統計検定１級 「数理統計」
過去問とその解答例

茨城大学 工学部 情報工学科

新納浩幸

2012 年 数理統計 問 1

以下の各問に答えよ。

[1] 連続型確率変数 Z の累積分布関数 $F(z) = Pr(Z \leq z)$ が狭義単調増加であるとき、$U = F(Z)$ は区間 $(0, 1)$ 上の一様分布に従うことを示せ。

[2] U_1, U_2 および U_3 を互いに独立に区間 $(0, 1)$ 上の一様分布にしたがう確率変数とし、X_1 をそれらのうち最も小さいもの、X_2 を 2 番目に小さいもの、そして X_3 を最も大きいものとする。このとき、$j = 1, 2, 3$ に対し、X_j の確率密度関数 $g_j(x)$ を求め、それらのグラフを描け。

[3] $j = 1, 2, 3$ に対し、上問 [2] の確率変数 X_j の期待値 $E[X_j]$ を求めよ。

2012 年 数理統計 問 1 ポイント

- 一様分布からの n 個の値を生成し、それらをソートした場合の第 i 番目の確率変数の分布の求め方

2012 年 数理統計 問 1 解答例

[1]

$$\forall a, b \in (0, 1) \quad a \leq b \qquad \exists z_0, z_1 \quad a = F(z_0), b = F(z_1)$$

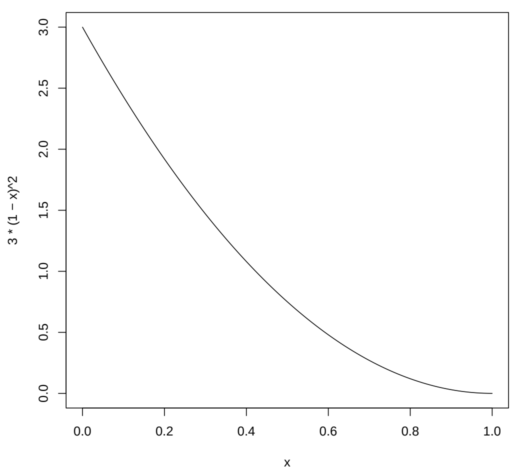
$$P(a < U < b) = P(z_0 < Z < z_1) = F(z_1) - F(z_0) = b - a$$

以上より U は $(0, 1)$ 上の一様分布であることが示された。

[2]

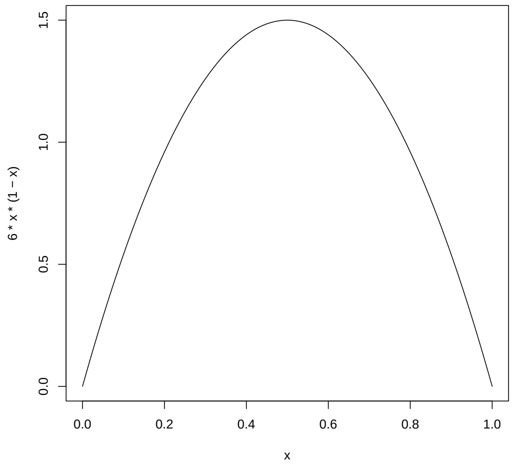
U_1, U_2, U_3 の最小値が x となる確率は、U_i の中のいずれかが x の値を取り（その確率は dx）、それ以外の U_i が x 以上の値を取れば（その確率は $1 - x$）よいので $g_1(x)$ は以下となる。

$$g_1(x) = 3(1 - x)^2$$



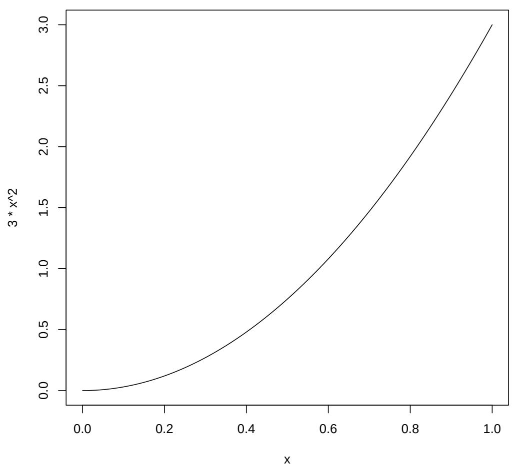
U_1, U_2, U_3 で 2 番目に小さい値が x となる確率は、U_i の中のいずれかが x の値を取り（その確率は dx)、それ以外の U_i の一方が x 以下の値を取り（その確率は x)、もう一方が x 以上の値を取れば（その確率は $1-x$）よいので $g_2(x)$ は以下となる。

$$g_2(x) = 6x(1-x)$$



U_1, U_2, U_3 で最大値が x となる確率は、U_i の中のいずれかが x の値を取り（その確率は dx)、それ以外の U_i が x 以下の値を取れば（その確率は x）よいので $g_3(x)$ は以下となる。

$$g_3(x) = 3x^2$$



[3]

$$E(X_1) = \int_0^1 3x(1-x)^2 dx = 3\left[\frac{x^4}{4} - \frac{x^3}{3} + \frac{x^2}{2}\right]_0^1 = \frac{1}{4}$$

$$E(X_2) = \int_0^1 6x^2(1-x) dx = 6\left[-\frac{x^4}{4} + \frac{x^3}{3}\right]_0^1 = \frac{1}{2}$$

$$E(X_3) = \int_0^1 3x^3 dx = 3\left[\frac{x^4}{4}\right]_0^1 = \frac{3}{4}$$

2012 年 数理統計 問 2

自由度 m のカイ２乗分布の確率密度関数は

$$f(x) = \begin{cases} \frac{1}{2^{m/2}\Gamma(m/2)} x^{(m/2)-1} e^{-x/2} & (x > 0), \\ 0 & (x \leq 0) \end{cases}$$

で与えられる。ただし $\Gamma(\cdot)$ はガンマ関数である。以下の各問に答えよ。

[1]　X が自由度 m のカイ２乗分布にしたがうとき、X のモーメント母関数 $M_X(t)$ は

$$M_X(t) = (1-2t)^{-m/2}$$

で与えられることを示せ。

[2]　X と Y が互いに独立にそれぞれ自由度 m および自由度 n のカイ２乗分布にしたがうとき、$Z = X + Y$ は自由度 $m + n$ のカイ２乗分布にしたがうことを示せ。ただし、確率分布とモーメント母関数が一対一に対応することは証明なしに用いてもよい。

[3]　X と Y が互いに独立にそれぞれ自由度 m および自由度 n のカイ２乗分布にしたがうとき、$W = X/(X+Y)$ の確率密度関数 $f_W(w)$ を求めよ。

[4]　$m, n > 2$ のとき、上問 [3] で求めた確率密度関数 $f_W(w)$ の最大値を与える w を求めよ。

2012 年 数理統計 問 2 ポイント

- モーメント母関数の定義
- モーメント母関数が同じなら確率分布は同じ（再生性を証明する際に利用）
- ２つの変数変換から同時確率密度関数を導く（独立性を示すことにも利用可能）

2012 年 数理統計 問 2 解答例

[1]

$$\int f(x)dx = \frac{1}{2^{m/2}\Gamma(m/2)} \int x^{(m/2)-1} e^{-x/2} dx = 1$$

より

$$\int x^{(m/2)-1} e^{-x/2} dx = 2^{m/2}\Gamma(m/2)$$

なので $m/2 = n$ として、

$$\int x^{n-1} e^{-x/2} dx = 2^n \Gamma(n)$$

$$M_X(t) = E(e^{tX}) = \int e^{tx} \frac{1}{2^{m/2}\Gamma(m/2)} x^{(m/2)-1} e^{-x/2} dx = \frac{1}{2^{m/2}\Gamma(m/2)} \int x^{(m/2)-1} e^{-\frac{(1-2t)x}{2}} dx$$

$y = (1-2t)x$ とおくと $dy = (1-2t)dx$ より

$$M_X(t) = \frac{1}{2^{m/2}\Gamma(m/2)} \int (\frac{y}{1-2t})^{(m/2)-1} e^{-y/2} (1-2t)^{-1} dy$$

$$= \frac{1}{2^{m/2}\Gamma(m/2)} \cdot \frac{1}{(1-2t)^{m/2}} 2^{m/2}\Gamma(m/2) = (1-2t)^{-m/2}$$

[2]

$Z = X + Y$ のモーメント母関数を調べる。X と Y が独立であるので、

$$M_Z(t) = E(e^{tZ}) = E(e^{t(X+Y)}) = E(e^{tX}e^{tY}) = E(e^{tX})E(e^{tY})$$

[1] より $M_Z(t) = (1-2t)^{-m/2} \cdot (1-2t)^{-n/2} = (1-2t)^{-(m+n)/2}$ これは [1] より自由度 $m+n$ のカイ 2 乗分布のモーメント母関数と一致する。よって Z は自由度 $m+n$ のカイ 2 乗分布に従う。

[3]

X と Y は独立なので、その同時確率密度関数 $f_{XY}(x,y)$ は以下となる。

$$f_{XY}(x,y) = \frac{1}{2^{\frac{m+n}{z}}\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} x^{m/2-1}y^{n/2-1}e^{-\frac{x+y}{2}}$$

Z と W の同時確率密度関数 $f_{ZW}(z,w)$ を考える。$Z = X+Y$、$W = X/(X+Y)$ より、$X = ZW$、$Y = (1-W)Z$ なのでヤコビアン $J = z$ となる。よって、

$$f_{ZW}(z,w) = zf_{XY}(zw,(1-w)z) = \frac{1}{2^{\frac{m+n}{2}-1}\Gamma(\frac{m+n}{2})} z^{\frac{m+n}{2}-1}e^{-z/2} \cdot \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} w^{m/2-1}(1-w)^{n/2-1}$$

これは Z が自由度 $m+n$ のカイ 2 乗分布に従い、W が自由度 $(\frac{m}{2}, \frac{n}{2})$ のベータ分布に従うことを示している。

以上より、

$$f_W(w) = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} w^{m/2-1}(1-w)^{n/2-1}$$

[4]

