数理統計学演習問題

□□□□□ □□□□□□ □□□□□□ □□□□□□□□□□□□ □□□□□

2013 年 8 月 16 日版

ファイル名：ash-stat322.tex

フォルダ：/統計学テキスト/stat-by-Ash

参考：http://www.math.uiuc.edu/~r-ash/Stat.html

□□□□□ □□□□□□ □□□□□□ □□□□□□□□□□□□ □□□□□

目次

第 1 章

確率と確率分布

1.1 数え上げの手法

順列と組み合わせなどの基本事項: n, r などを自然数 $\{1, 2, 3, \cdots\}$ とします。

べき乗 $n^r = \overbrace{n \times n \times \cdots \times n}^{r}$

階乗 $n! = n(n-1)\cdots 3 \cdot 2 \cdot 1, \quad 0! = 1$

減少乗積 (順列) $(n)_r = n(n-1)\cdots(n-r+2)(n-r+1)$

増加乗積 $[n]^r = n(n+1)\cdots(n+r-2)(n+r-1) = (n+r-1)_r$

2 項係数：$\dbinom{n}{r} = \dfrac{n!}{r!(n-r)!} = \dfrac{(n)_r}{r!} = \dfrac{n(n-1)\cdots(n-r+1)}{r!} = \dbinom{n}{n-r}$

$$(a+b)^2 = (a^2+ab)+(ab+b^2) = a^2+2ab+b^2$$
$$(a+b)^3 = (a^2+2ab+b^2)(a+b) = a^3+3a^2b+3ab^2+b^3$$
$$(a+b)^4 = a^4+4a^3b+6a^2b^2+4ab^3+b^4$$

多項式の展開：

$$\begin{aligned}(a+b+c)^2 &= (a^2+ab+ac)+(ab+b^2+bc)+(ac+bc+c^2)\\ &= a^2+b^2+c^2+2ab+2bc+2ac\end{aligned}$$

$$\begin{aligned}&(a+b+c)^3\\ &= \tfrac{3!}{3!0!0!}a^3+\tfrac{3!}{0!3!0!}b^3+\tfrac{3!}{0!0!3!}c^3+\tfrac{3!}{2!1!0!}a^2b+\tfrac{3!}{1!2!0!}ab^2+\tfrac{3!}{1!0!2!}ac^2+\tfrac{3!}{2!0!1!}a^2c+\tfrac{3!}{0!1!2!}bc^2+\tfrac{3!}{0!2!1!}b^2c+\tfrac{3!}{1!1!1!}abc\\ &= a^3+b^3+c^3+3a^2b+3ab^2+3ac^2+3a^2c+3bc^2+3b^2c+6abc\end{aligned}$$

条件つきの和計算 $\sum_{i<j} a_{ij} = a_{12}+a_{13}+a_{14}+a_{15}+a_{16}+\cdots$
$$+a_{23}+a_{24}+a_{25}+a_{26}+a_{27}+\cdots$$
$$+a_{34}+a_{35}+a_{36}+a_{37}+a_{38}+\cdots$$
$$+a_{45}+a_{46}+a_{47}+a_{48}+a_{49}+\cdots$$

□□□□□ 練習問題 □□□□□

問 1.1.1
$(a+b)^5$ を求めよ。

問 1.1.2
$\sum_{i<j}(i-j)^2$ を求めよ。

問 1.1.3
(1) $\sum_{i=1}^{n} i^2$ (2) 増加乗積 $\sum_{i=1}^{n}[i]^2$ (3) 増加乗積 $\sum_{i=1}^{n}[i]^r$ を求めよ。

1.2　確率と確率変数

確率のまとめ ３つの構成要素で定める確率空間 $(\Omega, \mathcal{F}, P)$:

標本空間 Ω;　起りえる結果の集合すべて。有限集合、可算集合、非可算集合
$\{\omega_i; i \in \mathcal{Z} = \{1, 2 \cdots\}\}, \{\omega_\alpha; \lambda \in \Lambda \subset \mathcal{R}\}$

事象の集まり (族)$\mathcal{F}$;　和事象、積事象、補事象, $A, B \in \mathcal{F} \Rightarrow A \cup B, A \cap B, A^c = \overline{A} = \Omega \setminus A \in \mathcal{F}$
$A_i \in \mathcal{F}; i = 1, 2, \cdots \Rightarrow \cup_{i=1}^{\infty} A_i \in \mathcal{F}$

確率 (確率測度) P;　$0 \le P(A) \le 1$, (事象から数値を対応させる)
$P(A \cup B) = P(A) + P(B) - P(A \cap B), A, B \in \mathcal{F}$

確率 (probability) と分布関数 (distribution function)：確率変数 (random variable) X の分布 (distribution), 密度 (density) とは、

$$
\begin{aligned}
&F(x) = F_X(x) = P(X \le x) \quad \text{値 } x \text{ 以下 (等号含む) となる結果事象の確率}\\
&P(a < X \le b) = P\left(\{X \le b\} \cap \overline{\{X \le a\}}\right) = F(b) - F(a), \ (a < b)\\
&\quad \text{ここで } \{a < X \le b\} := \{\omega \in \Omega \,|\, a < X(\omega) \le b\} \in \mathcal{F} \text{ などを表す事象}\\
&\text{連続型密度 (微分)} \quad f(x) = f_X(x) = \frac{d}{dx} F_X(x) \quad F_X(x) \text{ の導関数}\\
&\qquad f_X(x)\, dx = dF_X(x) = \lim_{h \downarrow 0}\{F(x) - F(x-h)\}\\
&\text{離散型密度 (差分)} \quad p(x) = p_X(x) = P(X = x)\\
&\qquad \text{ただし } p(x) = F(x) - F(x-0) = F(x) - \lim_{h \downarrow 0} F(x-h) \text{ (点 } x \text{ の左側極限との差)}
\end{aligned}
\tag{1.2.1}
$$

期待値 (expectation)：絶対収束とは $E|X| = E[X^+] + E[X^-] < \infty$ という絶対値の期待値が有限な値となる場合をいう。ここで $a^+ = \max\{a, 0\}, a^- = \max\{-a, 0\}$ とする。もし非負値、$X \ge 0$ ならば、

$$
\begin{aligned}
E[X] &= \int_\Omega X(\omega) dP(\omega)\\
&= \int_0^\infty x\, dF_X(x) = \int_0^\infty (1 - F_X(x))\, dx = \begin{cases} \displaystyle\int_0^\infty x\, f_X(x) dx \\ \displaystyle\sum_{x_i \ge 0} x_i\, p_X(x_i) \end{cases}
\end{aligned}
\tag{1.2.2}
$$

もし変数の値が変化する領域が非負のみに限らなければ、場合分けし分割してから計算すればよい。

確率変数 X から関数 $y = h(x)$ によって定められる確率変数 $h(X)$ の期待値は

$$
E[h(X)] = \int_{-\infty}^{\infty} h(x)\, dF_X(x) = \begin{cases} \displaystyle\int_{-\infty}^{\infty} h(x)\, f_X(x)\, dx & \text{連続型} \\ \displaystyle\sum_i h(x_i) p_X(x_i) & \text{離散型} \end{cases}
$$

結合 (joint) 分布（同時分布）と結合密度 (同時密度)：（独立な場合は積となる）

$$
\begin{aligned}
&F_{X,Y}(x, y) = P(\{X \le x\} \cap \{Y \le y\}) = P(X \le x, Y \le y),\\
&f_{X,Y}(x, y) = \frac{\partial^2}{\partial x\, \partial y} F_{X,Y}(x, y) \quad x, y \text{ についての偏微分したもの}
\end{aligned}
\tag{1.2.3}
$$

周辺密度 (marginal density)：(２変量から各変量の分布へ)

$$
f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx \tag{1.2.4}
$$

ひとつの分布 X から派生する新たな分布 Y を求めたい。$F_X(x) = P(X \le x) = P(Y \le y) = F_Y(y)$ の関係、$f_Y\, dy = f_X\, dx$ を定める。微分積分における逆関数やその微分がキーポイントとなる。

単変量の変換：与えられた確率変数 X の密度関数を $f_X(x)$、分布関数を $F_X(x) = P(X \le x)$ とおく。与えられた関数 $y = h(x)$ が逆関数 $x = h^{-1}(y)$ をもつとするとき、変換された確率変数 $Y = h(X)$ の密度関数 $f_Y(y)$ と分布関数 $F_Y(y)$ は

$$f_Y(y) = f_X(h^{-1}(y))\Big|\frac{1}{(dy/dx)_{x=h^{-1}(y)}}\Big| \tag{1.2.5}$$

便宜的な記法で $Y = h(X)$ の逆変換 $X = h^{-1}(X)$ だから

$$f_Y(y)dy = f_X(x)dx = f_X(x)\frac{1}{dy/dx}dy = f_X(h^{-1}(y))\Big|\frac{1}{(dy/dx)_{x=h^{-1}(y)}}\Big| \tag{1.2.6}$$

と書いてほうが理解しやすいかもしれない。積分を変換して $t = h(x),\, x = h^{-1}(t)$, また逆関数の微分公式で $dx/dt = \dfrac{1}{dt/dx}, \Big|\dfrac{dx}{dt}\Big| = \Big|\dfrac{1}{(dt/dx)_{x=h^{-1}(t)}}\Big|$ とするから、

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P\left(h(X) \le y\right) \\ &= \int_{\{x\,;\,h(x)\le y\}} f_X(x)\,dx = \int_{\{t\,;\,t\le y\}} f_X(h^{-1}(t))\,\Big|\frac{dx}{dt}\Big|\,dt \\ &= \int_{\{t\,;\,t\le y\}} f_X(h^{-1}(t))\,\Big|\frac{1}{(dt/dx)_{x=h^{-1}(t)}}\Big|\,dt \end{aligned} \tag{1.2.7}$$

2 変量の変換：$(U,V) = (h(X,Y), k(X,Y)),\ (X,Y) = (h^{-1}(U,V), k^{-1}(U,V))$ ならば、つまり $\begin{cases} U = h(X,Y) \\ V = k(X,Y) \end{cases},\quad \begin{cases} X = h^{-1}(U,V) \\ Y = k^{-1}(U,V) \end{cases}$ ならば、同様に $f_{U,V}dudv = f_{X,Y}dxdy$、$f_{U,V} = f_{X,Y}\dfrac{1}{dudv/dxdy}$ と考え、

$$f_{U,V}(u,v) = f_{X,Y}\left(h^{-1}(u,v), k^{-1}(u,v)\right)\Big|\frac{1}{(\partial(u,v)/\partial(x,y))_{(x,y)=(h^{-1}(u,v),k^{-1}(u,v))}}\Big| \tag{1.2.8}$$

ここで $(\partial(u,v)/\partial(x,y)) = \dfrac{\partial(u,v)}{\partial(x,y)}\Big|_{(x,y)=(x^{-1}(u,v),y^{-1}(u,v))} = \det\begin{vmatrix} u_x & u_y \\ v_x & v_y \end{vmatrix}_{(x,y)=(x^{-1}(u,v),y^{-1}(u,v))}$

2 変量の場合にはこのように変数が増えるが、基本的には同じ計算で、$f_{U,V}(u,v)\,dudv = f_{X,Y}(x,y)\,dxdy$ と考え、$\dfrac{dudv}{dxdy}$ がヤコビアンといわれるものに相当する。密度関数の積分が分布関数であるから

$$\begin{aligned} F_{U,V}(u,v) &= P(U \le u,\, V \le v) = \iint_{\{w\le u, z\le v\}} f_{U,V}(w,z)dwdz = \iint_{\{s\le u, t\le v\}} f_{X,Y}(s,t)\,dsdt \\ &= \int_{\{s\le u,\, t\le v\}} f_{X,Y}\left(h^{-1}(w,z), k^{-1}(w,z)\right)\Big|\frac{1}{(\partial(w,z)/\partial(s,t))_{(s,t)=(h^{-1}(w,z),k^{-1}(w,z))}}\Big|dsdt \end{aligned} \tag{1.2.9}$$

２変数の期待値： $(X,Y) \sim f_{X,Y}(x,y)$ のとき、

$$E\left[g(X)\,h(Y)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)\,h(y)\ f_{X,Y}(x,y)\,dxdy \tag{1.2.10}$$

もし独立ならば、$f_{X,Y}(x,y) = f_X(x)\,f_Y(y), \forall x,y$ より、

$$E\left[g(X)\,h(Y)\right] = E[g(X)]\,E[h(Y)]$$

なぜならば、$E\left[g(X)\,h(Y)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)\,f_X(x)f_Y(y)\,dxdy = \int_{-\infty}^{\infty} g(x)f_X(x)\,dx \times \int_{-\infty}^{\infty} h(y)f_Y(y)\,dy$ であるから。

条件つき期待値：条件つき密度をもちいて、条件つき平均と条件つき分散を定める。これらの値は確率変数 $Y = y$ の値によって定まるものとなる。条件つき平均とは

$$E[X\,|\,Y=y] = \begin{cases} \displaystyle\int_{-\infty}^{\infty} x\,f_{X|Y}(x|y)\,dx \\ \displaystyle\sum_i x_i\,p_{X|Y}(x_i\,|\,y) \end{cases} \tag{1.2.11}$$

ここで 条件つき密度とは、

$$f_{X|Y}(x\,|\,y)=\frac{f_{X,Y}(x,y)}{f_Y(y)},\ f_Y(y)\neq 0,\qquad p_{X|Y}(x\,|\,y)=\frac{p_{X,Y}(x,y)}{p_Y(y)},\ p_Y(y)\neq 0$$

とする。また

$$E[h(X,Y)\,|\,Y=y]=\begin{cases}\displaystyle\int_{-\infty}^{\infty}h(x,y)\,f_{X|Y}(x|y)\,dx\\ \displaystyle\sum_i h(x_i,y)\,p_{X|Y}(x_i|y)\end{cases}$$

となり、条件つき分散は、

$$\mathrm{var}(X\,|\,Y=y)=E[X^2\,|Y=y]-(E[X\,|\,Y=y])^2 \tag{1.2.12}$$

と定める。条件つきの部分の表し方では $E[X\,|\,Y=y]$ は y の関数であり、y の変化とみなすことで、これを確率変数と考えて、単に $E[X\,|\,Y]$ として、事象 $\{Y=y\}$ の上では、値 $E[X\,|\,Y=y]$ をとるような確率変数とし考える。

$$E[X\,|\,Y]\ =\ E[X\,|\,Y=y]\quad \text{on}\quad \{Y=y\}$$

条件つき平均 $\mu_{X|Y}$, 条件つき分散 $\sigma^2_{X|Y}$ は確率変数として Y の表れ方によって変化するものである。

□□□□□　練習問題　□□□□□

問 1.2.1
X,Y,Z を独立同一分布 (iid と略す) にしたがう確率変数とする。その密度を $f(x)=6x^5,(0\le x\le 1),=0($その他) とするとき、$\max\{X,Y,Z\}$ の分布関数と密度関数を求めよ。

問 1.2.2
X,Y は独立で、同じ密度 $f(x)=e^{-x},x\ge 0$ をもつとする。$Z=Y/X$ の分布を求めよ。以降では「分布を求めよ」とは、分布関数または密度関数を求めよという意味である。

問 1.2.3
X は離散型確率変数で、値 $x_1,x_2,\cdots,x_n$ をそれぞれ $1/n$ の確率でとる。任意の実数値関数 g に対して $Y=g(X)$ とおくとき、Y の確率密度関数 $p_Y(y)=P(Y=y)$ を g,x で表せ。

問 1.2.4
確率変数 X は区間 $[0,b]$ 上で密度関数 $f(x)=ax^2$ をもつ。$Y=X^3$ の密度関数をもとめよ。

問 1.2.5
コーシー分布とは密度関数が $f(y)=\dfrac{1}{\pi(1+y^2)}$, $-\infty<y<\infty$ で与えられるもの。もし X が $[-\pi/2,\pi/2]$ 上の一様分布であれば、$Y=\tan(X)$ はコーシー分布となることを示せ。

問 1.2.6
　結合分布が $f(x_1,x_2)=2e^{-x_1}e^{-x_2}, 0<x_1<x_2<\infty; f(x_1,x_2)=0$(その他) で与えられている。変換 $Y_1=2X_1$, $Y_2=X_2-X_1$ を行うとき、この結合分布をもとめよ。さらに独立であることを確かめよ。

問 1.2.7
　結合分布が $f(x_1,x_2)=8x_1x_2, 0<x_1<x_2<1; f(x_1,x_2)=0$(その他) で与えられている。変換 $Y_1=X_1/X_2$, $Y_2=X_2$ を行うとき、この結合分布をもとめよ。

問 1.2.8
　3 個の確率変数 $X_i, i=1,2,3$ が独立で同じ密度 $f(x)=e^{-x}$, $x>0; f(x)=0$(その他) で与えられている。変換 $Y_1=X_1/(X_1+X_2)$, $Y_2=(X_1+X_2)/(X_1+X_2+X_3)$, $Y_3=X_1+X_2+X_3$ を行うとき、この結合分布をもとめよ。この分布は独立になるだろうか？

補足 (1) Cauchy 分布には、平均、分散などが存在しない。両裾の確率が比較的高くて、積分値が有限確定値にはならない。しかし中央値や最頻値はいわゆる中心として存在する。不思議な性質がある。２つの独立な標準正規分布の比率（割合）の分布はコーシー分布となるし、Cauchy 分布の逆数も Cauchy 分布にしたがう。

イタリアの数学研究者による「Witch of Agnesi(アニージョの魔女)」と呼ばれる曲線である。彼女は貿易で利益をあげた裕福な実業家かつ教養ある名士の娘（父は３度の結婚、２１人の子）で、フランス、ラテン、ギリシャ、ヘブライ、ドイツ、スペイン語など７ケ国語に堪能という優れた才能をもち、微分積分法の統一的な解説をした教科書を書き、オイラーの影に隠れてしまうが何か国語にも訳され、ヨーロッパでは人気の教科書であったという。魔女と呼ばれるのは、誤訳のせいである。１７０３年グイド・グランディは、ベルソニアという名前をこのグラフにつけたが、ラテン語の「ひっくり返す」を意味する vertere に由来するが、イタリア語系の versiera をもちいた。イギリスのケンブリッジ大学教授、コルソンは、イタリア語から英訳するとき、la versiera を l'avversiera 「悪魔の妻、神に逆らう女」と取り違え、「魔女」と訳した。この間違いは、曲線のもつ不可思議な性質から、英語圏では残ってしまった。本来ならば、アニェージの三次曲線とかアニェジエンヌと呼ぶべきである。１７５２年父が亡くなると、貧しい人々への奉仕にその後の人生を捧げた。イタリアの数学界での評判も確立していたが、法王から送られた金のメダルや皇后マリア・テレジアからの贈り物のダイヤや水晶の箱をうり、ミラノに養老院を設立した。晩年は視力、聴力も失い、水腫にもなり、健康にすぐれず、自身の老人施設で無一文のまま、１７９９年１月９日８１歳で亡くなった。[1] Shirley B. Gray;"Maria Gaetana Agnesi", Calfornia State University, LA, [2] http://instructional1.calstatela.edu/sgray/agnesi. また Wikipedia:Witch of Agnesi を参照。

２項係数：

$$\binom{n}{r} = \binom{n}{n-r} = \frac{n!}{r!\,(n-r)!} = \frac{n(n-1)\cdots(n-r+2)(n-r+1)}{r!} = \frac{(n)_r}{r!}$$
$$= \frac{n(n-1)\cdots(r+2)(r+1)}{(n-r)!} = \frac{(n)_{n-r}}{(n-r)!}$$

ただし、$n = 1, 2, 3, \cdots, r = 1, 2, \cdots$

実数 α に対して $\binom{\alpha}{r} = \dfrac{\alpha(\alpha-1)\cdots(\alpha-r+2)(\alpha-r+1)}{r!} = \dfrac{(\alpha)_r}{r!}$ とおく。

２項展開：

(1) $(x+1)^n = \binom{n}{0}x^n + \binom{n}{1}x^{n-1} + + \binom{n}{2}x^{n-2} + \cdots + \binom{n}{n-1}x^1 + \binom{n}{n}x^0$

(2) $(x+1)^\alpha = x^\alpha + \binom{\alpha}{1}x^{\alpha-1} + + \binom{\alpha}{2}x^{\alpha-2} + \cdots$

(3) $(x+y)^n = \sum_{\{(i,j):i+j=n\}} \binom{n}{i,j} x^i y^j$　ただし、$\binom{n}{i,j} = \binom{n}{i,n-i} = \binom{n}{i}$

３項係数：$\binom{n}{r_1,\ r_2,\ r_3} = \dfrac{n!}{r_1!\,r_2!\,r_3!}$,　$r_1 + r_2 + r_3 = n$

３項展開：$(x+y+z)^n = \sum_{\{(i,j,k):i+j+k=n\}} \binom{n}{i,\ j,\ k} x^i y^j z^k$

1.3　確率分布、積率母関数

確率分布: 密度関数あるいは確率密度の名称、分布の略記号、関数形ととり得る値

1. ベルヌーイ (Bernoulli) 分布 : $\mathrm{Bern}(p) \sim p^x(1-p)^{1-x}$, $(x = 0, 1)$
2. ２項 (Binomial) 分布: $\mathrm{Binom}(n, p) \sim \binom{n}{x} p^x (1-p)^{n-x}$, $(x = 0, 1, \cdots, n)$

3. ポアソン (Poisson) 分布: $\mathrm{Poi}(\lambda) \sim \dfrac{\lambda^x}{x!}\, e^{-\lambda}, (x = 0, 1, \cdots)$
4. 正規 (Normal) 分布: $\mathrm{N}(\mu, \sigma^2) \sim \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, (-\infty < x < \infty)$
5. 標準正規 (Standard Normal) 分布: $\mathrm{N}(0, 1) \sim \dfrac{1}{\sqrt{2\pi}} e^{-x^2/2}, (-\infty < x < \infty)$
6. 一様 (Uniform) 分布: $\mathrm{U}([a, b]) \sim \dfrac{1}{b-a}, (a < x < b)$
7. 一様分布 (離散型): $\mathrm{U}(\{1, 2, \cdots, m\}) \sim \dfrac{1}{m}, (x \in \{1, 2, \cdots, m\})$
8. 指数 (Exponential) 分布: $\mathrm{Expo}(\lambda) \sim \dfrac{1}{\lambda} e^{-x/\lambda}, (x > 0)$
9. 幾何 (Geometric) 分布: $\mathrm{Geom}(p) \sim pq^{x-1}, (x = 1, 2, \cdots)$
10. ガンマ (Gamma) 分布: $\mathrm{Gam}(\alpha, \theta) \sim \dfrac{x^{\alpha-1}}{\Gamma(\alpha)\theta^\alpha} e^{-x/\theta},\ (x > 0)$
11. カイ 2 乗 (Chi-square) 分布: $\mathrm{Chi}^2(m) = \chi^2(m) = \mathrm{Gam}(m/2, 2) \sim \dfrac{x^{(m-2)/2}}{\Gamma(m/2)2^{m/2}}\, e^{-x/2},\ (x > 0)$
12. スチューデントの t 分布: $\mathrm{T}(n) \sim \dfrac{\Gamma((n+1)/2))}{\Gamma(1/2)\Gamma(n/2)} \dfrac{1}{\sqrt{n}} (1 + x^2/n)^{-(n+1)/2}, (-\infty < x < \infty)$
13. 負の 2 項 (Negative Binomial) 分布: ($x = 0$ の値から始まるとき) $\mathrm{NegBinom}(p, r) \sim \dbinom{r+x-1}{x} p^r q^x = \dbinom{-r}{x} p^r (-q)^x = \dfrac{[r]^x}{x!} p^r (-q)^x,\ (x = 0, 1, 2, \cdots)$
14. ($x = r$ の値から始まるとき) $\mathrm{NegBinom}(p, r) \sim \dbinom{x-1}{r-1} p^r q^{x-r},\ (x = r, r+1, r+2, \cdots), q = 1 - p$
15. ベータ (Beta) 分布: $\mathrm{Beta}(\alpha, \beta) \sim \dfrac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}, (0 < x < 1)$
16. コーシー (Cauchy) 分布: $\mathrm{Cauchy} = \mathrm{T}(1) \sim \dfrac{1}{\pi(1+x^2)},\ (-\infty < x < \infty)$

積率母関数（モーメント母関数, Moment Generating Function）：

$$M_X(t) = E[e^{Xt}] = E[\exp(Xt)] = \begin{cases} \sum_k e^{kt} P(X = k) \\ \int e^{xt} f_X(x) dx \end{cases} \tag{1.3.1}$$

2 変量の場合

$$M_{X,Y}(t, s) = E[e^{Xt+Ys}] = E[\exp(Xt + Ys)] \tag{1.3.2}$$

ベクトルの内積をもちいれば、$Xt + Ys = (t,\, s) \cdot \begin{pmatrix} X \\ Y \end{pmatrix}$ であり、積と内積が対応。

1. もし X, Y が独立であれば

$$M_{X,Y}(t, s) = M_X(t)\, M_Y(s) \tag{1.3.3}$$

2. もし X, Y が独立であれば, 線形な一次式: $aX + bY$ は

$$M_{aX+bY}(t) = M_X(at)\, M_Y(bt) \tag{1.3.4}$$

3. もし $X_i, i = 1, 2, \cdots$ が独立で、すべて X と同一分布であれば: $X_1 + X_2 + \cdots + X_n$ は

$$M_{X_1+X_2+\cdots+X_n}(t) = \{M_X(t)\}^n \tag{1.3.5}$$

r 次のモーメント (積率)　$\mu_r = E[X^r], r = 1, 2, \cdots$ との関係は

$$\mu_1 = \left.\frac{dM_X(t)}{dt}\right|_{t=0} = M'_X(0), \qquad \mu_2 = M''_X(0), \qquad \mu_r = M_X^{(r)}(0) \tag{1.3.6}$$

で与えられる。よって平均 μ_X、分散 σ_X^2 は

$$\mu_X = \mu_1, \quad \sigma_X^2 = \mu_2 - (\mu_1)^2 \tag{1.3.7}$$

として求められる。

分布の平均、分散とモーメント母関数：

分布	平均	分散	m g f $M_X(t)$
ベルヌーイ分布 Ber(p)= Binom(1, p)	p	pq $(q = 1-p)$	$pe^t + q$
2 項分布 Binom(n, p)	np	npq	$(pe^t + q)^n$
ポアソン分布 Po(λ)	λ	λ	$\exp[\lambda(e^t - 1)]$
正規分布 N(μ,σ^2)	μ	σ^2	$\exp[\mu t + \frac{\sigma^2}{2}t^2]$
指数分布 Expo(λ)=Gam(1,λ)	λ	λ^2	$1/(1-\lambda t)$
幾何分布 Geom(p)	q/p	q/p^2	$p/(1-qe^t)$
ガンマ分布 Gam(α,θ)	$\alpha\theta$	$\alpha\theta^2$	$(1-\theta t)^{-\alpha}$
カイ 2 乗分布 $\chi^2(r)$=Gam$\left(\frac{r}{2}, 2\right)$	r	$2r$	$(1-2t)^{-r/2}$
負の 2 項分布 NegBinom(r,p)	rq/p	rq/p^2	$\{p/(1-qe^t)\}^r$

定理 1.3.1

(1) $X \sim N(\mu, \sigma^2)$ ならば、$Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$ (分散は b に依存しない)

(2) 独立な $X_1, X_2, \cdots, X_n$ はすべて正規分布 $X_i \sim N(\mu_i, \sigma_i^2)$ ならば、$Y = \sum_i X_i \sim N(\mu, \sigma^2)$, ここで $\mu = \sum_i \mu_i, \sigma^2 = \sum_i \sigma^2$.

