章末問題解答例

竹村彰通

2026 年 4 月 17 日版

ここでは『現代数理統計学』の新装改訂版 (2020 年 11 月発行) の章末問題の解答例を示す．解答例の誤りや不完全な部分については，今後も改訂していく．

第 2 章

問 2.1　$(X-\mu)^k = \sum_{i=0}^{k} \binom{k}{i} (-1)^{k-i} X^i \mu^{k-i}$ の両辺の期待値をとれば，

$$\mu_k = \sum_{i=0}^{k} \binom{k}{i} (-1)^{k-i} \mu'_i \mu^{k-i}$$

である．同様に $X^k = ((X-\mu)+\mu)^k = \sum_{i=0}^{k} \binom{k}{i} (X-\mu)^i \mu^{k-i}$ の両辺の期待値をとれば

$$\mu'_k = \sum_{i=0}^{k} \binom{k}{i} \mu_i \mu^{k-i}$$

である．

問 2.2　各 y について y の「逆像」を $A_y = \{x \mid g(x) = y\}$ とおくと $p_Y(y) = P(Y=y) = \sum_{x\in A_y} p(x)$ となる．

$$E[g(X)] = \sum_x g(x)p(x) = \sum_y \sum_{x\in A_y} g(x)p(x) = \sum_y y p_Y(y)$$

より一致することがわかる．

問 2.3　連続変数の場合で示す．また $0 < h < k$ は正整数とする．$|x| \le 1$ のときは $|x|^h$ も $|x|^k$ も 1 以下であるから，この範囲での積分は有限である．次に $x > 1$ で $x \to \infty$ のときを考えると $x^h/x^k = 1/x^{k-h} \le 1$ であるから，$\int_1^\infty x^k f(x)\,dx < \infty$

ならば $\displaystyle\int_1^\infty x^h f(x)\,dx < \infty$ である．$x < -1$ のときは絶対値をとって議論すればよい．従って k 次のモーメントが存在すれば h 次のモーメントも存在する．

問 2.4　$0 < u < 1$ に対し $x_L = {F_L}^{-1}(u), x_R = {F_R}^{-1}(u)$ とおくと，$x_L \le x_R$ は本文に示されている．また (2.25) 式より $x \ge x_L \Leftrightarrow F(x) \ge u$ である．同様に，分布の右裾を考えれば $x \le x_R \Leftrightarrow P(X \ge x) \ge 1 - u$ である．これより $I_u = [x_L, x_R]$ となることがわかる．

問 2.5　X の符号を変えて $-X$ を考えれば，下側確率と上側確率が入れ替わる．また U を 0 と 1 の間の一様分布に従う確率変数とすれば，$1 - U$ も同じ分布に従う．このことから X_R も F に従うことがわかる．

$X_L \le X_R$ は常に成り立つ．ここで背理法を用いて，もし $P(X_L < X_R) > 0$ と仮定すると，ある x が存在して $P(X_L \le x < X_R) > 0$ となるが，このとき

$$P(X_L \le x) = P(X_L \le x, X_R \le x) + P(X_L \le x < X_R) > P(X_L \le x, X_R \le x)$$

$$P(X_R \le x) = P(X_L \le x, X_R \le x) + P(X_R \le x < X_L) = P(X_L \le x, X_R \le x)$$

より $P(X_L \le x) > P(X_R \le x)$ となり，分布が等しいことに矛盾する．従って確率 1 で $X_R = X_L$ である．

後半は，測度論的を用いて，以下のように議論することもできる．示すべきことは「$x_L = {F_L}^{-1}(u) < {F_R}^{-1}(u) = x_R$ となるような u が実現する確率が 0」と同値であるから，こちらを示す．$x_L < x_R$ のとき，開区間 (x_L, x_R) で F が平らであることを示そう．本文で $P(X \le x_L) \ge u$, $P(X \ge x_R) \ge 1 - u$ が示されている ((2.25) 式，(2.26) 式)．ところで $x_L < x_R$ より $(-\infty, x_L]$ と $[x_R, \infty)$ は排反であるから，これらの和集合の確率は 1 以下である．このことから上の 2 つの不等式はいずれも等号でなければならず，$P(X \le x_L) = u, P(X \ge x_R) = 1 - u$ であり F は開区間 (x_L, x_R) で平らである．つまりこの区間で $u = F(x)$ の値は一定であり，U が一様分布であればこのような特定の u が実現する確率は 0 である．また $x_L < x_R$ となるような区間はとびとびにしかないから (厳密には可算個しかないから)，このような区間をすべて考えても，${F_L}^{-1}(u) < {F_R}^{-1}(u)$ となるような u が実現する確率は 0 であることがわかる．

問 2.6　$F({F_L}^{-1}(u)) \geq u$ は (2.25) 式である．${F_R}^{-1}(u) \geq {F_L}^{-1}(u)$ より $F({F_R}^{-1}(u)) \geq u$ も成り立つ．$F({F_L}^{-1}(u)-) \leq u$ も (2.25) 式よりわかる．$F({F_R}^{-1}(u)-) \leq u$ は上の問 2.5 の解答例よりわかる．もし F が連続ならばこれらの不等式はすべて等式となる．そして (2.29) 式より ${F_L}^{-1}(U)$ は分布 F に従うことから，F が連続ならば $F(X) \sim \mathrm{U}[0,1]$ となることがわかる．

X が例えばベルヌーイ分布であれば $F(X)$ は $0, 1/2, 1$ の 3 個の値しかとらないので，連続分布である一様分布とはならない．

問 2.7　ポアソン分布の確率母関数 $E\left[s^X\right] = e^{\lambda(s-1)}$ を微分することにより $E[X] = \mathrm{Var}[X] = \lambda$ であることがわかる．小数法則では $\lambda = np$ を一定として $n \to \infty$ とするが，このとき $p \to 0$ であるから，2 項分布の分散も $np(1-p) \to \lambda$ となり，整合性が確認できる．

問 2.8　(2.58) 式の $(1-q)^{-r}$ のテーラー展開を用いれば $G(s) = p^r/(1-qs)^r$ となることが容易にわかる．これを微分すれば期待値及び分散が (2.62) 式となることも確かめられる．

問 2.9　標準正規分布に関する第 1 式は $\phi(x)$ が偶関数であり $x\phi(x)$ が奇関数から明らかである．第 2 式は

$$
\begin{aligned}
\int_{-\infty}^{\infty} x^2\phi(x)\,dx &= 2\int_0^{\infty} x^2\phi(x)\,dx = 2\int_0^{\infty} x\cdot(-\phi(x))'\,dx \\
&= 2\Big[-x\phi(x)\Big]_0^{\infty} + 2\int_0^{\infty}\phi(x)\,dx \\
&= 0 + 2\times\frac{1}{2} = 1
\end{aligned}
$$

よりわかる．あとは $X \sim \mathrm{N}(0,1)$ として $Y = \mu + \sigma X$ の期待値と分散よりわかる．

問 2.10　平均は

$$
E[X] = \frac{1}{B(a,b)}\int_0^1 x^a(1-x)^{b-1}\,dx = \frac{B(a+1,b)}{B(a,b)} = \frac{a}{a+b}
$$

である．同様に $E[X^2] = a(a+1)/((a+b)(a+b+1))$ となるから

$$
\mathrm{Var}[X] = \frac{a(a+1)}{(a+b)(a+b+1)} - \Big(\frac{a}{a+b}\Big)^2 = \frac{ab}{(a+b)^2(a+b+1)}
$$

である．

問 2.11

$$-\log(1-\theta) = \theta + \frac{1}{2}\theta^2 + \cdots + \frac{1}{k}\theta^k + \cdots = \sum_{k=1}^{\infty} \frac{1}{k}\theta^k$$

より $c(\theta) = 1/(-\log(1-\theta))$ である．積率母関数は

$$G(s) = E\left[s^X\right] = \frac{-\log(1-\theta s)}{-\log(1-\theta)}$$

である．

$$E[X] = G'(1) = \frac{1}{-\log(1-\theta)}\frac{\theta}{1-\theta}$$

$$\begin{aligned}\mathrm{Var}[X] &= G''(1) + E[X] - (E[X])^2 \\ &= \frac{1}{-\log(1-\theta)}\left(\frac{\theta^2}{(1-\theta)^2} + \frac{\theta}{1-\theta} - \frac{\theta^2}{(1-\theta)^2}\frac{1}{-\log(1-\theta)}\right) \\ &= \frac{\theta}{(-\log(1-\theta))^2(1-\theta)^2}(-\log(1-\theta)-\theta)\end{aligned}$$

問 2.12　正規分布の積率母関数を求めると同様の計算をおこなうと次を得る．

$$E[X] = e^{\mu+\sigma^2/2}, \quad \mathrm{Var}[X] = e^{2\mu+\sigma^2}(e^{\sigma^2}-1)$$

問 2.13　Λ を密度関数 $g(\lambda)$ を持つ確率変数とし，$\mu = E[\Lambda], \sigma^2 = \mathrm{Var}[\Lambda]$ とおく．Y の (無条件の) 期待値は $E[Y] = E[E[Y\,|\,\Lambda]] = E[\Lambda] = \mu$ である．また Y の (無条件の) 分散は (3.52) 式より

$$\mathrm{Var}[Y] = E[\mathrm{Var}[Y\,|\,\Lambda]] + \mathrm{Var}[E[Y\,|\,\Lambda]] = E[\Lambda] + \mathrm{Var}[\Lambda] = \mu + \sigma^2$$

となる．従って Y の周辺分布においては，$\mathrm{Var}[Y] - E[Y] = \mathrm{Var}[\Lambda] > 0$ であり，過分散である．

問 2.14　2 項分布の成功確率の確率密度を $g(\cdot)$ とし，$g(\cdot)$ に従う確率変数を P とする．$E[P] = E[Y/n] = p$ である． $\mathrm{Var}[P] = \sigma^2$ と表す．Y の分散は (3.52) 式より

$$\begin{aligned}\mathrm{Var}[Y] &= E[\mathrm{Var}[Y\,|\,P]] + \mathrm{Var}[E[Y\,|\,P]] = E[nP(1-P)] + \mathrm{Var}[nP] \\ &= E[nP] - nE[P^2] + n^2\sigma^2 = np - nE[P^2] + n^2\sigma^2\end{aligned}$$

である．右辺から $np(1-p)$ を引くと

$$np - nE[P^2] + n^2\sigma^2 - np + np^2 = n^2\sigma^2 - n[E[P^2] - p^2]$$

$$= n^2\sigma^2 - n\sigma^2 = n(n-1)\sigma^2 > 0$$

であり，過分散である．

第 3 章

問 3.1　平面上に図示し，包除原理のように考えると，容易にわかる．また多次元に議論を一般化するには，定義関数を用いて

$$I_{(x,x+\Delta x]} = I_{(-\infty,x+\Delta x]} - I_{(-\infty,x]}$$

を各座標について掛け合わせ，展開してから期待値をとればよい．2 次元の場合だと

$$(I_{(-\infty,x+\Delta x]} - I_{(-\infty,x]})(I_{(-\infty,y+\Delta y]} - I_{(-\infty,y]})$$

を展開したものに

$$F(x,y) = E[I_{(-\infty,x]\cap(-\infty,y]}(X,Y)] = E[I_{(-\infty,x]}(X)I_{(-\infty,y]}(Y)]$$

を適用すればよい．

問 3.2　$n=2$ の場合に確認する．連続分布の場合には (3.17) 式を x,y でそれぞれ微分すれば (3.14) 式を得る．離散分布のときは (3.9) 式のように差分をとれば (3.6) 式がわかる．

問 3.3　(3.29) 式より r と θ が独立で θ が 0 と 2π の間の一様分布に従うことは明らかである．$v = r^2$ とおいて変数変換すると $r = \sqrt{v}$ で $dr/dv = 1/(2\sqrt{v})$ より v の密度が $e^{-v/2}$ に比例し，これが Ex(2) の密度関数であることがわかる．

問 3.4　密度関数の変換の公式より

$$f_Y(y) = f_X(g^{-1}(y))dx/dy = f_X(g^{-1}(y))/g'(g^{-1}(y))$$

であるから，Y の密度関数を用いた Y の期待値は

$$\int_{-\infty}^{\infty} y f_X(g^{-1}(y)) \frac{1}{g'(g^{-1}(y))}\, dy$$

であるが，積分の変数変換の公式よりこれは

$$\int_{-\infty}^{\infty} g(x) f_X(x)\, dx$$

に等しい．

問 3.5　期待値記号の線形性より

$$E[Z] = a_1 E[X_1] + \cdots + a_n E[X_n]$$

であり，期待値からの偏差は

$$Z - E[Z] = a_1(X_1 - E[X_1]) + \cdots + a_n(X_n - E[X_n])$$

である．このことから, 各 X_i から期待値を引くことにより, $E[X_i] = 0, i = 1, \ldots, n$ のときに示せばよい．

$$Z^2 = (a_1 X_1 + \cdots + a_n X_n)^2 = \sum_{i=1}^{n} {a_i}^2 {X_i}^2 + 2 \sum_{i<j} a_i a_j X_i X_j$$

であるから，両辺の期待値をとれば，右辺第 1 項は分散の和，第 2 項が共分散の和となり (3.37) 式を得る．

問 3.6　両辺の要素が等しいことを示せばよい．$a + BX$ の第 i 要素は $a_i + \sum_{j=1}^{n} b_{ij} X_j$ である．この期待値をとると $a_i + \sum_{j=1}^{n} b_{ij} E[X_j]$ であるが，これは $a + BE[X]$ の第 i 要素に一致する．

問 3.7　(3.42) 式を示すには，両辺の (i,j) 要素を考えればよい．左辺で $\mathrm{Var}[X]$ の (i,j) 要素は $\sigma_{ij} = \mathrm{Cov}[X_i, X_j]$ である．右辺の (i,j) 要素は $(X - \mu)$ の第 i 要素 $X_i - \mu_i$ と $(X - \mu)^\top$ の第 j 要素 $X_j - \mu_j$ の積であり，その期待値は σ_{ij} である．従って両辺の要素が一致する．

(3.43) 式を示すには，まず定数ベクトル a での移動は期待値ベクトルと相殺するため $a = 0$ とおいてよいことに注意する．このとき $\mu = E[X]$ とすると $E[BX] = B\mu$ より (3.42) 式を用いて

$$\begin{aligned}\mathrm{Var}[BX] &= E\left[B(X - \mu)(B(X - \mu))^\top\right] = E\left[B(X - \mu)(X - \mu)^\top B^\top\right] \\ &= BE\left[(X - \mu)(X - \mu)^\top\right] B^\top = B\mathrm{Var}[X]\, B^\top\end{aligned}$$

を得る．

問 3.8　それぞれの分布の確率母関数あるいは積率母関数の積を確認すればよい．

問 3.9　連続分布の場合に示す．x, y の同時密度関数を $f(x, y)$ とし $f_X(x)$ を X の周辺密度関数とする．このとき

$$
\begin{aligned}
E[g(X,Y)] &= \iint g(x,y)f(x,y)\,dxdy = \int\left(\int g(x,y)\frac{f(x,y)}{f_X(x)}\,dy\right)f_X(x)\,dx \\
&= \int\left(\int g(x,y)f_{Y|X=x}(y)dy\right)f_X(x)\,dx \\
&= \int E[g(X,Y)\,|\,X=x]\,f_X(x)\,dx = E^X[E[g(X,Y)|X]]
\end{aligned}
$$

である.

問 3.10　$E[(Z-c)^2] = c^2 - 2cE[Z] + E[Z^2]$ は c の 2 次関数であり，下に凸だから，微分して 0 とおくことにより，最小にする c は $c = E[Z]$ で与えられることがわかる.

問 3.11　(3.63) 式の a を (3.62) 式に代入した上で，確率変数の線形結合の分散の式 ((3.37) 式) を係数ベクトル $(1, -b_1, \ldots, -b_n)$ に適用すればよい.

問 3.12　1 から n まで番号のついた n 個のボールを投げ，i 番目の箱に y_i 個 $(i = 1, \ldots, k)$ がはいる組合せの総数考えると，$n!$ 個の順列のうちそれぞれの箱にあるボールの順列は区別できないため，総数は $n!/(y_1! \cdots y_k!)$ となる．どの組合せでも確率は $p_1^{y_1} \cdots p_k^{y_k}$ であるから，多項分布の確率は (3.68) 式で与えられる.

問 3.13　条件つき確率関数を書いてみれば容易に確認できる.

問 3.14　多変量正規分布の密度関数の指数部分に $\theta^\top Y$ を加えて平方完成すると

$$
\begin{aligned}
\theta^\top Y - \frac{1}{2}(Y-\mu)^\top \Sigma^{-1}(Y-\mu) &= \theta^\top\mu + \theta^\top(Y-\mu) - \frac{1}{2}(Y-\mu)^\top\Sigma^{-1}(Y-\mu) \\
&= \theta^\top\mu + \frac{1}{2}\theta^\top\Sigma\theta - \frac{1}{2}(Y-\mu-\Sigma\theta)^\top\Sigma^{-1}(Y-\mu-\Sigma\theta)
\end{aligned}
$$

となり，右辺第 3 項は積分で消えるから，積率母関数は $\exp(\theta^\top\mu + \theta^\top\Sigma\theta/2)$ となる.

問 3.15　$x = B^{-1}(y-a)$ において x_i を y_j で微分すれば B^{-1} の (i,j) 要素となるから $J(\partial x/\partial y) = B^{-1}$ である.

問 3.16　(3.18) 式は

$$
\frac{f(x,y,z)}{f_Z(z)} = \frac{f_{X,Z}(x,z)}{f_Z(z)} \times \frac{f_{Y,Z}(x,z)}{f_Z(z)}
$$

と書けるが，両辺を右辺の第 2 項 $f_{Y,Z}(y,z)/f_Z(z)$ で割ると

$$\frac{f(x,y,z)}{f_{Y,Z}(y,z)} = \frac{f_{X,Z}(x,z)}{f_Z(z)}$$

となり (3.19) 式となる．操作を逆にすることもできるから，(3.18) 式と (3.19) 式は同値である．

問 3.17　条件つき独立性は

$$\frac{f(x,y,z)}{f_Z(z)} = \frac{f_{X,Z}(x,z)}{f_Z(z)} \frac{f_{Y,Z}(y,z)}{f_Z(z)}$$

と書けるが，両辺に $f_Z(z)$ をかけると

$$f(x,y,z) = f_{X,Z}(x,z)\frac{f_{Y,Z}(y,z)}{f_Z(z)}$$

となるから，例えば $g(x,z) = f_{X,Z}(x,z)$, $h(y,z) = f_{Y,Z}(y,z)/f_Z(z)$ とおけばよい．

逆に $f(x,y,z) = g(x,z)h(y,z)$ と書けたとする．$g_Z(z) = \displaystyle\int g(x,z)\,dx$, $h_Z(z) = \displaystyle\int h(y,z)\,dy$ とおくと，$f_{X,Z}(x,z) = g(x,z)h_Z(z)$, $f_{Y,Z}(y,z) = g_Z(z)h(y,z)$, $f_Z(z) = g_Z(z)h_Z(z)$ となる．従って

$$\begin{aligned}\frac{f(x,y,z)}{f_Z(z)} = \frac{g(x,z)h(y,z)}{g_Z(z)h_Z(z)} &= \frac{g(x,z)h_Z(z)}{g_Z(z)h_Z(z)} \times \frac{g_Z(z)h(y,z)}{g_Z(z)h_Z(z)}\\ &= \frac{f_{X,Z}(x,z)}{f_Z(z)} \times \frac{f_{Y,Z}(y,z)}{f_Z(z)}\end{aligned}$$

となり，条件つき独立性が成り立つ．

問 3.18　$(X_iY_i, (1-X_i)Y_i, X_i(1-Y_i), (1-X_i)(1-Y_i))$ のベクトルは 1 個の要素のみ 1，他の 3 個の要素は 0 となるから，4 次元ベルヌーイ試行である．これらを $i = 1,\ldots,n$ まで加えたものが $(Z, X-Z, Y-Z, n-X-Y+Z)$ となるから，これは 4 項分布に従う．成功確率のベクトルは $(p_1p_2, p_1(1-p_2), (1-p_1)p_2, (1-p_1)(1-p_2))$ である．多項確率は

$$\begin{aligned}&\frac{n!}{z!\,(x-z)!\,(y-z)!\,(n-x-y+z)!}\times\\ &\times (p_1p_2)^z(p_1(1-p_2))^{x-z}((1-p_1)p_2)^{y-z}((1-p_1)(1-p_2))^{n-x-y+z}\\ &\quad= \frac{n!}{z!\,(x-z)!\,(y-z)!\,(n-x-y+z)!}{p_1}^x(1-p_1)^{n-x}{p_2}^y(1-p_2)^{n-y}\end{aligned}$$

である．また X と Y は独立に 2 項分布に従い，それぞれの周辺確率は

$$\binom{n}{x} {p_1}^x (1-p_1)^{n-x}, \ \binom{n}{y} {p_2}^y (1-p_2)^{n-y}$$

である．多項確率を周辺確率の積で割ると求める条件つき分布となるが,

$$\frac{n!}{z!\,(x-z)!\,(y-z)!\,(n-x-y+z)!} \times \frac{1}{\binom{n}{x}\binom{n}{y}}$$

$$= \frac{x!\,(n-x)!\,y!\,(n-y)!}{n!\,z!\,(x-z)!\,(y-z)!\,(n-x-y+z)!}$$

は超幾何分布の確率関数である.

問 3.19　逆変換は $x = \sqrt{uv}$, $y = \sqrt{u/v}$ である．ヤコビアンは

$$J = \frac{1}{2}\begin{pmatrix} \sqrt{v/u} & \sqrt{u/v} \\ 1/\sqrt{uv} & -u^{1/2}/v^{3/2} \end{pmatrix}, \quad |\det J| = \frac{1}{4}|-1/v - 1/v| = \frac{1}{2v}$$

である．また (u,v) の動ける範囲は $0 \le uv = x^2 \le 1$, $0 \le u/v = y^2 \le 1$ より $0 \le uv \le 1$, $0 \le u \le v$ となる．(x,y) の同時密度は xy 平面の単位正方形上で 1, その他でゼロであるから，(u,v) の同時密度は

$$f(u,v) = \frac{1}{2v}, \qquad 0 \le uv \le 1, \ 0 \le u \le v$$

である．これを u で積分すると，u の範囲は $0 \le u \le \min\{v, 1/v\}$ であるから u で積分すると v の周辺密度は

$$f(v) = \frac{1}{2v}\min\{v, 1/v\} = \frac{1}{2}\min\{1, 1/v^2\}$$

である.

問 3.20　Z が連続な密度関数 $f(z)$ を持つ場合に示す.

$$E[|Z-c|] = \int_{-\infty}^{c} (c-x)f(x)\,dx + \int_{c}^{\infty} (x-c)f(x)\,dx$$

であるが, これを c で微分すると, $x=c$ での被積分関数の値がゼロになることから, 被積分関数の微分のみが残り $(d/dc)(c-x)f(x) = f(x)$, $(d/dc)(x-c)f(x) = -f(x)$ より

$$\frac{d}{dc}E[|Z-c|] = \int_{-\infty}^{c} f(x)\,dx - \int_{c}^{\infty} f(x)\,dx = P(Z \le c) - P(Z > c)$$

となる．これが 0 となる c は $1/2 = P(Z \le c) = P(z > c)$ となる c，つまりメディアンである．Z が離散の場合も同様に示すことができる.

