

Bayesian Data Analysis

Chapter 2: Single-parameter models

長島 健悟

城西大学

2009年4月22日, 2009年5月26日
2009年7月3日, 2009年10月30日

- 1.2 交換可能性 (Exchangeability)
 - 2.1 二項データに対する確率の推定 (p.33)
 - 2.2 データと事前情報の結合としての事後分布 (p.36)
 - 2.3 事後推定の要約 (p.37)
 - 2.4 情報事前分布 (p.39)
 - 2.5 例: 前置胎盤時の女性胎児の出生確率の推定 (p.43)
 - 2.6 母分散が既知の場合の正規分布の母平均の推定 (p.46)
 - 2.7 その他の1-パラメータモデル (p.49)
 - 2.8 例: がん死亡率推定に対する情報事前分布と多重階層構造 (p.55)
 - 2.9 無情報事前分布 (p.61)
 - 2.11 練習問題 (p.67)
- A.1 ラプラス近似 (付録)

交換可能の定義

- 交換可能性を満たす場合, n 個の確率変数 X_i ($i = 1, 2, \dots, n$) の同時確率 (密度) 関数について以下の関係が成り立つ.

$$p(x_1, x_2, \dots, x_n) = p(x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)}) \quad (1)$$

ただし, σ は集合 $\{1, 2, \dots, n\}$ に対する置換である.

- n 文字の置換 σ は全部で $n!$ 通り存在する.

確率変数 X_1, X_2, X_3 が 0 または 1 の値を取る場合

$$p(0, 0, 0) = p_{000}$$

$$p(1, 0, 0) = p(0, 1, 0) = p(0, 0, 1) = p_{100}$$

$$p(1, 1, 0) = p(1, 0, 1) = p(0, 1, 1) = p_{110}$$

$$p(1, 1, 1) = p_{111}$$

$$p_{000} + 3p_{100} + 3p_{110} + p_{111} = 1$$

$p(x_1), p(x_2), p(x_3)$ の積で同時分布を表現できなくてもよい

de Finetti の定理[†] (ベルヌーイ試行の場合)

確率変数 X_i ($i = 1, 2, \dots$) はそれぞれ 0 または 1 の値を取るとする. このとき, 交換可能な無限系列 $\{X_1, X_2, \dots, X_i, \dots, X_n, \dots, X_m, \dots\}$ において, n 個の確率変数の和を $Y = \sum_{i=1}^n X_i$ とすると,

$$\Pr(Y = y) = \int_0^1 \binom{n}{y} \theta^y (1 - \theta)^{n-y} p(\theta) d\theta$$

で, かつ確率 1 で,

$$\lim_{n \rightarrow \infty} \frac{Y}{n} = \theta$$

が成り立つ.

[†]渡部洋. ベイズ統計学入門^[11] の pp.98–100 による

de Finetti の定理の導出 1 (ベルヌーイ試行の場合)

- 交換可能な無限系列 $\{X_1, X_2, \dots, X_i, \dots, X_n, \dots, X_m, \dots\}$, $X_i = \{0, 1\}$ から, $p(x_1, x_2, \dots, x_n)$ や $p(x_1 + x_2 + \dots + x_n)$ を導く問題^[5,7]

m 個の交換可能な確率変数の和を $R = \sum_{i=1}^m X_i$ とおく. n 個 ($n \leq m$) までの交換可能な確率変数列を X_n と書く. X_n の同時確率は,

$$\begin{aligned} p(X_n = \mathbf{x}_n) &= \sum_{r=0}^m p(X_n = \mathbf{x}_n, R = r) \\ &= \sum_{r=0}^m p(X_n = \mathbf{x}_n \mid R = r)p(R = r) \end{aligned} \quad (2)$$

また, n 個の交換可能な確率変数の和 $Y = \sum_{i=1}^n X_i$ の確率関数は,

$$p(Y = y) = \binom{n}{y} p(X_n = \mathbf{x}_{\sigma(n)}) = \binom{n}{y} p(x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)}) \quad (3)$$

de Finetti の定理の導出 2 (ベルヌーイ試行の場合)

(2) 式, (3) 式より

$$\begin{aligned}
 p(X_n = \mathbf{x}_n) &= \sum_{r=y}^m \binom{n}{y}^{-1} p(Y = y | R = r)p(R = r) \\
 &= \sum_{r=y}^m \binom{n}{y}^{-1} \frac{\binom{m-n}{r-y} \binom{n}{y}}{\binom{m}{n}} p(R = r) = \sum_{r=y}^m \frac{r^y (m-r)^{n-y}}{m^n} p(R = r)
 \end{aligned} \tag{4}$$

ただし, $m^n = \frac{m!}{(m-n)!} = \prod_{i=0}^{n-1} (m-i)$ は下降階乗。

$p(Y = y | R = r)$ は, 「 m 個の確率変数のうち r 個が 1 であったという条件の下で m 個の中から n 個を取ってきたら y 個が 1 であった」という確率で, これは超幾何分布の確率関数である。

de Finetti の定理の導出 3 (ベルヌーイ試行の場合)

(4) 式について $\Theta = R/m$ という変換を考える. $\theta = r/m$ とおくと,

$$\begin{aligned}
 p(X_n = \mathbf{x}_n) &= \sum_{\theta=y/m}^1 \frac{(\theta m)^y \{(1-\theta)m\}^{n-y}}{m^n} p(R = \theta m) \\
 &= \sum_{\theta=y/m}^1 \frac{(\theta m)^y \{(1-\theta)m\}^{n-y}}{m^n} [p(R \leq \theta m) - p(R \leq (\theta - 1/m)m)]
 \end{aligned}
 \tag{5}$$

\sum_{θ} では y/m から 1 まで $1/m$ 区切りで和を求める. $p(R \leq \theta m)$ は確率変数 R の分布関数で, m ステップの階段関数であり,

$$Q_m(\theta) = p(R \leq \theta m) = p\left(\frac{R}{m} \leq \theta\right) = p(\Theta \leq \theta)$$

と書き換えることができる

de Finetti の定理の導出 4 (ベルヌーイ試行の場合)

(5) 式において, $m \rightarrow \infty$ のとき,

$$\begin{aligned} \lim_{m \rightarrow \infty} \sum_{\theta=y/m}^1 \frac{(\theta m)^y \{(1-\theta)m\}^{n-y}}{m^n} \{Q_m(\theta) - Q_m(\theta - 1/m)\} \\ = \int_0^1 \theta^y (1-\theta)^{n-y} dQ(\theta) = \int_0^1 \prod_{i=1}^n p(x_i | \theta) dQ(\theta) \end{aligned}$$

$m \rightarrow \infty$ においても分布関数 Q が存在する (必ず存在するわけではない)

$$\lim_{m \rightarrow \infty} Q_m(\theta) = Q(\theta)$$

Helly の選出定理 (Helly's theorem) により示す事ができる ([7] 等での証明).
下降階乗乗の極限を求めるために, スターリングの公式を用いた

$$\Gamma(x) \simeq \sqrt{2\pi} x^{x-\frac{1}{2}} \exp(-x), \quad \lim_{x \rightarrow \infty} \frac{\Gamma(x)}{\sqrt{2\pi} x^{x-\frac{1}{2}} \exp(-x)} = 1, \quad n! = n\Gamma(n)$$

de Finetti の定理 (ベルヌーイ試行の場合)

- 無限母集団からの抽出が仮定できるような場合, 交換可能性を満たすならば, i.i.d. と見なしたベイズ流の解析をしてもよい
分布関数 Q が階段関数か, 連続微分可能な場合ならば明らかに

$$p(x_1, x_2, \dots, x_n | \theta) = \prod_{i=1}^n p(x_i | \theta) = \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i}$$

- $E \left[\left(\frac{Y}{n} - \frac{R}{m} \right)^2 \right] \rightarrow 0, \theta = \lim_{n \rightarrow \infty} \frac{Y}{n}$ a.s.

マルチンゲール理論を使った証明^[8] など

- $\lim_{m \rightarrow \infty} Q_m = Q, \lim_{n \rightarrow \infty} \Pr \left(\frac{Y}{n} \leq \theta \right) = Q(\theta)$

Θ の分布 Q は, Y/n の極限 ($n \rightarrow \infty$) の分布でもある

- 交換可能性を満たしても i.i.d. にならない状態は, 現実ではどういう状況で, 何がうれしいのか? → 具体例は挙げにくい, 独立性より制約が緩いという事が重要?

リーマン・スチルチェス積分 [9, 10]

閉区間 $I = [a, b]$ を $a = x_0 < x_1 < \dots < x_n = b$ を分点とした n 個の小区間に分けることを分割 (partition), 各小区間の最大値 $\Delta = \max_{1 \leq i \leq n} (x_i - x_{i-1})$ を分割の幅 (mesh) と呼ぶ. 閉区間 I で有界な実関数 $f(x)$, $\varphi(x)$ について, 各小区間 $[x_{i-1}, x_i]$ 内に任意に点 ξ_i をとり, 和

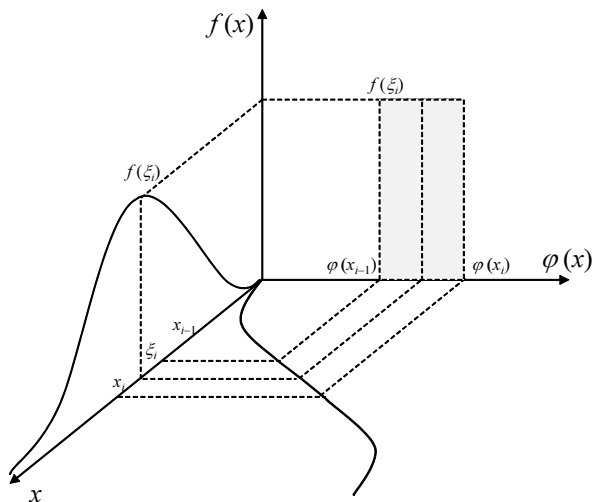
$$S = \sum_{i=1}^n f(\xi_i) \{\varphi(x_i) - \varphi(x_{i-1})\}$$

を考える. $\Delta \rightarrow 0$ ($n \rightarrow \infty$) において, 分点のとり方や ξ_i の取り方によらずに一定の極限

$$J = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(\xi_i) \{\varphi(x_i) - \varphi(x_{i-1})\} = \int_a^b f(x) d\varphi(x) \in \mathbb{R}$$

が存在するとき, J を f の φ に関する I 上のリーマン・スチルチェス積分とよぶ. リーマン・スチルチェス積分が存在するための十分条件は $f(x)$ が連続で, $g(x)$ が有界変動であることである.

リーマン・スチルチェス積分



$\varphi(x)$ が微分可能な場合と階段関数の場合

- $\varphi(x)$ が微分可能な場合

$$S = \sum_{i=1}^n f(\xi_i) \{\varphi(x_i) - \varphi(x_{i-1})\} = \sum_{i=1}^n f(\xi_i) \frac{\{\varphi(x_i) - \varphi(x_{i-1})\}}{(x_i - x_{i-1})} (x_i - x_{i-1})$$

$$\xrightarrow{n \rightarrow \infty} \int_a^b f(x) \varphi'(x) dx$$

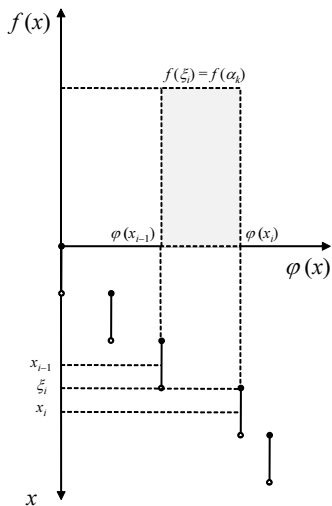
$\varphi(x) = x$ ならば $\varphi'(x) = 1$ であり、リーマン積分に一致

- K ステップの階段関数の場合

$$\int_a^b f(x) d\varphi(x) = \sum_{i=1}^n f(\xi_i) \delta_i = \sum_{k=1}^K f(\alpha_k) c_k$$

$$\delta_i = \varphi(x_i) - \varphi(x_{i-1}) = \begin{cases} c_k & (\varphi(x_{i-1}) < \varphi(\alpha_k) < \varphi(x_i)) \\ 0 & (\varphi(x_{i-1}) = \varphi(x_i)) \end{cases}$$

階段関数の場合



Chapter 2: Single-parameter models

- パラメータ θ がスカラー (1 次元) のモデル
- 二項モデル, 正規モデル, ポアソンモデル, 指数モデル
 - $p(y | \theta)$ がそれぞれ二項分布, 正規分布, ...
- ベイズ流データ解析の枠組みと, 数値計算法

二項データに対する確率の推定 (p.33)

- Bayes (1763)^[2] で扱われたシンプルだが重要な問題
- n 回のベルヌーイ試行の反復結果 $y_1, y_2, \dots, y_i, \dots, y_n$ ($y_i \in \{0, 1\}$) から試行の成功確率を表わすパラメータ θ を推測
 - ① 1 回の試行の結果は 2 通りである
 - ② それぞれの試行は独立である
 - ③ 成功確率 θ は繰り返しを通じて同一である
- 交換可能性を満たす n 回の反復においても, n 回中に成功した回数 y を使ってデータを要約できる, 結局は尤度が二項分布 → 二項モデル

$$p(y | \theta) = \text{Bin}(y | n, \theta) = \binom{n}{y} \theta^y (1 - \theta)^{n-y}$$

- n は試験デザインの一部と考えられるため, 以降固定されているものとして考える

二項モデルのベイズ推定 (p.34)

- $p(\theta)$ の設定: 今回は一様分布 $p(\theta) = U(0, 1)$

$$p(\theta | y) = \frac{p(\theta)p(y | \theta)}{p(y)}$$

$$p(\theta | y) \propto \theta^y (1 - \theta)^{n-y}$$

- 基準化定数を省略しない場合は

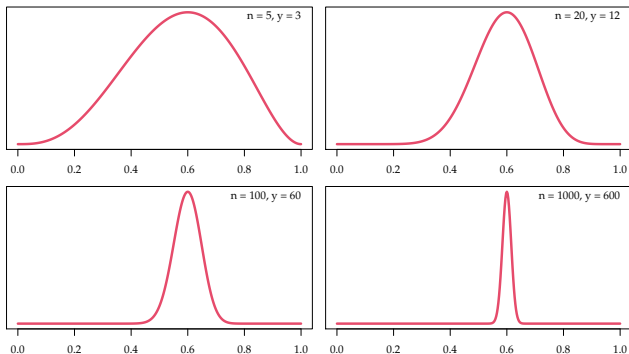
$$p(\theta | y) = \frac{\binom{n}{y} \theta^y (1 - \theta)^{n-y}}{\int_0^1 \binom{n}{y} \theta^y (1 - \theta)^{n-y} d\theta} = (n + 1) \binom{n}{y} \theta^y (1 - \theta)^{n-y}$$

$$= \frac{\Gamma(n + 2)}{\Gamma(y + 1) \Gamma(n - y + 1)} \theta^y (1 - \theta)^{n-y}$$

$$= \frac{\theta^y (1 - \theta)^{n-y}}{B(y + 1, n - y + 1)} = \text{Beta}(y + 1, n - y + 1)$$

二項モデルのベイズ推定 (p.34)

- 他にもいくつかの典型的なモデルも $p(\theta | y)$ を閉形式で書ける
 - 事後モードぐらいなら簡単に計算できるように...(傾きが0になる所を求める)
 - 事前分布が定数の場合, 事後モードは最尤推定量と等しい
- パラメータが1次元なら図示しやすい (下図は $\text{Beta}(y + 1, n - y + 1)$)



分母の計算 (p.35)

$$\begin{aligned}
 \binom{n}{y} \int_0^1 \theta^y (1 - \theta)^{n-y} d\theta &= \binom{n}{y} B(y+1, n-y+1) \\
 &= \binom{n}{y} \left\{ \left[\frac{\theta^{y+1}}{y+1} (1 - \theta)^{n-y} \right]_0^1 + \frac{(n-y)}{(y+1)} \int_0^1 \theta^{y+1} (1 - \theta)^{n-y-1} d\theta \right\} \\
 &= 0 + \binom{n}{y} \frac{(n-y)}{(y+1)} \int_0^1 \theta^{y+1} (1 - \theta)^{n-y-1} d\theta \\
 &= \dots \\
 &= \binom{n}{y} \frac{(n-y)!}{n^{n-y}} \int_0^1 \theta^n d\theta = \binom{n}{y} \frac{(n-y)!}{n^{n-y}} \frac{1}{n+1} \\
 &= \frac{n(n-1) \cdots (y+1)y \cdots 1}{y!(n-y)!} \frac{(n-y)!}{n(n-1) \cdots (y+1)} \frac{1}{n+1} = \frac{1}{n+1}
 \end{aligned}$$

予測 (p.36)

- n 回試行を行って y 回成功した, というデータが得られたもとの
- さらにもう 1 回試行を行ったとき (その結果を \tilde{y} とおく), 成功する確率を予測したい

$$\Pr(\tilde{y} = 1 | y) = \int_0^1 \Pr(\tilde{y} = 1 | \theta, y) p(\theta | y) d\theta$$

- ベルヌーイ試行を 1 回分追加しただけなので,

$$\Pr(\tilde{y} = 1 | \theta, y) = \Pr(\tilde{y} = 1 | \theta) = \binom{1}{1} \theta^1 (1 - \theta)^0 = \theta$$

$$\Pr(\tilde{y} = 1 | y) = \int_0^1 \theta p(\theta | y) d\theta = E(\theta | y) = \frac{y + 1}{n + 2}$$

歴史的な流れ (p.36)

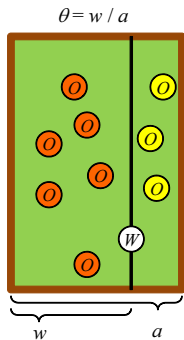
- 17-18 世紀前半は $p(y | \theta)$ に関する議論が中心
- 後に Bayes (1763) と Laplace (1774) が独立に**ベイズの定理**を示した

$$p(\theta \in (\theta_1, \theta_2) | y) = \frac{\int_{\theta_1}^{\theta_2} \binom{n}{y} \theta^y (1 - \theta)^{n-y} d\theta}{\int_0^1 \binom{n}{y} \theta^y (1 - \theta)^{n-y} d\theta} \quad (6)$$

- 当時は (6) 式の分子, 不完全ベータ関数の計算が問題
- Laplace (1786) では, パリで 1745 年-1770 年の間に生まれた女兒が 241,945 人, 男児が 251,527 人というデータを与えたもとで, 女兒が生まれる確率 θ が 0.5 よりも大きくなる確率を求めている

$$\Pr(\theta > 0.5 | y = 241945, n = 493472) = 1.15 \times 10^{-42}$$

- 後半 (p.340-) で出てくる Laplace 近似もここで示された



事前・事後分布の平均に関する性質 (p.36)

- $p(\theta)$ と $p(\theta | y)$ について言える一般的な関係

$$E(\theta) = E(E(\theta | y)) = \int E(\theta | y)p(y) dy \quad (7)$$

- 事前分布の平均 $E(\theta)$ と、事後分布の平均 $E(\theta | y)$ の (取り得るデータ全てに対する) 期待値が等しい
- θ の事前分布が $U(0, 1)$, 尤度が $\text{Bin}(y | \theta, n)$ の場合

$$E(\theta) = \frac{1}{2}, \quad E(E(\theta | y)) = \sum_{y=0}^n \left(\frac{y+1}{n+2} \right) \frac{1}{n+1} = \frac{1}{2}$$

- 事後平均 $E(\theta | y)$ にデータが得られる確率 $p(y)$ をかけて均す

事前・事後分布の分散に関する性質 (p.36)

- $p(\theta)$ と $p(\theta | y)$ について言える一般的な関係

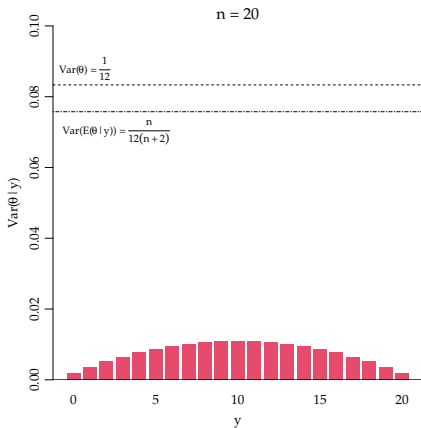
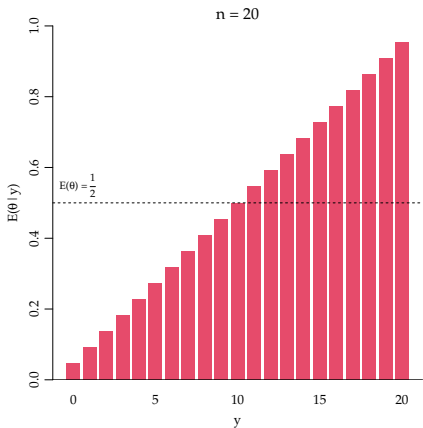
$$\text{Var}(\theta) = \text{E}(\text{Var}(\theta | y)) + \text{Var}(\text{E}(\theta | y)) \quad (8)$$

- 事後分布の分散 $\text{Var}(\theta | y)$ は、平均的にみて事前分布の分散 $\text{Var}(\theta)$ より小さい
- 事後平均の分散 $\text{Var}(\text{E}(\theta | y))$ の大きさに依存する
- θ の事前分布が $U(0, 1)$, 尤度が $\text{Bin}(y | \theta, n)$ の場合

$$\text{Var}(\theta) = \frac{1}{12}, \quad \text{E}(\text{Var}(\theta | y)) = \sum_{y=0}^n \left\{ \frac{(n-y+1)(y+1)}{(n+2)^2(n+3)} \right\} \frac{1}{n+1} = \frac{1}{6(n+2)}$$

$$\text{Var}(\text{E}(\theta | y)) = \sum_{y=0}^n \left(\frac{y+1}{n+2} - \frac{1}{2} \right)^2 \frac{1}{n+1} = \frac{n}{12(n+2)} \quad (9)$$

事後分布の平均と事後分布の分散 (p.37)



データと事前情報の結合としての事後分布 (p.37)

- (7) 式, (8) 式はあくまで期待値に関して言えること
- モデルを誤特定 (尤度と事前分布) したりすると成り立たない事がある
- $\frac{y+1}{n+2}$ (事後平均) = $\frac{1}{2}$ (事前平均) + $\frac{y}{n}$ (データ)
- サンプルサイズが増えるにしたがって事前分布の影響は, データの影響に比べて小さくなってゆく
- 事前情報も組み込めるが, データによるところが非常に大きい
- compromise: 妥協, 歩み寄り → 結合?