(3) 独立な $X_1, X_2, \cdots, X_n$ はすべて同一の正規分布（同じ平均と分散）$X_i \sim N(\mu, \sigma^2)$ ならば、$Y = \sum_i X_i^2 \sim \chi^2(n)$

分布の和 (畳み込み)：条件付き密度 $f_{Y|X}(y|x)$,$f_{X|Y}(x|y)$ が与えられるとき、、２つの確率変数 X, Y： $f_{X,Y}(x,y) = f_{Y|X}(y|x)\, f_X(x) = f_{X|Y}(x|y)\, f_Y(y)$ があるならば、和の分布 $Z = X + Y$ は

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_{Y|X}(z-x|x)\, dx \int_{-\infty}^{\infty} f_{X|Y}(z-y|y) f_Y(y)\, dy \tag{1.3.8}$$

X, Y が独立ならば,

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\, dy \tag{1.3.9}$$

ガンマ (Gamma) 関数: 実数 $\alpha > 0$ に対して、

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx = \lim_{n\to\infty} \frac{n!\, n^\alpha}{\alpha(\alpha+1)\cdots(\alpha+n)} = \int_0^1 \left(ln\left(\frac{1}{x}\right)\right)^{\alpha-1} dx$$

ここで $ln(a)$ は自然対数 $\log_e a$ とする。

(a) $\Gamma(\alpha+1) = \alpha\, \Gamma(\alpha)$　(b) $\Gamma(n+1) = n!, n = 0, 1, 2, \cdots$

(c) $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} = \int_0^\infty x^{-1/2} e^{-x}\, dx = \sqrt{2}\int_0^\infty e^{-x^2/2}\, dx$　(d) $\Gamma(x)\Gamma(1-x) = \dfrac{\pi}{\sin(\pi x)}$

ベータ (Beta) 関数：実数 $a, b > 0$ に対して

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

とくに $B(n,m) = \dfrac{\Gamma(n)\Gamma(m)}{\Gamma(n+m)} = \dfrac{1}{(n+m-1)\binom{n+m-2}{n-1}}$,　$n\binom{n-1}{k-1} = \dfrac{1}{B(k, n-k+1)}$

□□□□□　練習問題　□□□□□

問 1.3.1

X_1, X_2 は独立で、X_1 は自由度 r_1 のカイ 2 乗分布 $\chi^2(r_1)$、$X_1 + X_2$ が自由度 $r\,(r > r_1)$ のカイ 2 乗分布にしたがうとする。このとき、X_2 は自由度 $r_2 = r - r_1$ のカイ 2 乗分布 $\chi^2(r_2)$ にしたがうことを示せ。

問 1.3.2

X_1, X_2 は独立で、X_i はパラメータ $(\alpha_i, \beta_i), i = 1, 2$ のガンマ分布にしたがうとする。正の定数 c_1, c_2 から、$c_1 X_1 + c_2 X_2$ をつくる。$\beta_1 c_1 = \beta_2 c_2$　のとき、この分布を定めよ。

問 1.3.3

$X_i, i = 1, 2, \cdots, n$ は独立で、それぞれの mgf（モーメント母関数）を M_i とするとき、線形結合 $c_1 X_1 + c_2 X_2 + \cdots + c_n X_n$ に対する mgf M を定めよ。

問 1.3.4

$X_i, i = 1, 2, \cdots, n$ は独立で、それぞれパラメータ λ_i のポアソン分布をする。このとき、和 $Y = \sum_i X_i$ もやはりポアソン分布となることを示せ。

問 1.3.5

偏りのないコインを n_1 回投げ、さらに続けて n_2 回投げた。最初の試行で表の出た回数を X_1 とし, X_2 は 2 回目の試行で裏の出た回数とする。mgf の計算でもなく、直感的に $X_1 + X_2$ の分布はどうなるか？ またこの結果を mgf により確かめてみよ。

第 2 章

正規分布とその関連分布

2.1 正規分布

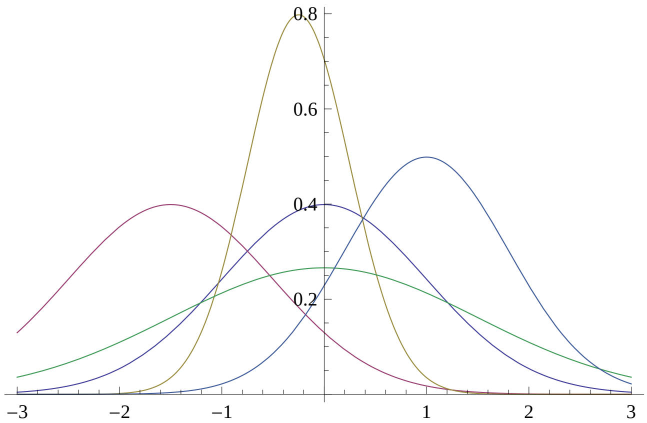
データから定める統計量：

1. 標本の大きさ、サイズ (data size): n, データ：$X_1, X_2, \cdots, X_n$
2. 標本平均 (sample average, sample mean): $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
3. 標本分散 (sample variance): $S_X^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \dfrac{\sum_{i=1}^{n} X_i^2}{n} - \overline{X}^2$
4. 標本不偏分散 (unbiased sample variance): $u_X^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \dfrac{n}{n-1}S_X^2$

確率分布として、正規分布が重要な働きをする。

正規分布 $N(\mu, \sigma^2)$ の密度関数と分布関数、５個のグラフ概形

(i) $\mathrm{N}(0,1)$　(ii) $\mathrm{N}(-1.5,1)$　(iii) $\mathrm{N}(-1/4,1/2)$　(iv) $\mathrm{N}(0,1.5)$　(v) $\mathrm{N}(1,0.8)$



密度関数 $\phi(x) = \dfrac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$,　分布関数 $\Phi(x) = \dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-\infty}^{x} e^{-\frac{1}{2}t^2}dt$

一般の平均 μ , 分散 σ^2 をもつ密度関数 $X \sim N(\mu, \sigma^2)$:

$$f_X(x) = \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

正規分布からの派生される分布 (とくに断らない限り独立を仮定)

● 正規分布の再生性：正規分布の和は正規分布

$$\begin{aligned} &X_i \sim \mathrm{N}(\mu_i, \sigma_i^2), i = 1, 2, \cdots, r \Rightarrow Y = \textstyle\sum_{i=1}^{r} a_i X_i \sim \mathrm{N}(\sum a_i\mu_i, \sum a_i^2\sigma_i^2) \\ &a\mathrm{N}(\mu_1, \sigma_1^2) + b\mathrm{N}(\mu_2, \sigma_2^2) \sim \mathrm{N}(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2) \end{aligned} \tag{2.1.1}$$

- カイ二乗分布:独立な標準正規分布の 2 乗和

$$X_i \sim \mathrm{N}(0,1), i=1,2,\cdots,r \Rightarrow Y=\sum_{i=1}^{r} X_i^2 : \quad \sum_{1:r}(\mathrm{N}(0,1))^2 \sim \chi^2(r) \tag{2.1.2}$$

- t 分布: 標準正規分布と自由度 r のカイ 2 乗分布の平方根の比

$$\left.\begin{array}{l} X_1 \sim \mathrm{N}(0,1) \\ X_2 \sim \chi^2(r) \end{array}\right\} \Rightarrow Y=\frac{X_1}{\sqrt{X_2/r}} : \quad \frac{\mathrm{N}(0,1)}{\sqrt{\chi^2(r)/r}} \sim T(r) \tag{2.1.3}$$

- F 分布: 2 つのカイ 2 乗分布の比

$$\left.\begin{array}{l} X_1 \sim \chi^2(m) \\ X_2 \sim \chi^2(n) \end{array}\right\} \Rightarrow Y=\frac{X_1/m}{X_2/n} : \quad \frac{\chi^2(m)/m}{\chi^2(n)/n} \sim \mathrm{F}(m,n) \tag{2.1.4}$$

- ベータ分布: 2 つのガンマ分布から派生

$$\left.\begin{array}{l} X_1 \sim \Gamma(\alpha,\lambda) \\ X_2 \sim \Gamma(\beta,\lambda) \end{array}\right\} \Rightarrow Y=\frac{X_1}{X_1+X_2} : \quad \frac{\Gamma(\alpha,\lambda)}{\Gamma(\alpha,\lambda)+\Gamma(\beta,\lambda)} \sim \mathrm{Beta}(\alpha,\beta) \tag{2.1.5}$$

- コーシー分布:標準正規分布の比

$$\left.\begin{array}{l} X_1 \sim N(0,1) \\ X_2 \sim N(0,1) \end{array}\right\} \Rightarrow Y=\frac{X_1}{X_2} : \quad \frac{N(0,1)}{N(0,1)} \sim \mathrm{Cauchy} \tag{2.1.6}$$

□□□□□　練習問題　□□□□□

問 2.1.1

$\{X_i\}$ が同じ平均 μ、分散 σ^2 をもち、独立であれば、

(1) $E[\overline{X}]=\mu$　(2) $\mathrm{Var}(\overline{X})=\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}X_i=\frac{\sigma^2}{n}\to 0\,(n\to\infty)$　(3) $E[u_X^2]=\sigma^2$

となることを示せ。ヒント : $E[(X_i-\overline{X})^2]=E[\{(X_i-\mu)-(\overline{X}-\mu)\}^2]=\mathrm{var}(X_i)+\mathrm{var}(\overline{X})-2E[(X_i-\mu)(\overline{X}-\mu)]$, さらに $E[(X_i-\mu)(\overline{X}-\mu)]=E[(X_i-\mu)\frac{1}{n}\sum_i(X_i-\mu)]=\frac{1}{n}E[(X_i-\mu)^2]$ をもちいよ。

問 2.1.2

$\overline{X}$ は $\{X_i+X_j\,;\,i,j\}$ から定められ、u_X^2 は $\{X_i-X_j;i,j\}$ から定められることを示せ。つまり

$$\overline{X}=\frac{1}{\binom{n}{2}}\sum_{i<j}\frac{X_i+X_j}{2}, \quad u_X^2=\frac{1}{\binom{n}{2}}\sum_{i<j}\frac{(X_i-X_j)^2}{2}$$

ここで 2 重和 $\sum_{i<j}=\sum_{\{(i,j);1\le i<j\le n\}}=\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}=\sum_{j=2}^{n}\sum_{i=1}^{j-1}$ の意味。和の個数は $\binom{n}{2}$ 個ある。

問 2.1.3

(1) X,Y はともに独立で、標準正規分布にしたがうとするとき、$Z=X+Y, W=X-Y$ の同時分布を求めよ。またこれらが独立となることを示せ。　(2) この関係式から、正規分布では標本平均と標本分散が独立であることを示せ。

問 2.1.4

n 個の $X_i, i=1,2,\cdots,n$ は iid で正規分布 $N(\mu,\sigma^2)$ にしたがい、$\overline{X}$ を標本平均とする。適当な定数 c で n を大きくし $P(\mu-c<\overline{X}<\mu+c)\ge 0.954$ となるようにしたい。分散 σ^2 と定数 c をもちいて (μ には関係しない、この理由も考えよ)、最小の n をもとめよ。

問 2.1.5
２つのグループ $\{X_i\}(i=1,2\cdots,n_1),\{Y_j\}(j=1,2,\cdots,n_2)$ があり、これらはすべて独立で、$X_i\sim N(\mu_1,\sigma_1^2)$ かつ $Y_i\sim N(\mu_2,\sigma_2^2)$ とし、それぞれの標本平均を $\overline{X},\overline{Y}$ とする。このとき、$P(\overline{X}>\overline{Y})$ の確率はどのようにして求めたらよいか。

問 2.1.6
$X_i\sim N(\mu,\sigma^2),(i=1,2\cdots,n)$ は iid とし、標本分散 $S^2=\frac{1}{n}\sum_i(X_i-\overline{X})^2$ とする。確率 $P(a<S^2<b)$ をどう計算するか述べよ。

問 2.1.7
$X_i\sim N(\mu,\sigma^2),(i=1,2\cdots,n)$ は iid とする。標本分散 $S^2=\frac{1}{n}\sum_i(X_i-\overline{X})^2$ の mgf を計算せよ。これから、S^2 の分布がガンマ分布になることを確かめよ。そのパラメータはどうなるか？

もし表計算ソフトを用いるならば、標準正規分布 (standard normal distribution)、一般正規分布 (normal distribution) に対する (1) 値から確率、 (2) 確率から値　は次の組み込み関数で求められる。

```
(1)  NORM.S.DIST(X，関数形式（0/1））：標準正規の場合
     NORM.DIST(X，平均，標準偏差，関数形式)：一般正規の場合
     書式の引数:
       値 X        分布を評価するための数値を指定。
       関数形式 TRUE("1"に対応) の場合は、(累積) 分布関数の値。
                 FALSE("0"に対応) の場合は、確率密度関数の値。
(2) NORM.S.INV(確率)
    NORM.INV(確率，平均，標準偏差)
    書式の引数: 返される値は分布関数の逆関数による値。
       確率        確率の値を指定。
       平均、標準偏差（分散の平方根）のパラメータ。
```

2.2　t 分布, カイ２乗分布とＦ分布

密度関数と平均、分散はつぎで与えられる。

	$f(x)$	$\mu=E(X)$	$\sigma^2=E(X^2)-\mu^2$
t 分布 $\mathrm{T}(k)\ (k>0)$	$\dfrac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})}\dfrac{1}{\sqrt{k\pi}}\dfrac{1}{\left[1+\frac{x^2}{k}\right]^{(k+1)/2}}$	$\mu=0\ (k>1)$	$\sigma^2=\dfrac{k}{k-2}\ (k>2)$
F 分布 $\mathrm{F}(m,n)\ (m,n>0)$	$\dfrac{1}{\mathrm{B}(\frac{m}{2},\frac{n}{2})}\left(\dfrac{m}{n}\right)^{\frac{m}{2}}\dfrac{x^{(m-2)/2}}{[1+\frac{m}{2}x]^{(m+n)/2}}$	$\dfrac{n}{n-2}\ (n>2)$	$\dfrac{2n^2(m+n-2)}{m(n-2)^2(n-4)}\ (n>4)$
カイ２乗分布 $\chi^2(k)$	$\dfrac{1}{\Gamma(k/2)}\left(\frac{1}{2}\right)^{k/2}x^{k/2-1}e^{-x/2}$	$\mu=k$	$\sigma^2=2k$

もし表計算ソフトを用いるならば、t 分布 (スチューデントの t-distribution) に対する (1) 値から確率、 および (2) 確率から値、は次の組み込み関数で求められる。

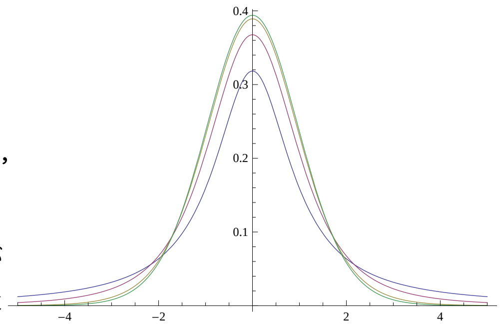
```
(1)  T.DIST(x，自由度，関数形式（0/1））
     書式の引数:
```

値 x　分布を評価するための数値を指定。
自由度　自由度を示す整数を指定。
関数形式 TRUE("1"に対応) の場合は、(累積) 分布関数の値。
FALSE("0"に対応) の場合は、確率密度関数の値。

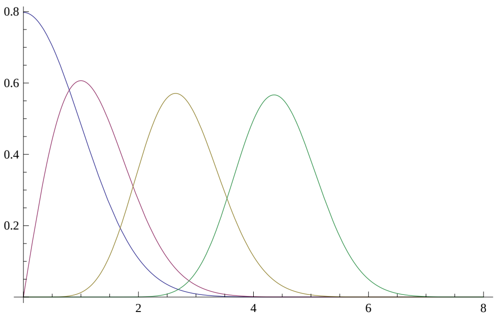
(2) T.INV(確率, 自由度)
書式の引数: 返される値は分布関数の逆関数による値。
確率　確率を指定。
自由度　分布の自由度。

以下ではグラフの概形を描いてみる。

[1] スチューデントの t-分布の概形:
自由度 df {\nu, {1, 3, 10, 20}} 4 本のグラフ
描画範囲 range = {x, -5, 5}
描画命令 Plot[Evaluate@
  Table[PDF[StudentTDistribution[\[Nu]], x],
   {\[Nu], {1, 3, 10, 20}}], {x, -5, 5}



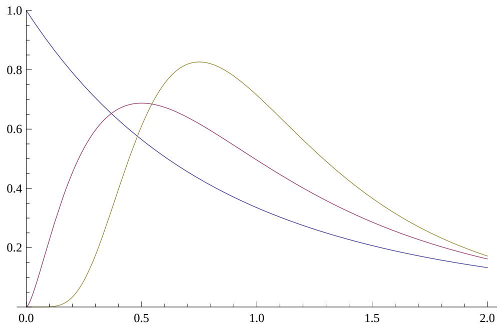
自由度が小さいときには広がり（裾の重みが大きい）があるが、大きくすると正規分布に近づく。自由度が大きいと正規分布表で確率の計算ができる。



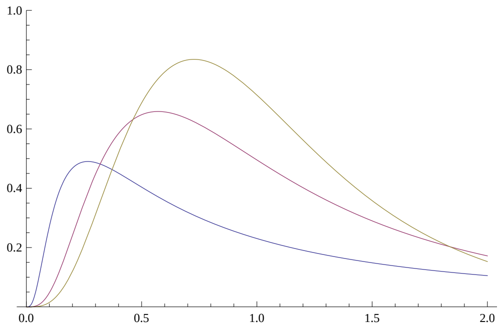
[2] カイ 2 乗分布の概形:
自由度 df {\nu, {1, 2, 8, 20}} 4 本のグラフ
描画範囲 range = {x, 0, 8}
描画命令 Plot[Evaluate@
  Table[PDF[ChiDistribution[\[Nu]], x],
  {\[Nu], {1,2,8,20}}], {x, 0, 8}]

自由度を大きくすると右側に動いて、形は正規分布に近づいている。

[3.1] F 分布の概形（その１）:
自由度 df {n=10, m={1, 2, 8, 20}} 4 本のグラフ
描画範囲 range = {x, 0, 2}
描画命令 Plot[Evaluate@
  Table[PDF[FRatioDistribution[n, 10], x],
  {n, {2, 5, 20}}], {x, 0, 2},
  PlotRange -> {0, 1}, Exclusions -> None]



[3.2] F 分布の概形（その２）:
自由度 df {n={1, 5,20}, m=10} 3 本のグラフ
描画範囲 range = {x, 0, 2}
描画命令
Plot[Evaluate@
  Table[PDF[FRatioDistribution[10, m], x],
  {m, {1, 5, 20}}], {x, 0, 2},
  PlotRange -> {0, 1}, Exclusions -> None]



□□□□□　練習問題　□□□□□

問 2.2.1
X はパラメータ (α, β) のベータ分布 $\mathrm{Beta}(\alpha, \beta)$ にしたがう。平均と分散を求めよ。

問 2.2.2
T は自由度 15 の t 分布にしたがう。$P(-c < T < c) = 0.95$ となるような c の値をもとめよ。

もし表計算ソフトを用いるならば、カイ 2 乗 (χ^2, chi-square) 分布に対する (1) 値から確率　(2) 確率から値　は次のように組み込み関数で求められる。この分布は正の値しかとらないし、y 軸での対称性をもたないから、右側（上側）{RT } と左側（下側）のそれぞれパーセント点を定める命令が用意されている。

(1)　CHISQ.DIST(x，自由度，関数形式（0/1））
　　書式の引数:
　　　値 x　　　分布を評価するための数値を指定。
　　　自由度　　自由度を示す整数を指定。
　　　関数形式 TRUE("1"に対応) の場合は、（累積）分布関数の値。
　　　　　　　FALSE("0"に対応) の場合は、確率密度関数の値。
(2)　CHISQ.DIST.RT(x, 自由度)
　　書式の引数: 返される値は右側{RT}確率の値。
　　　値 x　　　分布を評価するための数値を指定。
　　　自由度　　分布の自由度。
(3)　CHISQ.INV(確率, 自由度)
　　書式の引数: 返される値は分布関数の逆関数による値。
　　　確率　　　確率を指定。
　　　自由度　　分布の自由度。
(4)　CHISQ.INV.RT(確率, 自由度)
　　書式の引数: 返される値は分布関数の逆関数による右側{RT}確率の値。
　　　確率　　　確率を指定。
　　　自由度　　分布の自由度。

問 2.2.3
W は自由度 $(m.n)$ の F 分布にしたがうとする。これを $W \sim F(m.n)$ と表す。このとき逆数 $1/W$ の分布をもとめよ。

問 2.2.4
F 分布表には $c = 0.9, 0.95, 0.975, 0.99$ などの値に対する確率 $P(W < c)$ の値が掲載されている。これに対して $c = 0.1, 0.05, 0.025, 0.01$ などの値を求めるにはどうしたらよいだろうか？

F 分布 (スネディッカーの F-distribution) に対する (1) 値から確率、 (2) 確率から値　も同様に次の組み込み関数で求められる。

(1) F.DIST(X，自由度_1，自由度_2，関数形式（0/1））
　　関数形式 TRUE("1"に対応) の場合は、分布関数の値。
　　　　　　FALSE("0"に対応) の場合は、確率密度関数（pdf）の値。
(2) F.INV(確率，自由度_1，自由度_2)

問 2.2.5
$X \sim T(n)$ ならば、$X^2 \sim F(1, n)$ を示せ。(記号的には $T^2(n) = F(1, n)$)

問 2.2.6

$X \sim \mathrm{Expo}(1)$ つまり指数密度 $f_X(x) = e^{-x}, x > 0$ ならば、$X^2 \sim \chi^2(2)$ となることを示せ。さらに $X \sim \mathrm{Expo}(1), Y \sim \mathrm{Expo}(1)$ で独立ならば、$X/Y \sim F(2,2)$ が成り立つことを示せ。

第 3 章

データ解析と極限定理

3.1 順序統計量

多項 (Multinomial) 分布の密度関数 ($r=2$ は 2 項分布):

$X=(X_1,X_2,\cdots,X_r)\sim \mathrm{Mulnorm}(n;p_1,p_2,\cdots,p_r)$ ($p_1+p_2+\cdots+p_r=1$, $p_i>0$) とは

$$f_X(x_1,x_2,\cdots,x_r)=\binom{n}{x_1,x_2,\cdots,x_r}p_1^{x_1}p_2^{x_2}\cdots p_r^{x_r},\quad (x_1+x_2+\cdots+x_r=n)$$

ここで多項係数 $\displaystyle\binom{n}{x_1,x_2,\cdots,x_r}=\frac{n!}{x_1!\,x_2!\,\cdots x_r!}$, $(x_1+x_2+\cdots+x_r=n)$ とする。とくに 2 項係数の場合には他の一つは省略する。

順序統計量 (order statistics): 大きさの順に並び替える。タイ（同じ値）となるデータはないとする。

元のデータ	$x_1,$		$x_2,$		$x_3,$		$\cdots$		$x_{n-1},$		x_n
大きさの順に	y_1	$<$	y_2	$<$	y_3	$<$	$\cdots$	$<$	y_{n-1}	$<$	y_n

各 $X_i(i=1,2,\cdots,n)$ (連続型確率変数を仮定) が独立で、同一分布 つまり密度関数 $f(x)$、分布関数 $F(x)$ をもつならば、$\{X_i\}$ の順序統計量 $Y=(Y_1,Y_2,\cdots,Y_n):Y_1<Y_2<\cdots<Y_n$ の密度関数 f_Y は

$$f_Y(y_1,y_2,\cdots,y_n)=n!f(y_1)f(y_2)\cdots f(y_n),\quad y_1<y_2<\cdots<y_n \tag{3.1.1}$$

で与えられる。

パーセント点 : $100p\%$ 点、$(0<p<1)$、$F^{-1}(p)=\sup\{x;F(x)\le p\}=\inf\{x;p<F(x)\}$。2 つの値が等しい場合は連続型確率変数のとき。つまり分布関数の逆関数で定める。

□□□□□ 練習問題 □□□□□

問 3.1.1

3 つの確率変数 X_1,X_2,X_3 を大きさの順に並べた $Y_1<Y_2<Y_3$ について、もし X_i が独立で、単位区間 $[0,1]$ 上に一様分布であるとき、$Z=Y_3-Y_1$ の密度関数を求めよ。

問 3.1.2

確率変数 $X_1,X_2,\cdots,X_n$ は iid で、密度を $f(x)$, 分布関数を $F(x)$ で与えられたとする。この順序統計量を $Y_1<Y_2<\cdots<Y_n$ とおく。$k=1,2,\cdots,n$ に対して

$$\begin{aligned}f_{Y_k}(x)&=\frac{n!}{(k-1)!(n-k)!}F(x)^{k-1}\,[1-F(x)]^{n-k}f(x)\\&=n\binom{n-1}{k-1,n-k}F(x)^{k-1}\,[1-F(x)]^{n-k}f(x)\end{aligned}$$

を導け。ただし係数は 2 項係数とする。また同時分布として $j<k$, $(x<y)$ に対し、

$$\begin{aligned} f_{Y_j,Y_k}(x,y) &= \frac{n!}{(j-1)!(k-j-1)!(n-k)!}F(x)^{j-1}\left[F(y)-F(x)\right]^{k-j-1}\left[1-F(y)\right]^{n-k}f(x)f(y) \\ &= n(n-1)\binom{n-2}{j-1,k-j-1,n-k}F(x)^{j-1}\left[F(y)-F(x)\right]^{k-j-1}\left[1-F(y)\right]^{n-k}f(x)f(y) \end{aligned}$$

を導け。ただし係数は 3 項係数とする。

この式は、離散型確率変数では成り立たない。その理由を考えよ。

問 3.1.3

確率変数 $X_1, X_2, \cdots, X_n$ は iid で、密度を $f(x)$ は単位区間 $[0,1]$ 上の一様分布とする。この順序統計量を $Y_1<Y_2<\cdots<Y_n$ について、$Y_k \sim \mathrm{Beta}(\alpha,\beta)$ となり、このパラメータを k,n で表せ。

問 3.1.4

確率変数 $X_1, X_2, \cdots, X_n$ は iid で、密度を $f(x)$ は単位区間 $[0,1]$ 上の一様分布とする。この順序統計量 $Y_1<Y_2<\cdots<Y_n$ について、確率 $P(Y_k>p)$ （ただし $0<p<1$）を求めよ。この値を p,k,n をもちいた和の形で表せ。

問 3.1.5

確率変数 $X_1, X_2, \cdots, X_n$ は iid で、密度を $f(x)$ は単位区間 $[0,1]$ 上の一様分布とする。この順序統計量 $Y_1<Y_2<\cdots<Y_n$ について、(i) 平均 $E(Y_k)$,　(ii) 分散 $\mathrm{var}(Y_k)$, (iii) 共分散 $\mathrm{cov}(Y_j,Y_k)$ $(j<k)$ を求めよ。

3.2 大数の弱法則

確率についての不等式：

(a) **マルコフの不等式**：確率変数 $X>0$, 定数 $a>0$ に対し、$P(X\geq a)\leq \dfrac{E(X)}{a}$

(b) 任意の実数 a と整数 m, 実数 $d>0$ に対し、$P(|X-a|\geq d)\leq \dfrac{E(|X-a|^m)}{d^m}$

(c) **チェビシェフの不等式**：確率変数 X は有限な平均 μ, 分散 σ^2 をもつならば、

$$P(|X-\mu|\geq k\,\sigma)\leq \frac{1}{k^2} \tag{3.2.1}$$

確率収束する ($\{X_n\}$ converges to X in probability)：記号で $X_n \xrightarrow{P} X$, あるいは $X_n \to X$ in prob とは, 任意の $\epsilon>0$ に対して、$P(|X_n-X|>\epsilon)\to 0$ が成り立つとき。言いかえると, (i) $X_n \xrightarrow{P} 0 \;\leftrightarrow\; P(|X_n|>\epsilon)\to 0, \forall\epsilon>0$ と定め、(ii) $X_n \xrightarrow{P} X \;\leftrightarrow\; X_n - X \xrightarrow{P} 0$ とする。

分布収束 ($\{X_n\}$ converges to X in distribution): もし分布関数 $F_n(x)\to F(x)$ $(n\to\infty)$ という命題が $F(x)$ の連続点 $\{x; F(x)=F(x-)\}$ で成り立つとき。記号で $X_n \xrightarrow{D} X$ と表すとする。ただし分布の収束は確率変数をもちいることで明確に性質を表す表現とはいえない。簡単に分布関数の値に対する極限の形をもちいることが多い。

大数の弱法則 (weak Law of Large Numbers) （標本平均に対する母平均への確率収束）: 確率変数列 $\{X_1, X_2, \cdots\}$ が独立で、同一の分布に従い、有限な平均 μ, 分散 σ^2 をもつとする。$S_n = X_1+X_2+\cdots+X_n$ とおくとき、

$$\frac{S_n}{n}\to \mu \text{ in prob} \tag{3.2.2}$$

つまり、任意の $\epsilon>0$ に対し、$P\left(\left|\dfrac{S_n}{n}-\mu\right|\geq\epsilon\right)\to 0 \quad (n\to\infty)$ となること。

□□□□□　練習問題　□□□□□

問 3.2.1

X_n は区間 $(0, 1/n)$ 上の一様分布にしたがうとする。この確率変数列は 0 へ確率収束するが、$x = 0$ の点では分布収束しない。これを示せ。また確率収束するならば、分布収束することを示せ。

確率収束するための十分条件：

(a) $E[(X_n - X)^2] \to 0\,(n \to \infty)$ ならば、$X_n \xrightarrow{P} X$ （２次の平均収束は確率収束より強い）

(b) $E[X_n] - E[X] \to 0, \mathrm{var}(X_n - X) \to 0$ ならば、$X_n \xrightarrow{P} X$ (平均の収束と分散からの条件)

上の命題の証明：(a)　チェビシェフの不等式から、$\epsilon > 0$ とし、$P(|X_n - X| \geq \epsilon) \leq \dfrac{E[(X_n - X)^2]}{\epsilon^2} \to 0$ であるから。(b) $E[(X_n - X)^2] = \mathrm{Var}(X_n - X) + [E(X_n) - E(X)]^2 \to 0$ であるから。

「注意」もし確率変数 X が定数 c であるならば、$\mathrm{Var}(X_n - X) = \mathrm{Var}X_n, E(X_n) \to c, \mathrm{Var}X_n \to 0$ より、$X_n \xrightarrow{P} c$ である。

問 3.2.2

独立で同一の正規分布 $N(\mu, \sigma^2)$ から定めた標本分散 S_n^2 については $S_n^2 \xrightarrow{P} \sigma^2$ が成り立つ。

問 3.2.3

確率変数 $X_1, \cdots, X_n$ は独立であるが、必ずしも同一分布ではないとする。それぞれ X_i は平均 μ_i, 分散 σ_i^2 をもつとし、分散が一様有界である：$\forall i, \exists M, \sigma_i^2 \leq M$, とする。このとき、確率収束

$$\frac{S_n - E(S_n)}{n} \xrightarrow{P} 0$$

を示せ。ただし $E(S_n) = \sum_i \mu_i$ とする。

問 3.2.4

もし同一分布であるとすると、$\mu_i = \mu, \sigma_i^2 = \sigma^2$ であり、$\dfrac{S_n}{n} \xrightarrow{P} \mu$ と確率収束することを示せ。

問 3.2.5

$X_n \sim F_n,\ Y_n \sim G_n$ とし、$X_n - Y_n \xrightarrow{P} 0, G_n \to G$ ならば、$F_n \to G$ となる。

問 3.2.6

確率変数 $X_1, X_2, \cdots, X_n$ は iid で有限な平均 μ, 分散 σ^2 をもつとし、標本平均 $\overline{X}_n = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$ とおく。このとき、$\overline{X}_n$ の極限分布をもとめよ。つまり、$\overline{X}_n \xrightarrow{P} X$ となる確率変数 X をもとめよ。

