「データ解析のための数理統計入門」

更新日：2025 年 2 月 13 日

–演習問題の解答例–

このファイルでは，各章末にある演習問題の解答例が与えられています。一つの解き方ですので，別の解き方やわかりやすい解き方など工夫して下さい。また，補足事項をまとめた別のファイルには，追加の演習問題や割愛した演習問題など，発展的な問題も含めて取り上げていますので，興味のある方は参照して下さい。

初めて演習問題を解いてみると，わからないことや「この解き方は思いつかない」と感ずる方が多いと思います。多くの方からそのような声を聞いています。しかし，演習問題は慣れです。解答の 2 週目に入ると，かなりの問題が解けるようになっていると思います。2 週目で解けなかった問題にはチェックを入れ，3 週目にはそのチェックを入れた問題だけ解くのもいいと思います。この本の演習問題を制覇した後は,「現代数理統計学の基礎」の演習問題に挑戦してみて下さい。それらが制覇できれば，相当な力が付いていますので，1 級試験の数理と応用にトライすることをお勧めします。

なお，多くの読者の皆さまから，解答例の誤りについてのご指摘を頂きました。大変にありがとうございます。ここに御礼申し上げます。

解答例については随時修正し更新させて頂きますので，どうぞ宜しくお願い致します。

第1章　確率モデル

問 1　大学生の就職活動に関して，ある企業にエントリーした学生は男性が 60% をしめていたが，採用を内定された割合は，男性は男性応募者の 40%, 女性は女性応募者の 30% であった。

(1) エントリーした学生数を n とすると，エントリーした男性は $0.6n$ 人，女性は $0.4n$ 人になる。内定数は，男性が $0.4 \times (0.6n) = 0.24n$ 人，女性が $0.3 \times (0.4n) = 0.12n$ 人になるので，男性の採用内定率は 24%，女性の採用内定率は 12% になる。

(2) この企業の男女合わせた全体の内定数は $0.24n + 0.12n = 0.36n$ 人になるので，採用内定率は 36% になる。

(3) $\mathrm{P}(\text{男性} \cap \text{内定}) = 0.24$, $\mathrm{P}(\text{女性} \cap \text{内定}) = 0.12$ であり，$\mathrm{P}(\text{内定}) = 0.36$ であることに注意する。採用内定者の名簿から一人をランダムに選んだ場合，それが女性である確率は，ベイズの定理を用いると

$$\mathrm{P}(\text{女性} \mid \text{内定}) = \frac{\mathrm{P}(\text{女性} \cap \text{内定})}{\mathrm{P}(\text{内定})} = \frac{0.12}{0.36} = \frac{1}{3}$$

となる。

問 2　週末の天気予報で，土曜日に雨の降る事象を $A = \text{土曜日} \cap \text{雨}$ とし，日曜日に雨の降る事象を $B = \text{日曜日} \cap \text{雨}$ とすると，$\mathrm{P}(A) = \mathrm{P}(B) = 0.25$ である。求めたい確率は週末の土日のどこかで雨の降る確率なので，$\mathrm{P}(A \cup B)$ と書ける。従って

$$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B) = 0.5 - \mathrm{P}(A \cap B)$$

となる。$-\mathrm{P}(A \cap B)$ を見落としているので，50% という確率は誤りである。仮に A と B が独立であるとすると，$0.5 - 0.25 \times 0.25$ となる。

問 3　ある大学のある学部は A 学科と B 学科の 2 学科からなり，それぞれの志願者数と合格率が男女別に教科書の表で与えられている。その表から，A 学科と B 学科の 2 学科を合算したときの男女の合格率を計算すると男性の合格率は

$$\frac{500 \times 0.6 + 400 \times 0.1}{500 + 400} = \frac{340}{900} = \frac{17}{45}$$

となり，女性の合格率は

$$\frac{30 \times 0.7 + 300 \times (a/100)}{30 + 300} = \frac{21 + 3a}{330} = \frac{7 + a}{110}$$

となる。従って，シンプソンのパラドクスが生ずるためには，a は $a > 10$ で

$$\frac{17}{45} > \frac{7 + a}{110}$$

を満たせばよいので，a の条件は

$$10 < a < 34.6$$

となる。

問 4 (誕生日問題) 両親と 2 人の子どもがいる 4 人家族において，同じ誕生日の人がいる確率は

$$1 - \frac{365 \times 364 \times 363 \times 362}{365^4} = 0.016$$

となるので，1.6% である。

問 5 (1) 不等式 $\mathrm{P}(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n \mathrm{P}(A_i)$ については，数学的帰納法で示す。$n = 2$ のときには

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

となり，成り立つことがわかる。n のとき成り立つと仮定すると $\mathrm{P}(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n \mathrm{P}(A_i)$ が成り立つ。そのとき $n+1$ のときに成り立つことを示せばよい。

$$\begin{aligned}\mathrm{P}(\bigcup_{i=1}^{n+1} A_i) =& \mathrm{P}((\bigcup_{i=1}^{n} A_i) \cup A_{n+1}) \leq \mathrm{P}(\bigcup_{i=1}^{n} A_i) + \mathrm{P}(A_{n+1}) \\ &\leq \sum_{i=1}^{n} \mathrm{P}(A_i) + \mathrm{P}(A_{n+1}) = \sum_{i=1}^{n+1} \mathrm{P}(A_i)\end{aligned}$$

となり，$n+1$ のときにも成り立つことがわかる。

(2) $\mathrm{P}(A \cap B) \geq \mathrm{P}(A) + \mathrm{P}(B) - 1$ については

$$\mathrm{P}(A \cap B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cup B) \geq \mathrm{P}(A) + \mathrm{P}(B) - 1$$

となることからわかる。

問 6 条件付き確率について不等式 $\mathrm{P}(A|D) \geq \mathrm{P}(B|D)$, $\mathrm{P}(A|D^c) \geq \mathrm{P}(B|D^c)$ が成り立つことから

$$\begin{aligned}\mathrm{P}(A) =& \mathrm{P}(A|D)\mathrm{P}(D) + \mathrm{P}(A|D^c)\mathrm{P}(D^c) \\ \geq& \mathrm{P}(B|D)\mathrm{P}(D) + \mathrm{P}(B|D^c)\mathrm{P}(D^c) = \mathrm{P}(B)\end{aligned}$$

となる。

問 7 (1) 多項式 $(x+y)^6$ における x^3y^3 の係数は

$$\begin{pmatrix} 6 \\ 3 \end{pmatrix} = \frac{6!}{3!3!} = 20$$

となる。

(2) 多項式 $(x+y+z)^6$ における $x^2y^2z^2$ の係数は

$$\begin{pmatrix} & 6 & \\ 2 & 2 & 2 \end{pmatrix} = \frac{6!}{2!2!2!} = 90$$

となる。

問 8 正しいコインを 3 回投げるとき，現れる事象を列挙すると，HHH, HHT, HTH, THH, TTH, THT, HTT, TTT の 8 通りである。

(1) 少なくとも 1 回は表が出る事象の個数は 7 であり，2 回以上表の出る事象の個数は 4 であるから，少なくとも 1 回は表が出る条件のもとで 2 回以上表の出る確率は 4/7 になる。

(2) 少なくとも 1 回は裏が出る事象の個数は 7 であり，少なくとも 1 回は裏が出てしかも 2 回以上表の出る事象の個数は 3 であるから，少なくとも 1 回は裏が出るという条件のもとで 2 回以上表の出る確率は 3/7 となる。

問9 2人の子どものいる夫婦について，第1子が男，第2子が女である事象を (男, 女) と書くと，(男, 男)，(男, 女)，(女, 男)，(女, 女) の4通りである。

第1子が女の子であるという条件のもとで2人とも女の子である確率は 1/2 になる。

子どもの一人が女の子である事象の個数は3であるから，子どもの一人が女の子であるという条件のもとで2人とも女の子である確率は 1/3 となるよ。

問10 (1) A と B が独立なので $\mathrm{P}(A\cap B)=\mathrm{P}(A)\mathrm{P}(B)$ である。このとき，$\boldsymbol{A}\cap B^c=\boldsymbol{A}\backslash(A\cap B)$ より

$$\begin{aligned}\mathrm{P}(A\cap B^c) =&\mathrm{P}(A)-\mathrm{P}(A\cap B)=\mathrm{P}(A)-\mathrm{P}(A)\mathrm{P}(B)\\ =&\mathrm{P}(A)\{1-\mathrm{P}(B)\}=\mathrm{P}(A)\mathrm{P}(B^c)\end{aligned}$$

となる。また $A^c\cap B^c=(A\cup B)^c$ より

$$\begin{aligned}\mathrm{P}(A^c\cap B^c) =&\mathrm{P}((A\cup B)^c)=1-\mathrm{P}(A\cup B)=1-\{\mathrm{P}(A)+\mathrm{P}(B)-\mathrm{P}(A\cap B)\}\\ =&1-\mathrm{P}(A)-\mathrm{P}(B)+\mathrm{P}(A)\mathrm{P}(B)=\{1-\mathrm{P}(A)\}\{1-\mathrm{P}(B)\}=\mathrm{P}(A^c)\mathrm{P}(B^c)\end{aligned}$$

となる。

(2) A と B が独立で，B と C が独立であっても，A と C は独立になるとは限らない。例えば，$\Omega=\{1,2,3,4\}$, $\mathrm{P}(\{1\})=\mathrm{P}(\{2\})=\mathrm{P}(\{3\})=\mathrm{P}(\{4\})=1/4$ とし，$A=\{1,2\}$, $B=\{2,3\}$, $C=\{3,4\}$ とする。このとき，$\mathrm{P}(A\cap B)=1/4=\mathrm{P}(A)\mathrm{P}(B)$, $\mathrm{P}(B\cap C)=1/4=\mathrm{P}(B)\mathrm{P}(C)$ となるが，$\mathrm{P}(A\cap C)=\mathrm{P}(\emptyset)=0$, $\mathrm{P}(A)\mathrm{P}(C)=1/4$ となり，A と C は独立でない。

(3) A と B が排反であっても A と B が独立になるとは限らない。理由は，A と B が排反であるから $A\cap B=\emptyset$ より $\mathrm{P}(A\cap B)=0$ となるが，$\mathrm{P}(A)\mathrm{P}(B)$ は0になるとは限らないからである。例えば，上の(2)の例を用いると，$A=\{1,2\}$, $B=\{3,4\}$ については，$\mathrm{P}(A\cap B)=0\neq\mathrm{P}(A)\mathrm{P}(B)=(1/2)^2=1/4$ となる。

(4) $A\subset B$ であるとき，$\mathrm{P}(A\cap B)=\mathrm{P}(A)\neq\mathrm{P}(A)\mathrm{P}(B)$ より，A と B は独立にならない。

問11 事前確率が $\mathrm{P}(A_1)=0.5$, $\mathrm{P}(A_2)=0.5$ であるから，陽性反応が出たという条件のもとで感染している確率は，ベイズの定理を用いると

$$\mathrm{P}(A_1|B_1)=\frac{\mathrm{P}(B_1|A_1)\mathrm{P}(A_1)}{\mathrm{P}(B_1)}=\frac{0.9\times0.5}{0.9\times0.5+0.01\times0.5}=\frac{0.45}{0.455}=0.989$$

となり，98.9% の確率で感染していることになる。

問12 正しいコインを2回投げるとき，1回目に表が出る事象を A, 2回目に表の出る事象を B, 2回のうち1回だけ表の出る事象を C とする。従って，$A=\{HH,HT\}$, $B=\{HH,TH\}$, $C=\{HT,TH\}$ であり，$\mathrm{P}(A)=\mathrm{P}(B)=\mathrm{P}(C)=1/2$ である。

(1) $\mathrm{P}(A\cap B)=\mathrm{P}(\{HH\})=1/4=\mathrm{P}(A)\mathrm{P}(B)$, $\mathrm{P}(B\cap C)=\mathrm{P}(\{TH\})=1/4=\mathrm{P}(B)\mathrm{P}(C)$, $\mathrm{P}(C\cap A)=\mathrm{P}(\{HT\})=1/4=\mathrm{P}(C)\mathrm{P}(A)$ であるから，ペアワイズ独立である。

(2) $\mathrm{P}(A\cap B\cap C)=\mathrm{P}(\emptyset)=0\neq\mathrm{P}(A)\mathrm{P}(B)\mathrm{P}(C)=1/8$ となって，互いに独立ではないことがわかる。

問13 (モンティ・ホール問題) 3つの箱を A, B, C とすると，どれか1つだけに宝物が入っているので，宝物の当たる確率は，$\mathrm{P}(A)=\mathrm{P}(B)=\mathrm{P}(C)=1/3$ である。この設定のもとで，挑戦者は3つのうち1つの箱を選択し，司会者は残りの2つのうち一つの箱を開けて空であることを明らかにする。残された2つの箱のどちらかに宝物が入っているが，挑戦者にとって，(a) 選んだ箱を変えない，(b) 別の箱に変更する，のどちらの選択の方が確率が高いかを調べる。

いま，挑戦者が A を選択したとする。この確率は $\mathrm{P}(A)=1/3$ だから，(a) の選択をした場合，この確率は変わらない。(b) を選択した場合，$\mathrm{P}(B\cup C)=\mathrm{P}(B)+\mathrm{P}(C)=2/3$ であるから，別の箱に変更する方が確率が高くなる。

司会者が残りの 2 つのうち一つの箱を開けて空であることを明らかにすると，箱は 2 つになるので確率は 1/2 であると勘違いしてしまうことに注意する。

問 14 2 項係数の関係式 (1.2) については

$$\begin{aligned}\binom{n-1}{k-1}+\binom{n-1}{k} =& \frac{(n-1)!}{(k-1)!(n-k)!}+\frac{(n-1)!}{k!(n-k-1)!}\\ =& \frac{(k+n-k)(n-1)!}{k!(n-k)!}=\binom{n}{k}\end{aligned}$$

より示される。

2 項定理 (1.1) を示すには数学的帰納法を用いる。$n=1$ のときには

$$\binom{1}{0}a^0b^1+\binom{1}{1}a^1b^0=a+b$$

となるので成り立つことがわかる。

n のとき

$$(a+b)^n=\sum_{k=0}^{n}\binom{n}{k}a^kb^{n-k}$$

が成り立つと仮定する。

$n+1$ のときには

$$\begin{aligned}(a+b)^{n+1} =& \sum_{k=0}^{n}\binom{n}{k}a^kb^{n-k}(a+b)\\ =& \sum_{k=0}^{n}\binom{n}{k}a^{k+1}b^{n-k}+\sum_{k=0}^{n}\binom{n}{k}a^kb^{n-k+1}\\ =& a^{n+1}+\sum_{k=0}^{n-1}\binom{n}{k}a^{k+1}b^{n-k}+\sum_{k=1}^{n}\binom{n}{k}a^kb^{n+1-k}+b^{n+1}\\ =& a^{n+1}+\sum_{k=0}^{n-1}\left\{\binom{n}{k}+\binom{n}{k+1}\right\}a^{k+1}b^{n-k}+b^{n+1}\end{aligned}$$

と書ける。ここで (1.2) の関係式を用いると

$$\begin{aligned}(a+b)^{n+1} =& a^{n+1}+\sum_{k=0}^{n-1}\binom{n+1}{k+1}a^{k+1}b^{n-k}+b^{n+1}\\ =& a^{n+1}+\sum_{k=1}^{n}\binom{n+1}{k}a^kb^{n+1-k}+b^{n+1}\\ =& \sum_{k=0}^{n+1}\binom{n+1}{k}a^kb^{n+1-k}\end{aligned}$$

と変形することができる。以上より，2 項定理が示せた。

第2章　確率変数と確率分布

問1　2項分布 $Bin(n,p)$ の確率関数を $p_k = \mathrm{P}(X=k),\ k=0,\dots,n,$ とする。

(1) p_k と p_{k-1} の確率関数を書き下していくと

$$p_k = \frac{n+1-k}{k}\frac{p}{1-p}p_{k-1}$$

と表されることがわかる。

p_k のモードを求める。

$$p_k \geq p_{k-1} \Longleftrightarrow k \leq (n+1)p$$
$$p_{k+1} \leq p_k \Longleftrightarrow k+1 \geq (n+1)p$$

であることに注意すると，k がモードになるためには，$(n+1)p-1 \leq k \leq (n+1)p$ を満たす整数である必要がある。従って，$(n+1)p$ が整数のときには，モードは $k=(n+1)p-1/2$ であり，$(n+1)p$ が整数でないときには，モードは $[(n+1)p]$ である。ここで $[x]$ は x の整数部分を表す。

(2) $p=1/2$ のとき，$n=10, k=9$ のときの確率と $n=20, k=18$ のときの確率を比較すると

$$\binom{10}{9}\left(\frac{1}{2}\right)^{10} > \binom{20}{18}\left(\frac{1}{2}\right)^{20} \Longleftrightarrow 2^{10} > 19$$

となるので，$n=10, k=9$ のときの確率の方が大きい。

```
> dbinom(9, 10, 0.5)
[1] 0.009765625
> dbinom(18, 20, 0.5)
[1] 0.0001811981
```

問2　ポアソン分布 $Po(\lambda)$ の確率関数を $p_k = \mathrm{P}(X=k),\ k=0,1,2,\dots,$ とする。

(1) 確率関数 p_k と p_{k-1} を書き下すことによって

$$p_k = \frac{\lambda}{k}p_{k-1}$$

と表されることがわかる。

p_k のモードを求める。

$$p_k \geq p_{k-1} \Longleftrightarrow k \leq \lambda$$
$$p_{k+1} \leq p_k \Longleftrightarrow k \geq \lambda - 1$$

であることに注意すると，k がモードになるためには，$\lambda-1 \leq k \leq \lambda$ を満たす整数である必要がある。従って，λ が整数のときには，モードは $k=\lambda-1/2$ であり，λ が整数でないときには，モードは $[\lambda]$ である。

(2) 少なくとも1回以上起こる確率は

$$\mathrm{P}(X \geq 1) = 1 - \mathrm{P}(X = 0) = 1 - e^{-\lambda}$$

であるから，$\mathrm{P}(A \geq 1) = 1 - e^{-\lambda} \geq 0.9$ を解くと，$\lambda \geq \log(10)$ となる。

問3 3個の正しいコインを同時に投げ，3枚とも表か3枚とも裏がでるまで投げ続ける実験を行い，投げ続けた回数を $X+1$ で表すことにする。

3枚とも表がでる確率は $1/8$, 3枚とも裏がでる確率も $1/8$ なので，3枚とも表か3枚とも裏がでる確率は $1/8 + 1/8 = 1/4$ となる。従って X は，$p = 1/4$ の幾何分布 $Geo(1/4)$ に従う。確率関数は $p(x) = (1/4)(3/4)^x$ である。

$\mathrm{P}(X \leq k)$ の確率を求めると

$$\mathrm{P}(X \leq k) = \frac{1}{4} \cdot \frac{1 - (3/4)^{k+1}}{1 - 3/4} = 1 - \left(\frac{3}{4}\right)^{k+1}$$

となる。従って，$\mathrm{P}(X \leq k) \geq 0.9$ を解くと，$(k+1)\log(3/4) \leq -\log(10)$，すなわち

$$k \geq \frac{\log(10)}{\log(4/3)} - 1$$

となる。

問4 U は一様分布 $U(0,1)$ に従う確率変数とし，自然数 n に対して $X = [nU]$ とおくと，

$$X = [nU] = k \Longleftrightarrow k \leq nU < k+1 \Longleftrightarrow \frac{k}{n} \leq U < \frac{k+1}{n}$$

となるので，

$$p(k) = \mathrm{P}(X = k) = \mathrm{P}\Big(\frac{k}{n} \leq U < \frac{k+1}{n}\Big) = \frac{1}{n},\ k = 0, \ldots, n-1$$

となる。これは $\{0, 1, 2, \ldots, n-1\}$ 上の離散一様分布である。

問5 Y は指数分布 $Ex(\lambda)$ に従う確率変数とし，$X = [Y]$ とおくと，

$$X = [Y] = k \Longleftrightarrow k \leq Y < k+1$$

より，

$$\begin{aligned} p(k) =& \mathrm{P}(X = k) = \int_k^{k+1} \lambda e^{-\lambda y} dy = [-e^{-\lambda y}]_k^{k+1} \\ =& e^{-\lambda k} - e^{-\lambda(k+1)} = (1 - e^{-\lambda})(e^{-\lambda})^k \end{aligned}$$

となる。これは $p = 1 - e^{-\lambda}$ の幾何分布 $Geo(1 - e^{-\lambda})$ に従う。

問6 X が正規分布 $\mathcal{N}(\mu, \sigma^2)$ に従うとき，$Z = (X - \mu)/\sigma \sim \mathcal{N}(0,1)$ に従うことに注意する。

(1) $\mathrm{P}(\mu - c \leq X \leq \mu + c) = 0.95$ となる c の値を求める。

$$\mathrm{P}(\mu - c \leq X \leq \mu + c) = \mathrm{P}\Big(-\frac{c}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{c}{\sigma}\Big) = 1 - 2\mathrm{P}(Z > c/\sigma)$$

より，$1 - 2\mathrm{P}(Z > c/\sigma) = 0.95$，すなわち $\mathrm{P}(Z > c/\sigma) = 0.025$ と書ける。よって $c/\sigma = z_{0.025}$ より，$c = \sigma z_{0.025} = 1.96\sigma$ となる。

(2) $\mathrm{P}(\mu - 2\sigma \leq X \leq \mu + \sigma)$ の確率については

$$\begin{aligned} \mathrm{P}(\mu - 2\sigma \leq X \leq \mu + \sigma) =& \mathrm{P}(-2 \leq Z \leq 1) = \Phi(1) - \Phi(-2) \\ =& \Phi(1) - \{1 - \Phi(2)\} = \Phi(1) + \Phi(2) - 1 = 0.84 + 0.98 - 1 = 0.82 \end{aligned}$$

となる。また $\mathrm{P}(X > d) = 1 - \Phi((d - \mu)/\sigma) = 0.95$ を解けばよい。

(3) $Y = aX + b$ とおくとき，Y の確率分布は $\mathcal{N}(a\mu + b, a^2\sigma^2)$ になる。

問7 (1) $X \sim \mathcal{N}(0, \sigma^2)$ とし，分布関数と密度関数を $F(x)$, $f(x)$ とする。$Y = X^2$ とおくとき Y の確率密度関数は

$$\mathrm{P}(Y \le y) = \mathrm{P}(X^2 \le y) = \mathrm{P}(-\sqrt{y} \le X \le \sqrt{y}) = F(\sqrt{y}) - F(-\sqrt{y})$$

より，

$$f_Y(y) = \frac{d}{dy}\mathrm{P}(Y \le y) = \frac{1}{2\sqrt{y}}f(\sqrt{y}) + \frac{1}{2\sqrt{y}}f(-\sqrt{y})$$

と書ける。$f(x) = f(-x)$ であることに注意し正規分布の確率密度を代入すると

$$f_Y(y) = \frac{1}{\sqrt{y}}f(\sqrt{y}) = \frac{1}{\sqrt{\pi}}\Big(\frac{1}{2\sigma^2}\Big)^{1/2} y^{1/2-1}\exp\Big\{ -\frac{y}{2\sigma^2}\Big\}$$

となる。$\Gamma(1/2) = \sqrt{\pi}$ より，$Y \sim Ga(1/2, 2\sigma^2)$ に従うことがわかる。$Ga(1/2, 2)$ がカイ2乗分布であるから，$Y/\sigma^2 \sim \chi_1^2$ となる。

次に，$Z = |X|$ の確率分布を求めてみよう。

$$\mathrm{P}(Z \le z) = \mathrm{P}(|X| \le z) = \mathrm{P}(-z \le X \le z) = F(z) - F(-z)$$

より，

$$f_Z(z) = \frac{d}{dz}\mathrm{P}(Z \le z) = f(z) + f(-z)$$

と書ける。$f(z) = f(-z)$ であることに注意し正規分布の確率密度を代入すると，Z の確率密度関数 $f_Z(z)$ は

$$f_Z(z) = 2f(z) = \frac{2}{\Gamma(1/2)}\Big(\frac{1}{2\sigma^2}\Big)^{1/2}\exp\Big\{ -\frac{z^2}{2\sigma^2}\Big\}$$

となる。

(2) $U \sim U(-1, 1)$ のとき，U に確率密度は $f(u) = 2^{-1}I(-1 < u < 1)$ であるから，(1) と同様にして $Y = U^2$ の確率密度は

$$f_Y(y) = \frac{1}{\sqrt{y}}f(\sqrt{y}) = \frac{1}{2\sqrt{y}}I(0 < y < 1)$$

となる。

(3) $\Theta \sim U(-\pi/2, \pi/2)$ のとき，U の確率密度は $f(\theta) = \pi^{-1}I(-\pi/2 < \theta < \pi/2)$ である。$Y = \tan\Theta$ とおくと，$\theta = \tan^{-1}(y)$ より $d\theta = (1 + y^2)^{-1}dy$ より，Y の確率密度は

$$f_Y(y) = \frac{1}{\pi}\frac{1}{1+y^2}I(-\pi/2 < \tan^{-1}(y) < \pi/2) = \frac{1}{\pi}\frac{1}{1+y^2}I(-\infty < y < \infty)$$

となり，コーシー分布に従う。

問8 一様乱数を U とする。

(1) X の確率密度関数が $f(x) = \alpha x^{-\alpha-1}$, $x \ge 1$; $f(x) = 0$, $x < 1$, で与えられるから，分布関数は

$$F(x) = \int_1^x \frac{\alpha}{y^{\alpha+1}}dy = \Big[-\frac{1}{y^\alpha}\Big]_1^x = 1 - \frac{1}{x^\alpha}$$

となる。$1 - X^{-\alpha} = U$ を解くと，$X = (1 - U)^{-1/\alpha}$ となる。

(2) X の確率密度関数が $|\alpha| \leq 1$ に対して $f(x) = (1-\alpha x)/2$, $|x| \leq 1$; $f(x) = 0$, $|x| > 1$, で与えられるから，分布関数は

$$F(x) = \int_{-1}^{x} \frac{1-\alpha y}{2} dy = \frac{1}{2}\Big[y - \frac{\alpha}{2}y^2\Big]_{-1}^{x} = \frac{1}{2}\Big(x - \frac{\alpha}{2}x^2 + 1 + \frac{\alpha}{2}\Big)$$

となる。$F(X) = U$ を解くと

$$X = \frac{1}{\alpha}\Big(1 \pm \sqrt{(1+\alpha)^2 - 4\alpha U}\Big)$$

となる。ここで，$-1 \leq X \leq 1$ を満たすことから，解は

$$X = \frac{1}{\alpha}\Big(1 - \sqrt{(1+\alpha)^2 - 4\alpha U}\Big)$$

となる。

問 9 $f(x) = e^{-x}(1+e^{-x})^{-2}$, $-\infty < x < \infty$, に従う確率変数を X とおく。

(1) 確率密度関数になるためには，$f(x) \geq 0$ であるから，$\int_{-\infty}^{\infty} f(x)dx = 1$ を示せばよい。これは,

$$\int_{-\infty}^{\infty} \frac{e^{-x}}{(1+e^{-x})^2} dx = \Big[\frac{1}{1+e^{-x}}\Big]_{-\infty}^{\infty} = 1$$

となることからわかる。

分布関数は，(1) の計算から

$$F(x) = \int_{-\infty}^{x} \frac{e^{-y}}{(1+e^{-y})^2} dy = \Big[\frac{1}{1+e^{-y}}\Big]_{-\infty}^{x} = \frac{1}{1+e^{-x}}$$

で与えられる。$F(X) = U$ を解くと,

$$X = \log\Big(\frac{U}{1-U}\Big)$$

となる。

(2) $Y = |X|$ とおくとき，Y の確率密度は

$$f_Y(y) = 2f(y) = \frac{2e^{-y}}{(1+e^{-y})^2}, \quad y > 0$$

となる。分布関数は

$$F_Y(y) = 2\Big[\frac{1}{1+e^{-t}}\Big]_0^{y} = \frac{2}{1+e^{-y}} - 1 = \frac{e^y - 1}{e^y + 1}$$

と書ける。

問 10 $f(x) = \alpha(1+x)^{-\alpha-1}$, $x > 0$, $\alpha > 0$, が確率密度関数になることは

$$\int_0^{\infty} f(t)dt = \int_0^{\infty} \frac{\alpha}{(1+t)^{\alpha+1}} dt = \Big[-\frac{1}{(1+t)^{\alpha}}\Big]_0^{\infty} = 1$$

となることからわかる。同様な計算から分布関数は

$$F(x) = \int_0^{x} f(x)dx = \Big[-\frac{1}{(1+t)^{\alpha}}\Big]_0^{x} = 1 - \frac{1}{(1+x)^{\alpha}}$$

となる。これよりハザード関数は

$$h(x) = \frac{f(x)}{1-F(x)} = \frac{\alpha}{1+x}$$

となる。死亡率を扱う問題では x の増加とともに死亡リスクが減衰していくことは考えにくいので，生存分析を扱う文脈では巾のオーダーの確率分布は不適切であることがわかる。指数分布のときがハザードが定数になるので，それより速いオーダーで減衰するような確率分布が相応しい。

平均については，$\alpha > 1$ とし，部分積分を用いると

$$
\begin{aligned}
\mathrm{E}[X] =& \int_0^\infty x \frac{\alpha}{(1+x)^{\alpha+1}} dx = \Big[- \frac{x}{(1+x)^\alpha} \Big]_0^\infty + \int_0^\infty \frac{1}{(1+x)^\alpha} dx \\
=& 0 + \Big[- \frac{1}{\alpha-1} \frac{1}{(1+x)^{\alpha-1}} \Big]_0^\infty = \frac{1}{\alpha-1}
\end{aligned}
$$

となる。また分散については，$\alpha > 2$ とし，部分積分を用いると

$$
\begin{aligned}
\mathrm{E}[X^2] =& \int_0^\infty x^2 \frac{\alpha}{(1+x)^{\alpha+1}} dx = \Big[- \frac{x^2}{(1+x)^\alpha} \Big]_0^\infty + 2 \int_0^\infty \frac{x}{(1+x)^\alpha} dx \\
=& 0 + 2 \Big[- \frac{1}{\alpha-1} \frac{x}{(1+x)^{\alpha-1}} \Big]_0^\infty + \frac{2}{\alpha-1} \int_0^\infty \frac{1}{(1+x)^{\alpha-1}} dx \\
=& 0 + 0 + \frac{2}{\alpha-1} \Big[- \frac{1}{\alpha-2} \frac{1}{(1+x)^{\alpha-2}} \Big]_0^\infty = \frac{2}{(\alpha-1)(\alpha-2)}
\end{aligned}
$$

と書けるので，

$$
\mathrm{Var}(X) = \frac{2}{(\alpha-1)(\alpha-2)} - \frac{1}{(\alpha-1)^2} = \frac{\alpha}{(\alpha-1)^2(\alpha-2)}
$$

となる。

問 11　$f(x) = 2^{-1} e^{-|x|}$, $-\infty < x < \infty$, が確率密度関数になることは，対称性から

$$
\int_{-\infty}^\infty f(x) dx = 2 \int_0^\infty f(x) dx = \int_0^\infty e^{-x} dx = 1
$$

となることからわかる。分布の対称性から平均は $\mathrm{E}[X] = 0$ である。またガンマ関数を用いると

$$
\mathrm{E}[X^2] = \int_0^\infty x^2 e^{-x} dx = \Gamma(3) = 2! = 2
$$

となるので，$\mathrm{Var}(X) = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = 2$ となる。

分布関数 $F(x)$ は

$$
F(x) = \frac{1}{2} \int_{-\infty}^x e^{-|u|} du
$$

であり，$x \le 0$ と $x > 0$ の場合に分けて計算する。$x \le 0$ の場合は

$$
F(x) = \frac{1}{2} \int_{-\infty}^x e^u du = \frac{1}{2} \Big[e^u \Big]_{-\infty}^x = \frac{1}{2} e^x
$$

となり，$x > 0$ の場合は

$$
\begin{aligned}
F(x) =& \frac{1}{2} \int_{-\infty}^0 e^u du + \frac{1}{2} \int_0^x e^{-u} du = \frac{1}{2} + \frac{1}{2} \Big[- e^{-u} \Big]_0^x \\
=& \frac{1}{2} + \frac{1}{2} \Big(1 - e^{-x} \Big) = 1 - \frac{1}{2} e^{-x}
\end{aligned}
$$

となる。したがって

$$
F(x) = \frac{1}{2} e^x I(x \le 0) + \Big(1 - \frac{1}{2} e^{-x} \Big) I(x > 0)
$$

と表される。

問 12 確率変数 X が幾何分布 $Geo(\theta/n)$ に従うとき

$$\mathrm{P}(X \leq n) = \sum_{k=0}^{n} \frac{\theta}{n}\Big(1-\frac{\theta}{n}\Big)^k = \frac{\theta}{n}\frac{1-(1-\theta/n)^{n+1}}{\theta/n} = 1-\Big(1-\frac{\theta}{n}\Big)^{n+1}$$

より

$$\lim_{n\to\infty} \mathrm{P}(X \leq n) = 1 - \lim_{n\to\infty}\Big(1-\frac{\theta}{n}\Big)^{n+1} = 1-e^{-\theta}$$

となる。

問 13　ポアソン分布が確率分布になることを示す問題である。通常は，e^λ をテーラー展開することにより示される。ここでは等式 $\sum_{k=0}^{\infty}\lambda^k/k! = e^\lambda$ を次のようにして解くことを考える。

(1) $\sum_{k=0}^{\infty}\lambda^k/k! = g(\lambda)$ とおき，両辺を λ で微分すると

$$g'(\lambda) = \sum_{k=1}^{\infty}\frac{k}{k!}\lambda^{k-1} = \sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = g(\lambda)$$

となるので，微分方程式 $g'(\lambda) = g(\lambda)$ が導かれる。また $g(0) = 0^0/0! = 1$ である。

(2) 微分方程式

$$\frac{g'(\lambda)}{g(\lambda)} = 1$$

を解くと，$\log g(\lambda) = \lambda + C'$ となり $g(\lambda) = Ce^\lambda$ と書ける。ここで C', C は適当な定数である。$g(0) = 1$ より $C = 1$ となり，$g(\lambda) = e^\lambda$ が得られる。

問 14　負の 2 項分布が確率分布になることを示す問題である。通常は,「現代数理統計学の基礎」にあるように，$\sum_{k=0}^{\infty} q^k = 1/(1-q)$ の両辺を q に関して微分を繰り返していくことによって示される。ここでは，問 13 と同様に，等式 $\sum_{k=0}^{\infty} C_r(k)(1-p)^k = 1/p^r$ を微分方程式を用いて示すことを考える。ただし $C_r(k)$ は (2.5) で与えられている。

(1) まず等式 $\sum_{k=0}^{\infty} kC_r(k)(1-p)^{k-1} = (r/p)\sum_{k=0}^{\infty} C_r(k)(1-p)^k$ を示そう。

$$\begin{aligned}\sum_{k=0}^{\infty} kC_r(k)(1-p)^{k-1} =& \sum_{k=0}^{\infty} k\frac{(r-1+k)!}{k!(r-1)!}(1-p)^{k-1} = \sum_{k=1}^{\infty}\frac{(r-1+k)(r-1+k-1)!}{(k-1)!(r-1)!}(1-p)^{k-1}\\ =&(1-p)\sum_{k=0}^{\infty}\frac{(r+k)(r-1+k)!}{k!(r-1)!}(1-p)^{k-1} = (1-p)\sum_{k=0}^{\infty}(r+k)C_r(k)(1-p)^{k-1}\\ =&(1-p)\Big\{r\sum_{k=0}^{\infty}C_r(k)(1-p)^{k-1} + \sum_{k=0}^{\infty}kC_r(k)(1-p)^{k-1}\Big\}\end{aligned}$$

となる。この式は

$$p\sum_{k=0}^{\infty}kC_r(k)(1-p)^{k-1} = r\sum_{k=0}^{\infty}C_r(k)(1-p)^k$$

を示している。

(2) $\sum_{k=0}^{\infty} C_r(k)(1-p)^k = g_r(p)$ とおき，両辺を p で微分すると

$$g_r'(p) = -\sum_{k=0}^{\infty}kC_r(k)(1-p)^{k-1}$$

となり，(1) の結果を用いると微分方程式

$$g_r'(p) = -\frac{r}{p}g_r(p)$$

が導かれる。

(3) 微分方程式

$$\frac{g_r'(p)}{g_r(p)} = -\frac{r}{p}$$

を解くと，$\log g_r(p) = -r\log p + C'$ となり，$g_r(p) = C/p^r$ と書けることがわかる。$g_r(1) = C_r(0) = 1$ より $C = 1$ となるので，

$$g_r(p) = \frac{1}{p^r}$$

が得られる。

問 15　ガンマ分布 $Ga(\alpha, \beta)$ の確率密度は

$$f(x) = \frac{1}{\Gamma(\alpha)}\frac{1}{\beta^\alpha}x^{\alpha-1}e^{-x/\beta}$$

である。

(1) $\alpha > 1$ のとき，分布のモードを求める。$\log f(x) = -\log\Gamma(\alpha) - \alpha\log\beta + (\alpha-1)\log x - x/\beta$ を x で微分すると

$$\frac{d}{dx}\log f(x) = \frac{\alpha-1}{x} - \frac{1}{\beta}$$

となるので，モードは $x = (\alpha-1)\beta$ となる。

(2) n が奇数, すなわち $n = 2m+1$ のとき,

$$\Gamma\Big(m+\frac{1}{2}\Big) = \frac{\sqrt{\pi}(2m-1)!}{2^{2m-1}(m-1)!}$$

のような表現式が成り立つことを示す。ただし m は 1 以上の整数とする。

$$\Gamma\Big(m+\frac{1}{2}\Big) = \Big(m-\frac{1}{2}\Big)\Big(m-\frac{3}{2}\Big)\times\cdots\times\frac{1}{2}\Gamma\Big(\frac{1}{2}\Big)$$

と書ける。ここで右辺の $m-1/2$ から $1/2$ までの項の個数は m であり，$\Gamma(1/2) = \pi$ であるから

$$\begin{aligned}\Gamma\Big(m+\frac{1}{2}\Big) =&(2m-1)(2m-3)\times\cdots\times(2m-(2m-1))\frac{1}{2^m}\sqrt{\pi}\\ =&\frac{(2m-1)!}{2^{m-1}(m-1)!}\frac{1}{2^m}\sqrt{\pi} = \frac{(2m-1)!}{2^{2m-1}(m-1)!}\sqrt{\pi}\end{aligned}$$

となる。

問 16　$f(x) = \pi^{-1}(1+x^2)^{-1}$, $-\infty < x < \infty$, とする。

(1) 分布関数は,

$$F(x) = \frac{1}{\pi}\int_{-\infty}^{x}\frac{1}{1+t^2}dt = \frac{1}{\pi}\Big[\tan^{-1}(t)\Big]_{-\infty}^{x} = \frac{1}{\pi}\Big\{\tan^{-1}(x) + \frac{\pi}{2}\Big\} = \frac{1}{\pi}\tan^{-1}(x) + \frac{1}{2}$$

となる。$\lim_{x\to\infty}\tan^{-1}(x) = \pi/2$ となることから，$\lim_{x\to\infty}F(x) = 1$ となるので確率分布になることがわかる。

(2) 平均が存在しないことは

$$\begin{aligned}\mathrm{E}[|X|] =&\frac{2}{\pi}\int_0^\infty\frac{x}{1+x^2}dx \geq \frac{2}{\pi}\int_1^\infty\frac{x}{1+x^2}dx\\ =&\frac{2}{\pi}\int_1^\infty\frac{1}{x^{-1}+x}dx \geq \frac{2}{\pi}\int_1^\infty\frac{1}{1+x}dx\\ =&\frac{2}{\pi}[\log(1+x)]_1^\infty = \infty\end{aligned}$$

となることからわかる。

問 17 (2.9) で与えられる超幾何分布を考える。$K \leq M \leq N$, $K \leq N-M$ であることに注意する。

(1) $(a+b)^M(a+b)^{N-M}=(a+b)^N$ より

$$\sum_{x=0}^{M}\binom{M}{x}a^x b^{M-x}\sum_{\ell=0}^{N-M}\binom{N-M}{\ell}a^{\ell}b^{N-M-\ell}=\sum_{K=0}^{N}\binom{N}{K}a^K b^{N-K}$$

と書ける。両辺の $a^K b^{N-K}$ の係数を比較することを考える。左辺は

$$\sum_{x=0}^{M}\sum_{\ell=0}^{N-M}\binom{M}{x}\binom{N-M}{\ell}a^{x+\ell}b^{N-x-\ell}$$

と書けるので，$a^K b^{N-K}$ の係数は $x+\ell=K$, すなわち $\ell=K-x$ の制約を入れればよいことがわかる。$\ell=K-x\geq 0$ より $x\leq K$ であり，$K\leq M$ であることに注意すると

$$\sum_{x=0}^{K}\binom{M}{x}\binom{N-M}{K-x}=\binom{N}{K}$$

が得られる。このことは確率関数になることを示している。

(2) 平均については

$$\begin{aligned}\sum_{x=0}^{K}x\binom{M}{x}\binom{N-M}{K-x}&=\sum_{x=1}^{K}\frac{M!}{(x-1)!(M-x)!}\binom{N-M}{K-x}\\&=M\sum_{y=0}^{K-1}\binom{M-1}{y}\binom{(N-1)-(M-1)}{K-1-y}\end{aligned}$$

となる。ただし，$y=x-1$ とおいたことに注意する。一方,

$$\binom{N}{K}=\frac{N}{K}\frac{(N-1)!}{(K-1)!(N-1-(K-1))!}=\frac{N}{K}\binom{N-1}{K-1}$$

と書けるので

$$\mathrm{E}[X]=\sum_{x=0}^{K}xp(x)=\frac{M}{N/K}\sum_{y=0}^{K-1}\binom{M-1}{y}\binom{N-M}{K-1-y}\Big/\binom{N-1}{K-1}=\frac{MK}{N}$$

となる。

分散については，$\mathrm{E}[X(X-1)]$ を計算する。上の計算と同様にして

$$\sum_{x=0}^{K}x(x-1)\binom{M}{x}\binom{N-M}{K-x}=M(M-1)\sum_{y=0}^{K-2}\binom{M-2}{y}\binom{(N-2)-(M-2)}{K-2-y}$$

となる。ただし，$y=x-2$ とおいたことに注意する。一方,

$$\binom{N}{K}=\frac{N(N-1)}{K(K-1)}\frac{(N-1)!}{(K-2)!(N-2-(K-2))!}=\frac{N(N-1)}{K(K-1)}\binom{N-2}{K-2}$$

と書けるので

$$\begin{aligned}\mathrm{E}[X(X-1)]&=\sum_{x=0}^{K}x(x-1)p(x)\\&=\frac{M(M-1)}{N(N-1)/K(K-1)}\sum_{y=0}^{K-2}\binom{M-2}{y}\binom{(N-2)-(M-2)}{(K-2)-y}\Big/\binom{N-2}{K-2}\\&=\frac{M(M-1)K(K-1)}{N(N-1)}\end{aligned}$$

となる。従って

$$\begin{aligned}\mathrm{Var}(X) =&\mathrm{E}[X(X-1)]+\mathrm{E}[X]-(\mathrm{E}[X])^2=\frac{MK}{N}\left\{\frac{(M-1)(K-1)}{N-1}+1-\frac{MK}{N}\right\}\\ =&\frac{MK}{N}\frac{(N-M)(N-K)}{N(N-1)}\end{aligned}$$

となる。

$p=M/N$ とおくと

$$\mathrm{E}[X]=Kp,\quad \mathrm{Var}(X)=\frac{N-K}{N-1}Kp(1-p)$$

と表される。

(3) K を固定して $N\to\infty$, $M/N\to p$ とするとき，超幾何分布の確率関数を書き下すと

$$\begin{aligned}p(x)=&\binom{M}{x}\binom{N-M}{K-x}\Big/\binom{N}{K}\\ =&\frac{M!}{x!(M-x)!}\frac{(N-M)!}{(K-x)!(N-M-K+x)!}\frac{K!(N-K)!}{N!}\\ =&\binom{K}{x}\frac{M(M-1)\cdots(M-x+1)\times(N-M)\cdots(N-M-K+x+1)}{N(N-1)\cdots(N-K+1)}\end{aligned}$$

と書ける。$M(M-1)\cdots(M-x+1)$ は x 個，$(N-M)\cdots(N-M-K+x+1)$ は $K-x$ 個，$N(N-1)\cdots(N-K+1)$ は K 個あり，$M/N\to p$ より

$$\frac{M(M-1)\cdots(M-x+1)\times(N-M)\cdots(N-M-K+x+1)}{N(N-1)\cdots(N-K+1)}\to p^x(1-p)^{K-x}$$

に収束することがわかる。従って，超幾何分布は 2 項分布に収束する。

第3章　2変数の同時確率分布

問1　正四面体の面に 1, 2, 3, 4 の番号が書かれていて，転がしたとき各面は等確率で現れるとする。正四面体を 2 回転がして，出た目の和を X，差を Y で表すと，X は 2 から 8 の値をとり，Y は 0 から 3 の値をとる。

X 和

1回目 \ 2回目	1	2	3	4
1	2	3	4	5
2	3	4	5	6
3	4	5	6	7
4	5	6	7	8

Y 差

1回目 \ 2回目	1	2	3	4
1	0	1	2	3
2	1	0	1	2
3	2	1	0	1
4	3	2	1	0

(X, Y) の度数

$Y \backslash X$	2	3	4	5	6	7	8	合計
0	1	0	1	0	1	0	1	4
1	0	2	0	2	0	2	0	6
2	0	0	2	0	2	0	0	4
3	0	0	0	2	0	0	0	2
合計	1	2	3	4	3	4	1	16

(1) (X, Y) の同時確率関数 $f(x, y)$ について，上の表から $f(4,0) = 1/16,\ f(5,3) = 2/16 = 1/8$ である。

(2) X と Y の周辺確率関数について，上の表の合計欄から $f_X(6) = 3/16,\ f_Y(2) = 4/16 = 1/4$ である。

(3) X を与えたときの Y の条件付き確率関数について

$$f_{Y|X}(2|4) = \frac{f_{X,Y}(4,2)}{f_X(4)} = \frac{2/16}{3/16} = \frac{2}{3}$$

となる。また Y を与えたときの X の条件付き確率関数について

$$f_{X|Y}(5|1) = \frac{f_{X,Y}(5,1)}{f_Y(1)} = \frac{2/16}{6/16} = \frac{1}{3}$$

となる。

問2　多項分布 (3.2) についての設問である。m 個の面からなる多面体を n 回転がして，i 番目の面が X_i 回出るとする。また i 番目の面が出る確率を p_i とする。

(1) X_i の周辺分布は，i 番目の面が出るか出ないかに着目することにより，$X_i \sim Bin(n, p_i)$ となる。また，$X_i + X_j$ の周辺分布は，i 番目もしくは j 番目が出るか出ないかに着目することにより，$X_i + X_j \sim Bin(n, p_i + p_j)$ となる。

(2) (X_1, X_2) の同時確率と X_1 の周辺確率は

$$\begin{aligned} \mathrm{P}(X_1 = x_1, X_2 = x_2) =& \frac{n!}{x_1!x_2!(n - x_1 - x_2)!} p_1^{x_1} p_2^{x_2} (1 - p_1 - p_2)^{n - x_1 - x_2} \\ \mathrm{P}(X_1 = x_1) =& \frac{n!}{x_1!(n - x_1)!} p_1^{x_1} (1 - p_1)^{n - x_1} \end{aligned}$$

であることから，$X_1 = x_1$ を与えたときの $X_2 = x_2$ の条件付き確率は

$$\begin{aligned}
\mathrm{P}(X_2 = x_2 | X_1 = x_1) =& \frac{\mathrm{P}(X_1 = x_1, X_2 = x_2)}{\mathrm{P}(X_1 = x_1)} \\
=& \frac{x_1!(n-x_1)!}{x_1!x_2!(n-x_1-x_2)!} \frac{p_1^{x_1} p_2^{x_2} (1-p_1-p_2)^{n-x_1-x_2}}{p_1^{x_1}(1-p_1)^{n-x_1}} \\
=& \frac{(n-x_1)!}{x_2!(n-x_1-x_2)!} \Big(\frac{p_2}{1-p_1}\Big)^{x_2} \Big(1 - \frac{p_2}{1-p_1}\Big)^{n-x_1-x_2}
\end{aligned}$$

となるので，

$$X_2 | X_1 = x_1 \sim Bin\Big(n - x_1, \frac{p_2}{1-p_1}\Big)$$

に従う。

(3) $X_1 + X_2$ を与えたときの X_2 の条件付き確率を求めることを考える。$X_1 + X_2 = z$ が与えられたときの $X_2 = x_2$ となる確率を考えるとき，$X_1 = z - x_2$ として自動的に X_1 の値が決まることになるので，$x_1 = z - x_2$ とおくと，$(X_1 + X_2, X_2)$ の同時確率は $\mathrm{P}(X_1 + X_2 = x_1 + x_2, X_2 = x_2) = \mathrm{P}(X_1 = x_1, X_2 = x_2)$ となる。このことから，条件付き確率は

$$\begin{aligned}
\mathrm{P}(X_2 = x_2 | X_1 + X_2 = x_1 + x_2) =& \frac{(x_1+x_2)!(n-x_1-x_2)!}{x_1!x_2!(n-x_1-x_2)!} \frac{p_1^{x_1} p_2^{x_2} (1-p_1-p_2)^{n-x_1-x_2}}{(p_1+p_2)^{x_1+x_2}(1-p_1-p_2)^{n-x_1-x_2}} \\
=& \frac{(x_1+x_2)!}{x_1!x_2!} \Big(\frac{p_1}{p_1+p_2}\Big)^{x_1} \Big(\frac{p_2}{p_1+p_2}\Big)^{x_2}
\end{aligned}$$

と書けるので，

$$X_2 | X_1 + X_2 = z \sim Bin\Big(z, \frac{p_2}{p_1+p_2}\Big)$$

に従う。

問 3　X, Y を独立な確率変数で，X がポアソン分布 $Po(\lambda_1)$，Y が $Po(\lambda_2)$ に従うとする。

(1) $Z = X + Y$ とおくとき，Z の分布は

$$\begin{aligned}
\mathrm{P}(X + Y = n) =& \sum_{x=0}^{n} \mathrm{P}(X = x)\mathrm{P}(Y = n - x) = \sum_{x=0}^{n} \frac{\lambda_1^x}{x!} e^{-\lambda_1} \frac{\lambda_2^{n-x}}{(n-x)!} e^{-\lambda_2} \\
=& \frac{(\lambda_1+\lambda_2)^n}{n!} e^{-(\lambda_1+\lambda_2)} \sum_{x=0}^{n} \frac{n!}{x!(n-x)!} \Big(\frac{\lambda_1}{\lambda_1+\lambda_2}\Big)^x \Big(\frac{\lambda_2}{\lambda_1+\lambda_2}\Big)^{n-x}
\end{aligned}$$

より，

$$X + Y \sim Po(\lambda_1 + \lambda_2)$$

に従うことがわかる。

(2) $X + Y = n$ に固定した条件のもとで，X の条件付き分布を求める。$\mathrm{P}(X + Y = n, X = x) = \mathrm{P}(X + x, Y = n - x)$ と書けるので，

$$\begin{aligned}
\mathrm{P}(X = x | X + Y = n) =& \frac{\mathrm{P}(X = x, Y = n - x)}{\mathrm{P}(X + Y = n)} \\
=& \frac{(\lambda_1^x / x!) e^{-\lambda_1} (\lambda_2^{n-x}/(n-x)!) e^{-\lambda_2}}{((\lambda_1+\lambda_2)^n / n!) e^{-(\lambda_1+\lambda_2)}} \\
=& \frac{n!}{x!(n-x)!} \Big(\frac{\lambda_1}{\lambda_1+\lambda_2}\Big)^x \Big(\frac{\lambda_2}{\lambda_1+\lambda_2}\Big)^{n-x}
\end{aligned}$$

となる。従って，条件付き確率は

$$X | X + Y = n \sim Bin\Big(n, \frac{\lambda_1}{\lambda_1+\lambda_2}\Big)$$

に従う。

問 4 $(0,\infty)\times(0,\infty)$ 上の確率変数 (X,Y) について，同時確率関数が $f(x,y)=xe^{-x(y+1)}$, $x>0$, $y>0$, で与えられるとする。

X 及び Y の周辺確率密度関数を求める。

$$f_X(x)=\int_0^\infty xe^{-xy}e^{-x}dy=\Big[-e^{-xy}\Big]_{y=0}^\infty e^{-x}=e^{-x}$$
$$f_Y(y)=\int_0^\infty xe^{-x(y+1)}dx=\frac{1}{(y+1)^2}\int_0^\infty \frac{1}{\Gamma(2)}(y+1)^2x^{2-1}e^{-(y+1)x}dx=\frac{1}{(y+1)^2}$$

となる。ただし，2 番目の式の積分の中身は $Ga(2,1/(y+1))$ の確率密度である。

条件付き確率 $f_{X|Y}(x|y)$, $f_{Y|X}(y|x)$ は

$$f_{X|Y}(x|y)=\frac{xe^{-x(y+1)}}{1/(y+1)^2}=\frac{1}{\Gamma(2)}(y+1)^2x^{2-1}e^{-(y+1)x}$$
$$f_{Y|X}(y|x)=\frac{xe^{-x(y+1)}}{e^{-x}}=xe^{-xy}$$

と書けるので，

$$X|Y=y\sim Ga\Big(2,\frac{1}{y+1}\Big),\quad Y|X=x\sim Ex(x)$$

に従う。

確率 $\mathrm{P}(XY>1)$ を計算すると

$$\mathrm{P}(XY>1)=\int_0^\infty\int_{1/x}^\infty xe^{-x(y+1)}dydx=\int_0^\infty xe^{-x}\Big[-\frac{1}{x}e^{-xy}\Big]_{y=1/x}^\infty dx$$
$$=\int_0^\infty e^{-x}e^{-1}dx=\frac{1}{e}$$