事後推定の要約 (p.37)

- 事後分布全体を結果として示す事も理想的な方法
例えば, グラフを描く (分布図・等高線図・散布図)
- ベイズ的アプローチ (シミュレーションによる実装) の重要な利点は, 複雑な変換を行った後でも事後推測を要約できる柔軟性
- 位置の要約
 - 平均値, 中央値, 最頻値
- ばらつきの要約
 - 標準偏差, 四分位範囲
- それぞれ解釈が異なる
 - モデルの仮定の下で...
 - 平均値: パラメータの事後分布の期待値
 - 中央値: パラメータの事後分布のもっとも起こりやすい値

事後推定の要約 (p.38)

- 正規近似して平均値と標準偏差を使う事も多々ある
- θ の分布の近似が悪そうな場合は, 分布が対称になるように変換したり, 工夫する
- θ の事前分布が $U(0, 1)$, 尤度が $\text{Bin}(y | \theta, n)$ の場合

$$(\text{事後平均}) = E(\theta | y) = \frac{y + 1}{n + 2}$$

$$(\text{事後モード}) = \operatorname{argmax} p(\theta | y) = \operatorname{argmax} \log p(\theta | y) = \frac{y}{n}$$

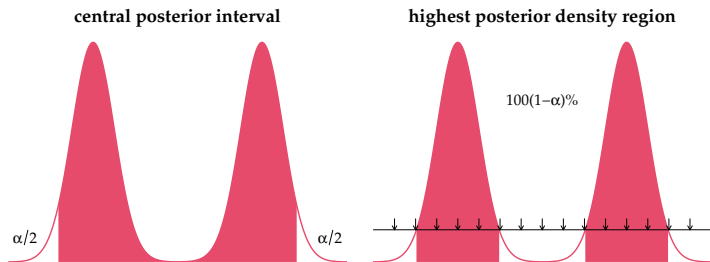
$$\log p(\theta | y) = y \log \theta + (n - y) \log(1 - \theta) - \log B(y + 1, n - y + 1)$$

$$\frac{d \log p(\theta | y)}{d\theta} = \frac{y}{\theta} - \frac{n - y}{1 - \theta}, \quad \frac{y}{\hat{\theta}} - \frac{n - y}{1 - \hat{\theta}} = 0, \quad \hat{\theta} = \frac{y}{n}$$

事後分布の分位点と区間 (p.38)

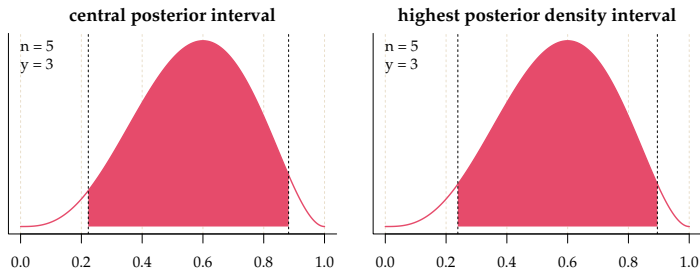
- 信用区間 (central posterior interval)
 - 事後分布について, 分布の下限から区間の下限まで, 区間の上限から分布の上限までの積分は $100(\alpha/2)\%$
- 最高事後密度領域 (highest posterior density region)
 - 事後分布の $100(1 - \alpha)\%$ を占め, 領域内の密度は領域外の密度より小さくならない
 - 事後分布が単峰で対称な分布なら信用区間と一致
- 信用区間の方が好まれる
 - 事後分布の分位点を使って直接解釈でき, 推定対称に一对一変換を行っても不変で, 計算がラク

事後分布の分位点と区間 (p.38)



- 信用区間は確率が非常に低い領域 (真ん中) を含む
- この例では区間が 1 つになってしまう信用区間は望ましくない
 - 事後分布が多峰性の分布 \Rightarrow 尤度か事前分布が多峰性の分布
- 事後分布の非対称性が強い場合も大きく違う

2つの区間の数値例 (p.38)



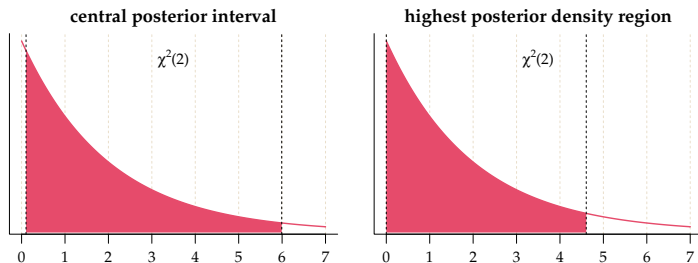
- 信用区間

$$\Pr(\theta \leq \theta_L | y) = \alpha/2, \quad \Pr(\theta \geq \theta_U | y) = \alpha/2, \quad [0.22, 0.88] \quad (10)$$

- 最高事後密度区間

$$\Pr(\theta'_L | y) = \Pr(\theta'_U | y), \quad \Pr(\theta'_L \leq \theta \leq \theta'_U | y) = 1 - \alpha, \quad [0.24, 0.90] \quad (11)$$

その他の例



- 非対称性が強く、

$$\Pr(\theta'_L | y) = \Pr(\theta'_U | y)$$

を満たす (θ_L, θ_U) が見つからない場合は最高時後密度区間の求め方が違う

情報事前分布 (p.39)

- 事前分布の解釈
 - 母集団に基づく解釈
事前分布は、興味のあるパラメータ θ の母集団分布 (取り得る値)
 - 持っている知識に基づく解釈
 θ に関する私たちが持つ知識 (や不確かさ) を, θ が事前分布からのランダムな実現値だと見なして表現する
- とは言ったものの, 新規に計画する試験など, 多くの場合において仮説以外には θ の母集団を決めるのは難しい
- 事前分布は, θ の値としてもっともらしい値を全て含むべきだが, どんなに適切な事前分布を定めても結局はデータに含まれる θ の情報が遥かに勝る事になるので, 真値の周りに集中している必要はない
- 二項モデルで事前分布として $U(0, 1)$ を用いた例
 - 一応正当化できる
 - Laplace の理由不十分の原理 (principle of insufficient reason)
 - Section 2.9 の無情報事前分布で軽く触れる

二項モデルに別の事前分布を設定した例 (p.40)

- 事後分布が数学的に扱い易くなるように, 事前分布を決めたい
- 尤度が

$$p(y | \theta) \propto \theta^y (1 - \theta)^{n-y}$$

なので, 事前分布も同じ形の式にすると, 事後分布も同じ形になる

$$p(\theta) \propto \theta^{\alpha-1} (1 - \theta)^{\beta-1}$$

- α は成功, β は失敗に関するパラメータ[†], 事前分布のパラメータをハイパーパラメータと呼ぶ
- 事後分布は

$$\begin{aligned} p(\theta | y) &\propto \theta^y (1 - \theta)^{n-y} \theta^{\alpha-1} (1 - \theta)^{\beta-1} \\ &\propto \theta^{y+\alpha-1} (1 - \theta)^{n-y+\beta-1} \end{aligned}$$

[†]適切な α, β の値の与え方については Chapter 5

事前分布がベータ分布の時の二項モデル (p.40)

$$p(y) = \int_0^1 \frac{\binom{n}{y} \theta^{y+\alpha-1} (1-\theta)^{n-y+\beta-1}}{B(\alpha, \beta)} d\theta = \frac{\binom{n}{y} B(\alpha+y, \beta+n-y)}{B(\alpha, \beta)}$$

$$p(\theta | y) = \frac{\binom{n}{y} \theta^{y+\alpha-1} (1-\theta)^{n-y+\beta-1}}{B(\alpha, \beta)} \bigg/ \frac{\binom{n}{y} B(\alpha+y, \beta+n-y)}{B(\alpha, \beta)}$$

$$= \text{Beta}(\theta | \alpha+y, \beta+n-y)$$

$$E(\theta | y) = \int_0^1 \theta \frac{\theta^{y+\alpha-1} (1-\theta)^{n-y+\beta-1}}{B(\alpha+y, \beta+n-y)} d\theta = \frac{B(\alpha+y+1, \beta+n-y)}{B(\alpha+y, \beta+n-y)}$$

$$= \frac{\alpha+y}{\alpha+\beta+n}$$

$$\text{Var}(\theta | y) = E(\theta^2 | y) - \{E(\theta | y)\}^2 = \frac{(\alpha+y+1)(\alpha+y)}{(\alpha+\beta+n+1)(\alpha+\beta+n)} - \left(\frac{\alpha+y}{\alpha+\beta+n}\right)^2$$

$$= \frac{(\alpha+y)(\beta+n-y)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)} = \frac{E(\theta | y)\{1 - E(\theta | y)\}}{\alpha+\beta+n+1}$$

事前分布がベータ分布の時の二項モデル (p.41)

- $y, n - y$ が充分大きいとき

$$E(\theta | y) = \frac{y}{n}, \quad \text{Var}(\theta | y) = \frac{1}{n} \frac{y}{n} \left(1 - \frac{y}{n}\right)$$

に近づいていく事が分かる (詳しくは Chapter 4 で中心極限定理を...)

- また,

$$\left(\frac{\theta - E(\theta | y)}{\sqrt{\text{Var}(\theta | y)}} \mid y \right) \rightarrow N(0, 1)$$

から, 事後分布を正規分布で近似することも良く行われる

- $\theta \in [0, 1]$ なので, logit 変換 $\log\{\theta/(1 - \theta)\} \in (-\infty, \infty)$ を行うと正規分布への近似が良くなる

共役事前分布・非共役事前分布 (p.41)

共役事前分布

- 事後分布が事前分布と同じ分布族に属する
- メリット
 - パラメータの解析解, 良いアルゴリズム, 良い近似
 - データを追加した際にモデルを書き直さなくてよい
- 共役事前分布族が尤度関数を含む分布族の場合は 自然共役事前分布
 - 尤度関数をパラメータの関数としてみると, 共役事前分布族に含まれる
 - $\binom{n}{y}\theta^y(1-\theta)^{n-y}$ も θ の関数として見ればベータ分布に属する

非共役事前分布

- 共役ではない情報事前分布は非共役事前分布
- その場限りの対応が多く, 計算が複雑になる
- 共役事前分布を試すか, 混合共役事前分布を試してからの方がよい

指数型分布族と十分統計量 (p.42)

n 個の独立な確率変数 Y がパラメータ θ を持つ指数型分布に従うとすると、尤度は以下のように書く事ができる

$$p(y_i | \theta) = f(y_i)g(\theta) \exp \left\{ \sum_{j=1}^k \phi_j(\theta)h_j(y_i) \right\}, \quad y_i \in E$$
$$g(\theta)^{-1} = \int_E f(y_i) \exp \left\{ \sum_{j=1}^k \phi_j(\theta)h_j(y_i) \right\} dy_i \quad (12)$$

$$p(\mathbf{y} | \theta) = \left[\prod_{i=1}^n f(y_i) \right] g(\theta)^n \exp \left\{ \sum_{j=1}^k \phi_j(\theta) \left(\sum_{i=1}^n h_j(y_i) \right) \right\}$$

$\phi_j(\theta)$ は自然パラメータ (k 次元) とよばれ, $t_j(\mathbf{y})$ は $\phi_j(\theta)$ の十分統計量になる

$$t_j(\mathbf{y}) = \sum_{i=1}^n h_j(y_i)$$

指数型分布族と共役事前分布 (p.42)

よって、事前分布を

$$p(\boldsymbol{\theta}) \propto g(\boldsymbol{\theta})^\eta \exp \left\{ \sum_{j=1}^k \phi_j(\boldsymbol{\theta}) v_j \right\}$$

とおけば、事後分布は

$$p(\boldsymbol{\theta} | \mathbf{y}) = K(\mathbf{v})^{-1} g(\boldsymbol{\theta})^{n+\eta} \exp \left\{ \sum_{j=1}^k \phi_j(\boldsymbol{\theta}) (v_j + t_j(\mathbf{y})) \right\}, \quad \boldsymbol{\theta} \in \Theta$$

$$K(\mathbf{v}) = \int_{\Theta} g(\boldsymbol{\theta})^{n+\eta} \exp \left\{ \sum_{j=1}^k \phi_j(\boldsymbol{\theta}) (v_j + t_j(\mathbf{y})) \right\} d\boldsymbol{\theta}$$

のように書くことができるので、尤度が指数型分布族の場合、自然共役事前分布が必ず存在する

二項モデルの例 (p.43)

尤度が二項分布 $\text{Bin}(y | \theta)$ のとき、自然パラメータは $\text{logit}(\theta)$ で、十分統計量は y

$$p(y | \theta) = \binom{n}{y} (1 - \theta)^n \exp \left\{ \log \left(\frac{\theta}{1 - \theta} \right) y \right\}$$
$$(1 - \theta)^{-n} = \sum_{y=0}^n \binom{n}{y} \exp \left\{ \log \left(\frac{\theta}{1 - \theta} \right) y \right\}$$

事前分布をベータ分布 $\text{Beta}(\theta | \alpha, \beta)$ とすると

$$p(\theta) = \frac{1}{B(\alpha, \beta)} (1 - \theta)^{\alpha-1} \theta^{\beta-1} \exp \left\{ \log \left(\frac{\theta}{1 - \theta} \right) (\alpha - 1) \right\} \quad (13)$$

$$p(\theta | y) = \frac{1}{B(\alpha + y, \beta + n - y)} (1 - \theta)^{n + \alpha - 1} \theta^{\beta - 1} \exp \left\{ \log \left(\frac{\theta}{1 - \theta} \right) (y + \alpha - 1) \right\}$$

例: 前置胎盤時の女性胎児の出生確率の推定 (p.43)

前置胎盤と性比の関連を調べたい

- ドイツで行われた研究において、980 人中 437 人が女兒という結果が得られた
- はたして、前置胎盤時の女性胎児の出生確率は 0.485(一般の集団の女性胎児の出生確率) 未満なのか?

前置胎盤[†]

- 胎盤は普通は子宮の上部に形成されますが、ときに子宮の下の方に位置する子宮頸部(子宮口)上やその近くにできることがあります。これを前置胎盤といい、胎盤が完全に子宮口をふさいでいるものから一部だけふさいでいるものまであります。前置胎盤は約 200 回の出産に 1 回の割合で起こり、経産婦や、子宮筋腫など子宮の異常がある人に起こりやすい傾向があります。

[†]メルクマニュアル (Web 版) より; <http://mmh.banyu.co.jp/mmhe2j/>

事前分布に一様分布を用いた解析; 解析解を使う (p.43)

- 尤度は $\text{Bin}(Y = 437 | \theta, n = 980)$, 事前分布は一様分布で $\text{Beta}(1, 1)$

$$p(\theta | y) = \text{Beta}(438, 544)$$

$$E(\theta | y) = \frac{y + 1}{n + 2} = 0.446, \quad \sqrt{\text{Var}(\theta | y)} = \sqrt{\frac{(n - y + 1)(y + 1)}{(n + 2)^2(n + 3)}} = 0.016$$

$$\text{central 95\% posterior interval} = [0.415, 0.477]$$

信用区間は正規近似しても十分な精度であった

事前分布に一様分布を用いた解析; シミュレーション (p.43)

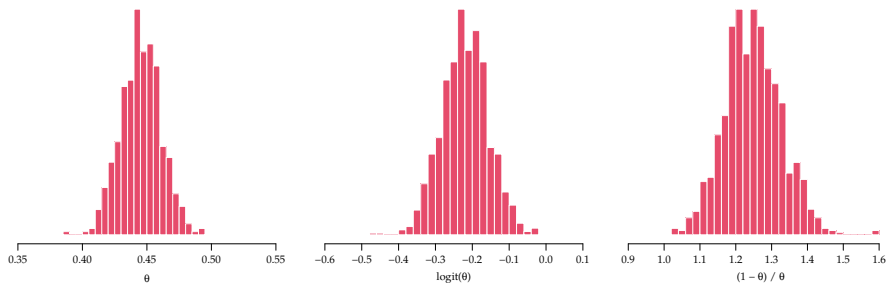
- 事後分布を $\text{Beta}(438, 544)$ として, 1000 回乱数を発生させた結果を使って事後分布を要約する

```
set.seed(8769453)
theta <- rbeta(1000, 438, 544)
quantile(theta, c(0.025, 0.5,
                 0.975))
  2.5%
0.4159787
  50%
0.4452324
  97.5%
0.4760559
```

```
mean(theta)
[1] 0.4454034
sd(theta)
[1] 0.01556588
mean(theta)+qnorm(0.025)*sd(theta)
[1] 0.4148949
mean(theta)+qnorm(0.975)*sd(theta)
[1] 0.475912
```

- 解析解にかなり近い結果になった

事前分布に一樣分布を用いた解析; ヒストグラム (p.44)



median と central 95% posterior interval を求める

- そのままの θ では, 0.445 [0.415, 0.476]
- $\text{logit}(\theta)$ で変換した場合, 正規近似して逆変換すると 0.445 [0.415, 0.476] (元々のスケールで近似が良いので差が分からない)
- 性比 $(1 - \theta)/\theta$ で変換した場合, 1.25 [1.10, 1.40] (一般的なヨーロッパ人の性比は 1.06)

共役事前分布を用いた解析 (p.45)

- 事前分布と事後分布

$$p(\theta) = \text{Beta}(\theta | \alpha, \beta), \quad p(\theta | y) = \text{Beta}(\theta | \alpha + y, \beta + n - y)$$

$\frac{\alpha}{\alpha + \beta}$	$\alpha + \beta$	posterior median	central 95% posterior interval
0.500	2	0.466	[0.415, 0.477]
0.485	2	0.466	[0.415, 0.477]
0.485	5	0.466	[0.415, 0.477]
0.485	10	0.466	[0.415, 0.477]
0.485	20	0.447	[0.416, 0.478]
0.485	100	0.450	[0.420, 0.479]
0.485	200	0.453	[0.424, 0.481]

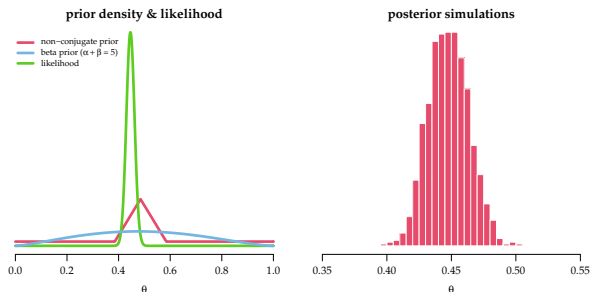
- 1 列目は事前分布の期待値
- 2 列目は事前分布の情報量; (13) 式
- 0.485 は一般的集団の女性胎児の出生確率
- 事前分布の情報量が 100, 200 人相当でも信用区間は 0.485 を含まなかった

