第 8 章「統 計 的 推 測」

1．統計的決定理論（Statistical Decision Theory）

A. Wald により 1940 年代に体系化された，ゲーム理論の観点からの統計的推測理論.

標本空間 $\mathcal{X}$ **(Sample Space)：**
確率変数の実現値（観測値）からなる集まり

行動空間 $\mathcal{A}$ **(Action Space)：**
分析者が取りうる行動の全体

決定空間 $\mathcal{D}$ **(Decision Space)：**
標本空間から行動空間への関数である**決定関数** $\delta(x) : \mathcal{X} \to \mathcal{A}$ の全体

母数空間 Θ **(Parameter Space)：**
パラメータが取りうる値の全体

・損失関数 (Loss Function)
パラメータ θ に対して，行動 a を取るときの損失を表す.

$$w(\theta, a) : \Theta \times \mathcal{A} \to \mathcal{R}$$

・危険関数（Risk Function）

$$r(\theta, \delta) = E[w(\theta, \delta(\mathbf{X}))] = \int w(\theta, \delta(\mathbf{x}))\, f_n(\mathbf{x}|\theta)\, d\mathbf{x}$$

2．推定理論（Estimation Theory）

この場合の行動は，パラメータを推定することだから，行動空間＝母数空間．また，決定関数 $T_n : \mathcal{X} \to \Theta$ は**推定量** **(estimator)** と呼ばれる.

$$w(\theta, T_n) = (T_n - \theta)^2$$

$$r(\theta, T_n) = E(|T_n(\mathbf{X}) - \theta|^2) = \int |T_n(\mathbf{x}) - \theta|^2\, f_n(\mathbf{x}|\theta)\, d\mathbf{x}$$

$E(T_n(\mathbf{X})) = \theta$ ：不偏推定量 (unbiased estimator)

$b(\theta) = E(T_n(\mathbf{X})) - \theta$ ：偏り (bias)

クラメル・ラオの定理

不偏推定量 T_n の分散は，密度関数に関する適当な正則条件のもとで，次の下限をもつ.

$$V(T_n) \geq \frac{1}{I_n(\theta)} = \frac{1}{n\, I_1(\theta)}$$

ここで，$I_n(\theta)$ はサイズ n の標本に基づくフィッシャー情報量であり，等号は，微分方程式

$$\frac{d\log f_n(\mathbf{x}|\theta)}{d\theta} = I_n(\theta)\,(T_n-\theta)$$

をみたすときに成り立つ．

証明　まず，次のことが成り立つ．

$$E\left(\frac{d\log f_n(\mathbf{x}|\theta)}{d\theta}\right)=0, \qquad Cov\left(T_n, \frac{d\log f_n(\mathbf{x}|\theta)}{d\theta}\right)=1$$

したがって，コーシー・シュワルツの不等式から，

$$1 = Cov^2\left(T_n, \frac{d\log f_n(\mathbf{x}|\theta)}{d\theta}\right) \le V(T_n)\,V\left(\frac{d\log f_n(\mathbf{x}|\theta)}{d\theta}\right) = V(T_n)\,I_n(\theta)$$

となるから，定理の不等式が証明される．等号は，

$$\frac{d\log f_n(\mathbf{x}|\theta)}{d\theta} = a + b\,T_n$$

のとき成り立つから，両辺の期待値を取ると

$$0 = a + b\,\theta \qquad \Leftrightarrow \qquad a = -b\,\theta \qquad \Leftrightarrow \qquad \frac{d\log f_n(\mathbf{x}|\theta)}{d\theta} = b\,(T_n-\theta)$$

となる．これから，

$$1 = Cov\left(\frac{d\log f_n(\mathbf{x}|\theta)}{d\theta}, T_n\right) = b\,V(T_n) = \frac{b}{I_n(\theta)}$$