問 3.2.7

$\overline{X}_n$ を区間 $[n, n+1]$ 上の一様分布とする。この分布の極限分布は存在しないことを示せ。直感的には確率が無限大の方へ逃げてしまうからである。

3.3　中心極限定理

確率変数列 $\{X_i\}$ が同じ平均 μ, 分散 σ^2 をもつ独立な分布（正規分布でなくてもよいし、離散型あるいは連続型でもよい）であるならば、標準化した変数は標準正規分布に分布収束する。

$$Y_n = \frac{\sum_{i=1}^{n} X_i - n\,\mu}{\sqrt{n}\sigma} \xrightarrow{D} N(0,1) \tag{3.3.1}$$

任意の $a, b(a < b)$ に対して

$$\lim_n P(a < Y_n < b) = \int_a^b \frac{1}{\sqrt{2\pi}} \exp\{-\frac{x^2}{2}\} dx$$

□□□□□　練習問題　□□□□□

問 3.3.1

$X_n \sim \mathrm{Gam}(n, \beta)$ とする。つまり確率変数 X_n はパラメータ n, β のガンマ分布にしたがうとする。この確率変数 X_n は n 個の独立な指数分布の和であることを示せ。この結果から、$\dfrac{X_n}{n}$ の極限分布をもとめよ。

問 3.3.2

自由度 n のカイ 2 乗分布 $\chi^2(n)$ は, n が十分に多いければ、正規分布 $N(n, 2n)$ で近似されることを示せ。

問 3.3.3

n 個の確率変数 $X_1, X_2, \cdots, X_n$ は iid で, その密度を f とする。Y_n で区間 (a, b) 内にあった個数を表すとする。正規近似をもちいて、Y_n に関する確率を計算するためにはどのようにしたらよいか。

問 3.3.4

3 個のデータ 6.45, 3.14, 4.93 を観測したとき、その値を丸めて、(最も近い整数値にする) 6, 3, 5 とし、これらの和を求めると、14 となるが、元データでは $14.52 = 0.45 + 0.14 + (-0.07)$ であった。いま $X_i, i = 1, 2, \cdots, n$ を i 番目のデータに関する丸め誤差を表すとし、X_i は iid で、区間 $(-\frac{1}{2}, \frac{1}{2})$ 上の一様分布と仮定する。丸め誤差の総和 $Y_n = \sum_{i=1}^{n} X_n$ に関する確率を、正規近似で計算するためにはどのようにしたらよいか。上の例では $Y_3 = 14.52,\ X_1 = 0.45,\ X_2 = 0.14,\ X_3 = -0.07$ であった。

問 3.3.5

$X_1, X_2, \cdots, X_n$ は iid で、連続型確率変数で分布関数 F をもつとする。この順序統計量を $Y_1 < Y_2 < \cdots < Y_n$ とおく。したがって、確率変数 $F(Y_1), F(Y_2), \cdots F(Y_n)$ は iid で、$[0, 1]$ 上の一様分布にしたがう。このとき、$n\,(1 - F(Y_n))$ の極限分布は指数分布であることを示せ。

第 4 章

推測理論

4.1 統計量と推定

推定の重要項目

1．最尤法 (MLE)：尤度がもっとも高くなる推定量。
2．信頼区間：多くの実践的に用いる手法である。データから区間（データの値によって変わる）を定め、未知パラメータがこの中に含まれる確率（精度）を保証するもの。
3．一様最小分散不偏推定量 (Uniformly Minimum Unbiased Estimators, UMVUE's)：大数の法則から数学的に定めるが、よく知られているように、不偏ではなくてもよい推定を行うことができる。分散の小さいことと、偏りのないことが同時には成り立たない。一様とは母数がどんな値であってもという意味。
4．ベイズ推定未知母数（パラメータ）について予めの知識や経験から、その分布を仮定できるという考え方から出発して、漸次ベイズの定理により、母数の確率を改訂していく。

最尤 (ML) 法とは、データ $(x_1, x_2, \cdots x_n)$ から、未知の母数 θ の尤度関数 $lik(\theta) = L(\theta) = f_\theta(x) = f(x_1, x_2, \cdots x_n \mid \theta) = \Pi_{i=1}^n f(x_i \mid \theta)$ を最大にするような $\hat{\theta}$:

$$L(\hat{\theta}) = \max_\theta L(\theta) \tag{4.1.1}$$

尤度関数は、標本データの取り出しが、独立、同一分布とすれば、上のような積の形となる。いいかえると、ラプラスの逆確率に相当する。データを得た後の母数の起こる可能性、事後確率が計算できるという考え方である。データ x のもとでの、未知母数 θ の確率 $P(\theta \mid x)$ は母数の事前確率 $P(\theta)$ を仮定すると、条件付き確率の定義式から

$$P(\theta \mid x) = \frac{1}{\text{const}} P(\theta)P(x \mid \theta) \approx P(\theta)P(x \mid \theta)$$

（ここで $\text{const} = \int P(\theta)P(x \mid \theta)\, d\theta$ とすれば、全確率の値が 1 になるよう定められる）の形に表されるから、未知母数の選択する基準として、$P(x \mid \theta)$ が大きくなるものを選べばよい。

$$\text{事前確率 } P(\theta) \quad \Rightarrow \quad \text{観測データ } x \quad \Rightarrow \quad \text{事後確率 } P(\theta \mid x)$$

これがベイズ統計学の最尤法である。多くの場合対数をとった対数尤度：

$$L(\theta) = \sum_i \log f(x_i \mid \theta)$$

が用いられる。対数関数の単調性から、この 2 つは同値であるから、同じ推定量を定める。最大値を求めるためには、もし滑らかならば、微分をして

$$\frac{\partial}{\partial \theta} L(\theta) = 0 \tag{4.1.2}$$

微分できないときには、アドホック的に考察する。

最尤法に関する一般定理：

定理 4.1.1
関数 h に対し、θ の MLE $\hat{\theta}$ であるならば、$h(\theta)$ の MLE は $\widehat{h(\theta)} = h(\hat{\theta})$ で与えられる。また h が連続ならば、一致性が保たれる。つまり、$\hat{\theta} \xrightarrow{P} \theta$ ならば、$h(\hat{\theta}) \xrightarrow{P} h(\theta)$ が成り立つ。

モーメント法：データの k 次積率 $\frac{1}{n}\sum_i x_i^k$ と原点のまわりの k 次積率（期待値）EX^k（未知母数 θ を含む）とを等しいとおいてつくる方程式から母数を定める。

例題. ガンマ分布 Gam(α,β) の場合ならば、$EX = \alpha\beta$, $\mathrm{var}(X) = \alpha\beta^2$ であり、一方、データの平均 $\overline{X} = \frac{1}{n}\sum_i X_i$、分散 $S^2 = \frac{1}{n}\sum_i (X_i - \overline{X})^2$ から次の対応式

$$\overline{X} = \alpha\beta, \quad S^2 = \alpha\beta^2$$

であるから、

$$\alpha = \frac{S^2}{\overline{X}}, \quad \beta = \frac{\overline{X}^2}{S^2}$$

がモーメント法によるそれぞれ定められるパラメータの推定量である。

□□□□□　練習問題　□□□□□

以下の問題では $X_1, \cdots, X_n$ は iid で、その密度関数 $f_\theta(x)$（連続型の場合）確率密度 $p_\theta(x)$(離散型の場合) として答えよ。

問 4.1.1
次の分布に対する θ の MLE をもとめよ。
(a) パラメータ $\theta > 0$ のポアソン分布
(b) $f_\theta(x) = \theta x^{\theta-1}, 0 < x < 1$ $(\theta > 0)$
(c) パラメータ θ の指数分布、$f_\theta(x) = \dfrac{1}{\theta}e^{-x/\theta}$, $x > 0$
(d) $f_\theta(x) = \dfrac{1}{2}e^{-|x-\theta|}$, θ, x は任意の実数。
(e) $f_\theta(x) = e^{-(x-\theta)}$, $x > \theta$

問 4.1.2
$X_1, X_2, \cdots, X_n$ は iid で、それぞれは区間 $(\theta - \frac{1}{2}, \theta + \frac{1}{2})$ 一様分布にしたがうとする。θ の MLE をひとつ以上もとめよ。(MLE は必ずしも一意とは限らないことに注意）

問 4.1.3
問 4.1.1 において, それぞれの $E(X_i)$ を計算し、標本平均を真の平均とおくことで得られるモーメント法により、推定量を導け。それぞれの場合について、推定量は一致性をもつことを確かめよ。

問 4.1.4
X はパラメータ θ の指数分布とする。$r > 0$ に対する確率 $P(X \le r)$ の MLE をもとめよ。

問 4.1.5
$X \sim \mathrm{Binom}(n, \theta)$ とし、a, b は整数で $0 \le a \le b \le n$ とする。$P(a \le X \le b)$ の MLE をもとめよ。

4.2　信頼区間

比率の推定：２項母集団 の成功確率を推定。

２項母集団での未知母数：成功の確率 p, 失敗の確率 $q = 1 - p$; 観測データ；サイズ (大きさ) n, 表の回数 X_n。成功（表の回数）: X_n と失敗（裏の回数）: $Y_n = n - X_n$ から、p の推定値として $\hat{p} = \frac{X_n}{n}$ を用いる。

比率 p の信頼区間 : 信頼係数 (推定値が未知母数を含む確率): $1-\alpha$、精度の上界、(偏差、推定値と未知母数とのずれの上限) : e, を与えると、$E[X_n] = np, \mathrm{var}[X_n] = npq$ であるから、$E[\hat{p}] = p, \quad \mathrm{var}[\hat{p}] = \dfrac{pq}{n} \le \dfrac{1}{4n}$ が成り立つ。

$$P\left\{|\hat{p} - p| < e\right\} > 1 - \alpha \qquad \Rightarrow \qquad \hat{p} - e < p < \hat{p} + e \quad \text{with Prob.}\, 1 - \alpha$$

の関係であるから、２項分布の正規近似より、

$$\begin{aligned} P\left\{|\hat{p} - p| < e\right\} &= P\left\{\left|\frac{X_n - np}{\sqrt{np(1-p)}}\right| < \frac{e\sqrt{n}}{\sqrt{p(1-p)}}\right\} \\ &\approx \Phi\left(\frac{e\sqrt{n}}{\sqrt{p(1-p)}}\right) - \Phi\left(-\frac{e\sqrt{n}}{\sqrt{p(1-p)}}\right) = 2\Phi\left(\frac{e\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1 \\ &= 1 - \alpha \end{aligned}$$

標準正規分布の両側 α パーセント点 $z_\alpha = \Phi(1 - \frac{\alpha}{2}), -z_\alpha = \Phi(\frac{\alpha}{2})$ より、$z_\alpha = \dfrac{e\sqrt{n}}{\sqrt{p(1-p)}}$ を変形すると、$e = z_\alpha\sqrt{\dfrac{p(1-p)}{n}}$ となる。信頼係数を定めるには、この確率を計算すればよい。しかしこの確率は値 p が未知であるから、p の代わりに標本値 $\hat{p}$ と近似する $\hat{e} = z_\alpha\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$ として、

$$\hat{p} - e < p < \hat{p} + e \tag{4.2.1}$$

必要なサイズ数：n_α の決定は p の値が未知であるから、$p(1-p) \le 1/4$ より、

$$n > \left(\frac{z_\alpha}{e}\right)^2 \times \frac{1}{4} \tag{4.2.2}$$

たとえば、$e = 0.03$, 推定の精度を $\pm 3\%$ で 95 パーセントの信頼区間にするためには、$n > \left(\dfrac{1.96}{0.03}\right)^2 \times \dfrac{1}{4} = 1067$ とする。

正規分布の母数 正規母集団の平均と分散の推定をするときには、(i) 分散が既知の場合と、(ii) 未知の場合にもちいる統計量が異なる。$X_i, i = 1, 2, \cdots, n \sim N(\mu.\sigma^2)$ として、μ の推定をおこなう。

(i) σ^2 は既知だから、$\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ より、

$$\overline{X} - b_0\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + b_0\frac{\sigma}{\sqrt{n}} \tag{4.2.3}$$

ここで、b_0 は正規分布の両側 α パーセント点。

(ii) σ^2 の代わりに標本不偏分散 S^2 をもちい、$\dfrac{\overline{X} - \mu}{S/\sqrt{n-1}} \sim T(n-1)$ より、

$$\overline{X} - b_0\frac{S}{\sqrt{n-1}} < \mu < \overline{X} + b_0\frac{S}{\sqrt{n-1}} \tag{4.2.4}$$

ここで、b_0 は t 分布の自由度 $n-1$ の α パーセント点。

このように信頼区間は分布のパーセント点をもちいるが、それぞれの分布表から求められる。もし表計算ソフトが利用できれば直接、信頼区間を求める組み込み関数が定義されている：(1) 正規母集団で分散既知の場合（正規分布をもちいる)、(2) 分散未知の場合（標本分散をつかい、ｔ－分布をもちいる）

(1)　CONFIDENCE.NORM(α，標準偏差，標本数)

書式の引数:正規分布をもちいる

値　α　　信頼係数を指定。
標準偏差　分散既知の場合。
標本数　　標本サイズの指定。

(2)　CONFIDENCE.T(α，標準偏差，標本数)
書式の引数:スチューデントの t 分布をもちいる
値　α　　信頼係数を指定。
標準偏差　標本不偏分散から求める場合。
標本数　　標本サイズの指定。

２項分布の差の信頼区間（大標本の場合）：
Y_1, Y_2 がそれぞれ 2 項分布 $binom(n_1, p_1)$ $binom(n_2, p_2)$ にしたがうとすれば、$E(Y_i/n_i) = p_i$, $\mathrm{var}(Y_i/n_i) = p_i q_i/n_i$, $i = 1, 2$ であるから、差の推定は標本比率の差で、分散が和になることに注意すれば、

$$\frac{(Y_1/n_1 - Y_2/n_2) - (p_1 - p_2)}{\sqrt{p_1 q_1/n_1 + p_2 q_2/n_2}} \sim N(0,1)$$

をもちいれば信頼区間が定まる。

分散の信頼区間： 正規母集団 $N(\mu, \sigma^2)$ の分散に対する信頼区間は

$$W = nS^2/\sigma^2 = \sum_{i=1}^{n} \left(\frac{X_i - \overline{X}}{\sigma} \right)^2 \sim \chi^2(n-1)$$

を利用する。確率のパーセント点は $P(b_1 < W < b_2) = \int_{b_1}^{b_2} f_W(t)dt = 1 - \alpha$ となる b_1, b_2 を下側 $\alpha/2$, 上側 $1 - \alpha/2$ と自由度 $n-1$ のカイ 2 乗分布表から定めればよい。これにより、

$$\frac{nS^2}{b_2} < \sigma^2 < \frac{nS^2}{b_1} \tag{4.2.5}$$

と定められる。

□□□□□　練習問題　□□□□□

問 4.2.1

$X_1, \cdots, X_n$ は iid とし、$N(\mu, \sigma^2)$ とする。σ^2 は既知と仮定する。μ の信頼区間の長さを計算するにはどうしたらよいだろうか。

問 4.2.2

$X_1, \cdots, X_n$ は iid とし、$N(\mu, \sigma^2)$ とする。もし σ^2 は未知であるときには、μ の信頼区間の長さを計算するにはどうしたらよいだろうか。このときには標本標準偏差 S を用いよ。

問 4.2.3

$X_1, \cdots, X_n$ は iid とし、$N(\mu, \sigma^2)$ とする。もし σ^2 は未知であるときには、μ の信頼区間の長さを計算するにはどうしたらよいだろうか。ただし未知の標準偏差 σ を用いよ。このときは区間の長さは情報の少ないことにより、区間の長さは長くなる。

問 4.2.4

$X_1, \cdots, X_n$ は iid とし、ガンマ分布 $\mathrm{Gam}(\alpha, \beta)$ とする。もし α が既知であるならば、平均 $\mu = \alpha\beta$ に対する信頼区間を計算するにはどうしたらよいだろうか。

問 4.2.5

2 項母集団における平均の推定で、信頼区間の長さと信頼係数を与えて、n の最小となるような値を計算するにはどうしたらよいか。

4.3　２標本母集団での信頼区間

2 標本 $N(\mu_1, \sigma_1^2)$, $N(\mu_2, \sigma_2^2)$ に対する平均の差 $\mu_1 - \mu_2$ の比較問題を考える。$X_1, \cdots, X_n$ は iid で $N(\mu_1, \sigma_1^2)$ として $Y_1, \cdots, Y_m$ は iid で $N(\mu_2, \sigma_2^2)$ で独立とする。標本分散をそれぞれ $S_1^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2$, $S_2^2 = \frac{1}{m}\sum_{i=1}^m (Y_i - \overline{Y})^2$ とする。

(1) $\sigma_1^2 = \sigma_2^2 = \sigma^2$ が未知の場合（等しいことは分かっているが、その値が分からない）: $R^2 = \left(\dfrac{nS_1^2 + mS_2^2}{n+m-2}\right)\left(\dfrac{1}{n} + \dfrac{1}{m}\right)$ とおくとき、

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{R} \sim t(n+m-2)$$

を適用して、$\mu_1 - \mu_2$ の区間推定を行うことができる。ここで $t(n+m-2)$ は自由度 $n+m-2$ の t 分布。なぜならば、$Z_j \sim N(\mu, \sigma^2), j = 1, 2, \cdots, k$ が独立ならば、変数の標準化やカイ 2 乗分布の加法性から $\dfrac{Z_j - \mu}{\sigma} \sim N(0, 1)$, $\sum_{j=1}^k \left(\dfrac{Z_j - \mu}{\sigma}\right)^2 \sim \chi^2(k)$。またとくに正規分布であれば、$\overline{Z}$ と S^2 が独立になることから独立な２つの比：標準正規分布とカイ 2 乗分布の平方根との比は t 分布となるから結論が得られる。記号的に書くと,

$$\frac{N(0,1)}{\sqrt{\chi^2(k)}} \sim t(k-1) \tag{4.3.1}$$

(2) σ_1^2, σ_2^2 がともに既知の場合：$R_0^2 = \left(\dfrac{n\sigma_1^2 + m\sigma_2^2}{n+m}\right)\left(\dfrac{1}{n} + \dfrac{1}{m}\right) = \dfrac{\sigma_1^2}{m} + \dfrac{\sigma_2^2}{n}$,（$n, m$ の順序に注意）とおくとき、

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{R_0} \sim N(0, 1) \tag{4.3.2}$$

を適用して、$\mu_1 - \mu_2$ の区間推定を行うことができる。

こんどは分散の比率に関する信頼区間を定めるもちいる統計量は F 分布である。なぜなら正規分布から得られる 2 乗和がカイ 2 乗分布であるから、この比率は F 分布になることをもちいる。つまり標本不偏分散 $V^2 = \dfrac{n}{n-1}S^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^n (X_i - \overline{X})^2$ とすれば $V^2/\sigma^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^n \left(\dfrac{X_i - \overline{X}}{\sigma}\right)^2 \sim \chi^2(n-1)$ であるから、2 つを分散 σ_1^2, σ_2^2 として, 対応する標本不偏分散 V_1, V_2 の比率

$$\frac{V_2^2/\sigma_2^2}{V_1^2/\sigma_1^2} \sim F(n_2 - 1, n_1 - 1) \tag{4.3.3}$$

から求める。

□□□□□　練習問題　□□□□□

問 4.3.1

分散が未知で等しくない場合にこの方法 (4.3 節) を適用できない理由は何か。（この問題はベーレン=フィシャーの問題とよばれる）

問 4.3.2

分散の値は未知であるが比の値が既知であるとき、つまり既知の定数 c で, $\sigma_1^2 = c\sigma_2^2$ である条件のもとで信頼区間を構成せよ。

4.4 仮説検定

統計的仮説検定 (Hypothesis Testing) とは、1 つあるいは複数の確率データの分布に関する命題について、真偽の判断を下そうとするもの。命題を棄却できるかどうかの規則、手順を定めることが目的である。分布に関する命題を仮説とよび、帰無仮説 (null hypothesis) とは、もう一方の対峙する対立仮説 (alternate hypothesis) を支持して、棄却されるあるいは破棄されることを意図して立てられるもの。このように 2 者択一の立場に立つものである。(参照: 8 章検定 115 ページ「統計学演習」村上/安田、培風館)

仮説の検定で “ある判断”を下すならば、2 種類の誤りが必ず生じる。帰無仮説が正しいにも関わらず、正しくないと棄却すること, **第 1 種の過誤** (Error of Type I) 、および対立仮説が正しくないにも関わらず、それを正しいと考えてしまうこと**第 2 種の過誤** (Error of Type II) である。ネイマン・ピアソン流の考えでは第 1 種の過誤の確率に対して上限（有意水準とよぶ）を決めておき、第 2 種の過誤の確率を最小とする、あるいは過誤を犯さない確率 [$= 1 -$ (第 2 種の過誤確率)]（**検出力 (power)** とよび）を最大にするような判断の方法を求める。

[ネイマン・ピアソンの補題] 帰無仮説 $H_0 : f = f_0$, 対立仮説 $H_1 : f = f_1$ において、有意水準（サイズともいう）α とするとき、第 2 種の過誤の確率を最小とする検定は次で与えられる : 定数 k と尤度 $L(\theta)$ の比からデータの領域（**棄却域** reject region）

$$C_k = \left\{ x = (x_1, x_2, \cdots, x_n) : \frac{\Pi_i f_1(x_i)}{\Pi_i f_0(x_i)} = \frac{L(\theta_1)}{L(\theta_0)} > k \right\}$$

とおくとき、得られたデータ x がもし $x \in C_k$ ならば、帰無仮説を棄却し (**reject**)、そうでないとき、つまり $x \notin C_k$ であれば、帰無仮説を棄却しない、採択する (**accept**) ものと考える。データ x を棄却すること “1”と、棄却しないこと “0”として、これを関数 $\phi(x) = 1, 0$ に対応させる。この関数を検定とよび、$\phi(x) = 1_{C_k}(x)$ の形であり、ただし定数 k は、あらかじめ与える第 1 種の過誤確率

$$\alpha = P(X \in C_k \,|\, H_0) = \int_{C_k} f_0(x) dx$$

を満たすとする。$\alpha = 0.05, 0.01$ がよく用いられる。水準が 5%、 1% とする。簡略化のために $(f_0(x) = \Pi_i f_0(x_i), dx = dx_1 dx_2 \cdots dx_n)$ としている。C_k は棄却域であり、帰無仮説と対立仮説を尤度比によって比較し、分子にある対立仮説のほうが可能性が高ければ帰無仮説を棄却するという意図である。また帰無仮説 H_0 をひとつの値 θ_0 としているが、この場合を単純 (simple) 帰無仮説とよぶ。より一般には集合 $\Theta \ni \theta_0$ を考える複合 (composite) 帰無仮説ときもある。このときには $\alpha = \sup_{\theta \in \Theta_0} P(X \in C_k \,|\, \theta)$ が第 1 種の過誤確率（検定の**サイズ** (size)）とする。

証明のために補題はつぎの問題の形になる : 与えられた関数 $f_i, g_i; i = 1, 2, \cdots, n$ と定数 $A, B \geq 0$ が $A < B$ であるとき、変数 w_i が $w_i : A \leq w_i \leq B,\ i = 1, 2, \cdots, n$ で動くとき、g_i についての条件 $\sum_i g_i w_i = c$ のもとで、f_i に関する目的値 $\sum_i f_i w_i \to \max_{w=(w_i)}$ と最大にするという問題。

この答えは, ある値 k を定めて、$C = \{i; f_i > k g_i\}$ とするとき、$w_i = \begin{cases} B, & i \in C \\ A, & i \notin C \end{cases}$ とすればよい。これを $w_C(i)$ とおく。等号はどちらでも同じ。証明は、$C = \{i; f_i > k g_i\}$ から決める $w_C(i) = B\mathbf{1}_C(i) + A\mathbf{1}_{\overline{C}}(i)$ と任意の $A < w_i < B$ の差を比較する。$\sum_i f_i w_C(i) = \max_{w=(w_i)} \sum_i f_i w_i$ を示せばよい。もし i が $f_i - k g_i > 0$ であれば、$w_C(i) - w_i = B - w_i > 0$ であり、逆の場合、$f_i - k g_i < 0$ であれば、$w_C(i) - w_i = A - w_i < 0$ と同符号であるから、積は $(f_i - k g_i)(w_C(i) - w_i) > 0$ この和をとると、$\sum_i (f_i - k g_i)(w_C(i) - w_i) = \sum_i (f_i w_C(i) - f_i w_i - k g_i w_C(i) + k g_i w_i) = \sum_i f_i w_C(i) - \sum_i f_i w_i - k \sum_i g_i w_C(i) + k \sum_i g_i w_i = \sum_i f_i w_C(i) - \sum_i f_i w_i - kc + kc = \sum_i f_i w_C(i) - \sum_i f_i w_i > 0$ ゆえに $\sum_i f_i w_C(i) > \sum_i f_i w_i,\ \forall w_i$, つまり、$\sum_i f_i w_C(i) = \max_{w=(w_i)} \sum_i f_i w_i$ となる。(終)

以上をいいかえると、未知母数 θ とし、帰無仮説 $H_0 : \theta = \theta_0$, 対立仮説 $H_1 : \theta = \theta_1$ において、$f(x|\theta) = \dfrac{dF(x|\theta)}{dx}$(連続型), $f(x|\theta) = F(x|\theta) - F(x - |\theta) = P(X = x|\theta)$(離散型), $X \sim F(x\,|\,\theta)$ のとき、$E_{\theta_0}\phi(X) = \alpha$（有意水準を一定値 α）のもとで、第 2 種の過誤確率を犯さない、検出力 $E_{\theta_1}\phi(X)$ を大きくする $\to \max_\phi$ 検定 ϕ を定めること。

[Neyman-Pearson の補題による最良検定] 検定 ϕ が、任意の ψ に対して

$$E_{\theta_0}\phi(X) = E_{\theta_0}\psi(X) = \alpha$$

を満たすとき、

$$E_{\theta_1}\phi(X) \geq E_{\theta_1}\psi(X) \quad \forall\psi$$

となるような最良の検定 ϕ を求めると、$E_{\theta_0}\phi(X) = \alpha$ で、$\phi(x) = \begin{cases} 1 & \text{if } f(x|\theta_1) > kf(x|\theta_0) \\ 0 & \text{if } f(x|\theta_1) < kf(x|\theta_0) \end{cases}$ と定めた場合で与えられる。

標準的な仮説検定問題では、どのような仮説（帰無仮説、対立仮説）について、どんな検定統計量、棄却域が用いられるかという問いに対しては、このネイマン・ピアソンの補題によって解いた結果を表にまとめられている。（参照: 8 章検定 116 ページから 118 ページまでの一覧表「統計学演習」村上/安田、培風館）

定理 4.4.1

尤度比検定 (Likelyhood Ratio Tests)LRT とは、つぎのそれぞれの仮説のもとで密度関数の比率: $L(x) = \dfrac{f_1(x)}{f_0(x)} = \dfrac{f(x\,|\,H_1)}{f(x\,|\,H_0)}$ の大きさによって、判断を下すもの。比率の分母には帰無仮説を中心として考える。

$L(x) > k : \phi_k(x) = 1 \quad \Rightarrow x \in C_k, x$ は棄却域の点 $\Rightarrow H_0$:棄却

$L(x) < k : \phi_k(x) = 0 \quad \Rightarrow x \notin C_k, x$ は棄却域の点でない $\Rightarrow H_0$:採択,H_1:棄却

この棄却域の定め方が尤度比を用いるから、このように意味のある検定を定める方法は尤度比検定とよばれる。また検定では、使われる分布からその名称を定めることも多い。

第 1 種の過誤確率は、検定の有意水準とよばれ、検定のサイズともよばれることもある。複合帰無仮説 $H_0 : \theta \in \Theta$、棄却域 C であれば、

$$\alpha = \sup_{\theta\in\Theta_0} P(X \in C\,|\,\theta)$$

となるが、尤度比検定での 観測値 x に対する p 値とは、

$$p^* = \sup_{\theta\in\Theta_0} P_\theta(L_X(H_0, H_1) \geq L_X(H_0, H_1)) \tag{4.4.1}$$

ここでデータ x に対して、仮説 $H : \theta = \theta^*$ における尤度は $L_x(H) = f_X(x\,|\,\theta = \theta^*)$ で複合仮説であれば、$L_x(H) = \sup_{\theta\in\Theta} f_X(x\,|\,\theta)$ とし、2 つの仮説における尤度比とは

$$L_x(H_0, H_1) = \frac{L_x(H_1)}{L_x(H_0)}$$

とおく。「注意」$T(x)$ が θ の十分統計量であれば、$L_x(H_0, H_1)$ は分解定理により、$T(x)$ のみの関数となる。

さらに歴史的にはフィシャーのいう有意性検定が知られています。対立仮説が正しい確率は直接計算することができないため、帰無仮説の正しい確率を見極めることを主眼として、これを「有意確率」または「p 値 (probability)」とよびます。そしてこの有意確率が小さな基準とした値、例えば 0.05(5 %) 以下になった時、つまり対立仮説の正しい確率が 0.95(95 %) 以上になった時、帰無仮説を棄却して対立仮説を採択します。このことを、統計用語では「有意水準 5 %で有意である」または「有意水準 5 %で帰無仮説を棄却 (reject) す

