現代数理統計学の基礎
章末演習問題答案

橋本　龍二

久保川先生著『現代数理科学の基礎』の章末問題の答案を書く。略解は先生がはしがきに載せてくださっているが、略されすぎていることがあるので途中式、説明をくどめに作った。必要に応じて方針、使う定理や公式の証明、コメントも記載した。付した難易度は主観である。

Contents

1 確率

1.1 易

(方針)
ベン図を書く。

(答案)

$$\begin{aligned} P[A\Delta B] &= P[(A\cup B)\backslash(A\cap B)] \\ &= P[A\cup B]-P[A\cap B] \\ &= P[A]+P[B]-2P[A\cap B] \end{aligned}$$

1.2 易

$$\begin{aligned} & P[A_1\cup A_2\cup A_3] \\ = \ & P[(A_1\cup A_2)\cup A_3] \\ = \ & P[A_1\cup A_2]+P[A_3]-A[(A_1\cup A_2)\cap A_3] \\ = \ & P[A_1]+P[A_2]-P[A_1\cap A_2]+P[A_3]-P[(A_1\cap A_3)\cup(A_2\cap A_3)] \\ = \ & P[A_1]+P[A_2]+P[A_3]-P[A_1\cap A_2]-P[A_1\cap A_3]-P[A_2\cap A_3]+P[A_1\cap A_2\cap A_3] \end{aligned}$$

1.3 易

(ベイジアンネットワークのアイデアとなる変形)
$A_1 \cap ... \cap A_k = X_k$ と置くと、

$$\begin{aligned}
P[A_1 \cap ... \cap A_n] &= P[A_n \cap X_{n-1}] \\
&= P[A_n|X_{n-1}]P[X_{n-1}] \\
&= P[A_n|X_{n-1}]P[A_{n-1} \cap X_{n-2}] \\
&= P[A_n|X_{n-2}]P[A_{n-1}|X_{n-2}]P[X_{n-2}] \\
&\vdots \\
&= P[A_n|X_{n-1}]...P[A_k|X_{k-1}]...P[A_2|X_1]P[X_1] \\
&= P[A_n|A_1 \cap ... \cap A_{n-1}]...P[A_2|A_1]P[A_1]
\end{aligned}$$

1.4 奇

(ちょっと出題意図が分からない)

1. $B \subset A$ とすると、$P[A|B^c] \neq 0,\ P[A|B] = 1$ より題意を満たす。
2. $A \subset C$ とすると、$P[C|A] = 1$ より題意を満たす。

1.5 易

(方針)
「0 を送信する」という事象を A、「1 を送信する」という事象を A^c と置き、「0 を受信する」という事象を B、「1 を受信する」という事象を B^c と置き、ベイズの定理を用いて求める。

1.

$$\begin{aligned}
P[B] &= P[A \cap B] + P[A^c \cap B] \\
&= \frac{1}{3}\frac{9}{10} + \frac{2}{3}\frac{1}{10} \\
&= \frac{11}{30}
\end{aligned}$$

2.

$$
\begin{aligned}
P[A^c|B] &= \frac{P[A^c \cap B]}{P[B]} \\
&= \frac{\frac{2}{3}\frac{1}{10}}{\frac{11}{30}} \\
&= \frac{2}{11} \\
P[A|B^c] &= \frac{P[A \cap B^c]}{P[B^c]} \\
&= \frac{\frac{1}{3}\frac{1}{10}}{\frac{19}{30}} \\
&= \frac{1}{19}
\end{aligned}
$$

1.6　易

(方針)

偽陽性についての問題。実際に、「病気にかかっているのに陰性が出る (偽陰性)」確率を優先的に低く保ちつつ「病気にかかっていないのに陽性が出る (偽陽性)」確率ができるだけ低くなるように閾値が設定される。p 値と検出力の話に通じる。この確率は、実験によって「疾患がある人に検査をして陽性 (陰性) になる割合」と「疾患のない人に検査をして陽性 (陰性) になる割合」を知ることができれば求められる。

ここでは「陽性反応が出る」という事象を A、「陰性反応が出る」という事象を A^c とし、「疾患がある」という事象を B、「疾患がない」という事象を B^c と置いてベイズの定理を用いて求める。

(答案)

$$
\begin{aligned}
P[A|B^c] &= \frac{2}{10} \\
P[A^c|B] &= \frac{1}{10} \\
P[B] &= \frac{1}{10}
\end{aligned}
$$

であるから、

$$\begin{aligned} p[B|A] &= \frac{P[A\cap B]}{P[A]} \\ &= \frac{P[A\cap B]}{P[A\cap B]+P[A\cap B^c]} \\ &= \frac{(1-P[A^c|B])P[B]}{(1-P[A^c|B])P[B]+P[A|B^c]P[B^c]} \\ &= \frac{(1-\frac{1}{10})\frac{1}{10}}{(1-\frac{1}{10})\frac{1}{10}+\frac{2}{10}\frac{9}{10}} \\ &= \frac{1}{3} \end{aligned}$$

1.7　易

(方針)
　確率変数 X, Y が無相関 $\Leftrightarrow E[XY]=E[X]E[Y]$
　確率変数 X, Y が互いに独立 $\Leftrightarrow P[XY]=P[X]P[Y]$

1. 　$A_1=1, A_2=0, B_1=1, B_2=0$ とダミー変数を対応させる。$E[AB]=a,\ E[A]E[B]=bc$ より、

$$\begin{cases} a=bc \\ a+\frac{1}{9}=c \\ a+\frac{4}{9}=b \end{cases}$$

$$\begin{aligned} \Rightarrow\quad & a=(a+\frac{4}{9})(a+\frac{1}{9}) \\ & (a-\frac{2}{9})^2=0 \qquad \therefore a=\frac{2}{9}\ b=\frac{2}{3}\ c=\frac{1}{3} \end{aligned}$$

$(d=(1-c)(1-b)$ で d の値も求められる。)

2.

$$\begin{cases} a+\frac{4}{9}=b \\ a+\frac{1}{9}=c \\ \frac{1}{9}+d=1-b \\ \frac{4}{9}+d=1-c \end{cases}$$

これを連立して解くと、$a=\frac{4}{9}-d,\ b=\frac{8}{9}-d\ c=\frac{5}{9}-d$

1.8　標準

(方針)

$$P[\cup_{k=1}^{\infty} A_k^c] \leq \sum_{k=1}^{\infty} P[A_k^c]$$

を使う。

(答案)

$$\begin{aligned} P[\cap_{k=1}^{\infty} A_k] &= 1 - P[\cup_{k=1}^{\infty} A_k^c] \\ &\geq 1 - \sum_{k=1}^{\infty} P[A_k^c] \end{aligned}$$

2　確率分布と期待値

2.1　標準

1.

$$\begin{aligned} \int_0^2 Cx^3 dx &= [\frac{C}{4}x^4]_0^2 \\ &= 4C = 1 \qquad \therefore C = \frac{1}{4} \\ F(x) &= \int_0^x \frac{1}{4}t^3 dt \\ &= [\frac{1}{16}t^4]_0^x = \frac{1}{16}x^4 \end{aligned}$$

2.

$$\begin{aligned} \int_{-\infty}^{\infty} Ce^{-|x|} dx &= \int_0^{\infty} Ce^{-x} dx + \int_{-\infty}^0 Ce^x dx \\ &= [-Ce^{-x}]_0^{\infty} + [Ce^x]_{-\infty}^0 \\ &= 2C = 1 \qquad \therefore C = \frac{1}{2} \\ F(x) &= \int_{-\infty}^x \frac{1}{2}e^{-|t|} dt \qquad (1) \end{aligned}$$

(i)$0 < x$ の時、

$$\begin{aligned}(1) &= \int_0^x \frac{1}{2}e^{-t}dt + \int_{-\infty}^0 \frac{1}{2}e^{t}dt \\ &= [-\frac{1}{2}e^{-t}]_0^x + [\frac{1}{2}e^{t}]_{-\infty}^0 \\ &= -\frac{1}{2}e^{-x} + 1\end{aligned}$$

(ii)$x < 0$ の時、

$$\begin{aligned}(1) &= \int_{-\infty}^x \frac{1}{2}e^{t}dt \\ &= [\frac{1}{2}e^{t}]_{-\infty}^x \\ &= \frac{1}{2}e^{x}\end{aligned}$$

(i),(ii) より、

$$F(x) = \begin{cases} -\frac{1}{2}e^{-x} + 1 & 0 < x \\ \frac{1}{2}e^{x} & x < 0 \end{cases}$$

3.

$$\begin{aligned}\int_0^\infty Ce^{-2x}dx &= [-\frac{C}{2}e^{-2x}]_0^\infty \\ &= \frac{C}{2} = 1 \qquad \therefore C = 2 \\ F(x) &= \int_0^x 2e^{-2t}dt \\ &= [-e^{-2t}]_0^x \\ &= -e^{-2x} + 1\end{aligned}$$

4.

$$\int_0^\infty Ce^{-x}e^{-e^{-x}}dx \qquad (1)$$

$u = e^{-x}$ と置くと、$du = -e^{-x}dx$ より、$dx = -\frac{1}{u}du$ で、$x : 0 \to \infty$ の時、$u : 1 \to 0$ なので、

$$\begin{aligned}
(1) &= \int_1^0 -Ce^{-u}du \\
&= [Ce^{-u}]_1^0 \\
&= C - Ce^{-1} = 1 \qquad \therefore C = \frac{1}{1-e^{-1}} \\
F(x) &= \int_0^x \frac{1}{1-e^{-1}} e^{-t} e^{-e^t} dt \qquad (2)
\end{aligned}$$

$u = e^{-t}$ と置くと、$dt = -\frac{1}{u}$ で、$t : 0 \to x$ の時、$u : 1 \to e^{-x}$ なので、

$$\begin{aligned}
(2) &= \int_1^{e^{-x}} -\frac{1}{1-e^{-1}} e^{-u} du \\
&= [\frac{1}{1-e^{-1}} e^{-u}]_1^{e^{-x}} \\
&= \frac{e^{-e^{-x}} - e^{-1}}{1-e^{-1}}
\end{aligned}$$

2.2 易

1.

$$\lim_{x\to\infty} \frac{1}{1+e^{-x}} = 1, \quad \lim_{x\to-\infty} \frac{1}{1+e^{-x}} = 0$$

より、F は分布関数。

$$\begin{aligned}
f(x) &= \frac{d}{dx} \frac{1}{1+e^{-x}} \\
&= \frac{e^{-x}}{(1+e^{-x})^2}
\end{aligned}$$

2.

$$\lim_{x\to\infty} 1 - \frac{1}{x^2} = 1, \quad \lim_{x\to 1} 1 - \frac{1}{x^2} = 0$$

より、F は分布関数。

$$\begin{aligned}
f(x) &= \frac{d}{dx}(1 - \frac{1}{x^2}) \\
&= \frac{2}{x^3}
\end{aligned}$$

3.

$$
\begin{aligned}
\lim_{x\to\infty}\frac{\log x}{1+\log x} &= \lim_{x\to\infty}\frac{1/x}{1+1/x} \\
&= 1 \\
\lim_{x\to 1}\frac{\log x}{1+\log x} &= 0
\end{aligned}
$$

より、F は分布関数。

$$
\begin{aligned}
f(x) &= \frac{d}{dx}\frac{\log x}{1+\log x} \\
&= \frac{\frac{1}{2}(1+\log x)-\frac{1}{x}\log x}{(1+\log x)^2} \\
&= \frac{1}{x(1+\log x)^2}
\end{aligned}
$$

4.

$$
\lim_{x\to\infty} 1-e^{-\frac{x^2}{2}} = 1, \quad \lim_{x\to 0} 1-e^{-\frac{x^2}{2}} = 0
$$

より、F は分布関数。

$$
\begin{aligned}
f(x) &= \frac{d}{dx}(1-e^{-\frac{x^2}{2}}) \\
&= xe^{-\frac{x^2}{2}}
\end{aligned}
$$

2.3 標準

$$
g(x) = \begin{cases} \frac{f(x)}{1-F(a)} & if\ a \leq x \\ \\ 0 & if\ x < a \end{cases}
$$

の時、g の台において非負性は成り立つ。全確率 1 について、

$$
\begin{aligned}
\int_{-\infty}^{\infty}\frac{f(x)}{1-F(a)}dx &= \frac{1}{1-F(a)}\{\int_{-\infty}^{\infty} f(x) - \int_{-\infty}^{a} f(x)dx\} \\
&= \frac{1}{1-F(a)}(1-F(a)) \\
&= 1
\end{aligned}
$$

ゆえ、$g(x)$ は確率密度関数である。

$f(x) = e^{-x},\ 0 < x$ とすると、

$$\begin{aligned} F(x) &= \int_0^x e^{-t} dt \\ &= [-e^{-t}]_0^x \\ &= -e^{-x} + 1 \end{aligned}$$

なので、$a = 1$ で打ち切った分布は、

$$\begin{aligned} g(x) &= \frac{e^{-x}}{1 - (e^{-1} + 1)} \\ &= e^{-(x-1)} \end{aligned}$$

2.4 標準

(方針)
前半は平方完成、後半は微分して 0 と置く。定積分の微分について、以下であることを用いる。

$$\frac{d}{dx} \int_a^x f(t)dt = f(x)$$

(答案)

$$\begin{aligned} E[(X-t)^2] &= E[X^2] - 2tE[X] + t^2 \\ &= (t - E[X])^2 + E[X^2] - (E[X])^2 \end{aligned}$$

より、$E[(X-t)^2]$ を最小化する定数 t は、$t = E[X]$

$$\begin{aligned} E[|X-t|] &= \int_{-\infty}^{\infty} |x-t| f(x)dx \\ &= \int_t^{\infty} (x-t)f(x)dx - \int_{-\infty}^t (x-t)f(x)dx \\ &= \int_t^{\infty} xf(x)dx - t\int_t^{\infty} f(x)dx - \int_{-\infty}^t xf(x) + t\int_{-\infty}^t f(x)dx \\ &= (\int_t^{\infty} xf(x) - \int_{-\infty}^t xf(x)dx) - t(\int_t^{\infty} f(x)dx - \int_{-\infty}^t f(x)dx) \\ &= (E[X] - 2\int_{-\infty}^t xf(x)dx) - t(1 - 2\int_{-\infty}^t f(X)dx) \end{aligned}$$

と変形できる。これを t について微分して 0 と置く。

$$\begin{aligned} \frac{\partial E[|X-t|]}{\partial t} &= -2tf(t) - (1 - 2\int_{-\infty}^{t} f(x)dx) + 2tf(t) = 0 \\ \Rightarrow \int_{-\infty}^{t} f(x)dx &= \frac{1}{2} \end{aligned}$$

ゆえに、$E[|X-t|]$ を最小化する t は $P[X \leq \frac{1}{2}] = t$ となるような点、すなわち X の分布の中央値である。

2.5 難

(方針)
「k 次モーメントが存在」とは、言い換えれば $E[|X|^k] < \infty$ である。解法は思いつき難いが結果は重要。

(答案)
　$0 < h < k$ の時、任意の x に対して、

$$|x|^h \leq 1 + |x|^k$$

($\because$ $|x| \leq 1$ ならば $|x|^h \leq 1$ であるし、$1 < |x|$ ならば $|x|^h \leq |x|^k$ である) が成り立つので、

$$\begin{aligned} E[|X|^h] &\leq E[1 + |X|^k] \\ &= 1 + E[|X|^k] < \infty \end{aligned}$$

(別解の方針)
　凸関数について。別解は問題文の制約に加えて、「$E[X]$ が存在する」という条件が必要になるので、今回の問題の解答としては不適だと思う。

Def.(凸関数)

　区間 I において関数 $\phi(x)$ が凸であるとは、全ての $x, y \in I$ と $\gamma \in (0,1)$ に対し、

$$\phi(\gamma x + (1-\gamma)y) \leq \gamma\phi(x) + (1 - \gamma(y))$$

が成り立つ、ということである。

※ ϕ 上の点 $(x, \phi(x))$ は、ϕ 上のどの点における接戦よりも上にある。($y = x^2$

を考えてみればわかる。) つまり、任意の z について、$(z, \phi(z))$ における ϕ の接線は、

$$y = \phi'(z)(x - z) + \phi(z)$$

であって、全ての x について、

$$\phi(x) \geq \phi'(z)(x - z) + \phi(z)$$

が成り立つ。このことは、後の Jensen の不等式の証明で用いる。

凸関数の判定方法

$\phi(x)$ が凸であることと、定義域内の全ての x について、

$$\phi''(x) \geq 0$$

が成り立つことは同値である。
例えば、$\phi(x) = -\log(x)$ について、$\phi''(x) = 1/x^2 > 0,\ \forall x > 0$ より、ϕ は凸。

Jensen の不等式

確率変数 X について、X は 1 次モーメントが存在するとする。ϕ を確率空間上で凸な関数とすると、

$$\phi(E[X]) \leq E[\phi(X)]$$

が成り立つ。(等号成立条件は X が定数か、ϕ が直線であること。)

(証明)
$\mu = E[X]$ と置く。$(\mu, \phi(\mu))$ における ϕ の接線は、

$$y = \phi'(x)(x - \mu) + \phi(\mu)$$

である。この時、全ての x で、

$$\begin{aligned} \phi(x) &\geq \phi'(\mu)(X - \mu) + \phi(\mu) \\ \therefore E[\phi(X)] &\geq \phi'(\mu)E[X - \mu] + \phi(\mu) = \phi(\mu) = \phi(E[X]) \end{aligned}$$

(答案)

$\phi(x) = x^k$ と置く。これは、$\phi''(x) = k(k-1)x^{(k-1)(k-2)} > 0, \quad \forall x \quad (\because (k-1)(k-2)$ は任意の自然数 k について偶数) が成り立つので凸関数である。Jensen の不等式により、

$$(E[X^h])^{k/h} \leq E[X^{h \times k/h}] = E[X^k]$$

が成り立つ。今、$E[X^k] < \infty$ より、$E[X^h] < \infty$ である。

2.6 難

(方針)

二重和についての公式が必要になるので、シグマの公式や二重和の意味について確認しておく。

シグマ公式 (平行移動)

$$\sum_{i=1}^{n} a_i = \sum_{i=k+1}^{n+k} a_{i-k}$$

上端、下端を同方向にずらして添字を逆方向にずらす。

(二項分布の期待値を導出する際にも使う基本的な変形)

二重和の意味

1.

$$\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij} = \sum_{i=1}^{m}(a_{i1} + ... + a_{in})$$

これは、全ての要素 a_{ij}, $\forall i, j$ の和。上端が自明であれば、$\sum_{i,j} a_{ij}$ と書く。

2.

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{i} a_{ij} &= \sum_{i=1}^{n}(a_{i1} + ... + a_{ii}) \\ &= (a_{11}) + (a_{21} + a_{22}) + (a_{31} + a_{32} + a_{33}) + ... \\ &\quad +(a_{n1} + ... + a_{nn})\end{aligned}$$

これは、$i \geq j$ なる全ての要素 a_{ij}, $\forall i \geq j$ の和。上端が自明であれば、$\sum_{i\geq j} a_{ij}$ と書く。

二重和の計算

1.

$$\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}=\sum_{j=1}^{n}\sum_{i=1}^{m}a_{ij}$$

a_{ij} の総和なので可換。

2.

$$\sum_{i=1}^{\infty}\sum_{j=1}^{i}a_{ij}=\sum_{j=1}^{\infty}\sum_{i=j}^{\infty}a_{ij}$$

$i \geq j$ なる全ての a_{ij}(ただし、要素は無限) の総和なのだから、上のように可換。問題ではこの入れ替え方法を用いる。

3.

$$\sum_{i=1}^{m}\sum_{j=1}^{n}a_ib_j=\sum_{i=1}^{m}a_i\sum_{j=1}^{n}b_j$$

(答案)

$$\begin{aligned}
E[X] &= \sum_{k=0}^{k} kP[X=k] \\
&= \sum_{k=0}^{\infty}(\sum_{i=0}^{k-1}1)P[X=k] \\
&\quad (\because k = 1 + ... + 1 = \sum_{i=0}^{k-1}1) \\
&= \sum_{k=0}^{\infty}\sum_{i=0}^{k-1}P[X=k] \\
&= \sum_{i=0}^{\infty}\sum_{k=i+1}^{\infty}P[X=k] \qquad (1)
\end{aligned}$$

これは、$0 \le i$ なる全ての整数 i について、$i+1 \le k$ なる整数 k を取り、$P[X=k]$ の和を取る操作である。つまり、

$$\begin{aligned}(1) &= \sum_{i=0}^{\infty}\{P[X=i+1]+...\} \\ &= \sum_{i=0}^{\infty}(1-P[X\le i]) \\ &= \sum_{k=0}^{\infty}(1-F(k))\end{aligned}$$

2.7 難

1. (方針)

$$x = \begin{cases} \int_0^x dt & if\ 0<x \\ \\ -\int_x^0 dt & if\ x<0 \end{cases}$$

を用いる。

(答案)

$$\begin{aligned}E[X] &= \int_0^{\infty} xf(x)dx + \int_{-\infty}^{0} xf(x)dx \\ &= \int_0^{\infty}\{\int_0^x dtf(x)\}dx + \int_{-\infty}^{0}\{-\int_x^0 dtf(x)\}dx \\ &= \int_0^{\infty} dx \int_0^x dtf(x) - \int_{-\infty}^{0} dx \int_x^0 dtf(x) \\ &= \int_0^{\infty} dt \int_t^{\infty} dxf(x) - \int_{-\infty}^{0} dt \int_{-\infty}^{t} dxf(x) \\ &= \int_0^{\infty}\{\int_t^{\infty} f(x)dx\}dt - \int_{-\infty}^{0}\{\int_{-\infty}^{t} f(x)dx\}dt \\ &= \int_0^{\infty}\{1-F(x)\}dx - \int_{-\infty}^{0} F(x)dx\end{aligned}$$

2. (方針)

$$\begin{aligned} x = F^{-1}(t) &\Leftrightarrow t = F(x) \\ &\Leftrightarrow t = P[X \leq x] \\ \therefore \frac{dt}{dx} = f(x) &\Rightarrow dt = f(x)dx \end{aligned}$$

であることに注意する。

(答案)
$x = F^{-1}(t)$ とおくと、$dt = f(x)dx$ で、$t : 0 \to 1$ の時、$x : -\infty \to \infty$ なので、

$$\int_0^1 F^{-1}(t)dt = \int_{-\infty}^{\infty} xf(x)dx = E[X]$$

2.8 易

$y = x - \mu$ と置くと、$f(y) = f(-y)$ より、$E[Y] = 0$ なので、$E[X] = \mu$

2.9 易

Y の上側 α 分位点を y_α と置くと、

$$\begin{aligned} P[Y \leq y_\alpha] &= P[\sigma X + \mu \leq y_\alpha] \\ &= P[X \leq \frac{y_\alpha - \mu}{\sigma}] = 1 - \alpha \\ \therefore \frac{y_\alpha - \mu}{\sigma} = x_\alpha &\Rightarrow \sigma x_\alpha + \mu \end{aligned}$$

2.10 標準

(方針)
積率母関数の k 階微分に 0 を代入したものは k 次モーメントである。

1.

$$\begin{aligned} M_X(t) &= \int_0^\infty e^{x(t-1)}dx \\ &= [\frac{1}{t-1}e^{x(t-1)}]_0^\infty \\ &= \frac{1}{1-t} \end{aligned}$$

2.

$$
\begin{aligned}
\frac{d}{dx}M_X(t) &= \frac{1}{(1-t)^2} \\
\frac{d^2}{dx^2}M_X(t) &= \frac{2(1-t)}{(1-t)^4} = \frac{2}{(1-t)^3} \\
\frac{d^3}{dx^3}M_X(t) &= \frac{3(1-t)^2 \times 2}{(1-t)^4} = \frac{3!}{(1-t)^4} \\
\frac{d^4}{dx^4}M_X(t) &= \frac{4(1-t)^2 \times 3!}{(1-t)^8} = \frac{4!}{(1-t)^5} \\
&\vdots \\
\frac{d^k}{dx^k}M_X(t) &= \frac{k(1-t)^{k-1}(k-1)!}{(1-t)^{2(k-1)}} = \frac{k!}{(1-t)^{k+1}} \\
\therefore E[X^k] &= \frac{d^k}{dx^k}M_X(0) = k!
\end{aligned}
$$

3.

$$
\begin{aligned}
F_Y(y) &= P[Y \leq y] \\
&= P[\sigma X + \mu \leq y] \\
&= P[X \leq \frac{y-\mu}{\sigma}] \\
&= \int_0^{\frac{y-\mu}{\sigma}} e^{-x}dx \\
&= [-e^{-x}]_0^{\frac{y-\mu}{\sigma}} \\
&= -e^{-\frac{y-\mu}{\sigma}} + 1 \quad (\mu < y) \\
f_Y(y) &= \frac{d}{dy}F_Y(y) \\
&= \frac{1}{\sigma}e^{-\frac{y-\mu}{\sigma}} \quad (\mu < y)
\end{aligned}
$$

2.11 標準

(方針)
確率母関数の k 階微分に 1 を代入したものは k 次階乗モーメントである。

(答案)

$$\sum_{k=0}^{\infty} f(k) = \frac{1/2}{1-1/2} = 1$$

より、f は確率密度関数である。今、確率変数 X の密度関数が f であるとすると、

$$\begin{aligned} G_X(s) &= \sum_{x=0}^{\infty} \frac{1}{2}(\frac{s}{2})^x \\ &= \frac{1/2}{1-s/2} \\ &= \frac{1}{2-s} \end{aligned}$$

である。確率変数 X の k 次階乗モーメントは、確率母関数の k 階微分を計算することで得られる。

$$\begin{aligned} \frac{d}{ds}G_X(s) &= \frac{1}{(2-s)^2} \\ \frac{d^2}{ds^2}G_X(s) &= \frac{2}{(2-s)^3} \\ \frac{d^3}{ds^3}G_X(s) &= \frac{3!}{(2-s)^4} \\ \frac{d^4}{ds^4}G_X(s) &= \frac{4!}{(2-s)^5} \\ &\vdots \\ \frac{d^k}{ds^k}G_X(s) &= \frac{k!}{(2-s)^{k+1}} \\ E[X(X-1)...(X-k+1)] &= \frac{d^k}{ds^k}G_X(1) = k! \end{aligned}$$

2.12 標準

1. (方針)

   平均と分散は一応積率母関数から求めた。極限の計算にはロピタルの定理を用いている。

(答案)

$$\begin{aligned} M_X(t) &= \int_0^1 e^{tx}dx \\ &= [\frac{1}{t}e^{tx}]_0^1 \\ &= \frac{1}{t}(e^t-1) \end{aligned}$$

より、モーメント母関数を一階微分することで X の平均を求める。

$$\begin{aligned} \frac{d}{dx}M_X(t) &= -\frac{1}{t^2}(e^t-1)+\frac{1}{t}e^t \\ &= \frac{-e^t+1+te^t}{t^2} \\ \therefore E[X] &= \lim_{t\to 0}\frac{d}{dx}M_X(t) \\ &= \lim_{t\to 0}\frac{te^t}{2t} \\ &= \lim_{t\to 0}\frac{e^t(t+1)}{2} \\ &= \frac{1}{2} \end{aligned}$$

次に、モーメント母関数の一、二階微分から X の分散を求める。

$$\begin{aligned} \frac{d^2}{dx^2}M_X(t) &= \frac{(-e^t+e^t+te^t)t^2-2t(-e^t+1+te^t)}{t^4} \\ &= \frac{t^2e^t+2e^t-2-2te^t}{t^3} \\ \therefore E[X^2] &= \lim_{t\to 0}\frac{d^2}{dx^2}M_X(t) \\ &= \lim_{t\to 0}\frac{t^2e^t}{3t^2} \\ &= \lim_{t\to 0}\frac{2te^t+t^2e^t}{6t} \\ &= \lim_{t\to 0}\frac{2e^t+4te^t+t^2e^t}{6} \\ &= \frac{1}{3} \\ \therefore Var(X) &= \frac{1}{3}-\frac{1}{4}=\frac{1}{12} \end{aligned}$$

2.

$$
\begin{aligned}
F_Y(y) &= P[Y \le y] \\
&= P[X^2 \le y] \\
&= P[X \le \sqrt{y}] \\
&= \int_0^{\sqrt{y}} f(x)dx \\
&= [x]_0^{\sqrt{y}} \\
&= \sqrt{y} \\
\therefore f_Y(y) &= \frac{1}{2\sqrt{y}} \\
E[Y] &= \int_0^1 \frac{y}{2\sqrt{y}}dy \\
&= [\frac{1}{2}\frac{2}{3}y^{3/2}]_0^1 \\
E[Y^2] &= \int_0^1 \frac{y^2}{2\sqrt{y}}dy \\
&= [\frac{1}{2}\frac{2}{5}y^{5/2}]_0^1 \\
&= \frac{1}{5} \\
\therefore Var(Y) &= \frac{1}{5} - \frac{1}{9} = \frac{4}{45}
\end{aligned}
$$

3. ($y = e^{-x}$ なのだから Y は指数分布に従う)

$$
\begin{aligned}
P[Y \le y] &= P[-\log X \le y] \\
&= P[\log X \ge -y] \\
&= P[X \ge e^{-y}] \\
&= \int_{e^{-y}}^1 f(x)dx \\
&= [x]_{e^{-y}}^1 \\
&= 1 - e^{-y} \\
\therefore f_Y(y) &= e^{-y}
\end{aligned}
$$

$$
\begin{aligned}
E[Y] &= \int_0^\infty y \exp(-y) dy \\
&= \int_0^\infty y\{-\exp(-y)\}' dy \\
&= [-y \exp(-y)]_0^\infty - \int_0^\infty -\exp(-y) dy \\
&= -[\exp(-y)]_0^\infty \\
&= 1 \\
E[Y^2] &= \int_0^\infty y^2 \exp(-y) dy \\
&= \int_0^\infty y^2\{-\exp(-y)\}' dy \\
&= [-y^2 \exp(-y)]_0^\infty - \int_0^\infty -2y \exp(-y) dy \\
&= \int_0^\infty 2y\{-\exp(-y)\}' dy \\
&= [-2y \exp(-y)]_0^\infty - \int_0^\infty -2 \exp(-y) dy \\
&= -[2 \exp(-y)]_0^\infty \\
&= 2 \\
\therefore Var(Y) &= 2 - 1 = 1
\end{aligned}
$$

4.

$$
Y = \sigma X + \mu \Leftrightarrow X = \frac{Y - \mu}{\sigma}
$$

なる一対一変換を考える。$\frac{dx}{dy} = 1/\sigma$ より、

$$
\begin{aligned}
f_Y(y) &= f(\frac{y-\mu}{\sigma})\frac{1}{\sigma} \\
&= \frac{1}{\sigma} \quad (\mu < y < \sigma + \mu) \\
N_Y(t) &= \int_{\mu}^{\sigma+\mu} \frac{1}{\sigma} e^{ty} dy \\
&= [\frac{\sigma t^{ty}}{e}]_{\mu}^{\sigma+\mu} \\
&= \frac{1}{\sigma t}(e^{t(\sigma+\mu)} - e^{t\mu}) \\
E[Y] &= \int_{\mu}^{\sigma+\mu} \frac{y}{\sigma} dy \\
&= [\frac{1}{2\sigma} y^2]_{\mu}^{\sigma+\mu} \\
&= \frac{\sigma}{2} + \mu \\
E[Y^2] &= \int_{\mu}^{\sigma+\mu} \frac{y^2}{\sigma} dy \\
&= [\frac{1}{3\sigma} y^3]_{\mu}^{\sigma+\mu} \\
\therefore Var(Y) &= \frac{\sigma^2}{12}
\end{aligned}
$$

2.13 標準

(方針)

近似方法を確認しておく。テイラーの定理は後にデルタ法などでも使う定理である。

テイラーの定理、有限テイラー展開

$f(x)$ が開区間 I で n 階微分可能だとする。I の任意の 2 点 a, b に対し、

$$f(b) = f(a) + f'(a)(b-a) + ... + \frac{f^{(n-1)}(a)}{(n-1)!}(b-a)^{n-1} + \frac{f^{(n)}(c)}{n!}(b-a)^n$$

を満たす点 c が a, b 間に存在する。(「$a = b$ 周りでテイラーの定理を適用する」という)

$f(x)$ が開区間 I について n 階微分可能とする。I 上の点 a を固定すると、各 $x \in I$ に対して、

$$f(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(a)}{k!}(x-a)^k + \frac{f^{(n)}(a+\theta(x-a))}{n!}(x-a)^n$$

を満たす $\theta (0 < \theta < 1)$ が存在する。これを有限テイラー展開と呼ぶ ($a = 0$ の時、有限マクローリン展開)

ビッグオー 『$f(x)$ の発散スピードは $g(x)$ の定数倍よりも遅い』

実直線上に定義された関数 $f(x), g(x)$ について、「$x > x_0$ ならば $|f(x)| \leq C|g(x)|$」となるような定数 x_0, C が存在するとき、

$$f(x) = O(g(x))$$

と書く。

(Ex)

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

である。つまり、x が大であるとき、$2x^3 < Cx^3$ が常に成り立つような定数 C は存在する。他にも、

$$52x^6 + x^5 + 1000x + 4 = O(x^6)$$
$$e^x = 1 + x + \frac{x^2}{2} + O(x^3)$$

が言える。

スモールオー『$x \to 0$ の時 $f(x)$ の 0 への収束スピードは $g(x)$ より速い』

$x=0$ 近傍に定義された関数 $f(x), g(x)$ に対して、

$$\lim_{x \to 0} \frac{f(x)}{g(x)} = 0$$

が成り立つ時、$f(x) = o(g(x))$ と書く。

(Ex)

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

が言える。実際、定義の極限式を満たす。

スモールオーの計算

$$\begin{aligned}
&x^{n+1} = o(x^n) \\
&co(x^n) = o(x^n) \\
&x^n o(x^m) = o(x^{n+m}) \\
&o(x^n)o(x^m) = o(x^{n+m}) \\
&o(x^n) + o(x^m) = o(x^n) \quad if\ n \leq m
\end{aligned}$$

漸近展開

$$\begin{aligned}
f(x) &= f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + ... + \frac{f^{(n)}(0)}{n!}x^n + o(x^n) \\
&= \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!}x^k + o(x^n)
\end{aligned}$$

(Ex)

$$\begin{aligned}
\exp(x) &= 1 + x + \frac{x^2}{2} + o(x^2) \\
\log(1+x) &= x - \frac{x^2}{2} + o(x^2) \\
\cos x &= 1 - \frac{x^2}{2} + o(x^2) \\
\sin x &= x + o(x^2)
\end{aligned}$$

1.

$$M_X(t) = \int_{-1}^{1} \frac{1}{2} e^{tx} dx = \frac{1}{2t}(e^t - e^{-t})$$

である。ここで、e^t, e^{-t} の無限マクローリン展開を考えることにより、

$$\frac{1}{2t}(e^t - e^{-t}) = \frac{1}{2t}\{\sum_{k=0}^{\infty} \frac{t^k}{k!} - \sum_{k=0}^{\infty} \frac{(-t)^k}{k!}\} \quad (1)$$

$$\begin{aligned}
\sum_{k=0}^{\infty} \frac{t^k}{k!} &= 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \frac{t^4}{4!} + \frac{t^5}{5!} + \ldots & (2) \\
-\sum_{k=0}^{\infty} \frac{(-t)^k}{k!} &= -1 + t - \frac{t^2}{2!} + \frac{t^3}{3!} - \frac{t^4}{4!} + \frac{t^5}{5!} + \ldots & (3) \\
\therefore (2) + (3) &= 2(t + \frac{t^3}{3!} + \frac{t^5}{5!} + \ldots) \\
\therefore (1) &= \frac{1}{2t} 2(t + \frac{t^3}{3!} + \frac{t^5}{5!} + \ldots) \\
&= 1 + \frac{t^2}{3!} + \frac{t^4}{5!} + o(t^4)
\end{aligned}$$

最後に漸近展開を行った。以上より、

$$\begin{aligned}
\frac{d}{dt} M_X(t) &= \frac{t}{3} + o(t) \\
\therefore E[X] &= 0 \\
\frac{d^2}{dt^2} M_X(t) &= \frac{1}{3} + o(1) \\
\therefore Var(X) &= \frac{1}{3}
\end{aligned}$$

(別解)

$e^t - e^{-t}$ のところで漸近展開して、

$$\begin{aligned}
e^t &= 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \frac{t^4}{4!} + \frac{t^5}{5!} + o(t^5) \\
-e^{-t} &= -1 + t - \frac{t^2}{2!} + \frac{t^3}{3!} - \frac{t^4}{4!} + \frac{t^5}{5!} + o(t^5) \\
\therefore e^t - e^{-t} &= 2(t + \frac{t^3}{3!} + \frac{t^5}{5!}) + o(t^5)
\end{aligned}$$

としたほうがやりやすいかも。

2.

$$\begin{aligned}
P[Y \leq y] &= P[X^2 \leq y] \\
&= P[-\sqrt{y} \leq X \leq \sqrt{y}] \\
&= \int_{-1}^{\sqrt{y}} f(x)dx - \int_{-1}^{-\sqrt{y}} f(x)dx \\
\therefore f_Y(y) &= f(\sqrt{y})\frac{1}{2\sqrt{y}} - f(-\sqrt{y})(-\frac{1}{2\sqrt{y}}) \\
&= f(\sqrt{y})\frac{1}{\sqrt{y}} \\
&= \frac{1}{2\sqrt{y}} \qquad (0 < y < 1) \\
E[Y] &= \int_0^1 \frac{y}{2\sqrt{y}}dy \\
&= [\frac{1}{2}\frac{2}{3}y^{3/2}]_0^1 \\
&= \frac{1}{3} \\
E[Y^2] &= \int_0^1 \frac{y^2}{2\sqrt{y}}dy \\
&= [\frac{1}{2}\frac{2}{5}y^{5/2}]_0^1 \\
&= \frac{1}{5} \\
\therefore Var(Y) &= \frac{1}{5} - \frac{1}{9}
\end{aligned}$$

3.

$$
\begin{aligned}
P[Y \leq y] &= P[-\log|X| \leq y] \\
&= P[\log|X| \geq -y] \\
&= P[|X| \geq e^{-y}] \\
&= 1 - P[-e^{-y} \leq X \leq e^{-y}] \\
&= 1 - \int_{-1}^{e^{-y}} f(x)dx + \int_{-1}^{-e^{-y}} f(x)dx \\
\therefore f_Y(y) &= -f(e^{-y}) \times (-e^{-y}) + f(-e^{-y}) \times e^{-y} \\
&= -\frac{1}{2} \times (-e^{-y}) + \frac{1}{2} \times e^{-y} \\
&= e^{-y} \qquad (0 < y < \infty)
\end{aligned}
$$

ゆえに、$Y \sim Exp(1)$ なので、$E[Y] = 1, Var(Y) = 1$

2.14 標準

$$
\begin{aligned}
F_Y(y) &= P[Y \leq y] \\
&= P[-\sqrt{y} \leq X \leq \sqrt{y}] \\
&= \int_{-1}^{\sqrt{y}} f(x)dx - \int_{-1}^{-\sqrt{y}} f(x)dx \\
\therefore f_Y(y) &= f(\sqrt{y})\frac{1}{2\sqrt{y}} - f(-\sqrt{y})(-\frac{1}{2\sqrt{y}}) \\
&= \frac{1+\sqrt{y}}{2\sqrt{y}} + \frac{1-\sqrt{y}}{2\sqrt{y}} \\
&= \frac{1}{2\sqrt{y}} \quad 0 < Y < 1 \\
E[Y] &= \int_0^1 \frac{y}{2\sqrt{y}}dy \\
&= [\frac{1}{2}\frac{2}{3}y^{3/2}]_0^1 \\
&= \frac{1}{3} \\
E[Y^2] &= \int_0^1 \frac{y^2}{2\sqrt{y}}dy \\
&= [\frac{1}{2}\frac{2}{5}y^{5/2}]_0^1 \\
&= \frac{1}{5} \\
\therefore Var(Y) &= \frac{1}{5} - \frac{1}{9} = \frac{4}{45}
\end{aligned}
$$

2.15 標準

(方針)

$P[Y \leq y]$ が答案のように y の範囲によって二通りに表されることに注意する。

(答案)

$$
\begin{aligned}
P[Y \leq y] &= P[X^2 \leq y] \\
&= \begin{cases} P[-\sqrt{y} \leq X \leq \sqrt{y}] & if\ 0 \leq y \leq 1 \\ P[X \leq \sqrt{y}] & if\ 1 \leq y \leq 4 \end{cases} \\
&= \begin{cases} \int_{-1}^{\sqrt{y}} f_X(x)dx - \int_{-1}^{-\sqrt{y}} f_X(x)dx & if\ 0 \leq y \leq 1 \\ \\ \int_{-1}^{\sqrt{y}} f_X(x)dx & if\ 1 \leq y \leq 4 \end{cases}
\end{aligned}
$$

(i)$0 \leq y \leq 1$ の時、

$$
\begin{aligned}
f_Y(y) &= f_X(\sqrt{y})\frac{1}{2\sqrt{y}} - f_X(-\sqrt{y})(-\frac{1}{2\sqrt{y}}) \\
&= \frac{\sqrt{y}+1}{9\sqrt{y}} + \frac{-\sqrt{y}+1}{9\sqrt{y}} \\
&= \frac{2}{9\sqrt{y}}
\end{aligned}
$$

(ii)$1 \leq y \leq 4$ の時、

$$
\begin{aligned}
f_Y(y) &= f_X(\sqrt{y})\frac{1}{2\sqrt{y}} \\
&= \frac{\sqrt{y}+1}{9\sqrt{y}}
\end{aligned}
$$

2.16　難

1. (方針)

$$
\begin{aligned}
E[X] &= \int_0^\infty \{1 - F(x)\}dx - \int_{-\infty}^0 F(x)dx \\
&= \int_0^\infty P[t \leq X]dt - \int_{-\infty}^0 \{1 - P[t \leq X]\}dt
\end{aligned}
$$

を用いる。

(答案)

$$
\begin{aligned}
E[X]-E[Y] &= \int_0^{\infty} P[t \le X]dt - \int_{-\infty}^{0}\{1-P[t \le X]\}dt \\
&\quad - \int_0^{\infty} P[t \le Y]dt + \int_{-\infty}^{0}\{1-P[t \le Y]\}dt \\
&= \int_0^{\infty}\{P[t \le X]-P[t \le Y]\}dt + \int_{-\infty}^{0}\{P[t \le X]-P[t \le Y]\}dt \\
&= \int_{-\infty}^{\infty}\{P[t \le X]-P[t \le Y]\}dt
\end{aligned}
$$

今、任意の実数 t について $P[t \le X] \ge P[t \le Y]$ が成り立つのだから、$E[X]-E[Y] \ge 0$

2. (方針)

$$
\begin{aligned}
x = F_X^{-1}(t) &\Leftrightarrow t = F_X(x) \\
&\Leftrightarrow t = P[X \le x] \\
y = F_Y^{-1}(t) &\Leftrightarrow t = P[Y \le y]
\end{aligned}
$$

であることに注意する。

(答案)
$x = F_X^{-1}(t),\ y = F_Y^{-1}(t)$ と置く。今、任意の定数 c について $P[X \le c] \le P[Y \le c]$ が成り立つのだから、任意の $t = P[X \le x] = P[Y \le y]$ について、$x \ge y$ と言える。

2.17 難

1.

$$
\begin{aligned}
h(t) &= \log A(t) \\
&= \frac{1}{t}\log E[X^t]
\end{aligned}
$$