を得るので，$b = I_n(\theta)$ となり，定理は証明された．

有効推定量 (Efficient Estimator) クラメル・ラオの不等式の下限を達成するような推定量を有効推定量という．

例 1　$X \sim B(n,p)$ のとき，$\bar{X}$ は p の有効推定量である．実際，$E(\bar{X}) = p, V(\bar{X}) = p(1-p)/n$ であり，また，

$$\log f_n(X|p) = \log {}_nC_X + X\log p + (n-X)\log(1-p)$$

であるから，

$$\frac{d\log f}{dp} = \frac{X}{p} - \frac{n-X}{1-p} = \frac{X-np}{p(1-p)}$$

となり，フィッシャー情報量 $I_n(p)$ は　$n/(p(1-p))$ となる．この逆数は，$V(\bar{X})$ であるから，$\bar{X}$ は有効推定量である．なお，

$$\frac{d\log f}{dp} = \frac{X-np}{p(1-p)} = \frac{n}{p(1-p)}\,(\bar{X}-p) = I_n(p)\,(\bar{X}-p)$$

が成り立つことからも，有効性がいえる．

例２　$X_1, \cdots, X_n$ は $\mathrm{N}(\mu, 1)$ からのランダム標本であるとする．このとき，$\bar{X}$ は μ の有効推定量である．実際，$V(\bar{X}) = 1/n$，$I_n(\mu) = n$ となる．

一様最小分散不偏推定量（UMVUE）　θ の不偏推定量 $\hat{\theta}$ の分散が，他のいかなる不偏推定量の分散よりも θ に関して一様に小さいとき，$\hat{\theta}$ は UMVUE であるという．

例　$X_1, \cdots, X_n$ が平均 μ，分散 σ^2 の独立，同一分布に従うとき，μ の推定量のクラスを $X_1, \cdots, X_n$ の線形関数に限れば，$\bar{X}$ は UMVUE となる．実際，

$$\tilde{\mu} = c_1 X_1 + \cdots c_n X_n$$

が不偏であることから，$c_1 + \cdots + c_n = 1$ であり，

$$V(\tilde{\mu}) = \sum_{i=1}^{n} c_i^2 \, \sigma^2 = \sum_{i=1}^{n} (c_i - 1/n + 1/n)^2 \, \sigma^2 = \sum_{i=1}^{n} (c_i - 1/n)^2 \sigma^2 + \sigma^2/n$$

となることから，$c_i = 1/n \ (i = 1, \cdots, n)$ のとき，最小の分散をもつ．

・有効推定量と UMVUE との関係

有効推定量ならば UMVUE であるが，逆は必ずしもいえない．例えば，正規母集団 $\mathrm{N}(\mu, \sigma^2)$ からのランダム標本 $X_1, \cdots, X_n$ に対して，$S^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2/(n-1)$ は σ^2 の UMVUE であるが，有効ではない．

３．検定理論（Testing Theory）

行動空間は，$\mathcal{A} = \{a_0, a_1\}$ であり，

a_0: 帰無仮説 H_0 を受容する行動

a_1: 帰無仮説 H_0 を棄却する行動

決定関数 δ には棄却域と受容域が対応する．

$$W = \{\mathbf{x} : \delta(\mathbf{x}) = a_1\} \qquad \text{棄却域}$$

$$W^c = \{\mathbf{x} : \delta(\mathbf{x}) = a_0\} \qquad \text{受容域}$$

・第１種の誤りと第２種の誤り

パラメータ θ は，帰無仮説のもとで θ_0，対立仮説のもとで θ_1 の値を取るものとする．このとき，

第１種の誤りの確率:

$$\alpha = r(\theta_0, \delta) = \int_{\mathcal{X}} w(\theta_0, \delta(\mathbf{x})) \, f_n(\mathbf{x}|\theta_0) \, d\mathbf{x} = \int_{W} f_n(\mathbf{x}|\theta_0) \, d\mathbf{x}$$