問 3.21　実はこの問題は i,j,k,l がすべて異なる場合について示せばよい．その理由は，すべて異なる場合に示しておけば，例えば $i=j$ のときには X_i と X_j の相関係数が 1 と思って $\rho_{ij}=1$ を代入して，分散行列が退化する特殊ケースとして扱えばよいからである．さらに $(i,j,k,l)=(1,2,3,4)$ としても一般性を失わない．さて，積率母関数

$$\phi(\theta)=E\left[e^{\theta^\top X}\right]=e^{\theta^\top \Sigma\theta/2}=1+\theta^\top\Sigma\theta/2+\frac{1}{2!}(\theta^\top\Sigma\theta/2)^2+\cdots$$

において，右辺 3 項目の θ の 4 乗項 $(\theta^\top\Sigma\theta)^2/8$ のみに注目する．

$$(\theta^\top\Sigma\theta)(\theta^\top\Sigma\theta)$$

の積の第 1 項から例えば $\theta_1\theta_2$ をとると，第 2 項からは $\theta_3\theta_4$ をとらなければならない．

$$\theta^\top\Sigma\theta=2\theta_1\theta_2\sigma_{12}+\text{他の項}$$

に注意して，また積の 1 項目と 2 項目の対称性に注意すると $(\theta^\top\Sigma\theta)^2/8$ において $\theta_1\theta_2\theta_3\theta_4$ の係数は

$$\sigma_{12}\sigma_{34}+\sigma_{13}\sigma_{24}+\sigma_{14}\sigma_{23}$$

であることがわかる．

第 4 章

問 4.1　$X\sim\mathrm{N}(0,1)$ として X^2 の積率母関数を求めると，$\theta<1/2$ について

$$E\left[e^{\theta X^2}\right]=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-x^2(1-2\theta)/2}\,dx=(1-2\theta)^{-1/2}$$

を得る．独立性より $E\left[e^{\theta({X_1}^2+\cdots+{X_n}^2)}\right]=(1-2\theta)^{-n/2}$ を得る．ガンマ分布の積率母関数と比較すれば $Y={X_1}^2+\cdots+{X_n}^2$ の分布は $\mathrm{Ga}(n/2,2)$ である．

問 4.2　G の 2 行目以降の行については，要素の和が 0 となるから，1 行目と直交している．$2\le i<j$ として第 i 行と第 j 行の内積を考える．第 j 行は第 1 要素から第 i 要素まで定数であるから，内積の計算では第 i 行の要素の和をとることになり，やはり 0 となる．従って G のすべての行は互いに直交している．さらに容易にわかるように，各行は長さが 1 に基準化されている．これより G は直交行列である．

問 4.3　t 統計量において $\sum_{i=1}^{n}(x_i-\bar{x})^2$ を明示的に示すと

$$\frac{\sqrt{n}(\bar{x}-\mu)}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2/(n-1)}}=t=\frac{\sqrt{n-1}(\bar{x}-\mu)}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2/n}}$$

となることから，s^2 として n で割ったものを用いれば，分子には $\sqrt{n-1}$ が現れる．

問 4.4　$a=(1+t^2/m)/2$ とおくと，

$$\begin{aligned}\int_0^{\infty} v^{(m+1)/2-1}e^{-va}\,dv &= (1/a)^{(m+1)/2}\int_0^{\infty}(va)^{(m+1)/2-1}e^{-va}\,d(va)\\ &= (1/a)^{(m+1)/2}\Gamma\Big(\frac{m+1}{2}\Big)\end{aligned}$$

である．

問 4.5　$(1+t^2/m)^{-(m+1)/2}\to e^{-t^2/2}$ である．またスターリングの公式において

$$\lim_{m\to\infty}\frac{((m+1)/2)^{(m+1)/2}}{(m/2)^{m/2}}=\frac{1}{\sqrt{2}}\lim_{m\to\infty}(1+1/m)^{m/2}=\frac{1}{\sqrt{2}}e^{1/2}$$

を用いると

$$\frac{\Gamma((m+1)/2)}{\sqrt{m}\Gamma(m/2)}\to\frac{1}{\sqrt{2}}$$

となることがわかる．これらより t 分布の密度関数が標準正規分布の密度関数に収束することがわかる．

問 4.6　$V=lY$ とおくと，V の密度関数は

$$f(v)=\frac{m^{m/2}\Gamma((l+m)/2)}{\Gamma(l/2)\Gamma(m/2)}\frac{v^{l/2-1}}{(m+v)^{(l+m)/2}}$$

である．ここで $v^{l/2-1}/\Gamma(l/2)$ は目標とするガンマ分布の密度関数の一部であるから，これを外し

$$\lim_{m\to\infty}\frac{m^{m/2}}{(m+v)^{m/2}}=e^{-v/2}$$

に注意すれば

$$\lim_{m\to\infty}\frac{\Gamma((l+m)/2))}{\Gamma(m/2)(m+v)^{l/2}}=\frac{1}{2^{l/2}}$$

を示せばよい．これは前問と同様の計算で示される．

問 4.7　非心度が $\sum_{i=1}^{m}(\mu_i-\bar{\mu})^2$ となることを確かめればよい．Helmert 変換の議論に従い $Y=GX$ とおくと期待値ベクトルについて $E[Y]=GE[X]=G\mu$ が成り立つ．さて求める非心度は $E[Y_2]^2+\cdots+E[Y_m]^2$ である．ここで

$$
\begin{aligned}
E[Y_2]^2+\cdots+E[Y_m]^2 &= E[Y]^\top E[Y]-E[Y_1]^2 \\
&= \mu^\top\mu - m\bar{\mu}^2 = \sum_{i=1}^{m}(\mu_i-\bar{\mu})^2
\end{aligned}
$$

となり，確かめられる．

問 4.8　項別積分において積分すべき j 番目の項は

$$
v^{j/2}v^{(m+1)/2-1}e^{-v(1+t^2/m)/2}
$$

であり，これを 0 から ∞ まで積分すれば

$$
\Gamma((m+1+j)/2)\left(\frac{2}{1+t^2/m}\right)^{m+1+j}
$$

となり，これを代入すれば (4.26) 式が確かめられる．

問 4.9　単一の順序統計量の分布を導いたときのように，3 項分布の確率を操作することによってきちんと示されるが，やや面倒なので，以下のような直感的な議論を与える．

特定の X_k が $X_{(i)}$ に一致し x の近傍 (長さ dx) に落ち，かつ特定の X_l が $X_{(j)}$ に一致し y の近傍 (長さ dy) に落ちる確率を考えると，まず確率要素として $f(x)dxf(y)dy$ が現れる．また残りの $n-2$ 個の確率変数については 3 項分布で考えると

$$
\frac{(n-2)!}{(i-1)!\,(j-i-1)!\,(n-j)!}F(x)^{i-1}(F(y)-F(x))^{j-i-1}(1-F(y))^{n-j}
$$

の確率が現れる．X_k, X_l の取り方が $n(n-1)$ 通りあることから，全体の確率は

$$
\begin{aligned}
&n(n-1)f(x)dxf(y)dy \\
&\times\frac{(n-2)!}{(i-1)!\,(j-i-1)!\,(n-j)!}F(x)^{i-1}(F(y)-F(x))^{j-i-1}(1-F(y))^{n-j}
\end{aligned}
$$

であり，$dxdy$ で割ることによって密度関数が (4.51) 式の形になることが理解される．

問 4.10　$t>1$ ならば $t^2/(1+t^2) \geq 1/2$ より

$$t \times \frac{1}{1+t^2} \geq \frac{1}{2t}, \quad t > 1$$

である．$\displaystyle\int_1^\infty \frac{1}{t}\,dt = \Big[\log t\Big]_1^\infty = \infty$ であるからコーシー分布の期待値は存在しない．期待値が存在しないため，分散も存在しない．

次に $\tan x$ の微分が

$$(\tan x)' = \frac{1}{\cos^2 x} = 1 + \tan^2 x$$

であるから，$\tan^{-1}(t)$ の微分は

$$(\tan^{-1}(t))' = \frac{1}{1+t^2}$$

であり，変数変換の公式より $X = \tan(\pi(U-1/2))$ の密度関数は

$$f(x) = \frac{1}{\pi(1+x^2)}$$

となり，コーシー分布であることがわかる．さらに密度関数の積分が $\tan^{-1}$ で表せることから，累積分布関数も確認される．

問 4.11　どのような順番で抽出しても，最後まで抽出すれば $X_1 + \cdots + X_N = a_1 + \cdots + a_N$ となり，これは定数である．従って $\mathrm{Var}[X_1 + \cdots + X_N] = 0$ である．一方で $\mathrm{Var}[X_1 + \cdots + X_N] = N\mathrm{Var}[X_1] + N(N-1)\mathrm{Cov}[X_1, X_2]$ であるから $0 = N\mathrm{Var}[X_1] + N(N-1)\mathrm{Cov}[X_1, X_2]$ より

$$\mathrm{Cov}[X_1, X_2] = -\frac{1}{N-1}\sigma^2$$

を得る．

問 4.12　(1) $i \leq j$ として $\mathrm{Cov}[W_i, W_j] = \mathrm{Var}[W_j] = n - j + 1$ となる．行列の形で書くと

$$\begin{pmatrix} n & n-1 & n-2 & \dots & 1 \\ n-1 & n-1 & n-2 & \dots & 1 \\ n-2 & n-2 & n-2 & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & \dots & 1 \end{pmatrix}$$

となる.

(2) $1 \le k < n$ として $(W_1, \ldots, W_k)$ の分散共分散行列 Σ_k は

$$\Sigma_k = \begin{pmatrix} n & n-1 & n-2 & \ldots & n-k+1 \\ n-1 & n-1 & n-2 & \ldots & n-k+1 \\ n-2 & n-2 & n-2 & \ldots & n-k+1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ n-k+1 & n-k+1 & n-k+1 & \ldots & n-k+1 \end{pmatrix}$$

である．この逆行列を求めるのに三角分解を用いるとよい．上三角部分がすべて 1 の $k \times k$ 行列を

$$U_k = \begin{pmatrix} 1 & 1 & 1 & \ldots & 1 \\ 0 & 1 & 1 & \ldots & 1 \\ 0 & 0 & 1 & \ldots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix}$$

とおく．D を $k \times k$ 対角行列 $\mathrm{diag}(1, \ldots, 1, n-k+1)$ とすると容易に確認できるように

$$\Sigma_k = U_k D {U_k}^\top$$

である．また ${U_k}^{-1}$ は対角要素が 1，その右上が -1 で，残りの要素はすべて 0 の上三角行列である:

$$ {U_k}^{-1} = \begin{pmatrix} 1 & -1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & -1 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & -1 \\ 0 & 0 & 0 & \ldots & 0 & 1 \end{pmatrix}$$

これを用いると

$${\Sigma_k}^{-1} = ({U_k}^{-1})^\top D^{-1} {U_k}^{-1}$$

と書ける．これは 3 重対角行列となるが，要素は省略する.

(3) 上の結果と多変量正規分布の条件つき分布の結果から確認できる．ここでは以下

のより簡便な議論を与える．問 3.16 にある (3.19) 式と (3.18) 式の同値性を用いると，W_k を与えたときに W_{k+1} と $W_1,\dots,W_{k-1}$ が条件つき独立であることを示せばよい．ところで W_k を与えると $W_1,\dots,W_{k-1}$ の条件つき分布は $X_{n-k+2},\dots,X_n$ の分布によって定まるから，これらは $W_k = X_1+\cdots+X_{n-k+1}$ を与えたときの $W_{k+1} = W_k - X_{n-k+1}$ とは条件つき独立である．なお，以上の性質は「マルコフ性」とよばれる．

(4)

$$\begin{aligned} W_{k+1} - E[W_{k+1}\,|\,W_k] &= W_{k+1} - W_k\frac{\mathrm{Cov}[W_k, W_{k+1}]}{\mathrm{Var}[W_k]} = W_{k+1} - \frac{n-k}{n-k+1}W_k \\ &= (x_1+\cdots+x_{n-k}) - \frac{n-k}{n-k+1}(x_1+\cdots+x_{n-k}+x_{n-k+1}) \\ &= \frac{1}{n-k+1}((x_1+\cdots+x_{n-k}) - (n-k)x_{n-k+1}) \end{aligned}$$

から確認できる．

問 4.13　(1) 各 $\widetilde{X}_i$ が x_1 以外となる確率は $1-1/n$ である．独立性より $\widetilde{X}_i$, $i=1,\dots,n$ がすべて x_1 以外となる確率は $(1-1/n)^n$ である．

(2) 上と同様に考えれば $\widetilde{X}_i$, $i = 1,\dots,n$ がすべて x_1, x_2 以外となる確率は $(1-2/n)^n$ である．

(3) $E[I_i] = P(I_i = 1) = (1-1/n)^n$ であるから $E[S_n] = n(1-(1-1/n)^n)$ である．また

$$\mathrm{Var}[S_n] = \mathrm{Var}[n - S_n] = n\mathrm{Var}[I_1] + n(n-1)\mathrm{Cov}[I_1, I_2]$$

であるが，$\mathrm{Var}[I_1] = (1-1/n)^n(1-(1-1/n)^n)$ である．また

$$\mathrm{Cov}[I_1, I_2] = E[I_1I_2] - E[I_1]\,E[I_2] = (1-2/n)^n - (1-1/n)^{2n}$$

である．これより

$$\mathrm{Var}[S_n] = n(1-1/n)^n(1-(1-1/n)^n) + n(n-1)((1-2/n)^n - (1-1/n)^{2n})$$

である．$E[S_n/n] = E[S_n]/n$, $\mathrm{Var}[S_n/n] = \mathrm{Var}[S_n]/n^2$ である．

(4) $n\to\infty$ のとき，$\mathrm{Cov}[I_1, I_2] = E[I_1I_2] - E[I_1]\,E[I_2] \to e^{-2} - e^{-2} = 0$ に注意すると，$\mathrm{Var}[S_n/n] \to 0$ である．また $E[S_n/n] = 1-(1-1/n)^n \to 1-e^{-1}$ である．これよりチェビシェフの不等式から $n\to\infty$ のとき S_n/n は $1-e^{-1}$ に確率収束する．

問 4.14 0 と 1 の間の一様分布について解答を示す．$X_{(i)}$ の分布はベータ分布であるから，その平均と分散は問 2.10 より，$a = i, b = n - i + 1$ とおいて

$$E[X_{(i)}] = \frac{i}{n+1}, \quad \mathrm{Var}[X_{(i)}] = \frac{i(n-i+1)}{(n+1)^2(n+2)}$$

である．次に

$$f(x,y) = \frac{n!}{(i-1)!\,(j-i-1)!\,(n-j)!} x^{i-1}(y-x)^{j-i-1}(1-y)^{n-j}$$

が同時密度関数になっていることから $E[X_{(i)}(X_{(j)} - X_{(i)})] = E[X(Y-X)]$ の計算において $i \to i+1,\ j \to j+2,\ n \to n+2$ と調整すると

$$E[X(Y-X)] = \frac{i(j-i)}{(n+1)(n+2)}$$

である．また問 2.10 の解答例にあるように $E[X^2] = i(i+1)/((n+1)(n+2))$ より $E[XY] = i(j+1)/((n+1)(n+2))$ となり，共分散は

$$\begin{aligned}\mathrm{Cov}[X,Y] &= E[XY] - E[X]\,E[Y] \\ &= \frac{i(j+1)}{(n+1)(n+2)} - \frac{ij}{(n+1)^2} = \frac{i(n+1-j)}{(n+1)^2(n+2)}\end{aligned}$$

となる．

なお，$X_{(i)}$ と $X_{(j)}$ の同時分布は「ディリクレ分布」とよばれる分布であり，多項分布の共役事前分布の形をしている．

第 5 章

問 5.1 δ が許容的とする．$R(\theta,\tilde{\delta}) \le R(\theta,\delta), \forall\theta$ が成り立つとして，ある θ_0 について $R(\theta_0,\tilde{\delta}) < R(\theta_0,\delta)$ であると，$\tilde{\delta}$ は δ を優越し矛盾であるから，そのような θ_0 は存在せず $R(\theta,\tilde{\delta}) = R(\theta,\delta), \forall\theta$ である．

逆に $R(\theta,\tilde{\delta}) \le R(\theta,\delta), \forall\theta \Rightarrow R(\theta,\tilde{\delta}) = R(\theta,\delta), \forall\theta$ ならば δ を優越する $\tilde{\delta}$ は存在しないから，δ は許容的である．

問 5.2 δ が許容的とする．ある θ_0 について $R(\theta_0,\delta) > R(\theta_0,\tilde{\delta})$ として，すべての θ_1 について $R(\theta_1,\delta) \ge R(\theta_1,\tilde{\delta})$ とすると $\tilde{\delta}$ が δ を優越し，矛盾である．したがって $R(\theta_1,\delta) < R(\theta_1,\tilde{\delta})$ となる θ_1 が存在する．

問 5.3 U をデータと独立に成功確率 α のベルヌーイ試行を表す確率変数とすると d_α の損失は

$$L(\theta, d_\alpha(x)) = I_{U=1}L(\theta, d_0(X)) + I_{U=0}L(\theta, d_*(X))$$

と書ける．両辺の期待値をとれば (5.11) 式を得る．

問 5.4　図より点 M の座標を

$$\alpha\begin{pmatrix}0\\1\end{pmatrix} + (1-\alpha)\begin{pmatrix}1/2\\1/3\end{pmatrix} = \begin{pmatrix}c\\c\end{pmatrix}$$

と置くと $c = (1-\alpha)\dfrac{1}{2} = \alpha + (1-\alpha)\dfrac{1}{3}$ より $\alpha = 1/7$, $c = 3/7$ を得る．ミニマックス決定方式は d_0 を確率 1/7，d^* を確率 6/7 で用いる確率化決定方式である．

問 5.5　2 つの決定関数 (確率化を含んでもよい) δ_1, δ_2 に対して問 5.3 と同様に考えると，データとは無関係に確率 α で δ_1 を選び確率 $1-\alpha$ で δ_2 を選ぶ (確率化) 決定関数 δ_α のリスク関数は

$$R(\theta, \delta_\alpha) = \alpha R(\theta, \delta_1) + (1-\alpha)R(\theta, \delta_2)$$

である．従ってリスク集合は凸集合である．

問 5.6　問ではリスクを「表にせよ」とあるが，実際に作業してみると「式で」書いたほうが簡潔となるので，以下ではシンボリックに表してみる．

$i, j \in \{1,2,3\}$ として $\delta(0) = i, \delta(1) = j$ である非確率化決定関数を δ_{ij} と書く．このように書けば，非確率化決定関数は $9 = 3 \times 3$ 通りあることも明確になる．ここからは $i = j$ のときは δ_{ii} と書くこととし，i, j, k の 3 個の添え字はそれぞれ異なるものとする．すなわち (i, j, k) は $(1, 2, 3)$ を並べ替えたものである．

- まず 0-1 損失で δ_{ii} の場合，リスクは

$$R(p_i, \delta_{ii}) = 0, \quad R(p_j, \delta_{ii}) = R(p_k, \delta_{ii}) = 1$$

　である．

- 次に 0-1 損失で δ_{ij} の場合は

$$R(p_i, \delta_{ij}) = p_i, \quad R(p_j, \delta_{ij}) = 1 - p_j, \quad R(p_k, \delta_{ij}) = 1$$

　である．

- 次に 2 乗損失で δ_{ii} の場合，リスクは

$$R(p_i, \delta_{ii}) = 0, \quad R(p_j, \delta_{ii}) = (p_j - p_i)^2, \quad R(p_k, \delta_{ii}) = (p_k - p_i)^2$$

　である．

- 次に 2 乗損失で δ_{ij} の場合は

$$
\begin{aligned}
R(p_i, \delta_{ij}) &= p_i(p_j - p_i)^2, \\
R(p_j, \delta_{ij}) &= (1-p_j)(p_i - p_j)^2, \\
R(p_k, \delta_{ij}) &= p_k(p_j - p_k)^2 + (1-p_k)(p_i - p_k)^2
\end{aligned}
$$

　である．

なお，以下で議論するように，2 乗損失の場合には $\delta(x)$ の値は必ずしも p_1, p_2, p_3 に一致しなくてもよい．$\delta(0), \delta(1) \in [0,1]$ とした一般の非確率化決定関数のリスクは

$$R(p_i, \delta) = (1-p_i)(\delta(0) - p_i)^2 + p_i(\delta(1) - p_i)^2, \quad i = 1,2,3 \tag{1}$$

である．

ベイズ決定関数

さて 14 章で論じているようにベイズ決定関数は各 x に対して事後リスクを最小にする決定を選べばよいため，ベイズ決定関数は容易に求められる．そこでまずベイズ決定関数を議論する．

0-1 損失の場合には，事後確率を最大とする p_i を選ぶのがベイズ決定関数である．事前確率が $(1/3, 1/3, 1/3)$ である場合，$x = 0$ のときの p_1, p_2, p_3 の事後確率はそれぞれ $(1-p_1), (1-p_2), (1-p_3)$ に比例したものとなる．この中で最大のものは $1 - p_1$ であるから $\delta(0) = 0$ である．同様に $x = 1$ のときの p_1, p_2, p_3 の事後確率はぞれぞれ p_1, p_2, p_3 に比例するから，$\delta(1) = p_3$ である．

次に 2 乗損失を議論する．2 乗損失の場合には $\delta(x)$ の値は必ずしも p_1, p_2, p_3 にぴったり一致しなくても「近い値ならよい」と考えられるから，とりあえず決定空間 D を $[0,1]$ とする．このとき，14 章の議論により，事後の 2 乗損失を最小にするには事後平均をとればよい．従って

$$\delta(0) = \frac{1}{\sum_{i=1}^{3}(1-p_i)} \sum_{i=1}^{3} p_i(1-p_i), \quad \delta(1) = \frac{1}{\sum_{i=1}^{3} p_i} \sum_{i=1}^{3} {p_i}^2$$

である．また決定空間 D を $\{p_1, p_2, p_3\}$ と限定するならば，それぞれ $\delta(x)$ に最も近い値を選べばよい．

ミニマックス決定関数

次にミニマックス決定関数について考える．考え方としては，14 章で論じているようにベイズ決定関数の中で，リスクが定数となるようなものを探索するとよい．事前確率を (π_1, π_2, π_3) とする．事後確率は，$x = 0$ のときにはそれぞれ

$$\pi_1(1-p_1) : \pi_2(1-p_2) : \pi_3(1-p_3) \tag{2}$$

に比例し，$x = 1$ のときには

$$\pi_1 p_1 : \pi_2 p_2 : \pi_3 p_3 \tag{3}$$

に比例する．$p_1 < p_2 < p_3$ より

$$\frac{p_1}{1-p_1} < \frac{p_2}{1-p_2} < \frac{p_3}{1-p_3}$$

であるから，$x = 0$ のときと比べて $x = 1$ のときのほうが分布が右側にシフトする．このことから $\delta(0) > \delta(1)$ とするような，つまり観測値と逆向きに決定する非確率的な決定関数は不合理であり，ベイズ決定関数にならないことが理解される．ただしそれだけではすぐにはミニマックス決定関数の候補はしぼれない．

ここでまず 0-1 損失を検討する．0-1 損失に対するミニマックス決定方式は，図 5.6 と同様に確率化が必要となると予想される．一方でベイズ決定関数は (2) 式及び (3) 式で最大のものを選ぶから，確率化が必要となるのは (2) 式及び (3) 式で比が 1 となるものが現れる場合である．この条件により π_1, π_2, π_3 が等式で制約されることになる．