非共役事前分布を用いた解析 (p.45)

- 区間 $[0.385, 0.585]$ に含まれない部分の確率質量が 40%, 期待値が 0.493, 標準偏差が 0.21 になるような事前分布

$$p(\theta) = 0.5 \times U(\theta | 0, 1) + 0.5 \times \text{Triangle}(\theta | 0.385, 0.585, 0.485)$$

- θ を $(0.000, 0.001, \dots, 1.000)$ と細かく分割し, 二項尤度との積を取って事後分布を近似する. 近似した事後分布から 1000 回サンプリングし, 中央値と 95%信用区間を求めた, 0.448 $[0.419, 0.480]$



非共役事前分布を用いた解析 (p.45)

- $\alpha + \beta = 5$ のベータ分布に近い形状をしているので似たような結果に
- シミュレーション誤差??

```
cdf <- function(x) {
  a <- 0.385; b <- 0.585; c <- (a + b) / 2
  if (0 > x) return(0)
  else if (0 <= x & x <= a) return(0.5 * x)
  else if (a < x & x <= c) return(0.5 * x + 0.5 * (x - a)^2 / ((b - a) * (c - a)))
  else if (c < x & x <= b) return(0.5 * x + 0.5 * (1 - (b - x)^2 / ((b - a) * (b - c))))
  else if (b <= x & x <= 1) return(0.5 * x + 0.5)
  else if (1 > x) return(1)
}

set.seed(11451355); y <- NULL; b <- NULL; S <- seq(0, 1, 0.001)
for (i in S) {
  y <- append(y, cdf(i) - cdf(i - 0.001))
  b <- append(b, dbinom(437, 980, i))
}
z <- sample(S, 1000, replace = TRUE, prob = y*b/sum(y*b))

par(family="serif")
hist(
  z, breaks = 25, col = "#E64B6B", border = "white",
  yaxt = "n", xlim = c(0.35, 0.55), ylab = NA,
  main = "posterior simulations", xlab = expression(theta)
)
quantile(z, c(0.5, 0.025, 0.975))
```

母平均が未知・母分散が既知の正規モデル (p.46)

まずはデータが一つの場合を考え、 $y \sim N(\theta, \sigma^2)$ とする。ただし、 σ^2 が既知、 θ は未知とする。

$$p(y | \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{y^2}{2\sigma^2}\right) \exp\left(-\frac{\theta^2}{2\sigma^2}\right) \exp\left(\frac{y\theta}{\sigma^2}\right)$$

事前分布は $N(\theta | \mu_0, \tau_0^2)$

$$p(\theta) \propto \exp\left(-\frac{\theta^2}{2\tau_0^2}\right) \exp\left(\frac{\mu_0\theta}{\tau_0^2}\right)$$

事後分布は

$$p(\theta | y) \propto \exp\left\{-\frac{\theta^2}{2} \left(\frac{1}{\sigma^2} + \frac{1}{\tau_0^2}\right)\right\} \exp\left\{\theta \left(\frac{y}{\sigma^2} + \frac{\mu_0}{\tau_0^2}\right)\right\} \quad (14)$$

中心極限定理を使って正規近似できる場合は扱いやすい。正規分布の当てはまりが良くない場合は、 t 分布や有限混合分布を使う、後の Chapter で解説...

母平均が未知・母分散が既知の正規モデル (p.47)

(14) 式について,

$$\mu_1 = \frac{\frac{1}{\sigma^2}y + \frac{1}{\tau_0^2}\mu_0}{\frac{1}{\sigma^2} + \frac{1}{\tau_0^2}}, \quad \frac{1}{\tau_1^2} = \frac{1}{\sigma^2} + \frac{1}{\tau_0^2} \quad (15)$$

とおけば $N(\theta | \mu_1, \tau_1^2)$

- 事後分布の母平均 μ_1 は, y (データ) と μ_0 (事前分布の母平均) に精度 (分散の逆数) で重みを付けた重み付き平均
- 事後分布の精度 $\frac{1}{\tau_1^2}$ は, $\frac{1}{\sigma^2}$ (データの精度) と $\frac{1}{\tau_0^2}$ (事前分布の精度) の和
- $\frac{1}{\sigma^2} \gg \frac{1}{\tau_0^2}$ のとき $\mu_1 \simeq y$, $\frac{1}{\sigma^2} \ll \frac{1}{\tau_0^2}$ のとき $\mu_1 \simeq \mu_0$
- $y = \mu_0$ のときは $\mu_1 = y = \mu_0$

事後予測分布 (p.48)

データ $\tilde{y} \sim N(\theta, \sigma^2)$ を追加した場合の事後予測分布 $p(\tilde{y} | y)$ は,

$$\begin{aligned}
 p(\tilde{y} | y) &= \int_{-\infty}^{\infty} p(\tilde{y} | \theta) p(\theta | y) d\theta \\
 &\propto \exp\left\{-\frac{\tilde{y}^2}{2\sigma^2}\right\} \int_{-\infty}^{\infty} \exp\left\{-\frac{\theta^2}{2}\left(\frac{1}{\sigma^2} + \frac{1}{\tau_1^2}\right)\right\} \exp\left\{\theta\left(\frac{\tilde{y}}{\sigma^2} + \frac{\mu_1}{\tau_1^2}\right)\right\} d\theta \\
 &\propto \frac{\exp\left\{-\frac{\tilde{y}^2}{2\sigma^2}\right\}}{\exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2} + \frac{1}{\tau_1^2}\right)\left(\frac{\frac{1}{\sigma^2}\tilde{y} + \frac{1}{\tau_1^2}\mu_1}{\frac{1}{\sigma^2} + \frac{1}{\tau_1^2}}\right)^2\right\}} = \exp\left\{-\frac{\tilde{y}^2 - 2\mu_1\tilde{y} - \frac{\mu_1^2\sigma^2}{\tau_1^2}}{2(\sigma^2 + \tau_1^2)}\right\}
 \end{aligned} \tag{16}$$

これも正規分布 $N(\tilde{y} | \mu_1, \sigma^2 + \tau_1^2)$ にしたがう

事後予測分布 (p.48)

事後予測分布を直接求めたが, $E(\tilde{y} | \theta) = \theta$, $\text{Var}(\tilde{y} | \theta) = \sigma^2$ が既知なので, 事後予測分布の期待値は, 以下で求めると簡単

$$E(\tilde{y} | y) = E(E(\tilde{y} | \theta, y) | y) = E(\theta | y) = \mu_1$$

$$\begin{aligned}\text{Var}(\tilde{y} | y) &= E(\text{Var}(\tilde{y} | \theta, y) | y) + \text{Var}(E(\tilde{y} | \theta, y) | y) \\ &= E(\sigma^2 | y) + \text{Var}(\theta | y) \\ &= \sigma^2 + \tau_1^2\end{aligned}$$

データが n 個の場合 (p.48)

- データが n 個, $\mathbf{y} = (y_1, y_2, \dots, y_n)$ の場合に拡張する
 n 個の観測値が i.i.d. の場合

$$\begin{aligned} p(\theta | \mathbf{y}) &= p(\theta) \prod_{i=1}^n p(y_i | \theta) \\ &\propto \exp\left(-\frac{\theta^2}{2\tau_0^2}\right) \exp\left(\frac{\mu_0\theta}{\tau_0^2}\right) \prod_{i=1}^n \exp\left(-\frac{\theta^2}{2\sigma^2}\right) \exp\left(\frac{y_i\theta}{\sigma^2}\right) \quad (17) \\ &= \exp\left\{-\frac{\theta^2}{2}\left(\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}\right)\right\} \exp\left\{\theta\left(\frac{n\bar{y}}{\sigma^2} + \frac{\mu_0}{\tau_0^2}\right)\right\} \end{aligned}$$

ただし, $\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$

データが n 個の場合 (p.49)

(17) 式について,

$$\mu_n = \frac{\frac{n}{\sigma^2}\bar{y} + \frac{1}{\tau_0^2}\mu_0}{\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}}, \quad \frac{1}{\tau_n^2} = \frac{n}{\sigma^2} + \frac{1}{\tau_0^2} \quad (18)$$

とおけば $N(\theta | \mu_n, \tau_n^2)$

- 1 個の場合とほぼ同じ形式に
- ちなみに, 1 個ずつ追加していった場合も $p(\theta | \mathbf{y})$ は変わらない
- n を固定して $\tau_0 \rightarrow \infty$ か, τ_0 を固定して $n \rightarrow \infty$ のとき,

$$p(\theta | \mathbf{y}) \simeq N(\theta | \bar{y}, \sigma^2/n)$$

事前分布が全く当てにならない場合や, データ数が多いときは, ほぼ尤度のみで結果が決まる

その他の1-パラメータモデル (p.49)

- 解析解が得られないモデルは, $p(y) = \int p(y | \theta)p(\theta) d\theta$ の積分が困難なので, なるべく閉形式で解けるモデルを使った方が良い
- 二項モデル
- 正規モデル (平均未知), 正規モデル (分散未知)
 - 交換可能な測定値 (数) の和や, 対数を取ると和で扱えるような (積で影響する様な) データに適用できる
- ポアソンモデル, 指数モデル
 - 時間に対して独立で, 一定の発生確率を持つイベントの
 - 観測回数と待ち時間
- もっと難しいモデルは, 上の基本的な分布を組み合わせた分布により構築する事が多い
 - Section 18.4 では正規分布を二項分布で混合したモデルを扱う

母分散が未知・母平均が既知の正規モデル (p.50)

- 母分散だけを推定することはあまりないが、複雑なモデルを作る足がかりとしては有用
- 母平均・母分散の両方が未知の場合は Section 3.2–3.4
- 尺度パラメータの推定の導入としてもちょうど良い

n 個の観測値 y_1, y_2, \dots, y_n が i.i.d の場合を考える。ただし、 $y_i \sim N(\theta, \sigma^2)$ で θ が既知、 σ^2 は未知とする

$$\begin{aligned} p(\mathbf{y} \mid \sigma^2) &\propto (\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \theta)^2 \right\} \\ &= (\sigma^2)^{-n/2} \exp \left\{ -\frac{n}{2\sigma^2} t(\mathbf{y}) \right\} \end{aligned} \quad (19)$$

ただし、 $t(\mathbf{y}) = \frac{1}{n} \sum_{i=1}^n (y_i - \theta)^2$ は十分統計量である

母分散が未知・母平均が既知の正規モデル (p.50)

尤度が (19) 式するとき, 事前分布は,

$$p(\sigma^2) \propto (\sigma^2)^{-(\nu_0/2+1)} \exp\left\{-\frac{\nu_0\sigma_0^2}{2\sigma^2}\right\}$$

ハイパーパラメータ (ν_0, σ_0^2) を持つ, scaled inverse- χ^2 distribution[†] に従うと考えられる

以上から, 事後分布は

$$\begin{aligned} p(\sigma^2 | \mathbf{y}) &\propto (\sigma^2)^{-n/2} \exp\left\{-\frac{n}{2\sigma^2}t(\mathbf{y})\right\} (\sigma^2)^{-(\nu_0/2+1)} \exp\left\{-\frac{\nu_0\sigma_0^2}{2\sigma^2}\right\} \\ &= (\sigma^2)^{-\{(\nu_0+n)/2+1\}} \exp\left\{-\frac{1}{2\sigma^2}(\nu_0\sigma_0^2 + nt(\mathbf{y}))\right\} \end{aligned} \quad (20)$$

[†] $\chi_{\nu_0}^2$ 分布 (自由度 ν_0) に従う確率変数 X について, 変数変換 $Y = \frac{\nu_0\sigma_0^2}{X}$ を行えば, $Y \sim \text{Scale-inv-}\chi^2(\nu_0, \sigma_0^2)$ である. ν_0 は自由度, σ_0^2 は非心パラメータ

母分散が未知・母平均が既知の正規モデル (p.50)

(20) 式より,

$$p(\sigma^2 | \mathbf{y}) = \text{Scale-inv-}\chi^2 \left(\nu_0 + n, \frac{\nu_0 \sigma_0^2 + nt(\mathbf{y})}{\nu_0 + n} \right)$$

やっぱり事後分布のパラメータは

- 自由度は和になる $\nu_0 + n$
- 尺度パラメータは自由度で重みを付けた重み付き平均 $\frac{\nu_0 \sigma_0^2 + nt(\mathbf{y})}{\nu_0 + n}$
- データと事前分布の情報を合わせた形になる

point spreads の例; 正規モデル (p.50)

- Section 1.6 のフットボールの point spread のデータを使って, 分散未知の正規モデルでベイズ推定を行う
- ハンデ戦の結果 d_i ($\text{outcome}_i - \text{point spread}_i$) は交換可能で $d_i \sim N(0, \sigma^2)$ と考えて良さそう

解析の例

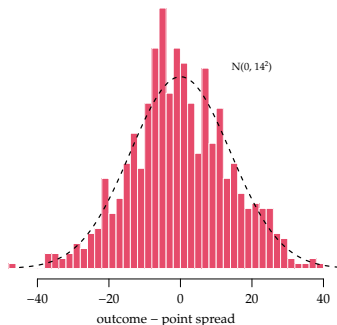
- 尤度以外に事前情報が全くないので, 事前分布は情報が無い状態に相当する $\nu_0 = 0$ の scaled inverse- χ^2 distribution

$$p(\sigma^2) \propto (\sigma^2)^{-1}$$

- $\int_0^{+\infty} p(\sigma^2) d\sigma$ は発散するので厳密にいうと密度関数になっていない, が, 事後分布としては問題がない

→Section 2.9

Figure 1.2 (b)

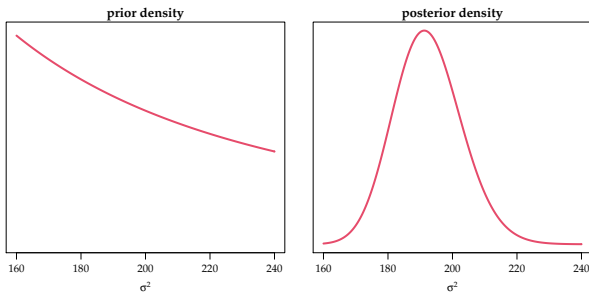


point spreads の例; 正規モデル (p.50)

- 事後分布は, Scale-inv- $\chi^2(n, t(\mathbf{d}))$

$$p(\sigma^2 | \mathbf{d}) \propto (\sigma^2)^{-(n/2+1)} \exp\left\{-\frac{1}{2\sigma^2}nt(\mathbf{d})\right\}, \quad t(\mathbf{d}) = \frac{1}{n} \sum_{i=1}^n d_i^2$$

- $n = 672, t(\mathbf{d}) = 13.85^2$



point spreads の例; 正規モデル (p.50)

- σ の 95%信用区間 (スライド 54 参照)

$$\left[\sqrt{\frac{nt(\mathbf{d})}{\chi_n^2(0.975)}}, \sqrt{\frac{nt(\mathbf{d})}{\chi_n^2(0.025)}} \right] = [13.1, 14.6]$$

- σ のシミュレーション

$$\text{median [2.5\%点, 97.5\%点]} = 13.8 [13.2, 14.5] \quad (21)$$

```
n <- 672
td <- 13.85^2
# 95% posterior interval of sigma^2
signif((n * td) / qchisq(0.975, n), 4)
signif((n * td) / qchisq(0.025, n), 4)
# 95% posterior interval of sigma
signif(sqrt((n * td) / qchisq(0.975, n)), 3)
signif(sqrt((n * td) / qchisq(0.025, n)), 3)
# posterior simulations of sigma
sigma <- sqrt((n * td) / rchisq(1000, n))
signif(quantile(sigma, c(0.5, 0.025, 0.975)), 3)
```

ポアソンモデル (p.51)

- 尤度がポアソン分布 $\text{Poisson}(y | \theta)$ の場合 (θ は平均発生回数)
- 疫学では, 人-時間サンプリングの時の発生率を調べる

n 個の観測値 y_1, y_2, \dots, y_n が i.i.d. のとき, 尤度は

$$\begin{aligned} p(\mathbf{y} | \theta) &= \prod_{i=1}^n \frac{\theta^{y_i}}{y_i!} \exp(-\theta) \\ &= \left(\prod_{i=1}^n \frac{1}{y_i!} \right) \exp(-n\theta) \exp\{t(\mathbf{y}) \log(\theta)\} \end{aligned}$$

自然パラメータ $\log(\theta)$ に対する十分統計量は $t(\mathbf{y}) = \sum_{i=1}^n y_i$

自然共役事前分布は $\text{Gamma}(\alpha, \beta)$ で, 事後分布は $\text{Gamma}(\alpha + t(\mathbf{y}), \beta + n)$,

$$\begin{aligned} p(\theta) &\propto \exp(-\beta\theta) \exp\{(\alpha - 1) \log(\theta)\} \\ p(\theta | \mathbf{y}) &\propto \exp\{-(\beta + n)\theta\} \exp\{(\alpha + t(\mathbf{y}) - 1) \log(\theta)\} \end{aligned}$$

負の二項分布 (p.51)

- 共役事前分布を用いている場合, 事前分布と事後分布が等しくなることが分かっているので, 積分することなく事前予測分布が求められる

$$p(y) = \frac{p(y | \theta)p(\theta)}{p(\theta | y)} \left(= \int_{\theta \in \Theta} p(y | \theta)p(\theta) d\theta \right)$$

観測値が 1 つの場合のポアソンモデルでは

$$\begin{aligned} p(y) &= \frac{\text{Poisson}(y | \theta)\text{Gamma}(\theta | \alpha, \beta)}{\text{Gamma}(\theta | \alpha + y, \beta + 1)} \\ &= \frac{\theta^y \exp(-\theta)}{y!} \frac{\exp(-\beta\theta) \beta^\alpha \theta^{\alpha-1}}{\Gamma(\alpha)} \bigg/ \frac{\exp\{-(\beta+1)\theta\} (\beta+1)^{\alpha+y} \theta^{\alpha+y-1}}{\Gamma(\alpha+y)} \quad (22) \\ &= \frac{\Gamma(\alpha+y)\beta^\alpha}{y!\Gamma(\alpha)(\beta+1)^{\alpha+y}} = \binom{\alpha+y-1}{y} \left(\frac{\beta}{\beta+1}\right)^\alpha \left(\frac{1}{\beta+1}\right)^y \end{aligned}$$

これは, 負の二項分布 $\text{Neg-bin}(y | \alpha, \beta)$ の確率関数

負の二項分布 (p.52)

(22) 式より,

$$\int_0^{+\infty} \text{Poisson}(y | \theta) \text{Gamma}(\theta | \alpha, \beta) d\theta = \text{Neg-bin}(y | \alpha, \beta)$$

負の二項分布は、ポアソン分布の無限混合分布になっている (重みがガンマ分布)

- Section 17.2 ではポアソン分布に代わるロバストな方法として再登場します

ポアソンモデルを rate と exposure でパラメトライズする場合 (p.53)

- 疫学で、人-時間サンプリングのデータを解析するとき
- x_i を暴露期間 (exposure; 人年など) として固定すると θ が発生率 (rate) になる

$$p(y_i | \theta) = \text{Poisson}(y_i | x_i \theta) \quad (23)$$

- y_i ではなく, $(y, x)_i$ のペアに関して exchangeable n 個の観測値 y_1, y_2, \dots, y_n のとき, 尤度は

$$p(\mathbf{y} | \theta) \propto \exp \left\{ - \left(\sum_{i=1}^n x_i \right) \theta \right\} \exp \{ t(\mathbf{y}) \log(\theta) \}$$

$t(\mathbf{y}) = \sum_{i=1}^n y_i$. 事前分布は $\text{Gamma}(\alpha, \beta)$ になるので, 事後分布は

$$p(\theta | \mathbf{y}) = \text{Gamma} \left(\theta \mid \alpha + t(\mathbf{y}), \beta + \sum_{i=1}^n x_i \right) \quad (24)$$

ポアソンデータにおける率の推定; ぜんそくの例 (p.53)

- 人口 20 万人の都市で, 1 年間にぜんそくが原因で死亡する人の数を調べたところ 3 人であった

$$\text{粗死亡率} = \frac{1.5 \text{ 人}}{10 \text{ 万人年}} (=10 \text{ 万人年あたり } 1.5 \text{ 人})$$

