入門ベイズ統計
１．５確率の更新～１．８事前分布

９月１９日

田中洸一

内容

- 1.5 確率の更新
- 1.6 多数の原因
- 1.7 事後分布
- 1.8 事前分布

1.5 確率の更新(1)

- ”標的と射手”の問題について、標的Xにもう一度射たれた場合

前回の事後確率が今回の事前確率となる。

$$P(H_1) = \frac{8}{17}, P(H_2) = \frac{9}{17}$$

$$P(H_1 \mid F') = \frac{\left(\frac{8}{17}\right) \bullet (0.8)}{\left(\frac{8}{17}\right) \bullet (0.8) + \left(\frac{9}{17}\right) \bullet (0.3)} = \frac{64}{91} = 0.703$$

$$F' = \{2\text{回目にも標的X が射たれる}\}$$

1.5 確率の更新（2）

- 前回の事後確率が、今回の事前確率となる根拠

$$P(H_1) = 0.25, P(H_2) = 0.75$$

$$P(FF' \mid H_1) = (0.8)^2, P(FF' \mid H_2) = (0.3)^2$$

※$F \cap F' = FF'$

$$P(H_1 \mid FF') = \frac{(0.25) \bullet (0.8)^2}{(0.25) \bullet (0.8)^2 + (0.75) \bullet (0.3)^2} = \frac{64}{91} \qquad (1.5.3)$$

(1.5.2)と一致する

1.6 多数の原因（1）

・原因が3つ以上であってもベイズの定理は適用可能

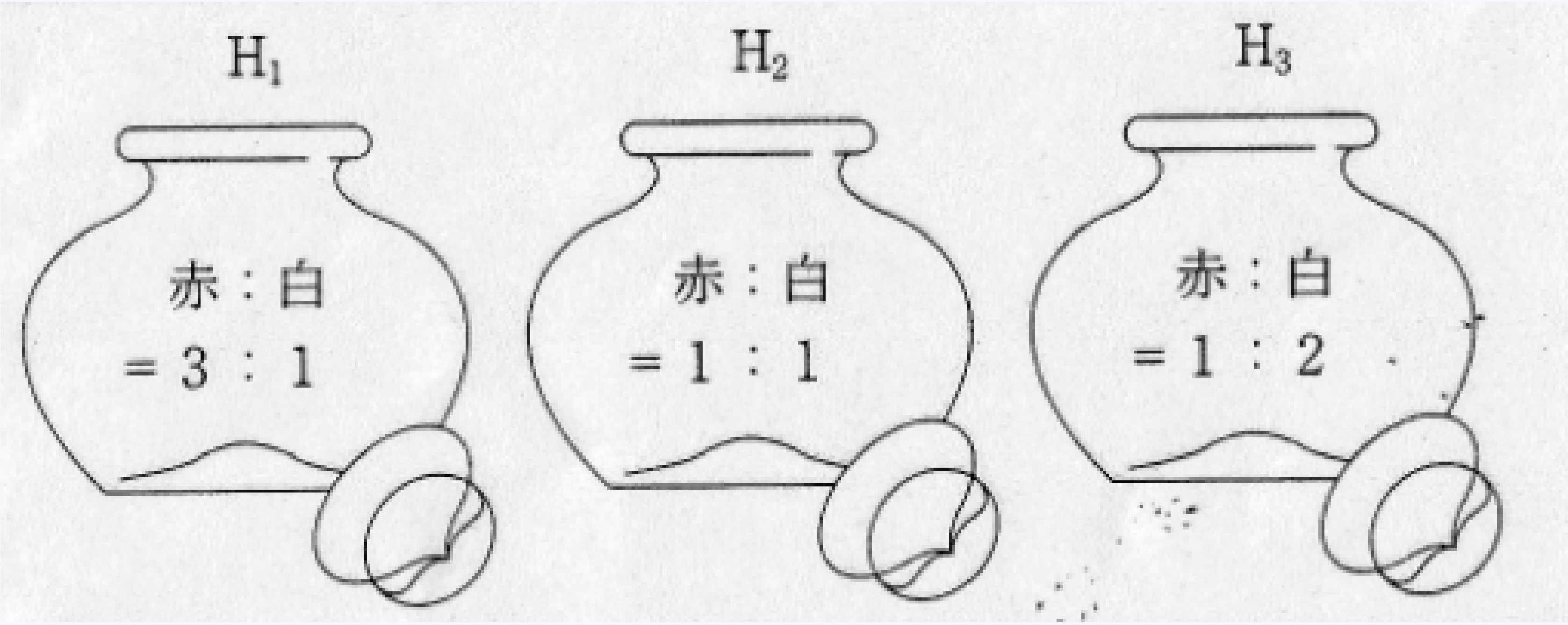


図1.2 壷から玉を抜き出す実験

1.6 多数の原因（2）

事象R：赤の玉が取り出される

$$P(R \mid H_1) = \frac{3}{4} = 0.75,\ P(R \mid H_2) = \frac{1}{2} = 0.50,\ P(R \mid H_3) = \frac{1}{3} = 0.33$$

