

基礎からのベイズ統計学

豊田秀樹（朝倉書店, 2015 年）

4 メトロポリス・ヘイスティングス法

4.1 事後分布からの乱数の発生

- 確率分布が評価できても、それに従う乱数を発生させられるとは限らない。
- 正規化定数が評価できないことが多い。

→ 事後分布のカーネルだけを使い、乱数を発生させたい。

4.5 メトロポリス・ヘイスティングス法

詳細釣り合い条件 マルコフ連鎖が定常分布に収束するための十分条件

$f(\theta)$ が目標分布、 $f(\theta'|\theta)$ が遷移核であるとする、

$$f(\theta|\theta')f(\theta') = f(\theta'|\theta)f(\theta) \quad (4.18)$$

が成り立っていること^{*1}。このとき、

$$f(\theta) = \int f(\theta|\theta')f(\theta')d\theta' \quad (4.19)$$

となる。

遷移核 $q(\cdot|\cdot)$ を適当に設定すると、

$$q(\theta|\theta')f(\theta') > q(\theta'|\theta)f(\theta) \quad (4.20)$$

のように詳細釣り合い条件は満たされなくなる。メトロポリス・ヘイスティングス法では、確率補正

$$0 \leq r \left(= \frac{q(\theta'|\theta)f(\theta)}{q(\theta|\theta')f(\theta')} \right) \leq 1 \quad (4.24)$$

により詳細釣り合い条件が満たされるようにする。

^{*1} ある時点で θ が θ' の a 倍起こりやすいとすると、 $\theta' \rightarrow \theta$ は $\theta \rightarrow \theta'$ より a 倍起こりやすい。

提案分布 $q(a|\theta^{(t)})$ において

$$q(a|\theta^{(t)})f(\theta^{(t)}) > q(\theta^{(t)}|a)f(a) \quad (4.27)$$

であれば、確率

$$r = \frac{q(\theta^{(t)}|a)f(a)}{q(a|\theta^{(t)})f(\theta^{(t)})} \quad (4.28)$$

で $\theta^{(t+1)} = a$ とし、確率 $1 - r$ で $\theta^{(t+1)} = \theta^{(t)}$ とする。

そうでなければ、常に $\theta^{(t+1)} = a$ とする。

ここで、一般に $f(\cdot)$ は事後分布であるから、

$$r = \frac{q(\theta^{(t)}|\theta_a)f(\theta_a|x)}{q(\theta_a|\theta^{(t)})f(\theta^{(t)}|x)} \quad (4.29)$$

$$= \frac{q(\theta^{(t)}|\theta_a) \frac{f(x|\theta_a)f(\theta_a)}{f(x)}}{q(\theta_a|\theta^{(t)}) \frac{f(x|\theta^{(t)})f(\theta^{(t)})}{f(x)}} \quad (4.30)$$

$$= \frac{q(\theta^{(t)}|\theta_a)f(x|\theta_a)f(\theta_a)}{q(\theta_a|\theta^{(t)})f(x|\theta^{(t)})f(\theta^{(t)})} \quad (4.31)$$