る」といいます。「有意水準 (significance level)」は、「危険率 (critical rate)」または「α エラー」とも呼ばれています。この有意という用語は帰無（仮説）が否定される理由で使われると考えられます。このような p 値（第 1 種の過誤）を中心とする検定の考え方はフィッシャー流の検定で、厳密には「有意性検定 (test of significance)」と呼ばれています。しかし現在では、ネイマン (Jerzy Neyman)・ピアソン (Egon Pearson) 流の「仮説検定論」(test of hypothesis, statistical hypothesis testing) 手法が数理的解析（条件付き最適化問題）として明快であり、主流になっています。結論が同じという場合でも出発の考えが異なるので、両者の間では厳しい論争が繰り広げられていました。有意性検定は、仮説の正しい確率が非常に高い時だけ結論として採用する手法であり、仮説の正しい確率が低い時には結論を保留します。このため、検定結果が有意にならない時は対立仮説を採択すること（結論の採用）ができません。

多数のデータが得られる場合には、近似の計算が確立されていましたが、さまざまな実験などでは、多数のデータが求められるとは限りません。繰り返しをすることのできない場合もあります。そこで少ないデータ数であっても、正確な分布を求める理論が求められました。小標本論（精密標本論）がフィッシャーによって次第に形成されつつあった時代でしたが、老ピアソンの小標本論に対する態度は著しく保守的でありました。しかし、責任はフィッシャーにもあり、頑強に自説を守り、妥協を許さぬきびしさは無収束の論争を引き起こし、終生相容れませんでした。

- デイヴィッド サルツブルグ: 統計学を拓いた異才たち―経験則から科学へ進展した一世紀 ￥2,310 2006/3/1 日本経済新聞社 (2006/03).
- 悪名高き論争： `http://www.geocities.jp/ikuro_kotaro/koramu/akumei.htm`

[補記] 一様最強力検定 (Uniformly most powerful test)(URT): 帰無仮説に対する第 1 種の過誤が有意水準以下で、対立仮説のついて検出力がすべての値で最大となるもの。しかしこのような検定は一般に存在しないことが多いが、尤度比検定では UMP が求められることもある。

□□□□□ 練習問題 □□□□□

問 4.4.1

正規母集団 $N(\theta, \sigma^2)$ の平均に関する仮説検定 : 帰無仮説 $H_0 : \theta = \theta_0$, 対立仮説 $H_1 : \theta > \theta_0$ を考え、σ^2 は既知とする。標本の大きさ n を固定して、検定の棄却域を「$\overline{X} > c$ ならば、H_0 を棄却」とするならば、この検定は一様最強力であることを示せ。つまり、与えられた条件 $\theta_1 > \theta_0$ のもとで、仮説 $H_0 : \theta = \theta_0, \quad H_1 : \theta = \theta_1$ の第 1 種の過誤確率 α のもとで、第 2 種の過誤確率 β を最小とするものは、上の形で与えられることを示せ。

問 4.4.2

あるサイコロの目の表れ方を調べる。帰無仮説は「このサイコロは偏りがない」とし、対立仮説は「目の表れ方が一方に有利になっている」とする。具体的に $1, 2$ の目は $\dfrac{1}{4}, 3, 4, 5, 6$ の目は $\dfrac{1}{8}$ であるものとする。観測値としては、一回投げた目の表れ方とする。有意水準 $\alpha = 0.1$ での最強力検定を定めよ。またこのときの第 2 種の過誤確率 β をもとめよ。

問 4.4.3

2 項母集団において、標本を $n = 400$ 個抽出し、仮説 $H_0 : p = \frac{1}{2}, H_1 : p > \frac{1}{2}$ を検定する。観測値 X から $Y = \dfrac{X - np}{\sqrt{n\,p\,(1-p)}} = \dfrac{X - 200}{10}$ とおくと、帰無仮説のもとでは $N(0, 1)$ にしたがう。いま検定はもし $Y > c$ ならば、H_0 を棄却するものと定める。たとえば $\alpha = 0.05$ ならば、$c = 1.645$ となるから、$X > 216.45$ である。もし実験をおこなった結果、$X = 220$ であれば、H_0 は棄却される。このとき、与えられた n, p について、帰無仮説を採択するような最小の α の値（これを p 値とよばれる）を定めよ。

4.5　カイ 2 乗検定

多項分布（k 項分布：起こり得る場合 $A_i, i = 1, 2, \cdots, k$ が k 通りある）に従う確率変数を $X = (X_1, X_2, \cdots, X_k) \sim \mathrm{Multinom}(n, p)$, ここで $\sum_i X_i = n$, $p = (p_1, p_2, \cdots, p_n) \geq 0, \sum_i p_i = 1$ とする。

$$P(X_1 = n_1, X_2 = n_2, \cdots, X_k = n_k) = \binom{n}{n_1, n_2, \cdots, n_k} p^{n_1} p^{n_2} \cdots p^{n_k} \sum_i n_i = n, n_i \geq 0$$

とくに $k = 2$ では 2 項分布の場合となる。表にまとめると

事象 A	A_1	A_2	$\cdots$	A_k	計
確率	p_1	p_2	$\cdots$	p_k	1
観測度数	X_1	X_2	$\cdots$	X_k	n
期待度数	np_1	np_2	$\cdots$	np_k	n

観測度数と期待度数のギャップをはかるために尤度比 $L_x(H_0, H_1) = \dfrac{L(H_1)}{L(H_0)}$ を考え、とくに対数尤度として $\log L_x(H_0, H_1) = \log \dfrac{L(H_1)}{L(H_0)}$ とする。順序あるいは分母と分子に注意。

いま n 回の独立試行を行ったとき、可能な k 種類の事象が結果データ：$(x_1, \cdots, x_k), \sum_j x_j = n$ として得られたものとする。また結果 i の起こる確率は p_i であるとする。結果データと起こる確率を比較して、実際にこのような生起確率にもとづく試行であったのだろうかを検定する。

[ピアソンのカイ 2 乗統計量] 適合度検定に用いる統計量：近似的に

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}$$

は自由度 $k-1$ のカイ 2 乗分布に従う。分子分母の次数が異なることに注意。ここで可能な事象は k 個あり、$i = 1, 2, \cdots, k$、観測値 $o_i = x_i$ は「データ」で、期待値 $e_i = np_i$ は、総回数に各確率を掛ければ各事象の起こりえる期待数となる。

このような形を得る理由は、対数尤度を考えると、つぎのように導かれる。$H_0 : p_i = p_i(\theta_0), H_1 : p_i(\text{任意})$

$$\begin{aligned}
2\log L_x(H_0, H_1) &= 2\log \frac{L(H_1)}{L(H_0)} = 2\sum x_i \log \hat{p}_i - 2\sum x_i \log p_i(\hat{\theta}) \\
&= 2\sum x_i \log\left(\frac{\hat{p}_i}{p_i(\theta)}\right) = 2\textstyle\sum_i o_i \log\left(\dfrac{o_i}{e_i}\right) \quad (\delta_i = o_i - e_i) \\
&= 2\textstyle\sum_i (\delta_i + e_i) \log\left(\dfrac{\delta_i + e_i}{e_i}\right) = 2\textstyle\sum_i (\delta_i + e_i)\left(\dfrac{\delta_i}{e_i} - \dfrac{\delta_i^2}{2e_i} + \cdots\right) \\
&= \textstyle\sum_i \dfrac{\delta_i^2}{e_i} = \textstyle\sum_i \dfrac{(o_i - e_i)^2}{e_i}
\end{aligned}$$

これがピアソンのカイ 2 乗統計量である。

$$\chi^2 = \text{カイ 2 乗統計量} = \sum \frac{(\text{観測度数} - \text{期待度数})^2}{\text{期待度数}} \tag{4.5.1}$$

(注)$\log(1 + x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 + \cdots$ を用いている。

□□□□□　練習問題　□□□□□

問 4.5.1

帰無仮説として、確率変数 X の分布は $P(X = 1) = 0.5$, $P(X = 2) = 0.3$, $P(X = 3) = 0.2$ とする。いま 100 個の観測をおこない、

X の値	1	2	3
度数	40	33	27

であったとする。有意水準 0.05 で帰無仮説が棄却されるかどうかの検定をおこなえ。

問 4.5.2

カイ 2 乗検定によって、つぎの 2 つの標本が同じ分布にしたがうものかどうかを有意水準 0.05 で調べよ。

	結果 A	結果 B	結果 C
標本 1	33	147	114
標本 2	67	153	86

問 4.5.3

独立性を調べるための 2×2 分割表の数値が

	属性 B_1	属性 B_2
属性 A_1	a	b
属性 A_2	c	d

で与えられるとき、カイ 2 乗統計量は

$$\frac{(ad-bc)^2\,(a+b+c+d)}{(a+b)(c+d)(a+c)(b+d)}$$

となることを示せ。

補足 また各マスの度数が少ないのものが含まれる場合には、イェーツの修正が知られている。あるいは Fisher の直接確率計算法では、周辺度数を決めておいて、実現度数より偏ったものが起こる確率を計算する方法がある。

4.6　十分統計量とラオ・ブラックウェルの定理

n 個の確率変数（観測データ）$X=(X_1,X_2,\cdots,X_n)$ から統計量 $Y=Y(X)$ をつくり、未知の母数 θ を推定する。$f_\theta(x)=f(x|\theta)$ をパラメータ母数 θ のもとでの X の密度関数とする。

定義: Y が θ に対して**十分統計量** (sufficient) であるとは、つぎの形に密度関数が分解できるときをいう。

$$f_\theta(x)\ = f(x|\theta) =\ g(\theta,u(x))\,h(x) \tag{4.6.1}$$

ここで $y=u(x)$ で、 $g(\theta,y),h(x)$ はある関数とする。

□□□□□　　練習問題　　□□□□□

以下の問題 (問 4.6.1-4.6.6) では統計量 $u(X)=u(X_1,\cdots,X_n)$ がパラメータ θ の十分統計量であることを示し、分解定理の適当な g,h を定めよ。

問 4.6.1

$X_i,\ i=1,2,\cdots,n$ はポアソン分布 $Po(\theta)$: $P(X_i=x)=\dfrac{\theta^x}{x!}e^{-\theta x}$, $u(X)=X_1+X_2+\cdots+X_n$ とする。

問 4.6.2

$X_i,\ i=1,2,\cdots,n$ の密度関数の形が、$f_\theta(x)=A(\theta)\,B(x),\ 0<x<\theta,=0$(その他)。ただし $\theta>0$。$u(X)=\max_{i=1,2,\cdots,n}X_i$ とする。

特別な場合の単位区間上の一様分布として、$X_i,\ i=1,2,\cdots,n$ の密度関数が $A(\theta)=1/\theta, B(x)=1,\ 0<x<\theta$ の場合。

問 4.6.3

$X_i,\ i=1,2,\cdots,n$ が幾何分布、$P(X_i=x)=\theta\,(1-\theta)^{x-1},\ \ x=1,2,\cdots$。つまり、$\theta$ がベルヌーイ試行の成功の確率で、$x-1$ 回続けて失敗が起こり、x 回目に初めて成功する確率。いいかえると初めて成功した回数を表す確率密度。$u(X)=\sum_{i=1}^n X_i$ とする。

問 4.6.4

$X_i,\ i=1,2,\cdots,n$ の指数分布 $\mathrm{Expo}(\theta)$; $f_\theta(x)=\dfrac{1}{\theta}e^{-\theta x},\ \ x>0$。$u(X)=\sum_{i=1}^n X_i$ とする。

問 4.6.5

$X_i,\ i=1,2,\cdots,n$ がベータ分布 $\mathrm{Beta}(\theta,2)$; $f_\theta(x)=\dfrac{1}{\mathrm{B}(\theta,2)}x^{\theta-1}(1-x),\ \ 0<x<1$。$u(X)=\Pi_{i=1}^n X_i$。

問 4.6.6

$X_i,\ i=1,2,\cdots,n$ がガンマ分布 $\mathrm{Gamma}(\theta)$; $f_\theta(x)=\dfrac{1}{\Gamma(\theta)}x^{\theta-1}e^{-x},\ \ 0<x<\infty$。$u(X)=\Pi_{i=1}^n X_i$。

問 4.6.7

十分統計量の概念は、個々の観測値 $X_1,X_2,\cdots,X_n$ にもとづく決定をおこなう場合と、一方、和 $Y=X_1+X_2+\cdots+X_n$ の値による決定を行う場合と比べて同等な情報を得ることができることを意味する。独立同一分布のベルヌーイ試行の場合に証明せよ。つまり条件つき確率の計算で $P_\theta(X_1=x_1,\cdots,X_n=x_n|X_1+\cdots+X_n=y)=\dfrac{P_\theta(X_1=x_1,\cdots,X_n=x_n,Y=y)}{P_\theta(Y=y)}=\dfrac{P_\theta(Y=y)}{P_\theta(Y=y)}=1$

ラオ・ブラックウェルの定理

$X_i,i=1,2,\cdots,n$ は独立同一分布で, その密度 $f_\theta(x)$ とする。$Y_1=u_1(X_1,X_2,\cdots,X_n)$ を θ の十分統計量とし、$Y_2=u_2(X_1,X_2,\cdots,X_n)$ を θ の不偏推定量であるならば、

1. $\mathrm{Var}[E(Y_2\,|\,Y_1)]\le \mathrm{Var}Y_2$
2. $E[E(Y_2\,|\,Y_1)]=\theta$

が成り立つ。言いかえると、θ の最小分散不偏推定量を求めるには、十分統計量 Y_1 の関数のみに制限してよいことを示している。

$Y_1=u_1(X_1,X_2,\cdots,X_n)$ を θ の十分統計量とする。θ の最尤推定量 $\hat\theta$ が一意であるならば、この推定量 $\hat\theta$ は Y_1 の関数となる

問 4.6.8

条件付き期待値の性質；

$$
\begin{aligned}
&(1)\ E[Y]=E\left[E[Y\,|\,X]\right]\\
&(2)\ E\left[\{Y-E[Y\,|\,X]\}\{E[Y\,|\,X]-E[Y]\}\right]=0\\
&(3)\ \mathrm{Var}(X)=E[\mathrm{Var}(X\,|\,Y)+\mathrm{Var}(E[X\,|\,Y])]\\
&(4)\ \mathrm{Var}(Y)\ge \mathrm{Var}\left[E[Y\,|\,X]\right]\quad \text{等号はある関数 } h \text{ で } Y=h(X) \text{ のとき。}
\end{aligned}
$$

を示せ。

定義： 母数 θ の十分推定量 Y が**完備** (complete) であるとは、自明でない不偏推定量が存在しないとき。いいかえると、$E[g(Y)]=0,\forall\theta$ ならば、$P_\theta(g(Y)=0)=1,\ \forall\theta$ となること。さらに言いかえると、2 つの不偏推定量 $\phi(Y),\psi(Y)$ があるとき、$E_\theta[\phi(Y)-\psi(Y)]=0\ \forall\theta$ ならば、$\phi(Y)=\psi(Y)$ である。

レーマン・シェッフェの定理： $Y=u(X_1,X_2,\cdots,X_n)$ が θ の完備十分統計量であるとする。Y による θ の推定量 $\phi(Y)$ が不偏であるならば、すべての可能な θ の不偏推定量の中で、$\phi(Y)$ が最小分散となる。つまり

一様最小分散不偏推定量 (uniformly minimum variance unbiased estimator, UMVUE) である。一様の用語はすべての θ についての意味である。

定義：密度関数が**指数族** (exponential family) とはつぎの形をしているとき：

$$f_\theta(x) = a(\theta)b(x)\exp\left[\sum_{j=1}^{m} p_j(\theta)K_j(x)\right]$$

問 4.6.9

密度関数 $f_\theta(x)$ が指数族 (簡単のために $m=1$) であるとき、$x=(x_1,x_2,\cdots,x_n)$ において、$Y=\sum_{j=1}^{n} K(X_j)$ は完備十分統計量であることを示せ。

問 4.6.10

確率変数 X はすべての θ で平均が 0 であるとする。たとえば、区間 $(-\theta,\theta)$ 上の一様分布であったり、正規分布 $N(0,\theta)$ が例である。θ の十分統計量であるが、完備ではないものの例を挙げよ。

問 4.6.11

$f_\theta(x) = \exp[-(x-\theta)], \theta < x < \infty, = 0,$ その他 とする。$Y_1 = \min_i X_i$ は θ の完備十分統計量である。また θ の UMVUE（uniformly minimum variance unbiased estimator, 一様最小分散不偏推定量、ラオ=ブラックウェルの定理とレーマン=シェッフェの定理参照）をもとめよ。

問 4.6.12

$\theta >$ に対して、$f_\theta(x) = \theta x^{\theta-1},\ 0<x<1$ とおく。$u(X_1,\cdots,X_n) = \left[\Pi_{i=1}^{n} X_i\right]^{1/n}$ は θ の完備十分統計量であり、最尤推定量 $\hat\theta$ は $u(X_1,\cdots,X_n)$ の関数である。

問 4.6.13

$\theta >$ に対して、$f_\theta(x) = \theta^2 x \exp[-x\theta],\ x>0$ とおくとこれは指数族に属し、$Y=\sum_{i=1}^{n} X_i$ は θ の完備十分統計量となる。θ のもとで、$1/Y$ の期待値を計算せよ。この結果から、θ の UMVUE をもとめよ。

問 4.6.14

$X_i \sim \mathrm{Binom}(1,\theta), i=1,2,\cdots,n$, $Y_1 = \sum_{i=1}^{n} X_i \sim \mathrm{Binom}(n,\theta)$ はいずれも指数族である。さらに Y_1 は θ の完備十分統計量であり、Y_1/n は θ の UMVUE であることを示せ。また $Y_2 = \dfrac{X_1+X_2}{2}$ とするとき、$E(Y_2 \mid Y_1)$ をもとめよ。

問 4.6.15

$X_i \sim N(0,\theta), i=1,2,\cdots,n$ は独立であるとすると、$Y=\sum X_i^2$ は θ の完備十分統計量であることを示せ。Y/θ の分布をもとめ、これから θ^2 の UMVUE をもとめよ。ヒント：レーマン=シェッフェの定理

問 4.6.16

$X_1,X_2,\cdots,X_n$ は iid で, ポアソン分布 $\mathrm{Po}(\theta), \theta>0$ とする。このとき $Y=\sum_{i=1}^{n} X_i$ は θ の完備十分統計量となる。示性関数 $I = I_{\{X_1\le 1\}}$ について

(a) $E(I \mid Y)$ は、$P(X_1\le 1) = (1+\theta)\exp(-\theta)$ の UMVUE であることを示せ。
(b) もし $y>0$ ならば、$Y=y$ が与えられたとき、X_1 の条件付き分布は Binom $\left(y, \dfrac{1}{n}\right)$ となる。
(c)

$$E[I \mid Y] = \left(\frac{n-1}{n}\right)^{Y}\left[1+\frac{Y}{n-1}\right]$$

問 4.6.17

$\theta = (\theta_1,\theta_2)$ とおき、$f_\theta(x) = \dfrac{1}{\theta_2}\exp(\dfrac{x-\theta_1}{\theta_2}), x>\theta_1$ とする。ただし θ_1 は任意の実数で、$\theta_2>0$ とする。このとき、統計量 $(\min_i X_i, \sum_{i=1}^{n} X_i)$ は θ の十分統計量であることを示せ。

4.7　ベイズ推定

数理統計学はベイズ統計学と非ベイズ統計学に分けられる。前者は,Tomas Bayes(1702〜1761) らによる考え方で、特徴として未知パラメータ θ の分布（これを事前分布とよぶ）および観測値 x によるパラメータの尤度の両者を、「ベイズの定理」として結合させ、パラメータの事後分布を導出する。つまり何らの予めの情報があれば、観測値を得ることでより正確な予測ができると考える。一方非ベイズ統計学は、このような事前の情報などを仮定できないとみなすものである。伝統的統計学、古典的統計学ともよばれる。著名な統計学者 K・ピアソン（1857〜1936）と R・A・フィシャー（1890〜1962）のいずれも非ベイズ統計学の立場をとっている。

まず確率のベイズの定理を述べる。事象 $A_1, A_2, \cdots, A_n$ と B において、条件 1) $A_i \cap A_j = \emptyset (i \neq j), \cup_i A_i = \Omega$ で、同時に起こることはなく、互いに素なる事象。条件 2) 確率 $P(A_i)$ と条件付き確率 $P(B|A_i)$, $i = 1, 2, \cdots, n$ が与えられている。この条件 1）, 2）から、条件付き確率 $P(A_i|B)$ が次の式で計算できる；

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_j P(A_j)P(B|A_j)}, \quad i = 1, 2, \cdots, n$$

このベイズの定理では、条件付き確率の計算であるが、未知のパラメータに対する推測では、つぎの連続型密度関数で表現される：θ を未知とし、その事前確率密度を $p(\theta)$ とするとき、観測値 $x = (x_1, x_2, \cdots, x_n)$ に対する θ の事後分布は

$$p(\theta|x) = \frac{f(x|\theta)p(\theta)}{\int f(x|\theta')p(\theta')d\theta'} \quad \text{または} \quad p(\theta|x) \propto f(x|\theta)p(\theta)$$

上式の前半で、分母に当たる部分は、確率が 1 になるように調整するだけであるから、これを省略して、後半のように比例記号 ($\propto$) を用いて表すこともある。

パラメータ推定におけるベイズ統計学による接近法は、**損失関数** (loss function)$L(\theta, a)$ をもちいる。未知パラメーラ θ に値 a で推定するとき、正確さの基準尺度として、この損失関数を小さくしようと考える。もし $\hat{\theta}$ を $\mathbb{E}[L(\theta, \hat{\theta})]$ (ここで期待値 $\mathbb{E}$ は事後分布 $p(\theta|x)$ に関してとる）が最小となるならば、ベイズ危険 (Bayes risk) を最小とする**ベイズ推定**とよぶ。一般に事後分布の計算は簡単ではない。

損失関数の期待値：

$$E[L(\theta,\, a)] = \int L(\underline{\theta}, a)\, p(\underline{\theta}\,|x_1, \cdots, x_n) d\underline{\theta}$$

(1)　2 乗損失関数：$L(\theta,\, a) = (\theta - a)^2$

$$E[L(\theta,\, a)] = \int (a - \underline{\theta})^2 p(\underline{\theta}\,|x_1, \cdots, x_n) d\underline{\theta}$$

(2)　絶対損失関数：$L(\theta,\, a) = |\theta - a| = \begin{cases} \theta - a, & \text{it}\, \theta \geq a \\ a - \theta, & \text{it}\, \theta < a \end{cases}$

$$E[L(\theta,\, a)] = \int_a^{\infty} (\underline{\theta} - a) p(\underline{\theta}\,|x_1, \cdots, x_n) d\underline{\theta} + \int_{-\infty}^{a} (a - \underline{\theta}) p(\underline{\theta}\,|x_1, \cdots, x_n) d\underline{\theta}$$

未知パラメータ θ を推定する場合には、観測値 $X = x$ を得るとき、$X \sim f_\theta(x)$ であり、θ を推定する決定に $\delta(x)$ と表し、未知のパラメータと決定とのズレに非負値の損失 $L(\theta, \delta(x))$ を考える。パラメータ θ の密度関数を $h(\theta)$ とする。つぎのベイズリスク、あるいは平均損失

$$B(\delta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\theta) f_\theta(x) L(\theta, \delta(x))\, d\theta\, dx$$

を最小にする解 δ をベイズ解とよぶ。ここで θ と x の同時密度は $h(\theta)f_\theta(x) = h(\theta)f(x|\theta) = f(x)f(\theta|x)$ となるから、

$$B(\delta) = \int_{-\infty}^{\infty} f(x) \left[\int_{-\infty}^{\infty} L(\theta, \delta(x)) f(\theta\,|\,x)\, d\theta \right] dx$$

と書き直せる。密度関数 $f(x) \geq 0$ だから、ベイズリスクを求めるには各 x で積分 $\int_{-\infty}^{\infty} L(\theta, \delta(x)) f(\theta\,|\,x)\, d\theta$ を最小にすればよい。この決定 δ を θ のベイズ推定とよぶ。同様に θ の関数と考えて $\gamma(\theta)$ として

$\int_{-\infty}^{\infty} L(\gamma(\theta), \delta(x)) f(\theta \mid x)\, d\theta$ を最小とする。簡潔に言い表すならば、X を観測して、関数 $g(X)$ によって確率変数 Y を推定しようとし、最小化基準として $E[L(Y, g(X))]$ を考える。

(1)　2 乗損失関数の場合：もし損失関数として $L(Y, g(X)) = (Y - g(X))^2$ であれば、期待値は

$$E[(Y - g(X))^2] = \int_{-\infty}^{\infty} E[(Y - g(X))^2 \mid X = x]\, f(x)\, dx$$

で、これを最小化する $g(X)$ を探すことである。2 乗損失関数は $z = g(x)$ とおくと、z の 2 次式だから、$z^2 - 2E[Y \mid X = x]z + E[Y^2 \mid X = x] = (z - E[Y \mid X = x])^2 - (E[Y \mid X = x])^2 + E[Y^2 \mid X = x]$ の最小化であり、この変形から、条件つき期待値 $z = E[Y \mid X = x]$ が最小値を与える。

したがってベイズ推定量は、統計的決定問題の形からは

$$\delta(x) = E[\theta \mid X = x] = \int_{-\infty}^{\infty} \theta\, f(\theta \mid x)\, d\theta$$

そして

$$f(\theta \mid x) = \frac{f(\theta, x)}{f(x)} = \frac{h(\theta)\, f(x \mid \theta)}{f(x)} = \frac{h(\theta)\, f_\theta(x)}{f(x)},$$
$$f(x) = \int_{-\infty}^{\infty} f(\theta, x)\, d\theta = \int_{-\infty}^{\infty} h(\theta)\, f(x \mid \theta)\, d\theta = \int_{-\infty}^{\infty} h(\theta)\, f_\theta(x)\, d\theta$$

したがって

$$\delta(x) = \frac{1}{f(x)} \int_{-\infty}^{\infty} \theta\, h(\theta)\, f_\theta(x)\, d\theta = \frac{\int_{-\infty}^{\infty} \theta\, h(\theta)\, f_\theta(x)\, d\theta}{\int_{-\infty}^{\infty} h(\theta)\, f_\theta(x)\, d\theta}$$

が得られる。

(2)　絶対損失関数の場合：$L(Y, g(X)) = |Y - g(X)|$ であれば、期待値は、$c = g(x)$ とおいて

$$E[|Y - g(X)|] = \int_{-\infty}^{\infty} E\big[|Y - g(X)| \bigm| X = x\big]\, f(x)\, dx$$
$$= \int_{-\infty}^{c} E[Y - g(X) \mid X = x]\, f(x)\, dx + \int_{c}^{\infty} E[-Y + g(X) \mid X = x]\, f(x)\, dx$$

より簡単に表現するために 目的関数として $E[|W - c|]$ において c を動かして最小化する問題としてみよう。すなわち

$$E[|W - c|] = \int_{-\infty}^{c} (c - w) f(w)\, dw + \int_{c}^{\infty} (w - c) f(w)\, dw$$

これを c で微分して、極値を求める。$(c - c)f(c) + \int_{-\infty}^{c} f(w)\, dw + (-1)(c - c)f(c) + \int_{c}^{\infty} (-1) f(w)\, dw = \int_{-\infty}^{c} f(w)\, dw - \int_{c}^{\infty} f(w)\, dw = 0$ から、$\int_{-\infty}^{c} f(w)\, dw = \int_{c}^{\infty} f(w)\, dw$ 分布関数では $F(c) = 1 - F(c)$, $F(c) = \frac{1}{2}$, $c = F^{-1}(\frac{1}{2})$ であるから、ちょうど分布関数の中間 50% 値、つまり中央値が極値、最小値を与えることがわかる。したがって今の場合には $E[|Y - g(X)|]$ を最小にするような $g(x)$ は、$X = x$ を与えた Y の条件つき分布関数 $F_{Y|X}$ の中央値 $g(x) = F_{Y|X}^{-1}\big(\frac{1}{2} \bigm| X = x\big)$ である。

□□□□□　練習問題　□□□□□

問 4.7.1

偏りのあるコイン投げを n 回繰り返す。 $x_i = 1, 0$ を表か裏とする。確率に関しては全く情報をもたないとするから、$p(\theta) = 1, 0 \leq \theta \leq 1$ と仮定する。このとき表の出た回数を $t = \sum_{i=1}^{n} x_i$ とするとき、θ の事後分布を求めよ。

問 4.7.2

未知の真値 θ を値 a で推定するとき、損失関数 $L(\theta, a)$ が次で与えられるとき、ベイズ危険を求めよ。(1) 2 乗誤差損失 $L(\theta, a) = (a - \theta)^2$　(2) 絶対誤差損失 $L(\theta, a) = |a - \theta|$　(ヒント、(1) 事後平均と (2) 事後中央値)

問 4.7.3

$X_1, \cdots, X_n$ をパラメータ λ のポアソン分布 $\mathrm{Po}(\lambda)$、λ をパラメータ 1 の指数分布 $\mathrm{Expo}(1)$ とする。(1) λ の事後分布は $(\sum_i X_i + 1, n+1)$ のガンマ分布になることを示せ。　(2) 2 乗誤差損失では $\hat{\theta} = (\sum_i x_i + 1)/(n+1)$ で与えられることを示せ。　(3) 絶対誤差損失ではどうなるか。

問 4.7.4

$X \sim \mathrm{Binom}(n, \theta)$ とし、事前分布は $\theta \sim \mathrm{Beta}(r, s)$ で、$h(\theta) = \dfrac{\theta^{r-1}(1-\theta)^{s-1}}{\mathrm{B}(r,s)}$ とする。2 乗損失によるベイズ推定は

$$\delta(x) = \frac{r+x}{r+s+n}, \quad x = 0, 1, 2, \cdots, n$$

となることを示せ。

問 4.7.5

リスク関数 $R_\delta(\theta)$ を、パラメータが θ のとき、δ を用いた平均損失とするとき、推定に対するリスク関数が

$$\frac{1}{(r+s+n)^2}\left[\left((r+s)^2 - n\right)\theta^2 + (n - 2r\,(r+s))\,\theta + r^2\right]$$

となることを示せ。

問 4.7.6

もし $r = s = \dfrac{\sqrt{n}}{2}$ ならば、$R_\delta(\theta)$ は定数で、θ と無関係となる。

問 4.7.7

ベイズ推定量 δ が定数であるならば、*minmax* である。つまり δ は $\max_\theta R_\delta(\theta)$ を最小にするものである。

第 5 章

多変量のデータ解析

5.1 相関係数

定義: X, Y は有限な平均、分散をもつとし、$\mu_X = EX, \mu_Y = E[Y], \sigma_X^2 = \mathrm{Var}(X), \sigma_Y^2 = \mathrm{Var}(Y)$ とする。このとき、$E(XY)$ は有限な値となる。なぜなら $2XY$ は $X^2 + Y^2$ で評価できるから。

共分散 (covariance) とは

$$\mathrm{Cov}(X, Y) = \sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - \mu_X \mu_Y \tag{5.1.1}$$

で定める。(X, Y) 組の**相関係数** (correlation coefficient) とは

$$\rho_{XY} = \rho = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\sigma_X^2 \sigma_Y^2}} = \mathrm{Cov}(\tilde{X}, \tilde{Y}) = E[\tilde{X}\tilde{Y}] = \rho_{\tilde{X}\tilde{Y}} \tag{5.1.2}$$

ここで $\tilde{X} = \dfrac{X - \mu_X}{\sqrt{\sigma_X^2}} = \dfrac{X - \mu_X}{\sigma_X}, \tilde{Y} = \dfrac{Y - \mu_Y}{\sqrt{\sigma_Y^2}} = \dfrac{Y - \mu_Y}{\sigma_Y}$. また $\sigma_{XY} = \rho\sigma_X\sigma_Y$ である.