- これをポアソンモデルで解析します

θ を, 10 万人年あたりのぜんそく死亡率のパラメータとすると, 尤度は (23) 式における $y = 3, x = 2$ に相当し

$$p(y | \theta) = \text{Poisson}(3 | 2.0\theta)$$

事前分布について

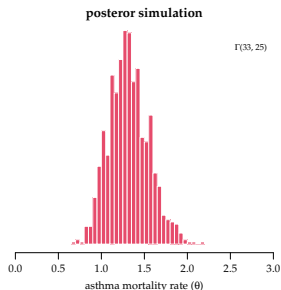
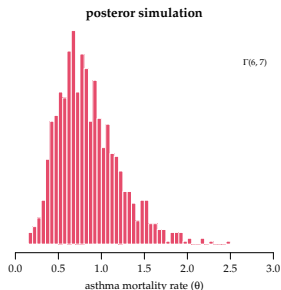
- 西洋諸国の標準的なぜんそく死亡率は 0.6 [人/10 万人年] 程度
- 事前情報を試行錯誤の後, 共役事前分布で表わすと $\text{Gamma}(3.0, 5.0)$
- 事前平均は 0.6 (事前モードは 0.4)

ポアソンデータにおける率の推定; ぜんそくの例 (p.53)

事後分布は (24) 式の結果から

$$\text{Gamma}(\theta \mid 3 + 3.0, 2.0 + 5.0)$$

- 事後平均は 0.86 (事後モードは 0.71)
- ぜんそく死亡率が 1.0 [人/10 万人年] を超える確率は, 0.3



ポアソンデータにおける率の推定; ぜんそくの例 (p.54)

- 他の条件を全て固定し, 観察期間が 10 年になったと仮定した場合の結果を検討
- 観察期間が 10 年なので $y = 30$ になる, 人口 20 万人を固定した時の尤度は

$$p(y | \theta) = \text{Poisson}(30 | 20.0\theta)$$

- 事前分布は同じ $p(y | \theta) = \text{Poisson}(3 | 2.0\theta)$

事後分布は

$$\text{Gamma}(\theta | 30 + 3.0, 20.0 + 5.0)$$

- 事後平均は 1.32 (事後モードは 1.28)
- ぜんそく死亡率が 1.0 [人/10 万人年] を超える確率は, 0.93

指数モデル (p.55)

- 尤度が指数分布 $\text{Exponential}(y | \theta)$ の場合
- 指数モデルは待ち時間が観測される場合, 発生率 θ が興味の対象

観測値 y が 1 つの時, 尤度は

$$p(y | \theta) = \theta \exp(-y\theta) = \text{Gamma}(y | 1, \theta)$$

$\theta = 1/E(y | \theta) > 0$ が発生率を表わすパラメータになる

- 指数分布の重要な特徴に"無記憶性"[†]があり, 以下の式が成立することである

$$\forall s, t \in (0, \infty) \quad \Pr(y > t + s | y > s, \theta) = \Pr(y > t | \theta)$$

- 生存時間解析のモデルで用いられる

[†]連続の場合は指数分布, 離散の場合は幾何分布のみが無記憶性をもつ

指数モデル (p.55)

n 個の観測値 y_1, y_2, \dots, y_n のとき, 尤度は

$$p(\mathbf{y} | \theta) = \theta^n \exp \{-t(\mathbf{y})\theta\}, \quad t(\mathbf{y}) = \sum_{i=1}^n y_i$$

観測値が n 個でもガンマ分布の形をしているので, 共役事前分布は $\text{Gamma}(\alpha, \beta)$, 事後分布は

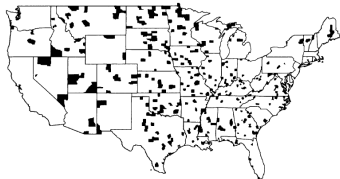
$$p(\theta | \mathbf{y}) = \text{Gamma} \left(\theta \mid \alpha + n, \beta + t(\mathbf{y}) \right)$$

- ポアソンモデルの尤度の部分がひっくり返った形

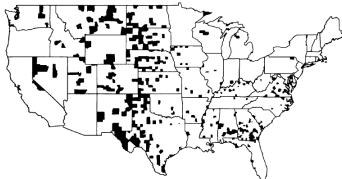
例: がん死亡率推定に対する情報事前分布と多重階層構造 (p.55)

- 1980年–1989年の間に腎臓がんで死亡した白人のアメリカ人男性の人数を郡別に調査
- Figure 2.7, Figure 2.8 は, 年齢調整死亡率が上位 10%, 下位 10% だった郡に色を付けたもの
- 海岸沿いでは死亡率が高くも低くもない, 内陸部, 特にグレートプレインズの辺りに死亡率が高い郡と低い郡が集中

Highest kidney cancer death rates



Lowest kidney cancer death rates



何故でしょうか?

死亡率が高い, 低い郡が内陸部に集中する理由 (p.56)

- 腎臓がんによる死亡が非常に起こりにくく, 10年の追跡で1件もイベントがない郡もあり得る
- グレートプレインズ周辺は人口密度が非常に低い地域で, 人口の少ない郡が多い
- 元々死亡率が低い上に人口が少ないため, イベント0の郡が多くなりやすく, イベントが少なくても他の地域より死亡率が高く見積もられやすい
- 郡毎のサンプルサイズの違いによって, 誤ったパターンが見えていた

がん死亡率のベイズ推定 (p.57)

- このような問題に対処するために、ベイズ推定を行う事にする
- j は郡を表わす添え字, y_j は 1980 年–1989 年の間の死亡数, n_j は郡の人口 ($10n_j$ で観測人年), θ_j が死亡率を表わすパラメータとすると, 尤度は

$$p(y_j | \theta_j) = \text{Poisson}(10n_j\theta_j)$$

- 事前分布にはポアソン尤度に対して共役事前分布になるように $p(\theta) = \text{Gamma}(20, 430000)$ を用いた
- 事後分布とその期待値・分散は,

$$p(\theta_j | y_j) = \text{Gamma}(20 + y_j, 430000 + 10n_j)$$

$$E(\theta_j | y_j) = \frac{20 + y_j}{430000 + 10n_j}, \quad \text{Var}(\theta_j | y_j) = \frac{20 + y_j}{(430000 + 10n_j)^2}$$

n_j が小さい郡はどうなるか (p.58)

- $n_j = 1000$ のとき, データ y_j が得られる確率をまず評価してみよう

データ y_j が得られる確率は, 事前予測分布により評価できる. 尤度がポアソン分布, 事前分布がガンマ分布のときの事前予測分布は Section 2.7 から $\text{Neg-bin}(y_j | \alpha, \frac{\beta}{10n_j})$ である.

負の二項分布の確率を直接計算しても良いが, 下記の様なシミュレーションにより確認するとより簡便だろう

- ① $\text{Gamma}(20, 430000)$ に従う乱数を 500 個発生させ θ_j を得る
- ② θ_j それぞれに対して $\text{Poisson}(10n_j\theta_j)$ に従う乱数を 1 個ずつ発生させ, 事前予測分布に従う乱数を 500 個を得る

```
theta <- 10 * 1000 * rgamma(500, 20, 430000)
negbin <- sapply(theta, function(theta) rpois(1, theta))
table(negbin)
```

結果: $y_j = 0$ が 319 回, $y_j = 1$ が 141 回, $y_j = 2$ が 33 回, $y_j = 3$ が 5 回

n_j が小さい郡はどうなるか (p.58)

ほとんどの場合, $y_j = 0 \sim 3$ になる事が分かったので, そのときの粗死亡率と事後平均を求めてみると,

y_j	raw death rate ($y_j/10n_j$)	posterior mean
0	0	4.55×10^{-5}
1	1.00×10^{-4}	4.77×10^{-5}
2	2.00×10^{-4}	5.00×10^{-5}
3	3.00×10^{-4}	5.23×10^{-5}

- $y_j = 0$ の時は事後平均が大きめに, $y_j \geq 1$ の時は事後平均が小さめの値になる, prior 支配的な結果

n_j が大きい郡はどうなるか (p.59)

- $n_j = 10^6$ のとき, データ y_j が得られる確率

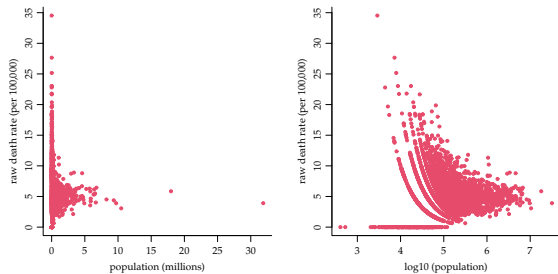
```
theta <- 10 * 10^6 * rgamma(500, 20, 430000)
negbin <- sapply(theta, function(theta) rpois(1, theta))
quantile(negbin, c(0.25, 0.50, 0.75))
```

median は 473, 50% 区間は [393, 545]. この時の粗死亡率と事後平均は,

y_j	raw death rate ($y_j/10n_j$)	posterior mean
393	3.93×10^{-5}	3.96×10^{-5}
473	4.73×10^{-5}	4.73×10^{-5}
545	5.45×10^{-5}	5.42×10^{-5}

- 粗死亡率と事後平均はほとんど同じで, データ支配的な結果

n_j と粗死亡率 (p.59)



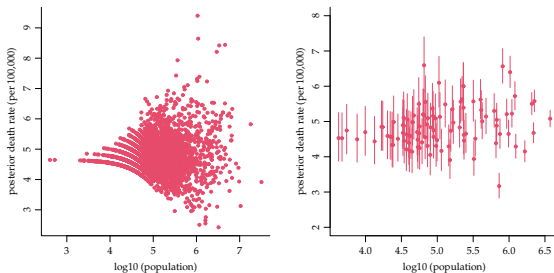
- ポアソンモデル $\text{Poisson}(y_j | 10n_j\theta_j)$ のとき, raw death rate の分散は n_j に反比例する (サンプルサイズと標準誤差)

$$\text{Var}\left(\frac{Y_j}{10n_j}\right) = \frac{10n_j\theta_j}{(10n_j)^2} = \frac{\theta_j}{10n_j}$$

- 実際, Figure 2.9 においてもそれが確認できる

横軸が観測人年とその対数, 縦軸は raw death rate

n_j と死亡率の事後分布 (p.59)



- Figure 2.10a は事後平均 $E(\theta_j | y_j)$ を示したもので、かなり変動が小さくなっている (軸を見る)
- Figure 2.10b は、事後分布の 25%, 50%, 75% 点をプロットしたもので、観測人年が大きくなるほど区間が狭くなっていることが分かる

横軸が観測人年の対数, 縦軸は事後平均または事後分布の 25%, 50%, 75%点

事前分布の選択を再考 (p.60)

Gamma(20, 430000) をどのように設定したのか

- 今回は事前予測分布に基づいて設定した. 事前予測分布のモーメントと, 3071 個のデータのモーメントが一致するパラメータを導く

$$p(y_j) = \int_0^{\infty} p(y_j | \theta_j) p(\theta_j) d\theta_j = \text{Neg-bin}(\alpha, \beta / (10n_j))$$

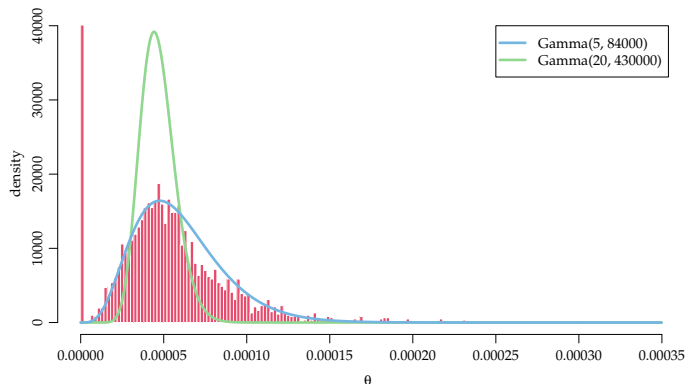
$$E(y_j) = 10n_j \frac{\alpha}{\beta} \tag{25}$$

$$\text{Var}(y_j) = 10n_j \frac{\alpha}{\beta} + (10n_j)^2 \frac{\alpha}{\beta^2}$$

- 本当は y_j よりも $y_j / (10n_j)$ を使った方が良く, 年齢調整等もするべきだが, パラメータを全データから推定するコンセプトの一つとして示している
- この方法は結構ずさんなやり方なので, Chapter 5 や 16 の階層モデルの話題の時により詳細な議論をする

事前分布の選択を再考 (p.60)

- 実際にやってみたが, $\text{Gamma}(20, 430000)$ は導けなかった, グラフを見るとばらつきが小さすぎる?
- θ の推定量 y_j/n_j からパラメータの最尤推定量を求め, それと (25) 式を用いて導くと, $\text{Gamma}(5, 84000)$ が得られた



無情報事前分布 (p.61)

- 母集団の事前情報として何の根拠も持っていない場合にどうするか?
 - 事前の情報が全くないことを表わすために、事後分布に最小の影響しか与えない事を保証できる事前分布が必要
 - vague prior (漠然事前分布), diffuse prior (拡散事前分布), noninformative prior (無情報事前分布) 等々複数の呼び方がある

変則でない事前分布と変則事前分布 (p.61)

- スライド 51, 平均 θ が未知の正規モデルの例

事後分布は $N\left(\theta \mid \frac{\frac{n}{\sigma^2}\bar{y} + \frac{1}{\tau_0^2}\mu_0}{\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}}, \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}}\right)$ であったから, $\tau_0^2 = \infty$ のとき,

$$p(\theta \mid \mathbf{y}) \simeq N(\theta \mid \bar{y}, \sigma^2/n) \quad (26)$$

事後分布は事前分布の影響を全くうけなくなる

- $\tau_0^2 = \infty$ のとき的事前分布は

$$p(\theta) \propto \exp\left(-\frac{\theta^2}{2\tau_0^2}\right) \exp\left(\frac{\mu_0\theta}{\tau_0^2}\right) = \text{const.}, \quad \Theta = \{\theta \in (-\infty, \infty)\} \quad (27)$$

であり, この分布は不可積分である

$$\int_{\Theta} p(\theta) d\theta = +\infty$$

変則でない事前分布と変則事前分布 (p.61)

- (27) 式の前分布は確率の定義 $\Pr(\Theta) = 1$ を満たさない, このような分布を変則分布 (improper distribution) と呼ぶ
- もう一つ無情報事前分布の例として, 分散未知の正規モデルを考える事後分布は $\text{Scale-inv-}\chi^2\left(\sigma^2 \mid \nu_0 + n, \frac{\nu_0\sigma_0^2 + nt(\mathbf{y})}{\nu_0 + n}\right)$ であったから, $\nu_0 = 0$ のとき, 事前分布の影響を全くうけなくなる

$$p(\sigma^2 \mid \mathbf{y}) = \text{Scale-inv-}\chi^2\left(\sigma^2 \mid n, t(\mathbf{y})\right) \quad (28)$$

$\nu_0 = 0$ の時の事前分布は

$$p(\sigma^2) \propto (\sigma^2)^{-(\nu_0/2+1)} \exp\left\{-\frac{\nu_0\sigma_0^2}{2\sigma^2}\right\} = (\sigma^2)^{-1}, \quad \Theta = \{\sigma^2 \in (0, \infty)\}$$

であり, この分布も変則分布である

変則事前分布の事後分布は? (p.62)

- (26) 式, (28) 式の例では事後分布は可積分で, ちゃんとした確率分布になる
- 事前分布が変則であっても, 大抵の場合は事後分布が可積分. ただし, 事後分布が変則分布にならぬようにチェックは必ず行った方がよい
- データ支配的な状況の極限とも解釈出来る

Jeffreys のルール; Jeffreys' invariance principle (p.62)

- 何らかの対一変換 $\phi = h(\theta)$ を行ったパラメータの分布が一様分布になるような無情報事前分布を構築する方法
- 確率変数変換の公式を使う. $p(\phi)$ を一様にするためには, $p(\theta)$ を以下のようにする必要がある

$$p(\phi) = p(\theta) \left| \frac{d\theta}{d\phi} \right| \propto \text{const.}, \quad p(\theta) \propto \left| \frac{d\phi}{d\theta} \right| \quad (29)$$

- $p(\theta) \propto J(\theta)^{1/2}$ となる (Fisher 情報量に比例する) ような事前分布を定めればよい \rightarrow (Box & Tiao, 1979)

$$J(\theta) = E \left[\left(\frac{d \log p(\mathbf{y} | \theta)}{d\theta} \right)^2 \middle| \theta \right] = -E \left[\frac{d^2 \log p(\mathbf{y} | \theta)}{d\theta^2} \middle| \theta \right]$$

$$J(\phi) = -E \left[\frac{d^2 \log p(\mathbf{y} | \theta = h(\phi)^{-1})}{d\theta^2} \left| \frac{d\theta}{d\phi} \right|^2 \middle| \theta \right] = J(\theta) \left| \frac{d\theta}{d\phi} \right|^2 \quad (30)$$

Jeffreys のルール; Jeffreys' invariance principle (p.63)

- (29) 式, (30) 式より

$$p(\theta) = \left| \frac{d\phi}{d\theta} \right| \propto J(\theta)^{1/2}$$

と選ぶことで,

$$J(\phi) = J(\theta) \left| \frac{d\theta}{d\phi} \right|^2 \propto \left| \frac{d\phi}{d\theta} \right|^2 \left| \frac{d\theta}{d\phi} \right|^2 = \text{const.}$$

が定数になるような変換を選んでいる事も確認できる

- Fisher 情報行列を用いることで多変量の場合に容易に拡張できる;
 $p(\theta) \propto |J(\theta)|^{-1/2}$
- Jeffreys のルールは, (期待値は取るものの) 尤度に基づく推測だから, パラメータ変換に関して共変的, $p(\phi)$ に一様分布を仮定した推測である事に関して不変 (invariant) である? ; 損失関数が不変?