となり、事後分布についてはカーネルのみ利用すれば良いことがわかる。

4.6 独立 MH 法

提案分布を単に無条件分布とすることも可能。この場合、補正係数は

$$r = \frac{q(\theta^{(t)})f(a)}{q(a)f(\theta^{(t)})} \quad (4.33)$$

となる。もし提案分布が区間 $[0, 1]$ の一様分布なら、補正係数は

$$r = \frac{f(a)}{f(\theta^{(t)})} \quad (4.35)$$

と簡略化される。

4.7 ランダムウォーク MH 法

候補の提案を、対象な分布からの実現値 e を利用して、

$$a = \theta^{(t)} + e \quad (4.36)$$

とする。この場合、提案分布が

$$q(a|\theta^{(t)}) = q(\theta^{(t)}|a) \quad (4.37)$$

となるため、補正係数は常に

$$r = \frac{f(a)}{f(\theta^{(t)})} \quad (4.38)$$

となる。

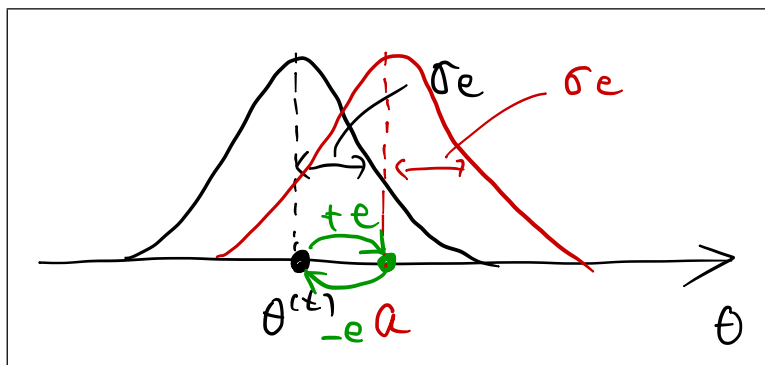


図 4.1 対象分布により $q(a|\theta^{(t)}) = q(\theta^{(t)}|a)$ となる理由

4.8 生成量・研究仮説が正しい確率

生成量 事後分布からのサンプル $\theta^{(t)}$ の関数 $g(\theta^{(t)})$

研究仮説 U について

$$u^{(t)} = g(\theta^{(t)}) = \begin{cases} 1 & U \text{ が成立} \\ 0 & U \text{ が不成立} \end{cases} \quad (4.44)$$

とすると、 $u^{(t)}$ の平均は $E[U|\mathbf{x}]$ の EAP 推定量で、研究仮説が正しい確率の評価を与えてくれる。

5 ハミルトニアンモンテカルロ法

5.4 ハミルトニアン

質量 $m = 1$ 、重力加速度 $g = 1$ の下で、運動エネルギーとポテンシャルエネルギーはそれぞれ

$$K(p) = \frac{1}{2}p^2 \quad (5.34)$$

$$U(\theta) = h(\theta) \quad (5.35)$$

となるので*2、ハミルトニアンは

$$H(\theta, p) = U(\theta) + K(p) = h(\theta) + \frac{1}{2}p^2 \quad (5.36)$$

となる。

位相空間 位置と運動量を座標とする空間

位相空間においては、

- 可逆
- 体積保存

が成り立つ。

ここで、物体はハミルトニアンが常に一定になるように運動する。つまりハミルトニアンを時間 τ の関数とすると、

$$\frac{dH(\tau)}{d\tau} = 0 \quad (5.18)$$

が成り立つ。ここから、物体の動きを予測するハミルトンの運動方程式

$$\frac{dp(\tau)}{d\tau} = -h'(\theta(\tau)) \quad (5.27)$$

$$\frac{d\theta(\tau)}{d\tau} = p(\tau) \quad (5.28)$$

が導かれる。

リーフフロッグ法 ハミルトンの運動方程式を数值的に解く方法

$$p\left(\tau + \frac{1}{2}\right) = p(\tau) - \frac{\epsilon}{2}h'(\theta(\tau)) \quad (5.31)$$

$$\theta(\tau + 1) = \theta(\tau) + \epsilon p\left(\tau + \frac{1}{2}\right) \quad (5.32)$$

$$p(\tau + 1) = p\left(\tau + \frac{1}{2}\right) - \frac{\epsilon}{2}h'(\theta(\tau + 1)) \quad (5.33)$$

5.5 HMC 法

HMC 法 事後分布 $f(\theta|\mathbf{x})$ と、それとは独立な標準正規分布 $f(p)$ との同時分布

$$f(\theta, p|\mathbf{x}) = f(\theta|\mathbf{x})f(p) \quad (5.37)$$

を考え、この同時分布から乱数を発生させる*3。

*2 θ は位置、 $p = mv$ は運動量。

*3 $f(\theta|\mathbf{x})$ と $f(p)$ は独立であるから、 $f(\theta, p|\mathbf{x})$ から θ を発生させるのは $f(\theta|\mathbf{x})$ から θ を発生させるのと同じこと。