上の分布のシフトの議論を念頭において，まず次のような確率化決定関数を考えよう．

$$\begin{aligned} &\delta(0) \text{ は } p_2 \text{ を確率 } \alpha,\ p_1 \text{ を確率 } 1-\alpha \text{ でとる．} \\ &\delta(1) \text{ は } p_2 \text{ を確率 } \beta,\ p_3 \text{ を確率 } 1-\beta \text{ でとる．} \end{aligned} \tag{4}$$

$\delta(0)$ が p_1 と p_2 を確率化してとるのは，(2) 式において

$$\pi_1(1-p_1) = \pi_2(1-p_2)$$

となっている場合であるから π_1 と π_2 の比が

$$\frac{\pi_1}{\pi_2} = \frac{1-p_2}{1-p_1}$$

と定まる．同様に $\delta(1)$ について考えると

$$\frac{\pi_3}{\pi_2} = \frac{p_2}{p_3}$$

と定まる．また，(2) 式及び (3) 式は比の形であるから，π_1, π_2, π_3 を正の定数倍しても，決定方式は変わらないことに注意すると，$((1-p_2)/(1-p_1), 1, p_2/p_3)$ に比例する (すなわち和が 1 となるように基準化した) 事前確率に対して，(4) 式の δ はベイズ決定関数であることがわかる．ここで $(1-p_2)/(1-p_1)$, p_2/p_3 はいずれも 1 より小であるから，事前確率としては π_2 が π_1, π_2, π_3 の中で一番大きくなっている．

さて (4) 式の δ のリスクを計算すると

$$\begin{aligned} R(p_1, \delta) &= (1-p_1)\alpha + p_1 \\ R(p_2, \delta) &= (1-p_2)(1-\alpha) + p_2(1-\beta) \\ R(p_3, \delta) &= (1-p_3) + p_3\beta \end{aligned}$$

であるから，これらが等しいとする線形連立方程式

$$R(p_1, \delta) = R(p_3, \delta), \quad R(p_1, \delta) = R(p_2, \delta)$$

を解いてみる．

さて，この問題での数値では $(p_1, p_2, p_3) = (0.3, 0.4, 0.7)$ としており，$1 = p_1 + p_3$ という関係が成り立っている．つまり $p_3 = 1 - p_1$, $1 - p_3 = p_1$ である．この特殊な場合には，$R(p_3, \delta)$ は

$$R(p_3, \delta) = p_1 + (1-p_1)\beta$$

と置き換えられるが，これは $R(p_1, \delta)$ で α に β を代入したものに等しい．そこで $R(p_1, \delta) = R(p_3, \delta)$ から $\alpha = \beta$ が従う．このとき $R(p_2, \delta) = 1 - \alpha$ となるから，$R(p_1, \delta) = R(p_2, \delta)$ は

$$(1-p_1)\alpha + p_1 = 1 - \alpha$$

となり

$$\alpha = \frac{1-p_1}{2-p_1}$$

と解ける．$0 < \alpha < 1$ となるから，確率化が定まることとなる．

以上より $(p_1,p_2,p_3)=(0.3,0.4,0.7)$ の場合には，(4) 式において $\alpha=\beta=7/17$ とおいたものが，リスクが一定でかつ，事前分布

$$(\pi_1,\pi_2,\pi_3)\propto(6/7,1,4/7)$$

に対するベイズ決定関数なので，補題 14.5 よりミニマックス決定関数である．14 章の用語を用いると，これが最も不利な分布である．

最後に 2 乗損失で，かつ決定空間 D を $[0,1]$ とした場合について，ミニマックス決定関数を求める．上と同様にリスクが一定のベイズ決定関数を求めればよいと思われるが，実は以下で詳しく見るように，それではうまくいかないようである．

しかしベイズ決定関数の形を確認しておくことは有用である．与えられた事前分布 π に対して，事後期待値を求めるとベイズ決定関数は

$$\delta_\pi(0)=\frac{\sum_{i=1}^{3}\pi_i(1-p_i)p_i}{\sum_{i=1}^{3}\pi_i(1-p_i)},\qquad \delta_\pi(1)=\frac{\sum_{i=1}^{3}\pi_i{p_i}^2}{\sum_{i=1}^{3}\pi_ip_i}$$

で与えられる．以下記法の簡単のために

$$a=\delta_\pi(0),\quad b=\delta_\pi(1)$$

とおく．$\pi\mapsto(a,b)$ の写像は $\mathbb{R}^2$ から $\mathbb{R}^2$ への写像であるから，特殊な点を除いては 1 対 1 になると考えられる．以下では π を動かすかわりに a,b を動かして，リスクの最大値を考察しよう．

さて，2 乗損失のもとでは確率化を考慮する必要はない．これは 2 乗 $(f(x)=x^2)$ が強凸であることから，もし確率化があったとしても，$X=x$ $(x=0,1)$ を観測したときに決定関数の条件つき期待値で置き換えればリスクが小さくなるからである．そこで上と同様に決定関数を $a=\delta(0),b=\delta(1)$ として考えればよい．与えられた数値 $(p_1,p_2,p_3)=(0.3,0.4,0.7)$ では $0.3\le a\le b\le 0.7$ となる範囲で探せばよい．

リスクはすでに (1) 式で示したが，成功確率 p を連続的に考えるために

$$R(p,\delta)=(1-p)(a-p)^2+p(b-p)^2=(1+2a-2b)p^2+(b^2-a^2-2a)p+a^2 \quad (5)$$

とおく．実際にリスクを評価する点は $p=p_1,p_2,p_3$ である．ところで (5) 式で $0\le b-a\le 0.4$ を考慮すると p^2 の係数 $(1-2(b-a))\ge 0.2>0$ であるから，(5) 式は p の関数として下に凸である．下に凸な関数の区間での最大値は，区間の端の

いずれかで達成されることに注意しよう．このことから「リスクが一定のベイズ決定関数を求める」という方針ではうまくいかないことがわかる．いずれにしても，以下では

$$
\begin{aligned}
&\max\{R(p_3,\delta),R(p_1,\delta)\}\\
&=\max\{(1+2a-2b){p_1}^2+(b^2-a^2-2a)p_1+a^2,\\
&\qquad (1+2a-2b){p_3}^2+(b^2-a^2-2a)p_3+a^2\}
\end{aligned}
$$

を考察することとなる．両端の値の差を見ると

$$R(p_3,\delta)-R(p_1,\delta)=(1+2a-2b)({p_3}^2-{p_1}^2)+(b^2-a^2-2a)(p_3-p_1)$$

であるが，右辺を $p_3-p_1>0$ で割ると

$$(1+2a-2b)(p_3+p_1)+(b^2-a^2-2a)$$

である．問題の数値では $p_1+p_3=1$ であるから，さらに差は

$$1+2a-2b+(b^2-a^2-2a)=b^2-2b+1-a^2=(b-1)^2-a^2=(b-1-a)(b-1+a)$$

となる．$b-1-a<b-1<0$ であるから差の符号は $a+b-1$ の符号で定まることがわかる．つまり

$$R(p_3,\delta)\geq R(p_1,\delta) \iff a+b-1\leq 0$$

ここで $R(p_3,\delta)\neq R(p_1,\delta)$ であるならば，a,b を微妙に摂動すれば，より高い方を少し低くできるはずであると考えられるから，最大値が最小になるのは $a+b=1$ となるときであることが了解される．そこで $b=1-a$ とおくとリスクの最大値は

$$
\begin{aligned}
&(1+2a-2(1-a)){p_1}^2+((1-a)^2-a^2-2a)p_1+a^2\\
&\quad=(4a-1){p_1}^2+(1-4a)p_1+a^2\\
&\quad=(4a-1)p_1(p_1-1)+a^2\\
&\quad=(a+2p_1(p_1-1))^2-p_1(p_1-1)-4{p_1}^2(p_1-1)^2
\end{aligned}
$$

となり，これを a について最小化すれば，最小値を与える a は

$$a=2p_1(1-p_1)$$

となる．$p_1=0.3$ を代入すれば $a=0.42$ を得る．$b=1-a=0.58$ である．$0.3<a<b<0.7$ の範囲になったので，これでミニマックス決定方式が見つけら

れた．またリスクの最大値の最小値は

$$0.3 \times 0.7 - 4(0.3 \times 0.7)^2 = 0.21(1 - 0.84) = 0.0336$$

である．対応するベイズ決定関数の事前分布は，実は

$$\pi = (0.5, 0, 0.5)$$

であり，これが最も不利な分布である．

実際この π に対応するベイズ決定関数は

$$\delta_\pi(0) = \frac{\frac{1}{2}((1-p_1)p_1 + (1-p_3)p_3)}{\frac{1}{2}((1-p_1)+(1-p_3))} = 2 \times 0.3 \times 0.7 = 0.42,$$

$$\delta_\pi(1) = \frac{\frac{1}{2}({p_1}^2 + {p_3}^2)}{\frac{1}{2}(p_1 + p_3)} = {p_1}^2 + {p_3}^2 = 0.09 + 0.49 = 0.58$$

となり，上で求めた (a, b) に一致する．$\pi_2 = 0$ となるから，最も不利な分布は確率分布の集合の端点にある．

また 0-1 損失と 2 乗損失で，最も不利な分布が非常に異なる性質を示していることも興味深い．

問 5.7　$\widetilde{\mu}$ の平均二乗誤差は $1/2$ である．以下 $\hat{\mu}$ の平均二乗誤差を計算する．$\phi(x)$ を $\mathrm{N}(0, 1)$ の密度関数とする．

$$\begin{aligned} R(\hat{\mu}, \mu) &= \int_{-c}^{c} (x_1 - \mu)^2 \phi(x_1 - \mu)\, dx_1 + P(|X_1| \geq c) E[(X_2 - \mu)^2] \\ &= \int_{-c-\mu}^{c-\mu} {x_1}^2 \phi(x_1)\, dx_1 + P(|X_1| \geq c) \end{aligned}$$

であるが，部分積分を用いると右辺第 1 項は以下のように書き換えられる．

$$\begin{aligned} \int_{-c-\mu}^{c-\mu} {x_1}^2 \phi(x_1)\, dx_1 &= \Big[-x_1 \phi(x_1)\Big]_{-c-\mu}^{c-\mu} + \int_{-c-\mu}^{c-\mu} \phi(x_1)\, dx_1 \\ &= -(c-\mu)\phi(c-\mu) + (-c-u)\phi(-c-\mu) + \int_{-c}^{c} \phi(x_1 - \mu)\, dx_1 \\ &= -(c-\mu)\phi(c-\mu) - (c+\mu)\phi(c+\mu) + P(|X_1| < c) \end{aligned}$$

これより

$$R(\hat{\mu}, \mu) = -(c-\mu)\phi(c-\mu) - (c+\mu)\phi(c+\mu) + 1$$

である．対称性から，与えられた c に対してリスクが最小となる μ は $\mu = 0$ であり，そのときのリスクは

$$R(\hat{\mu}, 0) = -2c\phi(c) + 1$$

これを $c > 0$ の範囲で最小化すると，$c\phi(c)$ の c での微分が $(1-c^2)\phi(c)$ であることから $c = 1$ でリスクが最小化される．そのリスクは

$$1 - 2\phi(1) = 1 - \frac{1}{\sqrt{2\pi}}e^{-1/2} = 0.75803 > 0.5$$

これより $\hat{\mu}$ のリスクは常に $\widetilde{\mu}$ のリスクより大である．

第 6 章

問 6.1　同時確率関数が

$$p^n(1-p)^{\sum_{i=1}^n x_i}$$

と書けるから，分解定理より $T = \sum_{i=1}^n X_i$ が十分統計量である．

問 6.2　同時密度関数

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

は $t = \sum_{i=1}^n {x_i}^2$ 与えると，半径 $\sqrt{t}$ の超球上一定である．このことから，条件つき分布は半径 $\sqrt{t}$ の超球上の一様分布である．

問 6.3　$\mathbb{R}^n$ の点 (ベクトル) $P = (X_1 - \bar{X}, \ldots, X_n - \bar{X})$ を考えると，これは $(\bar{X}, \ldots, \bar{X})$ と直交している．前問と同様に考えると，$t_2 = \sum_{i=1}^n (X_i - \bar{X})^2$ とおくと，点 P は $\mathbb{R}^n$ の半径 $\sqrt{t_2}$ の超球のうち $(\bar{X}, \ldots, \bar{X})$ と直交する集合 (次元が 1 つ下がった超球) 上の一様分布に従うことが理解される．

問 6.4　同時密度関数が

$$f_{\theta_1,\theta_2}(x) = \frac{1}{(\theta_2 - \theta_1)^n} I_{[\min_i X_i \geq \theta_1]}(x) I_{[\max_i X_i \leq \theta_2]}(x)$$

と書けることより，分解定理により $T = (\min_i X_i, \max_i X_i)$ が (θ_1, θ_2) に関する 2 次元の十分統計量であることがわかる．

条件つき分布については，$X_1, \ldots, X_n$ の中から最小値になる X_i と最大値となる X_j が無作為に選ばれ，その後残りの X_k は独立に一様分布 $\mathrm{U}[\min_i X_i, \max_i X_i]$ に従う．

問 6.5　2 項分布については

$$f(x,p) = \binom{n}{x} \exp\left(x \log \frac{p}{1-p} + n \log(1-p) \right)$$

と書けば指数型分布族の形となる．正規分布については基準化定数を指数の部分に取り込み，$\sum_{i=1}^{n} (x_i - \mu)^2 = \sum_{i=1}^{n} {x_i}^2 - 2n\mu\bar{x} + n\mu^2$ と展開すれば

$$f(x, \mu, \sigma^2) = \exp\left(\frac{n\mu}{\sigma^2}\bar{x} - \frac{1}{2\sigma^2} \sum_{i=1}^{n} {x_i}^2 - \frac{n\mu^2}{2\sigma^2} - \frac{n}{2}\log(2\pi\sigma^2) \right)$$

と書けば指数型分布族の形となる．

問 6.6　これらの分布のそれぞれについて以下のように指数型分布族の形に書ける．ただし負の 2 項分布においては r は既知とする．

$$f(x, \lambda) = \frac{\lambda^x}{x!} \exp(-\lambda) = \frac{1}{x!} \exp(x \log \lambda - \lambda)$$

$$f(x, p) = \frac{\Gamma(r+x)}{\Gamma(r)x!} (1-p)^x p^r = \frac{\Gamma(r+x)}{\Gamma(r)x!} \exp(x \log(1-p) + r \log p)$$

$$f(x, \nu, \alpha) = \frac{1}{\alpha^\nu \Gamma(\nu)} x^{\nu-1} e^{-x/\alpha} = \exp((\nu - 1)\log x - x/\alpha - \log(\alpha^\nu \Gamma(\nu)))$$

問 6.7　前問のように指数型分布族の形に書けば，完備性は容易に確認できる．

問 6.8　T が強い意味で最小十分とし，$U = g(T)$ が十分とする．強い意味の最小十分性から $T = h(U)$ とも書けるから，g に逆関数が存在し g は 1 対 1 となる．従って T は弱い意味でも最小十分である．

次に存在を仮定された強い意味での最小十分統計量を T とし，S を弱い意味で最小十分とする．S は十分であるから $T = h(S)$ と書ける．しかしながら S は弱い意味で最小十分であるからこの h は 1 対 1 でなければならない．従って T と S は 1 対 1 対応し，このことから S も強い意味で最小十分であることがわかる．

問 6.9　容易に確かめられる．

問 6.10　$f(x)=h(g(x))$ と書けるとき $g(x)=g(x')$ ならば $f(x)=f(x')$ であるから g による特定の同値類は f のある同値類の部分集合となる．従って P_f は P_g より粗い分割となる．

逆に P_f が P_g より粗い分割となるとき $a=g(x)=g(x')$ となる P_g の特定の同値類は，ある b について $b=f(x)=f(x')$ となる P_f の特定の同値類の部分集合となるから，このような a,b について $h(a)=b$ とする．これを g の値域のすべての a についておこなえば $f(x)=h(g(x))$ となる h が定まる．

問 6.11　この結果は，前問と強い最小十分統計量の定義から得られる．

問 6.12　同値関係であることは，与えられた条件から

$$h(x,x)=1,\quad h(x,x')=1/h(x',x),\quad h(x,x'')=h(x,x')h(x',x'')$$

となることからわかる．離散分布の場合に考えると，任意の特定の同値類の 2 点での確率の比が $h(x,x')$ によって与えられ θ には依存しないから，その同値類での条件つき分布は θ に依存しない．従ってこの同値関係によって導かれる分割は十分である．また $U(x)$ が十分で確率関数 (あるいは密度関数) が $f(x,\theta)=q(x)g_\theta(U(x))$ と書けるときは $h(x,x')=q(x')/q(x)$ とおけばよい．

問 6.13　問 6.4 と同様に

$$f_{a,b}(x)=\frac{1}{(b-a)^n}I_{[\min_i X_i\ge a]}(x)I_{[\max_i X_i\le b]}(x)$$

であるから，$h\equiv 1$ とすればよい．

問 6.14　(4.52) 式より $(t_1,t_2)=(\min_i X_i,\max_i X_i)$ の同時密度関数は

$$f(t_1,t_2)=\frac{n(n-1)}{(\theta_2-\theta_1)^n}(t_2-t_1)^{n-2}I_{[\theta_1\le t_1\le t_2\le\theta_2]}(t_1,t_2)$$

である．測度論的な議論を避けるた $g(t_1,t_2)$ を連続関数として

$$\begin{aligned}&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}g(t_1,t_2)f(t_1,t_2)\,dt_1dt_2\\ &\qquad=\frac{n(n-1)}{(\theta_2-\theta_1)^n}\int_{\theta_1\le t_1\le t_2\le\theta_2}g(t_1,t_2)(t_2-t_1)^{n-2}\,dt_1dt_2\\ &\qquad=0,\quad\forall\,(\theta_1,\theta_2)\ \ (\theta_1<\theta_2)\end{aligned}$$

と仮定する．$n(n-1)/(\theta_2-\theta_1)^n$ の部分を外して考えると

$$\int_{\theta_1 \le t_1 \le t_2 \le \theta_2} g(t_1, t_2)(t_2 - t_1)^{n-2}\, dt_1 dt_2 = 0, \quad \forall(\theta_1, \theta_2) \;\; (\theta_1 < \theta_2)$$

であるが，Δt_2 を小として積分領域を図示してみるとわかるように

$$\begin{aligned}
&\int_{\theta_1 \le t_1 \le t_2 \le \theta_2 + \Delta\theta_2} g(t_1, t_2)(t_2 - t_1)^{n-2}\, dt_1 dt_2 \\
&- \int_{\theta_1 \le t_1 \le t_2 \le \theta_2} g(t_1, t_2)(t_2 - t_1)^{n-2}\, dt_1 dt_2 \\
&\quad = \Delta\theta_2 \times \int_{\theta_1}^{\theta_2} g(t_1, \theta_2)(\theta_2 - t_1)^{n-2}\, dt_1 + o(\Delta\theta_2)
\end{aligned}$$

である．従って

$$\begin{aligned}
0 &= \frac{\partial}{\partial\theta_2} \int_{\theta_1 \le t_1 \le t_2 \le \theta_2} g(t_1, t_2)(t_2 - t_1)^{n-2}\, dt_1 dt_2 \\
&= \int_{\theta_1}^{\theta_2} g(t_1, \theta_2)(\theta_2 - t_1)^{n-2}\, dt_1, \quad \forall(\theta_1, \theta_2) \; (\theta_1 < \theta_2)
\end{aligned}$$

が成り立つ．　これを θ_1 で微分することにより

$$0 = g(\theta_1, \theta_2)(\theta_2 - \theta_1)^{n-2}, \quad \forall(\theta_1, \theta_2) \; (\theta_1 < \theta_2)$$

であり $g \equiv 0$ を得る．これより $(\min_i X_i, \max_i X_i)$ の完備性が示された．

なお，測度論的な議論を用いることにより，g は任意の可測関数でもよいことがわかる．

問 6.15

$$\frac{f(x, \psi)}{f(x', \psi)} = \frac{h(x)}{h(x')} \exp\left(\sum_{j=1}^{k} (T_j(x) - T_j(x'))\psi_j \right)$$

が ψ に依存しないための必要十分条件は $T_j(x) = T_j(x')$, $j = 1, \ldots, k$ であるから，$(T_1(x), \ldots, T_k(x))$ が最小十分統計量であることがわかる．

問 6.16　同時密度関数は

$$f(x, \alpha, \theta) = \frac{1}{\alpha^n} I_{[\min_i x_i \ge \theta]}(x) e^{-\sum_{i=1}^{n} x_i/\alpha + n\theta/\alpha}$$

である．これより $(\min_i X_i, \sum_{i=1}^{n} X_i)$ が 2 次元の十分統計量である．あるいはこれと 1 対 1 変換の $(T_1, T_2) = (\min_j X_j, \sum_{i=1}^{n}(X_i - \min_j X_j))$ を以下では用いる．

完備性も次のように示される．実は指数分布の性質より $T_1 = \min_j X_j$ は左端が θ で尺度母数が α/n の指数分布

$$f(t_1) = \frac{n}{\alpha} e^{-n(t_1-\theta)/\alpha}, \quad t_1 \geq \theta$$

$\sum_{i=1}^{n}(X_i - \min_j X_j)$ は左端が 0 で尺度母数が α の $n-1$ 個の指数分布の畳み込み，すなわちガンマ分布 $\mathrm{Ga}(n-1, \alpha)$

$$f(t_2) = \frac{1}{\alpha^{n-1}\Gamma(n-1)} {t_2}^{n-2} e^{-t_2/\alpha}$$

に従い，かつ T_1 と T_2 は互いに独立であることが示される．従って，基準化定数を無視して

$$\int_\theta^\infty \int_0^\infty g(t_1, t_2) e^{-n(t_1-\theta)/\alpha} {t_2}^{n-2} e^{-t_2/\alpha}\, dt_2 dt_1 = 0, \quad \forall\,(\theta, \alpha)$$

の条件を考察すればよい．g が連続関数の場合に示せばよい．これを θ で微分すれば

$$\int_0^\infty g(\theta, t_2) {t_2}^{n-2} e^{-t_2/\alpha}\, dt_2 = 0, \quad \forall\,(\theta, \alpha)$$

を得る．t_2 については指数型分布族の形をしているので，$g \equiv 0$ となり，完備性が示される．

第 7 章

問 7.1　ガンマ分布 $\mathrm{Ga}(\nu, 1)$ について

$$E[X] = \frac{\Gamma(\nu+1)}{\Gamma(\nu)} = \nu, \quad E[X^2] = \frac{\Gamma(\nu+2)}{\Gamma(\nu)} = \nu(\nu+1),$$

$$\mathrm{Var}[X] = \nu(\nu+1) - \nu^2 = \nu$$

である．これよりガンマ分布 $\mathrm{Ga}(\nu, \alpha)$ の期待値は $\alpha\nu$, 分散は $\alpha^2\nu$ である．自由度 k のカイ二乗分布は $\mathrm{Ga}(\nu/2, 2)$ であるから，分散は $2k$ であることがわかる．