と置く。$h(t)$ が t の増加関数であるためには、$h'(t) \ge 0$ であれば良い。

$$
h'(t) = -\frac{1}{t^2}\log E[X^t] + \frac{1}{t}\frac{1}{E[X^t]}E[X^t \log X]
$$

であるから、

$$\begin{aligned} h'(t) \geq 0 \quad &\Leftrightarrow \quad t\frac{E[X^t \log X]}{E[X^t]} \geq \log E[X^t] \\ &\Leftrightarrow \quad E[X^t \log X^t] \geq E[X^t] \log E[X^t] \qquad (1) \end{aligned}$$

である。(X の期待値を取っているので、$tE[X^t \log X] = E[X^t t \log X] = E[X^t \log X^t]$ の操作が可能。)　今、$Y = X^t$ と置くと、(1) は、

$$E[Y \log Y] \geq E[Y] \log E[Y] \qquad (2)$$

と書ける。さて、$g(y) = y \log y$ なる関数を考えると、$g''(y) = 1/y > 0,\ \forall y > 0$ より g は凸関数であるから、Jensen の不等式 (2.5 参照) により、(2) 式は成立する。

2. (方針)
1 を利用する。H,M は $t = -1, 1$ を代入したものなので $A(t)$ は増加関数であることより $H \geq M$ はすぐ分かる。実は G は $t = 0$ を代入したもの。

(答案)

$$\begin{aligned} \lim_{t\to 0} h(t) &= \lim_{t\to 0} \frac{\log E[X^t]}{t} \\ &= \lim_{t\to 0} \frac{E[X^t \log X]}{E[X^t]} \\ &= E[\log X] \end{aligned}$$

なので、

$$\begin{aligned} \lim_{t\to 0} (E[X^t])^{1/t} &= \exp(E[\log X]) \\ &= G \end{aligned}$$

であるから、$A(t)$ は増加関数より、$H \leq G \leq H$ が成り立つ。

2.18 易

$$
\begin{aligned}
M_X(t) &= \int_0^1 (m+1)x^m e^{tx} dx \\
&= \int_0^1 (m+1)x^m \sum_{k=0}^{\infty} \frac{(tx)^k}{k!} dx \\
&= \int_0^1 (m+1) \sum_{k=0}^{\infty} \frac{t^k x^{m+k}}{k!} dx \\
&= [(m+1) \sum_{k=0}^{\infty} \frac{t^k x^{m+k+1}}{(m+k+1)k!}]_0^1 \\
&= \sum_{k=0}^{\infty} \frac{m+1}{(m+k+1)k!} t^k
\end{aligned}
$$

したがって、

$$
\begin{aligned}
\frac{d}{dt} M_X(t) &= \sum_{k=0}^{\infty} \frac{m+1}{(m+k+1)k!} k t^{k-1} \\
&= \frac{m+1}{m+2} + \sum_{k=3}^{\infty} \frac{m+1}{(m+k+1)k!} k t^{k-1} \\
\therefore E[X] &= \frac{m+1}{m+2} \\
\frac{d^2}{dt^2} M_X(t) &= \sum_{k=0}^{\infty} \frac{m+1}{(m+k+1)k!} k(k-1) t^{k-2} \\
&= \frac{m+1}{m+3} + \sum_{k=4}^{\infty} \frac{m+1}{(m+k+1)k!} k(k-1) t^{k-2} \\
\therefore Var(X) &= \frac{m+1}{m+3} - \frac{(m+1)^2}{(m+2)^2} \\
&= \frac{m+1}{(m+2)^2(m+3)}
\end{aligned}
$$

3 代表的な確率分布

主要な確率分布について期待値、分散など導出する。問題と内容が被る部分もあるが基本中の基本なので丁寧に書く。どの分布もパッと導出イメージが湧くように練習しよう！

- 離散一様分布
  有限集合の全ての値について、等しく確からしい場合。すなわち異なる N 通りの状態をとりうる確率変数 X が各状態を取る確率が等しく $\frac{1}{N}$ であるような分布。確率密度関数は、

$$f_X(x) = \frac{1}{N} \quad x = 1, ..., N$$

確率母関数は、

$$\begin{aligned}
G_X(S) &= E[s^X] \\
&= \sum_{x=1}^{N} \frac{s^x}{N} \\
&= \frac{s(1 + s + ... + s^{N-1})}{N} \\
&= \frac{s(1 - s^N)}{N(1 - s)} \\
&= \frac{s(1 - s)(1 + s + ... + s^{N-1})}{N(1 - s)} \\
&= \frac{1}{N} \sum_{k=1}^{N} s^k
\end{aligned}$$

期待値は、

$$\begin{aligned}
E[X] &= \sum_{x=1}^{N} \frac{x}{N} \\
&= \frac{1 + ... + N}{N} \\
&= \frac{N(N + 1)}{2N} \\
&= \frac{N + 1}{2}
\end{aligned}$$

分散は、

$$
\begin{aligned}
Var(X) &= \sum_{x=1}^{N} \frac{(x - \frac{N+1}{2})^2}{N} \\
&= \frac{1}{N} \sum_{x=1}^{N} \{x^2 - (N+1)x + \frac{(N+1)^2}{4}\} \\
&= \frac{1}{N} \{\frac{1}{6} N(N+1)(2N+1) - \frac{N(N+1)^2}{2} + \frac{N(N+1)^2}{4}\} \\
&= \frac{2(N+1)(2N+1) - 3(N+1)^2}{12} \\
&= \frac{(N+1)(N-1)}{12}
\end{aligned}
$$

問 1 は確率母関数から平均、分散を導出させる問題。

- 二項分布
  ベルヌーイ試行を n 回行った時の成功回数に関する確率分布。確率密度関数は、

$$f_X(x) = {}_nC_x p^x (1-p)^{n-x} \quad x = 0, ..., n$$

積率母関数は、

$$
\begin{aligned}
M_X(t) &= E[e^{tX}] \\
&= \sum_{x=0}^{n} {}_nC_x (pe^t)^x (1-p)^{n-x} \\
&= (pe^t + 1 - p)^n
\end{aligned}
$$

最終行では問 2 で証明する二項定理を用いた。積率母関数を漸近展開す

ることで二項分布が正規分布に収束することがわかる (問 8)。期待値は、

$$
\begin{aligned}
E[X] &= \sum_{x=0}^{n} x \times {}_nC_x p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} \frac{n!}{(n-x)!(x-1)!} p^x (1-p)^{n-x} \\
&= np \sum_{x=1}^{n} \frac{(n-1)!}{(n-x)!(x-1)!} p^{x-1} (!-p)^{n-x} \\
&= np \sum_{x'=0}^{n-1} \frac{(n-1)!}{(n-1-x')!x'!} p^{x'} (1-p)^{n-1-x'} \\
& \quad (x' = x-1) \\
&= np
\end{aligned}
$$

分散は、

$$
\begin{aligned}
E[X^2] &= \sum_{x=0}^{n} x^2 \times {}_nC_x p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} x \times \frac{n!}{(n-x)!(x-1)!} p^x (1-p)^{n-x} \\
&= np \sum_{x=1}^{n} x \times \frac{(n-1)!}{(n-x)!(x-1)!} p^{x-1} (1-p)^{n-x} \\
&= np \sum_{x'=0}^{n-1} (x'+1) \times \frac{(n-1)!}{(n-1-x')!x'!} p^{x'} (1-p)^{n-1-x'} \\
&= np\{(n-1)p+1\} \\
\therefore Var(X) &= np(np-p+1) - n^2p^2 \\
&= np(1-p)
\end{aligned}
$$

- ポアソン分布
  単位時間内に平均 λ 回起こる事象が起こる回数を示す確率分布。確率密度関数は、

$$
f_X(x) = \frac{\lambda^x}{x!} \exp(-\lambda) \quad x = 0, ...
$$

積率母関数は、

$$\begin{aligned}
M_X(t) &= E[e^{tX}] \\
&= \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \exp(tx - \lambda) \\
&= \exp(-\lambda) \sum_{x=0}^{\infty} \frac{(e^t \lambda)^x}{x!} \\
&= \exp\{\lambda(e^t - 1)\}
\end{aligned}$$

二項定理の積率母関数において、$\lambda = np$ とし、$n \to \infty$ を考えると、

$$\begin{aligned}
\lim_{n\to\infty} (\frac{\lambda}{n} e^t + 1 - \frac{\lambda}{n})^n &= \lim_{n\to\infty} \{1 + \frac{\lambda(e^t - 1)}{n}\}^n \\
&= \exp\{\lambda(e^t - 1)\}
\end{aligned}$$

より、ポアソン分布は二項分布で $n \to \infty$ とした形であることがわかる。また積率母関数の漸近展開によってポアソン分布が正規分布に収束することが示せる (問 9)。期待値は、

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

分散は、

$$\begin{aligned}
E[X^2] &= \sum_{x=0}^{\infty} x^2 \frac{\lambda^2}{x!} \exp(-\lambda) \\
&= \lambda \sum_{x=1}^{\infty} x \frac{\lambda^{x-1}}{(x-1)!} \exp(-\lambda) \\
&= \lambda \sum_{x'=0}^{\infty} (x'+1) \frac{\lambda^{x'}}{x'!} \exp(-\lambda) \\
&= \lambda(\lambda + 1) \\
\therefore Var(X) &= \lambda(\lambda + 1) - \lambda^2 = \lambda
\end{aligned}$$

- 幾何分布
ベルヌーイ試行を繰り返す時、一回成功するまでに失敗した回数に関する確率変数。確率密度関数は、

$$f_X(x) = p(1-p)^x \quad x = 0, \ldots$$

積率母関数は、

$$\begin{aligned} G_X(s) &= \sum_{x=0}^{\infty} p\{s(1-p)\}^x \\ &= \frac{p}{1-s(1-p)} \end{aligned}$$

等比数列の和の公式を用いた。期待値は、

$$G'_X(s) = \frac{1-s(1-p)+(1-p)p}{\{1-s(1-p)\}^2}$$

より、

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

- 負の二項分布
ベルヌーイ試行を繰り返す時、r 回成功するまでに失敗した回数に関する確率分布。確率密度関数は、

$$f_X(x) = {}_{x+r-1}C_r p^r (1-p)^x \quad x = 0, \ldots$$

積率母関数は、

$$\begin{aligned} M_X(t) &= \sum_{x=0}^{\infty} {}_{x+r-1}C_r p^r \{e^t(1-p)\}^x \\ &= \frac{p^r}{\{1-e^t(1-p)\}^r} \sum_{x=0}^{\infty} {}_{x+r-1}C_r \{1-e^t(1-p)\}^r \{e^t(1-p)\}^x \\ &= \{\frac{p}{1-e^t(1-p)}\}^r \end{aligned}$$

$Y = 2pX$ として積率母関数を $p \to 0$ と漸近展開することで Y はカイ二乗分布に収束する (問 7)。

- 超幾何分布
赤玉 M 個、白玉 $N-M$ 個の箱から n 個取り出す時、赤玉が X 個である、という確率変数。確率密度関数は、

$$f(x) = \frac{{}_MC_x \times {}_{N-M}C_{n-x}}{{}_NC_n} \quad x = 0, ..., n$$

問 3 で期待値、分散を導出している。とても難しい。

- 多項分布
k 個の面からなるサイコロを n 回振る時、各面の出る回数に関する確率変数の多次元分布。確率密度関数は、

$$f_{X_1,...,X_k}(x_1, ..., x_k) = \frac{n!}{x_1!...x_k!} p_1^{x_1}...p_k^{x_k}$$

多項分布は第 4 章問 23 で扱う。

- 連続一様分布
確率変数 X が区間 $[a, b]$ においていずれの値を取る確率も同様に確からしい時、X は一様分布に従うという。確率密度関数は、

$$f_X(x) = \frac{1}{b-a} \qquad a \leq x \leq b$$

期待値、分散は、

$$\begin{aligned} E[X] &= \int_a^b \frac{x}{b-a} dx \\ &= \frac{1}{b-a} [\frac{x^2}{2}]_a^b = \frac{a+b}{2} \\ Var(X) &= \frac{(b-a)^2}{12} \end{aligned}$$

- 正規分布
$X \sim N(\mu, \sigma^2)$ の確率密度関数は、

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

標準正規分布の全確率 1 を証明する。

$$I = \int_{-\infty}^{\infty} \exp(-\frac{x^2}{2}) dx$$

と置くと、

$$I^2 = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \exp(-\frac{x^2+y^2}{2}) \tag{1}$$

$x = r\cos\theta,\ y = r\sin\theta$ と置くと、

$$\frac{\partial(x,y)}{\partial(r,\theta)} = det\begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\sin\theta \end{pmatrix} = r$$

で、$x : -\infty \to \infty,\ y : -\infty \to \infty$ のとき $\theta : 0 \to 2\pi,\ r : 0 \to \infty$ なので、

$$\begin{aligned} (1) &= \int_0^{2\pi} d\theta \int_0^{\infty} dr \exp(-\frac{r^2}{2})r \\ &= \int_0^{2\pi} [-\exp(-\frac{r^2}{2})]_0^{\infty} d\theta \\ &= 2\pi \end{aligned}$$

ゆえ、$I = \sqrt{2\pi}$ となり、全確率 1 が証明された。

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}dx \tag{2}$$

に対しても、$z = \frac{x-\mu}{\sigma}$ と置くと、$dx = \sigma dz$ なので、

$$(2) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{z^2}{2})dz = 1$$

で、一般の正規分布に対して全確率 1 が成り立つ。積率母関数は、

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(x-\mu t-\sigma^2 t)^2}{2\sigma^2} + \mu t + \frac{\sigma^2 t}{2}\}dx \\ &= \exp(\mu t + \frac{\sigma^2 t}{2}) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(x-\mu t-\sigma^2 t)^2}{2\sigma^2}\}dx \\ &= \exp(\mu t + \frac{\sigma^2 t}{2}) \end{aligned}$$

期待値、分散は μ, σ^2。これも全確率 1 と同様に、標準正規分布の期待値、分散が $0, 1$ であることを求めてから変数変換によって求められる。

- ベータ分布
  ベータ分布と次のディリクレ分布はベイズ統計の文脈。ベータ分布は二項分布のパラメータ p の共役事前分布である。確率密度関数は、

$$\pi(p) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1}$$

全確率 1 の証明は次のとおり。ちょっと大変。

$$\Gamma(a)\Gamma(b) = \int_0^1 \Gamma(a+b)p^{a-1}(1-p)^{b-1}dp$$

を示す。左辺について、

$$\begin{aligned}\Gamma(a)\Gamma(b) &= \int_0^\infty x^{a-1}\exp(-x)dx \int_0^\infty y^{b-1}\exp(-y)dy \\ &= \int_0^\infty dx \int_0^\infty dy x^{a-1}y^{b-1}\exp\{-(x+y)\} \qquad (1)\end{aligned}$$

x を固定して $t = x + y$ と置く。$y : 0 \to \infty$ のとき、$t : x \to \infty$ で、$dy = dt$ なので、

$$\begin{aligned}(1) &= \int_0^\infty dx \int_x^\infty dt x^{a-1}(t-x)^{b-1}\exp(-t) \\ &= \int_0^\infty dt \int_0^t dx x^{a-1}(t-x)^{b-1}\exp(-t) \qquad (2)\end{aligned}$$

t を固定して $p = \frac{x}{t}$ と置く。$x : 0 \to t$ のとき $p : 0 \to 1$ で $dp = \frac{1}{t}dx$ なので、

$$\begin{aligned}(2) &= \int_0^\infty dt \int_0^1 dp (tp)^{a-1}\{t(1-p)\}^{b-1}\exp(-t)\frac{1}{t} \\ &= \int_0^1 \{\int_0^\infty t^{a+b-1}\exp(-t)dt\}p^{a-1}(1-p)^{b-1}dp \\ &= \int_0^1 \Gamma(a+b)p^{a-1}(1-p)^{b-1}dp\end{aligned}$$

期待値は、

$$\begin{aligned}E[p] &= \int_0^1 \frac{1}{B(a,b)}p^a(1-p)^{b-1}dp \\ &= \frac{1}{B(a,b)}B(a+1,b)\int_0^1 \frac{1}{B(a+1,b)}p^a(1-p)^{b-1}dp \\ &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} \\ &= \frac{a}{a+b}\end{aligned}$$

- ディィリクレ分布
  ディィリクレ分布は多項分布のパラメータ p_k の共役事前分布である。確率密度関数は、

$$\pi(p_1,...,p_K)=\frac{\Gamma(a_1+...+a_K)}{\Gamma(a_1)...\Gamma(a_K)}\Pi_{k=1}^{K}p_k^{a_k-1}$$

- ガンマ分布
  めっちゃ有名なやつ。確率密度関数は、

$$f_X(x)=\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\exp(-\beta x)$$

  積率母関数は、

$$\begin{aligned}M_X(t) &= \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\exp\{(t-\beta)x\}dx\\ &= \frac{\beta^\alpha}{\Gamma(\alpha)}\frac{\Gamma(\alpha)}{(\beta-t)^\alpha}\int_0^\infty \frac{(\beta-t)^\alpha}{\Gamma(\alpha)}x^{\alpha-1}\exp\{-(\beta-t)x\}dx\\ &= (\frac{\beta}{\beta-t})^\alpha\end{aligned}$$

- 逆ガンマ分布
  X が逆ガンマ分布に従う時、$Y=\frac{1}{X}$ とすれば Y はガンマ分布に従う。すなわち確率密度関数は

$$f_X(x)=\frac{\beta^\alpha}{\Gamma(\alpha)}x^{-\alpha-1}\exp(-\frac{\beta}{x})$$

- カイ二乗分布
  $\chi_m^2=Ga(\frac{m}{2},\frac{1}{2})$ である。カイ二乗分布については 5.1 で死ぬほど詳しく説明した。

- 指数分布
  問 5 参照。

3.1 易

(方針)
離散一様分布の確率密度関数は

$$f_X(x)=\frac{1}{N} \qquad x=1,...,N$$

である。この確率母関数は、

$$
\begin{aligned}
G_X(s) &= E[s^X] \\
&= \sum_{x=1}^{N} \frac{s^x}{N} \\
&= \frac{s(1-s^N)}{N(1-s)}
\end{aligned}
$$

である。ここからの計算には

$$1 - s^N = (1-s)(1+s+...+s^{N-1})$$

を用いる。確率母関数の k 階微分が k 次階乗モーメントになることを用いて平均、分散を求める。

(答案)

$$
\begin{aligned}
G_X(s) &= \sum_{x=1}^{N} \frac{s^x}{N} \\
&= \frac{s(1-s^N)}{N(1-s)} \\
&= \frac{s(1-s)(1+s+...+s^{N-1})}{N(1-s)} \\
&= \frac{1}{N}\sum_{k=1}^{N} s^k \\
\frac{d}{ds}G_X(s) &= \frac{1}{N}\sum_{k=1}^{N} ks^{k-1} \\
\therefore E[X] &= \frac{d}{ds}G_X(1) \\
&= \frac{1}{N}\sum_{k=1}^{N} k \\
&= \frac{1}{N}\frac{N(N+1)}{2} \\
&= \frac{N+1}{2} \\
\frac{d^2}{ds^2}G_X(s) &= \frac{1}{N}\sum_{k=1}^{N} k(k-1)s^{k-2} \\
\therefore E[X(X-1)] &= \frac{d^2}{ds^2}G_X(1) \\
&= \frac{1}{N}\sum_{k=1}^{N} k(k-1) \\
&= \frac{1}{N}\{\frac{1}{6}N(N+1)(2N+1) - \frac{N(N+1)}{2}\} \\
&= \frac{(N+1)(2N+1)}{6} - \frac{N+1}{2} \\
Var(X) &= E[X(X-1)] + E[X] - \{E[X]\}^2 \\
&= \frac{(N+1)(2N+1)}{6} - \frac{N+1}{2} + \frac{N+1}{2} - \frac{(N+1)^2}{4} \\
&= \frac{(N+1)\{2(2N+1)-3(N+1)\}}{12} \\
&= \frac{(N+1)(N-1)}{12}
\end{aligned}
$$

3.2　易

(方針)

二項係数と二項定理について一応書いておく。

二項係数

1.
$$ {}_nC_r = {}_nC_{n-r} $$
n 個から選ばれた r 個の組み合わせの総数は、選ばれなかった $n-r$ 個の組み合わせの総数に等しい。

2.
$$ {}_{n+1}C_r = {}_nC_{r-1} + {}_nC_r $$
A さんを含む $n+1$ 人から r 人選ぶ組み合わせの総数は、「A さんを選ぶ時、A さん以外の n 人から残りの $r-1$ 人を選ぶ組み合わせの総数」と「A さんを選ばない時、A さん以外の n 人から残りの r 人を選ぶ組み合わせの総数」の和に等しい。

3.
$$ r\ {}_nC_r = n\ {}_{n-1}C_{r-1} $$
n 人から r 人選んでその中から代表者をひとり選ぶ組み合わせの総数と、n 人から代表者をひとり選んで残りの $r-1$ 人を選ぶ組み合わせの総数は等しい。

二項定理

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

答案では、二項係数の公式に加えて 2.6 で見た和の公式 (平行移動) も使っている。

(答案)

$$\begin{aligned} {}_nC_r &= \frac{n!}{r!(n-r)!} \\ {}_nC_{n-r} &= \frac{n!}{(n-r)!r!} \end{aligned}$$

より ${}_nC_r = {}_nC_{n-r}$ が成り立つ。また、

$$\begin{aligned} {}_{n-1}C_{r-1} + {}_{n-1}C_r &= \frac{(n-1)!}{(r-1)!(n-r)!} + \frac{(n-1)!}{r!(n-1-r)!} \\ &= \frac{r(n-1)! + (n-r)(n-1)!}{r!(n-r)!} \\ &= \frac{n!}{r!(n-r)!} \\ &= {}_nC_r \end{aligned}$$

が成り立つ。次に、数学的帰納法によって二項定理を証明する。
(i)$n = 1$ のとき、

$$a + b = {}_1C_0 a^0 b^1 + {}_1C_1 a^1 b^0 = a + b$$

より成立。
(ii)$n = m$ の時、

$$(a+b)^m = \sum_{k=0} {}_mC_k a^k b^{m-k}$$

が成り立つと仮定する。両辺に $(a+b)$ を掛けて、

$$\begin{aligned}
(a+b)^{m+1} &= (a+b)\{\sum_{k=0}^{m} {}_mC_k a^k b^{m-k}\} \\
&= \sum_{k=0}^{m} {}_mC_k a^{k+1} b^{m-k} + \sum_{k=0}^{m} {}_mC_k a^k b^{m-k+1} \\
&= a^{m+1} + \sum_{k=0}^{m-1} {}_mC_k a^{k+1} b^{m-k} + b^{m+1} + \sum_{k=1}^{m} {}_mC_k a^k b^{m-k+1} \\
&= a^{m+1} + \sum_{k=1}^{m} {}_mC_{k-1} a^k b^{m-k+1} + b^{m+1} + \sum_{k=1}^{m} {}_mC_k a^k b^{m-k+1} \\
&= a^{m+1} + b^{m+1} + \sum_{k=1}^{m} {}_{m+1}C_k a^k b^{m-k+1} \\
&= \sum_{k=0}^{m+1} {}_{m+1}C_k a^k b^{m-k+1}
\end{aligned}$$

より成立。

3.3　難

(方針？)
　さっぱりわからなかったので最初から略解読んだ。大変面白かった。 問 3 は初見殺しで問 4 は計算がゴツい。

(答案)

$$f(x) = \frac{{}_MC_x \ {}_{N-M}C_{K-x}}{{}_NC_K}$$

であるから、

$$\sum_{x=0}^{K} \frac{{}_MC_x \ {}_{N-M}C_{K-x}}{{}_NC_K} = 1 \tag{1}$$

を示せば良い。(1) は、

$${}_NC_K = \sum_{x=0}^{K} {}_MC_x \ {}_{N-M}C_{K-x} \tag{2}$$

と変形できる。さて、$(a+b)^N=(a+b)^M(a+b)^{N-M}$ を二項展開して、

$$\sum_{K=0}^{N} {}_NC_K a^K b^{N-K} = \sum_{x=0}^{M} {}_MC_x a^x b^{M-x} \sum_{y=0}^{N-M} {}_{N-M}C_y a^y b^{N-M-y}$$

これは、a,b を変数として持つ恒等式であるから、$a^K b^{N-K}$ の係数比較により、

$$\begin{aligned} {}_NC_K &= \sum_{x+y=K,0<x,y<K} {}_MC_x\ {}_{N-M}C_y \\ &= \sum_{x=0}^{K} {}_MC_x\ {}_{N-M}C_{K-x} \end{aligned}$$

が成り立つので、(2) が証明された。
　次に、平均を求める。

$$\begin{aligned} E[X] &= \sum_{x=0}^{K} x \frac{{}_MC_x\ {}_{N-M}C_{K-x}}{{}_NC_K} \\ &= \sum_{x=1}^{K} x \frac{{}_MC_x\ {}_{N-M}C_{K-x}}{{}_NC_K} \\ &= \frac{KM}{N} \sum_{x=1}^{K} \frac{{}_{M-1}C_{x-1}\ {}_{(N-1)-(M-1)}C_{(K-1)-(x-1)}}{{}_{N-1}C_{K-1}} \\ & \quad (\because x\ {}_MC_x = M\ {}_{M-1}C_{x-1},\quad {}_NC_K = \frac{N}{K}\ {}_{N-1}C_{K-1} \\ &= \frac{KM}{N} \sum_{y=0}^{K-1} \frac{{}_{M-1}C_y\ {}_{(N-1)-(M-1)}C_{(K-1)-y}}{{}_{N-1}C_y}) \\ & \quad (Y = X-1, Y \sim HGe(N-1, M-1, K-1)) \\ &= \frac{KM}{N} \end{aligned}$$

同様に、

$$\begin{aligned}
E[X(X-1)] &= \sum_{x=0}^{K} x(x-1)\frac{{}_MC_x\ {}_{N-M}C_{K-x}}{{}_NC_K} \\
&= \sum_{x=2}^{K} x(x-1)\frac{{}_MC_x\ {}_{N-M}C_{K-x}}{{}_NC_K} \\
&= \frac{K(K-1)M(M-1)}{N(N-1)}\sum_{x=2}^{K}\frac{{}_{M-2}C_{x-2}\ {}_{(N-2)-(M-2)}C_{(K-2)-(x-2)}}{{}_{N-2}C_{K-2}} \\
&\quad (\because\ \ x(x-1)\ {}_MC_x = M(x-1)\ {}_{M-1}C_{x-1} = M(M-1)\ {}_{M-2}C_{x-2} \\
&\quad {}_NC_K = \frac{N}{K}\ {}_{N-1}C_{K-1} = \frac{N(N-1)}{K(K-1)}\ {}_{n-2}C_{K-2} \qquad) \\
&= \frac{K(K-1)M(M-1)}{N(N-1)}\sum_{y=0}^{K}\frac{{}_{M-2}C_y\ {}_{(N-2)-(M-2)}C_{(K-2)-(y)}}{{}_{N-2}C_{K-2}} \\
&\quad Y = X-2,\ Y \sim HGe(N-2, M-2, K-2) \\
&= \frac{K(K-1)M(M-1)}{N(N-1)}
\end{aligned}$$

となるので、

$$Var(X) = \frac{N-K}{N-1}K\frac{M}{N}(1-\frac{M}{N})$$

3.4　難

(方針)
　密度関数を階乗表現を使って書き下すと、${}_KC_x$ が見えるので残りの部分をなんとかする。なんとかするところでスターリングの公式を使った。(忙しい人は飛ばしましょう)

スターリングの公式

実数 k が十分大きい時、

$$k! \approx \sqrt{2\pi} k^{k+a-\frac{1}{2}} \exp(-k)$$

(答案)

$$\begin{aligned}
f(x) &= \frac{{}_MC_x \ {}_{N-M}C_{K-x}}{{}_NC_K} \\
&= \frac{M!}{x!(M-x)!} \frac{(N-M)!}{(K-x)!(N-M-K+x)!} \frac{K!(N-K)!}{N!} \\
&= \frac{K!}{(K-x)!x!} \frac{M!(N-M)!(N-K)!}{(M-x)!(N-M-K+x)!N!} \\
&= {}_KC_x \frac{Np!(N-Np)!(N-K)!}{(Np-x)!(N-Np-K+x)!N!} \qquad (1)
\end{aligned}$$

今、$N \to \infty$ より、(1) の分数部分にスターリングの公式を適用すると、$\sqrt{2\pi}$ と exp の内部は消去されるので、

$$\frac{(Np)^{Np+\frac{1}{2}}(N-Np)^{N-Np+\frac{1}{2}}(N-K)^{N-K+\frac{1}{2}}}{(Np-x)^{Np-x+\frac{1}{2}}(n-Np-K+x)^{n-Np-K+x+\frac{1}{2}}N^{N+\frac{1}{2}}} \qquad (2)$$

べき乗の数が分母分子で等しいので、分母分子を N で割ることは各要素の中身を N で割ることに等しい。よって、

$$\begin{aligned}
(2) &= \frac{p^{Np+\frac{1}{2}}(1-p)^{N-Np+\frac{1}{2}}(1-\frac{K}{N})^{N-K+\frac{1}{2}}}{(p-\frac{x}{N})^{Np-x+\frac{1}{2}}(1-p-\frac{K}{N}+\frac{x}{N})^{N-Np-K+x+\frac{1}{2}}1^{N+\frac{1}{2}}} \\
&\approx p^{Np+\frac{1}{2}-Np+x-\frac{1}{2}}(1-p)^{N-Np+\frac{1}{2}-N+Np+K-x-\frac{1}{2}} \\
&= p^x(1-p)^{K-x} \\
\therefore (1) &\approx {}_KC_x p^x(1-p)^{K-x}
\end{aligned}$$

3.5 易

(方針)

指数分布について書く。指数分布は「単位時間あたり平均 λ 回発生する事象が、ある時点から次に発生するまでに x 単位時間かかる」確率を示す分布である。$Exp(\lambda) = Ga(1, \lambda)$ と考えれば確率密度関数を覚えなくて済む。確率変数 X が指数分布に従う時、$P[s \leq X]$ は、「ある時点から s 単位時間以上、

事故が起こらずに (故障せずに) 生存する (動作する) 確率」である。

確率密度関数：

$$f(x) = \lambda \exp(-\lambda x) \quad 0 < x$$

分布関数：

$$\begin{aligned} F_X(x) &= \int_0^x \lambda \exp(-\lambda t) dt \\ &= [-\exp(-\lambda t)]_0^x \\ &= -\exp(-\lambda x) + 1 \end{aligned}$$

生存確率：

$$P[s \leq X] = 1 - (-\exp(-\lambda s) + 1) = e^{-\lambda s}$$

条件付き生存確率 (s 単位時間生存したとすると、そこからさらに t 単位時間生存する確率)：

$$\begin{aligned} p[X \geq s + t \ |X \geq s] &= \frac{P[X \geq s + t \quad \wedge X \geq s]}{P[X \geq s]} \\ &= \frac{P[X \geq s + t]}{P[X \geq s]} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} \\ &= e^{-\lambda t} \\ &= P[X \geq t] \end{aligned}$$

これらのことに関連して、ハザード関数を定義する。ハザード関数は、x 単位時間生存した、という条件のもとで次の瞬間に故障する確率である。微小単位時間を Δ と置くと、

$$\begin{aligned} \lambda(x) &= \frac{P[x \leq X \leq x + \Delta \quad \wedge \quad x \leq X]}{P[x \leq X]} \\ &= \frac{P[x < X \leq x + \Delta]}{P[x \leq X]} \\ &\approx \frac{f(x)}{1 - F(x)} \end{aligned}$$

条件付き生存確率において確認した指数分布の無記憶性をハザード関数によって示そう。

$$\lambda(x) = \frac{\lambda \exp(-\lambda x)}{\exp(-\lambda x)} = \lambda$$

より、ハザード関数は時間 x に依存しない (何時間ここまで動作したかは次の瞬間故障する確率に関係しない)。

(答案)

$$
\begin{aligned}
\int_0^x \lambda(t)dt &= \int_0^x \frac{f(t)}{1-F(t)}dt \\
&= \int_0^x \frac{F'(t)}{1-F(t)}dt \\
&= [-\log(1-F(t))]_0^x \\
&= -\log(1-F(x))
\end{aligned}
$$

より、

$$
\begin{aligned}
F(x) &= 1-\exp\{-\int_0^x \lambda(t)dt\} \\
f(x) &= \frac{d}{dx}F(x) \\
&= -\exp\{-\int_0^x \lambda(t)dt\}(-\lambda(x)) \\
&= \lambda(x)\exp\{-\int_0^x \lambda(t)dt\}
\end{aligned}
$$

3.6 標準

(方針)
幾何分布について書く。$X \sim Geo(p)$ の時、X は「ベルヌーイ試行によって確率 p で起きる事象が、ある時点から初めて起きるまでに x 回失敗する」という確率変数である。問題は幾何分布が指数分布の離散バージョンであることを確認するもの。
確率密度関数：

$$f(x) = p(1-p)^x$$

分布関数：

$$\begin{aligned} F_X(x) &= \sum_{t=0}^{x} p(1-p)^t \\ &= p + p(1-p) + ... + p(1-p)^x \\ &= \frac{p\{1-(1-p)^{x+1}\}}{1-(1-p)} \\ &= 1-(1-p)^{x+1} \end{aligned}$$

生存確率 (s 回以上事象が起こらない確率):

$$\begin{aligned} P[s \leq X] &= 1 - P[X \leq s-1] \\ &= 1 - \{1-(1-p)^s\} \\ &= (1-p)^s \end{aligned}$$

条件付き生存確率 (s 回連続で失敗したとすると、そこからさらに t 回失敗する確率):

$$\begin{aligned} P[X \geq s+t | X \geq s] &= \frac{P[X \geq s+t]}{P[X \geq s]} \\ &= \frac{(1-p)^{s+t}}{(1-p)^s} \\ &= (1-p)^t \\ &= P[X \geq t] \end{aligned}$$

ハザード関数 (x 回失敗する条件のもとで、次に事象が起きる確率)：

$$\begin{aligned} \lambda(x) &= \frac{P[X = x]}{P[x \leq X]} \\ &= \frac{p(1-p)^x}{(1-p)^x} \\ &= p \end{aligned}$$

また、ここでの $[X]$ は通常 $\lfloor X \rfloor$ と書かれるもので、「X のフロア」と呼ぶ。このことについて書いておく。

フロア、シーリング

任意の実数 x に対して、x 以下の最大の整数を $\lfloor x \rfloor$(床、フロア) と表し、x 以上の最大の整数を $\lceil x \rceil$(天井、シーリング) と表す。(例えば、$\lfloor 1.5 \rfloor = 1$, $\lceil 1.5 \rceil = 2$。1.5 から見た床は 1 で、天井は 2。)

1. 任意の実数 x に対して、

$$x - 1 < \lfloor x \rfloor \leq x \leq \lceil x \rceil < x + 1$$

(x の天井は $x+1$ よりは低く、床は $x-1$ よりは高い。)

2. 任意の整数 n に対して、

$$\lceil n/2 \rceil + \lfloor n/2 \rfloor = n$$

($\lfloor a/2 \rfloor$ は a を 2 で割った時の整除 (小数点以下を切り捨てた商)、python で言う a//2 みたいなこと。$\lceil a/2 \rceil$ は a//2+1)

1 を書いたので若干混乱するが、

$$\lfloor X \rfloor = y \Leftrightarrow y \leq X < y$$

である。つまり、「実数 X の床が整数 y である」とは、「実数 X が y 以上 $y+1$ 未満である」ことと同値。このことを用いて答案を書いた。

(答案)

$$\begin{aligned}
P[Y=y] &= P[\lfloor X \rfloor = y] \\
&= P[y \leq X < y+1] \\
&= F(y+1) - F(y) \\
&= \{-\exp(-\lambda(y+1))\} - \{-\exp(-\lambda y)\} \\
&= -\exp\{-\lambda(y+1)\} + \exp(-\lambda y) \\
&= \exp(-\lambda y)\{1 - \exp(-\lambda)\} \\
&= \{1 - \exp(-\lambda)\}\{1 - (1 - \exp(-\lambda))\}^y
\end{aligned}$$

より、$Y \sim Geo(1 - \exp(-\lambda))$ である。無記憶性については方針の通り。

3.7 標準

(方針)

負の二項分布について書く。負の二項分布は、ベルヌーイ試行が r 回成功

するまでに x 回失敗する確率を示す。確率密度関数でコンビネーションの左下の添え字に-1 があることに注意。(一番最後は必ず成功)

確率密度関数：

$$f(x) \quad = \quad {}_{x+r-1}C_x p^r (1-p)^x$$

モーメント母関数：

$$\begin{aligned} M_X(t) &= \sum_{x=0}^{\infty} {}_{x+r-1}C_x p^r \{e^t(1-p)\}^x \\ &= p^r\{1-e^t(1-p)\}^{-r} \sum_{x=0}^{\infty} {}_{x+r-1}C_x \{1-e^t(1-p)\}^r \{e^t(1-p)\}^x \\ &= \{\frac{p}{1-e^t(1-p)}\}^r \end{aligned}$$

問題は、負の二項分布が変数変換によってカイ二乗分布に収束することを示すものである。自由度 m のカイ二乗分布に従う確率密度関数 Y のモーメント母関数は、

$$\begin{aligned} M_Y(t) &= \int_0^{\infty} \frac{(\frac{1}{2})^{m/2}}{\Gamma(\frac{m}{2})} y^{\frac{m}{2}-1} \exp\{(t-\frac{1}{2})y\}dy \\ &= \frac{(\frac{1}{2})^{m/2}}{\Gamma(\frac{m}{2})} \frac{\Gamma(\frac{m}{2})}{(\frac{1}{2}-t)^{m/2}} \int_0^{\infty} \frac{(\frac{1}{2}-t)^{m/2}}{\Gamma(\frac{m}{2})} y^{\frac{m}{2}-1} \exp\{-(\frac{1}{2}-t)y\}dy \\ &= (\frac{1}{1-2t})^{m/2} \end{aligned}$$

である。負の二項分布のモーメント母関数には e が含まれているのにカイ二乗分布のそれには含まれていないことから、e は漸近展開 (2.13 参照) を使うだろうな、と見通しを立てる。
(分布収束を示す問題は大体 5.10 で説明している CLT を使うか積率母関数でゴリゴリ計算するかのどっちか。CLT も積率母関数で証明する。)

(答案)

Y と、χ^2_{2r} に従う確率変数 Z の積率母関数をそれぞれ求める。

$$
\begin{aligned}
M_Y(t) &= E[e^{tY}] \\
&= E[e^{t2pX}] \\
&= \{\frac{p}{1-e^{t2p}(1-p)}\}^r \\
&= \{\frac{1}{\frac{1}{p}-e^{t2p}(\frac{1}{p}-1)}\}^r \qquad (1) \\
M_Z(t) &= E[e^{tZ}] \\
&= \int_0^{\infty} \frac{(\frac{1}{2})^r}{\Gamma(r)} z^{r-1} \exp\{-(\frac{1}{2}-t)z\} dz \\
&= (\frac{1}{1-2t})^r \qquad (2)
\end{aligned}
$$

$p \to 0$ のとき、(1) $\approx$ (2) を示ばよい。$e^{t2p} = 1 + t2p + o(p)$ より、

$$
\begin{aligned}
(1) &= \{\frac{1}{\frac{1}{p}-(1+t2p+o(p))(\frac{1}{p}-1)}\}^r \\
&= \{\frac{1}{1-2t+o(p)}\}^r \\
&\approx (\frac{1}{1-2t})^r \qquad as \;\; p \to 0
\end{aligned}
$$

3.8 標準〜難

(方針)

問 9 を先にやるべき。

二項分布に従う確率変数 X のモーメント母関数：

$$
\begin{aligned}
M_X(t) &= E[e^{tX}] \\
&= \sum_{x=0}^{n} {}_nC_x (e^t p)^x (1-p)^{n-x} \\
&= (e^t p + 1 - p)^n
\end{aligned}
$$

標準正規分布に従う確率変数 Y のモーメント母関数：

$$\begin{aligned}
M_Y(t) &= E[e^{tY}] \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{1\pi}} \exp(ty - \frac{y^2}{2})dy \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp[-\frac{1}{2}\{(y-t)^2 - t^2\}]dy \\
&= e^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{(y-t)^2}{2}\}dy \\
&= e^{t^2/2}
\end{aligned}$$

以上を用いて、「変数変換後のモーメント母関数を求める」「母関数の式をいじって導出したい分布の母関数にする」と言う手順で答案を書く。漸近展開で、

$$\begin{aligned}
e^x &= 1 + x + \frac{x^2}{2} + o(x^2) \\
\log(1+x) &= x - \frac{x^2}{2} + o(x^2)
\end{aligned}$$

を使う。オーダー記法による計算をわかりやすくするためにかなりインフォーマルに答案を書いた。

(答案)

$$\begin{aligned}
M_Y(t) &= E[e^{tY}] \\
&= \exp(-\frac{tnp}{\sqrt{np(1-p)}})E[e^{\frac{t}{\sqrt{np(1-p)}}X}] \\
&= \exp(-\frac{tnp}{\sqrt{np(1-p)}})(e^{\frac{t}{\sqrt{np(1-p)}}} \times p + 1 - p)^n
\end{aligned}$$

$\psi(t) = \log M_Y(t)$ とする。$n \to \infty$ のとき $\psi(t) \approx \frac{t^2}{2}$ を示ばよい。

$$\psi(t) = -\frac{tnp}{\sqrt{np(1-p)}} + n\log(e^{\frac{t}{\sqrt{np(1-p)}}} \times p + 1 - p)^n \tag{1}$$

と書ける。e の漸近展開により、

$$e^{\frac{t}{\sqrt{np(1-p)}}} = 1 + \frac{t}{\sqrt{np(1-p)}} + \frac{t^2}{2np(1-p)} + o(\frac{1}{n}) \qquad as\ n \to \infty$$

なので、

$$
\begin{aligned}
(1) &= -\frac{tnp}{\sqrt{np(1-p)}} + n\log\{p(1+\frac{t}{\sqrt{np(1-p)}} + \frac{t^2}{2np(1-p)} + o(\frac{1}{n})) + 1 - p\} \\
&= -\frac{tnp}{\sqrt{np(1-p)}} + n\log\{1 + (\frac{tp}{\sqrt{np(1-p)}} + \frac{t^2}{2n(1-p)} + o(\frac{1}{n}))\} \qquad (2)
\end{aligned}
$$

第二項を $n\log(1+A)$ と置くと、漸近展開によって

$$
\begin{aligned}
n\log(1+A) &= n(A + \frac{A^2}{2} + o(A^2)) \\
&= n(\frac{tp}{\sqrt{np(1-p)}} + \frac{t^2}{2n(1-p)} + o(\frac{1}{n}) + \frac{A^2}{2} + o(A^2)) \\
&= \frac{tnp}{\sqrt{np(1-p)}} + \frac{t^2}{2(1-p)} + o(1) + \frac{nA^2}{2} + o(1) \qquad (3) \\
\frac{nA^2}{2} &= (\frac{tp}{\sqrt{np(1-p)}} + \frac{t^2}{2n(1-p)})^2/2 \\
&= (\frac{t^2p^2}{p(1-p)} + \frac{pt^3}{\sqrt{np(1-p)}(1-p)} + \frac{t^4}{4n(1-p)^2})/2 \\
&= \frac{t^2p^2}{2p(1-p)} + o(\frac{1}{\sqrt{n}}) + o(\frac{1}{n}) \\
&= \frac{t^2p^2}{2p(1-p)} + o(\frac{1}{\sqrt{n}}) \\
\therefore (3) &= \frac{tnp}{\sqrt{np(1-p)}} + \frac{t^2}{2(1-p)} + \frac{t^2p^2}{2p(1-p)} + o(1)
\end{aligned}
$$