となる。

問 5 $(0,\infty)\times(0,\infty)$ 上の確率変数 (X,Y) について，同時確率関数は $f(x,y)=Ce^{-x-y}$ であるが，$D=\{(x,y)|0<x<y<\infty\}$ の上で正の確率密度をもつとする。

(1) 基準化定数 C の値は

$$C\int_0^\infty\int_0^y e^{-x-y}dxdy=C\int_0^\infty e^{-y}\Big[-e^{-x}\Big]_0^y dy=C\int_0^\infty e^{-y}(1-e^{-y})dy$$
$$=C\int_0^\infty e^{-y}dy-\frac{C}{2}\int_0^\infty 2e^{-2y}dy=C-\frac{C}{2}=\frac{C}{2}$$

より，$C=2$ となる。

X 及び Y の周辺確率密度関数は

$$f_X(x)=\int_x^\infty 2e^{-x-y}dy=2e^{-x}\Big[-e^{-y}\Big]_x^\infty=2e^{-2x}$$
$$f_Y(y)=\int_0^y 2e^{-x-y}dx=2e^{-y}\Big[-e^{-x}\Big]_0^y=2e^{-y}-2e^{-2y}$$

となる。

(2) 条件付き確率 $f_{X|Y}(x|y)$, $f_{Y|X}(y|x)$ は

$$f_{X|Y}(x|y)=\frac{2e^{-x-y}}{2e^{-y}(1-e^{-y})}=\frac{e^{-x}}{1-e^{-y}},\ 0<x<y$$
$$f_{Y|X}(y|x)=\frac{2e^{-x-y}}{2e^{-x}}=e^{-(y-x)},\ y>x$$

となる。

(3) $\mathrm{P}(Y < X + d) = 0.9$ を満たす d の値を求める。

$$\begin{aligned}\mathrm{P}(Y < X + d) &= \int_0^\infty \int_x^{x+d} 2e^{-x-y} dy dx = 2\int_0^\infty 2e^{-x}\Big[-e^{-y}\Big]_x^{x+d} dx \\ &= \int_0^\infty 2e^{-2x}dx - \int_0^\infty 2e^{-2x}dx e^{-d} = 1 - e^{-d}\end{aligned}$$

となるので，$1 - e^{-d} = 0.9$ を解いて，$d = \log(10)$ となる。

問6 2つの確率変数 X, Y について，Y は区間 $[0, 1]$ 上を一様に分布し，$Y = y$ が与えられたとき X は区間 $[0, y]$ 上を一様に分布するので，

$$f_{X|Y}(x|y) = \frac{1}{y}I(0 < x < y), \quad f_Y(y) = I(0 < y < 1)$$

と書ける。

(1) (X, Y) の同時確率密度関数は

$$f_{X,Y}(x, y) = \frac{1}{y}I(0 < x < y < 1)$$

で与えられる。

X の周辺確率密度関数は

$$f_X(x) = \int_x^1 \frac{1}{y}dy = \Big[\log y\Big]_x^1 = -\log x,\ 0 < x < 1$$

となる。

(2) $Z = -\log X$ の確率分布を求める。$z = -logx$，すなわち $x = e^{-z}$ であるから，$dx = -e^{-z}dz$ となり，変数変換公式より Z の確率密度は

$$f_Z(z) = ze^{-z} = \frac{1}{Ga(2)}z^{2-1}e^{-z},\ z > 0$$

で与えられる。これは $Z \sim Ga(2, 1)$ を示している。

問7 確率変数 (X, Y, Z) の同時確率密度関数がディリクレ分布

$$f(x, y, z) = \frac{\Gamma(a + b + c)}{\Gamma(a)\Gamma(b)\Gamma(c)}x^{a-1}y^{b-1}z^{c-1}, \quad 0 < x, y, z < 1,\ x + y + z = 1,$$

に従うとする。ただし，a, b, c は正の実数である。

(1) 簡単のために，$C = \Gamma(a + b + c)/\{\Gamma(a)\Gamma(b)\Gamma(c)\}$ とおくと，確率密度関数は

$$f(x, y) = Cx^{a-1}y^{b-1}(1 - x - y)^{c-1}$$

と表される。$V = X + Y$，$W = Y$ とおくと，$x = v - w$, $y = w$ より, ヤコビアンは

$$\left|\frac{\partial(x, y)}{\partial(v, w)}\right| = \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} = 1$$

となるので，(V, W) の同時確率密度は,

$$f_{V,W}(v, w) = C(v - w)^{a-1}w^{b-1}(1 - v)^{c-1},\ 0 < w < v$$

と書ける。このとき V の周辺密度は

$$f_V(v) = C(1-v)^{c-1}\int_0^v (v-w)^{a-1}w^{b-1}dw$$

を計算すればよい。ここで，変数変換 $vs = w$, $dw = vds$，を用いると

$$\begin{aligned} f_V(v) =& C(1-v)^{c-1}\int_0^1 (v-vs)^{a-1}(vs)^{b-1}vds = Cv^{a+b-1}(1-v)^{c-1}\int_0^1 (1-s)^{a-1}s^{b-1}ds \\ =& CB(a,b)v^{a+b-1}(1-v)^{c-1} \end{aligned}$$

と書けることがわかる。

$$CB(a,b) = \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)}\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \frac{\Gamma(a+b+c)}{\Gamma(a+b)\Gamma(c)} = \frac{1}{B(a+b,c)}$$

より，$V = X + Y$ は

$$X + Y \sim Beta(a+b, c)$$

に従う。

(2) $T = Y$, $W = X/(1-Y)$ とおくと，$X = W(1-T)$, $Y = T$ であるから，ヤコビアンは

$$\left|\frac{\partial(x,y)}{\partial(w,t)}\right| = \begin{vmatrix} 1-t & -w \\ 0 & 1 \end{vmatrix} = 1-t$$

となり，t, w の積分範囲は $0 < t < 1$, $0 < w < 1$ となることを示すことができる。従って，(T,W) の同時確率密度は

$$\begin{aligned} f_{T,W}(t,w) =& C\{w(1-t)\}^{a-1}t^{b-1}\{1-w(1-t)-t\}^{c-1}(1-t) \\ =& Cw^{a-1}(1-w)^{c-1}t^{b-1}(1-t)^{a+c-1} \\ =& \frac{\Gamma(a+c)}{\Gamma(a)\Gamma(c)}w^{a-1}(1-w)^{c-1}\frac{\Gamma(a+b+c)}{\Gamma(b)\Gamma(a+c)}t^{b-1}(1-t)^{a+c-1} \end{aligned}$$

と表される。この式は，T と W は独立で，

$$W \sim Beta(a,c), \quad T \sim Beta(b, a+c)$$

に従うことを示している。

第4章　期待値と積率母関数

問1　確率変数 X が次の分布に従うとき，積率母関数 $M_X(t)$ を求める。これを用いて $k=1,2,3,4$ に対して $\mathrm{E}[X^k]$ の値と歪度，尖度を計算する。

(1) ポアソン分布 $Po(\lambda)$ については，確率関数が $p(x)=(\lambda^x/x!)e^{-\lambda}$ であるから，積率母関数は

$$M_X(t)=\sum_{x=0}^{\infty}\frac{\lambda^x}{x!}e^{-\lambda}e^{tx}=\sum_{x=0}^{\infty}\frac{(\lambda e^t)^x}{x!}e^{-\lambda e^t}\times e^{\lambda e^t-\lambda}=e^{\lambda e^t-\lambda}$$

となる。

$$\begin{aligned}
M'_X(t)=&\lambda e^t e^{\lambda e^t-\lambda}\\
M''_X(t)=&(\lambda^2e^{2t}+\lambda e^t)e^{\lambda e^t-\lambda}\\
M^{(3)}_X(t)=&(\lambda^3e^{3t}+3\lambda^2e^{2t}+\lambda e^t)e^{\lambda e^t-\lambda}\\
M^{(4)}_X(t)=&(\lambda^4e^{4t}+6\lambda^3e^{3t}+7\lambda^2e^{2t}+\lambda e^t)e^{\lambda e^t-\lambda}
\end{aligned}$$

より，$\mathrm{E}[X]=\lambda$, $\mathrm{E}[X^2]=\lambda^2+\lambda$, $\mathrm{E}[X^3]=\lambda^3+3\lambda^2+\lambda$, $\mathrm{E}[X^4]=\lambda^4+6\lambda^3+7\lambda^2+\lambda$ となる。これより，$\mu=\lambda$, $\sigma^2=\lambda$,

$$\begin{aligned}
\mathrm{E}[(X-\lambda)^3]=&\mathrm{E}[X^3-3X^2\lambda+3X\lambda^2-\lambda^3]=\lambda^3+3\lambda^2+\lambda-3\lambda(\lambda^2+\lambda)+3\lambda^3-\lambda^3=\lambda\\
\mathrm{E}[(X-\lambda)^4]=&\mathrm{E}[X^4-4X^3\lambda+6X^2\lambda^2-4X\lambda^3+\lambda^4]\\
=&\lambda^4+6\lambda^3+7\lambda^2+\lambda-4(\lambda^3+3\lambda^2+\lambda)\lambda+6(\lambda^2+\lambda)\lambda^2-4\lambda^4+\lambda^4\\
=&3\lambda^2+\lambda
\end{aligned}$$

となるので，歪度と尖度は

$$\beta_1=\frac{\lambda}{\lambda^{3/2}}=\frac{1}{\sqrt{\lambda}},\quad \beta_2=\frac{3\lambda^2+\lambda}{\lambda^2}=3+\frac{1}{\lambda}$$

となる。

(2) 確率変数 X が正規分布 $\mathcal{N}(\mu,\sigma^2)$ に従うとき，$Z=(X-\mu)/\sigma$ は標準正規分布 $\mathcal{N}(0,1)$ に従うので，計算間違いをしないためにも Z に基づいて計算していく方が良さそうである。Z の積率母関数は

$$M_Z(t)=\mathrm{E}[e^{tZ}]=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-z^2/2+tz}dz=e^{t^2/2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-(z-t)^2/2}dz=e^{t^2/2}$$

となる。$X=\sigma Z+\mu$ であるから，

$$M_X(t)=\mathrm{E}[e^{tX}]=\mathrm{E}[e^{(t\sigma)Z}]e^{t\mu}=e^{(t\sigma)^2/2+t\mu}=e^{\mu t+\sigma^2t^2/2}$$

となる。

$$\begin{aligned}
M'_Z(t)=&te^{t^2/2}\\
M''_Z(t)=&(t^2+1)e^{t^2/2}\\
M^{(3)}_Z(t)=&(t^3+3t)e^{t^2/2}\\
M^{(4)}_Z(t)=&(t^4+6t^2+3)e^{t^2/2}
\end{aligned}$$

より，$\mathrm{E}[Z] = \mathrm{E}[Z^3] = 0$, $\mathrm{E}[Z^2] = 1$, $\mathrm{E}[Z^4] = 3$ となる。X については

$$
\begin{aligned}
X =& \sigma Z + \mu \\
X^2 =& \sigma^2 Z^2 + 2\sigma\mu Z + \mu^2 \\
X^3 =& \sigma^3 Z^3 + 3\sigma^2\mu Z^2 + 3\sigma\mu^2 Z + \mu^3 \\
X^4 =& \sigma^4 Z^4 + 4\sigma^3\mu Z^3 + 6\sigma^2\mu^2 Z^2 + 4\sigma\mu^3 Z + \mu^4
\end{aligned}
$$

より，$\mathrm{E}[X] = \mu$, $\mathrm{E}[X^2] = \sigma^2 + \mu^2$, $\mathrm{E}[X^3] = (3\sigma^2 + \mu^2)\mu$, $\mathrm{E}[X^4] = 3\sigma^4 + 6\sigma^2\mu^2 + \mu^4$ となる。これより，平均は μ，分散は σ^2 であり，歪度と尖度は

$$\beta_1 = \mathrm{E}[Z^3] = 0, \quad \beta_2 = \mathrm{E}[Z^4] = 3$$

となる。

(3) 確率変数 X が指数分布 $Ex(\lambda)$ に従うとき，$Y = \lambda X$ は $Ex(1)$ に従うので，Y に基づいて計算する。Y の積率母関数は

$$M_Y(t) = \mathrm{E}[e^{tY}] = \int_0^\infty e^{-(1-t)y} dy = \frac{1}{1-t}$$

となる。$X = Y/\lambda$ より

$$M_X(t) = \mathrm{E}[e^{(t/\lambda)Y}] = \frac{1}{1-t/\lambda} = \frac{\lambda}{\lambda - t}$$

と書ける。

$$M_Y'(t) = \frac{1}{(1-t)^2},\ M_Y''(t) = \frac{2}{(1-t)^3},\ M_Y^{(3)}(t) = \frac{6}{(1-t)^4}, \quad M_Y^{(4)}(t) = \frac{24}{(1-t)^5}$$

より，$\mathrm{E}[Y] = 1$, $\mathrm{E}[Y^2] = 2$, $\mathrm{E}[Y^3] = 6$, $\mathrm{E}[Y^4] = 24$ となる。$X = Y/\lambda$ より

$$\mathrm{E}[X] = \frac{1}{\lambda},\ \mathrm{E}[X^2] = \frac{2}{\lambda^2},\ \mathrm{E}[X^3] = \frac{6}{\lambda^3},\ \mathrm{E}[X^4] = \frac{24}{\lambda^4}$$

となることがわかる。これより，平均と分散は $\mu = 1/\lambda$, $\sigma^2 = 1/\lambda^2$ となる。歪度と尖度については

$$
\begin{aligned}
\mathrm{E}[(Y-1)^2] =& \mathrm{E}[Y^2 - 2Y + 1] = 2 - 2 + 1 = 1 \\
\mathrm{E}[(Y-1)^3] =& \mathrm{E}[Y^3 - 3Y^2 + 3Y - 1] = 6 - 6 + 3 - 1 = 2 \\
\mathrm{E}[(Y-1)^4] =& \mathrm{E}[Y^4 - 4Y^3 + 6Y^2 - 4Y + 1] = 24 - 24 + 12 - 4 + 1 = 9
\end{aligned}
$$

となることから

$$\beta_1 = 2, \quad \beta_2 = 9$$

となる。

問 2　多項分布 (3.2) について次の問に答えよ。

(1) $i \neq j$ のとき $\mathrm{Cov}(X_i, X_j) = -np_i p_j$ を示す。始めに

$$
\begin{aligned}
\mathrm{E}[X_1 X_2] =& \sum_{x_1,\ldots,x_m} \frac{n!}{x_1! x_2! x_3! \cdots x_m!} x_1 x_2 p_1^{x_1} p_1^{x_2} p_3^{x_3} \cdots p_m^{x_m} \\
=& n(n-1)p_1 p_2 \sum_{x_1,\ldots,x_m} \frac{(n-2)!}{(x_1-1)!(x_2-1)! x_3! \cdots x_m!} p_1^{x_1-1} p_2^{x_2-1} p_3^{x_3} \cdots p_m^{x_m} \\
=& n(n-1)p_1 p_2
\end{aligned}
$$

となることから，一般に

$$\mathrm{E}[X_iX_j] = n(n-1)p_ip_j$$

と書ける。$\mathrm{E}[X_i] = np_i$ より，共分散は

$$\mathrm{Cov}(X_i, X_j) = \mathrm{E}[X_iX_j] - \mathrm{E}[X_i]\mathrm{E}[X_j] = n(n-1)p_ip_j - n^2p_ip_j = -np_ip_j$$

となる。また $\mathrm{Var}(X_i) = np_i(1-p_j)$ より，相関係数は

$$\mathrm{Corr}(X_i, X_j) = \frac{\mathrm{Cov}(X_i, X_j)}{\sqrt{\mathrm{Var}(X_i)\mathrm{Var}(X_j)}} = -\frac{\sqrt{p_ip_j}}{\sqrt{(1-p_i)(1-p_j)}}$$

で与えられる。

(2) $r \le m$ に対して $S_r = \sum_{i=1}^r X_i$ とおくとき，S_r の平均と分散を求める。$\mathrm{E}[S_r] = \sum_{i=1}^r \mathrm{E}[X_i] = n\sum_{i=1}^r p_i$ であり，分散は

$$\begin{aligned}\mathrm{Var}(S_r) =& \mathrm{E}\Big[\Big\{\sum_{i=1}^r (X_i - np_i)\Big\}^2\Big] = \sum_{i=1}^r \mathrm{Var}(X_i) + \sum_{i\neq j} \mathrm{Cov}(X_i, X_j)\\ =& \sum_{i=1}^r np_i(1-p_i) - \sum_{i\neq j} np_ip_j\end{aligned}$$

と書ける。ここで

$$\sum_{i\neq j} p_ip_j = \sum_{i=1}^r p_i \sum_{j=1, j\neq i}^r p_j = \sum_{i=1}^r p_i\Big(\sum_{j=1}^r p_j - p_i\Big) = \Big(\sum_{i=1}^r p_i\Big)^2 - \sum_{i=1}^r p_i^2$$

と書けるので，

$$\begin{aligned}\mathrm{Var}(S_r) =& n\sum_{i=1}^r p_i - n\sum_{i=1}^r p_i^2 - n\Big(\sum_{i=1}^r p_i\Big)^2 + n\sum_{i=1}^r p_i^2\\ =& n\sum_{i=1}^r p_i\Big(1 - \sum_{i=1}^r p_i\Big)\end{aligned}$$

特に $r = m$ のときには，$S_m = n$ となるので，平均は n，分散は 0 になる。

（別解）$r < m$ のときには，$S_r \sim Bin(n, \sum_{i=1}^r p_i)$ となることから直接平均と分散を求めてもよい。

問 3　確率変数 X がポアソン分布 $Po(\lambda)$ に従うとする。

$$\begin{aligned}\mathrm{E}\Big[\frac{1}{X+1}\Big] =& \sum_{x=0}^\infty \frac{1}{x+1}\frac{\lambda^x}{x!}e^{-\lambda} = \frac{1}{\lambda}\sum_{x=0}^\infty \frac{\lambda^{x+1}}{(x+1)!}e^{-\lambda}\\ =& \frac{1}{\lambda}\sum_{x=1}^\infty \frac{\lambda^x}{x!}e^{-\lambda} = \frac{1}{\lambda}(1-e^{-\lambda})\end{aligned}$$

となる。また

$$\mathrm{E}\Big[\frac{1}{X+2}\Big] = \mathrm{E}\Big[\frac{X+1}{(X+2)(X+1)}\Big] = \mathrm{E}\Big[\frac{X+2-1}{(X+2)(X+1)}\Big] = \mathrm{E}\Big[\frac{1}{X+1}\Big] - \mathrm{E}\Big[\frac{1}{(X+2)(X+1)}\Big]$$

と書ける。上の計算と同様にして

$$\mathrm{E}\Big[\frac{1}{(X+2)(X+1)}\Big] = \frac{1}{\lambda^2}\sum_{x=0}^\infty \frac{\lambda^{x+2}}{(x+2)!}e^{-\lambda} = \frac{1}{\lambda^2}(1 - e^{-\lambda} - \lambda e^{-\lambda})$$

となるので,

$$\mathrm{E}\Big[\frac{1}{X+2}\Big] = \frac{1}{\lambda}(1-e^{-\lambda}) - \frac{1}{\lambda^2}(1-e^{-\lambda}-\lambda e^{-\lambda}) = \frac{1}{\lambda^2}(\lambda - 1 + e^{-\lambda})$$

となる。

問 4 ある連続確率変数 X の (累積) 分布関数が次で与えられている。

$$F(x) = \begin{cases} 0, & -\infty < x < 0 \\ 1-(x-2)^2/4 & 0 \le x < 2, \\ 1, & 2 \le x \end{cases}$$

(1) X の確率密度関数 $f(x)$ は

$$f(x) = \begin{cases} 0, & -\infty < x < 0 \\ (2-x)/2 & 0 \le x < 2, \\ 0, & 2 \le x \end{cases}$$

となる。

(2) $\mathrm{E}[X^k]$ は

$$\begin{aligned} \mathrm{E}[X^k] =& \frac{1}{2}\int_0^2 x^k(2-x)dx = \int_0^2 x^k dx - \frac{1}{2}\int_0^2 x^{k+1}dx \\ =& \frac{2^{k+1}}{k+1} - \frac{2^{k+2}}{2(k+2)} = \frac{2^{k+1}}{(k+1)(k+2)} \end{aligned}$$

となるので,

$$\mathrm{E}[X] = \frac{2^2}{2\times 3} = \frac{2}{3}, \quad \mathrm{E}[X^2] = \frac{2^3}{3\times 4} = \frac{2}{3}$$

となる。よって

$$\mathrm{Var}(X) = \frac{2}{3} - \Big(\frac{2}{3}\Big)^2 = \frac{2}{9}$$

となる。

(3) この確率分布のメディアン x_{med} は,

$$1 - \frac{(x_{\mathrm{med}}-2)^2}{4} = \frac{1}{2}$$

を解いて, $x_{\mathrm{med}} = 2-\sqrt{2}$ となる。またモードは 0 である。

問 5 確率変数 X が指数分布 $Ex(\lambda)$ に従い, $\mu = \mathrm{E}[X] = 1/\lambda$, $\sigma^2 = \mathrm{Var}(X) = 1/\lambda^2$ とする。チェビシェフの不等式より,

$$\mathrm{P}(|X-\mu| > k\sigma) \le \frac{1}{k^2}$$

となる。一方, この確率を直接評価することを考える。

$$\mathrm{P}(|X-1/\lambda| > k/\lambda) = 1 - \mathrm{P}(|X-1/\lambda| \le k/\lambda) = 1 - \mathrm{P}(-k+1 \le \lambda X \le k+1)$$

と書けることに注意する。$\lambda X \sim Ex(1)$ であるから,

$$\mathrm{P}(|X-\mu| > k\sigma) = 1 - \int_{\min(-k+1,0)}^{k+1} e^{-y}dy$$

となる。k は $k=2,3,4$ のときには $\min(-k+1,0)=0$ であるから

$$\mathrm{P}(|X-\mu|>k\sigma)=1-\int_0^{k+1}e^{-y}dy=1-[-e^{-y}]_0^{k+1}=e^{-k-1}$$

として，正確な値が求まる。e^{-k-1} と $1/k^2$ を比較すると次のようになる。

	e^{-k-1}	k^{-2}
$k=2$	0.050	0.250
$k=3$	0.018	0.111
$k=4$	0.007	0.063

問 6　X を非負の整数上で確率をもつ離散型確率変数とし，非負の整数 k に対して $F(k)=P(X\le k)=\sum_{x=0}^{k}f(x)$ とする。$x=\sum_{k=0}^{x-1}1$ を代入すると

$$\mathrm{E}[X]=\sum_{x=0}^{\infty}xf(x)=\sum_{x=0}^{\infty}\Big(\sum_{k=0}^{x-1}1\Big)f(x)=\sum_{x=0}^{\infty}\sum_{k=0}^{x-1}f(x)$$

と表される。ここで和をとる範囲 $\{0\le x<\infty,\ 0\le k\le x-1\}$ を $\{k+1\le x<\infty,\ 0<k<\infty\}$ と書き直すことができるので，和をとる順番を交換して

$$\mathrm{E}[X]=\sum_{k=0}^{\infty}\sum_{x=k+1}^{\infty}f(x)=\sum_{k=0}^{\infty}\{1-F(k)\}$$

と書けることがわかる。

X が幾何分布 $Geo(\theta)$ に従うときには，$1-F(k)=\sum_{x=k+1}^{\infty}pq^x=q^{k+1}$ であるから，X の平均は

$$\mathrm{E}[X]=\sum_{k=0}^{\infty}q^{k+1}=\frac{q}{1-q}=\frac{q}{p}$$

となる。

問 7　X を正の実数直線上の連続な確率変数とし，その分布関数を $F(x)$，確率密度関数を $f(x)$ とする。$x=\int_0^x 1dy$ を代入すると

$$\mathrm{E}[X]=\int_0^{\infty}xf(x)dx=\int_0^{\infty}\Big(\int_0^x 1dy\Big)f(x)dx=\int_0^{\infty}\int_0^x f(x)dydx$$

と書ける。ここで，積分範囲 $\{0\le x<\infty,\ 0\le y\le x\}$ を $\{0\le y<\infty,\ y\le x<\infty\}$ に変え，積分の順序を交換すると

$$\mathrm{E}[X]=\int_0^{\infty}\int_y^{\infty}f(x)dxdy=\int_0^{\infty}\{1-F(x)\}dx$$

が成り立つことがわかる。

X が指数分布 $Ex(\lambda)$ に従うとき，$1-F(x)=\int_x^{\infty}\lambda e^{-\lambda y}dy=e^{-\lambda x}$ であるから，X の平均として

$$\mathrm{E}[X]=\int_0^{\infty}e^{-\lambda x}dx=\Big[-\frac{1}{\lambda}e^{-\lambda x}\Big]_0^{\infty}=\frac{1}{\lambda}$$

が導かれる。

問 8 離散確率変数 X の確率関数を $f(x)$, $x = 0, 1, 2, \ldots,$ については

$$\mathrm{E}[g(X)|X \geq k] = \sum_{x=k}^{\infty} g(x) \frac{f(x)}{\mathrm{P}(X \geq k)}$$

を計算する。連続確率変数 X の確率密度関数を $f(x)$, $-\infty < x < \infty$, とするとき,

$$\mathrm{E}[g(X)|X > a] = \int_a^{\infty} g(x) \frac{f(x)}{\mathrm{P}(X > a)} dx$$

を計算する。

(1) X がポアソン分布 $Po(\lambda)$ に従うとき，$\mathrm{P}(X \geq 1) = 1 - e^{-\lambda}$ より

$$\mathrm{E}[X|X \geq 1] = \frac{\sum_{k=1}^{\infty} x f(x)}{1 - e^{-\lambda}} = \frac{\sum_{k=0}^{\infty} x f(x)}{1 - e^{-\lambda}} = \frac{\lambda}{1 - e^{-\lambda}}$$

となる。

(2) X が幾何分布 $Geo(\theta)$ に従うとき，$\mathrm{P}(X \geq k) = q^k$ より

$$\mathrm{E}[X|X \geq k] = \frac{\sum_{x=k}^{\infty} x p q^x}{q^x}$$

と書ける。ここで，この分子を計算する。

$$g(p) = \sum_{x=k}^{\infty} p q^x = \frac{p q^k}{1 - q} = q^k$$

となるので，両辺を p で微分すると

$$g'(p) = \sum_{x=k}^{\infty} \left(q^x - x p q^{x-1} \right) = -k q^{k-1}$$

となる。この両辺に q を掛け，$\sum_{x=k}^{\infty} q^x = q^k/p$ に注意すると

$$\sum_{x=k}^{\infty} x p q^x = \frac{q^{k+1}}{p} + k q^k$$

となることがわかる。従って

$$\mathrm{E}[X - k|X \geq k] = \mathrm{E}[X|X \geq k] - k = \frac{q^{k+1}/p + k q^k}{q^k} - k = \frac{q}{p}$$

となる。幾何分布については平均余寿命は常に一定であることがわかる。これは無記憶性と関係している。

(3) X が正規分布 $\mathcal{N}(0, \sigma^2)$ に従うとき，$\mathrm{P}(X \geq 0) = 1/2$ より

$$\mathrm{E}[X|X \geq 0] = 2 \int_0^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/(2\sigma^2)} dx$$

を計算すればよい。$y = x^2/(2\sigma^2)$ とおくと $dy = x dx/\sigma^2$ より

$$\mathrm{E}[X|X \geq 0] = 2 \int_0^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-y} \sigma^2 dy = \frac{\sqrt{2}\sigma}{\sqrt{\pi}} \int_0^{\infty} e^{-y} dy = \frac{\sqrt{2}\sigma}{\sqrt{\pi}}$$

となる。

(4) X が指数分布 $Ex(\lambda)$ に従うときには，$\mathrm{P}(X \geq a) = \int_a^\infty \lambda e^{-\lambda x} dx = e^{-a\lambda}$ より

$$\mathrm{E}[X - a | X \geq a] = e^{a\lambda} \int_a^\infty x\lambda e^{-\lambda x} dx - a$$

と書ける。部分積分を用いると

$$\int_a^\infty x\lambda e^{-\lambda x} dx = \Big[- xe^{-\lambda x} \Big]_a^\infty + \int_a^\infty e^{-\lambda x} dx = ae^{-a\lambda} + \frac{1}{\lambda} e^{-a\lambda}$$

と書けるので

$$\mathrm{E}[X - a | X \geq a] = e^{a\lambda} \Big(ae^{-a\lambda} + \frac{1}{\lambda} e^{-a\lambda} \Big) - a = \frac{1}{\lambda}$$

となる指数分布については平均余寿命は常に一定であることがわかる。これは無記憶性と関係している。

問 9　X が正規分布 $\mathcal{N}(\mu, \sigma^2)$ に従うとき, $Y = e^X$ の分布が対数正規分布である。$x = \log y$ より $dx = y^{-1} dy$ より Y の確率密度は

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{y} \exp\Big\{ - \frac{1}{2\sigma^2} (\log y - \mu)^2 \Big\}$$

と書ける。X の積率母関数は $\mathrm{E}[e^{tX}] = e^{\mu t + (\sigma^2/2)t^2}$ であるから,

$$\mathrm{E}[Y] = \mathrm{E}[e^X] = e^{\mu + \sigma^2/2}, \quad \mathrm{E}[Y^2] = \mathrm{E}[e^{2X}] = e^{2\mu + 2\sigma^2}$$

と表されるので,

$$\mathrm{Var}(Y) = e^{2\mu + 2\sigma^2} - e^{2\mu + \sigma^2} = e^{2\mu + \sigma^2} \Big(e^{\sigma^2} - 1 \Big)$$

となる。メディアンを c とすると

$$\int_0^c \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{y} e^{-(\log y - \mu)^2/(2\sigma^2)} dy = \frac{1}{2}$$

を満たす c を求めればよい。変数変換 $y = e^x$ を行うと

$$\int_{-\infty}^{\log c} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)} dx = \frac{1}{2}$$

と書けるので，$\log c = \mu$ が解となる。従って，メディアンは $c = e^\mu$ である。

問 10　確率変数 (X, Y) の同時確率密度関数が

$$f(x, y) = \begin{cases} 2, & 0 < x < y < 1 \text{ のとき}, \\ 0, & \text{その他} \end{cases}$$

で与えられているとする。

(1) X, Y の周辺分布は

$$\begin{aligned} f_X(x) =& 2 \int_x^1 dy = 2(1 - x), \ 0 < x < 1 \\ f_Y(y) =& 2 \int_0^y dx = 2y, \ 0 < y < 1 \end{aligned}$$

であり，それぞれの平均は

$$\begin{aligned} \mathrm{E}[X] =& 2\int_0^1 x(1-x)dx = 2\Big[\frac{x^2}{2} - \frac{x^3}{3}\Big]_0^1 = \frac{1}{3} \\ \mathrm{E}[Y] =& 2\int_0^1 y^2 dy = 2\Big[\frac{y^3}{3}\Big]_0^1 = \frac{2}{3} \end{aligned}$$

となる。

(2) X, Y の共分散 $\mathrm{Cov}(X, Y)$ については

$$\mathrm{E}[XY] = 2\int_0^1 y \int_0^y x dx dy = \int_0^1 y^3 dy = \frac{1}{4}$$

であるから，

$$\mathrm{Cov}(X, Y) = \frac{1}{4} - \frac{1}{3}\frac{2}{3} = \frac{1}{36}$$

となる。

(3) $Y = y$ を与えたときの X の条件付き確率密度関数は

$$f_{X|Y}(x|y) = \frac{f(x, y)}{X_Y(y)} = \frac{2I(0 < x < y)}{2y} = \frac{1}{y} I(0 < x < y)$$

と書ける。

(4) $X|Y = y \sim U(0, y)$ であり，一様分布の平均と分散から，

$$\mathrm{E}[X|Y = y] = \frac{y}{2}, \quad \mathrm{Var}(X|Y = y) = \frac{y^2}{12}$$

となることがわかる。

問 11　コーシー・シュバルツの不等式 (4.5) を証明する。$h(t) = E[\{t(Y - \mu_Y) - (X - \mu_X)\}^2]$ とおくと，$h(t) \geq 0$ より，

$$\mathrm{Var}(Y)t^2 - 2\mathrm{Cov}(X, Y)t + \mathrm{Var}(X) \geq 0$$

が任意の t に関して成り立つので，(判別式)≤ 0 の条件を求めれば，コーシー・シュバルツの不等式が導かれる。

等号条件については，$(Y - \mu_Y)t - (X - \mu_X) = 0$ が成る立つとき等号が成立することは容易に確かめることがわかる。逆に，$\mathrm{Cov}(X, Y) = \sqrt{\mathrm{Var}(Y)\mathrm{Var}(X)}$ のときには，

$$h(t) = \mathrm{Var}(Y)t^2 - 2\sqrt{\mathrm{Var}(Y)\mathrm{Var}(X)}t + \mathrm{Var}(X) = \{\sqrt{\mathrm{Var}(Y)}t - \sqrt{\mathrm{Var}(X)}\}^2$$

と書けるので，

$$t_0 = \sqrt{\mathrm{Var}(X)}/\sqrt{\mathrm{Var}(Y)}$$

とおくと，$h(t_0) = 0$ となる。ここで，$h(t_0) = 0$ は，

$$h(t_0) = E[\{(Y - \mu_Y)t_0 - (X - \mu_X)\}^2] = 0$$

と書き直すことができ, 非負関数の期待値が 0 になることから，

$$(Y - \mu_Y)t_0 - (X - \mu_X) = 0$$

が成り立つ。$\mathrm{Cov}(X, Y) = -\sqrt{\mathrm{Var}(Y)\mathrm{Var}(X)}$ のときにも，$t_0 = -\sqrt{\mathrm{Var}(X)}/\sqrt{\mathrm{Var}(Y)}$ とおくことにより $h(t_0) = 0$ となるので，$(Y - \mu_Y)t_0 + (X - \mu_X) = 0$ が成り立つ。

問 12　X, Y を確率変数とする。

(1) X と $Y - \mathrm{E}[Y|X]$ の共分散は

$$\mathrm{Cov}(X, Y - \mathrm{E}[Y|X]) = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y|X])] = \mathrm{E}[(X - \mathrm{E}[X])\mathrm{E}[\{Y - \mathrm{E}[Y|X]\}|X]]$$

と表される。ここで $\mathrm{E}[\{Y - \mathrm{E}[Y|X]\}|X] = \mathrm{E}[Y|X] - \mathrm{E}[Y|X] = 0$ となるので，X と $Y - \mathrm{E}[Y|X]$ は無相関になる。

(2) $\mathrm{Var}(Y - \mathrm{E}[Y|X])$ は

$$\mathrm{Var}(Y - \mathrm{E}[Y|X]) = \mathrm{E}[(Y - \mathrm{E}[Y|X])^2] = \mathrm{E}[\mathrm{E}[(Y - \mathrm{E}[Y|X])^2|X]]$$

と表される。この中身は，X を与えたときの Y の条件付き分散であるから $\mathrm{E}[(Y - \mathrm{E}[Y|X])^2|X] = \mathrm{Var}(Y|X)$ となり，等式 $\mathrm{Var}(Y - \mathrm{E}[Y|X]) = \mathrm{E}[\mathrm{Var}(Y|X)]$ が成り立つ。

(3) $\mathrm{E}[(Y - X)^2]$ について

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

となるので，不等式 $\mathrm{E}[(Y - X)^2] \geq \mathrm{E}[\mathrm{Var}(Y|X)]$ が成り立つ。

(4) X と Y が独立であることに注意すると

$$\mathrm{Var}(XY) = \mathrm{E}[X^2Y^2] - (\mathrm{E}[XY])^2 = \mathrm{E}[X^2]\mathrm{E}[Y^2] - (Ex[X])^2(\mathrm{E}[Y])^2$$

と表される。$\sigma_x^2 = \mathrm{E}[X^2] - \mu_x^2$, $\sigma_y^2 = \mathrm{E}[Y^2] - \mu_y^2$ であるから，

$$\mathrm{Var}(XY) = (\sigma_x^2 + \mu_x^2)(\sigma_y^2 + \mu_y^2) - \mu_x^2\mu_y^2 = \sigma_x^2\sigma_y^2 + \mu_x^2\sigma_y^2 + \mu_y^2\sigma_x^2$$

となる。

問 13　2 つの確率変数 X, Y について，Y は区間 $[0, 1]$ 上を一様に分布し，$Y = y$ が与えられたとき X は区間 $[0, y]$ 上を一様に分布するので，

$$X|Y = y \sim U(0, y), \quad Y \sim U(0, 1)$$

と表される。条件付き分布から $\mathrm{E}[X|Y] = Y/2$, $\mathrm{Var}(X|Y) = Y^2/12$ である。命題 4.10 を用いると，X の平均と分散は

$$\begin{aligned}
\mathrm{E}[X] =& \frac{1}{2}\mathrm{E}[Y] = \frac{1}{4} \\
\mathrm{Var}(X) =& \mathrm{E}[\mathrm{Var}(X|Y)] + \mathrm{Var}(\mathrm{E}[X|Y]) = \frac{1}{12}\mathrm{E}[Y^2] + \mathrm{Var}\left(\frac{Y}{2}\right) \\
=& \frac{1}{12}\{\mathrm{Var}(Y) + (\mathrm{E}[Y])^2\} + \frac{1}{4}\mathrm{Var}(Y) = \frac{1}{12}\left(\frac{1}{12} + \frac{1}{4}\right) + \frac{1}{4}\frac{1}{12} = \frac{7}{144}
\end{aligned}$$

となる。

問 14　確率変数 $X_1, \ldots, X_n$ が独立に分布し $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ とする。$a_1, \ldots, a_n$ を実数とし $Y = \sum_{i=1}^n a_i X_i$ とおく。

Y の積率母関数 $M_Y(t)=\mathrm{E}[e^{tY}]$ は

$$\begin{aligned}M_Y(t)=&\mathrm{E}\Big[e^{(\sum_{i=1}^n a_iX_i)t}\Big]=\prod_{i=1}^n \mathrm{E}\Big[e^{(a_it)X_i}\Big]\\=&\prod_{i=1}^n e^{\mu_ia_it+\sigma_i^2a_i^2t^2/2}=\exp\Big\{\Big(\sum_{i=1}^n a_i\mu_i\Big)t+\frac{1}{2}\Big(\sum_{i=1}^n a_i^2\sigma_i^2\Big)t^2\Big\}\end{aligned}$$

と書ける。

上で求めた積率母関数は，Y が

$$Y\sim\mathcal{N}\Big(\sum_{i=1}^n a_i\mu_i,\sum_{i=1}^n a_i^2\sigma_i^2\Big)$$

に従うことを示している。従って，Y の平均と分散は，$\sum_{i=1}^n a_i\mu_i$, $\sum_{i=1}^n a_i^2\sigma_i^2$ である。

問 15 確率変数 (X,Y,Z) が (3.16) のディリクレ分布に従うとし，

$$C_{a,b,c}=\frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)}$$

とおく。同時確率密度は $f(x,y,z)=C_{a,b,c}x^{a-1}y^{b-1}z^{c-1}$, $z=1-x-y$, であるから

$$\mathrm{E}[X]=C_{a,b,c}\int\int x^{a+1-1}y^{b-1}z^{c-1}dxdy=\frac{C_{a,b,c}}{C_{a+1,b,c}}=\frac{a}{a+b+c}$$

であり，同様にして $\mathrm{E}[Y]=b/(a+b+c)$ となる。

$$\mathrm{E}[XY]=C_{a,b,c}\int\int x^{a+1-1}y^{b+1-1}z^{c-1}dxdy=\frac{C_{a,b,c}}{C_{a+1,b+1,c}}=\frac{ab}{(a+b+c)(a+b+c+1)}$$

であるから，X と Y の共分散は

$$\mathrm{Cov}(X,Y)=\frac{ab}{(a+b+c)(a+b+c+1)}-\frac{ab}{(a+b+c)^2}=-\frac{ab}{(a+b+c)^2(a+b+c+1)}$$

となる。

問 16 2 次元の連続な確率変数 (X,Y) の確率密度関数が $f(x,y)=e^{-y}$, $0<x<y$, で与えられる。(X,Y) の積率母関数を $M_{X,Y}(s,t)=\mathrm{E}[e^{sX+tY}]$ は, $|s+t|<1$, $|t|<1$ に対して

$$\begin{aligned}M_{X,Y}(s,t)=&\int_0^\infty\int_0^y e^{sx}dxe^{ty-y}dy=\frac{1}{s}\int_0^\infty(e^{sy}-1)e^{(t-1)y}dy\\=&\frac{1}{s}\int_0^\infty e^{(s+t-1)y}dy-\frac{1}{s}\int_0^\infty e^{(t-1)y}dy\\=&\frac{1}{s}\Big[\frac{1}{s+t-1}e^{(s+t-1)y}\Big]_0^\infty-\frac{1}{s}\Big[\frac{1}{t-1}e^{(t-1)y}\Big]_0^\infty\\=&\frac{1}{s}\Big\{\frac{1}{1-s-t}-\frac{1}{1-t}\Big\}=\frac{1}{(1-t)(1-s-t)}\end{aligned}$$

で与えられる。

積率母関数を s, t に関して偏微分すると

$$\begin{aligned}\frac{\partial}{\partial s}M_{X,Y}(s,t)=&\frac{1}{(1-t)(1-t-s)^2}\\\frac{\partial}{\partial t}M_{X,Y}(s,t)=&\frac{1}{(1-t)^2(1-s-t)}+\frac{1}{(1-t)(1-t-s)^2}\\\frac{\partial^2}{\partial s\partial t}M_{X,Y}(s,t)=&\frac{1}{(1-t)^2(1-s-t)^2}+\frac{2}{(1-t)(1-t-s)^3}\end{aligned}$$

と書ける。従って，$\mathrm{E}[X]=1, \mathrm{E}[Y]=2, \mathrm{E}[XY]=3$ となるので，共分散は

$$\mathrm{Cov}(X,Y)=3-2=1$$

となる。

問 17 確率ベクトル $\boldsymbol{X}$ が $\mathrm{E}[\boldsymbol{X}]=\boldsymbol{0}, \mathbf{Cov}(\boldsymbol{X})=\boldsymbol{\Sigma}$ とする。このとき，行列 $\boldsymbol{A}, \boldsymbol{B}$ に対して，$\boldsymbol{AX}$, $\boldsymbol{BX}$ の共分散は

$$\mathrm{E}[\boldsymbol{AX}(\boldsymbol{BX})^\top]=\boldsymbol{A}\mathrm{E}[\boldsymbol{XX}^\top]\boldsymbol{B}^\top=\boldsymbol{A\Sigma B}$$

と書ける。これが $\boldsymbol{0}$ になることが無相関になるための必要十分条件なので，$\boldsymbol{AX}$ と $\boldsymbol{BX}$ が無相関になるための必要十分条件として $\boldsymbol{A\Sigma B}^\top=\boldsymbol{0}$ が得られる。

問 18 平均が 0 で分散が 1 の 4 つの確率変数 Z_1,Z_2,Z_3,Z_4 について，$i\neq j$ のとき $\mathrm{Cov}(Z_i,Z_j)=\rho$ とする。$\boldsymbol{Z}=(Z_1,Z_2,Z_3,Z_4)^\top$ とおくと，$\boldsymbol{Z}$ の共分散行列は

$$\boldsymbol{\Sigma}=\mathrm{E}[\boldsymbol{ZZ}^\top]=\begin{pmatrix}1&\rho&\rho&\rho\\ \rho&1&\rho&\rho\\ \rho&\rho&1&\rho\\ \rho&\rho&\rho&1\end{pmatrix}$$

と表される。また，$U_1=Z_1+Z_2+Z_3+Z_4=(1,1,1,1)\boldsymbol{Z}$，$U_2=Z_1+Z_2-Z_3-Z_4=(1,1,-1,-1)\boldsymbol{Z}$ と表される。U_1 と U_2 の共分散は, $U_2=\boldsymbol{Z}^\top(1,1,-1,-1)^\top$ と表されることに注意すると

$$\begin{aligned}\mathrm{Cov}(U_1,U_2)=&(1,1,1,1)\mathrm{E}[\boldsymbol{ZZ}^\top]\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix}\\ =&(1,1,1,1)\begin{pmatrix}1&\rho&\rho&\rho\\ \rho&1&\rho&\rho\\ \rho&\rho&1&\rho\\ \rho&\rho&\rho&1\end{pmatrix}\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix}\\ =&(1+3\rho,1+3\rho,1+3\rho,1+3\rho)\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix}=0\end{aligned}$$

となり，U_1 と U_2 が無相関になることが示される。

問 19　$X_1,\ldots,X_n$ を独立な確率変数とし，$\mathrm{E}[X_i]=\mu_i, \mathrm{Var}(X_i)=\sigma^2$ とする。$\boldsymbol{X}=(X_1,\ldots,X_n)^\top$, $\boldsymbol{\mu}=(\mu_1,\ldots,\mu_n)^\top$ とする。

(1) $Y=\sum_{i=1}^n a_iX_i$, $Z=\sum_{i=1}^n b_iX_i$ の共分散を求める。これは，$\boldsymbol{a}=(a_1,\ldots,a_n)^\top$, $\boldsymbol{b}=(b_1,\ldots,b_n)^\top$ とおくと，$Y=\boldsymbol{a}^\top\boldsymbol{X}$, $Z=\boldsymbol{b}^\top\boldsymbol{X}$ と表される。$\mathrm{E}[Y]=\boldsymbol{a}^\top\boldsymbol{\mu}$, $\mathrm{E}[Z]=\boldsymbol{b}^\top\boldsymbol{\mu}$ であるから

$$\mathrm{Cov}(Y,Z)=\mathrm{E}[\boldsymbol{a}^\top(\boldsymbol{X}-\boldsymbol{\mu})\boldsymbol{b}^\top(\boldsymbol{Y}-\boldsymbol{\mu})]=\boldsymbol{a}^\top\mathrm{E}[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{Y}-\boldsymbol{\mu})^\top]\boldsymbol{b}=\boldsymbol{a}^\top\sigma^2\boldsymbol{I}_n\boldsymbol{b}=\sigma^2\boldsymbol{a}^\top\boldsymbol{b}$$

と書ける。ここで $\boldsymbol{I}_n$ は $n\times n$ の単位行列である。

$\boldsymbol{a}$ が任意の与えられたとき，Y と Z が無相関になるためには $\boldsymbol{a}^\top\boldsymbol{b}=0$ となるように $\boldsymbol{b}$ をとればよい。例えば,

$$\boldsymbol{b}=\boldsymbol{a}-\frac{\boldsymbol{a}^\top\boldsymbol{a}}{\boldsymbol{a}^\top\boldsymbol{1}_n}\boldsymbol{1}_n$$

ととれば,

$$\boldsymbol{a}^\top \boldsymbol{b} = \boldsymbol{a}^\top \boldsymbol{a} - \frac{\boldsymbol{a}^\top \boldsymbol{a}}{\boldsymbol{a}^\top \mathbf{1}_n} \boldsymbol{a}^\top \mathbf{1}_n = 0$$

となる。ここで $\mathbf{1}_n = (1, \ldots, 1)^\top$ はすべての成分が 1 の n 次元縦ベクトルである。

(2) $\sum_{i=1}^n X_i = \mathbf{1}_n^\top \boldsymbol{X}$ と書けるので,

$$\begin{aligned}\sum_{i=1}^n \sum_{j=1}^n X_i X_j =& (\mathbf{1}_n^\top \boldsymbol{X})^2 = \{\mathbf{1}_n^\top (\boldsymbol{X} - \boldsymbol{\mu}) + \mathbf{1}_n^\top \boldsymbol{\mu}\}^2 \\ =& \{\mathbf{1}_n^\top (\boldsymbol{X} - \boldsymbol{\mu})\}^2 + (\mathbf{1}_n^\top \boldsymbol{\mu})^2 + 2\mathbf{1}_n^\top (\boldsymbol{X} - \boldsymbol{\mu}) \mathbf{1}_n^\top \boldsymbol{\mu}\end{aligned}$$

と書き直すことができる。従って

$$\mathrm{E}\Big[\sum_{i=1}^n \sum_{j=1}^n X_i X_j\Big] = \mathrm{E}[\{\mathbf{1}_n^\top (\boldsymbol{X} - \boldsymbol{\mu})\}^2] + (\mathbf{1}_n^\top \boldsymbol{\mu})^2$$

となる。ここで

$$\mathrm{E}[\{\mathbf{1}_n^\top (\boldsymbol{X} - \boldsymbol{\mu})\}^2] = \mathrm{E}[\mathbf{1}_n^\top (\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^\top \mathbf{1}_n^\top] = \sigma^2 \mathbf{1}_n^\top \mathbf{1}_n = n\sigma^2$$

であるから

$$\mathrm{E}\Big[\sum_{i=1}^n \sum_{j=1}^n X_i X_j\Big] = n\sigma^2 + \Big(\sum_{i=1}^n \mu_i\Big)^2$$

となる。

問 20 X を $\mathcal{N}(\mu, 1)$ に従う確率変数とし, $\phi(\cdot)$, $\Phi(\cdot)$ を標準正規分布の確率密度関数と分布関数とする。このとき,

$$\begin{aligned}\mathrm{E}[\phi(X)] =& \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2} dx = \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\mu^2/4} \int_{-\infty}^{\infty} \frac{\sqrt{2}}{\sqrt{2\pi}} e^{-(x-\mu/2)^2} dx \\ =& \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\mu^2/4} = \frac{1}{\sqrt{2}} \phi(\mu/\sqrt{2})\end{aligned}$$

と書ける。また

$$\mathrm{E}[\Phi(X)] = \int_{-\infty}^{\infty} \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2} dx$$

と書ける。ここで変数変換 $y = t - x$, $dy = dt$, を行うと

$$\mathrm{E}[\Phi(X)] = \int_{-\infty}^{\infty} \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} e^{-(y+x)^2/2} dt \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2} dx$$

と書き直すことができる。

$$(y + x)^2 + (x - \mu)^2 = 2\Big(x + \frac{1}{2}(y - \mu)\Big)^2 + \frac{1}{2}(y + \mu)^2$$

より

$$\begin{aligned}\mathrm{E}[\Phi(X)] =& \int_{-\infty}^{0} \Big\{ \int_{-\infty}^{\infty} \frac{\sqrt{2}}{\sqrt{2\pi}} e^{-(x+(y-\mu)/2)^2} dx \Big\} \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-(y+\mu)^2/4} dy \\ =& \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-(y+\mu)^2/4} dy = \int_{-\infty}^{\mu/\sqrt{2}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \\ =& \Phi(\mu/\sqrt{2})\end{aligned}$$

となる。ただし，変数変換 $z = (y+\mu)/\sqrt{2}$ を用いた。

問 21 (最適予測問題) 2 次元の確率変数 (X, Y) に対して Y を X の関数 $h(X)$ で予測する問題において

$$\mathrm{MSPE}(h) = \mathrm{E}[\{Y - h(X)\}^2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{y - h(x)\}^2 f(x,y)dydx$$

を考える。平均 2 乗予測誤差を最小にする関数 $h(\cdot)$ を求めるために，繰り返し期待値の法則を用いると

$$\mathrm{E}\big[\{Y - h(X)\}^2\big] = \mathrm{E}\big[\mathrm{E}[\{Y - h(X)\}^2|X]\big]$$

となる。内側の条件付き期待値を展開し $h(X)$ に関して平方完成すると

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

と書ける。従って，平均 2 乗予測誤差を最小にする最良予測量は $h^*(X) = E[Y|X]$ で与えられる。また，$\mathrm{E}[Y^2|X] - (\mathrm{E}[Y|X])^2 = \mathrm{Var}(Y|X)$ に注意すると，$\mathrm{MSPE}(h^*) = \mathrm{E}[\mathrm{Var}(Y|X)]$ となる。

確率変数 (X, Y) が 2 次元正規分布に従うときには，最良予測量は

$$h^*(X) = \mathrm{E}[Y|X] = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X)$$

で与えられる。

第5章　統計モデルとデータの縮約

問1　確率変数 $U_1, \ldots, U_n$ が独立に一様分布 $U(0,1)$ に従うとし，順序統計量を $U_{(1)} \leq \ldots \leq U_{(n)}$ とする。

$(U_{(1)}, U_{(n)})$ の同時確率密度は

$$f_{U_{(1)},U_{(n)}}(x,y) = n(n-1)(y-x)^{n-2}, \quad 0 < x < y < 1$$

と書ける。変数変換 $r = y - x,\ t = x$ を考えると，$x = t,\ y = r + t$ よりヤコビアンは 1 となる。積分範囲は，$0 < x < y < 1$ より $0 < t < r + t < 1$ と書けるので，$0 < t < 1 - r,\ 0 < r < 1$ となる。従って，(R, T) の同時確率密度は

$$f_{T,R}(t,r) = n(n-1)r^{n-2}, \quad 0 < t < 1 - r,\ 0 < r < 1$$

と書ける。従って，R の周辺確率密度は

$$f_R(r) = \int_0^{1-r} n(n-1)r^{n-2}dt = n(n-1)r^{n-2}(1-r) = \frac{1}{B(n-1,2)} r^{(n-1)-1}(1-r)^{2-1},\ 0 < r < 1$$

と書けることがわかる。これは R がベータ分布

$$R \sim Beta(n-1, 2)$$

に従うことを示している。

$U_{(1)}, U_{(n)}$ それぞれの確率密度は

$$\begin{aligned} f_{U_{(1)}}(x) =& n(1-x)^{n-1} \sim Beta(1,n) \\ f_{U_{(n)}}(y) =& ny^{n-1} \sim Beta(n,1) \end{aligned}$$

で与えられるので，$\mathrm{E}[U_{(1)}] = 1/(n+1),\ \mathrm{E}[U_{(n)}] = n/(n+1)$ となる。従って, レンジの期待値は

$$\mathrm{E}[U_{(n)} - U_{(1)}] = \frac{n}{n+1} - \frac{1}{n+1} = \frac{n-1}{n+1}$$