理由不十分の原理から$P(H_1) = P(H_2) = P(H_3) = \dfrac{1}{3} = 0.33$

$$P(H_1 \mid R) = \frac{\left(\frac{1}{3}\right)\bullet\left(\frac{3}{4}\right)}{\left(\frac{1}{3}\right)\bullet\left(\frac{3}{4}\right)+\left(\frac{1}{3}\right)\bullet\left(\frac{1}{2}\right)+\left(\frac{1}{3}\right)\bullet\left(\frac{1}{3}\right)} = \frac{9}{19} = 0.47 \qquad (1.6.1)$$

1.6 多数の原因（3）

他も同様に、

$$P(H_2 \mid R) = \frac{6}{19} = 0.32, P(H_3 \mid R) = \frac{4}{19} = 0.21$$

表1.2　　壺H_1, H_2, H_3の確率

壺	実験前の確率	実験後の条件付確率	
		赤の玉	白の玉
H_1	0.33	0.47(++)	0.18(--)
H_2	0.33	0.32(-)	0.35(+)
H_3	0.33	0.21(--)	0.47(++)

1.6 多数の原因（4）

- 確率の更新の原理によって、2回目が赤のときの事後確率

$$P(H_1 \mid RR') = \frac{\left(\frac{9}{19}\right)\bullet\left(\frac{3}{4}\right)}{\left(\frac{9}{19}\right)\bullet\left(\frac{3}{4}\right)+\left(\frac{6}{19}\right)\bullet\left(\frac{1}{2}\right)+\left(\frac{4}{19}\right)\bullet\left(\frac{1}{3}\right)} = \frac{81}{133} = 0.61$$

$$P(H_2 \mid RR') = \frac{36}{133} = 0.27, P(H_3 \mid RR') = \frac{16}{133} = 0.12$$

表1.3　　赤の多い壺H_1に対する予想の上昇

壺（比率）	H_1(3：1)	H_2(1：1)	H_3(1：2)
1 回目赤	0.47	0.32	0.21
2 回目赤	0.61	0.27	0.12
3 回目赤	0.72	0.21	0.06
4 回目赤	0.81	0.16	0.03
5 回目赤	0.87	0.11	0.02

1.7 事後分布(1)

- 原因(A)をθ, 結果(B)をZ,
- 事前確率$P(A_i)$を$w(\theta_i)$, 事後確率$P(A_i \mid B)$を$w'(\theta_i \mid z)$,
- 確率$P(z \mid \theta_i)$をzの関数として$p(z \mid \theta_i)$と表記

ベイズの定理

$$w'(\theta_i \mid z) = \frac{w(\theta_i)\,p(z \mid \theta_i)}{\sum w(\theta_j)\,p(z \mid \theta_j)} \tag{1.7.1}$$

θが連続なら

$$w'(\theta \mid z) = \frac{w(\theta)\,p(z \mid \theta)}{\int_\Theta w(\theta)\,p(z \mid \theta)} \tag{1.7.2}$$

Θはθ全体の集合

1.7 事後分布（2）

- （1.7.1）,（1.7.2）は統計学的な考えによって、

・θは母集合の母数(パラメータ), zは標本,

$p(z\,|\,\theta)$は母数がθのときの標本の確率分布（尤度）

※(1.7.1),(1.7.2)の分母はθが含まれず定数、よって分子のみ表示

比例関数表記 $$w'(\theta\,|\,z) \propto w(\theta) \bullet p(z\,|\,\theta) \qquad (1.7.3)$$

事後に有する情報（＝事前に関する情報）×（情報　標本の出方）の情

1.8 事前分布

- ベイズ統計学では事前分布の選び方に、さまざまな方法がある。ここでは、「共役事前分布」について紹介。

- 例題「ある両親から、連続して男の子が3人生まれた。次の子が女の子である確率はどれほどか？」

- 男の子＝１、女の子＝０とおく。

$w_i = (i = 1, \cdots, n)$ は独立な確率変数

$x_i = 1$(確率θ), $x_i = 0$(確率$1-\theta$)

$z = (x_1, \cdots, x_n)$ の尤度 $$p(z|\theta) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i} \qquad (1.8.1)$$

1.8 事前分布

この形に合わせて, $w(\theta)$を母数α, βのベータ分布の形にとる

$$w(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1} \qquad (1.8.2)$$

比例定数は$\int_0^1 w(\theta)\,d\theta = 1$とすればよいので

$$B(\alpha,\beta) = \int_0^1 u^{\alpha-1}(1-\theta)^{\beta-1}\,du \qquad (1.8.3)$$