$g(w) = w^{m/2-1}(1-w)^{n/2-1}$ とおくと、$f_W(w)$ を最大にする $w = \hat{w}$ は $g(w)$ を最大にする $w = \hat{w}$ である。

$$g'(w) = w^{m/2-2}(1-w)^{n/2-2}\left((\frac{m}{2}-1)(1-w) - (\frac{n}{2}-1)w\right)$$

$g'(\hat{w}) = 0$ より

$$\hat{w} = \frac{m-2}{m+n-4}$$

2012 年 数理統計 問 3

パラメータ λ の指数分布の確率密度関数は

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & (x \geq 0), \\ 0 & (x < 0) \end{cases}$$

である。この分布にしたがう互いに独立な n 個の確率変数を $X_1, \cdots, X_n$ とし、それらの和を $T = X_1 + \cdots + X_n$、標本平均を $\bar{X} = T/n$ とするとき、以下の各問に答えよ。

[1] 標本平均 $\bar{X}$ の期待値と分散を求めよ。

[2] この分布からの互いに独立な観測値を $x_1, \cdots, x_n$ とし、それらの和を $t = x_1 + \cdots + x_n$ とするとき、$x_1, \cdots, x_n$ に基づくパラメータ λ の最尤推定量 $\hat{\lambda}$ を求めよ。

[3] n 個の独立な観測値に基づくパラメータ λ のフィッシャー情報量 $i_n(\lambda)$ を求めよ。

[4] デルタ法を用いて最尤推定量 $\hat{\lambda}$ の漸近分散を求めよ。

2012 年 数理統計 問 3 ポイント

- フィッシャー情報量の定義
- デルタ法の定義

2012 年 数理統計 問 3 解答例

[1]

母集団 X の平均 $E(X) = 1/\lambda$、分散 $V(X) = 1/\lambda^2$ である。

$$E(\bar{X}) = E(X) = \frac{1}{\lambda}$$

$$V(\bar{X}) = V(X)/n = \frac{1}{n\lambda^2}$$

[2]

対数尤度関数 $l = \sum_{i=1}^{n}(\log\lambda - \lambda x_i)$

$l' = \sum_{i=1}^{n}(\frac{1}{\lambda} - x_i)$ よって、$\frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$

$$\hat{\lambda} = \frac{n}{\sum x_i} = \bar{x}^{-1}$$

[3]

[2] より $l' = \sum_{i=1}^{n}(\frac{1}{\lambda} - x_i)$

$$l'' = \sum_{i=1}^{n} -\frac{1}{\lambda^2} = -\frac{n}{\lambda^2}$$

$$i_n(\lambda) = -E(l'') = \frac{n}{\lambda^2}$$

[4]

$g(x) = 1/x$ とおくと、$V(\hat{\lambda}) = V(g(\bar(X))$ $g'(x) = -1/x^2$ とデルタ法から、

$$V(g(\bar(X)) = (g'(E(\bar{X})))^2 V(\bar{X}) = \lambda^4 \frac{1}{n\lambda^2} = \frac{\lambda^2}{n}$$

以上より、$\hat{\lambda}$ の漸近分散は $\frac{\lambda^2}{n}$

2012 年 数理統計 問 4

平均 μ 分散 1.0 の正規母集団からの大きさ n の無作為標本から求めた標本平均を $\bar{x}$ とする。標準正規分布の上側 100% 点を $z(\alpha)$ とするとき、以下の各問に答えよ。

[1]　母平均 μ の信頼係数を $100(1-\alpha)\%$ の信頼区間は $(\bar{x} - z(\alpha/2)/\sqrt{n}, \bar{x} + z(\alpha/2)/\sqrt{n})$ で与えられることを示せ。

[2]　帰無仮説 $H_0 : \mu = 0$ に対して対立仮説 $H_1 : \mu \neq 0$ を検定するとき、有意水準 $100\alpha\%$ の一様最強力検定の棄却域を定めよ。

[3]　$\mu = 0$ が上問 [1] で求めた信頼区間に入っていることと、上問 [2] で与えた帰無仮説 H_0 を棄却しないことが同値であることを示せ。

2012 年 数理統計 問 4 ポイント

- ネイマンピアソンの定理から尤度比検定が最強力検定となるということ
- 尤度比検定からの検定統計量の設定方法

2012 年 数理統計 問 4 解答例

[1]
$\bar{X} \sim N(\mu, 1/n)$ なので

$$Z = \frac{\bar{X} - \mu}{\sqrt{1/n}} \sim N(0,1)$$

$P(|Z| < z(\alpha/2)) = \alpha$ より、$P(-z(\alpha/2) < Z < z(\alpha/2)) = 1 - \alpha$
つまり $Z = \frac{\bar{X}-\mu}{\sqrt{1/n}}$ が区間 $(-z(\alpha/2), z(\alpha/2))$ に入る確率が $1-\alpha$ なので、μ が
区間 $(\bar{X} - \sqrt{1/n}z(\alpha/2), \bar{X} + \sqrt{1/n}z(\alpha/2))$ に入る確率が $1-\alpha$ となる。
よって μ の信頼係数を $100(1-\alpha)\%$ とした信頼区間は $(\bar{x} - z(\alpha/2)/\sqrt{n}, \bar{x} + z(\alpha/2)/\sqrt{n})$ となる。
[2]
ネイマンピアソンの定理から尤度比検定が最強力検定となる。
今、尤度は

$$L = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x_i-\mu)^2}{2}\right)$$

H_0 の下での最大尤度は $L_0 = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_i^2}{2}\right)$、$H_1$ の下での最大尤度は $L_1 = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x_i-\bar{x})^2}{2}\right)$
よって尤度比 L は以下となる。

$$\begin{aligned} L = \frac{L_1}{L_0} &= \prod_{i=1}^{n} \exp(x_i^2 - (x_i - \bar{x})^2) = \exp(\sum_{i=1}^{n}(2x_i\bar{x} - \bar{x}^2)) \\ &= \exp(2\bar{x}\sum_{i=1}^{n} x_i - n\bar{x}^2) = \exp(n\bar{x}^2) \end{aligned}$$

これは $\bar{X}$ のみの関数となっているので、$\bar{X}$ に基づく検定が一様最強力検定であることがわかる。H_0 の下で $\bar{X} \sim N(0, 1/n)$ なので、$Z = \frac{\bar{X}}{\sqrt{1/n}} \sim N(0, 1)$ 以上より有意水準 $100\alpha\%$ の一様最強力検定の棄却域 R は

$$R = \{z||z| > z(\alpha/2)\} = \{\bar{x}||\bar{x}| > \sqrt{1/n}z(\alpha/2)\}$$

[3]

$\mu = 0$ が [1] の信頼区間に入っていれば $\bar{x} - z(\alpha/2)/\sqrt{n} < 0 < \bar{x} + z(\alpha/2)/\sqrt{n}$ これは $|\bar{x}| < \sqrt{1/n}z(\alpha/2)$ 同値であり、$|\bar{x}| < \sqrt{1/n}z(\alpha/2)$ は $\bar{x}$ が [2] の棄却域に入っていないことを意味する。

2012 年 数理統計 問 5

(X, Y, Z) が平均ベクトル $(0, 0, 0)$、分散共分散行列を正値対称行列

$$\begin{pmatrix} 1 & \rho_{xy} & \rho_{xz} \\ \rho_{xy} & 1 & \rho_{yz} \\ \rho_{xz} & \rho_{yz} & 1 \end{pmatrix}$$

とする３変量正規分布にしたがうとき、次の各問に答えよ。

[1]　$X = x$ を与えた下での Y の条件付き期待値を $E[Y|x] = \beta_x x$ としたときの x の係数 β_x を求めよ。また、$X = x$ を与えた下での Y の条件付き分散 $V[Y|x]$ を求めよ。

[2]　$X = x$ および $Z = z$ を与えた下での Y の条件付き期待値を $E[Y|x, z] = \alpha_x x + \alpha_z z$ としたときの x の係数 α_x と z の係数 α_z を求めよ。また、$X = x$ および $Z = z$ を与えた下での Y の条件付き分散 $V[Y|x, z]$ を求めよ。

[3]　上問 [1] と [2] で求めた x の係数について、$\beta_x = \alpha_x$ となるための必要十分条件を示せ。

[4]　上問 [1] と [2] で求めた Y の条件付き分散について、$V[Y|x] = V[Y|x, z]$ となるための必要十分条件を示せ。

2012 年 数理統計 問 5 ポイント

- ２変量正規分布の確率密度関数
- 条件付き分布の定義
- 条件付き分布の平均（条件付き期待値）と分散（条件付き分散）の公式
- 偏相関係数の公式
- 正規分布では相関がなければ独立（逆は常に真）

2012 年 数理統計 問 5 解答例

[1]

$X \sim N(\mu_x, \sigma_x^2)$、$Y \sim N(\mu_y, \sigma_y^2)$ また X と Y との相関係数を ρ とおくと、(X, Y) の確率密度関数 $f(x, y)$ は以下で表せる。

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_x)^2}{\sigma_x^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right)\right\}$$

今 $\mu_x = \mu_y = 0$、$\sigma_x^2 = \sigma_y^2 = 1$、$\rho = \rho_{xy}\sigma_x\sigma_y = \rho_{xy}$ より、

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho_{xy}^2}} \exp\left\{-\frac{1}{2(1-\rho_{xy}^2)}\left(x^2 - 2\rho_{xy}xy + y^2\right)\right\}$$

また $X \sim N(0, 1)$ より X の確率密度関数 $f_X(x)$ は以下である。

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$$

以上より、$Y|x$ の確率密度関数 $f_{Y|x}(y)$ は以下である。

$$f_{Y|x}(y) = \frac{f(x,y)}{f_X(x)} = \frac{1}{\sqrt{2\pi(1-\rho_{xy}^2)}} \exp\left(-\frac{(y-\rho_{xy}x)^2}{2(1-\rho_{xy}^2)}\right)$$

これより $Y|x$ は平均 $\rho_{xy}x$、分散 $1-\rho_{xy}^2$ の正規分布に従うことがわかる。よって $\beta_x = \rho_{xy}$、$V[Y|x] = 1-\rho_{xy}^2$ となる。