このとき、 $-h(\theta) = \log(f(\theta|\mathbf{x}))$ とおくと、

$$f(\theta, p|\mathbf{x}) = \left(-h(\theta) - \frac{1}{2}p^2 \right) \quad (5.44)$$

$$= \exp(-H(\theta, p)) \quad (5.45)$$

が導かれる。

事後分布とそれとは独立な標準正規分布との同時事後分布の確率密度関数は、負のハミルトニアンに等しい。

ここで、位相空間の可逆性^{*4}と体積保存性^{*5}より、詳細釣り合い条件

$$f(\theta, p|\theta', p')f(\theta', p'|\mathbf{x}) = f(\theta', p'|\theta, p)f(\theta, p|\mathbf{x}) \quad (5.48)$$

が成立するが、リープフロッグ法に導入した ϵ により数値計算上は完全には成立しない。そこで、補正係数

$$r = \frac{f(\theta^{(a)}, p^{(a)}|\mathbf{x})}{f(\theta^{(t)}, p^{(t)}|\mathbf{x})} = \exp(H(\theta^{(t)}, p^{(t)}) - H(\theta^{(a)}, p^{(a)})) \approx 1 \quad (5.49)$$

を用いる^{*6}。

HMC 法アルゴリズム

1. 初期値 $\theta^{(1)}, \epsilon, L, T$ を定める。^a
2. $p^{(t)}$ を標準正規分布から発生
3. リープフロッグ法により $\theta^{(t)}, p^{(t)}$ から $\theta^{(a)}, p^{(a)}$ に移動
4. 確率 r で受容 ($\theta^{(t+1)} = \theta^{(a)}$)
5. $T = t$ で終了

^a L はリープフロッグ法の遷移時間。

標準正規分布から乱数 p 発生（運動量）、その時点での位置 $\theta^{(t)}$ と運動量 p からハミルトンの運動方程式に従って時間 L 後の物体の位置 $\theta^{(a)}$ と運動量 $p^{(a)}$ を予測。ハミルトニアンの比に従ってそれを $\theta^{(t+1)}$ として受容。 p はその都度使い捨て。

^{*4} 遷移確率 $f(\theta', p'|\theta, p) = f(\theta, p|\theta', p')$

^{*5} $f(\theta, p|\mathbf{x}) = f(\theta', p'|\mathbf{x})$

^{*6} 遷移確率が等しいため、ランダムウォーク MH 法のように補正係数から提案分布の部分は消える。

5.6 多次元の場合

HMC 法アルゴリズム（母数が d 次元）

1. 初期値 $\boldsymbol{\theta}^{(1)}, \epsilon, L, T$ を定める。^a
2. $\mathbf{p}^{(t)}$ を標準正規分布から発生（独立な d 個）
3. リーフフロッグ法により $\boldsymbol{\theta}^{(t)}, \mathbf{p}^{(t)}$ から $\boldsymbol{\theta}^{(a)}, \mathbf{p}^{(a)}$ に移動
4. 確率 r で受容 ($\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(a)}$)
5. $T = t$ で終了

^a L はリーフフロッグ法の遷移時間。

ただし、ハミルトニアンは

$$H(\boldsymbol{\theta}, \mathbf{p}) = h(\boldsymbol{\theta}) + \frac{1}{2} \sum_{i=1}^d p_i^2 \quad (5.55)$$

であり、

$$h'(\boldsymbol{\theta}) = \left(\frac{dh(\boldsymbol{\theta})}{d\theta_1}, \dots, \frac{dh(\boldsymbol{\theta})}{d\theta_d} \right) \quad (5.57)$$

$$r = \exp \left(H(\boldsymbol{\theta}^{(t)}, \mathbf{p}^{(t)}) - H(\boldsymbol{\theta}^{(a)}, \mathbf{p}^{(a)}) \right) \quad (5.58)$$