$1/B(\alpha,\beta)$として求 める

(1.7.3)の表記は、 $w'(\theta|z) \propto \theta^{\alpha'-1}(1-\theta)^{\beta'-1} \qquad (1.8.4)$

※$\alpha' = \alpha + \sum x_i, \beta' = \beta + n - \sum x_i \qquad (1.8.5)$

すなわ ち、ベータ分布$B(\alpha',\beta')$となる

1.8 事前分布

- 事前分布のα, βにおける、確率変数θの分布の期待値,
  分散, モ ード（最頻値）

期待値
$$E(\theta) = \frac{\alpha}{\alpha + \beta}$$

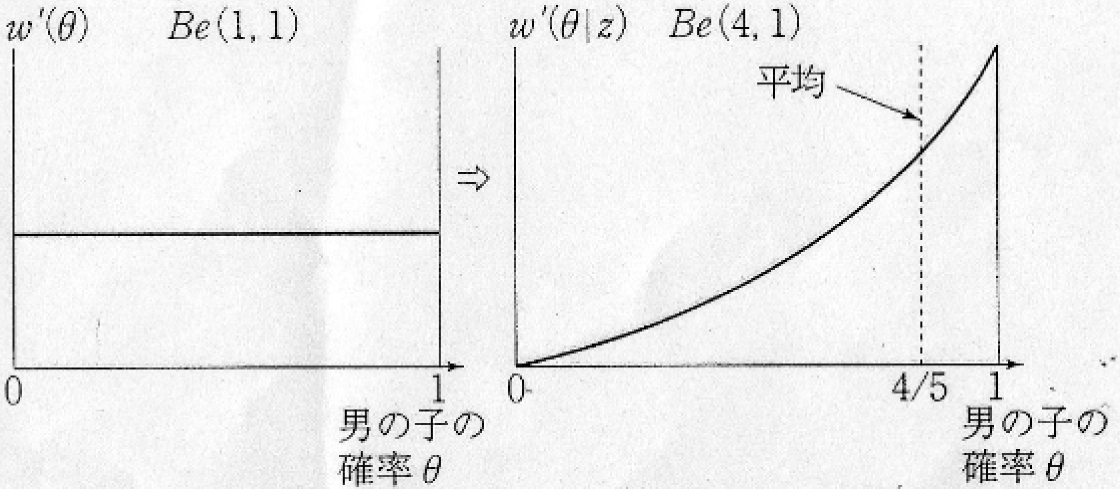
分 散
$$V(\theta) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

モ ード（最頻値）
$$Mode(\theta) = \frac{\alpha - 1}{(\alpha - 1) + (\beta - 1)}$$

1.8 事前分布

- 例題　一人目が男の子である確率 θ は、理由不十分の原理から $\frac{1}{2}$ 。つまり$\alpha=\beta$。

$\alpha=\beta=1$としてθに一様分布を仮定



生まれる前の事前分布

最初は θ については何の情報もなく，どの θ も一様に可能.

3 人の男の子の後の事後分布

この段階では，男の子の確率 θ はかなり高まったことになる.

1.8 事前分布

・ベイズ統計学による分析は、現象の尤度 $p(z|\bullet)$ に対して、
適切な事前分布 $w(\bullet)$ を選ぶことから始まる。

・適切な $w(\bullet)$ を選ぶには、$p(z|\theta)$ の関数型に応じて、
(1.7.3)の変換 $w \to w'$ が円滑にいく分布をとればよい。

↓

この分布を $p(z|\theta)$ の「自然な共役事前分布」という。

例：二項分布の共役事前分布はベータ分布

1.8 事前分布

- 正規分布の共役事前分布

$z=(x_1,\cdots,x_n)$ の尤度　$p(z|\theta)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x_i-\theta)^2}{2\sigma^2}\right\}$

$$=\exp\left\{-\frac{(\bar{x}-\theta)^2}{\frac{2\sigma^2}{n}}\right\}\bullet\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n\exp\left\{-\frac{\sum(x_i-\bar{x})^2}{2\sigma^2}\right\} \quad (1.8.7)$$

※$\bar{x}=\sum\frac{x_i}{n}$ （$\bar{x}$はθに対して十分であるという）

- θ の関数として必要な情報は第一因数のみ。
  よって、(1.7.3)より

$$w'(\theta\mid z)\propto w(\theta)\bullet\exp\left\{-\frac{(\bar{x}-\theta)^2}{\frac{2\sigma^2}{n}}\right\} \quad (1.8.8)$$

1.8 事前分布

- 事前確率分布は(1.8.8)の関数形に着目すれば、正規分布 $N(u,\tau^2)$ が適当

$$w(\theta) = \frac{1}{\sqrt{2\pi\tau}} \exp\left\{ -\frac{(\theta-\mu)^2}{2\tau^2} \right\} \qquad (1.8.9)$$