□□□□□　練習問題　□□□□□

問 5.1.1

いま Y を $aX + b$ の形で推定する。実際 $Y - \mu_Y$ を $c(X - \mu_X) + d$ で推定するとき、c, d を変化させて最も期待 2 乗偏差 (mean square deviation);

$$E[(\{Y - \mu_Y\} - \{c(X - \mu_X) + d\})^2] \to \min_{c,d}$$

が小さくなる c, d の値とそのときの最小値を求めよ。

問 5.1.2

(1) X, Y が独立であれば、無相関 (uncorrelated)$\rho = 0$ であるが、(2) 逆は必ずしも成り立たない。(3) 成り立つ場合はどんな場合か？ (1) の命題を示し、(2) の反例として、$X = \cos\theta, Y = \sin\theta, \theta \sim U[0, 2\pi]$ を確認せよ。(3) は正規分布が該当する。

問 5.1.3

(1) 平面上に n 個の点 $(x_i, y_i); i = 1, 2, \cdots, n$ があるとき、ここに直線 $y = ax + b$ を引き、$\sum_i (y_i - (ax_i + b))^2$ を最小にするような係数 a, b を定めよ。(2) $P(X = x_i) = 1/n, i = 1, 2, \cdots, n$ のとき直線が

$$y - \mu_Y = \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_X)$$

となることを示せ。これは最小 2 乗法とよばれる。

問 5.1.4

次の問 は X, Y が独立で、$\mu_X = E(X), \mu_Y = E(Y), \sigma_X^2 = Var(X), \sigma_Y^2 = Var(Y)$ として解け。また X と Y との相関係数 ρ は既知であるとせよ。

(1) 積 XY の分散をもとめよ。
(2) a, b を任意の実数とするとき、$aX + bY$ の分散を計算せよ。
(3) X と $X + Y$ との共分散をもとめよ。
(4) X と $X + Y$ との相関係数をもとめよ。
(5) X と XY との共分散をもとめよ。

問 5.1.5

コーシー=シュワルツの不等式：

$$\left(\sum_{i=1}^{n} x_i y_i\right)^2 \leqq \left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right)$$

はどんな条件のときに等号が成り立つか？ またこのときは点 $(x_i, y_i), i = 1, 2, \cdots, n$ は散布図ではどう配置されているか？

確率変数の場合に対しては

(1) $(E[XY])^2 \leq E[X^2]\,E[Y^2]$
(2) 相関係数の値は絶対値が 1 を超えることはない。$-1 \leq \rho_{X,Y} \leq 1$

がシュワルツの不等式とよばれるが、確率変数がどういう関係にあるときか？

多変量データを分析するためには、ベクトル、行列をもちいる。そのために必要な用語のまとめをしておく。

行列： $A = (a_{ij}), i = 1, 2, \cdots, n, j = 1, 2, \cdots, m, (i, j)$ 成分を a_{ij}、 m 行, $\times\, n$ 列、

和、積、スカラー倍： $A + B = (a_{ij} + b_{ij}), AB = (\sum_k a_{ik} b_{kj}), \alpha A = (\alpha a_{ij}),$

n 次正方行列： $A = (a_{ij}), i = j = n, I_n = (\delta_{ij}); \delta_{ij} = 1/0 (i = j / i \neq j)$、単位行列

逆行列: $A^{-1} A = A A^{-1} = I_n,$

正則（非特異）: A^{-1} が定まる

正則でない（特異）: 逆行列が定まらない

行列式: 2 次の正方行列では $\det \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$, 3 次の正方行列では

$$\det \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = a_1 b_2 c_3 + a_2 b_3 c_1 + a_3 b_1 c_2 - a_1 b_3 c_2 - a_2 b_1 c_3 - a_3 b_2 c_1,$$

対称行列、 $A = {}^T A = A'$ 　(転置行列)

２次形式、

$$ax^2 + 2bxy + cy^2 = (x, y)\begin{pmatrix} a & b \\ b & c \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \boldsymbol{x}' A \boldsymbol{x} = {}^T\boldsymbol{x} A \boldsymbol{x}$$

定非負値行列、 2 次形式がすべての変数に対して、負にならない。

定正値行列、 2 次形式がすべての変数に対して、負にならなず、ゼロとなるのは、変数がすべてゼロの場合のみ。

ヘッセ行列、 多変数スカラー値関数の 2 階偏導関数からつくられる行列、もし関数が連続であれば、対称行列となる。

5.2　多変量正規分布

2 変量標準正規分布 $N(\mathbf{0},\boldsymbol{\Sigma})$ 平均ベクトル $\mathbf{0}=\begin{pmatrix}0\\0\end{pmatrix}$, 分散共分散行列 $\boldsymbol{\Sigma}=\begin{pmatrix}1&\rho\\\rho&1\end{pmatrix}$:

$$\begin{aligned}\text{密度関数}\,\phi(x,y)\quad&=\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}\left(x^2-2\rho xy+y^2\right)\right]\\&=\frac{1}{2\pi\sqrt{\det|\boldsymbol{\Sigma}|}}\exp\left[-\frac{1}{2}\,(x,y)\,\boldsymbol{\Sigma}^{-1}\begin{pmatrix}x\\y\end{pmatrix}\right]\end{aligned}\tag{5.2.1}$$

ここで $\det|A|=\det\begin{vmatrix}a&b\\c&d\end{vmatrix}=ad-bc$ は行列式（実数値）、$A^{-1}=\dfrac{1}{ad-bc}\begin{pmatrix}d&-b\\-c&a\end{pmatrix}$ は逆行列で、$(x,y)\;A^{-1}\begin{pmatrix}x\\y\end{pmatrix}$ は 2 次形式。

2 変量正規分布 $N(\boldsymbol{\mu},\boldsymbol{A})$:平均ベクトル $\boldsymbol{\mu}=\begin{pmatrix}\mu_x\\\mu_y\end{pmatrix}$, 分散共分散行列 ; $\boldsymbol{A}=\begin{pmatrix}\sigma_x^2&\rho\sigma_x\sigma_y\\\rho\sigma_x\sigma_y&\sigma_y^2\end{pmatrix}$ では標準の場合での (x,y) 値を $(x-\mu_x)/\sigma_x,(y-\mu_y)/\sigma_y$ に置き換えたもの。この確率母関数は

$$M_{X,Y}(t,s)=E[\exp(tX+sY)]=\exp\left(t\mu_x+s\mu_y+\frac{1}{2}\{\sigma_x^2+2\rho\sigma_x\sigma_y+\sigma_y^2\}\right).$$

また条件つき密度関数は

$$\phi_{Y|X}(y|x)=\frac{\phi(x,y)}{\phi(x)}=\frac{1}{\sqrt{2\pi(1-\rho^2)\sigma_Y^2}}\exp\left[-\frac{(y-\rho\frac{\sigma_Y}{\sigma_X}\,x)^2}{2(1-\rho^2)\sigma_Y^2}\right].$$

これから、条件つき平均と分散は

$$E[Y\,|\,X]=\rho\frac{\sigma_Y}{\sigma_X}X,\qquad \mathrm{Var}(Y\,|\,X)=(1-\rho^2)\sigma_Y^2,:\ (X\text{ への依存は}\rho\text{に含まれる})$$

一般の n 次元ベクトル正規変量（正規分布に従う確率変数）については、平均は n 次元ベクトル、分散共分散行列は $n\times n$ 行列で定めたもの。$X=(X_1,X_2,\cdots,X_n)$ では、平均の第 i 成分は $E[X_i]$, 分散共分散の第 (i,j) 成分は $E[X_i\,X_j]$ とおいたもの。

定義：同時（結合）分布の確率母関数を次で定められる。確率ベクトル $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)$ に対し、$\underline{t}=(t_1,t_2,\cdots,t_n),\underline{t}'$ は $\underline{t}$ の転置ベクトル。

$$M_{\boldsymbol{X}}(\underline{t})=M(t_1,t_2,\cdots,t_n)=E[\exp(t_1X_1+t_2X_2+\cdots+t_nX_n)]=E[\exp(\underline{t}'\boldsymbol{X})]$$

多変量正規分布 $\mathrm{N}(\boldsymbol{\mu},\boldsymbol{A})$ あるいは結合ガウス分布 $\boldsymbol{X}$ については

$$M_{\boldsymbol{X}}(t)=\exp(\underline{t}'\boldsymbol{\mu})\exp\left(\frac{1}{2}\underline{t}'A\,\underline{t}\right).$$

□□□□□　練習問題　□□□□□

問 5.2.1

任意の確率変数 $X_1,X_2,\cdots,X_n$ に対する分散共分散行列 K は非特異と仮定する。この行列は対称行列で、定正値であることを示せ。もし K が特異である場合には、どうなるか？

問 5.2.2

n 次元ベクトルの正規変量 $\underline{X}$ にたいして、非特異定数行列 A をもちいて、$\underline{Y}=A\underline{X}$ をつくると、やはり正規変量であることを示せ。

問 5.2.3

n 次元ベクトルの (結合) 正規変量 $X_1, X_2, \cdots, X_n$ にたいして、$m\,(m \le n)$ 次元ベクトルの周辺分布 $X_1, X_2, \cdots, X_m$ は (結合) 正規変量となることを示せ。　多変量正規分布の部分系も正規分布となる。

問 5.2.4

n 次元ベクトルの (結合) 正規変量 $X_1, X_2, \cdots, X_n$ にたいして、これらの線形結合でつくった分布 $c_1X_1 + c_2X_2 + \cdots + c_nX_n$ は正規分布にしたがうことを示せ。

問 5.2.5

X, Y は 2 変量正規分布とする。$\mu_X = -1, \sigma_X = 2$ とし、仮定として与えられた X にもとづく Y の最良推定、つまり平均 2 乗誤差を最小にするものは, $3X + 7$ で与えられるとする。この推定にたいする Y の平均 2 乗誤差は 28 であり、μ_Y, σ_Y と、さらに X と Y の相関係数 ρ をもとめよ。

問 5.2.6

2 変量正規分布は指数族に属することを示し、対応する完備十分統計量をもとめよ。

問 5.2.7

確率変数 X が密度関数 $f(x)$ をもつとき、確率変数 $f(X)$ の分布をどのようにして求めたらよいか。

問 5.2.8

$X \sim f_\theta(x)$ で、$Y = g(X), E_\theta(Y) = k(\theta), k'(\theta) = \dfrac{d\,k(\theta)}{d\theta}$ とする。クラメル=ラオの不等式：

$$\mathrm{Var}_\theta f(X) \geqq \frac{[k'(\theta)]^2}{E_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f_\theta(X)\right)^2\right]}$$

をもちいて、標本平均が真の平均に関する UMVUE であることをつぎの 3 つの場合について証明せよ。
(1) ベルヌーイ分布 $\mathrm{Bern}(p) = \mathrm{Binom}(1, p) = \mathrm{Binom}(1, \theta)$
(2) 分散 σ^2 が既知の正規分布 $\mathrm{N}(\mu, \sigma^2) = \mathrm{N}(\theta, \sigma^2)$
(3) ポアソン分布 $\mathrm{Po}(\lambda) = \mathrm{Po}(\theta)$

5.3　ノンパラ統計量

統計理論の解析では、母集団分布を正規分布や 2 項分布とすることがよく前提とされるが、このような具体的な分布を想定とはしない理論がノンパラメトリック (母数によらない) 推測理論である。母集団の性質をより広い範疇で適用することをめざす。日本語でやや短く呼称するときにはノンパラとすることが多い。

ウィルコクソンの検定 (1 標本問題) とは

- 中央値の検定
- 分布の仮定では密度関数が、中央値で対称であること。
- 2 つの集団を比較した正負符号とデータ差の絶対値の順位をもとに統計量を計算

分布パーセント点 ξ_α とは、分布関数 $F_X(\xi_\alpha) = P(X \le \xi_\alpha) = \alpha$, いいかえると逆関数より、$\xi_\alpha = F_X^{-1}(\alpha)$ となる点で、$\alpha = \frac{1}{2} = 0.5$ 、$\xi_{0.5}$ が中央値に他ならない。この値に関する検定を調べる。平均ではなく、中央値は頑健性があり、ベイズ推定においても絶対偏差を最小とする優れた性質をもつ。

帰無仮説を中央値の値が m であること：H_0: $\xi_{.5} = m$,　　対立仮説は片側検定として m 以上であること：H_1: $\xi_{.5} > m$

観測データ： $X_1, \cdots, X_n$

検定統計量:

$$W = \sum_{i=1}^{n} Z_i R_i$$

ここで R_i は m との差の絶対値 $|X_i - m|$ を並び替えた順位、Z_i は $X_i - m$ の符号、正負で ± 1 とした値。

たとえば、標本数が $n = 5$ で, 各データと m との差を計算し、

$$X_1 - m = 2.7,\ X_2 - m = -1.3,\ X_3 - m = -0.3,\ X_4 - m = -3.2,\ X_5 - m = 2.4$$

であれば、この値の絶対値の大きさで全てを並び替えて $|-0.3| < |-1.3| < |2.4| < |2.7| < |-3.2|$ となっているから、

$$|X_3 - m| < |X_2 - m| < |X_5 - m| < |X_1 - m| < |X_4 - m|$$

順位は、

$$R_3 = 1, \quad R_2 = 2, \quad R_5 = 3, \quad R_1 = 4, \quad R_4 = 5$$

で、その差値の正負の符号は、

$$Z_3 = -1, \quad Z_2 = -1, \quad Z_5 = 1, \quad Z_1 = 1, \quad Z_4 = -1$$

したがって

$$W_n = \sum_{i=1}^{n} Z_i R_i = (-1) + (-2) + 3 + 4 + (-5) = -1$$

帰無仮説のもとでは

$$E(W) = 0, \qquad \mathrm{Var} W = \frac{n(n+1)(2n+1)}{6} \tag{5.3.1}$$

となる。

もし組となる 2 集団 $(X_i, Y_i), i = 1, 2, \cdots, n$ を比較する場合であれば、

$$W_n = \sum_i R_i^+ I\{X_i - Y_i > 0\} = nU_n^{(1)} + \binom{n}{2} U_n^{(2)}$$

$U_n^1 = n^{-1} \sum_i I(X_i - Y_i > 0)$, $U_n^2 = \binom{n}{2}^{-1} \sum_{i<j} I((X_i - Y_i) + (X_j - Y_j) > 0)$ とする。cf:T.S.Ferguson,"U-statistics", Notes for stattistics 200C, Springer 2005. ここで R_i^+ は絶対値 $|X_i - Y_i|$ の値についての n 個中の順位で小さいときには全ての場合を計算できる：

(1) $n = 3$ のとき、

W 値	-6	-4	-2	0	2	4	6	計
場合の数	1	1	1	2	1	1	1	$2^3 = 8$

(2) $n = 4$ のとき、

W 値	-10	-8	-6	-4	-2	0	2	4	6	8	10	計
場合の数	1	1	1	2	2	2	2	2	1	1	1	$2^4 = 16$

(3) $n = 5$ のとき、

W 値	-15	-13	-11	-9	-7	-5	-3	-1	1	3	5	7	9	11	13	15	計
場合の数	1	1	1	2	2	3	3	3	3	3	3	2	2	1	1	1	$2^5 = 32$

正規近似：もし標本数が大きければ、中心極限定理が適用できる（Liapounov 条件）から、

$$\frac{W - E(W)}{\sqrt{\mathrm{Var} W}} \sim N(0,1)$$

により、検定をおこなう。

定理 5.3.1

２標本問題では、２つの母集団（第 1 集団 X、第 2 集団 Y）において、大きさを m, n とし、絶対値の値を並び代えて、それぞれの順位の和を T_X, T_Y とする。順位データは $1, 2, \cdots, m+n$ である。T_X で第 1 集団の順位和、T_Y で第 2 集団の順位和とすると、

(1) $T_X + T_Y = \dfrac{(m+n)(m+n+1)}{2}$

(2) $E[T_X] = \dfrac{m\,(m+n+1)}{2}$

(3) $\mathrm{Var}[T_X] = \dfrac{mn\,(m+n+1)}{12}$

が成り立つ。

2 母集団分布 F_X, F_Y を比較する検定は、

$$\text{帰無仮説 } H_0 : \text{すべての } z \text{ で } F_X(z) = F_Y(z)$$
$$\text{対立仮説 } H_1 : \text{ある } z \text{ で } F_X(z) \neq F_Y(z)$$

分布に関する厳しい限定をせず、この統計量 T_X をもちい、

$$H_0 \text{ を棄却} \iff |T_X - E[T_X]| \text{ が大きい} \iff |T_X - E[T_X]| \geq k$$

ここで k の値は T_X の漸近的正規性から、与えられる有意水準と正規分布のパーセント点をもちいて定めばよい。

□□□□□　練習問題　□□□□□

問 5.3.1

中央値 m に関する符号検定で、$n = 12$ 個の観測値を得、帰無仮説 $H_0 : m = 40$ と対立仮説 $H_1 : m > 40$ を考える。統計値として $Y = \{40$ 以下である観測値の個数 $\}$ とし、$Y \leqq c$ ならば、H_0 を棄却する。c と $p = F(40)$ を用いて、検出力 $K(p)$ をもとめ、さらに $c = 2$ のときの第 1 種過誤の確率 α をもとめよ。ここで F は観測値の分布関数で連続、強単調増加とする。

問 5.3.2

中央値 m のまわりで対称な確率変数を考える。ウィルコクソンの検定をもちい、$n = 16$ 個の観測値から仮説 $H_0 : m = 160, H_1 : m > 160$ を調べる。観測値は以下の通り：

176.9	158.3	152.1	158.8
172.4	169.8	159.7	162.7
156.6	174.5	184.4	165.2
147.8	177.8	160.1	160.5

有意水準 0.05(第 1 種の過誤確率) のもとで検定をおこなえ。

問 5.3.3

ウィルコクソンの検定は n が小さい場合には厳密な統計量 W の分布を求めることができる。$n = 1, 2, 3$ について、これをもとめよ。

参考文献

A: ホーエル,GH 「統計学入門」(浅井晃、村上正康訳)、培風館、

A: 前園宣彦 「概説確率統計」サイエンス社、１９９９年

A: 村上正康、安田正實 「統計学演習」培風館、

B: 北川敏男、稲葉三男 「統計学通論」共立出版、

B: ホーエル,GH., ポート,SC., ストーン、CJ.,「確率、統計、確率論」3 部作安田正實、柳川尭、大和元、岩本誠一訳（絶版）東京図書

B: 竹内啓 「数理統計学」東洋経済新報社、1963 年

B: 稲垣宣生 「数理統計学（改訂版）」裳華房、2003 年

C: ザルツブルグ,D 「統計学を拓いた異才たち」日経ビジネス文庫、2010 年竹内恵行、熊谷悦生訳

D: Cramer,H. Mathematical methods of Statistics, Princeton Univ Press, 1946.

D: Mood,A. Graybill,F. Boes,DC., Introduction to the theory of Statistics, MacGrow-Hill.

D: Feller,W.; Introduction to Probability Theory and Its Applications, Vol. I, 1968 , Vol.II, 1966, John Wiley & Sons,Inc.

まとめ

● 統計的推測の推定量、統計量

○ 正規母集団の母平均

○ ２標本正規母集団の母平均の差

○ 正規母集団の母分散

○ ２標本正規母集団の分散の比の区間推定、等分散の検定

○ ２次元正規母集団の母相関係数の点推定、区間推定

○ 回帰直線に関する統計的推測

○ ２項母集団の母比率の点推定、区間推定、大標本近似

○ ２標本での２項母集団の母比率の差

＝＝＝＝＝　練習問題の解答　＝＝＝＝＝

第 1.1 節（1.1 ページ）==

問 1.1.1　(略)

問 1.1.2　(略)

問 1.1.3　(略)

第 1.2 節（1.2 ページ）==

問 1.2.1　密度関数 $f(x)$ から、その分布関数 $F(x)$ を求めておく。分布関数は $F(x) = \int_{-\infty}^{t} f(x)\,dx = \int_0^t f(x)\,dx = \int_0^t 6x^5\,dx = t^6\,(0 \le t \le 1)$. また $F(t) = 0$, $(-\infty < t < 0)$, $F(t) = 1$, $(1 \le t < \infty)$ が得られる。事象 $\max\{X, Y, Z\} \le t$ は同時事象 $\{X \le t\} \cap \{Y \le t\} \cap \{Z \le t\} = \{X \le t, Y \le t, Z \le t\}$ と同じである。また独立であることから、同時事象はそれぞれの確率の積になり、また同一分布という仮定から $P(X \le t) = P(Y \le t) = P(Z \le t) = F(t)$ である。したがって分布関数 $F_M(t) = P(\max\{X, Y, Z\} \le t)$ は $P(\max\{X, Y, Z\} \le t) = F_M(t)^3 = (t^6)^3 = t^{18}$, $(0 \le t < 1)$ となる。ただし $F_M(t) = 0\,(-\infty < t < 1)$, $= 1\,(1 \le t\infty)$.

(注) 最小値の場合には、補事象を考え、$P(\min\{X, Y, Z\} \le t) = 1 - P(\overline{\min\{X, Y, Z\} \le t}) = 1 - P(\min\{X, Y, Z\} \ge t) = 1 - P(X \ge t, Y \ge t, Z \ge t) = 1 - \{1 - F(t)\}^3$ として計算する。

問 1.2.2　事象 $Z \le t$ は $X > 0$ に注意して、$Y \le tX$ と同じ。もしそうでない場合には正と負の場合に分けて考える。正負の値によって、不等号の向きが異なるので十分に注意して計算しなければならない。したがって (X, Y) の同時分布が独立性から、$f_{X,Y}(x, y) = f_X(x) f_Y(y) = e^{-x} e^{-y}$, $x, y > 0$ となるから、$F_Z(t) = P(Z \le t) = \iint_{Y/X \le t} f_{X,Y}\,dxdy = \iint_{Y \le tX} f_{X,Y}\,dxdy = \int_0^\infty dx \int_0^{tx} f_{X,Y}\,dy = \int_0^\infty e^{-x} dx \int_0^{tx} e^{-y}\,dy = \int_0^\infty e^{-x}(1 - e^{-tx})dx = 1 - (1 + x)^{-1}$, $(t \ge 0)$. ただし, $F_Z(t) = 0, (t < 0)$ これを微分すれば、密度関数 $f_Z(t) = (t + 1)^2$, $(t \ge 0), = 0\,(t < 0)$ が得られる。

問 1.2.3　X のとり得る値から,Y のとり得る値を定める。このためには逆関数の関係をもちいる。事象の関係：$\{Y = y\} = \{g(X) = y\} = \{X \in g^{-1}(y)\}$. よって $p_Y(y) = P(Y = y) = P(X \in g^{-1}(y)) = \sum_{\{x : g(x) = y\}} p_X(x) = \sum_{\{x_i : g(x_i) = y\}} 1/n = \frac{1}{n} \times$ [集合 $\{x_i : g(x_i) = y\}$ を満たす個数]

問 1.2.4　密度関数の積分値は確率の全確率すなわち 1 になるから、$1 = \int_{-\infty}^{\infty} f(x)\,dx = \int_0^b ax^2\,dx = \frac{ab^3}{3}$ から、$ab^3 = 3$ という関係がある。いま $Y = X^3$ の密度関数 $f_Y(y)$ をもとめるから、$y = x^3$ から　$x = y^{1/3}$, つまり, $dx = \frac{1}{3} y^{1/3-1}\,dy = \frac{dy}{3y^{2/3}}$ から、$f_Y(y)dy = f(x)dx = f(y^{1/3}) \frac{dy}{3y^{2/3}} = a(y^{1/3})^2 \frac{dy}{3y^{2/3}} = \frac{ay^{2/3}}{3y^{2/3}} dy = 1/b^3$, 範囲は $(0 < x = y^{1/3} < b)$, つまり、$0 < y < b^3$ で新しい密度関数 $f_Y(y)$ は区間 $[0, b^3]$ 上の一様分布に等しい。

問 1.2.5　与えられた条件は $f_X(x) = 1/\pi$, $(-\pi/2 \le x \le \pi/2), = 0$(その他) である。ここでは正接関数 $y = \tan x$ の逆関数 $x = \tan^{-1} y = \arctan y$,（ただし y 変域 $-\infty < y < \infty$ に対して、x 変域 $-\pi/2 < x < \pi/2$)、三角関数の正接関数の主値をもちいる。微分 $y' = \frac{dy}{dx} = (\tan x)' = \frac{1}{\cos^2 x} = \sec^2 x = 1 + \tan^2 x = 1 + y^2$ であるから、その逆関数の微分は $\frac{dx}{dy} = (\tan^{-1} y)' = (\arctan y)' = \frac{1}{y'} = \frac{1}{1 + y^2}$。したがって $f_Y(y) = \frac{f_X(\tan^{-1} y)}{|dy/dx|_{x = \tan^{-1} y}} = \frac{1/\pi}{1 + y^2} = \frac{1}{\pi\,(1 + y^2)}$, $(-\infty < y < \infty)$.