となる。

例 5.2 より，$U_{(j)} \sim Beta(j, n-j+1)$ に従うことから

$$\mathrm{E}[U_{(j)}] = \frac{j}{n+1}$$

となることに注意する。

$n = 2k+1$ のときには，メディアンは $U_{\mathrm{med}} = U_{(k+1)}$ であるから，

$$\mathrm{E}[U_{\mathrm{med}}] = \mathrm{E}[U_{(k+1)}] = \frac{k+1}{n+1} = \frac{k+1}{2k+1+1} = \frac{1}{2}$$

となる。

$n = 2k$ のときには，メディアンは $U_{\mathrm{med}} = \{U_{(k)} + U_{(k+1)}\}/2$ であるから，

$$\mathrm{E}[U_{\mathrm{med}}] = \frac{\mathrm{E}[U_{(k)}] + \mathrm{E}[U_{(k+1)}]}{2} = \frac{k+k+1}{2(n+1)} = \frac{2k+1}{2(2k+1)} = \frac{1}{2}$$

となる。

問 2 問 1 と同じ設定を考える。

$(U_{(k-1)}, U_{(k)})$ の同時確率密度関数は

$$f_{U_{(k-1)},U_{(k)}}(x,y) = \frac{n!}{(k-2)!(n-k)!} x^{k-2}(1-y)^{n-k}, \quad 0 < x < y < 1$$

で与えられる。

$W_k = U_{(k)} - U_{(k-1)}$ とおくとき，W_k の確率密度関数を求める。変数変換 $w = y - k,\ t = x$ を考えると，$x = t,\ y = w + t$ より，ヤコビアンは 1 になるので，(W_k, T) の同時確率密度は

$$f_{W_k,T}(w,t) = \frac{n!}{(k-2)!(n-k)!} t^{k-2}(1-t-w)^{n-k}, \quad 0 < t < 1-w,\ 0 < w < 1$$

と書ける。これより，W_k の周辺確率密度は

$$f_{W_k}(w) = \frac{n!}{(k-2)!(n-k)!} \int_0^{1-w} t^{k-2}(1-t-w)^{n-k} dt$$

で与えられる。ここで変数変換 $t = (1-w)s,\ dt = (1-w)ds,$ により

$$\begin{aligned}
f_{W_k}(w) =& \frac{n!}{(k-2)!(n-k)!} \int_0^1 \{(1-w)s\}^{k-2}\{1-w-(1-w)s\}^{n-k}(1-w)ds \\
=& \frac{n!}{(k-2)!(n-k)!}(1-w)^{n-1} \int_0^1 s^{k-2}(1-s)^{n-k} ds \\
=& \frac{n!}{(k-2)!(n-k)!}(1-w)^{n-1} B(k-1, n-k+1) \\
=& n(1-w)^{n-1}, \quad 0 < w < 1
\end{aligned}$$

となる。この分布は k に依存しないことがわかる。

また

$$\mathrm{E}[U_{(k+1)} - U_{(k)}] = \frac{k+1}{n+1} - \frac{k}{n+1} = \frac{1}{n+1}$$

となる。

問 3 $X_1, \ldots, X_n$ が独立に指数分布 $Ex(\lambda)$ に従うとする。確率密度関数は $f(x) = \lambda \exp(-\lambda x)$, $x > 0$, であるから，分布関数は

$$F(x) = \int_0^x \lambda \exp(-\lambda t) dt = \Big[-\exp(-\lambda t) \Big]_0^x = 1 - \exp(-\lambda x)$$

となる。これより，$1 - F(x) = \exp(-\lambda x)$ と書けることに注意する。

(1) $X_{(1)},\ X_{(n)}$ の周辺確率密度関数 $f_{X_{(1)}}(x),\ f_{X_{(n)}}(y)$ は

$$\begin{aligned}
f_{X_{(1)}}(x) =& nf(x)\{1-F(x)\}^{n-1} \\
=& n\lambda e^{-\lambda x}(e^{-\lambda x})^{n-1} = n\lambda e^{-n\lambda x} \\
f_{X_{(n)}}(y) =& nf(y)\{F(y)\}^{n-1} \\
=& n\lambda e^{-\lambda y}(1-e^{-\lambda y})^{n-1}
\end{aligned}$$

で与えられる。特に，$X_{(1)}$ は指数分布 $Ex(n\lambda)$ に従う。

(2) $(X_{(1)}, X_{(n)})$ の同時確率密度関数は

$$f_{X_{(1)},X_{(n)}}(x,y) = n(n-1)\lambda^2 e^{-\lambda(x+y)}\left(e^{-\lambda x} - e^{-\lambda y}\right)^{n-2}, \quad 0 < x < y$$

で与えられる。

$(X_{(1)}, \ldots, X_{(n)})$ の同時確率密度関数は

$$f_{X_{(1)},\ldots,X_{(n)}}(x_1,\ldots,x_n) = n!\lambda^n e^{-\lambda(x_1+\cdots+x_n)}, \quad 0 < x_1 < x_2 < \cdots < x_n$$

と書ける。

問 4 $X_1, \ldots, X_n$ を独立な確率変数とし，$i = 1, \ldots, n$ に対して X_i が連続な分布関数 $F_i(x)$ に従っているとする。$Y_i = -2\log\{F_i(X_i)\}$ とおく。命題 2.17 より $U_i = F_i(X_i) \sim U(0,1)$ であり，$-\log(U_i) \sim Ex(1)$ であることに注意する。変数変換により，

$$Y_i = -2\log\{F_i(X_i)\} \sim Ex\left(\frac{1}{2}\right)$$

となることがわかるが，これはガンマ分布を用いて $Ga(1,2)$ と表される。

(1) ガンマ分布の再生性から，$S = \sum_{i=1}^n Y_i \sim Ga(n,2)$ に従う。よって $\mathrm{E}[S] = 2n$ になる。

(2) $Y_{(1)} = \min\{Y_1, \ldots, Y_n\}$ の確率密度関数は

$$f_{X_{(1)}}(x) = n\frac{1}{2}e^{-x/2}\left(e^{-x/2}\right)^{n-1} = \frac{n}{2}e^{-(n/2)x}$$

で与えられる。すなわち

$$Y_{(1)} \sim Ex\left(\frac{n}{2}\right)$$

であるから，その平均と分散は

$$\mathrm{E}[Y_{(1)}] = \frac{2}{n}, \quad \mathrm{Var}(Y_{(1)}) = \frac{4}{n^2}$$

となる。

(3) $U = \sum_{j=1}^k Y_j$, $V = \sum_{j=k+1}^n Y_j$ とおくと，U と V は独立で，$U \sim Ga(k,2)$, $V \sim Ga(n-k,2)$ に従う。例 3.9 と同様にして，W は

$$W = \frac{U}{U+V} \sim Beta(k, n-k)$$

に従うことがわかる。従って，W の確率密度は

$$f_W(w) = \frac{1}{B(k,n-k)}w^{k-1}(1-w)^{n-k-1}$$

と書ける。また $\mathrm{E}[W] = k/n$ となる。

問 5 ガンマ・ポアソン分布の平均と分散を積率母関数から直接計算することを考える。

$$X|Y \sim Po(Y), \quad Y \sim Ga(\alpha, \beta)$$

なる階層モデルにおいて X の周辺分布の積率母関数を $M_X(t)$ とする。

$M_X(t)$ は

$$M_X(t) = \mathrm{E}[e^{tX}] = \mathrm{E}[\mathrm{E}[e^{tX}|Y]] = \mathrm{E}\left[e^{(e^t-1)Y}\right] = \frac{1}{\{1-\beta(e^t-1)\}^\alpha} = \frac{1}{(1+\beta-\beta e^t)^\alpha}$$

と書ける。

$M_X(t)$ を t に関して微分すると

$$
\begin{aligned}
M'_X(t) =& \frac{\alpha\beta e^t}{(1+\beta-\beta e^t)^{\alpha+1}} \\
M''_X(t) =& \frac{\alpha\beta e^t}{(1+\beta-\beta e^t)^{\alpha+1}} + \frac{\alpha(\alpha+1)\beta^2 e^{2t}}{(1+\beta-\beta e^t)^{\alpha+2}}
\end{aligned}
$$

となるので，$\mathrm{E}[X] = M'_X(0) = \alpha\beta$, $\mathrm{E}[X^2] = M''_X(0) = \alpha\beta + \alpha(\alpha+1)\beta^2$ となる。従って，X の分散は

$$
\mathrm{Var}(X) = \alpha\beta + \alpha(\alpha+1)\beta^2 - (\alpha\beta)^2 = \alpha\beta(\beta+1)
$$

となる。

問 6　X と N を離散確率変数とし，N を与えたときの X の条件付き分布が $Bin(N,\theta)$ に従い，また N の周辺分布が $Bin(m,\lambda)$ に従うので

$$
X|N \sim Bin(N,\theta), \quad N \sim Bin(m,\lambda)
$$

と書ける。(X,N) の同時確率は

$$
f(x,n) = \frac{n!}{x!(n-x)!}\theta^x(1-\theta)^{n-x}\frac{m!}{n!(m-n)!}\lambda^n(1-\lambda)^{m-n}
$$

であるから，X の周辺確率は

$$
\begin{aligned}
f_X(x) =& \frac{\theta^x}{(1-\theta)^x}\sum_{n=x}^{m}\frac{m!}{x!(n-x)!(m-n)!}\{\lambda(1-\theta)\}^n(1-\lambda)^{m-n} \\
=& \frac{\theta^x}{x!(1-\theta)^x}\sum_{k=0}^{m-x}\frac{m!}{k!(m-k-x)!}\{\lambda(1-\theta)\}^{k+x}(1-\lambda)^{m-k-x} \\
=& \frac{m!}{x!(m-x)!}(\lambda\theta)^x(1-\lambda\theta)^{m-x}\sum_{k=0}^{m-x}\frac{(m-x)!}{k!(m-x-k)!}\Big(\frac{\lambda(1-\theta)}{1-\lambda\theta}\Big)^k\Big(\frac{1-\lambda}{1-\lambda\theta}\Big)^{m-x-k} \\
=& \frac{m!}{x!(m-x)!}(\lambda\theta)^x(1-\lambda\theta)^{m-x}\Big(\frac{\lambda(1-\theta)}{1-\lambda\theta}+\frac{1-\lambda}{1-\lambda\theta}\Big)^{m-x} \\
=& \frac{m!}{x!(m-x)!}(\lambda\theta)^x(1-\lambda\theta)^{m-x}
\end{aligned}
$$

となる。従って，X の周辺分布は

$$
X \sim Bin(m,\lambda\theta)
$$

となる。

問 7　X_i と N の積率母関数を $M_X(t)$，$M_N(t)$ とすると，S_N の積率母関数は

$$
\begin{aligned}
M_S(t) =& \mathrm{E}[e^{tS_N}] = \mathrm{E}[\mathrm{E}[e^{tX_1+\cdots+tX_N}|N]] \\
=& \mathrm{E}[\{M_X(t)\}^N] = \mathrm{E}[e^{N\log M_X(t)}] = M_N(\log M_X(t))
\end{aligned}
$$

と表される。$M_S(t)$ の t に関する微分は次のように書ける。

$$
\begin{aligned}
M'_S(t) =& \mathrm{E}\Big[Ne^{N\log M_X(t)}\Big]\frac{M'_X(t)}{M_X(t)} \\
M''_S(t) =& \mathrm{E}\Big[N^2e^{N\log M_X(t)}\Big]\Big(\frac{M'_X(t)}{M_X(t)}\Big)^2 \\
&+ \mathrm{E}\Big[Ne^{N\log M_X(t)}\Big]\frac{M''_X(t)M_X(t)-\{M'_X(t)\}^2}{\{M_X(t)\}^2}
\end{aligned}
$$

X_i の平均を μ, 分散を σ^2 とすると，$M'_S(0) = \mathrm{E}[N]\mu$, $M''_S(0) = \mathrm{E}[N^2]\mu^2 + \mathrm{E}[N]\sigma^2$ となるので，S_N の平均と分散は

$$\mathrm{E}[S_N] = \mathrm{E}[N]\mu, \quad \mathrm{Var}(S_N) = \mathrm{E}[N]\sigma^2 + \mathrm{Var}(N)\mu^2$$

となる。複合ポアソン分布の場合は，$\mu = \sigma^2 = \lambda$, $\mathrm{E}[N] = \mathrm{Var}(N) = \nu$ であるから，$\mathrm{E}[S_N] = \nu\lambda$, $\mathrm{Var}(S_N) = \nu\lambda(\lambda + 1)$ となる。

問 8 等式 $\mathrm{Cov}(X, Y) = \mathrm{E}[\mathrm{Cov}(X, Y|Z)] + \mathrm{Cov}(\mathrm{E}[X|Z], \mathrm{E}[Y|Z])$ を示す。

$$\begin{aligned}
\mathrm{Cov}(X, Y) =& \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])] \\
=& \mathrm{E}[(X - \mathrm{E}[X|Z] + \mathrm{E}[X|Z] - \mathrm{E}[X])(Y - \mathrm{E}[Y|Z] + \mathrm{E}[Y|Z] - \mathrm{E}[Y])] \\
=& \mathrm{E}[(X - \mathrm{E}[X|Z])(Y - \mathrm{E}[Y|Z])] + \mathrm{E}[(\mathrm{E}[X|Z] - \mathrm{E}[X])(\mathrm{E}[Y|Z] - \mathrm{E}[Y])] \\
&+ \mathrm{E}[(\mathrm{E}[X|Z] - \mathrm{E}[X])(Y - \mathrm{E}[Y|Z])] + \mathrm{E}[(\mathrm{E}[Y|Z] - \mathrm{E}[Y])(X - \mathrm{E}[X|Z])]
\end{aligned}$$

ここで，$\mathrm{E}[(\mathrm{E}[X|Z] - \mathrm{E}[X])(Y - \mathrm{E}[Y|Z])|Z] = 0$, $\mathrm{E}[(\mathrm{E}[Y|Z] - \mathrm{E}[Y])(X - \mathrm{E}[X|Z])|Z] = 0$ であり，$\mathrm{E}[(X - \mathrm{E}[X|Z])(Y - \mathrm{E}[Y|Z])|Z] = \mathrm{Cov}(X, Y|Z)$, $\mathrm{E}[(\mathrm{E}[X|Z] - \mathrm{E}[X])(\mathrm{E}[Y|Z] - \mathrm{E}[Y])] = \mathrm{Cov}(\mathrm{E}[X|Z], \mathrm{E}[Y|Z])$ より，上の等式が成り立つ。

問 9 N をポアソン分布 $Po(\lambda)$ に従う確率変数，$X_1, X_2, \ldots$ を独立な確率変数で，各 X_i は $\mathcal{N}(\mu, \sigma^2)$ に従い N とは独立であるとする。$\overline{X}_{N+1} = (N+1)^{-1}\sum_{i=1}^{N+1} X_i$ とおく。

$$\mathrm{E}[\overline{X}_{N+1}|N] = \frac{1}{N+1}\sum_{i=1}^{N+1}\mathrm{E}[X_i] = \mu$$

となるので，$\overline{X}_{N+1}$ の平均は $\mathrm{E}[\overline{X}_{N+1}] = \mathrm{E}[\mathrm{E}[\overline{X}_{N+1}|N]] = \mathrm{E}[\mu] = \mu$ となる。

命題 4.10 を用いると，$\overline{X}_{N+1}$ の分散は,

$$\mathrm{Var}(\overline{X}_{N+1}) = \mathrm{E}[\mathrm{Var}(\overline{X}_{N+1}|N)] + \mathrm{Var}(\mathrm{E}[\overline{X}_{N+1}|N])$$

と表される。ここで,

$$\mathrm{Var}(\overline{X}_{N+1}|N) = \frac{\sigma^2}{N+1}$$

であり，$\mathrm{E}[\overline{X}_{N+1}|N] = \mu$ であるから

$$\mathrm{Var}(\overline{X}_{N+1}) = \mathrm{E}\Big[\frac{\sigma^2}{N+1}\Big]$$

となる。$N \sim Po(\lambda)$ であるから

$$\mathrm{E}\Big[\frac{1}{N+1}\Big] = \sum_{x=0}^{\infty}\frac{\lambda^x}{(x+1)x!}e^{-\lambda} = \frac{1}{\lambda}\sum_{x=0}^{\infty}\frac{\lambda^{x+1}}{(x+1)!}e^{-\lambda} = \frac{1}{\lambda}(1 - e^{-\lambda})$$

となる。以上より,

$$\mathrm{Var}(\overline{X}_{N+1}) = \frac{\sigma^2}{\lambda}(1 - e^{-\lambda})$$

となることがわかる。

問 10 U が一様分布 $U(0, 1)$ に従う確率変数で，U を与えたとき X と Y は条件付き独立であり，$X|U \sim Ber(U)$, $Y|U \sim Ber(1 - U)$ のようなベルヌーイ分布に従うとすると

$$f_{X|U}(x|u) = u^x(1-u)^{1-x}, \quad f_{Y|U}(y|u) = (1-u)^y u^{1-y}, \quad f_U(u) = 1$$

と書ける。(X,Y,U) の同時確率密度は, 条件付き独立性から

$$f_{X|U}(x|u)f_{Y|U}(y|u)f_U(u) = u^{x+1-y}(1-u)^{y+1-x}$$

となる。従って，(X,Y) の同時確率関数は

$$\begin{aligned} f_{X,Y}(x,y) = & \int_0^1 f_{X|U}(x|u)f_{Y|U}(y|u)f_U(u)du = \int_0^1 u^{x+1-y}(1-u)^{y+1-x}du \\ =& B(x-y+2, y-x+2) = \frac{\Gamma(x-y+2)\Gamma(y-x+2)}{\Gamma(4)} \\ =& \frac{(x-y+1)!(y-x+1)!}{6} \end{aligned}$$

で与えられる。この同時確率関数を表で表すと次のようになる。

$X \backslash Y$	0	1	合計
0	1/6	1/3	1/2
1	1/3	1/6	1/2
合計	1/2	1/2	1

この表から，$\mathrm{E}[X] = 1/2$, $\mathrm{E}[Y] = 1/2$, $\mathrm{E}[XY] = 1/6$ であり，共分散は

$$\mathrm{Cov}(X,Y) = \frac{1}{6} - \frac{1}{4} = -\frac{1}{12}$$

となる。

問 11 X，Y，V を確率変数とし，V を与えたとき X と Y は条件付き独立であり，条件付き確率分布は $X|V \sim Ex(V)$, $Y|V \sim Ex(V)$ に従い，V の周辺分布は指数分布 $Ex(1)$ に従うとする。

(1) V を与えたときの (X,Y) の条件付き同時確率密度は

$$f_{(X,Y)|V}(x,y|v) = v^2 e^{-v(x+y)}$$

で与えられる。$Z = X+Y$, $W = X/(X+Y)$ とおくと，$X = ZW$, $Y = Z(1-W)$ であり，ヤコビアンは

$$\left|\frac{\partial(x,y)}{\partial(z,w)}\right| = \begin{vmatrix} w & z \\ 1-w & -z \end{vmatrix} = -wz - z(1-w) = -z$$

となるので，V を与えたときの (Z,W) の条件付き同時確率密度は

$$f_{(Z,W)|V}(z,w|v) = v^2 e^{-vz} z = \frac{1}{\Gamma(2)} v^2 z^{2-1} e^{-vz} \times I(0 < w < 1)$$

と書ける。このことは，V を与えたとき，Z と W は条件付き独立になり

$$Z|V \sim Ga\Big(2, \frac{1}{V}\Big), \quad W|V \sim U(0,1)$$

に従うことを示している。また V を与えたときの W の条件付き分布が V に依存しないこともわかる。従って,

$$f_{(Z,W)|V}(z,w|v) = f_{Z|V}(z|v) \times f_W(w)$$

と書けていることになる。V の確率分布 $f_V(v) \sim Ex(1)$ で積分すると

$$f_{Z,W}(z,w) = \int_0^\infty f_{Z|V}(z|v)f_V(v)dv \times f_W(w)$$

と書ける。従って，W の周辺分布は一様分布 $W \sim U(0,1)$ である。また Z の周辺密度は

$$
\begin{aligned}
f_Z(z) =& \int_0^\infty v^2 z e^{-vz} e^{-v} dv = z \int_0^\infty v^2 e^{-(z+1)v} dv \\
=& \frac{z\Gamma(3)}{(z+1)^3} \int_0^\infty \frac{1}{\Gamma(3)} (z+1)^3 v^{3-1} e^{-(z+1)v} dv = \frac{2z}{(z+1)^3}
\end{aligned}
$$

となる。

(2) 上の計算から $f_{Z,W}(z,w) = f_Z(z) \times f_W(w)$ と書けることがわかるので，Z と W は独立になる。

第6章　大数の法則と中心極限定理

問 1　確率変数 $X_1, X_2, \ldots$ が独立に分布し，各 X_i は 2 項分布 $Bin(k_i, \theta_i)$ に従うとする。$\mu = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n k_i\theta_i$ と $\lim_{n\to\infty} n^{-1}\sum_{i=1}^n k_i$ が存在する。標本平均 $\overline{X}_n$ に対して

$$\overline{X}_n - \frac{1}{n}\sum_{i=1}^n k_i\theta_i = \frac{1}{n}\sum_{i=1}^n (X_i - k_i\theta_i)$$

と書けるので

$$\mathrm{E}\Big[\Big(\overline{X}_n - \frac{1}{n}\sum_{i=1}^n k_i\theta_i\Big)^2\Big] = \frac{1}{n^2}\sum_{i=1}^n \mathrm{E}[(X_i - k_i\theta_i)^2] = \frac{1}{n^2}\sum_{i=1}^n k_i\theta_i(1-\theta_i)$$

となる。ここで $\theta_i(1-\theta_i) \le 1/4$ より

$$\mathrm{E}\Big[\Big(\overline{X}_n - \frac{1}{n}\sum_{i=1}^n k_i\theta_i\Big)^2\Big] \le \frac{1}{4}\frac{\sum_{i=1}^n k_i}{n^2}$$

と評価することができる。仮定より $\sum_{i=1}^n k_i/n^2 \to 0$, $n^{-1}\sum_{i=1}^n k_i\theta_i - \mu \to 0$ であるから

$$\mathrm{E}[(\overline{X}_n - \mu)^2] = \mathrm{E}\Big[\Big(\overline{X}_n - \frac{1}{n}\sum_{i=1}^n k_i\theta_i\Big)^2\Big] + \Big(\frac{1}{n}\sum_{i=1}^n k_i\theta_i - \mu\Big)^2 \to 0$$

となることがわかる。これは $\overline{X}_n$ が μ に確率収束することを示している。

問 2　$G = \int_{-\infty}^{\infty} x^2 e^{-x^2/2}dx/\sqrt{2\pi}$ の積分を考える。

$X_1, \ldots, X_n$ を独立に正規分布 $\mathcal{N}(0,1)$ に従う確率変数とするとき，G のモンテカルロ積分は

$$\overline{G}_n = \frac{1}{n}\sum_{i=1}^n X_i^2$$

で与えられる。

大数の法則から $\overline{G}_n$ は $\mathrm{E}[X_1^2] = 1$ に確率収束する。

X_1^2 の分散は $\mathrm{Var}(X_1^2) = \mathrm{E}[X_1^4] - (\mathrm{E}[X_1^2])^2 = 3 - 1 = 2$ となるので，

$$\sqrt{n}(\overline{G}_n - 1) \to_d \mathcal{N}(0,2)$$

に収束する。

```
> n <- 10; X <- rnorm(n); mean(X^2)
[1] 0.6077762
> n <- 1000; X <- rnorm(n); mean(X^2)
[1] 1.045525
> n <- 100000; X <- rnorm(n); mean(X^2)
[1] 1.003366
```

中心極限定理が成り立つ様子は次のようにして調べることができる。

```
> n <- 100; K <- 10000; sqX <- numeric(K)
> for( i in 1:K){X <- rnorm(n); sqX[i] <- mean(X^2)}
> hist(sqrt(n)*(sqX-1))
```

問 3 積分 $I = \int_0^\infty \sin(x)e^{-x^2}dx$ の値を数値的に求めたい。一様分布 $U(0,1)$ に従う独立な確率変数を $U_1, \ldots, U_n$ とするとき，これらに基づいて I のモンテカルロ積分を与えることを考える。

変数変換 $y = x^2$ を考えると，$x = \sqrt{y}$, $dx = (2\sqrt{y})^{-1}dy$, より

$$I = \int_0^\infty \sin(x)e^{-x^2}dx = I = \int_0^\infty \sin(\sqrt{y})\frac{1}{2\sqrt{y}}e^{-y}dy$$

と書き直すことができる。$Y_i = -\log U_i$ とおくと，$Y_i \sim Ex(1)$ であるから，I のモンテカルロ積分は

$$\frac{1}{n}\sum_{i=1}^n \sin\left(\sqrt{-\log U_i}\right)\frac{1}{2\sqrt{-\log U_i}}$$

で与えられる。

```
> n <- 100; U <- runif(n); X <- sqrt(-log(U)); mean(sin(X)/(2*X))
[1] 0.4260547
> n <- 10000; U <- runif(n); X <- sqrt(-log(U)); mean(sin(X)/(2*X))
[1] 0.4246022
```

問 4 確率変数 $X_1, \ldots, X_n$ が独立で分布関数 $F(x)$ に従うとする。経験分布関数は $\hat{F}(x) = \sum_{i=1}^n I(X_i \leq x)/n$ で与えられる。このとき，

$$n\hat{F}(x) \sim Bin(n, F(x))$$

に従うことに注意する。

2 項分布の分散から $\mathrm{E}[\{\hat{F}(x) - E(x)\}^2] = n^{-1}F(x)\{1 - F(x)\}$ となり，これは 0 に収束するので，$\hat{F}(x)$ が $F(x)$ に確率収束することがわかる。

$\sqrt{n}\{\hat{F}(x) - F(x)\} \to_d \mathcal{N}(0, F(x)\{1 - F(x)\})$ に収束する。

問 5 確率変数 X が負の 2 項分布 $NBin(m,p)$ に従っているものとする。$m \to \infty$ とするときの漸近的性質を考える。いま，$Y_1, \ldots, Y_m$ を独立で幾何分布 $Geo(p)$ に従う確率変数とすると，命題 4.20 の分布の再生性から

$$Y_1 + \cdots + Y_m \sim NBin(m,p)$$

に従うので，X は $\sum_{i=1}^m Y_i$ と同じ分布に従うと考えてよい。また，$\mathrm{E}[Y_i] = q/p$, $\mathrm{Var}(Y_i) = q/p^2$ であることに注意する。

X/m が $\mathrm{E}[X/m] = q/p$ に確率収束することは, 任意の $c > 0$ に対して

$$\mathrm{P}\Big(\Big|\frac{X}{m} - \frac{q}{p}\Big| > c\Big) = \mathrm{P}\Big(\Big|\frac{1}{m}\sum_{i=1}^m Y_i - \frac{q}{p}\Big| > c\Big)$$

が 0 に収束することを示せばよいが，これは大数の法則から成り立つ。

$(X - \mathrm{E}[X])/\sqrt{m}$ の漸近分布については,

$$\mathrm{P}\Big(\frac{X - \mathrm{E}[X]}{\sqrt{m}} < x\Big) = \mathrm{P}\Big(\sqrt{m}\Big(\frac{1}{m}\sum_{i=1}^m Y_i - \frac{q}{p}\Big) < x\Big)$$

であり，中心極限定理からこれが

$$\mathcal{N}\Big(0, \frac{q}{p^2}\Big)$$

の分布関数に収束することがわかる。

問 6　確率変数 X が 2 項分布 $Bin(n,p)$ に従うとき，$np=\lambda$ のもとで $n\to\infty, p\to 0$ とすると，X の分布はポアソン分布 $Po(\lambda)$ に分布収束することを積率母関数を用いて示す。

X の積率母関数は

$$M_X(t) = \mathrm{E}[e^{tX}] = (pe^t+1-p)^n = \Big(1+\frac{\lambda(e^t-1)}{n}\Big)^n \to e^{\lambda(e^t-1)}$$

に収束することがわかる。これはポアソン分布の積率母関数であるから，定理 6.5(連続性定理) より 2 項分布がポアソン分布に収束することがわかる。

問 7　確率変数 $X_1,\dots,X_n$ が独立で，各 X_i の確率密度関数は $f(x)=2x, 0<x<1,$ とする。

$$\mathrm{E}[X_1] = \int_0^1 2x^2dx = \frac{2}{3}, \quad \mathrm{E}[X_1^2] = \int_0^1 2x^3dx = \frac{1}{2},\ \mathrm{Var}(X_1) = \frac{1}{2}-\frac{4}{9}=\frac{1}{18}$$

であることに注意する。$S_n = X_1+\cdots X_n$ とおくと，中心極限定理を用いると

$$\sqrt{n}3\sqrt{2}\Big(\frac{1}{n}S_n-\frac{2}{3}\Big) = \frac{\sqrt{2}}{\sqrt{n}}(3S_n-2n) \to_d \mathcal{N}(0,1)$$

となる。この近似を用いると

$$\mathrm{P}(S_n\le a) = \mathrm{P}\Big(\frac{\sqrt{2}}{\sqrt{n}}(3S_n-2n)\le\frac{\sqrt{2}}{\sqrt{n}}(3a-2n)\Big) \approx \Phi\Big(\frac{\sqrt{2}}{\sqrt{n}}(3a-2n)\Big)$$

で近似できることがわかる。

問 8　確率変数 $X_1,\dots,X_n$ が独立で，$X_i\sim Po(\lambda)$ に従うとする。また $g(\cdot)$ を連続微分可能な関数とし $g'(\lambda)\neq 0$ とする。$g(\overline{X})$ を $\overline{X}=\lambda$ の周りで 1 回テーラー展開すると

$$g(\overline{X}) \approx g(\lambda)+g'(\lambda)(\overline{X}-\lambda)$$

と書ける。また中心極限定理より $\sqrt{n}(\overline{X}-\lambda)\to_d \mathcal{N}(0,\lambda)$ に収束することがわかる。

(1) $n\to\infty$ のとき

$$\sqrt{n}\{g(\overline{X})-g(\lambda)\} = g'(\lambda)\sqrt{n}(\overline{X}-\lambda) \to_d \mathcal{N}\big(0,\lambda\{g'(\lambda)\}^2\big)$$

となる。

(2) $\lambda\{g'(\lambda)\}^2=1$ を満たす関数 $g(\cdot)$ を求める。

$$g'(\lambda) = \frac{1}{\sqrt{\lambda}}$$

となるので，両辺を λ に関して不定積分すると

$$g(\lambda) = 2\sqrt{\lambda}$$

となる。従って，$n\to\infty$ のとき $2\sqrt{n}\{\sqrt{\overline{X}}-\sqrt{\lambda}\}\to_d\mathcal{N}(0,1)$ となる。

これらを数値的に比較するには次のようにすればよい。

```
> n <- 20; la <- 3
> ppois(4*n,n*la)-ppois(2.5*n-1,n*la)
[1] 0.9099617
> pnorm(sqrt(n)*(4-la)/sqrt(la))-pnorm(sqrt(n)*(2.5-la)/sqrt(la))
[1] 0.8967356
> pnorm(2*sqrt(n)*(sqrt(4)-sqrt(la)))-pnorm(2*sqrt(n)*(sqrt(2.5)-sqrt(la)))
[1] 0.9031858
```

問 9 $X_1,\ldots,X_n$ が独立に平均 μ, 分散 σ^2 の分布に従うとき，標本分散 S^2 について，$n\to\infty$ のときの漸近的な性質を考える。

$\sum_{i=1}^n(X_i-\overline{X})^2=\sum_{i=1}^n(X_i-\mu)^2-n(\overline{X}-\mu)^2$ より，$W_n=n^{-1}\sum_{i=1}^n(X_i-\mu)^2$ とおくと

$$S^2=\frac{1}{n}\sum_{i=1}^n(X_i-\overline{X})^2=W_n-(\overline{X}-\mu)^2$$

と書ける。$(X_1-\mu)^2,\ldots,(X_n-\mu)^2$ は互いに独立に同一分布に従い，$\mathrm{E}[(X_1-\mu)^2]=\sigma^2$ であるから，大数の法則により $W_n\to_p\sigma^2$ に確率収束する。また $\overline{X}\to_p\mu$ より $(\overline{X}-\mu)^2\to_p 0$ である。$h(W_n,\overline{X})=W_n-(\overline{X}-\mu)^2$ とおくと，これは連続関数なので，定理 6.9 の連続写像定理より，$h(W_n,\overline{X})\to_p h(\sigma^2,\mu)$ に確率収束する。このことは S^2 が σ^2 に確率収束することを示している。

上の式展開を用いると

$$\sqrt{n}(S^2-\sigma^2)=\sqrt{n}(W_n-\sigma^2)-\frac{1}{\sqrt{n}}\{\sqrt{n}(\overline{X}-\mu)\}^2$$

と表すことができる。$(X_1-\mu)^2$ の分散は

$$\mathrm{Var}((X_1-\mu)^2)=\mathrm{E}\Big[\Big\{(X_1-\mu)^2-\sigma^2\Big\}^2\Big]=\mathrm{E}[(X_1-\mu)^4]-\sigma^4=\mu_4-\sigma^4$$

と書ける。ここで $\mu_4=\mathrm{E}[(X_1-\mu)^4]$ である。中心極限定理より

$$\sqrt{n}(W_n-\sigma^2)\to_d\mathcal{N}(0,\mu_4-\sigma^4)$$

に分布収束する。また $\sqrt{n}(\overline{X}-\mu)$ は正規分布に収束するから，

$$\frac{1}{\sqrt{n}}\{\sqrt{n}(\overline{X}-\mu)\}^2\to_p 0$$

に確率収束する。従って，定理 6.10(スラツキーの定理) より

$$\sqrt{n}(S^2-\sigma^2)=\sqrt{n}(W_n-\sigma^2)-\frac{1}{\sqrt{n}}\{\sqrt{n}(\overline{X}-\mu)\}^2\to_d\mathcal{N}(0,\mu_4-\sigma^4)$$

となる。例えば，$X_i\sim Ex(\lambda)$ のときには $\sigma^2=1/\lambda^2$ であるから，中心極限定理が成り立つ様子は次のようにして調べることができる。

```
> n <- 50; K <- 10000; S2 <- numeric(K); la <- 2
> for(i in 1:K){ X <- rexp(n, la); S2[i] <- var(X)*(n-1)/n}
> hist(sqrt(n)*(S2-1/la^2))
```

問 10 $U_1,\ldots,U_n$ を独立な確率変数で，各 U_i は区間 $(0,\theta)$ 上の一様分布 $U(0,\theta)$ に従うとすると

$$f(x)=\frac{1}{\theta}I(0<x<\theta),\quad F(x)=\frac{x}{\theta}I(0<x<\theta)+I(x\ge\theta)$$

と書ける。$U_{(n)} = \max(U_1, \dots, U_n)$ とする。

(1) $U_{(n)}$ の確率密度関数は

$$f_{U_{(n)}}(x) = \frac{n}{\theta}\Big(\frac{x}{\theta}\Big)^{n-1}$$

で与えられる。

(2) $n \to \infty$ のとき，$U_{(n)}$ が θ に確率収束することを示す。

$$\mathrm{E}[(U_{(n)} - \theta)^2] = \int_0^\theta (x-\theta)^2 \frac{n}{\theta}\Big(\frac{x}{\theta}\Big)^{n-1} dx$$

において，変数変換 $y = x/\theta$，$x = \theta y$，$dx = \theta dy$, を用いると

$$\begin{aligned}\mathrm{E}[(U_{(n)} - \theta)^2] =& n\theta^2 \int_0^1 (y-1)^2 y^{n-1} dy = n\theta^2 \int_0^1 \{y^{n+1} - 2y^n + y^{n-1}\} dy \\ =& n\theta^2\Big(\frac{1}{n+2} - \frac{2}{n+1} + \frac{1}{n}\Big) = \frac{2\theta^2}{(n+1)(n+2)} \to 0\end{aligned}$$

となり，$U_{(n)}$ は θ に平均2乗収束する。従って $U_{(n)}$ は θ に確率収束する。

(3) $n \to \infty$ のとき，$n(\theta - U_{(n)})$ の漸近分布を求める。

$$\begin{aligned}\mathrm{P}(n(\theta - U_{(n)}) > c) =& \mathrm{P}\Big(U_{(n)} < \theta - \frac{c}{n}\Big) = \int_0^{\theta - c/n} \frac{n}{\theta}\Big(\frac{x}{\theta}\Big)^{n-1} dx \\ =& \int_0^{1 - c/(n\theta)} n y^{n-1} dy = \Big(1 - \frac{c}{n\theta}\Big)^n \to e^{-c/\theta}\end{aligned}$$

となる。これは

$$n(\theta - U_{(n)}) \to_d Ex\Big(\frac{1}{\theta}\Big)$$

に分布収束することを示している。この収束の様子は次のようにして調べることができる。

```
> n <- 10; K <- 10000; maxU <- numeric(K); th <- 5
> for(i in 1:K){ U <- runif(n, 0, th); maxU[i] <- max(U)}
> hist(n*(th-maxU))
```

問 11　確率変数の列 $\{U_n\}$, $n = 1, 2, \dots,$ について，$k = 1, \dots, n$ に対して U_n の確率が $P(U_n = k/n) = 1/n$, $k = 1, \dots, n,$ で与えられるとする。$n \to \infty$ のとき，U_n は区間 $(0,1)$ 上の一様分布 U に分布収束することを示す。

U_n の積率母関数は

$$M_{U_n}(t) = \mathrm{E}[e^{tU_n}] = \frac{1}{n}\sum_{k=1}^n e^{(t/n)k} = \frac{1}{n}\frac{e^{t/n}(1-e^t)}{1-e^{t/n}}$$

と書ける。ここで

$$e^{t/n} = 1 + \frac{t}{n} + \frac{t^2}{2n^2} + o(n^{-2})$$

より

$$M_{U_n}(t) = \frac{1}{n}\frac{e^{t/n}}{t/n + t^2/(2n^2) + o(n^{-2})}(e^t - 1) = \frac{e^{t/n}}{t + t^2/(2n) + o(n^{-1})}(e^t - 1) \to \frac{e^t - 1}{t}$$

となる。一方，$U \sim U(0,1)$ の積率母関数は

$$M_U(t) = \mathrm{E}[e^{tU}] = \int_0^1 e^{tx} dx = \Big[\frac{1}{t}e^{tx}\Big]_0^1 = \frac{e^t - 1}{t}$$

と書けるので，定理 6.5(連続性定理) より，$U_n \to_d U$ に収束することがわかる。

問 12　確率変数 X がガンマ分布 $Ga(\alpha,\beta)$ に従うとする。β を固定し $\alpha\to\infty$ とするとき，$(X-\alpha\beta)/\sqrt{\alpha\beta^2}$ が $\mathcal{N}(0,1)$ に分布収束することを示す。

$X\sim Ga(\alpha,\beta)$ より $\mathrm{E}[X]=\alpha\beta$, $\mathrm{Var}(X)=\alpha\beta^2$ である。$Z=(X-\alpha\beta)/\sqrt{\alpha\beta^2}$ とおくと，この積率母関数は, X の積率母関数を用いて

$$\begin{aligned}M_Z(t) =&\mathrm{E}[e^{t(X-\alpha\beta)/\sqrt{\alpha\beta^2}}]=\mathrm{E}[e^{(t/\sqrt{\alpha\beta^2})X}]e^{-t\sqrt{\alpha}}=\frac{e^{-t\sqrt{\alpha}}}{(1-\beta t/\sqrt{\alpha\beta^2})^{\alpha}}\\ =&\frac{1}{e^{t\sqrt{\alpha}}(1-t/\sqrt{\alpha})^{\alpha}}=\frac{1}{e^{(t/\sqrt{\alpha})\alpha}(1-t/\sqrt{\alpha})^{\alpha}}=\frac{1}{\{e^{(t/\sqrt{\alpha})}(1-t/\sqrt{\alpha})\}^{\alpha}}\\ =&\frac{1}{\{e^{t/\sqrt{\alpha}}-(t/\sqrt{\alpha})e^{t/\sqrt{\alpha}}\}^{\alpha}}\end{aligned}$$

と表すことができる。ここで, α が大きいとき

$$e^{t/\sqrt{\alpha}}=1+\frac{t}{\sqrt{\alpha}}+\frac{t^2}{2\alpha}+o(\alpha^{-1})$$

で近似できるので

$$e^{t/\sqrt{\alpha}}-\frac{t}{\sqrt{\alpha}}e^{t/\sqrt{\alpha}}=1-\frac{t^2}{2\alpha}+o(\alpha^{-1})$$

と書ける。これを代入すると

$$\frac{1}{(e^{t/\sqrt{\alpha}}-(t/\sqrt{\alpha})e^{t/\sqrt{\alpha}})^{\alpha}}=\frac{1}{\{1-t^2/(2\alpha)+o(\alpha^{-1})\}^{\alpha}}\to\frac{1}{e^{-t^2/2}}=e^{t^2/2}$$

に収束することがわかる。これは標準正規分布の積率母関数であるから，定理 6.5(連続性定理) より, $\alpha\to\infty$ のとき

$$(X-\alpha\beta)/\sqrt{\alpha\beta^2}\to_d\mathcal{N}(0,1)$$

に分布収束する。

この収束の様子は次のようにして調べることができる。

```
> al <- 100; be <- 1; K <- 10000
> X <- rgamma(K, al, 1/be)
> hist( (X-al*be)/sqrt(al*be^2))
```

第7章　正規分布から導かれる分布

問1　確率変数 $X_1, \ldots, X_9$ が独立で，各 X_i が $\mathcal{N}(\mu, 1)$ に従うので，標本平均は $\overline{X} \sim \mathcal{N}(\mu, 1/9)$ に従う。従って，$Z = 3(\overline{X} - \mu) \sim \mathcal{N}(0, 1)$ に従う。

(1) については，$\Phi(1) = 0.84$ より

$$\mathrm{P}\Big(\mu - \frac{1}{3} \leq \overline{X} \leq \mu + \frac{1}{3}\Big) = \mathrm{P}(-1 \leq Z \leq 1) = \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 = 0.68$$

となる。

(2) については，$z_{0.025} = 1.96$ より

$$\mathrm{P}\Big(\mu - \frac{c}{3} \leq \overline{X} \leq \mu + \frac{c}{3}\Big) = \mathrm{P}(-c \leq Z \leq c) = 2\Phi(c) - 1 = 1 - 2\mathrm{P}(Z > c)$$

となり，$1 - 2\mathrm{P}(Z > c) = 0.95$，すなわち $\mathrm{P}(Z > c) = 0.025$ となる c の値は $c = z_{0.025}$ であるから $c = 1.96$ になる。

```
> pnorm(1)-pnorm(-1)
[1] 0.6826895
> c <- qnorm(0.975)
> c
[1] 1.959964
```

問2　確率変数 $X_1, \ldots, X_9$ が独立で，各 X_i が $\mathcal{N}(\mu, \sigma^2)$ に従うとする。$\overline{X}$ を標本平均，V^2 を不偏分散とすると，$T = 3(\overline{X} - \mu)/V$ は自由度 8 の t 分布に従う。。

(1) については，$\mathrm{P}(T \leq 1) = 0.83$ より

$$\mathrm{P}\Big(\mu - \frac{V}{3} \leq \overline{X} \leq \mu + \frac{V}{3}\Big) = \mathrm{P}(-1 \leq T \leq 1) = 2\mathrm{P}(T \leq 1) - 1 = 0.65$$

となる。

(2) については，$t_{8,0.025} = 2.31$ より

$$\mathrm{P}\Big(\mu - c\frac{V}{3} \leq \overline{X} \leq \mu + c\frac{V}{3}\Big) = \mathrm{P}(-c \leq T \leq c) = 1 - 2\mathrm{P}(T > c))$$

となり，$1 - 2\mathrm{P}(T > c)) = 0.95$，すなわち $\mathrm{P}(T > c) = 0.025$ を満たす c は $c = t_{8,0.025}$ であるから $c = 2.31$ になる。

```
> pt(1, 8)-pt(-1, 8)
[1] 0.6534065
> c <- qt(0.975, 8)
> c
[1] 2.306004
```

問3　確率変数 $X_1, \ldots, X_n$ が独立で，各 X_i が $\mathcal{N}(\mu, \sigma^2)$ に従うとし，V^2 を不偏分散とする。このとき，$Y = (n-1)V^2/\sigma^2$ とおくと，$Y \sim \chi^2_{n-1}$ に従う。

$V^2 = \sigma^2 Y/(n-1)$ と表しておく。$Y \sim \chi^2_{n-1}$ より

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

となることに注意する。これより $\mathrm{E}[V^2] = \sigma^2$ となる。また

$$\mathrm{E}[V^4] = \frac{\sigma^4}{(n-1)^2}\mathrm{E}[Y^2] = \sigma^4 \frac{n+1}{n-1}$$

より

$$\mathrm{Var}(V^2) = \sigma^4 \frac{n+1}{n-1} - \sigma^4 = \frac{2}{n-1}\sigma^4$$

となる。

問 4 確率変数 X, Y が独立で，ともに指数分布 $Ex(1)$ に従うので，同時確率密度は

$$f(x,y) = e^{-(x+y)}$$

と書ける。変数変換 $w = x/y,\ t = y$ を考えると，$x = wt,\ y = t$ より，ヤコビアンは

$$\left|\frac{\partial(x,y)}{\partial(w,t)}\right| = \begin{vmatrix} t & w \\ 0 & 1 \end{vmatrix} = t$$

となるので，(W,T) の同時確率密度は

$$f_{W,T}(w,t) = e^{-(w+1)t}t$$

となり，W の周辺密度は

$$f_W(w) = \int_0^\infty t e^{-(w+1)t} dt = \frac{1}{(w+1)^2}\int_0^\infty \frac{1}{\Gamma(2)}(w+1)^2 t^{2-1} e^{-(w+1)t} dt = \frac{1}{(w+1)^2}$$

となる。この分布関数は

$$F_W(w) = \int_0^w \frac{1}{(x+1)^2}dx = \Big[-\frac{1}{x+1}\Big]_0^w = 1 - \frac{1}{w+1} = \frac{w}{w+1}$$

となる。

問 5 確率変数 $X_1, \ldots, X_n$ が独立に，$X_i \sim \mathcal{N}(\mu, \sigma^2)$ に従うとし，$\overline{X}$, S^2 を標本平均，標本分散とする。$Z = \sqrt{n}(\overline{X}-\mu)/\sigma$, $Y = nS^2/\sigma^2$ とおくと，Z と Y は独立に分布し，$Z \sim \mathcal{N}(0,1)$, $Y \sim \chi^2_{n-1}$ に従う。$\overline{X} = \mu + \sigma Z/\sqrt{n}$, $S = \sigma\sqrt{Y/n}$ と書けることと $a + (n-1)/2 > 0$ を満たす実数 a に対して

$$\mathrm{E}[Y^a] = 2^a \frac{\Gamma(a+(n-1)/2)}{\Gamma((n-1)/2)}$$

であることに注意する。

$\overline{X} + \sqrt{n}S = \mu + \sigma Z/\sqrt{n} + \sigma\sqrt{Y}$ であるから

$$\mathrm{E}[\overline{X} + \sqrt{n}S] = \mu + \sigma\mathrm{E}[Y^{1/2}] = \mu + \sigma\sqrt{2}\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}$$

となる。また $\overline{X} + (\sqrt{n}S)^{-1} = \mu + \sigma Z/\sqrt{n} + (\sigma\sqrt{Y})^{-1}$ であるから

$$\mathrm{E}[\overline{X} + (\sqrt{n}S)^{-1}] = \mu + \frac{1}{\sigma}\mathrm{E}[Y^{-1/2}] = \mu + \frac{1}{\sigma}\frac{1}{\sqrt{2}}\frac{\Gamma(n/2-1)}{\Gamma((n-1)/2)}$$

となる。

$\overline{X}+\sqrt{n}S$ と $\overline{X}+(\sqrt{n}S)^{-1}$ の共分散については

$$\begin{aligned}&\mathrm{E}[\{\overline{X}+\sqrt{n}S\}\{\overline{X}+(\sqrt{n}S)^{-1}\}]=\mathrm{E}[\overline{X}^2]+\mathrm{E}[\sqrt{n}S+(\sqrt{n}S)^{-1}]\mathrm{E}[\overline{X}]+1\\ =&\mu^2+\frac{\sigma^2}{n}+\sigma\mathrm{E}[Y^{1/2}]\mu+\frac{1}{\sigma}\mathrm{E}[Y^{-1/2}]\mu+1\end{aligned}$$

と書けることに注意し，また

$$\begin{aligned}&\mathrm{E}[\overline{X}+\sqrt{n}S]\mathrm{E}[\overline{X}+(\sqrt{n}S)^{-1}]=\{\mu+\sigma\mathrm{E}[Y^{1/2}]\}\{\mu+\frac{1}{\sigma}\mathrm{E}[Y^{-1/2}]\}\\ =&\mu^2+\sigma\mathrm{E}[Y^{1/2}]\mu+\frac{1}{\sigma}\mathrm{E}[Y^{-1/2}]\mu+\mathrm{E}[Y^{1/2}]\mathrm{E}[Y^{-1/2}]\end{aligned}$$

と書けるので

$$\begin{aligned}&\mathrm{Cov}(\overline{X}+\sqrt{n}S,\overline{X}+(\sqrt{n}S)^{-1})=\frac{\sigma^2}{n}+1-\mathrm{E}[Y^{1/2}]\mathrm{E}[Y^{-1/2}]\\ =&\frac{\sigma^2}{n}+1-\frac{\Gamma(n/2)\Gamma(n/2-1)}{\{\Gamma((n-1)/2)\}^2}\end{aligned}$$

となる。

問 6　自然数 n に対して X を χ^2_n に従う確率変数とする。$Z_1,\dots,Z_n$ を独立に $\mathcal{N}(0,1)$ に従う確率変数とすると，命題 6.2 より

$$X=\sum_{i=1}^n Z_i^2$$

$n\to\infty$ とするとき，漸近的な性質を調べる。

(1) $\mathrm{E}[Z_i^2]=1$ より

$$\frac{X}{n}=\frac{1}{n}\sum_{i=1}^n Z_i^2\to_p 1$$

となる。

$\mathrm{Var}(Z_i^2)=\mathrm{E}[Z_i^4]-1=3-1=2$ より

$$\sqrt{n}(X/n-1)=\sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^n Z_i^2-1\Big)\to_d\mathcal{N}(0,2)$$

となる。

(2) $\sqrt{X}-\sqrt{n}=\sqrt{n}(\sqrt{X/n}-1)$ と表せることに注意する。$g(y)=\sqrt{y}$ とおくと，$g'(y)=1/(2\sqrt{y})$ であり，

$$\sqrt{n}(g(X/n)-g(1))\to_d\mathcal{N}(0,2\{g'(1)\}^2)$$

に収束するので，$\sqrt{X}-\sqrt{n}\to_d\mathcal{N}(0,1/2)$ となる。

収束の様子は次のようにして調べることができる。

```
> n <- 100; X <- rchisq(K,n);
Y1 <-sqrt(n)*(X/n-1);
Y2 <- sqrt(X)-sqrt(n);
> hist(Y1)
> hist(Y2)
```

問 7 X と Y は独立な確率変数で，$X \sim \mathcal{N}(0,1)$, $\mathrm{P}(Y=-1)=\mathrm{P}(Y=1)=1/2$ とする。$Z=XY$ とおく。

Z の積率母関数は

$$M_Z(t)=\mathrm{E}[e^{tXY}]=\frac{1}{2}\mathrm{E}\Big[e^{tX}+e^{-tX}\Big]=\frac{1}{2}\Big\{\mathrm{E}[e^{tX}]+\mathrm{E}[e^{-tX}]\Big\}$$

と書ける。ここで $\mathrm{E}[e^{tX}]=\mathrm{E}[e^{-tX}]=e^{t^2/2}$ であるから，$M_Z(t)=e^{t^2/2}$ となり，$Z \sim \mathcal{N}(0,1)$ が示される。

X と Z は独立でないことを示すには積率母関数を用いればよい。まず

$$\begin{aligned}M_{X,Z}(s,t)=&\mathrm{E}[e^{sX+tZ}]=\mathrm{E}[e^{(s+tY)X}]=\frac{1}{2}\mathrm{E}[e^{(s+t)X}+e^{(s-t)X}]\\=&\frac{1}{2}\Big\{e^{(s+t)^2/2}+e^{(s-t)^2/2}\Big\}=\frac{1}{2}\Big\{e^{(s^2+t^2)/2+st}+e^{s^2+t^2)/2-st}\Big\}\\=&\frac{1}{2}e^{(s^2+t^2)/2}\Big(e^{st}+e^{-st}\Big)\end{aligned}$$

と書けることがわかる。一方，X，Z ともに標準正規分布に従うので

$$M_X(s)M_Z(t)=e^{s^2/2}e^{t^2/2}=e^{(s^2+t^2)/2}$$

となる。これは，$M_{X,Z}(st)\neq M_X(s)M_Z(t)$ を示しており，X と Z が独立でないことがわかる。

X と Z の共分散については

$$\mathrm{E}[XZ]=\mathrm{E}[X^2Y]=\mathrm{E}[X^2]\Big\{-\frac{1}{2}+\frac{1}{2}\Big\}=0$$

より, $\mathrm{Cov}(X,Z)=\mathrm{E}[XZ]-\mathrm{E}[X]\mathrm{E}[Z]=0$ となり，無相関になることがわかる。

問 8 2 つの連続な確率変数 X と Y が独立で同一分布に従うとし，$Z=\min(X,Y)$ とおく。

(1) X, Y の確率分布が $\mathcal{N}(0,1)$ であるとき，順序統計量の節で学んだ内容を思い出すと，Z の分布関数は

$$\begin{aligned}F_Z(z)=&1-\mathrm{P}(Z>z)=1-\mathrm{P}(\min(X,Y)>z)=1-\mathrm{P}(X>z,Y>z)\\=&1-\mathrm{P}(X>z)\mathrm{P}(Y>z)=1-(1-\Phi(z))^2\end{aligned}$$