以上を (2) に代入することで、

$$
\begin{aligned}
\psi(t) &= \frac{t^2}{2(1-p)} + \frac{t^2p^2}{2p(1-p)} + o(1) \\
&\approx \frac{t^2}{2} \quad as\ n \to \infty
\end{aligned}
$$

3.9 標準

(方針)

ポアソン分布に従う確率変数 X のモーメント母関数：

$$\begin{aligned} M_X(t) &= \sum_{x=0}^{\infty} \frac{(e^t\lambda)^x}{x!} \exp(-\lambda) \\ &= \exp(-\lambda)\exp(e^t\lambda) \\ &= \exp\{\lambda(e^t-1)\} \end{aligned}$$

(答案)

$$\begin{aligned} M_Y(t) &= E[e^{tY}] \\ &= E[e^{\frac{t(X-\lambda)}{\sqrt{\lambda}}}] \\ &= E[\exp(\frac{t}{\sqrt{\lambda}}X)\exp(-t\sqrt{\lambda})] \\ &= \exp(-t\sqrt{\lambda})\exp\{\lambda(e^{\frac{t}{\sqrt{\lambda}}}-1)\} \end{aligned}$$

ここで、$\psi(t) = \log M_Y(t)$ と置くと、

$$\begin{aligned} \psi(t) &= -t\sqrt{\lambda} + \lambda(e^{\frac{t}{\sqrt{\lambda}}} - 1) \\ &= -t\sqrt{\lambda} + \lambda\{(1 + \frac{t}{\sqrt{\lambda}} + \frac{t^2}{2\lambda} + o(\frac{1}{\lambda})) - 1\} \\ &\approx \frac{t^2}{2} \quad as \quad \lambda \to \infty \end{aligned}$$

3.10 標準

$$E[X] = \int_0^{\infty} \frac{x}{\sqrt{2\pi}x} \exp\{-\frac{(\log x)^2}{2}\}dx \tag{1}$$

$t = \log x$ と置くと、$x = e^t$ で、$dx = e^t dt$ である。また、$x : 0 \to \infty$ の時、$t : -\infty \to \infty$ なので、

$$
\begin{aligned}
(1) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-\frac{t^2}{2} + t)dt \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{(t+1)^2}{2} + \frac{1}{2}\}dt \\
&= e^{1/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{(t+1)^2}{2}\}dt \\
&= e^{1/2}
\end{aligned}
$$

次に、$Y = \log X$ なる変数変換を考える。

$$
\begin{aligned}
P[Y \leq y] &= P[\log X \leq y] \\
&= P[X \leq e^y] \\
&= \int_0^{e^y} f_X(x)dx \\
\therefore f_Y(y) &= f_X(e^y)e^y \\
&= \frac{1}{\sqrt{2\pi}e^y} \exp\{\frac{-(\log e^y)^2}{2}\}e^y \\
&= \frac{1}{\sqrt{2\pi}} \exp(-\frac{y^2}{2})
\end{aligned}
$$

以上より、$Y \sim N(0,1)$

3.11 標準

$$
\begin{aligned}
E[X] &= \int_{\alpha}^{\infty} \frac{\beta\alpha^{\beta}}{x^{\beta}} dx \\
&= \beta\alpha^{\beta} \int_{\alpha}^{\infty} x^{-\beta} dx \\
&= \beta\alpha^{\beta} [-\frac{1}{\beta - 1} x^{-(\beta-1)}]_{\alpha}^{\infty} \\
&= \frac{\alpha\beta}{\beta - 1} \qquad (1 < \beta) \\
E[X^2] &= \int_{\alpha}^{\infty} \frac{\beta\alpha^{\beta}}{x^{\beta-1}} dx \\
&= \beta\alpha^{\beta} \int_{\alpha}^{\infty} x^{-(\beta-1)} dx \\
&= \beta\alpha^{\beta} [-\frac{1}{\beta - 2} x^{-(\beta-2)}]_{\alpha}^{\infty} \\
&= \frac{\alpha^2\beta}{\beta - 2} \\
\therefore Var(X) &= \frac{\alpha^2\beta}{\beta - 2} - (\frac{\alpha\beta}{\beta - 1})^2 \\
&= \frac{\alpha^2\beta}{(\beta - 2)(\beta - 1)^2} \qquad (2 < \beta)
\end{aligned}
$$

次に、$Y = \log X$ なる変数変換を行うと、

$$
\begin{aligned}
f_Y(y) &= f(e^y)e^y \\
&= \frac{\beta\alpha^{\beta}}{e^{(\beta+1)y}} e^y \\
&= \beta\alpha^{\beta} \exp(-\beta y) \qquad (1)
\end{aligned}
$$

ここで、$\beta = \frac{1}{\beta}$, $\alpha = e^y$ を代入すると

$$
(1) = \frac{1}{\sigma} \exp(-\frac{y - \mu}{\sigma})
$$

3.12 標準

$$
\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} \frac{x}{2\sigma} \exp\{-\frac{|x-\mu|}{\sigma}\}dx \\
&= \int_{-\infty}^{\mu} \frac{x}{2\sigma} \exp(\frac{x-\mu}{\sigma})dx + \int_{\mu}^{\infty} \frac{x}{2\sigma} \exp(-\frac{x-\mu}{\sigma})dx \\
&= \int_{-\infty}^{\mu} \frac{x}{2}\{\exp(\frac{x-\mu}{\sigma})\}'dx + \int_{\mu}^{\infty} \frac{x}{2}\{-\exp(-\frac{x-\mu}{\sigma})\}'dx \\
&= [\frac{x}{2}\exp(\frac{x-\mu}{\sigma})]_{-\infty}^{\mu} - \int_{-\infty}^{\mu} \frac{1}{2}\exp(\frac{x-\mu}{\sigma})dx \\
&\quad +[-\frac{x}{2}\exp(-\frac{x-\mu}{\sigma})]_{\mu}^{\infty} + \int_{\mu}^{\infty} \frac{1}{2}\exp(-\frac{x-\mu}{\sigma})dx \\
&= \mu - [\frac{\sigma}{2}\exp(\frac{x-\mu}{\sigma})]_{-\infty}^{\mu} + [-\frac{\sigma}{2}\exp(-\frac{x-\mu}{\sigma})]_{\mu}^{\infty} \\
&= \mu - \frac{\sigma}{2} + \frac{\sigma}{2} \\
&= \mu
\end{aligned}
$$

$$
\begin{aligned}
Var(X) &= \int_{-\infty}^{\infty} \frac{(x-\mu)^2}{2\sigma} \exp\{-\frac{|x-\mu|}{\sigma}\}dx \\
&= \int_{-\infty}^{\mu} \frac{(x-\mu)^2}{2\sigma} \exp(\frac{x-\mu}{\sigma})dx + \int_{\mu}^{\infty} \frac{(x-\mu)^2}{2\sigma} \exp(-\frac{x-\mu}{\sigma})dx \qquad (1)
\end{aligned}
$$

ここで、$y = \frac{x-\mu}{\sigma}$ と置くと、$dx = \sigma dy$。また、$x : -\infty \to \mu$ の時、$y : -\infty \to 0$,$x : \mu \to \infty$ の時、$y : 0 \to \infty$ であるから、

$$
\begin{aligned}
(1) &= \int_{-\infty}^{0} \frac{\sigma^2}{2} y^2 \exp y dy + \int_{0}^{\infty} \frac{\sigma^2}{2} y^2 \exp(-y) dy \\
&= \frac{\sigma^2}{2} \int_{-\infty}^{0} y^2 (\exp y)' dy + \frac{\sigma^2}{2} \int_{0}^{\infty} y^2 \{-\exp(-y)\}' dy \\
&= \frac{\sigma^2}{2} \{[y^2 \exp y]_{-\infty}^{0} - \int_{-\infty}^{0} 2y \exp y dy\} \\
&\quad + \frac{\sigma^2}{2} \{[-y^2 \exp(-y)]_{0}^{\infty} - \int_{0}^{\infty} -2y \exp(-y) dy\} \\
&= -\frac{\sigma^2}{2} \{[2y \exp y]_{-\infty}^{0} - \int_{-\infty}^{0} 2 \exp y dy\} \\
&\quad - \frac{\sigma^2}{2} \{[2y \exp(-y)]_{0}^{\infty} - \int_{0}^{\infty} 2 \exp(-y) dy\} \\
&= \frac{\sigma^2}{2} [2 \exp y]_{-\infty}^{0} + \frac{\sigma^2}{2} [-2 \exp(-y)]_{0}^{\infty} \\
&= 2\sigma^2
\end{aligned}
$$

3.13 標準

(方針)
ロジスティックシグモイド関数は、機械学習の分類問題でよく使われる。二ク

ラス分類を考える。入力 a が与えられたときクラス C_1 に属する事後確率は、

$$\begin{aligned}
P[C_1|a] &= \frac{P[a \wedge C_1]}{P[a]} \\
&= \frac{P[a|C_1]P[C_1]}{P[a|C_1]P[C_1] + P[a|C_2]P[C_2]} \\
&= \frac{1}{1 - \frac{P[a|C_2]P[C_2]}{P[a|C_1]P[C_1]}} \\
&= \frac{1}{1 + \exp\{-\log \frac{P[a|C_2]P[C_2]}{P[a|C_1]P[C_1]}\}} \\
&= \frac{1}{1 + \exp\{-\log \frac{P[a|C_1]P[C_1]}{P[a|C_2]P[C_2]}\}} \\
&= \frac{1}{1 + \exp(-t)} \\
& t = \log \frac{P[a|C_1]P[C_1]}{P[a|C_2]P[C_2]}
\end{aligned}$$

このように、

$$\sigma(t) = \frac{1}{1 + \exp(-t)}$$

で定義される $\sigma(t)$ をロジスティックシグモイド関数という。これは分布関数である。微分すると、

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

となる。このようにロジスティックシグモイド関数は確率的生成モデルとして導出できる。また実数空間の入力を [0,1] 区間に押し込む性質からロジスティック回帰や活性化関数などにも用いられる。今回の問題では、ロジスティックシグモイド関数 (分布関数) の微分が確率密度関数になっていることや、分布が左右対象になっていることなどを説明する。

(答案)

1.

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

2.

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} \frac{e^{-t}}{(1+e^{-t})^2} dt \\ &= [\frac{1}{1+e^{-t}}]_{-\infty}^{x} \\ &= \frac{1}{1+e^{-x}} \end{aligned}$$

また、

$$\begin{aligned} \frac{e^{-x}}{(1+e^{-x})^2} - \frac{e^{x}}{(1+e^{x})^2} &= \frac{e^{-x}(1+e^{x})^2 - e^{x}(1+e^{-x})^2}{(1+e^{-x})^2(1+e^{x})^2} \\ &= \frac{(e^{-x}+1+e^{x}) - (e^{x}+2+e^{-x})}{(1+e^{-x})^2(1+e^{x})^2} \\ &= 0 \end{aligned}$$

より、任意の実数 x について $f(x) = f(-x)$ が成り立つので、$f(x)$ は y 軸について対称。

3.

$$\begin{aligned} U &= \frac{e^{-x}}{1+e^{-x}} \\ &= 1 - \frac{1}{1+e^{-x}} \\ &= 1 - F_X(x) \end{aligned}$$

より、U は X のついての単調減少関数で、$0 \leq U \leq 1$ である。また、$X = F_X^{-1}(1-U)$ なので、

$$\begin{aligned} \frac{dx}{du} &= (\frac{du}{dx})^{-1} \\ &= -\frac{1}{f_X(x)} \\ &= -\frac{1}{f_X(F_X^{-1}(1-U))} \end{aligned}$$

以上より、

$$\begin{aligned} f_U(u) &= f_X(F_X^{-1}(1-U))\{-(-\frac{1}{f_X(F_X^{-1}(1-U))})\} \\ &= 1 \quad 0 \leq u \leq 1 \end{aligned}$$

ゆえに、$U \sim U(0,1)$

4.

$$
\begin{aligned}
P[Y \le y] &= P[|X| \le y] \\
&= P[-y \le X \le y] \\
&= \int_{-\infty}^{y} f_X(x)dx - \int_{-\infty}^{-y} f_X(x)dx \\
\therefore f_Y(y) &= f_X(y) - f_X(-y)(-1) \\
&= 2f_X(y) \\
&= \frac{2e^{-y}}{(1+e^{-y})^2} \qquad 0 < y \\
F_Y(y) &= \int_0^y \frac{2e^{-t}}{(1+e^{-t})^2} dt \\
&= [\frac{2}{(1+e^{-t})}]_0^y \\
&= \frac{2}{1+e^{-y}} - 1 \\
&= \frac{1-e^{-y}}{1+e^{-y}} \\
\therefore \lambda(y) &= \frac{f_Y(y)}{1-F_Y(y)} \\
&= \frac{2e^{-y}}{(1+e^{-y})^2} \frac{1+e^{-y}}{2e^{-y}} \\
&= \frac{1}{1+e^{-y}}
\end{aligned}
$$

5.

$$
\begin{aligned}
P[Y \le y] &= P[\sigma X + \mu \le y] \\
&= \begin{cases} P[X \le \frac{y-\mu}{\sigma}] & if\ 0 < \sigma \\ P[X \ge \frac{y-\mu}{\sigma}] & if\ \sigma < 0 \end{cases} \\
&= \begin{cases} \int_{-\infty}^{\frac{y-\mu}{\sigma}} f_X(x)dx \\ 1 - \int_{-\infty}^{\frac{y-\mu}{\sigma}} f_X(x)dx \end{cases}
\end{aligned}
$$

より、

$$
\begin{aligned}
f_Y(y) &= \begin{cases} f_X(\frac{y-\mu}{\sigma})\frac{1}{\sigma} & if\ 0 < \sigma \\ -f_X(\frac{y-\mu}{\sigma})\frac{1}{\sigma} & if\ \sigma < 0 \end{cases} \\
&= \begin{cases} \frac{e^{-(y-\mu)/\sigma}}{\sigma(1+e^{-(y-\mu)/\sigma})^2} & if\ 0 < \sigma \\ -\frac{e^{-(y-\mu)/\sigma}}{\sigma(1+e^{-(y-\mu)/\sigma})^2} & if\ \sigma < 0 \end{cases}
\end{aligned}
$$

3.14 標準

1.

$$
\begin{aligned}
E[X^2] &= \int_{-\infty}^{\infty} \frac{x^2}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})dx \\
&= \int_{-\infty}^{\infty} \frac{x}{\sqrt{2\pi}} \{-\exp(-\frac{x^2}{2})\}' dx \\
&= [-\frac{x}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})dx \\
&= 1
\end{aligned}
$$

2.

$$
\begin{aligned}
P[Y \leq y] &= P[X^2 \leq y] \\
&= P[-\sqrt{y} \leq X \leq \sqrt{y}] \\
&= \int_{-\infty}^{\sqrt{y}} f_X(x)dx - \int_{-\infty}^{-\sqrt{y}} f_X(x)dx
\end{aligned}
$$

より、

$$
\begin{aligned}
f_Y(y) &= f_X(\sqrt{y})\frac{1}{2\sqrt{y}} - f_X(-\sqrt{y})(-\frac{1}{2\sqrt{y}}) \\
&= \frac{1}{\sqrt{y}} f_X(\sqrt{y}) \\
&= \frac{1}{\sqrt{2\pi y}} \exp(-\frac{y}{2})
\end{aligned}
$$

である。したがって、

$$
E[Y] = \int_0^{\infty} \sqrt{\frac{y}{2\pi}} \exp(-\frac{y}{2})dy \tag{1}
$$

ここで、$\sqrt{y} = t$ と置くと、$\frac{dt}{dy} = \frac{1}{2\sqrt{y}} \Rightarrow dy = 2tdt$ で、$y : 0 \to \infty$ の時、$t : 0 \to \infty$ なので、

$$\begin{aligned}
(1) &= \int_0^\infty \frac{2}{\sqrt{2\pi}} t^2 \exp(-\frac{t^2}{2}) dt \\
&= \int_0^\infty \frac{2}{\sqrt{2\pi}} t\{-\exp(-\frac{t^2}{2})\}' dt \\
&= [-\frac{2}{\sqrt{2\pi}} t \exp(-\frac{t^2}{2})]_0^\infty + \int_0^\infty \frac{2}{\sqrt{2\pi}} \exp(-\frac{t^2}{2}) dt \\
&= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} \exp(-\frac{t^2}{2}) dt \\
&= 1
\end{aligned}$$

3.

$$\begin{aligned}
P[Y \leq y] &= P[|X| \leq y] \\
&= P[-y \leq X \leq y] \\
&= \int_{-\infty}^{y} f(x)dx - \int_{-\infty}^{-y} f(x)dx \\
\therefore f_Y(y) &= f_X(y) - f_X(-y)(-1) \\
&= 2f_X(y) \\
&= \frac{2}{\sqrt{2\pi}} \exp(-\frac{y^2}{2})
\end{aligned}$$

より、

$$
\begin{aligned}
E[Y] &= \int_0^\infty \frac{2y}{\sqrt{2\pi}} \exp(-\frac{y^2}{2}) dy \\
&= [-\frac{2}{\sqrt{2\pi}} \exp(-\frac{y^2}{2})]_0^\infty \\
&= \sqrt{\frac{2}{\pi}} \\
E[Y^2] &= \int_0^\infty \frac{2y^2}{\sqrt{2\pi}} \exp(-\frac{y^2}{2}) dy \\
&= \int_0^\infty \frac{2y}{\sqrt{2y}} \{-\exp(-\frac{y^2}{2})\}' dy \\
&= [-\frac{2y}{\sqrt{2\pi}} \exp(-\frac{y^2}{2})]_0^\infty + \int_0^\infty \frac{2}{\sqrt{2\pi}} \exp(-\frac{y^2}{2}) dy \\
&= 1 \\
\therefore Var(Y) &= 1 - \frac{2}{\pi}
\end{aligned}
$$

3.15 標準

(方針)
奇関数の積分は 0 である。

偶関数、奇関数

実数直線状に定義される関数 $f(x)$ が

$$f(-x) = f(x)$$

を満たす時、f は偶関数であるといい、f のグラフは y 軸で対称となる。
また、$f(x)$ が

$$f(-x) = -f(x)$$

を満たす時、f は奇関数であるといい、f のグラフは原点で対称となる。

(答案)

$$f(x) = \frac{x^k}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$$

と置くと、k が奇数の時 $f(-x) = -f(x)$ となるので、f は奇関数。奇関数の積分は 0 であるから、$E[X^k] = 0$。

k が偶数の時、$Y = X^2$ と置くと、$Y \sim \chi_1^2$ より、

$$\begin{aligned}
f_Y(y) &= \frac{(\frac{1}{2})^{1/2}}{\Gamma(\frac{1}{2})} y^{\frac{1}{2}-1} \exp(-\frac{1}{2}y) \\
\therefore E[X^k] &= E[Y^{k/2}] \\
&= \int_0^\infty \frac{(\frac{1}{2})^{1/2}}{\Gamma(\frac{1}{2})} y^{\frac{k+1}{2}-1} \exp(-\frac{1}{2}y) \\
&= \frac{(\frac{1}{2})^{1/2}}{\Gamma(\frac{1}{2})} \frac{\Gamma(\frac{k+1}{2})}{(\frac{1}{2})^{\frac{k+1}{2}}} \int_0^\infty \frac{(\frac{1}{2})^{\frac{k+1}{2}}}{\Gamma(\frac{k+1}{2})} y^{\frac{k+1}{2}-1} \exp(-\frac{1}{2}y) dy \\
&= 2^{k/2} \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{1}{2})}
\end{aligned}$$

3.16 標準

1. (方針)

　そのままでは一対一変換にならないので $Y = |X|$ の確率密度関数を求めてから $Z = Y^k$(一対一) を考える。

　(答案)

　$Y = |X|$ と置くと、

$$\begin{aligned}
P[Y \leq y] &= P[|X| \leq y] \\
&= P[-y \leq X \leq y] \\
&= \int_{-\infty}^{y} f_X(x)dx - \int_{-\infty}^{-y} f_X(x)dx \\
\therefore f_Y(y) &= f_X(y) - f_X(-y)(-1) \\
&= 2f_X(y)
\end{aligned}$$

次に $Z = Y^k$ と置くと、

$$\begin{aligned} P[Z \le z] &= P[Y^k \le z] \\ &= P[Y \le z^{1/k}] \\ &= \int_0^{z^{1/k}} f_Y(y)dy \\ \therefore f_Z(z) &= f_Y(z^{1/k})(\frac{1}{k} z^{\frac{1}{k}-1}) \\ &= 2f_X(z^{1/k})(\frac{1}{k} z^{\frac{1}{k}-1}) \\ &= \frac{2}{k\sqrt{2\pi}} z^{\frac{1}{k}-1} \exp(-\frac{z^{2/k}}{2}) \end{aligned}$$

2. $Z = |X|^{2m}$ と置くと、

$$f_Z(z) = \frac{1}{m\sqrt{2\pi}} z^{\frac{1}{2m}-1} \exp(-\frac{z^{1/m}}{2})$$

であるから、

$$\begin{aligned} E[X^{2m}] &= E[|X|^{2m}] \\ &= E[Z] \\ &= \int_0^\infty \frac{1}{m\sqrt{2\pi}} z^{\frac{1}{2m}} \exp(-\frac{z^{1/m}}{2}) dz \end{aligned} \tag{1}$$

ここで、$t = z^{1/m}$ と置くと、$dz = mt^{m-1}dt$ なので、

$$\begin{aligned} (1) &= \int_0^\infty \frac{1}{\sqrt{2\pi}} t^{(\frac{1}{2}+m)-1} \exp(-\frac{t}{2}) dt \\ &= \frac{(\frac{1}{2})^{1/2}}{\Gamma(\frac{1}{2})} \frac{\Gamma(\frac{1}{2}+m)}{(\frac{1}{2})^{\frac{1}{2}+m}} \int_0^\infty \frac{(\frac{1}{2})^{\frac{1}{2}+m}}{\Gamma(\frac{1}{2}+m)} t^{(\frac{1}{2}+m)-1} \exp(-\frac{1}{2}t) dt \\ &= \frac{2^m \Gamma(\frac{1}{2}+m)}{\Gamma(\frac{1}{2})} \end{aligned}$$

次に、$Z = |X|^{2m+1}$ と置くと、

$$f_Z(z) = \frac{2}{(2m+1)\sqrt{2\pi}} z^{\frac{1}{2m+1}-1} \exp(-\frac{z^{\frac{2}{2m+1}}}{2})$$

であるから、

$$\begin{aligned}
E[|X|^{2m+1}] &= E[Z] \\
&= \int_0^\infty \frac{2}{(2m+1)\sqrt{2\pi}} z^{\frac{1}{2m+1}} \exp(-\frac{z^{\frac{2}{2m+1}}}{2}) dz \qquad (2)
\end{aligned}$$

ここで、$t = z^{\frac{2}{2m+1}}$ と置く。$t^{\frac{2m+1}{2}} = z$ より、$dz = \frac{2m+1}{2} t^{\frac{2m+1}{2}-1} dt$ なので、

$$\begin{aligned}
(2) &= \int_0^\infty \frac{1}{\sqrt{2\pi}} t^{(m+1)-1} \exp(-\frac{1}{2}t) dt \\
&= \frac{(\frac{1}{2})^{1/2}}{\Gamma(\frac{1}{2})} \frac{\Gamma(m+1)}{(\frac{1}{2})^{m+1}} \int_0^\infty \frac{(\frac{1}{2})^{m+1}}{\Gamma(m+1)} t^{(m+1)-1} \exp(-\frac{1}{2}t) dt \\
&= \frac{2^{m+\frac{1}{2}} \Gamma(m+1)}{\Gamma(\frac{1}{2})}
\end{aligned}$$

3.17 標準

(方針)
絶対値は場合分けするより丸々置き換えた方が単純になる場合がある。また、exp 内に次数の高い項があったら大体置換積分する。

(答案)
$Y = |X|$ と置く。

$$\begin{aligned}
P[Y \leq y] &= P[|X| \leq y] \\
&= P[-y \leq X \leq y] \\
&= \int_{-\infty}^{y} f_X(x) dx - \int_{-\infty}^{-y} f_X(x) dx \\
\therefore f_Y(y) &= f_X(y) - f_X(-y)(-1) \\
&= 2C(\alpha) \exp(-y^\alpha) \qquad (0 < y < \infty)
\end{aligned}$$

より、

$$\int_0^\infty f_Y(y) dy = \int_0^\infty 2C(\alpha) \exp(-y^\alpha) dy \qquad (1)$$

$t = y^\alpha$ と置くと、$dy = \frac{1}{\alpha} t^{\frac{1}{\alpha}-1} dt$ で、$y : 0 \to \infty$ の時、$t : 0 \to \infty$ なので、

$$\begin{aligned}
(1) &= \int_0^\infty \frac{2C(\alpha)}{\alpha} t^{\frac{1}{\alpha}-1} \exp(-t) dt \\
&= \frac{2C(\alpha)}{\alpha} \Gamma(\frac{1}{\alpha}) \int_0^\infty \frac{1}{\Gamma(\frac{1}{\alpha})} t^{\frac{1}{\alpha}-1} \exp(-t) dt \\
&= \frac{2C(\alpha)}{\alpha} \Gamma(\frac{1}{\alpha}) = 1 \\
\therefore C(\alpha) &= \frac{\alpha}{2\Gamma(\frac{1}{\alpha})}
\end{aligned}$$

次に、

$$\begin{aligned}
E[|X|^\nu] &= E[Y^\nu] \\
&= \int_0^\infty \frac{\alpha}{\Gamma(\frac{1}{\alpha})} y^\nu \exp(-y^\alpha) dy \qquad (2)
\end{aligned}$$

$t = y^\alpha$ と置くと、$dy = \frac{1}{\alpha} t^{\frac{1}{\alpha}-1} dt,\ y^\nu = t^{\nu/\alpha}$ なので、

$$\begin{aligned}
(2) &= \int_0^\infty \frac{1}{\Gamma(\frac{1}{\alpha})} t^{\frac{\nu+1}{\alpha}-1} \exp(-t) dt \\
&= \frac{\Gamma(\frac{\nu+1}{\alpha})}{\Gamma(\frac{1}{\alpha})} \int_0^\infty \frac{1}{\Gamma(\frac{\nu+1}{\alpha})} t^{\frac{\nu+1}{\alpha}-1} \exp(-t) dt \\
&= \frac{\Gamma(\frac{\nu+1}{\alpha})}{\Gamma(\frac{1}{\alpha})}
\end{aligned}$$

3.18 易

1.

$$\begin{aligned}
E[X^\nu] &= \int_0^\infty \frac{(\frac{1}{2})^{n/2}}{\Gamma(\frac{n}{2})} x^{\nu+\frac{n}{2}-1} \exp(-\frac{1}{2}x) dx \\
&= \frac{(\frac{1}{2})^{n/2}}{\Gamma(\frac{n}{2})} \frac{\Gamma(\nu+\frac{n}{2})}{(\frac{1}{2})^{\nu+\frac{n}{2}}} \int_0^\infty \frac{(\frac{1}{2})^{\nu+\frac{n}{2}}}{\Gamma(\nu+\frac{n}{2})} x^{\nu+\frac{n}{2}-1} \exp(-\frac{1}{2}x) dx \\
&= \frac{2^\nu \Gamma(\nu+\frac{n}{2})}{\Gamma(\frac{n}{2})}
\end{aligned}$$

2. $X \sim \chi_n^2$ の時、 X の確率密度関数を $f_n(x)$ と表すことにする。

$$
\begin{aligned}
f_{n+2}(x) &= \frac{(\frac{1}{2})^{\frac{n}{2}+1}}{\Gamma(\frac{n}{2}+1)} x^{n/2} \exp(-\frac{1}{2}x) \\
&= \frac{(\frac{1}{2})x}{(\frac{n}{2})} \frac{(\frac{1}{2})^{n/2}}{\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} \exp(-\frac{1}{2}x) \\
&= \frac{x}{n} f_n(x) \\
\therefore f_n(x) &= \frac{n}{x} f_{n+2}(x) \\
f_n(x) &= \frac{x}{n-2} f_{n-2}(x)
\end{aligned}
$$

と表せる。したがって、

$$
\begin{aligned}
E[h(\chi_n^2)] &= \int_0^\infty h(x) \frac{n}{x} f_{n+2}(x) dx \\
&= n \int_0^\infty \frac{h(x)}{x} f_{n+2}(x) dx \\
&= nE[\frac{h(\chi_{n+2}^2)}{x}] \\
E[h(\chi_n^2)] &= \int_0^\infty h(x) \frac{x}{n-2} f_{n-2}(x) dx \\
&= \frac{1}{n-2} \int_0^\infty x h(x) f_{n-2}(x) dx \\
&= \frac{1}{n-2} E[\chi_{n-2}^2 h(\chi_{n-2}^2)]
\end{aligned}
$$

3.19 標準

(方針)
ベータ分布について。$X \sim Beta(\alpha, \beta)$
確率密度：

$$
\begin{aligned}
f(x) &= \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} \\
& \quad (B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt)
\end{aligned}
$$

全確率１：

$$
\begin{aligned}
f(x) &= \int_0^1 \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}dx \\
&= \frac{1}{B(\alpha,\beta)} \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx \\
&= \frac{B(\alpha,\beta)}{B(\alpha,\beta)} \\
&= 1
\end{aligned}
$$

分布関数：

$$
\begin{aligned}
F(y) &= \int_0^y \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}dx \\
&= \frac{1}{B(\alpha,\beta)} \int_0^y x^{\alpha-1}(1-x)^{\beta-1}dx
\end{aligned}
$$

上式の定積分の部分を不完全ベータ関数といい、$B_y(\alpha,\beta)$ と書く。すると分布関数は正規化された不完全ベータ関数であって、

$$
F(y) = \frac{B_y(\alpha,\beta)}{B(\alpha,\beta)}
$$

期待値、分散：

$$\begin{aligned}
E[X] &= \int_0^1 \frac{1}{B(\alpha,\beta)} x^{\alpha}(1-x)^{\beta-1}dx \\
&= \frac{B(\alpha+1,\beta)}{B(\alpha,\beta)} \int_0^1 \frac{1}{B(\alpha+1,\beta)} x^{\alpha}(1-x)^{\beta-1}dx \\
&= \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \\
&= \frac{\alpha}{\alpha+\beta} \\
E[X^2] &= \int_0^1 \frac{1}{B(\alpha,\beta)} x^{\alpha+1}(1-x)^{\beta-1}dx \\
&= \frac{B(\alpha+1,\beta)}{B(\alpha,\beta)} \\
&= \frac{\Gamma(\alpha+2)\Gamma(\beta)}{\Gamma(\alpha+\beta+2)} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \\
&= \frac{\alpha(\alpha+1)}{(\alpha+\beta+1)(\alpha+\beta)} \\
Var(X) &= \frac{\alpha(\alpha+1)}{(\alpha+\beta+1)(\alpha+\beta)} - \frac{\alpha^2}{(\alpha+\beta)^2} \\
&= \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}
\end{aligned}$$

(答案)

$$\begin{aligned}
F_Y(1-p) &= \int_0^{1-p} \frac{1}{B(n-x,x+1)} y^{n-x-1}(1-y)^x dy \\
&= \frac{n!}{(n-x-1)!x!} \int_0^{1-p} \frac{1}{n-x} (y^{n-x})'(1-y)^x dy \\
&= \frac{n!}{(n-x)!x!} \{[y^{n-x}(1-y)^x]_0^{1-p} + x\int_0^{1-p} y^{n-x}(1-y)^{x-1}dy\} \\
&= {}_nC_x p^x(1-p)^{n-x} + \frac{n!}{(n-x)!(x-1)!} \int_0^{1-p} y^{n-x}(1-y)^{x-1}dy \\
&= {}_nC_x p^x(1-p)^{n-x} + \int_0^{1-p} \frac{1}{B(n-x+1,x)} y^{n-x}(1-y)^{x-1}dy
\end{aligned}$$

ここで、数式の 1 行目を、$(1-y)$ の次数を添字として P_x と置くと、

$$\begin{aligned} P_x &= {}_nC_x p^x (1-p)^{n-x} + P_{x-1} \\ &= {}_nC_x p^x (1-p)^{n-x} + {}_nC_{x-1} p^{x-1} (1-p)^{n-x+1} + P_{x-2} \\ &\vdots \\ &= \sum_{k=0}^{x} {}_nC_k p^k (1-p)^{n-k} \\ &= F_X(x) \end{aligned}$$

以上より、題意が従う。

3.20 標準

$$\begin{aligned} P[\lambda \leq Y] &= \int_\lambda^\infty \frac{1}{\Gamma(x)} y^{x-1} \exp(-y) dy \\ &= -\frac{1}{(x-1)!} \int_\lambda^\infty y^{x-1} \{\exp(-y)\}' dy \\ &= -\frac{1}{(x-1)!} \{[y^{x-1} \exp(-y)]_\lambda^\infty - (x-1) \int_\lambda^\infty y^{x-2} \exp(-y) dy\} \\ &= \frac{\lambda^{x-1}}{(x-1)!} \exp(-\lambda) - \frac{1}{(x-2)!} \int_\lambda^\infty y^{x-2} \{\exp(-y)\}' dy \end{aligned}$$

ここで、数式の 2 行目を、y の次数を添字として P_{x-1} と置くと、

$$\begin{aligned} P_{x-1} &= \frac{\lambda^{x-1}}{(x-1)!} \exp(-\lambda) + P_{x-2} \\ &= \frac{\lambda^{x-1}}{(x-1)!} \exp(-\lambda) + \frac{\lambda^{x-2}}{(x-2)!} \exp(-\lambda) + P_{x-3} \\ &\vdots \\ &= \sum_{k-0}^{x-1} \frac{\lambda^k}{k!} \exp(-\lambda) \\ &= P[X \leq x-1] \end{aligned}$$

3.21 難

1. (方針)
   第 2 章問 7 で、

$$x = \int_0^x dt$$

なる小技を使った。また、指数分布については第 3 章問 7 に詳しい。

(答案)

$$
\begin{aligned}
r(t) &= E[X-t|X\geq t] \\
&= \int_t^\infty (x-t)P[X=x|X\geq t]dx \\
&= \int_t^\infty (x-t)\frac{f(x)}{P[X\geq t]}dx \\
&= \frac{1}{1-F(t)}\int_t^\infty (x-t)f(x)dx \\
&= \frac{1}{1-F(t)}\int_t^\infty dx\int_t^x dy\ f(x) \\
&= \frac{1}{1-F(t)}\int_t^\infty\ dy\int_y^\infty dx\ f(x) \\
&= \frac{1}{1-F(t)}\int_t^\infty (1-F(x))dx
\end{aligned}
$$

指数分布の分布関数は $F(x)=1-e^{-\lambda x}$ なので、

$$
\begin{aligned}
r(t) &= \frac{1}{1-(1-e^{-\lambda t})}\int_t^\infty \{1-(1-e^{-\lambda x})\}dx \\
&= \frac{1}{e^{-\lambda t}}\int_t^\infty e^{-\lambda x}dx \\
&= \frac{1}{e^{-\lambda t}}[-\frac{1}{\lambda}e^{-\lambda x}]_t^\infty \\
&= \frac{1}{\lambda}
\end{aligned}
$$

2. (方針)

$$
\begin{aligned}
x &= \int_0^x dt \\
x^2 &= 2\int_0^x tdt
\end{aligned}
$$

を使って、(1) を二通りに変形 ((2)、(3)) すると題意が従う。。

(答案)

$$
\begin{aligned}
E[X^2] &= \int_0^\infty x^2 f(x)dx && (1)\\
&= \int_0^\infty dt \int_0^x dx\ xf(x)\\
&= \int_0^\infty dt \int_t^\infty dx\ xf(x)\\
&= \int_0^\infty \{\int_t^\infty xf(x)dx\}dt\\
&= \int_0^\infty \{P[X \geq t]\int_t^\infty (x-t)\frac{f(x)}{P[X \geq t]}dx + \int_t^\infty tf(x)dx\}dt\\
&= \int_0^\infty (1-F(t))r(t) + t(1-F(t))dt\\
&= \int_0^\infty (1-F(t))r(t)dt + \int_0^\infty t(1-F(t))dt && (2)\\
(1) &= \int_0^\infty dx \int_0^x dt2tf(x)\\
&= \int_0^\infty dt \int_t^\infty dx\ 2tf(x)\\
&= \int_0^\infty \{2t\int_t^\infty f(x)dx\}dt\\
&= \int_0^\infty 2t(1-F(t))dt && (3)
\end{aligned}
$$

ここで、(2) = (3) より、

$$\int_0^\infty 2t(1-F(t))dt = \int_0^\infty (1-F(t))r(t)dt$$

が成り立つので、これを (2) に代入して、

$$(2) = 2\int_0^\infty (1-F(t))r(t)dt$$

より、題意が従う。また、$E[X] = r(0)$ で、$F(0) = 0$ なので、

$$2\int_0^\infty (1-F(t))r(t)dt - \{\int_t^\infty (1-F(x))dx\}^2$$

4　多次元確率変数の分布

4.1　標準

(前半までは覚えておくべき)

$$
\begin{aligned}
h(t) &= E[\{(X-\mu_X)-t(Y-\mu_Y)\}^2] \\
&= E[(X-\mu_X)^2]-2tE[(X-\mu_X)(Y-\mu_Y)]+t^2E[(Y-\mu_Y)^2] \quad (1)
\end{aligned}
$$

$h(t) \geq 0,\ E[(Y-\mu_Y)^2] > 0$ より、二次方程式 (1) の判別式を D と置くと、$D \leq 0$(解は多くとも一つ)。

$$
\begin{aligned}
&D = 4(E[(X-\mu_X)(Y-\mu_Y)])^2 - 4E[(X-\mu_X)^2]E[(Y-\mu_Y)^2] \leq 0 \\
&\therefore \{Cov(X,Y)\}^2 \leq Var(X)Var(Y)
\end{aligned}
$$

次に、等号成立条件を考える。

$(X-\mu_X)-t(Y-\mu_Y)=0$ ならば、$h(t)=E[0]=0$ より、判別式の等号は自明に成立。

一方、$Cov(X,Y)=\sqrt{Var(X)Var(Y)}$ ならば、

$$
\begin{aligned}
h(t) &= Var(X)-2tCov(X,Y)+t^2Var(Y) \\
&= Var(Y)t^2-2\sqrt{Var(X)Var(Y)}t+Var(X) \\
&= (\sqrt{Var(Y)}t-\sqrt{Var(X)})^2
\end{aligned}
$$

より、$t_0=\sqrt{Var(X)}/\sqrt{Var(Y)}$ と置くと、

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

非負関数の期待値が 0 なのだから、$(X-\mu_X)-t_0(Y-\mu_Y)=0$。$Cov(X,Y)=-\sqrt{Var(X)Var(Y)}$ でも同様。

4.2 易

1.

$$\begin{aligned}
P[Z \le z] &= P[WX + (1-W)Y \le z] \\
&= P[W=1,\ X \le z] + P[W=0,\ Y \le z] \\
&= pP[X \le z] + (1-p)P[Y \le z] \\
\therefore f_Z(z) &= pf(z) + (1-p)g(z) \\
E[Z] &= \int z f_Z(z)dz \\
&= \int z\{pf(z) + (1-p)g(z)\}dz \\
&= p\int zf(z)dz + (1-p)\int zg(z)dz \\
&= p\mu + (1-p)\mu = \mu \\
Var(Z) &= \int (z-\mu)^2 f_Z(z)dz \\
&= \int (z-\mu)^2\{pf(z) + (1-p)g(z)\}dz \\
&= p\int (z-\mu)^2 f(z)dz + (1-p)\int (z-\mu)^2 g(z)dz \\
&= p\sigma^2 + (1-p)\sigma^2 = \sigma^2
\end{aligned}$$

2.

$$\begin{aligned}
Var(U) &= Var(wX + (1-w)Y) \\
&= w^2Var(X) + 2w(1-w)Cov(X,Y) + (1-w)^2Var(Y) \\
&= w^2\sigma^2 + 2w(1-w)\rho\sigma^2 + (1-w)^2\sigma^2 \\
&= 2\sigma^2(1-\rho)w(1-w) + \sigma^2
\end{aligned}$$

より、$2\sigma^2(1-\rho) > 0$ なのだから、$Var(U)$ を最小化する w は、$w = \frac{1}{2}$

4.3 標準

1.

$$\begin{aligned}
Cov(X, Y - E[Y|X]) &= E[X(Y - E[Y|X])] - E[X]E[Y - E[Y|X]] \\
&= E[XY] - E[XE[Y|X]] - E[X]\{E[Y] - E[E[Y|X]]\} \qquad (1) \\
E[XE[Y|X]] &= \int xE[Y|X]f_X(x)dx \\
&= \int x\{\int yf_{Y|X}(y|x)dy\}f_X(x)dx \\
&= \int\int xyf_{X,Y}(x,y)dxdy \\
&= E[XY] \\
E[E[Y|X]] &= \int E[Y|X]f_X(x)dx \\
&= \int\{\int yf_{Y|X}(y|x)dy\}f_X(x)dx \\
&= \int\int yf_{X,Y}(x,y)dxdy \\
&= \int yf_Y(y)dy \\
&= E[Y] \\
\therefore (1) &= 0
\end{aligned}$$

2. (方針)

答案2行目が条件付き分散の形に近いことに気づく。

条件付き期待値の法則により、「$g(X)$ の期待値は、X で条件づけた $g(X)$ の期待値に対して X で期待値を取ったものに等しい」ので、$g(X) = (Y - E[Y|X])^2$ として考える。

(答案)

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

4.4 標準

(方針)

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\sqrt{(1-\rho^2)\sigma_1^2\sigma_2^2}} \quad \exp \quad \{-\frac{1}{2(1-\rho^2)}[(\frac{x_1-\mu_1}{\sigma_1})^2$$
$$-2\rho\frac{x_1-\mu_1}{\sigma_1}\frac{x_2-\mu_2}{\sigma 2}+(\frac{x_2-\mu_2}{\sigma_2})^2]\}$$

である。今、$X_2 = x_2$ が与えられているので、x_2 を定数扱いすることによって条件付き分布を求める。

ところで、「条件付き $\sim$」の文脈で、「$X = x$ が与えられた時」と言われたら x は定数であり、1 とか 2 とか、陽性とか陰性とか、具体的な値だと思えば良い。例えば、「$X = x$ のもとでの $g(x,Y)$ の条件付き期待値;

$$E[g(x,Y)|X=x] = \int g(x,y)f_{Y|X}(y|x)dy$$

では、x は条件づけられた値であるから、期待値は「定数」である。一方、X が与えられた時の $g(X,Y)$ の条件付き期待値;

$$E[g(X,Y)|X] = \int g(x,y)f_{Y|X}(y|x)dy$$

では、期待値は確率変数なので、期待値や分散を考えることができる。このように、いずれの「条件付き」でも変数を定数扱いするのに変わりはないが、条件づける値が具体的な定数なのか、実際は変数だがひとまず定数として見て条件づけているのか、で若干意味合いが異なる。そのため「$X = x$ を与えた時」「X を与えた時」で区別されている。

最後に、条件付き分布を求める時、

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} \\ &= \frac{f_{X,Y}(x,y)}{\int f_{X,Y}(x,y)dy} \end{aligned}$$