問 1.2.6　変換式は $y_1 = 2\,x_1$, $y_2 = x_2 - x_1$ と与えられているから、$x_1 = \frac{y_1}{2}$, $x_2 = \frac{y_1}{2} + y_2$ となる。変域は $0 < x_1 < x_2 < \infty$ には $0 < y_1$, $0 < y_2$ が対応する。ヤコビアン $\frac{\partial(y_1,\, y_2)}{\partial(x_1,\, x_2)} = \det \begin{vmatrix} 2 & 0 \\ -1 & 1 \end{vmatrix} = 2$。よって $f_{Y_1,Y_2}(y_1, y_2) = \frac{f_{X_1,X_2}(x_1, x_2)}{|2|} = 2e^{-x_1} e^{-x_2} \times \frac{1}{2} = \exp[-y_1/2 - (y_1/2 + y_2)] = e^{-y_1} e^{-y_2}$ $(0 < y_1, 0 <$

y_2)。この関係から、$f_{Y_i}(y) = e^{-y}, (y > 0), i = 1, 2$ とおけば、任意の y_1, y_2 に対して、$f_{Y_1,Y_2}(y_1, y_2) = f_{Y_1}(y_1) f_{Y_2}(y_2)$ となるから、2つの確率変数は独立である。

問 1.2.7 変換式は $y_1 = x_1/x_2,\ y_2 = x_2$ とするから、$x_1 = y_1 y_2,\ x_2 = y_2$ となる。また領域は $0 < x_1 < x_2 < 1$ に対して、$0 < y_1 < 1, 0 < y_2 < 1$。さらにヤコビアンは $\left|\frac{\partial(y_1,y_2)}{\partial(x_1,x_2)}\right|_{x_1=y_1y_2,\,x_2=y_2} = \left|\det \begin{vmatrix} 1/x_2 & -x_1/x_2^2 \\ 0 & 1 \end{vmatrix}\right|_{x_1=y_1y_2,\,x_2=y_2} = \left|\frac{1}{x_2}\right|_{x_1=y_1y_2,\,x_2=y_2} = 1/y_2$。したがって $f_{Y_1,Y_2}(y_1 y_2) = f_{X_1,X_2}(x_1,x_2)\left|\frac{1}{\partial(y_1,y_2)/\partial(x_1,x_2)}\right|_{x_1=y_1y_2,\,x_2=y_2} = 8y_1y_2y_2 \times y_2 = 8y_1y_2^3 = 2y_1 \times 4y_2^3$ 2つの密度関数 $f_{Y_1}(y) = 2y, f_{Y_2}(y) = 4y^3$ とすると積に分離されるから、独立であることがわかる。

問 1.2.8 変換式のヤコビアンを求める。非負となるのは明らかだから、絶対値は省いてみると、また逆関数の微分は求めた結果の逆数をとればよい。したがって $y_1 = \frac{x_1}{x_1+x_2},\ y_2 = \frac{x_1+x_2}{x_1+x_2+x_3},\ y_3 = x_1+x_2+x_3$ より、$x_1 = y_1y_2y_3,\ x_2 = y_2y_3 - y_1y_2y_3,\ x_3 = y_3 - y_2y_3,$ したがって $\frac{\partial(x_1,x_2,x_3)}{\partial(y_1,y_2,y_3)}$ を求めて、その逆数をかければよい。偏微分 $\partial x_i/\partial y_j,\ i,j = 1,2,3$ はそれぞれ計算してみると、$\begin{vmatrix} y_2y_3 & y_1y_3 & y_1y_2 \\ -y_2y_3 & y_3 - y_1y_3 & y_2 - y_1y_2 \\ 0 & -y_3 & 1-y_2 \end{vmatrix} = \begin{vmatrix} y_2y_3 & y_1y_3 & y_1y_2 \\ 0 & y_3 & y_2 \\ 0 & -y_3 & 1-y_2 \end{vmatrix} = \begin{vmatrix} y_2y_3 & y_1y_3 & y_1y_2 \\ 0 & y_3 & y_2 \\ 0 & 0 & 1 \end{vmatrix} = y_2y_3^2$ したがって $f_{Y_1,Y_2,Y_3}(y_1,y_2,y_3) = e^{-x_1}e^{-x_2}e^{-x_3}|_{x_1=y_1y_2y_3, x_2=y_2y_3-y_1y_2y_3, x_3=y_3-y_2y_3-y_1y_2y_3} \times y_2y_3^2 = \exp(-y_3)y_2y_3^2$ となり、$f_{Y_1}(y) = 1, (0 < y < 1),\ f_{Y_2}(y) = 2y, (0 < y < 1),\ f_{Y_3}(y) = \frac{1}{2}y^2e^{-y}, (y > 0)$ と分解できるから、3つは独立であることが示される。

第 1.3 節(1.3 ページ)==

問 1.3.1 $Y = X_1 + X_2$ とすると、仮定からモーメント母関数は $M_{X_1}(t) = (1-2t)^{r_1/2},\ M_Y(t) = (1-2t)^{r/2}$ である。また X_1, X_2 の独立性から $M_Y(t) = M_{X_1+X_2}(t) = M_{X_1}(t)\,M_{X_2}(t)$ したがって関係式として $(1-2t)^{r/2} = (1-2t)^{r_1/2}\,M_{X_2}(t)$ が得られる。よって $M_{X_2}(t) = (1-2t)^{r/2} / (1-2t)^{r_1/2} = (1-2t)^{(r-r_1)/2}$ つまり $X_2 \sim \chi^2(r - r_1)$ である。

問 1.3.2 $Y = c_1X_1 + c_2X_2$ とおくと、ガンマ分布のｍｇｆは $M_{X_i}(t) = (1-\beta_i t)^{-\alpha_i},\ i = 1,2$ だから、$M_Y(t) = E[e^{tY}] = E[e^{t(c_1X_1+c_2X_2)}] = E[e^{tc_1X_1}]\,E[e^{tc_2X_2}] = (1-\beta_1c_1t)^{-\alpha_1}(1-\beta_2c_2t)^{-\alpha_2}$ もし $\beta_1c_1 = \beta_2c_2 = \beta$ とおくならば、$M_Y(t) = E[e^{tY}] = (1-\beta t)^{-(\alpha_1+\alpha_2)}$ となる。よって $Y \sim \mathrm{Gam}(\alpha_1+\alpha_2, \beta)$ ここで、$\beta = \beta_1c_1 = \beta_2c_2$ とする。

問 1.3.3 定義により、$M = E[e^{(c_1X_1+c_2X_2+\cdots+c_nX_n)t}] = E[e^{c_1X_1t}]E[e^{c_2X_2t}] \times \cdots \times E[e^{c_nX_nt}] = M_{X_1}(c_1t)M_{X_2}(c_2t)\cdots M_{X_n}(c_nt) = \Pi_{i=1}^n M_{X_i}(c_it)$

問 1.3.4 パラメータ λ_i の ポアソン分布に対するｍｇｆは $M_{X_i}(t) = E[e^{X_it}] = \exp[\lambda_i(e^t-1)]$ であるから、その独立な確率変数和 $Y = \sum_i X_i$ に対する mgf $M_Y(t) = E[e^{Yt}] = E[\Pi_i e^{X_it}] = \Pi_i E[e^{X_it}] = \Pi_i \exp[\lambda_i(e^t - 1)] = \exp[\sum_i \lambda_i(e^t-1)]$。すなわち $Y \sim \mathrm{Po}(\sum_i \lambda_i)$。もし同一な分布、パラメータが $\lambda_i = \lambda (i = 1, 2, \cdots, n)$ ならば $Y \sim \mathrm{Po}(n\lambda)$

問 1.3.5 第 1 段階での和と第 2 段階での和は合計すれば、その和に関する値は、平均 $(n_1+n_2)p$, 分散 $(n_1+n_2)pq$ となる。仮定から $M_{X_i}(t) = E[e^{X_it}] = (pe^t+q)^{n_i},\ (i = 1,2)$。したがって $M_{X_1+X_2}(t) = E[e^{(X_1+X_2)t}] = E[e^{X_1t}]E[e^{X_2t}] = (pe^t+q)^{n_1} \times (pe^t+q)^{n_2} = (pe^t+q)^{n_1+n_2}$ つまり、$Y \sim \mathrm{Binom}(n_1+n_2, p)$

第 2.1 節(2.1 ページ) ==

問 2.1.1 $E[(X_i - \overline{X})^2] = E[(\{X_i-\mu) - (\overline{X}-\mu)\}^2] = E[(X_i-\mu)^2 - 2(X_i-\mu)(\overline{X}-\mu) + (\overline{X}-\mu)^2\}] = \mathrm{var}(X_i) + \mathrm{var}(\overline{X}) - 2E[(X_i-\mu)(\overline{X}-\mu)]$。ここで $E[(X_i-\mu)(\overline{X}-\mu)] = E[(X_i-\mu)\,\frac{1}{n}\sum_j(X_j-\mu)] = \frac{1}{n}\sum_j E[(X_i-\mu)(X_j-\mu)] = \frac{1}{n}\{\sum_{j\neq i} E[X_i-\mu]\,E[X_j-\mu] + E[(X_i-\mu)^2]\} = \frac{1}{n}E[(X_i-\mu)^2]$ であるから、

問 2.1.2　期待値（平均）については、$E\left[\dfrac{X_i+X_j}{2}\right]=\dfrac{E[X_i]+E[X_j]}{2}=\dfrac{\mu+\mu}{2}=\mu$ より、$E\left[\overline{X}\right]=\dfrac{1}{\binom{n}{2}}\sum_{i<j}E\left[\dfrac{X_i+X_j}{2}\right]=\mu\times\dfrac{1}{\binom{n}{2}}\sum_{i<j}1=\mu$。また不偏分散とよばれる由縁であるが、2 次モーメント（不偏分散）の期待値は $E\left[\dfrac{(X_i-X_j)^2}{2}\right]=\frac{1}{2}E[(X_i-\mu)^2+(X_j-\mu)^2]=\frac{1}{2}\{\mathrm{var}(X_i)+\mathrm{var}(X_j)-2E[(X_i-\mu)(X_j-\mu)]\}=\frac{1}{2}\{\sigma^2+\sigma^2+0\}=\sigma^2\ (i\neq i)$。したがってすべての $i,j\,(i<j)$ に共通であり、これらの和をとるならば、条件 $\{(i,j);i<j\}$ を満たすものがすべて一定値 σ^2 となるから、不偏性が示された。

問 2.1.3　(1) 独立であるから、同時分布は積で表され、変換のヤコビアンを計算すればよい。

(2) この関係式から、和 Z, 差の 2 乗 W^2 も独立となるから、正規分布においては標本平均と標本分散が独立になることがわかる。

問 2.1.4　標準正規分布の分布関数を $\Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\exp\{-t^2/2\}\,dt$ とおくと、$\Phi(-x)=\int_{-\infty}^{-x}\frac{1}{\sqrt{2\pi}}\exp\{-t^2/2\}\,dt=1-\Phi(x)=\int_{x}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\{-t^2/2\}\,dt,\ (x>0)$ だから、$P(\mu-c<\overline{X}<\mu+c)=P\left(-c\sqrt{\dfrac{n}{\sigma^2}}<\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}<c\sqrt{\dfrac{n}{\sigma^2}}\right)=\Phi(c\sqrt{n}/\sigma)-\Phi(-c\sqrt{n}/\sigma)=2\Phi(c\sqrt{n}/\sigma)-1\geq 0.954$。$\Phi(c\sqrt{n}/\sigma)\geq\dfrac{1.954}{2}=0.977$。正規分布表から、$c\sqrt{n}/\sigma\geq\Phi^{-1}(0.977)=2.00$。$n\geq\dfrac{4\sigma^2}{c^2}$。ここで、(a) $X_i\sim N(\mu,\sigma^2)\Rightarrow\overline{X}\sim N\left(\mu,\dfrac{\sigma^2}{n}\right)\Rightarrow\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$。(b) $\texttt{NORMS.S.INV}(\alpha)=\texttt{NORMS.INV}(\alpha,0,1)=\Phi^{-1}(\alpha),\ (0<\alpha<1)$ を用いる。

問 2.1.5　事象 $\{X>Y\}$ は事象 $\{X-Y>0\}$ に等しいことに注意する。いま $Z=\overline{X}-\overline{Y}$ とおけば、$\overline{X}\sim N(\mu_1,\sigma_1^2/n_1)$, $\overline{Y}\sim N(\mu_2,\sigma_2^2/n_2)$ の差の分布はやはり正規分布で、分散が和になることから、$Z\sim N(\mu,\sigma^2)$, $\sigma^2=\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}$,(2 つの分散の和)。したがって $P(Z>0)=P\left(\dfrac{Z-\mu}{\sigma}>\dfrac{-\mu}{\sigma}\right)=1-\Phi\left(\dfrac{-\mu}{\sigma}\right)=\Phi\left(\dfrac{\mu}{\sigma}\right)$, 不等号の向きに注意。

問 2.1.6　確率変数 $\dfrac{S^2}{\sigma^2/n}=nS^2/\sigma^2\sim\chi^2(n-1)$ をもちいる。$P(a<S^2<b)=P\left(\dfrac{n\,a}{\sigma^2}<nS^2/\sigma^2<\dfrac{n\,b}{\sigma^2}\right)=F(\dfrac{n\,b}{\sigma^2})-F(\dfrac{n\,a}{\sigma^2})=\texttt{CHISQ.DIST}(\dfrac{n\,b}{\sigma^2},n-1,1)-\texttt{CHISQ.DIST}(\dfrac{n\,a}{\sigma^2},n-1,1)$ で求められる。

問 2.1.7　モーメント母関数の定義から、$E[\exp\{tS^2\}]=E\left[\exp\left(\dfrac{n\,S^2}{\sigma^2}\dfrac{t\,\sigma^2}{n}\right)\right]=E\left[\exp\left(X\,\dfrac{t\,\sigma^2}{n}\right)\right]$。ここで $X\sim\chi^2(n-1)$ より、$M_X(t)=(1-2\,t)^{-(n-1)/2}$ である。したがって $M_{S^2}(t)=\left(1-\dfrac{2\,t\sigma^2}{n}\right)^{-(n-1)/2}=(1-\beta\,t)^{-\alpha}$ ここで $\alpha=\dfrac{n-1}{2}$, $\beta=2\sigma^2/n$ とおいた。すなわち、これをパラメータとするガンマ分布、$S^2\sim\mathrm{Gam}\left(\dfrac{n-1}{2},\dfrac{2\sigma^2}{n}\right)$ である。

第 2.2 節（2.2 ページ）===

問 2.2.1　確率変数が $X\sim\mathrm{Beta}(a,b)$ だから、密度関数は $f_X(x)=\dfrac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}x^{a-1}\,(1-x)^{b-1},\ (0<x<1)$ となるから、平均は $EX=\int_0^1 x\,f_X(x)\,dx=\dfrac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\int_0^1 x^a\,(1-x)^{b-1}dx=\dfrac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\times\dfrac{\Gamma(a+1)\,\Gamma(b)}{\Gamma(a+b+1)}=\dfrac{a}{a+b}$。また分散は $E[X^2]=\int_0^1 x^2\,f_X(x)\,dx=\dfrac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\int_0^1 x^{a+1}\,(1-x)^{b-1}dx=\dfrac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\times\dfrac{\Gamma(a+2)\,\Gamma(b)}{\Gamma(a+b+2)}=\dfrac{a(a+1)}{(a+b)(a+b+1)}$ より、$\mathrm{var(X)}=E[X^2]-(EX)^2$ より、$\mathrm{var(X)}=\dfrac{a(a+1)}{(a+b)(a+b+1)}-\left(\dfrac{a}{a+b}\right)^2=\dfrac{ab}{(a+b)^2\,(a+b+1)}$ で計算される。ここで $\Gamma(\alpha)=(\alpha-1)!$ に注意。

問 2.2.2　確率分布の t 分布も正規分布と同様に y 軸対称であることに注意して、自由度 (degree of freedom) $\texttt{df}=15$ としているから、$P(-c<T<c)=F_T(c)-F_T(-c)=F_T(c)-(1-F_T(c))=2\,F_T(c)-1=0.95$。ゆえに $F_T(c)=1.95/2=0.975$, $\texttt{df}=15$. $\therefore c=F_T^{-1}(0.975)=\texttt{T.INV}(0.975,15)=2.131$ が得られる。t 分

布表から読み取る練習も必要。

問 2.2.3 仮定から, $W \sim F(m,n)$ つまり、独立なカイ 2 乗分布の比率が F 分布にしたがうことから、$X \sim \chi^2(m), Y \sim \chi(n)$ とするとき、$W = \dfrac{Y/n}{X/m} \sim F(m,n)$、(分母分子の順に注意) であるから、逆数の分布も F 分布で、$1/W = \dfrac{X/m}{Y/n} \sim F(n,m)$ となる。

問 2.2.4 F 分布の確率について、上側 (right-hand-side)RT の確率は逆数の関係も用いる。たとえば、もし $W \sim F(m,n)$ で $P(W \le c) = 0.05$ を求める場合であれば、$1/W \sim F(n,m)$ であり、$P(1/W \ge 1/c) = 0.05$ より、$P(1/W \le 1/c) = 1 - 0.05 = 0.95$ よって $1/c = \texttt{F.INV}(0.95, n, m)$ から、$c = 1/\texttt{F.INV}(0.95, n, m)$。

問 2.2.5 自由度 n の t 分布は標準正規分布 N と自由度 n のカイ 2 乗分布から、$T(n) = \dfrac{N(0,1)}{\sqrt{\chi^2(n)/n}}$ と定められるものである。したがって 2 乗値は $T^2(n) = \dfrac{N(0,1)^2}{\chi^2(n)/n} = \dfrac{\chi^2(1)}{\chi^2(n)/n} = F(1,n)$。ここで正規分布の 2 乗は自由度 1 のカイ 2 乗分布で、２つのカイ 2 乗分布の比率は F 分布になることをもちいた。

問 2.2.6 確率変数 X から、変換した確率変数 $Y = 2X$ を求めればよい。$f_Y(y) = f_X(x)\left|\dfrac{1}{dy/dx}\right|$ をもちいるから、$x = y/2, dx/dy = 1/2, (y > 0)$ より、$f_Y(y) = e^{-y/2}\left|\dfrac{1}{2}\right| = \dfrac{1}{2}e^{-y/2}, (y > 0)$。$Y \sim Gam(1,2) = \chi^2(2)$。ここで $Z \sim \mathrm{Gam}(\alpha, \beta) : f_Z(z) = \dfrac{z^{\alpha-1}}{\Gamma(\alpha)\,\beta^\alpha}\exp\{-z/\beta\}\mathbf{1}_{z>0}(z)$ におけるパラメータが $\alpha = 1, \beta = 2$ とすればよい。

また $X \sim \mathrm{Expo}(1), Y \sim \mathrm{Expo}(1)$ より、$2X \sim \chi^2(2), 2Y \sim \chi^2(2)$ であり、独立となることは明らか。したがってその比率は $\dfrac{X}{Y} = \dfrac{(2X)/2}{(2Y)/2} \sim \dfrac{\chi^2(2)/2}{\chi^2(2)/2} = F(2,2)$ となる。

第 3 章

第 3.1 節（3.1 ページ）==

問 3.1.1 3 個のデータ $Y_i, i = 1,2,3$ について $f_X(x) = 1, (0 < x < 1), F_X(x) = x, (0 < x < 1)$ より、$f_{Y_1,Y_3}(y_1, y_3) = 6(y_3 - y_1), (0 < y_1 < y_3 < 1)$ となる。したがって変換 $Z = Y_3 - Y_1, W = Y_3$ を考えれば、ヤコビアンの $\dfrac{\partial(y_1, y_3)}{\partial(z, w)}$ の絶対値は 1 となるから、$f_{Z,W}(z, w) = f_{Y_1,Y_3}(y_1, y_3) = 6(y_3 - y_1) = 6z$, $0 < z < w < 1$. したがって $f_Z(z) = \int_z^1 6z\,dw = 6z(1 - z), 0 < z < 1.$

問 3.1.2 順序統計量の同時（結合）分布は、対称性と独立性から n 変量では、$f_Y^n(y_1, \cdots, y_n) = n!\,f(y_1)\cdots f(y_n)$, $-\infty < y_1 < \cdots < y_n < \infty$ であるから、この分布から y_n を $y_{n-1} < y_n < \infty$ の範囲で積分することで、周辺分布を求めていけばよい。すなわち $f_Y^n(y_1, \cdots, y_n)$ から $f_Y^{n-1}(y_1, \cdots, y_{n-1})$ は

$$\begin{aligned} f_Y^{n-1}(y_1, \cdots, y_{n-1}) &= \int_{y_{n-1}}^{\infty} f_Y^n(y_1, \cdots, y_n)\,dy_n = \\ &= n!\,f(y_1)\cdots f(y_{n-1})\int_{y_{n-1}}^{\infty} f(y_n)\,dy_n = n!\,f(y_1)\cdots f(y_{n-1})(1 - F(y_{n-1})) \end{aligned}$$

変数 y_{n-1} を範囲 $y_{n-2} < y_{n-1} < \infty$ で積分すると

$$\begin{aligned} f_Y^{n-2}(y_1, \cdots, y_{n-2}) &= \int_{y_{n-2}}^{\infty} f_Y^{n-1}(y_1, \cdots, y_{n-1})\,dy_{n-1} \\ &= n!\,f(y_1)\cdots f(y_{n-2})\int_{y_{n-2}}^{\infty} f(y_{n-1})(1 - F(y_{n-1}))\,dy_{n-1} \\ &= n!\,f(y_1)\cdots f(y_{n-2})\frac{(1 - F(y_{n-2}))^2}{2} \end{aligned}$$

と順次に上から下げていく。

$$\begin{aligned} f_Y^{n-3}(y_1, \cdots, y_{n-3}) &= \int_{y_{n-3}}^{\infty} f_Y^{n-2}(y_1, \cdots, y_{n-2})\,dy_{n-2} \\ &= n!\,f(y_1)\cdots f(y_{n-3})\int_{y_{n-3}}^{\infty} f(y_{n-2})\frac{(1 - F(y_{n-2}))^2}{2}\,dy_{n-2} \\ &= n!\,f(y_1)\cdots f(y_{n-3})\frac{(1 - F(y_{n-3}))^3}{2\cdot 3} \end{aligned}$$

あるいは中間の変数 y_i を範囲 $y_{i-1} < y_i < y_{i+1}$ で積分すると

$$\begin{aligned} f_Y^{(-i)}(y_1,\cdots,y_{i-1},y_{i+1},\cdots,y_n) &= f_Y^{(-i)}(y_1,\cdots,\overline{(y_i)},\cdots,y_n) = \int_{y_{i-1}}^{y_{i+1}} n!\, f(y_1)\cdots f(y_n)\,dy_i \\ &= n!\, f(y_1)\cdots f(y_{i-2})\{F(y_{i+1})-F(y_{i-1})\}f(y_{i+2})\cdots f(y_n) \end{aligned}$$

などと計算していけばよいから、n 変量から、次数の少ないものへと帰着できる。

この方法では離散型の場合には成り立たない。それはタイのデータ、同順位の場合が起り得るから、このような密度関数からの積分ができないからである。

問 3.1.3　Y_k の密度関数は $f_{Y_k}(x) = \dfrac{n!}{(k-1)!(n-k)!}x^{k-1}(1-x)^{n-k}, (0 < x < 1)$ であり、$n! = \Gamma(n+1) = \Gamma(k+n-k+1), (k-1)! = \Gamma(k), (n-k)! = \Gamma(n-k+1)$ の関係に注意すると $\alpha = k, \beta = n-k+1$ となる。

問 3.1.4　事象 $\{Y_k > p\}$ が起こることは、多くとも $k-1$ 個の観測データ Y_i が区間 $[0,p]$ に含まれること。一つの観測データがこの範囲に含まれる確率は p で、起らない確率は $1-p$ であるから、$P(Y_k > p) = \displaystyle\sum_{i=0}^{k-1} p^i(1-p)^{n-i}$ となる。

問 3.1.5　一様分布であるから、$F(x) = x, f(x) = 1 (0 < x < 1)$, より、$f_{Y_k}(x) = n\begin{pmatrix} n-1 \\ k-1, n-k\end{pmatrix}x^{k-1}(1-x)^{n-k}$, $f_{Y_j,Y_k}(x,y) = n(n-1)\begin{pmatrix} n-2 \\ j-1, k-j-1, n-k\end{pmatrix}x^{j-1}(y-x)^{k-j-1}(1-y)^{n-k}, (j<k)$ であるから、平均、分散、共分散の定義にしたがって計算する。(i)　$E[Y_k] = \displaystyle\int_0^1 x f_{Y_k}(x)\,dx = n\begin{pmatrix} n-1 \\ k-1, n-k\end{pmatrix}\int_0^1 x^k(1-x)^{n-k}\,dx = n\begin{pmatrix} n-1 \\ k-1, n-k\end{pmatrix}B(k+1, n-k+1) = n\frac{(n-1)!}{(k-1)!(n-k)!}\frac{k!(n-k)!}{(k+1+n-k+1-1)!} = \frac{k}{n+1}$,
(ii)　$E[Y_k^2] = \displaystyle\int_0^1 x^2 f_{Y_k}(x)\,dx = n\begin{pmatrix} n-1 \\ k-1, n-k\end{pmatrix}\int_0^1 x^{k+1}(1-x)^{n-k}\,dx = n\begin{pmatrix} n-1 \\ k-1, n-k\end{pmatrix}B(k+2, n-k+1) = n\frac{(n-1)!}{(k-1)!(n-k)!}\frac{(k+1)!(n-k)!}{(k+2+n-k+1-1)!} = \frac{k(k+1)}{(n+1)(n+2)}$, したがって分散は $\mathrm{var}(Y_k) = \dfrac{k(k+1)}{(n+1)(n+2)} - \left(\dfrac{k}{n+1}\right)^2 = \dfrac{k\,(n-k+1)}{(n+1)^2(n+2)}$

$$\begin{aligned} \text{(iii)}\quad E[Y_j\,Y_k] &= \iint_{0<x<y<1} xy\, f_{Y_j,Y_k}(x,y)\,dxdy \\ &= \iint_{0<x<y<1} xy\ n(n-1)\begin{pmatrix} n-2 \\ j-1, k-j-1, n-k\end{pmatrix}x^{j-1}(y-x)^{k-j-1}(1-y)^{n-k}\,dxdy \\ &= n(n-1)\begin{pmatrix} n-2 \\ j-1, k-j-1, n-k\end{pmatrix}\int_0^y x^j(y-x)^{k-j-1}\,dx\int_0^1 y(1-y)^{n-k}\,dy \\ &= n(n-1)\begin{pmatrix} n-2 \\ j-1, k-j-1, n-k\end{pmatrix}\int_0^1 t^j(1-t)^{k-j-1}\,dt\int_0^1 y^k y(1-y)^{n-k}\,dy \\ &= n(n-1)\begin{pmatrix} n-2 \\ j-1, k-j-1, n-k\end{pmatrix}B(j+1, k-j)\,B(k+2, n-k+1) = \frac{j\,(k+1)}{(n+1)(n+2)} \end{aligned}$$

したがって $\mathrm{cov}(Y_j, Y_k) = \dfrac{j\,(k+1)}{(n+1)(n+2)} - \dfrac{j}{n+1}\dfrac{k}{n+1} = \dfrac{j\,(n-k+1)}{(n+1)^2(n+2)}$

第 3.2 節（3.2 ページ）===

問 3.2.1　確率収束することは、任意の $a > 0$ に対して、$P(|X_n| > a) \to 0\ (a \to \infty)$ であるから。しかし分布収束については、極限の分布関数 $\lim_n F_n(x) = F(x)$ が、$F(x) = 0, x < 0 \quad = 1, x \geq 0$ であるから、$x = 0$ では、$F(0) = 1$ であり、$\lim_n F_n(0) = 0$ となって等号が成り立たない。このような点があるから、したがって分布収束していない。

確率収束から分布収束を導く。まず $F_n(x) = P(X_n \leq x) = P(X_n \leq x, X > x+\epsilon) + P(X_n \leq x, X \leq x+\epsilon) \leq P(\,|X_n - X|\ > \epsilon) + P(X \leq x+\epsilon) = P(\,|X_n - X|\ > \epsilon) + F(x+\epsilon)$ より、$F_n(x) - F(x+\epsilon) \leq P(\,|X_n - X| > \epsilon)$。同様に $F(x-\epsilon) = P(X \leq x-\epsilon) = P(X \leq x-\epsilon, X_n \leq x) + P(X \leq x-\epsilon, X_n \leq x) \leq P(\,|X_n - X| > \epsilon) + P(X_n \leq x) = P(\,|X_n - X|\ > \epsilon) + F_n(x)$ より、$F(x-\epsilon) - F_n(x) \leq P(\,|X_n - X|\ > \epsilon)$。

$$\max\{F_n(x) - F(x+\epsilon),\ F(x-\epsilon) - F_n(x)\} \leq P(\,|X_n - X|\ > \epsilon)。$$

仮定により、$\lim_n P(\,|X_n - X| > \epsilon) = 0$ だから、$\max\{\lim_n F_n(x) - F(x+\epsilon), F(x-\epsilon) - \lim_n F_n(x)\} = 0.$ つまり、$F(x-\epsilon) \le \lim_n F_n(x) \le F(x+\epsilon), \forall\epsilon > 0$。したがって $F(x) = F(x-)$ となる（X の連続点）x では、$\lim_n F_n(x) = F(x)$ を得る。

問 3.2.2 カイ 2 乗分布の平均分散をもちいて、$nS_n^2/\sigma^2 \sim \chi^2(n-1)$ であるから、$E[nS_n^2/\sigma^2] = n-1$, $\mathrm{var}(nS_n^2/\sigma^2) = 2(n-1)$, $E[S_n^2] = (n-1)\sigma^2/n \to \sigma^2, \mathrm{var}(S_n^2) = 2(n-1)\sigma^4/n^2) \to 0$ したがって $S_n^2 \xrightarrow{P} \sigma^2$

問 3.2.3 $W_n = \dfrac{S_n - E(S_n)}{n}$ とおけば、$E(W_n) = 0$ であり、$\mathrm{Var}(W_n) = \dfrac{\mathrm{Var}(S_n)}{n^2} = \dfrac{1}{n^2}\sum_i \sigma_i^2 \le \dfrac{1}{n^2}nM = \dfrac{M}{n} \to 0$ となる。したがって、$W_n \xrightarrow{P} 0$。

問 3.2.4 もし独立同一分布ならば、$\dfrac{S_n - E(S_n)}{n} \xrightarrow{P} 0$ は $\dfrac{S_n - n\mu}{n} = \dfrac{S_n}{n} - \mu \xrightarrow{P} 0$ より $\dfrac{S_n}{n} \xrightarrow{P} \mu$ を得る。

問 3.2.5 $|X_n - Y_n| > \epsilon$ を 2 つの場合、$X_n - Y_n < -\epsilon$, $X_n - Y_n > \epsilon$ に分けて考える。$\{X_n \le x\} \subset \{Y_n \le x+\epsilon\} \cup \{X_n - Y_n \le -\epsilon\}$ から $F_n(x) = P(X_n \le x) \le P(Y_n \le x+\epsilon) + P(X_n - Y_n \le -\epsilon) = G_n(x+\epsilon) + P(X_n - Y_n \le -\epsilon)$ となる。同様に、$\{X_n > x\} \subset \{Y_n > x-\epsilon\} \cup \{X_n - Y_n > \epsilon\}$ から $1 - F_n(x) = P(X_n > x) \le 1 - G_n(x-\epsilon) + P(X_n - Y_n > \epsilon)$, $F_n(x) \ge G_n(x-\epsilon) - P(X_n - Y_n > \epsilon)$ 。この 2 つから、$G_n(x-\epsilon) - P(X_n - Y_n > \epsilon) \le F_n(x) \le G_n(x+\epsilon) + P(X_n - Y_n \le -\epsilon)$, 極限をとると確率収束しているから、$\lim_n G_n(x-\epsilon) = G(x-\epsilon) \le \lim_n F_n(x) \le \lim_n G_n(x+\epsilon) = G(x+\epsilon)$ $(\forall\epsilon > 0)$ ゆえに $\lim_n F_n(x) = G(x)$。

問 3.2.6 大数の弱法則から $\overline{X}_n \to \mu$ in prob ですなわち分布収束する。この定数 μ の分布関数は $F(x) = 0$ if $x < \mu, = 1$ if $\mu \le x$.

問 3.2.7 $X_n \sim F_n$ とすると、十分大きな n では、$F_n(x) = 0$ となり、恒常的にゼロになる関数は分布関数になれないから、極限分布は存在しない。

第 3.3 節（3.3 ページ）==

問 3.3.1 モーメント母関数の極限と確率収束の関係をもちいる。仮定から $M_{X_n}(t) = 1/(1-\beta t)^n$ で関数 $1/(1-\beta t)$ は指数分布 $\mathrm{Expo}(\beta) = \mathrm{Gam}(1,\beta)$. 平均は β であるから、大数の弱法則から $X_n/n \xrightarrow{P} \beta$, $X_n/n \xrightarrow{D} \mathrm{Expo}(\beta)$.