第２種の誤りの確率:

$$\beta = r(\theta_1, \delta) = \int_{\mathcal{X}} w(\theta_1, \delta(\mathbf{x})) \, f_n(\mathbf{x}|\theta_1) \, d\mathbf{x} = \int_{W^c} f_n(\mathbf{x}|\theta_1) \, d\mathbf{x}$$

ここで，w は次のように定義される損失関数である．

$$w(\theta_0, \delta(\mathbf{x})) = \begin{cases} 1 & (\delta(\mathbf{x}) = a_1) \\ 0 & (\delta(\mathbf{x}) = a_0) \end{cases} \qquad w(\theta_1, \delta(\mathbf{x})) = \begin{cases} 1 & (\delta(\mathbf{x}) = a_0) \\ 0 & (\delta(\mathbf{x}) = a_1) \end{cases}$$

・**検定関数（Test function）**

$$\phi(\mathbf{x}) = \begin{cases} 1 & (\mathbf{x} \in W) \\ 0 & (\mathbf{x} \in W^c) \end{cases}$$

このとき，

$$\alpha = \int_{\mathcal{X}} \phi(\mathbf{x})\, f_n(\mathbf{x}|\theta_0)\, d\mathbf{x}, \qquad \beta = \int_{\mathcal{X}} \{1 - \phi(\mathbf{x})\}\, f_n(\mathbf{x}|\theta_1)\, d\mathbf{x}$$

・**検出力（Power）**

$= 1 - \beta$（1から第2種の誤りの確率をひいたもの）

・**Neyman-Pearson's fundamental lemma**

有意水準を α とする（＝第1種の誤りの確率を α 以下とする）とき，最強力検定は

$$\phi^*(x) = \begin{cases} 1 & (f_n(\mathbf{x}|\theta_1)/f_n(\mathbf{x}|\theta_0) > k) \\ 0 & (f_n(\mathbf{x}|\theta_1)/f_n(\mathbf{x}|\theta_0) < k) \\ p & (f_n(\mathbf{x}|\theta_1)/f_n(\mathbf{x}|\theta_0) = k) \end{cases}$$

で与えられる．ここで，定数 k と確率 p は，

$$\int \phi^*(\mathbf{x})\, f_n(\mathbf{x}|\theta_0)\, d\mathbf{x} = \alpha$$

をみたすように決められる．

証明　有意水準 α の任意の検定を ϕ とするとき，

$$\begin{aligned} \int \phi(\mathbf{x})\, f_n(\mathbf{x}|\theta_1)\, d\mathbf{x} - k\alpha &\leq \int \phi(x) f_n(\mathbf{x}|\theta_1)\, d\mathbf{x} - k \int \phi(\mathbf{x})\, f_n(\mathbf{x}|\theta_0)\, d\mathbf{x} \\ &= \int \phi(x)\, (f_n(\mathbf{x}|\theta_1) - k f(\mathbf{x}|\theta_0))\, d\mathbf{x} \\ &\leq \int \phi^*(x)\, (f_n(\mathbf{x}|\theta_1) - k f(\mathbf{x}|\theta_0))\, d\mathbf{x} \\ &= \int \phi^*(\mathbf{x})\, f_n(\mathbf{x}|\theta_1)\, d\mathbf{x} - k\alpha \end{aligned}$$

となる．したがって，

$$\int \phi^*(\mathbf{x})\, f_n(\mathbf{x}|\theta_1)\, d\mathbf{x} \geq \int \phi(\mathbf{x})\, f_n(\mathbf{x}|\theta_1)\, d\mathbf{x}$$

となり，証明された．

（注）　$\phi^*(\mathbf{x}) = p$（確率 p で帰無仮説を棄却）が必要なのは離散的な場合であり，上の証明の中で最後の等号を保証するものである．