を考えたくなるが、$f_{X,Y}(x,y)$ の密度関数を持っているなら y を定数扱いすれば良いのでわざわざ積分したりする必要はない。類題：問 17

(答案)

$$
\begin{aligned}
& (\frac{x_1-\mu_1}{\sigma_1})^2 - 2\rho\frac{x_1-\mu_1}{\sigma_1}\frac{x_2-\mu_2}{\sigma 2} + (\frac{x_2-\mu_2}{\sigma_2})^2 \\
= \;& \frac{x_1^2 - 2\mu_1 x_1 + \mu_1^2}{\sigma_1^2} - 2\rho\frac{x_2-\mu_2}{\sigma_1\sigma_2}x_1 + 2\rho\frac{x_2-\mu_2}{\sigma_1\sigma_2} + (\frac{x_2-\mu_2}{\sigma_2})^2 \\
= \;& \frac{1}{\sigma_1^2}\{x_1^2 - 2(\mu_1 + \frac{\rho\sigma_1}{\sigma_2}(x_2-\mu_2))x_1\} + \frac{\mu_1^2}{\sigma_1^2} + 2\rho\frac{x_2-\mu_2}{\sigma_1\sigma_2}\mu_2 + (\frac{x_2-\mu_2}{\sigma_2})^2 \\
= \;& \frac{1}{\sigma_1^2}\{x_1 - (\mu_1 + \frac{\rho\sigma_1}{\sigma_2}(x_2-\mu_2))\}^2 \\
& -\frac{1}{\sigma_1^2}(\mu_1 + \frac{\rho\sigma_1}{\sigma_2}(x_2-\mu_2))^2 + \frac{\mu_1^2}{\sigma_1^2} + 2\rho\frac{x_2-\mu_2}{\sigma_1\sigma_2}\mu_2 + (\frac{x_2-\mu_2}{\sigma_2})^2 \\
= \;& \frac{1}{\sigma_1^2}\{x_1 - (\mu_1 + \frac{\rho\sigma_1}{\sigma_2}(x_2-\mu_2))\}^2 + \frac{1-\rho^2}{\sigma_2^2}(\frac{x_2-\mu_2}{\sigma_2})^2
\end{aligned}
$$

ゆえに、$f_{X_1,X_2}(x_1,x_2)$ の exp の中身は次のように変形できる。

$$
-\frac{\{x_1 - (\mu_1 + \frac{\rho\sigma_1}{\sigma_2}(x_2-\mu_2))\}^2}{2(1-\rho^2)\sigma_1^2} + (1-\rho^2)\frac{(x_2-\mu_2)^2}{2\sigma_2^2}
$$

したがって、$X_2 = x_2$ が与えられたときの X_1 の確率密度関数は、

$$
\begin{aligned}
f_{X_1|X_2}(x_1|x_2) &= \frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\{-\frac{\{x_1 - (\mu_1 + \frac{\rho\sigma_1}{\sigma_2}(x_2-\mu_2))\}^2}{2(1-\rho^2)\sigma_1^2}\} \\
&\quad \times\frac{1}{\sqrt{(1-\rho^2)\sigma_2^2}}\exp\{(1-\rho^2)\frac{(x_2-\mu_2)^2}{2\sigma_2^2}\}
\end{aligned}
$$

$\times$ 以降は定数なので、題意が従う。また、X_1, X_2 が無相関、つまり $\rho = 0$ の時、$X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$ が成り立つ。

4.5 易

(畳み込み)

$$
\begin{cases} Z = X + Y \\ W = Y \end{cases} \Leftrightarrow \begin{cases} X = Z - W \\ Y = W \end{cases}
$$

なる一対一変換を考える。

$$
|\frac{\partial(X,Y)}{\partial(Z,W)}| = det\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = 1
$$

より、

$$
\begin{aligned}
f_{Z,W}(z,w) &= f_{X,Y}(z-w,z)\times 1 \\
&= \frac{1}{2\pi\sigma_1^2}\exp\{-\frac{(z-w-\mu_1)^2}{2\sigma_1^2}\}\frac{1}{2\pi\sigma_2^2}\exp\{-\frac{(w-\mu_2)^2}{2\sigma_2^2}\} \\
&= \frac{1}{2\pi\sigma_1\sigma_2}\exp[-\frac{1}{2}\{\frac{(z-w-\mu_1)^2}{\sigma_1^2}+\frac{(w-\mu_2)^2}{\sigma_2^2}\}]
\end{aligned}
$$

ここで、

$$
\begin{aligned}
\frac{(z-w-\mu_1)^2}{\sigma_1^2}+\frac{(w-\mu_2)^2}{\sigma_2^2} &= \frac{w^2+2(\mu_1-z)w+(z-\mu_1)^2}{\sigma_1^2}+\frac{w^2-2\mu_2 w+\mu_2^2}{\sigma_2^2} \\
&= (\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2})w^2+2(\frac{\mu_1-z}{\sigma_1^2}-\frac{\mu_2}{\sigma_2^2})w+\frac{(z-\mu_1)^2}{\sigma_1^2}+\frac{\mu_2^2}{\sigma_2^2} \\
&= \frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2}\{w^2+2\frac{\sigma_2^2(\mu_1-z)-\sigma_1^2\mu_2}{\sigma_1^2+\sigma_2^2}\}+\frac{(z-\mu_1)^2}{\sigma_1^2}+\frac{\mu_2^2}{\sigma_2^2} \\
&= \frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2}\{w+\frac{\sigma_2^2(\mu_1-z)-\sigma_1^2\mu_2}{\sigma_1^2+\sigma_2^2}\}^2 \\
& \quad \frac{\sigma_2^4(\mu_1-z)^2-2\sigma_1^2\sigma_2^2\mu_2(\mu_1-z)+\sigma_1^4\mu_2^2}{\sigma_1^2\sigma_2^2(\sigma_1^2+\sigma_2^2)} \\
& \quad +\frac{(z-\mu_1)^2(\sigma_1^2+\sigma_2^2)}{\sigma_1^2(\sigma_1^2+\sigma_2^2)}+\frac{\mu_2^2(\sigma_1^2+\sigma_2^2)}{\sigma_2^2(\sigma_1^2+\sigma_2^2)} \\
&= \frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2}\{w-\frac{\sigma_2^2(z-\mu_1)+\sigma_1^2\mu_2}{\sigma_1^2+\sigma_2^2}\}^2+\frac{(z-\mu_1-\mu_2)^2}{\sigma_1^2+\sigma_2^2}
\end{aligned}
$$

$$
\begin{aligned}
\therefore f_{Z,W}(z,w) &= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2}}\exp\{-\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2\sigma_2^2}(w-\frac{\sigma_2^2(z-\mu_1)+\sigma_1^2\mu_2}{\sigma_1^2+\sigma_2^2})^2\} \\
& \quad \times\frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}\exp\{-\frac{(z-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}\}
\end{aligned}
$$

$$
\begin{aligned}
f_Z(z) &= \int f_{Z,W}(z,w)dw \\
&= \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}\exp\{-\frac{(z-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}\}
\end{aligned}
$$

したがって、$Z=X+Y\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$

(積率母関数)

$$
\begin{aligned}
M_{X+Y}(t) &= M_X(t)M_Y(t) \\
&= \exp(\mu_1 t + \frac{\sigma_1^2 t^2}{2})\exp(\mu_2 t + \frac{\sigma_2^2 t^2}{2}) \\
&= \exp\{(\mu_1+\mu_2)t + \frac{(\sigma_1^2+\sigma_2^2)t^2}{2}\} \\
\therefore X+Y &\sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)
\end{aligned}
$$

4.6 標準

(方針)
　前問で見た通り、和が有名分布になる場合は積率母関数を使う方が計算が楽だが、ここでは畳み込みで求める。

(答案)

$$
\begin{cases} Z = X+Y \\ V = Y \end{cases} \Leftrightarrow \begin{cases} X = Z-V \\ Y = V \end{cases}
$$

なる一対一変換を考える。

$$
\begin{aligned}
f_{Z,V}(z,v) &= f_X(z-v)f_Y(v) \\
\therefore f_Z(z) &= \int f_X(z-v)f_Y(v)dv = \int dv \\
& \quad (0<z-v<1,\ 0<v<1)
\end{aligned}
$$

(i)$0<z<1$ の時
　v の範囲は $0<v<z$ より、

$$
f_Z(z) = \int_0^z dv = z
$$

(ii)$1<z<2$ の時、
　v の範囲は $z-1<v<1$ より、

$$
f(z) = \int_{z-1}^1 dv = 2-z
$$

次に、

$$\begin{cases} W = XY \\ U = Y \end{cases} \Leftrightarrow \begin{cases} X = W/U \\ Y = U \end{cases}$$

なる一対一変換を考える。この時、

$$|\frac{\partial(X,Y)}{\partial(W,U)}| = det\begin{pmatrix} 1/u & -w/u^2 \\ 0 & 1 \end{pmatrix} = 1/u$$

$$\begin{aligned} f_{W,U}(w,u) &= f_X(\frac{w}{u})f_Y(u)\frac{1}{u} \\ \therefore f_W(w) &= \int f_X(\frac{w}{u})f_Y(u)\frac{1}{u}du = \int \frac{1}{u}du \\ & (0 < \frac{w}{u} < 1,\ 0 < u < 1) \end{aligned}$$

ゆえ、u の範囲は $w < u < 1$ であるから、

$$\begin{aligned} f_W(w) &= \int_w^1 \frac{1}{u}du \\ &= -\log w \end{aligned}$$

4.7　易

$$\begin{aligned} M_{X_1}(t) &= E[e^{tX_1}] \\ &= \exp(-\lambda_1)\sum_{x_1=0}^{\infty}\frac{(e^t\lambda_1)_1^x}{x_1!} \\ &= \exp\{\lambda_1(e^t-1)\} \end{aligned}$$

$$\begin{aligned} \therefore M_{X_1+...+X_n}(t) &= M_{X_1}(t)...M_{X_n}(t) \\ &= \exp\{\lambda_1(e^t-1)\}...\exp\{\lambda_n(e^t-1)\} \\ &= \exp\{(\lambda_1+...+\lambda_n)(e^t-1)\} \end{aligned}$$

以上より、$X_1 + ... + X_n \sim Po(\lambda_1 + ... + \lambda_n)$

4.8 標準

(方針)

Box-Muller 変換。2 段階に変数変換を行なっている。
逆関数の微分について。

逆関数の微分

関数 $y = f(x)$ は区間 I において微分可能で単調な関数とする。$f'(x) \neq 0$ $(x \in I)$ ならば逆関数 $x = f^{-1}(y)$ は区間 $J = f(I)$ で微分可能で

$$\frac{dx}{dy} = (\frac{dy}{dx})^{-1}$$

(証明)

$a \in I, b = f(a), y = f(x)$ と置くと、$x = f^{-1}(y)$

$$\begin{aligned} \frac{dx}{dy} &= \lim_{y \to b} \frac{f^{-1}(y) - f^{-1}(b)}{y - b} \\ &= \lim_{y \to b} (\frac{y - b}{f^{-1}(y) - f^{-1}(b)})^{-1} \\ &= \lim_{y \to b} (\frac{f(x) - f(a)}{x - a}) \end{aligned}$$

f は単調より、$y \to b$ は $X \to a$ と同値であるから、

$$\frac{dx}{dy} = (\frac{dy}{dx})^{-1}$$

が示された。

例として $Sin^{-1}x, Cos^{-1}x, Tan^{-1}x$ の微分を求める。

$$\begin{aligned} x = \sin\theta \quad &\Rightarrow \quad \frac{dx}{d\theta} = \cos\theta \\ \frac{d}{dx}Sin^{-1}x \quad &= \quad \frac{d\theta}{dx} = (\frac{dx}{d\theta})^{-1} = (\cos\theta)^{-1} \\ &= \quad \frac{1}{\cos\theta} = \frac{1}{\sqrt{1 - \sin^2\theta}} = \frac{1}{\sqrt{1 - x^2}} \end{aligned}$$

$$
\begin{aligned}
x = \tan\theta \quad &\Rightarrow \quad \frac{dx}{d\theta} = \frac{1}{\cos^2\theta} \\
\therefore \frac{d}{dx}Tan^{-1}x \quad &= \quad \frac{d\theta}{dx} = (\frac{dx}{d\theta})^{-1} = (\frac{1}{\cos^2\theta})^{-1} \\
&= \quad (1+\tan^2\theta)^{-1} = \frac{1}{1+x^2}
\end{aligned}
$$

(答案)

$$
\begin{cases} r = \sqrt{-2\log U_1} \\ \theta = 2\pi U_2 \end{cases} \Leftrightarrow \begin{cases} U_1 = \exp(-\frac{r^2}{2}) \\ U_2 = \frac{\theta}{2\pi} \end{cases}
$$

なる一対一変換を考える。

$$
\frac{\partial(U_1, U_2)}{\partial(r,\theta)} = det\begin{pmatrix} -r\exp(-\frac{r^2}{2}) & 0 \\ 0 & \frac{1}{2\pi} \end{pmatrix} = \frac{r}{2\pi}\exp(-\frac{r^2}{2})
$$

より、

$$
\begin{aligned}
f_{r,\theta}(r,\theta) \quad &= \quad f_{U_1}(\exp(-\frac{r^2}{2}))f_{U_2}(\frac{\theta}{2\pi})\frac{r}{2\pi}\exp(-\frac{r^2}{2}) \\
&= \quad \frac{r}{2\pi}\exp(-\frac{r^2}{2})
\end{aligned}
$$

次に、

$$
\begin{aligned}
\begin{cases} X = r\cos\theta \\ Y = r\sin\theta \end{cases} \quad &\Leftrightarrow \quad \begin{cases} X^2 + Y^2 = r^2(\sin^2\theta + \cos^2\theta) \\ \frac{Y}{X} = \frac{\sin\theta}{\cos\theta} = \tan\theta \end{cases} \\
&\Leftrightarrow \quad \begin{cases} r = \sqrt{X^2+Y^2} \\ \theta = Tan^{-1}(\frac{Y}{X}) \end{cases}
\end{aligned}
$$

なる一対一変換を考える。

$$
\begin{aligned}
\frac{\partial(r,\theta)}{\partial(X,Y)} \quad &= \quad det\begin{pmatrix} \frac{2X}{2\sqrt{X^2+Y^2}} & \frac{2Y}{2\sqrt{X^2+Y^2}} \\ \frac{-(\frac{Y}{X^2})}{(\frac{Y}{X})^2+1} & \frac{\frac{1}{X}}{(\frac{Y}{X})^2+1} \end{pmatrix} \\
&= \quad det\begin{pmatrix} \frac{X}{\sqrt{X^2+Y^2}} & \frac{Y}{\sqrt{X^2+Y^2}} \\ -\frac{Y}{X^2+Y^2} & \frac{X}{X^2+Y^2} \end{pmatrix} \\
&= \quad \frac{X^2+Y^2}{(X^2+Y^2)\sqrt{X^2+Y^2}} = \frac{1}{\sqrt{X^2+Y^2}}
\end{aligned}
$$

より、

$$\begin{aligned} f_{X,Y}(x,y) &= f_{r,\theta}(\sqrt{x^2+y^2}, Tan^{-1}(5-\frac{y}{x}))\sqrt{1}\sqrt{x^2+y^2} \\ &= \frac{\sqrt{x^2+y^2}}{2\pi}\exp(-\frac{x^2+y^2}{2})\frac{1}{\sqrt{x^2+y^2}} \\ &= \frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2})\frac{1}{\sqrt{2\pi}}\exp(-\frac{y^2}{2}) \end{aligned}$$

したがって、X, Y は独立に標準正規分布に従う。

4.9 標準

$$\begin{cases} X = r\cos\theta \\ Y = r\sin\theta \end{cases} \Leftrightarrow \begin{cases} r^2 = X^2+Y^2 \\ \theta = Tan^{-1}(\frac{y}{x}) \end{cases}$$

なる一対一変換を考える。

$$\begin{aligned} f_{r^2,\theta}(r^2,\theta) &= \frac{(\frac{1}{2})}{\Gamma(1)}(r^2)^{1-1}\exp(-\frac{1}{2}r^2)\frac{1}{2\pi}I(0<\theta<2\pi) \\ &= \frac{1}{2}\exp(-\frac{1}{2}r^2)\frac{1}{2\pi}I(0<\theta<2\pi) \end{aligned}$$

より、

$$\begin{aligned} \frac{\partial(r^2,\theta)}{\partial(X,Y)} &= det\begin{pmatrix} 2x & 2y \\ -\frac{y}{x^2+y^2} & \frac{x}{x^2+y^2} \end{pmatrix} \\ &= 2 \end{aligned}$$

$$\begin{aligned} \therefore f_{X,Y}(x,y) &= f_{r^2,\theta}(x^2+y^2, Tan^{-1}(\frac{y}{x}))2 \\ &= \frac{1}{2}\exp\{-\frac{1}{2}(x^2+y^2)\}\frac{1}{2\pi}2 \\ &= \frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2})\frac{1}{\sqrt{2\pi}}\exp(-\frac{y^2}{2}) \end{aligned}$$

4.10 易

$$\begin{cases} U = X+Y \\ V = X-Y \end{cases} \Leftrightarrow \begin{cases} X = \frac{U+V}{2} \\ Y = \frac{U-V}{2} \end{cases}$$

なる一対一変換を考える。

$$\begin{aligned}\frac{\partial(X,Y)}{\partial(U,V)} &= det\begin{pmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & -\frac{1}{2}\end{pmatrix}\\ &= -\frac{1}{2}\end{aligned}$$

より、

$$\begin{aligned}f_{U,V}(u,v) &= f_{X,Y}(\frac{u+v}{2},\frac{u-v}{2})\frac{1}{2}\\ &= \frac{1}{4\pi\sigma^2}\exp\{-\frac{(\frac{u+v}{2}-\mu)^2+(\frac{u-v}{2}-\xi)^2}{2\sigma^2}\}\end{aligned}$$

ここで、

$$(\frac{u+v}{2}-\mu)^2+(\frac{u-v}{2}-\xi)^2 = \frac{1}{2}\{u-(\mu+\xi)\}^2+\frac{1}{2}\{v-(\mu-\xi)\}^2$$

なので、

$$\begin{aligned}f_{U,V}(u,v) &= \frac{1}{4\pi\sigma^2}\exp\{-\frac{\{u-(\mu+\xi)\}^2+\{v-(\mu-\xi)\}^2}{4\sigma^2}\}\\ &= \frac{1}{\sqrt{2\pi(2\sigma^2)}}\exp[-\frac{\{u-(\mu+\xi)\}^2}{2(2\sigma^2)}]\frac{1}{\sqrt{2\pi(2\sigma^2)}}\exp[-\frac{\{v-(\mu-\xi)\}^2}{2(2\sigma^2)}]\end{aligned}$$

従って $U\sim N(\mu+\xi,2\sigma^2), V\sim N(\mu-\xi,2\sigma^2)$

4.11 難

(略解の方法は思いつけません)

(答案)

$$\begin{cases}W=X^2+Y^2\\ Z=Z/\sqrt{W}\end{cases}\Leftrightarrow\begin{cases}X=Z\sqrt{W}\\ Y=\pm\sqrt{W(1-Z^2)}\end{cases}$$

なる変換を考える。

$$\begin{aligned}P[W\le w,Z\le z] &= P[X\le z\sqrt{w},\ -\sqrt{w(1-z^2)}\le Y\le\sqrt{w(1-z^2)}]\\ &= \int_{-\infty}^{z\sqrt{w}}dx\int_{-\infty}^{\sqrt{w(1-z^2)}}dyf_{X,Y}(x,y)-\int_{-\infty}^{z\sqrt{w}}dx\int_{-\infty}^{-\sqrt{w(1-z^2)}}dyf_{X,Y}(x,y)\end{aligned}$$

より、

$$f_{W,Z}(w,z) = f_{X,Y}(z\sqrt{w}, \sqrt{w(1-z^2)})|\frac{\partial(X,Y)}{\partial(W,Z)}| + f_{X,Y}(z\sqrt{w}, -\sqrt{w(1-z^2)})|\frac{\partial(X,Y)}{\partial(W,Z)}|$$

である。

$$\begin{aligned}
f_{X,Y}(x,y) &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{x^2}{2})\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{y^2}{2}) \\
\frac{\partial(X,Y)}{\partial(W,Z)} &= det\begin{pmatrix} \sqrt{w} & \frac{z}{2\sqrt{w}} \\ \frac{-2wz}{2\sqrt{w(1-z^2)}} & \frac{1-z^2}{2\sqrt{w(1-z^2)}} \end{pmatrix} \\
&= \frac{1-z^2}{2\sqrt{1-z^2}} + \frac{z^2}{2\sqrt{1-z^2}} \\
&= \frac{1}{2\sqrt{1-z^2}}
\end{aligned}$$

より、

$$\begin{aligned}
f_{W,Z}(w,z) &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{wz^2}{2\sigma^2})\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{w(1-z^2)}{2\sigma^2}\}\frac{1}{\sqrt{1-z^2}} \\
&= \frac{1}{2\sigma^2}\exp(-\frac{1}{2\sigma^2}w)\frac{1}{\pi}\frac{1}{\sqrt{1-z^2}}
\end{aligned}$$

4.12 難

(方針)
$X, Y : iid \sim N(0,1)$ の時、これらの比 X/Y はコーシー分布に従う。

(証明)
$Z = \frac{X}{Y}$ と置くと、

$$\begin{aligned}
P[Z \leq z] &= P[\frac{X}{Y} \leq z,\ 0 < X,\ 0 < Y] + P[\frac{X}{Y} \leq z,\ 0 < X,\ Y < 0] \\
&\quad + P[\frac{X}{Y} \leq z,\ X < 0,\ 0 < Y] + P[\frac{X}{Y} \leq z,\ X < 0,\ Y < 0] \\
&= 2P[\frac{X}{Y} \leq z,\ 0 < X,\ 0 < Y] + 2P[\frac{X}{Y} \leq z,\ X < 0,\ 0 < Y] \quad (1) \\
&\quad (\because P[0 < Y] = P[Y < 0])
\end{aligned}$$

(i)$0 < z$ の時

$X < 0$ かつ $0 < Y$ ならば、$\frac{X}{Y} \leq z$ は自明に成り立つので、

$$\begin{aligned} P[\frac{X}{Y} \leq z,\ X < 0,\ 0 < Y] &= P[X < 0, 0 < Y] \\ &= P[X < 0]P[0 < Y] \\ &= \frac{1}{2} \times \frac{1}{2} \\ &= \frac{1}{4} \\ \therefore (1) &= 2P[\frac{X}{Y} \leq z,\ 0 < X,\ 0 < Y] + \frac{1}{2} \end{aligned}$$

ここで、

$$\begin{aligned} P[\frac{X}{Y} \leq z,\ 0 < X,\ 0 < Y] &= P[0 < X < zY, 0 < Y] \\ &= \int_0^\infty dy \int_0^{zy} dx \frac{1}{2\pi} \exp(-\frac{x^2 + y^2}{2}) \end{aligned}$$

今、$w = \frac{x}{y} \Rightarrow x = yw$ なる変換を行うと、$dx = ydw$ で、$x : 0 \to zy$ の時 $w : 0 \to z$ で、

$$\begin{aligned} \int_0^\infty dy \int_0^{zy} dx \frac{1}{2\pi} \exp(-\frac{x^2 + y^2}{2}) &= \int_0^\infty dy \int_0^z dw \frac{1}{2\pi} \exp\{-\frac{1}{2}(1 + w^2)y^2\}y \\ &= \int_0^z [-\frac{1}{2\pi} \frac{1}{1 + w^2} \exp\{-\frac{1}{2}(1 + w^2)y^2\}]_0^\infty dw \\ &= \int_0^z \frac{1}{2\pi} \frac{1}{1 + w^2} dw \end{aligned}$$

したがって、

$$\begin{aligned} F_Z(z) &= 2\int_0^z \frac{1}{2\pi} \frac{1}{1 + w^2} dw + \frac{1}{2} \\ &= \int_{-\infty}^z \frac{1}{\pi} \frac{1}{1 + w^2} dw \\ f_Z(z) &= F_Z'(z) = \frac{1}{\pi} \frac{1}{1 + z^2} \end{aligned}$$

(ii)$z \leq 0$ の時、
$0 < X$ かつ $0 < Y$ ならば、$\frac{X}{Y} \leq z$ は自明に成り立たないので、

$$\begin{aligned}
P[\frac{X}{Y} \leq z,\ 0 < X,\ 0 < Y] &= 0 \\
\therefore (1) &= 2P[\frac{X}{Y} \leq z,\ X < 0,\ 0 < Y] \\
&= 2P[X \leq zY, 0 < Y] \\
&= 2\int_0^\infty dy \int_{-\infty}^{zy} dx \frac{1}{2\pi} \exp(-\frac{x^2+y^2}{2}) \\
&= 2\int_0^\infty dy \int_{-\infty}^{z} dw \frac{1}{2\pi} \exp\{-\frac{1}{2}(1+w^2)y^2\} y \\
&= \int_{-\infty}^{z} \frac{1}{\pi}\frac{1}{1+w^2} \\
\therefore f_Z(z) &= F_Z'(z) = \frac{1}{\pi}\frac{1}{1+z^2}
\end{aligned}$$

(i),(ii) より、一般に、$X, Y : iid \sim N(0,1)$ の時、$Z = \frac{X}{Y}$ はコーシー分布に従う。

(答案)
$Z = \frac{Y}{X}$ とすると、$\frac{X}{X+Y} = \frac{1}{1+Z}$ となる。

$$W = \frac{1}{1+Z} \Leftrightarrow Z = \frac{1-W}{W}$$

なる一対一変換を考える。この時、

$$P[W \leq w] = \begin{cases} P[\frac{1-w}{w} \leq Z < -1] & (if\ w < 0) \\ \\ P[Z < -1, \vee\ \frac{1-w}{w} \leq Z] & (if\ 0 < w) \end{cases}$$

(i)$w < 0$ の時

$$\begin{aligned}
P[\frac{1-w}{w} \leq Z < -1] &= \int_{\frac{1-w}{w}}^{-1} f_Z(z)dz \\
&= \int_{-\infty}^{-1} f_Z(z)dz - \int_{-\infty}^{\frac{1-w}{w}} f_Z(z)dz
\end{aligned}$$

ゆえに、$\frac{dz}{dw} = -\frac{1}{w^2}$ に注意すると、

$$\begin{aligned} f_W(w) &= -f_Z(\frac{1-w}{w})(-\frac{1}{w^2}) \\ &= \frac{1}{\pi}\frac{1}{2w^2-2w+1} \end{aligned}$$

(ii)$0 < w$ の時

$$P[Z < -1] + P[\frac{1-w}{w} \leq Z] = \int_{-\infty}^{-1} f_Z(z)dz + 1 - \int_{-\infty}^{\frac{1-w}{w}} f_Z(z)dz$$

なので、結局

$$\begin{aligned} f_W(w) &= -f_Z(\frac{1-w}{w})(-\frac{1}{w^2}) \\ &= \frac{1}{\pi}\frac{1}{2w^2-2w+1} \end{aligned}$$

4.13 標準

(方針)

W は 2 値なので、Z, W の同時分布とは $P[Z \leq z, W = 1]$ と $P[Z \leq z, W = 0]$ に分けられる。さらに今回は Z と W は関係しているので、同値変形によって簡単化できる。

(答案)

$$
\begin{aligned}
P[Z \leq z, W=1] &= P[Z \leq z, Z = X] \\
&= P[X \leq z, X \leq Y] \\
&= \int_0^z dx \int_x^\infty dy \lambda \exp(-\lambda x) \mu \exp(-\mu y) \\
&= \int_0^z \{\lambda \exp(-\lambda x) \int_x^\infty \mu \exp(-\mu y) dy\} dx \\
&= \int_0^z \lambda \exp(-\lambda x) [-\exp(-\mu y)]_x^\infty dx \\
&= \int_0^z \lambda \exp\{-(\lambda+\mu)x\} dx \\
&= [-\frac{\lambda}{\lambda+\mu} \exp\{-(\lambda+\mu)x\}]_0^z \\
&= \frac{\lambda}{\lambda+\mu}[1-\exp\{-(\lambda+\mu)z\}] \\
P[Z \leq z, W=0] &= P[Z \leq z, Z = Y] \\
&= P[Y \leq z, Y \leq X] \\
&= \int_0^z dy \int_y^\infty dx \lambda \exp(-\lambda x) \mu \exp(-\mu x) \\
&= \frac{\mu}{\mu+\lambda}[1-\exp\{-(\mu+\lambda)z\}]
\end{aligned}
$$

4.14 易

(方針)

ベータ分布の導出を誘導付きで行う問題。ガンマ分布にそれぞれ独立に従う確率変数からベータ分布が導出できる、という事実はF分布の導出 (7章序盤) でも使うので重要。

(答案)

$$
\begin{cases} Z = X + Y \\ W = \frac{X}{X+Y} \end{cases} \Leftrightarrow \begin{cases} X = WZ \\ Y = Z(1-W) \end{cases}
$$

なる一対一変換を考える。

$$
\begin{aligned}
\frac{\partial(X,Y)}{\partial(W,Z)} &= det \begin{pmatrix} z & w \\ -z & 1-w \end{pmatrix} \\
&= z
\end{aligned}
$$

より、

$$
\begin{aligned}
f_{Z,W}(z,w) &= f_{X,Y}(wz, z(1-w))z \\
&= \frac{(\frac{1}{2})^{m_1/2}}{\Gamma(\frac{m_1}{2})}(wz)^{\frac{m_1}{2}-1}\exp(-\frac{1}{2}wz)\frac{(\frac{1}{2})^{m_2/2}}{\Gamma(\frac{m_2}{2})}\{z(1-w)\}^{\frac{m_2}{2}-1}\exp\{-\frac{1}{2}z(1-w)\} \\
&= \frac{\Gamma(\frac{m_1}{2}+\frac{m_2}{2})}{\Gamma(\frac{m_1}{2})\Gamma(\frac{m_2}{2})}w^{\frac{m_1}{2}-1}(1-w)^{\frac{m_2}{2}-1}\frac{(\frac{1}{2})^{\frac{m_1}{2}+\frac{m_2}{2}}}{\Gamma(\frac{m_1}{2}+\frac{m_2}{2})}z^{\frac{m_1}{2}+\frac{m_2}{2}-1}\exp(-\frac{1}{2}z) \\
&= \frac{1}{B(\frac{m_1}{2}+\frac{m_2}{2})}w^{\frac{m_1}{2}-1}(1-w)^{\frac{m_2}{2}-1}\frac{(\frac{1}{2})^{\frac{m_1}{2}+\frac{m_2}{2}}}{\Gamma(\frac{m_1}{2}+\frac{m_2}{2})}z^{\frac{m_1}{2}+\frac{m_2}{2}-1}\exp(-\frac{1}{2}z)
\end{aligned}
$$

したがって、W, Z は独立で、それぞれ $W \sim Beta(\frac{m_1}{2}, \frac{m_2}{2}),\ Z \sim Ga(\frac{m_1}{2} + \frac{m_2}{2}, \frac{1}{2})$

4.15 標準

1.

$$
\begin{cases}
Z_1 = \frac{X_1}{X_1+X_2} \\
Z_2 = \frac{X_1+X_2}{X_1+X_2+X_3} \\
Z_3 = X_1 + X_2 + X_3
\end{cases}
\Leftrightarrow
\begin{cases}
X_1 = Z_1 Z_2 Z_3 \\
X_2 = Z_2 Z_3 (1 - Z_1) \\
X_3 = Z_3 (1 - Z_2)
\end{cases}
$$

なる一対一変換を考える。

$$
\begin{aligned}
\frac{\partial(X_1, X_2, X_3)}{\partial(Z_1, Z_2, Z_3)} &= det\begin{pmatrix} z_2z_3 & z_1z_3 & z_1z_2 \\ -z_2z_3 & z_3(1-z_1) & z_2(1-z_1) \\ 0 & -z_3 & 1-z_2 \end{pmatrix} \\
&= z_2z_3^2(1-z_1)(1-z_2) + z_1z_2^2z_3^2 + z_1z_2z_3^2(1-z_2) + z_2^2z_3^2(1-z_1) \\
&= z_2z_3^2
\end{aligned}
$$

より、

$$
\begin{aligned}
f_{Z_1,Z_2,Z_3}(z_1,z_2,z_3) &= f_{X_1,X_2,X_3}(z_1z_2z_3, z_2z_3(1-z_1), z_3(1-z_2))z_2z_3^2 \\
&= \lambda^3\exp\{-\lambda(z_1z_2z_3 + z_2z_3 - z_1z_2z_3 + z_3 - z_2z_3)\}z_2z_3^2 \\
&= \lambda^3\exp(-\lambda z_3)z_2z_3^2
\end{aligned}
$$

同時密度が積で表されるので、確率変数 Z_1, Z_2, Z_3 は独立。

2. 同時密度より、$f_{Z_1}(z_1) \propto 1$ が成り立つ。ここで、$Z_1 = \frac{X_1}{X_1+X_2}$, $0 < X_1$, $0 < X_2$ より、$0 < Z_1 < 1$ であるから、

$$
f_{Z_1}(z_1) = I(0 < Z_1 < 1)
$$

同時密度より $f_{Z_2}(z_2) \propto z_2$ が成り立つ。ここで、$Z_1 = \frac{X_1+X_2}{X_1+X_2+X_3}$, $0 < X_1$, $0 < X_2$, $0 < X_3$ より、$0 < Z_2 < 1$ であるから、正規化定数を C と置くと、

$$\begin{aligned}\int_0^1 Cz_2dz_2 &= [\frac{C}{2}z_2^2]_0^1 \\ &= \frac{C}{2} = 1 \\ \therefore f_{Z_2}(z_2) &= 2z_2I(0 < Z_2 < 1)\end{aligned}$$

3.

$$\begin{aligned}f_{Z_1,Z_2,Z_3}(z_1,z_2,z_3) &= f_{Z_1}(z_1)f_{Z_2}(z_2)f_{Z_3}(z_3) \\ &= 2z_2f_{Z_3}(z_3) = \lambda^3\exp(-\lambda z_3)z_2z_3^2 \\ \therefore f_{Z_3}(z_3) &= \frac{\lambda^3}{2}z_3^2\exp(-\lambda z_3)\end{aligned}$$

$$\begin{aligned}\int_0^\infty \frac{\lambda^3}{2}z_3^2\exp(-\lambda z_3)dz &= \frac{\lambda^3}{2}\int_0^\infty z_3^2\{-\frac{1}{\lambda}\exp(-\lambda z_3)\}' \\ &= \frac{\lambda^2}{2}[-z_3^2\exp(-\lambda z_3)]_0^\infty - \frac{\lambda^2}{2}\int_0^\infty 2z_3\{\frac{1}{\lambda}\exp(-\lambda z_3)\}'dz \\ &= -\frac{\lambda}{2}\{[2z_3\exp(-\lambda z_3)]_0^\infty - \int_0^\infty 2\exp(-\lambda z_3)dz\} \\ &= \lambda[-\frac{1}{\lambda}\exp(-\lambda z)]_0^\infty = 1\end{aligned}$$

より、確かにこれは密度関数になっている。

4.16 易

1.

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

2.

$$\begin{aligned}Cov(X,Y) &= E[Cov(X,Y|Z)] + Cov(E[X|Z],E[Y|Z]) \\ &= E[0] + Var(Z) \\ &= 1\end{aligned}$$

4.17 標準

1.

$$
\begin{aligned}
Var(Y) &= E[Var(Y|X)] + Var(E[Y|X]) \\
&= E[1] + Var(\alpha + \beta X) \\
&= 1 + \beta^2
\end{aligned}
$$

2.

$$
\begin{aligned}
Var(X) &= 1 \\
Var(Y) &= 1 + \beta^2 \\
E[X] &= \mu \\
E[Y] &= E[E[Y|X]] \\
&= E[\alpha + \beta X] \\
&= \alpha + \beta\mu \\
E[XY] &= \int\int xyf_{X,Y}(x,y)dxdy \\
&= \int\int xyf_{Y|X}(y|x)f_X(x)dxdy \\
&= \int\{\int yf_{Y|X}(y|x)dy\}xf_X(x)dx \\
&= \int(\alpha + \beta x)xf_X(x)dx \\
&= E[\alpha X + \beta X^2] \\
&= \alpha E[X] + \beta E[X^2] \\
&= \alpha\mu + \beta(1 + \mu^2) \\
&= \beta\mu^2 + \alpha\mu + \beta \\
\therefore \rho_{X,Y} &= \frac{E[XY] - E[X]E[Y]}{\sqrt{Var(X)}\sqrt{Var(Y)}} \\
&= \frac{\beta}{\sqrt{\beta^2 + 1}}
\end{aligned}
$$

3. (方針)
$f_{X,Y}(x,y)$ を持っているのだから Y を頑張って分離する。久保川先生の略解で謎の ν が出現しているのは μ との誤植だと思う。

(答案)

$$f_{X,Y}(x,y) \;=\; \frac{1}{2\pi}\exp[-\frac{1}{2}\{(y-\alpha-\beta x)^2+(x-\mu)^2\}]$$

より、y を定数と見て exp の中身を x で平方完成する。

$$\begin{aligned}
&(y-\alpha-\beta x)^2+(x-\mu)^2\\
=\;&(1+\beta^2)x-2(\beta y-\alpha\beta+\mu)x+(y-\alpha)^2+\mu^2\\
=\;&(1+\beta^2)\{x^2-2\frac{\beta y-\alpha\beta+\mu}{1+\beta^2}x\}+(y-\alpha)^2+\mu^2\\
=\;&(1+\beta^2)(x-\frac{\beta y-\alpha\beta+\mu}{1+\beta^2})^2-\frac{(\beta y-\alpha\beta+\mu)^2}{1+\beta^2}+(y-\alpha)^2+\mu^2\\
=\;&(1+\beta^2)(x-\frac{\beta y-\alpha\beta+\mu}{1+\beta^2})^2+\frac{(y-\mu\beta-\alpha)^2}{1+\beta^2}
\end{aligned}$$

なので、

$$\begin{aligned}
f_{X,Y}(x,y) \;=\;& \frac{1}{\sqrt{2\pi/(1+\beta^2)}}\exp\{-\frac{(x-\frac{\beta y-\alpha\beta+\mu}{1+\beta^2})^2}{2/(1+\beta^2)}\}\\
&\times\frac{1}{\sqrt{2\pi(1+\beta^2)}}\exp\{\frac{(y-\mu\beta-\alpha)^2}{2(1+\beta^2)}\}
\end{aligned}$$

と表せる。$\times$ 以下は定数なので、$Y|X \sim N(\frac{\beta y-\alpha\beta+\mu}{1+\beta^2}, \frac{1}{1+\beta^2})$

4.18 標準

1.

$$\begin{cases} X = r\cos\theta \\ Y = r\sin\theta \end{cases}$$

なる一対一変換を考える。ただし、A において $0<\theta<\frac{\pi}{2}$, $0<r$ である。

$$\frac{\partial(X,Y)}{\partial(r,\theta)} \;=\; det\begin{pmatrix}\cos\theta & -r\sin\theta\\ \sin\theta & r\cos\theta\end{pmatrix}=r$$

より、

$$\begin{aligned}
f_{r,\theta}(r,\theta) \;=\;& f_{X,Y}(r\cos\theta, r\sin\theta)\times r\\
=\;& C\times\frac{g(\sqrt{r^2(\sin^2\theta+\cos^2\theta)})}{\sqrt{r^2(\sin^2\theta+\cos^2\theta)}}\times r\\
=\;& Cg(r)
\end{aligned}$$

なので、($g(r)$ は一変数関数に見えるが、$f_{r,\theta}$ の変形なので θ にも依存する。)

$$\begin{aligned}\int_0^\infty dr \int_0^{\pi/2} d\theta C g(r) &= \int_0^\infty [Cg(r)\theta]_0^{\pi/2} dr \\ &= \frac{\pi}{2} C \int_0^\infty g(r) dr \\ &= \frac{\pi}{2} C = 1 \\ \therefore C &= \frac{2}{\pi}\end{aligned}$$

2. (方針)
   W の周辺化のための積分がスーパー難しい。

   (答案)

$$\begin{cases} W = \frac{X^2}{X^2+Y^2} \\ Z = X^2 + Y^2 \end{cases} \Leftrightarrow \begin{cases} X = \sqrt{WZ} \\ Y = \sqrt{Z(1-W)} \end{cases}$$

なる一対一変換を考える。

$$\begin{aligned}\frac{\partial(X,Y)}{\partial(W,Z)} &= det \begin{pmatrix} \frac{1}{2}\sqrt{\frac{w}{z}} & \frac{1}{2}\sqrt{\frac{z}{w}} \\ \frac{1}{2}\sqrt{\frac{1-w}{z}} & -\frac{1}{2}\sqrt{\frac{z}{1-w}} \end{pmatrix} \\ &= -\frac{1}{4\sqrt{w(1-w)}}\end{aligned}$$

より、

$$\begin{aligned} f_{W,Z}(w,z) &= f_{X,Y}(\sqrt{wz}, \sqrt{z(1-w)}) \frac{1}{4\sqrt{w(1-w)}} \\ &= \frac{2}{\pi} \frac{g(\sqrt{z})}{\sqrt{z}} \frac{1}{4\sqrt{w(1-w)}} \\ &= \frac{1}{2\pi} \frac{g(\sqrt{z})}{\sqrt{z}} \frac{1}{\sqrt{w(1-w)}}\end{aligned}$$

同時密度が積で書けるのだから、W, Z は互いに独立である。W の周辺分布を求める。同時密度より、

$$f_W(w) \propto \frac{1}{\sqrt{w(1-w)}}$$

が成り立つ。

$$\int_0^1 \frac{1}{\sqrt{w-w^2}}dw = \int_0^1 \frac{1}{\sqrt{(\frac{1}{2})^2-(w-\frac{1}{2})^2}}dw \tag{1}$$

$w-\frac{1}{2}=\frac{1}{2}\sin\theta$ と置くと、$dw=\frac{1}{2}\cos\theta d\theta$ で、$w:0\to 1$ の時 $\theta:-\frac{\pi}{2}\to\frac{\pi}{2}$ なので、

$$\begin{aligned}(1) &= \int_{-\pi/2}^{\pi/2}\frac{1}{\sqrt{(\frac{1}{2})^2-(\frac{1}{2}\sin\theta)^2}}(\frac{1}{2}\cos\theta)d\theta \\ &= \int_{-\pi/2}^{\pi/2}\frac{1}{\frac{1}{4}(1-\sin^2\theta)}(\frac{1}{2}\cos\theta)d\theta \\ &= \int_{-\pi/2}^{\pi/2}\frac{1}{\frac{1}{2}\cos\theta}\frac{1}{2}\cos\theta d\theta \\ &= [\theta]_{-\pi/2}^{\pi/2}=\pi\end{aligned}$$

より、正規化定数を C と置くと、$\pi C=1\Rightarrow C=\frac{1}{\pi}$。したがって、

$$\begin{aligned}f_W(w) &= \frac{1}{\pi}\frac{1}{\sqrt{w(1-w)}} \\ f_Z(z) &= \frac{g(\sqrt{z})}{\sqrt{z}}\end{aligned}$$

4.19 易

(ベイズ推定で二項分布のパラメータの事後分布を求めるとき、パラメータの事前分布はベータ分布で与える。)

次のベータ・二項モデル；

$$\begin{aligned}X|Y &\sim Bin(n,Y) \\ Y &\sim Beta(\alpha,\beta)\end{aligned}$$

について、X の平均と分散を求める。

$$\begin{aligned} E[X] &= E[E[X|Y]] \\ &= E[nY] \\ &= \frac{n\alpha}{\alpha+\beta} \\ Var(X) &= E[Var(X|Y)] + Var(E[X|Y]) \\ &= E[nY(1-Y)] + Var(nY) \\ &= n(E[Y] - Var(Y) - E[Y]^2) + n^2 Var(Y) \\ &= nE[Y](1-E[Y]) + nVar(Y)(n-1) \\ &= n\frac{\alpha}{\alpha+\beta}\frac{\beta}{\alpha+\beta} + n(n-1)\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \\ &= \frac{n\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)} \end{aligned}$$