これにより $(n-1)s^2/\sigma^2$ の分散は $2(n-1)$ であり，s^2 の分散は

$$2(n-1)\frac{\sigma^4}{(n-1)^2} = \frac{2\sigma^4}{n-1}$$

と確かめられる．

問 7.2　確率関数 $f(x) = \dfrac{\lambda^x}{x!}e^{-\lambda}$ から，$n=1$ として，対数尤度関数は

$$\ell(\lambda) = x\log\lambda - \lambda - \log(x!)$$

でこれを 2 回微分すると

$$\ell''(\lambda) = -\frac{x}{\lambda^2}$$

よりフィッシャー情報量は $I(\lambda) = E[-\ell''(\lambda)] = E[X]/\lambda^2 = 1/\lambda$ である．一方 $\hat{\lambda} = \bar{X}$ の分散は $\lambda/n = 1/(nI(\lambda))$ であるから，クラメル・ラオの不等式により $\bar{X}$ は λ の UMVU であることがわかる．

問 7.3　(7.27) 式にあるように σ^2 のフィッシャー情報量は $1/(2\sigma^4)$ である．一方 $\sum_{i=1}^{n}(X_i-\mu)^2/n$ は σ^2 の不偏推定量で，その分散は問 7.1 にあるように $2\sigma^4/n$ である．従ってクラメル・ラオの不等式により UMVU であることが確認される．

問 7.4　$\theta = 1$ の場合で考えると

$$E[T] = \int_0^1 tnt^{n-1}\,dt = n\left[\frac{t^{n+1}}{n+1}\right]_0^1 = \frac{n}{n+1},$$

$$E[T^2] = \frac{n}{n+2},$$

$$\mathrm{Var}[T] = \frac{n}{n+2} - \frac{n^2}{(n+1)^2} = \frac{n}{(n+1)^2(n+2)}$$

となる．これより $(n+1)T/n$ が θ の不偏推定量であることがわかる．その分散は

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

と確かめられる．

問 7.5　平方根 $f(x) = \sqrt{x}$ は厳密に上に凸であるから，ジェンセンの不等式より

$$E[s] = E\left[\sqrt{s^2}\right] < \sqrt{E[s^2]} = \sqrt{\sigma^2} = \sigma$$

である．

問 7.6　$X \sim \mathrm{Ga}(\nu,\alpha)$ のとき，X^a の期待値 (ただし a は整数とは限らない) は

$$E[X^a] = \int_0^\infty x^a\frac{x^{\nu-1}}{\alpha^\nu\Gamma(\nu)}e^{-x/\alpha}\,dx = \alpha^a\frac{\Gamma(\nu+a)}{\Gamma(\nu)}$$

である．$\nu=(n-1)/2,\ \alpha=2$ の場合に応用すると $s\sqrt{n-1}\Gamma((n-1)/2)/[\sqrt{2}\Gamma(n/2)]$ が σ の不偏推定量であることが容易にわかる．またこの推定量は完備十分統計量の関数であるから UMVU でもある．

問 7.7 $\bar{x}$ と s^2 は独立であるから，特にそれらの動く範囲も直積 (互いに制約がない) である．従って $\bar{X}$ が 0 に近く，かつ s^2 が大きく，(7.38) 式の $\hat{\mu}^2$ が負となる確率も正である．

問 7.8 対偶を示す．$P(X<0)=0$ すなわち X が非負の確率変数ならば，T で条件つけても $X<0$ となる条件つき確率はやはり 0 である．

問 7.9 幾何分布は負の 2 項分布の特殊ケースであるから，(2.62) 式 (問 2.8 も参照) より平均及び分散は $(1-p)/p,\ (1-p)/p^2$ である．対数尤度関数は $\ell(p)=x\log(1-p)+\log p$ で，これを p で 2 回微分すると

$$-\ell''(p)=\frac{x}{(1-p)^2}+\frac{1}{p^2}$$

であり，この期待値をとることによりフィッシャー情報量は

$$I(p)=\frac{1-p}{p}\frac{1}{(1-p)^2}+\frac{1}{p^2}=\frac{1}{p^2(1-p)}$$

と求まる．

問 7.10 1 から n までの番号のついた区別できる箱の中に，区別のつかない y 個の玉を入れ，x_i を箱 i に入った玉の数とすれば $y=x_1+\cdots+x_n$ となる．玉の入れ方とこのような表し方は 1 対 1 に対応している．箱 i に X_i 個の玉を入れることを，箱 i を x_i 回選ぶことと考えると，表し方の総数は「n 個の箱を重複を許して合計 y 回選ぶ選び方の総数」すなわち「重複組合せ」となり，$\binom{n+y-1}{y}$ である．

問 7.11 幾何分布は $r=1$ の場合の負の 2 項分布である．(3.51) 式あるいは問 3.8 にあるように，負の 2 項分布は畳み込みについて閉じているので，成功確率が p の幾何分布を n 回畳み込んだものは負の 2 項分布 $\mathrm{NB}(n,p)$ である．

問 7.12 $X_1,\dots,X_n$ の同時確率は

$$\prod_{i=1}^{n}p(1-p)^{x_i}=p^n(1-p)^{\sum_{i=1}^{n}x_i}$$

であるから $Y = \sum_{i=1}^{n} X_i$ が十分統計量である．完備性は問 6.7 で示されている．$Y = y$ を与えたときの $X_1, \ldots, X_n$ の条件つき分布は，同時確率を Y の確率関数で割って求められるので

$$P(X_1 = x_1, \ldots, X_n = x_n \mid Y = y) = 1 \Big/ \binom{n+y-1}{y}$$

で与えられる．

問 7.13

$$\binom{n+y-1}{y} p^n (1-p)^y$$

を最大化する p は 2 項分布の場合と同様に求められるから最尤推定量は $\hat{p} = n/(n+y)$ である．(7.62) 式は次のように示される．

$$\begin{aligned}
&\sum_{y=0}^{\infty} \frac{1}{(n+y-1)(n+y-2)\ldots(n+y-k)} \binom{n+y-1}{y} p^n (1-p)^y \\
&= \sum_{y=0}^{\infty} \frac{1}{(n+y-1)(n+y-2)\ldots(n+y-k)} \frac{(n+y-1)!}{y!\,(n-1)!} p^n (1-p)^y \\
&= \frac{p^k}{(n-1)(n-2)\ldots(n-k)} \sum_{y=0}^{\infty} \frac{(n-k+y-1)!}{y!\,(n-k-1)!} p^{n-k} (1-p)^y \\
&= \frac{p^k}{(n-1)(n-2)\ldots(n-k)}
\end{aligned}$$

問 7.14　Y を与えたもとで $P(X_1 = 0 | Y = y)$ の条件つき確率を求めればよい．問 7.10 の設定で，箱 1 に玉がはいらない場合の数は n を $n-1$ で置き換えればよいから $\binom{n+y-2}{y}$ である．そこで

$$P(X_1 = 0 | Y = y) = \frac{\binom{n+y-2}{y}}{\binom{n+y-1}{y}} = \frac{n-1}{n+y-1}$$

となり一致する．

問 7.15　容易に確かめられる．

問 7.16　$Y = \sum_{i=1}^{n} X_i$ は $\mathrm{Po}(n\lambda)$ に従うから

$$e^{-2\lambda} = \sum_{y=0}^{\infty} \delta(y) \frac{(n\lambda)^y}{y!} e^{-n\lambda}$$

あるいは

$$e^{(n-2)\lambda} = \sum_{y=0}^{\infty} \delta(y) \frac{(n\lambda)^y}{y!}$$

を考察する．$e^{(n-2)\lambda} = \sum_{y=0}^{\infty} ((n-2)\lambda)^y / y!$ と λ の級数として一致するから

$$\delta(y) = \left(\frac{n-2}{n}\right)^y$$

を得る．

問 7.17　$(\partial/\partial b)(b^\top Ab) = 2Ab$，すなわち 2 次形式 $b^\top Ab$ を b の各要素で偏微分して列ベクトルは $2Ab$ となること，及び関数の比の微分の公式から

$$0 = \frac{\partial}{\partial b} \frac{(a^\top J(\theta)b)^2}{b^\top I(\theta)b} = \frac{2}{(b^\top I(\theta)b)^2}((b^\top I(\theta)b)J(\theta)^\top a - (a^\top J(\theta)b)I(\theta)b)$$

を解くと，$b^\top I(\theta)b$ 及び $a^\top J(\theta)b$ はスカラーだから $b \propto I(\theta)^{-1}J(\theta)^\top a$ を得る．

なお，この問題では $(a^\top J(\theta)b)^2/b^\top I(\theta)b$ は b のスケール倍に依存しないから，分母を 1 に制約して，ラグランジュ乗数法を用い

$$(a^\top J(\theta)b)^2 - \lambda(b^\top I(\theta)b - 1)$$

を最大化するのが通常である．

問 7.18　略．13 章の (13.60) 式の導出と同様に議論すればよい．

問 7.19　p_3 が既知かどうかの違いを見るため，以下では $p_2 = 1 - p_1 - p_3$ として考える．多項係数の部分を省略して，対数尤度は

$$\ell(p_1, p_3) = X_1 \log p_1 + X_2 \log(1 - p_1 - p_3) + X_3 \log p_3$$

である．この 2 階微分の符号を変えたものは，それぞれ

$$-\frac{\partial^2}{\partial {p_1}^2}\ell = \frac{X_1}{{p_1}^2} + \frac{X_2}{(1-p_1-p_3)^2},$$

$$-\frac{\partial^2}{\partial {p_3}^2}\ell = \frac{X_3}{{p_3}^2} + \frac{X_2}{(1-p_1-p_3)^2},$$

$$-\frac{\partial^2}{\partial p_1 \partial p_3}\ell = \frac{X_2}{(1-p_1-p_3)^2}$$

となり，これらの期待値をとればフィッシャー情報行列は

$$I(\theta) = \begin{pmatrix} \dfrac{1}{p_1} + \dfrac{1}{1-p_1-p_3} & \dfrac{1}{1-p_1-p_3} \\ \dfrac{1}{1-p_1-p_3} & \dfrac{1}{p_3} + \dfrac{1}{1-p_1-p_3} \end{pmatrix}$$

$$= \begin{pmatrix} \dfrac{1}{p_1} + \dfrac{1}{p_2} & \dfrac{1}{p_2} \\ \dfrac{1}{p_2} & \dfrac{1}{p_3} + \dfrac{1}{p_2} \end{pmatrix}$$

となる．行列式は

$$\det I(\theta) = \left(\frac{1}{p_1} + \frac{1}{p_2}\right)\left(\frac{1}{p_3} + \frac{1}{p_2}\right) - \frac{1}{{p_2}^2} = \frac{1}{p_1p_2p_3}(p_1+p_2+p_3) = \frac{1}{p_1p_2p_3}$$

となり，逆行列は

$$I(\theta)^{-1} = \begin{pmatrix} p_1(p_2+p_3) & -p_1p_3 \\ -p_1p_3 & p_3(p_1+p_2) \end{pmatrix} = \begin{pmatrix} p_1(1-p_1) & -p_1p_3 \\ -p_1p_3 & p_3(1-p_3) \end{pmatrix}$$

である．p_3 が未知の場合は最尤推定量は $\hat{p}_1 = X_1/n$ であり，その分散が $p_1(1-p_1)/n$ であるからこれは $I(\theta)^{-1}$ の $(1,1)$ 要素に対応している．

一方で p_3 が既知ならば，最尤推定量の分散は近似的に

$$\frac{1}{n}\frac{1}{\frac{1}{p_1}+\frac{1}{p_2}} = \frac{1}{n}\frac{p_1p_2}{1-p_3}$$

となる．なお p_3 が既知の場合の p_1 の最尤推定量 $\widetilde{p}_1$ は対数尤度を p_1 で微分して 0 とおくことにより

$$0 = \frac{X_1}{p_1} - \frac{X_2}{1-p_1-p_3}$$

から

$$\widetilde{p}_1 = (1-p_3)\frac{X_1}{X_1+X_2}$$

と求められる．デルタ法により $\widetilde{p}_1$ の分散が上の近似値であることが確かめられる．

漸近分散の n 倍の差を見ると

$$p_1(1-p_1) - \frac{p_1p_2}{p_1+p_2} = \frac{p_1}{1-p_3}((1-p_1)(1-p_3) - (1-p_1-p_3))$$

$$= \frac{{p_1}^2p_3}{1-p_3} > 0$$

である．

問 7.20

$$\mathrm{Var}[\hat{\mu}^2] = \mathrm{Var}[\bar{X}^2] = E[\bar{X}^4] - E[\bar{X}^2]^2$$

を求めればよい．$X \sim \mathrm{N}(\mu, \sigma^2)$ とすると $X = \mu + \sigma Z$, $Z \sim \mathrm{N}(0,1)$ と表されるから，$E[Z^4] = 3$ を用いると

$$E[X^2] = \mu^2 + \sigma^2, \quad E[X^4] = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$$

より

$$\mathrm{Var}[X^2] = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 - (\mu^2+\sigma^2)^2 = 4\mu^2\sigma^2 + 2\sigma^4$$

である．問題の設定では $\bar{X} \sim \mathrm{N}(\mu, 1/n)$ より

$$\mathrm{Var}[\bar{X}^2] = \frac{4\mu^2}{n} + \frac{2}{n^2}$$

となり (7.98) 式が得られる．

次に $\tau = \mu^2$ とおくと，$\mu = \sqrt{\tau}$ であり

$$\frac{d\mu}{d\tau} = \frac{1}{2\sqrt{\tau}} = \frac{1}{2\mu}$$

であり，また $I(\mu) = 1$ より母数の変換にともなうフィッシャー情報量の変化の議論より

$$I(\tau) = \left(\frac{1}{2\mu}\right)^2 = \frac{1}{4\tau}$$

となり (7.99) 式が得られる．

(7.100) 式，(7.101) 式は容易に確かめられる．

問 7.21　確率母関数 $G(s) = e^{\lambda(s-1)}$ より $E[X(X-1)] = G''(1) = \lambda^2$ となるので $X(X-1)$ は λ^2 の不偏推定量である．同様に $E[X(X-1)\dots(X-k+1)] = \lambda^k$ である．$X, X(X-1), X(X-1)(X-2), \dots$ と $X, X^2, X^3, \dots$ の関係は線形であるから，これらからポアソン分布の任意のモーメントが計算できる．例えば

$$\mathrm{Var}[X(X-1)] = 4\lambda^3 + 2\lambda^2$$

となることが示される．一方，尤度関数 $L = (\lambda^x/x!)e^{-\lambda}$ を λ で微分すると

$$L' = \left(\frac{x}{\lambda} - 1\right)L, \quad L'' = \left(\left(\frac{x}{\lambda} - 1\right)^2 - \frac{x}{\lambda}\right)L$$

よりバッタチャリアの不等式の右辺が計算できる．ただし，この右辺の計算はやや面倒なため，以下のような簡便な議論を示す．

3.3 節の確率変数の間の回帰分析の考え方を用いると，$X(X-1)$ を L'/L 及び L''/L の線形結合で予測したときに，予測誤差の二乗の期待値の最小値は $\sigma_{YY} - {\sigma_{XY}}^\top {\Sigma_{XX}}^{-1} \sigma_{XY}$ で与えられる．バッタチャリアの不等式も $\sigma_{YY} \geq {\sigma_{XY}}^\top {\Sigma_{XX}}^{-1} \sigma_{XY}$ の形で与えられている．従ってバッタチャリアの不等式が等式で成り立つことを示すには，$X(X-1)$ が定数，$L'/L, L''/L$ の線形結合で表されることを示せばよい．実際

$$X(X-1) = \lambda^2 \left(\left(\frac{x}{\lambda} - 1 \right)^2 - \frac{x}{\lambda^2} \right) + 2\lambda^2 \left(\frac{x}{\lambda} - 1 \right) + \lambda^2$$

となることから，バッタチャリヤの不等式により $X(X-1)$ が λ^2 の UMVU であることが示される．

問 7.22　問 6.14 より $(\min_i X_i, \max_i X_i)$ は完備十分統計量であるから，これらの関数で不偏推定量を求めればよい．U[0, 1] からの標本の最大値の期待値が $n/(n+1)$ であったことから，位置尺度変換すれば

$$E\left[\max_i X_i\right] = \theta_1 + \frac{n}{n+1}(\theta_2 - \theta_1) = \frac{1}{n+1}\theta_1 + \frac{n}{n+1}\theta_2$$

である．左右を反転して考えれば

$$E\left[\min_i X_i\right] = \theta_2 - \frac{n}{n+1}(\theta_2 - \theta_1) = \frac{n}{n+1}\theta_1 + \frac{1}{n+1}\theta_2$$

である．これから線形連立方程式を解けば，θ_1, θ_2 の UMVU は

$$\hat{\theta}_1 = \frac{n \min_i X_i - \max_i X_i}{n-1}, \quad \hat{\theta}_2 = \frac{n \max_i X_i - \min_i X_i}{n-1}$$

である．

問 7.23　問 2.11 より確率関数は $p(x) = \theta^x / (x(-\log(1-\theta)))$ で，単一の観測値の対数尤度は $\ell(\theta) = x \log \theta - \log(-\log(1-\theta)) - \log x$ である．

$$\ell'(\theta) = \frac{x}{\theta} - \frac{1}{(-\log(1-\theta))(1-\theta)}$$

で問 2.11 の分散の計算より

$$I(\theta) = \frac{\mathrm{Var}[X]}{\theta^2} = \frac{1}{(-\log(1-\theta))^2 \theta (1-\theta^2)} \left(-\log(1-\theta) - \theta \right)$$

$X_1, \ldots, X_n$ の同時確率関数は

$$\frac{1}{\prod_{i=1}^{n} x_i} \frac{1}{(-\log(1-\theta))^n} \theta^{\sum_{i=1}^{n} x_i}$$

で対数尤度関数は

$$\ell(\theta) = \sum_{i=1}^{n} x_i \log\theta - n\log(-\log(1-\theta)) - \sum_{i=1}^{n} \log x_i$$

で，これを微分して 0 とおくと尤度方程式は

$$\bar{x} = E_\theta[X] = \frac{1}{-\log(1-\theta)} \frac{\theta}{1-\theta}$$

を得る．尤度方程式は明示的には解けないが，増減表を作ることにより，右辺は 1 から ∞ まで増加する関数であることがわかるから最尤推定値 $\hat{\theta}$ は一意に定まる．最尤推定量の漸近分布は理論通り $\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} \mathrm{N}(0, 1/I(\theta))$ である．

問 7.24　$X_1, \ldots, X_n$ を対数正規分布からの無作為標本とする．$Y_i = \log X_i$ は正規分布 $\mathrm{N}(\mu, \sigma^2)$ に従う．対数正規分布の密度関数からも求められるが，Y_i に変換してしまえば正規分布の十分統計量 $(T_1, T_2) = (\bar{Y}, \sum_{i=1}^{n}(Y_i - \bar{Y})^2)$ に基づいて推定をおこなえばよい．正規分布の中央値は μ であるから，対数正規分布の中央値は e^μ である．最尤推定量はパラメータの変換について不変であるから，e^μ の最尤推定量 $e^{\bar{Y}}$ である．

e^μ の不偏推定については次のように議論する．(T_1, T_2) は完備十分統計量だから，これらに基づく不偏推定量を見つければそれが UMVU である．まず，正規分布の積率母関数から

$$E\left[e^{\bar{Y}}\right] = e^{\mu+\sigma^2/(2n)}$$

となることがわかる．$e^{\sigma^2/(2n)}$ の部分がバイアスとなっているから，ここをキャンセルするように修正したい．そのために T_2 を用いる．T_1 と T_2 が独立であることから T_2 の関数で，その期待値が $e^{-\sigma^2/(2n)}$ となるものを求め，それを $e^{\bar{Y}}$ にかければよい．そのような関数がすぐには見つからないので，級数展開を利用しよう．

$$e^{-\sigma^2/(2n)} = \sum_{k=0}^{\infty} \frac{1}{k!}\left(-\frac{1}{2n}\right)^k \sigma^{2k}$$

より σ^{2k} の不偏推定量を求めればよい．ここで

$$E\left[{T_2}^k\right] = \sigma^{2k}(n-1)(n+1)(n+3)\cdots(n+2k-3)$$

であるから $c_{n,k}=(n-1)(n+1)(n+3)\cdots(n+2k-3)$ とおけば ${T_2}^k/c_{n,k}$ が σ^{2k} の不偏推定量となる．これより

$$\widehat{e^{-\sigma^2/(2n)}}=\sum_{k=0}^{\infty}\frac{1}{k!\,c_{n,k}}\left(-\frac{1}{2n}\right)^k {T_2}^k$$

が $e^{-\sigma^2/(2n)}$ の不偏推定量である．e^{μ} の UMVU はこれに $e^{\bar{Y}}$ をかけたものである．

第 8 章

問 8.1　c を増加させると棄却域が小さくなるから $\alpha(c,r)$ は c の減少関数である．また r を増加させると $\delta_{c,r}$ も (尤度比が c となる点で) 増加するから α は r の増加関数である．$c=\infty, r=0$ とすると常に受容で $\alpha(\infty,0)=0$，また $c=0, r=1$ とすると常に棄却となり $\alpha(0,1)=1$ より，最小値は 0 最大値は 1 である．

問 8.2　ネイマン・ピアソンの補題の証明の不等式において等式が成り立つためには $\delta_{c,r}(x)=1$ のときに $\delta(x)=1$ も成り立たなければならず，$\delta_{c,r}(x)=0$ のときに $\delta(x)=0$ も成り立たなければならないことから (8.23) 式が成り立つ．

問 8.3　$a=p_1(1-p_0)/(p_0(1-p_1))>1$, $b=c(1-p_0)^n/(1-p_1)^n$ とおくと左辺は $a^x>b$ と表せるから，対数をとって $x>\log b/\log a$ を得る．右辺が整数でないときは整数に切り下げてもよい．

問 8.4　$f(x)$ を X の密度関数とすれば補題 8.4 は

$$E[g(T(X))]\,E[h(T(X)]\le E[g(T(X))h(T(X))]$$

を主張している．T の密度関数を $f_T(t)$ とすれば不等式は

$$\int_{-\infty}^{\infty}g(t)f_T(t)\,dt\int_{-\infty}^{\infty}h(t)f_T(t)\,dt\le\int_{-\infty}^{\infty}g(t)h(t)f_T(t)\,dt$$

となり，$T(x)=x$ の場合に帰着する．

問 8.5　ネイマン・ピアソンの補題を用いる．一様最強力検定は

$${X_1}^2+\cdots+{X_n}^2>\chi^2_n(\alpha)\ \Rightarrow\ \text{reject}$$

で与えられる．

問 8.6　ネイマン・ピアソンの補題を用いる．一様最強力検定は

$$X_1+\cdots+X_n>c_\alpha\ \Rightarrow\ \text{reject}$$

で与えられる．ただし c_α はパラメータ n のポアソン分布 $\mathrm{Po}(n)$ の上側 α 点である．離散分布なので水準が α に一致しないときは確率化をおこなう．

問 8.7　ネイマン・ピアソンの補題を用いる．一様最強力検定は

$$X_1 + \cdots + X_n > c_\alpha \ \Rightarrow\ \text{reject}$$

で与えられる．ただし c_α は負の 2 項分布 $\mathrm{NB}(nr, p_0)$ の上側 α 点である．離散分布なので水準が α に一致しないときは確率化をおこなう．

問 8.8　(8.54) と同値の (8.55) 式が点 P と点 Q で等式となればよい．このとき

$$\widetilde{c}_1 - c\widetilde{c}_2 = e^{-c\mu}$$

$$\widetilde{c}_1 + c\widetilde{c}_2 = e^{c\mu}$$

となり，これを解くと $\widetilde{c}_1 = (e^{c\mu} + e^{-c\mu})/2,\ \widetilde{c}_2 = (e^{c\mu} - e^{-c\mu})/(2c)$ を得る．