例 1　$X_1, \cdots, X_n$ を $\mathrm{N}(\mu, \sigma^2)$ からのランダム標本とする．ただし，σ^2 は既知として，検定問題

$$H_0:\ \mu = \mu_0 \qquad \text{vs.} \qquad H_1:\ \mu = \mu_1\ (> \mu_0)$$

を考える．このとき，

$$f_n(\mathbf{x}|\mu) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

より，

$$\frac{f_n(\mathbf{x}|\mu_1)}{f_n(\mathbf{x}|\mu_0)} = \exp\left[\frac{\mu_1-\mu_0}{\sigma^2}\sum_{i=1}^{n}\left(x_i - \frac{\mu_0+\mu_1}{2}\right)\right] \quad > \quad k \quad \Leftrightarrow \quad \bar{x} > k'$$

有意点 k' は，与えられた有意水準 α に対して，

$$\begin{aligned}\alpha &= P(\bar{X} > k'|\mu_0) \\ &= P\left(\frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma} > \frac{\sqrt{n}(k'-\mu_0)}{\sigma}\middle|\mu_0\right) = P\left(\frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma} > z_\alpha\middle|\mu_0\right)\end{aligned}$$

と標準化することによって求められる．z_α は N(0, 1) の上側 $100\alpha\%$ 点である．したがって，最強力検定の棄却域は

$$W = \left\{\bar{x} > \mu_0 + z_\alpha\frac{\sigma}{\sqrt{n}}\right\}$$

となる．検出力は，

$$\begin{aligned}1-\beta &= P\left(\bar{X} > \mu_0 + z_\alpha\frac{\sigma}{\sqrt{n}}\middle|\mu_1\right) \\ &= P\left(\frac{\sqrt{n}(\bar{X}-\mu_1)}{\sigma} > z_\alpha - \frac{\sqrt{n}(\mu_1-\mu_0)}{\sigma}\middle|\mu_1\right)\end{aligned}$$

となり，$\mu_1 \to \infty$ のとき，1 に近づく．

数値例　あるメーカーの従来の電球の寿命は平均 1,500 時間，標準偏差 200 時間の正規分布に従うとする．新製品は標準偏差は従来通りであるが，平均は 10 % アップして，1,650 時間であると言われている．このことを有意水準 5 % で検定したい．標本サイズが 10 のとき，最強力検定の棄却域と検出力は次のようになる．

棄却域は，

$$\bar{x} > \mu_0 + z_\alpha\frac{\sigma}{\sqrt{n}} = 1500 + 1.645 \times \frac{200}{\sqrt{10}} = 1604$$

検出力は，

$$P(\bar{X} > 1604|\mu_1 = 1650) = P\left(Z > \frac{\sqrt{10}(1604-1650)}{200}\right) = P(Z > -0.7273) = 0.7664$$

例 2　$X_1, \cdots, X_n$ はポアソン分布 $Po(\lambda)$ に従うランダム標本であるとして，検定問題

$$H_0:\ \lambda = \lambda_0 \qquad \text{vs.} \qquad H_1:\ \lambda = \lambda_1\ (< \lambda_0)$$

を考える．

$$f_n(\mathbf{x}|\lambda) = \prod_{i=1}^{n} \frac{\exp(-\lambda)\,\lambda^{x_i}}{x_i!} = \exp(-n\lambda)\prod_{i=1}^{n}\frac{\lambda^{x_i}}{x_i!}$$

であるから，最強力検定の棄却域は

$$\frac{f_n(\mathbf{x}|\lambda_1)}{f_n(\mathbf{x}|\lambda_0)} = \exp(n(\lambda_0-\lambda_1))\prod_{i=1}^{n}\left(\frac{\lambda_1}{\lambda_0}\right)^{x_i} \quad > \quad k \qquad \Leftrightarrow \qquad S_n = \sum_{i=1}^{n} x_i < k'$$

となる．ここで，$S_n \sim Po(n\lambda)$ である．この場合の検定は，一般に確率化検定となる．例えば，有意水準 $\alpha = 0.1$, $n\lambda_0 = 3$ のとき，

$$P(S_n = 0|n\lambda_0 = 3) = e^{-3} = 0.05, \qquad P(S_n = 1|n\lambda_0 = 3) = 3\,e^{-3} = 0.15$$