4.20 標準

次のガンマ・ポアソンモデル；

$$\begin{aligned} X|Y &\sim Po(Y) \\ Y &\sim Ga(\alpha, \beta) \end{aligned}$$

について、X の平均と分散を求める。

$$\begin{aligned} E[X] &= E[E[X|Y]] \\ &= E[Y] \\ &= \frac{\alpha}{\beta} \\ var(X) &= E[Var(X|Y)] + Var(E[X|Y]) \\ &= E[Y] + Var(Y) \\ &= \frac{\alpha(\beta+1)}{\beta^2} \end{aligned}$$

4.21 標準

(方針)
多変量正規分布について。$Z_1, ..., Z_n : iid \sim N(0,1)$ とすると、$Z_1, ..., Z_n$ の

同時密度は

$$
\begin{aligned}
f_{Z_1,\ldots,Z_n}(z_1,..,z_n) &= \Pi_{i=1}^{n}\frac{1}{\sqrt{2\pi}}\exp(-\frac{z_i^2}{2}) \\
&= \frac{1}{(2\pi)^{n/2}}\exp(-\frac{1}{2}\sum_{i=1}^{n} z_i^2) \\
&= \frac{1}{(2\pi)^{n/2}}\exp(-\frac{1}{2}\ {}^{T}\boldsymbol{z}\boldsymbol{z})
\end{aligned}
$$

で与えられる。今、変数変換 $X = AZ + \boldsymbol{\mu}$ を考える。

$$
\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \ldots & a_{nn} \end{pmatrix} \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}
$$

この時、A^{-1} が存在して $\boldsymbol{Z} = \boldsymbol{A}^{-1}(\boldsymbol{X} - \boldsymbol{\mu})$ で、ヤコビアンは

$$
\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}} = det(A^{-1}) = \frac{1}{|A|}
$$

なので、

$$
\begin{aligned}
f_X(\boldsymbol{x}) &= f_{\boldsymbol{Z}}(A^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))\frac{1}{|A|} \\
&= \frac{1}{(2\pi)^{n/2}}\exp[-\frac{1}{2}\ {}^{T}\{A^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\}\{A^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\}]\frac{1}{|A|} \\
&= \frac{1}{(2\pi)^{n/2}|A|}\exp\{-\frac{1}{2}\ {}^{T}(\boldsymbol{x}-\boldsymbol{\mu})\ {}^{T}(A^{-1})A^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\} \\
&= \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\exp\{-\frac{1}{2}\ {}^{T}(\boldsymbol{x}-\boldsymbol{\mu})\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\}
\end{aligned}
$$

より、$X \sim N_n(\boldsymbol{\mu}, \Sigma)$

(答案)

1.

$$
\begin{cases} U = X \\ V = Y - \rho X \end{cases} \Leftrightarrow \begin{cases} X = U \\ Y = V + \rho U \end{cases}
$$

なる一対一変換を考える。

$$\frac{\partial(X,Y)}{\partial(U,V)} = det\begin{pmatrix}1 & 0\\ \rho & 1\end{pmatrix} = 1$$

なので、

$$\begin{aligned}
f_{U,V}(u,v) &= f_{X,Y}(u, v+\rho u)\\
&= \frac{1}{2\pi\sqrt{1-\rho^2}}\exp[-\frac{1}{2(1-\rho^2)}\{u^2 - 2\rho u(\rho u + v) + (\rho u + v)^2\}]\\
&= \frac{1}{\sqrt{2\pi}}\exp(-\frac{u^2}{2})\frac{1}{\sqrt{2\pi(1-\rho^2)}}\exp\{-\frac{v^2}{2(1-\rho^2)}\}
\end{aligned}$$

より、U,V は互いに独立に分布し、$U \sim N(0,1),\ V \sim N(0,1-\rho^2)$

2.

$$\begin{aligned}
Cov(X,Y) &= Cov(U, V+\rho U)\\
&= E[U(V+\rho U)] - E[U]E[V+\rho U]\\
&= E[UV] + \rho E[U^2]\\
&= \rho\\
Var(X) &= Var(U)\\
&= 1\\
Var(Y) &= Var(V+\rho U)\\
&= Var(V) + \rho^2 Var(U) + 2\rho Cov(U,V)\\
&= 1\\
\therefore Corr(X,Y) &= \rho
\end{aligned}$$

$$
\begin{aligned}
Cov(X^2, Y^2) &= E[X^2Y^2] - E[X^2]E[Y^2] \\
E[X^2Y^2] &= E[U^2V^2] + 2\rho E[U^3V] + \rho^2 E[U^4] \\
&= E[U^2]E[V^2] + 2\rho E[U^3]E[V] + \rho^2 E[U^4] \quad (\because iid) \\
&= (1-\rho^2) + 3\rho^2 \\
&= 2\rho^2 + 1 \\
E[X^2]E[Y^2] &= E[U^2](E[V^2] + 2\rho E[VU] + \rho^2 E[U^2]) \\
&= 1 - \rho^2 + \rho^2 \\
&= 1 \\
\therefore Cov(X^2, Y^2) &= 2\rho^2 \\
Var(X^2) &= Var(U^2) \\
&= E[U^4] - (E[U^2])^2 \\
&= 3 - 1 = 2 \\
Var(Y^2) &= Var(V^2 + 2\rho UV + \rho^2 U^2) \\
&= Var(V^2) + 4\rho^2 Var(UV) + \rho^4 Var(U^2) \\
&= E[V^4] - (E[V^2])^2 + 4\rho^2(E[U^2]E[V^2] - E[UV]^2) + 2\rho^4 \\
&= 3(1-\rho^2)^2 - (1-\rho^2)^2 + 4\rho^2(1-\rho^2) + 2\rho^4 \\
&= 2 \\
\therefore Corr(X^2, Y^2) &= \rho^2
\end{aligned}
$$

3.

$$
\begin{aligned}
X^2 - 2\rho XY + Y^2 &= U^2 - 2\rho U(V + \rho U) + (V + \rho U)^2 \\
&= (1-\rho^2)U^2 + V^2 \\
\therefore \frac{X^2 - 2\rho XY + Y^2}{1-\rho^2} &= U^2 + \tfrac{V^2}{1-\rho^2}
\end{aligned}
$$

$U \sim N(0,1), \frac{V}{\sqrt{1-\rho^2}} \sim N(0,1)$ なので、$U^2 + \frac{V^2}{1-\rho^2} \sim \chi_2^2$

4.22 標準

1.

$$
\begin{aligned}
& \int_0^1 dx \int_0^{1-x} dy x^{a-1} y^{b-1} (1-x-y)^{c-1} \\
= \ & \int_0^1 x^{a-1}(1-x)^{c-1} \{ \int_0^{1-x} y^{b-1} (1 - \frac{y}{1-x})^{c-1} dy \} dx \qquad (1) \\
& (\because 1 - x - y = (1-x)(1 - \frac{y}{1-x}))
\end{aligned}
$$

ここで、$\frac{y}{1-x} = u$ と置くと、$dy = (1-x)du$ で、$y : 0 \to 1-x$ の時 $u : 0 \to 1$ なので、

$$\int_0^{1-x} y^{b-1}(1-\frac{y}{1-x})^{c-1}dy = \int_0^1 u^{b-1}(1-u)^{c-1}(1-x)^b du$$

$$\begin{aligned}
\therefore (1) &= \int_0^1 x^{a-1}(1-x)^{b+c-1}dx \int_0^1 u^{b-1}(1-u)^{c-1}du \\
&= B(a,b+c)B(b,c) \\
&= \frac{\Gamma(a)\Gamma(b+c)}{\Gamma(a+b+c)}\frac{\Gamma(b)\Gamma(c)}{\Gamma(b+c)} \\
&= \frac{\Gamma(a)\Gamma(b)\Gamma(c)}{\Gamma(a+b+c)}
\end{aligned}$$

2.

$$\begin{aligned}
f_X(x) &\propto \int_0^{1-x} x^{a-1}y^{b-1}(1-x-y)^{c-1}dy \\
&= \int_0^{1-x} x^{a-1}(1-x)^{c-1}y^{b-1}(1-\frac{y}{1-x})^{c-1}dy \qquad (1)
\end{aligned}$$

ここで、$u = \frac{y}{1-x}$ と置くと、$dy = (1-x)du$ で、$y : 0 \to 1-x$ の時、$u : 0 \to 1$ なので、

$$\begin{aligned}
(1) &= \int_0^1 x^{a-1}(1-x)^{b+c-1}u^{b-1}(1-u)^{c-1}du \\
&\propto x^{a-1}(1-x)^{b+c-1} \\
\therefore f_X(x) &= \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b+c)}x^{a-1}(1-x)^{b+c-1}
\end{aligned}$$

このように、$X \sim Beta(a,b+c)$ である。

$$\begin{aligned}
f_Y(y) &\propto \int_0^{1-y} x^{a-1}y^{b-1}(1-y-x)^{c-1}dx \\
&= \int_0^{1-y} x^{a-1}y^{b-1}(1-y)^{c-1}(1-\frac{x}{1-y})^{c-1}dx \qquad (2)
\end{aligned}$$

ここで、$t = \frac{x}{1-y}$ と置くと、$dx = (1-y)dt$ で、$x : 0 \to 1-y$ の時

$4t: 0 \to 1$ なので、

$$\begin{aligned}(2) &= \int_0^1 y^{b-1}(1-y)^{a+c-1}t^{a-1}(1-t)^{c-1}dt \\ &\propto y^{b-1}(1-y)^{a+c-1} \\ \therefore f_Y(y) &= \frac{\Gamma(a+b+c)}{\Gamma(b)\Gamma(a+c)}y^{b-1}(1-y)^{a+c-1}\end{aligned}$$

このように、$Y \sim Beta(b, a+c)$ である。

3.

$$\begin{aligned}f_{Y|X}(y|x) &= \frac{f_{X,Y}(x,y)}{f_X(x)} \\ &= \frac{\Gamma(a)\Gamma(b+c)\Gamma(a+b+c)x^{a-1}y^{b-1}(1-x-y)^{c-1}}{\Gamma(a+b+c)x^{a-1}(1-x)^{b+c-1}\Gamma(a)\Gamma(b)\Gamma(c)} \\ &= \frac{\Gamma(b+c)}{\Gamma(b)\Gamma(c)}\frac{1}{1-x}(\frac{y}{1-x})^{b-1}(1-\frac{y}{1-x})^{c-1}\end{aligned}$$

4. まずは $E[XY]$ を求める。$E[XY]$ の定数部分以外の積分は、

$$\begin{aligned}&\int_0^1 dx \int_0^{1-x} dy x^a y^b (1-x-y)^{c-1} \\ =& \int_0^1 \{\int_0^{1-x} x^a y^b (1-x)^{c-1}(1-\frac{y}{1-x})^{c-1}dy\}dx \qquad (3)\end{aligned}$$

ここで、$u = \frac{y}{1-x}$ と置くと、$dy = (1-x)du$ で、$y: 0 \to 1-x$ の時、$u: 0 \to 1$ であるから、

$$\begin{aligned}&\int_0^{1-x} x^a y^b (1-x)^{c-1}(1-\frac{y}{1-x})^{c-1}dy \\ =& \int_0^1 x^a u^b (1-x)^{b+c}(1-u)^{c-1}du\end{aligned}$$

となる。したがって

$$\begin{aligned}(3) &= \int_0^1 x^a(1-x)^{b+c}dx \int_0^1 u^b(1-u)^{c-1}du \\ &= \frac{\Gamma(a+1)\Gamma(b+c+1)}{\Gamma(a+b+c+2)}\frac{\Gamma(b+1)\Gamma(c)}{\Gamma(b+c+1)}\end{aligned}$$

以上より、

$$
\begin{aligned}
E[XY] &= \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)}\frac{\Gamma(a+1)\Gamma(b+1)\Gamma(c)}{\Gamma(a+b+c+2)} \\
&= \frac{ab}{(a+b+c+1)(a+b+c)} \\
E[X] &= \frac{a}{a+b+c} \\
E[Y] &= \frac{b}{a+b+c} \\
Cov(X,Y) &= \frac{ab}{(a+b+c+1)(a+b+c)} - \frac{ab}{(a+b+c)^2} \\
&= -\frac{ab}{(a+b+c)^2(a+b+c+1)} \\
Var(X) &= \frac{a(b+c)}{(a+b+c)^2(a+b+c+1)} \\
Var(Y) &= \frac{b(a+c)}{(a+b+c)^2(a+b+c+1)} \\
\therefore \rho_{X,Y} &= -\sqrt{\frac{ab}{(a+c)(b+c)}}
\end{aligned}
$$

4.23　標準

1. (方針)
　x_k を定数扱い作戦だと正規化が難しい。多項分布において、各 x_k の周辺分布は二項分布になっていることを使う。

   (答案)

$$
f_{X_k}(x_k) = \frac{n!}{x_k!(n-x_k)!}p_k^{x_k}(1-p_k)^{n-x_k}
$$

なので、

$$\begin{aligned}
& f_{X_1,...,X_{k-1}|X_k}(x_1,...,x_{k-1}|x_k) \\
= & \frac{f_{X_1,...,X_k}(x_1,...,c_k)}{f_{X_k}(x_k)} \\
= & \frac{n!p_1^{x_1}...p_k^{x_k}x_k!(n-x_k)!}{x_1!...x_k!n!p_k^{x_k}(1-p_k)^{n-x_k}} \\
= & \frac{(n-x_k)!}{x_1!...x_{k-1}!}\frac{p_1^{x_1}...p_{k-1}^{x_{k-1}}}{(1-p_k)^{n-x_k}} \\
= & \frac{(n-x_k)!}{x_1!...x_{k-1}!}(\frac{p_1}{1-p_k})^{x_1}...(\frac{p_{k-1}}{1-p_k})^{x_{k-1}} \\
& (\because x_1+...+x_k=n)
\end{aligned}$$

4.24 標準

(注)
行列 X の転置を ${}^TX$ と書く。
モーメント母関数で、

$$M_X(\boldsymbol{t}) = E[e^{{}^T\boldsymbol{t}X}]$$

である。

(方針)

$$\begin{aligned}
{}^T\boldsymbol{t}X &= \begin{pmatrix} t_1 & \dots & t_k \end{pmatrix}\begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix} \\
&= t_1X_1+...+t_kX_k
\end{aligned}$$

となるので、

$$\begin{aligned}
E[e^{{}^T\boldsymbol{t}X}] &= E[e^{t_1X_1}...e^{t_kX_k}] \\
&= \Pi_{i=1}^k E[e^{t_iX_i}]
\end{aligned}$$

ここで、$E[e^{t_iX_i}]$ は正規分布 $N(\mu_i,\sigma_i^2)$ に従う確率変数 X_i のモーメント母関数なので、

$$\Pi_{i=1}^k E[e^{t_iX_i}] = \Pi_{i=1}^k \exp(\mu_it_i+\frac{\sigma_i^2t_i}{2})$$

である。ところが、exp の中身に項が 2 つ出てしまってよくわからない。そこで、一番最初に X を標準化した確率変数ベクトル Z を導入 ($Z_1, ..., Z_k$ は標準正規分布に従うのでモーメント母関数の exp の中身が単純) する工夫を施している。

(答案)

$Z = \Sigma^{-1/2}(X - \boldsymbol{\mu})$ と置くと、$X = \Sigma^{1/2}Z + \boldsymbol{\mu}$ で、

$$Z \sim N(\Sigma^{-1/2}\boldsymbol{\mu} - \Sigma^{-1/2}\boldsymbol{\mu}, \Sigma^{-1/2}\Sigma^T\Sigma^{-1/2}) = N(\mathbf{0}, I_k)$$

であるから、

$$\begin{aligned} M_X(\boldsymbol{t}) &= E[e^{{}^T\boldsymbol{t}X}] \\ &= E[e^{{}^T\boldsymbol{t}\Sigma^{1/2}Z + {}^T\boldsymbol{t}\boldsymbol{\mu}}] \\ &= E[e^{{}^T\boldsymbol{t}\Sigma^{1/2}Z}]e^{\boldsymbol{t}\boldsymbol{\mu}} \end{aligned} \tag{1}$$

ここで、${}^T\boldsymbol{s} = {}^T\boldsymbol{t}\Sigma^{1/2}$ と置くと、

$$\begin{aligned} E[e^{{}^T\boldsymbol{s}Z}] &= E[e^{s_1Z_1 + ... + s_kZ_k}] \\ &= \Pi_{i=1}^k E[e^{s_iZ_i}] \\ &= \Pi_{i=1}^k \exp\frac{s_i^2}{2} \\ &= \exp(\sum_{i=1}^k \frac{s_i^2}{2}) \\ &= \exp(\frac{1}{2}\ {}^T\boldsymbol{s}\boldsymbol{s}) \\ &= \exp\{\frac{1}{2}\ {}^T\boldsymbol{t}\Sigma^{1/2}\ {}^T({}^T\boldsymbol{t}\Sigma^{1/2})\} \\ &= \exp(\frac{{}^T\boldsymbol{t}\Sigma\boldsymbol{t}}{2}) \\ \therefore (1) &= \exp(\frac{{}^T\boldsymbol{t}\Sigma\boldsymbol{t}}{2})\exp({}^T\boldsymbol{t}\boldsymbol{\mu}) \\ &= \exp({}^T\boldsymbol{\mu}\boldsymbol{t} + \frac{{}^T\boldsymbol{t}\Sigma\boldsymbol{t}}{2}) \end{aligned}$$

5 標本分布とその近似

5.1

(方針)

$E[\bar{X} + \sqrt{n}S] = E[\bar{X}] + \sqrt{n}E[S]$ としたいところだが、$E[S^2] = \sigma^2$ は知っ

ているものの S^2 の確率密度関数はわからないので、変数変換によって $E[S]$ を知ることはできない。そこで、分布がわかっている形 (カイ二乗分布) が使えるように変形する。カイ二乗分布について。

$$Z_1, ..., Z_n : iid \sim N(0,1) \Rightarrow Z_1^2 + ... + Z_n^2 \sim \chi_n^2$$

(証明)
まずは、$Z_1^2 \sim \chi_1^2$ を示す。$Y_1 = Z_1^2$ と置くと、

$$\begin{aligned} P[Y_1 \leq y] &= P[Z_1^2 \leq y] \\ &= P[-\sqrt{y} \leq Z_1 \leq \sqrt{y}] \\ &= \int_{-\infty}^{\sqrt{y}} f_{Z_1}(z_1) dz_1 - \int_{-\infty}^{-\sqrt{y}} f_{Z_1}(z_1) dz_1 \\ \therefore f_{Y_1}(y) &= f_{Z_1}(\sqrt{y}) \frac{1}{2\sqrt{y}} - f_{Z_1}(-\sqrt{y})(-\frac{1}{2\sqrt{y}}) \\ &= \frac{1}{\sqrt{y}} \frac{1}{\sqrt{2\pi}} \exp(-\frac{y}{2}) \\ &= \frac{(\frac{1}{2})^{1/2}}{\Gamma(\frac{1}{2})} y^{\frac{1}{2}-1} \exp(-\frac{1}{2}y) \end{aligned}$$

より、$Z_1^2 \sim Ga(\frac{1}{2}, \frac{1}{2}) = \chi_1^2$ が示された。カイ二乗分布の再生性 ($X_1, ..., X_n : iid \sim Ga(\alpha, \beta)$ として、$X_1 + ... + X_n$ のモーメント母関数を計算することで $X_1 + ... + X_n : iid \sim Ga(n\alpha, \beta)$ が分かる。よって $Z_1^2, ..., Z_n^2 = Y_1, ..., Y_n : iid \sim \chi_1^2$ の時、$Z_1^2 + ... + Z_n^2 = Y_1 + ... + Y_n : iid \sim \chi_n^2$ が言える) より、$Z_1^2 + ... + Z_n^2 \sim \chi_n^2$ となる。

標本がカイ二乗分布に従う例

1. 母平均 μ がわかっている場合

$$\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^2}=\sum_{i=1}^{n}(\frac{x_i-\mu}{\sigma^2})^2\sim\chi_n^2$$

2. 母平均 μ がわからない場合

$$\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma^2}=\frac{nS^2}{\sigma^2}=\frac{(n-1)V^2}{\sigma^2}\sim\chi_{n-1}^2$$

ただし、S^2 は標本分散、V^2 は不偏標本分散。

(答案)

$$M=\frac{\bar{X}-\mu}{\sigma},\ T=\frac{nS^2}{\sigma^2}$$

と置くと、

$$\begin{aligned}E[\bar{X}+\sqrt{n}S] &= E[\sigma M+\mu+\sigma\sqrt{T}]\\ &= \mu+\sigma E[\sqrt{T}]\end{aligned}$$

と表せる。$T\sim\chi_{n-1}^2$ より、

$$\begin{aligned}E[\sqrt{T}] &= \int_0^\infty\frac{(\frac{1}{2})^{n-1/2}}{\Gamma(\frac{n-1}{2})}t^{\frac{n}{2}-1}\exp(-\frac{1}{2}t)dt\\ &= \frac{(\frac{1}{2})^{n-1/2}}{\Gamma(\frac{n-1}{2})}\frac{\Gamma(\frac{n}{2})}{(\frac{1}{2})^{n/2}}\int_0^\infty\frac{(\frac{1}{2})^{n/2}}{\Gamma(\frac{n}{2})}t^{\frac{n}{2}-1}\exp(-\frac{1}{2}t)dt\\ &= \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}\sqrt{2}\end{aligned}$$

これを C と置くと、$E[\bar{X}+\sqrt{n}S]=\mu+\sigma C$。

5.2 難

(方針)
尺度混合とは、ある分布の分散が確率変数になっている場合のことである。

正規尺度混合

$$\begin{cases} X|Y = y \sim N(\mu, y) \\ Y \sim g(y),\ y > 0 \end{cases}$$

なる階層モデルを考えると、

$$f_X(x) = \int_0^\infty \frac{1}{\sqrt{2\pi y}} \exp\{-\frac{(x-\mu)^2}{2y}\} g(y) dy$$

(証明)

$$\begin{aligned} f_X(x) &= \int_0^\infty f_{X,Y}(x,y)dy \\ &= \int_0^\infty f_{X|Y}(x|y)g(y)dy \\ &= \int_0^\infty \frac{1}{\sqrt{2\pi y}} \exp\{-\frac{(x-\mu)^2}{2y}\} g(y) dy \end{aligned}$$

自由度 m の t 分布の密度関数は、次のような正規尺度混合によって導出できる。

$$\begin{cases} X|Y = y \sim N(0, \frac{m}{y}) \\ Y \sim \chi^2_m = Ga(\frac{m}{2}, \frac{1}{2}) \end{cases}$$

X の周辺分布が自由度 m の t 分布。

t 分布

$Z \sim N(0,1),\ Y \sim \chi^2_m$ であるとき、

$$\frac{Z}{\sqrt{Y/m}} \sim t_m$$

(答案)
$X = \frac{Z}{\sqrt{Y/m}}$ と置くと (= T だけど T で進めると表記が混乱するので X に

した)、

$$\begin{cases} X|Y = y \sim N(0, \frac{m}{y}) \\ Y \sim \chi^2_m \end{cases}$$

である。

$$\begin{aligned}
P[\frac{Z}{\sqrt{Y/m}} \leq t] &= P[X \leq t] \\
&= \int_0^t f_X(x)dx \\
&= \int_0^t \{\int_0^\infty f_{X,Y}(x,y)dy\}dx \\
&= \int_0^t \{\int_0^\infty f_{X|Y}(x|y)f_Y(y)dy\}dx \\
&= \int_0^\infty \{\int_0^t f_{X|Y}(x|y)f_Y(y)dx\}dy \\
&= \int_0^\infty \{\int_0^t P[X|Y=y]dx\}f_Y(y)dy \\
&= \int_0^\infty P[X \leq t|Y=y]f_Y(y)dy \\
&= \int_0^\infty P[Z \leq t\frac{\sqrt{y}}{\sqrt{m}}]f_Y(y)dy \qquad (1)
\end{aligned}$$

次に、(1) を微分して T $(= X)$ の確率密度関数を求める。

$$\begin{aligned}
(1) &= \int_0^\infty \{\int_{-\infty}^{\sqrt{\frac{y}{m}}t} \frac{1}{\sqrt{2\pi}} \exp(-\frac{z^2}{2}) dz\} f_Y(y) dy \\
&= \int_{-\infty}^{\sqrt{\frac{y}{m}}t} dz \int_0^\infty dy \frac{1}{\sqrt{2\pi}} \exp(-\frac{z^2}{2}) f_Y(y) \\
\therefore f_T(t) &= \frac{d}{dt}(1) \\
&= \int_0^\infty \sqrt{\frac{y}{2\pi m}} \exp(-\frac{t^2}{2m} y) \frac{(\frac{1}{2})^{m/2}}{\Gamma(\frac{m}{2})} y^{\frac{m}{2}-1} \exp(-\frac{1}{2} y) dy \\
&= \frac{(\frac{1}{2})^{m/2}}{\sqrt{\pi m}} \times \frac{1}{\Gamma(\frac{m}{2})} \int_0^\infty y^{\frac{m+1}{2}-1} \exp\{-(\frac{t^2}{2m} + \frac{1}{2}) y\} dy \\
&= \frac{(\frac{1}{2})^{m/2}}{\sqrt{\pi m}} \frac{\Gamma(\frac{m+1}{2})}{\Gamma(\frac{m}{2})} \frac{1}{(\frac{t^2}{2m} + \frac{1}{2})^{\frac{m+1}{2}}} \\
&= \frac{1}{\sqrt{\pi m}} \frac{\Gamma(\frac{m+1}{2})}{\Gamma(\frac{m}{2})} (\frac{t^2}{m} + 1)^{-\frac{m+1}{2}}
\end{aligned}$$

となる。これは自由度 m の t 分布の密度関数。

5.3 難

$$\begin{cases} X|Y = y \sim N(0, \frac{m}{y}) \\ Y \sim \chi^2_m = Ga(\frac{m}{2}, \frac{1}{2}) \end{cases}$$

なる正規尺度混合において、X の周辺分布が自由度 m の t 分布なので、

$$
\begin{aligned}
E[X] &= E^Y[E[X|Y]] \\
&= E^Y[0] \\
&= 0 \\
Var(X) &= E^Y[Var(X|Y)] + Var^Y(E[X|Y]) \\
&= E[\frac{m}{Y}] + Var(0) \\
&= mE[\frac{1}{Y}] \\
&= m\int_0^\infty \frac{(\frac{1}{2})^{m/2}}{\Gamma(\frac{m}{2})} y^{\frac{m}{2}-1-1}\exp(-\frac{1}{2}y)dy \\
&= m\frac{(\frac{1}{2})^{m/2}}{\Gamma(\frac{m}{2})}\frac{\Gamma(\frac{m}{2}-1)}{(\frac{1}{2})^{m/2-1}} \\
&= \frac{m}{2}\frac{1}{\frac{m}{2}-1} \\
&= \frac{m}{m-2}
\end{aligned}
$$

また、

$$
E[|X|^n] = E^Y[E[|X|^n|Y]]
$$

であるから、

$$
\begin{aligned}
E[|X|^n|Y=y] &= \int_{-\infty}^{\infty} |x|^n \frac{\sqrt{y}}{\sqrt{2\pi m}}\exp(-\frac{yx^2}{2m})dx \\
&= 2\int_0^\infty x^n \sqrt{\frac{y}{2\pi m}}\exp(-\frac{yx^2}{2m})dx \\
&= \frac{2y^{1/2}}{\sqrt{2\pi m}}\int_0^\infty x^n \exp(-\frac{y}{2m}x^2)dx \qquad (1)
\end{aligned}
$$

ここで、$z=x^2$ と置くと、$dx=\frac{1}{2}z^{-\frac{1}{2}}dz$ なので、

$$
\begin{aligned}
(1) &= \frac{2y^{1/2}}{\sqrt{2\pi m}}\int_0^\infty z^{\frac{n-1}{2}}\exp(-\frac{y}{2m}z)dz \\
&= \frac{2y^{1/2}}{\sqrt{2\pi m}}\int_0^\infty z^{\frac{n+1}{2}-1}\exp(-\frac{y}{2m}z)dz \\
&= \frac{2}{\sqrt{\pi}}(\frac{y}{2m})^{1/2}\frac{\Gamma(\frac{n+1}{2})}{(\frac{y}{2m})^{\frac{n+2}{2}}} \\
&= \frac{2\Gamma(\frac{n+1}{2})}{\sqrt{\pi}}(\frac{1}{2m})y^{-\frac{n}{2}}
\end{aligned}
$$

上式の $y^{-\frac{n}{2}}$ 以外の部分を c と置くと、$E[|X|^n]=E[cY^{-\frac{n}{2}}]$ であるから、

$$
\begin{aligned}
E[cY^{-\frac{n}{2}}] &= c\int_0^\infty \frac{(\frac{1}{2})^{m/2}}{\Gamma(\frac{m}{2})}y^{\frac{m-n}{2}-1}\exp(-\frac{1}{2}y)dy \\
&= \frac{c(\frac{1}{2})^{m/2}}{\Gamma(\frac{m}{2})}\int_0^\infty y^{\frac{m-n}{2}}\exp(-\frac{1}{2}y)dy
\end{aligned}
$$

積分部分は $Ga(\frac{m-n}{2},\frac{1}{2})$ の確率密度関数の核である。ガンマ分布の確率密度関数は、パラメータ α が正ならば収束するので、モーメント母関数が存在するためには、$m>n$(自由度 $>n$) である必要がある。

5.4 標準

(方針)
F 分布を導出する。$F_{m,n}$ は、$X\sim\chi^2_m,\ Y\sim\chi^2_n$ の時 $\frac{X/m}{Y/n}$ が従う分布である。(7 章でも導出した)

$$
\begin{cases} X\sim Ga(\alpha_1,\beta) \\ Y\sim Ga(\alpha_2,\beta) \end{cases} \Leftrightarrow \begin{cases} Z=X+Y\sim Ga(\alpha_1+\alpha_2,\beta) \\ W=\frac{X}{X+Y}\sim Beta(\alpha_1,\beta_1) \end{cases}
$$

である。今、$R=X/Y$ と置くと W は、

$$
W=\frac{X}{X+Y}=\frac{(\frac{X}{Y})}{(\frac{X}{Y})+1}=\frac{R}{R+1}, \qquad \frac{dW}{dR}=\frac{1}{(R+1)^2}
$$

なので、

$$\begin{aligned} f_R(r) &= f_W(\frac{r}{1+r})\frac{1}{(1+r)^2} \\ &= \frac{1}{B(\alpha_1,\alpha_2)}(\frac{r}{1+r})^{\alpha_1-1}(\frac{1}{1+r})^{\alpha_2-1}(\frac{1}{1+r})^2 \\ &= \frac{1}{B(\alpha_1,\alpha_2)}r^{\alpha_1-1}(1+r)^{-(\alpha_1+\alpha_2)} \end{aligned}$$

これは、Scaled beta distribution(以下 SB) の確率分布である。このように、$X \sim Ga(\alpha_1,\beta),\ Y \sim Ga(\alpha_2,\beta)$ の時、$\frac{X}{Y} \sim SB(\alpha_1,\alpha_2)$ である。このことを用いると、

$$\begin{cases} X \sim \chi^2_m = Ga(\frac{m}{2},\frac{1}{2}) \\ X \sim \chi^2_n = Ga(\frac{n}{2},\frac{1}{2}) \end{cases} \Rightarrow R = \frac{X}{Y} \sim SB(\frac{m}{2},\frac{n}{2})$$

今、$S = X/m,\ T = Y/n$ と置くと、

$$R = \frac{X}{Y} = \frac{mS}{nT} \sim SB(\frac{m}{2},\frac{n}{2})$$

今求めたいのは、$\frac{S}{T}$ の確率密度関数。

$$Z = \frac{S}{T} = \frac{n}{m}R, \quad \frac{dR}{dZ} = \frac{m}{n}$$

なる変数変換で、Z の確率密度関数がもとまる。

$$\begin{aligned} f_Z(z) &= f_R(\frac{m}{n}z)\frac{m}{n} \\ &= \frac{1}{B(\frac{m}{2},\frac{n}{2})}(\frac{m}{n}z)^{\frac{m}{2}-1}(1+\frac{n}{m}z)^{-(\frac{m}{2}+\frac{n}{2})}(\frac{m}{n}) \\ &= \frac{(\frac{m}{n})^{m/2}}{B(\frac{m}{2},\frac{n}{2})}z^{\frac{m}{2}-1}(1+\frac{m}{n}z)^{-(\frac{m}{2}+\frac{n}{2})} \end{aligned}$$

F 分布

$X \sim \chi_m^2,\ Y \sim \chi_n^2$ の時、

$$\frac{X/m}{Y/n} \sim F(m,n)$$

また、$X_1,...,X_m \sim N(\mu,\sigma^2),\ Y_1,...,Y_n \sim N(\mu,\sigma^2)$ の時、

$$\frac{V_X^2}{V_Y^2} = \frac{(m-1)\frac{V_X^2}{\sigma^2}/(m-1)}{(n-1)\frac{V_Y^2}{\sigma^2}/(n-1)} = \frac{\chi_{m-1}^2/(m-1)}{\chi_{n-1}^2/(n-1)} \sim F(m-1,n-1)$$

(答案)
$Z \sim \chi_m^2,\ W \sim \chi_n^2$ と置くと、

$$Y = \frac{Z/m}{W/n}$$

と書けるので、Y^r の期待値は、

$$\begin{aligned} E[Y^r] &= E[(\frac{Z/m}{W/n})^r] \\ &= (\frac{n}{m})^r E[(\frac{Z}{W})^r] \\ &= (\frac{n}{m})^r E[Z^r]E[\frac{1}{W^r}] \end{aligned}$$

ここで、

$$\begin{aligned} E[Z^r] &= \int_0^\infty \frac{(\frac{1}{2})^{m/2}}{\Gamma(\frac{m}{2})} z^{(\frac{m}{2}+r)-1} \exp(-\frac{1}{2}z)dz \\ &= \frac{(\frac{1}{2})^{m/2}}{\Gamma(\frac{m}{2})} \frac{\Gamma(\frac{m}{2}+r)}{(\frac{1}{2})^{\frac{m}{2}+r}} \\ &= \frac{\Gamma(\frac{m}{2}+r)}{\Gamma(\frac{m}{2})} 2^r \\ E[\frac{1}{W^r}] &= \int_0^\infty \frac{(\frac{1}{2})^{n/2}}{\Gamma(\frac{n}{2})} w^{(\frac{n}{2}-r)-1} \exp(-\frac{1}{2}w)dw \\ &= \frac{(\frac{1}{2})^{n/2}}{\Gamma(\frac{n}{2})} \frac{\Gamma(\frac{n}{2}-r)}{(\frac{1}{2})^{\frac{n}{2}-r}} \\ &= \frac{\Gamma(\frac{n}{2}-r)}{\Gamma(\frac{n}{2})} \frac{1}{2^r} \end{aligned}$$

であるから、

$$\begin{aligned}E[Y^r] &= (\frac{n}{m})^r \frac{\Gamma(\frac{m}{2}+r)}{\Gamma(\frac{m}{2})} 2^r \frac{\Gamma(\frac{n}{2}-r)}{\Gamma(\frac{n}{2})} \frac{1}{2^r} \\ &= (\frac{n}{m})^r \frac{\Gamma(\frac{m}{2}+\frac{n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \frac{\Gamma(\frac{m}{2}+r)\Gamma(\frac{n}{2}-r)}{\Gamma(\frac{m}{2}+\frac{n}{2})} \\ &= (\frac{n}{m})^r \frac{B(\frac{m}{2}+r,\frac{n}{2}-r)}{B(\frac{m}{2},\frac{n}{2})}\end{aligned}$$

5.5 標準

$$\begin{aligned}P[Z \le z] &= 1 - P[z \le Z] \\ &= 1 - P[z \le Z \wedge z \le Y] \\ &= 1 - P[z \le Z]P[z \le Y] \\ &= 1 - (1 - P[X \le z])^2 \\ \therefore f_Z(z) &= 2(1 - P[Z \le z])f_X(z)\end{aligned}$$

さて、$T = Z^2$ と置くと、

$$\begin{aligned}P[T \le t] &= P[Z^2 \le t] \\ &= \int_{-\infty}^{\sqrt{t}} f_Z(z)dz - \int_{-\infty}^{-\sqrt{t}} f_Z(z)dz \\ \therefore f_T(t) &= f_Z(\sqrt{t})\frac{1}{2\sqrt{t}} - f_Z(-\sqrt{t})(-\frac{1}{2\sqrt{t}}) \\ &= \frac{1}{\sqrt{t}}(1 - P[X \le \sqrt{t}])f_X(\sqrt{t}) + \frac{1}{\sqrt{t}}(1 - P[X \le -\sqrt{t}])f_X(-\sqrt{t}) \\ &= \frac{1}{\sqrt{t}} f_X(\sqrt{t}) \qquad \because P[X \le -\sqrt{t}] = 1 - P[X \le \sqrt{t}],\ f_X(-\sqrt{t}) = f_X(\sqrt{t}) \\ &= \frac{1}{\sqrt{t}}\frac{1}{\sqrt{2\pi}}\exp(-\frac{t}{2}) \\ &= \frac{(\frac{1}{2})^{1/2}}{\Gamma(\frac{1}{2})} t^{\frac{1}{2}-1}\exp(-\frac{1}{2}t)\end{aligned}$$

5.6 標準

(方針)

Marcov の不等式

$$
\begin{aligned}
Y \geq 0 \wedge E[Y] < \infty \quad &\Rightarrow \quad P[Y \geq c] \leq \frac{E[Y]}{c} \\
&\Leftrightarrow \quad E[Y] \geq cP[Y \geq c]
\end{aligned}
$$

(確率を下から押さえられる！)
(期待値を上から押さえられる！)

(証明)

$$
\begin{aligned}
E[Y] &= E[Y\{I(Y \geq c) + I(Y < c)\}] \\
&= E[YI(Y \geq c)] + E[YI(Y \leq c)] \\
&\geq E[YI(Y \geq c)] \\
&\geq E[cI(Y \geq c)] \\
&= cE[I(Y \geq c)] = cP[Y \geq c]
\end{aligned}
$$

チェビシェフの不等式

$E[X] = \mu$ かつ $Var(X) = \sigma^2$ の時、

$$P[|X - \mu| \geq k] \leq \frac{\sigma^2}{k^2}$$

$k = m\sigma$ とすると、

$$P[|X - \mu| \geq m\sigma] \leq \frac{1}{m^2}$$

(証明)
マルコフの不等式 $P[Y \geq c] \leq \frac{E[Y]}{c}$ において、$Y = |X - \mu|^2,\ c = k^2$ と置くと、

$$P[|X - \mu|^2 \geq k^2] \leq \frac{E[|X - \mu|^2]}{k^2} = \frac{\sigma^2}{k^2}$$

1.

$$\begin{aligned} E[|Z|I(|Z| \geq t)] &\geq E[tI(|Z| \geq t)] \\ &= tE[I(|Z| \geq t)] \\ &= tP[|Z| \geq t] \end{aligned}$$

より、

$$\begin{aligned} P[|Z| \geq t] \leq E[|Z|I(|Z| \geq t)] &= \frac{1}{t}\{\int_t^\infty z\frac{1}{\sqrt{2\pi}}\exp(-\frac{z^2}{2})dz + \int_{-\infty}^{-t} -z\frac{1}{\sqrt{2\pi}}\exp(-\frac{z^2}{2})dz \\ &= \frac{1}{t}\{[-\frac{1}{\sqrt{2\pi}}\exp(-\frac{z^2}{2})]_t^\infty - [-\frac{1}{\sqrt{2\pi}}\exp(-\frac{z^2}{2})]_{-\infty}^{-t}\} \\ &= \frac{2}{t}\phi(t) \end{aligned}$$

2. ($X > 0$ が条件に必要だと思う)

$$\begin{aligned} M_Z(t) &= E[e^{tX}] \\ &= E[e^{tX}\{I(X \geq a) + I(X \leq a)\}] \\ &\geq E[e^{tX}I(X \geq a)] \\ &\geq e^{at}E[I(X \geq a)] \\ &= e^{at}P[X \geq a] \end{aligned}$$

5.7 標準

$$\begin{aligned} E[g(X)] &= E[g(X)\{I(|X| \geq t) + I(t \leq |X|)\}] \\ &\geq E[g(X)I(|X| \geq t)] \\ &= E[g(X)I(X \leq -t) + g(X)I(t \leq X)] \\ &= E[g(X)I(X \leq -t)] + E[g(X)I(t \leq X)] \\ &\geq E[g(-t)I(X \leq -t)] + E[g(t)I(t \leq X)] \\ &= g(-t)E[I(X \leq -t)] + g(t)E[I(t \leq X)] \\ &= g(-t)P[X \leq -t] + g(t)P[t \leq X] \\ &= g(t)P[|X| \geq t] \end{aligned}$$

5.8 標準

(方針)

確率収束

確率変数の列 $U_1, ..., U_n$ が確率収束するとは、任意の $\epsilon > 0$ について、

$$\lim_{n\to\infty} P[|U_n - U| < \epsilon] = 1$$
$$(\lim_{n\to\infty} P[|U_n - U| > \epsilon] = 0)$$

が成り立つことを言い、

$$U_n \to_p U \ as \ n \to \infty$$

と書く。

(答案)

任意の ϵ について、

$$P[|\frac{2}{n(n+1)}\sum_{j=1}^{n} jX_j - \mu|^2 \geq \epsilon] \leq \frac{E[|\frac{2}{n(n+1)}\sum_{j=1}^{n} jX_j - \mu|^2]}{\epsilon}$$

であるから、

$$\lim_{n\to\infty} E[|\frac{2}{n(n+1)}\sum_{j=1}^{n} jX_j - \mu|^2] = 0$$

を示せば良い。

$$\begin{aligned}
E[(\frac{2}{n(n+1)}\sum_{j=1}^{n} jX_j - \mu)^2] &= E[\{\frac{2}{n(n+1)}(\sum_{j=1}^{n} jX_j - \frac{n(n+1)}{2}\mu)\}^2] \\
&= E[\frac{4}{n^2(n+1)^2}\{\sum_{j=1}^{n} j(X_j - \mu)\}^2] \\
&= \frac{4}{n^2(n+1)^2} E[\{\sum_{j=1}^{n} j(X_j - \mu)\}^2] \\
&= \frac{4}{n^2(n+1)^2} \sum_{j=1}^{n} j^2 E[(X_j - \mu)^2] \\
&= \frac{4\sigma^2}{n^2(n+1)^2} \frac{n(n+1)(2n+1)}{6} \\
&\to 0 \quad as \ \ n \to \infty
\end{aligned}$$

5.9 標本平均

(方針)
標本平均は確率変数列。

1.

$$
\begin{aligned}
\frac{X_{n+1}+n\bar{X}_n}{n+1} &= \frac{1}{n+1}\sum_{i=1}^{n+1}X_i \\
&= \bar{X}_{n+1}
\end{aligned}
$$