問 3.3.2 $\chi^2(n) = \sum_{i=1}^n X_i^2$, ここで $X_i \sim \mathrm{N}(0,1)$。自由度 n のカイ 2 乗分布の平均は n, 分散は $2n$ であるから、中心極限定理をもちいて正規近似される。

問 3.3.3 個数を求めると $Y_n = \sum_i \mathbf{1}_{\{\mathbf{a}<\mathbf{X_i}<\mathbf{b}\}}$ である。ここで $E[\mathbf{1}_{\{\mathbf{a}<\mathbf{X_i}<\mathbf{b}\}}] = \int_{\mathbf{a}}^{\mathbf{b}} \mathbf{f(x)\,dx}$ で、確率変数 $\mathbf{1}_{\{\mathbf{a}<\mathbf{X_i}<\mathbf{b}\}}$ は成功の確率パラメータが $p = \int_a^b f(x)\,dx$ のベルヌーイ分布に等しい。よって $Y_n \sim \mathrm{Binom}(n,p)$ で平均、分散が $E(Y_n) = np, \mathrm{var}(Y_n) = np(1-p)$ という正規分布で近似される。

問 3.3.4 $E(X_i) = 0, \mathrm{var}(X_n) = E(X_i^2) = \int_{-1/2}^{1/2} x^2\,dx = 2\int_0^{1/2} x^2\,dx = 1/12$ 中心極限定理より、Y_n は $\mathrm{N}(0, \dfrac{n}{12})$ で近似できる。

問 3.3.5 $W_n = n(1 - F(Y_n)$ とおくと、$P(W_n \ge w) = P(n(1 - F(Y_n) \ge w) = P\left(F(Y_n) \le 1 - \dfrac{w}{n}\right) = P\left(\max_i F(X_i) \le 1 - \dfrac{w}{n}\right) = \left(1 - \dfrac{w}{n}\right)^n$。ただし $0 < w < n$ とする。これから $\left(1 - \dfrac{w}{n}\right)^n \to e^{-w}$ でパラメータ 1 の指数分布。よって $W_n \to \mathrm{Expo}(1)$ in prob。

第 4 章第 4.1 節（4.1 ページ）===

問 4.1.1 (a) $f_\theta(x_i) = e^{-\theta}\dfrac{\theta^{x_i}}{x_i!}, i = 1,2,\cdots,n$ であるから、$f_\theta(x_1,x_2,\cdots,x_n) = \theta^{x_1+x_2+\cdots+x_n}\dfrac{e^{-n\theta}}{x_1!\cdots x_n!}$ は, $x = x_1 + x_2 + \cdots + x_n$ として $L(\theta) = \ln f_\theta(x_1,x_2,\cdots,x_n) = x\ln\theta - n\theta + \ln(x_1!\cdots x_n!)$ を θ で偏微分し、ゼロとおくと、$\dfrac{\partial}{\partial\theta}L(\theta) = \dfrac{x}{\theta} - n = 0$ だから、$\hat{\theta} = \dfrac{x}{n} = \dfrac{x_1 + x_2 + \cdots + x_n}{n} = \overline{X}$、すなわち、標本平均。

(b) $L(\theta) = \prod_i \theta x_i^{\theta-1} = \theta^n\left(\prod_i x_i\right)^{\theta-1}$, 対数をとって $\ln L(\theta) = n\theta + (\theta-1)\sum_i \ln x_i$ となるから、$\dfrac{\partial}{\partial\theta}\ln L(\theta) = \dfrac{n}{\theta} + \sum_i \ln x_i$ よって $\hat{\theta} = -\dfrac{n}{\sum_i \ln x_i}$ を得る。ここで $0 < x_i < 1$ より、$\ln x_i < 0$ だから、$\hat{\theta} > 0$ で

あることに注意。

(c) 指数分布だから、$L(\theta) = (1/\theta)^n \exp\{-\sum_i x_i/\theta\}$ より、$\frac{\partial}{\partial\theta}(-n\ln\theta - \frac{x}{\theta}) = -\frac{n}{\theta} + \frac{x^2}{\theta} = 0$ (ここで $x = \sum_i x_i$) として、$\hat{\theta} = \frac{x}{n} = \overline{X}$.

(d) 絶対値が含まれている両側指数分布であることに注意する。また $\min_\theta \sum_i |x_i - \theta|$ は $\{x_i\}$ の順序統計量のうち、中央値 (median) で最小となるから、 $\theta = M_d = \mathrm{med}\{x_i\}$、すなわち中央値が MLE となる。

(e) 密度関数が正 $f_\theta(x) > 0$ となる x の領域は $\{x : x > \theta\}$ であるから、$L(\theta) = \exp[-\sum_i x_i]e^{n\theta} \prod_i \mathbf{1}_{\{\theta:\theta<x_i\}}(\theta)$, したがって $\prod_i \mathbf{1}_{\{\theta:\theta<x_i\}}(\theta) \leq \mathbf{1}_{\{\theta:\theta<\min_i x_i\}}(\theta)$ となるから、不等号が成立するような θ の最大値としては、観測値 $\{X_i\}$ に対する最小値、$\hat{\theta} = \min_i X_i$ となる。

問 4.1.2　$f_\theta(x_1, x_2, \cdots, x_n) = 1$, if $\theta - 1/2 \leq x_i \leq \theta + 1/2, \forall i, \theta$ である。このように範囲内では最大値が一定であるから、$\theta - 1/2 \leq X_i, \forall i$ より、$\theta - 1/2 \leq \min_i X_i$, よって $\theta \leq \min_i X_i + 1/2$。同様に順序に注意して、$x_i \leq \theta + 1/2, \forall i$ より、$\max_i X_i - 1/2 \leq \theta$ となる。もし、ある関数 $h(x_1, x_2, \cdots, x_n)$ が $\max_i X_i - \frac{1}{2} \leq h(X_1, X_2, \cdots, X_n) \leq \min_i X_i + \frac{1}{2}$ であるならば、これは MLE となる。したがってこれを満たすような関数 h はたくさん作れる。例えば、(i) $h = \frac{\max_i X_i + \min_i X_i}{2}$ (ii) $h = \frac{\max_i X_i + 2\min_i X_i}{3}$ (iii) $h = \frac{4\max_i X_i + 2\min_i X_i - 1}{6}$ (vi) $h = \frac{2\max_i X_i + 4\min_i X_i + 1}{6}$ など。なぜならば、これら全て $\max_i X_i - \min_i X_i \leq 1$ を満たすから。

問 4.1.3　期待値と平均を等置させる方程式から、パラメータの推定値を定め、また一致性の証明には、大数の弱法則や確率収束の関連定理をもちいる。(a) $EX_i = \theta$ より、モーメント法では期待値と平均を等しくするよう定めるから、$\theta^* = \overline{X}$。一致性の証明は大数の弱法則から、明らか。(b) $EX_i = \int_0^1 x\theta x^{\theta-1}\,dx = \int_0^1 x\theta x^\theta\,dx = \frac{\theta}{\theta+1}$ となる。したがって $\frac{\theta}{\theta+1} = \overline{X}$ より、$\theta^* = \frac{\overline{X}}{\overline{X}+1}$。確率収束することとこの関数の連続性から、$\overline{X} \to \frac{\theta}{\theta+1}$ から、$\frac{\overline{X}}{1-\overline{X}} \xrightarrow{P} \frac{\theta/(\theta+1)}{1-[\theta/(\theta+1)]} = \theta$ となり、ゆえに推定量 θ^* は θ への一致性をもつ。(c) $EX_i = \theta$ より、$\theta^* = \overline{X}$。一致性は明らか。(d) 対称性から明らかに $EX_i = \theta$。よって $\theta^* = \overline{X}$。この一致性も明らか。(e) 期待値の計算は、$y = x - \theta$ とおけばよいから、$EX_i = \int_\theta^\infty xe^{-(x-\theta)}\,dx = \int_0^\infty (y+\theta)e^{-y}\,dy = 1 + \theta$。よって、$\theta^* = \overline{X} - 1$。一致性は $\theta^* = \overline{X} - 1 \to (1+\theta) - 1 = \theta$ であるから。

問 4.1.4　確率 $P(X \leq r) = \int_0^r \frac{1}{\theta}e^{-x/\theta}\,dx = 1 - e^{-r/\theta}$ となるから、θ の最尤推定量は $\theta^* = \overline{X}$ であった (4.1.1(c)) から、この確率 $P(X \leq r) = 1 - e^{-r/\theta}$ の最尤推定量は $1 - e^{-r/\overline{X}}$

問 4.1.5　2 項分布のパラメータ推定から、θ の推定量は X/n。つまり成功の回数における相対比率であった。求める確率 $P(a \leq X \leq b) = \sum_{k=a}^b \binom{n}{k}\theta^k(1-\theta)^{n-k}$ であるから、この式の θ に X/n を代入すればよい。

第 4.2 節（4.2 ページ） ==

問 4.2.1　$\frac{\overline{X}-\mu}{\sigma\sqrt{n}} \sim \mathrm{N}(0,1)$ から、$P\left(-b < \frac{\overline{X}-\mu}{\sigma\sqrt{n}} < b\right) = \Phi(b) - \Phi(-b) = \Phi(b) - (1 - \Phi(b)) = 2\Phi(b) - 1$ が信頼係数。信頼区間は $\overline{X} - b\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + b\frac{\sigma}{\sqrt{n}}$ で、その長さは $\overline{X} + b\frac{\sigma}{\sqrt{n}} - (\overline{X} - b\frac{\sigma}{\sqrt{n}}) = 2b\frac{\sigma}{\sqrt{n}}$ となる。

問 4.2.2　分散の値が未知であるから、正規分布の分布関数値 $\Phi(b)$ の代わりに、t 分布のパーセント値 $\mathrm{F}_T(b)$ を用いる。つまり、$2\Phi(b) - 1$ の代わりに $2\mathrm{F}_T(b) - 1$ とする。$P\left(-b < \frac{\overline{X}-\mu}{S\sqrt{n-1}} < b\right) = \mathrm{F}_T(b) - \mathrm{F}_T(-b) = \mathrm{F}_T(b) - (1 - \mathrm{F}_T(b)) = 2\mathrm{F}_T(b) - 1$ を信頼係数とする。ただし b は自由度 $n-1$ の t 分布表をもちいる。$\frac{\overline{X}-\mu}{S\sqrt{n-1}} \sim \mathrm{T}(n-1)$ で、信頼区間は $\overline{X} - b\frac{S}{\sqrt{n-1}} < \mu < \overline{X} + b\frac{S}{\sqrt{n-1}}$ であるから、その長さは $\overline{X} + b\frac{S}{\sqrt{n-1}} - (\overline{X} - b\frac{S}{\sqrt{n-1}}) = 2b\frac{S}{\sqrt{n-1}}$ となる。

問 4.2.3 未知の標準偏差 σ を用いる。$S = \dfrac{\sigma}{\sqrt{n}}\sqrt{\dfrac{nS^2}{\sigma^2}}$ と変形すると、$\dfrac{nS^2}{\sigma^2} \sim \chi^2(n-1)$ であるから、信頼区間の長さ $2bS/\sqrt{n-1}$ に対する期待値は $E\left[\dfrac{2b}{\sqrt{n-1}}S\right] = E\left[\dfrac{2b}{\sqrt{n-1}}\dfrac{\sigma}{\sqrt{n}}\sqrt{\chi^2(n-1)}\right] = \dfrac{2b}{\sqrt{n-1}}\dfrac{\sigma}{\sqrt{n}}\int_0^\infty \sqrt{x}\, f_{\chi^2}(x)\,dx$ ここで $f_{\chi^2}(x)$ は $\chi^2(n-1)$ の密度関数。さらに変数変換をおこない、計算する。

問 4.2.4 ガンマ分布に対しては、$E(X_i) = \alpha\beta$, $\mathrm{var}(X_i) = \alpha\beta^2$ であるから、中心極限定理により、$\dfrac{\overline{X}-\alpha\beta}{\sqrt{\alpha}\beta/\sqrt{n}} = \dfrac{\overline{X}-\mu}{\mu/\sqrt{\alpha n}} \sim \mathrm{N}(0,1)$ となる。したがって $c = 1/\sqrt{\alpha n}$ とおけば、$P\left(-b < \dfrac{\overline{X}-\mu}{c\mu} < b\right) = \Phi(b)-\Phi(-b) = 2\Phi(b)-1$。この確率が信頼係数の値 a になるから、$b = \Phi^{-1}\left(\dfrac{a+1}{2}\right)$。信頼区間は $-b < \dfrac{\overline{X}-\mu}{c\mu} < b$ を μ について解けばよいから、$(1-bc)\mu < \overline{X} < (1+bc)\mu$ としてから、、$\dfrac{\overline{X}}{1+bc} < \mu < \dfrac{\overline{X}}{1-bc}$ を得る。

問 4.2.5 信頼区間の長さを L, 信頼係数を a と与える。したがって 2 項母集団では $X = \sum_i X_i$, $X \sim \mathrm{Binom}(n,p)$ として $\left|\dfrac{X}{n} - p\right| < \dfrac{L}{2}$, $a = 2\Phi\left(\dfrac{L/2}{\sqrt{p(1-p)/n}}\right) - 1$。$\dfrac{L\sqrt{n}}{2\sqrt{p(1-p)}} > \Phi^{-1}\left(\dfrac{a+1}{2}\right)$。$p(1-p) \geq 1/4$ より、$\dfrac{L\sqrt{n}}{2\sqrt{1/4}} > \Phi^{-1}\left(\dfrac{a+1}{2}\right) = c$, よって $L\sqrt{n} > c$ の形になる。. $n > (c/L)^2$ という評価が得られる。

第 4.3 節（4.3 ページ）==

問 4.3.1 未知の分散 σ_1^2, σ_2^2 を両方消去できる統計量が求められていないから。正規分布：$\dfrac{Z}{\sqrt{\sigma_1^2/n+\sigma_2^2/m}} = \dfrac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n+\sigma_2^2/m}} \sim N(0,1)$, カイ 2 乗分布：$W = \dfrac{S_1^2}{\sigma_1^2/n} + \dfrac{S_2^2}{\sigma_2^2/m} \sim \chi^2(n+m-2)$, T 分布：$\dfrac{Z}{\sqrt{W/(n+m-2)}} \sim \mathrm{T}(n+m-2)$ である。ここで、$S_1^2 = \dfrac{1}{n}\sum_i(X_i-\overline{X})$, $S_2^2 = \dfrac{1}{m}\sum_i(Y_i-\overline{Y})$. このように、２つの未知数が入っているから、$\sigma_1^2,\sigma_2^2$ について解くことができない。

問 4.3.2 もし $\sigma_1^2 = c\sigma_2^2$ であれば、未知数が減り、1 個の σ_2^2 だけとできる。したがって $\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m} = c\sigma_2^2\left(\dfrac{1}{n}+\dfrac{1}{cm}\right)$, $\dfrac{n\,S_1^2}{\sigma_1^2} + \dfrac{m\,S_2^2}{\sigma_2^2} = \dfrac{n\,S_1^2+cm\,S_2^2}{c\,\sigma_2^2}$。このような式変形が可能となるから、比率 c がわかっているならば、σ_2^2 について解けば、分布のパーセント値から、これを含む区間、すなわち信頼区間が求められる。

第 4.4 節（4.4 ページ）==

問 4.4.1 与えられた検定は LRT（尤度比検定）によって、棄却域が定まり、この領域が最良であることは、ネイマン=ピアソンの補題と同様な条件つき最小化問題（第 1 種の過誤を一定にし、第 2 種の過誤を最小にする）で求められる。この領域が c の値だけで定まり、$\theta > \theta_0$ に依らないことから、一様最強力となる。

問 4.4.2 仮定によって、$f_0(x) = 1/6$, $x = 1,2,\cdots,6$ で、$f_1(x) = 1/4$, $x = 1,2$,　$f_1(x) = 1/8, 3,\cdots,6$ としている。よって尤度比 $L(x) = f_1(x)/f_0(x)$ は, $L(x) = \dfrac{1/4}{1/6} = 3/2$, $x = 1,2$ また $L(x) = \dfrac{1/8}{1/6} = 3/4$, $x = 3,4,5,6$ となる。尤度比の値 λ は $3/4, 3/2$ をとるから、$0 < \lambda < 3/4$ では H_0 を棄却、このときは $\alpha = 1, \beta = 0$。$3/4 < \lambda < 3/2$ では, $x = 1,2$ ならば H_0 を棄却で、$x = 3,4,5,6$ ならば、H_0 を採択とする。このときは $\alpha = 1/3, \beta = 1/2$。$0 < \lambda < 3/4$ ではすべて H_0 を採択。このときは $\alpha = 0, \beta = 1$。、

有意水準を $\alpha = 0.1$ とするから、$\lambda = 3/2$ として、採択は $x = 3,4,5,6$, 棄却は $x = 1,2$ で確率 a とおく。$\alpha = 1/3a = 0.1$ より、$a = 0.3$ となるから、$\beta = 1/2 + 1/2(1-a) = 1/2 + 1/2 \times 0.7 = 0.85$。

問 4.4.3 帰無仮説から $np = 200$, $np(1-p) = 100$ である。$X = 220$ で第 1 種の過誤の確率は $\alpha = 1 - \Phi\left(\dfrac{X-np}{\sqrt{np(1-p)}}\right) = 1 - \Phi(2) = 1 - 0.977 = 0.023$ であるから、この値が有意水準, p 値である。

第 4.5 節（4.5 ページ）==

問 4.5.1 $\sum_i \frac{(X_i - np_i)^2}{np_i}$ を計算するから、観測値 X_i を代入して、$\frac{(40-50)^2}{50} + \frac{(33-30)^2}{30} + \frac{(27-20)^2}{20} = 2.00 + 0.30 + 2.45 = 4.75$。カイ 2 乗分布表から、自由度 2 、有意水準 0.05 では $P(\chi^2(2) > 5.99) = 0.05$ より、棄却域は 5.99 以上だから、含まれない。したがって帰無仮説は棄却されない。採択と考えられる。

問 4.5.2 各マスの期待度数を周辺度数から求めると、

	A	B	C	計
標本 1	33	147	114	294
標本 2	67	153	86	306
計	100	300	200	600

$\Longrightarrow$

	A	B	C
標本 1	49	147	98
標本 2	51	153	86

この内訳計算は $(A,1) = 294 \times 100/600 = 49$, $(B,1) = 294 \times 300/600 = 147$, $(C,1) = 294 \times 200/600 = 98$, $(A,2) = 306 \times 100/600 = 51$, $(B,2) = 306 \times 300/600 = 153$, $(C,2) = 306 \times 200/600 = 102$ である。カイ 2 乗統計量は $\frac{(33-49)^2}{49} + \frac{(147-147)^2}{147} + \frac{(114-98)^2}{98} + \cdots + \frac{(86-102)^2}{102} = 5.224 + 0.0 + 2.612 + 5.02 + 0.0 + 2.510 = 15.366$ 自由度は $(2-1)(3-1) = 1 \times 2 = 2$, $P(\chi^2(2) > 5.99) = 0.05$, $15.366 > 5.94$ であるから、帰無仮説は棄却される。

問 4.5.3 分割表の期待度数は

	属性 B_1	属性 B_2
属性 A_1	e_{11}	e_{12}
属性 A_2	e_{21}	e_{22}

とおけば、行周辺度数が $a+b, c+d$, 列周辺度数が $a+c, b+d$ であるから、各セルの期待度数は $e_{11} = \frac{(a+b)(a+c)}{a+b+c+d}$, $e_{12} = \frac{(a+b)(b+d)}{a+b+c+d}$, $e_{11} = \frac{(a+c)(c+d)}{a+b+c+d}$, $e_{11} = \frac{(c+d)(b+d)}{a+b+c+d}$ である。。これから、$a - e_{11} = \frac{ad-bc}{a+b+c+d}$, $b - e_{12} = \frac{bc-ad}{a+b+c+d}$, $c - e_{21} = \frac{bc-ad}{a+b+c+d}$, $d - e_{22} = \frac{ad-bc}{a+b+c+d}$ となる。対称行列である。これをカイ 2 乗統計量に当てはめると、答えが得られる。ここで $(c+d)(b+d) + (a+c)(c+d) + (a+b)(b+d) + (a+b)(a+c) = (a+b+c+d)^2$ となることに注意する。

第 4.6 節（4.6 ページ）==

問 4.6.1 同時確率密度 $x = (x_1, x_2, \cdots, x_n)$ は

$$f_\theta(x) = \prod_{i=1}^{n} \frac{e^{-\theta^{x_i}}\theta\, x_i}{x_i!} = \frac{e^{-n\,\theta}\theta^{u(x)}}{x_1!x_2!\cdots x_n!}$$

となるから、$g(\theta, u(x)) = e^{-n\,\theta}\theta^{u(x)}$, また $h(x) = 1/(x_1!x_2!\cdots x_n!)$ とする。

問 4.6.2 $x = (x_1, x_2, \cdots, x_n)$ に対して、$f_\theta(x) = \begin{cases} \{A(\theta)\}^n B(x_1)B(x_2)\cdots B(x_n) & \text{all}\, i, 0 < x_i < \theta \\ 0 & otherwise \end{cases}$ 。この式を示性関数 $I_A(x) = 1, x \in A, = 0, x \notin A$ を用いれば、関数の積は積集合で $I_A(x)I_B(x) = I_{A\cap B}(x)$ となる。書き直しすれば、$f_\theta(x) = \{A(\theta)\}^n \prod_i B(x_i) I_{\max_i x_i < \theta}$ となるから、$g(\theta, u(x)) = \{A(\theta)\}^n I_{\max_i x_i < \theta}$, $h(x) = \prod_i B(x_i)$ とおけばよい。

問 4.6.3 $f_\theta(x) = \prod_i \theta(1-\theta)^{x_i} = \theta^n(1-\theta)^{u(x)}$ だから、$g(\theta, u(x)) = \theta^n(1-\theta)^{u(x)}$, $h(x) = 1$

問 4.6.4 $f_\theta(x) = \prod_i \frac{1}{\theta} e^{-\theta x_i} = \theta^{-n} \exp[-(\sum_i x_i)/\theta]$ だから、$g(\theta, u(x)) = \theta^{-n} \exp[(-u(x)/\theta]$, $h(x) = 1$

問 4.6.5 $\Gamma(\theta) = \theta - 1, \Gamma(2) = 1$ より、$\mathrm{B}(\theta, 2) = \Gamma(\theta+2)/[\Gamma(\theta)\Gamma(2)] = (\theta+1)\theta$ であるから、$f_\theta(x) = \prod_i \frac{1}{B(\theta,2)} x_i{}^{\theta-1}(1-x_i) = \prod_i (\theta+1)\theta x_i^{\theta-1}(1-x_i) = \{(\theta+1)\theta\}^n \left(\prod_i x_i\right)^{\theta-1} \prod_i (1-x_i)$ だから、$g(\theta, u(x)) = \{(\theta+1)\theta\}^n u(x)^{\theta-1}$, $h(x) = \prod_i (1-x_i)$.

問 4.6.6　$f_\theta(x) = \prod_i \frac{1}{\Gamma(\theta)} x_i^{\theta-1} e^{-x_i} = \frac{1}{\Gamma(\theta)^n} \prod_i x_i^{\theta-1} e^{-\sum_i x_i} = \frac{1}{\Gamma(\theta)^n} u(x)^{\theta-1} \exp[-\sum x_i]$. $g(\theta, u(x)) = \Gamma(\theta)^{-n} u(x)^{\theta-1}$, $h(x) = \exp[-\sum x_i]$.

問 4.6.7　$P_\theta(X_1' = x_1, \cdots, X_n' = x_n) = \prod_i \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum_i x_i}(1-\theta)^{n-\sum_i x_i} = \theta^y(1-\theta)^{n-y} = u(\theta, y)$, $X_1' + \cdots + X_n' = X_1 + \cdots + X_n = Y$ であれば、$P_\theta(X_1' = x_1, \cdots, X_n' = x_n) = P_\theta(X_1 = x_1, \cdots, X_n = x_n)$ が成り立つ。

問 4.6.8　(1) 連続型確率変数の場合を計算するが、離散型でも同様。同時分布と周辺分布の関係から $f_Y(y) = \int f_{X,Y}(x,y)dx$ と $f_{Y|X}(y|x) = f_{X,Y}(x,y)/f_X(x)$ が成り立っている。したがって $E[Y] = \int y\, f_Y(y)\, dy = \int y \int f_{X,Y}(x,y)\, dxdy = \int \left(\int y f_{Y|X}(y|x)\, dy \right) f_X(x) dx = \int (E[Y \mid X = x])\, f_X(x) dx = E[\, E[Y \mid X]\,]$. (2) $\{Y - E[Y \mid X]\}\{E[Y \mid X] - E[Y]\} = YE[Y] - E[Y \mid X]E[Y] + YE[Y \mid X] - \{E[Y \mid X]\}^2$ と展開して両辺の期待値をとると、$E[Y]$ は定数であることに注意して、(右辺)$= E[Y]E[Y] - E[E[Y \mid X]]E[Y] + E[YE[Y \mid X]] - E[\{E[Y \mid X]\}^2] = E[YE[Y \mid X]] - E[\{E[Y \mid X]\}^2] = E[E[Y \mid X]E[Y \mid X]] - E[\{E[Y \mid X]\}^2] = 0$. (3) 証明すべきの式 $\mathrm{Var}(X) = E[\mathrm{Var}(X \mid Y)] + \mathrm{Var}(E[X \mid Y])$ における右辺第 1 項は $E[\mathrm{Var}(X \mid Y)] = E\left[E[X^2 \mid Y] - \{E[X \mid Y]\}^2\right] = E\left[E[X^2 \mid Y]\right] - E\left[\{E[X \mid Y]\}^2\right] = E\left[X^2\right] - E\left[\{E[X \mid Y]\}^2\right]$ であり、また右辺第 2 項は、分散の定義から $\mathrm{Var}(E[X \mid Y]) = E\left[\{E[X \mid Y]\}^2\right] - \{E[E[X \mid Y]]\}^2 = E\left[\{E[X \mid Y]\}^2\right] - \{E[X]\}^2$ となる。加え合わすことで $E[\mathrm{Var}(X \mid Y)] + \mathrm{Var}(E[X \mid Y]) = E[X^2] - \{E[X]\}^2 = \mathrm{Var}(X)$ したがって関係式が証明された。(4) 条件つき分散は $\mathrm{Var}(X \mid Y) \geq 0$ であるから、常に $E[\mathrm{Var}(X \mid Y)] \geq 0$ となる。よって $\mathrm{Var}(E[X \mid Y]) \leq \mathrm{Var}(X)$ が得られる。条件を与えることで変動の要因が少なること、条件つきのほうが分散が小さくなることがあるを示している。この等号が成り立つときは $E[\mathrm{Var}(X \mid Y)] = 0$ のとき。つまり、$\mathrm{Var}(X \mid Y) \geq 0$ だから $\mathrm{Var}(X \mid Y) = 0$ でなければならない, Y の情報が完全に X を決定する、言いかえると、ある関数 h があって、$X = h(Y)$ のときに他ならない。

問 4.6.9　$m = 1$ とした指数族であるから、$f_\theta(x) = a(\theta)b(x)\exp[p(\theta)K(x)]$ として、$f_\theta(x_1, x_2, \cdots, x_n) = \prod_i f_\theta(x_i) = \prod_i a(\theta)b(x_i)\exp[p(\theta)K(x_i)] = \{a(\theta)\}^n \prod_i b(x_i) \exp[p(\theta)\sum_i K(x_i)]$ の形。指数関数を $\exp[p(\theta)\sum_i K(x_i)] = \exp[p(\theta)y] = g(\theta, u(x)) = g(\theta, y)$ とし、$h(x) = \{a(\theta)\}^n \prod_i b(x_i)$ と分解できるから、十分統計量である。完備性は、指数関数 $y = \exp(x), -\infty < x < \infty$ がゼロになることはないから。

問 4.6.10　（略）

問 4.6.11　$f_\theta(x) = \exp[-\sum_i(x_i - \theta)]I_{\min x_i > \theta}$,　$x = (x_1, \cdots, x_n)$ だから、$Y = \min X_i$ は十分統計量。完備であることは

$$P(Y_1 > y) = P(X_1 > y) \times \cdots \times P(X_n > y) = \left(\int_y^\infty \exp\{-(x-\theta)\}\, dx \right)^n = \exp\{-n(y-\theta)\}$$

したがって

$$F_{Y_1}(y) = 1 - \exp\{-n(y-\theta)\}, \quad f_{Y_1}(y) = ne^{n(y-\theta}I_{y>\theta}.$$

これからパラメータ θ のもとでの $g(Y_1)$ の期待値は

$$E_\theta[g(Y_1)] = \int_{-\infty}^\infty g(y) f_{Y_1}(y)\, dy = \int_\theta^\infty g(y) ne^{n(y-\theta)}\, dy$$

で、すべてのパラメータでゼロとなるときは, $ne^{-n\theta}$ で割ると $\int_y^\infty g(y)e^{ny}dy = 0$ となる。微分をすれば、$-g(\theta)\exp(n\theta) = 0$ となり、$g(\theta) = 0, \forall\theta$. すなわち完備であることが示された。

つぎに Y の推定量 $\phi(Y)$ で不偏統計量を求めればよいから、$\phi(y) = y - 1/n$ とすればよいことを示す。

$$E_\theta[Y] = \int_{-\infty}^\infty y f_{Y_1}(y)\, dy = \int_\theta^\infty yne^{n(y-\theta)}\, dy = \int_\theta^\infty (y-\theta)ne^{n(y-\theta)}\, dy + \theta \int_\theta^\infty ne^{n(y-\theta)}\, dy = \frac{1}{n} + \theta$$

すなわち不偏推定量 $E_\theta[\phi(Y)] = E[\phi(Y) \mid \theta] = E[Y - \frac{1}{n} \mid \theta] = \theta$.