であるから，$S_n = 0$ を棄却域とすれば検定のサイズは 0.05．また，$S_n = 0, 1$ を棄却域とすれば検定のサイズは 0.2 となってしまう．そこで，$S_n = 0$ のときには常に棄却，$S_n = 1$ のときには確率 1/3 で棄却する確率化検定を考えれば，そのサイズは

$$P(S_n = 0|n\lambda_0 = 3) + \frac{1}{3}\times P(S_n = 1|n\lambda_0 = 3) = e^{-3} + \frac{1}{3}\times 3\times e^{-3} = 0.1$$

となる．検出力は，例えば，$n\lambda_1 = 2$ に対しては，

$$P(S_n = 0|n\lambda_1 = 2) + \frac{1}{3}\times P(S_n = 1|n\lambda_1 = 2) = e^{-2} + \frac{1}{3}\times 2\times e^{-2} = 0.23$$

４．回帰理論（Regression Theory）

2 つの確率変数 X と Y があるとき，Y の値を X の関数 $\hat{Y} = h(X)$ によって推定することを，Y を X に回帰するといい，$h(X)$ を回帰関数（regression function）という．この場合，行動空間 $\mathcal{A}$ ＝標本空間 $\mathcal{X}$，決定関数 $\delta(X) = h(X)$ である．損失関数を 2 乗誤差で定義すると，危険関数は

$$r(h) = E[(Y - h(X))^2] = \int\int (y - h(x))^2\, f(x, y)\, dx\, dy$$

となる．

定理　上のリスクを最小にする $h(X)$ は，

$$h^*(X) = E[Y|X]$$

である．

証明

$$\begin{aligned} E[(Y - h(X))^2] &= E[\{Y - E(Y|X) - (h(X) - E(Y|X))\}^2] \\ &= E[(Y - E(Y|X))^2] + E[(h(X) - E(Y|X))^2] \\ &\qquad + 2E[(Y - E(Y|X))(h(X) - E(Y|X))] \end{aligned}$$

において，最後の項は，期待値の繰り返しの公式を使えば 0 になることがわかる．したがって，リスクは $h(X) = E(Y|X)$ のとき最小となる．

・最良線形回帰（Best Linear Regression）

条件付き期待値の計算が困難な場合，線形回帰 $Y_L = a + bX$ のクラスでリスクを最小にするものを考えることが多い．それは，

$$a = E(Y) - bE(X), \qquad b = Cov(X,Y)/V(X) = \rho_{XY}\sqrt{V(Y)/V(X)}$$

で与えられる．

例　確率変数 X と Y の同時密度関数が

$$f(x,y) = 2(x+2y)/3 \qquad (0 \le x \le 1,\ 0 \le y \le 1)$$

であたえられているとき，最良回帰と最良線形回帰を求めよう．まず，

$$f_1(x) = \int_0^1 f(x,y)\,dy = 2(x+1)/3$$

$$f_2(y|x) = f(x,y)/f_1(x) = (x+2y)/(x+1)$$

であるから，最良回帰は

$$E(Y|X) = \int_0^1 y\, f_2(y|x)\,dy = (3x+4)/(6x+6)$$

となる．このときのリスクは $E(V(Y|X))$ であるが，

$$V(Y|X) = E(Y^2|X) - E^2(Y|X) = \frac{1}{12} - \frac{1}{36(x+1)^2}$$

となることから，

$$E(V(Y|X)) = \frac{1}{12} - \frac{\log 2}{54} = 0.07049727$$

を得る．他方，最良線形回帰は，

$$E(X) = 5/9, \quad E(Y) = 11/18, \quad V(X) = 13/162, \quad V(Y) = 23/324,$$

$$Cov(X,Y) = -1/162, \quad \rho_{XY} = -\sqrt{2/299}$$

を使って，$b = Cov(X,Y)/V(X) = -1/13$，$a = E(Y) - bE(X) = 17/26$ を得るから，

$$Y_L^* = \frac{17-2x}{26}$$

となる．このときのリスクは，

$$V(Y)(1-\rho_{XY}^2) = \frac{23}{324}\left(1 - \frac{2}{299}\right) = 0.07051282$$