（別解）

$$E(Y|x) = E(Y) + \frac{Cov(X,Y)}{V(X)}(x - E(X)) = \rho_{xy}x$$

よって $\beta_x = \rho_{xy}$

$$V(Y|x) = V(Y) - \frac{(Cov(X,Y))^2}{V(X)} = 1 - \rho_{xy}^2$$

[2]

$Y' = Y|z$、$X' = X|z$ とおく Y' と X' の偏相関係数 $\rho(X',Y')$ は以下となる。

$$\rho(X',Y') = \frac{\rho_{xy} - \rho_{xz}\rho_{yz}}{\sqrt{1-\rho_{xz}^2}\sqrt{1-\rho_{yz}^2}}$$

[1] より $X' \sim N(\rho_{xz}z, 1-\rho_{xz}^2)$、$Y' \sim N(\rho_{yz}z, 1-\rho_{yz}^2)$ なので

$$Cov(X',Y') = \rho(X',Y')\ \sqrt{1-\rho_{xz}^2}\sqrt{1-\rho_{yz}^2} = \rho_{xy} - \rho_{xz}\rho_{yz}$$

$E(Y|x,z) = E(Y'|X'=x)$ なので

$$\begin{aligned} E(Y|x,z) &= E(Y') + \frac{Cov(X',Y')}{V(X')}(x - E(X')) = \rho_{yz}z + \frac{(\rho_{xy} - \rho_{xz}\rho_{yz})}{1-\rho_{xz}^2}(x - \rho_{xz}z) \\ &= \frac{\rho_{xy} - \rho_{xz}\rho_{yz}}{1-\rho_{xz}^2}x + \frac{\rho_{yz} - \rho_{xy}\rho_{xz}}{1-\rho_{xz}^2}z \end{aligned}$$

以上より、

$$\alpha_x = \frac{\rho_{xy} - \rho_{xz}\rho_{yz}}{1-\rho_{xz}^2}, \quad \alpha_z = \frac{\rho_{yz} - \rho_{xy}\rho_{xz}}{1-\rho_{xz}^2}$$

$V(Y|x,z) = V(Y'|X'=x)$ なので

$$\begin{aligned} V(Y|x,z) &= V(Y') - \frac{(Cov(X',Y'))^2}{V(X')} = 1 - \rho_{yz}^2 - \frac{(\rho_{xy} - \rho_{xz}\rho_{yz})^2}{1-\rho_{xz}^2} \\ &= \frac{1 - \rho_{xy}^2 - \rho_{yz}^2 - \rho_{xz}^2 + 2\rho_{xy}\rho_{yz}\rho_{xz}}{1-\rho_{xz}^2} \end{aligned}$$

[3]

$$\rho_{xy} = \frac{\rho_{xy} - \rho_{xz}\rho_{yz}}{1-\rho_{xz}^2}$$

より $\rho_{xz}(\rho_{yz} - \rho_{xy}\rho_{xz}) = 0$ よって $\rho_{xz} = 0$ または $\rho_{yz} - \rho_{xy}\rho_{xz} = 0$

正規分布では相関係数が 0 であることと独立であることが同値であるので、$\rho_{xz} = 0$ は X と Z が独立であることを意味する。

また [2] と同様にして $Cov(Y',Z') = \rho_{yz} - \rho_{xy}\rho_{xz}$ なので $\rho_{yz} - \rho_{xy}\rho_{xz} = 0$ は $Y' = Y|x$（ x を与えたときの Y の条件付き分布）と $Z' = Z|x$（ x を与えたときの Z の条件付き分布）が独立であることを意味する。

これらが求める必要十分条件である。

[4]

$$1-\rho_{xy}^2=\frac{1-\rho_{xy}^2-\rho_{yz}^2-\rho_{xz}^2+2\rho_{xy}\rho_{yz}\rho_{xz}}{1-\rho_{xz}^2}$$

より

$$\rho_{xy}^2\rho_{xz}^2-2\rho_{xy}\rho_{yz}\rho_{xz}+\rho_{yz}^2=(\rho_{xy}\rho_{xz}-\rho_{yz})^2=0$$

よって $\rho_{xy}\rho_{xz}-\rho_{yz}=0$。これは [3] と同じく、$Y'=Y|x$（$x$ を与えたときの Y の条件付き分布）と $Z'=Z|x$（x を与えたときの Z の条件付き分布）が独立であることを意味する。

これらが求める必要十分条件である。

2013 年 数理統計 問 1

周の長さ 1 の円の円周上に 1 点 A を固定し、次に同じ円周上で一様に分布するように、別の点 B を取る。そして、それら 2 点で区切られてできる 2 つの弧 AB のうち、短い方の長さを X とし、長い方の長さを Y とする。このとき、以下の各問に答えよ。

[1] X および Y の期待値と標準偏差をそれぞれ求めよ。また、X と Y の間の相関係数はいくらか。

[2] X と Y の比 $W = X/Y$ の累積分布関数 $F(w)$ および確率密度関数 $f(w)$ を求め、$F(w)$ と $f(w)$ の概形を描け。

[3] 比 W の期待値、および、その分布の中央値はいくらか。

2013 年 数理統計 問 1 ポイント

ポイントはない、サービス問題

2013 年 数理統計 問 1 解答例

[1]

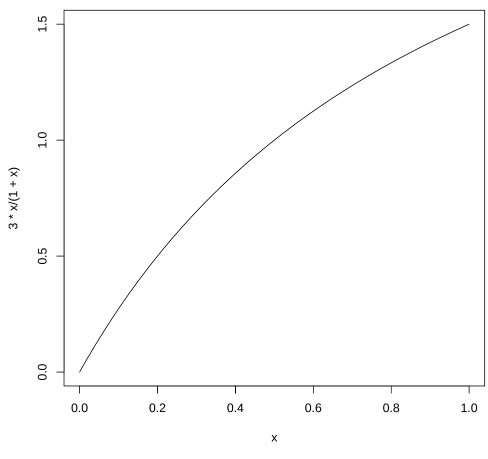
題意から X は $(0, 1/2)$ 上の一様分布に従う。よって $E(X) = 1/4$、$V(X) = 1/48$、よって X の標準偏差は $1/4\sqrt{3} = \sqrt{3}/12$。また $X+Y = 1$ より、$E(Y) = E(1-X) = 1-E(X) = 3/4$。$V(Y) = V(1-X) = V(X)$。よって Y の標準偏差は $1/4\sqrt{3} = \sqrt{3}/12$。

$Cov(X, Y) = Cov(X, 1-X) = E((X-1/4)(1-X-3/4)) = -E((X-1/4)^2) = -V(X) = -1/48$。よって X と Y の相関係数は $\frac{-1/48}{1/4\sqrt{3}\cdot 1/4\sqrt{3}} = -1$ である。

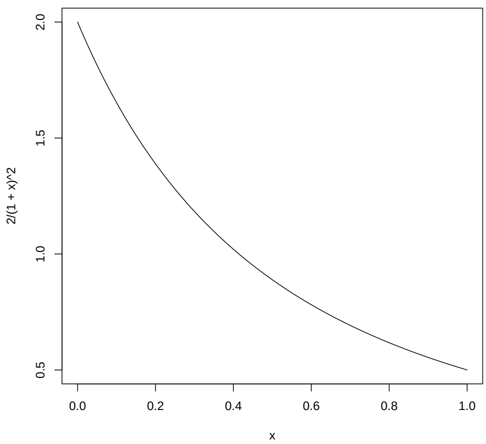
[2]

$W = \frac{X}{Y} = \frac{X}{1-X}$ である。

$$F(w) = P(W < w) = P\left(\frac{X}{1-X} < w\right) = P\left(X < \frac{w}{1+w}\right) = \int_0^{\frac{w}{1+w}} 2dx = \frac{2w}{1+w}$$



$$f(w) = F'(w) = \frac{2}{1+w} - \frac{2w}{(1+w)^2} = \frac{2}{(1+w)^2}$$



[3]

$$E(W) = \int_0^1 wf(w)dw = \int_0^1 \frac{2w}{(1+w)^2}dw$$

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

より

$$E(W) = 2\left[\log(1+w)\right]_0^1 + 2\left[\frac{1}{1+w}\right]_0^1 = 2\log 2 + 2(1/2-1) = 2\log 2 - 1$$

中央値 c は $F(c) = \frac{2c}{1+c} = \frac{1}{2}$ を解いて $c = 1/3$

2013 年 数理統計 問 2

確率変数 $X_1, \cdots, X_n$ が互いに独立に正規分布 $N(\mu, \sigma^2)$ に従うとする。以下の各問に答えよ。

[1]　総和 $Y_n = X_1 + \cdots + X_n$ と $X_1 - Y_n/n$ は互いに独立であることを証明せよ。

[2]　上問 [1] で、$Y_n = y_n$ を与えたという条件の下での、X_1 の条件付き分布を求めよ。

[3]　一般に、k 個の累積和を $Y_k = X_1 + \cdots + X_k$ $(k = 2, \cdots, n)$ と表わすとき、$Y_k = y_k$ を与えたという条件の下での、Y_{k-1} の条件付き分布を求めよ。

[4]　上問 [3] で求めた条件付き分布の確率密度関数を $f_{k-1}(y_{k-1}|y_k)$ とし、Y_n の周辺確率密度関数を $g_n(y_n)$ とするとき、積 $\{\prod_{k=2}^n f_{k-1}(y_{k-1}|y_k)\} \times g_n(y_n)$ はどのような関数になるか。

2013 年 数理統計 問 2 ポイント

- 正規分布では相関がなければ独立（逆は常に真）
- ２つの変数変換から同時確率密度関数を導く

2013 年 数理統計 問 2 解答例

[1]