2. (1) を用いて、

$$
\begin{aligned}
nV_{n+1}^2 &= \sum_{i=1}^{n+1}(X_i-\bar{X}_{n+1})^2 \\
&= \sum_{i=1}^{n+1}(X_i-\frac{X_{n+1}+n\bar{X}_n}{n+1})^2 \\
&= \sum_{i=1}^{n}(X_i-\frac{X_{n+1}+n\bar{X}_n}{n+1})^2+(X_{n+1}-\frac{X_{n+1}+n\bar{X}_n}{n+1})^2 \\
&= \sum\{(X_i-\bar{X}_n)-(\frac{X_{n+1}-\bar{X}_n}{n+1})\}^2+\{\frac{n}{n+1}(X_{n+1}-\bar{X}_n)\}^2 \\
&= \sum_{i=1}^{n}(X_i-\bar{X}_n)^2-2\frac{X_{n+1}-\bar{X}_n}{n+1}\sum_{i=1}^{n}(X_i-\bar{X}_n)+n(\frac{X_{n+1}-\bar{X}_n}{n+1})^2 \\
&\quad +\{\frac{n}{n+1}(X_{n+1}-\bar{X}_n)\}^2 \\
&= (n-1)V_n^2+\frac{n}{n+1}(X_{n+1}-\bar{X}_n)^2
\end{aligned}
$$

5.10 標準

(方針)
収束に関する使えるやつ。

大数法則 (LLN)

$$X_1, ..., X_n : iid \sim (\mu, \sigma^2) \ \Rightarrow \ \bar{X} \to_p \mu \ as \ n \to \infty$$

(証明)

$$P[|\bar{X} - \mu| \geq \epsilon] \leq \frac{E[|\bar{X} - \mu|^2]}{\epsilon^2} = \frac{Var(\bar{X})}{\epsilon^2} = \frac{1}{\epsilon^2}\frac{\sigma^2}{n} \to 0$$

分布収束

確率変数列 $U_1, ..., U_n$ が U に分布収束するとは、任意の U_n の台上の実数 x について、

$$P[U_n \leq x] \to P[U \leq x] \ as \ n \to \infty$$

が成り立つことをいう。この定義は、

$$M_{U_n}(t) \to M_U(t) \ as \ n \to \infty$$

と同値である。(CLT の証明で使う。)

中心極限定理 (CLT)

$$X_1, ..., X_n : iid \sim (\mu, \sigma^2) \ \Rightarrow \ \sqrt{n}(\bar{X} - \mu) \to_d Z, \ Z \sim N(0, \sigma^2)$$

(証明)

$\mu = 0$ の時、$\sqrt{n}\bar{X} \to_d Z \sim N(0, \sigma^2)$ を示す。$M_{X_1}(t) = ... = M_{X_n}(t) =$

$M(t)$ と置くと、

$$\begin{aligned} M_{\sqrt{n}\bar{X}}(t) &= E[e^{\sqrt{n}\bar{X}t}] \\ &= E[\exp(\frac{t}{\sqrt{n}}\sum_{i=1}^{n} X_i)] \\ &= \Pi_{i=1}^{n} E[e^{\frac{t}{\sqrt{n}} X_i}] \\ &= \{M(\frac{t}{\sqrt{n}})\}^n \end{aligned}$$

ここで、$M(x)$ に対して $x = 0$ まわりでテイラーの定理を適用すると、

$$M(x) = M(0) + M'(o)x + \frac{M''(o)}{2!}x^2 + \frac{M'''(x^*)}{3!}x^3$$

ただし $0 < x^* < x$。今、$M(0) = E[e^{0X}] = 1,\ M'(0) = \mu = 0,\ M'''(0) = \sigma^2 + \mu^2 = \sigma^2$ であるから、x を $\frac{t}{\sqrt{n}}$ に置き換えて、

$$\begin{aligned} M(\frac{t}{\sqrt{n}}) &= 1 + \frac{\sigma^2}{2}(\frac{t}{\sqrt{n}})^2 + \frac{M'''(x^*)}{3!}(\frac{t}{\sqrt{n}})^3 \\ &= 1 + \frac{\frac{\sigma^2}{2}t^2 + \frac{M'''(x^*)}{6}\frac{t^3}{\sqrt{n}}}{n} \\ \therefore \{M(\frac{t}{\sqrt{n}})\}^n &= \{1 + \frac{\frac{\sigma^2}{2}t^2 + \frac{M'''(x^*)}{6}\frac{t^3}{\sqrt{n}}}{n}\}^n \\ &\to \exp(\frac{\sigma^2}{2}t^2) \qquad as\ n \to \infty \end{aligned}$$

$\sqrt{n}\bar{X}$ のモーメント母関数が標準正規分布のモーメント母関数へと収束するので、$\sqrt{n}\bar{X}$ は標準正規分布へと分布収束する。

※ CLT より、

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

は自明 (定数倍しただけなので)。一方、S を標本分散として、

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

は自明でない (S は確率変数なので)。これを示すためには、次のスラツキーの定理が必要。

スラツキーの定理

$U_n \to_d U$、$V_n \to_p a$ とする。

1. $U_n + V_n \to_d U + a$
2. $V_n U_n \to_d aU$

$$\begin{cases} \sqrt{n}(\bar{X} - \mu) \to_d Z \sim N(0, \sigma^2) \\ S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \to_p \sigma^2 \end{cases}$$

を用いて、スラツキーの定理より、

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S} \to \frac{1}{\sigma}Z \sim N(0, 1)$$

本問は、S^2 が σ^2 に確率収束するのは本当？というもの。標本平均の形に近いが、中身に確率変数が二つ (X と $\bar{X}$) 含まれているのをまずはなんとかする。

(答案)

1.

$$\begin{aligned} V_n^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \\ &= \frac{1}{n-1}\{\sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2\} \\ &= \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 - \frac{n}{n-1}(\bar{X} - \mu)^2 \end{aligned}$$

最終行の第 2 項は大数法則により 0 に収束する。第 1 項については、$Y = (X - \mu)^2$ と置くと、

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

と言える。ここで、

$$\begin{aligned} E[Y] &= E[(X-\mu)^2] = \sigma^2 \\ Var(Y) &= E[(X-\mu)^4] - \{E[(X-\mu)^2]\}^2 = \mu_4 - \sigma^4 \end{aligned}$$

したがって、大数法則により $\bar{Y} \to_p \sigma^2$ なので、$V_n^2 \to_p \sigma^2$

2.

$$\sqrt{n}(V_n^2 - \sigma^2) = \sqrt{n}\{\frac{n}{n-1}\bar{Y} - \frac{n}{n-1}(\bar{X}-\mu)^2 - \sigma^2\}$$

右辺第 2 項は大数法則により 0 に収束する。ゆえ、小問 1 より $Y_1, ..., Y_n : iid \sim (\sigma^2, \mu_4 - \sigma^4)$ なので、

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

5.11 標本

(方針)
大数法則は $X_1, ..., X_n$ が iid でないと成り立たないのでここでは使えない。独立標本でなくてもサイズ n が大きくなって標本間の相関が小さくなれば平均値が母平均に近づく、ということを示す問題。大数法則の導出とやることは一緒。

(答案)
任意の $\epsilon > 0$ について、

$$P[|\bar{X} - \mu| > \epsilon] \le \frac{E[|\bar{X} - \mu|^2]}{\epsilon^2}$$

より、$E[(\bar{X}-\mu)^2] \to 0$ を示せば良い。

$$
\begin{aligned}
E[(\bar{X}_n-\mu)^2] &= E[(\frac{1}{n}\sum_{i=1}^{n} X_i-\mu)^2] \\
&= E[\{\frac{1}{n}(\sum_{i=1}^{n} X_i-n\mu)\}^2] \\
&= \frac{1}{n^2}E[(\sum_{i=1}^{n} X_i-n\mu)^2] \\
&= \frac{1}{n^2}E[\{\sum_{i=1}^{n}(X_i-\mu)\}^2] \\
&= \frac{1}{n^2}\sum_{i=1}^{n} E[(X_i-\mu)^2]+\frac{1}{n^2}\sum_{i=1}^{n}\sum_{i\neq j} E[(X_i-\mu)(X_j-\mu)] \\
&= \frac{1}{n^2}\sum_{i=1}^{n} E[(X_i-\mu)^2]+\frac{1}{n^2}\sum_{i=1}^{n}\sum_{i\neq j} Cov(X_i,X_j) \\
&= \frac{\sigma^2}{n}+\frac{n(n-1)\rho_n\sigma^2}{n^2} \\
&\to 0 \quad as\ n\to\infty\ (\because \rho_n\to 0)
\end{aligned}
$$

5.12 難

デルタ法について。

デルタ法

θ を定数とし、定数数列 a_n は無限大に発散するとする。また、$a_n(U_n-\theta)\to_d U$ で、g は連続微分可能かつ $g'(\theta)\neq 0$ とする。この時、

$$a_n(g(U_n)-g(\theta))\to_d g'(\theta)U$$

(証明)
$g(U_n)$ に $U_n=\theta$ 周りでテイラーの定理を適用する。

$$g(U_n)=g(\theta)+g'(U_n^*)(U_n-\theta)$$

ただし、U_n^* は U_n と θ の間。また、

$$\begin{aligned}
a_n(U_n-\theta) \to_d U \ \wedge \ a_n \to \infty &\Rightarrow U_n-\theta \to_p 0 \\
&\Rightarrow U_n \to_p \theta \\
&\Rightarrow U_n^* \to_p \theta \quad (n \to \infty)
\end{aligned}$$

なので、

$$\begin{aligned}
a_n(g(U_n)-g(\theta)) &= a_n\{g(\theta)+g'(U_n^*)(U_n-\theta)-g(\theta)\} \\
&= g'(U_n^*)a_n(U_n-\theta) \\
&\to_d g'(\theta)U
\end{aligned}$$

さて、本問について、$Y_n \to_p \log p$ を示すのはとっても難しいので置いておく。後半についての方針は、
$X_n \sim Bin(n,p)$ より、$\frac{X_n}{n} \sim Ber(p)$ なので、中心極限定理により、

$$\sqrt{n}(\frac{X_n}{n}-p) \to_d U \sim N(0,p(1-p))$$

今、$g(\frac{X_n}{n})=\log\frac{X_n}{n}$ と置き、g に $\frac{X_n}{n}=p$ まわりでテーラーの定理を適用すると、

$$g(\frac{X_n}{n})=g(p)+g'(\frac{X_n^*}{n})(\frac{X_n}{n}-p)$$

ただし、$\frac{X_n^*}{n}$ は p と $\frac{X_n}{n}$ の間。また、

$$\begin{aligned}
\sqrt{n}(\frac{X_n}{n}-p) \to_d U,\ \sqrt{n} \to \infty &\Rightarrow \frac{X_n}{n} \to_p p \\
&\Rightarrow \frac{X_n^*}{n} \to_p p
\end{aligned}$$

したがって、

$$\begin{aligned}
\sqrt{n}(g(\frac{X_n}{n})-g(p)) &= \sqrt{n}(g(p)+g'(\frac{X_n^*}{n})(\frac{X_n}{n}-p)-g(p)) \\
&= g'(\frac{X_n^*}{n})\sqrt{n}(\frac{X_n}{n}-p) \\
&\to_d g'(p)U
\end{aligned}$$

$g'(p) = \frac{1}{p}$ より、

$$\sqrt{n}(\log \frac{X_n}{n} - \log p) \to_d N(0, \frac{1-p}{p})$$

となる。ただし、今回は $\log \frac{X_n}{n}$ が $X_n = 0$ では定義されないことに気をつける。

(答案)
まずは、$Y_n \to_p \log p$、すなわち $\lim_{n\to\infty} P[|Y_n - \log p| > \epsilon] = 0$ を示す。

$$\begin{aligned}
\lim_{n\to\infty} P[|Y_n - \log p| > \epsilon] &= \lim_{n\to\infty} \{P[|\log \frac{X_n}{n} - \log p| > \epsilon,\ 1 \le X_n] \\
&\quad + P[|1 - \log p| > \epsilon,\ X_n = 0]\} \\
&= \lim_{n\to\infty} P[|\log \frac{X_n}{np}| > \epsilon,\ 1 \le X_n] \\
&\quad (\because P[X_n = 0] = \binom{n}{0} p^0 (1-p)^n \to 0)
\end{aligned}$$

ここで、

$$\begin{aligned}
|\log \frac{X_n}{np}| \le \epsilon &\Leftrightarrow -\epsilon \le \log \frac{X_n}{np} \le \epsilon \\
&\Leftrightarrow e^{-\epsilon} \le \frac{X_n}{np} \le e^{\epsilon} \\
&\Leftrightarrow pe^{-\epsilon} \le \frac{X_n}{n} \le pe^{\epsilon} \\
&\Leftrightarrow -p(1 - e^{-\epsilon}) \le \frac{X_n}{n} - p \le p(e^{\epsilon} - 1)
\end{aligned}$$

ゆえ、$\delta = \min\{p(1 - e^{-\epsilon}),\ p(e^{\epsilon} - 1)\}$ と置くと、

$$|\frac{X_n}{n} - p| \le \delta \Rightarrow |\log \frac{X_n}{np}| \le \epsilon$$

が成り立つ。対偶を取ることにより、

$$|\log \frac{X_n}{n} - \log p| > \epsilon \Rightarrow |\frac{X_n}{n} - p| > \delta$$

したがって、

$$
\begin{aligned}
P[|\log \frac{X_n}{n} - \log p| > \epsilon,\ 1 \le X_n] &\le P[|\frac{X_n}{n} - p| > \delta,\ 1 \le X_n] \\
&\le P[|\frac{X_n}{n} - p| > \delta|]
\end{aligned}
$$

X_n/n は p に収束するので、$P[|X_n/n - p| > \delta] \to 0$ ゆえ、Y_n は $\log p$ に確率収束する。

次に、$\sqrt{n}(Y_n - \log p) \to_d N(0, \frac{1-p}{p})$ を示す。

$$
\begin{aligned}
\sqrt{n}(Y_n - \log p) &= \sqrt{n}(\log \frac{X_n}{n} - \log p) I(1 \le X_n) \\
&\quad + \sqrt{n}(1 - \log p) I(X_n = 0) \\
&= \sqrt{n}(\log \frac{X_n}{n} - \log p) \quad (\because P[X_n = 0] \to 0)
\end{aligned}
$$

以下、(方針) の通りに題意は示される。

※デルタ法の応用：

分散安定化変換

$X_1, ..., X_n : iid \sim (\mu, \sigma^2)$ の時、中心極限定理により、

$$\sqrt{n}(\bar{X} - \mu) \to_d N(0, \sigma^2)$$

である。この時、σ^2 を漸近的分散と呼ぶ。今、デルタ法により

$$\sqrt{n}(g(\bar{X}) - g(\mu)) \to_d N(0, g'(\mu)^2 \sigma^2)$$

となる。漸近的分散 $g'(\mu)^2\sigma^2$ が定数になるように g を取ることを分散安定化変換と言う。

(Ex)

$$X_1, ..., X_n : iid \sim Po(\lambda) \ \Rightarrow \ \sqrt{n}(\bar{X} - \lambda) \to_d N(0, \lambda)$$

この時、λ の漸近的 $1-\alpha$ 信頼区間は、

$$\lambda \in [\bar{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}, \bar{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}] = [\bar{X} - z_{\alpha/2}\sqrt{\frac{\lambda}{n}}, \bar{X} + z_{\alpha/2}\sqrt{\frac{\lambda}{n}}\ ]$$

これは、信頼区間の中にパラメータ λ が入っているため解くのが大変。デルタ法により、

$$\sqrt{n}(g(\bar{X}) - g(\lambda)) \to_d N(0, g'(\lambda)^2\lambda)$$

さて、$g'(\lambda)^2\lambda = 1$ となるような g は $g(\lambda) = 2\sqrt{\lambda}$ であるから、

$$\sqrt{n}(2\sqrt{\bar{X}} - 2\sqrt{\lambda}) \to_d N(0,1)$$
$$\Leftrightarrow \sqrt{n}(\sqrt{\bar{X}} - \sqrt{\lambda}) \to_d N(0, \frac{1}{4})$$

したがって、λ の区間推定量は

$$\sqrt{\lambda} \in [\sqrt{\bar{X}} - \frac{z_{\alpha/2}}{2\sqrt{n}}, \sqrt{\bar{X}} + \frac{z_{\alpha/2}}{2\sqrt{n}}\ ]$$

5.13 標準

(方針)
順序統計量について。

最大値の分布

$$\begin{aligned}
F_{X_{(n)}}(x) &= P[X_{(n)} \le x] \\
&= P[X_{(1)} \le x, ..., X_{(n)} \le x] \\
&= \Pi_{i=1}^{n} P[X_i \le x] \\
&= F(x)^n \\
f_{X_{(n)}}(x) &= nF(x)^{n-1} f(x)
\end{aligned}$$

最小値の分布

$$\begin{aligned}
F_{X_{(1)}}(x) &= P[X_{(1)} \le x] \\
&= 1 - P[X_{(1)} > x] \\
&= 1 - P[X_{(1)} > x, ..., X_{(n)} > x] \\
&= 1 - (1 - F(x))^n \\
f_{X_{(1)}}(x) &= nf(x)(1 - F(x))^{n-1}
\end{aligned}$$

最大値、最小値の同時分布

$$
\begin{aligned}
F_{X_{(1)},X_{(n)}}(x,y) &= P[X_{(1)} \le x, X_{(n)} \le y] \\
&= P[X_{(n)} \le y] - P[X_{(1)} > x, X_{(n)} \le y] \\
&\quad (P[A \cap B] = P[B] - P[A^c \cap B]) \\
&= (F(y))^n - P[x \le X_{(1)} \le y, ..., x \le X_{(n)} \le y] \\
&= (F(y))^n - (F(y) - F(x))^n \\
f_{X_{(1)},X_{(n)}}(x,y) &= n(n-1)f(x)f(y)(F(y) - F(x))^{n-2}
\end{aligned}
$$

二次元連続分布の分布関数と密度関数には以下のような関係があるので、それを用いた。

$$
\begin{aligned}
F_{X,Y}(x,y) &= \int_{-\infty}^{y} dv \int_{-\infty}^{x} du f_{X,Y}(u,v) \\
f_{X,Y}(x,y) &= \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)
\end{aligned}
$$

最大値、最小値の同時分布は、例えば $X_{(n)} - X_{(1)}$(最大値と最小値の差、range) を調べたいときなどに使える。

$X_{(k)}$ の分布

「$X_{(k)} \leq x$」は、「$X_1, ..., X_n$ のうち少なくとも k 個が x 以下」であることと同値。これは、

ちょうど k 個が x 以下で、$n-k$ 個が x より大きい。${}_nC_k$ 通り
ちょうど $k+1$ 個が x 以下で $n-(k+1)$ 個が x より大きい。${}_nC_{k+1}$ 通り
$\vdots$
ちょうど n 個が x 以下。${}_nC_n$ 通り

と場合分けできるので、

$$
\begin{aligned}
F_{X_{(k)}}(x) &= P[X_{(k)} \leq x] \\
&= \sum_{j=k}^{n} {}_nC_j F(x)^j (1-F(x))^{n-j}
\end{aligned}
$$

1. $k=1, k=n$ を代入すると、$X_{(1)}, X_{(k)}$ の分布に一致する。
2. $X_{(k_1)}, X_{(k_2)}$ の同時分布などが計算できる。

分布関数を微分して確率密度関数を求める。

$$
\begin{aligned}
f_{X_{(k)}}(x) &= \frac{d}{dx}\sum_{j=k}^{n} {}_nC_j F(x)^j (1-F(x))^{n-j} \\
&= \sum_{j=k}^{n} {}_nC_j \{ jF(x)^{j-1}(1-F(x))^{n-j} f(x) \\
&\quad -(n-j)F(x)^j (1-F(x))^{n-j-1} f(x) \} \\
&= \sum_{j=k}^{n} {}_nC_j jF(x)^{j-1}(1-F(x))^{n-j} f(x) \\
&\quad -\sum_{j=k}^{n-1} {}_nC_j (n-j)F(x)^j (1-F(x))^{n-j-1} f(x) \\
&= {}_nC_k kF(x)^{j-1}(1-F(x))^{n-j} f(x) \\
&\quad +\sum_{j=k+1}^{n} {}_nC_j jF(x)^{j-1}(1-F(x))^{n-j} f(x) \\
&\quad -\sum_{j=k}^{n-1} {}_nC_j (n-j)F(x)^j (1-F(x))^{n-j-1} f(x) \\
&= \frac{n!}{(k-1)!(n-k)!} F(x)^{j-1}(1-F(x))^{n-j} f(x) \\
&\quad +\sum_{j=k}^{n-1} {}_nC_{j+1} (j+1)F(x)^j (1-F(x))^{n-j-1} f(x) \\
&\quad -\sum_{j=k}^{n-1} {}_nC_j (n-j)F(x)^j (1-F(x))^{n-j-1} f(x)
\end{aligned}
$$

最終行の第 2 項と第 3 項について、

$$
{}_nC_{j+1}(j+1) = \frac{n!}{j!(n-j-1)!}, \quad {}_nC_j(n-j) = \frac{n!}{j!(n-j-1)!}
$$

なので、

$$
f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!} F(x)^{j-1}(1-F(x))^{n-j} f(x)
$$

1. 最大値と最小値の同時密度は、

$$\begin{aligned} f_{X_{(1)},X_{(n)}}(x,y) &= n(n-1)\lambda\exp(-\lambda x)\lambda\exp(-\lambda y)(-\exp(-\lambda y)+1+\exp(-\lambda x)-1)^{n-2} \\ &= n(n-1)\lambda^2\exp\{-\lambda(x+y)\}(\exp(-\lambda x)-\exp(-\lambda y))^{n-2} \end{aligned}$$

また、$X_{(i)}$ の確率密度関数は、

$$f_{X_{(i)}}(x) = \frac{n!}{(i-1)!(n-i)!}\{-\exp(-\lambda x)+1\}^{j-1}\{-\exp(-\lambda x)\}^{n-j}\lambda\exp(-\lambda x)$$

2. (標本平均関係ないし CLT ではなさそうなので分布収束の定義に沿って計算する)

$$\begin{aligned} P[U_n \le x] &= P[\lambda X_{(n)} - \log n \le x] \\ &= P[X_{(n)} \le \frac{x+\log n}{\lambda}] \\ &= [-\exp\{-\lambda(\frac{x+\log n}{\lambda})\}+1]^n \\ &= \{-\exp(-x-\log n)+1\}^n \\ &= \{1+\frac{-\exp(-x)}{n}\}^n \\ &\to \exp\{-\exp(-x)\} \qquad as\ n\to\infty \end{aligned}$$

ゆえ $U_n \to_d Gumbel(0,1)$ である。ガンベル分布は任意の確率変数の最大値が漸近的に従う分布。

5.14 標準

(方針)
$X_{(1)},...,X_{(n)}$ の同時確率密度関数は、

$$f_{X_{(1)},...,X_{(n)}}(x_1,...,x_n) = n!f(x_1)...f(x_n)$$

(答案)

$$\begin{cases} Y_1 = nX_{(1)} \\ \\ Y_2 = (n-1)(X_{(2)} - X_{(1)}) \\ \\ Y_3 = (n-2)(X_{(3)} - X_{(2)}) \\ \vdots \\ Y_n = X_{(n)} - X_{(n-1)} \end{cases} \Leftrightarrow \begin{cases} X_{(1)} = \frac{Y_1}{n} \\ \\ X_{(2)} = \frac{Y_2}{n-1} + \frac{Y_1}{n} \\ \\ X_{(3)} = \frac{Y_3}{n-2} + \frac{Y_2}{n-1} + \frac{Y_1}{n} \\ \vdots \\ X_{(n)} = Y_n + \frac{Y_{n-1}}{2} + ... + \frac{Y_1}{n} \end{cases}$$

なる一対一変換を考える。

$$\begin{aligned} \frac{\partial(x_{(1)}, ..., x_{(n)})}{\partial(y_1, ..., y_n)} &= det \begin{pmatrix} \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ 0 & \frac{1}{n-1} & \frac{1}{n-1} & & \vdots \\ \vdots & 0 & \frac{1}{n-2} & & \\ & \vdots & 0 & \ddots & \\ & & \vdots & & \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \\ \\ &= \frac{1}{n!} \end{aligned}$$

したがって、

$$\begin{aligned} f_{Y_1,...,Y_n}(y_1, ..., y_n) &= f_{X_{(1)},...,X_{(n)}}(y_1, ..., y_n)\frac{1}{n!} \\ &= f(y_1)...f(y_n) \end{aligned}$$

5.15 標準

(ハザード関数については 3.5 に詳しい。)

1.

$$
\begin{aligned}
\lambda_1(x)+...+\lambda_n(x) &= \frac{nf(x)}{1-F(x)} \\
&= \frac{nf(x)(1-F(x))^{n-1}}{1-\{1-(1-F(x))^n\}} \\
&= \frac{f_{X_{(1)}}(x)}{1-F_{X_{(1)}}(x)} \\
&= \lambda_{X_{(1)}}(x)
\end{aligned}
$$

6 統計的推定

6.1 標準

(方針)

指数型分布族について。定義だけ眺めると ??? ってなるから本問でたくさん計算してなんとなく何がしたいのかわかってね、と言うことだと思う。

指数型分布族

確率関数が、

$$f(x|\theta) = h(x)c(\theta)\exp\{\sum_{i=1}^{k} w_i(\theta)t_i(x)\}$$

の形で表せるとき、確率変数 X の従う分布を指数型分布族と呼ぶ。
(x の関数)×(パラメータの関数)× exp{x の関数 × パラメータの関数 +...+ x の関数 × パラメータの関数 }

※知ってる分布はほとんど全て指数型分布族。分布が指数型分布族に入っていると、いろいろ嬉しいことがある (後述)。以後、とりあえず

$$f(x|\theta) = h(x)c(\theta)\exp\{w(\theta)t(x)\}$$

として話を進める。f は確率密度関数であるから、全確率 1 を満たす。また式の中で x に依存しない部分は $c(\theta)$ だけなので、

$$\begin{aligned}
\int f(x|\theta)dx &= \int h(x)c(\theta)\exp\{w(\theta)t(x)\}dx \\
&= c(\theta)\int h(x)\exp\{w(\theta)t(x)\}dx \\
&= 1 \\
\therefore \frac{1}{c(\theta)} &= \int h(x)\exp\{w(\theta)t(x)\}dx
\end{aligned}$$

さて、$w(\theta)=\eta$ と置くと、$c(\theta)$ はある関数 c^* によって $c^*(\eta)=c(\theta)$ と書けるので、

$$f(x|\theta)=h(x)c^*(\eta)\exp\{\eta t(x)\}$$

と書き換えることができる。このとき η を自然母数、$E[t(X)]$ を期待値母数と呼ぶ。(w を自然母数、t の期待値を期待値母数と呼ぶ！で良さそう)

分布が指数型分布族に属している場合、期待値母数の計算が簡単 (一般には積分計算) になる。全確率 1 より、

$$\int f(x|\theta)dx=\int h(x)c(\theta)\exp\{w(\theta)t(x)\}dx=1$$

両辺を θ で微分すると、

$$\frac{d}{d\theta}\int h(x)c(\theta)\exp\{w(\theta)t(x)\}dx=\int \frac{dh(x)c(\theta)\exp\{w(\theta)t(x)\}}{d\theta}dx=0 \quad (1)$$

さて、左辺の積分の中身の微分を計算する。

$$\begin{aligned}
\frac{d}{d\theta}h(x)c(\theta)\exp\{w(\theta)t(x)\} &= h(x)c'(\theta)\exp\{w(\theta)t(x)\}+w'(\theta)t(x)h(x)c(\theta)\exp\{w(\theta)t(x)\} \\
&= \{\frac{c'(\theta)}{c(\theta)}+t(x)w'(\theta)\}f(x|\theta)
\end{aligned}$$

これを (1) に代入して、

$$\begin{aligned}
\int\{\frac{c'(\theta)}{c(\theta)}+t(x)w'(\theta)\}f(x|\theta)dx &= \frac{c'(\theta)}{c(\theta)}+w'(\theta)E[t(x)]=0 \\
\therefore E[t(x)] &= -\frac{c'(\theta)}{w'(\theta)c(\theta)}
\end{aligned}$$

(答案)

1. 二項分布

$$\begin{aligned} f(x|p) &= {}_nC_x p^x (1-p)^{n-x} \\ &= {}_nC_x (1-p)^n (\frac{p}{1-p})^x \\ &= {}_nC_x (1-p)^n \exp\{x \log \frac{p}{1-p}\} \end{aligned}$$

より、指数型分布族に入る。自然母数 η は、

$$\eta = \log \frac{p}{1-p}$$

期待値母数は、

$$E[X] = np$$

※指数型分布族を考えるとき、観測値のサイズは固定されていなければならない。例えば、二項分布における試行回数 n は、ここでは既知定数とされる。負の二項分布の失敗回数 r なども同様。

2. ポアソン分布

$$\begin{aligned} f(x|\lambda) &= \frac{\lambda^x}{x!} \exp(-\lambda) \\ &= \frac{1}{x!} \exp(-\lambda) \exp(x \log \lambda) \end{aligned}$$

より、指数型分布族に入る。自然母数 η は、

$$\eta = \log \lambda$$

期待値母数は、

$$E[X] = \lambda$$

3. 幾何分布

$$\begin{aligned} f(x|p) &= p(1-p)^x \\ &= p \exp\{x \log(1-p)\} \end{aligned}$$

より、指数型分布族に入る。自然母数 η は、

$$\eta = \log(1-p)$$

期待値母数は、

$$E[X] = \frac{1-p}{p}$$

4. 負の二項分布

$$\begin{aligned} f(x|p) &= {}_{x+r-1}C_x p^r (1-p)^x \\ &= {}_{x+r-1}C_x p^r \exp\{x \log(1-p)\} \end{aligned}$$

より、指数型分布族に入る。自然母数 η は、

$$\eta = \log(1-p)$$

期待値母数は、

$$E[X] = \frac{r(1-p)}{p}$$

5. 正規分布

$$\begin{aligned} f(x|\mu,\sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2\sigma^2}(x^2 - 2\mu x + \mu^2)\} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{\mu^2}{2\sigma^2}) \exp(-\frac{x^2}{2\sigma^2} + \frac{\mu}{\sigma^2}x) \end{aligned}$$

より、指数型分布族に入る。自然母数 η は、

$$\eta_1 = -\frac{1}{2\sigma^2}, \quad \eta_2 = \frac{\mu}{\sigma^2}$$

期待値母数は、

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

6. ガンマ分布

$$\begin{aligned} f(x|\alpha,\beta) &= \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x) \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{x} \exp(\alpha \log x - \beta x) \end{aligned}$$

より、指数型分布族に入る。自然母数 η は、

$$\eta_1 = \alpha, \quad \eta_2 = -\beta$$

期待値母数 $E[\log X], E[X]$ を求める。対数尤度は、

$$\log f(x|\alpha,\beta) = \alpha \log \beta - \log \Gamma(\alpha) + (\alpha - 1) \log x - \beta x$$

であるから、α についてのスコア関数 (対数尤度の微分) は、

$$S(\alpha, X) = \log \beta - \frac{d}{d\alpha} \log \Gamma(\alpha) + \log X$$

となる。スコア関数の期待値は 0 であるから、

$$E[\log X] = \frac{d}{d\alpha} \log \Gamma(\alpha) - \log \beta$$

また、

$$E[X] = \frac{\alpha}{\beta}$$

(別解)

方針で示した期待値母数の期待値の公式に従えば、

$$\begin{aligned} E[\log X] &= -(\frac{\partial}{\partial \alpha} \frac{\beta^\alpha}{\Gamma(\alpha)}) \frac{\Gamma(\alpha)}{\beta^\alpha} \\ &= -\frac{\Gamma(\alpha)\beta^\alpha \log \beta - \beta^\alpha \Gamma(\alpha)'}{\Gamma(\alpha)^2} \frac{\Gamma(\alpha)}{\beta^\alpha} \\ &= -\log \beta + \frac{\Gamma(\alpha)'}{\Gamma(\alpha)} \\ &= \frac{d}{d\alpha} \log \Gamma(\alpha) \end{aligned}$$

7. ベータ分布

$$\begin{aligned} f(x|a,b) &= \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} \\ &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{1}{x(1-x)} \exp\{a\log x + b\log(1-x)\} \end{aligned}$$

より、指数型分布族に入る。自然母数 η は、

$$\eta_1 = a, \quad \eta_2 = b$$

期待値母数 $E[\log X], E[log(1-X)]$ を求める。対数尤度は、

$$\begin{aligned} \log f(x|a,b) &= \log\Gamma(a+b) - \log\Gamma(a) - \log\Gamma(b) \\ &\quad -\log x - \log(1-x) + a\log x + b\log(1-x) \end{aligned}$$

であるから、a,b についてのスコア関数 (対数尤度の微分) はそれぞれ、

$$\begin{aligned} S(a,x) &= \frac{d}{da}\log\Gamma(a+b) - \frac{d}{da}\log\Gamma(a) + \log x \\ S(b,x) &= \frac{d}{db}\log\Gamma(a+b) - \frac{d}{db}\log\Gamma(b) + \log(1-x) \end{aligned}$$

となる。スコア関数の期待値は 0 であるから、

$$\begin{aligned} E[\log X] &= -\frac{d}{da}\log\Gamma(a+b) + \frac{d}{da}\log\Gamma(a) \\ E[\log(1-X)] &= -\frac{d}{db}\log\Gamma(a+b) + \frac{d}{db}\log\Gamma(b) \end{aligned}$$

(別解)

方針で示した期待値母数の期待値の公式に従えば、

$$\begin{aligned} E[\log X] &= -\left(\frac{\partial}{\partial a}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right)\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \\ &= -\frac{\Gamma(a+b)\Gamma(a)\Gamma(b) - \Gamma(a)\Gamma(b) - \Gamma(a+b)\Gamma(b)\Gamma(a)'}{(\Gamma(a)\Gamma(b))^2}\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \\ &= -\frac{\Gamma(a+b)'}{\Gamma(a+b)} + \frac{\Gamma(a)'}{\Gamma(a)} \\ &= -\frac{d}{da}\log\Gamma(a+b) + \frac{d}{da}\log\Gamma(a) \end{aligned}$$

6.2 標準

十分統計量は、パラメータの点推定のために最低限必要な情報量のことである。$X_1, ..., X_n$ の一つ一つの値は保存しておかなくても標本平均と標本分散だけ残しておけばパラメータを推定できる、みたいなとき、標本平均と標本分散は十分統計量である。

十分統計量

尤度が

$$f_{\boldsymbol{X}}(\boldsymbol{x}|\theta) = h(\boldsymbol{x})g_\theta(t(x))$$

と書ける時、$t(\boldsymbol{x})$ を十分統計量という。
(尤度がパラメータに依存しない $h(\boldsymbol{x})$ とパラメータに依存し、十分統計量 t のみで表せる $g_\theta(t)$ に分割できる)

(Ex)
$X_1, ..., X_n : iid \sim N(\mu, \sigma^2)$ の時、μ, σ^2 の尤度関数は、

$$\begin{aligned} f_{\boldsymbol{X}}(\boldsymbol{x}|\mu, \sigma^2) &= \Pi_{i=1}^n f(x_i|\mu, \sigma^2) \\ &= (2\pi\sigma^2)^{-n/2} \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2\} \\ &= (2\pi\sigma^2)^{-n/2} \exp\{-\frac{nS^2 + n(\bar{x} - \mu)^2}{2\sigma^2}\} \end{aligned}$$

となる。もともとの n 個のデータ $x_1, ..., x_n$ は陽に登場しない。十分統計量は $S^2, \bar{x}$ である。つまり、尤度をベースに分析を行う場合、標本平均、分散のみ報告されていれば良い。

(Ex)
$X_1, ..., X_n : iid \sim Po(\lambda)$ の時、λ の尤度関数は、

$$f_{\boldsymbol{X}}(\boldsymbol{x}|\lambda) = \frac{\lambda^{\sum_{i=1}^n x_i}}{(\Pi_{i=1}^n x_i!)} \exp(-n\lambda)$$

で、十分統計量は観測値の和 $\sum_{i=1}^n x_i$。(標本平均でも良い。)$\Pi_{i=1}^n x_i!$ は十分統計量ではない。なぜなら、対数尤度；

$$\log f_{\boldsymbol{X}}(\boldsymbol{x}|\lambda) = (\sum_{i=1}^n x_i) \log\lambda - \log \Pi_{i=1}^n x_i - n\lambda$$

を考えた時、右辺第二項は λ で微分すると 0 になることから、尤度の大小に寄与しない部分であるからである。定義に沿った議論もできる。

$$h(\boldsymbol{x}) = \Pi_{i=1}^{n} x_i!,\ g_\lambda(t) = \lambda^t \exp(-n\lambda) \quad t = \sum_{i=1}^{n} x_i$$

とすれば、パラメータに関する議論は $g_\lambda(t)$ さえあれば十分であることが分かるので、十分統計量は t。

(答案)
　標本 $X_1, .., X_n$ が得られた時の尤度関数 $f(\boldsymbol{x}|\theta)$ は、

$$\begin{aligned} f(\boldsymbol{x}|\theta) &= \Pi_{i=1}^{n} \theta x_i^{\theta-1} \\ &= \theta^n (\Pi_{i=1}^{n} x_i)^{\theta-1} \end{aligned}$$

より、十分統計量は $\Pi_{i=1}^{n} x_i$ である。$\xi = 1/\theta$ と置くと、

$$\begin{aligned} f(\boldsymbol{x}|\xi) &= \Pi_{i=1}^{n} \frac{1}{\xi} x_i^{\frac{1}{\xi}-1} \\ &= \frac{1}{\xi^n} (\Pi_{i=1}^{n} x_i)^{\frac{1}{\xi}-1} \end{aligned}$$

となる。対数を取って、

$$\log f(\boldsymbol{x}|\xi) = -n \log \xi + (\frac{1}{\xi} - 1) \sum_{i=1}^{n} \log x_i$$

これを ξ で微分して 0 と置くと、最尤推定量が求まる。

$$\begin{aligned} \frac{\partial}{\partial \xi} \log f(\boldsymbol{x}|\xi) &= -\frac{n}{\xi} - \frac{1}{\xi^2} \sum_{i=1}^{n} \log x_i = 0 \\ \therefore \hat{\xi}^{ML} &= -\frac{1}{n} \sum_{i=1}^{n} \log x_i \end{aligned}$$

(標本平均の形にしたい！)$Y_1 = -\log X_1$ と置くと、最尤推定量 $\hat{\xi}^{ML}$ は

$$\hat{\xi}^{ML} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

と書ける。(Y の期待値、分散を求めれば中心極限定理により $\hat{\xi}^{ML}$ の分散の収束先がわかる。)

$$\begin{aligned}
E[Y] &= \int_0^1 -\log x\theta x^{\theta-1}dx \\
&= -\int_0^1 \log x(x^\theta)'dx \\
&= -[x^\theta \log x]_0^1 + \int_0^1 x^{\theta-1}dx \\
&= [\frac{1}{\theta}x^\theta]_0^1 \\
&= \frac{1}{\theta} = \xi \\
E[Y^2] &= \int_0^1 (\log x)^2\theta x^{\theta-1}dx \\
&= \int_0^1 (\log x)^2(x^\theta)'dx \\
&= [x^\theta(\log x)^2]_0^1 - \int_0^1 2x^{\theta-1}\log x dx \\
&= -\int_0^1 2\log x(\frac{1}{\theta}x^\theta)'dx \\
&= -[\frac{2}{\theta}x^\theta \log x]_0^1 + \int_0^1 \frac{2}{\theta}x^{\theta-1}dx \\
&= [\frac{2}{\theta^2}x^\theta]_0^1 \\
&= \frac{2}{\theta^2} = 2\xi^2 \\
\therefore Var(Y) &= 2\xi^2 - \xi^2 = \xi^2
\end{aligned}$$

したがって、$Y_1, ..., Y_n : iid \sim (\xi, \xi^2)$ となるので、中心極限定理より、

$$\hat{\xi}^{ML} = \frac{1}{n}\sum_{i=1}^n Y_i \to_d N(\xi, \frac{\xi^2}{n})$$

となるので、$\frac{\xi^2}{n} \to 0 \;\; as \; n \to \infty$ より $Var(\hat{\xi}^{ML})$ は 0 に収束する。

※標本平均は、母平均、母分散から平均や分散が求められる。

6.3 標準

標本 $X_1, ..., X_n$ が与えられた時の尤度関数 $f(\boldsymbol{x}|\theta)$ は、

$$\begin{aligned} f(\boldsymbol{x}|\theta) &= \Pi_{i=1}^n \theta(1+x_i)^{-(1+\theta)} \\ &= \theta^n \{\Pi_{i=1}^n (1+x_i)\}^{-(1+\theta)} \\ &= \theta^n \exp\{-(1+\theta)\sum_{i=1}^n \log(1+x_i)\} \end{aligned}$$

ゆえ、T は θ に対する十分統計量である。$Y_1 = \log(1+X_1)$ と置くと、

$$\begin{aligned} E[Y] &= \int_0^\infty \log(1+x)\frac{\theta}{(1+x)^{1+\theta}}dx \\ &= \int_0^\infty \{-\frac{1}{(1+x)^\theta}\}' \log(1+x)dx \\ &= [-\frac{1}{(1+x)^\theta}\log(1+x)]_0^\infty + \int_0^\infty \frac{1}{(1+x)^{1+\theta}}dx \\ &= [-\frac{1}{\theta}\frac{1}{(1+x)^\theta}]_0^\infty = \frac{1}{\theta} \\ \therefore E[T] &= nE[\frac{1}{n}\sum_{i=1}^n \log(1+x)] = \frac{n}{\theta} \\ E[Y^2] &= \int_0^\infty \{\log(1+x)\}^2 \frac{\theta}{(1+x)^{1+\theta}}dx \\ &= \int_0^\infty \{-\frac{1}{(1+x)^\theta}\}'\{\log(1+x)\}^2 dx \\ &= [-\frac{1}{(1+x)^\theta}\{\log(1+x)\}^2]_0^\infty + \int_0^\infty \frac{1}{(1+x)^{1+\theta}} 2\log(1+x)dx \\ &= \int_0^\infty \{-\frac{1}{\theta}\frac{1}{(1+x)^\theta}\}' 2\log(1+x)dx \\ &= [-\frac{1}{\theta}\frac{1}{(1+x)^\theta} 2\log(1+x)]_0^\infty + \int_0^\infty \frac{2}{\theta}\frac{1}{(1+x)^{(1+\theta)}}dx \\ &= [-\frac{2}{\theta^2}\frac{1}{(1+x)^\theta}]_0^\infty = \frac{2}{\theta^2} \\ \therefore Var(Y) &= \frac{2}{\theta^2} - \frac{1}{\theta^2} = \frac{1}{\theta^2} \\ \therefore Var(T) &= n^2 Var(\frac{1}{n}\sum_{i=1}^n \log(1+x)) \\ &= n^2 \frac{1}{n\theta^2} = \frac{n}{\theta^2} \end{aligned}$$

6.4 標準

(方針)

確率変数の台 (定義域) にパラメータが含まれている場合はしばしばイレギュラーなことになるので注意。その典型例が一様分布。

推定量の評価について。最尤法、モーメント法、ベイズ法などから得られたパラメータの推定量のうち、どれが最も適切か、という評価基準が必要である。

推定量の評価

1. 不偏性
   推定量 $\hat{\theta}(\boldsymbol{X})$ が不偏であるとは、$E[\hat{\theta}] = \theta$ が成り立つことである。

2. バイアス
   推定量 $\hat{\theta}(\boldsymbol{X})$ のバイアスとは、次のように定義される。

$$Bias_{\theta}(\hat{\theta}) \Leftrightarrow E[\hat{\theta}(\boldsymbol{X})] - \theta$$