を用いる。

平均 μ 、分散 σ^2 の正規分布について事後分布 $f(\mu, \sigma^2 | \mathbf{x})$ を推測したいとする。 μ と σ^2 のいずれにも一様事前分布を設定した場合、 $h(\boldsymbol{\theta})$ は

$$\log(\mu, \sigma^2 | \mathbf{x}) = \frac{-n}{2} \log \sigma^2 + \frac{-1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 + \text{const.} \quad (5.59)$$

の右辺第 2 項までにマイナスをかけたものを取れば良い。

また、 $h'(\boldsymbol{\theta})$ には

$$\frac{d}{d\mu} \log f(\mathbf{x} | \mu, \sigma^2) = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) \quad (5.60)$$

$$\frac{d}{d\sigma^2} \log f(\mathbf{x} | \mu, \sigma^2) = \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 \quad (5.61)$$

を利用する。

6 正規分布に関する推測

6.1 基本的な推測

6.1.3 分位に関する推測

上位 25% がボーダーラインだとする。受験者集団の得点データが得られており、ある生徒の過去の得点分布が既知の正規分布であるとき、この生徒がボーダーラインを超える確率を求める。

1. データから事後分布について $\mu^{(t)}$ と $\sigma^{2(t)}$ を発生
2. 事後 75% 分位点を $\xi_{0.75}^{(t)} = \mu^{(t)} + z_{0.75}\sigma^{(t)}$ で算出
3. ある既知の分布において、 $\xi_{0.75}^{(t)}$ が何 % 分位点にあたるか算出 (生成量 $\Psi(\xi_{0.75}^{(t)})$)
4. 生成量 $\Psi(\xi_{0.75}^{(t)})$ の平均を算出

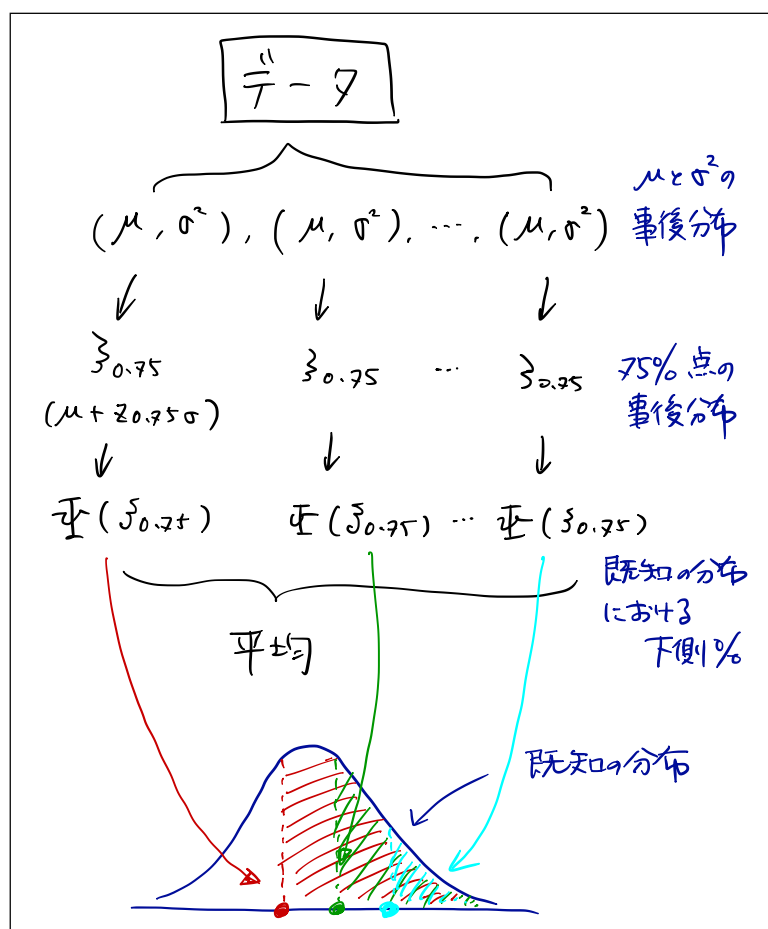


図 6.1 分位に関する推測

注 16 (本文 P.121) にあるように、 p の分散や ϵ 、 L は自動調節する必要がある。
第 6 章の章末問題を解くとわかるが、

```
data01 <- c(76230,73550,80750,71500,75420,74840,  
            71580,76920,68450,76990,64070,76200)
```