$T = X_1 - Y_n/n$ とおく。$Y_n \sim N(n\mu, n\sigma^2)$、$T \sim N(0, \frac{n-1}{n}\sigma^2)$ である。正規分布 Y_n と正規分布 T が独立であることは、Y_n と T が無相関であることと同値である。よって $Cov(Y_n, T) = 0$ を示せば良い。今、$E(T) = E(X_1) - E(Y_n)/n = 0$ であり、$Cov(Y_n, T) = E(Y_nT) - E(Y_n)E(T) = E(Y_nT)$ なので $E(Y_nT) = 0$ を示せば良い。

$$E(Y_nT) = E(X_1(X_1 + \cdots + X_n) - Y_n^2/n) = E(X_1^2) + E(X_1X_2) + \cdots E(X_1X_n) - E(Y_n^2)/2$$

$E(X_1^2) = V(X_1) + E(X_1)^2 = \sigma^2 + \mu^2$、$j \neq 1$ のとき $E(X_1X_j) = E(X_1)E(X_j) = \mu^2$、
$E(Y_n^2) = V(Y_n) + E(Y_n)^2 = n\sigma^2 + n^2\mu^2$ なので、

$$E(Y_nT) = \sigma^2 + \mu^2 + (n-1)\mu^2 - (\sigma^2 + n\mu^2) = 0$$

[2]

[1] より $T = X_1 - Y_n/n$ の分布は Y_n の値によって変化せず、しかも $T \sim N(0, \frac{n-1}{n}\sigma^2)$ なので、

$$X_1|y_n - y_n/n \sim N(0, \frac{n-1}{n}\sigma^2)$$

$E(X_1|y_n - y_n/n) = 0$ より $E(X_1|y_n) = y_n/n$ また $V(X_1|y_n - y_n/n) = V(X_1|y_n) = \frac{n-1}{n}\sigma^2$ 以上より

$$X_1|y_n \sim N(y_n/n, \frac{n-1}{n}\sigma^2)$$

[3]

[2] より $X_i|y_k \sim N(y_k/k, \frac{k-1}{k}\sigma^2)$ また $Y_{k-1}|y_k = y_k - X_k|y_k$ であり、$X_k|y_k$ が正規分布であることに注意すると、$Y_{k-1}|y_k$ は正規分布となり、その平均と分散は以下となる。

$$E(Y_{k-1}|y_k) = E(y_k - X_k|y_k) = y_k - E(X_k|y_k) = y_k - y_n/k = \frac{k-1}{k}y_k$$

$$V(Y_{k-1}|y_k) = V(y_k - X_k|y_k) = V(X_k|y_k) = \frac{k-1}{k}\sigma^2$$

以上より

$$Y_{k-1}|y_k \sim N\left(\frac{k-1}{k}y_n, \frac{k-1}{k}\sigma^2\right)$$

[4]

X_i の確率密度関数を $f(x)$ とおく。明らかに $g_1(y_1) = f(x_1)$ である。また (Y_{k-1}, Y_k) の同時密度関数を $h_{k-1}(y_{k-1}, y_k)$ とおくと、$f_{k-1}(y_{k-1}|y_k) = h_{k-1}(y_{k-1}, y_k)/g_k(y_k)$ である。(Y_{k-1}, Y_k) を (Y_{k-1}, X_k) に変数変換すると、変換のヤコビアンは 1 であることから、

$$h_{k-1}(y_{k-1}, y_k) = g_{k-1}(y_{k-1})f(y_k - y_{k-1}) = g_{k-1}(y_{k-1})f(x_k)$$

よって

$$f_{k-1}(y_{k-1}|y_k) = \frac{g_{k-1}(y_{k-1})f(x_k)}{g_k(y_k)}$$

よって

$$\begin{aligned}\prod_{k=2}^{n} f_{k-1}(y_{k-1}|y_k) \cdot g_n(y_n) &= \frac{f(x_2)g_1(y_1)}{g_2(y_2)} \cdot \frac{f(x_3)g_2(y_2)}{g_3(y_3)} \cdot \cdots \cdot \frac{f(x_n)g_{n-1}(y_{n-1})}{g_n(y_n)} \cdot g_n(y_n) \\ &= g_1(y_1)\prod_{k=2}^{n} f(x_k) \\ &= \prod_{k=1}^{n} f(x_k)\end{aligned}$$

2013 年 数理統計 問 3

試行回数 n、成功の確率 p の二項分布 $B(n,p)$ に従う確率変数を X とする。二項分布の確率関数は $p(x)=Pr(X=x)={}_nC_xp^x(1-p)^{n-x}$ である。以下の各問に答えよ。

[1] X の期待値は np であり、分散は $np(1-p)$ であることを示せ。

[2] $\hat{p}=X/n$ は p の最尤推定量であること、および $\hat{p}$ は p の不偏推定量であることを示せ。

[3] $\hat{p}$ の分散を、p の関数として $w_n(p)$ とする。この分散 $w_n(p)$ の推定量 $\hat{w}_n(p)=w_n(\hat{p})$ は、$w_n(p)$ の不偏推定量でないことを示せ。また $\hat{w}_n=c(n)\hat{w}_n$ が $w_n(p)$ の不偏推定量となるための関数 $c(n)$ を求めよ。

[4] 二項分布の正規近似に基づく p の近似的な 95% 信頼区間は $\hat{p}\pm1.96\sqrt{\hat{w}_n}$ で与えられる。この信頼区間と、$\hat{w}_n$ の代わりに上問 [3] で求めた不偏推定量 $\tilde{w}_n$ を用いた信頼区間 $\hat{p}\pm1.96\sqrt{\tilde{w}_n}$ とを、被服確率の観点から比較して論ぜよ。

2013 年 数理統計 問 3 ポイント

- 不偏推定量の定義
- 被覆率の定義

2013 年 数理統計 問 3 解答例

[1]

i 回目試行で成功すれば 1 成功しなければ 0 となる確率変数を X_i とすると、$X=\sum_{i=1}^n X_i$ となる。各 X_i は独立であり、$E(X_i)=p$、$V(X_i)=p(1-p)$ なので

$$E(X)=\sum_{i=1}^n E(X_i)=np$$

$$V(X)=\sum_{i=1}^n V(X_i)=np(1-p)$$

[2]

尤度関数は $L(p)={}_nC_xp^x(1-p)^{n-x}$、対数尤度関数は $l(p)=\log_n C_x+x\log p+(n-x)\log(1-p)$

$$l'(p)=\frac{x}{p}-\frac{n-x}{1-p}=\frac{x-np}{p(1-p)}$$

$l'(\hat{p})=0$ より $\hat{p}=X/n$

[3]

$$w_n(p)=V(X/n)=\frac{V(X)}{n^2}=\frac{p(1-p)}{n}$$

$$\hat{w}_n(p)=w_n(\hat{p})=\frac{\hat{p}(1-\hat{p})}{n}=\frac{X/n(1-X/n)}{n}=\frac{X(n-X)}{n^3}$$

$$
\begin{aligned}
E(\hat{w}_n(p)) &= E(\frac{X(n-X)}{n^3}) = \frac{E(nX-X^2)}{n^3} = \frac{nE(X)-E(X^2)}{n^3} \\
&= \frac{n^2p-(np(1-p)+(np)^2)}{n^3} = \frac{np-(p(1-p)+np^2)}{n^2} \\
&= \frac{np(1-p)-p(1-p)}{n^2} = \frac{n-1}{n^2}p(1-p) = \frac{n-1}{n}w_n(p)
\end{aligned}
$$

$E(\hat{w}_n(p)) \neq w_n(p)$ より $\hat{w}_n(p)$ は $w_n(p)$ の不偏推定量ではない。

また $c(n) = \frac{n}{n-1}$ であれば上記より $\hat{w}_n(p)$ は $w_n(p)$ の不偏推定量となる。

[4]

$\tilde{w}_n$ を用いる場合の信頼区間と $\hat{w}_n$ を用いる場合の信頼区間は、区間の中点は同じあるが、区間の幅が $\tilde{w}_n$ を用いる方が若干広くなる。このため $\tilde{w}_n$ を用いる方が被服確率が若干高くなる。ただしサンプル数 n が十分大きい場合には差はない。

2013 年 数理統計 問 4

2 変量連続型分布から大きさ n の無作為標本を $(x_1, y_1), \cdots, (x_n, y_n)$ とし、差を $z_i = y_i - x_i$ とする $(i = 1, \cdots, n)$。差 z_i の母集団分布 $F(z)$ の中央値を θ とし、帰無仮説 $H_0 : \theta = 0$ を、対立仮説 $H_1 : \theta > 0$ に対して、有意水準 5% で検定する。ここでは、母集団分布 $F(z)$ は正規分布とは限らないが、θ を中心に左右対称であること仮定する。また、n 個の z_i のなかに 0 であるものはなく、かつ、すべての z_i の絶対値は異なるもとする。n 個 z_i を絶対値の小さい順に並べたときの z_i の順位を R_i とし、z_i の符号が正のもののみの順位和 $T^+ = \sum_{z_i > 0} R_i$ を検定統計量とする。そして、検定の棄却限界値 c を、$Pr(T^+ \geq 0 | H_0) \leq 0.05$ を満たす最小の整数とし、$T^+ \geq c$ のとき H_0 を棄却する。このとき、以下の各問に答えよ。

[1]　すべての z_i が正でも、サンプルサイズ n によっては検定で有意にならないことがある。では、すべての z_i が正であるときに、この検定で 5% 有意となるためには、n はいくつ以上ででなければならないか。

[2]　$n = 7$ のとき T^+ の値の数え上げによって有意水準 5% で検定する場合の棄却限界値 c はいくらか。また、このときの実際の有意確率はいくらか。

[3]　H_0 の下での T^+ の期待値と分散は、それぞれ

$$E[T^+] = \frac{n(n+1)}{4}, \quad V[T^+] = \frac{n(n+1)(2n+1)}{24}$$

となることを示せ。

[4]　T^+ は、n が大きいとき近似的に、上問 [3] で求めた期待値と分散を持つ正規分布に従う。この正規近似を利用して検定する場合、$n = 7$ のときの棄却限界値を求めよ。

2013 年 数理統計 問 4 ポイント

ポイントはない。$\theta = 0$ を仮定すると $P(z_i > 0) = 1/2$ であることに気がつくかどうかだけの問題。

2013 年 数理統計 問 4 解答例

[1]