問 4.6.12　密度関数は $f_\theta(x) = \theta x^{\theta-1} = \theta e^{(\theta-1)\ln x}$ と変形できるから、この形から指数族に属する。$f_\theta(x_1, \cdots, x_n) = \prod_i f_\theta(x_i) = \prod_i \theta e^{(\theta-1)\ln x_i} = \theta^n e^{(\theta-1)\sum_i \ln x_i} = \theta^n e^{(\theta-1)\ln y} = \theta^n y^{\theta-1} = \{\theta u(x)^{(\theta-1)}\}^n$ こ

こで $y = \prod_i x_i$, $u(x) = y^{1/n} = (\prod_i x_i)^{1/n}$ した。このように分解できるから、$u(X) = y^{1/n} = (\prod_i X_i)^{1/n}$ が十分統計量である。完備性は指数族であるから成り立つ。最尤推定量は、同時密度関数の対数、すなわち対数尤度を計算してみればよい。対数をとると $\ln f_\theta(x) = n\theta + (\theta-1)\ln y$ だから、両辺を微分して、ゼロとおくことから、最尤推定量が求められる。すなわち $n\frac{1}{\theta} + \ln y = 0$ より、推定量を $\hat{\theta}$ として $\hat{\theta} = -n/\ln Y = -1/\ln u(X)$ となり、$u(X) = u(X_1, \cdots, X_n)$ の関数である。鍵となっている性質は、Y が十分統計量であれば、1 対 1 対応の関数 g による情報 $g(Y)$ も同じ情報を与え、逆もそう。

問 4.6.13　密度関数 f_θ の形から、$X \sim \mathrm{Gamma}(2, 1/\theta)$ である。よってその n 個の和の分布は $Y = \sum_i X_i \sim \mathrm{Gamma}(2n, 1/\theta)$ となる。期待値は

$$E_\theta\left[\frac{1}{Y}\right] = \int_0^\infty \frac{1}{y}\frac{1}{\Gamma(2n)(1/\theta)^{2n}} y^{2n-1} e^{\theta y}\,dy = \frac{\theta}{\Gamma(2n)}\Gamma(2n-1) = \frac{\theta}{2n-1}$$

ここで、途中の積分計算には $z = \theta y$ と変換すればよい。したがって $E_\theta\left[\frac{2n-1}{Y}\right] = \theta$ となるから、不偏推定量は $\frac{2n-1}{Y}$ でしかも UMVUE である。

問 4.6.14　ベルヌーイ分布の密度は、$f_\theta(x) = \theta^x(1-\theta)^{1-x} = (1-\theta)\left(\frac{\theta}{1-\theta}\right)^x = (1-\theta)e^{x\ln(\theta/1-\theta)}$, $x = 0, 1$ となるから、指数族である。2 項分布 $\mathrm{Binom}(n,\theta)$ も同様に、$f_\theta(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x} = \binom{n}{x}(1-\theta)^n\left(\frac{\theta}{1-\theta}\right)^x = (1-\theta)^n\binom{n}{x}e^{x\ln(\theta/1-\theta)}$, $x = 0, 1, \cdots, n$ となり、指数族である。ベルヌーイ試行の同時分布は $f_\theta(x_1, \cdots, x_n) = \theta^{x_1}(1-\theta)^{1-x_1} \times \cdots \times \theta^{x_n}(1-\theta)^{1-x_n} = \theta^{\sum_i x_i}(1-\theta)^{n-\sum_i x_i} = \theta^y(1-\theta)^{n-y} = (1-\theta)^n(\theta/1-\theta)^y$ から、完備十分統計量であり、不偏性も $E_\theta[Y] = E_\theta[X_1 + \cdots + X_n] = n\theta$ から $E_\theta[Y/n] = \theta$ から示され、よって Y/n が UMVUE となる。標本平均が母平均の推定量としてよく用いられる由縁である。

つぎに $Y_2 = (X_1 + X_2)/2$ の期待値は $E[Y_2] = E_\theta[Y_2] = (E_\theta X_1 + E_\theta X_2)/2 = (\theta+\theta)/2 = \theta$ であり、$E[Y_1] = E_\theta[Y_1] = E_\theta[X_1 + \cdots + X_n] = n\theta$ となっている。つまり、Y_1/n は θ の不偏推定量である。条件つき期待値の関係式から、一般に $E\left[E[Y_2 \mid Y_1]\right] = E[Y_2]$ が成り立つ。2 つの不偏推定量 $E[Y_2 \mid Y_1], Y_1/n$ があって $E\left[E[Y_2 \mid Y_1] - Y_1/n\right] = 0$ であるから、十分性と完備性の性質から、$E[Y_2 \mid Y_1] - Y_1/n = 0$. $E[Y_2 \mid Y_1] = Y_1/n$ が成り立つ。これから条件つき期待値で $E[X_1 + X_2 \mid X_1 + \cdots + X_n = k] = 2k/n\, k = 0, 1, 2, \cdots, n$.

問 4.6.15　仮定から $X_i \sim \mathrm{N}(0,\theta)$ を標準正規分布に直すと $X_i/\sqrt{\theta} \sim \mathrm{N}(0,1)$, $i = 1, 2, \cdots, n$ である。よってその 2 乗値の和 $Y/\theta = \sum_i (X_i/\sqrt{\theta})^2$ は n 個が独立であるから、自由度 n のカイ 2 乗分布 $\chi(n)$ となる。平均は n, 分散は $2n$。2 次モーメントは分散と平均の 2 乗の和であるから、$E[(Y/\theta)^2] = n^2 + 2n$ となっている。したがって $E[Y^2] = \theta^2(n^2 + 2n)$ より、不偏推定量は $E[Y^2/(n^2+2n)] = \theta^2$ となる。これが θ^2 に対する UMVUE である。

問 4.6.16　ポアソン分布の密度関数は $f_\theta(x) = \frac{\theta^x}{x!}e^{-\theta}$, $x = 0, 1, 2, \cdots$ であり、$f_\theta(x) = e^{-\theta}\frac{1}{x!}\exp\{x\ln\theta\}$ となるから指数族である。よって $Y = \sum_i X_i$ は完備十分統計量となる。

(a)　$E\left[E[I_{\{X_1 \le 1\}} \mid Y]\right] = E[I_{\{X_1 \le 1\}}] = P(X_1 \le 1) = P(X_1 = 0) + P(X_1 = 1) = \exp(-\theta)/0! + \theta\exp(-\theta)/1! = (1+\theta)\exp(-\theta)$, $P(X_1 \le 1, Y = y) = P(X_1 = 0, Y = y) + P(X_1 = 1, Y = y)$ であり、$P(X_1 = 0, Y = 0) = 1, P(X_1 = 0, Y = 1) = 0$, また同一分布であるから $P(X_1 = 1, Y = 0) = 0, P(X_1 = 1, Y = 1) = 1/n$.

(b)　$P(X_1 = r \mid Y = s) = P(X_1 = r \mid X_1 + \cdots + X_n = s) = \frac{P(X_1 = r,\, X_1 + \cdots + X_n = s)}{P(X_1 + \cdots + X_n = s)} = \frac{P(X_1 = r,\, X_2 + \cdots + X_n = s - r)}{P(X_1 + \cdots + X_n = s)} = \frac{P(X_1 = r)\,P(X_2 + \cdots + X_n = s - r)}{P(X_1 + \cdots + X_n = s)}$, ここで (分子) $= \frac{e^{-\theta}\theta^r}{r!} \times e^{-(n-1)\theta}\frac{\{(n-1)\theta\}^{s-r}}{(s-r)!)}$, (分母) $= \frac{\exp(-n\theta)(n\theta)^s}{s!}$ であるから、整理をすると、$\frac{s!}{r!(s-r)!}\frac{(n-1)^{s-r}}{n^s} = \binom{s}{r}\left(\frac{1}{n}\right)^r\left(\frac{n-1}{n}\right)^{s-r}$, $r = 1, 2, \cdots, s \sim \mathrm{Binom}\left(s, \frac{1}{n}\right)$. とくに $r = 0$ のときには、$\frac{1}{n}$.

(c) $y > 0$ ならば、$P(X_1 = 0 \mid Y = y) = \left(1 - \dfrac{1}{n}\right)^y$, $P(X_1 = 1 \mid Y = y) = y\left(\dfrac{1}{n}\right)\left(1 - \dfrac{1}{n}\right)^{y-1}$. 和をとると、$P(X_1 \leq 1 \mid Y = y) = \left(1 - \dfrac{1}{n}\right)^y + y\left(\dfrac{1}{n}\right)\left(1 - \dfrac{1}{n}\right)^{y-1} = \left(\dfrac{n-1}{n}\right)^y\left(1 + \dfrac{y}{n-1}\right)$. $y = 0$ のときは $E[X_1 \leq 1 \mid Y = 0] = P[X_1 = 0 \mid Y = 0] = 1$.

問 4.6.17 $\theta = (\theta_1, \theta_2)$ とすると、$\prod_i I_{\{x_i > \theta_1\}} = I_{\{\min_i x_i > \theta_1\}}$ となるから同時密度関数は

$$f_\theta(x_1, \cdots, x_n) = \frac{1}{\theta_2^n}\exp\left[-\sum_i \frac{x_i - \theta_1}{\theta_2}\right] I_{\{\min_i x_i > \theta_1\}} = \frac{1}{\theta_2^n}\exp\left[-\frac{1}{\theta_2}\left(\sum_i x_i - n\theta_1\right)\right] I_{\{\min_i x_i > \theta_1\}}$$

となり、$Y = (Y_1, Y_2)$, ただし $Y_1 = \min_i X_i$, $Y_2 = \sum_i X_i$ が十分統計量であることがこの形からわかる。

第 4.7 節（4.7 ページ） ==

問 4.7.1 （略）

問 4.7.2 （略）

問 4.7.3 （略）

問 4.7.4 $h(\theta) = \theta^{r-1}(1-\theta)^{s-1}$, $f_\theta(x) = \dbinom{n}{x}\theta^x(1-\theta)^{n-x}$ であり、$\delta(x)$ の分子は $\int_0^1 \theta h(\theta) f_\theta(x)\, d\theta = \mathrm{B}(r+x+1, n-x+s)$, また分母は $\int_0^1 h(\theta) f_\theta(x)\, d\theta = \mathrm{B}(r+x, n-x+s)$. したがって $\delta(x) = \dfrac{\Gamma(r+x+1)}{\Gamma(r+x)}\dfrac{\Gamma(r+s+n)}{\Gamma(r+s+n+1)} = \dfrac{r+x}{r+s+n}$.

問 4.7.5 リスク関数は $R_\theta(\theta) = E_\theta[(\delta(X) - \theta)^2]$ において、$\delta(X) = \dfrac{r+X}{r+s+n}$ を代入し、$E_\theta[X - n\theta] = 0$, $E_\theta[(X - n\theta)^2] = \mathrm{Var}(X) = n\theta(1-\theta)$ をもちいると

$$\begin{aligned} R_\delta(\theta) &= \frac{1}{(r+s+n)^2}\{n\theta(1-\theta) + (r - r\theta - s\theta)^2\} \\ &= \frac{1}{(r+s+n)^2}\left[\left((r+s)^2 - n\right)\theta^2 + \left(n - 2r\,(r+s)\right)\theta + r^2\right] \end{aligned}$$

問 4.7.6 $r = s = \dfrac{\sqrt{n}}{2}$ であれば、$(r+s)^2 - n = 0$, $n - 2r\,(r+s) = 0$ から、$R_\delta(\theta) = \dfrac{r^2}{(r+s+n)^2}$ で θ に依らない。

問 4.7.7 決定 δ をもちいたときの期待損失 $B(\delta) = \int_{-\infty}^{\infty} h(\theta) R_\delta(\theta)\, d\theta$ とする。一方 ϕ をより小さな決定とすると、$R_\phi(\theta) < R_\delta(\theta) = R_\delta, \forall\theta$. 期待損失が $B(\phi) < B(\delta)$ となるから、θ がベイズ推定であることに矛盾する。

第 5 章

第 5.1 節（5.1 ページ） ==

問 5.1.1 平均 μ_X, μ_Y, 分散 $\sigma_X^2 = \mathrm{Var}(X)$, $\sigma_Y^2 = \mathrm{Var}(Y)$、共分散 $\sigma_{XY} = \mathrm{Cov}(X, Y)$ とすると、

$$M(c, d) = E[(\{Y - \mu_y\} - \{c(X - \mu_X) + d\})^2] = \sigma_Y^2 + c^2\sigma_X^2 + d^2 - 2c\sigma_{XY}$$

これから、最小化を考えると $c = \dfrac{\sigma_{XY}}{\sigma_X^2} = \dfrac{\rho\sigma_X\sigma_Y}{\sigma_X^2} = \rho\dfrac{\sigma_Y}{\sigma_X}$, また $d = 0$ となる。最小値は $\sigma_Y^2(1 - \rho^2)$.

問 5.1.2 (1) $E[XY] = E[X]E[Y]$ より、共分散がゼロとなるから、相関係数もゼロ。(2) $X = \cos\theta$, $Y = \sin\theta$ であれば、$X^2 + Y^2 = 1$ の関係から、独立ではない。しかし $E[XY] - E[X]E[Y] = \dfrac{1}{2\pi}\int_0^{2\pi} \sin\theta\cos\theta\, d\theta - \dfrac{1}{2\pi}\int_0^{2\pi}\cos\theta\, d\theta \cdot \dfrac{1}{2\pi}\int_0^{2\pi}\sin\theta\, d\theta = 0$. (3) 2 変量正規分布では、相関係数が ρ であり、密度関数の形から、これがゼロであることが、独立である、すなわち 2 つの密度関数の積となるための必要十分条件である。

問 5.1.3 (1) $L(a, b) = \sum_i\{y_i - (ax_i + b)\}^2$ の 1 次、2 次微分係数を求める。$\frac{\partial L}{\partial a} = -2\sum_i\{y_i - (ax_i + b)\}x_i = -2(s_{xy} - as_{xx} - bs_x)$
$\frac{\partial L}{\partial b} = -2\sum_i\{y_i - (ax_i + b)\} = -2(s_y - as_x - nb)$ ここで、$s_x = \sum_i x_i$, $s_y = \sum_i y_i$, $s_{xx} = \sum_i x_i x_i$, $s_{xy} = \sum_i x_i y_i$ と、$\overline{x} = \frac{1}{n}s_x$, $\overline{y} = \frac{1}{n}s_y$, $\sigma_{xx} = \frac{1}{n}s_{xx} - (s_x/n)^2$ などとおいた。この微分係数をゼロとおいて得られる a, b に関

する連立方程式（正規方程式）：

$$\left\{\begin{array}{l} a\,s_{x^2} + b\,s_x \ = \ s_{xy} \\ a\,s_x + n\,b \ = \ s_y \end{array}\right. , \qquad \begin{pmatrix} s_{x^2} & s_x \\ s_x & n \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} s_{xy} \\ s_y \end{pmatrix}$$

を解いて

$$\left\{\begin{array}{l} a = \dfrac{ns_{xy} - s_x s_y}{ns_{xx} - (s_x)^2} = \dfrac{\sigma_{xy}}{\sigma_{xx}}, \\ b = -\overline{x}\left(\frac{\sigma_{xy}}{\sigma_{xx}}\right) + \overline{y} \end{array}\right.$$

となる。また $\dfrac{\partial^2 L}{\partial a^2} = 2\sum_i x_i^2$, $\dfrac{\partial^2 L}{\partial b^2} = 2\sum_i y_i^2$, $\dfrac{\partial^2 L}{\partial a \partial b} = 2n$. であるから、Hesse 行列の行列式は、$n > 0$, $s_{xx} > 0$, $ns_{xx} - (s_x)^2 = n^2(\sigma_{xx}) > 0$ で正定値であるから、極値が最小値となる。

問 5.1.4　(1) 独立性から、積を分解できるから、$\mathrm{Var}(XY) = E[(XY)^2] - (E[XY])^2 = E[X^2]E[Y^2] - (E[X]E[Y])^2 = (\sigma_X^2 + \mu_X^2)(\sigma_Y^2 + \mu_Y^2) - \mu_X^2\mu_Y^2 = \sigma_X^2\sigma_Y^2 + \mu_X^2\sigma_Y^2 + \mu_Y^2\sigma_X^2$

(2) $\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Y} + 2ab\mathrm{Cov}(X, Y) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\rho\sigma_X\sigma_Y$ 独立としているから、$\rho = 0$. よって $\mathrm{Var}(aX + bY) = \quad a^2\sigma_X^2 + b^2\sigma_Y^2$

(3) $\mathrm{Cov}(X, X + Y) = \mathrm{Cov}(X, X) + \mathrm{Cov}(X, Y) = \mathrm{Var}(X, X) + \mathrm{Cov}(X, Y) = \sigma_X^2 + \rho\sigma_X\sigma_Y = \sigma_X^2$.

(4) 相関係数の添え数に確率変数を明記してこなかった理由は、当然 (X, Y) という組が前提であったからでいまの場合 $(X, X + Y)$ であるから、これを記して、前述の結果をもちいると、

$$\rho_{X,X+Y} = \frac{\mathrm{Cov}(X, X + Y)}{\sigma_X\,\sigma_{X+Y}} = \frac{\sigma_X^2}{\sigma_X\sqrt{\sigma_X^2 + \sigma_Y^2}} = \frac{\sigma_X}{\sqrt{\sigma_X^2 + \sigma_Y^2}}$$

(5)

$$\mathrm{Cov}(X,\, XY) = E[(X \cdot XY)] - E[X]\,E[XY] = E[X^2]E[Y] - (E[X])^2 E[Y]$$
$$= (\sigma_X^2 + \mu_X^2)\,\mu_Y - \mu_X^2\,\mu_Y = \sigma_X^2\,\mu_Y$$

問 5.1.5　$S_x^2 = \sum_i x_i^2$, $S_y^2 = \sum_i y_i^2$, $S_{xy} = \sum_i x_i y_i$ とおく。データの和と差とのそれぞれの場合 $x_i \pm y_i$ を同時に考えて、$0 \le \sum_i (x_i \pm y_i)^2 = \sum_i x_i^2 \pm 2\sum_i x_i y_i + \sum_i y_i^2 = S_x^2 \pm 2S_{xy} + S_y^2 = (S_x \pm \frac{S_{xy}}{S_x})^2 + \dfrac{S_x^2 S_y^2 - S_{xy}^2}{S_x^2}$. したがって $S_x^2 S_y^2 - S_{xy}^2 \ge 0$. 等号は、$(x_i \pm y_i)^2 = 0, \forall i$ から、$x_i = \pm y_i$ のとき。直線 $y = x$ あるいは $y = -x$ 上に点 $(x_i, y_i), i = 1, 2, \cdots, n$ が並んでいるときである。

第 5.2 節（5.2 ページ）==

問 5.2.1　$Y_i = X_i - E[X_i]$ とすると、２乗モーメント値は 任意の $t = (t_1, \cdots, t_n)$ で $M_Y(t) = E\left[(\sum_i t_i Y_i)^2\right] \ge 0, \forall t$ となり、期待値は分散共分散行列 $K = (\sigma_{ij}), \sigma_{ij} = \mathrm{Cov}(Y_i, Y_j)$ をもちいて $M_Y(t) = E\left[(\sum_i t_i Y_i)(\sum_j t_j Y_j)\right] = \sum_{i,j} t_i \sigma_{ij} t_j = t' K t$, K は対称行列で、K の固有値 λ_i はすべて非負。もし K が正則であれば、固有値 $\lambda_i > 0$ より、定正値。特異であれば、$\lambda_i \le 0$.

問 5.2.2　$\underline{X} = C\underline{Z} + \underline{\mu}$ で、行列 C は非特異で、$\underline{Z}$ の平均ベクトルは 0 ベクトル。$\underline{Y} = A\underline{X} = AC\underline{Z} + A\underline{\mu}$ となって正規分布である。

問 5.2.3　$M_X(c_1 t, \cdots, c_n t) = \exp\left(t\sum_i c_i \mu_i + \dfrac{t^2}{2}\sum_{i,j} c_i a_{ij} c_j\right)$ において、m 次元の周辺分布の積率母関数は $c_{m+1}, \cdots, c_n = 0$ とすればよいから、やはり多変量正規分布の形である。

問 5.2.4　$Y = \sum_i c_i X_i$ のとき、$M_Y(t) = E[e^{tY}] = E[\exp(\sum_i c_i t X_i)] = M_X(c_1 t, \cdots, c_n t) = \exp\left(t\sum_i c_i \mu_i + \dfrac{t^2}{2}\sum_{i,j} c_i a_{ij} c_j\right)$. $W = c_1 X_1 + c_2 X_2 + \cdots + c_n X_n = \underline{c'}\underline{X} = \underline{c'}(A\underline{Y} + \underline{\mu}) = \underline{b'}\underline{Y} + \underline{c'}\underline{\mu}$ という形になるから、やはり多変量正規分布である。

問 5.2.5　$X = x$ のとき、Y の最良推定値は $y - \mu_Y = \rho\dfrac{\sigma_Y}{\sigma_X}(x - \mu_X)$ で, 最小平均偏差は $\sigma_Y^2(1 - \rho^2))$ である。このことから、$\rho\dfrac{\sigma_Y}{\sigma_X} = 3$, $\mu_Y - \rho\dfrac{\sigma_Y}{\sigma_X}\mu_X = 7$. これらの関係から、$\rho\sigma_Y = 6$, また $\sigma_Y^2(1 - \rho^2)) = 28$ より、$\sigma_Y = 8$, $\rho^2 = 36/64 = 9/16$, $\rho = 3/4 = 0.75$.

問 5.2.6 パラメータ θ を含む 2 変量分布関数 $f_\theta(x,y)$ は

$$f_\theta(x,y) = a(\theta)\, b(x,y)\, \exp\{p_1(\theta)x^2 + p_2(\theta)y^2 + p_3(\theta)xy + p_4(\theta)x + p_5(\theta)y\}$$

の形に変形できることを計算する。したがって

$$\left(\sum_i X_i^2,\ \sum_i Y_i^2,\ \sum_i X_iY_i,\ \sum_i X_i,\ \sum_i Y_i\right)$$

が完備十分統計量で、推定するパラメータ θ は

$$\theta = \left(\sigma_X^2,\ \sigma_Y^2,\ \rho_{XY},\ \sigma_Y^2,\ \mu_X,\ \mu_Y\right)$$

またそれぞれが 1 対 1 対応の完備十分推定量である。

問 5.2.7 確率変数 $Y = f(X)$ を考え、連続型とし、$A = \{x : a \le f(x) \le b\}$ とおけば $F_Y(b) - F_Y(a) = P(f(X) \le b) - P(f(X) \le a) = P(a < X \le b) = \int_A f(x)dx$ 連続型では不等号に等号が含まれていてもいなくても同じ確率であることに注意する。離散型の場合には、領域の範囲から対応する値を求めればよい。

問 5.2.8 (1) ベルヌーイ分布 $\mathrm{Bernoulli}(p) = \mathrm{Binom}(1,p)$: $f_\theta(x) = \theta^x\,(1-\theta)^{1-x},\ x = 0, 1$ だから

$$\frac{\partial}{\partial\theta}\ln f_\theta(x) = \frac{\partial}{\partial\theta}[x\ln\theta + (1-x)\ln(1-\theta)] = \frac{x}{\theta} - \frac{1-x}{1-\theta}$$
$$\frac{\partial^2}{\partial\theta^2}\ln f_\theta(x) = -\frac{x}{\theta^2} - \frac{1-x}{(1-\theta)^2}$$

よって

$$I_\theta = E_\theta\left[\frac{x}{\theta^2} + \frac{1-x}{(1-\theta)^2}\right] = \frac{1}{\theta} + \frac{1}{1-\theta} = \frac{1}{\theta\,(1-\theta)}$$

クラメル=ラオの不等式では

$$\mathrm{Var}_\theta Y \ge \frac{1}{nI(\theta)} = \frac{\theta(1-\theta)}{n}$$

標本平均 $\overline{X} = \frac{1}{n}\sum X_i$ の分散は

$$\mathrm{Var}(\overline{X}) = \frac{1}{n^2}n\theta(1-\theta) = \frac{\theta(1-\theta)}{n}$$

したがって $\overline{X}$ が θ の UMVUE である。

(2) 正規分布 $\mathrm{N}(\theta,\sigma^2)$: $f_\theta(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\dfrac{(x-\theta)^2}{2\sigma^2}\right),\ -\infty < x < \infty$

$$\frac{\partial}{\partial\theta}\ln_\theta f_\theta(x) = \frac{\partial}{\partial\theta}\left[-\frac{(x-\theta)^2}{2\sigma^2}\right] = \frac{x-\theta}{\sigma^2},\quad \frac{\partial^2}{\partial\theta^2}\ln_\theta f_\theta(x) = -\frac{1}{\sigma^2}.$$

よって $I(\theta) = \dfrac{1}{\sigma^2}$ を得る。

$$\mathrm{Var}_\theta(Y) \ge \frac{1}{nI(\theta)} = \frac{\sigma^2}{n}$$

であり、$\mathrm{Var}_\theta(\overline{X}) = \sigma^2/n$ から、これは UMVUE である。

(3) ポアソン分布: $f_\theta(x) = \dfrac{\theta^x}{x!}e^{-\theta},\ x = 0, 1, 2, \cdots$

$$\frac{\partial}{\partial\theta}\ln f_\theta(x) = \frac{\partial}{\partial\theta}(-\theta + x\ln\theta) = -1 + \frac{x}{\theta}$$
$$\frac{\partial^2}{\partial\theta^2}\ln f_\theta(x) = -\frac{x}{\theta^2}$$

よって $I(\theta) = E\left[\dfrac{X}{\theta^2}\right] = \theta/\theta^2 = 1/\theta$ を得る。

$$\mathrm{Var}_\theta(Y) \ge \frac{1}{nI(\theta)} = \frac{\theta}{n} = \mathrm{Var}_\theta(\overline{X})$$

であり、UMVUE である。

第 5.3 節（5.3 ページ）==

問 5.3.1　帰無仮説 H_0 のもとでは $p=1/2, c=2$ とするから検出力は $K(p)=\sum_{k=0}^{c}\binom{n}{k}p^k(1-p)^{n-k}=\binom{n}{0}(1-p)^n+\binom{n}{1}p(1-p)^{n-1}+\binom{n}{2}p^2(1-p)^{n-2}=(1-p)^{n-2}\{(1-p)^2+np(1-p)+\frac{n(n-1)}{2}p^2\}$.

$n=12$ を代入すると、$\alpha=(1/2)^{12}\{1+12+66\}=79/4096=0.019$

問 5.3.2　各データ $x_i-m(i=1,2,\cdots,16)$ を計算し、その順位を求める。

x_i	$\mathrm{abs}(x_i-m)$	rank	±1	wilcoxson
176.9	16.9	14	+1	14
158.3	1.7	5	-1	-5
152.1	7.9	9	-1	-9
158.8	1.2	4	-1	-4
172.4	12.4	12	+1	12
169.8	9.8	10	+1	10
159.7	0.3	2	-1	-2
162.7	2.7	6	+1	6
156.6	3.4	7	-1	-7
174.5	14.5	13	+1	13
184.4	24.4	16	+1	16
165.2	5.2	8	+1	8
147.8	12.2	11	-1	-9
177.8	17.8	15	+1	15
160.1	0.1	1	+1	1
160.5	0.5	3	+1	3
				$W=60$

帰無仮説 H_0 のもとでは、$E[W]=0, \mathrm{Var}(W)=n(n+1)(2n+1)/6=16(16+1)(2\cdot16+1)/6=1496, \sigma_W=\sqrt{\mathrm{Var}(W)}=\sqrt{1496}=38.678$. 正規近似をもちいると $\frac{W-E[W]}{\sigma_W}=W/38.678\sim N(0,1)$. 右側 95% 点は 1.645 であるから、$P(W>c)=P(W/\sigma_W>c/38.678)=0.05$ だから $c/38.678=1.645$, $c=63.626$. $W<c$ であるから、有意ではない。帰無仮説 H_0 は採択される。

問 5.3.3　V_j の確率母関数は $M_{V_j}(t)=1/2(e^{-jt}+e^{jt})$, W について $M_W(t)=\prod_j M_{V_j}(t)$ となる。$n=1$ では明らかに $W=\pm1$ で、$n=2$ では

$$M_W(t)=\frac{1}{2}(e^{-t}+e^{t})\frac{1}{2}(e^{-2t}+e^{2t})=\frac{1}{4}(e^{-3t}+e^{-t}+e^{t}+e^{3t})$$

これをまとめると

W 値	-3	-1	1	3	計
場合数	1	1	1	1	$2^2=4$

$n=3$ では

$$M_W(t)=\frac{1}{4}(e^{-3t}+e^{-t}+e^{t}+e^{3t})\times\frac{1}{2}(e^{-3t}+e^{3t})=\frac{1}{8}(e^{-6t}+e^{-4t}+e^{-2t}+1+1+e^{2t}+e^{4t}+e^{6t})$$

W 値	-6	-4	-2	0	2	4	6	計
場合数	1	1	1	2	1	1	1	$2^3=8$

$n=4$ では

$$M_W(t)=\tfrac{1}{8}(e^{-6t}+e^{-4t}+e^{-2t}+1+1+e^{2t}+e^{4t}+e^{6t})\times\tfrac{1}{2}(e^{-4t}+e^{4t})$$
$$=\tfrac{1}{2^4}(e^{-10t}+e^{-8t}+e^{-6t}+\cdots+e^{8t}+e^{10t})$$

W 値	-10	-8	-6	-4	-2	0	2	4	6	8	10	計
場合数	1	1	1	2	2	2	2	2	1	1	1	$2^4=16$

以上