で $M = 73875$, $SD = 4442$ のようなデータだと、 μ や σ^2 の事後分布をリープフロッグで更新しようとしても、標準正規分布から発生させた乱数では 1 ステップあたりの遷移が (データの単位に比べて) 少なすぎるため更新がほとんど進まない。

7 さまざまな分布を用いた推測

7.1 ポアソン分布を用いた推測

満たしているべき条件

1. 稀少性：同時に 2 回以上起こらない
2. 独立性：事象の生起は過去の事象に依存しない
3. 定常性：単位時間内の平均観測数は常に一定

7.2 2 つのポアソン分布を用いた推測

2 つの母数 λ_A, λ_B を比較する。生成量

$$\delta^{(t)} = g(\lambda_A^{(t)}, \lambda_B^{(t)}) = \lambda_A^{(t)} - \lambda_B^{(t)} \quad (7.3)$$

$$u_{\delta>0}^{(t)} = g(\delta^{(t)}) = \begin{cases} 1 & \delta^{(t)} > 0 \\ 0 & \text{それ以外} \end{cases} \quad (7.4)$$

を算出すると、 $\delta^{(t)}$ の 95% 確信区間によって母数の差 δ に関する推測を*7、 $u_{\delta>0}^{(t)}$ の平均によって仮説 $\lambda_B > \lambda_A$ が正しい確率の評価を行うことができる。

*7 区間の下側が 0 より大きければ、97.5% の確率で $\lambda_B > \lambda_A$ であると言える。

7.3 指数分布を用いた推測

ポアソン分布に従う確率事象が初めて観測されるまでの時間が従う分布。

$$f(x|\lambda) = \lambda e^{-\lambda x}, \quad 0 \leq x, \quad 0 < \lambda \quad (7.5)$$

$$F(x|\lambda) = 1 - e^{-\lambda x} \quad (7.6)$$

7.4 ガンマ分布を用いた推測

独立に指数分布に従う α 個の確率変数の和を表す分布。つまり、 α 個のポアソン事象が発生するまでにかかる合計時間の分布。

単位時間当たりの観測個数がデータとして得られているとき、次に α 個観測するまでの時間は？

データ発生モデル 母数 λ のポアソン分布（観測機会 N 回）

1. データから λ の事後分布を算出 → 単位時間 $\times \alpha/\lambda$ の事後分布（＝待ち時間の平均事後分布）
2. データから λ の事後分布を算出 → α 個観測するまでの時間をガンマ分布に従って発生させ、その分布を算出（＝待ち時間の事後予測分布）

テキストでは、本文では後者、stan スクリプトでは前者の方法を採用している。前者の方法は、あくまで待ち時間の平均が従う分布に関する推測である。待ち時間自体が従う分布について推測したいのであれば、後者の方法をとる必要がある。

7.6 対数正規分布を用いた推測

母数 μ, σ^2 の対数正規分布に従う確率変数 x の対数を取ったもの $\log x$ は、母数 μ, σ^2 の正規分布に従う。預金額や収入の分布としてよく用いられる。

データ (単位時間あたり観測回数)

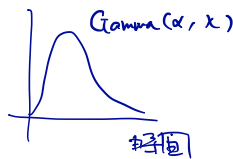
〇〇〇 〇

ポアソン分布に従っているといて、

単位時間あたり平均観測回数 λ を推定

↓
母数 α, λ の
ガンマ分布から

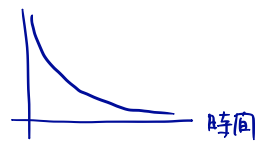
待ち時間と発生
(α 回起こるまでの時間 (単位))



↓
「 α 回起こるまでの
時間はどのような
分布に従う
か?」

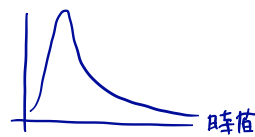
↓
 $1/\lambda$... 1 回起こるまでの待ち時間 (単位)