$\theta = 0$ を仮定すると題意より $P(z_i > 0) = 1/2$ である。このため n 個すべてで $z_i > 0$ である確率 p は $(1/2)^n$ である。$n = 4$ のとき $p = 0.0625 > 0.05$、$n = 5$ のとき $p = 0.03125 < 0.05$ なので $n = 5$ 以上でなければならない。

[2]

z_i の正負の並びは 2^7 通り存在し、$\theta = 0$ の仮定の下では $P(z_i > 0) = 1/2$ であるので、それらの出現する確率は等しい。よって $Pr(T^+ \geq c | H_0) \leq 0.05$ を満たす z_i のパターンは 6.4 通り以下でなければならない。T^+ の最大値は $1 + 2 + \cdots + 7 = 28$ であり、$T^+ = 28$ となる z_i の並びは 1 通り、$T^+ = 27$ となる z_i の並びは 1 通り、$T^+ = 26$ となる z_i の並びは 1 通り、$T^+ = 25$ となる z_i の並びは 2 通り、$T^+ = 24$ となる z_i の並びは 2 通り、以上より、$T^+ \geq 25$ となるのは 5 通り、$T^+ \geq 24$ となるのは 7 通りなので、$c = 25$ となる。このときの有意確率は $5/2^7 \approx 0.039$ である。

[3]

以下のような確率変数 X_i を定義する。

$$X_i = \begin{cases} i & (z_i > 0) \\ 0 & (z_i < 0) \end{cases}$$

H_0 の仮定の下では $X_i = i$ となる確率は 1/2 なので $E(X_i) = i/2$。明らかに $T^+ = \sum_{i=1}^{n} X_i$ なので、

$$E(T^+) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} \frac{i}{2} = \frac{n(n+1)}{4}$$

また $V(X_i) = E(X_i^2) - E(X_i)^2 = i^2/2 - (i/2)^2 = i^2/4$

$$V(T^+) = \sum_{i=1}^{n} V(X_i) = \sum_{i=1}^{n} \frac{i^2}{4} = \frac{n(n+1)(2n+1)}{24}$$

[4]

$\mu = \frac{7\cdot8}{4} = 14$ 、$\sigma^2 = \frac{7\cdot8\cdot15}{24} = 35$

$$Z = \frac{T^+ - \mu}{\sigma} \sim N(0,1)$$

$$P(T^+ > c) = P(\sigma Z + \mu > c) = P(Z > \frac{c-\mu}{\sigma}) = 0.05$$

$$c = z(0.05)\sigma + \mu = 1.64 \cdot \sqrt{35} + 14 = 23.7$$

以上より $c = 24$ となる。

2013 年 数理統計 問 5

多項分布に関する以下の推測問題を考える。離散型確率変数 $X_1, \cdots, X_m$ （ただし、$X_1 + \cdots + X_m = n$ であり、 n は与えられた正の整数）が、パラメータ $n, \theta_1, \cdots, \theta_m$ $(\theta_1 + \cdots + \theta_m = 1)$ の多項分布に従っているとする。これらの確率変数の 1 組の観測値 $x_1, \cdots, x_m$ が観測されたとして、以下の各問に答えよ。

[1]　この観測値に基づいた尤度関数を求めよ。

[2]　$\theta_{10}, \cdots, \theta_{m0}$ を与えられたパラメータ値とし、帰無仮説 $H_0 : \theta_1 = \theta_{10}, \cdots, \theta_m = \theta_{m0}$ を検定する。このとき、
(a) 尤度比検定統計量と、その帰無仮説の下での漸近分布を示せ。
(b) 適合度のカイ二乗検定統計量と、その帰無仮説の下での漸近分布を示せ。

[3]　確率変数 X がパラメータ n および p の二項分布 $B(n, p)$ に従うとする。p_0 をある与えられた値としたとき、X の正規近似による、帰無仮説 $H_0 : p = p_0$ の両側検定は、上問 [2] の (b) における $m = 2$ の場合の帰無仮説 $H_0 : \theta_1 = p_0 . \theta_2 = 1 - p_0$ の検定と一致することを示せ。

[4]　上問 [2] における (a) および (b) の検定法は、漸近的に同等であることを示せ。

2013 年 数理統計 問 5 ポイント

- 尤度比検定と適合度のカイ 2 乗検定 (2014-Q5)

[4] でテーラー展開による $\log(1 + x) \approx x - \frac{1}{2}x^2$ を用いているが、これを試験の場で使うのは無理。

2013 年 数理統計 問 5 解答例

[1]

$$L = \frac{n!}{x_1!x_2! \cdots x_m!} \theta_1^{x_1} \theta_2^{x_2} \cdots \theta_m^{x_m}$$

[2]

(a)

H_0 の下での尤度 L_0 は以下である。

$$L_0 = \frac{n!}{x_1!x_2! \cdots x_m!} \theta_{10}^{x_1} \theta_{20}^{x_2} \cdots \theta_m^{x_m}$$

一方、θ_i の最尤推定値は $\hat{\theta}_i = x_i/n$ なので、対立仮説の下での尤度 L_1 は以下である。

$$\begin{aligned} L_1 &= \frac{n!}{x_1!x_2! \cdots x_m!} \hat{\theta}_1^{x_1} \hat{\theta}_2^{x_2} \cdots \hat{\theta}_m^{x_m} \\ &= \frac{n!}{x_1!x_2! \cdots x_m!} (\frac{x_1}{n})^{x_1} (\frac{x_2}{n})^{x_2} \cdots (\frac{x_m}{n})^{x_m} \end{aligned}$$

よって尤度比 Λ は

$$\Lambda = \frac{L_1}{L_0} = (\frac{x_1}{n\theta_{10}})^{x_1}(\frac{x_2}{n\theta_{20}})^{x_2}\cdots(\frac{x_m}{n\theta_{m0}})^{x_m} = \prod_{i=1}^{m}(\frac{x_i}{n\theta_{i0}})^{x_i}$$

よって求める統計量は以下である。

$$\Lambda' = 2\log\Lambda = 2\sum_{i=1}^{m} x_i \log(\frac{x_i}{n\theta_{i0}})$$

H_0 でのパラメータ数は 0 個、対立仮説でのパラメータ数は $m-1$ 個なので、上記の統計量が自由度 $m-1$ のカイ 2 乗分布に従うことを利用して検定を行う。

(b)

H_0 の下で $E(X_i) = n\theta_{i0}$ なので、求める統計量は以下である。

$$Y = \sum_{i=1}^{m} \frac{(n\theta_{i0} - x_i)^2}{n\theta_{i0}}$$

H_0 でのパラメータ数は 0 個なので、上記の統計量が自由度 $m-0-1 = m-1$ のカイ 2 乗分布に従うことを利用して検定を行う。

[3]

H_0 の下で $m=2$ のとき $X_1 = X$、$X_2 = n - X$、$\theta_1 = p_0$、$\theta_2 = 1 - p_0$ となるので、

$$\begin{aligned} Y &= \frac{(np_0 - X)^2}{np_0} + \frac{(n(1-p_0) - (n-X))^2}{n(1-p_0)} \\ &= \frac{(X - np_0)^2}{np_0(1-p_0)} \\ &= \frac{(X/n - p_0)^2}{p_0(1-p_0)} \sim \chi_1^2 \end{aligned}$$

H_0 の下で X の正規近似により $X \sim N(np_0, np_0(1-p_0))$ なので正規化して

$$Z = \frac{X - np_0}{\sqrt{np_0(1-p_0)}} = \frac{X/n - p_0}{\sqrt{p_0(1-p_0)}} \sim N(0,1)$$

Z^2 は自由度 1 のカイ 2 乗分布に従い、しかも Y と等しい。これによって問題の 2 つの検定は一致する。

[4]

$$\begin{aligned} \Lambda' &= 2\sum_{i=1}^{m} x_i \log(\frac{x_i}{n\theta_{i0}}) \\ &= 2\sum_{i=1}^{m}(x_i - n\theta_{i0} + n\theta_{i0})\log(\frac{x_i - n\theta_{i0} + n\theta_{i0}}{n\theta_{i0}}) \\ &= 2\sum_{i=1}^{m}(x_i - n\theta_{i0} + n\theta_{i0})\log(1 + \frac{x_i - n\theta_{i0}}{n\theta_{i0}}) \end{aligned}$$

テーラー展開より $\log(1+x) \approx x - \frac{1}{2}x^2$

$$
\begin{aligned}
\Lambda' &\approx 2\sum_{i=1}^{m}(x_i - n\theta_{i0} + n\theta_{i0})\left(\frac{x_i - n\theta_{i0}}{n\theta_{i0}} - \frac{1}{2}(\frac{x_i - n\theta_{i0}}{n\theta_{i0}})^2\right) \\
&= 2\sum_{i=1}^{m}\left\{\left(\frac{(x_i - n\theta_{i0})^2}{n\theta_{i0}} - \frac{1}{2}\frac{(x_i - n\theta_{i0})^3}{(n\theta_{i0})^2}\right) + \left((x_i - n\theta_{i0}) - \frac{1}{2}\frac{(x_i - n\theta_{i0})^2}{n\theta_{i0}}\right)\right\} \\
&= \sum_{i=1}^{m}\frac{(x_i - n\theta_{i0})^2}{n\theta_{i0}} + \sum_{i=1}^{m}\left(-\frac{(x_i - n\theta_{i0})^3}{(n\theta_{i0})^2} + 2(x_i - n\theta_{i0})\right)
\end{aligned}
$$

ここで

$$
\sum_{i=1}^{m}(x_i - n\theta_{i0}) = \sum_{i=1}^{m}x_i - n\sum_{i=1}^{m}\theta_{i0} = n - n = 0
$$

$$
\frac{(x_i - n\theta_{i0})^3}{(n\theta_{i0})^2} \approx 0
$$

よって

$$
\Lambda' = \sum_{i=1}^{m}\frac{(x_i - n\theta_{i0})^2}{n\theta_{i0}}
$$

2014 年 数理統計 問 1

区間 $(0,1)$ 上の一様分布に従う互いに独立な確率変数 U,V,W に対し、α,β,γ を正の定数として、$X=U^{\alpha}, Y=V^{\beta}, Z=W^{\gamma}$ とする。このとき、以下の各問に答えよ。