問 8.9　以下の例にあるように，(8.60) 式を書き下すと r_a, r_b を未知数とする 2 元の連立線形方程式が得られるので，それを解けば r_a, r_b が定まる．$0 \le r_a, r_b \le 1$ となる解が得られるように a, b はいくつか試す必要がある．両端の確率が $\alpha/2$ に近い (a, b) の組を試せばよい．

2 項分布の確率関数を

$$f(x, p) = \binom{n}{x} p^x (1-p)^{n-x}$$

とする．これを p で微分すると

$$f'(x, p) = \binom{n}{x}\left(\frac{x}{p} - \frac{n-x}{1-p}\right) p^x (1-p)^{n-x} = \frac{x - np}{p(1-p)} f(x, p)$$

である．(8.60) 式は

$$\alpha = \sum_{x=0}^{a-1} f(x, p_0) + r_a f(a, p_0) + r_b f(b, p_0) + \sum_{x=b+1}^{n} f(x, p_0)$$

$$0 = \sum_{x=0}^{a-1} f'(x, p_0) + r_a f'(a, p_0) + r_b f'(b, p_0) + \sum_{x=b+1}^{n} f'(x, p_0)$$

であるから，これを解くと

$$\begin{pmatrix} r_a \\ r_b \end{pmatrix} = \begin{pmatrix} f(a, p_0) & f(b, p_0) \\ f'(a, p_0) & f'(b, p_0) \end{pmatrix}^{-1} \begin{pmatrix} \alpha - \sum_{x=0}^{a-1} f(x, p_0) - \sum_{x=b+1}^{n} f(x, p_0) \\ - \sum_{x=0}^{a-1} f'(x, p_0) - \sum_{x=b+1}^{n} f'(x, p_0) \end{pmatrix}$$

$(a, b) = (0, 3)$ として解くと $(r_a, r_b) = (0.302, 0.283)$ を得る．

問 8.10　前問と同様に解く．確率関数を $f(x, \lambda) = (\lambda^x/x!)e^{-\lambda}$ とし，その λ での微分を

$$f'(x, \lambda) = \left(\frac{x}{\lambda} - 1\right) f(x, \lambda)$$

とおく．(r_a, r_b) の式は前問の解で $n = \infty$ と置いたもので与えられる．(a, b) として $(2, 9)$ を選ぶと $(r_a, r_b) = (0.210, 0.277)$ を得る．

問 8.11　問 2.6 と同様の人工的な例であるが X をベルヌーイ変数として帰無仮説を $H_0 : p = 1/2$ とし，$T(X) = X \geq 1$ (すなわち $X = 1$) のときに帰無仮説を棄却することとする．この検定のサイズは $\alpha = 1/2$ であり，これを有意水準とする．X' を X と独立なベルヌーイ変数とし，$P_{1/2}(X' > x \mid X = x)$ を考えると，この確率は $x = 0$ のとき $1/2$，$x = 1$ のとき 0 となるので $P_{1/2}(X' > x \mid X = x) \leq \alpha = 1/2$ を基準とすると $x = 0, 1$ のいずれでも棄却，つまり常に棄却となりサイズは 1 となる．つまり有意水準 $\alpha = 1/2$ ではない．

問 8.12　棄却域は

$$\sqrt{n}\frac{\bar{X} - \mu_0}{\sigma} \geq z_\alpha$$

で表される．$\bar{X} \sim \mathrm{N}(\mu_1, \sigma^2)$ のときのこの確率が $1 - \beta$ より大きくなればよい．不等式を以下のように書き換える．

$$\sqrt{n}\frac{\bar{X} - \mu_1}{\sigma} \geq z_\alpha + \sqrt{n}\frac{\mu_0 - \mu_1}{\sigma}$$

H_1 の下で左辺が標準正規分布に従うことを用いれば，この確率が $1 - \beta$ より大きいことは

$$z_\alpha + \sqrt{n}\frac{\mu_0 - \mu_1}{\sigma} < z_{1-\beta} = -z_\beta$$

と同値である．不等式を変形すれば

$$\frac{\sigma(z_\alpha + z_\beta)}{\mu_1 - \mu_0} < \sqrt{n}$$

であり，これを 2 乗すれば解を得る．

問 8.13　棄却域は

$$\frac{X - np_0}{\sqrt{np_0(1 - p_0)}} > z_\alpha$$

である．不等式を変形して

$$X - np_1 > z_\alpha\sqrt{np_0(1-p_0)} + n(p_0 - p_1)$$

さらに

$$\begin{aligned}\frac{X - np_1}{\sqrt{np_1(1-p_1)}} &> \frac{z_\alpha\sqrt{np_0(1-p_0)} - n(p_1 - p_0)}{\sqrt{np_1(1-p_1)}} \\ &= \frac{z_\alpha\sqrt{p_0(1-p_0)} - \sqrt{n}(p_1 - p_0)}{\sqrt{p_1(1-p_1)}}\end{aligned}$$

となるから，右辺が $-z_\beta$ 以下であればよい．

$$\frac{z_\alpha\sqrt{p_0(1-p_0)} - \sqrt{n}(p_1 - p_0)}{\sqrt{p_1(1-p_1)}} < -z_\beta$$

より

$$\sqrt{n} > \frac{z_\alpha\sqrt{p_0(1-p_0)} + z_\beta\sqrt{p_1(1-p_1)}}{p_1 - p_0}$$

を得る．

問 8.14　F が密度関数 $f(x) = F'(x)$ を持つ場合に示す．$X < c$ を棄却域とするとき $F(c)$ は第 1 種の過誤の確率である．また $G(c)$ は H_1 のもとでの検出力となる．従って $(F(c), G(c))$ は ROC 曲線の点である．$u = F(c)$ とおいて c を動かせば u は 0 から 1 の間を動くから ROC 曲線は $G(F^{-1}(u))$, $u \in (0,1)$, と表される．AUC は面積 $\displaystyle\int_0^1 G(F^{-1}(u))\,du$ であるが，$x = F^{-1}(u)$ と変数変換すると

$$\begin{aligned}\int_0^1 G(F^{-1}(u))\,du &= \int_{-\infty}^{\infty} G(x)f(x)\,dx = E^X[G(X)] \\ &= E^X[P(Y < X \,|\, X)] = P(Y < X)\end{aligned}$$

である．

問 8.15　補題 8.4 で g, h の符号を変えると，f, g で共に単調減少でも補題 8.4 の不等式が成り立っている．ここで $T(x) = x$, $g(x) = I_{(-\infty,c]}(x)$, $h(x) = g(x)/f(x)$ とおくと

$$\int_{-\infty}^{\infty} I_{(-\infty,c]}f(x)\,dx \int_{-\infty}^{\infty} \frac{g(x)}{f(x)}f(x)\,dx \le \int_{-\infty}^{\infty} I_{(-\infty,c]}(x)\frac{g(x)}{f(x)}f(x)\,dx$$

であるが，$\displaystyle\int_{-\infty}^{\infty} \frac{g(x)}{f(x)} f(x)\, dx = \int_{-\infty}^{\infty} g(x)\, dx = 1$ より，任意の c に対して $F(c) \leq G(c)$ を得る．

凸性は

$$G(F^{-1}(u))' = \frac{g(F^{-1}(u))}{f(F^{-1}(u))}$$

が単調減少であることからわかる．

問 8.16　前半は $X \sim \mathrm{N}(\mu_0, \sigma^2)$, $Y \sim \mathrm{N}(\mu_1, \sigma^2)$，$X, Y$ は独立として $P(X > Y)$ を求めればよい．$X - Y \sim \mathrm{N}(\mu_0 - \mu_1, 2\sigma^2)$ より

$$P(X - Y > 0) = \Phi((\mu_0 - \mu_1)/\sqrt{2\sigma^2})$$

である．

後半は ${X_1}^2 + \cdots + {X_n}^2$ の分布，すなわち自由度 n のカイ二乗分布の定数倍を考える．U, V を互いに独立な自由度 n のカイ二乗分布に従う確率変数として $X = U{\sigma_0}^2$, $Y = V{\sigma_1}^2$ とすると $P(X > Y) = P(X/Y > 1) = P(U/V > {\sigma_1}^2/{\sigma_0}^2)$ が AUC となる．U/V は自由度 (n, n) の F 分布に従うから F 分布の分布関数を計算すればよい．

問 8.17　$\hat{p} = x/n$ とし

$$L = \frac{\binom{n}{x} \hat{p}^x (1 - \hat{p})^{n-x}}{\binom{n}{x} {p_0}^x (1 - p_0)^{n-x}}, \quad 2\log L = -2x \log \frac{p_0}{\hat{p}} - 2(n - x) \log \frac{1 - p_0}{1 - \hat{p}}$$

として $\hat{p}$ の周りで 2 次項までテーラー展開する．

$$\begin{aligned}
-2x \log \frac{p_0}{\hat{p}} &= -2x \log\left(1 - \frac{\hat{p} - p_0}{\hat{p}}\right) \\
&= -2n\hat{p}\left(-\frac{\hat{p} - p_0}{\hat{p}} - \frac{1}{2}\left(-\frac{\hat{p} - p_0}{\hat{p}}\right)^2\right) + o_p(1) \\
&= 2n(\hat{p} - p_0) + n\hat{p}\left(\frac{\hat{p} - p_0}{\hat{p}}\right)^2 + o_p(1)
\end{aligned}$$

$$-2(n - x) \log \frac{1 - p_0}{1 - \hat{p}} = -2n(1 - \hat{p}) \log\left(1 + \frac{\hat{p} - p_0}{1 - \hat{p}}\right)$$

$$= -2n(1-\hat{p})\left(\frac{\hat{p}-p_0}{1-\hat{p}} - \frac{1}{2}\left(\frac{\hat{p}-p_0}{1-\hat{p}}\right)^2\right) + o_p(1)$$

これらを加えると最右辺第1項はキャンセルして

$$2\log L = n\frac{(\hat{p}-p_0)^2}{\hat{p}} + n\frac{(\hat{p}-p_0)^2}{1-\hat{p}} = \frac{n(\hat{p}-p_0)^2}{\hat{p}(1-\hat{p})} + o(1)$$

となる．これを自由度1のカイ二乗分布の上側 α 点と比較する．帰無仮説のもとで分母の $\hat{p}$ は p_0 に確率収束するから，(8.62) 式と同値である．

第9章

問 9.1 任意の θ_0 について

$$P_{\theta_0}(X \in A(\theta_0)) = P_{\theta_0}(\theta_0 \in S(X)) \geq 1-\alpha$$

より (9.8) 式が成り立ち，$A(\theta_0)$ は有意水準 α の受容域である．

問 9.2 (μ, X) 平面において (9.11) 式及び (912) 式は原点を通る傾き1の直線に平行な2直線で囲まれる領域となっている．(9.11) 式はその領域を縦 (X 軸に平行) に切った断面であり，(9.12) 式は横 (μ 軸に平行) に切った断面である．

問 9.3 $S(X)$ が一様最強力不偏信頼域とし，S から導かれる受容域を $A(\theta)$ とする．$\widetilde{S}(X)$ 及び $\widetilde{A}(\theta)$ を任意の他の不偏信頼域及び対応する受容域とする．このとき，θ_0 を帰無仮説のもとでの値，$\theta \neq \theta_0$ を対立仮説のもとでの値とすると

$$P_\theta(X \in A(\theta_0)) = P_\theta(\theta_0 \in S(X)) \leq P_\theta(\theta_0 \in \widetilde{S}(X)) = P_\theta(X \in \widetilde{A}(\theta_0))$$

であるから $A(\theta_0)$ を受容域とする検定の検出力のほうが $\widetilde{A}(\theta_0)$ を受容域とする検定の検出力より大きい．このことから補題 8.2 の後半が示される．

問 9.4 ポアソン分布の確率関数 $p(x,\lambda) = (\lambda^x/x!)e^{-\lambda}$ より，対数尤度は $x!$ を無視して $\ell = -\lambda + x\log\lambda$ である．対数尤度の2階微分の符号を変えたものは

$$-\ell'' = \frac{x}{\lambda^2}$$

であり，この期待値をとるとフィッシャー情報量は $I(\lambda) = 1/\lambda$ となる．最尤推定量は標本平均 $\hat{\lambda} = \bar{X}$ である．9.5 節の議論より，近似的な信頼区間は

$$\hat{\lambda} \pm z_{\alpha/2}\sqrt{\frac{\hat{\lambda}}{n}}$$

で与えられる．

問 9.5　前問と同様に考えると，確率関数を $f(x,p) = p(1-p)^x$, $x = 0, 1, \ldots$ として対数尤度の 2 階微分の符号を変えたものは

$$-\ell'' = \frac{x}{(1-p)^2} + \frac{1}{p^2}$$

となり，この期待値をとるとフィッシャー情報量は

$$I(p) = \frac{1}{p^2(1-p)}$$

である．最尤推定量は

$$\hat{p} = \frac{1}{1+\bar{X}}$$

である．近似的な信頼区間は

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}^2(1-\hat{p})}{n}}$$

で与えられる．

問 9.6　$p = 2$ のときは

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \leq P(A_1) + P(A_2)$$

である．$p \geq 3$ については帰納法を用いて

$$\begin{aligned} P((A_1 \cup \cdots \cup A_{p-1}) \cup A_p) &\leq P(A_1 \cup \cdots \cup A_{p-1}) + P(A_p) \\ &\leq P(A_1) + \cdots + P(A_{p-1}) + P(A_p) \end{aligned}$$

のように議論すればよい．(9.32) 式は $1 - P(A_1 \cup \cdots \cup A_p) = P({A_1}^c \cup \cdots \cup {A_p}^c)$ より明らかである．

問 9.7　［**第 2 刷までの注**：問題文の $\alpha = 0.95$ は誤りで，以下 $\alpha = 0.05$ として解答する．］

H_0 のもとでの $\theta_1 = \theta_2$ の値を 0 として一般性を失わない．$Z_1 = \hat{\theta}_{1n}\sqrt{nI(\hat{\theta}_{1n})}$, $Z_2 = \hat{\theta}_{2n}\sqrt{nI(\hat{\theta}_{2n})}$ とおくと，Z_1, Z_2 は近似的に互いに独立に標準正規分布に従う．帰無仮説を棄却するのは $Z_1 + z_{\alpha/2} < Z_2 - z_{\alpha/2}$ あるいは $Z_2 + z_{\alpha/2} < Z_1 - z_{\alpha/2}$ の場合で，これらの事象は排反で確率も等しい．そこで前者の確率を考えると，$z_{0.025} = 1.96$ として $Z_2 - Z_1 > 1.96 \times 2$，あるいは基準化して

$$\frac{Z_2 - Z_1}{\sqrt{2}} > 1.96 \times \sqrt{2} = 2.772$$

である．正規分布表よりこの確率は 0.0028 であるから，この検定方式のサイズはこの 2 倍の 0.56％ 程度である．

第 10 章

問 10.1　V を自由度 k のカイ二乗分布に従う確率変数とすると，$1/V$ の期待値は

$$\int_0^\infty \frac{1}{v}\frac{v^{k/2-1}}{2^{k/2}\Gamma(k/2)}e^{-v/2}dv=\frac{2^{k/2-1}\Gamma(k/2-1)}{2^{k/2}\Gamma(k/2)}=\frac{1}{k-2}$$

である．(10.14) 式が不偏推定量を与えることは，(10.14) 式で分母分子が互いに独立であり，$1\Big/\sum_{i=1}^m(X_i-\bar{X})^2$ の期待値が $1/({\sigma_1}^2(m-3))$ となることからわかる．

問 10.2　$m\le n$ として一般性を失わない．$m-1\le f$ を示すには

$$\left(\frac{{s_1}^2}{m}+\frac{{s_2}^2}{n}\right)^2\ge\frac{{s_1}^4}{m^2}+\frac{{s_2}^4(m-1)}{n^2(n-1)}$$

を示せばよい．展開して左辺と右辺の差を整理すると

$$2\frac{{s_1}^2{s_2}^2}{mn}+\frac{{s_2}^4}{n^2}-\frac{{s_2}^4(m-1)}{n^2(n-1)}=2\frac{{s_1}^2{s_2}^2}{mn}+\frac{{s_2}^4}{n^2}\left(1-\frac{m-1}{n-1}\right)\ge 2\frac{{s_1}^2{s_2}^2}{mn}\ge 0$$

であるから不等式は成り立つ．

$f\le m+n-2$ を示すには

$$\left(\frac{{s_1}^2}{m}+\frac{{s_2}^2}{n}\right)^2\le\frac{{s_1}^4(m+n-2)}{m^2(m-1)}+\frac{{s_2}^4(m+n-2)}{n^2(n-1)}$$

を示せばよい．右辺は

$$\frac{{s_1}^4}{m^2}+\frac{{s_2}^4}{n^2}+\frac{{s_1}^4(n-1)}{m^2(m-1)}+\frac{{s_2}^4(m-1)}{n^2(n-1)}$$

であるから，右辺と左辺の差は

$$\frac{{s_1}^4(n-1)}{m^2(m-1)}+\frac{{s_2}^4(m-1)}{n^2(n-1)}-2\frac{{s_1}^2}{m}\frac{{s_2}^2}{n}=\left(\frac{{s_1}^2}{m}\sqrt{\frac{n-1}{m-1}}-\frac{{s_2}^2}{n}\sqrt{\frac{m-1}{n-1}}\right)^2\ge 0$$

となり，不等式は成り立つ．

問 10.3　$T(x)=-(x-\mu)^2/2,\psi=1/\sigma^2$ とおくと，密度関数は

$$f(x)=\frac{1}{(2\pi\sigma^2)^{1/2}}\exp\!\left(-\frac{(x-\mu)^2}{\sigma^2}\right)=\exp\!\left(\psi T(x)-\frac{1}{2}\log(2\pi/\psi)\right)$$

より，$h(x)=1,\ c(\psi)=\dfrac{1}{2}\log(2\pi/\psi)$ である．

問 10.4　$\Gamma(n/2)2^{n/2}$ の部分は無視してよいので，これを無視して示す．ψ で微分すると

$$\frac{\partial}{\partial\psi}\psi^{n/2}w^{n/2-1}e^{-\psi w/2} = \frac{n}{2}\psi^{n/2-1}w^{n/2-1}e^{-\psi w/2} - \frac{w}{2}\psi^{n/2}w^{n/2-1}e^{-\psi w/2}$$

となる．一方 $w \times \psi^{n/2}w^{n/2-1}e^{-\psi w/2}$ を w で微分すると

$$\frac{\partial}{\partial w}(w \times \psi^{n/2}w^{n/2-1}e^{-\psi w/2}) = \frac{n}{2}\psi^{n/2}w^{n/2-1}e^{-\psi w/2} - \frac{\psi}{2}\psi^{n/2}w^{n/2}e^{-\psi w/2}$$

となる．両者の関係は

$$\frac{\partial}{\partial w}(w \times \psi^{n/2}w^{n/2-1}e^{-\psi w/2}) = \psi\frac{\partial}{\partial\psi}\psi^{n/2}w^{n/2-1}e^{-\psi w/2}$$

である．これより ψ での微分を w での微分に置き換えることにより，w で不定積分できるので，(10.41) 式の第 2 式が導かれる．

問 10.5　最初の等式については，分母の $\Gamma(n/2)2^{n/2}$ は相殺するので

$$\frac{f_n(w, n/w)}{f_n(w, 1)} = \frac{(n/w)^{n/2}w^{n/2-1}e^{-(n/w)w/2}}{w^{n/2-1}e^{-w/2}} = \left(\frac{n}{w}\right)^{n/2}e^{w/2-n/2}$$

と確認できる．また第 2 式に $wf_n(w, 1)$ をかけると

$$\left(\frac{n}{w}\right)^{n/2}e^{w/2-n/2} \times w\frac{1}{\Gamma(n/2)2^{n/2}}w^{n/2-1}e^{-w/2} = \frac{n^{n/2}e^{-n/2}}{\Gamma(n/2)2^{n/2}}$$

と確認できる．

次に正規分布の同時密度関数 f_n で考えると

$$f_n(x, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}}e^{-w/(2\sigma^2)} = \frac{\psi^{n/1}}{(2\pi)^{n/2}}e^{-\psi w/2}, \quad w = \sum_{i=1}^{n}(x_i - \mu)^2$$

の形である．このとき

$$\frac{f_n(x, w/n)}{f_n(x, 1)} = \left(\frac{n}{w}\right)^{n/2}e^{-n/2+w/2}$$

は (10.42) 式と同じである．

「一山型」となるのは (10.42) 式の逆数の関数である．ここでは (10.42) 式の第 2 式において，定数を無視したものの対数をとった関数 $g(w)$ が下に凸であることを示す．実際 $g(w)$ を微分すると

$$g(w) = \frac{w}{2} - \frac{n}{2}\log w, \quad g'(w) = \frac{1}{2} - \frac{n}{2w}, \quad g''(w) = \frac{n}{2w^2} > 0$$

となり $g(w)$ は下に凸である．これより $-g(w)$ が一山型となることからわかる．

問 10.6　$U=\sum_{i=1}^{m}(X_i-\bar{X})^2, V=\sum_{i=1}^{n}(y_i-\bar{y})^2$ とおく．帰無仮説 $(H_0:\sigma^2={\sigma_1}^2={\sigma_2}^2)$ のもとでの最尤推定量は $\tilde{\sigma}^2=(U+V)/(m+n)$ であり，対立仮説のもとでの最尤推定値はそれぞれ ${\hat{\sigma}_1}^2=U/m, {\hat{\sigma}_2}^2=V/n$ である．正規分布の同時密度にこれらを代入すると，指数部分は定数となり，基準化定数中の σ^2 に最尤推定量を代入し，比をとることとなる．このことから，やはり定数部分を除いて，尤度比は

$$\frac{(\tilde{\sigma}^2)^{(m+n)/2}}{({\hat{\sigma}_1}^2)^{m/2}({\hat{\sigma}_2}^2)^{n/2}}$$

となり，これは (10.64) 式と比例的である．

問 10.7　$F={s_2}^2/{s_1}^2$ とおく．(10.65) 式の分母分子を $\left(\sum_{i=1}^{m}(X_i-\bar{X})^2\right)^{(m+n-2)/2}$ で割ると，(10.65) 式は

$$\frac{\left(1+\frac{n-1}{m-1}F\right)^{(m+n-2)/2}}{\left(\frac{n-1}{m-1}F\right)^{(n-1)/2}}\propto\frac{\left(1+\frac{n-1}{m-1}F\right)^{(m+n-2)/2}}{F^{(n-1)/2}}$$

となる．一方で (4.24) 式の F 分布の密度関数が定数を除いて

$$f_{n-1,m-1}(F)\propto\frac{F^{(n-1)/2-1}}{\left(1+\frac{n-1}{m-1}F\right)^{(m+n-2)/2}}$$

であるから

$$Ff_{n-1,m-1}(F)\propto\frac{F^{(n-1)/2}}{\left(1+\frac{n-1}{m-1}F\right)^{(m+n-2)/2}} \qquad (*)$$