   $\hat{\theta}$ が不偏であることと、$Bias(\hat{\theta}) = 0$ であることは同値。

3. 平均二乗誤差 MSE
   推定量 $\hat{\theta}(\boldsymbol{X})$ の MSE とは、次のように定義される。(1 行目と最終行どちらも重要)

$$\begin{aligned} MSE(\hat{\theta}, \theta) &= E[(\hat{\theta}(\boldsymbol{X}) - \theta)^2] \\ &= E[\hat{\theta}^2] - 2\theta E[\hat{\theta}] + \theta^2 \\ &= E[\hat{\theta}^2] - E[\hat{\theta}]^2 + E[\hat{\theta}]^2 - 2\theta E[\hat{\theta}] + \theta^2 \\ &= Var(\hat{\theta}) + (E[\hat{\theta}] - \theta)^2 \\ &= Var(\hat{\theta}) + Bias(\hat{\theta}) \end{aligned}$$

4. 一致性
   推定量 $\hat{\theta}(\boldsymbol{X})$ が一致性を持つとは、「$\hat{\theta}(\boldsymbol{X})$ がパラメータの真値 θ に確率収束する」ことである。

1.

$$\begin{aligned} f(\boldsymbol{x}|\theta) &= \Pi_{i=1}^{n}\frac{1}{\theta}I(0 < X_i < \theta) \\ &= \frac{1}{\theta^n}\Pi_{i=1}^{n}I(0 < X_i < \theta) \\ &= \frac{1}{\theta^n}I(0 < X_{(1)})I(X_{(n)}) \end{aligned}$$

より ($I(0 < X_{(1)})$ はパラメータ θ に依存しないので)、十分統計量は $X_{(n)}$

2. まずは、最大統計量 $X_{(n)}$ に基づいた不偏推定量 $\hat{\theta}^{U_1}$ を求める。

$$\begin{aligned} F_{X_{(n)}}(x) &= P[X_{(1)} \leq x, ..., X_{(n)} \leq x] = (\frac{x}{\theta})^n \\ \therefore f_{X_{(n)}}(x) &= \frac{n}{\theta^n}x^{n-1} \\ \therefore E[X_{(n)}] &= \int_0^\theta \frac{n}{\theta^n}x^n dx \\ &= [\frac{n}{\theta^n(n+1)}x^{n+1}]_0^\theta \\ &= \frac{n}{n+1}\theta \end{aligned}$$

したがって、

$$\hat{\theta}^{U_1} = \frac{n+1}{n}X_{(n)}$$

とすればこれは不偏推定量である。また、

$$E[\bar{X}] = \frac{\theta}{2}$$

(標本平均の期待値は母平均に一致) より、

$$\hat{\theta}^{U_2} = 2\bar{X}$$

とすれば、これは不偏推定量である。次に、$\hat{\theta}^{U_1}, \hat{\theta}^{U_2}$ の MSE を求める。

推定量が不偏なので、MSE は分散に一致する。

$$\begin{aligned}
E[(\hat{\theta}^{U_1})^2] &= E[(\frac{n+1}{n})^2 X_{(n)}^2] \\
&= (\frac{n+1}{n})^2 \int_0^\infty \frac{n}{\theta^n} x^{n+1} dx \\
&= (\frac{n+1}{n})^2 (\frac{n}{n+2}) \theta^2 \\
\therefore MSE(\theta, \hat{\theta}^{U_1}) &= (\frac{n+1}{n})^2 (\frac{n}{n+2})^2 \theta^2 - \theta^2 \\
&= \frac{(n+1)^2 - n(n+2)}{n(n+2)} \theta^2 \\
&= \frac{1}{n(n+2)} \theta^2 \\
E[(\hat{\theta}^{U_2})^2] &= 4E[\bar{X}^2] \\
&= 4\{Var(\bar{X}) + (E[\bar{X}])^2\} \\
&= 4(\frac{\theta^2}{12n} + \frac{\theta^2}{4}) \\
\therefore MSE(\theta, \hat{\theta}^{U_2}) &= 4(\frac{\theta^2}{12n} + \frac{\theta^2}{4}) - \theta^2 \\
&= \frac{\theta^2}{3n}
\end{aligned}$$

したがって、$MSE(\theta, \hat{\theta}^{U_2}) \geq MSE(\theta, \hat{\theta}^{U_1})$

3. θ に関する尤度関数は、

$$\begin{aligned}
f(\theta|\boldsymbol{X}) &= \Pi_{i=1}^n \frac{1}{\theta} I(0 < X_i < \theta) \\
&= \frac{1}{\theta^2} I(0 < X_{(1)}) I(X_{(n)} < \theta)
\end{aligned}$$

ゆえ、$f(\theta|\boldsymbol{X})$ を最大化する θ は $X_{(n)}$ である。($X_{(n)}$ 以上で f は 0 以上となるが、$X_{(n)}$ よりも大きくなるにつれて分母の値が大きくなって

しまう)

$$\begin{aligned}
Bias(\hat{\theta}^{ML}) &= E[X_n] - \theta = \frac{n}{n+1}\theta - \theta = -\frac{\theta}{n+1} \\
E[(X_{(n)})^2] &= \int_0^\infty \frac{n}{\theta^2} x^{n+1} dx = [\frac{n}{(n+2)\theta^n} x^{n+2}]_0^\theta = \frac{n\theta^2}{n+2} \\
\therefore Var(X_{(n)}) &= \frac{n\theta^2}{n+2} - (\frac{n}{n+1}\theta)^2 \\
\therefore MSE(\theta, \hat{\theta}^{ML}) &= \frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2} - (\frac{\theta}{n+1})^2 \\
&= \frac{n(n+1)^2\theta^2 - n^2(n+2)\theta^2 - (n+2)\theta^2}{(n+1)^2(n+2)} \\
&= \frac{2}{(n+1)(n+2)}\theta^2
\end{aligned}$$

6.5

$p_1, ..., p_n$ に関する尤度関数は、

$$f_{1,...,k}(p_1, ..., p_k|\boldsymbol{X}) = \frac{n!}{x_1!...x_k!} p_1^{x_1} ... p_k^{x_k}$$

である。対数を取って、

$\log f_{1,...,k}(p_1, ..., p_k|\boldsymbol{X}) = \log n! - \log x_1! - ... - \log x_k! + x_1 \log p_1 + ... + x_k \log p_k$

より、最尤推定量 $\hat{p}_i$ は次の最大化問題を解くことで得られる。

$$\max_{p_1,...,p_k} \log f_{1,...,k}(p_1, ..., p_k|\boldsymbol{X}) \ s.t. \ p_1 + ... + p_k = 1$$

したがって、λ を定数としてラグランジアン L を設定すると、

$$L = \log f_{1,...,k}(p_1, ..., p_k|\boldsymbol{X}) - \lambda(p_1 + ... + p_k - 1)$$

一階条件は、

$$\begin{cases}
\frac{\partial L}{\partial p_1} = \frac{x_1}{p_1} - \lambda = 0 \\
\vdots \\
\frac{\partial L}{\partial p_k} = \frac{x_k}{p_k} - \lambda = 0 \\
\\
\frac{\partial L}{\partial \lambda} = p_1 + ... + p_k - 1 = 0
\end{cases}$$

ゆえ、これらを連立して解くことにより、$\hat{p}_i = \frac{X_i}{n}$

6.6

(方針)

指数型分布族の期待値について。

パラメータの推定法の一つに、モーメント法がある。モーメント法は、母集団分布のモーメントと標本のモーメントが合うように推定する。母集団の k 次モーメントを $\mu_k = E[X^k]$ と置くと、

$$\begin{cases} \mu_1 = \frac{1}{n}\sum_{i=1}^n X_i \\ \mu_2 = \frac{1}{n}\sum_{i=1}^n X_i^2 \\ \vdots \\ \mu_k = \frac{1}{n}\sum_{i=1}^n X_i^k \end{cases}$$

上の連立方程式を解いて得られる推定量をモーメント推定量という。(最尤法が難しい時に採用する。2 つ目の例が有名。)

(Ex)

$X_1, ..., X_n : iid \sim Ber(p)$ の時、$\mu_1 = E[X] = p$ と置けるので、

$$\mu_1 = \frac{1}{n}\sum_{i=1}^n X_i \quad \therefore \hat{p} = \bar{X}$$

(Ex)

$X_1, ..., X_n : iid \sim Ga(\alpha, \beta)$ の時、パラメータ α, β を最尤法によって推定しようとすると、

$$\begin{aligned} f(\alpha, \beta|\boldsymbol{X}) &= \frac{\beta^{\alpha n}}{\Gamma(\alpha)^n}(\Pi_{i=1}^n x_i)^{\alpha-1}\exp\{-\beta\sum_{i=1}^n x_i\} \\ \log f(\alpha, \beta|\boldsymbol{X}) &= -n\log\Gamma(\alpha) + n\alpha\log(\beta) + (\alpha-1)\sum_{i=1}^n \log x_i - \beta\sum_{i=1}^n x_i \end{aligned}$$

より、対数尤度を偏微分して 0 と置くと、

$$\begin{aligned} \frac{\partial \log f(\alpha, \beta|\boldsymbol{X})}{\partial\alpha} &= -n\frac{\partial\log\Gamma(\alpha)}{\partial\alpha} + n\log(\beta) + \sum_{i=1}^n \log x_i = 0 \\ \frac{\partial \log f(\alpha, \beta|\boldsymbol{X})}{\partial\beta} &= \frac{n\alpha}{\beta} - \sum_{i=1}^n x_i = 0 \end{aligned}$$

これを解くのは一般にとても難しい。そこでモーメント法を用いる。母集団の一次、二次モーメントは、

$$E[X]=\frac{\alpha}{\beta}, Var(X)=\frac{\alpha}{\beta^2},\ E[X^2]=\frac{\alpha(\alpha+1)}{\beta^2}$$

であるから、これを標本モーメントと一致させて、

$$\begin{aligned}\frac{\alpha}{\beta} &= \frac{1}{n}\sum_{i=1}^{n}X_i=\bar{X}\\ \frac{\alpha(\alpha+1)}{\beta^2} &= \frac{1}{n}\sum_{i=1}^{n}X_i^2\\ \therefore \hat{\beta} &= \frac{\bar{X}}{S^2},\\ \therefore \hat{\alpha} &= S^2\end{aligned}$$

本問は、指数型分布族に属する確率変数について、自然母数 η の最尤推定量とモーメント推定量が一致することを示すものである。前半については、ぱっと見で log があって η_j で微分しているあたりから対数尤度関数の微分を求めるのかな、と考えて試してみる。(別解の方が体系的でよろしいかも)

(答案)
$\boldsymbol{\eta}$ に関する尤度関数は、

$$f(\boldsymbol{\eta}|x) = h(x)c^*(\boldsymbol{\eta})\exp\{\sum_{i=1}^{k}\eta_i t_i(x)\}$$

である。両辺に対数を取って、

$$\log f(\boldsymbol{\eta}|x) = \log h(x)+\log c^*(\boldsymbol{\eta})+\sum_{i=1}^{k}\eta_i t_i(x)$$

これを η_j で偏微分すると、

$$\frac{\partial}{\partial\eta_j}f(\boldsymbol{\eta}|x) = \frac{\partial}{\partial\eta_j}\log c^*(\boldsymbol{\eta})+t_j(x)$$

最後に両辺の期待値を取る。スコア関数 (左辺の対数尤度関数の微分) の期待値はゼロ、という性質を使うと、

$$E[t_j(x)] = -\frac{\partial}{\partial\eta_j}\log c^*(\boldsymbol{\eta})$$

次に、標本 $X_1, ..., X_n$ が得られた時の尤度 $f(\boldsymbol{\eta}|x)$ は、

$$\begin{aligned} f(\boldsymbol{\eta}|x) &= \Pi_{i=1}^{n} h(x_i) c^*(\boldsymbol{\eta}) \exp\{\sum_{j=1}^{k} \eta_j t_j(x_i)\} \\ &= \Pi_{i=1}^{n} h(x_i) \{c^*(\boldsymbol{\eta})\}^n \exp\{\sum_{i=1}^{n}\sum_{j=1}^{k} \eta_j t_j(x_i)\} \end{aligned}$$

対数をとって、

$$\log f(\boldsymbol{\eta}|x) = \sum_{i=1}^{n} \log h(x_i) + n \log c^*(\boldsymbol{\eta}) + \sum_{i=1}^{n}\sum_{j=1}^{k} \eta_j t_j(x_i)$$

各 j について対数尤度を偏微分して 0 と置くと、

$$\frac{\partial}{\partial \eta_j} \log f(\boldsymbol{\eta}|x) = n \frac{\partial}{\partial \eta_j} \log c^*(\boldsymbol{\eta}) + \sum_{i=1}^{n} t_j(x_i) = 0$$

より、パラメータ $\boldsymbol{\eta}$ の MLE は、

$$\sum_{i=1}^{n} t_j(x_i) = -n \frac{\partial}{\partial \eta_j} \log c^*(\boldsymbol{\eta}) \quad \forall j$$

また、前述の通り

$$E[t_j(X)] = -\frac{\partial}{\partial \eta_j} \log c^*(\boldsymbol{\eta})$$

より、$\boldsymbol{\eta}$ のモーメント推定量は、

$$\sum_{i=1}^{n} t_j(x_i) = -n \frac{\partial}{\partial \eta_j} \log c^*(\boldsymbol{\eta}) \quad \forall j$$

(前半の別解)
全確率 1 より、

$$\int f(x|\boldsymbol{\eta}) = \int h(x) c^*(\boldsymbol{\eta}) \exp\{\sum_{i=1}^{k} \eta_i t_i(x)\} dx = 1$$

両辺を η_j で偏微分して、

$$\frac{d}{d\eta_j}\int h(x)c^*(\boldsymbol{\eta})\exp\{\sum_{i=1}^{k}\eta_i t_i(x)\}dx = \int \frac{d}{d\eta_j}h(x)c^*(\boldsymbol{\eta})\exp\{\sum_{i=1}^{k}\eta_i t_i(x)\}dx = 0 \quad (1)$$

さて、積分の中身の微分を計算する。

$$\begin{aligned}
\frac{d}{d\eta_j}h(x)c^*(\boldsymbol{\eta})\exp\{\sum_{i=1}^{k}\eta_i t_i(x)\}dx &= \frac{d}{d\eta_j}c^*(\boldsymbol{\eta}) \times h(x)\exp\{\sum_{i=1}^{k}\eta_i t_i(x)\} \\
&\quad + t_j \times h(x)c^*(\boldsymbol{\eta})\exp\{\sum_{i=1}^{k}\eta_i t_i(x)\} \\
&= \{\frac{1}{c^*(\boldsymbol{\eta})}\frac{dc^*(\boldsymbol{\eta})}{d\eta_j} + t_j(x)\}f(x|\eta)
\end{aligned}$$

これを (1) に代入して、

$$\begin{aligned}
\int\{\frac{1}{c^*(\boldsymbol{\eta})}\frac{dc^*(\boldsymbol{\eta})}{d\eta_j} + t_j(x)\}f(x|\eta)dx &= \frac{1}{c^*(\boldsymbol{\eta})}\frac{dc^*(\boldsymbol{\eta})}{+}E[t_j(x)] = 0 \\
\therefore E[t_j(x)] &= -\frac{1}{c^*(\boldsymbol{\eta})}\frac{dc^*(\boldsymbol{\eta})}{d\eta_j} = -\frac{d}{d\eta_j}\log c^*(\boldsymbol{\eta})
\end{aligned}$$

6.7

1.

$$\begin{aligned}
\pi(p|X) &\propto {}_nC_x p^x(1-p)^{n-x}\frac{1}{B(\alpha,\beta)}p^{\alpha-1}(1-p)^{\beta-1} \\
&\propto p^{x+\alpha-1}(1-p)^{n-x+\beta-1}
\end{aligned}$$

より、事後分布は $p|X \sim Beta(X+\alpha, n-Y+\beta)$ で与えられる。

2.

$$
\begin{aligned}
\hat{p}^B &= E[p|X] \\
&= \int_0^1 \frac{1}{B(x+\alpha, n-x+\beta)} p^{x+\alpha}(1-p)^{n-x+\beta-1}dp \\
&= \int_0^1 \frac{1}{B(x+\alpha'-1, n-x+\beta)} p^{x+\alpha'-1}(1-p)^{n-x+\beta-1}dp \quad (\alpha = \alpha - 1) \\
&= \frac{x+\alpha'-1}{x+\alpha'-1+n-x+\beta} \int_0^1 \frac{1}{B(x+\alpha'-1, n-x+\beta)} 9^{x+\alpha'-1}(1-p)^{n-x+\beta-1}dp \\
&= \frac{x+\alpha}{x+\alpha+n-x+\beta} \\
&= \frac{n}{\alpha+\beta+n}\frac{x}{n} + \frac{\alpha+\beta}{\alpha+\beta+n}\frac{\alpha}{\alpha+\beta}
\end{aligned}
$$

6.8

(解答とは $Ga(\alpha,\beta)$ の定義が異なる (スケールパラメータ β が、解答では逆数になっている)。)

1.

$$
\begin{aligned}
\pi(\lambda|\boldsymbol{x}) &\propto \Pi_{i=1}^n \frac{\lambda^{x_i}}{x_i!}\exp(-\lambda) \times \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}\exp(-\beta\lambda) \\
&\propto \lambda^{\sum_{i=1}^n x_i+\alpha-1}\exp\{-(n+\beta)\lambda\}
\end{aligned}
$$

より、$\lambda|\boldsymbol{X} \sim Ga(\sum_{i=1}^n x_i + \alpha, \beta+n)$

2. ガンマ分布の平均・分散に代入。

6.9

1.

$$
\begin{aligned}
E[\sum_i a_i\hat{\theta}_i] &= a_1E[\hat{\theta}_1] + ... + a_kE[\hat{\theta}_k] \\
&= (a_1 + ... + a_k)\theta
\end{aligned}
$$

より、$\sum_i a_i\hat{\theta}_i$ は θ の不偏推定量であるための条件は、$a_1 + ... + a_k = 1$

2.

$$
\begin{aligned}
Var(\sum_i a_i\hat{\theta}_i) &= a_1^2 var(\hat{\theta}_1) + ... + a_k^2 Var(\hat{\theta}_k) + \sum_{i\neq j} a_ia_jCov(\hat{\theta}_i, \hat{\theta}_j) \\
&= a_1^2\sigma_1^2 + ... + a_k\sigma_k^2
\end{aligned}
$$

より、次の条件付き最小化問題を解く。

$$\min \quad a_1^2\sigma_1^2 + .. + a_k^2\sigma_k^2 \quad s.t. \quad a_1 + ... + a_k = 1$$

λ を定数としてラグランジアンを設定して、

$$L = a_1^2\sigma_1^2 + .. + a_k^2\sigma_k^2 + \lambda(a_1 + ... + a_k - 1)$$

一階条件は、

$$\begin{aligned}
\frac{\partial L}{\partial a_1} &= 2\sigma_1^2 a_1 + \lambda = 0 \\
&\vdots \\
\frac{\partial L}{\partial a_k} &= 2\sigma_k^2 a_k + \lambda = 0 \\
\frac{\partial L}{\partial \lambda} &= a_1 + ... + a_k - 1 = 0
\end{aligned}$$

となるので、この連立方程式を解くことで、$1 \leq j \leq k$ について

$$a_1 = \frac{1}{\sigma_j^2(\frac{1}{\sigma_1^2} + ... + \frac{1}{\sigma_k^2})}$$

が求める条件となる。

6.10

1. μ, σ についての尤度関数は、

$$\begin{aligned}
f(\mu,\sigma|\boldsymbol{X}) &= \Pi_{i=1}^n \frac{1}{\sigma}\exp(-\frac{X_i - \mu}{\sigma}) \\
&= \frac{1}{\sigma^n}\exp\{-\frac{1}{\sigma}(\sum_{i=1}^n X_i - n\mu)\} \\
&= \frac{1}{\sigma^n}\exp[-\frac{1}{\sigma}\{\sum_{i=1}^n (X_i - X_{(1)}) + n(X_{(1)} - \mu)\}] \\
&= \frac{1}{\sigma^n}\exp[-\frac{1}{\sigma}\{T + n(U - \mu)\}]
\end{aligned}$$

となるので、U, T は十分統計量となる。

2. (方針)

$X_{(1)},...,X_{(n)}$ の同時確率密度関数は、

$$f_{X_{(1)},...,X_{(n)}}(x_1,...,x_n) = n!f(x_1)...f(x_n)$$

(答案)

順序統計量の同時密度は、

$$\begin{aligned} f_{X_{(1)},...,X_{(n)}}(x_1,...,x_n) &= n!\frac{1}{\sigma}\exp(-\frac{x_1-\mu}{\sigma})...\frac{1}{\sigma}\exp(-\frac{x_n-\mu}{\sigma}) \\ &= n!\frac{1}{\sigma^2}\exp[-\frac{1}{\sigma}\{(x_1-\mu)+...+(x_n-\mu)\}] \qquad (1) \end{aligned}$$

ここで、exp の中身について、

$$\begin{aligned} (x_1-\mu)+...+(x_n-\mu) &= (x_1-\mu) \\ &\quad +\{(x_2-x_1)-(\mu-x_1)\}+...+\{(x_n-x_1)-(\mu-x_1)\} \\ &= (x_1-\mu)+\{(x_2-x_1)+...+(x_n-x_1)-n(\mu-x_1)\} \\ &= n(x_1-\mu)+(x_2-x_1)+...+(x_n-x_1) \end{aligned}$$

より、

$$(1) = n!\frac{1}{\sigma^n}\exp\{-\frac{n}{\sigma}(x_1-\mu)-\frac{1}{\sigma}\sum_{i=2}^{n}(x_i-x_1)\}$$

今、

$$\begin{cases} Z_1 = \frac{X_{(1)}-\mu}{\sigma} \\ \vdots \\ Z_n = \frac{X_{(n)}-\mu}{\sigma} \end{cases} \Leftrightarrow \begin{cases} X_{(1)} = \sigma Z_1+\mu \\ \vdots \\ X_{(n)} = \sigma Z_n+\mu \end{cases}$$

なる一対一変換を行うと、ヤコビアンは

$$\frac{\partial(X_{(1)},...,X_{(n)})}{\partial(Z_1,...,Z_n)} = det\begin{pmatrix} \sigma & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \sigma \end{pmatrix} = \sigma^n$$

となるので

$$\begin{aligned} f_{Z_1,...,Z_n}(z_1,...,z_n) &= n!\exp(-nz_1-z_2-...-z_n) \\ &= n!\exp(-nz_1)\exp(-z_2-...-z_n) \end{aligned}$$

同時密度が積で書けるので、

$$\begin{aligned}
nZ_1 &= \frac{n(X_{(1)}-\mu)}{\sigma} = \frac{n(U-\mu)}{\sigma} \\
Z_2 + ... + Z_n &= \frac{(X_{(2)}-\mu) + ... + (X_{(n)}-\mu)}{\sigma} \\
&= \frac{\sum_{i=1}^n (X_i - X_{(1)}) - n(\mu - X_{(1)})}{\sigma} \\
&= \frac{1}{2}\frac{2T}{\sigma} + \frac{n(\mu - X_{(1)})}{\sigma}
\end{aligned}$$

は独立に分布する。

6.11

1.

$$\begin{aligned}
f(p|x) &= {}_nC_x p^X (1-p)^{n-X} \\
\log f(p|x) &= \log {}_nC_x + X\log p + (n-X)\log(1-p) \\
\frac{d}{dp}\log f(p|X) &= \frac{X}{p} - \frac{n-X}{1-p} = 0
\end{aligned}$$

より、$\hat{p}^{ML} = \frac{X}{n}$ となる。$\theta = p(1-p)$ と置くと、

$$\hat{\theta}^{ML} = \frac{X}{n}(1 - \frac{X}{n}) = \frac{X(n-X)}{n^2}$$

次に θ の不偏推定量 $\hat{\theta}^{ML}$ を求める。$E[aX + bX^2] = p(1-p)$ となるような定数 a, b を求めれば良い。

$$\begin{aligned}
E[aX + bX^2] &= aE[X] + bE[X^2] \\
&= anp + b\{np(1-p) - n^2p^2\} \\
&= n(a+b)p - bn(1-n)p^2
\end{aligned}$$

より、

$$\begin{cases} n(a+b) = 1 \\ bn(1-n) = 1 \end{cases}$$

この連立方程式を解くと、

$$a = -\frac{1}{1-n}, \quad b = \frac{1}{n(1-n)}$$

となるので、

$$\hat{\theta}^U = -\frac{X}{1-n} + \frac{X^2}{n(1-n)} = \frac{X(n-X)}{n(n-1)}$$

2. まずは $\sqrt{n}(\hat{\theta}^{ML} - \theta)$ の漸近分布を求める。

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

より、$g(\frac{X}{n}) = \frac{X}{n}(1 - \frac{X}{n})$ と置くと、

$$\sqrt{n}(\frac{X(n-X)}{n^2} - p(1-p)) = \sqrt{n}(g(\frac{X}{n}) - g(p))$$

と書ける。今、$\sqrt{n}(\frac{X}{n} - p) \sim N(-, p(1-p))$ なので、デルタ法により、

$$\begin{aligned}
\sqrt{n}(g(\frac{X}{n}) - g(p)) &\to g'(p)N(0, p(1-p)) \\
&= (1-2p)N(0, p(1-p)) \\
&= N(0, p(1-p)(1-2p)^2)
\end{aligned}$$

次に、$\sqrt{n}(\hat{\theta}^U - \theta)$ の漸近分布を求める。

$$\begin{aligned}
\frac{X(n-X)}{n(n-1)} &= \frac{n}{n-1}\frac{X(n-X)}{n^2} \\
&= \frac{n}{n-1}\hat{\theta}^{ML}
\end{aligned}$$

より、

$$\begin{aligned}
\sqrt{n}(\hat{\theta}^U - p(1-p)) &= \sqrt{n}(\frac{n}{n-1}\hat{\theta}^{ML} - p(1-p)) \\
&= \sqrt{n}(\hat{\theta}^{ML} + \frac{1}{n-1}\hat{\theta}^{ML} - p(1-p)) \\
&= \sqrt{n}(\hat{\theta}^{ML} - p(1-p)) + \frac{\sqrt{n}}{n-1}\hat{\theta}^{ML}
\end{aligned}$$

と書ける。今、

$$\sqrt{n}(\hat{\theta}^{ML} - p(1-p)) \to_d N(0, p(1-p)(1-2p)^2), \quad \frac{\sqrt{n}}{n-1}\hat{\theta}^{ML} \to_p 0$$

なので、スラツキーの定理より、$\sqrt{n}(\hat{\theta}^U - p(1-p)) \to_d N(0, p(1-p)(1-2p)^2)$ である。

6.12

(方針)

最尤推定量について。以下、パラメータ θ の関数としての尤度を、サンプル数 n を強調するため $f_n(\theta|\boldsymbol{x})$ と書く。

スコア関数

スコア関数とは、対数尤度関数の微分であって、以下のように定義される。

$$S_n(\theta|\boldsymbol{X}) = \frac{d}{d\theta}\log f_n(\theta|\boldsymbol{X})$$

データ n 個のスコア関数は、データ 1 個のスコア関数の和として表される。つまり、

$$\begin{aligned} S_1(\theta|\boldsymbol{X}) &= \frac{d}{d\theta}\log f_1(\theta|X_i) \\ \Rightarrow S_n(\theta|\boldsymbol{X}) &= \frac{d}{d\theta}\log f_n(\theta|\boldsymbol{X}) \\ &= \sum_{i=1}^{n}\frac{d}{d\theta}\log f(\theta|X_i) \\ &= \sum_{i=1}^{n} S_1(\theta|X_i) \end{aligned}$$

Fisher 情報量

Fisher 情報量とは、スコア関数の 2 乗の期待値であって、以下のように定義される。

$$I_n(\theta) = E[S_n(\theta|\boldsymbol{X})^2]$$

命題

条件「サポートが θ を含まない」、「f が θ で微分可能であること」、「$0 < I_1(\theta) < \infty$」を満たす時、以下が成り立つ。

1. $E[S_1(\theta|X_i)] = 0$
2. $I_n(\theta) = nI_1(\theta)$
3. $I_n(\theta) = -E[\frac{d^2}{d\theta^2}\log f(\theta|X_i)]$

(3 つ目はとても嬉しい。次数を徒に増やす二乗よりも微分の方が計算な楽な場合が多い (6.13 でよくわかる))

クラメール・ラオ不等式

不偏推定量 $\hat{\theta}$ に対して、任意の θ で、

$$Var_\theta(\hat{\theta}) \geq \frac{1}{I_n(\theta)}$$

(Ex)
$X_1, ..., X_n : iid \sim N(\mu, 1)$ の時、データ 1 つのスコア関数を求める。

$$\begin{aligned}
f_1(\mu|X_i) &= \frac{1}{\sqrt{2\pi}}\exp\{-\frac{(X_i-\mu)^2}{2}\} \\
\log f_1(\mu|X_i) &= -\frac{1}{2}\log(2\pi) - \frac{1}{2}(X_i - mu)^2 \\
S_1(\mu, X_i) &= \frac{d}{d\mu}\log f_1(\mu, X_i)
\end{aligned}$$

スコア関数の期待値がゼロであることを確かめる。

$$E[S_1(\mu, X_i)] = E[X_i - \mu] = 0$$

データ一つのフィッシャー情報量を二通りの方法で求める。

$$\begin{aligned}
I_1(\mu) &= E[S_1(\mu, X_i)^2] = E[(X_i - \mu)^2] = Var(X_i) = 1 \\
I_1(\mu) &= -E[\frac{d^2}{d\mu^2}\log f(\mu|X_i)] = -E[\frac{d}{d\mu}(X_i - \mu)] = 1
\end{aligned}$$

今、不偏推定量として $\hat{\mu} = \bar{X}$ を取ると、推定量の分散は、

$$MSE(\hat{\mu}, \mu) = Var(\hat{\mu}) = \frac{1}{n}$$

である。CR 下限は

$$\frac{1}{I_n(\mu)} = \frac{1}{nI_1(\mu)} = \frac{1}{n}$$

なので、$\bar{X}$ は CR 下限を達成する。一方、$\hat{\mu}' = X_1$ を取ると (これは自明に不偏)、

$$Var(\hat{\mu}') = Var(X_1) = 1 > \frac{1}{n}$$

より、推定量の分散は CR 下限より大きい。

MLE の一致性、漸近正規性

パラメータの最尤推定量 $\hat{\theta}$ は、正則条件と「θ が識別可能 (θ の値が異なれば尤度も異なる)」、「真値 θ_0 がパラメータ空間 Θ の内点」なる条件を満たす時、次が成り立つ。

1. $\hat{\theta} \to_p \theta$
2. $\sqrt{n}(\hat{\theta} - \theta_0) \to_d N(0, \frac{1}{I_1(\theta_0)})$

(証明は別の問の方針で書く)
標本平均の漸近生規性は CLT。MLE の漸近正規性も言える。したがって最尤推定量が標本平均だったら両者は一緒。

(答案)

1. ポアソン分布に従う確率変数 X のモーメント母関数は、

$$M_X(t) \quad = \quad \exp\{\lambda(e^t - 1)\}$$

である。$X_1, ..., X_n : iid \sim Po(\lambda)$ で、$Y = \sum_{i=1}^n X_i$ とすると、

$$\begin{aligned} M_Y(t) &= E[\exp\{t(X_1 + ... + X_n)\}] \\ &= E[e^{tX_1 + ... + tX_n}] \\ &= E[e^{tX_1}]...E[e^{tX_n}] \quad (\because iid) \\ &= \exp\{n\lambda(e^t - 1)\} \end{aligned}$$

となるので、$Y \sim Po(n\lambda)$

2. データ一つの尤度関数は、

$$f_1(\lambda|X_i) = \frac{\lambda^{X_i}}{X_i!}\exp(-\lambda)$$

なので、対数を取って、

$$\log f_1(\lambda|X_i) = X_i \log\lambda - \log X_i! - \lambda$$

となる。スコア関数は、

$$S_1(\lambda|X_i) = \frac{d}{d\lambda}\log f_1(\lambda|X_i) = \frac{X_i}{\lambda} - 1$$

フィッシャー情報量を (二通りの方法で) 求める。

$$\begin{aligned} I_1(\lambda) &= E[S_1(\lambda|X_i)^2] \\ &= E[(\frac{X_i}{\lambda} - 1)^2] \\ &= \frac{1}{\lambda^2}E[X_i^2] - \frac{2}{\lambda}E[X_i] + 1 \\ &= \frac{1}{\lambda^2}(\lambda^2 + \lambda) - \frac{2}{\lambda}\lambda + 1 \\ &= \frac{1}{\lambda} \\ I_1(\lambda) &= -E[\frac{d^2}{d\lambda^2}\log f_1(\lambda|X_i)] \\ &= -E[\frac{d}{d\lambda}(\frac{X_i}{\lambda} - 1)] \\ &= -E[-\frac{X_i}{\lambda^2}] \\ &= \frac{1}{\lambda} \end{aligned}$$

以上より、CR 下限は、

$$Var(\hat{\lambda}) \geq \frac{1}{I_n(\lambda)} = \frac{1}{nI_1(\lambda)} = \frac{\lambda}{n}$$

(別解)
データ n 個の尤度関数は、

$$\begin{aligned} f_n(\lambda|\boldsymbol{X}) &= \Pi_{i=1}^n \frac{\lambda^{X_i}}{X_i!} \exp(-\lambda) \\ &= \frac{\lambda^{\sum_{i=1}^n X_i}}{\Pi_{i=1}^n X_i!} \exp(-n\lambda) \end{aligned}$$

対数を取って、

$$\log f_n(\lambda|\boldsymbol{X}) = \sum_{i=1}^n X_i \log \lambda - \log \Pi_{i=1}^n X_i! - n\lambda$$

となる。スコア関数は、

$$S_n(\lambda|\boldsymbol{X}) = \log_n(\lambda|\boldsymbol{X}) = \frac{\sum_{i=1}^n X_i}{\lambda} - n$$

フィッシャー情報量を (二通りの方法で) 求める。

$$\begin{aligned}
I_n(\lambda) &= E[S_n(\lambda|\boldsymbol{X})^2] \\
&= E[(\frac{\sum_{i=1}^n X_i}{\lambda} - n)^2] \\
&= \frac{1}{\lambda^2}E[(\sum_{i=1}^n X_i)^2] - \frac{2n}{\lambda}E[\sum_{i=1}^n X_i] + n^2 \\
&= \frac{n^2}{\lambda^2}E[\bar{X}] - \frac{2n^2}{\lambda}E[\bar{X}] + n^2 \\
&= \frac{n^2}{\lambda^2}(\frac{\lambda}{n} + \lambda^2) - \frac{2n^2}{\lambda}\lambda + n^2 \\
&= \frac{n}{\lambda} \\
I_n(\lambda) &= -E[\frac{d^2}{d\lambda^2}\log f_n(\lambda|\boldsymbol{X})] \\
&= -E[\frac{d}{d\lambda}(\frac{\sum_{i=1}^n X_i}{\lambda} - n)] \\
&= \frac{n}{\lambda^2}E[\bar{X}] \\
&= \frac{n}{\lambda}
\end{aligned}$$

以上より、CR 下限は、

$$Var(\hat{\lambda}) \geq \frac{1}{I_n(\lambda)} = \frac{\lambda}{n}$$

3. $S_n(\lambda|\boldsymbol{X}) = 0$ と置くと、

$$\hat{\lambda}^{ML} = \frac{1}{n}\sum_{i=1}^n X_i = \bar{X}$$

$Var(\bar{X}) = \frac{\lambda}{n}$ より、CR 下限を達成。

6.13

1. データ 1 個の尤度関数は、

$$f_1(\theta|X_i) = \frac{1}{\sqrt{2\pi\theta}}\exp(-\frac{X_i^2}{2\theta})$$

対数を取って、

$$\log f_1(\theta|X_i) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\theta - \frac{X_i^2}{2\theta}$$

ゆえ、フィッシャー情報量は、

$$\begin{aligned}
I_1(\theta) &= -E[\frac{d^2}{d\theta^2}\log f_1(\theta|X_i)] \\
&= -E[\frac{d}{d\theta}(-\frac{1}{2\theta} + \frac{X_i^2}{2\theta^2})] \\
&= -E[\frac{1}{2\theta^2} - \frac{X_i^2}{\theta^3}] \\
&= -\frac{1}{2\theta^2} + \frac{1}{\theta^3}E[X_i^2] \\
&= -\frac{1}{2\theta^2} + \frac{1}{\theta^2} \\
&= \frac{1}{2\theta^2} \\
\therefore I_n(\theta) &= \frac{n}{2\theta^2}
\end{aligned}$$

2. データ n 個の尤度関数は、

$$\begin{aligned}
f_n(\theta|\boldsymbol{X}) &= \Pi_{i=1}^n \frac{1}{\sqrt{2\pi}}\exp(-\frac{X_i^2}{2\theta}) \\
&= (2\pi\theta)^{-\frac{n}{2}}\exp(-\frac{1}{2\theta}\sum_{i=1}^n X_i^2)
\end{aligned}$$

対数を取って、

$$\log f_n(\theta|\boldsymbol{X}) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\theta - \frac{1}{2\theta}\sum_{i=1}^n X_i^2$$

θ に関する一階条件から最尤推定量を求める。

$$\begin{aligned}
-\frac{n}{2\theta} + \frac{1}{2\theta^2}\sum_{i=1}^n X_i^2 &= 0 \\
\therefore \hat{\theta}^{ML} &= \frac{1}{n}\sum_{i=1}^n X_i^2
\end{aligned}$$

次に、$\hat{\theta}^{ML}$ の期待値と分散を求める。(標本平均の形なので X_i^2 の期待値と分散を求めれば良い。)

$$
\begin{aligned}
E[\hat{\theta}^{ML}] &= E[\frac{1}{n}\sum_{i=1}^{n} X_i^2] = \theta \quad (\because E[X_i^2] = \theta) \\
E[X_i^4] &= \int \frac{1}{\sqrt{2\pi}} x^4 \exp(-\frac{x^2}{2\theta}) dx \\
&= \int \frac{1}{\sqrt{2\pi\theta}} x^3 \{-\theta \exp(-\frac{x^2}{2\theta})\}' dx \\
&= \theta \int \frac{1}{\sqrt{2\pi\theta}} 3x^2 \exp(-\frac{x^2}{2\theta}) dx \\
&= 3\theta^2 \\
\therefore Var(\hat{\theta}^{ML}) &= \frac{2\theta^2}{n}
\end{aligned}
$$

3. (5.1 の方針を参照)

$$\frac{n\hat{\theta}}{\theta} \sim \chi_n^2$$

4.

$$X_1^2, ..., X_n^2 : iid \sim (\theta, 2\theta^2) \Rightarrow_{CLT} \sqrt{n}(\hat{\theta} - \theta) \sim N(0, 2\theta)$$

5. 5.1 参照。

$$\frac{n\hat{\theta}}{\theta} \sim \chi_n^2$$

6.

$$
\begin{gathered}
X_1^2, ..., X_n^2 : iid \sim (\theta, 2\theta^2) \\
\Rightarrow_{CLT} \sqrt{n}(\hat{\theta} - \theta) \sim N(0, 2\theta^2)
\end{gathered}
$$

6.14

1.

$$\begin{aligned}
E[|X_i|] &= \int_{-\infty}^{\infty} |x| \frac{1}{\sqrt{2\pi\theta}} \exp(-\frac{x^2}{2\theta}) dx \\
&= 2\int_0^{\infty} \frac{x}{\sqrt{2\pi\theta}} \exp(-\frac{x^2}{2\theta}) dx \\
&= 2[-\sqrt{\frac{\theta}{2\pi}} \exp(-\frac{x^2}{2\theta})]_0^{\infty} \\
&= \sqrt{\frac{2\theta}{\pi}}
\end{aligned}$$

2. (方針)
MLE の漸近正規性について。正則条件が成り立つ時、パラメータの真の値を θ_0、最尤推定値を $\hat{\theta}$ と置くと、

$$\sqrt{n}(\hat{\theta} - \theta_0) \to_d N(0, \frac{1}{I_1(\theta_0)})$$

(漸近有効) となることを示す。対数尤度関数の微分 $l'(t)$ に、$t = \theta_0$ 周りで Taylor の定理を適用する。

$$l'(t) = l'(\theta_0) + l''(\theta_0)(t - \theta_0) + \frac{l'''(t^*)}{2}(t - \theta_0)^2$$

ただし、t^* は t と θ_0 の間。$t = \hat{\theta}$ とすると、$l'(\hat{\theta}) = 0$ なので、

$$l'(\hat{\theta}) = l'(\theta_0) + l''(\theta_0)(\hat{\theta} - \theta_0) + \frac{l'''(\theta^*)}{2}(\hat{\theta} - \theta_0)^2 = 0$$

(第二項から $\sqrt{n}(\hat{\theta} - \theta_0)$ の形が見える！)

$$\begin{aligned}
\therefore \sqrt{n}(\hat{\theta} - \theta_0) &= -\frac{\sqrt{n}}{l''(\theta_0)} l'(\theta_0) - \frac{\sqrt{n} l'''(\theta^*)}{2l''(\theta_0)}(\hat{\theta} - \theta_0)^2 \\
&= \frac{1}{-l''(\theta_0)/n} \frac{l'(\theta_0)}{\sqrt{n}} \quad (\because (\hat{\theta} - \theta_0) \to_p 0) \qquad (1)
\end{aligned}$$

$$\begin{aligned}
-\frac{1}{n} l'''(\theta_0) &= \frac{1}{n} \sum_{i=1}^{n} \{-\frac{\partial^2}{\partial\theta^2} \log f(X_i|\theta_0)\} \\
&\to_p I_1(\theta_0) \qquad (\because LLN) \qquad (2)
\end{aligned}$$

(上式は、

$$E[-\frac{\partial^2}{\partial\theta^2}\log f(X_i|\theta_0)] = I_1(\theta)$$

なることを用いた)

$$\begin{aligned}
\frac{l'(\theta_0)}{\sqrt{n}} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(X_i|\theta_0) \\
&= \sqrt{n}\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(X_i|\theta_0) \\
&= \sqrt{n}\bar{S}
\end{aligned}$$

ただし、S はデータ一個の対数尤度関数の微分。$S_1,...,S_n\ :\ iid$ で、$E[S_i]=0$、$Var(S_i)=E[S_i^2]=I_1(\theta_0)$ なので、

$$\begin{aligned}
& S_1,...,S_n : iid \sim (0, I_1(\theta_0)) \\
\Rightarrow\ & \sqrt{n}(\bar{S}-0) \to_d N(0, I_1(\theta_0)) \qquad (3)
\end{aligned}$$

(1),(2),(3) より、

$$\sqrt{n}(\hat{\theta}-\theta_0) \to_d \frac{1}{I_1(\theta_0)}N(0, I_1(\theta_0)) = N(0, \frac{1}{I_1(\theta_0)})$$

(答案)
標本平均と期待値を一致させることでモーメント推定量を求める。

$$\begin{aligned}
\sqrt{\frac{2\tilde{\theta}}{\pi}} &= \frac{1}{n}\sum_{i=1}^{n}|X_i| \\
\therefore \sqrt{\tilde{\theta}} &= \frac{1}{n}\sum_{i=1}^{n}\sqrt{\frac{\pi}{2}}|X_i|
\end{aligned}$$

$\sqrt{\frac{\pi}{2}}|X_1|,...,\sqrt{\frac{\pi}{2}}|X_n|$ の期待値 $(=\theta)$、分散がわかれば、その標本平均