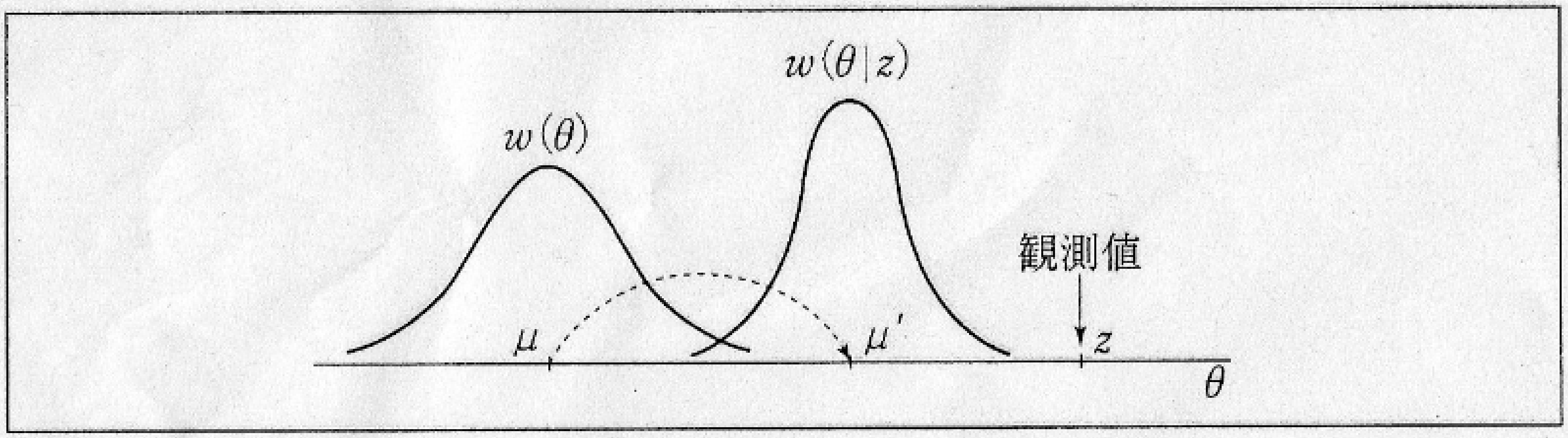


図 **1.5**　正規分布における事前確率分布から事後確率分布への更新

1.8 事前分布

- 二次式を操作すれば、事後確率は再び以下の形になり、$N\left(\mu', \tau'^2\right)$となる

$$w'\left(\theta \mid z\right) \propto \exp\left\{-\frac{(\theta-\mu)^2}{2\tau^2}\right\} \qquad (1.8.10)$$

※$\dfrac{1}{\tau'^2} = \dfrac{1}{\tau^2} + \dfrac{n}{\sigma^2}$　　※$\mu' = \dfrac{\left(\dfrac{1}{\tau^2}\right)\mu + \left(\dfrac{n}{\sigma^2}\right)\bar{x}}{\dfrac{1}{\tau^2} + \dfrac{n}{\sigma^2}}$

尤度が正規分布ならば、事前分布を<u>正規分布</u>にとれば、

事後分布も<u>正規分布</u>となる

1.8 事前分布

- ポアソン分布の共役事前分布

独立な確率変数 $x_i\left(i=1,\cdots,n\right)$

母数θ（平均θ）のポアソン分布に従う

$z=\left(x_1,\cdots,x_n\right)$ とする

尤度 $$p\left(z\,|\,\theta\right)=\prod_{i=1}^{n}e^{-\theta}\frac{\theta^{x_i}}{x_i!}\propto\theta^{-n\theta}\theta^{\sum x_i} \qquad (1.8.12)$$

$w\left(\theta\right)$を母数$\lambda,\ \alpha$のガンマ分布の形にとる

$$w\left(\theta\right)\propto\theta^{-\alpha\theta}\theta^{\lambda-1} \qquad (1.8.13)$$

比例定数は$\Gamma\left(s\right)=\int_0^\infty e^{-u}u^{s-1}du$(1.8.14) として $\dfrac{\alpha^\lambda}{\Gamma(\lambda)}$

1.8 事前分布

よって、

$$w'(\theta|z) \propto e^{-\alpha'\theta}\theta^{\lambda'-1} \quad (\alpha' = \alpha + n, \lambda' = \lambda + \sum x_i)$$

(1.8.15)

母数α', λ'のガンマ分布となる。

つまり、ポアソン　分布の共役　事前分布はガン　マ分布

α, λは以下の式で定　められる

$$E(\theta) = \frac{\lambda}{\alpha}, V(\theta) = \frac{\lambda}{\alpha^2} \quad (1.8.16)$$