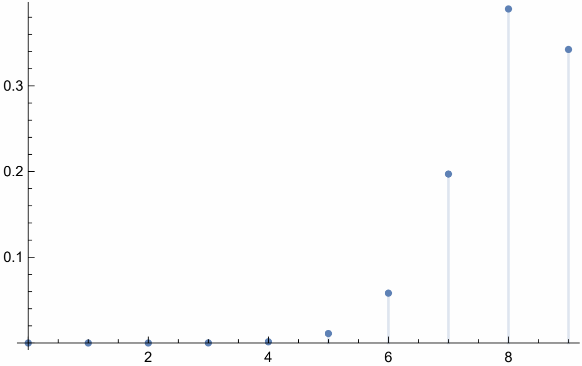

Figure: 収入評価の分布（500 万円の人）

トイモデルからの導出

$\pi = F(x)$ とおくと，二項分布 $\mathrm{Binomial}(m, \pi)$ における Y の確率質量関数は

$$\mathrm{Binomial}(y|m, \pi) = {}_mC_y \pi^y (1-\pi)^{m-y}$$

ここで，$\pi = F(x)$ は個人の評価次元の値を示す確率変数 X の関数なので，π も確率変数であり，$\mathrm{Binomial}(y|m, \pi)$ は確率変数 Y と π の同時確率関数と見なせる．そこで π で周辺化した y の確率関数 $p(y)$ を求めると（π の密度は一様分布となる．テキスト参照）

$$
\begin{aligned}
p(y) &= \int_0^1 \mathrm{Binomial}(y|m,\pi)d\pi \\
&= \int_0^1 {}_mC_y \pi^y (1-\pi)^{m-y} d\pi \\
&= {}_mC_y \int_0^1 \pi^{y+1-1}(1-\pi)^{m-y+1-1} d\pi \\
&= {}_mC_y \mathrm{B}(y+1, m-y+1)
\end{aligned}
$$

ただし，$\mathrm{B}(a,b)$ はベータ関数.

$y+1, m-y+1$ は整数なので，ベータ関数の公式より，

$$\begin{aligned} p(y) &= {}_mC_y B(y+1, m-y+1) \\ &= \frac{m!}{y!(m-y)!}\frac{y!(m-y)!}{(m+1)!} = \frac{1}{m+1} \end{aligned}$$

ベータ 2 項分布の特殊型

$\rightarrow$ しかし観測データは一様分布ではない!

ベイズ統計モデリング

null モデル

$$
\begin{gathered}
Y_i \sim \mathrm{Binomial}(9, \pi) \\
\pi = \mathrm{logistic}(a) \\
a \sim \mathrm{Uniform}(-\infty, \infty) \\
X_i \sim \mathrm{Lognormal}(\mu, \sigma) \\
\mu \sim \mathrm{Uniform}(0, 8) \qquad \sigma \sim \mathrm{Uniform}(0, \infty)
\end{gathered}
$$

線形バイアスモデル

$$Y_i \sim \text{Binomial}(9, \pi_i)$$

$$\pi_i = \text{logistic}(a + b * \text{logit}(F_\Lambda(X_i | \mu, \sigma)))$$

$$a \sim \text{Normal}(0, 10^2) \qquad b \sim \text{Normal}(1, 10^2)$$

$$X_i \sim \text{Lognormal}(\mu, \sigma)$$

$$\mu \sim \text{Uniform}(0, 8) \qquad \sigma \sim \text{Uniform}(0, \infty)$$

バイアスの意味

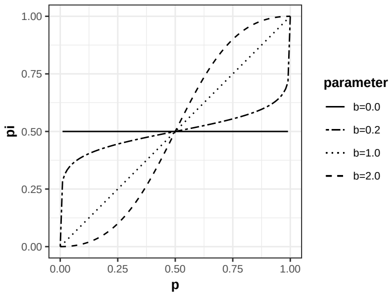

Figure: 線形バイアスモデルの $p = F_\Lambda$ と π の理論的関係 $(a = 0)$

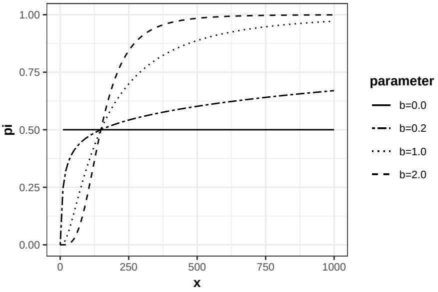

Figure: 線形バイアスモデルの x と π の理論的関係 $(a = 0)$

Table: 線形バイアスモデルの推定結果

	平均	2.50%	97.50%	$\hat{R}$
a	-0.404	-0.433	-0.376	1.000
b	0.187	0.170	0.204	1.000
μ	5.476	5.437	5.514	1.000
σ	1.023	0.996	1.051	1.000

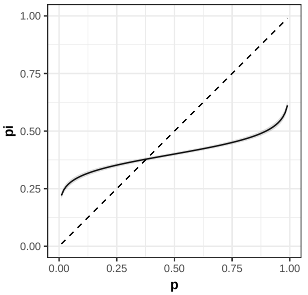

Figure: 線形バイアスモデルにおける $p = F_\Lambda$ と π の関係

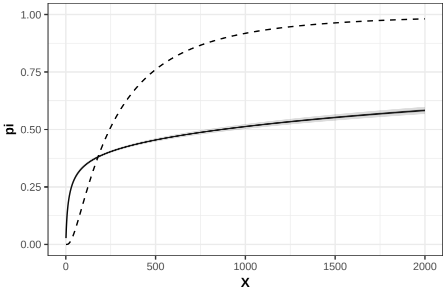

Figure: 線形バイアスモデルにおける x と π の関係

破線はバイアスがない場合を表す

null モデルと線形バイアスモデルの周辺尤度の対数比（対数ベイズファクター）は，

$$\log \mathrm{BF}(\mathrm{bias}, \mathrm{null}) = F_n(\mathrm{null}) - F_n(\mathrm{bias}) = 293.27$$

線形バイアスモデルは，null モデルに対して，データにより当てはまる

階層モデル

$$
\begin{gathered}
Y_i \sim \mathrm{Binomial}(9, \pi_i) \\
\pi_i = \mathrm{logistic}(a_{j(i)} + b_{j(i)} * \mathrm{logit}(F_\Lambda(X_i | \mu_{j(i)}, \sigma_{j(i)}))) \\
a_j \sim \mathrm{Normal}(a_0, \sigma_a) \qquad b_j \sim \mathrm{Normal}(b_0, \sigma_b) \\
a_0 \sim \mathrm{Normal}(0, 10^2) \qquad \sigma_a \sim \mathrm{Uniform}(0, \infty) \\
b_0 \sim \mathrm{Normal}(1, 10^2) \qquad \sigma_b \sim \mathrm{Uniform}(0, \infty) \\
X_i \sim \mathrm{Lognormal}(\mu_{j(i)}, \sigma_{j(i)}) \\
\mu_j \sim \mathrm{Uniform}(0, 8) \qquad \sigma_j \sim \mathrm{Uniform}(0, \infty)
\end{gathered}
$$

$j(i)$ は，個人 i の属する属性カテゴリ j を示す

線形バイアスモデルと階層モデルの対数ベイズファクターは,

$$\begin{aligned}\log \mathrm{BF}(\text{hierarchical}, \text{bias}) &= F_n(\text{bias}) - F_n(\text{hierarchical}) \\ &= 508.24\end{aligned}$$

階層モデルは線形バイアスモデルよりも，さらに高い当てはまりを示す