[1] $\alpha=2,\beta=3$ のとき、$U=u$ が与えられた下での条件付き確率 $P(X>Y|U=u)$ を求めよ。これを用いて $X>Y$ となる確率 $P(X>Y)$ を求めよ。

[2] 一般の α,β,γ に対し、X が X,Y,Z の中で最大となる確率 $P(X=\max(X,Y,Z))$ を求めよ。

[3] 上問 [2] を拡張し、$U_i(i=1,\cdots,n)$ を互いに独立な区間 $(0,1)$ 上の一様分布に従う n 個の確率変数とし、$\alpha_i(i=1,\cdots,n)$ を正の定数として、$X_i=U_i^{\alpha_i}(i=1,\cdots,n)$ とするとき、X_1 が $X_1,\cdots,X_n$ の中で最大となる確率 $P(X_1=\max(X_1,\cdots,X_n))$ を $\alpha_1,\cdots,\alpha_n$ の関数として求めよ。

2014 年 数理統計 問 1 ポイント

- 一様分布の最大値の確率の表現

2014 年 数理統計 問 1 解答例

[1]

$$P(X>Y|U=u)=P(V^3<u^2)=P(V<u^{2/3})=\int_0^{u^{2/3}}1\ dv=u^{2/3}$$
$$P(X>Y)=\int_0^1 P(X>Y|U=u)du=\int_0^1 u^{2/3}du=\left[\frac{3}{5}u^{5/3}\right]_0^1=\frac{3}{5}$$

[2]

$$\begin{aligned}P(X=\max(X,Y,Z)|U=u)&=P(X>Y,X>Z|U=u)=P(X>Y|U=u)P(X>Z|U=u)\\&=u^{\alpha/\beta}u^{\alpha/\gamma}=u^{\alpha/\beta+\alpha/\gamma}\end{aligned}$$

$$\begin{aligned}P(X=\max(X,Y,Z))&=\int_0^1 P(X=\max(X,Y,Z)|U=u)du\\&=(\alpha/\beta+\alpha/\gamma+1)^{-1}=(\alpha/\alpha+\alpha/\beta+\alpha/\gamma)^{-1}\\&=\frac{1/\alpha}{1/\alpha+1/\beta+1/\gamma}\end{aligned}$$

[3]

[2] から明らかに

$$P(X_1=\max(X_1,\cdots,X_n))=\frac{1/\alpha_1}{1/\alpha_1+1/\alpha_2+\cdots+1/\alpha_n}$$

2014 年 数理統計 問 2

連続型確率変数 X の確率密度関数が，$m > 0$ として

$$f(x) = \begin{cases} \frac{1}{\Gamma(m)} x^{m-1} e^{-x} & (x > 0), \\ 0 & (x \leq 0) \end{cases}$$

で与えられるとき，X はパラメータ m のガンマ分布に従うという。ここで $\Gamma(m)$ はガンマ関数を表し，m が正の整数ならば $\Gamma(m) = (m-1)!$ である。

[1]　パラメータ m のガンマ分布のモーメント母関数は，$t < 1$ に対して、$M_X(t) = (1-t)^{-m}$ であることを示せ。

[2]　パラメータ m のガンマ分布の平均と分散をそれぞれ求めよ。

[3]　$n+1$ 個の確率変数 $X_1, \cdots, X_{n+1}$ が、それぞれ互いに独立に $m = 1$ のガンマ分布に従うとき，$T = X_1 + \cdots + X_n + X_{n+1}$ および $Y_1 = X_1/T, Y_2 = (X_1 + X_2)/T, \cdots, Y_n = (X_1 + \cdots + X_n)/T$ とする。このとき，T の確率分布は何か。また，$(Y_1, \cdots, Y_n)$ と T は互いに独立であることを示せ。

[4]　上問 [3] で定義された $(Y_1, \cdots, Y_n)$ の同時分布は，互いに独立に区間 $(0,1)$ 上の一様分布に従う確率変数 $U_1, \cdots, U_n$ の順序統計量 $U_{(1)} < \cdots < U_{(n)}$ の同時分布に等しいことを示せ。

[5]　上問 [4] の結果を用い，第 j 番目の順序統計量 $U_{(j)}$ の期待値 $E[U_{(j)}]$ $(j = 1, \cdots, n)$ を求めよ。

2014 年 数理統計 問 2 ポイント

- モーメント母関数の定義
- ガンマ分布とベータ分布の定義
- ガンマ分布とベータ分布の関係
- ２つの変数変換から同時確率密度関数を導く（独立性を示すことにも利用可能）

2014 年 数理統計 問 2 解答例

[1]

$$M_X(t) = E[e^{tX}] = \int e^{tx} f(x) dx = \int_0^\infty \frac{1}{\Gamma(m)} x^{m-1} e^{-(1-t)x} dx$$

$y = (1-t)x$ で変数変換すると、$dy = (1-t)dx$ より

$$\begin{aligned} M_X(t) &= \int_0^\infty \frac{1}{\Gamma(m)} (\frac{y}{1-t})^{m-1} e^{-y} \frac{1}{1-t} dy \\ &= (\frac{1}{1-t})^{m-1} \frac{1}{1-t} \int_0^\infty \frac{1}{\Gamma(m)} y^{m-1} e^{-y} dy = (1-t)^{-m} \end{aligned}$$

[2]

$$M'_X(t) = m(1-t)^{-(m+1)}, M''_X(t) = m(m+1)(1-t)^{-(m+2)}$$

$$E(X) = M'_X(0) = m$$

$$V(X) = M''_X(0) - M'_X(0)^2 = m(m+1) - m^2 = m$$

[3]

T のモーメント母関数を求める。

$$\begin{aligned} M_T(t) = E[e^{tT}] &= E[e^{t(X_1+\cdots+X_{n+1})}] = E[e^{tX_1}e^{tX_2}\cdots e^{tX_{n+1}}] \\ &= E[e^{tX_1}]E[e^{tX_2}]\cdots E[e^{tX_{n+1}}] = (1-t)^{-1}(1-t)^{-1}\cdots(1-t)^{-1} \\ &= (1-t)^{-(n+1)} \end{aligned}$$

これは [1] より自由度 $n+1$ のガンマ分布のモーメント母関数と等しいので、T の確率分布は自由度 $n+1$ のガンマ分布である。

$U = X_1 + X_2 + \cdots + X_i$, $V = X_{i+1} + X_{i+2} + \cdots + X_{n+1}$ とおくと、U は自由度 i のガンマ分布、V は自由度 $n+1-i$ のガンマ分布となり、明らかに U と V は独立である。(U, V) の同時確率密度関数 $f_{UV}(u, v)$ は以下となる。

$$\frac{1}{\Gamma(i)}u^{i-1}e^{-u} \cdot \frac{1}{\Gamma(n+1-i)}v^{n-i}e^{-v}$$

また $T = U + V$,$Y_i = U/(U+V)$ となっている。ここで (Y_i, T) の同時確率密度関数 $f_{Y_iT}(y, t)$ を考える。今 $u = yt, v = (1-y)t$ よりヤコビアンは t なので、

$$\begin{aligned} f_{Y_iT}(y,t) &= tf_{UV}(yt, (1-y)t) \\ &= t\frac{1}{\Gamma(i)}(yt)^{i-1}e^{-yt} \cdot \frac{1}{\Gamma(n+1-i)}((1-y)t)^{n-i}e^{-(1-y)t} \\ &= \frac{1}{\Gamma(i)\Gamma(n+1-i)}y^{i-1}(1-y)^{n-i}t^n e^{-t} \\ &= \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n+1-i)}y^{i-1}(1-y)^{n-i}\frac{1}{\Gamma(n+1)}t^n e^{-t} \end{aligned}$$

これは Y_i と T が独立で、Y_i が自由度 $(i, n-i+1)$ のベータ分布に従い、T が自由度 $n+1$ のガンマ分布に従うことを示している。

[4]

[3] より Y_i が自由度 $(i, n-i+1)$ のベータ分布に従っているので、$U_{(i)}$ の確率密度関数 $f(x)$ が以下であることを示せばよい。

$$\frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n+1-i)}x^{i-1}(1-x)^{n-i}$$

$U_{(i)}$ の値が x となる確率は dx である。どの U_j が $U_{(i)}$ になるかで n 通り。残りの $n-1$ 個の U_j のうち $i-1$ 個が x 以下の値を取り、残りの $n-i$ 個が x 以上の値を取るので、その確率は $x^{i-1}(1-x)^{n-i}$ である。またどの $i-1$ 個の U_j が x 以下の値を取るかで、${}_{n-1}C_{i-1}$ 通りある。以上より、$U_{(i)}$ の確率密度関数は

$$f(x) = n \cdot {}_{n-1}C_{i-1} \cdot x^{i-1}(1-x)^{n-i} = \frac{n!}{(i-1)!(n-i)!}x^{i-1}(1-x)^{n-i}$$

また $\Gamma(n+1) = n!$ なので、

$$f(x) = \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n+1-i)}x^{i-1}(1-x)^{n-i}$$

[5]

$$E(Y_iT) = E(X_1 + X_2 + \cdots + X_i) = E(X_1) + E(X_2) + \cdots + E(X_i) = i$$

また $E(T) = n + 1$。次に Y_i と T は独立なので、$E(Y_i T) = E(Y_i)E(T)$。よって

$$E(Y_i) = \frac{i}{n+1}$$

2014 年 数理統計 問 3

ある液体の比重を求めるために n 個のサンプルを取り、それぞれの体積 (X) と重さ (Y) を量ったものを $(X_i, Y_i),\ i = 1, \cdots, n$ とする。ただし、単位はミリリットルおよびグラムである。すべての $i = 1, \cdots, n$ において、X_i と Y_i は独立であり、各 X_i, Y_i は互いに独立にそれぞれの分散 σ_X^2, σ_Y^2 の正規分布に従うとする。この液体の比重（重さ/体積）を β としたとき、次の各問に答えよ。