と書ける。よって，Z の確率密度は

$$f_Z(z)=2\phi(z)(1-\Phi(z))$$

となる。ただし $\phi(z)$, $\Phi(z)$ は標準正規分布の確率密度関数と分布関数を表している。$V=Z^2$ とおくと，命題 2.19 の平方変換公式を用いると

$$\begin{aligned}f_V(v)=&\frac{2\phi(\sqrt{v})(1-\Phi(\sqrt{v}))+2\phi(-\sqrt{v})(1-\Phi(-\sqrt{v}))}{2\sqrt{v}}\\=&\frac{\phi(\sqrt{v})\{1-\Phi(\sqrt{v})+1-(1-\Phi(-\sqrt{v}))\}}{\sqrt{v}}\\=&v^{-1/2}\phi(\sqrt{v})=\frac{1}{\Gamma(1/2)}\Big(\frac{1}{2}\Big)^{1/2}v^{1/2-1}e^{-v/2}\end{aligned}$$

となる。これは自由度 1 のカイ 2 乗分布である。

(2) (1) の結果は，$\{\min(X,Y)\}^2$ の分布は X^2 の分布に等しくなることを示している。この性質は，一般に，X，Y の確率密度関数 $f(x)$ が対称性 $f(x)=f(-x)$ を満たすならば，常に成り立つことを示す。

$Z = \min(X, Y) = XI(X < Y) + YI(X > Y)$ であるから，$Z^2 = X^2 I(X < Y) + Y^2 I(X > Y)$ である。従って

$$M_{Z^2}(t) = \mathrm{E}[e^{tZ^2}] = \mathrm{E}[e^{tX^2} I(X < Y) + e^{tY^2} I(X > Y)] = \mathrm{E}[e^{tX^2} I(X < Y)] + \mathrm{E}[e^{tY^2} I(X > Y)]$$

書ける。ここで X, Y の確率分布の対称性から，(X, Y) の分布は $(-X, -Y)$ の分布に等しい。このことから

$$\mathrm{E}[e^{tY^2} I(X > Y)] = \mathrm{E}[e^{t(-Y)^2} I(-X > -Y)] = \mathrm{E}[e^{tY^2} I(X < Y)] = \mathrm{E}[e^{tX^2} I(Y < X)]$$

となる。ただし，最後の等式は X と Y の記号を交換しただけである。従って

$$\begin{aligned} M_{Z^2}(t) =& \mathrm{E}[e^{tX^2} I(X < Y)] + \mathrm{E}[e^{tX^2} I(Y < X)] = \mathrm{E}[e^{tX^2}\{I(X < Y) + I(Y < X)\}] \\ =& \mathrm{E}[e^{tX^2}] = M_{X^2}(t) \end{aligned}$$

となることがわかる。このことは，Z^2 の分布と X^2 の分布が等しいことを示している。

第8章　パラメータの推定

問 1　例 8.2 では，時間間隔のデータを取り上げた。$n = 50$ として指数分布 $Ex(\lambda)$ を当てはめるとする。

$X_1, \ldots, X_n$ が独立に指数分布 $Ex(\lambda)$ に従うとき，λ の MLE は $\hat{\lambda} = 1/\overline{X}$ であるから，推定値は $1/19.334 = 0.052$ になる。

また，この指数分布の平均と分散は $\mathrm{E}[X_1] = 1/\lambda$, $\mathrm{Var}(X_1) = 1/\lambda^2$ であるから，それらの MLE は $\overline{X}$, $\overline{X}^2$ になり，推定値はそれぞれ $\overline{x} = 19.334$, $\overline{x}^2 = 373.8$ になる。

$\mathrm{P}(N_t = k) = \{(\lambda t)^k/k!\}e^{-\lambda t}$ より，$\mathrm{P}(N_t = 0) = e^{-\lambda t}$ であるから，この確率の MLE は $e^{-t/\overline{X}}$ となる。

$t = 24$, $\overline{x} = 19.334$ を代入すると，24 時間以内に消防車が出動しない確率の推定値は 0.29 になる。を与えよ。

```
> exp(-24/19.334)
[1] 0.2889977
```

問 2　例 8.3 では，東京都の 1 日当たりの交通事故による死亡数のデータを取り上げた。$n = 30$ としてポアソン分布 $Po(\lambda)$ を当てはめることにする。

$X_1, \ldots, X_n$ が独立に $Po(\lambda)$ に従うので，λ の MLE は $\hat{\lambda} = \overline{X}$ になる。この推定値は $\overline{x} = 0.3$ となる。

また 1 日に 1 件以上の死亡事故が起こる確率は $\mathrm{P}(X \geq 1) = 1 - \mathrm{P}(X = 0) = 1 - e^{-\lambda}$ となるので，この MLE の推定値は $1 - e^{-0.3} = 0.26$ になる。

```
> 1-exp(-0.3)
[1] 0.2591818
```

問 3　確率変数 $X_1, \ldots, X_n$ が独立にベルヌーイ分布 $Ber(p)$ に従うとする。$\theta = p(1-p)$ とおく。

(1) 同時確率関数は

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

と書けるので，$\overline{X}$ が p の十分統計量になる。$\mathrm{E}[X_i] = p$ より，$\overline{X}$ が p のモーメント推定量である。また，対数尤度と尤度方程式は

$$\ell(p) = n\overline{x}\log p + n(1-\overline{x})\log(1-p), \quad \ell'(p) = \frac{n\overline{x}}{p} - \frac{n(1-\overline{x})}{1-p} = 0$$

と書けるので，尤度方程式の解を求めると p の最尤推定量として $\hat{p}^M = \overline{X}$ が得られる。

(2) 命題 8.10 より θ の MLE は $\hat{\theta} = \overline{X}(1-\overline{X})$ となる。一方，不偏推定量を求めるために次の期

待値を計算してみる。$Y=\sum_{i=1}^n X_i$ とおくと $Y\sim Bin(n,p)$ であることに注意する。

$$\begin{aligned}\mathrm{E}[\overline{X}(1-\overline{X})] =&\frac{1}{n^2}\mathrm{E}[Y(n-Y)] = \frac{1}{n^2}\Big\{-\mathrm{E}[Y(Y-1)]+(n-1)\mathrm{E}[Y]\Big\}\\ =&\frac{1}{n^2}\{-n(n-1)p^2+n(n-1)p\} = \frac{n-1}{n}p(1-p)\end{aligned}$$

と書ける。この式から θ の不偏推定量として

$$\hat{\theta}^U = \frac{n}{n-1}\overline{X}(1-\overline{X})$$

が得られることがわかる。

θ の MLE は標本分散 S^2 に一致し，θ の不偏推定量は不偏分散 V^2 に一致することを示す。X_i は 0 と 1 の値しかとらないので，$X_i^2=X_i$ となることに注意すると

$$\begin{aligned}\sum_{i=1}^n (X_i-\overline{X})^2 =&\sum_{i=1}^n (X_i^2-2X_i\overline{X}+\overline{X}^2) = \sum_{i=1}^n (X_i-2X_i\overline{X}+\overline{X}^2)\\ =&n\overline{X}-2n\overline{X}^2+n\overline{X}^2 = n\overline{X}(1-\overline{X})\end{aligned}$$

となる。これは標本分散が

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

となることを示しているので，θ の MLE が S^2 になることがわかる。一方，不偏推定量については

$$\hat{\theta}^U = \frac{n}{n-1}\overline{X}(1-\overline{X}) = \frac{1}{n-1}\sum_{i=1}^n (X_i-\overline{X})^2 = V^2$$

となって，不偏分散に一致する。

問 4 確率変数 $X_1,\ldots,X_n$ が独立に $\mathcal{N}(\mu,\sigma^2)$ に従うとする。

(1) $\sigma^2=\sigma_0^2$ で既知とするとき，同時確率関数は

$$\frac{1}{(\sqrt{2\pi}\sigma_0)^n}e^{-\sum_{i=1}^n (x_i-\mu)^2/(2\sigma_0^2)} = \frac{1}{(\sqrt{2\pi}\sigma_0)^n}e^{-\sum_{i=1}^n (x_i-\overline{x})^2/(2\sigma_0^2)-n(\overline{x}-\mu)^2/(2\sigma_0^2)}$$

と書けるので，$\overline{X}$ が十分統計量になる。この式から対数尤度は

$$\ell(\mu) = -(n/2)\log(2\pi\sigma_0^2) - \frac{1}{2\sigma_0^2}\sum_{i=1}^n (x_i-\overline{x})^2 - \frac{1}{2\sigma_0^2/n}(\overline{x}-\mu)^2$$

より，

$$\ell'(\mu) = \frac{1}{\sigma_0^2/n}(\overline{x}-\mu) = 0$$

を解く MLE として $\widehat{\mu}^M=\overline{X}$ が導かれる。更に微分するとフィッシャー情報量

$$I_n(\mu) = -\mathrm{E}[\ell''(\mu)] = \frac{n}{\sigma_0^2}$$

が得られるので，μ のすべての不偏推定量 $\widehat{\mu}$ に対してクラメール・ラオの不等式は

$$\mathrm{Var}(\hat{\theta}) \geq \frac{\sigma_0^2}{n}$$

となる。MLE の分散は $\mathrm{Var}(\hat{\theta}^M)=\mathrm{Var}(\overline{X})=\sigma_0^2/n$ より，クラメール・ラオの不等式の下限に達することがわかる。

(2) メディアン $X_{\mathrm{med}}=\mathrm{med}(X_1,\ldots,X_n)$ が μ の不偏推定量であることを示す。$X_{\mathrm{med}}-\mu=\mathrm{med}(X_1-\mu,\ldots,X_n-\mu)$ であり，この期待値が 0 になることを示せばよい。分布の対称性から

$$\mathrm{E}[\mathrm{med}(X_1-\mu,\ldots,X_n-\mu)]=\mathrm{E}[\mathrm{med}(\mu-X_1,\ldots,\mu-X_n)]=-\mathrm{E}[\mathrm{med}(X_1-\mu,\ldots,X_n-\mu)]$$

と書けるので，$\mathrm{E}[\mathrm{med}(X_1-\mu,\ldots,X_n-\mu)]=0$ となることがわかる。すなわち，$\mathrm{E}[X_{\mathrm{med}}]=\mu$ となり，メディアンが不偏推定量になる。クラメール・ラオ不等式より

$$\mathrm{Var}(X_{\mathrm{med}})\geq\frac{\sigma_0^2}{n}=\mathrm{Var}(\hat{\theta}^M)$$

となり，$\mathrm{P}(X_{\mathrm{med}}\neq\hat{\theta}^M)>0$ であるから，$\mathrm{Var}(X_{\mathrm{med}})>\mathrm{Var}(\hat{\theta}^M)$ となる。

(3) $\mu=\mu_0$ で既知とするときには，同時確率関数は

$$\frac{1}{(\sqrt{2\pi})^2}\frac{1}{(\sigma^2)^{n/2}}e^{-\sum_{i=1}^n(x_i-\mu_0)^2/(2\sigma^2)}$$

と書けるので，$\sum_{i=1}^n(X_i-\mu_0)^2$ が σ^2 の十分統計量になる。この式から対数尤度は

$$\ell(\sigma^2)=-(n/2)\log(2\pi\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\mu_0)^2$$

より，

$$\ell'(\sigma^2)=-\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^n(x_i-\mu_0)^2=0$$

を解く MLE として $\hat{\sigma}^{2M}=n^{-1}\sum_{i=1}^n(X_i-\mu_0)^2$ が導かれる。更に微分すると

$$\ell''(\sigma^2)=\frac{n}{2\sigma^4}-\frac{1}{\sigma^6}\sum_{i=1}^n(x_i-\mu_0)^2$$

となり，$\mathrm{E}[\sum_{i=1}^n(X_i-\mu_0)^2]=n\sigma^2$ であるからフィッシャー情報量

$$I_n(\sigma^2)=-\mathrm{E}[\ell''(\sigma^2)]=-\frac{n}{2\sigma^4}+\frac{1}{\sigma^6}n\sigma^2=\frac{n}{2\sigma^4}$$

が得られるので，σ^2 のすべての不偏推定量 $\hat{\sigma}^2$ に対してクラメール・ラオの不等式は

$$\mathrm{Var}(\hat{\sigma}^2)\geq\frac{2\sigma^4}{n}$$

となる。$\hat{\sigma}^{2M}=(\sigma^2/n)Y,\ Y=\sum_{i=1}^n(X_i-\mu_0)^2/\sigma^2\sim\chi_n^2$，とすると，

$$\mathrm{E}[\hat{\sigma}^{2M}]=\frac{\sigma^2}{n}\mathrm{E}[Y]=\frac{\sigma^2}{n}n=\sigma^2$$

となり，$\hat{\sigma}^{2M}$ は σ^2 の不偏推定量である。このとき，その分散は

$$\mathrm{Var}(\hat{\sigma}^{2M})=\frac{\sigma^4}{n^2}\mathrm{E}[(Y-n)^2]=\frac{\sigma^4}{n^2}\mathrm{Var}(Y)=\frac{\sigma^4}{n^2}2n=\frac{2\sigma^4}{n}$$

となり，クラメール・ラオの不等式の下限に達することがわかる。

問 5　$X_1, \ldots, X_n$ が独立で，パレート分布

$$f(x|\alpha,\beta) = \beta\frac{\alpha^\beta}{x^{\beta+1}}, \quad x \geq \alpha$$

に従うとする。ここで，$\alpha > 0$, $\beta > 0$ である。

(1) 尤度関数は

$$\prod_{i=1}^{n} \frac{\beta\alpha^\beta}{x_i^{\beta+1}} I(x_i \geq \alpha) = \frac{\beta^n \alpha^{n\beta}}{(\prod_{i=1}^n x_i)^{\beta+1}} I(x_{(1)} \geq \alpha)$$

と書けるので，$(X_{(1)}, \prod_{i=1}^n X_i)$ が (α, β) の十分統計量である。ただし，$X_{(1)}$ は最小統計量である。

まず，この尤度を最大にする α を求めてみると，$\widehat{\alpha}^M = X_{(1)}$ であることがわかる。対数尤度関数に $\widehat{\alpha}^M = X_{(1)}$ を代入すると，

$$\ell(\beta) = n\log\beta + n\beta\log X_{(1)} - \sum_{i=1}^{n}\log X_i$$

と書けるので，尤度方程式は

$$\ell'(\beta) = \frac{n}{\beta} + n\log X_{(1)} - \sum_{i=1}^{n}\log X_i = 0$$

となり，これを解くと，β の MLE として

$$\widehat{\beta}^M = \frac{n}{\sum_{i=1}^n \log(X_i/X_{(1)})}$$

が得られる。

(2) 以下では，$\alpha = \alpha_0$ として既知とする。$\theta = 1/\beta$，$\alpha = \alpha_0$ とおいて尤度関数を書き直すと

$$\frac{1}{\theta^n}\frac{\alpha_0^{n/\theta}}{(\prod_{i=1}^n x_i)^{1/\theta+1}} I(x_i \geq \alpha_0)$$

と表されるので，対数尤度は

$$\ell(\theta) = -n\log\theta + \frac{n}{\theta}\log\alpha_0 - \Big(\frac{1}{\theta}+1\Big)\sum_{i=1}^{n}\log x_i$$

であり，尤度方程式は

$$\ell'(\theta) = -\frac{n}{\theta} - \frac{n}{\theta^2}\log\alpha_0 + \frac{1}{\theta^2}\sum_{i=1}^{n}\log x_i = 0$$

となり，この解を求めると，θ の MLE は

$$\hat{\theta}^M = \frac{1}{n}\sum_{i=1}^{n}\log(X_i/\alpha_0)$$

として与えられる。

$\hat{\theta}^M$ の期待値を計算するには，$\mathrm{E}[\log(X_i/\alpha_0)]$ を求める必要がある。これを求めるには，スコア関数について $\mathrm{E}[\ell'(\theta)] = 0$ が常に成り立っていることを利用する。すなわち

$$\mathrm{E}[\ell'(\theta)] = \mathrm{E}\Big[-\frac{n}{\theta} - \frac{n}{\theta^2}\log\alpha_0 + \frac{1}{\theta^2}\sum_{i=1}^{n}\log X_i\Big] = 0$$

であることから

$$\mathrm{E}\Big[\sum_{i=1}^{n}\log X_i\Big]=n\theta$$

が得られる。従って

$$\mathrm{E}[\hat{\theta}^M]=\frac{1}{n}\mathrm{E}\Big[\sum_{i=1}^{n}\log(X_i/\alpha_0)\Big]=\frac{1}{n}n\theta=\theta$$

となり，MLE $\hat{\theta}^{\mathrm{M}}$ が θ の不偏推定量になることがわかる。

(3) さらに θ で微分すると

$$\ell''(\theta)=\frac{n}{\theta^2}+\frac{2n}{\theta^3}\log\alpha_0-\frac{2}{\theta^3}\sum_{i=1}^{n}\log x_i$$

と書けるので，θ のフィッシャー情報量は

$$I_n(\theta)=\mathrm{E}[-\ell''(\theta)]=\mathrm{E}\Big[-\frac{n}{\theta^2}+\frac{2}{\theta^3}\sum_{i=1}^{n}\log(X_i/\alpha_0)\Big]=-\frac{n}{\theta^2}+\frac{2n\theta}{\theta^3}=\frac{n}{\theta^2}$$

となる。従って，$\hat{\theta}^M$ の漸近分布は

$$\sqrt{n}(\hat{\theta}^M-\theta)\to_d\mathcal{N}(0,\theta^2)$$

で与えられる。

(4) θ のすべての不偏推定量 $\hat{\theta}$ に対するクラメール・ラオの不等式は

$$\mathrm{Var}(\hat{\theta})\geq\frac{\theta^2}{n}$$

である。θ の MLE $\hat{\theta}^M$ の分散は

$$\mathrm{Var}(\hat{\theta}^M)=\frac{1}{n^2}\mathrm{E}\Big[\Big\{\sum_{i=1}^{n}\log(X_i/\alpha_0)-n\theta\Big\}^2\Big]$$

と表すことができる。この値を直接計算する代わりに，スコア関数を利用する。スコア関数は

$$S(\theta,\boldsymbol{X})=-\frac{n}{\theta}+\frac{1}{\theta^2}\sum_{i=1}^{n}\log(X_i/\alpha_0)=\frac{1}{\theta^2}\Big\{\sum_{i=1}^{n}\log(X_i/\alpha_0)-n\theta\Big\}$$

のように表すことができる。これと $\mathrm{Var}(\hat{\theta}^M)$ とを見比べると

$$\mathrm{Var}(\hat{\theta}^M)=\frac{\theta^4}{n^2}\mathrm{E}[\{S(\theta,\boldsymbol{X})\}^2]$$

と書けることがわかる。$\mathrm{E}[\{S(\theta,\boldsymbol{X})\}^2]$ はフィッシャー情報量 $I_n(\theta)$ に等しいので，

$$\mathrm{E}[\{S(\theta,\boldsymbol{X})\}^2]=\frac{n}{\theta^2}$$

となることから，

$$\mathrm{Var}(\hat{\theta}^M)=\frac{\theta^4}{n^2}\frac{n}{\theta^2}=\frac{\theta^2}{n}$$

となり，下限に達することがわかる。

問 6　確率変数 $X_1, \dots, X_n$ が独立で，両側指数分布

$$f(x|\theta) = \frac{1}{2\theta} \exp\Big\{ -\frac{|x|}{\theta} \Big\}$$

に従うとする。

(1) 尤度関数は

$$\prod_{i=1}^{n} \frac{1}{2\theta} e^{-|x_i|/\theta} = \frac{1}{(2\theta)^2} e^{-\sum_{i=1}^{n}|x_i|/\theta}$$

であるから，$\sum_{i=1}^{n} |X_i|$ が θ の十分統計量になる。

$$\mathrm{E}[|X_1|] = \int_{-\infty}^{\infty} |x| \frac{1}{2\theta} e^{-x/\theta} dx = \int_0^{\infty} \frac{x}{\theta} e^{-x/\theta} dx = \Gamma(2)\theta \int_0^{\infty} \frac{1}{\Gamma(2)} \frac{1}{\theta^2} x^{2-1} e^{-x/\theta} dx = \theta$$

となるので，θ のモーメント推定量は $n^{-1}\sum_{i=1}^{n} |X_i|$ となる。

対数尤度は

$$\ell(\theta) = -n\log(2\theta) - \frac{1}{\theta}\sum_{i=1}^{n} |x_i|$$

であるから，尤度方程式は

$$\ell'(\theta) = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} |x_i| = 0$$

となる。この解を求めると，$\hat{\theta}^M = n^{-1}\sum_{i=1}^{n} |X_i|$ となり，MLE とモーメント推定量が一致する。

(2) さらに θ で微分すると

$$\ell''(\theta) = \frac{n}{\theta^2} - 2\theta^3 \sum_{i=1}^{n} |x_i|$$

となるので，θ のフィッシャー情報量は

$$I_n(\theta) = -\frac{n}{\theta^2} + \frac{2}{\theta^2}\mathrm{E}\Big[\sum_{i=1}^{n} |X_i|\Big] = -\frac{n}{\theta^2} + \frac{2}{\theta^2} n\theta = \frac{n}{\theta^2}$$

で与えられる。従って，漸近分布は

$$\sqrt{n}(\hat{\theta}^M - \theta) \to_d \mathcal{N}(0, \theta^2)$$

となる。

(3) θ のすべての不偏推定量 $\hat{\theta}$ についてのクラメール・ラオの不等式は

$$\mathrm{Var}(\hat{\theta}) \geq \frac{\theta^2}{n}$$

となる。$\mathrm{E}[\hat{\theta}^M] = \theta$ となることが容易に確かめられるので，$\hat{\theta}^M$ は θ の不偏推定量である。

$$\ell'(\theta) = \frac{n}{\theta^2}(\hat{\theta}^M - \theta)$$

と表されるので，

$$\mathrm{Var}(\hat{\theta}^M) = \frac{\theta^4}{n^2}\mathrm{E}[\{\ell'(\theta)\}^2]$$

と書かれる。ここで，

$$\mathrm{E}[\{\ell'(\theta)\}^2] = I_n(\theta) = \frac{n}{\theta^2}$$

であるから

$$\mathrm{Var}(\hat{\theta}^M) = \frac{\theta^4}{n^2}\frac{n}{\theta^2} = \frac{\theta^2}{n}$$

となり，$\hat{\theta}^{\mathrm{M}}$ がクラメール・ラオ不等式の下限に達することがわかる。

問 7　(X_1, X_2, X_3) を 3 項分布 $Mult(n, p_1, p_2, p_3)$ に従う確率変数とする。ただし，p_1, p_2, p_3 は，$p_1 = 1-\theta$, $p_2 = 1-\theta$, $p_3 = 2\theta - 1$ で与えられていて，θ は $1/2 < \theta < 1$ を満たす未知母数とする。$Y_1 = X_1 + X_3$, $Y_2 = X_2 + X_3$ として以下の問に答えよ。

(1) $\theta = p_1 + p_3$ より，$Y_1 \sim Bin(n, \theta)$ に従うので，$\mathrm{E}[Y_1] = n\theta$, $\mathrm{Var}(Y_1) = n\theta(1-\theta)$ となる。

(2) Y_1 と Y_2 の共分散 $\mathrm{Cov}(Y_1, Y_2)$ を求める。まず, $\mathrm{Cov}(Y_1, Y_2) = \mathrm{E}[Y_1Y_2] - \mathrm{E}[Y_1]\mathrm{E}[Y_2]$ であり,

$$\mathrm{E}[Y_1Y_2] = \mathrm{E}[(X_1 + X_3)(X_2 + X_3)] = \mathrm{E}[X_1X_2] + \mathrm{E}[X_1X_3] + \mathrm{E}[X_2X_3] + \mathrm{E}[X_3^2]$$

と書ける。ここで

$$\mathrm{E}[X_1X_2] = \mathrm{Cov}(X_1, X_2) + \mathrm{E}[X_1]\mathrm{E}[X_2] = -np_1p_2 + n^2p_1p_2 = n(n-1)p_1p_2$$

となり，同様にして $\mathrm{E}[X_1X_3] = n(n-1)p_1p_3$, $\mathrm{E}[X_2X_3] = n(n-1)p_2p_3$ となる。また，$X_3 \sim Bin(n, p_3)$ より，$\mathrm{E}[X_3^2] = \mathrm{Var}(X_3) - (\mathrm{E}[X_3])^2 = np_3(1-p_3) + n^2p_3^2 = np_3 + n(n-1)p_3^2$ となる。従って

$$\begin{aligned}\mathrm{E}[Y_1Y_2] =& n(n-1)\{p_1p_2 + p_1p_3 + p_2p_3 + p_3^2\} + np_3 \\ =& n(n-1)(p_1 + p_3)(p_2 + p_3) + np_3 = n(n-1)\theta^2 + n(2\theta - 1)\end{aligned}$$

と書けることがわかる。以上より

$$\mathrm{Cov}(Y_1, Y_2) = n(n-1)\theta^2 + n(2\theta - 1) - n^2\theta^2 = -n(1-\theta)^2$$

となる。

(3) $\hat{\theta}^U = (Y_1 + Y_2)/(2n)$ とおくとき，$\mathrm{E}[Y_1] = n\theta$ より，$\mathrm{E}[\hat{\theta}^U] = \theta$ となり，θ の不偏推定量であることがわかる。

$$\begin{aligned}\mathrm{Var}(\hat{\theta}^U) =& \frac{1}{4n^2}\mathrm{E}[(Y_1 + Y_2 - 2n\theta)^2] = \frac{1}{4n^2}\{\mathrm{Var}(Y_1) + \mathrm{Var}(Y_2) + 2\mathrm{Cov}(Y_1, Y_2)\} \\ =& \frac{1}{4n^2}\{2n\theta(1-\theta) - 2n(1-\theta)^2\} = \frac{1}{2n}(1-\theta)(2\theta - 1)\end{aligned}$$

となる。一方，Y_1/n も θ の不偏推定量であり,

$$\mathrm{Var}(Y_1/n) = \frac{1}{n^2}\mathrm{Var}(Y_1) = \frac{n\theta(1-\theta)}{n^2} = \frac{\theta(1-\theta)}{n}$$

となる。この両者の分散を比較すると

$$\mathrm{Var}(Y_1/n) - \mathrm{Var}(\hat{\theta}^U) = \frac{1-\theta}{2n} > 0$$

となり，$\hat{\theta}^U$ は Y_1/n より常に分散が小さくなる。

問 8　確率変数 $X_1, \ldots, X_n$ が独立で，指数分布

$$f(x|\theta) = e^{-(x-\theta)}, \quad x > \theta$$

に従うとする。

(1) 尤度関数が

$$\prod_{i=1}^{n} e^{-(x_i-\theta)} I(X_i > \theta) = e^{-\sum_{i=1}^{n}(x_i-\theta)} I(x_{(1)} > \theta) = e^{-\sum_{i=1}^{n}(x_i-x_{(1)})-n(x_{(1)}-\theta)} I(x_{(1)} > \theta)$$

と書けるので，$X_{(1)}$ が θ の十分統計量である。また，尤度関数を最大にする θ を考えると，θ の MLE は $\hat{\theta}^M = X_{(1)}$ になる。

(2) $X_{(1)}$ の確率密度関数は, $F(x) = \int_\theta^x e^{-(t-\theta)}dt = 1 - e^{-(x-\theta)}$ より

$$f_{X_{(1)}}(x) = ne^{-(x-\theta)}\Big\{e^{-(x-\theta)}\Big\}^{n-1} = ne^{-n(x-\theta)},\ x > \theta$$

で与えられるので，変数変換 $y = x - \theta$ により

$$\begin{aligned}\mathrm{E}[X_{(1)}] =& n\int_\theta^\infty xe^{-n(x-\theta)}dx = n\int_0^\infty (y+\theta)e^{-ny}dy \\ =& n\frac{1}{n^2}\int_0^\infty \frac{1}{\Gamma(2)}y^{2-1}n^2e^{-ny}dy + n\theta\Big[-\frac{1}{n}e^{-ny}\Big]_0^\infty = \frac{1}{n}+\theta\end{aligned}$$

となる。よって，$X_{(1)}$ に基づいた θ の不偏推定量は

$$\hat{\theta}^U = X_{(1)} - \frac{1}{n}$$

となる。

(3) 両者の MSE を計算する。$\hat{\theta}^M$ の MSE は

$$\begin{aligned}\mathrm{MSE}(\hat{\theta}^M) =& \mathrm{E}[(X_{(1)}-\theta)^2] = \int_\theta^\infty (x-\theta)^2 ne^{-n(x-\theta)}dx = \int_0^\infty y^2ne^{-ny}dy \\ =& \frac{\Gamma(3)}{n^2}\int_0^\infty \frac{1}{\Gamma(3)}n^3y^{3-1}e^{-ny}dy\frac{2}{n^2}\end{aligned}$$

となり，$\hat{\theta}^U$ の MSE は

$$\begin{aligned}\mathrm{MSE}(\hat{\theta}^U) =& \mathrm{E}\Big[\Big(X_{(1)}-\theta-\frac{1}{n}\Big)^2\Big] = \mathrm{E}[(X_{(1)}-\theta)^2] - \frac{2}{n}\mathrm{E}[X_{(1)}-\theta] + \frac{1}{n^2} \\ =& \frac{2}{n^2} - \frac{2}{n^2} + \frac{1}{n^2} = \frac{1}{n^2}\end{aligned}$$

となる。$\mathrm{MSE}(\hat{\theta}^M) > \mathrm{MSE}(\hat{\theta}^U)$ より，$\hat{\theta}^U$ の MSE の方が小さいことがわかる。

(4) $c > 0$ に対して

$$\begin{aligned}\mathrm{P}(|X_{(1)}-\theta| > c)\mathrm{P}(X_{(1)}-\theta > c) =& \int_{c+\theta}^\infty ne^{-n(x-\theta)}dx \\ =& \int_c^\infty ne^{-ny}dy = [-e^{-ny}]_c^\infty = e^{-nc} \to 0\end{aligned}$$

となり，$X_{(1)} \to_p \theta$ が示される。この式において $c = x/n$ とおくと

$$\mathrm{P}(n(X_{(1)}-\theta) > x) = \mathrm{P}\Big(X_{(1)}-\theta > \frac{x}{n}\Big) = e^{-n(x/n)} = e^{-x}$$

と書けることがわかる。このことは

$$n(X_{(1)}-\theta) \sim Ex(1)$$

に従うことを意味する。正確に指数分布に従うので，n が大きくても同じ分布になる。

問 9 確率変数 $X_1, \dots, X_n$ が独立に確率密度関数

$$f(x|\theta) = \frac{\theta}{x^2}, \quad 0 < \theta < x < \infty$$

に従うとする。ただし，$n \geq 3$ とする。

(1) X_1 の平均については

$$\mathrm{E}[X_1] = \int_\theta^\infty \frac{\theta}{x} dx = \theta[\log x]_\theta^\infty \to \infty$$

となり，平均が存在しない。

(2) 尤度関数は

$$\prod_{i=1}^n \frac{\theta}{x_i^2} I(x_i > \theta) = \frac{\theta^n}{\prod_{i=1}^n x_i^2} I(x_{(1)} > \theta)$$

と書けるので，$X_{(1)}$ が θ の十分統計量になる。また尤度関数を最大にする θ を考えると，θ の MLE は $\hat{\theta}^M = X_{(1)}$ になることがわかる。

$X_{(1)}$ の確率密度は，$F(x) = \int_\theta^x \theta/t^2 dt = 1 - \theta/x$ より

$$f_{X_{(1)}}(x) = n\frac{\theta}{x^2}\Big(\frac{\theta}{x}\Big)^{n-1} = \frac{n\theta^2}{x^{n+1}},\ x > \theta$$

と書けるので，

$$\begin{aligned}\mathrm{E}[X_{(1)}] =& \int_\theta^\infty \frac{n\theta^n}{x^n} dx = n\theta^n\Big[-\frac{1}{n-1}\frac{1}{x^{n-1}}\Big]_\theta^\infty \\ =& n\theta^2\frac{1}{n-1}\frac{1}{\theta^{n-1}} = \frac{n}{n-1}\theta\end{aligned}$$

となる。これより，θ の不偏推定量として

$$\hat{\theta}^U = \frac{n-1}{n} X_{(1)}$$

が得られる。

(3) まず，

$$\mathrm{E}[X_{(1)}^2] = \int_\theta^\infty \frac{n\theta^n}{x^{n-1}} dx = n\theta^n\Big[-\frac{1}{n-2}\frac{1}{x^{n-2}}\Big]_\theta^\infty = \frac{n}{n-2}\theta^2$$

となることに注意する。$\hat{\theta}^M$ の MSE は

$$\begin{aligned}\mathrm{MSE}(\hat{\theta}^M) =& \mathrm{E}[(X_{(1)} - \theta)^2] = \mathrm{E}[X_{(1)}^2 - 2\theta X_{(1)} + \theta^2] \\ =& \frac{n}{n-2}\theta^2 - 2\theta\frac{n}{n-1}\theta + \theta^2 = \frac{2\theta^2}{(n-1)(n-2)}\end{aligned}$$

であり，$\hat{\theta}^U$ の MSE は

$$\mathrm{MSE}(\hat{\theta}^U) = \frac{(n-1)^2}{n^2}\mathrm{E}[X_{(1)}^2] - \theta^2 = \frac{(n-1)^2}{n^2}\frac{n}{n-2}\theta^2 - \theta^2 = \frac{\theta^2}{n(n-2)}$$

である。両者の MSE を比較すると，$\mathrm{MSE}(\hat{\theta}^U) < \mathrm{MSE}(\hat{\theta}^M)$ となり，不偏推定量の MSE の方が小さい。

(4) $c > 0, \theta > 0$ に対して，変数変換 $y = x/\theta$ を用いると

$$\mathrm{P}(|X_{(1)} - \theta| > c\theta) = \mathrm{P}(X_{(1)} - \theta > c\theta) = \int_{\theta+c\theta}^{\infty} \frac{n\theta^n}{x^{n+1}} dx = \int_{1+c}^{\infty} \frac{n}{y^{n+1}} dy = \frac{1}{(1+c)^n} \to 0$$

となり，$X_{(1)} \to_p \theta$ に確率収束することがわかる。
　また，上の式で $c\theta = x/n$ とおくと

$$\mathrm{P}(n(X_{(1)} - \theta) > x) = \mathrm{P}\Big(X_{(1)} - \theta > \frac{x}{n}\Big) = \frac{1}{(1 + x/(n\theta))^n} \to e^{-x/\theta}$$

となる。従って

$$n(\hat{\theta}^{\mathrm{M}} - \theta) \to_d Ex(1/\theta)$$

に収束する。

問 10　$X_1, \ldots, X_n, \ldots$ が $E[X_i] = \mu$, $\mathrm{Var}(X_i) = \sigma^2$ を満たす確率変数の列とし，$i \neq j$ に対して $\mathrm{Cov}(X_i, X_j) = \rho_{ij}\sigma^2$ であるとする。$\boldsymbol{X} = (X_1, \ldots, X_n)^\top$, $\mathbf{1}_n = (1, \ldots, 1)^\top$ とおくと，$\overline{X} = n^{-1}\mathbf{1}_n^\top \boldsymbol{X}$ であり，この分散は

$$\begin{aligned}\mathrm{Var}(\overline{X}) =& \mathrm{E}[(n^{-1}\mathbf{1}_n^\top \boldsymbol{X} - \mu)^2] = \mathrm{E}[(n^{-1}\mathbf{1}_n^\top \boldsymbol{X} - n^{-1}\mathbf{1}_n^\top \mathbf{1}_n \mu)^2] \\ =& \frac{1}{n^2}\mathrm{E}[\mathbf{1}_n^\top(\boldsymbol{X} - \mu\mathbf{1}_n)\mathbf{1}_n^\top(\boldsymbol{X} - \mu\mathbf{1}_n)] = \frac{1}{n^2}\mathrm{E}[\mathbf{1}_n^\top(\boldsymbol{X} - \mu\mathbf{1}_n)(\boldsymbol{X} - \mu\mathbf{1}_n)^\top\mathbf{1}_n] = \frac{1}{n^2}\mathbf{1}_n^\top \boldsymbol{\Sigma} \mathbf{1}_n\end{aligned}$$

と書ける。ただし, $\boldsymbol{\Sigma}$ は $\boldsymbol{X}$ の共分散行列で $\boldsymbol{\Sigma} = (\rho_{ij}\sigma^2)$，$\rho_{ii} = 1$ である。
　(a) ρ_{ij} が，$\rho_{i,i+1} = b > 0$ で，$|i - j| \geq 2$ に対して $\rho_{ij} = 0$ であるとき，

$$\boldsymbol{\Sigma} = \sigma^2 \begin{pmatrix} 1 & b & 0 & 0 & \cdots & 0 \\ b & 1 & b & 0 & \cdots & 0 \\ 0 & b & 1 & b & \cdots & 0 \\ 0 & 0 & b & 1 & \cdots & 0 \\ \vdots & \vdots & & & \ddots & \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

と書けるので，

$$\mathrm{Var}(\overline{X}) = \frac{1}{n^2}\mathbf{1}_n^\top \boldsymbol{\Sigma} \mathbf{1}_n = \frac{\sigma^2}{n^2}(1+b, 1+2b, \ldots, 1+2b, 1+b)\mathbf{1}_n = \frac{\sigma^2}{n^2}(n + 2(n-1)b)$$

となる。b が定数なので，$\mathrm{Var}(\overline{X}) \to 0$ となり，$\overline{X}$ は μ の一致推定量である。
　(b) $\rho_{ij} = \rho^{|i-j|}$, $0 < \rho < 1$，の場合は

$$\boldsymbol{\Sigma} = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \rho^2 & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \rho & \cdots & \rho^{n-3} \\ \rho^3 & \rho^2 & \rho & 1 & \cdots & \rho^{n-4} \\ \vdots & \vdots & & & \ddots & \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \rho^{n-4} & \cdots & 1 \end{pmatrix}$$

と書けるので，

$$\mathrm{Var}(\overline{X}) = \frac{1}{n^2}\mathbf{1}_n^\top \boldsymbol{\Sigma} \mathbf{1}_n = \frac{\sigma^2}{n^2}\Big\{n + 2\sum_{i=1}^{n-1}\sum_{j=1}^{n-i}\rho^j\Big\}$$

となる。ここで

$$\sum_{i=1}^{n-1}\sum_{j=1}^{n-i}\rho^j = \sum_{i=1}^{n-1}\frac{\rho-\rho^{n-i+1}}{1-\rho} = \frac{(n-1)\rho}{1-\rho} - \frac{\rho}{1-\rho}\sum_{i=1}^{n-1}\rho^{n-i}$$
$$= \frac{(n-1)\rho}{1-\rho} - \frac{\rho}{1-\rho}\frac{\rho-\rho^n}{1-\rho} = \frac{\rho(n-1-n\rho+\rho^n)}{(1-\rho)^2}$$

となることから，

$$\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n^2}\left\{n + 2\frac{\rho(n-1-n\rho+\rho^n)}{(1-\rho)^2}\right\} = \frac{\sigma^2}{n^2(1-\rho)^2}\{n(1-\rho^2) - 2\rho(1-\rho^n)\}$$

となり，$|\rho| < 1$ より $\mathrm{Var}(\overline{X}) \to 0$ となる。

問 11 $\mathrm{E}[X_1] = \alpha\beta$, $\mathrm{Var}(X_1) = \alpha\beta^2$ より，$\alpha\beta = \overline{X}$, $\alpha\beta^2 = S^2$ の解がモーメント推定量になるので，

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

となる。

対数尤度は

$$\ell(\alpha,\beta) = -n\log\Gamma(\alpha) - n\alpha\log\beta + (\alpha-1)\sum_{i=1}^{n}\log X_i - \frac{n}{\beta}\overline{X}$$

と書けるので，MLE を求めるための尤度方程式は

$$\frac{\partial\ell}{\partial\alpha} = -n\psi(\alpha) - n\log\beta + \sum_{i=1}^{n}\log X_i = 0$$
$$\frac{\partial\ell}{\partial\beta} = -n\frac{\alpha}{\beta} + \frac{n}{\beta^2}\overline{X} = 0$$

で与えられる。ただし，

$$\psi(\alpha) = \frac{\partial}{\partial\alpha}\log\Gamma(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$$

であり，ダイガンマ関数とよばれる。$n=1$ のときスコア関数は

$$\frac{\partial\ell}{\partial\alpha} = -\psi(\alpha) - \log\beta + \log X_1 = 0$$

であり，この期待値は 0 になるので，$\mathrm{E}[\log X_1] = \psi(\alpha) + \log\beta$ となる。

問 12 確率変数 $X_1,\ldots,X_n$ が独立で，両側指数分布

$$f(x|\theta) = \frac{1}{2}e^{-|x-\theta|}, \quad -\infty < x < \infty$$

に従うとする。尤度関数，対数尤度関数は

$$L(\theta) = \frac{1}{2^n}e^{-\sum_{i=1}^{n}|x_i-\theta|}, \quad \ell(\theta) = -n\log 2 - \sum_{i=1}^{n}|x_i-\theta|$$

と書けるので，θ の最尤推定量は $\sum_{i=1}^{n}|x_i-\theta|$ を最小にする解で与えられる。この解はメディアンになる。いま，n が奇数，すなわち $n = 2m+1$ の場合を考えるので，θ の MLE は $\hat{\theta}^M = X_{(m+1)}$ になる。

$X_{(m+1)}$ が θ に確率収束することは，7 章の問 13 と同様にして示すことができる。まず，$X_1 - \theta, \ldots, X_n - \theta$ のメディアンは $X_{(m+1)} - \theta$ になるので，一般性を失うことなく $\theta = 0$ としてよい。$c > 0$ を任意にとると，両側指数分布 $2^{-1}\exp(-|x|)$ の分布関数を $F(x)$ とするとこれは 0 に関して対称な分布なので，$F(c) > 1/2$ となることに注意する。また，$X_i > c$ となる X_i の個数を S_n とすると

$$X_{(m+1)} > c \Longleftrightarrow S_n \geq m + 1 = \frac{n+1}{2}$$

となることがわかる。$p = 1 - F(c)$ に対して $S_n \sim Bin(n, p)$ に従い，$n^{-1}S_n \to_p p$ に確率収束することに注意する。このとき

$$\mathrm{P}(X_{(m+1)} > c) = \mathrm{P}\Big(S_n \geq \frac{n+1}{2}\Big) = \mathrm{P}\Big(\frac{S_n}{n} - p \geq \frac{1}{2} - p + \frac{1}{2n}\Big)$$

と書ける。$p < 1/2$ であることに注意すると

$$\mathrm{P}\Big(\frac{S_n}{n} - p \geq \frac{1}{2} - p + \frac{1}{2n}\Big) \leq \mathrm{P}\Big(\Big|\frac{S_n}{n} - p\Big| \geq \frac{1}{2} - p\Big) \to 0$$

となる。以上から，$X_{(m+1)}$ が θ に確率収束することが示される。

第9章　仮説検定と信頼区間

問 1　例 8.1 では男子大学生 50 人の身長のデータを取り上げ，正規分布 $\mathcal{N}(\mu, \sigma^2)$ を当てはめた。20 歳男性の身長の全国平均が $\mu_0 = 170.4$ であるとき，この集団の学生の身長が全国平均と等しいかを調べたい。仮説検定 $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$ について有意水準 5% で検定する。$\overline{x} = 172$, $v = \sqrt{18.0} = 4.24$, $n = 50$ より

$$\frac{\sqrt{n}|\overline{x} - \mu_0|}{v} = 2.67 > t_{49,0.025} = 2.01$$

となり，有意となる。

　また μ の信頼係数 95% の信頼区間は

$$\overline{x} \pm \frac{v}{\sqrt{n}} t_{49,0.025} = [170.8, 173.2]$$

となる。

```
> sqrt(50)*abs(172-170.4)/4.24
[1] 2.668327
> qt(0.975, 49)
[1] 2.009575
> 172-(4.24/sqrt(50))*qt(0.975,49)
[1] 170.795
> 172+(4.24/sqrt(50))*qt(0.975,49)
[1] 173.205
```

問 2　例 8.2 では，時間間隔のデータを取り上げ，$n = 50$ として指数分布 $Ex(1/\theta)$ を当てはめた。消防車の出動が起きる時間間隔が 24 時間程度か否かを検定するため，$\theta_0 = 24$ に対して仮説検定 $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$ を考える。有意水準 5% で検定する。$\theta = 1/\lambda$ とおくと，$\mathrm{E}[X] = \theta$, $\mathrm{Var}(X) = \theta^2$ より，

$$\sqrt{n}(\hat{\theta} - \theta)/\hat{\theta} \to_d \mathcal{N}(0, 1)$$

で近似できることを使う。$\hat{\theta} = 19.334$ より

$$\sqrt{n}|\hat{\theta} - \theta_0|/\hat{\theta} = \sqrt{50}|19.334 - 24|/19.334 = 1.76 < z_{0.025} = 1.96$$

より，$H_0 : \theta = 24$ は有意でない。

　θ の信頼係数 95% の信頼区間は

$$\hat{\theta} \pm \frac{\hat{\theta}}{\sqrt{n}} z_{0.025} = [13.97, 24.69]$$

となる。

```
> sqrt(50)*abs(19.334-24)/19.334
[1] 1.706507
> qnorm(0.975)
[1] 1.959964
> 19.334-(19.334/sqrt(50))*qnorm(0.975)
[1] 13.97499
> 19.334+(19.334/sqrt(50))*qnorm(0.975)
[1] 24.69301
```

問3 例8.3では，東京都の1日当たりの交通事故による死亡数のデータを取り上げ，$n = 30$としてポアソン分布$Po(\lambda)$を当てはめた。1日1件の死亡事故が起こるか否かの検定$H_0 : \lambda = 1$ vs. $H_1 : \lambda \neq 1$について有意水準5%で検定する。

$\sqrt{n}(\overline{X} - \lambda)/\sqrt{\overline{X}} \to_d \mathcal{N}(0, 1)$であるから

$$\sqrt{n}\frac{|\overline{X} - 1|}{\sqrt{\overline{X}}} = \sqrt{30}\frac{|0.3 - 1|}{\sqrt{0.3}} = 7 > z_{0.025} = 1.96$$

より有意となる。

また，λの信頼係数95%の信頼区間は

$$\hat{\lambda} \pm \frac{\sqrt{\hat{\lambda}}}{\sqrt{n}} z_{0.025} = [0.104, 0.496]$$

となる。

```
> sqrt(30)*abs(0.3-1)/sqrt(0.3)
[1] 7
> 0.3-(sqrt(0.3)/sqrt(30))*qnorm(0.975)
[1] 0.1040036
> 0.3+(sqrt(0.3)/sqrt(30))*qnorm(0.975)
[1] 0.4959964
```

問4　あるハンバーガーショップで販売されているフライドポテトについて，2つの店舗A, BでMサイズの重さを調べてみた。店舗Aでは，$m = 10$個のフライドポテトをランダムに購入した結果，重さの標本平均は$\overline{x}_A = 132(\mathrm{g})$, 標本分散は$s_A^2 = 8^2$であった。同様にして，店舗Bでは，$n = 15$個について，$\overline{x}_B = 135$, $s_B = 10^2$であった。店舗Aの母集団平均と分散をμ_A, σ_A^2とし，店舗Bの母集団平均と分散をμ_B, σ_B^2とする。

(1) $\sigma_A^2 = \sigma_B^2 = \sigma^2$という設定のもとで仮説検定$H_0 : \mu_A = \mu_B$ vs. $H_1 : \mu_A \neq \mu_B$を考える。有意水準5%で検定する。

$\sigma_A^2 = \sigma_B^2 = \sigma^2$より，$\sigma^2$のプールされた推定値は

$$\hat{\sigma}^2 = \frac{ms_A^2 + ns_B^2}{m + n - 2} = \frac{10 \times 8^2 + 15 \times 10^2}{10 + 15 - 2} = 9.65$$

となる。

$$\overline{X}_A - \overline{X}_B \sim \mathcal{N}\Big(\mu_A - \mu_B,\ \Big(\frac{1}{m} + \frac{1}{n}\Big)\sigma^2\Big)$$

より

$$\frac{(\overline{X}_A - \mu_A) - (\overline{X}_B - \mu_B)}{\sqrt{m^{-1} + n^{-1}}\hat{\sigma}} \sim t_{m+n-2}$$

に従う。$H_0 : \mu_A = \mu_B$ vs. $H_1 : \mu_A \neq \mu_B$ の検定については

$$\frac{|\overline{x}_A - \overline{x}_B|}{\sqrt{m^{-1}+n^{-1}}\hat{\sigma}} = 0.76 < t_{23,0.025} = 2.07$$

となり，有意でない。

(2) 同じ設定のもとで，$\mu_A - \mu_B$ の信頼係数 95% の信頼区間を求める。

$$\overline{x}_A - \overline{x}_B \pm \sqrt{m^{-1}+n^{-1}}\hat{\sigma} t_{m+n-2,\alpha/2} = [-11.15, 5.15]$$

となる。

(3) σ_A^2 と σ_B^2 が等しいという仮定を外した場合，(1) の仮説検定を有意水準 5% で行う。この場合，ベーレンス・フィッシャー問題が起こるので，ウェルチの t 検定を用いる。

$$T_W = \frac{\overline{x}_A - \overline{x}_B}{\sqrt{s_A^2/(m-1) + s_B^2/(n-1)}} = -0.795$$

となり，自由度は

$$\nu = \Big[\frac{\{s_A^2/(m-1)+s_B^2/(n-1)\}^2}{(s_A^2/(m-1))^2/(m-1) + (s_B^2/(n-1))^2/(n-1)}\Big] = [21.93] = 21$$

となるので，$t_{21,0.025} = 2.08$ となり，有意でない。

問 5 $X_1, \ldots, X_n$ をベルヌーイ分布 $Ber(p)$ からのランダム標本とする。

(1) 信頼係数 99% の p の近似的な信頼区間を与える。$\sqrt{n}(\overline{X}-p) \to_d \mathcal{N}(0, p(1-p))$ より，

$$\overline{X} \pm \frac{\sqrt{\overline{X}(1-\overline{X})}}{\sqrt{n}} z_{\alpha/2},\ \alpha = 0.01$$

となる。

(2) 与えられた精度で p を推定するために必要なサンプルサイズ n を求める問題を考える。c を正の定数とし，99% の確率で $|\overline{X}-p| \le c$ を満たすように n を定める。

$$\mathrm{P}(|\overline{X}-p| \le c) = \mathrm{P}\Big(\frac{\sqrt{n}|\overline{X}-p|}{\sqrt{p(1-p)}} \le \frac{\sqrt{n}c}{\sqrt{p(1-p)}}\Big) \approx \mathrm{P}(|Z| \le \sqrt{n}c/\sqrt{p(1-p)})$$

と書ける。ただし $Z \sim \mathcal{N}(0,1)$ である。従って，

$$\frac{\sqrt{n}c}{\sqrt{p(1-p)}} \ge z_{\alpha/2},\ \text{すなわち}\ n \ge \Big(\frac{z_{\alpha/2}}{c}\Big)^2 p(1-p)$$

を満たすような n をとれば，$\mathrm{P}(|\overline{X}-p| \le c) = 1-\alpha$ が成り立つ。ここで，$p(1-p) \le 1/4$ であるから

$$n \ge \Big(\frac{z_{\alpha/2}}{c}\Big)^2 \frac{1}{4} = \Big(\frac{z_{\alpha/2}}{2c}\Big)^2$$

を満たすように n をとれれば，$1-\alpha$ の確率で $|\overline{X}-p| \le c$ が成り立つことになる。

$c = 0.04$，$\alpha = 0.01$ のときには

$$n \ge \Big(\frac{z_{\alpha/2}}{2c}\Big)^2 = 1036.7$$

となるので，$n \ge 1037$ となる。

問 6 確率変数 $(X_1, X_2, X_3, X_4, X_5)$ が多項分布 $Mult_5(n, p_1, p_2, p_3, p_4, p_5)$ に従うとする。$X_1 + \cdots + X_5 = n$, $p_1 + \cdots + p_5 = 1$ を満たす。

(1) 仮説検定 $H_0: p_1 = 1/2$ vs. $H_1: p_1 \neq 1/2$ について，有意水準 α の尤度比検定とワルド型検定を与える。$X_1 \sim Bin(n, p_1)$ であり，$\hat{p}_1 = X_1/n$ である。尤度比は, $p_{10} = 1/2$ に対して

$$\Lambda = \frac{L(p_{10})}{L(\hat{p}_1)} = \frac{p_{10}^{x_1}(1-p_{10})^{n-x_1}}{\hat{p}_1^{x_1}(1-\hat{p}_1)^{n-x_1}} = \frac{(1/2)^{x_1}(1/2)^{n-x_1}}{(x_1/n)^{x_1}(1-x_1/n)^{n-x_1}} = \frac{(n/2)^n}{x_1^{x_1}(n-x_1)^{n-x_1}}$$

と書けるので，尤度比検定は

$$-2\log\Lambda = -2n\log(n/2) + 2x_1\log(x_1) + 2(n-x_1)\log(n-x_1) > \chi_{1,\alpha} \Longrightarrow H_0\text{を棄却}$$

となる。また,

$$\sqrt{n}(X_1/n - p_1) \to_d \mathcal{N}(0, p_1(1-p_1))$$

より，ワルド検定は

$$\frac{\sqrt{n}|X_1/n - 1/2|}{\sqrt{(X_1/n)(1-X_1/n)}} = \frac{\sqrt{n}|x_1 - n/2|}{\sqrt{x_1(n-x_1)}} > z_{\alpha/2} \Longrightarrow H_0\text{を棄却}$$