$\sqrt{\tilde{\theta}}$ の漸近分布も CLT より得られる。

$$\begin{aligned} Var(\sqrt{\frac{\pi}{2}}|X_i|) &= \frac{\pi}{2}\{E[X_i^2]-(E[|X_i|])^2\} \\ &= \frac{\pi}{2}(\theta-\frac{2\theta}{\pi}) \\ &= (\frac{\pi}{2}-1)\theta \\ \therefore \sqrt{n}(\sqrt{\tilde{\theta}}-\sqrt{\theta}) &\to_d N(0,(\frac{\pi}{2}-1)\theta) \end{aligned}$$

さて、θ の最尤推定量とデータ 1 個のフィッシャー情報量は問 13(2) より、

$$\hat{\theta}=\frac{1}{n}\sum_{i=1}^{n}X_i^2,\ I_1(\theta)=\frac{1}{2\theta^2}$$

なので、

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

ここで、$g(t)=\sqrt{n}$ とすると、$g(\hat{\theta})=\sqrt{\hat{\theta}}$ なので、デルタ法により、

$$\begin{aligned} \sqrt{n}(\sqrt{\hat{\theta}}-\sqrt{\theta}) &\to_d g'(\theta)N(0,2\theta^2) \\ &= \frac{1}{2\sqrt{\theta}}N(0,2\theta^2) \\ &= N(0,\frac{\theta}{2}) \end{aligned}$$

6.15

(方針)
CR 下限の証明。不偏推定量 $\hat{\theta}$ に対して、任意の θ で

$$Var_\theta(\hat{\theta})\geq\frac{1}{I_n(\theta)}$$

が成り立つことを示す。
データ n 個のスコア関数を $S_n(\theta,\boldsymbol{X})$ とおく。$\hat{\theta}-\theta$ と $S_n(\theta,\boldsymbol{X})$ に Cauchy-Schwarz の不等式を適用したい。

$$E[\hat{\theta}-\theta]=0,\ E[S_n(\theta,\boldsymbol{X})]=0$$

より、

$$\begin{aligned}\therefore E[(\hat{\theta}-\theta)S_n(\theta,\boldsymbol{X})]^2 &\le E[(\hat{\theta}-\theta)^2]E[S_n(\theta,\boldsymbol{X})^2]\\ &= Var_\theta(\hat{\theta})I_n(\theta) \quad (1)\end{aligned}$$

最左辺が 1 であれば良い。

$$\begin{aligned}E[(\hat{\theta}-\theta)S_n(\theta,\boldsymbol{X})] &= E[\hat{\theta}S_n(\theta,\boldsymbol{X})]-\theta E[S_n(\theta,\boldsymbol{X})]\\ &= E[\hat{\theta}S_n(\theta,\boldsymbol{X})]\\ &= \int ... \int \hat{\theta}(\boldsymbol{X})S_n(\theta,\boldsymbol{X})f_n(\boldsymbol{X},\theta)dX_1...dX_n\\ &= \int ... \int \hat{\theta}(\boldsymbol{X})\frac{d}{d\theta}\log f_n(\boldsymbol{X},\theta)\times f_n(\boldsymbol{X},\theta)dX_1...dX_n\\ &= \int ... \int \hat{\theta}(\boldsymbol{X})\frac{\frac{d}{d\theta}f_n(\boldsymbol{X},\theta)}{f_n(\boldsymbol{X},\theta)}f_n(\boldsymbol{X},\theta)dX_1...dX_n\\ &= \frac{d}{d\theta}\int ... \int \hat{\theta}(\boldsymbol{X})f_n(\boldsymbol{X},\theta)dX_1...dX_n\\ &= \frac{d}{d\theta}E[\hat{\theta}]=1 \quad (2)\end{aligned}$$

(1)、(2) より、CR 下限が示された。これは本問題の $h(\theta)=\theta$ のケース ($h'(\theta)=1$)。

(答案)
$\hat{h}-h(\theta)$ と $S_n(\theta,\boldsymbol{X})$ に Cauchy-Schwarz の不等式を適用したい。

$$E[\hat{h}-h(\theta)]=0,\ E[S_n(\theta,\boldsymbol{X})]=0$$

より、

$$\begin{aligned}E[\hat{h}-h(\theta)S_n(\theta,\boldsymbol{X})]^2 &\le E[(\hat{\theta}-h(\theta))^2]E[S_n(\theta,\boldsymbol{X})^2]\\ &= Var_\theta(\hat{h})nI_1(\theta)\end{aligned}$$

最左辺を $h'(\theta)^2$ にしたい。

$$\begin{aligned}E[\hat{h}-h(\theta)S_n(\theta,\boldsymbol{X})] &= E[\hat{h}S_n(\theta,\boldsymbol{X})]-h(\theta)E[S_n(\theta,\boldsymbol{X})]\\ &= E[\hat{h}S_n(\theta,\boldsymbol{X})]\\ &= \int ... \int \hat{\theta}\frac{\frac{d}{d\theta}f_n(\boldsymbol{X},\theta)}{f_n(\boldsymbol{X},\theta)}f_n(\boldsymbol{X},\theta)dX_1...dX_n\\ &= \frac{d}{d\theta}\int ... \int \hat{h}f_n(\boldsymbol{X},\theta)dX_1...dX_n\\ &= \frac{d}{d\theta}E[\hat{h}]=h'(\theta)\end{aligned}$$

したがって、CR 下限が示された。

6.16

1. データ n 組の尤度関数は、

$$\begin{aligned} f_n(\boldsymbol{X},\boldsymbol{Y}|\boldsymbol{\mu},\sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{(X_1-\mu_1)^2}{2\sigma^2}\}\times...\times\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{(X_n-\mu_n)^2}{2\sigma^2}\} \\ &\quad \times\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{(Y_1-\mu_1)^2}{2\sigma^2}\}\times...\times\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{(Y_n-\mu_n)^2}{2\sigma^2}\} \\ &= (2\pi\sigma^2)^{-n}\exp[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\{(X_i-\mu_i)^2+(Y_i-\mu_i)^2\}] \end{aligned}$$

対数を取って、

$$\log f_n(\boldsymbol{X},\boldsymbol{Y}|\boldsymbol{\mu},\sigma^2) = -n\log(2\pi)-n\log\sigma^2-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\{(X_i-\mu_i)^2+(Y_i-\mu_i)^2\}$$

これを偏微分して 0 と置くことでパラメータの最尤推定量を得る。

$$\begin{aligned} \frac{\partial\log f_n}{\partial\mu_i} &= \frac{1}{\sigma^2}\{2(X_i-\mu_i)+2(Y_i-\mu_i)\}=0 \\ &\Rightarrow\ \hat{\mu}_i=\frac{X_i+Y_i}{2} \\ \frac{\partial\log f_n}{\partial\sigma^2} &= -\frac{n}{\sigma^2}-\frac{1}{2\sigma^4}\sum_{i=1}^{n}\{(X_i-\mu_i)^2+(Y_i-\mu_i)^2\}=0 \\ &\Rightarrow\ \hat{\sigma}^2=\frac{1}{4n}\sum_{i=1}^{n}(X_i-Y_i)^2 \end{aligned}$$

2. $X_i,Y_i: iid\sim N(\mu_i,\sigma^2)$ より、$X_i-Y_i\sim N(0,2\sigma^2)$ なので、$E[(X_i-Y_i)^2]=2\sigma^2$ である。大数法則により、

$$\frac{1}{n}\sum_{i=1}^{n}(X_i-Y_i)^2\to_p 2\sigma^2$$

であるから、

$$\hat{\sigma}^2\to_p\frac{\sigma^2}{2}$$

となり、最尤推定量は一致性を持たない。一致推定量は

$$2\hat{\sigma}^2=\frac{1}{2n}\sum_{i=1}^{n}(X_i-Y_i)^2$$

7 統計的仮説検定

仮説検定と区間推定に関しては知識が 2 年生で止まっているので復習がてら知ってることを書いていく。
$X_1, ..., X_n : iid \sim f(x|\theta)$ なる標本が得られたときに、パラメータの推定量 $\hat{\theta}(\boldsymbol{X})$ を考えるのが点推定であった。しかし、$\hat{\theta}$ の実現値が真の θ とちょうど同じ値になることはほぼあり得ない。そこで、代わりに θ を含む区間 $[L, U]$ を報告することを考える。

Def. 区間推定量

$X_1, ..., X_n : iid \sim f(x|\theta)$ とする。二つの統計量 $L(\boldsymbol{X}), U(\boldsymbol{X})$ について、$L(\boldsymbol{X}) \leq U(\boldsymbol{X})$ で、全ての θ に対して

$$P_\theta[\theta \in (L(\boldsymbol{X}), U(\boldsymbol{X}))] \geq 1 - \alpha$$

となる区間 (L, U) を信頼係数が $1 - \alpha$ の信頼区間という。
上の P_θ をカバレッジ確率という。区間 (L, U) が真のパラメータの値 θ を含む確率。

$X_1, ..., X_n : iid \sim (\mu, \sigma^2)$ の時、平均 μ の区間推定を考える。点推定量は $\bar{X}$ なので、カバレッジ確率はある c を用いて

$$P[\bar{X} - c < \mu < \bar{X} + c]$$

と表せる。この c を求めよう。$P[\bar{X} - c < \mu < \bar{X} + c] = 1 - \alpha$ は $P[\mu < L \vee U < \mu] = \alpha$ と同値なので、これを変形する。$\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \sim N(0,1)$ を用いる。

$$\begin{aligned}
P[\mu < L \vee U < \mu] &= P[\mu < \bar{X} - c] + P[\bar{X} + c < \mu] \\
&= P[\mu + c < \bar{X}] + P[\bar{X} < \mu - c] \\
&= P[\frac{\sqrt{n}\{(\mu + c) - \mu\}}{\sigma} < \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}] \\
&\quad + P[\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} < \frac{\sqrt{n}\{(\mu - c) - \mu\}}{\sigma}] \\
&= P[\sqrt{n}\frac{c}{\sigma} < \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}] \\
&\quad + P[\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} < -\sqrt{n}\frac{c}{\sigma}] \\
&= 2P[\sqrt{n}\frac{c}{\sigma} < \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}] = \alpha
\end{aligned}$$

よって $\sqrt{n}\frac{c}{\sigma}$ は $N(0,1)$ の上側 $\frac{\alpha}{2}$ 点。これを $z_{\alpha/2}$ と表記することにすると、$c=\frac{\sigma}{\sqrt{n}}z_{\alpha/2}$ となる。したがって、

$$L=\bar{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2}, U=\bar{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}$$

命題

母集団が正規分布の iid 標本において、平均 μ の区間推定量

$$(\bar{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2},\ \bar{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2})$$

は $1-\alpha$ 信頼区間である。
(サンプル数 n が大きいと区間は短くなる。)

同様に、σ^2 が未知の場合は不偏分散 V を用いて $\frac{\sqrt{n}(\bar{X}-\mu)}{V}\sim t_{n-1}$ から、

命題

$X_1,...,X_n : iid \sim N(\mu,\sigma^2)$ で、σ^2 が未知の時、平均 μ の区間推定量

$$(\bar{X}-\frac{V}{\sqrt{n}}t_{n-1,\alpha/2},\ \bar{X}+\frac{V}{\sqrt{n}}t_{n-1,\alpha/2})$$

は $1-\alpha$ 信頼区間である。

さらに、正規分布以外の一般の分布に対しても CLT とスラツキーの定理 (5.10 に詳しい) により、$X_1,...,X_n : iid \sim (\mu,\sigma^2)$ の時

$$\sqrt{n}\frac{\bar{X}-\mu}{\sigma}\to_d N(0,1)$$
$$\sqrt{n}\frac{\bar{X}-\mu}{V}\to_d N(0,1)$$

が言えるのだから、

命題

一般の iid 標本において、平均 μ の区間推定量

$$(\bar{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}, \bar{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2})$$

を考える。$n \to \infty$ の時、カバレッジ確率は $1 - \alpha$ に近づく。

命題

一般の iid 標本において、平均 μ の区間推定量

$$(\bar{X} - \frac{V}{\sqrt{n}} z_{\alpha/2}, \bar{X} + \frac{V}{\sqrt{n}} z_{\alpha/2})$$

を考える。$n \to \infty$ の時、カバレッジ確率は $1 - \alpha$ に近づく。

※正規分布で分散未知の場合では t 分布を用いたが、n が大ならば正規分布に分布収束するため t 分布でなく正規分布を用いて区間推定してもよい。
※区間推定の注意事項について述べる。μ の 95% 信頼区間が $(3.897, 4.003)$ と計算されたとしよう。この時、「μ が $(3.897, 4.003)$ の間に含まれる確率は 95% である」と主張することはできない。これは文章に登場する $\mu, 3.897, 4.003$ が全て確率的に振る舞わない定数 (確率変数でない) だから。すなわち、$L(\boldsymbol{X}), U(\boldsymbol{X})$ は間に μ が含まれる確率が 95% になるように振る舞う確率変数であるが、一度観測されて信頼区間が割り出されたら「そこに μ が含まれる確率は 95%」とは言えないのである。
(信頼区間の実現値に対してパラメータが確率的に含まれたり含まれなかったりするような記述をしてはいけない、ということ。)
これに対し、パラメータを確率変数とみなしてその事後分布を推定するベイズ推定では、パラメータの「信用区間」という、「μ が $(3.897, 4.003)$ の間に含まれる確率は 95% である」と主張できる区間を求めることができる。

ここからは検定の話。統計的仮説検定の概観を示す。教科書では一標本問題に一切触れずに別書への誘導だけが書かれている。簡単すぎるということだろうか。。。

正規分布からの標本

- μ の検定 $\rightarrow$ 分散既知 (z 検定)、分散未知 (t 検定)
- μ の差の検定 (二標本)$\rightarrow$ 分散既知 (等分散を仮定)、分散未知 (等分散を仮定、t 検定)
- σ^2 の差の検定 (二標本 F 検定)

一般の分布からの標本

- 尤度比検定
- 最尤推定量の漸近正規性の利用 $\rightarrow$ ワルド検定、スコア検定
- 標本平均の漸近正規性の利用

上の検定について順に書いていく。個人的に一番難しいのは F 検定。

＜正規母集団の平均の検定 (分散既知、z 検定) ＞
$X_1, ..., X_n : iid \sim N(\mu, \sigma^2)$ とし、$H_0 : \mu = \mu_0 \quad H_1 : \mu \neq \mu_0$ なる検定を考える。検定統計量 W は、

$$W = \frac{|\bar{X} - \mu_0|}{\sigma}$$

とする。この時、誤って帰無仮説を棄却する確率を α にするように臨海値を定める。すなわち、

$$P[\frac{|\bar{X} - \mu_0|}{\sigma} > C|\ H_0 : \mu = \mu_0] = \alpha$$

を満たす C を求める。帰無仮説のもとで $\frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} \sim N(0, 1)$ なので、

$$\begin{aligned}
\alpha &= P[\frac{|\bar{X} - \mu_0|}{\sigma} > C|H_0 : \mu = \mu_0] \\
&= P[\sqrt{n}\frac{|\bar{X} - \mu_0|}{\sigma} > \sqrt{n}C|H_0 : \mu = \mu_0] \\
&= P[\sqrt{n}\frac{\bar{X} - \mu_0}{\sigma} < -\sqrt{n}C,\ \sqrt{n}C < \sqrt{n}\frac{\bar{X} - \mu_0}{\sigma}|H_0 : \mu = \mu_0] \\
&= 2P[\sqrt{n}C < \sqrt{n}\frac{\bar{X} - \mu_0}{\sigma}|H_0 : \mu = \mu_0]
\end{aligned}$$

より、$\sqrt{n}C$ は標準正規分布の上側 $\frac{\alpha}{2}$ 点 $z_{\alpha/2}$ なので、$C = \frac{z_{\alpha/2}}{\sqrt{n}}$ となる。したがって、求める有意水準 α の検定は、

$$\frac{|\bar{X} - \mu_0|}{\sigma} > \frac{z_{\alpha/2}}{\sqrt{n}} \ \Rightarrow \ reject\ H_0$$

＜正規母集団の平均の検定 (分散未知、t 検定) ＞
$X_1, ..., X_n : iid \sim N(\mu, \sigma^2)$ とし、$H_0 : \mu = \mu_0 \quad H_1 : \mu \neq \mu_0$ なる検定を考える。検定統計量 W は、

$$W = \frac{|\bar{X} - \mu_0|}{V}$$

とする。この時、誤って帰無仮説を棄却する確率を α にするように臨海値を定める。すなわち、

$$P[\frac{|\bar{X} - \mu_0|}{V} > C|\ H_0 : \mu = \mu_0] = \alpha$$

を満たす C を求める。帰無仮説のもとで $\frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma} \sim t_{n-1}$ なので、

$$\begin{aligned}
\alpha &= P[\frac{|\bar{X} - \mu_0|}{V} > C|H_0 : \mu = \mu_0] \\
&= P[\sqrt{n}\frac{\bar{X} - \mu_0}{V} < -\sqrt{n}C, \sqrt{n}C < \sqrt{n}\frac{\bar{X} - \mu_0}{V}|H_0 : \mu = \mu_0] \\
&= 2P[\sqrt{n}C < \sqrt{n}\frac{\bar{X} - \mu_0}{V}|H_0 : \mu = \mu_0]
\end{aligned}$$

より、$\sqrt{n}C$ は自由度 $n-1$ の t 分布の上側 $\frac{\alpha}{2}$ 点なので、$C = \frac{t_{n-1,\alpha/2}}{\sqrt{n}}$ となる。したがって、求める有意水準 α の検定は、

$$\frac{|\bar{X} - \mu_0|}{V} > C \ \Rightarrow \ reject\ H_0$$

＜一般の平均の検定 (標本平均の漸近正規性を利用) ＞
CLT により $\hat{\sigma}$ を σ の一致推定量とすれば $\frac{\sqrt{n}(\bar{X}-\mu)}{\hat{\sigma}} \to_d N(0,1)$ となるので、これを用いて有意水準 α の近似的な検定は次のように定められる。

$$\sqrt{n}\frac{|\bar{X} - \mu_0|}{\hat{\sigma}} > z_{\alpha/2} \ \Rightarrow \ reject\ H_0$$

※ $\hat{\sigma}$ に関して、σ が未知の場合でも帰無仮説 $H : \mu = \mu_0$ のもとで既知になる場合がある。例えばベルヌーイ分布のパラメータ p に関する検定では、帰無仮説 $H : p = p_0$ のもとで分散は $p(1-p)$ で既知と言える。

＜ μ の差の検定 (分散既知、等分散を仮定) ＞
$X_1, ..., X_m : iid \sim N(\mu_1, \sigma^2)$、$Y_1, ..., Y_n : iid \sim N(\mu_2, \sigma^2)$ かつ $X_1, ..., X_m, Y_1, ..., Y_n$ は互いに独立であるとする。平均の差の検定を $H_0 : \mu_1 = \mu_2,\ H_1 : \mu_1 \neq \mu_2$ という仮説を用いて行う。検定統計量 W は、

$$W = |\bar{X} - \bar{Y}|$$

$\bar{X} \sim N(\mu_1, \frac{\sigma^2}{m}), \bar{Y} \sim N(\mu_2, \frac{\sigma^2}{n})$ なので、

$$\bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, (\frac{1}{m} + \frac{1}{n})\sigma^2)$$

となるのだから、帰無分布は

$$\bar{X} - \bar{Y} \sim N(0, (\frac{1}{m} + \frac{1}{n})\sigma^2)$$

で与えられる。これを標準化した Z を定義する。

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

したがって有意水準 α の検定は

$$|Z| > z_{\alpha/2} \ \Rightarrow \ reject\ H_0$$

で与えられる。

＜ μ の差の検定 (分散未知、等分散を仮定) ＞

$$Z \sim N(0, 1), U \sim \chi^2_\nu \Rightarrow \frac{Z}{\sqrt{U/\nu}} \sim t_\nu$$

を用いる。分散が未知なのでとりあえず σ^2 の推定量 $\hat{\sigma}^2$ を使うことにしよう (後できちんと与える)。T を下のように定義する。

$$\begin{aligned} T &= \frac{\bar{X} - \bar{Y}}{\sqrt{(\frac{1}{m} + \frac{1}{n})\hat{\sigma}^2}} \\ &= \frac{\bar{X} - \bar{Y}}{\sqrt{(\frac{1}{m} + \frac{1}{n})\sigma^2}} \frac{1}{\sqrt{\frac{\hat{\sigma}^2}{\sigma^2}}} \end{aligned}$$

右辺の左部分は $Z \sim N(0,1)$ である。$\nu\frac{\hat{\sigma}^2}{\sigma^2} = U$ としたとき U/ν が χ^2_ν に従うように $\hat{\sigma}^2$ を作る。このような $\hat{\sigma}^2$ とはプールされた統計量

$$\begin{aligned}\hat{\sigma}^2 &= \frac{m-1}{m+n-2}V_X^2 + \frac{m-1}{m+n-2}V_Y^2 \\ &= \frac{1}{m+n-2}\{\sum_{i=1}^{n}(X_i - \bar{X})^2 + \sum_{j=1}^{m}(Y_j - \bar{Y})^2\}\end{aligned}$$

である (V_X^2, V_Y^2 はそれぞれ X, Y の不偏分散)。すると、$\nu = m+n-2$ で、

$$\begin{aligned}U &= (m+n-2)\frac{\hat{\sigma}^2}{\sigma^2} = (m-1)\frac{V_X^2}{\sigma^2} + (n-1)\frac{V_Y^2}{\sigma^2} \sim \chi^2_{m+n-2} \\ &\because (m-1)\frac{V_X^2}{\sigma^2} \sim \chi^2_{m-1}, (n-1)\frac{V_Y^2}{\sigma^2} \sim \chi^2_{n-1}\end{aligned}$$

まとめると (ここから先を抑えれば十分)、

$$\begin{aligned}Z &= \frac{\bar{X} - \bar{Y}}{\sqrt{(\frac{1}{m} + \frac{1}{n})\sigma^2}} \sim N(0,1) \\ U &= (m+n-2)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{m+n-2} \\ \Rightarrow T &= \frac{Z}{\sqrt{U/(m+n-2)}} = \frac{\bar{X} - \bar{Y}}{\sqrt{(\frac{1}{m} + \frac{1}{n})\hat{\sigma}^2}} \sim t_{m+n-2}\end{aligned}$$

以上より、有意水準 α の検定は

$$|T| > z_{\alpha/2} \Rightarrow reject H_0$$

で与えられる。

$<$ σ^2 の差の検定 (F 検定) $>$

まずは F 分布の導出をする。F 分布は「カイ二乗分布に従う二つの変数の比が従う分布」である。なお F 分布は第二種ベータ分布という「ガンマ分布に従う二つの変数の比が従う分布」の応用版なのでそこから導出する。

$X \sim Ga(\alpha_1, \beta), Y \sim Ga(\alpha_2, \beta)$ とした時、その比 $R = \frac{X}{Y}$ が従う分布を知りたい。4.14 で見た通り、

$$\begin{cases} Z = X + Y \sim Ga(\alpha_1 + \alpha_2, \beta) \\ W = \frac{X}{X+Y} \sim Beta(\alpha_1, \alpha_2) \end{cases}$$

なので、W をさらに変形することで、

$$\begin{aligned} W &= \frac{X}{X+Y} = \frac{X}{X+Y} = \frac{(\frac{X}{Y})}{(\frac{X}{Y})+1} = \frac{R}{R+1} \\ \frac{dW}{dR} &= \frac{1}{(1+R)^2} \end{aligned}$$

よって、R の密度関数がもとまる。

$$\begin{aligned} f_R(r) &= f_W(\frac{r}{1+r}) \times \frac{1}{(1+r)^2} \\ &= \frac{1}{B(\alpha_1,\alpha_2)}(\frac{r}{1+r})^{\alpha_1-1}(\frac{1}{1+r})^{\alpha_2-1}(\frac{1}{1+r})^2 \\ &= \frac{1}{B(\alpha_1,\alpha_2)} r^{\alpha_1-1}(1+r)^{-(\alpha_1+\alpha_2)} \end{aligned}$$

これは第二種ベータ分布の密度関数であり、$R \sim SB(\alpha_1,\alpha_2)$ と表記する。(第二種ベータ分布は、「ベータ分布に従う W を $R=\frac{W}{1-W}$ と変数変換したときに R が従う分布」とも言える)
さて、$\alpha_1=\frac{m}{2},\alpha_2=\frac{n}{2},\beta=\frac{1}{2}$ なる特殊ケースについてもう一度考えよう。つまり、

$$X \sim \chi^2_m = Ga(\frac{m}{2},\frac{1}{2}),\ Y \sim \chi^2_n = Ga(\frac{n}{2},\frac{1}{2})$$

の時

$$R = \frac{X}{Y} \sim SB(\frac{m}{2},\frac{n}{2})$$

である。今、$S=\frac{Y}{m},\ T=\frac{Y}{n}$ として $Z=\frac{S}{T}$ の分布が知りたい。

$$\begin{aligned} Z &= \frac{S}{T} = \frac{X/m}{Y/n} = \frac{n}{m}R \\ \frac{dR}{dZ} &= \frac{m}{n} \end{aligned}$$

より、

$$\begin{aligned} f_Z(z) &= f_R(\frac{m}{n}z)\frac{m}{n} \\ &= \frac{1}{B(\frac{m}{2},\frac{n}{2})}(\frac{m}{n}z)^{\frac{m}{2}-1}(1+\frac{n}{m}z)^{-(\frac{m}{2}+\frac{n}{2})}\frac{m}{n} \\ &= \frac{(\frac{m}{n})^{m/2}}{B(\frac{m}{2},\frac{n}{2})} z^{\frac{m}{n}-1}(1+\frac{m}{n}z)^{-(\frac{m}{2}+\frac{n}{2})} \end{aligned}$$

この分布を自由度 m,n の F 分布といい、$Z \sim F_{m,n}$ とかく。まとめると、

- $X \sim Ga(\alpha_1, \beta),\ Y \sim Ga(\alpha_2, \beta)$ の時、$W = \frac{X}{X+Y} \sim Beta(\alpha_1, \alpha_2)$
- W の変数変換によって、$R = \frac{W}{1-W} = \frac{X}{Y} \sim SB(\alpha_1, \alpha_2)$
- 特に、$X \sim \chi^2_m,\ Y \sim \chi^2_n$ の時 $R = \frac{X}{Y} \sim SB(\frac{m}{2}, \frac{n}{2})$
- R の変数変換によって、$Z = \frac{n}{m}R = \frac{X/m}{Y/n} \sim F_{m,n}$

検定の話に戻る。二つの正規母集団からそれぞれ標本分散が得られたとすると、帰無仮説のもとでその比は F 分布に従う。
$X_1, ..., X_m : iid \sim N(\mu_1, \sigma_1^2),\ Y_1, ..., Y_n : iid \sim N(\mu_1, \sigma_2^2)$ で、X, Y は互いに独立であるとする。$H_0 : \sigma_1^2 = \sigma_2^2,\ H_1 : \sigma_1^2 \neq \sigma_2^2$ なる検定問題を考えよう。検定は、$c_L < c_U$ なる定数 c_L, c_U を用いて次のように表せる。

$$\frac{\hat{\sigma_1}^2}{\hat{\sigma_2}^2} < c_L \quad \vee \quad c_U < \frac{\hat{\sigma_1}^2}{\hat{\sigma_2}^2} \quad \Rightarrow \quad reject \quad H_0$$

普遍分散を用いて $\hat{\sigma_1}^2 = V_X,\ \hat{\sigma_2}^2 = V_Y$ とすると、

$$(m-1)\frac{V_X^2}{\sigma_1^2} \sim \chi^2_{m-1}, \quad (n-1)\frac{V_Y^2}{\sigma_2^2} \sim \chi^2_{n-1}$$

であるから、帰無仮説 $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma^2$ が正しいとき、検定統計量は

$$\frac{V_X^2}{V_Y^2} = \frac{(m-1)\frac{V_X^2}{\sigma^2}/(m-1)}{(n-1)\frac{V_Y^2}{\sigma^2}/(n-1)} = \frac{\chi^2_{m-1}/(m-1)}{\chi^2_{n-1}/(n-1)} \sim F_{m-1,n-1}$$

ゆえ、帰無分布は F 分布であるから、c_L は $F_{m-1,n-1}$ の下側 $\frac{\alpha}{2}$ 分位点で、c_U は上側 $\frac{\alpha}{2}$ 分位点。

＜尤度比検定＞
漸近正規性を必要としない検定。$X_1, ..., X_n : iid \sim f(x|\theta),\ \theta \in \Theta$ とする。

尤度比検定

$H_0 : \theta \in \Theta_0,\ H_1 : \theta \in \Theta_0$ の時、尤度比検定統計量 LRTS は次のように定義される。

$$\lambda(\boldsymbol{X}) = \frac{\sup_{\Theta_0} L(\theta|\boldsymbol{X})}{\sup_{\Theta} L(\theta|\boldsymbol{X})}$$

すると、検定は

$$\lambda(\boldsymbol{X}) < c \ \Rightarrow \ reject\ H_0\ (c < 1)$$

で与えられる。

特殊ケースとして帰無仮説が単純仮説 $H_0 : \theta = \theta_0$ で、最尤推定量が

$$\hat{\theta} = \arg\max_{\theta \in \Theta} L(\theta|\boldsymbol{X})$$

である場合を考える。この時 LRTS は、

$$\lambda(\boldsymbol{X}) = \frac{L(\theta_0|\boldsymbol{X})}{L(\hat{\theta}|\boldsymbol{X})}$$

となる。では棄却のための定数 c はどのように決めるのだろうか。実は、CRTS は分布収束する。

CRTS の漸近的な帰無分布

θ はスカラーで、仮説は単純仮説だとする。ある仮定をおけば、H_0 のもとで

$$-2 \log \lambda(\boldsymbol{X}) \to_d \chi_1^2$$

これを用いれば、サイズ α の漸近的な検定は

$$-2 \log \lambda(\boldsymbol{X}) \to_d \chi_{1,\alpha}^2 \ \Rightarrow \ reject\ H_0$$

＜スコア検定＞

最尤推定量の漸近正規性を用いる。すなわち、一定の条件のもとで

$$\sqrt{n}(\hat{\theta}_n - \theta) \to_d N(0, \frac{1}{I_1(\theta)})$$

であることを使う。$H_0 : \theta = \theta_0$ のもとで、

$$\frac{\sqrt{n}(\hat{\theta}_n - \theta_0)}{\sqrt{1/I_1(\theta_0)}} = \sqrt{nI_1(\theta_0)}(\hat{\theta}_n - \theta_0) \to_d N(0,1) \tag{1}$$

なのだから、検定は

$$|\hat{\theta}_n - \theta_0|\sqrt{nI_1(\theta_0)} > z_{\alpha/2} \ \Rightarrow \ reject\ H_0$$

(1) は

$$(\hat{\theta}_n - \theta_0)^2 I_n(\theta_0) \to_d \chi^2_1$$

ともかけるので、検定は

$$(\hat{\theta}_n - \theta_0)^2 I_n(\theta_0) > \chi^2_{1,\alpha} \ \Rightarrow \ reject\ H_0$$

と書くこともある。

＜ワルド検定＞
スコア検定では分散を帰無仮説のもとで既知 $(1/I_1(\theta_0))$ としたが、ワルド検定ではこれを一致推定する。MLE の一致性により、$I_1(\hat{\theta}_n) \to_p I_1(\theta)$ なので、

$$\frac{\sqrt{n}(\hat{\theta}_n - \theta_0)}{\sqrt{1/I_1(\hat{\theta}_n)}} = \sqrt{nI_1(\hat{\theta}_n)}(\hat{\theta}_n - \theta_0) \to_d N(0,1)$$

したがって、検定は以下の二通りで表される。

$$|\hat{\theta}_n - \theta_0|\sqrt{nI_1(\hat{\theta}_n)} > z_{\alpha/2} \ \Rightarrow \ reject\ H_0$$
$$(\hat{\theta}_n - \theta_0)^2 I_n(\hat{\theta}_n) > \chi^2_{1,\alpha} \ \Rightarrow \ reject\ H_0$$

7.1 標準

1. (フィッシャー情報量に μ が登場しないのでスコア検定とワルド検定は一致する)

(a) 尤度比検定

$$
\begin{aligned}
L_n(\mu|\boldsymbol{X}) &= \Pi_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(X_i-\mu)^2}{2\sigma^2}\} \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-\mu)^2\}
\end{aligned}
$$

で、$\hat{\mu} = \bar{X}$ なので、CRTS は

$$\lambda(\boldsymbol{X}) = \exp[-\frac{1}{2\sigma^2}\{\sum_{i=1}^n (X_i-\mu_0)^2 - \sum_{i=1}^n (X_i-\bar{X})^2\}]$$

ここで、

$$
\begin{aligned}
\sum_{i=1}^n (X_i-\mu_0)^2 - \sum_{i=1}^n (X_i-\bar{X})^2 &= \sum_{i=1}^n (X_i^2 - 2\mu_0 X_i + \mu_0^2) - \sum_{i=1}^n (X_i^2 - 2\bar{X}X_i + \bar{X}^2) \\
&= -2n\bar{X}(\mu_0 - \bar{X}) + n(\mu_0^2 - \bar{X}^2) \\
&= n(\mu_0 - \bar{X})^2
\end{aligned}
$$

より、

$$\lambda(\boldsymbol{X}) = \exp\{-\frac{n}{2\sigma^2}(\mu_0 - \bar{X})^2\}$$

さて、帰無仮説のもとで

$$-2\log\lambda(\boldsymbol{X}) = \frac{n}{\sigma^2}(\mu_0 - \bar{X})^2 \to_d \chi_1^2$$

なので、検定は

$$\frac{n}{\sigma^2}(\mu_0 - \bar{X})^2 \chi_{1,\alpha}^2 \ \Rightarrow\ reject\ H_0$$

(b) スコア検定

スコア関数からデータ n 個のフィッシャー情報量を求める。

$$\begin{aligned}
S_n(\mu|\boldsymbol{X}) &= \frac{d}{d\mu}\log L_n(\mu|\boldsymbol{X}) \\
&= \frac{d}{d\mu}\{-\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2\} \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu) \\
\therefore I_n(\mu) &= E[-\frac{d}{d\mu}S_n(\mu|\boldsymbol{X})] \\
&= E[\frac{n}{\sigma^2}] \\
&= \frac{n}{\sigma^2}
\end{aligned}$$

ゆえに、$I_n(\mu_0) = \frac{n}{\sigma^2}$ なので

$$\frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma} \to_d N(0,1)$$

なのだから、検定は

$$\frac{\sqrt{n}|\bar{X}-\mu_0|}{\sigma} > z_{\alpha/2} \ \Rightarrow \ reject\ H_0$$

(c) ワルド検定
$I_n(\hat{\mu}) = \frac{n}{\sigma^2}$ なので、検定はスコア検定と同じく

$$\frac{\sqrt{n}|\bar{X}-\mu_0|}{\sigma} > z_{\alpha/2} \ \Rightarrow \ reject\ H_0$$

2. (a) 尤度比検定

$$L(\sigma^2|\boldsymbol{X}) = (2\pi\sigma^2)^{-\frac{n}{2}}\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2\}$$

μ が既知の時、σ^2 の推定量 $\hat{\sigma}^2$ は

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$$

なので、CRTS は

$$\begin{aligned}\lambda(\boldsymbol{X}) &= \{\frac{1}{\sigma_0^2 n}\sum_{i=1}^{n}(X_i-\mu)^2\}^{n/2}\exp\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(X_i-\mu)^2+\frac{n}{2}\} \\ &= (\frac{\hat{\sigma}^2}{\sigma_0^2})^{n/2}\exp\{-\frac{n}{2}(\frac{\hat{\sigma}^2}{\sigma_0^2}-1)\}\end{aligned}$$

となる。したがって、帰無仮説のもとで

$$-2\log\lambda(\boldsymbol{X}) = \frac{n}{2}(\frac{\hat{\sigma}^2}{\sigma_0^2}-\log\frac{\hat{\sigma}^2}{\sigma_0^2}-1) \to_d \chi_1^2$$

よって、検定は

$$\frac{n}{2}(\frac{\hat{\sigma}^2}{\sigma_0^2}-\log\frac{\hat{\sigma}^2}{\sigma_0^2}-1) > \chi_{1,\alpha}^2 \;\Rightarrow\; reject\ H_0$$

(b) スコア検定
スコア関数からフィッシャー情報量を求める。

$$\begin{aligned}S_n(\sigma^2|\boldsymbol{X}) &= \frac{d}{d\sigma^2}\log L(\sigma^2|\boldsymbol{X}) \\ &= \frac{d}{d\sigma^2}\{-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log(\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2\} \\ &= -\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}(X_i-\mu)^2 \\ \therefore I_n(\sigma^2) &= E[-\frac{d}{d\sigma^2}S_n(\sigma^2|\boldsymbol{X})] \\ &= E[-\frac{n}{2\sigma^4}+\frac{1}{\sigma^6}\sum_{i=1}^{n}(X_i-\mu)^2] \\ &= -\frac{n}{2\sigma^4}+\frac{n}{\sigma^6}E[\frac{1}{n}\sum_{i=1}^{n}\sum_{i=1}^{n}(X_i-\mu)^2] \\ &= \frac{n}{2\sigma^4}\end{aligned}$$

より、$I_n(\sigma_0^2)=\frac{n}{2\sigma_0^2}$ で、

$$\frac{\sqrt{n}(\hat{\sigma}^2-\sigma_0^2)}{\sqrt{2}\sigma_0^2} \to N(0,1)$$

であるから、検定は

$$\frac{\sqrt{n}(\hat{\sigma}^2-\sigma_0^2)}{\sqrt{2}\sigma_0^2} > z_{\alpha/2} \;\Rightarrow\; reject\ H_0$$

(c) ワルド検定
$I_n(\hat{\sigma}^2)=\frac{n}{2\hat{\sigma}^2}$ なので、検定は

$$\frac{\sqrt{n}(\hat{\sigma}^2-\sigma_0^2)}{\sqrt{2}\hat{\sigma}^2} > z_{\alpha/2} \;\Rightarrow\; reject\ H_0$$

7.2 標準

1. 尤度関数は

$$\begin{aligned} L(\mu|\boldsymbol{X}) &= \Pi_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{(X_i-\mu)^2}{2\sigma^2}\} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}}\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-\mu)^2\} \end{aligned}$$

Θ 全域で尤度を最大化する最尤推定量は $\bar{X}$。今、尤度関数は $\bar{X}$ を頂点とする放物線なので、Θ_0、すなわち $\mu<\mu_0$ なる μ においては、μ_0 が $\bar{X}$ の右にあるか左にあるかで最大値を与える μ は変わってくる。つまり、$\hat{\mu}_0=\min(\bar{X},\mu 0)$
CRTS は、

$$\begin{aligned} \lambda(\boldsymbol{X}) &= \exp[-\frac{1}{2\sigma^2}\{\sum_{i=1}^n (X_i-\hat{\mu}_0)^2-\sum_{i=1}^n (X_i-\bar{X})^2\}] \\ &= \exp\{-\frac{n}{2\sigma^2}(\bar{X}-\hat{\mu}_0)^2\} \\ \therefore -2\log\lambda(\boldsymbol{X}) &= \begin{cases} 1 & if\ \ \bar{X}<\mu_0 \\ \frac{n}{\sigma^2}(\bar{X}-\mu_0)^2 & otherwise \end{cases} \end{aligned}$$

より、検定は

$$\begin{cases} \bar{X}<\mu_0 \;\Rightarrow\; accept\ H_0 \\ \frac{n}{\sigma^2}(\bar{X}-\mu 0)^2 > c \;\Rightarrow\; reject\ H_0 \end{cases}$$

で与えられる。誤棄却の確率が α となるように棄却域を設定したい。期無分布 $N(\mu_0,\sigma^2)$ のもとで、

$$\frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma} \sim N(0,1)$$

が成り立つのだから、

$$\frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma} > z_\alpha$$

を棄却域とすれば、有意水準 α の検定となる。

2. (略解がいくつが誤植しているので注意)
尤度関数は、

$$L(\sigma^2|\boldsymbol{X}) = (2\pi\sigma^2)^{-\frac{n}{2}}\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2\}$$

μ が既知の時、Θ 全域で尤度を最大化する最尤推定量は $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$。今、尤度関数は $\hat{\sigma}^2$ を頂点とする放物線なので、Θ_0、すなわち $\sigma^2 < \sigma_0^2$ なる σ^2 においては、σ_0^2 が $\hat{\sigma}^2$ の右にあるか左にあるかで最大値を与える σ^2 は変わってくる。つまり、$\hat{\sigma}_0^2 = \min(\hat{\sigma}^2, \sigma_0^2)$。
CRTS は、

$$\begin{aligned}\lambda(\boldsymbol{X}) &= (\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2})^{n/2}\exp\{\frac{n}{2}(1-\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2})\} \\ &= \begin{cases} 1 & if\ \hat{\sigma}^2 < \sigma_0^2 \\ (\frac{\hat{\sigma}^2}{\sigma_0^2})^{n/2}\exp\{\frac{n}{2}(1-\frac{\hat{\sigma}^2}{\sigma_0^2})\} & otherwise \end{cases}\end{aligned}$$

ゆえに、

$$-2\log\lambda(\boldsymbol{X}) = \begin{cases} 0 & if\ \hat{\sigma}^2 < \sigma_0^2 \\ \frac{n\hat{\sigma}^2}{\sigma_0^2} - n\log\frac{\hat{\sigma}^2}{\sigma_0^2} - n & if\ \sigma_0^2 < \hat{\sigma}^2 \end{cases}$$

$t > 0$ において $nt - \log t - n$ は単調増加より、検定は

$$\frac{n\hat{\sigma}^2}{\sigma_0^2} > C \ \Rightarrow\ reject\ H_0$$

で与えられる。期無分布 $N(0,1)$ のもとで

$$\frac{n\hat{\sigma}^2}{\sigma_0^2} \sim \chi_n^2$$

を棄却域とすれば、有意水準 α の検定となる。

7.3 難

1. Θ 全域において μ, σ^2 の最尤推定量は

$$\hat{\mu} = \bar{X},\ \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Θ_0 においては

$$\hat{\mu}_0 = \mu_0,\ \hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_0)^2$$

今、尤度関数が

$$L(\mu, \sigma^2|\boldsymbol{X}) = (2\pi\sigma^2)^{-\frac{n}{2}}\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2\}$$

で与えられるので、CRTS は

$$\begin{aligned}\lambda(\boldsymbol{X}) &= (\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2})^{n/2}\exp(-\frac{n\sigma_0^2}{2\hat{\sigma}_0^2} + \frac{n\hat{\sigma}^2}{2\hat{\sigma}^2}) \\ &= (\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2})^{n/2}\end{aligned}$$

したがって、検定は

$$(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2})^{n/2} < C \Leftrightarrow \frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} > C' \ \Rightarrow\ reject\ H_0$$

今、

$$\begin{aligned}\hat{\sigma}_0^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_0)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\{(X_i - \bar{X})^2 - (\mu_0 - \bar{X})\}^2 \\ &= \hat{\sigma}^2 + (\bar{X} - \mu_0)^2\end{aligned}$$

より、検定は

$$\frac{(\bar{X} - \mu_0)^2}{\hat{\sigma}^2} > C''$$

と変形できる。

$$V^2 = \frac{n}{n-1}\hat{\sigma}^2$$

とすると、帰無仮説のもとで $\frac{\sqrt{n}(\bar{X}-\mu_0)}{V} \sim t_{n-1}$ が成り立つので、

$$\frac{\sqrt{n}|\bar{X}-\mu_0|}{V} > t_{n-1,\alpha/2}$$

を棄却域とすればこれは有意水準 α の検定である。