となる．

５．ベイズ解とミニマックス解

・ベイズ解

パラメータ θ が確率変数の実現値であるという立場に立つとき，θ として実現する前の分布（**事前分布**）として $\pi(\theta)$ を考えることができる．このとき，リスクの事前分布に関する期待値

$$r_B(\delta) = \int r(\theta,\delta)\,\pi(\theta)\,d\theta$$

をベイズ・リスクといい，ベイズ・リスクを最小にするような決定関数，すなわち

$$r_B(\delta_B) = \min\{r_B(\delta) : \delta \in \mathcal{D}\}$$

もみたす δ_B をベイズ解という．

・事後分布と事後危険（Posterior Distribution and Risk）

$f_n(\mathbf{x},\theta) = \pi(\theta)\,f_n(\mathbf{x}|\theta)$ を $\mathbf{x}$ と θ の同時密度関数とするとき，

$$\pi(\theta|\mathbf{x}) = \frac{f_n(\mathbf{x},\theta)}{\int f_n(\mathbf{x},\theta)\,d\theta}$$

を，θ の事後分布という．損失関数の事後分布による期待値

$$r(\delta(\mathbf{x})|\mathbf{x}) = \int w(\theta,\delta(\mathbf{x}))\,\pi(\theta|\mathbf{x})\,d\theta$$

を $\delta(\mathbf{x})$ の事後リスクという．事後リスクは $\mathbf{X}=\mathbf{x}$ の関数であるが，ここでさらに X の分布に関する期待値をとれば，

$$\begin{aligned}
\int r(\delta(\mathbf{x})|\mathbf{x})\,f_n(\mathbf{x})\,d\mathbf{x} &= \int\int w(\theta,\delta(\mathbf{x}))\,\pi(\theta|\mathbf{x})\,f_n(\mathbf{x})\,d\theta\,d\mathbf{x} \\
&= \int\int w(\theta,\delta(\mathbf{x}))\,\frac{\pi(\theta)\,f_n(\mathbf{x}|\theta)}{f_n(\mathbf{x})}\,f_n(\mathbf{x})\,d\theta\,d\mathbf{x} \\
&= \int r(\theta,\delta)\,\pi(\theta)\,d\theta \\
&= r_B(\theta)
\end{aligned}$$

となる．すなわち，事後リスクの期待値はベイズ・リスクである．このことから，次の定理が成立する．

定理　事後リスクを最小にする決定関数 δ^* はベイズ解である．

証明　任意の決定関数に対して，

$$r(\delta(\mathbf{x})|\mathbf{x}) \geq r(\delta^*(\mathbf{x})|\mathbf{x})$$

が成り立つから，両辺を X の分布に関して平均すれば $r_B(\delta) \geq r_B(\delta^*)$ を得る．

定理　損失関数が 2 乗誤差 $w(\theta,T) = (T-\theta)^2$ で与えられるとき，ベイズ解は事後平均

$$E(\theta|\mathbf{x}) = \int \theta\,\pi(\theta|\mathbf{x})\,d\theta$$

で与えられる．

証明　この場合の事後リスクは，

$$r(T(\mathbf{x})|\mathbf{x}) = \int (\theta - T(\mathbf{x}))^2\,\pi(\theta|\mathbf{x})\,d\theta$$

であり，これを最小にする $T(\mathbf{x})=E(\theta|\mathbf{x})$，すなわち事後平均である．

・ミニマックス解　リスク $r(\theta,\delta)$ の θ に関する最大値を最小とするような決定関数 δ_M をミニマックス解という．すなわち，

$$r_M(\delta_M)=\min_{\delta}\ \max_{\theta}\ r(\theta,\delta)$$

定理　ある事前分布 π に対して，ベイズ解 δ_B のリスク $r(\theta,\delta_B)$ が定数（θ に無関係）ならば，δ_B はミニマックス解である．