で与えられる。

(2) ワルド型検定に基づいて，信頼係数 $1-\alpha$ の p_1 の信頼区間は

$$\frac{X_1}{n} \pm \frac{\sqrt{X_1(n-X_1)}}{n\sqrt{n}} z_{\alpha/2}$$

となる。

(3) 仮説検定 $H_0: p_1 = p_2 = p_3$ vs. $H_1:$「$p_1 = p_2 = p_3$でない」について，有意水準 α の尤度比検定の棄却域を求める。

p_i の MLE は $\hat{p}_i = X_i/n$ である。一方，$H_0: p_1 = p_2 = p_3 = p$ のもとでは，$X = X_1 + X_2 + X_3$ とおくと，$X + X_4 + X_5 = n$, $p + p_4 + p_5 = 1$ であり，$(X, X_4, X_5) \sim Mult_3(n, p, p_4, p_5)$ に従い，$\hat{p} = X/n$ である。従って尤度比は

$$\Lambda = \frac{L(\hat{p}, \hat{p}_4, \hat{p}_5)}{L(\hat{p}_1, \ldots, \hat{p}_5)} = \frac{\frac{n!}{x!x_4!x_5!}\hat{p}^x\hat{p}_4^{x_4}\hat{p}_5^{x_5}}{\frac{n!}{x_1!x_2!x_3!x_4!x_5!}\hat{p}_1^{x_1}\hat{p}_2^{x_2}\hat{p}_3^{x_3}\hat{p}_4^{x_4}\hat{p}_5^{x_5}} = \Big[\frac{x!}{x_1!x_2!x_3!}\Big(\frac{\hat{p}_1}{\hat{p}}\Big)^{x_1}\Big(\frac{\hat{p}_2}{\hat{p}}\Big)^{x_2}\Big(\frac{\hat{p}_3}{\hat{p}}\Big)^{x_3}\Big]^{-1}$$

と書けるので，尤度比検定は

$$-2\log\Lambda = 2\log\frac{(x_1+x_2+x_3)!}{x_1!x_2!x_3!} + 2\sum_{i=1}^{3} x_i \log\frac{x_i}{x_1+x_2+x_3} > \chi^2_{2,\alpha} \Longrightarrow H_0\text{を棄却}$$

となる。

問 7　確率変数 $X_1, \ldots, X_n$ 及び $Y_1, \ldots, Y_n$ が互いに独立に分布し，$X_i \sim Po(\lambda)$，$Y_j \sim Po(\mu)$ に従うとする 2 標本問題を考える。$S = \sum_{i=1}^n X_i$, $T = \sum_{i=1}^n Y_i$ とすると，$S \sim Po(n\lambda)$, $T \sim Po(n\mu)$ に従う。

(1) $H_0: \lambda = \mu$ vs. $H_1: \lambda \neq \mu$ なる仮説検定問題を考える。尤度関数は

$$L(\lambda, \mu) = \prod_{i=1}^{n}\Big\{\frac{\lambda^{x_i}}{x_i!}e^{-\lambda}\frac{\mu^{y_i}}{y_i!}e^{-\mu}\Big\} = \frac{\lambda^S\mu^T}{\prod_{i=1}^n(x_i!y_i!)}e^{-n(\lambda+\mu)}$$

であり，λ, μ の MLE は $\hat{\lambda} = S/n$, $\widehat{\mu} = T/n$ である。一方，$H_0: \lambda = \mu = \beta$ のもとでは,

$$L(\beta, \beta) = \frac{\beta^{S+T}}{\prod_{i=1}^n(x_i!y_i!)}e^{-2n\beta}$$

となり，β の MLE は $\widehat{\beta} = (S+T)/(2n)$ となる。尤度比は

$$\Lambda = \frac{L(\widehat{\beta}, \widehat{\beta})}{L(\hat{\lambda}, \widehat{\mu})} = \frac{(\frac{S+T}{2n})^{S+T} e^{-(S+T)}}{(S/n)^S (T/n)^T e^{-(S+T)}}$$

と書けるので，有意水準 α の尤度比検定の棄却域は

$$-2\log\Lambda = 2S\log S + 2T\log T - 2(S+T)\log\frac{S+T}{2} > \chi^2_{1,\alpha} \Longrightarrow H_0\text{を棄却}$$

となる。

(2) $S+T=m$ を与えたときの S の条件付き確率は $Bin(m,\theta)$, $\theta = \lambda/(\lambda+\mu)$, となることを示す。

$$\mathrm{P}(S=x|S+T=k) = \frac{\mathrm{P}(S=x, S+T=k)}{\mathrm{P}(S+T=k)} = \frac{\mathrm{P}(S=x, T=k-x)}{\mathrm{P}(S+T=k)} = \frac{\mathrm{P}(S=x)\mathrm{P}(T=k-x)}{\mathrm{P}(S+T=k)}$$

と書ける。$S+T \sim Po(n\lambda+n\mu)$ であるから

$$\mathrm{P}(S=x|S+T=k) = \frac{\frac{n\lambda)^x}{x!}e^{-n\lambda}\frac{(n\mu)^{k-x}}{(k-x)!}e^{-n\mu}}{\frac{(n\lambda+n\mu)^k}{k!}e^{-n(\lambda+\mu)}} = \frac{k!}{x!(k-x)!}\Big(\frac{\lambda}{\lambda+\mu}\Big)^x\Big(\frac{\mu}{\lambda+\mu}\Big)^{k-x}$$

と書けるので，条件付き確率は

$$S|S+T=k \sim Bin(k,\theta), \quad \theta = \frac{\lambda}{\lambda+\mu}$$

となる。

(1) の検定問題は，$H_0: \theta = 1/2$ vs. $H_1: \theta \neq 1/2$ と表すことができる。条件付き確率に基づいて，この検定の有意水準 α の検定の近似的な棄却域は, k が大きいとき

$$\mathrm{P}\Big(\frac{\sqrt{k}(S/k-\theta)}{\sqrt{\theta(1-\theta)}} \le c\Big|S+T=k\Big) \approx \Phi(c)$$

より

$$\frac{\sqrt{k}|S/k-1/2|}{\sqrt{(S/k)(1-S/k)}} = \frac{\sqrt{k}|S-k/2|}{\sqrt{S(k-S)}} > z_{\alpha/2} \Longrightarrow H_0\text{を棄却}$$

となる。

問 8　2 次元の確率変数 $(X_1,Y_1),\ldots,(X_n,Y_n)$ が独立で，2 変量正規分布に従い，X_i, Y_i の平均は $E[X_i]=\mu_1$, $E[Y_i]=\mu_2$, 分散は $\mathrm{Var}(X_i)=\sigma_1^2$, $\mathrm{Var}(Y_i)=\sigma_2^2$，共分散は $\mathrm{Cov}(X_i,Y_i)=\rho\sigma_1\sigma_2$ であるとする。$Z_i = X_i - Y_i$ とおくと，$Z_1,\ldots,Z_n$, i.i.d. $\sim \mathcal{N}(\mu_1-\mu_2,\sigma^2)$ となる。ここで，$\sigma^2 = \sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2$ である。

$$\overline{Z} = \frac{1}{n}\sum_{i=1}^n Z_i, \quad V^2 = \frac{1}{n-1}\sum_{i=1}^n (Z_i-\overline{Z})^2$$

とおく。

(1) $H_0: \mu_1-\mu_2=0$ v.s. $H_1: \mu_1-\mu_2 \neq 0$ なる両側検定の棄却域は

$$\sqrt{n}\frac{|\overline{Z}|}{V} > t_{n-1,\alpha/2} \Longrightarrow H_0\text{を棄却}$$

で与えられる。

(2) $\mu_1-\mu_2$ の信頼係数 $1-\alpha$ の信頼区間は

$$\overline{Z}\pm\frac{V}{\sqrt{n}}t_{n-1,\alpha/2}$$

で与えられる。

(3) $H_0:\mu_1=\mu_2$ vs $H_1:\mu_1>\mu_2$ なる片側検定の棄却域は

$$\sqrt{n}\frac{\overline{Z}}{V}>t_{n-1,\alpha}\Longrightarrow\ H_0\text{を棄却}$$

で与えられる。

問 9　確率変数 $X_1,\ldots,X_n$ が独立に確率密度 $\sim f(x|\theta)$ に従うとする。既知の θ_0 に対して，$H_0:\theta=\theta_0$ vs. $H_1:\theta\neq\theta_0$ を検定する問題を考える。$\boldsymbol{X}=(X_1,\ldots,X_n)$ とし，有意水準 α の検定の棄却域が $W(\boldsymbol{X})>c_\alpha$ で与えられるとする。ただし，c_α は $\mathrm{P}_{\theta_0}(W(\boldsymbol{X})>c_\alpha)=\alpha$ となる定数である。

(1) $\boldsymbol{X}$ の実現値 $\boldsymbol{x}$ に対して P 値 $p(\boldsymbol{x})$ は

$$p(\boldsymbol{x})=\mathrm{P}_{\theta_0}\Big(W(\boldsymbol{X})\geq W(\boldsymbol{x})\Big)$$

で定義される。

(2) $H_0:\theta=\theta_0$ のもとで，$F_{\theta_0}(\cdot)$ を $W(\boldsymbol{X})$ の分布関数とすると，$p(\boldsymbol{x})=1-F_{\theta_0}\big(W(\boldsymbol{x})\big)$ と表され，$p(\boldsymbol{X})$ は一様分布 $U(0,1)$ に従う。そこで $U=p(\boldsymbol{X})$ とおくと

$$\mathrm{P}_{\theta_0}\Big(p(\boldsymbol{X})\leq\alpha\Big)=\mathrm{P}(U\leq\alpha)=\alpha$$

となる。

(3) 有意水準 α の棄却域が $W(\boldsymbol{x})>c_\alpha$ で与えられているので，$\mathrm{P}_{\theta_0}(W(\boldsymbol{X})>c_\alpha)=\alpha$ であることに注意する。

$$\begin{aligned}p(\boldsymbol{x})<\alpha &\Longleftrightarrow 1-F_{\theta_0}(W(\boldsymbol{x}))<\alpha\\ &\Longleftrightarrow \mathrm{P}_{\theta_0}\Big(W(\boldsymbol{X})>W(\boldsymbol{x})\Big)<\alpha=\mathrm{P}_{\theta_0}(W(\boldsymbol{X})>c_\alpha)\\ &\Longleftrightarrow W(\boldsymbol{x})>c_\alpha\end{aligned}$$

となる。

問 10　確率変数 $X_1,\ldots,X_n$ が独立に $\mathcal{N}(0,\theta)$ に従うとする。

(1) 尤度関数は

$$\frac{1}{(2\pi\theta)^{n/2}}e^{-\sum_{i=1}^n x_i^2/(2\theta)}$$

であり，対数尤度と尤度方程式は

$$\ell(\theta)=-\frac{n}{2}\log(2\pi\theta)-\frac{1}{2\theta}\sum_{i=1}^n x_i^2,\ \ell'(\theta)=-\frac{n}{2\theta}+\frac{1}{2\theta^2}\sum_{i=1}^n x_i^2=0$$

と書けるので，θ の MLE は $\hat{\theta}^M=n^{-1}\sum_{i=1}^n X_i^2$ となる。

仮説検定 $H_0:\theta=\theta_0$ vs. $H_1:\theta\neq\theta_0$ について，尤度比は

$$\Lambda=\frac{\theta_0^{-n/2}e^{-\sum_{i=1}^n x_i^2/(2\theta_0)}}{(\hat{\theta}^M)^{-n/2}e^{-n/2}}=\Big(\frac{\hat{\theta}^M}{\theta_0}\Big)^{n/2}e^{-n\hat{\theta}^M/(2\theta_0)+n/2}$$

であるから，有意水準 α の尤度比検定の棄却域は

$$-2\log\Lambda = -n\log\left(\frac{\hat{\theta}^M}{\theta_0}\right) + n\frac{\hat{\theta}^M}{\theta_0} - n > \chi^2_{1,\alpha} \Longrightarrow H_0\text{を棄却}$$

で与えられる。

(2) 上で与えられた尤度比検定は，$c_1 < c_2$ に対して

$$n\frac{\hat{\theta}^M}{\theta_0} < c_1 \text{ もしくは } n\frac{\hat{\theta}^M}{\theta_0} > c_2 \Longrightarrow H_0\text{を棄却}$$

のような形をしている。そこで，$n\hat{\theta}^M/\theta_0$ に基づいた正確な検定を求めてみる。$H_0 : \theta = \theta_0$ のもとで

$$n\frac{\hat{\theta}^M}{\theta_0} = \frac{\sum_{i=1}^n X_i^2}{\theta_0} \sim \chi^2_n$$

に従うので，第 1 種の誤りの確率を両側に $\alpha/2$ ずつとるような検定を求めると

$$n\frac{\hat{\theta}^M}{\theta_0} < \chi^2_{n,1-\alpha/2} \text{ もしくは } n\frac{\hat{\theta}^M}{\theta_0} > \chi^2_{n,\alpha/2} \Longrightarrow H_0\text{を棄却}$$

が有意水準 α の正確な検定の棄却域になる。

(3) (2) で求めた検定に基づいて，θ に関する信頼係数 $1-\alpha$ の信頼区間は

$$\left[n\frac{\hat{\theta}^M}{\chi^2_{n,\alpha/2}},\ n\frac{\hat{\theta}^M}{\chi^2_{n,1-\alpha/2}}\right]$$

で与えられる。

第10章　カイ２乗適合度検定と応用例

問 1　あるプログラムから 10 個の一様乱数を発生させてみた。0.21, 0.35, 0.46, 0.51, 0.55, 0.62, 0.79, 0.80, 0.83, 0.92 これは一様分布 $U(0,1)$ からの乱数と見なせるかを検定する。

	0〜0.25	0.25〜0.5	0.5 〜0.75	0.75 〜 1.0
観測度数	1	2	3	4
期待度数	2.5	2.5	2.5	2.5

この表から，カイ 2 乗検定統計量を計算すると

$$Q = \frac{(1-2.5)^2}{2.5} + \frac{(2-2.5)^2}{2.5} + \frac{(3-2.5)^2}{2.5} + \frac{(4-2.5)^2}{2.5} = 2$$

となる。$\chi^2_{3,0.05} = 7.81$ より，有意でない。$U(0,1)$ からの乱数であることは否定できない。

```
> X <- c(1,2,3,4)
> Y <- c(2.5,2.5,2.5,2.5)
> Q <- sum( (X-Y)*(X-Y)/Y)
> Q
[1] 2
> qchisq(0.95, 3)
[1] 7.814728
```

問 2　あるプログラムから 10 個の正規乱数を発生させてみた。-2.03, -1.21,-0.5, -0.22, -0.13, 0.03, 0.39, 0.78, 1.52, 2.14 これは標準正規分布 $\mathcal{N}(0,1)$ からの乱数と見なせるかを検定する。

	〜-1.0	-1.0〜0.0	0.0 〜1.0	1.0 〜
観測度数	2	3	3	2
期待度数	1.59	3.41	3.41	1.59

この表から，カイ 2 乗検定統計量を計算すると, $Q = 0.32$ となる。$\chi^2_{3,0.05} = 7.81$ より，有意でない。$\mathcal{N}(0,1)$ からの乱数であることは否定できない。

```
> n <- 10
> X <- c(2,3,3,2)
> Y <- n*c(pnorm(-1), pnorm(0)-pnorm(-1), pnorm(1)-pnorm(0), 1-pnorm(1))
> Q <- sum( (X-Y)*(X-Y)/Y)
> Q
[1] 0.3156407
```

問 3　エンドウ豆の分類に関する観測データとメンデルの法則についての問題である。

(1) それぞれの表現型の期待度数は理論確率から計算されるので次の表で与えられる。

表現型	黄色丸型	黄色しわ型	緑色丸型	緑色しわ型	合計
観測度数	315	101	108	32	556
理論確率	9/16	3/16	3/16	1/16	1
期待度数	313	104	104	35	556

(2) カイ 2 乗検定統計量を計算すると $Q = 0.47$ になり，$\chi^2_{3,0.05} = 7.81$ より，メンデルの法則は棄却できない。

(3) $\mathrm{P}(\chi^2_3 > 0.47) = 0.93$ で，P 値が 0.93 とかなり高い値を示しており，この観測データはメンデルの法則に合い過ぎている。このことから，データが実際に観測されたものなのかという疑義を生ずることになる。

```
> X <- c(315,101,108,32)
> Y <- 556*c(9/16, 3/16, 3/16, 1/16)
> Q <- sum( (X-Y)*(X-Y)/Y)
> Q
[1] 0.470024
```

問 4 2021 年の東京都の月別自殺者数が警察庁のホームページから公開されている。季節性があるかを調べたい。

(1) 季節性が無いとした場合の各月の期待度数は，$2292/12 = 191$ になる。

月	1	2	3	4	5	6	7	8	9	10	11	12	合計
自殺者数	195	190	187	197	201	216	189	203	173	179	193	169	2292
期待度数	191	191	191	191	191	191	191	191	191	191	191	191	2292

(2) 季節性があるかについてカイ 2 乗検定統計量は

$$Q = \frac{(195-191)^2}{191} + \cdots + \frac{(169-191)^2}{191} = 9.94$$

となり，$\chi^2_{11,0.05} = 19.68$ より，有意でないことがわかる。すなわち，季節性が無いことは否定できない。

```
> X <- c(195,190,187,197,201,216,189,203,173,179,193,169)
> n <- sum(X)
> n/12
[1] 191
> Y <- 191*c(rep(1, 12))
> Q <- sum( (X-Y)*(X-Y)/Y)
> Q
[1] 9.937173
> qchisq(0.95, 11)
[1] 19.67514
```

問 5 ハーディー・ワインベルグ平衡によると，遺伝子型 AA, Aa, aa の出現確率はそれぞれ $(1-\theta)^2$, $2\theta(1-\theta)$, θ^2 で与えられる。いま，1000 人の赤血球抗原の遺伝子型 MM,MN, NN を調べたところ次の表のようになった。ハーディー・ワインベルグ平衡が成り立つかを検証したい。

遺伝子型	MM	MN	NN	合計
観測度数	330	490	180	1000

(1) MM, MN, NN の度数を確率変数 X_1, X_2, X_3 で表すと，$X_1 + X_2 + X_3 = 1000$ を満たす。ハーディー・ワインベルグ平衡が成り立つときには (X_1, X_2, X_3) は，多項分布

$$(X_1, X_2, X_3) \sim Mult_3(n, p_1, p_2, p_3) = \frac{n!}{x_1!x_2!x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3}$$

に従う。ただし，$n = 1000$, $x_1 = 330$, $x_2 = 490$, $x_3 = 180$, $p_1 = (1-\theta)^2$, $p_2 = 2\theta(1-\theta)$, $p_3 = \theta^2$ である。

(2) 対数尤度は

$$\ell(\theta) = \log \frac{n!}{x_1!x_2!x_3!} + 2x_1 \log(1-\theta) + x_2 \log\{2\theta(1-\theta)\} + 2x_3 \log\theta$$

であるから，尤度方程式は

$$\ell'(\theta) = -\frac{2x_1}{1-\theta} + \frac{x_2}{\theta} - \frac{x_2}{1-\theta} + \frac{2x_3}{\theta} = 0$$

と書ける。この解を求めると，θ の MLE は

$$\hat{\theta} = \frac{x_2 + 2x_3}{2x_1 + 2x_2 + 2x_3} = \frac{x_2 + 2x_3}{2n}$$

となる。ここに数値を代入すると θ の最尤推定値は

$$\hat{\theta} = \frac{490 + 2 \times 180}{2 \times 1000} = 0.425$$

となる。

(3) ハーディー・ワインベルグ平衡が成り立つときの MM, MN, NN の期待度数は，それぞれ

$$\begin{aligned} 1000 \times (1-\hat{\theta})^2 =& 330.6 \\ 1000 \times 2\hat{\theta}(1-\hat{\theta}) =& 488.8 \\ 1000 \times \hat{\theta}^2 =& 180.6 \end{aligned}$$

で与えられるので，次の表のようになる。

遺伝子型	MM	MN	NN	合計
観測度数	330	490	180	1000
期待度数	330.6	488.8	180.6	1000

この表から，カイ 2 乗検定統計量は $Q = 0.006$ となる。また，カイ 2 乗分布の自由度は未知母数 θ を 1 つ含んでいることから，$2 - 1 = 1$ となり，$\chi^2_{1,0.05} = 3.84$ より，この検定は有意でない。すなわち，ハーディー・ワインベルグ平衡が成り立つことは否定できない。

```
> X <- c(330,490,180)
> Y <- c(330.6,488.8,180.6)
> Q <- sum( (X-Y)*(X-Y)/Y)
> Q
[1] 0.006028275
> qchisq(0.95,1)
[1] 3.841459
```

問 6　$X_i \sim Po(\lambda_i)$ より $\mathrm{E}[\overline{X}] = \overline{\lambda} = n^{-1}\sum_{i=1}^n \lambda_i$ である。

$$\sum_{i=1}^n (X_i - \overline{X})^2 = \sum_{i=1}^n \{(X_i - \lambda_i) - (\overline{X} - \overline{\lambda}) + (\lambda_i - \overline{\lambda})\}^2$$

より

$$\begin{aligned} \sum_{i=1}^n \mathrm{E}[(X_i - \overline{X})^2] =& \sum_{i=1}^n \mathrm{E}[(X_i - \lambda_i)^2 + (\overline{X} - \overline{\lambda})^2 + (\lambda_i - \overline{\lambda})^2 - 2(X_i - \lambda_i)(\overline{X} - \overline{\lambda})] \\ =& n\overline{\lambda} + \overline{\lambda} + \sum_{i=1}^n (\lambda_i - \overline{\lambda})^2 - 2\overline{\lambda} = (n-1)\overline{\lambda} + \sum_{i=1}^n (\lambda_i - \overline{\lambda})^2 \end{aligned}$$

となる。

第11章　回帰分析-単回帰モデル-

問1 13人の親子の身長を調べ，親から子への要因分析を行ってみた。x, y をそれぞれ父親，息子の身長とする。y を x で説明する回帰分析をRを用いて行ったところ, 次の分析結果を得た。

(1) 残差平方和の自由度は11である。

(2) 回帰直線は $y = 47.8 + 0.75x$ である。

(3) 決定係数は $R^2 = 0.78$ である。$\hat{\sigma} = 2.71$ であるから，標準化残差の最小値と最大値は

$$\frac{\text{Min of Residuals}}{\hat{\sigma}} = -1.30, \quad \frac{\text{Max of Residuals}}{\hat{\sigma}} = 1.66$$

となり，± 2 を超えてないので外れ値はないと思われる。

(4) 父親の身長は息子の身長を説明するのに，x の係数のP値が0.000056なので，有意水準5%で有意といえる。

Residuals:	Min	1Q	Median	3Q	Max
	-3.5103	-2.3164	-0.2991	1.5070	4.5070

Coefficients:		Estimate	Std. Error	t value	Pr(> |t|)
	(Intercept)	47.8025	19.8113	2.413	0.0344
	x	0.7543	0.1192	6.328	5.61e-05

Residual standard error: 2.71 on 11 degrees of freedom

Multiple R-squared: 0.7845, Adjusted R-squared: 0.7649

F-statistic: 40.04 on 1 and 11 DF, p-value: 5.615e-05

問2 次の表はある県の15の市について人口 (x) と一般行政職員数 (y) を調べたデータである。ただし人口の単位は10,000人，職員数の単位は100人である。

番号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
人口	7.7	4.7	14.4	7.8	5.3	5.2	8.1	4.6	20.0	5.7	11.2	11.1	4.0	4.9	6.6
職員数	3.4	2.1	7.1	3.9	2.7	2.7	4.4	2.6	11.5	3.3	6.6	6.7	2.5	3.1	4.3

(1) 単回帰モデルを当てはめるとき，α, β, σ^2 について不偏推定量による推定値は

$$\widehat{\alpha} = -0.074, \quad \widehat{\beta} = 0.561, \quad \hat{\sigma}^2 = 0.256$$

となる。

(2) 決定係数は $R^2 = 0.963$ である。

(3) x の値を横軸にとって残差をプロットしたのが図11.1である。残差に傾向性が見られないので回帰分析に問題があるとは考えにくい。

(4) 仮説検定 $H_0 : \beta = 0$ vs $H_1 : \beta \neq 0$ の検定については，x の係数のP値が 1.10×10^{-10} となるので，有意水準5%で有意になる。

(5) β について信頼係数95%の信頼区間は

$$\widehat{\beta} \pm \frac{\hat{\sigma}}{\sqrt{nS_{xx}}} t_{n-2,\alpha/2} = 0.561 \pm \frac{0.5062}{\sqrt{15 \times 19.66}} \times 2.16$$

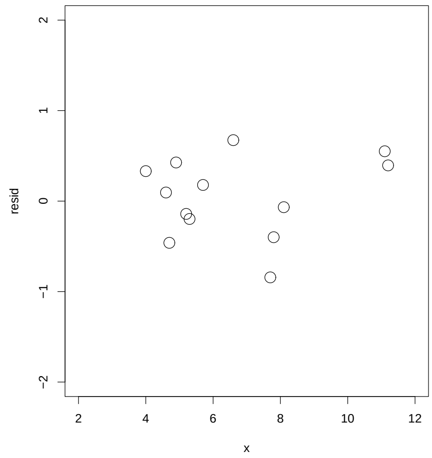

図 11.1: 人口 (x 軸) と残差 (y 軸) のプロット

で与えられるので，$[0.50, 0.63]$ になる。

```
> population <- c(7.7,4.7,14.4,7.8,5.3,5.2,8.1,4.6,20.0,
+ 5.7,11.2,11.1,4.0,4.9,6.6)
> staff <- c(3.4,2.1,7.1,3.9,2.7,2.7,4.4,2.6,11.5,3.3,6.6,6.7,2.5,3.1,4.3)
> staff.lm <- lm(staff ~ population)
> summary(staff.lm)

Call:
lm(formula = staff ~ population)

Residuals:
     Min 1Q Median 3Q Max
-0.89952 -0.29848 0.09477 0.37771 0.67349

Coefficients:
           Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.07372 0.27920 -0.264 0.796
population 0.56064 0.03051 18.376 1.1e-10 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 ' '. 0.1 ' '1

Residual standard error: 0.5062 on 13 degrees of freedom
Multiple R-squared: 0.9629, Adjusted R-squared: 0.9601
F-statistic: 337.7 on 1 and 13 DF, p-value: 1.103e-10
```

問 3 標準化残差をプロットしたところ次のような傾向性が見られた。どのように対処したらよい。

(1) 正の値になると正の値が続き，負の値に変わると負の値が続いている。$\Longrightarrow$ 系列相関があると思われるので，ダービン・ワトソン検定を行い，系列相関の存在が確認されたら，コクラン・オーカット法を用いて解析する。

(2) 説明変数の値が大きくなるにつれて，残差のバラツキが大きくなる。$\Longrightarrow$ 分散の不均一性があるので，重み付き最小 2 乗法を用いる。

(3) 標準化残差の値が 2 を超える点が 1 つあり，クックの距離を調べたところ他に比べてかなり大きな値を示している。$\Longrightarrow$ 標準化残差が 2 を超える点を除いて分析する。

(4) 残差が，U 字型をしており，x 軸の両側で正で大きく中央付近で負で小さい。$\Longrightarrow$ 対数変換を行って分析する。

(5) 残差を正規 Q-Q プロットすると，x 軸の両端近くで直線 $y = x$ よりも上側にプロットされている。$\Longrightarrow$ 正規性が疑われるので，正規分布を前提としない分析を行う必要がある。

問 4 予測量の分散が (11.7) で与えられることを示す。まず

$$\hat{y}_0 - y_0 = (\widehat{\alpha} - \alpha) + (\widehat{\beta} - \beta)x_0 - u_0$$

より

$$\begin{aligned}\mathrm{Var}(\hat{y}_0 - y_0) =& \mathrm{Var}(\widehat{\alpha}) + x_0^2\mathrm{Var}(\widehat{\beta}) + 2x_0\mathrm{Cov}(\widehat{\alpha}, \widehat{\beta}) + \sigma^2 = \sigma^2\Big\{\frac{\sum_{i=1}^n x_i^2}{n^2 S_{xx}} + \frac{x_0^2}{n s_{xx}} - 2x_0\frac{\overline{x}\sigma^2}{nS_{xx}}\Big\} + \sigma^2 \\ =& \sigma^2\frac{\sum_{i=1}^n x_i^2 + nx_0^2 - 2nx_0\overline{x}}{n^2 S_{xx}} + \sigma^2\end{aligned}$$

と書ける。ここで，$\sum_{i=1}^n x_i^2 = nS_{xx} + n\overline{x}^2$ であるから，

$$\sum_{i=1}^n x_i^2 + nx_0^2 - 2nx_0\overline{x} = nS_{xx} + n\overline{x}^2 + nx_0^2 - 2nx_0\overline{x} = nS_{xx} + n(\overline{x} - x_0)^2$$

となり

$$\mathrm{Var}(\hat{y}_0 - y_0) = \sigma^2\Big\{1 + \frac{1}{n} + \frac{(\overline{x} - x_0)^2}{nS_{xx}}\Big\}$$

となる。

問 5 $Y_1, \ldots, Y_n$ を独立な確率変数とし，各 i に対して $Y_i \sim \mathcal{N}(\alpha + \beta x_i, \sigma^2)$ とする。定数 $x_1, \ldots, x_n$ は $\sum_{i=1}^n x_i = 0$, $\sum_{i=1}^n x_i^2 = n$ を満たすものとする。このとき，$\overline{x} = 0$, $S_{xx} = 1$ である。

(1) α, β の最小 2 乗推定量 (LSE) $\widehat{\alpha}$, $\widehat{\beta}$ を求める。$h(\alpha, \beta) = \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2$ とおくと

$$\begin{aligned}\frac{\partial h}{\partial \alpha} =& -2\sum_{i=1}^n (y_i - \alpha - \beta x_i) = -2n(\overline{y} - \alpha) \\ \frac{\partial h}{\partial \beta} =& -2\sum_{i=1}^n x_i(y_i - \alpha - \beta x_i) = -2\Big(\sum_{i=1}^n x_i y_i - n\beta\Big)\end{aligned}$$

と書けるので，$\widehat{\alpha} = \overline{y}$, $\widehat{\beta} = n^{-1}\sum_{i=1}^n x_i y_i$ となる。

(2) $\widehat{\alpha} = \overline{y} = \alpha + \overline{\varepsilon}$,

$$\widehat{\beta} = \frac{1}{n}\sum_{i=1}^n x_i(\alpha + \beta x_i + \varepsilon_i) = \beta + \frac{1}{n}\sum_{i=1}^n x_i\varepsilon_i$$

と書けるので，$\mathrm{E}[\widehat{\alpha}] = \alpha$, $\mathrm{E}[\widehat{\beta}] = \beta$,

$$\begin{aligned}\mathrm{Var}(\widehat{\alpha}) =& \frac{\sigma^2}{n},\ \mathrm{Var}(\widehat{\beta}) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{E}[x_i^2\varepsilon_i^2] = \frac{\sigma^2}{n}, \\ \mathrm{Cov}(\widehat{\alpha}, \widehat{\beta}) =& \frac{1}{n^2}\mathrm{E}\Big[\sum_{i=1}^n \varepsilon_i\sum_{j=1}^n x_j\varepsilon_j\Big] = \frac{1}{n^2}\sum_{i=1}^n \mathrm{E}[x_i^2\varepsilon_i^2] = 0\end{aligned}$$

となる。以上より，定数 x に対して $y = \alpha + \beta x$ を $\hat{y} = \widehat{\alpha} + \widehat{\beta}x$ で推定するとき，

$$\begin{aligned}\mathrm{E}[\hat{y}] =& \mathrm{E}[\widehat{\alpha} + \widehat{\beta}x] = \alpha + \beta x = y \\ \mathrm{Var}(\hat{y}) =& \mathrm{Var}(\widehat{\alpha}) + x^2\mathrm{Var}(\widehat{\beta}) + 2x\mathrm{Cov}(\widehat{\alpha}, \widehat{\beta}) = \frac{\sigma^2}{n} + \frac{x^2\sigma^2}{n} = \frac{(1 + x^2)\sigma^2}{n}\end{aligned}$$

となり，$\hat{y}$ が y の不偏推定量であることがわかる。

(3) $\hat{y}$ の確率分布は

$$\hat{y}-y=(\widehat{\alpha}-\alpha)+(\widehat{\beta}-\beta)x=\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i+\frac{x}{n}\sum_{i=1}^{n}x_i\varepsilon_i=\frac{1}{n}\sum_{i=1}^{n}(1+xx_i)\varepsilon_i$$

と書けるので，

$$\hat{y}-y\sim\mathcal{N}\Big(0,\frac{\sigma^2}{n^2}\sum_{i=1}^{n}(1+xx_i)^2\Big)$$

となる。$\sum_{i=1}^{n}(1+xx_i)^2=n+2x\sum_{i=1}^{n}x_i+x^2\sum_{i=1}^{n}x_i^2$ より

$$\hat{y}\sim\mathcal{N}\Big(y,\frac{(1+x^2)\sigma^2}{n}\Big)$$

となる。

$\hat{\sigma}^2=(n-2)^{-1}\sum_{i=1}^{n}(y_i-\widehat{\alpha}-x_i\widehat{\beta})^2$ とおくと，$(n-2)\hat{\sigma}^2/\sigma^2\sim\chi^2_{n-2}$ に従うので，y の信頼係数 $1-\alpha$ の信頼区間は

$$\hat{y}\pm\frac{\sqrt{1+x^2}\hat{\sigma}}{\sqrt{n}}t_{n-2,\alpha/2}$$

で与えられる。

問 6　$Y_1,\dots,Y_n$ を独立な確率変数とし，各 i に対して $\mathrm{E}[Y_i]=\beta x_i$, $\mathrm{Var}(Y_i)=\sigma^2$ とし，正規分布を仮定しない。定数 $x_1,\dots,x_n$ は $\sum_{i=1}^{n}x_i^2>0$ を満たすものとする。$Y_i=\beta x_i+\varepsilon_i$, $\mathrm{E}[\varepsilon_i]=0$, $\mathrm{Var}(\varepsilon_i)=\sigma^2$ と書ける。

$h(\beta)=\sum_{i=1}^{n}(Y_i-\beta x_i)^2$ とおくと，

$$h'(\beta)=-2\sum_{i=1}^{n}x_i(Y_i-\beta x_i)=0$$

より

$$\widehat{\beta}^L=\sum_{i=1}^{n}x_iY_i/\sum_{i=1}^{n}x_i^2$$

となる。

$\widehat{\beta}^L$ は

$$\widehat{\beta}^L=\frac{\sum_{i=1}^{n}x_i(\beta x_i+\varepsilon_i)}{\sum_{i=1}^{n}x_i^2}=\beta+\frac{\sum_{i=1}^{n}x_i\varepsilon_i}{\sum_{i=1}^{n}x_i^2}$$

より，$\mathrm{E}[\widehat{\beta}^L]=\beta$,

$$\mathrm{Var}(\widehat{\beta}^L)=\frac{1}{(\sum_{i=1}^{n}x_i^2)^2}\sum_{i=1}^{n}x_i^2\mathrm{E}[\varepsilon_i^2]=\frac{\sigma^2}{\sum_{i=1}^{n}x_i^2}$$

となる。

問 7　問 6 と同じ設定で考える。

(1) 定数 $c_1,\dots,c_n$ に対して $\sum_{i=1}^{n}c_iY_i$ の線形推定量が β の不偏推定量になるための $c_1,\dots,c_n$ の条件は

$$\sum_{i=1}^{n}c_iY_i=\sum_{i=1}^{n}c_ix_i\beta+\sum_{i=1}^{n}c_i\varepsilon_i$$

より，$\mathrm{E}[\sum_{i=1}^{n}c_iY_i]=\sum_{i=1}^{n}c_ix_i\beta$ となるので，これが β になるので，条件は

$$\sum_{i=1}^{n}c_ix_i=1$$

となる。

(2) 線形不偏な推定量のクラスの中で分散を最小にするものを求める。(1) で求めた条件を用いると，$\sum_{i=1}^n c_i Y_i - \beta = \sum_{i=1}^n c_i(Y_i - x_i\beta) = \sum_{i=1}^n c_i\varepsilon_i$ と表される。従って

$$\mathrm{Var}\Big(\sum_{i=1}^n c_i Y_i\Big) = \mathrm{E}\Big[\Big(\sum_{i=1}^n c_i\varepsilon_i\Big)^2\Big] = \sum_{i=1}^n c_i^2 \mathrm{E}[\varepsilon_i^2] = \sigma^2 \sum_{i=1}^n c_i^2$$

と書ける。ラグランジュの未定乗数法を用いて，$\sum_{i=1}^n c_i x_i = 1$ の条件のもとで $\sum_{i=1}^n c_i^2$ を最小にする解を求める。

$$g(c_1, \ldots, c_n) = \sum_{i=1}^n c_i^2 - \lambda\Big(\sum_{i=1}^n c_i x_i - 1\Big)$$

とおくと

$$\frac{\partial g}{\partial c_i} = 2c_i - \lambda x_i = 0,\ i = 1, \ldots, n$$

より，$c_i = \lambda x_i/2$ となる。これを $\sum_{i=1}^n c_i x_i = 1$ に代入すると，$\lambda = 2/\sum_{i=1}^n x_i^2$ となる。従って

$$c_i = \frac{x_i}{2}\frac{2}{\sum_{j=1}^n x_j^2} = \frac{x_i}{\sum_{j=1}^n x_j^2}$$

となる。よって，β の最小 2 乗推定量は

$$\widehat{\beta}^B = \sum_{i=1}^n c_i Y_i = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{j=1}^n x_j^2}$$

で与えられる。分散は $\mathrm{Var}(\sum_{i=1}^n c_i Y_i) = \sigma^2 \sum_{i=1}^n c_i^2$ に，上で求めた c_i を代入して求められるので，

$$\mathrm{Var}(\widehat{\beta}^B) = \sigma^2 \frac{\sum_{i=1}^n x_i^2}{(\sum_{j=1}^n x_j^2)^2} = \frac{\sigma^2}{\sum_{j=1}^n x_j^2}$$

となる。

(3) $x_1, \ldots, x_n$ は 0 でない定数とするときには，$\widehat{\beta}^A = \{\sum_{i=1}^n (Y_i/x_i)\}/n$ なる形の推定量を考えることができる。

$$\widehat{\beta}^A = \frac{1}{n}\sum_{i=1}^n \frac{Y_i}{x_i} = \frac{1}{n}\sum_{i=1}^n \frac{\beta x_i + \varepsilon_i}{x_i} = \beta + \frac{1}{n}\sum_{i=1}^n \frac{\varepsilon_i}{x_i}$$

と書けるので，$\mathrm{E}[\widehat{\beta}^A] = \beta$ となり，不偏性が示される。分散は

$$\mathrm{Var}(\widehat{\beta}^A) = \frac{1}{n^2}\sum_{i=1}^n \frac{\mathrm{E}[\varepsilon_i^2]}{x_i^2} = \frac{\sigma^2}{n^2}\sum_{i=1}^n \frac{1}{x_i^2}$$

で与えられる。

$$\mathrm{Var}(\widehat{\beta}^B) = \frac{\sigma^2}{\sum_{i=1}^n x_i^2} \le \frac{\sigma^2}{n^2}\sum_{i=1}^n \frac{1}{x_i^2} = \mathrm{Var}(\widehat{\beta}^A)$$

は

$$n^2 \le \sum_{i=1}^n x_i^2 \sum_{i=1}^n \frac{1}{x_i^2}$$

のように書き直すことができる。これはコーシー・シュヴァルツの不等式から成り立つ。実際，

$$n^2 = \Big(\sum_{i=1}^n x_i \times \frac{1}{x_i}\Big)^2 \le \sum_{i=1}^n x_i^2 \sum_{i=1}^n \frac{1}{x_i^2}$$

が成り立つことがわかる。

問 8　$Y_1, \ldots, Y_n$ を独立な確率変数とし，$i = 1, \ldots, n$ に対して $Y_i \sim \mathcal{N}(\beta x_i, \sigma^2)$ に従うとする。ここで，$x_1, \ldots, x_n$ は定数で $\sum_{i=1}^n x_i^2 > 0$ とし，β と σ^2 は未知母数とする。

(1) 尤度関数は

$$L(\beta, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(Y_i - \beta x_i)^2} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (Y_i - \beta x_i)^2}$$

であり，$\sum_{i=1}^n (Y_i - \beta x_i)^2 = \sum_{i=1}^n Y_i^2 - 2\sum_{i=1}^n x_i Y_i \beta + \sum_{i=1}^n x_i^2 \beta^2$ と書けるので，

$$L(\beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\Big[-\frac{1}{2\sigma^2}\Big\{ \sum_{i=1}^n Y_i^2 - 2\beta \sum_{i=1}^n x_i Y_i - \beta^2 \sum_{i=1}^n x_i^2 \Big\}\Big]$$

と表される。従って，$(\sum_{i=1}^n Y_i^2, \sum_{i=1}^n x_i Y_i)$ が (β, σ^2) の十分統計量になる。

対数尤度は

$$\ell(\beta, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\Big\{ \sum_{i=1}^n Y_i^2 - 2\beta \sum_{i=1}^n x_i Y_i - \beta^2 \sum_{i=1}^n x_i^2 \Big\}$$

で与えられるので

$$\begin{aligned}
\frac{\partial \ell}{\beta} &= \frac{1}{\sigma^2}\Big(\sum_{i=1}^n x_i Y_i - \beta \sum_{i=1}^n x_i^2 \Big) \\
\frac{\partial \ell}{\beta} &= -\frac{n}{2\sigma^2} - \frac{1}{2\sigma^4}\Big\{ \sum_{i=1}^n Y_i^2 - 2\beta \sum_{i=1}^n x_i Y_i - \beta^2 \sum_{i=1}^n x_i^2 \Big\}
\end{aligned}$$

となる。これより, β の MLE は

$$\widehat{\beta}^M = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}$$

となり，σ^2 の MLE は

$$\hat{\sigma}^{2M} = \frac{1}{n}\Big\{ \sum_{i=1}^n Y_i^2 - 2\widehat{\beta}^M \sum_{i=1}^n x_i Y_i - (\widehat{\beta}^M)^2 \sum_{i=1}^n x_i^2 \Big\} = \frac{1}{n}\sum_{i=1}^n (Y_i - x_i \widehat{\beta}^M)^2$$

となる。

(2) σ^2 が既知であるとする。

$$\ell''(\beta) = -\frac{1}{\sigma^2}\sum_{i=1}^n x_i^2$$

であるから，β の Fisher 情報量は

$$I_n(\beta) = \frac{\sum_{i=1}^n x_i^2}{\sigma^2}$$

になる。よって，クラメール・ラオ不等式は，任意の不偏推定量 $\widehat{\beta}$ に対して

$$\mathrm{Var}(\widehat{\beta}) \geq \frac{\sigma^2}{\sum_{i=1}^n x_i^2}$$

となる。$\mathrm{Var}(\widehat{\beta}^M) = \sigma^2 / \sum_{i=1}^n x_i^2$ であるから，$\widehat{\beta}^M$ は不偏推定量の中で分散を最小にすることがわかる。

(3) $\sum_{i=1}^n (Y_i - x_i\widehat{\beta}^M)^2/\sigma^2 \sim \chi^2_{n-1}$ に従うので

$$\mathrm{E}[\hat{\sigma}^{2M}] = \frac{n-1}{n}\sigma^2$$

となり，不偏でないことがわかる。σ^2 の不偏推定量は

$$\frac{n}{n-1}\hat{\sigma}^{2M} = \frac{1}{n-1}\sum_{i=1}^n (Y_i - x_i\widehat{\beta}^M)^2$$

で与えられる。

問 9　問 8 と同じ設定で，$\widehat{\beta} = \sum_{i=1}^n x_iY_i/\sum_{i=1}^n x_i^2$, $V^2 = (n-1)^{-1}\sum_{i=1}^n (Y_i - \widehat{\beta}x_i)^2$ とおく。$Y_i = \beta x_i + \varepsilon_i$, $\varepsilon_i \sim \mathcal{N}(0,\sigma^2)$ と表される。

(1) $\widehat{\beta}$ は

$$\widehat{\beta} = \frac{\sum_{i=1}^n x_i(\beta x_i + \varepsilon_i)}{\sum_{i=1}^n x_i^2} = \beta + \frac{\sum_{i=1}^n x_i\varepsilon_i}{\sum_{i=1}^n x_i^2}$$

と表される。

$$\sum_{i=1}^n \frac{x_i}{\sum_{j=1}^n x_j^2}\varepsilon_i \sim \mathcal{N}\Big(0, \frac{\sum_{i=1}^n x_i^2\sigma^2}{(\sum_{j=1}^n x_j^2)^2}\Big)$$

より，$\widehat{\beta}$ の確率分布は

$$\widehat{\beta} \sim \mathcal{N}\Big(\beta, \frac{\sigma^2}{\sum_{i=1}^n x_i^2}\Big)$$

に従う。

(2) $\widehat{\beta}$ と V^2 は独立になることを示す。いま簡単のために，

$$c_i = \frac{x_i}{\sum_{j=1}^n x_j^2}$$

とおくと，$\widehat{\beta} = \sum_{i=1}^n c_iY_i = \beta + \sum_{i=1}^n c_i\varepsilon_i$ と書ける。また，$Y_i - \widehat{\boldsymbol{\beta}}x_i = \beta x_i + \varepsilon_i - \beta x_i - x_i\sum_{j=1}^n c_j\varepsilon_j = \varepsilon_i - x_i\sum_{j=1}^n c_j\varepsilon_j$ と書ける。ここで

$$\begin{aligned}
&\mathrm{Cov}\Big(\sum_{j=1}^n c_j\varepsilon_j, \varepsilon_i - x_i\sum_{j=1}^n c_j\varepsilon_j\Big) = \mathrm{E}\Big[\sum_{j=1}^n c_j\varepsilon_j\Big(\varepsilon_i - x_i\sum_{j=1}^n c_j\varepsilon_j\Big)\Big] \\
=&c_i\sigma^2 - x_i\mathrm{E}\Big[\Big(\sum_{j=1}^n c_j\varepsilon_j\Big)^2\Big] = c_i\sigma^2 - x_i\sum_{j=1}^n c_j^2\sigma^2 \\
=&\frac{x_i}{\sum_{j=1}^n x_j^2}\sigma^2 - x_i\frac{\sum_{j=1}^n x_j^2}{(\sum_{j=1}^n x_j^2)^2}\sigma^2 = 0
\end{aligned}$$

となり，$\sum_{i=1}^n c_i\varepsilon_i$ と $\{\varepsilon_i - x_i\sum_{j=1}^n c_j\varepsilon_j, i = 1,\ldots,n\}$ は無相関になることがわかる。正規分布を仮定しているので，これらは独立になる。よって，$\widehat{\beta}$ と $\{Y_i - \widehat{\beta}x_i, i = 1,\ldots,n\}$ は独立になる。$V^2 = (n-1)^{-1}\sum_{i=1}^n (Y_i - \widehat{\beta}x_i)^2$ であるから，$\widehat{\beta}$ と V^2 は独立になる。

(3) $(n-1)V^2/\sigma^2 \sim \chi^2_{n-1}$ を示す。$(n-1)V^2 = \sum_{i=1}^n (\varepsilon_i - x_i\sum_{j=1}^n c_j\varepsilon_j)^2$ と書けることに注意する。

$$\begin{aligned}
&\sum_{i=1}^n (\varepsilon_i - x_i\sum_{j=1}^n c_j\varepsilon_j)^2 = \sum_{i=1}^n \varepsilon_i^2 - 2\sum_{i=1}^n x_i\varepsilon_i\sum_{j=1}^n c_j\varepsilon_j + \sum_{i=1}^n x_i^2\Big(\sum_{j=1}^n c_j\varepsilon_j\Big)^2 \\
=&\sum_{i=1}^n \varepsilon_i^2 - \sum_{i=1}^n x_i^2\Big(\sum_{j=1}^n c_j\varepsilon_j\Big)^2
\end{aligned}$$

と書ける。この式から

$$\sum_{i=1}^{n}\varepsilon_i^2/\sigma^2=\sum_{i=1}^{n}(\varepsilon_i-x_i\sum_{j=1}^{n}c_j\varepsilon_j)^2/\sigma^2+\sum_{i=1}^{n}x_i^2\Big(\sum_{j=1}^{n}c_j\varepsilon_j\Big)^2/\sigma^2$$

が得られる。$\sum_{i=1}^{n}\varepsilon_i^2/\sigma^2\sim\chi_n^2$，$\sum_{i=1}^{n}x_i^2\Big(\sum_{j=1}^{n}c_j\varepsilon_j\Big)^2/\sigma^2\sim\chi_1^2$ である。

$$X=\sum_{i=1}^{n}x_i^2\Big(\sum_{j=1}^{n}c_j\varepsilon_j\Big)^2/\sigma^2,\quad Y=\sum_{i=1}^{n}(\varepsilon_i-x_i\sum_{j=1}^{n}c_j\varepsilon_j)^2/\sigma^2$$

とおくと，X と Y は独立で，$X\sim\chi_1^2$, $X+Y\sim\chi_n^2$ であることから，定理 7.8 の証明で使った方法を利用すると，$Y\sim\chi_{n-1}^2$ であることが示される。

(4) y_i と x_i の因果関係に関する仮説 $H_0:\ \beta=0$ vs. $H_1:\ \beta\neq 0$ についての検定は

$$\Big(\sum_{i=1}^{n}x_i^2\Big)^{1/2}\frac{|\widehat{\beta}|}{V}>t_{n-1,\alpha/2}\Longrightarrow\ H_0\text{を棄却}$$

で与えられる。

問 10　n 個の 2 次元確率変数 $(X_1,Y_1),\ldots,(X_n,Y_n)$ が互いに独立に分布していると仮定する。また各 $i=1,\ldots,n$ に対して，X_i を与えたときの Y_i の条件付き分布及び X_i の周辺分布が

$$\begin{aligned}Y_i|X_i\sim&\mathcal{N}(\alpha+\beta X_i,1),\\X_i\sim&\mathcal{N}(\mu,1)\end{aligned}$$

で与えられていると仮定する。ただし，α,β,μ は実数値をとる未知母数であり，$\boldsymbol{X}=(X_1,\ldots,X_n)$, $\boldsymbol{x}=(x_1,\ldots,x_n)$ とする。

(1) 尤度関数は

$$\prod_{i=1}^{n}\Big\{\frac{1}{2\pi}e^{-(y_i-\alpha-\beta x_i)^2/2-(x_i-\mu)^2/2}\Big\}$$

と書けるので，α,β,μ の MLE は $\widehat{\alpha}=\overline{y}-\widehat{\beta}\overline{x}$, $\widehat{\beta}=S_{xy}/S_{xx}$, $\widehat{\mu}=\overline{x}$ となる。

(2) $\boldsymbol{X}=\boldsymbol{x}$ を与えたときの $\widehat{\beta}$ の条件付き分布は

$$\widehat{\beta}|\boldsymbol{X}=\boldsymbol{x}\sim\mathcal{N}\Big(\beta,\frac{1}{\sum_{i=1}^{n}(x_i-\overline{x})^2}\Big)$$

となり，条件付き分散は

$$\mathrm{Var}(\widehat{\beta}\mid\boldsymbol{X}=\boldsymbol{x})=\frac{1}{\sum_{i=1}^{n}(x_i-\overline{x})^2}$$

で与えられる。

(3) $W=\sum_{i=1}^{n}(X_i-\overline{X})^2\sim\chi_{n-1}^2$ に従う。W を与えたときの $\widehat{\beta}-\beta$ の条件付き分布は $\mathcal{N}(0,1/W)$ となるので，$\widehat{\beta}$ と W の同時確率密度を W に関して積分すると

$$\begin{aligned}&\int_0^\infty\frac{1}{\sqrt{2\pi}}w^{1/2}e^{-w(\widehat{\beta}-\beta)^2/2}\frac{1}{\Gamma((n-1)/2)}w^{(n-1)/2-1}\frac{1}{2^{(n-1)/2}}e^{-w/2}dw\\=&\frac{1}{\sqrt{2\pi}\Gamma((n-1)/2)2^{(n-1)/2}}\int_0^\infty w^{n/2-1}e^{-w\{(\widehat{\beta}-\beta)^2+1\}/2}dw\\=&\frac{1}{\sqrt{2\pi}\Gamma((n-1)/2)2^{(n-1)/2}}2^{n/2}\Gamma(n/2)\frac{1}{\{1+(\widehat{\beta}-\beta)^2\}^{n/2}}\\=&\frac{\Gamma(n/2)}{\sqrt{\pi}\Gamma((n-1)/2)}\frac{1}{\{1+(\widehat{\beta}-\beta)^2\}^{n/2}}\end{aligned}$$

となり，$\widehat{\beta}$ の周辺分布が得られる。

分散は

$$\mathrm{Var}(\widehat{\beta}) = \mathrm{E}\Big[\mathrm{Var}(\widehat{\beta}|\boldsymbol{X})\Big] + \mathrm{Var}\Big(\mathrm{E}[\widehat{\beta}|\boldsymbol{X}]\Big) = \mathrm{E}\Big[\frac{1}{\sum_{i=1}^n (X_i - \overline{X})^2}\Big] = \mathrm{E}[\frac{1}{W}] = \frac{1}{n-3}$$

となる。

問 11 (X, Y) が 2 次元正規分布に従うとき，標本相関係数を $R_{XY} = S_{XY}/(S_X S_Y)$ に対して

$$T = \sqrt{n-2}R_{XY}/\sqrt{1-R_{XY}^2}$$

が，$H_0 : \rho = 0$ のもとで t-分布 t_{n-2} に従うことを示す。

命題 3.4 より，$X = x$ を与えたときの Y の条件付き分布は

$$Y|X = x \sim \mathcal{N}(\mu_y + \rho(\sigma_y/\sigma_x)(x-\mu_x), \sigma_y^2(1-\rho^2))$$

と書ける。ここで，$\alpha = \mu_y - \rho(\sigma_y/\sigma_x)\mu_x$, $\beta = \rho(\sigma_y/\sigma_x)$, $\sigma^2 = \sigma_y^2(1-\rho^2)$ とおくと

$$Y_i = \alpha + \beta X_i + u_i, \quad i = 1, \ldots, n$$

と書けて，$u_i \sim \mathcal{N}(0, \sigma^2)$ と表される。これは単回帰モデルであり，$H_0 : \rho = 0$ についての検定は $H_0 : \beta = 0$ の検定に等しい。β の最小 2 乗推定量は

$$\widehat{\beta} = S_{XY}/S_X^2$$

であり，σ^2 の残差平方和は

$$\begin{aligned}\mathrm{RSS} =& \sum_{i=1}^n (Y_i - \overline{Y} - \widehat{\beta}(X_i - \overline{X}))^2 = n(X_Y^2 + \widehat{\beta}^2 S_X^2 - 2\widehat{\beta}S_{XY}) \\ =& n(S_X^2 S_Y^2 - S_{XY}^2)/S_X^2 = n(1-R_{XY}^2)S_Y^2\end{aligned}$$