二項モデルの例 (p.63)

- Jeffreys のルールによる事前分布は

$$\log p(y | \theta) = \text{const.} + y \log \theta + (n - y) \log(1 - \theta)$$

$$J(\theta) = -E \left[\frac{d^2 \log p(\mathbf{y} | \theta)}{d\theta^2} \middle| \theta \right] = \frac{n}{\theta(1 - \theta)}$$

$$J(\theta)^{1/2} \propto \theta^{-1/2}(1 - \theta)^{-1/2} = \text{Beta}\left(\frac{1}{2}, \frac{1}{2}\right)$$

- Bayes と Laplace が用いた事前分布 (一様分布) は $\text{Beta}(1, 1)$
- natural parameter; $\text{logit}(\theta)$ の分布が一様分布になるような事前分布は $\text{Beta}(0, 0) \rightarrow \text{Exercise 7}$
 - 3 つとも事後分布に与える影響は通常小さいため, 特異な場合を除けばどれを使っても結果はほとんど変わらない
 - この中で $\text{Beta}(0, 0)$ は変則分布で, この場合は $y = 0, n$ が観測されると事後分布も変則分布になってしまう

Pivotal quantities (p.64)

- 二項モデルの例で見たように、どんなパラメータに対して一様とするかによって、 $p(\theta)$ が変わる。一体どういう変換が良いのだろうか？
- 尤度関数が、あるパラメータについて、特定の形をしている場合については、以下のことがいえる

pure location parameter

- 尤度関数が $p(y - \theta | \theta) = f(u | \theta)$ のように、データとパラメータの差 $u = y - \theta$ の関数に変換できるとき、 $y - \theta$ を pivotal quantity, θ を pure location parameter と呼ぶ
- このとき、線型変換 $X = Y + a$ について、 $p(x - \phi | \phi) = p(u | \theta)$, $\phi = \theta + a$ を構成できる。 Y と X は、位置パラメータだけが異なる「同一の」分布に従う。すべての a について、 $p(\theta) = p(\theta + a)$ を満たすためには、 $p(\theta) \propto \text{const.}$ となる必要がある^[6]
- 当然、事後分布は、 $p(\theta | y - \theta) \propto p(y - \theta | \theta)p(\theta) = p(y | \theta)$,
 $p(\phi | x - \phi) \propto p(x - \phi | \phi)p(\phi) = p(x | \phi)$

Pivotal quantities (p.64)

pure scale parameter

- 尤度関数が $p(y/\theta | \theta) = g(v)$ のように, データとパラメータの比 $v = y/\theta$ の関数に変換できるとき, y/θ を pivotal quantity, θ を pure scale parameter と呼ぶ ($\theta \in (0, +\infty)$)
- ただし, Y の分布は $p(y | \theta) = p(v | \theta) \left| \frac{dv}{dy} \right| = p(v | \theta)/\theta$. 今度は定数倍 $X = bY$ について, $p(x/\phi | \phi) = p(v | \theta)$, $\phi = b\theta$ を構成できる. Y と X は, 尺度パラメータだけが異なる「同一の」分布に従う. すべての $b(> 0)$ について, $p(\theta) = p(b\theta)$ を満たすためには, $p(\log \theta) \propto \text{const.}$; $p(\theta) \propto \theta^{-1}$ となる必要がある^[6]
- 当然, 事後分布は, $p(\theta | y/\theta) \propto p(y/\theta | \theta)p(\theta) = p(y | \theta)$, $p(\phi | x/\phi) \propto p(x/\phi | \phi)p(\phi) \propto p(x | \phi)$
- また, 対数変換したときに, データ y の変化に対し位置のみが変動すると解釈出来る

ちょっとした実例

$$p(y | \sigma) = N(y | \mu, \sigma^2), \quad p(\sigma) = \sigma^{-1}, \quad v = y/\sigma$$

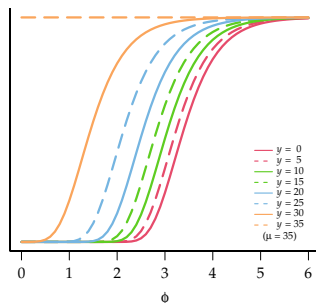
$$p(v | \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(v - \mu/\sigma)^2}{2}\right\} \left|\frac{dy}{dv}\right| = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(v - \mu/\sigma)^2}{2}\right\}$$

$$p(\sigma | v) \propto p(v | \sigma)p(\sigma) \propto \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(v - \mu/\sigma)^2}{2}\right\} = N(\sigma | y, \mu)$$

- $\phi = \log \sigma$ とおくと、パラメータ σ を対数変換した分布は、

$$p(\phi | v) \propto \exp\left\{-\frac{(v - \mu/\exp(\phi))^2}{2}\right\}$$

- 一様分布を与えたスケールに戻すと、異なるデータに対して平行移動する様子が見える



Pivotal quantities (p.64)

注意

- 変則事前分布を用いても, (ほとんどの場合) 事後分布が変則分布にならない, という状況を期待して用いる事が多いが, 例えば, 階層分散パラメータの対数に一樣分布を用いてはならない!→Section 5.4
- 事後分布の確認を怠ってはいけない

data-translated likelihood (Box and Tiao 1979)^[12]

- 尤度関数を $p(y | \theta) = g[h(\theta) - f(y)]$ の様な形に変換できるとき, その尤度関数を data-translated likelihood と呼ぶ
- 近似的に location parameter の様な関係を持つ,
- 変換 ϕ を見つけるために, 対数尤度関数 $L(\phi | y) = \log p(y | \phi)$ を最尤推定量 $\hat{\phi}$ の周りで 2 次まで Taylor 展開すると (1 次の項は 0)

$$L(\phi | y) \simeq L(\hat{\phi} | y) - \frac{n}{2} \left\{ -\frac{1}{n} \frac{d^2 L(\phi | y)}{d\phi^2} \Big|_{\phi=\hat{\phi}} \right\} (\phi - \hat{\phi})^2$$

この式を見ると, $\left\{ -\frac{1}{n} \frac{d^2 L(\phi | y)}{d\phi^2} \Big|_{\phi=\hat{\phi}} \right\}$ が定数であれば, (対数) 尤度関数が近似的に $(\phi - \hat{\phi})$ の関数になり, $g[h(\theta) - f(y)]$ の形が得られる

data-translated likelihood (Box and Tiao 1979)^[12]

$\left\{ -\frac{1}{n} \frac{d^2 L(\phi | y)}{d\phi^2} \Big|_{\phi=\hat{\phi}} \right\}$ について, n が十分大きいとき, 期待値 (Fisher 情報量 $I(\hat{\phi})$) で近似することができる

$$-\frac{1}{n} \frac{d^2 L(\phi | y)}{d\phi^2} \Big|_{\phi=\hat{\phi}} = -\frac{1}{n} \sum_{i=1}^n \frac{d^2 L(\phi | y_i)}{d\phi^2} \Big|_{\phi=\hat{\phi}} \simeq I(\hat{\phi})$$

(30) 式より,

$$I(\hat{\phi}) = I(\hat{\theta}) \left| \frac{d\theta}{d\phi} \right|_{\theta=\hat{\theta}}^2 \propto \text{const.}$$

となるような変換を選べばよい. ϕ と θ は一対一変換だから,

$$\left| \frac{d\theta}{d\phi} \right| \propto I(\theta)^{-1/2}, \quad \left| \frac{d\phi}{d\theta} \right| \propto I(\theta)^{1/2}$$

これに基づいて無情報事前分布を決める事を Jeffreys のルールと呼ぶ

無情報事前分布の問題点 (p.64)

- ① 本当にデータ支配的であれば, ある程度フラットな事前分布であればよい. 何でもかんでも無情報事前分布を使っていると, 誤用を誘発する可能性がある
 - ② あるパラメトライズをした分布が一様分布でも, 他のパラメトライズした分布は違う.
 - ③ 色々な無情報事前分布を持った競合するモデルを平均化する時にも問題が →Section 6.7
- 問題点はあるにしても, 適切に用いれば (モデルの妥当性さえ検証すれば) 有用である
 - ほぼ全ての問題点は, データ解析を行う人が, 都合の良い情報をモデルに取り込もうとするため生じる
 - 便利なモデルを作りたいが, 複雑だとうまく事前情報を取り込めない, 計算が難しくなる, また余計な情報が入る可能性がある. こういったもののバランスを取るようにするとよい

練習問題 1

- コインの表が出る確率を θ とし, θ の事前分布が $\text{Beta}(\theta | 4, 4)$ であるとする. 10 回のコイントスをして表が出た回数が 3 回未満だった, というデータを得たときの θ に関する事後分布を求めて図示してください

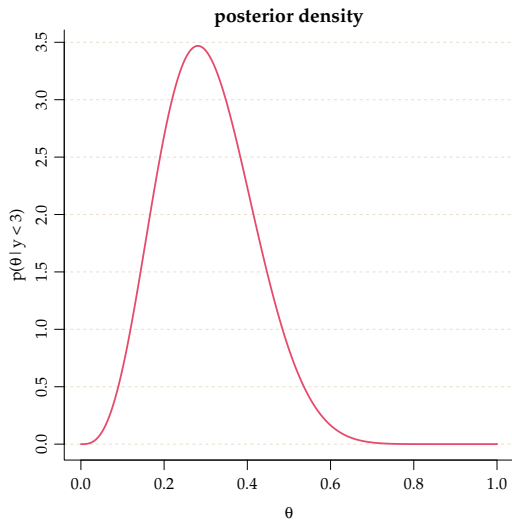
$$\Pr(y < 3 | \theta) = \sum_{i=0}^2 \binom{10}{i} \theta^i (1 - \theta)^{10-i}, \quad p(\theta) = \text{Beta}(\theta | 4, 4) = \frac{\theta^3 (1 - \theta)^3}{B(4, 4)}$$

ただし, $B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt$ はベータ関数.

事後分布は

$$\begin{aligned} \Pr(\theta | y < 3) &= \frac{\Pr(y < 3 | \theta) p(\theta)}{\int_0^1 \Pr(y < 3 | \theta) p(\theta) d\theta} \\ &= \frac{7735}{8} \{ \theta^3 (1 - \theta)^{13} + 10\theta^4 (1 - \theta)^{12} + 45\theta^5 (1 - \theta)^{11} \} \end{aligned}$$

練習問題 1 のグラフ



練習問題 2

- C_1, C_2 というコインがあり, 表が出る確率をそれぞれ $\Pr(\text{heads} | C_1) = 0.6, \Pr(\text{heads} | C_2) = 0.4$ とする. どちらかのコインをランダムに選んで, 2 回投げたところ 2 回とも裏が出たとする. この情報を得たもとで, 表が出るまでコインを投げ続けるとすると, 表が出るまでにコインを投げる回数の期待値はいくつか.

まずはコインをどちらかランダムに選び, 2 回投げたところ 2 回とも裏が出たときの事後確率を求める

$$\Pr(y = 0 | \pi) = \binom{2}{0} \pi^0 (1 - \pi)^{2-0} = (1 - \pi)^2$$

$$p(\pi | y = 0) = \frac{0.5(1 - \pi)^2}{0.5(1 - 0.4)^2 + 0.5(1 - 0.6)^2} = \frac{(1 - \pi)^2}{0.6^2 + 0.4^2}$$

練習問題 2

表が出るまでにコインを投げる回数 Z の分布を求める (幾何分布)

$$p(z | \pi) = \pi(1 - \pi)^{z-1}$$

$$\begin{aligned} p(z | y = 0) &= \sum_{\pi=0.4, 0.6} p(z | \pi)p(\pi | y = 0) \\ &= \frac{0.6^2}{0.6^2 + 0.4^2} p(z | \pi = 0.4) + \frac{0.4^2}{0.6^2 + 0.4^2} p(z | \pi = 0.6) \end{aligned}$$

$p(z | y = 0)$ は混合幾何分布

$$\begin{aligned} E(z | y = 0) &= \sum_{z=1}^{\infty} z \frac{0.6^2}{0.6^2 + 0.4^2} p(z | 0.4) + \sum_{z=1}^{\infty} z \frac{0.4^2}{0.6^2 + 0.4^2} p(z | 0.6) \\ &= \frac{0.6^2}{0.6^2 + 0.4^2} \frac{1}{0.4} + \frac{0.4^2}{0.6^2 + 0.4^2} \frac{1}{0.6} = 2.24 \end{aligned}$$

幾何分布の期待値

負の二項分布[†]

- 確率 π のベルヌーイ試行を行って、 z 回目の試行で r 回成功する確率
- $z \in \mathbb{Z}^+$; $r \in \mathbb{Z}^+$; $\pi \in (0, 1)$

$$\Pr(Z = z \mid r, \pi) = \binom{z-1}{r-1} \pi^r (1-\pi)^{z-r}$$

$$\begin{aligned} E(Z) &= \sum_{z=1}^{\infty} z \frac{(z-1)!}{(r-1)! (z-r)!} \pi^r (1-\pi)^{z-r} \\ &= \sum_{z=1}^{\infty} r \frac{z!}{r! (z-r)!} \frac{\pi^{r+1}}{\pi} (1-\pi)^{(z+1)-(r+1)} = \frac{r}{\pi} \end{aligned}$$

$$E(Z-r) = \frac{r(1-\pi)}{\pi} \text{ を求めることもある; } r=1 \text{ の時, } E(Z) = \frac{1}{\pi}$$

[†]幾何分布は負の二項分布の特殊な場合

練習問題 3

- 正確なサイコロを 1000 回振ったとき、6 の面が上を向いた回数を y とする。

(a) y の分布を正規近似してプロットしましょう

6 の面が上を向く確率を θ とすると、正確なサイコロであるから、

$$\Pr\left(\theta = \frac{1}{6}\right) = 1$$

また、6 の面が上になるか、6 以外の面が上になるか 2 つの結果しか存在しないため、

$$p(y | \theta) = \binom{1000}{y} \theta^y (1 - \theta)^{1000-y}$$

したがって、

$$p(y) = \sum_{\theta} p(y | \theta) p(\theta) = \binom{1000}{y} \left(\frac{1}{6}\right)^y \left(\frac{5}{6}\right)^{1000-y} \times 1$$

練習問題 3

$$E(y) = n\theta = \frac{1000}{6} = 166.6, \text{Var}(y) = n\theta(1 - \theta) = \frac{5000}{36} = 11.8^2$$

正規近似を行うと, $y \sim N(166.6, 11.8^2)$

(b) y の分布を正規近似をしたときの 5%, 25%, 50%, 75%, 95% 点を求めましょう

標準正規分布の各%点は, $-1.645, -0.674, 0, 0.674, 1.645$

$$5\% \text{点: } 166.7 - 1.645 \times 11.8 = 147.3$$

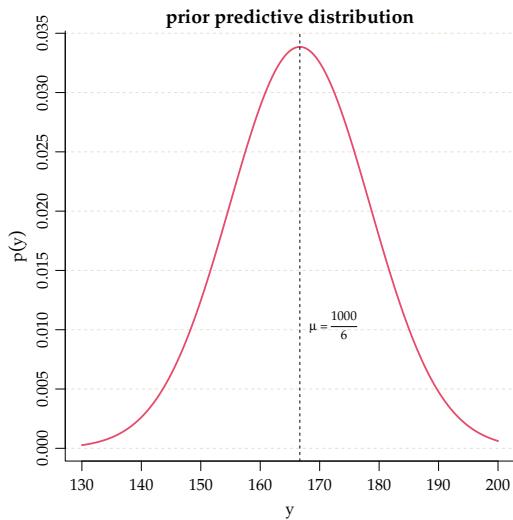
$$25\% \text{点: } 166.7 - 0.674 \times 11.8 = 158.7$$

$$50\% \text{点: } 166.7$$

$$75\% \text{点: } 166.7 + 0.674 \times 11.8 = 174.6$$

$$95\% \text{点: } 166.7 + 1.645 \times 11.8 = 186.1$$

練習問題 3 のグラフ



練習問題 4

- イカサマの疑いがあるサイコロを 1000 回振ったとき、6 の面が上を向いた回数を y とする。6 の面が上を向く確率を θ とし、 θ に対する事前分布を以下のように仮定する。

$$\Pr\left(\theta = \frac{1}{12}\right) = 0.25, \Pr\left(\theta = \frac{1}{6}\right) = 0.5, \Pr\left(\theta = \frac{1}{4}\right) = 0.25$$

(a) $p(y | \theta)$ を正規近似して、 $p(y)$ を図示しましょう

$p(y | \theta)$ を正規近似するため、期待値と分散を求めると、

$$E\left(y \mid \theta = \frac{1}{12}\right) = 83.3, \text{Var}\left(y \mid \theta = \frac{1}{12}\right) = 8.74^2$$

$$E\left(y \mid \theta = \frac{1}{6}\right) = 166.7, \text{Var}\left(y \mid \theta = \frac{1}{6}\right) = 11.8^2$$

$$E\left(y \mid \theta = \frac{1}{4}\right) = 250, \text{Var}\left(y \mid \theta = \frac{1}{4}\right) = 13.7^2$$

練習問題 4

$p(y)$ は混合二項分布になる

$$p(y) = 0.25 \binom{1000}{y} \left(\frac{1}{12}\right)^y \left(\frac{11}{12}\right)^{1000-y} + 0.5 \binom{1000}{y} \left(\frac{1}{6}\right)^y \left(\frac{5}{6}\right)^{1000-y} \\ + 0.25 \binom{1000}{y} \left(\frac{1}{4}\right)^y \left(\frac{3}{4}\right)^{1000-y}$$

正規近似を行うと,

$$p(y) \approx 0.25N(83.3, 8.74^2) + 0.5N(166.7, 11.8^2) + 0.25N(250, 13.7^2)$$

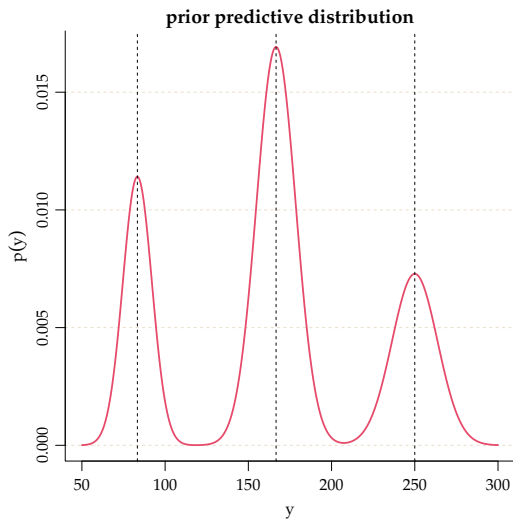
(b) $p(y)$ を正規近似したときの 5%, 25%, 50%, 75%, 95%点を求めましょう

5%点: 76.0, 25%点: 118.0

50%点: 166.6, 75%点: 206.5

95%点: 261.5

練習問題 4 のグラフ



練習問題 5

- n 回コインを投げて、表が出た回数を y とし、コインを投げたときに表が出る確率を θ とする.

(a) 事前分布が一様分布 $U(0, 1)$ のとき、事前予測分布を求めましょう

(18) 式参照

$$p(y) = \int_0^1 p(y | \theta)p(\theta) d\theta = \int_0^1 \binom{n}{y} \theta^y (1 - \theta)^{n-y} d\theta = \frac{1}{n+1}$$

(b) 事前分布が $\text{Beta}(\theta | \alpha, \beta)$ のとき、 θ の事後平均が、常に事前平均 $\frac{\alpha}{\alpha+\beta}$ と観測値の相対度数 $\frac{y}{n}$ の間にある事を代数的に示してください

ベータ分布の定義より $\alpha, \beta > 0$, 二項分布の尤度より $n \geq y, n > 0, y \geq 0$
 まずは、 $\frac{\alpha}{\alpha+\beta} < \frac{y}{n}$ の場合を考える,

$$\frac{\alpha}{\alpha+\beta} < \frac{y}{n}, \quad \alpha + y < \frac{y(\alpha + \beta)}{n} + y, \quad \frac{\alpha + y}{\alpha + \beta + n} < \frac{y}{n}$$

練習問題 5

$$\frac{\alpha}{\alpha + \beta} < \frac{y}{n}, \quad n \left(\frac{\alpha}{\alpha + \beta} \right) + \alpha < y + \alpha, \quad \frac{\alpha}{\alpha + \beta} < \frac{y + \alpha}{\alpha + \beta + n}$$

$\frac{\alpha}{\alpha + \beta} > \frac{y}{n}$ の場合は不等式が逆転するだけである。 □

(c) 事前分布が一様分布 $U(0, 1)$ のとき, θ の事後分散が常に θ の事前分散よりも小さくなることを示してください

(9) 式参照

$$\text{Var}(\theta) = \frac{1}{12}, \quad \text{Var}(\theta | y) = \frac{(n - y + 1)(y + 1)}{(n + 2)^2(n + 3)}$$

$\text{Var}(\theta | y)$ が最大となる値 $y^* = n/2$ で置き換えると

$$\text{Var}(\theta | y^*) = \frac{1}{12 + 4n}$$

$n > 0$ であるから, $\text{Var}(\theta) > \text{Var}(\theta | y^*)$ は必ず成立する。 □

練習問題 5

- (d) 事前分布がベータ分布 $\text{Beta}(\theta | \alpha, \beta)$ のとき, θ の事後分散が事前分散よりも大きくなる場合を示して下さい

$$\text{Var}(\theta) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}, \quad \text{Var}(\theta | y) = \frac{(\alpha + y)(\beta + n - y)}{(\alpha + \beta + n)^2(\alpha + \beta + n + 1)}$$

例えば

$$\alpha = 1, \beta = 10, n = 20, y = 10$$

$$\text{Var}(\theta) = 0.00689, \quad \text{Var}(\theta | y) = 0.00715 \quad (31)$$

など

練習問題 6

- 事前分布が $\text{Gamma}(\theta_j | \alpha, \beta)$, 尤度が $\text{Poisson}(y_j | 10n_j\theta_j)$ のとき, (7) 式, (8) 式を用いて事後予測分布の期待値と分散を求めてください

期待値は

$$\begin{aligned} E(y_j) &= \int_0^{\infty} E(y_j | \theta_j) p(\theta_j) d\theta_j = \int_0^{\infty} 10n_j\theta_j \frac{\beta^\alpha}{\Gamma(\alpha)} \theta_j^{\alpha-1} \exp(-\beta\theta_j) d\theta_j \\ &= 10n_j \frac{\alpha}{\beta} \int_0^{\infty} \frac{\beta^{\alpha+1}}{\Gamma(\alpha+1)} \theta_j^{(\alpha+1)-1} \exp(-\beta\theta_j) d\theta_j = 10n_j \frac{\alpha}{\beta} \end{aligned}$$

分散は

$$\begin{aligned} \text{Var}(y_j) &= E(\text{Var}(y_j | \theta_j)) + \text{Var}(E(y_j | \theta_j)) \\ E(\text{Var}(y_j | \theta_j)) &= \int_0^{\infty} \text{Var}(y_j | \theta_j) p(\theta_j) d\theta_j = 10n_j \frac{\alpha}{\beta} \end{aligned}$$