となり，定数を除いて (10.65) 式の逆数に一致している．

(10.65) 式の不等式の領域が区間 $(\underline{F},\overline{F})$ で与えられることは，$(*)$ 式の右辺が一山型であることを示せばよい．記法の簡便のため

$$y=\frac{n-1}{m-1}F,\quad b=\frac{n-1}{2},\quad c=\frac{m-1}{2}$$

とおき

$$g(y)=\log\frac{y^b}{(1+y)^{b+c}}=b\log y-(b+c)\log(1+y)$$

を微分すると

$$g'(y)=\frac{b}{y}-\frac{(b+c)}{1+y}=\frac{b(1+y)-(b+c)y}{y(1+y)}=\frac{b-cy}{y(1+y)}$$

となり，$g'(y)$ は $y<b/c$ で正，$y>b/c$ で負であるから，$g(y)$ は一山型である．

問 10.8　帰無仮説のもとでの $\mu = \mu_1 = \cdots = \mu_k$ の最尤推定量は $\widetilde{\mu} = \bar{\bar{X}}$ であり，対立仮説のもとでの最尤推定量は $\hat{\mu}_i = \bar{X}_i$, $i = 1, \ldots, k$ である．また分散の最尤推定量はそれぞれ $\widetilde{\sigma}^2 = W_{\mathrm{T}}/n$, $\hat{\sigma}^2 = W_{\mathrm{E}}/n$ である．これらを正規分布の密度関数に代入すると，指数部分は定数となり，基準化定数の σ^2 にそれぞれ $\widetilde{\sigma}^2, \hat{\sigma}^2$ を代入すればよい．これより尤度比は $(\widetilde{\sigma}^2/\hat{\sigma}^2)^{n/2}$ に比例し，従って $(W_{\mathrm{T}}/W_{\mathrm{E}})^{n/2}$ に比例する．

問 10.9　U の無条件の期待値は $\bar{\bar{x}}$ であり，無条件の分散は W_{T}/n である．また $V = i$ を与えたときの U の条件つき期待値は $\bar{x}_i$ であり，条件つき分散は $\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2/n_i$ である．そこで条件つき期待値の分散は

$$\sum_{i=1}^{k} \frac{n_i}{n}(\bar{x}_i - \bar{\bar{x}})^2 = \frac{1}{n} W_{\mathrm{H}}$$

であり，条件つき分散の期待値は

$$\sum_{i=1}^{k} \frac{n_i}{n} \frac{\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2}{n_i} = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2 = \frac{1}{n} W_{\mathrm{E}}$$

である．これより，1 元配置分散分析の平方和の分解と一致することがわかる．

問 10.10　十分性については，(T_1, T_2) を与えたときの $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ の条件つき分布は，$X_1, \ldots, X_m$ と $Y_1, \ldots, Y_n$ が独立となり，それぞれ θ_2 及び θ_1 には依存しないことからわかる．完備性については，T_1, T_2 が密度関数 $f_{\theta_1}(t_1), f_{\theta_2}(t_2)$ を持つ場合について示す．$g(t_1, t_2)$ の期待値を考えると

$$0 = \iint g(t_1, t_2) f_{\theta_1}(t_1) f_{\theta_2}(t_2)\, dt_1 dt_2, \quad \forall\,(\theta_1, \theta_2)$$

が考慮する条件である．$h(t_2, \theta_1) = \int g(t_1, t_2) f_{\theta_1}(t_1)\, dt_1$ とおくと

$$\int h(t_2, \theta_1) f_{\theta_2}(t_2)\, dt_2 = 0, \quad \forall\,(\theta_1, \theta_2)$$

であるが，T_2 の完備性より

$$h(t_2, \theta_1) = \int g(t_1, t_2) f_{\theta_1}(t_1)\, dt_1 = 0, \quad \forall\,(t_2, \theta_1)$$

となる．ここで T_1 の完備性より $g(t_1, t_2) = 0, \forall\,(t_1, t_2)$ となる．従って (T_1, T_2) は完備である．

問 10.11　$\hat{p}_2 - \hat{p}_1$ は不偏であり，$(T_1, T_2) = (X, Y)$ の関数であるので，前問より UMVU であることがわかる．

問 10.12　多項分布の確率関数は

$$\frac{n!}{\prod_{i,j} y_{ij}!} \prod_{i,j} {p_{ij}}^{y_{ij}}$$

であるが，ここに $p_{ij} = p_{i\bullet} p_{\bullet j}$ を代入すると

$$\frac{n!}{\prod_{i,j} y_{ij}!} \prod_{i,j} (p_{i\bullet} p_{\bullet j})^{y_{ij}} = \frac{n!}{\prod_{i,j} y_{ij}!} \prod_{i=1}^{r} {p_{i\bullet}}^{y_{i\bullet}} \times \prod {p_{\bullet j}}^{y_{\bullet j}}$$

となり，(多項係数の部分を除けば) 確率関数が 2 つの多項分布の確率関数の積になる．そこで多項分布の最尤推定量の形から (10.100) 式が導かれる．

第 11 章

問 11.1

$$f(y) = \exp\left(-\frac{1}{2\sigma^2} y^\top y + \frac{1}{\sigma^2} \beta^\top X^\top y - \frac{\beta^\top X^\top X \beta}{2\sigma^2} - (n/2) \log(2\pi\sigma^2) \right)$$

と書いて，$\psi_1 = -1/(2\sigma^2)$, $\psi_2 = \beta/\sigma^2$ とおくと指数型分布族の形に書け，$(y^\top y, X^\top y)$ が完備十分統計量であることがわかる．ψ_2 はベクトルであることに注意する．

問 11.2　$a^\top b = \sum_{i=1}^{q} a_i b_i$ であるから，これを a_i で偏微分すれば b_i である．これをベクトル表記すればよい．

次に $a^\top C a = \sum_{i=1}^{q} {a_i}^2 c_{ii} + 2 \sum_{i<j} a_i a_j c_{ij}$ であるから，これを a_i で偏微分すれば

$$2a_i c_{ii} + 2 \sum_{j \neq i} c_{ij} a_j = 2 \sum_{j=1}^{q} c_{ij} a_j$$

である．これをベクトル表記すればよい．

問 11.3　前問を用いると (11.15) 式の $Q(\beta)$ を β で微分すれば

$$-2X^\top y + 2X^\top X \beta$$

である．

問 11.4　X の列が一次従属であることと，あるゼロベクトルでない a が存在して $Xa=0$ となることは同値である．さらに $Xa=0$ となることと Xa の長さの二乗すなわち $a^\top X^\top Xa$ がゼロとなること $(a^\top X^\top Xa=0)$ は同値である．さらに非負定値行列 $X^\top X$ が正定値でないことと，あるゼロベクトルでない a が存在して $a^\top X^\top Xa=0$ となることは同値である．これらより

$$X^\top X \text{ が正定値でない} \iff X \text{ の列が一次従属}$$

が成り立つ．

問 11.5　集約尤度関数は

$$f(y)=\frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left(-\frac{e^\top e}{2\sigma^2}\right)$$

である．$\tau=\sigma^2$ とおき，対数をとると

$$-\frac{n}{2}\log(2\pi\tau)-\frac{e^\top e}{2\tau}$$

である．これを τ で微分して 0 とおくと

$$0=-\frac{n}{2}\frac{1}{\tau}+\frac{e^\top e}{2\tau^2}$$

となり，これを解くと $\tau=e^\top e/n$ を得る．

問 11.6　行列のトレースについて $\mathrm{tr}(AB)=\mathrm{tr}(BA)$ となることなどを用いると，等式を順次確認できる．

問 11.7　$X^\top X$ は n_i を対角要素とする対角行列であり，$X^\top Y$ は $\sum_{j=1}^{n_i} Y_{ij}$ を要素とするベクトルであることから (11.33) 式が確認できる．

問 11.8　(11.36) 式の制約については，(11.37) 式のように μ, α_i が表せることから $\mu_1,\dots,\mu_k$ から $\mu,\alpha_1,\dots,\alpha_k$ が定まる．逆方向はもちろん $\mu_i=\mu+\alpha_i$ で定まる．これより 1 対 1 対応がわかる．

(11.39) 式の制約のもとでは，$\mu_i=\mu+\alpha_i$ に n_i をかけて加えると

$$\sum_{i=1}^{k} n_i\mu_i=(n_1+\cdots+n_k)\mu+\sum_{i=1}^{k} n_i\alpha_i=n\mu$$

より

$$\mu = \frac{\sum_{i=1}^{k} n_i \mu_i}{n}$$

で表され，この μ を用いて $\alpha_i = \mu_i - \mu$ と表される．このことから $\mu_1, \ldots, \mu_k$ から $\mu, \alpha_1, \ldots, \alpha_k$ が定まる．逆方向は上と同様に $\mu_i = \mu + \alpha_i$ であり，1 対 1 対応がわかる．

問 11.9 (11.58) の第 1 式の μ については，(11.55) 式で μ_{ij} をすべての i, j について加えると (11.57) 式の制約により $\alpha_i, \beta_j, \gamma_{ij}$ の和はゼロであることからわかる．α_i については，この μ を用いて (11.55) 式を j について加えると $\sum_{j=1}^{b} \beta_j = 0$ 及び $\sum_{j=1}^{b} \gamma_{ij} = 0$ からわかる．β_j については i について加えることで同様にわかる．γ_{ij} は (11.55) 式が等式で成り立つように定まる．

問 11.10 (11.67) 式の右辺第 1 項は，$1_a \otimes 1_b = 1_{a \times b}$ であることから，μ がすべての y_{ij} に共通に現れることに対応している．$(I_a \otimes 1_b)\alpha$ は各 α_i が b 回繰り返されて $y_{ij}, j = 1, \ldots, b$ に現れることに対応している．$(1_a \otimes I_b)\beta$ も同様である．第 4 項は $I_a \otimes I_b = I_{ab}$ であることから，各 y_{ij} に γ_{ij} が現れることに対応している．

問 11.11 $J_a = 1_a 1_a^\top$ と書けることに注意すると $1_a^\top J_a = 1_a^\top 1_a 1_a^\top = a 1_a^\top$ であり

$$1_a^\top \frac{1}{a} J_a = 1_a^\top$$

が成り立つ．これより

$$1_a^\top \left(I_a - \frac{1}{a} J_a \right) = 1_a^\top - 1_a^\top = 0$$

である．これより (11.68) 式の α について $1_a^\top \alpha = 0$ すなわち $\sum_{i=1}^{a} \alpha_i = 0$ である．β についても同様である．

次に γ について考える．γ の要素を 1 列に並べたとき γ_{ij} の i に関する和 (あるいは j に関する和) は左から $1_a^\top \otimes I_b$ (あるいは $I_a \otimes 1_b^\top$) をかけることにあたる．行列のサイズが整合的であれば $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$ となることから

$$(1_a^\top \otimes I_b) \left(I_a - \frac{1}{a} J_a \right) \otimes \left(I_b - \frac{1}{b} J_b \right) = 0,$$

$$(I_a \otimes 1_b{}^\top)\left(I_a - \frac{1}{a}J_a\right) \otimes \left(I_b - \frac{1}{b}J_b\right) = 0$$

であり，これより $\sum_{i=1}^{a} \gamma_{ij} = 0, \forall j$ 及び $\sum_{j=1}^{b} \gamma_{ij} = 0, \forall i$ が成り立つ．

問 11.12　(11.68) 式の α 及び (11.69) 式の β, γ を (11.67) 式に代入して，クロネッカー積の部分に $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$ を適用すれば，(11.70) 式が得られる．

問 11.13　$G_1 = (g_1, \ldots, g_p)$ を M の正規直交基底を列ベクトルとする行列とし，$H_1 = (h_1, \ldots, h_p)$ を別の正規直交基底を列ベクトルとする行列とすると，$p \times p$ 正則行列 Q を用いて $G_1 = H_1 Q$ と書ける．このとき $I_p = G_1{}^\top G_1 = H_1{}^\top H_1$ であるが，$G_1 = H_1 Q$ を代入すれば $I_p = Q^\top H_1{}^\top H_1 Q = Q^\top Q$ である．これより Q は $p \times p$ 直交行列であり，$QQ^\top = I_p$ も成り立つ．したがって

$$G_1 G_1{}^\top = H_1 Q Q^\top H_1{}^\top = H_1 H_1{}^\top$$

となり (11.82) 式は正規直交基底の取り方に依存しないことがわかる．

問 11.14　本文にあるように $X = G_1 A$ と表す．このとき

$$\begin{aligned} X(X^\top X)^{-1}X^\top &= G_1 A(A^\top G_1{}^\top G_1 A)^{-1} A^\top G_1{}^\top = G_1 A(A^\top A)^{-1} A^\top G_1 \\ &= G_1 A A^{-1} (A^\top)^{-1} A^\top G_1{}^\top = G_1 G_1{}^\top \end{aligned}$$

である．

問 11.15　対称行列である B を対角化すればすぐわかるように $B^2 = B$ より B の固有値は 0 か 1 である．$r = \mathrm{rank} B$ として固有値 1 に属する正規直交なベクトルを $g_1, \ldots, g_r$ とすると B のスペクトル分解は

$$B = (G_1, G_2)\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}(G_1, G_2)^\top = G_1 G_1{}^\top,$$

$$G_1 = (g_1, \ldots, g_r), \quad (G_1, G_2) : \text{直交行列}$$

となる．これは G_1 の列ベクトルのはる空間 M への直交射影子である．あとは，$B = G_1 G_1{}^\top$ に右から任意のベクトルをかければわかるように，B の列ベクトルのはる空間と G_1 の列ベクトルのはる空間が等しいことに気が付けばよい．

問 11.16　M について考えると説明変数行列 X は (11.31) 式で与えられている．$X^\top X$ は n_i を対角要素とする対角行列であるから，その逆行列は $1/n_i$ を対角要素とする対角行列となる．また $X^\top y$ は各群での Y の要素の和からなるベクトルである．このことから

$$y^\top P_M y = y^\top X(X^\top X)^{-1}X^\top y = \sum_{i=1}^{k} \frac{1}{n_i}\left(\sum_{j=1}^{n_i} y_{ij}\right)^2 = \sum_{i=1}^{k} n_i \bar{y}_i^2$$

である．また $y^\top P_{M_0} y = n\bar{\bar{y}}^2$ はすぐにわかる．これより (11.101) 式の最初の等号がわかる．2 つ目の等号は右辺を展開して整理すればわかる．

問 11.17　$1_a{}^\top(I_a - J_a/a) = 1_a{}^\top - 1_a{}^\top = 0$，$(I_a - J_a/a)1_a = 0$，$(A \otimes B)^\top = (A^\top) \otimes (B^\top)$，$(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$ を用いると (11.102) 式の直交性が確かめられる．(11.103) がそれぞれの空間に対する直交射影行列であることは，対称性とべき等性 ($P_{M_a}{}^\top = P_{M_a}, P_{M_a}{}^2 = P_{M_a}$ など) を確かめ，問 11.15 を用いてこれらの行列の列空間を確認すればよい．

問 11.18　$0 \neq c_1 + \cdots + c_k$ とする．$\mu_i = \mu + \alpha_i$ の両辺に c_i をかけて加えれば

$$\sum_{i=1}^{k} c_i\mu_i = (c_1 + \cdots + c_k)\mu + \sum_{i=1}^{k} c_i\alpha_i = (c_1 + \cdots + c_k)\mu$$

より

$$\mu = \frac{\displaystyle\sum_{i=1}^{k} c_i\mu_i}{\displaystyle\sum_{i=1}^{k} c_i}$$

と定まり，この μ を用いて $\alpha_i = \mu_i - \mu$ と定まる．

逆に $0 = c_1 + \cdots + c_k$ のときは μ に c を加えて，各 α_i から c を引くと，$\mu_i = \mu + \alpha_i$ は不変であるし，$0 = c_1\alpha_1 + \cdots + c_k\alpha_k$ の制約も成り立ったままである．このため $\mu, \alpha_1, \ldots, \alpha_k$ は一意には定まらない．

問 11.19　問 11.4 の解答にあるように $\ker X$ と $\ker(X^\top X)$ は等しい．そこで (11.106) 式と (11.107) 式を比べれば，正規方程式を満たす任意の $\hat{\beta}$ について $X\hat{\beta} = \hat{\mu}$ となることがわかる．

問 11.20　まずは測り方を数え上げることを考える．場合分けのために，2 つとも載せて測る回数を A，1 つ目のボールだけを載せて測る回数を B，2 つ目のボールだけを載せて測る回数を C とする．A, B, C とも $0, 1, 2$ のいずれかの値をとる．ここで A の値で場合分けをおこなう．$A = 0$ のときは $B = 1, C = 2$ あるいは $B = 2, C = 1$ が考えられる．問題の対称性から $B = 1, C = 2$ の場合を考えればよい．つまり $B = 2, C = 1$ の場合は，$B = 1, C = 2$ の場合と平均二乗誤差が等しいからである．$A = 1$ のときは，やはり対称性から $B = 1, C = 1$ のときと $B = 2, C = 0$ の場合を考えればよい．また $A = 2$ のときは $B = 1, C = 0$ の場合を考えればよい．以上 4 通りの場合について平均二乗誤差を求めて比較すればよい．

3 回の計測 Y_1, Y_2, Y_3 の順序は，記法の簡便のために，2 つとも載せるのを最初の A 回おこない，次に 1 つ目のボールを載せるのを B 回おこない，最後に 2 つ目のボールを載せるのを C 回おこなうとする．以下でそれぞれの場合の推定量を示すが，いずれも不偏推定量であり，完備十分統計量の関数であるから UMVU である．

$A = 0, B = 1, C = 2$ のときは，$\hat{\mu}_1 = Y_1, \hat{\mu}_2 = (Y_2 + Y_3)/2$ である．平均二乗誤差は分散の和であるから $(1 + 1/2)\sigma^2 = (3/2)\sigma^2$ である．

$A = 1, B = 1, C = 1$ のときは $\hat{\mu}_1 = (Y_1 + Y_2 - Y_3)/2$ でありこの推定量の分散は $(3/4)\sigma^2$ である．対称性から $\hat{\mu}_2$ の分散も同様であり，分散の和である平均二乗誤差は $(3/2)\sigma^2$ である．

$A = 1, B = 2, C = 0$ のときは $\hat{\mu}_1 = (Y_2 + Y_3)/2$ でありその分散は $\sigma^2/2$ である．また $\hat{\mu}_2 = Y_1 - \hat{\mu}_1 = Y_1 - (Y_2 + Y_3)/2$ でありその分散は $(1 + 1/2)\sigma^2 = (3/2)\sigma^2$ である．これらの分散の和は $2\sigma^2$ である．

$A = 2, B = 1, C = 0$ のときは $\hat{\mu}_1 = Y_3$ でありその分散は σ^2，$\hat{\mu}_2 = (Y_1 + Y_2)/2 - Y_3$ であり，その分散は $(1/2 + 1)\sigma^2$ である．これらの分散の和は $(5/2)\sigma^2$ である．

以上より，最小の平均二乗誤差は $(3/2)\sigma^2$ であり，これらは $A = 0, B = 1, C = 2$ 及び $A = 1, B = 1, C = 1$ の場合に達成される．

問 11.21　推定値 θ が θ_0 に近いとして $f(x_i, \theta)$ を θ_0 の回りで展開すると，線形近似は

$$f(x_i, \theta) \doteq f(x_i, \theta_0) + (\theta - \theta_0)^\top \dot{f}(x_i, \theta_0)$$

である．そこで最小にする式は

$$\sum_{i=1}^{n}(Y_i - f(x_i,\theta_0) - (\theta-\theta_0)^\top \dot{f}(x_i,\theta_0))^2 = \sum_{i=1}^{n}(\varepsilon_i - (\theta-\theta_0)^\top \dot{f}(x_i,\theta_0))^2 \quad (*)$$

となる．ただし $\varepsilon_i = Y_i - f(x_i,\theta_0)$, $i=1,\ldots,n$ は互いに独立に $\mathrm{N}(0,\sigma^2)$ に従う．$(*)$ 式を最小にする $\hat{\theta}$ は

$$\hat{\theta}-\theta_0 = \left(\sum_{i=1}^{n}\dot{f}(x_i,\theta_0)\dot{f}(x_i,\theta_0)^\top\right)^{-1}\sum_{i=1}^{n}\dot{f}(x_i,\theta_0)\varepsilon_i$$

となり，右辺は $\mathrm{N}\left(0,\sigma^2\left(\sum_{i=1}^{n}\dot{f}(x_i,\theta_0)\dot{f}(x_i,\theta_0)^\top\right)^{-1}\right)$ に従う．

第 12 章

問 12.1　$P(X\le x)) = E[I_{[X\le x]}]$ に期待値の繰り返しの公式を用いると

$$P(X\le x) = E^Y[E[I_{[X\le x]}|Y]] = E^Y[P(X\le x|Y)]$$

であるが $P(X\le x|Y) = F(x)$ が Y に依存しないならば $E^Y[F(x)] = F(x)$ であるから，$F(x)$ が X の周辺分布関数であることがわかる．また独立性に関する (3.6) 式や (3.14) 式について述べたように，$P(X\le x|Y)=F(x)$ が Y に依存しないことと，X と Y が互いに独立であることは同値である．

問 12.2　前問と (12.8) 式のあとの議論により $\varepsilon_1,\ldots,\varepsilon_n$ の周辺分布の結果及び $R_1,\ldots,R_n$ との独立性は明らかである．また，$X_1,\ldots,X_n$ の *i.i.d.* 性から，$(1,\ldots,n)$ の任意の並べかえ $(\pi_1,\ldots,\pi_n)$ について $(|X_{\pi_1}|,\ldots,|X_{\pi_n}|)$ の分布は $(|X_1|,\ldots,|X_n|)$ の分布と等しいから，$(R_1,\ldots,R_n)$ は任意の並べかえを同様に確からしくとることがわかる．

問 12.3　$(R_1,\ldots,R_{m+n})$ が $(1,\ldots,m+n)$ の任意の並べかえを等確率でとることは，前問と同様に確認できる．この場合，最初の n 個の順位の集合 $\{R_1,\ldots,R_n\}$ は $\{1,\ldots,m+n\}$ のサイズ n の部分集合を等しい確率でとることも明らかである．

問 12.4　順位 a が k 人のタイとなっていると，中間順位は

$$\frac{a+\cdots+(a+k-1)}{k} = a + \frac{0+1+\cdots+(k-1)}{k} = a+\frac{k-1}{2}$$
$$= \frac{1}{2}+(a-1)+\frac{k}{2}$$

である．(12.17) 式の右辺において $a-1$ は和の第 1 項目に対応し，$k/2$ は

$$\sum_{j=1}^{n} \frac{1}{2} I_{[x_j=x_i]}$$

に対応している．

また k 人の中間順位の和は，タイがない場合の対応する順位の和 $a+\cdots+(a+k-1)$ であるから，中間順位を用いれば

$$\sum_{i=1}^{n} R_i = 1+2+\cdots+n = \frac{n(n+1)}{2}$$

が成り立つ．

問 12.5　$\sum_{i=1}^{m+n} R_i = (m+n)(m+n+1)/2$ とし，$R_1,\ldots,R_{m+n}$ からなる有限母集団を考える．この母集団の平均と分散は

$$\mu = \frac{m+n+1}{2}, \quad \sigma^2 = \frac{1}{m+n} \sum_{i=1}^{m+n} (R_i - \bar{R})^2$$