と書けるので，σ^2 の不偏推定量は $\hat{\sigma}^2 = n(1-R_{XY}^2)S_Y^2/(n-2)$ となる。$\widehat{\beta}$ と $\hat{\sigma}^2$ の分布は

$$\widehat{\beta} \sim \mathcal{N}(\beta, \sigma^2/(nS_X^2)), \quad (n-2)\hat{\sigma}^2/\sigma^2 \sim \chi_{n-2}^2$$

となる。従って，$H_0 : \beta = 0$ のもとで

$$\sqrt{n}S_X\widehat{\beta}/\hat{\sigma} \sim t_{n-2}$$

に従うことがわかる。ここで

$$\frac{\sqrt{n}S_X\widehat{\beta}}{\hat{\sigma}} = \frac{\sqrt{n}S_X S_{XY}/S_X^2}{\sqrt{n(1-R_{XY}^2 S_Y^2/(n-2}} = \sqrt{n-2}\frac{R_{XY}}{\sqrt{1-R_{XY}^2}} = T$$

となるので，結局，$H_0 : \hat{\rho} = 0$ のもとで $T \sim t_{n-2}$ となる。

第12章　重回帰モデル

問1　多項式回帰モデル (12.20) について次の問に答えよ。

(1) このモデルの行列表現は

$$\boldsymbol{y}=\begin{pmatrix}y_1\\ \vdots\\ y_n\end{pmatrix},\ \boldsymbol{X}=\begin{pmatrix}1 & x_1 & x_1^2 & \dots & x_1^{k-1}\\ \vdots & \vdots & \vdots & & \vdots\\ 1 & x_n & x_n^2 & \dots & x_n^{k-1}\end{pmatrix},\ \boldsymbol{\beta}=\begin{pmatrix}\beta_0\\ \vdots\\ \beta_{k-1}\end{pmatrix},\ \boldsymbol{u}=\begin{pmatrix}u_1\\ \vdots\\ u_n\end{pmatrix}$$

とおくとき，

$$\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{u}$$

と表される。

(2) 仮説検定 $H_0:\beta_1=\cdots=\beta_{k-1}=0$ vs. $H_1:\beta_1,\dots,\beta_{k-1}$ のどれかが 0 でない，を考える。有意水準 α の F 検定は，式 (12.13) を用いればよいので，

$$\mathrm{RSS}_k=\boldsymbol{y}^\top(\boldsymbol{I}-\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top)\boldsymbol{y},\quad \mathrm{RSS}_1=\sum_{i=1}^n(y_i-\overline{y})^2$$

に対して

$$F=\frac{(\mathrm{RSS}_1-\mathrm{RSS}_k)/(k-1)}{\mathrm{RSS}_k/(n-k)}>F_{k-1,n-k,\alpha}\Longrightarrow\ H_0\text{を棄却}$$

で与えられる。

(3) 表 11.3 の土地価格と通勤時間のデータを (12.13) の多項式モデル M_3 を当てはめるとき，$H_0:\beta_1=\beta_2=0$ を有意水準 1% で検定する。$\mathrm{RSS}_1=645983.9$，$\mathrm{RSS}_3=69981.1$ より

$$F=\frac{(645983.9-69981.1)/2}{69981.1/37}=152.3>F_{2,37,0.01}=7.37$$

となり，有意となる。

問2　線形回帰モデル $\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{u}$ において，$\boldsymbol{u}$ の平均と共分散行列が $\mathrm{E}[\boldsymbol{u}]=\boldsymbol{0}$, $\mathbf{Cov}(\boldsymbol{u})=\sigma^2\boldsymbol{I}$ で与えられるとする。説明変数行列 $\boldsymbol{X}$ を $\boldsymbol{X}=(\boldsymbol{x}_{(1)},\dots,\boldsymbol{x}_{(k)})$ と表すとき，$i\neq j$ に対して列ベクトル $\boldsymbol{x}_i$, $\boldsymbol{x}_j$ は直交する，すなわち $\boldsymbol{x}_i^\top\boldsymbol{x}_j=0$ を満たすとして以下の問に答えよ。

(1) $\boldsymbol{\beta}=(\beta_1,\dots,\beta_k)^\top$ の最小 2 乗推定量 $\widehat{\boldsymbol{\beta}}=(\widehat{\beta}_1,\dots,\widehat{\beta}_k)^\top$ は $\widehat{\boldsymbol{\beta}}=(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}$ で与えられる。ここで

$$\boldsymbol{X}^\top\boldsymbol{X}=\begin{pmatrix}\boldsymbol{x}_{(1)}^\top\\ \vdots\\ \boldsymbol{x}_{(k)}^\top\end{pmatrix}(\boldsymbol{x}_{(1)},\dots,\boldsymbol{x}_{(k)})$$

であり，$i\neq j$ に対して $\boldsymbol{x}_{(i)}^\top\boldsymbol{x}_{(j)}=0$ より

$$\boldsymbol{X}^\top\boldsymbol{X}=\mathrm{diag}\,(\|\boldsymbol{x}_{(1)}\|^2,\dots,\|\boldsymbol{x}_{(k)}\|^2)$$

となる。また

$$\boldsymbol{X}^\top \boldsymbol{y} = \begin{pmatrix} \boldsymbol{x}_{(1)}^\top \boldsymbol{y} \\ \vdots \\ \boldsymbol{x}_{(k)}^\top \boldsymbol{y} \end{pmatrix}$$

と書けるので

$$\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_1, \ldots, \widehat{\beta}_k)^\top, \quad \widehat{\beta}_i = \frac{\boldsymbol{x}_{(i)}^\top \boldsymbol{y}}{\|\boldsymbol{x}_{(i)}\|^2}$$

となる。

(2) $\widehat{\boldsymbol{\beta}}$ の共分散行列は

$$\mathbf{Cov}(\widehat{\boldsymbol{\beta}}) = \sigma^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1} = \sigma^2 \mathrm{diag}\left(\frac{1}{\|\boldsymbol{x}_{(1)}\|^2}, \ldots, \frac{1}{\|\boldsymbol{x}_{(k)}\|^2}\right)$$

と表されるので，$i \neq j$ のとき $\widehat{\beta}_i$ と $\widehat{\beta}_j$ は無相関であることがわかる。

(3) σ^2 の不偏推定量は $\hat{\sigma}^2 = (n-k)^{-1} \boldsymbol{y}^\top (\boldsymbol{I} - \boldsymbol{P}) \boldsymbol{y}$ で与えられる。ここで

$$\begin{aligned} \boldsymbol{P} =& \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top = (\boldsymbol{x}_{(1)}, \ldots, \boldsymbol{x}_{(k)}) \mathrm{diag}\left(\frac{1}{\|\boldsymbol{x}_{(1)}\|^2}, \ldots, \frac{1}{\|\boldsymbol{x}_{(k)}\|^2}\right) \begin{pmatrix} \boldsymbol{x}_{(1)}^\top \\ \vdots \\ \boldsymbol{x}_{(k)}^\top \end{pmatrix} \\ =& \sum_{i=1}^k \frac{1}{\|\boldsymbol{x}_{(i)}\|^2} \boldsymbol{x}_{(i)} \boldsymbol{x}_{(i)}^\top \end{aligned}$$

より

$$\begin{aligned} \hat{\sigma}^2 =& \frac{1}{n-k} \boldsymbol{y}^\top \Big(\boldsymbol{I} - \sum_{i=1}^k \frac{1}{\|\boldsymbol{x}_{(i)}\|^2} \boldsymbol{x}_{(i)} \boldsymbol{x}_{(i)}^\top \Big) \boldsymbol{y} \\ =& \frac{1}{n-k} \Big\{ \|\boldsymbol{y}\|^2 - \sum_{i=1}^k \frac{(\boldsymbol{x}_{(i)}^\top \boldsymbol{y})^2}{\|\boldsymbol{x}_{(i)}\|^2} \Big\} \end{aligned}$$

と書ける。

(4) $\boldsymbol{u} \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$ を仮定すると，

$$\widehat{\beta}_i \sim \mathcal{N}\Big(\beta_i, \ \frac{\sigma^2}{\|\boldsymbol{x}_{(i)}\|^2}\Big)$$

となり，

$$\frac{\|\boldsymbol{x}_{(i)}\|(\widehat{\beta}_i - \beta_i)}{\hat{\sigma}} \sim t_{n-k}$$

に従うことがわかる。従って，仮説検定 $H_0 : \beta_j = \beta_{j0}$ vs. $H_1 : \beta_j \neq \beta_{j0}$ について，有意水準 α の検定は

$$\frac{\|\boldsymbol{x}_{(i)}\| |\widehat{\beta}_i - \beta_{j0}|}{\hat{\sigma}} > t_{n-k,\alpha/2} \Longrightarrow H_0 \text{を棄却}$$

で与えられる。また，信頼係数 $1-\gamma$ の β_j の信頼区間は

$$\widehat{\beta}_j \pm \frac{\hat{\sigma}}{\|\boldsymbol{x}_{(i)}\|} t_{n-k,\alpha/2}$$

で与えられる。

問 3　線形回帰モデル $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$ において $\boldsymbol{X}$ は $n \times k$ のフルランク行列とし，$\boldsymbol{u}$ の平均と共分散行列が $\mathrm{E}[\boldsymbol{u}] = \boldsymbol{0}$, $\mathbf{Cov}(\boldsymbol{u}) = \sigma^2 \boldsymbol{I}$ で与えられるとする。$\widehat{\boldsymbol{\beta}}$ を $\boldsymbol{\beta}$ の最小 2 乗推定量とすると，共分散行列は

$$\mathbf{Cov}(\widehat{\boldsymbol{\beta}}) = \sigma^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1}$$

となる。2 つのベクトル $\boldsymbol{a}, \boldsymbol{b}$ に対して $\boldsymbol{a}^\top \widehat{\boldsymbol{\beta}}$ と $\boldsymbol{b}^\top \widehat{\boldsymbol{\beta}}$ の共分散は

$$\begin{aligned}\mathrm{Cov}(\boldsymbol{a}^\top \widehat{\boldsymbol{\beta}}, \boldsymbol{b}^\top \widehat{\boldsymbol{\beta}}) =& \mathrm{E}[\boldsymbol{a}^\top (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \boldsymbol{b}^\top (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})] \\ =& \mathrm{E}[\boldsymbol{a}^\top (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \boldsymbol{b}] = \sigma^2 \boldsymbol{a}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{b}\end{aligned}$$

となるので，$\boldsymbol{a}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{b} = 0$ のとき，$\boldsymbol{a}^\top \widehat{\boldsymbol{\beta}}$ と $\boldsymbol{b}^\top \widehat{\boldsymbol{\beta}}$ は無相関になる。

問 4 $\mathcal{N}(0, 1)$ に従う独立な確率変数 Z_1, Z_2 に対して $U_1 = Z_1 + Z_2$, $U_2 = Z_1 - Z_2$ とおく。U_1^2, U_2^2 が独立に χ_1^2 分布に従うことをコクランの定理を用いて示す。

$U_1^2 = Z_1^2 + Z_2^2 + 2Z_1Z_2$, $U_2^2 = Z_1^2 + Z_2^2 - 2Z_1Z_2$ より, $\boldsymbol{Z} = (Z_1, Z_2)^\top$ に対して

$$\begin{aligned}Z_1^2 + Z_2^2 =& \frac{1}{2}U_1^2 + \frac{1}{2}U_2^2 \\ =& \boldsymbol{Z}^\top \boldsymbol{A} \boldsymbol{Z} + \boldsymbol{Z}^\top \boldsymbol{B} \boldsymbol{Z}\end{aligned}$$

と表される。ただし

$$\boldsymbol{A} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad \boldsymbol{B} = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

である。ここで，$\boldsymbol{A}^2 = \boldsymbol{A}$, $\boldsymbol{B}^2 = \boldsymbol{B}$ より，$\boldsymbol{A}$, $\boldsymbol{B}$ ともに巾等行列であり，$\mathrm{rank}\,(\boldsymbol{A}) = \mathrm{rank}\,(\boldsymbol{B}) = 1$ である。ランクについては

$$2 = 1 + 1 = \mathrm{rank}\,(\boldsymbol{A}) + \mathrm{rank}\,(\boldsymbol{B})$$

となるので，定理 12.5 のコクランの定理より，$\boldsymbol{Z}^\top \boldsymbol{A} \boldsymbol{Z} = U_1^2/2$ と $\boldsymbol{Z}^\top \boldsymbol{B} \boldsymbol{Z} = U_2^2/2$ は独立で

$$\boldsymbol{Z}^\top \boldsymbol{A} \boldsymbol{Z} \sim \chi_1^2, \quad \boldsymbol{Z}^\top \boldsymbol{B} \boldsymbol{Z} \sim \chi_1^2$$

に従うことがわかる。

問 5　線形回帰モデル $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$ において $\boldsymbol{u} \sim \mathcal{N}_n(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$ とする。$\boldsymbol{X}$ は $n \times k$ のフルランクな行列とし，$\boldsymbol{C}$ を $r \times k$ のフルランクな行列とし，$r \leq k$ とする。このとき，仮説検定 $H_0 : \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{d}$ vs. $H_1 : \boldsymbol{C}\boldsymbol{\beta} \neq \boldsymbol{d}$ を考える。ただし，$\boldsymbol{d}$ は r 次元の既知のベクトルである。

$$\begin{aligned}\widehat{\boldsymbol{\beta}} =& (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1}) \\ \hat{\sigma}^2 =& \frac{1}{n-k} \|\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\|^2 \sim \frac{\sigma^2}{n-k} \chi_{n-k}^2\end{aligned}$$

(1) H_0 のもとでの $\boldsymbol{\beta}$, σ^2 の MLE を求めるには，$\boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{d}$ のもとで対数尤度を最大化する必要がある。ラグランジュの未定乗数法により

$$g(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 - \boldsymbol{\lambda}^\top (\boldsymbol{C}\boldsymbol{\beta} - \boldsymbol{d})$$

を考える。このとき

$$\begin{aligned}\frac{\partial g}{\partial \boldsymbol{\beta}} =& \frac{1}{\sigma^2} \boldsymbol{X}^\top (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) - \boldsymbol{C}^\top \boldsymbol{\lambda} = \boldsymbol{0} \\ \frac{\partial g}{\partial \sigma^2} =& -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 = 0\end{aligned}$$

となり，最初の等式から

$$\boldsymbol{\beta} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}(\boldsymbol{X}^\top \boldsymbol{y} - \sigma^2 \boldsymbol{C}^\top \boldsymbol{\lambda})$$

となる。これを $\boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{d}$ に代入すると $\boldsymbol{C}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}(\boldsymbol{X}^\top \boldsymbol{y} - \sigma^2 \boldsymbol{C}^\top \boldsymbol{\lambda}) = \boldsymbol{d}$ より

$$\boldsymbol{\lambda} = \frac{1}{\sigma^2}\{\boldsymbol{C}(\boldsymbol{X}^\rightarrow \boldsymbol{X})^{-1}\boldsymbol{C}^\top\}^{-1}(\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{d})$$

が得られる。これを $\boldsymbol{\beta} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}(\boldsymbol{X}^\top \boldsymbol{y} - \sigma^2 \boldsymbol{C}^\top \boldsymbol{\lambda})$ に代入すると

$$\widehat{\boldsymbol{\beta}}_{H_0} = \widehat{\boldsymbol{\beta}} - (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{C}^\top\{\boldsymbol{C}(\boldsymbol{X}^\rightarrow \boldsymbol{X})^{-1}\boldsymbol{C}^\top\}^{-1}(\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{d})$$

となる。これを上の 2 番目の等式に代入すると

$$\begin{aligned}\hat{\sigma}^2_{H_0} =& \frac{1}{n}\|\boldsymbol{y} - \widehat{\boldsymbol{\beta}} + (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{C}^\top\{\boldsymbol{C}(\boldsymbol{X}^\rightarrow \boldsymbol{X})^{-1}\boldsymbol{C}^\top\}^{-1}(\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{d})\|^2 \\ =& \frac{1}{n}\|\boldsymbol{y} - \widehat{\boldsymbol{\beta}}\|^2 + \frac{1}{n}(\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{d})^\top\{\boldsymbol{C}(\boldsymbol{X}^\rightarrow \boldsymbol{X})^{-1}\boldsymbol{C}^\top\}^{-1}(\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{d})\end{aligned}$$

となる。$\widehat{\boldsymbol{\beta}}_{H_0}$, $\hat{\sigma}^2_{H_0}$ が H_0 のもとでの $\boldsymbol{\beta}$, σ^2 の MLE である。

(2) 有意水準 α の尤度比検定は

$$\begin{aligned}-2\log\Lambda =& 2\ell(\widehat{\boldsymbol{\beta}}, \hat{\sigma}^2) - 2\ell(\widehat{\boldsymbol{\beta}}_{H_0}, \sigma^2_{H_0}) = -n\log\sigma^2 + n\log\sigma^2_{H_0} \\ =& n\log\left(\frac{\sigma^2_{H_0}}{\hat{\sigma}^2}\right) > C'\end{aligned}$$

のとき H_0 を棄却する方法になる。棄却域は $\sigma^2_{H_0}/\hat{\sigma}^2 > C''$ のように書けるので，尤度比検定は F 分布を用いて

$$F = \frac{(\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{d})^\top\{\boldsymbol{C}(\boldsymbol{X}^\rightarrow \boldsymbol{X})^{-1}\boldsymbol{C}^\top\}^{-1}(\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{d})/r}{\hat{\sigma}^2} > F_{r,n-k,\alpha} \Longrightarrow H_0\text{を棄却}$$

のような形で表すことができる。

問 6 線形回帰モデル $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$ において $\boldsymbol{X}$ は $n \times k$ のフルランク行列とし，$\boldsymbol{u}$ の平均と共分散行列が $\mathrm{E}[\boldsymbol{u}] = \boldsymbol{0}$, $\mathbf{Cov}(\boldsymbol{u}) = \sigma^2\boldsymbol{I}$ で与えられるとする。重相関係数を R とするとき，$\mathrm{R}^2 = 1 - \sum_{i=1}^n(y_i - \hat{y}_i)^2/\sum_{i=1}^n(y_i - \overline{y})^2$ と書けることを示す。

線形回帰モデルは第 1 項に定数項を含むので行列 $\boldsymbol{X}$ は $\boldsymbol{X} = (\mathbf{1}_n, \boldsymbol{X}_2)$ のように表すことができる。$\boldsymbol{X}^\top(\boldsymbol{I} - \boldsymbol{P}) = \boldsymbol{0}$，すなわち $\mathbf{1}_n^\top(\boldsymbol{I} - \boldsymbol{P}) = \boldsymbol{0}$, $\boldsymbol{X}_2^\top(\boldsymbol{I} - \boldsymbol{P}) = \boldsymbol{0}$ を意味する。これより，

$$\mathbf{1}_n^\top\boldsymbol{P} = \mathbf{1}_n^\top, \quad \text{もしくは} \quad \boldsymbol{P}\mathbf{1}_n = \mathbf{1}_n$$

が成り立つ。$\hat{\boldsymbol{y}}$ は $\hat{\boldsymbol{y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{P}\boldsymbol{y}$ と書けるので，

$$\mathbf{1}_n^\top\hat{\boldsymbol{y}} = \mathbf{1}^\top\boldsymbol{P}\boldsymbol{y} = \mathbf{1}^\top\boldsymbol{u} = n\overline{y}$$

となることがわかる。記号の簡単のために $\boldsymbol{u}^\top\boldsymbol{u} = \|\boldsymbol{u}\|^2$ という記号を用いて表すことにする。$\|\cdot\|$ をノルムとよぶ。$\sum_{i=1}^n(y_i - \overline{y})^2 = (\boldsymbol{y} - \overline{y}\mathbf{1}_n)^\top(\boldsymbol{y} - \overline{y}\mathbf{1}_n) = \|\boldsymbol{y} - \overline{y}\mathbf{1}_n\|^2$ であり，

$$\begin{aligned}(\boldsymbol{y} - \overline{y}\mathbf{1}_n)^\top(\boldsymbol{P}\boldsymbol{y} - \overline{y}\mathbf{1}_n) =& \{(\boldsymbol{y} - \boldsymbol{P}\boldsymbol{y}) + (\boldsymbol{P}\boldsymbol{y} - \overline{y}\mathbf{1}_n)\}^\top(\boldsymbol{P}\boldsymbol{y} - \overline{y}\mathbf{1}_n) \\ =& \|\boldsymbol{P}\boldsymbol{y} - \overline{y}\mathbf{1}_n\|^2 + \boldsymbol{y}^\top(\boldsymbol{I} - \boldsymbol{P})(\boldsymbol{P}\boldsymbol{y} - \overline{y}\mathbf{1}_n)\end{aligned}$$

と書き直す。ここで

$$\begin{aligned}(\boldsymbol{I} - \boldsymbol{P})(\boldsymbol{P}\boldsymbol{y} - \overline{y}\mathbf{1}_n) =& (\boldsymbol{I} - \boldsymbol{P})(\boldsymbol{P}\boldsymbol{y} - \overline{y}\boldsymbol{P}\mathbf{1}_n) = (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{P}(\boldsymbol{y} - \overline{y}\mathbf{1}_1) \\ =& (\boldsymbol{P} - \boldsymbol{P}\boldsymbol{P})(\boldsymbol{y} - \overline{y}\mathbf{1}_n) = \boldsymbol{0}\end{aligned}$$

となることに注意する。よって，等式

$$(\boldsymbol{y}-\overline{y}\mathbf{1}_n)^\top(\boldsymbol{P}\boldsymbol{y}-\overline{y}\mathbf{1}_n)=\|\boldsymbol{P}\boldsymbol{y}-\overline{y}\mathbf{1}_n\|^2$$

が成り立つ。従って

$$\begin{aligned}R^2=&\frac{\{\sum_{i=1}^n(y_i-\overline{y})(\hat{y}_i-\overline{y})\}^2}{\sum_{i=1}^n(y_i-\overline{y})^2\sum_{i=1}^n(\hat{y}_i-\overline{y})^2}=\frac{\{(\boldsymbol{y}-\overline{y}\mathbf{1}_n)^\top(\boldsymbol{P}\boldsymbol{y}-\overline{y}\mathbf{1}_n)\}^2}{\|\boldsymbol{y}-\overline{y}\mathbf{1}_n\|^2\|\boldsymbol{P}\boldsymbol{y}-\overline{y}\mathbf{1}_n\|^2}\\=&\frac{\|\boldsymbol{P}\boldsymbol{y}-\overline{y}\mathbf{1}_n\|^4}{\|\boldsymbol{y}-\overline{y}\mathbf{1}_n\|^2\|\boldsymbol{P}\boldsymbol{y}-\overline{y}\mathbf{1}_n\|^2}=\frac{\|\boldsymbol{P}\boldsymbol{y}-\overline{y}\mathbf{1}_n\|^2}{\|\boldsymbol{y}-\overline{y}\mathbf{1}_n\|^2}\end{aligned}$$

と書ける。ここで

$$(\boldsymbol{y}-\boldsymbol{P}\boldsymbol{y})^\top(\boldsymbol{P}\boldsymbol{y}-\overline{y}\mathbf{1}_n)=\boldsymbol{y}^\top(\boldsymbol{I}-\boldsymbol{P})(\boldsymbol{P}\boldsymbol{y}-\overline{y}\boldsymbol{P}\mathbf{1}_n)=\boldsymbol{y}^\top(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{P}(\boldsymbol{y}-\overline{y}\mathbf{1}_n)=0$$

に注意すると

$$\begin{aligned}\|\boldsymbol{y}-\overline{y}\mathbf{1}_n\|^2=&\|(\boldsymbol{y}-\boldsymbol{P}\boldsymbol{y})+(\boldsymbol{P}\boldsymbol{y}-\overline{y}\mathbf{1}_n)\|^2\\=&\|\boldsymbol{y}-\boldsymbol{P}\boldsymbol{y}\|^2+\|\boldsymbol{P}\boldsymbol{y}-\overline{y}\mathbf{1}_n\|^2+2(\boldsymbol{y}-\boldsymbol{P}\boldsymbol{y})^\top(\boldsymbol{P}\boldsymbol{y}-\overline{y}\mathbf{1}_n)\\=&\|\boldsymbol{y}-\boldsymbol{P}\boldsymbol{y}\|^2+\|\boldsymbol{P}\boldsymbol{y}-\overline{y}\mathbf{1}_n\|^2\end{aligned}$$

となる。この等式を用いると

$$R^2=1-\frac{\|\boldsymbol{y}-\boldsymbol{P}\boldsymbol{y}\|^2}{\|\boldsymbol{y}-\overline{y}\mathbf{1}_n\|^2}=1-\frac{\sum_{i=1}^n(y_i-\hat{y}_i)^2}{\sum_{i=1}^n(y_i-\overline{y})^2}$$

と書けることがわかる。

第13章　ロジスティック回帰とポアソン回帰

問1　連続な確率変数 X の確率密度関数が $f(x) = e^{-x}(1+e^{-x})^{-2}$ で与えられるときロジスティック分布とよぶ。

(1) 原点に関して対称であることは

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

からわかる。また

$$F(x) = \frac{1}{1+e^{-x}}$$

となることは，微分して $F'(x) = f(x)$ となることで確かめることができる。

$$F(x) = p \Longleftrightarrow e^{-x} = \frac{1-p}{p} \Longleftrightarrow x = \log\frac{p}{1-p}$$

より

$$F^{-1}(p) = \log\left\{\frac{p}{1-p}\right\}$$

である。

(2) 積率母関数 $M(t) = \mathrm{E}[e^{tX}]$ は

$$M(t) = \int_{-\infty}^{\infty} e^{tx}\frac{e^{-x}}{(1+e^{-x})^2}dx$$

であり，$u = 1/(1+e^{-x})$ とおいて変数変換を行うと，

$$du = \frac{e^{-x}}{(1+e^{-x})^2}dx, \quad x = \log(u) - \log(1-u)$$

より

$$\begin{aligned}M(t) &= \int_0^1 e^{t\{\log u - \log(1-u)\}}du = \int_0^1 \left(\frac{u}{1-u}\right)^t du = \int_0^1 u^t(1-u)^{-t}du \\ &= \int_0^1 u^{t+1-1}(1-u)^{1-t-1}du = B(t+1, 1-t) = \frac{\Gamma(t+1)\Gamma(1-t)}{\Gamma(2)} \\ &= \Gamma(t+1)\Gamma(1-t), \ |t| < 1\end{aligned}$$

となる。

(3) $M(t)$ を微分すると

$$M'(t) = \Gamma'(t+1)\Gamma(1-t) - \Gamma(t+1)\Gamma'(1-t)$$

であるから，$\mathrm{E}[X] = M'(0) = \Gamma'(1)\Gamma(1) - \Gamma(1)\Gamma'(1) = 0$ となる。

問2　n 個の2次元データ $(y_1, x_1), \ldots, (y_n, x_n)$ が観測され，各 y_i は0か1の2値をとり，x_i は y_i の共変量であるとする。

(1) ロジスティック回帰モデルの尤度関数は,

$$p_i = \frac{1}{1+e^{-\alpha-\beta x_i}}$$

に対して

$$L(\alpha,\beta) = \prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{1-y_i} = \prod_{i=1}^{n} \frac{e^{-(\alpha+\beta x_i)(1-y_i)}}{1+e^{-(\alpha+\beta x_i)}} = \prod_{i=1}^{n} \frac{e^{(\alpha+\beta x_i)y_i}}{1+e^{\alpha+\beta x_i}}$$

であり，対数尤度は

$$\ell(\alpha,\beta) = \sum_{i=1}^{n} \Big\{(\alpha+\beta x_i)y_i - \log(1+e^{\alpha+\beta x_i})\Big\}$$

となる。尤度方程式は

$$\begin{aligned}
\frac{\partial \ell(\alpha,\beta)}{\partial \alpha} &= \sum_{i=1}^{n} \Big(y_i - \frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}\Big) = 0 \\
\frac{\partial \ell(\alpha,\beta)}{\partial \beta} &= \sum_{i=1}^{n} x_i\Big(y_i - \frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}\Big) = 0
\end{aligned}$$

で与えられる。

(2) プロビット回帰モデルの場合は，標準正規分布の分布関数 $\Phi(\cdot)$ を用いて

$$p_i = \Phi(\alpha+\beta x_i)$$

で与えられるから，尤度関数は

$$L(\alpha,\beta) = \prod_{i=1}^{n} \Big\{\Phi(\alpha+\beta x_i)\Big\}^{y_i} \Big\{1-\Phi(\alpha+\beta x_i)\Big\}^{1-y_i}$$

であり，対数尤度は

$$\ell(\alpha,\beta) = \sum_{i=1}^{n} \Big[y_i \log\Phi(\alpha+\beta x_i) + (1-y_i)\log\{1-\Phi(\alpha+\beta x_i)\}\Big]$$

となる。尤度方程式は，標準正規分布の確率密度関数 $\phi(\cdot)$ を用いて

$$\begin{aligned}
\frac{\partial \ell(\alpha,\beta)}{\partial \alpha} &= \sum_{i=1}^{n} \Big\{\frac{y_i}{\Phi(\alpha+\beta x_i)} - \frac{1-y_i}{1-\Phi(\alpha+\beta x_i)}\Big\}\phi(\alpha+\beta x_i) = 0 \\
\frac{\partial \ell(\alpha,\beta)}{\partial \beta} &= \sum_{i=1}^{n} x_i\Big\{\frac{y_i}{\Phi(\alpha+\beta x_i)} - \frac{1-y_i}{1-\Phi(\alpha+\beta x_i)}\Big\}\phi(\alpha+\beta x_i) = 0
\end{aligned}$$

で与えられる。

$$\frac{y_i}{\Phi(\alpha+\beta x_i)} - \frac{1-y_i}{1-\Phi(\alpha+\beta x_i)} = \frac{y_i-\Phi(\alpha+\beta x_i)}{\Phi(\alpha+\beta x_i)\{1-\Phi(\alpha+\beta x_i)\}}$$

と書けるので,

$$\begin{aligned}
&\sum_{i=1}^{n} \frac{y_i-\Phi(\alpha+\beta x_i)}{\Phi(\alpha+\beta x_i)\{1-\Phi(\alpha+\beta x_i)\}}\phi(\alpha+\beta x_i) = 0 \\
&\sum_{i=1}^{n} \frac{x_i\{y_i-\Phi(\alpha+\beta x_i)\}}{\Phi(\alpha+\beta x_i)\{1-\Phi(\alpha+\beta x_i)\}}\phi(\alpha+\beta x_i) = 0
\end{aligned}$$

と書ける。

問 3 ある保険会社では自動車事故による高額請求の要因の一つとして年齢を考え，ある都市の交通事故データから因果関係を調べることを計画している。年齢階級を 18〜22 歳，23〜30 歳，30〜50 歳，51〜65 歳，>66 歳の 5 段階に分け，それぞれの年齢階級の登録ドライバー数を n_1, n_2, n_3, n_4, n_5 とし，高額請求の件数を y_1, y_2, y_3, y_4, y_5 とする。

(1) 高額請求の件数を表す確率変数を Y_i, $i=1,\ldots,5$, とすると，希な現象に対応するのでポアソン分布を当てはめて

$$Y_i \sim Po(n_i\lambda_i), \quad i=1,\ldots,5$$

とするのが自然である。年齢階級の 18〜22 歳，23〜30 歳，30〜50 歳，51〜65 歳，>66 歳の 5 段階を小さい順に $1,2,3,4,5$ とし，この変数を x_i とする。高額請求の生起率と年齢とを関係づけるために

$$\log\lambda_i = \alpha + \beta x_i$$

のようなリンク関数を用いると, $\lambda_i = e^{\alpha+\beta x_i}$ と書ける。このとき，尤度関数は

$$\prod_{i=1}^{5}\frac{1}{y_i!}n_i^{y_i}e^{(\alpha+\beta x_i)y_i}e^{-n_ie^{\alpha+\beta x_i}}$$

で与えられる。

（補足）5 つの集計データ $y_1,\ldots,y_5$ に基づいてモデルを立てるので，パラメータの数を多くとることができないので，単純な線形式で記述してみた。集計データの数が多ければ，$\log\lambda_i = \alpha+\beta_1x_i+\beta_2x_i^2$ のような多項式を考えてよい。

(2) 対数尤度は

$$\ell(\alpha,\beta) = \sum_{i=1}^{5}\Big\{-\log y_i! + y_i\log n_i + (\alpha+\beta x_i)y_i - n_ie^{\alpha+\beta x_i}\Big\}$$

と書けるので，尤度方程式は

$$\frac{\partial\ell}{\partial\alpha} = \sum_{i=1}^{5}\Big(y_i - n_ie^{\alpha+\beta x_i}\Big) = 0$$
$$\frac{\partial\ell}{\partial\beta} = \sum_{i=1}^{5}x_i\Big(y_i - n_ie^{\alpha+\beta x_i}\Big) = 0$$

で与えられる。この解が MLE になり，それを $\widehat{\alpha}$, $\widehat{\beta}$ と記すことにする。

$H_0:\beta=0$ を検定するので，H_0 の元での α の MLE を求めると，対数尤度が $\ell(\alpha)\propto\sum_{i=1}^{5}(y_i\alpha - n_ie^{\alpha})$ で与えられるので, 尤度方程式は

$$\ell'(\alpha) = \sum_{i=1}^{5}(y_i - n_ie^{\alpha}) = 0$$

となり，解は $\widehat{e^{\alpha}} = \sum_{i=1}^{5}y_i/\sum_{i=1}^{5}n_i$ となり，これを $\overline{y}$ で表す。このとき尤度比は

$$\Lambda = \frac{\prod_{i=1}^{5}(\overline{y})^{y_i}e^{-n_i\overline{y}}}{\prod_{i=1}^{5}e^{(\widehat{\alpha}+\widehat{\beta}x_i)y_i}e^{-n_ie^{\widehat{\alpha}+\widehat{\beta}x_i}}}$$

と書かれるので，年齢階級との因果関係に関して有意水準 α の尤度比検定は

$$-2\log\Lambda = 2\sum_{i=1}^{5}\Big\{(\widehat{\alpha}+\widehat{\beta}x_i)y_i - n_ie^{\widehat{\alpha}+\widehat{\beta}x_i} - y_i\log\overline{y} + n_i\overline{y}\Big\} > \chi^2_{1,\alpha} \Longrightarrow H_0\text{を棄却}$$

で与えられる。

第14章　ベイズ統計とMCMC法

問1　T が十分統計量のとき，尤度関数は $f_n(\boldsymbol{x}|th) = g_\theta(T(\boldsymbol{x}))h(\boldsymbol{x})$ と書ける。$\pi(\theta)$ を事前分布とすると，事後分布は

$$\pi(\theta|\boldsymbol{x}) = \frac{g_\theta(T(\boldsymbol{x}))h(\boldsymbol{x})\pi(\theta)}{\int g_\theta(T(\boldsymbol{x}))h(\boldsymbol{x})\pi(\theta)d\theta} = \frac{g_\theta(T(\boldsymbol{x}))\pi(\theta)}{\int g_\theta(T(\boldsymbol{x}))\pi(\theta)d\theta}$$

となり，$T(\boldsymbol{x})$ の関数になることがわかる。従って，ベイズ推定量やベイズ信用区間は十分統計量の関数になる。

問2　確率変数 $X_1, \ldots, X_n$ が独立に一様分布分布 $U(0, \theta)$ に従うとする。$\theta > 0$ として事前分布 $\pi(\theta) = 1/\theta^2$ を考える。

(1) この事前分布は

$$\int_0^\infty \frac{1}{\theta^2}d\theta = \Big[-\frac{1}{\theta}\Big]_0^\infty = \infty$$

となり，正則でない。$(\boldsymbol{X}, \theta)$ の同時確率密度は

$$f_n(\boldsymbol{x}|\theta)\pi(\theta) = \prod_{i=1}^n \Big\{\frac{1}{\theta}I(0 < x_i < \theta)\Big\} \times \frac{1}{\theta^2} = \frac{1}{\theta^{n+2}}I(x_{(n)} < \theta)$$

と書けるので，事後分布は

$$\pi(\theta|\boldsymbol{x}) = \frac{\theta^{-(n+2)}I(x_{(n)} < \theta)}{\int_0^\infty \theta^{-(n+2)}I(x_{(n)} < \theta)d\theta}$$

であるが，

$$\int_0^\infty \theta^{-(n+2)}I(x_{(n)} < \theta)d\theta = \int_{x_{(n)}}^\infty \theta^{-(n+2)}d\theta = \Big[-\frac{1}{n+1}\frac{1}{\theta^{n+1}}\Big]_{x_{(n)}}^\infty = \frac{1}{n+1}\frac{1}{(x_{(n)})^{n+1}}$$

と書けるので，事後分布は

$$\pi(\theta|\boldsymbol{x}) = \frac{(n+1)(x_{(n)})^{n+1}}{\theta^{n+2}}I(x_{(n)} < \theta)$$

で与えられる。

(2) θ のベイズ推定量は

$$\mathrm{E}[\theta|x_{(n)}] = (n+1)(x_{(n)})^{n+1}\int_{x_{(n)}}^\infty \frac{\theta}{\theta^{n+2}}d\theta = (n+1)(x_{(n)})^{n+1}\Big[-\frac{1}{n\theta^n}\Big]_{x_{(n)}}^\infty = \frac{n+1}{n}x_{(n)}$$

より，

$$\hat{\theta}^B = \frac{n+1}{n}X_{(n)}$$

となる。MLE は $\hat{\theta}^M = X_{(n)}$ である。不偏推定量を求めるには $X_{(n)}$ の確率密度を与える必要があり，

$$f_{X_{(n)}}(x) = \frac{n}{\theta}\Big(\frac{x}{\theta}\Big)^{n-1}I(0 < x < \theta)$$

より, $y = x/\theta$ の変数変換を用いると

$$\mathrm{E}[X_{(n)}] = n\int_0^\theta \frac{x^n}{\theta^n}d\theta = n\theta\int_0^1 y^n dy = \frac{n}{n+1}\theta$$

となるので，θ の不偏推定量は

$$\hat{\theta}^U = \frac{n+1}{n}\boldsymbol{X}_{(n)}$$

となり，この場合はベイズ推定量に一致することがわかる。

(3) θ の事後分布の分散は

$$\mathrm{E}[\theta^2|x_{(n)}] = (n+1)(x_{(n)})^{n+1}\int_{x_{(n)}}^\infty \frac{\theta^2}{\theta^{n+2}}d\theta = \frac{n+1}{n-1}(x_{(n)})^2$$

より

$$\mathrm{Var}(\theta|\boldsymbol{x}) = \mathrm{E}[\theta^2|\boldsymbol{x}] - (\mathrm{E}[\theta|\boldsymbol{x}])^2 = \left\{\frac{n+1}{n-1} - \frac{(n+1)^2}{n^2}\right\}(x_{(n)})^2 = \frac{n+1}{(n-1)n^2}(x_{(n)})^2$$

となる。

(4) 信用係数 $1-\gamma$ の θ のベイズ信用区間を考える。(1) で与えられた事後確率関数は $\theta = x_{(n)}$ のときに最も大きくて，θ が大きくなるにつれて小さくなるので，最高事後密度信用区間 (HPD) は $[X_{(n)}, c]$ のような形をする。この c の値は信用確率に基づいて決める必要があるので

$$\int_{x_{(n)}}^c \pi(\theta|x_{(n)})d\theta = (n+1)(x_{(n)})^{n+1}\int_{x_{(n)}}^c \frac{1}{\theta^{n+2}}d\theta = (x_{(n)})^{n+1}\left\{\frac{1}{(x_{(n)})^{n+1}} - \frac{1}{c^{n+1}}\right\} = 1-\gamma$$

を満たすように c の値を定めればよい。上の方程式を解くと

$$c = \frac{x_{(n)}}{\gamma^{1/(n+1)}}$$

となるので，信用係数 $1-\gamma$ の θ のベイズ信用区間は

$$\left[X_{(n)},\ X_{(n)}\gamma^{-1/(n+1)}\right]$$

で与えられる。

問 3 確率変数 $X_1, \ldots, X_n$ が独立に指数分布 $f(x|\theta) = \exp\{-(x-\theta)\}I(x>\theta)$ に従うとする。θ の分布として，すべての θ に対して $\pi(\theta) = 1$ であるような非正則な事前分布を仮定する。

(1) 尤度関数は

$$\left\{\prod_{i=1}^n e^{-(x_i-\theta)}I(x_i>\theta)\right\} = e^{-\sum_{i=1}^n(x_i-\theta)}I(x_{(1)}>\theta) = e^{-\sum_{i=2}^n(x_{(i)}-x_{(1)})}e^{-n(x_{(1)}-\theta)}I(x_{(1)}>\theta)$$

と書けるので，θ の十分統計量は $X_{(1)}$ である。この尤度を最大にする θ は $\hat{\theta}^M = X_{(1)}$ であり，これが MLE となる。

不偏推定量については，$X_{(1)}$ の確率密度関数は, $F(x) = \int_\theta^x e^{-(t-\theta)}dt = 1 - e^{-(x-\theta)}$ より

$$f_{X_{(1)}}(x) = ne^{-(x-\theta)}\left\{e^{-(x-\theta)}\right\}^{n-1} = ne^{-n(x-\theta)},\ x>\theta$$

で与えられるので，変数変換 $y = x-\theta$ により

$$\begin{aligned}\mathrm{E}[X_{(1)}] =& n\int_\theta^\infty xe^{-n(x-\theta)}dx = n\int_0^\infty (y+\theta)e^{-ny}dy \\ =& n\frac{1}{n^2}\int_0^\infty \frac{1}{\Gamma(2)}y^{2-1}n^2e^{-ny}dy + n\theta\left[-\frac{1}{n}e^{-ny}\right]_0^\infty = \frac{1}{n}+\theta\end{aligned}$$

となる。よって，$X_{(1)}$ に基づいた θ の不偏推定量は

$$\hat{\theta}^U = X_{(1)} - \frac{1}{n}$$

となる。

(注)　この設問は 8 章問 9 と同じである。

(2) $\pi(\theta)=1$ に対する θ のベイズ推定量は, $z=n(x_{(1)}-\theta)$, $\theta = x_{(1)} - z/n$ のような変数変換を用いると

$$\mathrm{E}[\theta|x_{(1)}] = \frac{\int \theta e^{-n(x_{(1)}-\theta)} I(x_{(1)}>\theta)d\theta}{\int e^{-n(x_{(1)}-\theta)} I(x_{(1)}>\theta)d\theta} = \frac{\int_0^\infty (x_{(1)}-z/n)e^{-z}dz}{\int_0^\infty e^{-z}dz} = x_{(1)} - \frac{1}{n}\frac{\int_0^\infty ze^{-z}dz}{\int_0^\infty e^{-z}dz} = x_{(1)} - \frac{1}{n}$$

となるので，ベイズ推定量は

$$\hat{\theta}^B = X_{(1)} - \frac{1}{n}$$

となる。これは不偏推定量に一致することがわかる。

(3) 信用係数 $1-\gamma$ の θ のベイズ信用区間を求める。事後確率関数は

$$\pi(\theta|x_{(1)}) = ne^{-n(x_{(1)}-\theta)},\ \theta < x_{(1)}$$

で与えられ，θ に関して単調に増加し $\theta = x_{(1)}$ で最大となる。従って，最高事後密度信用区間 (HPD) は $[c, X_{(1)}]$ のような形をする。この c の値は信用確率に基づいて決める必要があるので

$$\int_c^{x_{(1)}} \pi(\theta|x_{(1)})d\theta = \int_c^{x_{(1)}} ne^{-n(x_{(1)}-\theta)}d\theta = \Big[e^{-n(x_{(1)}-\theta)}\Big]_c^{x_{(1)}} = 1 - e^{-n(x_{(1)}-c)} = 1-\gamma$$

を満たすように c の値を定めればよい。上の方程式を解くと

$$c = x_{(1)} + \frac{\log\gamma}{n}$$

となるので，信用係数 $1-\gamma$ の θ のベイズ信用区間は

$$\Big[X_{(1)} + \frac{\log\gamma}{n},\ X_{(1)}\Big]$$

で与えられる。

問 4　確率変数 $X_1,\dots,X_n$ が独立にポアソン分布 $Po(\lambda)$ に従うとする。

(1) λ にガンマ分布 $Ga(a,b)$ を仮定したときの λ の事後分布は

$$\Big(\prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!}e^{-\lambda}\Big)\frac{1}{\Gamma(a)}\frac{1}{b^a}\lambda^{a-1}e^{-\lambda/b} \propto \lambda^{n\overline{x}}e^{-n\lambda}\lambda^{a-1}e^{-\lambda/b} = \lambda^{(n\overline{x}+a)-1}e^{-(n+1/b)\lambda}$$

より,

$$\pi(\lambda|\boldsymbol{x}) \sim Ga(n\overline{x}+a, (n+1/b)^{-1})$$

となる。この分布から，事後分布の平均と分散は

$$\mathrm{E}[\lambda|\boldsymbol{X}] = \frac{n\overline{x}+a}{n+1/b}, \quad \mathrm{Var}(\lambda|\boldsymbol{X}) = \frac{n\overline{x}+a}{(n+1/b)^2}$$

になる。

(2) 信用係数 $1-\gamma$ の λ のベイズ信用区間を求める。ガンマ分布に基づいて最高事後密度信用区間 (HPD) を求めることは数値的に解く必要があるため，ここでは両側に $\gamma/2$ の確率で λ を含まない領域を持つような簡便な信用区間を求める。適当な $c>0$ に対して

$$\int_0^c \pi(\lambda|\boldsymbol{x})d\lambda = \frac{1}{\Gamma(n\overline{x}+a)}\left(n+\frac{1}{b}\right)^{n\overline{x}+a}\int_0^c \lambda^{n\overline{x}+a-1}e^{-(n+1/b)\lambda}d\lambda$$

を計算してみる。ここで $x=2(n+1/b)\lambda$ の変数変換を行うと

$$\int_0^c \pi(\lambda|\boldsymbol{x})d\lambda = \int_0^{2(n+1/b)c}\frac{1}{\Gamma(2(n\overline{x}+a)/2)}\frac{1}{2^{n\overline{x}+a}}x^{n\overline{x}+a-1}e^{-x/2}dx$$

と書ける。これは $2(n\overline{x}+a)$ が自然数のときには，自由度 $2(n\overline{x}+a)$ のカイ 2 乗分布に関する積分で表される。そこで，$2(n\overline{x}+a)$ の整数部分をガウス記号を用いて $[2(n\overline{x}+a)]$ で表し，

$$\int_0^c \pi(\lambda|\boldsymbol{x})d\lambda \approx \mathrm{P}\left[\chi^2_{[2(n\overline{x}+a)]} < 2\left(n+\frac{1}{b}\right)c\right]$$

で近似することにする。この確率が $\gamma/2$ になるように c を定めればよいので

$$2\left(n+\frac{1}{b}\right)c = \chi^2_{[2(n\overline{x}+a)],1-\gamma/2}$$

の解

$$c = \frac{\chi^2_{[2(n\overline{x}+a)],1-\gamma/2}}{2(n+1/b)}$$

が得られる。上側についても同様にして解を求めると，

$$\left[\frac{\chi^2_{[2(n\overline{x}+a)],1-\gamma/2}}{2(n+1/b)},\ \frac{\chi^2_{[2(n\overline{x}+a)],\gamma/2}}{2(n+1/b)}\right]$$

が信用係数 $1-\gamma$ の λ のベイズ信用区間になる。

(3) $H_0:\lambda=\lambda_0$ vs. $H_1:\lambda\neq\lambda_0$ の検定について，ベイズファクターによる検定方法を求める。そのために，$\boldsymbol{X}$ の周辺確率密度を求めると，

$$m(\boldsymbol{x}|a,b) = \frac{1}{\Gamma(a)b^a\prod_{i=1}^n x_i!}\int_0^\infty \lambda^{(n\overline{x}+a)-1}e^{-(n+1/b)\lambda}d\lambda = \frac{\Gamma(n\overline{x}+a)}{\Gamma(a)b^a\prod_{i=1}^n x_i!}\frac{1}{(n+1/b)^{n\overline{x}+a}}$$

と書けるので，ベイズファクターは

$$\mathrm{BF}_{01} = \frac{[\Gamma(a)b^a\prod_{i=1}^n x_i!]^{-1}\lambda_0^{(n\overline{x}+a)-1}e^{-(n+1/b)\lambda_0}}{[\Gamma(a)b^a\prod_{i=1}^n x_i!]^{-1}\Gamma(n\overline{x}+a)(n+1/b)^{-(n\overline{x}+a)}} = \frac{\{\lambda_0(n+1/b)\}^{n\overline{x}+a}}{\Gamma(n\overline{x}+a)\lambda_0}e^{-(n+1/b)\lambda_0}$$

となり，$\mathrm{BF}_{01}>1$ なら H_0 を選択し，$\mathrm{BF}_{01}<1$ なら H_1 を選択する。

また (2) のベイズ信用区間を反転させて，信用確率 γ の検定を求めると

$$\lambda_0 \notin \left[\frac{\chi^2_{[2(n\overline{x}+a)],1-\gamma/2}}{2(n+1/b)},\ \frac{\chi^2_{[2(n\overline{x}+a)],\gamma/2}}{2(n+1/b)}\right] \Longrightarrow H_0\text{を棄却}$$

のようになる。

問 5 例 8.3 では，東京都の 1 日当たりの交通事故による死亡数のデータを取り上げた。$n=30$ としてポアソン分布 $Po(\lambda)$ を当てはめることにする。λ にガンマ分布 $Ga(1,100)$ を仮定したときの λ のベイズ推定値は，$\overline{x}=0.3$, $a=1$, $b=100$ より

$$\hat{\lambda}^B = \frac{30\times 0.3+1}{30+1/100} = \frac{10}{30.01} = 0.33$$

となり，MLE は $\hat{\lambda}^M = \overline{x} - 0.3$ であるから，近い値を与えている。

信用係数 95% の λ のベイズ信用区間については，$2(n\overline{x}+a) = 20$ になるので，χ^2_{20} 分布を用いることができる。

$$\left[\frac{\chi^2_{20,1-0.025}}{2 \times 30.01},\ \frac{\chi^2_{20,0.025}}{2 \times 30.01}\right] = [0.16, 0.57]$$

となる。

仮説検定 $H_0 : \lambda = 1$ vs. $H_1 : \lambda \neq 1$ の検定について，ベイズファクターを計算すると

$$\mathrm{BF}_{01} = \frac{1}{\Gamma(10)}(30.01)^{10} e^{-30.01} = 0.00015 < 1$$

となるので，H_1 を選択することになる。

またベイズ信用区間を反転させて作った信用確率 5% の検定方法を用いて $H_0 : \lambda = 1$ を検定すると

$$1 \notin [0.16, 0.57]$$

であるから，信用確率 5% で H_0 は棄却される。

(補足) もし仮説検定 $H_0 : \lambda = 0.3$ vs. $H_1 : \lambda \neq 0.3$ の検定を考えてみると，ベイズファクターは

$$\mathrm{BF}_{01} = \frac{1}{\Gamma(10) \times 0.3}(30.01 \times 0.3)^{10} e^{-30.01 \times 0.3} = 3.95 > 1$$

となり，H_0 を選択することになる。

またベイズ信用区間を反転させて作った信用確率 5% の検定方法を用いて $H_0 : \lambda = 0.3$ を検定すると

$$0.3 \in [0.16, 0.57]$$

であるから，信用確率 5% で H_0 は棄却されない。

```
> (30*0.3+1)/(30+1/100)
[1] 0.3332223
> qchisq(0.025, 20)/(2*30.01)
[1] 0.159793
> qchisq(0.975, 20)/(2*30.01)
[1] 0.5693037
> (1/gamma(10))*(30.01)^10*exp(-30.01)
[1] 0.0001512584
```

問 6　確率変数 $X_1, \ldots, X_n$ が独立に指数分布 $Ex(\lambda)$ に従うとする。

(1) λ にガンマ分布 $Ga(a, b)$ を仮定したときの λ の事後分布は

$$\Big(\prod_{i=1}^{n} \lambda e^{-\lambda x_i}\Big) \frac{1}{\Gamma(a)} \frac{1}{b^a} \lambda^{a-1} e^{-\lambda/b} = \frac{1}{\Gamma(a) b^a} \lambda^{n+a-1} e^{-(n\overline{x}+1/b)\lambda}$$

より

$$\pi(\lambda|\boldsymbol{x}) \sim Ga\Big(n+a, \frac{1}{n\overline{x}+1/b}\Big)$$

となる。これより，事後分布の平均と分散は

$$\mathrm{E}[\lambda|\boldsymbol{x}] = \frac{n+a}{n\overline{x}+1/b}, \quad \mathrm{Var}(\lambda|\boldsymbol{x}) = \frac{n+a}{(n\overline{x}+1/b)^2}$$

となる。

(2) 信用係数 $1-\gamma$ の λ のベイズ信用区間を求める。ガンマ分布に基づいて最高事後密度信用区間 (HPD) を求めることは数値的に解く必要があるため，ここでは両側に $\gamma/2$ の確率で λ を含まない領域を持つような簡便な信用区間を求める。適当な $c>0$ に対して

$$\int_0^c \pi(\lambda|\boldsymbol{x})d\lambda = \frac{1}{\Gamma(n+a)}\Big(n\overline{x}+\frac{1}{b}\Big)^{n+a}\int_0^c \lambda^{n+a-1}e^{-(n\overline{x}+1/b)\lambda}d\lambda$$

を計算してみる。ここで $u=2(n\overline{x}+1/b)\lambda$ の変数変換を行うと

$$\int_0^c \pi(\lambda|\boldsymbol{x})d\lambda = \int_0^{2(n\overline{x}+1/b)c}\frac{1}{\Gamma(n+a)}\frac{1}{2^{n+a}}u^{n+a-1}e^{-u/2}du$$