練習問題 6

$$\begin{aligned}\text{Var}(E(y_j | \theta_j)) &= \int_0^{\infty} \{E(y_j | \theta_j) - E(E(y_j | \theta_j))\}^2 p(\theta_j) d\theta_j \\ &= \int_0^{\infty} \left(10n_j\theta_j - 10n_j\frac{\alpha}{\beta}\right)^2 p(\theta_j) d\theta_j \\ &= (10n_j)^2 \int_0^{\infty} \theta_j^2 \frac{\beta^\alpha}{\Gamma(\alpha)} \theta_j^{\alpha-1} \exp(-\beta\theta_j) d\theta_j - \left(10n_j\frac{\alpha}{\beta}\right)^2 \\ &= (10n_j)^2 \frac{\alpha}{\beta^2} \\ \text{Var}(y_j) &= 10n_j\frac{\alpha}{\beta} + (10n_j)^2 \frac{\alpha}{\beta^2}\end{aligned}$$

練習問題 7

- (a) 尤度が $\text{Bin}(y | n, \theta)$ のとき, $p(\theta) \propto \theta^{-1}(1 - \theta)^{-1}$ が指数型分布族の自然パラメータに対して一様な事前分布である事を示してください

スライド 38 より, 自然パラメータは $\phi = \text{logit}(\theta)$ であり, ϕ に対して一様な事前分布を設定する

$$p(\phi) = c$$

(29) 式のように, 確率変数の変換を行えば,

$$\begin{aligned} p(\theta) &= p(\phi) \left| \frac{d\phi}{d\theta} \right| = c \frac{d}{d\theta} \log \left(\frac{\theta}{1 - \theta} \right) \\ &\propto \theta^{-1}(1 - \theta)^{-1} \end{aligned}$$

練習問題 7

(b) $y = 0$ または n のとき, 事後分布が積分不能な分布であることを示しなさい

$p(\theta) \propto \theta^{-1}(1 - \theta)^{-1}$ は Beta(0, 0) なので, 事後分布は Beta($\theta \mid y, n - y$)
 $y = 0$ のとき

$$\begin{aligned} \int_0^1 p(\theta \mid y = 0) d\theta &\propto \int_0^1 (1 - \theta)^{n-1} \theta^{-1} d\theta \\ &= \left[(1 - \theta)^{n-1} \log(\theta) \right]_0^1 - (n - 1) \int_0^1 (1 - \theta)^{n-2} \log(\theta) d\theta \end{aligned} \quad (32)$$

上式の右辺の第 2 項は,

$$\begin{aligned} \int_0^1 (1 - \theta)^{n-2} \log(\theta) d\theta &= \left[(1 - \theta)^{n-2} \{-\theta + \theta \log(\theta)\} \right]_0^1 - \\ &\quad (n - 2) \int_0^1 (1 - \theta)^{n-3} \{-\theta + \theta \log(\theta)\} d\theta \end{aligned} \quad (33)$$

練習問題 7

ロピタルの定理より,

$$\lim_{\theta \rightarrow 0} \theta^k \log(\theta) = \lim_{\theta \rightarrow 0} \frac{\{\log(\theta)\}'}{\left(\frac{1}{\theta^k}\right)'} = \lim_{\theta \rightarrow 0} -\frac{\theta^k}{k} = 0$$

だから, (33) 式は

$$\int_0^1 (1-\theta)^{n-2} \log(\theta) d\theta = (0-0) - (n-2) \int_0^1 (1-\theta)^{n-3} \{-\theta + \theta \log(\theta)\} d\theta$$

$\log(\theta)$ を θ について k 回積分すれば, a_k, b_k を定数とすると
 $-a\theta^k + b\theta^k \log(\theta)$ であり, (32) 式は

$$\int_0^1 p(\theta | y=0) d\theta \propto \left[(1-\theta)^{n-1} \log(\theta) \right]_0^1 - 0 = +\infty$$

$y = n$ のときは $p(\theta | y = n) = \theta^{n-1}(1-\theta)^{-1}$ であるため置換積分を用いれば
 $y = 0$ の時と同じである

練習問題 8

- ある母集団から n 人の生徒をランダムにサンプリングし、体重を測定したところ、平均値は $\bar{y} = 150$ ポンドであった。母集団において体重が正規分布 $N(y | \theta, 20^2)$ にしたがって、 θ の事前分布に $N(\theta | 180, 40^2)$ を与える事にします

(a) θ に対する事後分布を求めてください

尤度は正規分布で、事前分布も正規分布だから、(17) 式、(18) 式より事後分布は $p(\theta | y) = N(\theta | \mu_n, \tau_n^2)$

$$\mu_n = \frac{\frac{n}{\sigma^2}\bar{y} + \frac{1}{\tau_0^2}\mu_0}{\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}} = \frac{60(10n + 3)}{4n + 1}, \quad \frac{1}{\tau_n^2} = \frac{n}{\sigma^2} + \frac{1}{\tau_0^2} = \frac{4n + 1}{1600}$$

(b) 同じ母集団から 1 人の生徒を追加でランダムにサンプリングし \tilde{y} が測定された。事後予測分布 $p(\tilde{y} | y)$ を求めてください

$$p(\tilde{y} | y) = \int_{-\infty}^{+\infty} p(\tilde{y} | y, \theta)p(\theta | y) d\theta = \int_{-\infty}^{+\infty} p(\tilde{y} | \theta)p(\theta | y) d\theta$$

練習問題 8

(16) 式より

$$p(\tilde{y} | y) = \int_{-\infty}^{+\infty} N(\tilde{y} | \theta, 20^2) N(\theta | \mu_n, \tau_n^2) d\theta = N(\tilde{y} | \mu_n, 20^2 + \tau_n^2)$$

- (c) $n = 10$ の時, θ の 95% 信用区間と, \tilde{y} の 95% 事後予測区間を求めてください
- (d) $n = 100$ の時も求めてください

	95% 信用区間	95% 事後予測区間
$n = 10$:	[138.5, 163.0]	[109.7, 191.8]
$n = 100$:	[146.2, 154.0]	[110.7, 189.5]

```
ex8 <- function(n) {
  mu <- 60 * (10 * n + 3) / (4 * n + 1)
  tau <- 1600 / (4 * n + 1)
  cat("\nn =", n, "\n")
  cat("central interval:", qnorm(c(0.025, 0.975), mu, sqrt(tau)), "\n")
  cat("predictive interval:", qnorm(c(0.025, 0.975), mu, sqrt(20^2 + tau)), "\n\n")
}
ex8( 10); ex8(100)
```


練習問題 9

- いま, カリフォルニア人が死刑制度に賛成する確率 θ についてしらべるために, 事前分布として期待値が 0.6, 標準偏差が 0.3 のベータ分布を与えたとします
- (a) θ に対する事前分布のハイパーパラメータ α, β を求め, グラフを描いてください

期待値が 0.6, 標準偏差が 0.3 のベータ分布だから

$$\begin{cases} \frac{\alpha}{\alpha + \beta} = 0.6 \\ \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} = 0.3^2 \end{cases}$$

これを解くと $\alpha = 1, \beta = 2/3$ より, 事前分布は Beta(1, 2/3) である

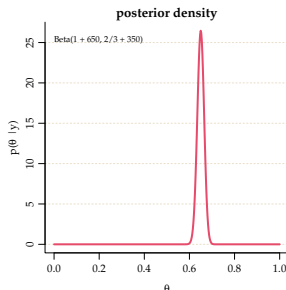
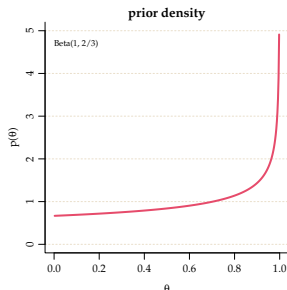
- (b) 1000 人のカリフォルニア人をランダムにサンプリングしたところ, 65% が死刑制度に賛成であった. θ に対する事後分布の期待値・分散を求め, グラフを描いてください

練習問題 9

二項分布に従うデータ $n = 1000$ 人分得て、賛成の人数 $y = 650$ 人であったので、事後分布は $\text{Beta}(1 + 650, 2/3 + 350)$ である
期待値と分散は

$$E(\theta | y) = \frac{\alpha + y}{\alpha + \beta + n} = 0.65$$

$$\text{Var}(\theta | y) = \frac{(\alpha + y)(\beta + n - y)}{(\alpha + \beta + n)^2(\alpha + \beta + n + 1)} = 0.00023$$



練習問題 10

- サンフランシスコには N 台のケーブルカーがあり、1 から N まで連番で番号が付けられています。あなたは 203 番のケーブルカーを見た後に、総数 N に関する推測を考えています

(a) N の事前分布に $E(N) = 100$ の幾何分布

$$p(N) = \frac{1}{100} \left(\frac{99}{100} \right)^{N-1}, \text{ for } N = 1, 2, \dots$$

を与えたとき、 N の事後分布を求めてください

N 台のケーブルカーが観測される確率が等しいと仮定してしまうと、尤度は

$$\Pr(Y = 203 | N) = 1 - \left(\sum_{i=1}^{202} p_i + \sum_{j=204}^N p_j \right) = \frac{1}{N}$$

練習問題 10

事後分布は

$$\begin{aligned}
 p(N | Y = 203) &= \frac{\Pr(Y = 203 | N)p(N)}{\sum_{k=203}^{+\infty} \Pr(Y = 203 | k)p(k)} = \frac{\left(\frac{99}{100}\right)^{N-1} / N}{\sum_{k=203}^{+\infty} \left(\frac{99}{100}\right)^{k-1} / k} \\
 &= \frac{\left(\frac{99}{100}\right)^N / N}{\left\{ \log(100) - \sum_{k=1}^{202} \left(\frac{99}{100}\right)^k / k \right\}} = \frac{21.47 \left(\frac{99}{100}\right)^N}{N}
 \end{aligned}$$

分母の計算には対数級数 $-\log(1-r) = \sum_{k=1}^{+\infty} \frac{r^k}{k}$; $|r| < 1$ を使う

(b) N の事後分布の期待値と標準偏差を求めてください

期待値は

$$E(N) = \sum_{k=203}^{+\infty} k \times \frac{21.47 \left(\frac{99}{100}\right)^k}{k} = 21.47 \frac{\left(\frac{99}{100}\right)^{203}}{1 - \left(\frac{99}{100}\right)} = 279.1$$

練習問題 10

分散および標準偏差は

$$\begin{aligned} \text{Var}(N) &= \left\{ \sum_{k=203}^{+\infty} k^2 \times \frac{21.47 \left(\frac{99}{100}\right)^k}{k} \right\} - \{\text{E}(N)\}^2 \\ &= \left\{ \frac{\left(\frac{99}{100}\right)}{\left\{1 - \left(\frac{99}{100}\right)\right\}^2} - \sum_{k=1}^{202} 21.47k \left(\frac{99}{100}\right)^k \right\} - \{\text{E}(N)\}^2 = 6394.33 \end{aligned}$$

$$\text{SD} = 79.96$$

- (c) N の適切な無情報事前分布を与えたときの、 N の事後分布、事後分布の期待値・標準偏差を求めてください

N は離散的なパラメータであり、事前分布にも離散的な一様分布

$p(N) = \frac{1}{N}$ を与えることにする

練習問題 10

すると、事後分布は

$$p(N | Y = 203) = \frac{1/N^2}{\sum_{k=203}^{+\infty} 1/k^2} = \left(\xi(2) - \sum_{k=1}^{202} \frac{1}{k^2} \right)^{-1} \frac{1}{N^2} = \frac{202.5}{N^2}$$

$$\text{ゼータ関数: } \xi(s) = \sum_{k=1}^{+\infty} \frac{1}{k^s}, \quad \xi(2) = \frac{\pi^2}{6}$$

期待値は発散し[†],

$$E(N) = \sum_{k=203}^{+\infty} k \frac{202.5}{k^2} = 202.5 \left(\xi(1) - \sum_{k=1}^{202} \frac{1}{k} \right) = +\infty$$

[†] 調和級数は発散する, $\xi(1) = \sum_{k=1}^{+\infty} \frac{1}{k} > \lim_{k \rightarrow +\infty} \int_1^{k+1} \frac{1}{x} dx = +\infty$

練習問題 10

分散および標準偏差も発散する

$$\text{Var}(N) = \sum_{k=203}^{+\infty} k^2 \frac{202.5}{k^2} - \{E(N)\}^2 = +\infty$$

ただし、以下を示す必要がある

$$\sum_{k=203}^{+\infty} 202.5 = \lim_{k \rightarrow +\infty} 202.5(k - 202) > \left\{ 202.5 \left(\xi(1) - \sum_{k=1}^{202} \frac{1}{k} \right) \right\}^2$$

これは、 $a = 202.5$, $b = 202$, $c = \sum_{i=1}^{202} \frac{1}{i}$ とおくと、

$$\lim_{k \rightarrow +\infty} \{a(\log(k) - c)\}^2 > \{a(\xi(1) - c)\}^2$$

$$\lim_{k \rightarrow +\infty} a(k - b) - \{a(\log(k) - c)\}^2 = +\infty$$

以上より示す事ができる

□

練習問題 11

- y_1, y_2, \dots, y_5 は互いに独立でコーシー分布 $p(y_i | \theta) \propto 1/(1 + (y_i - \theta)^2)$ にしたがうとし, $(y_1, y_2, \dots, y_5) = (-2, -1, 0, 1.5, 2.5)$ が得られたとします. 単純化するために, 事前分布には一様分布 $U(0, 1)$ を用います

- (a) $p(y | \theta)p(\theta)$ を $\theta = 0, \frac{1}{m}, \frac{2}{m}, \dots, 1$ (m には充分大きな値をとる) に分割して近似し (grid approach; pp.45–46), それを標準化して $p(\theta | y)$ の分布を図示してください

```

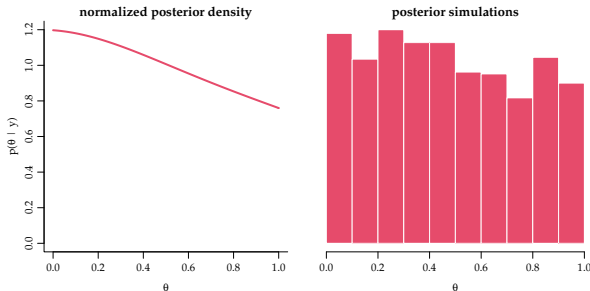
y <- NULL; m <- 1000; d <- c(-2.5, -1, 0, 1.5, 2.5)
u <- seq(0, 1, 1/m)
for (i in u) {
  y <- append(y, prod(dcauchy(d, i, 1)))
}
par(bty="l", family="serif")
plot(
  u, (m+1)*y/sum(y), type = "l", col = "#E64B6B",
  ylim = c(0, 1.2), xlab = expression(theta), lwd = 6,
  ylab = expression(paste(p, "(", theta, " | ", y, ")")),
  main = "normalized posterior density"
)

```


練習問題 11

- (b) 事後分布から 1000 回サンプリングを行って、ヒストグラムを描いてください

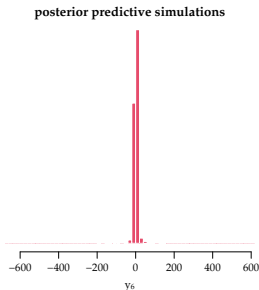
```
set.seed(11451355);
z <- sample(u, m, replace = TRUE, prob = y/sum(y))
hist(
  z, col = "#E64B6B", border = "white",
  yaxt = "n", ylab = NA, main = "posterior simulations",
  xlab = expression(theta)
)
```



練習問題 11

- (c) 前問でサンプリングした 1000 個の θ を使い, y_6 を追加した時の事後予測分布から 1000 回サンプリングし, ヒストグラムを描いてください

```
set.seed(45687993);
y6 <- rcauchy(m, z, 1)
hist(
  y6, breaks = 50, col = "#E64B6B", border = "white",
  yaxt = "n", ylab = NA, main = "posterior predictive simulations",
  xlab = expression(y[6])
)
```



練習問題 12

- 尤度が $p(y | \theta) = \text{Poisson}(y | \theta)$ のとき, Jeffreys' prior を見つけてください. また, 適当な α と β を見つけると, $\text{Gamma}(\alpha, \beta)$ になる事を示してください

$p(y | \theta) = \frac{\theta^y \exp(-\theta)}{y!}$ の Fisher 情報量は

$$J(\theta) = -E \left(\frac{d^2 \log p(y | \theta)}{d\theta^2} \middle| \theta \right) = \theta^{-1}$$

ジェフリーズのルールを用いると, $p(\theta) \propto \theta^{-1/2}$ である

ガンマ分布 $\frac{\beta^\alpha \theta^{\alpha-1}}{\Gamma(\alpha)} \exp(-\beta\theta)$ に当てはめれば, $\alpha = 1/2, \beta = 0$ の時の場合に
あたる

練習問題 13

- Table 2.2 は、航空機の定期便において 1 年間に起こった致命的事故の件数と死亡者数のデータである。死亡率の単位は 100 万人マイルとする。
- 表とは別に、1986 年のデータがあり、致命的事故の件数が 22、死亡者 546 名、死亡率 0.06 であった。

Table 2.2

Year	Fatal accidents	Passenger deaths	Death rate	Year	Fatal accidents	Passenger deaths	Death rate
1976	24	734	0.19	1981	21	362	0.06
1977	25	516	0.12	1982	26	764	0.13
1978	31	754	0.15	1983	20	809	0.13
1979	31	877	0.16	1984	16	223	0.03
1980	22	814	0.14	1985	22	1066	0.15

練習問題 13

- (a) 各年の致命的事故の件数は独立で $\text{Poisson}(y | \theta)$ に従うとします. 事前分布および事後分布は 1976 年から 1985 年のデータのみに基づくとします (無情報事前分布を用いる). このモデルのもとで, 1986 年の 95% 予測区間を求めてください.

練習問題 12 の結果から, 事前分布は $p(\theta) \propto \theta^{-1/2}$ であり, 事後分布は,

$$p(\theta | \mathbf{y}) = \frac{\theta^{t(\mathbf{y})} \exp(-n\theta) \times \theta^{-1/2}}{\int_0^{+\infty} \theta^{t(\mathbf{y})} \exp(-n\theta) \times \theta^{-1/2} d\theta} = \text{Gamma}\left(\theta \mid t(\mathbf{y}) + \frac{1}{2}, n\right)$$

ただし $t(\mathbf{y}) = \sum_{i=1}^n y_i$, 事後予測分布は,

$$\begin{aligned} p(\tilde{y} | \mathbf{y}) &= \int_0^{+\infty} \text{Poisson}(\tilde{y} | \theta) \times \text{Gamma}\left(\theta \mid t(\mathbf{y}) + \frac{1}{2}, n\right) d\theta \\ &= \frac{1}{\tilde{y}!} \frac{n^{t(\mathbf{y}) + \frac{1}{2}}}{\Gamma\left(t(\mathbf{y}) + \frac{1}{2}\right)} \frac{\Gamma\left(\tilde{y} + t(\mathbf{y}) + \frac{1}{2}\right)}{(n+1)^{\tilde{y} + t(\mathbf{y}) + \frac{1}{2}}} \end{aligned}$$

練習問題 13

したがって, 1986 年の致命的事故の件数の 95% 予測区間は,

```
ppda <- function(yt, y, n) {
  a <- y + 1/2
  exp( a*log(n) + lgamma(a+yt) - lgamma(a) -
        lgamma(yt+1) - (a+yt)*log(n+1) )
}
```

```
pred_inta <- function(upper, bound, srange,
  y, n) {
  tmp <- 0
  for(i in srange) {
    tmp <- tmp + ppda(i, y, n)
    if (tmp > bound) {
      return(
        i + upper + (bound - tmp + ppda(i, y, n)
          )
      )
    }
  }
}
```

```
* ppda(i, y, n)^(-1)
)
}
}
}

y <- sum(c(24, 25, 31, 31, 22, 21, 26, 20,
  16, 22))
n <- 10
L <- pred_inta(-1, 0.025, seq(1, 10000, 1),
  y, n)
U <- pred_inta( 1, 0.025, seq(10000, 1, -1),
  y, n)
L; U
```

95% predictive interval = [13.9, 35.1]

練習問題 13

- (b) 各年の致命的事故の件数は, x_i を暴露した人マイルとした Poisson($y_i \mid x_i\theta$) に従うとします. 先ほどと同様に無情報事前分布を用いて, 8×10^{11} 人マイルが暴露したと仮定した場合の 1986 年の 95% 予測区間を求めてください. 人マイルは Table 2.2 から求め, 丸め誤差は無視してください.
- (a) と同様に無情報事前分布を仮定すると, 事後分布は,

$$p(\theta \mid \mathbf{y}) = \text{Gamma}\left(\theta \mid t(\mathbf{y}) + \frac{1}{2}, n\bar{x}\right)$$