である．W は標本平均の n 倍であるから，有限母集団からの非復元無作為抽出に関する結果より

$$\begin{aligned} E[W\,|\,R_1,\ldots,R_{m+n}] &= \frac{n(m+n+1)}{2}, \\ \mathrm{Var}[W\,|\,R_1,\ldots,R_{m+n}] &= n^2 \frac{1}{n} \frac{m+n-n}{m+n-1} \frac{1}{m+n} \sum_{i=1}^{m+n} (R_i - \bar{R})^2 \\ &= \frac{mn}{(m+n)(m+n-1)} \sum_{i=1}^{m+n} (R_i - \bar{R})^2 \end{aligned}$$

である．もしタイがなければ

$$\begin{aligned} &\sum_{i=1}^{m+n} (R_i - \bar{R})^2 \\ &\quad = \sum_{i=1}^{m+n} {R_i}^2 - (m+n)\bar{R}^2 \\ &\quad = \frac{1}{6}(m+n)(m+n+1)(2m+2n+1) - (m+n)\frac{(m+n+1)^2}{4} \\ &\quad = \frac{(m+n)(m+n+1)}{12}\left(2(2m+2n+1) - 3(m+n+1)\right) \\ &\quad = \frac{(m+n)(m+n+1)}{12}(m+n-1) \end{aligned}$$

より

$$\mathrm{Var}[W \,|\, R_1, \ldots, R_{m+n}] = \frac{mn(m+n+1)}{12}$$

を得る．

問 12.6　$X_1, \ldots, X_m$ の累積分布関数を $F(x)$ とし，$Y_1, \ldots, Y_n$ の累積分布関数を $F(y-\Delta)$ とする．(12.12) 式の帰無仮説では $\Delta = 0$ としているが，信頼区間を考えるときには $\Delta = \Delta_0$ とし，かつ Δ_0 を動かして仮説検定の族を考えなければならない．そして

$$Z_i = Y_i - \Delta_0,\ i = 1, \ldots, n, \quad Z_{i+n} = X_i,\ i = 1, \ldots, m$$

とおき，$R_i, i = 1, \ldots, m+n$ を $Z_i, i = 1, \ldots, m+n$ の順位として，検定統計量は $W = \sum_{j=1}^{n} R_j$ である．以下では議論の簡単のためタイのない場合を考える．$Y_1 < \cdots < Y_n$ とおいて一般性を失わない．また Δ_0 はどの $U_{ij} = Y_j - X_i$ とも一致しない値として考える．このとき R_j は $Y_1, \ldots, Y_{j-1}$ の $(j-1)$ 個と，X_i, $i = 1, \ldots, m$ の中で $Y_j - \Delta_0$ より小さいものの個数の和であるから

$$R_j = j - 1 + \sum_{i=1}^{m} I_{[Y_j - \Delta_0 > X_i]}$$

である．ここで

$$I_{[Y_j - \Delta_0 > X_i]} = I_{[Y_j - X_i > \Delta_0]} = I_{[U_{ij} > \Delta_0]}$$

に注意すると

$$R_j = j - 1 + \sum_{i=1}^{m} I_{[U_{ij} > \Delta_0]}$$

であり，

$$\begin{aligned} W = \sum_{j=1}^{n} R_j &= \frac{n(n-1)}{2} + \sum_{j=1}^{n} \sum_{i=1}^{m} I_{[U_{ij} > \Delta_0]} \\ &= \frac{n(n-1)}{2} + (U_{ij} > \Delta_0 \text{ となる } U_{ij} \text{ の個数}) \end{aligned}$$

と書ける．つまり $n(n-1)/2$ を除いて W は $U_{ij} > \Delta_0$ となる U_{ij} の個数に一致する．これは (12.28) 式と同じ形をしているから，やはり符号検定の議論と同様の議論となり，(12.29) 式が成り立つ．

問 12.7　記法の簡単のため $a = mn/(m+n)$, $b = (n+m-2)/(m+n-1)$ とおく．そして $U = Ts/\sqrt{a}$ を (12.37) 式に代入すると

$$\frac{|T|s}{\sqrt{a}} > z_{\alpha/2}\sqrt{\frac{b}{a}s^2 + \frac{1}{m+n-1}\frac{T^2s^2}{a}}$$

である．両辺で $s/\sqrt{a}$ がキャンセルするから不等式は

$$|T| > z_{\alpha/2}\sqrt{b + \frac{T^2}{m+n-1}}$$

となる．両辺正であるから 2 乗して b を元に戻すと

$$T^2 > {z_{\alpha/2}}^2\left(\frac{n+m-2}{m+n-1} + \frac{T^2}{m+n-1}\right)$$

であり，両辺に $(m+n-1)$ をかけて右辺第 2 項を移項すれば

$$T^2\left((m+n-1) - {z_{\alpha/2}}^2\right) > {z_{\alpha/2}}^2(n+m-2)$$

となる．正の平方根をとれば

$$|T| > z_{\alpha/2}\sqrt{\frac{n+m-2}{m+n-1-{z_{\alpha/2}}^2}}$$

を得る．

問 12.8　簡単のためにタイのない場合に示す．以下の解答では問 12.6 で $u_{ij} = y_j - x_i$ としていたものを $u_{ij} = x_i - y_j$ と符号を変えて $u_{ij} > 0$ となる個数を数えるものとする．ペアの総数 mn は固定されているから，符号を変えて数えても同値な統計量が得られる．

F_m は 0 から始まって各 x_i で $1/m$ 増加して 1 に達する．スケールを m 倍すれば 0 から m まで増加する．同様に G_n を n 倍すれば 0 から n まで増加する．$(F_m(c), G_n(c))$ の図を x 軸を m 倍，y 軸を n 倍すれば $(0,0)$ から (m,n) に達する折れ線が得られる．折れ線は，本文にあるように $x_1,\dots,x_m$, $y_1,\dots,y_n$ をあわせて $z_1,\dots,z_{m+n}$ とし，これを小さい順に見ていったときに，x であれば右に一歩，y であれば上に一歩進む形で得られる．図 1 は $m=5, n=4$ で

$$x_1 < y_1 < y_2 < x_2 < y_3 < x_3 < x_4 < y_4 < x_5$$

となる例である．この図で例えば y_2 とラベルしている縦棒の右横に 4 個の四角のセルがあるが，これは

$$y_2 < x_2, x_3, x_4, x_5$$

つまり，y_2 より大きい x が 4 個あることを示している．他の y_j についても同様である．このように太線で表した経験 ROC 曲線の右下のセルは $u_{ij} = x_i - y_j > 0$ となる (x_i, y_j) の組にそれぞれ対応している．これより経験 ROC 曲線の下の面積は $u_{ij} > 0$ となる組の数に対応しており，問 12.6 と同様にその数は 2 標本のウィルコクソン順位和検定統計量の値に対応している．

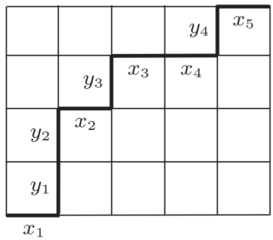

図 1　経験 ROC 曲線の例

第 13 章

問 13.1　$I(\theta_0, \theta) = \eta(\theta_0, \theta_0) - \eta(\theta_0, \theta)$ より $I(\theta_0, \theta_0) = 0$ である．また (13.11) 式より θ の関数として $I(\theta_0, \theta)$ は $\theta = \theta_0$ で極大となるから

$$\frac{\partial}{\partial\theta} I(\theta_0, \theta)\Big|_{\theta=\theta_0} = 0$$

である．さらに $I(\theta_0, \theta)$ を θ で 2 回微分すれば

$$\begin{aligned}\frac{\partial^2}{\partial\theta^2} I(\theta_0, \theta) &= -\frac{\partial^2}{\partial\theta^2}\eta(\theta_0, \theta) \\ &= -\int_{-\infty}^{\infty} \frac{\partial^2}{\partial\theta^2} \log(f(x,\theta)) f(x,\theta_0)\,dx\end{aligned}$$

であるから，これを $\theta = \theta_0$ で評価すれば

$$\frac{\partial^2}{\partial\theta^2} I(\theta_0, \theta)\Big|_{\theta=\theta_0} = -\int_{-\infty}^{\infty} \left(\frac{\partial^2}{\partial\theta^2} \log f(x,\theta)\right)\Big|_{\theta=\theta_0} f(x,\theta_0)\,dx = I(\theta_0)$$

を得る．

問 13.2　(13.22) 式及び (13.23) 式より

$$P_0\left(\frac{1}{n}\ell(0) - (\eta(0,0) - c_g) \geq \frac{1}{n}\sup_{|\theta|>M} \ell(\theta)\right) \to 1 \quad (n \to \infty)$$

となる．ここで $P()$ の中の事象が成り立てば，$|\theta| > M$ の領域では ℓ は最大とはならなく，したがって $|\hat{\theta}_n^{\mathrm{ML}}| \le M$ であるから

$$P_0(|\hat{\theta}_n^{\mathrm{ML}}| \le M) \to 1 \quad (n \to \infty)$$

であり，(13.24) 式が成り立つ．

問 13.3　(13.17) 式の仮定と (13.25) 式より, (13.27) 式の $\eta(0,\theta_i)$ について $\eta(0,\theta_i) \le \eta(0,0) - \delta$ であることに注意する．そこで (13.29) 式より 1 に収束する確率で

$$\sup_{\theta \in [\theta_j, \theta_{j+1}]} \frac{1}{n}\ell(\theta) < \eta(0,\theta_j) + \frac{\delta}{2} \le \eta(0,0) - \frac{\delta}{2}$$

となる．これより 1 に近づく確率で区間 $[\theta_j, \theta_{j+1}]$ には最尤推定値が落ちないことがわかる．したがって (13.30) 式が成り立つ．

問 13.4　$1 - P(A_n \cap B_n) = P((A_n \cap B_n)^c) = P({A_n}^c \cup {B_n}^c) \le P({A_n}^c) + P({B_n}^c) \to 0$ $(n \to \infty)$ より $P(A_n \cap B_n) \to 1$ がわかる．集合が k 個のときは k について帰納法を用いればよい．補集合で考えれば，$\lim_{n\to\infty} P(A_{i,n}) = 0,\ i = 1,\ldots,k$ のとき $\lim_{n\to\infty} P(A_{1,n} \cup \cdots \cup A_{k,n}) = 0$ である．このことを以下で用いる．

また前問より $2K$ 個の各区間について，最尤推定値がその区間に落ちる確率は 0 に収束するから，最尤推定値がこれらの $2K$ 個の和集合に落ちる確率も 0 に収束する．すなわち

$$P_0(\varepsilon \le |\hat{\theta}_n^{\mathrm{ML}}| \le M) \to 0 \quad (n \to \infty)$$

である．このことと (13.24) 式より (13.31) 式が成り立つ．

問 13.5　順次正則条件を確認する．以下，極限と積分の交換や微積の交換の正当化が必要であり，そのためには有界収束定理のような測度論的な道具が必要であるが，ここではその点の厳密な確認は省略している．

まず (13.17) 式の単調性を確認する．

$$\eta(0,\theta) = \int_{-\infty}^{\infty} \log\left(\frac{1}{\pi(1+(x-\theta)^2)}\right) \frac{1}{\pi(1+x^2)}\,dx$$

であるが，定数等を省略して $q(\theta) = \pi(\eta(0,\theta) + \log\pi)$ とおくと

$$q(\theta) = -\int_{-\infty}^{\infty} \log(1+(x-\theta)^2)\,\frac{1}{1+x^2}\,dx$$

であり，$q(\theta)$ の単調性を示せばよい．問題の対称性から q は偶関数 $(q(\theta) = q(-\theta))$ であり，正の θ のみを考えて，$q'(\theta) < 0$ を示せばよい．微分すると

$$q'(\theta) = 2\int_{-\infty}^{\infty} \frac{x-\theta}{1+(x-\theta)^2}\frac{1}{1+x^2}\,dx$$

であるが，ここで x を $x+\theta/2$ と変数変換すれば (ヤコビアンは 1 であり)，

$$q'(\theta) = 2\int_{-\infty}^{\infty} \frac{x-\theta/2}{(1+(x-\theta/2)^2)\times(1+(x+\theta/2)^2)}\,dx$$

となるが，分母は x の偶関数である．ここで分子の x は奇関数であるから，その積分は 0 となるため

$$q'(\theta) = -\theta\int_{-\infty}^{\infty} \frac{1}{(1+(x-\theta/2)^2)\times(1+(x+\theta/2)^2)}\,dx < 0$$

となり，単調性が示された．

さて, 次に進む前に $\theta \to \infty$ のときに $\eta(0,\theta) \to -\infty$ を示す．常に $\pi(1+(x-\theta)^2) > 1$ であるから，常に

$$\log\left(\frac{1}{\pi(1+(x-\theta)^2)}\right) < 0$$

が成り立つ．そこで，例えば積分範囲を $[-1,1]$ に限っても

$$\eta(0,\theta) < \int_{-1}^{1} \log\left(\frac{1}{\pi(1+(x-\theta)^2)}\right)\frac{1}{\pi(1+x^2)}\,dx, \quad x \in [-1,1]$$

である．そこで右辺が $-\infty$ に発散することを示せばよい．ここで $\theta > 1$ を考えれば $x \in [-1,1]$ に対して $x = 1$ で log 内が最大となるから

$$\log\left(\frac{1}{\pi(1+(x-\theta)^2)}\right) \le \log\left(\frac{1}{\pi(1+(1-\theta)^2)}\right), \quad x \in [-1,1]$$

であり，これより

$$\eta(0,\theta) < \log\left(\frac{1}{\pi(1+(1-\theta)^2)}\right)P(|X| \le 1)$$

が成り立つ．ただし X はコーシー分布に従う確率変数である．$\theta \to \infty$ で右辺は $-\infty$ に発散するから，$\eta(0,\theta) \to -\infty$ $(\theta \to \infty)$ が示された．

次に (13.19) 式を確認する．実は，すぐ上で示したことと同様に示せばよい．$M \to \infty$ のとき $E_0[g_M(X)] \to -\infty$ を示せば (18.19) 式は成り立つ．$g_M(x)$ についても，すべての x について $g_M(x) < 0$ であるから，今度は積分範囲を $[0,1]$ とすると

$$E_0[g_M(X)] = \int_{-\infty}^{\infty} g_M(x)\frac{1}{\pi(1+x^2)}\,dx < \int_0^1 g_M(x)\frac{1}{\pi(1+x^2)}\,dx$$

であるが，$M>1$ ならば，$x\in[0,1]$ に対して $g_M(x)$ は $\theta=M$ で最大化されるから

$$g_M(x) = \sup_{|\theta|\ge M}\log\left(\frac{1}{\pi(1+(x-\theta)^2)}\right) \le \log\left(\frac{1}{\pi(1+(x-M)^2)}\right),\ x\in[0,1]$$

である．さらにこれを $x\in[0,1]$ について最大化すれば，$x=1$ で最大化されるから，

$$g_M(x) \le \log\left(\frac{1}{\pi(1+(1-M)^2)}\right), \quad x\in[0,1]$$

である．これより

$$\int_0^1 g_M(x)\frac{1}{\pi(1+x^2)}\,dx \le \log\left(\frac{1}{\pi(1+(1-M)^2)}\right)\times P(0\le X\le 1)$$

となり，右辺は $M\to\infty$ で $-\infty$ に発散する．これより $E_0[g_M(X)]\to-\infty$ が示された．

最後に (13.20) 式を示す．h_M は M について単調増加であるから $M=\infty$ とおいて

$$h_\infty(x) = \sup_\theta\left|\frac{\partial}{\partial\theta}\log f(x,\theta)\right|$$

を考えればよいが，コーシー分布の場合はさらに x について最大化して

$$h_\infty(x) \le \sup_x\sup_\theta\left|\frac{\partial}{\partial\theta}\log f(x,\theta)\right| = C < \infty$$

を示せばよい．ここで $y=x-\theta$ とおけば

$$\sup_x\sup_\theta\left|\frac{\partial}{\partial\theta}\log f(x,\theta)\right| = 2\sup_y\frac{|y|}{1+y^2}$$

であるが $q(y)=y/(1+y^2)$ として $q(y)$ の増減を調べると，$0=\lim_{y\to\pm\infty}q(y)$ であり，また

$$q'(y) = \frac{1-y^2}{(1+y^2)^2}$$

より $q(y)$ は $y=-1$ で最小値 $-1/2$，$y=1$ で最大値 $1/2$ をとる．従って $C=1$ であり (13.20) 式も成り立つ．

問 13.6　X が p 次元標準正規分布に従うとき，B を $p\times p$ 正則行列として $Y=BX$ は p 次元正規分布 $\mathrm{N}_p(0,\varSigma), \varSigma=BB^\top$ に従う．また $\varSigma$ が正定値行列ならば $\varSigma$ は必ずこの形に書ける．ここで

$$Y^\top \Sigma^{-1} Y = X^\top B^\top (BB^\top)^{-1} BX = X^\top B^\top (B^\top)^{-1} B^{-1} BX = X^\top X$$

であるが，$X^\top X$ はカイ二乗分布の定義より自由度 p のカイ二乗分布に従う．

問 13.7　それぞれのブロック分割行列の積を計算していけばわかる．なお，逆行列であることを確かめるためには，行列の掛け算を実行して単位行列が得られることを示せばよい．

問 13.8　Helmert 変換を用いることにより $\mu = 0$ の場合で，かつ分散共分散行列を原点回りのもの ($s_{xx} = \dfrac{1}{n} \sum_{i=1}^{n} {X_i}^2$ 等) として示せばよい．$\sqrt{n}$ と $\sqrt{n-1}$ の違いは，これらの比が 1 に収束するため漸近分布には影響しない．

s_{xx}, s_{yy}, s_{xy} は確率変数の 2 乗を平均したものになっており，多次元の中心極限定理が応用できる．極限の 3 次元正規分布の分散共分散行列は $({X_1}^2, X_1 Y_1, {Y_1}^2)$ の分散共分散行列となる．この 3×3 の分散共分散行列の要素は以下である．

$$E\left[({X_1}^2)^2\right] - {\sigma_{xx}}^2,\ E\left[({Y_1}^2)^2\right] - {\sigma_{yy}}^2,\ E[(X_1 Y_1)^2] - {\sigma_{xy}}^2,$$

$$E\left[{X_1}^2 \times {Y_1}^2\right] - \sigma_{xx}\sigma_{yy},\ E\left[{X_1}^2 \times X_1 Y_1\right] - \sigma_{xx}\sigma_{xy},\ E\left[{Y_1}^2 \times X_1 Y_1\right] - \sigma_{yy}\sigma_{xy}$$

ここで，それぞれの 4 次のモーメントは問 3.21 より求められる．すなわち問 3.21 において i, j, k, l を，X_1 に対応するものは 1，Y_2 に対応するものは 2 とおき，さらに σ の下付き文字を 1 は x 及び 2 は y に書き換えればよい．これにより

$$E\left[{X_1}^4\right] = 3{\sigma_{xx}}^2,\ E\left[{Y_1}^4\right] = 3{\sigma_{yy}}^2,\ E\left[{X_1}^2{Y_1}^2\right] = \sigma_{xx}\sigma_{yy} + 2{\sigma_{xy}}^2,$$

$$E\left[{X_1}^2{Y_1}^2\right] = \sigma_{xx}\sigma_{yy} + 2{\sigma_{xy}}^2,\ E\left[{X_1}^3 Y_1\right] = 3\sigma_{xx}\sigma_{xy},\ E\left[X_1 {Y_1}^3\right] = 3\sigma_{xy}\sigma_{yy}$$

となる．これより $\sqrt{n}(s_{xx} - \sigma_{xx}, s_{yy} - \sigma_{yy}, s_{xy} - \sigma_{xy})$ の漸近分布は平均 0 の 3 次元正規分布で，その分散共分散行列が以下で与えられる．

$$\begin{pmatrix} 2{\sigma_{xx}}^2 & 2{\sigma_{xy}}^2 & 2\sigma_{xx}\sigma_{xy} \\ 2{\sigma_{xy}}^2 & 2{\sigma_{yy}}^2 & 2\sigma_{xy}\sigma_{yy} \\ 2\sigma_{xx}\sigma_{xy} & 2\sigma_{xy}\sigma_{yy} & \sigma_{xx}\sigma_{yy} + {\sigma_{xy}}^2 \end{pmatrix}$$

である．

次の問題のために $\sigma_{xx} = \sigma_{yy} = 1$, $\sigma_{xy} = \rho$ の場合に特殊化すれば分散共分散行列は

$$\Sigma = \begin{pmatrix} 2 & 2\rho^2 & 2\rho \\ 2\rho^2 & 2 & 2\rho \\ 2\rho & 2\rho & 1+\rho^2 \end{pmatrix} \qquad (*)$$

である.

問 13.9　標本相関係数 r の分布を考えるときには $\sigma_{xx} = \sigma_{yy} = 1$, $\sigma_{xy} = \rho$ としてよい．$f(u, v, w) = w/\sqrt{uv}$ とおくと，f の勾配 f' の要素は

$$\frac{\partial}{\partial u} f(u, v, w) = -\frac{1}{2u}\frac{w}{\sqrt{uv}}, \quad \frac{\partial}{\partial v} f(u, v, w) = -\frac{1}{2v}\frac{w}{\sqrt{uv}},$$

$$\frac{\partial}{\partial w} f(u, v, w) = \frac{1}{\sqrt{uv}}$$

であり，これを $(u, v, w) = (1, 1, \rho)$ で評価したものは $(-\rho/2, -\rho/2, 1)$ となる．これより，デルタ法を用いると，$\sqrt{n}(r - \rho)$ の漸近分布は期待値 0 の 1 次元正規分布で，その分散は前問の $(*)$ 式の Σ を用いて

$$\begin{aligned}
&(-\rho/2, -\rho/2, 1)\, \Sigma \begin{pmatrix} -\rho/2 \\ -\rho/2 \\ 1 \end{pmatrix} \\
&= \frac{\rho^2}{4}(2+2) + (1+\rho^2) + 2\frac{\rho^2}{4}(2\rho^2) - (2+2)\frac{\rho}{2}(2\rho) \\
&= \rho^2 + (1+\rho^2) + \rho^4 - 4\rho^2 = 1 - 2\rho^2 + \rho^4 \\
&= (1-\rho^2)^2
\end{aligned}$$

となる.

問 13.10　$f(\rho) = \dfrac{1}{2}(\log(1+\rho) - \log(1-\rho))$ とおけば

$$f'(\rho) = \frac{1}{2}\left(\frac{1}{1+\rho} + \frac{1}{1-\rho}\right) = \frac{1}{1-\rho^2}$$

である．デルタ法により $(f'(\rho))^2$ を前問の漸近分散にかけると 1 を得る．そこで $\sqrt{n}(z - \zeta)$ は標準正規分布に分布収束する.