と書ける。これは $2(n+a)$ が自然数のときには，自由度 $2(n+a)$ のカイ 2 乗分布に関する積分で表される。そこで，$2(n+a)$ の整数部分をガウス記号を用いて $[2(n+a)]$ で表し，

$$\int_0^c \pi(\lambda|\boldsymbol{x})d\lambda \approx \mathrm{P}\Big[\chi^2_{[2(n+a)]} < 2\Big(n\overline{x}+\frac{1}{b}\Big)c\Big]$$

で近似することにする。この確率が $\gamma/2$ になるように c を定めればよいので

$$2\Big(n\overline{x}+\frac{1}{b}\Big)c = \chi^2_{[2(n+a)],1-\gamma/2}$$

の解

$$c = \frac{\chi^2_{[2(n+a)],1-\gamma/2}}{2(n\overline{x}+1/b)}$$

が得られる。上側についても同様にして解を求めると，

$$\Big[\frac{\chi^2_{[2(n+a)],1-\gamma/2}}{2(n\overline{x}+1/b)},\ \frac{\chi^2_{[2(n+a)],\gamma/2}}{2(n\overline{x}+1/b)}\Big]$$

が信用係数 $1-\gamma$ の λ のベイズ信用区間になる。信用係数 $1-\gamma$ の λ のベイズ信用区間を与えよ。

(3) $H_0:\lambda=\lambda_0$ vs. $H_1:\lambda\neq\lambda_0$ の検定について，ベイズファクターによる検定方法を求める。そのために，$\boldsymbol{X}$ の周辺確率密度を求めると，

$$m(\boldsymbol{x}|a,b) = \frac{1}{\Gamma(a)b^a}\int_0^\infty \lambda^{(n+a)-1}e^{-(n\overline{x}+1/b)\lambda}d\lambda = \frac{\Gamma(n+a)}{\Gamma(a)b^a}\frac{1}{(n\overline{x}+1/b)^{n+a}}$$

と書けるので，ベイズファクターは

$$\mathrm{BF}_{01} = \frac{\lambda_0^{(n+a)-1}e^{-(n\overline{x}+1/b)\lambda_0}}{\Gamma(n+a)(n\overline{x}+1/b)^{-(n+a)}} = \frac{\{\lambda_0(n\overline{x}+1/b)\}^{n+a}}{\Gamma(n+a)\lambda_0}e^{-(n\overline{x}+1/b)\lambda_0}$$

となり，$\mathrm{BF}_{01}>1$ なら H_0 を選択し，$\mathrm{BF}_{01}<1$ なら H_1 を選択する。

また (2) のベイズ信用区間を反転させて，信用確率 γ の検定を求めると

$$\lambda_0 \notin \Big[\frac{\chi^2_{[2(n+a)],1-\gamma/2}}{2(n\overline{x}+1/b)},\ \frac{\chi^2_{[2(n+a)],\gamma/2}}{2(n\overline{x}+1/b)}\Big] \Longrightarrow H_0\text{を棄却}$$

のようになる。

問 7 例 8.2 では，時間間隔のデータを取り上げた。$n=50$ として指数分布 $Ex(\lambda)$ を当てはめるとする。λ にガンマ分布 $Ga(1,100)$ を仮定したときの λ のベイズ推定値は, $\overline{x}=19.334$ より

$$\hat{\lambda}^B = \frac{n+a}{n\overline{x}+1/b} = \frac{51}{966.71} = 0.053$$

となる。MLE は $\hat{\lambda}^M = 1/19.334 = 0.052$ となり，両者は近い値を与えている。

信用係数 95% の λ のベイズ信用区間は

$$\left[\frac{\chi^2_{102,1-0.025}}{2\times 966.71},\ \frac{\chi^2_{102,0.025}}{2\times 966.71}\right] = [0.039,\ 0.068]$$

仮説検定 $H_0 : \lambda = 1/24$ vs. $H_1 : \lambda \neq 1/24$ の検定について，ベイズファクターによる検定方法では

$$\mathrm{BF}_{01} = \frac{(966.71)^{51}}{\Gamma(51)} \frac{1}{24^{50}} e^{-966.71/24} = 18.33 > 1$$

より，H_0 を選択することになる。またベイズ信用区間を反転させて作った信用確率 5% の検定方法では

$$\lambda_0 = \frac{1}{24} = 0.042 \in [0.039,\ 0.068]$$

となるので，信用確率 5% で $H_0 : \lambda = 1/24$ は棄却できない。

```
> n <- 50; xo <- 19.334; a <- 1; b <- 100
> (n+a)/(n*xo+1/b)
[1] 0.05275626
> qchisq(0.025, 102)/(2*966.71)
[1] 0.0392805
> qchisq(0.975, 102)/(2*966.71)
[1] 0.06818877
> (966.71/24)^(50)*(966.71/gamma(51))*exp(-966.71/24)
[1] 18.3345
```

問 8 確率変数 $X_1, \ldots, X_n$ が独立にベルヌーイ分布 $Ber(\theta)$ に従うとする。θ にジェフリーズの事前分布 $\pi^{\mathrm{J}}(\theta)$ を仮定する。

(1) ベルヌーイ分布について 1 個のデータの尤度は $f(x|\theta) = \theta^x(1-\theta)^{1-x}$ であり，対数尤度は

$$\ell(\theta) = x\log\theta + (1-x)\log(1-\theta)$$

であるから

$$\ell'(\theta) = \frac{x}{\theta} - \frac{1-x}{1-\theta},\ \ell''(\theta) = -\frac{x}{\theta^2} - \frac{1-x}{(1-\theta)^2}$$

となるので，θ のフィッシャー情報量は

$$I(\theta) = \frac{1}{\theta} + \frac{1}{1-\theta} = \frac{1}{\theta(1-\theta)}$$

となる。従って，θ についてジェフリーズの事前分布は

$$\pi^J(\theta) = \frac{1}{\sqrt{\theta(1-\theta)}}$$

で与えられる。これは正則になる。

(2) θ の事後分布は

$$\left\{\prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i}\right\}\theta^{-1/2}(1-\theta)^{-1/2} = \theta^{n\overline{x}+1/2-1}(1-\theta)^{n-n\overline{x}+1/2-1}$$

より

$$\pi(\theta|\boldsymbol{x}) \sim Beta(n\overline{x}+1/2, n-n\overline{x}+1/2)$$

となる。

ベイズ推定量は

$$\hat{\theta}^B = \frac{n\overline{X}+1/2}{(n\overline{X}+1/2)+(n-n\overline{X}+1/2)} = \frac{n\overline{X}+1/2}{n+1}$$

となる。

(3) 仮説検定 $H_0: \theta = \theta_0$ vs. $H_1: \theta \neq \theta_0$ について，ベイズファクターによる検定方法を考える。$\boldsymbol{X}$ の周辺分布はベータ関数を用いて $B(n\overline{x}+1/2, n-n\overline{x}+1/2)$ で与えられるので，ベイズファクターは

$$\mathrm{BF}_{01} = \frac{\theta_0^{n\overline{x}}(1-\theta_0)^{n-n\overline{x}}}{B(n\overline{x}+1/2, n-n\overline{x}+1/2)}$$

となり，$\mathrm{BF}_{01} > 1$ なら H_0 を選択し，$\mathrm{BF}_{01} < 1$ なら H_1 を選択する。

(4) $h(\theta) = \theta(1-\theta)$ とおくとき，$h(\theta)$ のベイズ推定量は

$$\begin{aligned}\widehat{h(\theta)}^B =& \mathrm{E}[\theta(1-\theta)|\boldsymbol{X}] = \mathrm{E}[\theta-\theta^2|\boldsymbol{X}] = \mathrm{E}[\theta|\boldsymbol{X}](1-\mathrm{E}[\theta|\boldsymbol{x}]) - \{\mathrm{E}[\theta^2|\boldsymbol{X}] - (\mathrm{E}[\theta|\boldsymbol{X}])^2\} \\ =& \mathrm{E}[\theta|\boldsymbol{X}](1-\mathrm{E}[\theta|\boldsymbol{x}]) - \mathrm{Var}(\theta|\boldsymbol{X})\end{aligned}$$

と表される。$\theta|\boldsymbol{X} \sim Beta(n\overline{X}+1/2, n-n\overline{X}+1/2)$ より

$$\begin{aligned}\widehat{h(\theta)}^B =& \frac{n\overline{X}+1/2}{n+1}\Big(1-\frac{n\overline{X}+1/2}{n+1}\Big) - \frac{(n\overline{X}+1/2)(n-n\overline{X}+1/2)}{(n+1)^2(n+2)} \\ =& \frac{(n\overline{X}+1/2)(n-n\overline{X}+1/2)}{(n+1)(n+2)}\end{aligned}$$

で与えられる。$h(\theta)$ の MLE は，$\widehat{h(\theta)}^M = \overline{X}(1-\overline{X})$ である。

$$\widehat{h(\theta)}^B = \Big(\overline{X}+\frac{1}{2n}\Big)\Big(1-\overline{X}+\frac{1}{2n}\Big)\frac{n^2}{(n+1)(n+2)}$$

より，n が大きければ両者は近い推定値を与える。

問 9 確率変数 $X_1, \ldots, X_m$ が独立に幾何分布 $Geo(\theta)$ に従うとする。$Y = X_1 + \cdots + X_m$ とおく。θ にジェフリーズの事前分布 $\pi^J(\theta)$ を仮定する。

(1) 幾何分布の確率関数は $\theta(1-\theta)^x$ であり，対数尤度が $\ell(\theta) = \log\theta + x\log(1-\theta)$ であるから，

$$\ell'(\theta) = \frac{1}{\theta} - \frac{x}{1-\theta}, \quad \ell''(\theta) = -\frac{1}{\theta^2} - \frac{x}{(1-\theta)^2}$$

となる。$\mathrm{E}[X] = (1-\theta)/\theta$ よりフィッシャー情報量は

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

となるので，θ についてジェフリーズの事前分布は

$$\pi^J(\theta) = \frac{1}{\theta\sqrt{1-\theta}}$$

である。これは正則でない。

(2) θ の事後分布は, $y = \sum_{i=1}^m x_i$ とおくと

$$\Big\{\prod_{i=1}^m \theta(1-\theta)^{x_i}\Big\}\frac{1}{\theta\sqrt{1-\theta}} = \theta^m(1-\theta)^y\frac{1}{\theta\sqrt{1-\theta}} = \theta^{m-1}(1-\theta)^{y-1/2}$$

より

$$\pi(\theta|\boldsymbol{x}) \sim Beta(m, y+1/2)$$

となる

　これより，θ のベイズ推定量は

$$\hat{\theta}^B = \mathrm{E}[\theta|\boldsymbol{X}] = \frac{m}{m+Y+1/2}, \quad Y = \sum_{i=1}^{m} X_i$$

となる。MLE は，$m/\theta = Y/(1-\theta)$ を解いて

$$\hat{\theta}^M = \frac{m}{m+Y}$$

となる。m が大きければ両者は近い推定値を与える。

問 10　確率変数 $X_1, \ldots, X_n$ が独立に $\mathcal{N}(0,\theta)$ に従うとする。θ にジェフリーズの事前分布 $\pi^J(\theta)$ を仮定する。

　(1) 確率密度が $(2\pi\theta)^{-1/2}e^{-x^2/(2\theta)}$ であり，対数尤度が

$$\ell(\theta) = -\frac{1}{2\pi} - \frac{1}{2}\log\theta - \frac{x^2}{2\theta}$$

と書かれる。よって

$$\ell'(\theta) = -\frac{1}{2\theta} + \frac{x^2}{2\theta^2}, \quad \ell''(\theta) = \frac{1}{2\theta^2} - \frac{x^2}{\theta^3}$$

より，フィッシャー情報量は

$$I(\theta) = -\frac{1}{2\theta^2} + \frac{\mathrm{E}[X^2]}{\theta^3} = \frac{1}{2\theta^2}$$

で与えられる。従って，θ についてジェフリーズの事前分布は

$$\pi^J(\theta) = \frac{1}{\theta}$$

である。これは正則ではない。

　(2) θ の事後分布は

$$\left\{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\theta}} e^{-x_i^2/(2\theta)}\right\} \frac{1}{\theta} \propto \frac{1}{\theta^{n/2+1}} e^{-\sum_{i=1}^n x_i^2/(2\theta)}$$

であるから，ベイズ推定量は

$$\hat{\theta}^B = \mathrm{E}[\theta|\boldsymbol{x}] = \frac{\int_0^\infty \theta^{-n/2} e^{-\sum_{i=1}^n x_i^2/(2\theta)} d\theta}{\int_0^\infty \theta^{-n/2-1} e^{-\sum_{i=1}^n x_i^2/(2\theta)} d\theta}$$

と書ける。変数変換 $\lambda = 1/\theta$ を行うと，$d\theta = -\lambda^{-2}d\lambda$ より, $y = \sum_{i=1}^n x_i^2$ とおくと

$$\hat{\theta}^B = \frac{\int_0^\infty \lambda^{n/2-2} e^{-y\lambda/2} d\lambda}{\int_0^\infty \lambda^{n/2-1} e^{-y\lambda/2} d\lambda} = \frac{\Gamma(n/2-1)}{\Gamma(n/2)} \frac{y}{2} = \frac{y}{n-2}$$

となる。これより

$$\hat{\theta}^B = \frac{\sum_{i=1}^n X_i^2}{n-2}$$

となる。MLE は

$$\hat{\theta}^M = \frac{\sum_{i=1}^n X_i^2}{n}$$

となる。

(3) 信用係数 $1-\gamma$ の θ のベイズ信用区間を作る際，両側に $\gamma/2$ の確率をもつように作ることを考える。そこで $c>0$ に対して

$$\int_0^c \pi(\theta|\boldsymbol{x})d\theta = \int_{1/c}^\infty \frac{1}{\Gamma(n/2)}\Big(\frac{y}{2}\Big)^{n/2}\lambda^{n/2-1}e^{-(y/2)\lambda}d\lambda$$

を計算する。$u=y\lambda$ とおくと

$$\int_0^c \pi(\theta|\boldsymbol{x})d\theta = \int_{y/c}^\infty \frac{1}{\Gamma(n/2)}\Big(\frac{1}{2}\Big)^{n/2}u^{n/2-1}e^{-u/2}du = \mathrm{P}(\chi_n^2 > y/c)$$

と書ける。$\mathrm{P}(\chi_n^2 > y/c) = \gamma/2$ を満たすように c を定めるので，

$$\frac{y}{c} = \chi^2_{n,\gamma/2}, \quad c = \frac{y}{\chi^2_{n,\gamma/2}}$$

となる。従って，信用係数 $1-\gamma$ の θ の 95% ベイズ信用区間は

$$\Big[\frac{\sum_{i=1}^n X_i^2}{\chi^2_{n,\gamma/2}},\ \frac{\sum_{i=1}^n X_i^2}{\chi^2_{n,1-\gamma/2}}\Big]$$

で与えられる。

問 11　以下の問において，確率変数 U は一様分布 $U(0,1)$ に従い，$U_1,\ldots,U_n$ は独立に $U(0,1)$ に従うとする。

(1) 変数変換 $x=-\log u$ により，$u=e^{-x}$ となるので，$du=-e^{-x}dx$ より，X の確率密度は

$$f_X(x) = e^{-x}$$

となる。従って，$-\log U \sim Ex(1)$ となる。

$X_i = -\log U_i$ とおくと，$X_1,\ldots,X_n$ は独立に指数分布 $Ex(1)$ に従う。分布の再生性より

$$-\sum_{i=1}^n \log U_i = \sum_{i=1}^n X_i \sim Ga(n,1)$$

となる。

(2) 確率変数 X が $Ex(\lambda)$ の分布関数は $F(x) = 1-e^{-\lambda x}$ である。$F(X) \sim U(0,1)$ であるが，$1-F(X)$ も $U(0,1)$ に従う。$U = 1-F(X) = e^{-\lambda X}$ とおくと

$$X = -\frac{1}{\lambda}\log U$$

となる。

(3) 確率変数 X が $Ga(n,\beta)$ に従うとする。$X_1,\ldots,X_n$ が独立に指数分布 $Ex(1/\beta)$ に従うとすると，分布の再生性より，$X = X_1+\cdots+X_n \sim Ga(n,\beta)$ と書けることがわかる。$X_i = -\beta\log U_i$ と書けるので

$$X = -\beta\sum_{i=1}^n \log U_i$$

と表される。

問 12　確率変数 X が両側指数分布 $f(x) = 2^{-1}\exp(-|x|)$ に従うとする。一様分布 $U(0,1)$ に従う確率変数を U とするとき，X を U の関数として表す。分布関数 $F(x) = \int_{-\infty}^x f(t)dt$ は，

$$x<0 \text{ のとき } F(x) = \int_{-\infty}^x \frac{1}{2}e^t dt = \frac{1}{2}e^x$$

$$x>0 \text{ のとき } F(x) = \frac{1}{2} + \int_0^x \frac{1}{2}e^{-t}dt = \frac{1}{2} + \frac{1}{2}\Big[-e^{-t}\Big]_0^x = 1-\frac{1}{2}e^{-x}$$

で与えられる。$F(X)=U$ を解くと

$$U<\frac{1}{2}\ \text{のとき}\ e^X=2U,\quad X=\log(2U)$$
$$U>\frac{1}{2}\ \text{のとき}\ 1-\frac{1}{2}e^{-X}=U,\quad X=-\log\{2(1-U)\}$$

となる。

問 13　受容棄却法を用いて正規乱数を発生させる問題を考える。この場合，

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

である。

コーシー分布に従う乱数を用いて正規乱数を発生させる場合には，$h(x)=\pi^{-1}(1+x^2)^{-1}$ より

$$M_1=\max_x\left\{\frac{\pi(x)}{h(x)}\right\}=\frac{\sqrt{\pi}}{\sqrt{2}}\max_x\left\{(1+x^2)e^{-x^2/2}\right\}$$

である。$\pi(x)/h(x)$ は $x^2=1$ のときに最大化されるので

$$M_1=\frac{\sqrt{\pi}}{\sqrt{2}}2e^{-1/2}=\sqrt{2\pi}e^{-1/2}$$

で与えられる。

乱数発生のための受容棄却法によるアルゴリズムは次のようになる。

Step 1. $h(y)$ から乱数 y^* を発生させる。また $U\sim U(0,1)$ を発生させる。

Step 2. $U\le\pi(y^*)/\{M_1h(y^*)\}$ ならば y^* を $\pi(y)$ からの標本として受容して $Y=y^*$ とおき，そうでなければ棄却して Step 1 へ戻る。

両側指数分布に従う乱数を用いて正規乱数を発生させる場合には，$h(x)=2^{-1}e^{-|x|}$ より

$$M_2=\max_x\left\{\frac{\pi(x)}{h(x)}\right\}=\frac{2}{\sqrt{2\pi}}\max_x\left\{e^{-x^2/2+|x|}\right\}$$

である。$\pi(x)/h(x)$ は $|x|=1$ のときに最大化されるので

$$M_2=\frac{\sqrt{2}}{\sqrt{\pi}}e^{1/2}$$

で与えられる。

乱数発生のための受容棄却法によるアルゴリズムは次のようになる。

Step 1. $h(y)$ から乱数 y^* を発生させる。また $U\sim U(0,1)$ を発生させる。

Step 2. $U\le\pi(y^*)/\{M_2h(y^*)\}$ ならば y^* を $\pi(y)$ からの標本として受容して $Y=y^*$ とおき，そうでなければ棄却して Step 1 へ戻る。

受容棄却法で用いる M の値を比較すると，

$$\sqrt{2\pi}e^{-1/2}>\frac{\sqrt{2}}{\sqrt{\pi}}e^{1/2}\Longleftrightarrow\pi>e$$

より，$M_1>M_2$ となり，M_2 の方が小さい。M の値が小さい方が棄却率が小さくなるので，その方が望ましい。

問 14　ギブスサンプリングを用いて，次の周辺分布の乱数を発生させる。

(1) $X|Y \sim Bin(n, Y)$, $Y \sim Beta(\alpha_0, \beta_0)$ のとき，

$$
\begin{aligned}
f(x|y)f(y) =& \begin{pmatrix} n \\ x \end{pmatrix} y^x (1-y)^{n-x} \frac{1}{B(\alpha_0, \beta_0)} y^{\alpha_0 - 1}(1-y)^{\beta_0 - 1} \\
=& \begin{pmatrix} n \\ x \end{pmatrix} \frac{1}{B(\alpha_0, \beta_0)} y^{x+\alpha_0-1}(1-y)^{n-x+\beta_0-1}
\end{aligned}
$$

より

$$
Y|x \sim Beta(x+\alpha_0, n-x+\beta_0)
$$

に従う。

Step 1. $x_k \sim Bin(n, y_{k-1})$ を発生させる。

Step 2. $y_k \sim Beta(x_k+\alpha_0, n-x_k+\beta_0)$ を発生させて，Step 1 へ戻る。
このとき，k が大きいとき x_k は周辺分布からの乱数と見なすことができる。

(2) $X|Y \sim Po(Y)$, $Y \sim Ga(\alpha_0, \beta_0)$ のとき，

$$
\begin{aligned}
f(x|y)f(y) =& \frac{y^x}{x!} e^{-y} \frac{1}{\Gamma(\alpha_0)} \frac{1}{\beta_0^{\alpha_0}} y^{\alpha_0-1} e^{-y/\beta_0} \\
=& \frac{1}{x!\Gamma(\alpha_0)\beta_0^{\alpha_0}} y^{x+\alpha_0-1} e^{(1+1/\beta_0)y}
\end{aligned}
$$

より

$$
Y|x \sim Ga\Big(x+\alpha_0, \frac{\beta_0}{1+\beta_0}\Big)
$$

に従う。

Step 1. $x_k \sim Po(y_{k-1})$ を発生させる。

Step 2. $y_k \sim Ga(x_k+\alpha_0, (1+1/\beta_0)^{-1})$ を発生させて，Step 1 へ戻る。
このとき，k が大きいとき x_k は周辺分布からの乱数と見なすことができる。

問 15　メトロポリス・ヘイスティング・サンプリングを用いて，次の分布の乱数を発生させることを考える。

(1) 対数正規分布の確率密度は

$$
\pi(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-(\log x)^2/2}
$$

であり，

$$
\frac{\pi(y)}{\pi(x)} = \frac{x}{y} \exp\left\{ \frac{1}{2}(\log x)^2 - \frac{1}{2}(\log y)^2 \right\}
$$

と書ける。σ^2 の値を決める。

Step 1. $y_k \sim \mathcal{N}(x^{(k-1)}, \sigma^2)$ を発生させる。

Step 2. $U \sim U(0,1)$ を発生させる。

$$
\begin{aligned}
U \le \min\left\{1, \frac{\pi(y^{(k)})}{\pi(x^{(k-1)})}\right\} &\Longrightarrow x^{(k)} = y^{(k)} \\
U > \min\left\{1, \frac{\pi(y^{(k)})}{\pi(x^{(k-1)})}\right\} &\Longrightarrow x^{(k)} = x^{(k-1)}
\end{aligned}
$$

として，Step 1 へ戻る。

このとき，k が大きいとき x_k は周辺分布からの乱数と見なすことができる。

(2) $Beta(a,b)$ については

$$\pi(x) \propto x^{a-1}(1-x)^{b-1}$$

であり，

$$\frac{\pi(y)}{\pi(x)} = \left(\frac{y}{x}\right)^{a-1}\left(\frac{1-y}{1-x}\right)^{b-1}$$

と書ける。σ^2 の値を決める。

Step 1. $y_k \sim \mathcal{N}(x^{(k-1)}, \sigma^2)$ を発生させる。

Step 2. $U \sim U(0,1)$ を発生させる。

$$U \le \min\left\{1, \frac{\pi(y^{(k)})}{\pi(x^{(k-1)})}\right\} \Longrightarrow x^{(k)} = y^{(k)}$$
$$U > \min\left\{1, \frac{\pi(y^{(k)})}{\pi(x^{(k-1)})}\right\} \Longrightarrow x^{(k)} = x^{(k-1)}$$

として，Step 1 へ戻る。

このとき，k が大きいとき x_k は周辺分布からの乱数と見なすことができる。

問 16　確率変数 (X_1, X_2, X_3) が 3 項分布 $Mult_3(n, p_1, p_2, p_3)$ に従い，(p_1, p_2, p_3) の事前分布がディリクレ分布 $Dir(c_1, c_2, c_3)$ に従い，同時確率密度関数が

$$f(p_1, p_2, p_3) = \frac{\Gamma(c_1+c_2+c_3)}{\Gamma(c_1)\Gamma(c_2)\Gamma(c_3)} p_1^{c_1-1} p_2^{c_2-1} p_3^{c_3-1}$$

で与えられるとする。ただし c_1, c_2, c_3 は正の実数である。

(1) (p_1, p_2, p_3) の事後分布は

$$\frac{n!}{x_1!x_2!x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3} \frac{\Gamma(c_1+c_2+c_3)}{\Gamma(c_1)\Gamma(c_2)\Gamma(c_3)} p_1^{c_1-1} p_2^{c_2-1} p_3^{c_3-1} \propto p_1^{x_1+c_1-1} p_2^{x_2+c_2-1} p_3^{x_3+c_3-1}$$

より

$$\pi(p_1, p_2, p_3 | \boldsymbol{x}) \sim Dir(x_1+c_1, x_2+c_2, x_3+c_3)$$

となる。

p_i のベイズ推定量は

$$\hat{p}_i^B = \mathrm{E}[p_i | \boldsymbol{X}] = \frac{X_i + c_i}{n + c_1 + c_2 + c_3}$$

となる。

(2) (p_1, p_2, p_3) についてジェフリーズの事前分布 $\pi^{\mathrm{J}}(p_1, p_2, p_3)$ を求める。$p_3 = 1 - p_1 - p_2$ より，対数尤度は

$$\ell(p_1, p_2) \propto x_1 \log p_1 + x_2 \log p_2 + x_3 \log(1 - p_1 - p_2)$$

であるから

$$\frac{\partial \ell}{\partial p_1} = \frac{x_1}{p_1} - \frac{x_3}{1-p_1-p_2}, \quad \frac{\partial \ell}{\partial p_2} = \frac{x_2}{p_2} - \frac{x_3}{1-p_1-p_2},$$
$$\frac{\partial^2 \ell}{\partial p_1^2} = -\frac{x_1}{p_1^2} - \frac{x_3}{(1-p_1-p_2)^2}, \quad \frac{\partial^2 \ell}{\partial p_2^2} = -\frac{x_2}{p_2^2} - \frac{x_3}{(1-p_1-p_2)^2},$$
$$\frac{\partial^2 \ell}{\partial p_1 \partial p_2} = -\frac{x_3}{(1-p_1-p_2)^2}$$

と書ける。$\mathrm{E}[X_i] = np_i$ より，フィッシャー情報量を求めると

$$\begin{aligned} I_{11} =& \frac{np_1}{p_1^2} + \frac{n(1-p_1-p_2)}{(1-p_1-p_2)^2} = n\frac{p_3+p_1}{p_1p_3} = n\frac{1-p_2}{p_1p_3} \\ I_{12} =& n\frac{(1-p_1-p_2)}{(1-p_1-p_2)^2} = n\frac{1}{p_3} \\ I_{22} =& n\frac{1}{p_2} + n\frac{1}{p_3} = n\frac{1-p_1}{p_2p_3} \end{aligned}$$

より，フィッシャー情報量行列は

$$\boldsymbol{I}(p_1,p_2,p_3) = \frac{n}{p_1p_2p_3}\begin{pmatrix} p_2(1-p_2) & p_1p_2 \\ p_1p_2 & p_1(1-p_1) \end{pmatrix}$$

で与えられる。この行列式は

$$\begin{aligned} |\boldsymbol{I}(p_1,p_2,p_3)| =& \frac{n^2}{p_1^2p_2^2p_3^2}\{p_1(1-p_1)p_2(1-p_2) - p_1^2p_2^2\} \\ =& \frac{n^2}{p_1p_2p_3^2}\{(1-p_1)(1-p_2) - p_1p_2\} \\ =& \frac{n^2}{p_1p_2p_3^2}\{1-p_1-p_2\} = \frac{n^2}{p_1p_2p_3} \end{aligned}$$

となる。従って，ジェフリーズの事前分布は

$$\pi^J(p_1,p_2,p_3) = \frac{1}{\sqrt{p_1p_2p_3}}$$

で与えられる。これは，$Dir(1/2,1/2,1/2)$ に対応する。そのときの p_i のベイズ推定量は

$$\hat{p}_i^B = \frac{X_i+1/2}{n+3/2}$$

となる。

問 17 (14.11) については，$X_1,\ldots,X_n$ が独立に正規分布 $\mathcal{N}(\mu,\sigma_0^2)$ に従い，μ の事前分布を $\mathcal{N}(\mu_0,\tau^2)$ とし，$\tau^2 \sim \tau^{-2a}$ であるから,。このとき，尤度関数と事前分布の確率密度関数は

$$\begin{aligned} f_n(\boldsymbol{x}|\mu) =& \frac{1}{(2\pi\sigma_0^2)^{n/2}}\exp\Big[-\frac{1}{2\sigma_0^2}\Big\{\sum_{i=1}^n(x_i-\overline{x})^2 + n(\overline{x}-\mu)^2\Big\}\Big] \\ \pi(\mu,\tau^2) =& \frac{1}{(\tau^2)^a}\frac{1}{(2\pi\tau^2)^{1/2}}\exp\Big[-\frac{1}{2\tau^2}(\mu-\mu_0)^2\Big] \end{aligned}$$

で表されるので，同時確率関数 $f_n(\boldsymbol{x}|\mu)\pi(\mu,\tau^2)$ を書き直すために

$$\widehat{\mu}^{\mathrm{B}}(\tau^2) = \frac{n\tau^2\overline{x}+\sigma_0^2\mu_0}{n\tau^2+\sigma_0^2} = \overline{x} - \frac{\sigma_0^2\mu_0}{n\tau^2+\sigma_0^2}(\overline{x}-\mu_0)$$

とおくと，次の等式が成り立つことに注意する。

$$\frac{n}{\sigma_0^2}(\overline{x}-\mu)^2 + \frac{1}{\tau^2}(\mu-\mu_0)^2 = \Big(\frac{n}{\sigma_0^2}+\frac{1}{\tau^2}\Big)(\mu-\widehat{\mu}^{\mathrm{B}}(\tau^2))^2 + \frac{n}{\sigma_0^2+n\tau^2}(\overline{x}-\mu_0)^2$$

従って，$s^2 = n^{-1}\sum_{i=1}^n(x_i-\overline{x})^2$ とおくと

$$\begin{aligned} f_n(\boldsymbol{x}|\mu)\pi(\mu,\tau^2) =& \frac{1}{(\tau^2)^a}\frac{1}{(2\pi\sigma_0^2)^{n/2}(2\pi\tau^2)^{1/2}}\exp\Big\{-\frac{n/\sigma_0^2+1/\tau^2}{2}(\mu-\widehat{\mu}^{\mathrm{B}}(\tau^2))^2 \\ & -\frac{n}{2(\sigma_0^2+n\tau^2)}(\overline{x}-\mu_0)^2 - \frac{n}{2\sigma_0^2}s^2\Big\} \end{aligned}$$

と表されるので，μ の事後分布は $\pi(\mu|\boldsymbol{x},\tau^2)=\mathcal{N}(\widehat{\mu}^{\mathrm{B}}(\tau^2),(n/\sigma_0^2+1/\tau^2)^{-1})$ となる。階層ベイズ推定量は

$$\begin{aligned}\widehat{\mu}^{\mathrm{HB}}=&\mathrm{E}[\mu|\overline{X}]=\frac{\int_0^\infty \widehat{\mu}^{\mathrm{B}}(\tau^2)g(\tau^2)d\tau^2}{\int_0^\infty g(\tau^2)d\tau^2}\\ =&\overline{X}-\frac{\int_0^\infty \frac{\sigma_0^2}{n\tau^2+\sigma_0^2}g(\tau^2)d\tau^2}{\int_0^\infty g(\tau^2)d\tau^2}(\overline{X}-\mu_0)\end{aligned}$$

と表されることがわかる。ここで

$$g(\tau^2)=\frac{1}{\sqrt{n\tau^2+\sigma_0^2}}\frac{1}{(\tau^2)^{a+1/2}}\exp\Big\{-\frac{n}{2(n\tau^2+\sigma_0^2)}(\overline{X}-\mu_0)^2\Big\}$$

である。

(14.12) については，標本平均の分布が $\overline{X}\sim\mathcal{N}(\mu,\sigma_0^2/n)$ に従い，μ の階層事前分布が $\mu|\eta\sim\mathcal{N}(\mu_0,1/\eta)$, $\eta\sim\eta^a$ であるから，完全条件付き分布は

$$\begin{aligned}\pi_1(\mu|\eta,\overline{X})\sim&\mathcal{N}(\widehat{\mu}^{\mathrm{B}}(\eta),(n/\sigma_0^2+\eta)^{-1})\\ \pi_2(\eta|\mu)\sim&Ga(a+3/2,2/(\mu-\mu_0)^2)\end{aligned}$$

となる。ただし $\widehat{\mu}^{\mathrm{B}}(\eta)=\{(n/\sigma_0^2)\overline{X}+\eta\mu_0\}/(n/\sigma_0^2+\eta)$ である。

第15章　分散分析と多重比較

問1　フィッシャーは，実験計画を立てる上での3つの原則を提唱した。

(a) 反復，繰り返し。実験は誤差を伴うので推定誤差を小さくするためには反復が必要である。

(b) 無作為化，ランダム化。実験に伴う誤差に偶然誤差と系統誤差があり，系統誤差を緩和するためにランダムに割り付ける必要がある。

(c) 局所管理，均一条件下での実験。条件が均一になるように全体の実験をブロックに分け，ブロック内では均一になるように管理する必要がある。

農事試験においてAとBの2種類の肥料の効果について比較実験を行うときの問題点は次の通りである。

(1) 実験農地を2つに分け，どちらかをランダムに選んでAを与え，残りにBを与えることは，１つしかデータをとっておらず，反復の原則に反する。

(2) 実験農地を東側から均等に10区画に分け，最初の5区画にAを与え後の5区画にBを与えることは，AとBの割付に系統誤差が生まれる可能性があり，無作為化の原則に反する。。

(3) 区画内では地力が均質になるような区画を10個選び，その中からランダムに5区画を選んでAを与え，残りの5区画にBを与える場合，均質な区画の中にランダムにAとBを割り付けることができていないため，局所管理の原則に反する。

問2　5種類のコンクリート・ブロックの吸湿量の比較実験を行った。それぞれ6個のデータをとり，1元配置モデル(15.1)を用いて解析することを考える。各種類の吸湿量の平均$\overline{y}_{i\cdot}$と平方和$w_i = \sum_{j=1}^{6}(y_{ij} - \overline{y}_{i\cdot})^2$の値が次の表で与えられる。

コンクリートの種類	1	2	3	4	5
平均 ($\overline{y}_{i\cdot}$)	5.1	5.2	5.7	4.3	6.7
平方和 (w_i)	1.50	0.84	1.34	1.33	1.05

(1) 郡内平方和は，$\mathrm{SS_W} = w_1 + w_2 + w_3 + w_4 + w_5 = 6.06$となり，群間平方和は，$\mathrm{SS_B} = 6\sum_{i=1}^{5}(\overline{y}_{i\cdot} - \overline{y}_{\cdot\cdot})^2 = 18.72$となる。

(2) 平均の同等性の検定$H_0 : \mu_1 = \cdots = \mu_5$について，有意水準5%で検定する。

$$F = \frac{\mathrm{SS_B}/4}{\mathrm{SS_W}/25} = \frac{18.72/4}{6.06/25} = 19.31$$

となり，$F_{4,25,0.05} = 2.76$なので，有意になる。

```
> y <- c(5.1,5.2,5.7,4.3,6.7)
> w <- c(1.50,0.84,1.34,1.33,1.05)
> SSw <- sum(w)
> SSb <- 6*var(y)*4
> SSw
[1] 6.06
> SSb
[1] 18.72
> F <- (18.72/4)/(6.06/25)
> F
[1] 19.30693
> qf(0.95, 4,25)
[1] 2.75871
```

問 3　(15.10) の多重比較の問題を問 2 のデータについて考える。

(1) 有意水準 5% の多重比較検定をテューキー法を用いて行う。テューキー法は

$$|\overline{y}_{a\cdot} - \overline{y}_{b\cdot}| > \frac{\hat{\sigma}}{\sqrt{6}} q_{5,25}(0.05)$$

を満たす組 (a, b) が有意な差として検出される。$\hat{\sigma} = \sqrt{6.06/25} = 0.492$, $q_{5,25}(0.05) = 4.153$ より

$$|\overline{y}_{a\cdot} - \overline{y}_{b\cdot}| > 0.835$$

となり，$\mu_1 \neq \mu_5$, $\mu_2 \neq \mu_4$, $\mu_2 \neq \mu_5$, $\mu_3 \neq \mu_4$, $\mu_3 \neq \mu_5$, $\mu_4 \neq \mu_5$ の 6 組が有意な差として検出される。

(2) 有意水準 5% の多重比較検定をボンフェローニ法を用いて行う。ボンフェローニ法は

$$|\overline{y}_{a\cdot} - \overline{y}_{b\cdot}| > \frac{\sqrt{2}\hat{\sigma}}{\sqrt{6}} t_{25,0.05/40}$$

を満たす組 (a, b) が有意な差として検出される。$t_{25,0.05/40} = 3.361$ より

$$|\overline{y}_{a\cdot} - \overline{y}_{b\cdot}| > \frac{\sqrt{2} \times 0.492}{\sqrt{6}} \times 3.361 = 0.956$$

となり，$\mu_1 \neq \mu_5$, $\mu_2 \neq \mu_5$, $\mu_3 \neq \mu_4$, $\mu_3 \neq \mu_5$, $\mu_4 \neq \mu_5$ の 5 組が有意な差として検出される。ボンフェローニ法はテューキー法より保守的である。

```
> qtukey(0.95, 5, 25, nranges=1, lower.tail=TRUE, log.p=FALSE)
[1] 4.153363
> (sqrt(6.06/25)/sqrt(6))*4.153363
[1] 0.8348156
> qt(1- 0.05/40, 25)
[1] 3.360631
> (sqrt(2)*sqrt(6.06/25)/sqrt(6))*3.360631
[1] 0.9552708
```

問 4　ラットの毒性試験データが表 15.4 で与えられているが，別の薬剤について同様の実験を行った。平均 $\overline{y}_{ij\cdot}$ の値は次の表のようになった。ただし，残差平方和 $\mathrm{SS_E}$ の値は $32,000$ であり，繰り返し数は $K = 5$ である。

薬剤濃度	B_1	B_2	B_3	B_4	平均
雄 A_1	830	820	750	700	775
雌 A_2	750	750	740	708	737
平均	790	785	745	704	756

(1) $I=2,\ J=4,\ K=5$ であり，$\overline{y}_{1\cdot\cdot}=775,\ \overline{y}_{2\cdot\cdot}=737,\ \overline{y}_{\cdot1\cdot}=790,\ \overline{y}_{\cdot2\cdot}=785,\ \overline{y}_{\cdot3\cdot}=745,\ \overline{y}_{\cdot4\cdot}=704,\ \overline{y}_{\cdots}=756$ である。これらに基づいて平方和を計算する。

$$\mathrm{SS_A}=JK\sum_{i=1}^{I}(\overline{y}_{i\cdot\cdot}-\overline{y}_{\cdots})^2=14440$$

となり，自由度は $I-1=1$ である。

$$\mathrm{SS_B}=IK\sum_{j=1}^{J}(\overline{y}_{\cdot j\cdot}-\overline{y}_{\cdots})^2=48220$$

となり，自由度は $J-1=3$ である。

$$\mathrm{SS_{AB}}=K\sum_{i=1}^{I}\sum_{j=1}^{J}(\overline{y}_{ij\cdot}-\overline{y}_{\cdots})^2-\mathrm{SS_A}-\mathrm{SS_B}=76880-14440-48220=14220$$

となり，自由度は $(I-1)(J-1)=3$ である。

(2) $\mathrm{SS_E}=32000$ が与えられており，自由度が $IJ(K-1)=32$ より分散の不偏推定量の値は,

$$\hat{\sigma}^2=\frac{32000}{32}=1000$$

となる。$H_\alpha:\alpha_1=\alpha_2=0$ の検定については

$$\frac{\mathrm{SS_A}/(I-1)}{\hat{\sigma}^2}=\frac{14440/1}{1000}=14.4>F_{1,32,0.05}=4.15$$

より，有意水準 5% で有意になる。

(3) $H_\beta:\beta_1=\beta_2=\beta_3=\beta_4=0$ の検定については

$$\frac{\mathrm{SS_B}/(J-1)}{\hat{\sigma}^2}=\frac{48220/3}{1000}=16.07>F_{3,32,0.05}=2.90$$

より，有意水準 5% で有意になる。

(4) $H_\gamma:\gamma_{ij}=0,\ (i=1,2,\ j=1,2,3,4)$ の検定については

$$\frac{\mathrm{SS_{AB}}/(I-1)(J-1)}{\hat{\sigma}^2}=\frac{14220/3}{1000}=4.74>F_{3,32,0.05}=2.90$$

より，有意水準 5% で有意になる。

```
> I <- 2; J <- 4; K <- 5
> SSa <- J*K*var(c(775, 737))
> SSa
[1] 14440
> SSb <- I*K*var(c(790,785,745,704))*3
> SSb
[1] 48220
> X <- c(830,820,750,700,750,750,740,708)
> SSab <- K*var(X)*7-SSa-SSb
> SSab
[1] 14220
> K*var(X)*7
[1] 76880
> var <- 32000/32
> Fa <- (14440/1)/var
> Fa
[1] 14.44
> qf(0.95, 1,32)
[1] 4.149097
> Fb <- (48220/3)/var
> Fb
[1] 16.07333
> qf(0.95, 3,32)
[1] 2.90112
> Fab <- (14220/3)/var
> Fab
[1] 4.74
> qf(0.95, 3,32)
[1] 2.90112
```

問 5 (15.13) 実際，$\mathrm{E}[\mathrm{SS_{AB}}]$ 以外は補題 15.1 より容易に確かめられる。$\mathrm{E}[\mathrm{SS_{AB}}]$ については

$$
\begin{aligned}
\mathrm{SS_{AB}} =& K\sum_{i=1}^{I}\sum_{j=1}^{J}\{(\overline{y}_{ij\cdot}-\overline{y}_{\cdots})-(\overline{y}_{i\cdot\cdot}-\overline{y}_{\cdots})-(\overline{y}_{\cdot j\cdot}-\overline{y}_{\cdots})\}^2 \\
=& K\sum_{i=1}^{I}\sum_{j=1}^{J}(\overline{y}_{ij\cdot}-\overline{y}_{\cdots})^2-\mathrm{SS_A}-\mathrm{SS_B}
\end{aligned}
$$

と変形できるので，補題 15.1 を用いると

$$
\begin{aligned}
\mathrm{E}[\mathrm{SS_{AB}}] =& (IJ-1)\sigma^2+K\sum_{i=1}^{I}\sum_{j=1}^{J}(\alpha_i+\beta_j+\gamma_{ij})^2 \\
& -\left\{(I-1)\sigma^2+JK\sum_{i=1}^{I}\alpha_i^2\right\}-\left\{(J-1)\sigma^2+IK\sum_{j=1}^{J}\beta_j^2\right\}
\end{aligned}
$$

のようになる。

問 6 命題 15.3 を証明する。記号の簡単のために $z_{ij}=\overline{\varepsilon}_{ij\cdot}$, $z_{i\cdot}=\overline{\varepsilon}_{i\cdot\cdot}$, $z_{\cdot j}=\overline{\varepsilon}_{\cdot j\cdot}$ と書くことにすると，

$$
\begin{aligned}
\overline{y}_{i\cdot\cdot}-\overline{y}_{\cdots} =& \alpha_i+\overline{\varepsilon}_{i\cdot\cdot}-\overline{\varepsilon}_{\cdots}=\alpha_i+z_{i\cdot}-z_{\cdot\cdot} \\
\overline{y}_{\cdot j\cdot}-\overline{y}_{\cdots} =& \beta_j+\overline{\varepsilon}_{\cdot j\cdot}-\overline{\varepsilon}_{\cdots}=\beta_j+z_{\cdot j}-z_{\cdot\cdot} \\
\overline{y}_{ij\cdot}-\overline{y}_{i\cdot\cdot}-\overline{y}_{\cdot j\cdot}+\overline{y}_{\cdots} =& \gamma_{ij}+\overline{\varepsilon}_{ij\cdot}-\overline{\varepsilon}_{i\cdot\cdot}-\overline{\varepsilon}_{\cdot j\cdot}+\overline{\varepsilon}_{\cdots} \\
=& \gamma_{ij}+z_{ij}-z_{i\cdot}-z_{\cdot j}+z_{\cdot\cdot} \\
y_{ijk}-\overline{y}_{ij\cdot} =& \varepsilon_{ijk}-\overline{\varepsilon}_{ij\cdot}
\end{aligned}
$$

と表すことができる。

(1) については，すべての i,j,k に対して，$\varepsilon_{ijk}-\overline{\varepsilon}_{ij\cdot}$ と $\overline{\varepsilon}_{ij\cdot}$ は独立になるので，$\mathrm{SS_E}$ は，$\mathrm{SS_A}$, $\mathrm{SS_B}$, $\mathrm{SS_{AB}}$ と独立になる。後は，$\mathrm{SS_A}$, $\mathrm{SS_B}$, $\mathrm{SS_{AB}}$ が独立になることを示せばよい。そのためには，

$$X_1=\sum_{i=1}^I s_i(z_{i\cdot}-z_{\cdot\cdot}),\ X_2=\sum_{j=1}^J t_j(z_{\cdot j}-z_{\cdot\cdot}),\ X_3=\sum_{i=1}^I\sum_{j=1}^J u_{ij}(z_{ij}-z_{i\cdot}-z_{\cdot j}+z_{\cdot\cdot})$$

とおくとき, $\mathrm{E}[e^{X_1+X_2+X_3}]=\mathrm{E}[e^{X_1}]\mathrm{E}[e^{X_2}]\mathrm{E}[e^{X_3}]$ となることを示せばよい。まず

$$\begin{aligned}X_1+X_2+X_3=&\sum_{i=1}^I s_i(z_{i\cdot}-z_{\cdot\cdot})+\sum_{j=1}^J t_j(z_{\cdot j}-z_{\cdot\cdot})+\sum_{i=1}^I\sum_{j=1}^J u_{ij}(z_{ij}-z_{i\cdot}-z_{\cdot j}+z_{\cdot\cdot})\\=&\sum_{i=1}^I\sum_{j=1}^J D_{ij}z_{ij}\end{aligned}$$

と表す。ここで

$$D_{ij}=u_{ij}+\frac{1}{J}\Big(s_i-\sum_b u_{ib}\Big)+\frac{1}{I}\Big(t_j-\sum_a u_{aj}\Big)+\frac{1}{IJ}\Big(\sum_a\sum_b u_{ab}-\sum_a s_a-\sum_b t_b\Big)$$

である。$z_{ij}\sim\mathcal{N}(0,\sigma^2/K)$ に従うので，正規分布の積率母関数から，

$$\begin{aligned}\mathrm{E}[e^{X_1+X_2+X_3}]=&\mathrm{E}[e^{\sum_{i=1}^I\sum_{j=1}^J D_{ij}z_{ij}}]=\prod_{i=1}^I\prod_{j=1}^J\mathrm{E}[e^{D_{ij}z_{ij}}]=\prod_{i=1}^I\prod_{j=1}^J e^{D_{ij}^2\sigma^2/(2K)}\\=&\exp\Big\{\frac{\sigma^2}{2K}\sum_{i=1}^I\sum_{j=1}^J D_{ij}^2\Big\}\end{aligned}$$

と書けることがわかる。$\sum_{i=1}^I\sum_{j=1}^J D_{ij}^2$ を計算するので，D_{ij} を次のように変形するのがよい。

$$\begin{aligned}D_{ij}=&\frac{1}{IJ}\sum_{i=1}^I\sum_{j=1}^J\Big\{(Is_i-\sum_a s_a)+(Jt_j-\sum_b t_b)\\&+(IJu_{ij}-I\sum_b u_{ib}-J\sum_a u_{aj}+\sum_a\sum_b u_{ab})\Big\}\end{aligned}$$

この形を利用すると

$$\begin{aligned}\sum_{i=1}^I\sum_{j=1}^J D_{ij}^2=&\frac{1}{I^2J^2}\Big\{J\sum_{i=1}^I(Is_i-\sum_a s_a)^2+I\sum_{j=1}^J(Jt_j-\sum_b t_b)^2\\&+\sum_{i=1}^I\sum_{j=1}^J(IJu_{ij}-I\sum_b u_{ib}-J\sum_a u_{aj}+\sum_a\sum_b u_{ab})^2\Big\}\end{aligned}$$

と書けることがわかる。こうして，s_i の項，t_j の項，u_{ij} の項に分解できるので，

$$\begin{aligned}&\{z_{i\cdot}-z_{\cdot\cdot}, i=1,\ldots,I\},\ \{z_{\cdot j}-z_{\cdot\cdot}, j=1,\ldots,J\},\\&\{z_{ij}-z_{i\cdot}-z_{\cdot j}+z_{\cdot\cdot}, i=1,\ldots,I, j=1,\ldots,J\}\end{aligned}$$

は独立になる。従って，$\mathrm{SS_A}$, $\mathrm{SS_B}$, $\mathrm{SS_{AB}}$ は独立になることが示される。

(2) については，

$$\mathrm{SS_E}/\sigma^2=\sum_i\sum_j\sum_k(\varepsilon_{ijk}-\overline{\varepsilon}_{ij\cdot})^2/\sigma^2\sim\chi^2_{IJ(K-1)}$$

に従う。

(3) については，H_α のもとで，

$$\frac{\mathrm{SS_A}}{\sigma^2} = \frac{\sum_i (\bar{\varepsilon}_{i\cdot\cdot} - \bar{\varepsilon}_{\cdots})^2}{\sigma^2/JK}, \quad \bar{\varepsilon}_{i\cdot\cdot} \sim \mathcal{N}\Big(0, \frac{\sigma^2}{JK}\Big)$$

に注意すると，$\mathrm{SS_A}/\sigma^2 \sim \chi^2_{I-1}$ に従う。

(4) については，(3) と同様にして，H_β のもとで

$$\frac{\mathrm{SS_B}}{\sigma^2} = \frac{\sum_j (\bar{\varepsilon}_{\cdot j\cdot} - \bar{\varepsilon}_{\cdots})^2}{\sigma^2/IK}, \quad \bar{\varepsilon}_{\cdot j\cdot} \sim \mathcal{N}\Big(0, \frac{\sigma^2}{IK}\Big)$$

より，$\mathrm{SS_B}/\sigma^2 \sim \chi^2_{J-1}$ に従う。

(5) については，H_γ のもとで

$$\frac{\mathrm{SS_{AB}}}{\sigma^2} = \frac{\sum_i \sum_j (\bar{\varepsilon}_{ij\cdot} - \bar{\varepsilon}_{i\cdot\cdot} - \bar{\varepsilon}_{\cdot j\cdot} + \bar{\varepsilon}_{\cdots})^2}{\sigma^2/K}$$

を考える。

$$\sum_i \sum_j \sum_k (\varepsilon_{ijk} - \bar{\varepsilon}_{\cdots})^2/\sigma^2 = \mathrm{SS_A}/\sigma^2 + \mathrm{SS_B}/\sigma^2 + \mathrm{SS_{AB}}/\sigma^2 + \mathrm{SS_E}/\sigma^2$$

であり，

$$\frac{\sum_i \sum_j \sum_k (\varepsilon_{ijk} - \bar{\varepsilon}_{\cdots})^2}{\sigma^2} \sim \chi^2_{IJK-1},\ \frac{\mathrm{SS_A}}{\sigma^2} \sim \chi^2_{I-1},\ \frac{\mathrm{SS_B}}{\sigma^2} \sim \chi^2_{J-1},\ \frac{\mathrm{SS_E}}{\sigma^2} \sim \chi^2_{IJ(K-1)}$$

である。自由度については

$$(IJK-1) - (I-1) - (J-1) - IJ(K-1) = (I-1)(J-1)$$

となる。次の補題を用いれば，$\mathrm{SS_{AB}}/\sigma^2 \sim \chi^2_{(I-1)(J-1)}$ が導かれる。

補題 確率変数 X と Y が独立で，$X \sim \chi^2_m$, $X+Y \sim \chi^2_{m+n}$ に従うならば，$Y \sim \chi^2_n$ に従う。

この補題は，定理 6.8 の証明に中で使われていて

$$\begin{aligned} \mathrm{E}[e^{tY}] &= \mathrm{E}[e^{tY}]\frac{\mathrm{E}[e^{tX}]}{\mathrm{E}[e^{tX}]} = \frac{\mathrm{E}[e^{t(X+Y)}]}{\mathrm{E}[e^{tX}]} \\ &= \frac{(1-2t)^{m/2}}{(1-2t)^{(m+n)/2}} = \frac{1}{(1-2t)^{n/2}} \end{aligned}$$

となり，$Y \sim \chi^2_n$ が示される。

第16章　分布によらない推測法

問 1　ある会社の 12 人の月給が，24, 26, 28, 34, 34, 34, 36, 36, 40, 42, 50, 100(万円) とする。

(1) 標本平均 $\overline{x} = 40.3$，幾何平均 $G = 37.4$，調和平均 $H = 35.5$，メディアン $x_{\mathrm{med}} = 35$，$r = 1$ の刈り込み平均は 36 となる。

(2) 標本分散 $S_x^2 = 369.9$，メディアン絶対偏差は 6，範囲は $100 - 24 = 76$ である。分位点については (16.7) の定義を用いる。

$$q_{0.25} = \frac{x_{(3)} + x_{(4)}}{2} = \frac{28 + 34}{2} = 31,\ q_{0.75} = \frac{x_{(9)} + x_{(10)}}{2} = \frac{40 + 42}{2} = 41$$