ただし $\bar{x} = \sum_{i=1}^n x_i$, 事後予測分布は,

$$\begin{aligned} p(\tilde{y} \mid \mathbf{y}) &= \int_0^{+\infty} \text{Poisson}(\tilde{y} \mid \tilde{x}\theta) \times \text{Gamma}\left(\theta \mid t(\mathbf{y}) + \frac{1}{2}, n\bar{x}\right) d\theta \\ &= \frac{\tilde{x}^{\tilde{y}} (n\bar{x})^{t(\mathbf{y})+\frac{1}{2}} \Gamma\left(\tilde{y} + t(\mathbf{y}) + \frac{1}{2}\right)}{\tilde{y}! \Gamma\left(t(\mathbf{y}) + \frac{1}{2}\right) (n\bar{x} + \tilde{x})^{\tilde{y}+t(\mathbf{y})+\frac{1}{2}}} \end{aligned}$$

練習問題 13

したがって, 1986 年の致命的事故の件数の 95% 予測区間は,

```
ppdb <- function(yt, xt, y, x) {
  a <- y + 1/2
  exp( yt*log(xt) + a*log(x) + lgamma(a+yt)
        - lgamma(a) - lgamma(yt+1) - (a+yt)*
          log(x+xt) )
}

pred_intb <- function(upper, bound, srange,
  xt, y, x) {
  tmp <- 0
  for(i in srange) {
    tmp <- tmp + ppdb(i, xt, y, x)
    if (tmp > bound) {
      return(
        i + upper + (bound - tmp + ppdb(i, xt, y
          , x))
        * ppdb(i, xt, y, x)^(-1)
      )
    }
  }
}
```

```
}
}
}

y <- sum(c(24, 25, 31, 31, 22, 21, 26, 20,
  16, 22))
pd <- c(734, 516, 754, 877, 814, 362, 764,
  809, 223, 1066)
dr <- c(0.19, 0.12, 0.15, 0.16, 0.14, 0.06,
  0.13, 0.13, 0.03, 0.15)
x <- sum(pd * 10E+7 * dr^(-1))
xt <- 8E+11
L <- pred_intb(-1, 0.025, seq(1, 10000, 1)
  , xt, y, x)
U <- pred_intb( 1, 0.025, seq(10000, 1, -1)
  , xt, y, x)
L; U
```

95% predictive interval = [21.4, 47.4]

練習問題 13

(c) (a) と同じ事を, 死亡数について行ってください

```
y <- sum(c(734, 516, 754, 877, 814, 362, 764, 809, 223, 1066))
n <- 10
L <- pred_inta(-1, 0.025, seq(1, 10000, 1), y, n)
U <- pred_inta(1, 0.025, seq(10000, 1, -1), y, n)
L; U
```

95% predictive interval = [638.0, 748.9]

(d) (b) と同じ事を, 死亡数について行ってください

```
y <- sum(c(734, 516, 754, 877, 814, 362, 764, 809, 223, 1066))
pd <- c(734, 516, 754, 877, 814, 362, 764, 809, 223, 1066)
dr <- c(0.19, 0.12, 0.15, 0.16, 0.14, 0.06, 0.13, 0.13, 0.03, 0.15)
x <- sum(pd * 10E+7 * dr^(-1))
xt <- 8E+11
L <- pred_intb(-1, 0.025, seq(1, 10000, 1), xt, y, x)
U <- pred_intb(1, 0.025, seq(10000, 1, -1), xt, y, x)
L; U
```

95% predictive interval = [903.4, 1035.3]

練習問題 13

(e) Table 2.2 に限らず、一般的なデータについて、(a)–(d) のうちどのモデルがもっとも妥当だと思うか理由とともに教えてください。

- 死亡者数について

死亡は恐らく飛行機単位で起こる事 (各イベントが独立ではない) がほとんどだと思われるので、ポアソン分布に従わない事象と考えられる。少なくとも事故件数よりはフィッティングが良くないだろうし、実際に (c), (d) の結果を見ると過大に推定されている可能性がある。

- x_i の考慮について

(b) では暴露した人マイルを考慮しているが、飛行機事故に対する暴露としては少しズレていると思う。飛行機事故に対する暴露ならば、マイルだけを考慮すれば良いと思う。

- 結論

(a) が一番ましたが、予測区間の幅は結構広い

練習問題 14

- 平均が未知の正規モデル $p(\theta) = N(\theta | \mu_0, \tau_0)$, $p(y | \theta) = N(y | \theta, \sigma^2)$ において

- (a) 事後分布 $N(\theta | \mu_1, \tau_1)$ と $N(\theta | \mu_n, \tau_n)$ について, (μ_1, τ_1) , (μ_n, τ_n) を導出する過程を示してください

見たら分かるので省略します; (15) 式, (18) 式

$$p(\theta | y) \propto \exp \left\{ -\frac{\theta^2}{2} \left(\frac{n}{\sigma^2} + \frac{1}{\tau_0^2} \right) \right\} \exp \left\{ \theta \left(\frac{n\bar{y}}{\sigma^2} + \frac{\mu_0}{\tau_0^2} \right) \right\}, \quad \mu_n = \frac{\frac{n}{\sigma^2} \bar{y} + \frac{1}{\tau_0^2} \mu_0}{\frac{n}{\sigma^2} + \frac{1}{\tau_0^2}}, \quad \frac{1}{\tau_n^2} = \frac{n}{\sigma^2} + \frac{1}{\tau_0^2} \quad (34)$$

- (b) データを 1 つずつ追加して事後分布を更新していく過程を示し, n 個目のデータを追加したとき, $N(\theta | \mu_n, \tau_n)$ になる事示してください

1 つ目のデータ y_1 を与えたときの事後分布

$$p(\theta | y_1) \propto \exp \left\{ -\frac{\theta^2}{2} \left(\frac{1}{\sigma^2} + \frac{1}{\tau_0^2} \right) \right\} \exp \left\{ \theta \left(\frac{y_1}{\sigma^2} + \frac{\mu_0}{\tau_0^2} \right) \right\}$$

練習問題 14

2 つ目のデータ y_2 を追加したときの事後分布

$$\begin{aligned}
 p(\theta | y_1, y_2) &\propto p(y_2 | y_1, \theta)p(\theta | y_1) \\
 &\propto \exp\left(-\frac{\theta^2}{2\sigma^2}\right)\exp\left(\frac{y\theta}{\sigma^2}\right) \times \exp\left\{-\frac{\theta^2}{2}\left(\frac{1}{\sigma^2} + \frac{1}{\tau_0^2}\right)\right\} \exp\left\{\theta\left(\frac{y_1}{\sigma^2} + \frac{\mu_0}{\tau_0^2}\right)\right\} \\
 &= \exp\left\{-\frac{\theta^2}{2}\left(\frac{2}{\sigma^2} + \frac{1}{\tau_0^2}\right)\right\} \exp\left\{\theta\left(\frac{y_1 + y_2}{\sigma^2} + \frac{\mu_0}{\tau_0^2}\right)\right\}
 \end{aligned}$$

更新を繰り返し、 n 個目を追加すれば、(34) 式に一致することが分かる

練習問題 15

- 以下を用いて

$$\int_0^1 u^{\alpha-1}(1-u)^{\beta-1} du = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = B(\alpha, \beta)$$

- $Z \sim \text{Beta}(\alpha, \beta)$ のとき, $E(Z^m(1-Z)^n)$; $m, n \in \mathbb{Z}^{0,+}$ を求めてください, また, Z の期待値と分散も求めてください

Z はベータ分布に従うので,

$$p(z) = \frac{z^{\alpha-1}(1-z)^{\beta-1}}{B(\alpha, \beta)} \left(= \frac{z^{\alpha-1}(1-z)^{\beta-1}}{\int_0^1 z^{\alpha-1}(1-z)^{\beta-1} dz} \right)$$

よって,

$$E(Z^m(1-Z)^n) = \int_0^1 z^m(1-z)^n \frac{z^{\alpha-1}(1-z)^{\beta-1}}{B(\alpha, \beta)} dz = \frac{B(\alpha+m, \beta+n)}{B(\alpha, \beta)}$$

練習問題 15

期待値は $m = 1, n = 0$ の時なので,

$$E(Z) = \frac{B(\alpha + 1, \beta)}{B(\alpha, \beta)} = \frac{\alpha}{\alpha + \beta}$$

また, $m = 2, n = 0$ の時は

$$E(Z^2) = \frac{B(\alpha + 2, \beta)}{B(\alpha, \beta)} = \frac{\alpha(\alpha + 1)}{(\alpha + \beta)(\alpha + \beta + 1)}$$

従って, 分散は

$$\text{Var}(Z) = E(Z^2) - \{E(Z)\}^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

練習問題 16

- y が n を与えたもとの、未知パラメータ θ の二項分布にしたがうとします。また、 θ の事前分布を $\text{Beta}(\theta | \alpha, \beta)$ とします

(a) 事後予測分布 $p(y)$ を求めてください。

尤度が二項分布で、事前分布がベータ分布なので、

$$p(y) = \frac{\binom{n}{y}}{B(\alpha, \beta)} \int_0^1 \theta^{\alpha+y-1} (1-\theta)^{\beta+n-y-1} d\theta = \binom{n}{y} \frac{B(\alpha+y, \beta+n-y)}{B(\alpha, \beta)}$$

この離散分布はベータ二項分布として知られている

$$\text{Beta-bin}(y | n, \alpha, \beta) = \begin{cases} \binom{n}{y} \frac{B(\alpha+y, \beta+n-y)}{B(\alpha, \beta)} & y = 0, 1, \dots, n \\ 0 & \text{other} \end{cases} \quad (35)$$

$n \in \mathbb{Z}^+, \alpha, \beta \in (0, \infty)$

練習問題 16

- (b) ベータ二項分布の確率が全ての y について等になるとき, 事前分布が $\text{Beta}(\theta | 1, 1)$ になる必要がある事を示してください

ガンマ関数の性質 $\Gamma(z+1) = z\Gamma(z)$ より,

$$\frac{B(\alpha + y, \beta + n - y)}{B(\alpha, \beta)} = \frac{\left(\prod_{i=0}^{y-1} \alpha + i\right) \Gamma(\alpha) \left(\prod_{j=0}^{n-y-1} \beta + j\right) \Gamma(\beta) \Gamma(\alpha + \beta)}{\left(\prod_{k=0}^{n-1} \alpha + \beta + k\right) \Gamma(\alpha + \beta) \Gamma(\alpha) \Gamma(\beta)}$$

確率変数 Y の取り得る値は $n+1$ 種類だから, 全ての y について確率が等しいとき, 以下を満たす

$$\frac{n!}{y!(n-y)!} \frac{\left(\prod_{i=0}^{y-1} \alpha + i\right) \left(\prod_{j=0}^{n-y-1} \beta + j\right)}{\left(\prod_{k=0}^{n-1} \alpha + \beta + k\right)} = \frac{1}{n+1}$$

これを満たす (α, β) は明らかに $(1, 1)$ の場合だけである

上昇階乗冪 $x^{\bar{n}} = \prod_{i=0}^{n-1} x + i$ で書くこともあります

練習問題 17

- Laplace に倣い, ある母集団における男児の出生確率 θ の解析をしてみましょう. ただし, 事前に θ についてより明確な信念を持っており, 事前分布として一様分布ではなく $\text{Beta}(\theta | 100, 100)$ を与えたとします

- (a) 与えた事前分布の意味が " θ が 0.5 より大きいのか未満なのかどうかはどっちつかずだが, 0.4 ~ 0.6 である確率が 95% 以上だと確信している" として解釈できる事を示してください

(0.4, 0.6) は 0.5 を含むし,

$$\text{Beta}(\theta \in (0.4, 0.6) | 100, 100) = 0.996$$

となるから, 0.4 ~ 0.6 である確率が 95% 以上だと確信していると言えるだろう

練習問題 17

- (b) 1000 人をランダムサンプリングしたところ, 511 人が男性でした.
 $p(\theta > 0.5 | y)$ を求めてください

尤度が二項分布, 事前分布がベータ分布なので, 事後分布は
 $\text{Beta}(\theta | \alpha + y, \beta + n - y)$, $y = 511$, $n - y = 489$ だから,

$$\text{Beta}(\theta > 0.5 | 611, 589) = 0.737$$

練習問題 18

- ある科学者が測定を行い、9個の独立な測定値を得たとします。測定器の性能から、標準偏差は0.12である事が分かっています。測定値は平均 θ が未知の正規分布 $N(\theta, 0.12^2)$ に従うと仮定できるとします。また、科学者は θ の事前分布に対し、無情報事前分布を与えたいと考えています。9個の測定値の標本平均が17.653であったとき、10個目の測定値 \tilde{y} の事後予測分布の99%を含む区間を求めてください。このような区間のことを99%許容区間 (tolerance interval) といいます。

問題より、尤度は $p(y | \theta) = N(y | \theta, \sigma^2 = 0.12^2)$ であるから、Jeffreys のルールを用いると、

$$J(\theta) = \frac{1}{\sigma^2}$$

σ^2 は既知であるから $p(\theta) \propto c$ (定数) を与えるすると、事後分布は $p(\theta | y) \propto N(y | \theta, \sigma^2) \times c$ である事が分かる

練習問題 18

事後予測分布は,

$$\begin{aligned}
 p(\tilde{y} | \mathbf{y}) &= \int_{-\infty}^{+\infty} p(\tilde{y} | \theta, \mathbf{y}) p(\theta | \mathbf{y}) d\theta \\
 &\propto \exp\left(-\frac{\tilde{y}^2}{2\sigma^2}\right) \int_{-\infty}^{+\infty} \exp\left(-\frac{\theta^2}{\sigma^2}\right) \exp\left\{\frac{\theta(\tilde{y} + t(\mathbf{y}))}{\sigma^2}\right\} d\theta \\
 &\propto \frac{\exp\left(-\frac{\tilde{y}^2}{2\sigma^2}\right)}{\exp\left(-\frac{\frac{1}{2}(\tilde{y} + t(\mathbf{y}))^2}{2\sigma^2}\right)} = \exp\left(-\frac{\tilde{y}^2 - 2t(\mathbf{y})\tilde{y} - t(\mathbf{y})^2}{2(2\sigma^2)}\right)
 \end{aligned}$$

$N(\tilde{y} | t(\mathbf{y}), 2\sigma^2)$ である事が分かる. ただし, $t(\mathbf{y}) = \frac{1}{n} \sum_{i=1}^n y_i$
 99% 許容区間として以下に示す区間を求めると,

$$\Pr(\tilde{y} \leq \tilde{y}_L) = \alpha/2, \quad \Pr(\tilde{y}_U \leq \tilde{y}) = \alpha/2$$

$$[\tilde{y}_L, \tilde{y}_U] = [17.2, 18.1]$$

練習問題 19

- 信用区間と違って, 最高事後密度区間は変換に対して不変ではないことを示してください.

ここで, σ^2 が既知として nv/σ^2 が χ_n^2 にしたがうとします. また, 無情報 (変則) 事前分布 $p(\sigma) \propto \sigma^{-1}$, $\sigma > 0$ を用いるとします

- (a) σ^2 に対する事前分布が $p(\sigma^2) \propto \sigma^{-2}$ となることを示してください

$Z = t(\sigma)$ なる変換を考えると,

$$p(z) = p(t^{-1}(z)) \left| \frac{dt^{-1}(z)}{dz} \right|$$

$t(\sigma) = \sigma^2$ のとき, 逆関数とその微分は, $t^{-1}(Z) = \sqrt{Z}$, $\frac{dt^{-1}(Z)}{dZ} = (2\sqrt{Z})^{-1}$, したがって

$$p(\sigma^2) = p(\sqrt{\sigma^2}) (2\sqrt{\sigma^2})^{-1} = p(\sigma)(2\sigma)^{-1} \propto \sigma^{-2}$$

練習問題 19

(b) σ^2 の 95% 最高事後密度区間が, σ の 95% 最高事後密度区間を二乗したものと異なることを示してください

nv/σ^2 が自由度 n のカイ二乗分布 χ_n^2 にしたがうので, 尤度は

$$p(v | \sigma, n) = \frac{2^{-n/2}}{\Gamma(n/2)} \left(\frac{nv}{\sigma^2}\right)^{\frac{n}{2}-1} \exp\left(-\frac{nv}{2\sigma^2}\right)$$

σ の事後分布は

$$p(\sigma | v, n) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}-1} \exp\left(-\frac{nv}{2\sigma^2}\right) \sigma^{-1} = (\sigma^2)^{-\frac{n}{2}+\frac{1}{2}} \exp\left(-\frac{nv}{2\sigma^2}\right)$$

σ^2 の事後分布は

$$p(\sigma^2 | v, n) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}-1} \exp\left(-\frac{nv}{2\sigma^2}\right) \sigma^{-2} = (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{nv}{2\sigma^2}\right)$$

$[\sigma_{L1}, \sigma_{U1}]; \sigma_{L1} < \sigma_{U2}$ が σ の最高事後密度区間, $[\sigma_{L2}^2, \sigma_{U2}^2]; \sigma_{L1}^2 < \sigma_{U2}^2$ が σ^2 の最高事後密度区間を与えると仮定する

練習問題 19

定義より, $p(\sigma_{L1} | v, n) = p(\sigma_{U1} | v, n)$, $p(\sigma_{L2}^2 | v, n) = p(\sigma_{U2}^2 | v, n)$ が成り立つ

したがって, σ の最高事後密度区間を二乗したものが, σ^2 の最高事後密度区間と等しくなるためには, 以下の連立方程式が成り立つ必要がある

$$\left\{ \begin{aligned} (\sigma_L^2)^{-\frac{n}{2} + \frac{1}{2}} \exp\left(-\frac{nv}{2\sigma_L^2}\right) &= (\sigma_U^2)^{-\frac{n}{2} + \frac{1}{2}} \exp\left(-\frac{nv}{2\sigma_U^2}\right) & (i) \end{aligned} \right.$$

$$\left\{ \begin{aligned} (\sigma_L^2)^{-\frac{n}{2}} \exp\left(-\frac{nv}{2\sigma_L^2}\right) &= (\sigma_U^2)^{-\frac{n}{2}} \exp\left(-\frac{nv}{2\sigma_U^2}\right) & (ii) \end{aligned} \right.$$

(ii) を $(\sigma_U^2)^{\frac{1}{2}}$ 倍して (i) から引くと

$$(\sigma_L - \sigma_U) (\sigma_L^2)^{-\frac{n}{2}} \exp\left(-\frac{nv}{2\sigma_L^2}\right) = 0$$

$\sigma_L = \sigma_U$ が得られ, これは定義と矛盾する. したがって, σ の最高事後密度区間を二乗したものは, σ^2 の最高事後密度区間と等しくならない \square

練習問題 20

- 率と期間でパラメトライズした場合の、共役事前分布を用いたポアソンモデルの事後分布の導出過程を示してください

n 個の観測値 \mathbf{y} が i.i.d. で、対応する観察期間 \mathbf{x} を与えたとき、尤度は

$$\begin{aligned} p(\mathbf{y} | \theta, \mathbf{x}) &= \prod_{i=1}^n \frac{(x_i \theta)^{y_i}}{y_i!} \exp(-x_i \theta) \\ &= \left(\prod_{i=1}^n \frac{x_i^{y_i}}{y_i!} \right) \exp \left\{ - \left(\sum_{i=1}^n x_i \right) \theta \right\} \exp \{ t(\mathbf{y}) \log(\theta) \} \end{aligned}$$

したがって事前分布は $\text{Gamma}(\alpha, \beta)$

$$p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \exp\{-\beta\theta\} \exp\{(\alpha - 1) \log(\theta)\}$$

練習問題 20

事後分布は

$$\begin{aligned}
 p(\theta \mid \mathbf{y}, \mathbf{x}) &= \frac{p(\mathbf{y} \mid \theta, \mathbf{x})p(\theta)}{p(\mathbf{y})} \\
 &= \frac{\exp\left\{-\left(\beta + \sum_{i=1}^n x_i\right)\theta\right\} \theta^{\alpha+t(\mathbf{y})-1}}{\int_0^{\infty} \exp\left\{-\left(\beta + \sum_{i=1}^n x_i\right)\theta\right\} \theta^{\alpha+t(\mathbf{y})-1} d\theta} \\
 &= \frac{\left(\beta + \sum_{i=1}^n x_i\right)^{\alpha+t(\mathbf{y})}}{\Gamma(\alpha + t(\mathbf{y}))} \exp\left\{-\left(\beta + \sum_{i=1}^n x_i\right)\theta\right\} \theta^{\alpha+t(\mathbf{y})-1} \\
 &= \text{Gamma}\left(\theta \mid \alpha + t(\mathbf{y}), \beta + \sum_{i=1}^n x_i\right) \\
 \int_0^{\infty} \exp(-\eta\theta) \theta^{\nu-1} d\theta &= \left(\frac{1}{\eta}\right)^{\nu-1} \int_0^{\infty} \exp(-z) z^{\nu-1} \frac{dz}{\eta} = \frac{\Gamma(\nu)}{\eta^{\nu}} \quad (36)
 \end{aligned}$$