[1]　比重が β のとき、$U_i = Y_i - \beta X_i,\ i = 1, \cdots, n$ は互いに独立に平均 0 の正規分布に従うとして、帰無仮説 $H_0 : \beta = \beta_0$ の検定方式を

(i) σ_X^2, σ_Y^2 が既知の場合、(ii) σ_X^2, σ_Y^2 が未知の場合

のそれぞれについて求めよ。ただし、標準正規分布および自由度 f の t 分布の上側 $100\alpha\%$点をそれぞれ $z(\alpha)$,$t_f(\alpha)$ と表す。

[2]　上問 [1] の (i) および (ii) の 2 つの場合について、β の信頼係数 $1 - \alpha$ の信頼区間を求めよ。

2014 年 数理統計 問 3 ポイント

- 正規分布の平均の検定（分散既知）
- 正規分布の平均の検定（分散未知）

2014 年 数理統計 問 3 解答例

[1]

(i)

$U \sim N(0, \sigma^2)$ である。また

$$\sigma^2 = V(U) = V(Y - \beta X) = \sigma_Y^2 + \beta^2 \sigma_X^2$$

σ_X^2 と σ_Y^2 が既知なので分散既知の正規分布の平均の検定として、H_0 の下で検定統計量 Z を以下で定めればよい。

$$Z = \frac{\bar{U}}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\bar{Y} - \beta_0 \bar{X}}{\sqrt{\frac{\sigma_Y^2 + \beta_0^2 \sigma_X^2}{n}}} \sim N(0, 1)$$

有意水準 α とすると、棄却域 R は

$$R = \{z | |z| > z(\alpha/2)\}$$

(ii)

(i) と同様であるが、σ_X^2 と σ_Y^2 が未知なので分散未知の正規分布の平均の検定として、H_0 の下で検定統計量 T を以下で定めればよい。

$$T = \frac{\bar{U}}{\sqrt{\frac{S^2}{n-1}}} = \frac{\bar{Y} - \beta_0 \bar{X}}{\sqrt{\frac{S_Y^2 + \beta_0^2 S_X^2}{n-1}}} \sim t_{n-1}$$

ただし

$$S_X^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

$$S_Y^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

である。

有意水準 α とすると、棄却域 R は

$$R = \{t||t| > t_{n-1}(\alpha/2)\}$$

[2]

信頼係数 $1-\alpha$ の信頼区間は有意水準 α とした棄却域 R の補集合になるので、R の境界点が信頼区間の上下点となる。

(i)

信頼区間の上下点は

$$\frac{\bar{Y} - \beta_0\bar{X}}{\sqrt{\frac{\sigma_Y^2 + \beta_0^2\sigma_X^2}{n}}} = z(\alpha/2)$$

を β_0 の方程式として見た場合の解となる。上記の式は以下に変形できる。

$$\left(\bar{X}^2 - \frac{z(\alpha/2)^2\sigma_X^2}{n}\right)\beta_0^2 - 2\bar{X}\bar{Y}\beta_0 + \bar{Y}^2 - \frac{z(\alpha/2)^2\sigma_Y^2}{n} = 0$$

今、n がある程度大きければ確率 1 で $\bar{X}^2 - \frac{z(\alpha/2)^2\sigma_X^2}{n} > 0$ となるので、上記 2 次方程式を解くことで、信頼区間の上下点は以下となる。

$$\frac{\bar{X}\bar{Y} \pm \sqrt{(\bar{X}\bar{Y})^2 - \left(\bar{X}^2 - \frac{z(\alpha/2)^2\sigma_X^2}{n}\right)\left(\bar{Y}^2 - \frac{z(\alpha/2)^2\sigma_Y^2}{n}\right)}}{\bar{X}^2 - \frac{z(\alpha/2)^2\sigma_X^2}{n}}$$

(ii)

(i) と同様、信頼区間の上下点は

$$\frac{\bar{Y} - \beta_0\bar{X}}{\sqrt{\frac{S_Y^2 + \beta_0^2 S_X^2}{n-1}}} = t_{n-1}(\alpha/2)$$

を β_0 の方程式として見た場合の解となる。これは (i) の解において、n を $n-1$、σ_x^2 を S_x^2、σ_Y^2 を S_Y^2、$z(\alpha/2)$ を $t_{n-1}(\alpha/2)$ とみたものと同じであるので、求める信頼区間の上下点は以下となる。

$$\frac{\bar{X}\bar{Y} \pm \sqrt{(\bar{X}\bar{Y})^2 - \left(\bar{X}^2 - \frac{t_{n-1}(\alpha/2)^2 S_X^2}{n-1}\right)\left(\bar{Y}^2 - \frac{t_{n-1}(\alpha/2)^2 S_Y^2}{n-1}\right)}}{\bar{X}^2 - \frac{t_{n-1}(\alpha/2)^2 S_X^2}{n-1}}$$

2014 年 数理統計 問 4

ばね秤りを用いて物体の重さを量る。5 個の物体の重さを秤りを 10 回用いて量るとし、次の 2 通りの測定方法を考える。

(1) 1 つずつそれぞれ 2 回量る。

(2) 2 つずつ 10 通り（${}_5C_2 = 10$）の異なる組み合わせについて各 1 回量る。

測定誤差は互いに独立に平均 0、分散 σ^2 の正規分布に従うとして、それぞれの測定法につき、以下の各問に答えよ。

[1]　物体の重さの真値を $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_5)'$、測定値を $\boldsymbol{x} = (x_1, \cdots, x_{10})'$、計画行列を X, 誤差を $\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_{10})'$ として、線形モデル $\boldsymbol{x} = X\boldsymbol{\theta} + \boldsymbol{\epsilon}$ を仮定する。ここでプライム ($'$) は行列またはベクトルの転置を表す。このとき、X を具体的に与えよ。また θ_i の最小二乗推定量を $\hat{\theta}_i$ $(i = 1, \cdots, 5)$ として、それを導く正規方程式を書け。

[2]　推定量 $\hat{\theta}_1, \cdots, \hat{\theta}_5$ の分散を求めよ。

[3]　σ^2 は未知として、帰無仮説「$H_0 : \theta_1 = \cdots = \theta_5$」を対立仮説「$H_1 : \theta_1 = \cdots = \theta_5$ でない」に対して検定するための F 統計量を求めよ。なお F 統計量は、適宜、行列などを用いて表せ。

[4]　対立仮説 H_1 の下での F 統計量の非心度 λ を求めよ。ただし、$Z_i, i = 1, \cdots, f$ が互いに独立に $N(\mu_i, 1)$ に従うとき、$Y = Z_1^2 + \cdots + Z_f^2$ の分布を非心度 $\lambda = \mu_1^2 + \cdots + \mu_f^2$ で自由度が f の非心カイ二乗分布といい、Y_1 および Y_2 が互いに独立にそれぞれ非心度 λ で自由度 f_1 の非心カイ二乗分布、および自由度 f_2 のカイ二乗分布に従うとき、F 統計量 $F = (Y_1/f_1)/(Y_2/f_2)$ の非心度は λ であると定義される。

[5]　上問に対する結果から、測定法 (1) と (2) のどちらの量り方が効率的と考えられるか。なお、すべての対角要素が a であり、すべての非対角要素が b である正則な $k \times k$ 行列につうて次の結果を用いてよい。

$$\begin{pmatrix} a & b & \dots & b \\ b & a & \dots & b \\ \vdots & \vdots & \ddots & \vdots \\ b & b & \dots & a \end{pmatrix}^{-1} = \begin{pmatrix} \alpha & \beta & \dots & \beta \\ \beta & \alpha & \dots & \beta \\ \vdots & \vdots & \ddots & \vdots \\ \beta & \beta & \dots & \alpha \end{pmatrix}$$

$$\alpha = \frac{a + (k-2)b}{(a-b)\{a + (k-1)b\}}, \beta = \frac{-b}{(a-b)\{a + (k-1)b\}}$$

2014 年 数理統計 問 4 ポイント

難問。[4] の解答は公式解答と異なる。間違っているかも。

- 複数群の平均が等しいことを示す F 検定

2014 年 数理統計 問 4 解答例

[1]

(1)

$$X = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

(2)

$$X = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

$$f(\boldsymbol{\theta}) = \boldsymbol{\epsilon}^2 = (\boldsymbol{x} - X\boldsymbol{\theta})^2$$

とするとき $f'(\boldsymbol{\theta}) = X'(\boldsymbol{x} - X\boldsymbol{\theta})$ なので求める正規方程式は $X'(\boldsymbol{x} - X\boldsymbol{\theta}) = 0$

[2]

[1] より $\hat{\boldsymbol{\theta}} = (X'X)^{-1}X'\boldsymbol{x}$

(1)

$$X'X = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix} = 2I$$

$(X'X)^{-1} = (1/2)I$

$$\hat{\boldsymbol{\theta}} = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \\ x_{10} \end{pmatrix}$$

$\hat{\theta_1} = \frac{X_1+X_2}{2}$ より

$$V(\hat{\theta_1}) = \frac{1}{4}V(X_1+X_2) = \frac{1}{4}(V(X_1)+V(X_2)) = \frac{\sigma^2}{2}$$

以下同様にして、$V(\hat{\theta_i}) = \sigma^2/2$

(2)

$$X'X = \begin{pmatrix} 4 & 1 & 1 & 1 & 1 \\ 1 & 4 & 1 & 1 & 1 \\ 1 & 1 & 4 & 1 & 1 \\ 1 & 1 & 1 & 4 & 1 \\ 1 & 1 & 1 & 1 & 4 \end{pmatrix}$$