指数分布



↓
 α/λ ... α 回起こるまでの待ち時間 (単位)

ガンマ分布



↓
単位時間 $\times \alpha/\lambda$... α 回起こるまでの待ち時間 (意味)
「 α 回観測するまでの平均待ち時間
はどのような分布に従うか?」

※ 将来の値が従う分布と、その分布の母数が
従う分布とを区別しておかないと

図 7.1 予測分布と事後分布

8 比率・相関・信頼性

8.1 比率を用いた推測

コホート (前向き) 研究

要因への曝露を統制した上で、正反応の比率 (リスク) を比較する。

ケース・コントロール（症例対照・後ろ向き）研究

正反応が見られたサンプルと見られなかったサンプルとで、要因への曝露の有無を比較する。

リスク差 比率の差

リスク比 RR 比率の差は同じでも、比率によってその比は異なる。要因への曝露が、非曝露に対してリスクを何倍高めるか。

オッズ比 OR 前向き研究でも後ろ向き研究でも算出可能。リスクが極めて小さい場合は、リスク比とオッズ比は近い値をとる。

分割表の分析の場合、例えば曝露群と非曝露群における正反応の数をそれぞれ二項分布によってモデル化し、各群のリスクからリスク差、リスク比、オッズ比を生成量として算出すれば良い。

8.2 相関係数の差に関する推測

「自己・上司評価相関問題」で、転職群の相関が生え抜き群の相関よりも大きいか検証する。すると、stan を用いて以下のような結果が得られる。

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
rho_a	0.627	0.000	0.043	0.536	0.599	0.629	0.657	0.706	10000	1
rho_b	0.724	0.000	0.034	0.653	0.702	0.726	0.748	0.785	8848	1
diff_rho	0.097	0.001	0.055	-0.009	0.060	0.096	0.134	0.206	10000	1
p_over0	0.961	0.002	0.193	0.000	1.000	1.000	1.000	1.000	7969	1

このとき、

- 相関係数の差 `diff_rho` の 95% 確信区間は $[-0.009, 0.206]$ と下限がゼロを含んでいるため、B 群の相関と A 群の相関に差があるとは言いにくい。
- 一方、仮説 `diff_rho` が正しい確率 `p_over0` は 96.1% で、B 群の相関は A 群の相関より強いとかなりの確率で言える。

これらの結果は矛盾しているように思うが、どのように解釈すればよいのだろうか？
 確かな母数値の範囲と、母数値に関する仮説の確かさは別物。

8.4 切断データの相関係数に関する推測

切断データ ある基準値を超えると、もう一方の変数が必ず欠測するデータ

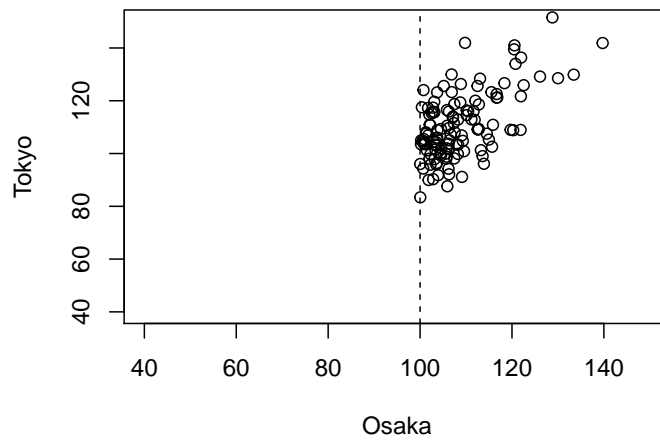


図 8.1 切断データ

ここで、

- y : 欠測を含む大阪東京双方のデータ
- x_1 : 大阪のみで販売された商品の（大阪での）データ
- x_2 : 東京でも販売された商品の大阪でのデータ

とすると、観測データ全体 $\{x, y\}$ の尤度は

$$\prod_{i=1}^{N_y} p(x_{2i}, y | \mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \sigma_{xy}) \prod_{i=1}^{N_{x_1}} p(x_{1j} | \mu_x, \sigma_x^2) \quad (8.14)$$