練習問題 21

- (a) 尤度が死亡率 θ の指数分布にしたがうとき, θ の共役事前分布がガンマ分布になる事と, そのときの事後分布を示してください. ただし y は iid とします

観測値が n 個の場合, 尤度とその共役事前分布は,

$$p(\mathbf{y} | \theta) = \theta^n \exp\{-t(\mathbf{y})\theta\}, \quad p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} \exp(-\beta\theta)$$

ただし, $t(\mathbf{y}) = \sum_{i=1}^n y_i$
したがって, 事後分布は

$$\begin{aligned} p(\theta | \mathbf{y}) &= \frac{\theta^{\alpha+n-1} \exp[-\{\beta + t(\mathbf{y})\}\theta]}{\int_0^{+\infty} \theta^{\alpha+n-1} \exp[-\{\beta + t(\mathbf{y})\}\theta] d\theta} \\ &= \frac{\{\beta + t(\mathbf{y})\}^{\alpha+n}}{\Gamma(\alpha + n)} \theta^{\alpha+n-1} \exp[-\{\beta + t(\mathbf{y})\}\theta] \\ &= \text{Gamma}(\alpha + n, \beta + t(\mathbf{y})) \end{aligned}$$

練習問題 21

- (b) (a) において、平均死亡時間 $\phi = 1/\theta$ についてパラメトライズした場合に、共役事前分布が逆ガンマ分布になる事と、そのときの事後分布を示してください。

観測値が n 個の場合、尤度とその共役事前分布は、

$$p(\mathbf{y} | \phi) = \phi^{-n} \exp\left\{-\frac{t(\mathbf{y})}{\phi}\right\}, \quad p(\phi) = \frac{\beta^\alpha}{\Gamma(\alpha)} \phi^{-(\alpha-1)} \exp\left(-\frac{\beta}{\phi}\right)$$

事後分布は、

$$\begin{aligned} p(\phi | \mathbf{y}) &= \frac{\phi^{-(\alpha+n-1)} \exp\left\{-\frac{\beta+t(\mathbf{y})}{\phi}\right\}}{\int_{+\infty}^0 \phi^{-(\alpha+n-1)} \exp\left\{-\frac{\beta+t(\mathbf{y})}{\phi}\right\} d\phi} \\ &= \frac{\{\beta + t(\mathbf{y})\}^{\alpha+n}}{\Gamma(\alpha + n)} \phi^{-(\alpha+n-1)} \exp\left\{-\frac{\beta + t(\mathbf{y})}{\phi}\right\} \\ &= \text{Inv-Gamma}(\alpha + n, \beta + t(\mathbf{y})) \end{aligned}$$

練習問題 21

- (c) ある製造工程で作られる電球の寿命が、未知の故障率 θ をもつ指数分布にしたがうとします。事前分布として $CV(\theta) = 0.5 (= \sqrt{\text{Var}(\theta)}/E(\theta))$ のガンマ分布を仮定します。電球をサンプリングし寿命を調べ、 $CV(\theta | y)$ を 0.1 減らすためには、何個の電球が必要になるでしょう。

Gamma($\theta | \alpha, \beta$) の期待値と分散は、

$$E(\theta) = \int_0^{+\infty} \theta \times \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} \exp(-\beta\theta) d\theta = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{\beta^{\alpha+1}} = \frac{\alpha}{\beta}$$

$$\text{Var}(\theta) = E(\theta^2) - \{E(\theta)\}^2 = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+2)}{\beta^{\alpha+2}} - \frac{\alpha^2}{\beta^2} = \frac{\alpha}{\beta^2}$$

よって、以下を用いれば、必要な n の数を求めることができる

$$CV(\theta) = \alpha^{-1/2} = 0.5, \quad CV(\theta | y) = (\alpha + n)^{-1/2} = 0.4$$

$$n = 2.25$$

したがって、3 個以上あればよい

練習問題 21

(d) (c) において、 ϕ についてパラメトライズした場合、答えはどうなるでしょうか

Inv-Gamma(α, β) の期待値と分散は、

$$E(\phi) = \int_{+\infty}^0 \phi \times \frac{\beta^\alpha}{\Gamma(\alpha)} \phi^{-(\alpha-1)} \exp\left(-\frac{\beta}{\phi}\right) d\phi = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha-1)}{\beta^{\alpha-1}} = \frac{\beta}{\alpha-1}$$

$$\text{Var}(\phi) = E(\phi^2) - \{E(\phi)\}^2 = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha-2)}{\beta^{\alpha-2}} - \frac{\beta^2}{(\alpha-1)^2} = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$$

よって、以下を用いれば、必要な n の数を求めることができる

$$\text{CV}(\theta) = (\alpha-2)^{-1/2} = 0.5, \quad \text{CV}(\theta | y) = (\alpha-2+n)^{-1/2} = 0.4$$

$$n = 2.25$$

したがって、3 個以上あればよく、これは θ のままで求めた場合と等しい

練習問題 22

- (a) 尤度が死亡率 θ の指数分布にしたがうとし、その周辺 (事前) 分布が $\text{Gamma}(\theta | \alpha, \beta)$ とする. 観測値が $y \geq 100$ であるという結果が得られたとき、事後分布 $p(\theta | y \geq 100)$ および期待値と分散を求めてください.

$y \geq 100$ の場合の事後分布は,

$$\begin{aligned} p(\theta | y \geq 100) &\propto \text{Gamma}(\theta | \alpha, \beta) \times \int_{100}^{+\infty} \theta \exp(-y\theta) dy \\ &= \text{Gamma}(\theta | \alpha, \beta) \times \{\exp(-100\theta) - 0\} \\ p(\theta | y \geq 100) &= \text{Gamma}(\theta | \alpha, \beta + 100) \end{aligned}$$

$y \geq 100$ の場合の事後分布の期待値と分散は,

$$E(\theta | y \geq 100) = \frac{\alpha}{\beta + 100}, \quad \text{Var}(\theta | y \geq 100) = \frac{\alpha}{(\beta + 100)^2}$$

練習問題 22

- (b) (a) において, $y = 100$ に変えた場合の事後分布, および期待値と分散を求めてください.

$y = 100$ の場合の事後分布は,

$$p(\theta | y = 100) \propto \text{Gamma}(\theta | \alpha, \beta) \times \theta \exp(-100\theta)$$

$$p(\theta | y = 100) = \text{Gamma}(\theta | \alpha + 1, \beta + 100)$$

$y = 100$ の場合の事後分布の期待値と分散は,

$$E(\theta | y = 100) = \frac{\alpha + 1}{\beta + 100}, \quad \text{Var}(\theta | y = 100) = \frac{\alpha + 1}{(\beta + 100)^2}$$

練習問題 22

- (c) 情報量が増えたにもかかわらず, (b) の事後分布の分散が (a) に比べて大きくなる理由を説明してください. また, これはなぜ (8) 式 (p.37, 2.8 式) と矛盾しないのでしょうか?
- $p(\theta | y \geq 100)$ は得られるデータに制限が与えられている状態であり, $p(\theta | y)$ を求める場合と標本空間 (サンプリングモデル) が異なる. $y \geq 100$ と制限する事は特殊な周辺化 ($y > 0$ なら周辺化に一致) をしているのと同じで, データによる変動が減ると解釈出来るため, 分散が小さくなる事は特におかしくないと思われる.
 - (8) 式については, $y \geq 100$ などの条件を設けると, 既に周辺化に近い事をしてしまっているため, 事前予測分布が不可積分になって, $E(\text{Var}(\theta | y \geq 100))$ 等は求められない.

練習問題 23

- Section 2.7 のフットボールのデータを用いて、 σ^2 の共役事前分布を用いた解析を行ってみましょう

(a) データを見る前の段階での、 σ^2 に対する事前の知識について、 σ の期待値と標準偏差を用いて表現してください。ただし、事前分布には逆ガンマ分布を用いてください。

分散未知の正規モデルを仮定すると、事前分布は

$$p(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-(\alpha+1)} \exp\left(-\frac{\beta}{\sigma^2}\right)$$

σ の期待値および標準偏差は、

$$E(\sigma) = \frac{\sqrt{\beta} \Gamma\left(\alpha - \frac{1}{2}\right)}{\Gamma(\alpha)}, \quad \sqrt{\text{Var}(\sigma)} = \sqrt{\beta \left(\frac{1}{\alpha - 1} - \frac{\Gamma\left(\alpha - \frac{1}{2}\right)^2}{\Gamma(\alpha)^2} \right)} \quad (37)$$

練習問題 23

(37) 式では α, β を解く事ができなかったのを,

$$E(\sigma^2) = \frac{\beta}{\alpha - 1}, \quad SD(\sigma^2) = \left\{ \frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)} \right\}^{1/2}$$

を用いる事にした.

事前知識は特に無いので, 適当に

$$E(\sigma^2) = 10^2, \quad SD(\sigma^2) = 5^2$$

と置いて解くと,

$$\alpha = 18, \beta = 1700$$

練習問題 23

- (b) 今度は事前の知識として、 σ の 95% が 3–20 の間に入ると仮定します。2.5–97.5 パーセンタイル値が大体 3–20 になるように、逆ガンマ分布のパラメータを求めてください。

```
library(psc1)
c1 <- 0.001; c2 <- 0.001
a <- seq(1.4, 1.6, .001)
b <- seq( 41,  42, .001)
x <- matrix(rep(0, length(a)*length(b)), nrow=length(a))
y <- matrix(rep(0, length(a)*length(b)), nrow=length(a))
for (i in 1:length(a)) {
  for (j in 1:length(b)) {
    x[i,j] <- qgamma(0.025, a[i], b[j])^(1/2)
    y[i,j] <- qgamma(0.975, a[i], b[j])^(1/2)
  }
}
z <- which(x>3-c1 & x<3+c1 & y>20-c2 & y<20+c2, arr.ind=TRUE)
a[z[,1]]; b[z[,2]]
x[x>3-c1 & x<3+c1 & y>20-c2 & y<20+c2]
y[x>3-c1 & x<3+c1 & y>20-c2 & y<20+c2]
# answer
qgamma(0.025, 1.485, 41.839)^(1/2)
qgamma(0.975, 1.485, 41.839)^(1/2)
```

$$\alpha = 1.485, \beta = 41.839$$

練習問題 23

- (c) 今度は scaled inverse- χ^2 分布としてパラメトライズして, (a), (b) と同様の事を行ってください.

scaled inverse- χ^2 分布は, 逆ガンマ分布について $\alpha = \frac{\nu_0}{2}$, $\beta = \frac{\nu_0}{2}\sigma_0^2$ と置いたものなので,

$$p(\sigma^2) = \frac{(\nu_0/2)^{\nu_0/2}}{\Gamma(\nu_0/2)} \sigma_0^{\nu_0} (\sigma^2)^{-(\nu_0/2+1)} \exp\left(-\frac{\nu_0\sigma_0^2}{2\sigma^2}\right)$$

$$E(\sigma^2) = \frac{\nu_0\sigma_0^2}{\nu_0 - 2}, \quad \text{SD}(\sigma^2) = \left\{ \frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)} \right\}^{1/2}$$

練習問題 23

(a) から, $\alpha = \frac{v_0}{2} = 18$, $\beta = \frac{v_0}{2}\sigma_0^2 = 1700$ を解くと

$$v_0 = 36, \sigma_0^2 = 94.4$$

(b) から, $\alpha = \frac{v_0}{2} = 1.485$, $\beta = \frac{v_0}{2}\sigma_0^2 = 41.839$ を解くと

$$v_0 = 2.97, \sigma_0^2 = 28.174$$

```
library(geoR)
s <- rinvcchisq(10000, 2.97, 28.174)^(1/2)
signif(quantile(s, c(0.025, 0.975)), 3)
```

ほぼ 3-20 になる

練習問題 23

- (d) それぞれの事前分布の場合について、 σ^2 の事後分布を導き、グラフと σ の 95% 信用区間を求めてください。

事後分布は、(55) 式より、

$$p(\sigma^2 | \mathbf{y}) = \text{Scale-inv-}\chi^2 \left(\nu_0 + n, \frac{\nu_0 \sigma_0^2 + nt(\mathbf{d})}{\nu_0 + n} \right)$$

および、逆ガンマ分布としてパラメータを変換したものである (結果は一致するため省略する)

Section 2.7 と同様に、 $n = 672$, $t(\mathbf{d}) = 13.85^2$ を用いると、

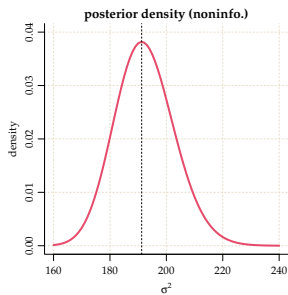
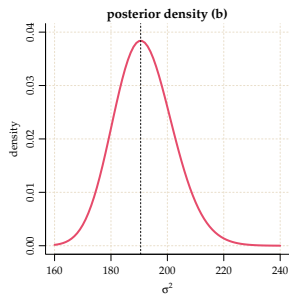
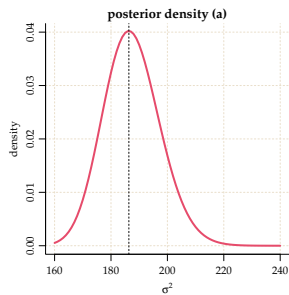
```
n <- 672; td <- 13.85^2;
nu0a <- 36; s0a <- 94.444
nu0b <- 2.97; s0b <- 28.174
nula <- nu0a + n
sla <- (nu0a * s0a + n * td) / nula
nulb <- nu0b + n
s1b <- (nu0b * s0b + n * td) / nulb
# 95% posterior interval of sigma (a)
signif(sqrt((nula * sla) / qchisq(0.975,
```

```
nula))), 3)
signif(sqrt((nula * sla) / qchisq(0.025,
nula))), 3)
# 95% posterior interval of sigma (b)
signif(sqrt((nulb * s1b) / qchisq(0.975,
nulb))), 3)
signif(sqrt((nulb * s1b) / qchisq(0.025,
nulb))), 3)
```

練習問題 23

σ の 95% 信用区間は

- (a) [13.0, 14.4]
 (b) [13.1, 14.6]
 (noninfo.) [13.1, 14.6]



練習問題 23

- (e) 以上の結果が無情報事前分布を用いた場合とどれくらい違うのか考察してください。
- 自分で適当に設定した事前分布を用いたとき、事後分布の形や位置が変わるものの、 σ の 95% 信用区間で見るとあまり変化していない
 - 事前分布の自由度が $\nu_0 = 36$ であったから、データのサンプルサイズ $n = 672$ に比べて小さくはあるものの、無情報事前分布に比べれば影響が大きいのではないかと考えられる
 - σ の 95% が 3–20 の間に入るような事前分布を用いたとき、結果はほとんど変わらない事が分かる
 - 事前分布の自由度が $\nu_0 = 2.97$ であり、データのサンプルサイズ $n = 672$ に比べて小さいため、無情報事前分布にかなり近い結果が得られたものと考えられる

参考文献 I

- [1] Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian Data Analysis. 2nd ed.* Chapman & Hall/CRC 2003.
- [2] Bayes T. An essay toward solving a problem in the doctrine of chances. *The Philosophical Transactions of the Royal Society* 1763; 53: 330–418. Reprinted by Barnard GA, in *Biometrika* 1958; 45: 293–315.
- [3] Laplace PS. *Mémoire sur les approximations des formules qui sont fonctions de très grands nombres.* Mémoires de l'Academie royale des sciences de Paris, année 1783; 1785; 1786.
- [4] Tierney L, Kadane JB. Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association* 1986; 81(393): 82–86.
- [5] De Finetti's theorem on exchangeable variables. *The American Statistician* 1976; 30(4): 188–189.
- [6] Syversveen AR. Noninformative bayesian priors. Interpretation and problems with construction and applications. Preprint Statistics 3, Department of Mathematical Sciences, NTNU, Trondheim, 1998.
- [7] Bernardo JM, Smith AFM. *Bayesian Theory.* John Wiley & Sons; 1994.
- [8] Galambos J. *Advanced Probability Theory.* Marcel Dekker Inc; 1988.
- [9] W. ルディン, 近藤基吉 (訳), 柳原二郎 (訳). 現代解析学. 共立出版 1971.

参考文献 II

- [10] 杉浦光夫. 解析入門 I. 東京大学出版会 1980.
- [11] 渡部洋. ベイズ統計学入門. 福村出版 1999.
- [12] 繁栞算男. ベイズ統計入門. 東京大学出版会 1985.

ラプラス近似^[4]

スカラー関数 $\exp\{h(\mathbf{b})\}$ の全領域の積分

$$\int_{\mathbf{b} \in R^q} \exp\{h(\mathbf{b})\} d\mathbf{b}$$

を正規近似により求める方法

関数 $h(\mathbf{b})$ を, $h(\mathbf{b})$ が最大値を取る値 $\hat{\mathbf{b}}$ の周りで 2 次まで Taylor 展開すると

$$\begin{aligned} h(\mathbf{b}) &\simeq h(\hat{\mathbf{b}}) + h'(\hat{\mathbf{b}})(\mathbf{b} - \hat{\mathbf{b}}) + \frac{1}{2}(\mathbf{b} - \hat{\mathbf{b}})' h''(\hat{\mathbf{b}})(\mathbf{b} - \hat{\mathbf{b}}) \\ &= h(\hat{\mathbf{b}}) + \frac{1}{2}(\mathbf{b} - \hat{\mathbf{b}})' h''(\hat{\mathbf{b}})(\mathbf{b} - \hat{\mathbf{b}}) \end{aligned}$$

1 階偏微分の項は $h'(\hat{\mathbf{b}}) = 0$

ラプラス近似

$$\begin{aligned}
 \int_{\mathbf{b} \in \mathbb{R}^q} \exp\{h(\mathbf{b})\} d\mathbf{b} &\simeq \int_{\mathbf{b} \in \mathbb{R}^q} \exp\left\{h(\hat{\mathbf{b}}) + \frac{1}{2}(\mathbf{b} - \hat{\mathbf{b}})^t h''(\hat{\mathbf{b}})(\mathbf{b} - \hat{\mathbf{b}})\right\} d\mathbf{b} \\
 &= \exp\{h(\hat{\mathbf{b}})\} \times \int_{\mathbf{b} \in \mathbb{R}^q} \exp\left\{\frac{1}{2}(\mathbf{b} - \hat{\mathbf{b}})^t h''(\hat{\mathbf{b}})(\mathbf{b} - \hat{\mathbf{b}})\right\} d\mathbf{b} \\
 &= \exp\{h(\hat{\mathbf{b}})\} \times (2\pi)^{q/2} | -h''(\hat{\mathbf{b}})^{-1} |^{1/2} \times \\
 &\quad \int_{\mathbf{b} \in \mathbb{R}^q} \frac{1}{(2\pi)^{q/2} | -h''(\hat{\mathbf{b}})^{-1} |^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{b} - \hat{\mathbf{b}})^t \{-h''(\hat{\mathbf{b}})^{-1}\}^{-1}(\mathbf{b} - \hat{\mathbf{b}})\right\} d\mathbf{b}
 \end{aligned}$$

最後の積分は多変量正規分布 $N(\hat{\mathbf{b}}, -h''(\hat{\mathbf{b}})^{-1})$ の密度関数の全領域を積分したものだから

$$\int_{\mathbf{b} \in \mathbb{R}^q} \exp\{h(\mathbf{b})\} d\mathbf{b} \simeq \exp\{h(\hat{\mathbf{b}})\} \times (2\pi)^{q/2} | -h''(\hat{\mathbf{b}})^{-1} |^{1/2} \times 1$$

多変量正規分布の積分を使って面積を近似するため、 $\exp\{h(\mathbf{b})\}$ の形状が正規分布に近くなければ精度が悪くなる

ラプラス近似

- 二項分布の分布関数を θ に関して積分した場合; $\int_0^1 \binom{n}{y} \theta^y (1 - \theta)^{n-y} d\theta$
- 右図の近似した場合は左端の余計な領域も積分している事に

$$\text{解析解} = \frac{1}{n+1}, \quad \text{ラプラス近似} = \frac{\sqrt{2\pi} \binom{n}{y} (y/n)^y (1 - y/n)^{n-y}}{\sqrt{n^3 / \{y(n-y)\}}}$$