より四分位範囲は $41 - 31 = 10$ となる。

(3) 下側 40% 点については，$n\alpha = 12 \times 0.4 = 4.8$ より，下から 5 番目以上で最小の値は 34 であり，上から 8 番目以下で最大の値も 34 なので，下側 40% 点は 34 になる。

上側 30% 点については，$n\alpha = 12 \times 0.7 = 8.4$ より，下から 9 番目以上で最小の値は 40 であり，上から 4 番目以下で最大の値も 40 なので，上側 30% 点は 40 になる。

```
> salary <- c(24,26,28,34,34,34,36,36,40,42,50,100)
> summary(salary)
   Min. 1st Qu. Median Mean 3rd Qu. Max.
  24.00 32.50 35.00 40.33 40.50 100.00
> exp(mean(log(salary)))
[1] 37.39429
> 1/mean(1/salary)
[1] 35.4565
> salary[2:11]
 [1] 26 28 34 34 34 36 36 40 42 50
> salary.trim <- mean(salary[2:11])
> salary.trim
[1] 36
> n <- length(salary)
> var(salary)*(n-1)/n
[1] 369.8889
> max(salary)-min(salary)
[1] 76
> median( abs(salary-median(salary)))
[1] 6
> quantile(salary, 0.75)-quantile(salary, 0.25)
75%
  8
> quantile(salary, 0.4)
40%
 34
> quantile(salary, 0.7)
 70%
38.8
```

問 2　5 人の生徒の数学と英語の得点が,

	1	2	3	4	5
数学	20	30	50	40	100
英語	40	30	60	50	0

であるから，ピアソンの標本相関係数は $R = -0.68$ になる。順位は

	1	2	3	4	5
数学	1	2	4	3	5
英語	3	2	5	4	1

より，スピアマンの順位相関係数は $\rho = -0.1$ となる。また，ケンドールの順位相関係数は

$$\tau = \frac{2}{5 \times 4}\{(-1+1+1-1)+(1+1-1)+(1-1)+(-1)\} = 0$$

となる。

　ピアソンの標本相関係数はデータ $(100, 0)$ の影響を受けるが，スピアマンの順位相関係数もケンドールの順位相関係数も影響を受けないことがわかる。

```
> A <- c(20,30,50,40,100)
> B <- c(40,30,60,50,0)
> cor(A,B)
[1] -0.6764223
> cor(A,B, method="spearman")
[1] -0.1
> cor(A,B, method="kendall")
[1] 0
```

問 3　例 8.1 で取り上げた身長データについて，平均 μ を標本平均 $\overline{X}$ で推定することを考える。

　(1) $B = 10000$ 個のブートストラップ標本を発生させ，$\overline{X}^*$ のヒストグラムを描くと，次のようになる。

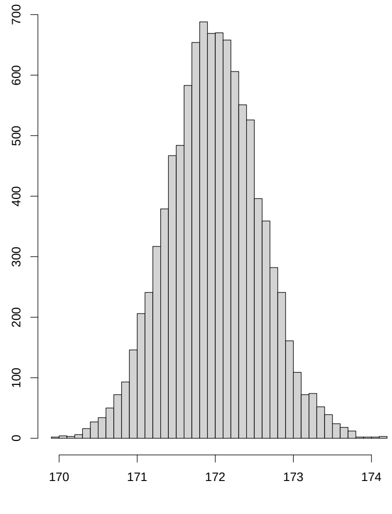

図 16.1: ブートストラップ標本による $\overline{X}^*$ の分布

(2) $\overline{X}$ のバイアスと分散のブートストラップ推定値は，それぞれ $-0.007, 0.343$ になる。ブートストラップ標本に基づいた μ の信頼係数 95% の信頼区間は，$[170.8, 173.2]$ になる。

```
> nsims<-10000
> x <- tall
> n <-length(x)
> tmpdata <- sample(x,nx*nsims, replace=TRUE)
> bootsp <- matrix(tmpdata, nrow=nx, ncol=nsims,byrow=T)
> xbarm <- 1/colMeans(bootsp)
> mean(xbarm)  #標本平均の平均値
[1] 0.005813781
> sd(xbarm) #標本平均の標準誤差
[1] 2.037755e-05
> mean(xbarm)-1/mean(x) #偏り（バイアス）
[1] -1.726714e-07
> quantile(xbarm, 0.025, type = 1)
       2.5%
0.005774339
> quantile(xbarm, 0.975, type = 1)
      97.5%
0.005852745
> quantile(xbarm, 0.025)
       2.5%
0.005774989
> quantile(xbarm, 0.975)
      97.5%
0.005852745
```

問 4　例 8.2 で取り上げたデータについて，平均 μ の逆数 $\lambda = 1/\mu$ を $\hat{\lambda} = 1/\overline{X}$ で推定することを考える。

(1) $B = 10000$ 個のブートストラップ標本を発生させ，$\hat{\lambda}^*$ のヒストグラムを描くと，次のようになる。

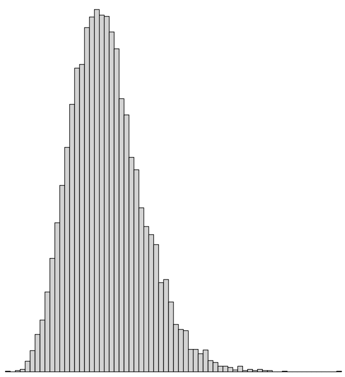

図 16.2: ブートストラップ標本による $\hat{\lambda}^*$ の分布

(2) $\hat{\lambda}$ のバイアスと分散のブートストラップ推定値は，それぞれ $0.0009, 0.0073^2$ になる。ブートストラップ標本に基づいた λ の信頼係数 95% の信頼区間は $[0.040, 0.069]$ になる。

```
> set.seed(123)
> nsims<-10000
> x <- crime
> nx<-length(x)
> tmpdata <- sample(x,nx*nsims, replace=TRUE)
> bootsp <- matrix(tmpdata, nrow=nx, ncol=nsims,byrow=T)
> xbarm <- 1/colMeans(bootsp)
> mean(xbarm)  #標本平均の平均値
[1] 0.05306791
> sd(xbarm) #標本平均の標準誤差
[1] 0.007358099
> mean(xbarm)-1/mean(x) #偏り（バイアス）
[1] 0.001478975
> quantile(xbarm, 0.025, type = 1)
      2.5%
0.04081299
> quantile(xbarm, 0.975, type = 1)
     97.5%
0.06916586
> quantile(xbarm, 0.025)
     2.5%
0.0408911
> quantile(xbarm, 0.975)
     97.5%
0.06916706
```

問5 次の表は，10人について禁煙前の体重と禁煙から5週間経過後の体重を調べたデータである。

個人	1	2	3	4	5	6	7	8	9	10
禁煙前 (y)	66	80	69	52	75	67	70	66	55	74
禁煙後 (x)	71.5	82.5	68.2	56	73	72	73	69.5	54.4	73.6
$x-y$	5.5	2.5	-0.8	4	-2	5	3	3.5	-0.6	-0.4
符号	$+$	$+$	$-$	$+$	$-$	$+$	$+$	$+$	$-$	$-$
$\|x-y\|$	5.5	2.5	0.8	4	2	5	3	3.5	0.6	0.4
順位	10	5	3	8	4	9	6	7	2	1

(1) 禁煙前後で変化があったか否かの片側検定を符号検定を用いて検定する。$S_n^+ = 6$ より

$$\left(6 - \frac{10}{2} - \frac{1}{2}\right)/\sqrt{10/4} = 0.32 < z_{0.05} = 1.64$$

より，5% で有意でない。

(2) ウィルコクソンの符号付き順位和検定を用いて片側検定する。上の表から

$$W_+ = 10 + 5 - 3 + 8 - 4 + 9 + 6 + 7 - 2 - 1 = 35$$

より

$$\left(35 - \frac{10 \times 11}{4} - \frac{1}{2}\right)\frac{\sqrt{24}}{\sqrt{10 \times 11 \times 21}} = 0.71 < z_{0.05} = 1.64$$

となり，5% で有意でない。

禁煙前後で変化があったか否かを符号検定を用いて検定する。

```
> A <- c(66,80,69,52,75,67,70,66,55,74)
> B <- c(71.5,82.5,68.2,56,73,72,73,69.5,54.4,73.6)
> C <- A-B
> C
 [1] -5.5 -2.5 0.8 -4.0 2.0 -5.0 -3.0 -3.5 0.6 0.4
> binom.test(4,10)

       Exact binomial test

data: 4 and 10
number of successes = 4, number of trials = 10, p-value =
0.7539
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.1215523 0.7376219
sample estimates:
probability of success
                 0.4
```

ウィルコクソンの符号付き順位和検定を用いて検定する。

```
> wilcox.test(C)

       Wilcoxon signed rank exact test

data: C
V = 10, p-value = 0.08398
alternative hypothesis: true location is not equal to 0
```

問 6 野球選手とバスケットボール選手の移動のときの手荷物の重量 (kg) を調べたところ，野球選手は，16.3, 20.1, 18.7, 18.2, 15.1, 13.1, 17.3, 16.6, 16.4, 13.7 であり，バスケットボール選手は, 16.0, 16.5, 15.2, 13.8, 12.8, 14.2, 18.6, 16.8, 17.7, 18.1 であった。差があるか否かをウィルコクソンの順位和検定を用いて両側検定する。全体の順位は

野球選手	9	20	19	17	6	2	14	12	10	3
バスケ選手	8	11	7	4	1	5	18	13	15	16

となる。

$$W = \sum_{i=1}^{n} R_i = 8 + 11 + 7 + 4 + 1 + 5 + 18 + 13 + 15 + 16 = 98$$

であるから

$$|T| = \left|98 - \frac{10 \times 21}{2} - \frac{1}{2}\right| \frac{\sqrt{12}}{\sqrt{10^2 \times 21}} = 0.57 < z_{0.025} = 1.96$$

となり，有意でない。

```
> A <- c(16.3, 20.1, 18.7, 18.2, 15.1, 13.1, 17.3, 16.6, 16.4, 13.7)
> B <- c(16.0, 16.5, 15.2, 13.8, 12.8, 14.2, 18.6, 16.8, 17.7, 18.1)
> wilcox.test(A,B)

       Wilcoxon rank sum exact test

data: A and B
W = 57, p-value = 0.6305
alternative hypothesis: true location shift is not equal to 0
```

問 7　問 6 に関連して，バレーボール選手については，15.5, 16.1, 14.7, 13.3, 12.4, 13.6 であった。野球選手，バスケットボール選手，バレーボール選手の間で差があるか否かをクラスカル・ウォリス検定を用いて検定する。全体の順位は

											合計
野球選手	15	26	25	23	10	3	20	18	16	6	162
バスケ選手	13	17	11	7	2	8	24	19	21	22	144
バレー選手	12	14	9	4	1	5					45

となる。$R_{1.} = 162,\ R_{2.} = 144,\ R_{3.} = 45,\ n_1 = 10,\ n_2 = 10,\ n_3 = 6,\ N = 26$ であるから

$$H = \frac{12}{26\times 27}\left\{10\left(\frac{162}{10} - \frac{27}{2}\right)^2 + 10\left(\frac{144}{10} - \frac{27}{2}\right)^2 + 6\left(\frac{45}{6} - \frac{27}{2}\right)^2\right\} = 5.08 < \chi^2_{2,0.05} = 5.99$$

となり，有意でない。

```
> C <- c(15.5, 16.1, 14.7, 13.3, 12.4, 13.6)
> length(A)
[1] 10
> length(B)
[1] 10
> length(C)
[1] 6
> kruskal.test( c(A,B,C) ~ rep(1:3,c(10,10,6)))

        Kruskal-Wallis rank sum test

data: c(A, B, C) by rep(1:3, c(10, 10, 6))
Kruskal-Wallis chi-squared = 5.0769, df = 2, p-value = 0.07899
```

問 8　2 つの事象 O と X が，OXOXOXOXOXOXOXOXOXOXOO のように起こるとき，これはランダムに起こると見なしてよいかの問である。連検定を用いると，$n = 20,\ r = 19$ より

$$\left|19 - 1 - \frac{2\times 11\times 9}{20}\right|\frac{\sqrt{20^2\times 19}}{\sqrt{2\times 11\times 9\times(2\times 11\times 9 - 20)}} = 3.76 > z_{0.025} = 1.96$$

となり，有意水準 5% で有意になる。

```
> r <- factor(c(0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,0))
> runs.test(r)

        Runs Test

data: r
Standard Normal = 3.7614, p-value = 0.000169
alternative hypothesis: two.sided
```

問 9　W は (16.12) のワイブル分布に従う確率変数とする。

(1) W の確率密度関数は

$$f_W(w) = abw^{b-1}e^{-aw^b}, \quad w > 0, a > 0, b > 0$$

で与えられるので，分布関数は

$$F_W(w) = \int_0^w f_W(x)dx = \Big[-e^{-ax^b}\Big]_0^w = 1 - e^{-aw^b}$$

となる。

$X = aW^b$ とおくとき，$W = a^{-1/b}X^{1/b}$ であり，$dw = (1/b)a^{-1/b}x^{1/b-1}dx$ より，X の確率密度は

$$f_X(x) = ab\Big(a^{-1/b}x^{1/b}\Big)^{b-1}e^{-x}\frac{1}{b}a^{-1/b}x^{1/b-1} = e^{-x}$$

となる。従って $X \sim Ex(1)$ に従うことがわかる。

(2) 一様乱数 $U \sim U(0,1)$ に対して

$$F_W(W) = 1 - e^{-aW^b} = U$$

を解くと，$-aW^b = \log(1-U)$，すなわち

$$W = \Big(-\frac{\log(1-U)}{a}\Big)^{1/b}$$

となり，これがワイブル分布に従う乱数になる。

問 10 非負の連続確率変数 X に対して平均余寿命関数 $r(t) = \mathrm{E}[X-t|X \geq t]$ を考える。

(1) X の分布関数を $F(x)$ とする。このとき

$$r(t) = \frac{\int_t^\infty (x-t)f(x)dx}{1-F(t)}$$

と表される。ここで $x - t = \int_t^x 1dy$ を代入すると，分子は

$$\int_t^\infty (x-t)f(x)dx = \int_t^\infty \Big(\int_t^x 1dy\Big)f(x)dx = \int_t^\infty \int_t^x f(x)dydx$$

と書ける。ここで，積分範囲 $\{t \leq x < \infty,\ t \leq y < x\}$ を $\{t \leq y < \infty,\ y < x < \infty\}$ と変えて積分の順序を交換すると

$$\int_t^\infty (x-t)f(x)dx = \int_t^\infty \int_y^\infty f(x)dxdy = \int_t^\infty \{1-F(y)\}dy$$

と表すことができる。従って，等式

$$r(t) = \frac{\int_t^\infty (1-F(x))dx}{1-F(t)}$$

が得られる。

(2) X が (16.12) のワイブル分布に従うとき，

$$1 - F(t) = \int_t^\infty abx^{b-1}e^{-ax^b}dx = \Big[-e^{-ax^b}\Big]_t^\infty = e^{-at^b}$$

より

$$r(t) = \frac{\int_t^\infty e^{-ax^b}dx}{e^{-at^b}}$$

となる。ロピタルの定理を用いると

$$\lim_{t\to\infty} r(t) = \lim_{t\to\infty}\frac{-e^{-at^b}}{-abt^{b-1}e^{-at^b}} = \frac{1}{ab}\lim_{t\to\infty}\frac{1}{t^{b-1}}$$

となるので，これが 0 に収束するためには $b > 1$ である条件が必要である。すなわち，ワイブル分布において平均余寿命が時間の経過とともに最終的に 0 になるためには $b > 1$ の条件が必要である。

問 11　分布関数が $F(t) = 1 - e^{-at^b}$ で与えられるワイブル分布について，寿命に関する観測データ $x_1, \ldots, x_n$ の分布がワイブル分布に従っているかを視覚的に調べてみたい。

(1) a, b の適当な一致推定量を $\hat{a}, \hat{b}$ とし，順序データを $x_{(1)} \leq \ldots \leq x_{(n)}$ とする。$F(X_1), \ldots, F(X_n)$ は i.i.d. で $U(0,1)$ に従う。順序統計量は $F(X_{(1)}) \leq \cdots \leq F(X_{(n)})$ は一様分布の順序統計量になるので，

$$\mathrm{E}[F(X_{(i)})] = \frac{i}{n+1}$$

となる。そこで，$i = 1, \ldots, n$ に対して

$$F(x_{(i)}) = \frac{i}{n+1} \Longleftrightarrow x_{(i)} = \exp\left\{\frac{\log(-\log(1-\frac{i}{n+1})) - \log a}{b}\right\}\Big)$$

となるので，a, b の適当な一致推定量を $\hat{a}, \hat{b}$ を代入して

$$\Big(x_{(i)},\ \exp\left\{\frac{\log(-\log(1-\frac{i}{n+1})) - \log \hat{a}}{\hat{b}}\right\}\Big), \quad i = 1, \ldots, n$$

をプロットしたときに，直線 $y = x$ に沿ってプロットされていれば，ワイブル分布が妥当であると示唆される。これが，ワイブル分布の Q-Q プロットである。

(2) 本文で取り上げた Gehan(1965) のデータについて，対照群の 21 人のデータがワイブル分布に従うかを Q-Q プロットを描いて調べる。ただし，$\hat{b} = 1.37$，$\log(\hat{a}) = -\hat{b} \times \log(9.48)$ を用いて計算する。データのワイブル分布への当てはめは良さそうである。

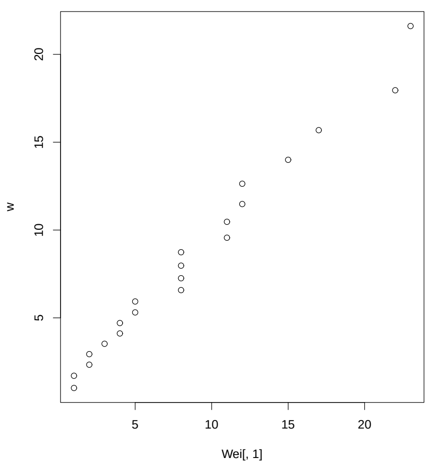

図 16.3: ワイブル分布の QQ プロット

```
> install.packages("fitdistrplus")
> library(fitdistrplus)
> Wei <- read.table(file="C:Weibull.txt", header=T)
> fit.wei <- fitdist(Wei[,1],"weibull")
> summary(fit.wei)
Fitting of the distribution '␣weibull␣' by maximum likelihood
Parameters :
      estimate Std. Error
shape 1.370430 0.2378693
scale 9.482705 1.5914768
Loglikelihood: -64.92011 AIC: 133.8402 BIC: 135.9293
Correlation matrix:
         shape scale
shape 1.0000000 0.3156842
scale 0.3156842 1.0000000

> b <- 1.37
> a1=1/(9.482705)^(1.37)
> I <- c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21)
> w <- exp((log(-log(1-I/22)))/b+log(9.4827))
> plot(Wei[,1],w)
> abline(0,1)
```

問 12　ある物質の発がん性を調べるため 10 匹の動物実験を行ったところ，物質投与から腫瘍発症までの日数が次のようであったとする。

40^+ 40^+ 45 45 45^+ 48^+ 50 50^+ 58　100

カプラン・マイヤー法による生存関数の推定値を与える。

t	45	50	58	100
x	2	1	1	1
n	8	4	2	1
推定値	0.75	0.56	0.28	0

$$1-\frac{x_1}{n_1}=\frac{3}{4}$$
$$\left(1-\frac{x_1}{n_1}\right)\left(1-\frac{x_2}{n_2}\right)=\frac{3}{4}\left(1-\frac{1}{4}\right)=\frac{9}{16}=0.56$$
$$\left(1-\frac{x_1}{n_1}\right)\left(1-\frac{x_2}{n_2}\right)\left(1-\frac{x_3}{n_3}\right)=\frac{9}{16}\left(1-\frac{1}{2}\right)=\frac{9}{32}=0.28$$

としてカプラン・マイヤー推定値が得られる。

問 13　不等式 (16.3) については，x_{med} をメディアンとし，すべての定数 a に対して次の不等式を示す。

$$\sum_{i=1}^{n}|x_i-a|\geq\sum_{i=1}^{n}|x_i-x_{\mathrm{med}}|$$

自然数 m に対して $n=2m+1$ のとき，メディアンは $x_{(m+1)}$ になる。このとき

$$\begin{aligned}\sum_{i=1}^{2m+1}|x_{(i)}-x_{(m+1)}| &= -\sum_{i=1}^{m}(x_{(i)}-x_{(m+1)})+\sum_{i=m+2}^{2m+1}(x_{(i)}-x_{(m+1)})\\ &= -\sum_{i=1}^{m}x_{(i)}+\sum_{m+2}^{2m+1}x_{(i)}\end{aligned}$$

となる。この値と $\sum_{i=1}^{2m+1}|x_{(i)}-a|$ の値との差を求めて，その差が非負であることを示せばよい。例えば，$x_{(m+1)} \le a < x_{(m+2)}$ の範囲にあるときには，

$$\begin{aligned}\sum_{i=1}^{2m+1}|x_{(i)}-a| &= -\sum_{i=1}^{m+1}(x_{(i)}-a)+\sum_{i=m+2}^{2m+1}(x_{(i)}-a)\\ &= -\sum_{i=1}^{m+1}x_{(i)}+\sum_{m+2}^{2m+1}x_{(i)}+a\end{aligned}$$

と書ける。従って，$\sum_{i=1}^{2m+1}|x_{(i)}-a|-\sum_{i=1}^{2m+1}|x_{(i)}-x_{(m+1)}| = a-x_{(m+1)} \ge 0$ が成り立つことがわかる。あらゆる範囲の a に対して同様の方法で不等式を示すことができる。また $n=2m$ のときにも同様にして示すことができるので確かめてほしい。

不等式 (16.4) については，正のデータ $x_1,\dots,x_n$ の幾何平均，調和平均を G_x, H_x とするとき，$\mathrm{H}_x \le \mathrm{G}_x \le \overline{x}$ が常に成り立つことを示す。

この不等式は凸関数に関するイェンセンの不等式が本質的である。$g(\cdot)$ を凸関数とすると

$$\frac{1}{n}\sum_{i=1}^{n}g(x_i) \ge g(\overline{x}), \quad \overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$$

のような不等式が成り立つ。証明は，点 $(\overline{x}, g(\overline{x}))$ を通る直線 $y=a(x-\overline{x})+g(\overline{x})$ を考えると，$g(x)$ は凸関数だから，a を適当にとって

$$g(x) \ge a(x-\overline{x})+g(\overline{x})$$

がすべての x に対して成り立つ。従って

$$\frac{1}{n}\sum_{i=1}^{n}g(x_i) \ge a\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})+g(\overline{x}) = g(\overline{x})$$

となり，求める不等式が得られる。

まず，

$$\Big(\prod_{i=1}^{n}x_i\Big)^{1/n} \le \frac{1}{n}\sum_{i=1}^{n}x_i = \overline{x}$$

を示そう。対数をとってマイナス符号を付けると

$$\frac{1}{n}\sum_{i=1}^{n}\{-\log(x_i)\} \ge -\log(\overline{x})$$

と書ける。ここで $g(x)=-\log x$ とおくと，$g'(x)=-1/x$, $g''(x)=1/x^2$ となり，凸関数であることがわかる。従って，示す不等式は上のイェンセンの不等式から得られることがわかる。

次に

$$\Big(\prod_{i=1}^{n}x_i\Big)^{1/n} \ge \frac{n}{\sum_{i=1}^{n}1/x_i}$$

を示そう。$y_i=1/x_i$ とおくと

$$\Big(\prod_{i=1}^{n}\frac{1}{y_i}\Big)^{1/n} \ge \frac{n}{\sum_{i=1}^{n}y_i} = \frac{1}{\overline{y}}, \quad \overline{y}=\frac{1}{n}\sum_{i=1}^{n}y_i$$

と書き直すことができる。両辺に対数をとると

$$\frac{1}{n}\sum_{i=1}^{n}\log\frac{1}{y_i} \ge \log\frac{1}{\overline{y}}$$

すなわち

$$\frac{1}{n}\sum_{i=1}^{n}\{-\log(y_i)\} \geq -\log(\overline{y})$$

と書ける。これは上で示されているので，この不等式も成り立つことがわかる。

問 14 不等式 (16.5) については，データ $x_1,\dots,x_n$ の標本平均，標本分散を $\overline{x}, S_x^2$ とする。正の定数 k に対して $|x_i-\overline{x}| \geq ks$ を満たす x_i の個数を n_k で表すとき，$n_k/n \leq 1/k^2$ が常に成り立つことを示せ。k を正の定数とし，$|x_i-\overline{x}| \geq kS_x$ を満たすようなデータ x_i の個数を n_k とする。このとき，常に $n_k/n \leq 1/k^2$ が成り立つ。

データに関するチェビシェフの不等式は，次のようにして確かめられる。$|x_i-\overline{x}| \geq kS_x$ を満たすようなインデックス i の集合を A とする。すなわち，$A=\{i : |x_i-\overline{x}| \geq kS_x\}$ である。集合 A に入る i の個数が n_k となる。A の補集合を A^c で表すと

$$nS_x^2=\sum_{i=1}^{n}(x_i-\overline{x})^2=\sum_{i\in A}(x_i-\overline{x})^2+\sum_{i\in A^c}(x_i-\overline{x})^2 \geq \sum_{i\in A}(x_i-\overline{x})^2$$

となる。ここで，A に属する i については $(x_i-\overline{x})^2 \geq k^2S_x^2$ が常に成り立つので

$$\sum_{i\in A}(x_i-\overline{x})^2 \geq \sum_{i\in A}k^2S_x^2 = n_kk^2S_x^2$$

となる。従って，$nS_x^2 \geq n_kk^2S_x^2$, 即ち $n \geq n_kk^2$ が成り立つので，上の不等式が得られる。

不等式 (16.6) については，メディアンを x_{med} とすると，$|\overline{x}-x_{\mathrm{med}}| < S_x$ が常に成り立つことを示す。

$$|\overline{x}-\mathrm{med}_x| = \left|\frac{1}{n}\sum_{i=1}^{n}(x_i-\mathrm{med}_x)\right| \leq \frac{1}{n}\sum_{i=1}^{n}|x_i-\mathrm{med}_x|$$

が成り立つことに注意する。ここでメディアンは$\sum_{i=1}^{n}|x_i-a|$ を最小する解を与えるので，$\sum_{i=1}^{n}|x_i-\mathrm{med}_x| \leq \sum_{i=1}^{n}|x_i-\overline{x}|$ なる不等式が成り立つことがわかる。さらに

$$\frac{1}{n}\sum_{i=1}^{n}|x_i-\overline{x}| \leq \left\{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2\right\}^{1/2}=S_x$$

なる不等式が成り立つことが確かめられる。この不等式は，任意の正の定数 $c_1,\dots,c_n$ に対して $(\sum_{i=1}^{n}c_i)^2 \leq n\sum_{i=1}^{n}c_i^2$ なる不等式に書き直すことができ，数学的帰納法を用いて示すことができる。

問 15 $0<\alpha<1$ に対して

$$q_\alpha^L = \min_x\{X_i \leq x \text{ となる } X_i\text{の個数が } n\alpha\text{以上}\}$$
$$q_\alpha^R = \max_x\{X_i \geq x \text{ となる } X_i\text{の個数が } n(1-\alpha)\text{ 以上}\}$$

とおくとき，下側 $100\alpha\%$ 分位点 q_α を

$$q_\alpha=\frac{q_\alpha^L+q_\alpha^R}{2}$$

で定義する。順序統計量 $X_{(1)}<\cdots<X_{(n)}$ について第 1 四分位点と第 3 四分位点をこの定義に沿って求めてみる。自然数 m に対して，$n=4m$ のときには

$$q_{0.25}=\frac{X_{(m)}+X_{(m+1)}}{2},\ q_{0.75}=\frac{X_{(3m)}+X_{(3m+1)}}{2}$$

となる。$n = 4m + 1$ のときには $q_{0.25} = X_{(m+1)}$，$q_{0.75} = X_{(3m+1)}$ となる。$n = 4m + 2$ のときには $q_{0.25} = X_{(m+1)}$，$q_{0.75} = X_{(3m+2)}$ となる。$n = 4m + 3$ のときには $q_{0.25} = X_{(m+1)}$，$q_{0.75} = X_{(3m+3)}$ となる。従って，IQR は，$n = 4m$ のとき

$$\mathrm{IQR} = \frac{X_{(3m+1)} - X_{(m+1)}}{2} + \frac{X_{(3m)} - X_{(m)}}{2}$$

となり，$n = 4m + 1$ のときには $\mathrm{IQR} = X_{(3m+1)} - X_{(m+1)}$，$n = 4m + 2$ のときには $\mathrm{IQR} = X_{(3m+2)} - X_{(m+1)}$，$n = 4m + 3$ のときには $\mathrm{IQR} = X_{(3m+3)} - X_{(m+1)}$ となる。

第17章　多変量解析手法

問1　相関係数に関する問である。

(1) 見かけの相関と対処法について説明する。3つの変数 x, y, z について，x と z が相関し y と z が相関するとき，x と y の間に見かけ上の相関が生まれる。これを見かけの相関とよぶ。

2つの単回帰モデル $x = \alpha_0 + \alpha_1 z$, $y = \beta_0 + \beta_1 z$ を考え，それぞれの残差を e_x, e_y とおくとき，e_x と e_y の相関係数を偏相関係数とよぶ。それぞれの残差は変数 z の影響が取り除かれているので，真の相関を見ることができる。

(2) 3変数 (x_1, x_2, x_3) の標本相関行列 $\boldsymbol{R}$ の成分が，$r_{12} = 0.9$，$r_{13} = 0.8$，$r_{23} = 0.7$ であるから

$$\boldsymbol{R} = \begin{pmatrix} 0 & 0.9 & 0.8 \\ 0.9 & 0 & 0.7 \\ 0.8 & 0.7 & 0 \end{pmatrix}$$

と書ける。x_1 の影響を取り除いた x_2 と x_3 の偏相関は

$$r_{23|1} = \frac{r_{23} - r_{12}r_{13}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{13}^2}} = \frac{0.7 - 0.9 \times 0.8}{\sqrt{1-0.9^2}\sqrt{1-0.9^2}} = -0.08$$

となる。

問2　n 個の2次元データ $(x_1, y_1), \ldots, (x_n, y_n)$ の主成分分析は，$w_1^2 + w_2^2 = 1$ を満たす w_1, w_2 について合成変量 $w_1(x - \overline{x}) + w_2(y - \overline{y})$ の分散 $n^{-1}\sum_{i=1}^n \{w_1(x_i - \overline{x}) + w_2(y_i - \overline{y})\}^2$ を最大化することにより求めることができる。

(1) ラグランジュの未定乗数法を用いて，最大化を与える解を求める。合成変量の分散は

$$g(w_1, w_2) = \frac{1}{n}\sum_{i=1}^n \{w_1(x_i - \overline{x}) + w_2(y_i - \overline{y})\}^2 = w_1^2 S_{xx} + 2w_1 w_2 S_{xy} + w_2^2 S_{yy}$$

と書ける。ただし，$S_{xx} = n^{-1}\sum_{i=1}^n (x_i - \overline{x})^2$, $S_{xy} = n^{-1}\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})$, $S_{yy} = n^{-1}\sum_{i=1}^n (y_i - \overline{y})^2$ である。ラグランジュの未定乗数法により

$$L(w_1, w_2) = g(w_1, w_2) - \lambda(w_1^2 + w_2^2 - 1)$$

の解は，次の連立方程式の解になる。

$$\begin{aligned} \frac{\partial L}{\partial w_1} &= 2w_1 S_{xx} + 2w_2 S_{xy} - 2\lambda w_1 = 0 \\ \frac{\partial L}{\partial w_2} &= 2w_2 S_{yy} + 2w_1 S_{xy} - 2\lambda w_2 = 0 \end{aligned}$$

すなわち

$$(S_{xx} - \lambda)w_1 = -w_2 S_{xy}, \quad (S_{yy} - \lambda)w_2 = -w_1 S_{xy}$$

と書ける。左式から $w_1 = -w_2 S_{xy}/(S_{xx} - \lambda)$ が得られるので，これを右式に代入すると

$$(S_{xx} - \lambda)(S_{yy} - \lambda) = S_{xy}^2$$

が得られる。すなわち

$$\lambda^2 - (S_{xx} + S_{yy})\lambda + (S_{xx}S_{yy} - S_{xy}^2) = 0$$

と書ける。この解は

$$\lambda = \frac{1}{2}\left\{S_{xx} + S_{yy} \pm \sqrt{(S_{xx} - S_{yy})^2 + 4S_{xy}^2}\right\}$$

であり，それらを $\lambda_+, \lambda_-, (\lambda_+ > \lambda_-)$, で表すことにする。一方，

$$w_1 = -\frac{S_{xy}/(S_{xx} - \lambda)}{\sqrt{S_{xy}^2/(S_{xx} - \lambda)^2 + 1}}, \quad w_2 = \frac{1}{\sqrt{S_{xy}^2/(S_{xx} - \lambda)^2 + 1}}$$

であるから，λ_+, λ_- を代入したものを $w_{1+}, w_{1-}, w_{2+}, w_{2-}$ としたものが解となる。

(2) 行列表現を与え，固有値と固有ベクトルを求めることによって解を求めよ。分散 $g(w_1, w_2)$ は

$$g(w_1, w_2) = (w_1, w_2)\begin{pmatrix} S_{xx} & S_{xy} \\ S_{xy} & S_{yy} \end{pmatrix}\begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$$

と書けるので，共分散行列の固有値を求めると

$$\begin{vmatrix} S_{xx} - \lambda & S_{xy} \\ S_{xy} & S_{yy} - \lambda \end{vmatrix} = (S_{xx} - \lambda)(S_{yy} - \lambda) - S_{xy}^2 = 0$$

となり，この解は (1) で求めた λ_+, λ_- となることがわかる。$\lambda = \lambda_+$ に対する固有ベクトルは

$$\begin{pmatrix} S_{xx} & S_{xy} \\ S_{xy} & S_{yy} \end{pmatrix}\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \lambda_+\begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$$

の解，すなわち

$$(S_{xx} - \lambda_+)w_1 = -S_{xy}w_2, \quad (S_{yy} - \lambda_+)w_2 = -S_{xy}w_1$$

となり，(1) と同じ解が得られることがわかる。

(3) 第 1 主成分は λ_+ で，その固有値は w_{1+}, w_{2+} であるから，i 番目のデータについて第 1 主成分のスコアは

$$w_{1+}(x_i - \overline{x}) + w_{2+}(y_i - \overline{y})$$

となる。

問 3　ある小学校の 6 年生の身体計測について，身長，体重，胸囲，座高のデータから標本相関係数を求めたところ，固有値と固有ベクトルが次の表で与えられている。

累積寄与率は，表の中で与えられている。固有値が 1 以上になるものは最初の 1 つだけなので，主要な主成分の個数は 1 である。

主成分	固有値	累積寄与率 (%)	身長	体重	胸囲	座高
第 1 主成分	3.55	89	0.49	0.51	0.48	0.50
第 2 主成分	0.32	97	-0.54	0.21	0.72	-0.36
第 3 主成分	0.07	99	-0.45	-0.46	0.17	0.74
第 4 主成分	0.06	100	-0.50	0.69	-0.46	0.23

第 1 主成分は，身長，体重，胸囲，座高のすべてに対して正で同じような値を示しているので，体全体の大きさを表していると解釈できる。

問 4　患者がある病気の罹患群に入るか否かを判断するために，多変量正規分布に基づいた判別分析を用いることにした。2 つの群を $\Pi_1 : \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$, $\Pi_2 : \mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ とする。

(1) $\mathbf{\Sigma}_1$ と $\mathbf{\Sigma}_2$ が等しくないときには 2 次判別を用いて

$$(\boldsymbol{x}-\boldsymbol{\mu}_2)^\top\mathbf{\Sigma}_2^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_2)-(\boldsymbol{x}-\boldsymbol{\mu}_1)^\top\mathbf{\Sigma}_1^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)+\log(|\mathbf{\Sigma}_2|/|\mathbf{\Sigma}_1|)>C\Longrightarrow\ \Pi_1\text{に判別する}$$

となる。ただし

$$C=\log\{C(1|2)/C(2|1)\}+\log(\pi_2/\pi_1)$$

である。

(2) $\mathbf{\Sigma}_1=\mathbf{\Sigma}_2$ のときにはフィッシャーの線形判別を用いて

$$\Big(\boldsymbol{x}-\frac{\boldsymbol{\mu}_1+\boldsymbol{\mu}_2}{2}\Big)^\top\mathbf{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)>C\Longrightarrow\ \Pi_1\text{に判別する}$$

となる。また Π_2 が正しいのに Π_1 に判別する誤判別確率は

$$\begin{aligned}&\mathrm{P}\Big(\Big(\boldsymbol{X}-\frac{\boldsymbol{\mu}_1+\boldsymbol{\mu}_2}{2}\Big)^\top\mathbf{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)>C\Big|\boldsymbol{X}\sim\mathcal{N}(\boldsymbol{\mu}_2,\mathbf{\Sigma})\Big)\\&=\mathrm{P}\Big((\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^\top\mathbf{\Sigma}^{-1/2}\boldsymbol{Z}-\frac{1}{2}D^2>C\Big|\boldsymbol{Z}\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\Big)\end{aligned}$$

と書ける。ただし，$D^2=(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^\top\mathbf{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)$ である。ここで

$$(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^\top\mathbf{\Sigma}^{-1/2}\boldsymbol{Z}-\frac{1}{2}D^2\sim\mathcal{N}\Big(-\frac{1}{2}D^2,D^2\Big)$$

であるから，上の誤判別確率は

$$1-\Phi\Big(\frac{C+D^2/2}{D}\Big)$$

と表される。同様にして，Π_1 が正しいのに Π_2 に判別する誤判別確率は

$$\begin{aligned}&\mathrm{P}\Big(\Big(\boldsymbol{X}-\frac{\boldsymbol{\mu}_1+\boldsymbol{\mu}_2}{2}\Big)^\top\mathbf{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)<C\Big|\boldsymbol{X}\sim\mathcal{N}(\boldsymbol{\mu}_1,\mathbf{\Sigma})\Big)\\&=\mathrm{P}\Big((\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^\top\mathbf{\Sigma}^{-1/2}\boldsymbol{Z}+\frac{1}{2}D^2<C\Big|\boldsymbol{Z}\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})\Big)\end{aligned}$$

と書ける。ここで

$$(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^\top\mathbf{\Sigma}^{-1/2}\boldsymbol{Z}+\frac{1}{2}D^2\sim\mathcal{N}\Big(\frac{1}{2}D^2,D^2\Big)$$

であるから，上の誤判別確率は

$$\Phi\Big(\frac{C-D^2/2}{D}\Big)$$

と表される。

(3) 線形判別は

$$(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^\top\mathbf{\Sigma}^{-1}\boldsymbol{x}-\frac{1}{2}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2)^\top\mathbf{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)>C\Longrightarrow\ \Pi_1\text{に判別する}$$

と書き直すことができる。ここで，

$$\boldsymbol{\alpha}=-\frac{1}{2}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2)^\top\mathbf{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2),\quad\boldsymbol{\beta}^\top=(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^\top\mathbf{\Sigma}^{-1}$$

とおくと，

$$\boldsymbol{\alpha}+\boldsymbol{\beta}^\top\boldsymbol{x}>C\Longrightarrow\ \Pi_1\text{に判別する}$$

と表すことができる。線形判別は，この $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ を正規母集団から推定することになる。

これに対して，ロジスティック判別は

$$\log\left\{\frac{p(\boldsymbol{x})}{1-p(\boldsymbol{x})}\right\}=\boldsymbol{\alpha}+\boldsymbol{\beta}^\top\boldsymbol{x}>C\Longrightarrow \Pi_1 \text{に判別する}$$

という方法で，書き換えると

$$p(\boldsymbol{x})=\frac{e^{\boldsymbol{\alpha}+\boldsymbol{\beta}^\top\boldsymbol{x}}}{1+e^{\boldsymbol{\alpha}+\boldsymbol{\beta}^\top\boldsymbol{x}}}>\frac{c^C}{1+e^C}\Longrightarrow \Pi_1 \text{に判別する}$$

となる。この判別方法自体は線形判別と同じであるが，ロジスティック判別は $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ を尤度

$$\prod_{i=1}^n\{p(\boldsymbol{x}_i)\}^{y_i}\{1-p(\boldsymbol{x}_i)\}^{1-y_i}$$

に基づいて推定する方法をとる。この点が線形判別と異なる点である。

問 5　ある病気の患者群 Π_1 と健康群 Π_2 について，その病気と関係の深い 2 種類の検査項目に関するデータがとられた。Π_1 から 20 個のデータ，Π_2 から 40 個のデータに基づいて，それぞれの標本平均と標本共分散行列が次のように与えられている。両群でコストと事前確率は等しいものとして判別関数を求めたい。

$$\overline{\boldsymbol{x}}_1=\begin{pmatrix}50\\6\end{pmatrix},\ \boldsymbol{S}_1=\begin{pmatrix}250&10\\10&2\end{pmatrix},\ \overline{\boldsymbol{x}}_2=\begin{pmatrix}70\\5\end{pmatrix},\ \boldsymbol{S}_2=\begin{pmatrix}160&8\\8&2\end{pmatrix}$$

(1) 2 次判別に基づいた判別関数は

$$(\boldsymbol{x}-\overline{\boldsymbol{x}}_2)^\top\boldsymbol{S}_2^{-1}(\boldsymbol{x}-\overline{\boldsymbol{x}}_2)>(\boldsymbol{x}-\overline{\boldsymbol{x}}_1)^\top\boldsymbol{S}_1^{-1}(\boldsymbol{x}-\overline{\boldsymbol{x}}_1)\Longrightarrow \Pi_1 \text{に判別する}$$

で与えられるので，これに数値を代入すればよい。

(2) $\boldsymbol{\Sigma}_1=\boldsymbol{\Sigma}_2=\boldsymbol{\Sigma}$ とする場合，$\boldsymbol{\Sigma}$ の推定値は

$$\boldsymbol{S}=\frac{1}{20+40}(20\boldsymbol{S}_1+40\boldsymbol{S}_2)=\begin{pmatrix}190&8.7\\8.7&2\end{pmatrix}$$

で与えられるので

$$\left(\boldsymbol{x}-\frac{\overline{\boldsymbol{x}}_1+\overline{\boldsymbol{x}}_2}{2}\right)^\top\boldsymbol{S}^{-1}(\overline{\boldsymbol{x}}_1-\overline{\boldsymbol{x}}_2)=\left(\boldsymbol{x}-\begin{pmatrix}60\\5.5\end{pmatrix}\right)^\top\boldsymbol{S}^{-1}\begin{pmatrix}-20\\1\end{pmatrix}>0\Longrightarrow \Pi_1 \text{に判別する}$$

となる。

問 6　因子分析に関する問である。

(1) 主成分分析は確率モデルを想定せず，大きなバラツキを与える合成変量をいくつか見出して，多変量データ全体を説明する手法である。因子分析は，いくつかの共通因子と誤差項からデータが生成されるという確率モデルを想定し，共通因子が相関構造を生み出すことから共通因子の意味を見出していく方法である。。

(2) 3 次元データを 2 因子で解析するモデルは

$$\begin{pmatrix}z_{1i}\\z_{2i}\\z_{3i}\end{pmatrix}=f_{1i}\boldsymbol{a}_1+f_{2i}\boldsymbol{a}_2+\boldsymbol{u}_i,\quad i=1,\ldots,n$$

で与えられる。ここで

$$f_{1i}, f_{2i} \sim \mathcal{N}(0,1), \quad \boldsymbol{u}_i \sim \mathcal{N}\big(\mathbf{0}, \mathrm{diag}\,(d_1^2, d_2^2, d_3^2)\big), \quad \boldsymbol{a}_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix},\ \boldsymbol{a}_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \end{pmatrix}$$

である。行列を用いて書くと

$$\boldsymbol{z}_i = \boldsymbol{A}\boldsymbol{f}_i + \boldsymbol{u}_i, \quad \boldsymbol{A} = (\boldsymbol{a}_1, \boldsymbol{a}_2),\ \boldsymbol{f}_i = \begin{pmatrix} f_{1i} \\ f_{2i} \end{pmatrix}$$

と表される。

(3) 因子負荷量の推定には回転による自由度がある。これは，$\boldsymbol{H}$ を適当な直交行列とすると，$\boldsymbol{H}\boldsymbol{H}^\top = \boldsymbol{I}$ より

$$\boldsymbol{z}_i = \boldsymbol{A}\boldsymbol{H}\boldsymbol{H}^\top \boldsymbol{f}_i + \boldsymbol{u}_i$$

と書ける。$\boldsymbol{A}^* = \boldsymbol{A}\boldsymbol{H}$, $\boldsymbol{f}_i^* = \boldsymbol{H}^\top \boldsymbol{f}_i$ とおくと

$$\boldsymbol{z}_i = \boldsymbol{A}^*\boldsymbol{f}_i^* + \boldsymbol{u}_i, \quad \boldsymbol{f}_i^* \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_2)$$

と表されるので，$\boldsymbol{A}$ の推定には直交行列（回転）の自由度がある。

問 7 5 人の生徒 1, 2, 3, 4, 5 の国語と英語の成績 $(x_1, y_1), \dots, (x_5, y_5)$ が 5 点満点で次の表で与えられた。生徒間の類似度を $d(i,j) = |x_i - x_j| + |y_i - y_j|$ で測ることにする。

距離行列は以下のようになる。

	1	2	3	4	5
1	0	2	8	4	3
2	2	0	6	4	3
3	8	6	0	4	5
4	4	4	4	0	1
5	3	3	5	1	0

最短距離法を用いてクラスターを構成していくと

	1	2	3	(4,5)
1	0	2	8	3
2	2	0	6	3
3	8	6	0	4
(4,5)	3	3	4	0

$\Longrightarrow$

	(1,2)	3	(4,5)
(1,2)	0	6	3
3	6	0	4
(4,5)	3	4	0

$\Longrightarrow$

	(1,2,4,5)	3
(1,2,4,5)	0	4
3	4	0

となる。従って，3 個のクラスターに分ける場合は，$\{(1,2), 3, (4,5)\}$ となり，2 つのクラスターに分ける場合は $\{(1,2,4,5), 3\}$ となる。

問 8 銀行の窓口に来る客の 1 時間当たりの人数を 1ヶ月間調べたところ，$x_1, \dots, x_n$ のデータが得られた。時間帯や曜日などによって混み具合が異なるので，m 個のポアソン分布 $Po(\lambda_1), \dots, Po(\lambda_m)$ を想定して混合分布のモデル $\pi_1 Po(\lambda_1) + \cdots + \pi_m Po(\lambda_m)$ を考えてみる。$\lambda_1, \dots, \lambda_m, \pi_1, \dots, \pi_m$ を推定するための EM アルゴリズムを求める。

$\boldsymbol{Z}_i = (Z_{i1}, \dots, Z_{im})$, $\mathrm{P}(Z_{ij} = 1) = \pi_j$, $j = 1, \dots, m$, とすると，$(x_1, \boldsymbol{Z}_1), \dots, (x_n, \boldsymbol{Z}_n)$ の同時確率は

$$\prod_{i=1}^n \prod_{j=1}^m \left\{ \frac{\lambda_j^{x_i}}{x_i!} e^{-\lambda_j} \pi_j \right\}^{Z_{ij}}$$

と表され，対数尤度は

$$\ell(\boldsymbol{\pi}, \boldsymbol{\lambda}) = \sum_{i=1}^n \sum_{j=1}^m Z_{ij} \Big\{ x_i \log \lambda_j - \lambda_j - \log x_i! + \log \pi_j \Big\}$$

と書ける。$\boldsymbol{Z}_i$ は観測できないので，それを条件付き期待値で置き換える。すなわち

$$\mathrm{E}[\ell(\boldsymbol{\pi},\boldsymbol{\lambda})|\boldsymbol{x}] = \sum_{i=1}^{n}\sum_{j=1}^{m}\mathrm{E}[Z_{ij}|x_i]\Big\{x_i\log\lambda_j - \lambda_j - \log x_i! + \log\pi_j\Big\}$$

であり，

$$\mathrm{E}[Z_{ij}|x_i] = \frac{\lambda_j^{x_i}e^{-\lambda_j}\pi_j}{\sum_{k=1}^{m}\lambda_k^{x_i}e^{-\lambda_k}\pi_k} \equiv P(j|x_i,\boldsymbol{\pi},\boldsymbol{\lambda})$$

となる。(17.12) とその下の議論と同様にして

$$\pi_j = \frac{1}{n}\sum_{i=1}^{n}P(j|x_i,\boldsymbol{\pi},\boldsymbol{\lambda})$$

となる。また

$$\sum_{i=1}^{n}P(j|x_i,\boldsymbol{\pi},\boldsymbol{\lambda})\Big\{\frac{x_i}{\lambda_j}-1\Big\} = 0$$

より

$$\lambda_j = \frac{\sum_{i=1}^{n}P(j|x_i,\boldsymbol{\pi},\boldsymbol{\lambda})x_i}{\sum_{i=1}^{n}P(j|x_i,\boldsymbol{\pi},\boldsymbol{\lambda})}$$

となる。以上より，EM アルゴリズムは

$$\begin{aligned}\pi_j^{(t+1)} =& \frac{1}{n}\sum_{i=1}^{n}P(j|x_i,\boldsymbol{\pi}^{(t)},\boldsymbol{\lambda}^{(t)})\\ \lambda_j^{(t+1)} =& \frac{\sum_{i=1}^{n}P(j|x_i,\boldsymbol{\pi}^{(t+1)},\boldsymbol{\lambda}^{(t)})x_i}{\sum_{i=1}^{n}P(j|x_i,\boldsymbol{\pi}^{(t+1)},\boldsymbol{\lambda}^{(t)})}\end{aligned}$$

で与えられる。

問 9　σ^2 が既知の正規分布 $\mathcal{N}(\mu,\sigma^2)$ からデータがとられるが，定数 c を超えるときにはデータは欠損され観測されないとする。簡単のため σ^2 を既知とし $a=(c-\mu)/\sigma$ とおく。また $\phi(\cdot)$, $\Phi(\cdot)$ を $\mathcal{N}(0,1)$ の確率密度関数と分布関数とする。いま $x_1,\ldots,x_m$ が観測され，$n-m$ 個が欠損する場合を考える。

(1) 欠損する変数を $y_{m+1},\ldots,y_n$ で表すと，$(X_1,\ldots,X_m)$ の尤度関数は

$$L(\mu) = \int_c^\infty\cdots\int_c^\infty\prod_{i=1}^{m}f(x_i)\prod_{j=m+1}^{n}f(y_j)d\boldsymbol{y} = \Big\{\prod_{i=1}^{m}f(x_i)\Big\}\{1-F(c)\}^{n-m}$$

と表される。$(Y_j-\mu)/\sigma\sim\mathcal{N}(0,1)$ に従うので

$$1-F(c) = \mathrm{P}(Y_j>c) = 1-\Phi((c-\mu)/\sigma) = \Phi((\mu-c)/\sigma)$$

と書ける。従って，対数尤度は

$$\ell(\mu) = -\frac{m}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}(x_i-\mu)^2 + (n-m)\log\Phi((\mu-c)/\sigma)$$

であり，尤度方程式は

$$\ell'(\mu) = \frac{1}{\sigma^2}\sum_{i=1}^{m}(x_i-\mu) + (n-m)\frac{\sigma^{-1}\phi((\mu-c)/\sigma)}{\Phi((\mu-c)/\sigma)} = 0$$

と書ける。

(2) $X \sim \mathcal{N}(\mu, \sigma^2)$ とし欠損する場合の平均は

$$\mathrm{E}[X|c < X] = \frac{\int_c^\infty y \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\mu)^2/\sigma^2} dy}{\int_c^\infty \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\mu)^2/\sigma^2} dy}$$

と書ける。$z = (y - \mu)/\sigma$ とおくと，$y > c \iff z > (c - \mu)/\sigma = a$ であるから

$$\mathrm{E}[X|c < X] = \frac{\int_a^\infty (\sigma z + \mu) \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz}{\int_a^\infty \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz} = \mu + \sigma \frac{\int_a^\infty z \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz}{\int_a^\infty \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz}$$

と表される。ここで $\int_a^\infty \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = 1 - \Phi(a) = \Phi(-a)$ であり

$$\int_a^\infty z \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = \frac{1}{\sqrt{2\pi}} \Big[-e^{-z^2/2} \Big]_a^\infty = \phi(a)$$

と書けるので，

$$h(c, \mu) = \mathrm{E}[X|c < X] = \mu + \frac{\phi(a)}{\Phi(-a)} \sigma$$

となる。

(3) μ を推定するための EM アルゴリズムを求める。$(X_1, \ldots, X_m, Y_{m+1}, \ldots, Y_n)$ を欠損のない完全な確率変数の組とすると，対数尤度は

$$\ell(\mu) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^m (X_i - \mu)^2 - \frac{1}{2\sigma^2} \sum_{j=m+1}^n (Y_j - \mu)^2$$

より，尤度方程式は

$$\sum_{i=1}^m (X_i - \mu) + \sum_{j=m+1}^n (Y_j - \mu) = 0$$

と書ける。$\overline{X}_m = m^{-1} \sum_{i=1}^m X_i$ とおくと

$$\mu = \frac{m}{n} \overline{X}_m + \frac{1}{n} \sum_{j=1}^m Y_j$$

が得られる。これが M ステップになる。Y_j は観測できないので条件付き期待値 $h(c, \mu) = \mathrm{E}[X|c < X]$ で置き換えると

$$\mu = \frac{m}{n} \overline{X}_m + \frac{n - m}{n} h(c, \mu)$$

と表される。これが E ステップになる。

以上から，μ を求めるための EM アルゴリズムは

$$\mu^{(t+1)} = \frac{m}{n} \overline{X}_m + \frac{n - m}{n} h(c, \mu^{(t)})$$

で与えられる。