問 13.11　付録 A.2 の連続写像定理により $\log Z_n$ の極限分布について考えればよい．$\log Z_n = \log X_1 + \cdots + \log X_n$ より $\log Z_n$ は $i.i.d.$ 確率変数の和であるから，中心極限定理を応用できる.

$$\mu = E[\log X_1] = p\log b + (1-p)\log a,$$
$$\sigma^2 = \mathrm{Var}[\log X_1] = p(\log b)^2 + (1-p)(\log a)^2 - \mu^2$$

である．これより $W_n = (\log Z_n - n\mu)/(\sqrt{n}\sigma)$ とおくと W_n が標準正規分布に分布収束する．

$$\log Z_n = \sqrt{n}\sigma W_n + n\mu, \quad Z_n = \exp(\sqrt{n}\sigma W_n + n\mu),$$
$$(Z_n e^{-n\mu})^{1/\sqrt{n}} = \exp(\sigma W_n)$$

より ${Z_n}^{1/\sqrt{n}} e^{-\sqrt{n}\mu}$ が対数正規分布に分布収束する．

問 13.12　$p \times p$ の Σ のランクが r であるとき，ある $p \times p$ 正則行列 B を用いて

$$\Sigma = BDB^\top, \quad D = \mathrm{diag}(\underbrace{1,\dots,1}_{r\text{ 個}},0,\dots,0) \tag{$**$}$$

と表される．ただし D の対角要素で 1 の個数は r である．$Z \sim \mathrm{N}_p(0,D)$ とし，$X = BZ$ とおけば $X \sim \mathrm{N}_p(0,\Sigma)$ となる．Z は最初の r 個の要素が互いに独立に標準正規分布に従い，残りの $p-r$ 個はすべて 0 であることに注意する．

ここで $\Sigma\Sigma^+\Sigma = \Sigma$ に $(**)$ 式を代入すると $BDB^\top\Sigma^+BDB^\top = BDB^\top$ であるが，最初の B 及び最後の $B^\top$ をキャンセルすれば

$$D\widetilde{\Sigma}^+D = D, \quad \widetilde{\Sigma}^+ = B^\top\Sigma^+B$$

を得る．

ここで

$$X^\top\Sigma^+X = (BZ)^\top\Sigma^+(BZ) = Z^\top\widetilde{\Sigma}^+Z$$

と書けることに注意する．さらに Z の最後の $p-r$ の要素は 0 であるから，Z_1 を Z の最初の r 個からなる部分ベクトルとし，$\widetilde{\Sigma}^+_{11}$ を $\widetilde{\Sigma}^+$ の左上の $r \times r$ ブロックとすれば

$$Z^\top\widetilde{\Sigma}^+Z = {Z_1}^\top\widetilde{\Sigma}^+_{11}Z_1$$

である．ここで $D\widetilde{\Sigma}^+D = D$ で左上の $r \times r$ ブロックを考慮すると，I_r を $r \times r$ の単位行列として

$$I_r\widetilde{\Sigma}^+_{11}I_r = I_r$$

であるから $\Sigma_{11}^{+} = I_r$ でなければならない．従って $X^\top \Sigma^+ X = Z_1^\top Z_1$ であり，これは自由度 r のカイ二乗分布に従う．

問 13.13　$Y = (Y_1, \ldots, Y_k)$ が多項分布 $\mathrm{Mn}(n, p_1, \ldots, p_k)$ に従うとき

$$\frac{1}{\sqrt{n}}(Y - n(p_1, \ldots, p_k)) = \frac{1}{\sqrt{n}}(Y - \hat{Y})$$

は近似的に平均 0，分散共分散行列が以下の $\Sigma = (\sigma_{ij})$ の多変量正規分布に従う：

$$\sigma_{ii} = p_i(1 - p_i), \quad \sigma_{ij} = -p_i p_j \ (i \neq j)$$

$p = (p_1, \ldots, p_k)^\top$ とおくと $\Sigma = \mathrm{diag}(p_1, \ldots, p_k) - pp^\top$ と書ける．これより Σ のランクが $k-1$ であることもわかる．ここで $A = \mathrm{diag}(1/p_1, \ldots, 1/p_k)$ とおくと

$$\Sigma A \Sigma = (A^{-1} - pp^\top)A(A^{-1} - pp^\top) = A^{-1} - 2pp^\top + pp^\top A pp^\top$$

であるが $p^\top A = (1, \ldots, 1)$ より $p^\top A p = 1$ となり

$$\Sigma A \Sigma = A^{-1} - pp^\top = \Sigma$$

であるから $A = \Sigma^+$ は一般化逆行列であることがわかる．これと前問より

$$\frac{1}{n}(Y - \hat{Y})^\top A (Y - \hat{Y}) = \sum_{i=1}^{k} \frac{(Y_i - \hat{Y}_i)^2}{\hat{Y}_i}$$

は近似的に自由度 $k-1$ のカイ二乗分布に従う．

第 14 章

問 14.1　議論の簡単のため各 c について $\{x \mid f(x) = c\}$ のルベーグ測度が 0 の場合に示す．$A = \{x \mid f(x) > c\}$ とおき，$P(B) = P(A)$ となる B を考える．$c > 0$ とする．以下 $l(C)$ で集合のルベーグ測度 (長さ) を表す．A と B が実質的に異なる場合を考えるために $l(A \cap B^c) > 0, l(B \cap A^c) > 0$ の場合を考える．$P(A) = P(B)$ より

$$0 = P(A) - P(B) = \int_A f(x)\,dx - \int_B f(x)\,dx = \int_{A \cap B^c} f(x)\,dx - \int_{B \cap A^c} f(x)\,dx$$

である．$A \cap B^c$ 上で $f(x) > c$ であるから $\displaystyle\int_{A \cap B^c} f(x)\,dx > c \times l(A \cap B^c)$ である．また最初の仮定より $B \cap A^c$ 上で $f(x) < c$ としてよいから $\displaystyle\int_{B \cap A^c} f(x)\,dx < c \times l(B \cap A^c)$ である．これより

$$l(A) = l(A\cap B) + l(A\cap B^c) < l(A\cap B) + l(B\cap A^c) = l(B)$$

となり，A の長さのほうが小さいことがわかる．

問 14.2　(14.13) 式の事前分布のもとで $p\in A$ と $X=x$ の同時確率は

$$\pi_0 I_{[p_0\in A]} \times \binom{n}{x} {p_0}^x (1-p_0)^{n-x} + \pi_1 \int_A \eta(p) \binom{n}{x} p^x (1-p)^{n-x}\, dp$$

であり，これは p の事後分布と比例的である．$A=[0,1]$ とすれば，第 1 項が帰無仮説に対応し，第 2 項が対立仮説に対応する．これより，帰無仮説を棄却するのは

$$\pi_0 \binom{n}{x} {p_0}^x (1-p_0)^{n-x} < \pi_1 \int_0^1 \eta(p) \binom{n}{x} p^x (1-p)^{n-x}\, dp$$

のときであり，両辺を $\pi_1 {p_0}^x (1-p_0)^{n-x}$ で割れば (14.14) 式を得る．

同様に一般の場合で議論すれば (14.16) 式を得る．

問 14.3　(14.20) 式の前の事前分布の記述により，$\mathrm{Ga}(\nu,\alpha)$ の密度と $\mathrm{N}(\eta, 1/(\psi\tau))$ の密度をかけることにより，(μ,τ) の同時事前密度は

$$\frac{1}{\Gamma(\nu)\alpha^\nu}\tau^{\nu-1}\exp\Big(-\frac{\tau}{\alpha}\Big)\frac{(\psi\tau)^{1/2}}{(2\pi)^{1/2}}\exp\Big(-\frac{\psi\tau}{2}(\mu-\eta)^2\Big)$$

である．これに $x_1,\dots,x_n$ の密度関数

$$\frac{\tau^{n/2}}{(2\pi)^{n/2}}\exp\Big(-\frac{\tau}{2}n(\bar{x}-\mu)^2 - \frac{\tau}{2}\sum_{i=1}^n (x_i-\bar{x})^2\Big)$$

をかけて，(μ,τ) の関数として眺める．事後分布を確認するのに不要な定数を無視すると，同時事前密度と密度関数の積の主要な部分は以下の形である．

$$\tau^{\nu-1+\frac{1}{2}+\frac{n}{2}}\exp\Big(-\frac{\tau}{\alpha} - \frac{\psi\tau}{2}(\mu-\eta)^2 - \frac{\tau}{2}n(\bar{x}-\mu)^2 - \frac{\tau}{2}\sum_{i=1}^n (x_i-\bar{x})^2\Big)$$

ここで exp の中で μ を含む部分を平方完成すると

$$-\frac{\tau}{2}(\psi+n)\left(\mu - \frac{\psi\eta+n\bar{x}}{\psi+n}\right)^2 + \frac{\tau}{2}\frac{(\psi\eta+n\bar{x})^2}{\psi+n}$$

となり，ここから (14.20) の第 2 式が確認される．なおここで正規分布の密度関数から $\tau^{1/2}$ は吸収される．残りは τ の周辺事後密度に対応するが，その部分を書き出すと

$$\tau^{\nu-1+\frac{n}{2}}\exp\Big(-\frac{\tau}{\alpha} - \frac{\psi\tau}{2}\eta^2 - \frac{\tau}{2}n\bar{x}^2 + \frac{\tau}{2}\frac{(\psi\eta+n\bar{x})^2}{\psi+n} - \frac{\tau}{2}\sum_{i=1}^n (x_i-\bar{x})^2\Big)$$

である．exp の中で $-\tau$ の係数を見ると

$$\begin{aligned}
&\frac{1}{\alpha}+\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2+\frac{1}{2(\psi+n)}\left(\psi\eta^2(\psi+n)+n\bar{x}^2(\psi+n)-(\psi\eta+n\bar{x})^2\right)\\
&\quad=\frac{1}{\alpha}+\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2+\frac{1}{2(\psi+n)}(\psi\eta^2 n+n\bar{x}^2\psi-2\psi\eta n\bar{x})\\
&\quad=\frac{1}{\alpha}+\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2+\frac{1}{2(\psi+n)}n\psi(\eta-\bar{x})^2
\end{aligned}$$

である．これから (14.20) の第 1 式が確認される．

(14.20) の第 2 式より τ, x を与えたもとで $\mu-(n\bar{x}+\psi\eta)/(n+\psi)$ は $\mathrm{N}(0,1/(\tau(n+\psi)))$ に従っている．ここで τ について積分すれば (x を条件とした) μ の周辺事後分布は「$\mu-(n\bar{x}+\psi\eta)/(n+\psi)$ の定数倍が自由度 $2\nu+n$ の t 分布に従う」という形であることがわかる．

問 14.4　ポアソン分布の確率関数

$$p(x)=\frac{\lambda^x}{x!}e^{-\lambda}$$

を λ の関数と見るとガンマ分布の密度関数の形をしている．そこで λ の事前分布として $\mathrm{Ga}(\nu,\alpha)$ を考えてその密度関数をかけると，一部の定数を省略して，積は

$$\lambda^{\nu-1}\exp\Bigl(-\frac{\lambda}{\alpha}\Bigr)\lambda^x\exp(-\lambda)=\lambda^{x+\nu-1}\exp\Bigl(-\lambda\Bigl(\frac{1}{\alpha}+1\Bigr)\Bigr)$$

となるから，事後分布は $\mathrm{Ga}(\nu+x,(1+1/\alpha)^{-1})$ である．サンプルサイズが n のときは確率関数が $\lambda^{\sum_{i=1}^n x_i}e^{-\lambda n}$ となることから，事後分布は $\mathrm{Ga}(\nu+\sum_{i=1}^{n}x_i,(n+1/\alpha)^{-1})$ である．これは x_1 から x_n まで順番に事後分布を更新しても同じ結果が得られる．

負の 2 項分布の確率関数

$$p(x)=\binom{r+x-1}{x}(1-p)^x p^r$$

を p の関数としてみるとベータ分布の密度関数の形をしている．そこで p の事前分布として $\mathrm{Be}(a,b)$ を考えると，一部の定数を省略して確率関数と密度関数の積は

$$(1-p)^x p^r p^{a-1}(1-p)^{b-1}=p^{r+a-1}(1-p)^{b+x-1}$$

であるから，事後分布は $\mathrm{Be}(a+r,b+x)$ である．サンプルサイズが n のときの事後分布は $\mathrm{Be}(a+nr,b+\sum_{i=1}^{n}x_i)$ である．

ガンマ分布の密度関数は

$$f(x)=\frac{1}{\Gamma(\nu)\alpha^\nu}x^{\nu-1}\exp\Bigl(-\frac{x}{\alpha}\Bigr)$$

である．ここでは尺度母数の逆数 $\beta = 1/\alpha$ に対して事前分布を考えることとする．密度関数は

$$f(x) = \frac{\beta^\nu}{\Gamma(\nu)} x^{\nu-1} \exp(-x\beta)$$

となりガンマ分布の形をしている．つまり α に戻ると「逆ガンマ分布」をしている．そこで β にガンマ分布 $\mathrm{Ga}(\kappa, 1/\gamma)$ を仮定すれば，事前密度と $f(x)$ の積は，主要な項のみを残すと

$$\beta^{\kappa-1} \exp(-\gamma\beta) \beta^\nu \exp(-x\beta)$$

の形をしている．これより事後分布は $\mathrm{Ga}(\kappa+\nu, 1/(x+\gamma)$ である．サンプルサイズが n のときの事後分布は $\mathrm{Ga}(\kappa + n\nu, 1/(\sum_{i=1}^n x_i + \gamma))$ である．

問 14.5　事後密度関数の分子は (14.21) 式に $f(x\,|\,\theta)$ をかけて

$$\pi(\theta) f(x\,|\,\theta) = \sum_{j=1}^k \alpha_j \pi(\theta, \xi_j) f(x\,|\,\theta) \tag{$*$}$$

である．次に j 番目の事前分布を単独で用いたときの事後密度 $\pi(\theta, \widetilde{\xi}_j(x))$ は

$$\pi(\theta, \widetilde{\xi}_j(x)) = \frac{\pi(\theta, \xi_j) f(x\,|\,\theta)}{\displaystyle\int \pi(\theta', \xi_j) f(x\,|\,\theta')\, d\theta'}$$

であるから，分母を払うと

$$\pi(\theta, \xi_j) f(x\,|\,\theta) = \int \pi(\theta', \xi_j) f(x\,|\,\theta')\, d\theta' \times \pi(\theta, \widetilde{\xi}_j(x))$$

となり，これを $(*)$ 式に代入すると

$$\pi(\theta) f(x\,|\,\theta) = \sum_{j=1}^k \alpha_j \int \pi(\theta', \xi_j) f(x\,|\,\theta')\, d\theta' \times \pi(\theta, \widetilde{\xi}_j(x))$$

である．x の周辺密度関数は $1 = \displaystyle\int \pi(\theta, \widetilde{\xi}_j(x))\, d\theta$ より

$$\int \pi(\theta) f(x\,|\,\theta)\, d\theta = \sum_{j=1}^k \alpha_j \int \pi(\theta', \xi_j) f(x\,|\,\theta')\, d\theta'$$

である．したがって

$$\pi(\theta\,|\,x) = \sum_{j=1}^k \frac{\alpha_j \displaystyle\int \pi(\theta', \xi_j) f(x\,|\,\theta')\, d\theta'}{\displaystyle\sum_{j'=1}^k \alpha_{j'} \int \pi(\theta', \xi_{j'}) f(x\,|\,\theta')\, d\theta'} \pi(\theta, \widetilde{\xi}_j(x))$$

$$= \sum_{j=1}^{k} \widetilde{\alpha}_j \pi(\theta, \widetilde{\xi}_j(x))$$

となり (14.22) 式が確認される．

問 14.6　(14.30) 式で示したようにベイズ決定関数は事後の平均損失を最小にすることによって得られる．帰無仮説に対応する Θ_0 の事後確率を $\pi(\Theta_0\,|\,x)$，対立仮説に対応する Θ_1 の事後確率を $\pi(\Theta_1\,|\,x)$ と表すと，決定関数 $\delta(x) \in \{0,1\}$ の事後平均損失は

$$I_{[\delta(x)=1]}\pi(\Theta_0\,|\,x) + I_{[\delta(x)=0]}\pi(\Theta_1\,|\,x)$$

である．これを最小にするには $\pi(\Theta_0\,|\,x) < \pi(\Theta_1\,|\,x)$ のときには $\delta(x) = 1$, $\pi(\Theta_0\,|\,x) > \pi(\Theta_1\,|\,x)$ のときには $\delta(x) = 0$ とすればよい．つまり事後確率のより大きな仮説を選べばよい．

問 14.7　$(\log\tau)' = 1/\tau$ で，$\tau \to 0$ のとき $\log\tau \to -\infty$，また $\tau \to \infty$ のとき $\log\tau \to \infty$ より

$$\int_0^\infty \frac{1}{\tau}\,d\tau = \Big[\log\tau\Big]_0^\infty = \infty$$

である．

また $\tau^2 = \zeta$ とおくと

$$\frac{d\zeta}{d\tau} = 2\tau \quad \text{あるいは} \quad 2\tau d\tau = d\zeta$$

であり，$2\tau d\tau = d\zeta$ の両辺を $\tau^2 = \zeta$ で割ると

$$2\frac{1}{\tau}\,d\tau = \frac{1}{\zeta}\,d\zeta$$

となり，$\zeta = \tau^2$ で表しても同じ事前密度の形となる．

問 14.8　補題 14.4 あるいは補題 14.5 を用いる．$\pi(\{-1\}) = \pi(\{1\}) = 1/2$ とする．$X = x$ を観測したときの事後確率は

$$\pi(\{-1\}\,|\,x) = \frac{e^{-(x+1)^2/2}}{e^{-(x+1)^2/2} + e^{-(x-1)^2/2}} = \frac{e^{-x}}{e^{-x} + e^{x}},$$

$$\pi(\{1\}\,|\,x) = \frac{e^{x}}{e^{-x} + e^{x}}$$

である．ベイズ推定量は事後期待値であるから

$$\delta_\pi = \pi(\{1\}\,|\,x) - \pi(\{-1\}\,|\,x) = \frac{e^x - e^{-x}}{e^{-x} + e^x}$$

となる．この設定では，$x \to -x, \theta \to -\theta$ のように符号を変えることに対する対称性があるから，$R(1, \delta_\pi) = R(-1, \delta_\pi)$ となり，(14.36) 式は $(1, \delta_\pi) = R(-1, \delta_\pi) = r(\pi, \delta_\pi)$ と等式で成り立つ．これより π は最も不利な分布であり，δ_π はミニマックスである．

問 14.9　X が $\mathrm{Bin}(n, p)$ に従うとして

$$\left(\frac{X+\alpha}{n+\alpha+\beta} - p\right)^2 = \frac{(X+\alpha-p(n+\alpha+\beta))^2}{(n+\alpha+\beta)^2}$$

の期待値を計算したい．全体に $(n+\alpha+\beta)^2$ をかけて分子のみを確認することとする．

$$X+\alpha-p(n+\alpha+\beta) = (X-pn) + ((1-p)\alpha - p\beta)$$

と書いて 2 乗して期待値をとると

$$\begin{aligned} E[(X+\alpha-p(n+\alpha+\beta))^2] &= np(1-p) + ((1-p)\alpha - p\beta)^2 \\ &= np(1-p) + (\alpha - p(\alpha+\beta))^2 \end{aligned}$$

であり，展開して p のべきをそろえて整理すれば

$$p^2(-n + (\alpha+\beta)^2) + p(n - 2\alpha(\alpha+\beta)) + \alpha^2$$

となり，(14.39) 式が確認される．

問 14.10　δ を任意の決定関数とする．また任意の $\varepsilon > 0$ に対して事前分布 π_ε を選び，その事前分布に対するベイズ決定関数を δ_{π_ε} とする．このとき

$$\sup_\theta R(\theta, \delta) \ge r(\pi_\varepsilon, \delta) \ge r(\pi_\varepsilon, \delta_{\pi_\varepsilon}) > c - \varepsilon = \sup_\theta R(\theta, \delta^*) - \varepsilon$$

である．ここで ε を 0 に近づければ

$$\sup_\theta R(\theta, \delta) \ge \sup_\theta R(\theta, \delta^*)$$

となるから，δ^* はミニマックス決定関数である．

問 14.11　δ_ψ の平均二乗誤差が

$$E\left[\left(\frac{n\bar{X}}{n+\psi} - \mu\right)^2\right] = \frac{1}{(n+\psi)^2} E\left[\left(n\bar{X} - \mu(n+\psi)\right)^2\right]$$

$$= \frac{1}{(n+\psi)^2} E\left[\left(n(\bar{X}-\mu)-\mu\psi\right)^2\right]$$
$$= \frac{1}{(n+\psi)^2}(n+\mu^2\psi^2)$$

である．これを μ の事前分布 $\mathrm{N}(0, 1/\psi)$ で期待値をとると μ^2 が $1/\psi$ で置き換わるから

$$r(\pi_\psi, \delta_\psi) = \frac{1}{(n+\psi)^2}(n+(1/\psi)\psi^2)) = \frac{1}{n+\psi}$$

を得る．

ここで ψ を 0 に近づければベイズリスクが $1/n$ に近づくため，補題 14.6 より $\bar{X}$ はミニマックス推定量である．

問 14.12　(14.3) 式で p を含む部分のみに着目すると

$$p^{\alpha+x-1}(1-p)^{\beta+n-x-1}$$

となり (14.5) 式の基準化定数がないものが得られている．p についての最大化は 2 項分布の最尤推定量と同様であるから，MAP 推定値は

$$\frac{\alpha+x-1}{\alpha+\beta+n-2}$$

で与えられる．

問 14.13　回帰モデルを $y \sim \mathrm{N}_n(X\beta, (1/\tau)I_n)$ とし，τ 及び β の事前分布を，τ の周辺事前分布が $\tau \sim \mathrm{Ga}(\nu, \alpha)$，また τ を与えたときの β の条件つき事前分布が $\beta|\tau \sim \mathrm{N}_p(\eta, (1/\tau)I_p)$ とする．事前密度と回帰モデルの密度の積で，β, τ に関係するところのみを残すと以下のようになる．

$$\tau^{n/2}\exp\left(-\frac{\tau}{2}(y-X\beta)^\top(y-X\beta)\right) \times \tau^{p/2}\exp\left(-\frac{\tau}{2}(\beta-\eta)^\top(\beta-\eta)\right)$$
$$\times\, \tau^{\nu-1}\exp(-\tau/\alpha)$$

まずは，事後分布において $\beta|\tau$ の分布を求めるために $\exp(\,)$ の中で β について平方完成をおこなうと，

$$-\frac{\tau}{2}\left(\beta^\top(I_p+X^\top X)\beta - 2\beta^\top X^\top y - 2\beta^\top\eta + y^\top y + \eta^\top\eta\right)$$
$$= -\frac{1}{2}\Big[(\beta-(I_p+X^\top X)^{-1}(X^\top y+\eta))^\top(\tau(I_p+X^\top X))$$

$$
\times (\beta - (I_p + X^\top X)^{-1}(X^\top y + \eta))
$$
$$
+ \tau \left(y^\top y + \eta^\top \eta - (X^\top y + \eta)^\top (I_p + X^\top X)^{-1}(X^\top y + \eta) \right) \Big]
$$

となるから，2 次形式の形から，$\beta | \tau$ の事後分布は

$$
\mathrm{N}_p((I_p + X^\top X)^{-1}(X^\top y + \eta), (\tau(I_p + X^\top X))^{-1})
$$

であることがわかる．なお $\tau^{p/2}$ はこの正規分布の密度に吸収されることがわかる．

次に残りの項で τ に関する部分を整理すると，べきの部分が $\tau^{n/2+\nu-1}$ であり，$\exp(\,)$ の中は

$$
-\frac{\tau}{2} \left(\frac{2}{\alpha} + y^\top y + \eta^\top \eta - (X^\top y + \eta)^\top (I_p + X^\top X)^{-1}(X^\top y + \eta) \right)
$$

であるから，τ の事後分布が

$$
\mathrm{Ga}\left(n/2 + \nu,\ 2 \left(\frac{2}{\alpha} + y^\top y + \eta^\top \eta - (X^\top y + \eta)^\top (I_p + X^\top X)^{-1}(X^\top y + \eta) \right)^{-1} \right)
$$

であることがわかる．