与えられた公式より、

$$(X'X)^{-1} = \frac{1}{24}\begin{pmatrix} 7 & -1 & -1 & -1 & -1 \\ -1 & 7 & -1 & -1 & -1 \\ -1 & -1 & 7 & -1 & -1 \\ -1 & -1 & -1 & 7 & -1 \\ -1 & -1 & -1 & -1 & 7 \end{pmatrix}$$

$$\hat{\boldsymbol{\theta}} = \frac{1}{24}\begin{pmatrix} 6 & 6 & 6 & 6 & -2 & -2 & -2 & -2 & -2 & -2 \\ 6 & -2 & -2 & -2 & 6 & 6 & 6 & -2 & -2 & -2 \\ -2 & 6 & -2 & -2 & 6 & -2 & -2 & 6 & 6 & -2 \\ -2 & -2 & 6 & -2 & -2 & 6 & -2 & 6 & -2 & 6 \\ -2 & -2 & -2 & 6 & -2 & -2 & 6 & -2 & 6 & 6 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \\ x_{10} \end{pmatrix}$$

$$\hat{\theta_1} = \frac{1}{24}(6X_1+6X_2+6X_3+6X_4-2X_5-2X_6-2X_7-2X_8-2X_9-2X_{10})$$

よって

$$V(\hat{\theta_1}) = \frac{1}{24^2}(36+36+36+36+4+4+4+4+4+4)\sigma^2 = \frac{7}{24}\sigma^2$$

以下同様にして、$V(\hat{\theta_i}) = \frac{7}{24}\sigma^2$

[3]

問題は 5 群の平均値が全て等しいかどうかの検定であり、F 検定が使える。

群内変動の不偏分散は

$$F_0 = \frac{||\boldsymbol{x} - X\hat{\boldsymbol{\theta}}||^2}{10-5} = \frac{||\boldsymbol{x} - X\hat{\boldsymbol{\theta}}||^2}{5} \sim \chi^2_5$$

$\bar{\theta} = (\hat{\theta_1}+\hat{\theta_2}+\hat{\theta_3}+\hat{\theta_4}+\hat{\theta_5})/5$, $\bar{\boldsymbol{\theta}} = (\bar{\theta},\bar{\theta},\bar{\theta},\bar{\theta},\bar{\theta})'$ とおく。H_0 を仮定の下で $\theta_0 = \theta_1 = \cdots = \theta_5$ とおくと、$\bar{\boldsymbol{\theta}} = (\theta_0,\theta_0,\theta_0,\theta_0,\theta_0)' = \boldsymbol{\theta}_0$ となるので、群間変動の不偏分散は

$$F_1 = \frac{||X\hat{\boldsymbol{\theta}} - X\bar{\boldsymbol{\theta}}_0||^2}{5-1} = \frac{||X\hat{\boldsymbol{\theta}} - X\boldsymbol{\theta}_0||^2}{4} \sim \chi^2_4$$

以上より求める F 統計量は (1) (2) とも

$$F = \frac{F_1}{F_0} = \frac{||X\hat{\boldsymbol{\theta}} - X\boldsymbol{\theta}_0||^2/4}{||\boldsymbol{x} - X\hat{\boldsymbol{\theta}}||^2/5} \sim F_5^4$$

[4]

対立仮説の下での F 統計量も以下で表せる。

$$F = \frac{F_1}{F_0} = \frac{||X\hat{\boldsymbol{\theta}} - X\bar{\boldsymbol{\theta}}||^2/4}{||\boldsymbol{x} - X\hat{\boldsymbol{\theta}}||^2/5}$$

今、(1) (2) ともに

$$\frac{||\boldsymbol{x} - X\hat{\boldsymbol{\theta}}||^2}{\sigma^2}$$

がカイ 2 乗分布に従うことから、

$$Y = \frac{||X\hat{\boldsymbol{\theta}} - X\bar{\boldsymbol{\theta}}||^2}{\sigma^2}$$

の分布の非心度を求めればよい。

(1)

$$Y = \frac{||X\hat{\boldsymbol{\theta}} - X\bar{\boldsymbol{\theta}}||^2}{\sigma^2} = \frac{2\sum_{i=1}^5 |\hat{\theta}_i - \bar{\theta}|^2}{\sigma^2}$$

今、

$$Y_i = \frac{\sqrt{2}(\hat{\theta}_i - \bar{\theta})}{\sigma}$$

とおくと $V(Y_i) = 2V(\hat{\theta}_i)/\sigma^2 = 1$ より

$$Y_i \sim N(\frac{\sqrt{2}(\theta_i - \bar{\theta})}{\sigma}, 1)$$

$Y = \sum_{i=1}^5 Y_i^2$ なので、Y の非心度は以下となる。

$$\sum_{i=1}^5 (\frac{\sqrt{2}(\theta_i - \bar{\theta})}{\sigma})^2 = \frac{2}{\sigma^2}\sum_{i=1}^5 (\theta_i - \bar{\theta})^2$$

(2)

$$Y = \frac{||X\hat{\boldsymbol{\theta}} - X\bar{\boldsymbol{\theta}}||^2}{\sigma^2} = \frac{4\sum_{i=1}^5 |\hat{\theta}_i - \bar{\theta}|^2}{\sigma^2}$$

今、

$$Y_i = \frac{2(\hat{\theta}_i - \bar{\theta})}{\sigma}$$

とおくと $V(Y_i) = 4V(\hat{\theta}_i)/\sigma^2 = \frac{7}{6}$ より

$$Y_i \sim N(\frac{2(\theta_i - \bar{\theta})}{\sigma}, \frac{7}{6})$$

$Y = \sum_{i=1}^5 Y_i^2$ なので、Y の非心度は以下となる。

$$\sum_{i=1}^5 (\frac{2(\theta_i - \bar{\theta})}{\sqrt{\frac{7}{6}}\sigma})^2 = \frac{24}{7\sigma^2}\sum_{i=1}^5 (\theta_i - \bar{\theta})^2$$

[5]

(2) の測定方法による推定量 $\hat{\boldsymbol{\theta}}^{(2)}$ と、(1) の測定方法による推定量 $\hat{\boldsymbol{\theta}}^{(1)}$ を比較すると、どちらも不偏推定量になっているが、$V(\hat{\boldsymbol{\theta}}^{(2)}) < V(\hat{\boldsymbol{\theta}}^{(1)})$ であるため $\hat{\boldsymbol{\theta}}^{(2)}$ の方が有効である。さらに対立仮説を仮定した F 統計量の非心度も $\hat{\boldsymbol{\theta}}^{(2)}$ を使った方が $\hat{\boldsymbol{\theta}}^{(1)}$ を使った方が大きい。これらの点から (2) の測定方法の方が効率的な測り方と言える。

2014 年 数理統計 問 5

毎週 1 回同じ時間帯に放送されるテレビ番組の 4 回の視聴において、番組を全回視聴したした n 人を無作為に選び、番組に満足したかどうかのアンケートを行ない、満足した回数を集計して下記の表を得たとする。

回数	0	1	2	3	4	合計
人数	n_0	n_1	n_2	n_3	n_4	n

このとき、各人数 $n_j(j=0,1,2,3,4)$ は試行回数 n, 確率 q_j $(q_0+\cdots q_4=1)$ の多項分布に従う。特に、それぞれの人について、各番組に対して満足するか否かは互いに独立で、満足する確率はすべて一定値 p であるとすると、多項確率 q_j は、試行回数 4、二項確率 p の二項分布により

$$q_j = q_j(p) ={}_4 C_j p^j (1-p)^{4-j},\ j=0,\cdots,4 \tag{1}$$

と表される。この二項分布モデルのデータへの適合を調べたいものとする。以下の各問に答えよ。

[1]　モデル (1) の下での p の尤度関数を求め、p の最尤推定値が $\hat{p}=\sum_{j=0}^{4} jn_j/(4n)$ となることを示せ。

[2]　モデル (1) の適合度検定のための適合度カイ二乗統計量とその自由度を求めよ。

[3]　モデル (1) の適合度検定のための尤度比統計量を求めよ。

[4]　モデル (1) のデータへの適合を判断する手順を示せ。モデルがデータに適合しないという結果になった場合、何が考えられるか。

2014 年 数理統計 問 5 ポイント

- 尤度比検定と適合度のカイ 2 乗検定 (2013-Q5)

2014 年 数理統計 問 5 解答例

[1]

尤度関数は $L=\prod_{j=0}^{4} q_j^{n_j}$ であり、対数尤度関数は以下である。

$$l=\sum_{j=0}^{4} n_j \log q_j = \sum_{j=0}^{4} n_j(\log_4 C_j + j\log p + (4-j)\log(1-p))$$

$$l' = \sum_{j=0}^{4} n_j\left(\frac{j}{p}-\frac{4-j}{1-p}\right) = \sum_{j=0}^{4} n_j \frac{j-4p}{p(1-p)}$$

よって $\sum_{j=0}^{4} jn_j = \sum_{j=0}^{4} 4n_j\hat{p} = 4n\hat{p}$。以上より、$\hat{p}=\sum_{j=0}^{4} jn_j/(4n)$

[2]

$$Y=\sum_{j=0}^{4}\frac{(n\hat{q}_j - n_i)^2}{n\hat{q}_j}$$

自由度は 5 - 1 - 1 = 3 である。

[3]

尤度関数は $L = \prod_{j=0}^{4} q_j^{n_j}$ なので、モデルが適合するという帰無仮説の下での最大尤度は $L_0 = \prod_{j=0}^{4} \hat{q_j}^{n_j}$ となる。一方、対立仮説（モデルを仮定しない）の場合、尤度関数を最大化する q_j は $q_j = n_j/n$ であり、最大尤度は $L_0 = \prod_{j=0}^{4} (\frac{n_j}{n})^{n_j}$ となる。よって尤度比は以下となる。

$$L = \frac{L_1}{L_0} = \prod_{j=0}^{4} (\frac{n_j}{n\hat{q_j}})^{n_j}$$

よって尤度比統計量 Y' は

$$Y' = 2\log L = 2\sum_{j=0}^{4} n_j \log(\frac{n_j}{n\hat{q_j}})$$

であり、これは自由度 $4-1=3$ のカイ二乗分布に従う。

[4]

まず有意水準を定め、次に Y あるいは Y' が自由度 3 のカイ二乗分布に従うことを利用して棄却域を定める。次に観測データから Y や Y' の実現値を求め、それが棄却域に属するかどうかでモデルが適合しているかどうかを判定する。棄却された場合はモデルが適合していないことを意味する。

もしも棄却された場合、設定したモデルでは各番組に対する満足する確率が同じ値であるが、この確率が番組に依存していることが考えられる。