証明　δ_B のリスク $r(\theta,\delta_B)$ が定数で，事前分布 π に対するベイズ解ならば，

$$\min_{\delta}\int r(\theta,\delta)\,\pi(\theta)\,d\theta=\int r(\theta,\delta_B)\,\pi(\theta)\,d\theta=c$$

明らかに，

$$\max_{\theta} r(\theta,\delta)\geq\int r(\theta,\delta)\,\pi(\theta)\,d\theta$$

であるから，

$$\min_{\delta}\max_{\theta} r(\theta,\delta)\geq\min_{\delta}\int r(\theta,\delta)\,\pi(\theta)\,d\theta=c$$

他方，δ_B のリスクは θ に無関係な定数 c であることから，

$$c=\max_{\theta} r(\theta,\delta_B)\geq\min_{\delta}\max_{\theta} r(\theta,\delta)$$

以上から，

$$\max_{\theta} r(\theta,\delta_B)=\min_{\delta}\max_{\theta} r(\theta,\delta)=c$$

となり，δ_B はミニマックス解となる．

例　成功の確率 p を推定するために，n 回の独立なベルヌーイ試行 $X_1,\cdots,X_n$ を行うとき（各 X_i が取る値は 0, 1），尤度関数は

$$f_n(\mathbf{x}|p)=p^x\,(1-p)^{n-x},\qquad x=\sum_{i=1}^{n}x_i$$

である．推定量 $\bar{X}$ は不偏であり，2 乗誤差を損失関数とするとき，リスクは

$$r(p,\bar{X})=\frac{p(1-p)}{n}$$

である．次に，ベイズ解を考えよう．p の事前分布を

$$\pi(p)=B(\alpha,\beta)^{-1}\,p^{\alpha-1}\,(1-p)^{\beta-1}$$

とするとき，$(\mathbf{x},p)$ の同時分布は，

$$f_n(\mathbf{x},p)=B(\alpha,\beta)^{-1}\,p^{x+\alpha-1}\,(1-p)^{n-x+\beta-1}$$

となる．したがって，$\mathbf{x}$ の周辺密度は，

$$f_n(\mathbf{x}) = \int f_n(\mathbf{x}, p)\, dp = \frac{B(x+\alpha, n-x+\beta)}{B(\alpha,\beta)}$$

となる．また，p の事後分布は

$$\pi(p|\mathbf{x}) = \frac{f_n(\mathbf{x}, p)}{f_n(\mathbf{x})} = B(x+\alpha, n-x+\beta)^{-1}\, p^{x+\alpha-1}\,(1-p)^{n-x+\beta-1}$$

となる．すなわち，p の事前分布$= B_E(\alpha,\beta)$ は事後分布$= B_E(x+\alpha, n-x+\beta)$ となる．2 乗誤差を損失関数とする場合，ベイズ推定量（ベイズ解）は事後分布 $B_E(x+\alpha, n-x+\beta)$ の平均

$$T_B = \frac{x+\alpha}{n+\alpha+\beta} = \frac{\bar{x} + \frac{\alpha}{n}}{1 + \frac{\alpha+\beta}{n}}$$

であり，そのリスクは，$X = \sum_{i=1}^n X_i \sim B(n,p)$ であることから，

$$\begin{aligned} E(T_B - p)^2 &= E\left(\frac{X+\alpha}{n+\alpha+\beta} - p\right)^2 \\ &= \frac{((\alpha+\beta)^2 - n)p^2 + (n - 2\alpha(\alpha+\beta))p + \alpha^2}{(n+\alpha+\beta)^2} \end{aligned}$$

となる．ここで，$\alpha = \beta = \sqrt{n}/2$（$=$事前分布が $B_E(\sqrt{n}/2, \sqrt{n}/2)$）ならば，ベイズ解のリスクは定数となるから，このときはミニマックス解でもある．