と表せる（ただし、欠測は x_1 にのみ依存する MAR）。この尤度を最大化する σ_{xy} は一致推定量となる。

尤度を最大化する σ_{xy} が一致推定量になったとしても、 ρ は結局 σ_x, σ_y で割るわけで、相関係数自体がよい推定量かどうかは別物ではないのか？

もし N_y セットのデータしか用いなければ、尤度は前半部分のみとなり、また母数は $\sigma_{x_2}^2, \sigma_{x_2y}$ とせざるを得ない。

ペアになっているデータは結局 y しかないのに、そこにペアを持っていない片方だけのデータを加えて尤度を構成することで、 $\hat{\rho}$ が変わるのは不思議ではある。

8.5 級内相関

級内相関 測定したい特性の分散が全分散に占める割合

Shrout & Fleiss (1979) I 人を J 人が評定するとして、

1. I 人それぞれ異なる J 人が評価する場合
2. 評価者がランダムサンプルと考えられる場合（変量効果モデル）
3. 評価者が最初から固定されている場合（混合効果モデル）

受験者 i を採点者 j が採点する。

$$x_{ij} = \mu_k + \alpha_{ki} + \beta_{kj} + e_{kij}, \quad e_{kij} \sim N(0, \sigma_{ek}^2) \quad (8.15)$$

- μ_k : 全平均
- α_{ki} : 受験者 i の効果 $\mu_{ki} - \mu_k, \alpha_k \sim N(0, \sigma_{\alpha k}^2)$
- β_{kj} : 評価者 j の効果 $\mu_{kj} - \mu_k, \beta_k \sim N(0, \sigma_{\beta k}^2)$ （変量効果モデルの場合）

このとき、分散構造は

変量効果モデル $\sigma_x^2 = \sigma_{\alpha k}^2 + \sigma_{\beta k}^2 + \sigma_{ek}^2$

混合効果モデル $\sigma_x^2 = \sigma_{\alpha k}^2 + \sigma_{ek}^2$

となる。

このとき、評定者が n 人のときの評定平均に関する級内相関は、それぞれ

$$ICC = \frac{\sigma_{\alpha k}^2}{\sigma_{\alpha k}^2 + (\sigma_{\beta k}^2 + \sigma_{ek}^2)/n} \quad (8.16)$$

$$ICC = \frac{\sigma_{\alpha k}^2}{\sigma_{\alpha k}^2 + \sigma_{ek}^2/n} \quad (8.17)$$

と定義される。後者は α 係数に一致。

付録 B 補足資料

B.1 収束判定指標 \hat{R}

Gelman & Rubin (1992); Gelman (1996) K 本のマルコフ連鎖 $\theta_k (= \{\theta_k^{(t)}\})$ が得られているとき、連鎖間分散 $\text{Var}_B(\theta)$ と連鎖内分散 $\text{Var}_W(\theta)$ を推定する。

仮に強い定常性が成り立っている場合、これらの重み付き和により $\text{Var}(\theta)$ の不偏推定量を構成することができる。

このとき、

$$\hat{R} = \sqrt{\frac{\widehat{\text{Var}}(\theta)}{\widehat{\text{Var}}_W(\theta)}} \quad (\text{付録 B.1})$$

により収束判定を行う。一般に、 $\hat{R} < 1.1$ ないし < 1.2 で収束したと判断する。

B.2 非効率性因子と Effective Sample Size

非効率性因子

系列内相関のある標本が無相関な標本から得られる平均と同じ精度を達成するにはその何倍の標本が必要であることを示す*⁸。系列内相関がある場合の標本平均の分散と無相関である場合の標本平均の分散との比を算出する。

また、この逆数を相対数値的効率性という (Geweke, 1992)。

Effective Sample Size

系列内相関のある標本のサンプルサイズが、無相関の標本ではいくつのサンプルサイズに相当するか評価する*⁹。

*⁸ 平均の分散は、系列内相関が強いほど大きくなる。

*⁹ 系列内相関が高いと MCMC の相対効率が下がる。