中国科学技术大学

2021—2022学年第二学期考试试卷

考试科目 概率论与数理统计　　　得分 ________

一.

(1) 知识点：条件分布，二项分布.

记 A 为"出现正面次数为 k"，B 为"第1次为正面"

X_i 为第 i 次抛出正面，$Y := \sum_{i=1}^{n} X_i \sim B(n, \frac{1}{2})$

则 $P(B|A) = \dfrac{P(AB)}{P(A)} = \dfrac{P(Y=k, X_1=1)}{P(Y=k)} = \dfrac{\frac{1}{2} \cdot C_{n-1}^{k-1} (\frac{1}{2})^{n-1}}{C_n^k \cdot (\frac{1}{2})^n} = \dfrac{k}{n}$

(2) 概率运算. $P(A) = P(AB) + P(A\bar{B})$

$P(\max\{X,Y\}=1) = P(X=1, X \geq Y) + P(Y=1, X < Y)$

$P(\min\{X,Y\}=1) = P(X=1, X \leq Y) + P(Y=1, X > Y)$

上下相加 $\Rightarrow P(X=1) + P(Y=1) = p_2 + p_3$

$P(Y=1) = p_2 + p_3 - P(X=1) = p_2 + p_3 - p_1$

(3) 二元正态参数性质. 卡方分布

$(X,Y) \sim N(0,0;1,1;0)$. $\rho=0$ 说明 X,Y 相互独立. $\Rightarrow Z := X^2 + Y^2 \sim \chi^2(2)$

$E(\sqrt{Z}) = \int_0^{\infty} \sqrt{z} \cdot \frac{1}{2} e^{-\frac{z}{2}} dz = \frac{1}{2} \int_0^{\infty} \sqrt{2t} \cdot e^{-t} \cdot 2\,dt = \sqrt{2}\,\Gamma(\frac{3}{2}) = \frac{\sqrt{2}}{2} \Gamma(\frac{1}{2}) = \frac{\sqrt{2\pi}}{2}$

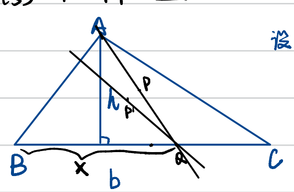
$\chi^2(n)$ 的 pdf：$f(z) = \dfrac{(1/2)^{\frac{n}{2}}}{\Gamma(n/2)} z^{\frac{n}{2}-1} e^{-\frac{z}{2}}$

$= Ga(\frac{n}{2}, \frac{1}{2})$　　　$Ga(\alpha, \lambda)$ pdf: $f(x) = \dfrac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$　$x > 0$

(4) 重期望公式

$E(X) = E[E(X|Y)] = E[Y^2] = Var(Y) + [E(Y)]^2 = 2 + 2^2 = 6$

(5) 几何概型



设 $|BC| = b$，$|BQ| = x$，则 $X \sim U(0,b)$，"相交"为事件 A

$P(A) = \int_0^b P(A|X=x) \cdot \frac{1}{b} dx$

$= \int_0^1 \dfrac{S_{\triangle ABQ}}{S_{\triangle ABC}} \cdot \frac{1}{b} dx$

$= \int_0^b \frac{x}{b^2} dx = \frac{1}{2}$

(6) F分布的构造

$X \sim N(0,1)$. $Y \sim N(0,1)$ 且 X, Y 相互独立.

$X+Y \sim N(0,2)$. $X-Y \sim N(0,2)$

由 $\bar{X}$ 与 S^2 独立 可推 $(X+Y)^2$ 与 $(X-Y)^2$ 独立. 或利用 $Cov(X+Y, X-Y)=0$ 也可推独立.

$(X+Y)^2/(X-Y)^2 = (\frac{X+Y}{2})^2/(\frac{X-Y}{2})^2 \sim F_{1,1}$

(7) 卡方分布的构造

$X_1, X_2, \cdots X_n \sim N(0, \sigma^2)$ $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(0, \frac{\sigma^2}{n})$. $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

(A) ✗ $\sum_{i=1}^{n} (\frac{X_i}{\sigma})^2 \sim \chi^2_n$

(B) X_i 与 S^2 不独立

$\bar{X}$ 与 S^2 独立. 且 $\frac{\sqrt{n}\bar{X}}{\sigma} \sim N(0,1)$ $\frac{n(\bar{X})^2}{\sigma^2} = n \cdot \frac{(\frac{1}{n}\sum_{i=1}^{n} X_i)^2}{\sigma^2} = \frac{1}{n}(\frac{\sum_{i=1}^{n} X_i}{\sigma})^2 \sim \chi^2(1)$

$\therefore \frac{1}{n}(\sum_{i=1}^{n} \frac{X_i}{\sigma})^2 + \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_n$ (D) 正确

(8) 估计量的无偏性. 有效性 $E(X+Y)$, $Var(X+Y)$ 的运算.

$E(\alpha\hat{\mu}_1 + \beta\hat{\mu}_2) = (\alpha+\beta)\mu$

$Var(\alpha\hat{\mu}_1 + \beta\hat{\mu}_2) = \alpha^2 + 2\beta^2 + 2\alpha\beta \cdot (\sqrt{2} \cdot \frac{\sqrt{2}}{4}) = \alpha^2 + 2\beta^2 + \alpha\beta$

$f(\alpha) = \alpha^2 + 2(1-\alpha)^2 + \alpha(1-\alpha) = 2\alpha^2 - 3\alpha + 1$. $\alpha_{min} = \frac{3}{4}$. $\beta_{min} = \frac{1}{4}$ $\alpha\beta = \frac{3}{16}$

(9) 置信区间. 置信上下限

$X_1, X_2, \cdots X_n \sim N(\mu, \sigma^2)$ σ^2 未知   则 $\sqrt{n}(\bar{X}-\mu)/S \sim t(n-1)$

置信区间 $[\bar{X} - t_{\frac{\alpha}{2}}(n-1) S/\sqrt{n},\ \bar{X} + t_{\frac{\alpha}{2}}(n-1) S/\sqrt{n}]$

置信下限 $\bar{X} - t_\alpha(n-1) S/\sqrt{n}$

置信上限 $\bar{X} + t_\alpha(n-1) S/\sqrt{n}$

(10) 显著性水平

拒绝域为 W. 根据显著性水平的定义. $P_0(X \in W) \le \alpha$.

α 从 0.05 ↓ 0.01. W 范围变小. ⇒ 接受 H_0

二. $(X,Y) \sim f(x,y) = cx(y-x)e^{-y}$, $0 \le x \le y < \infty$

(1) 二元联合密度的性质 $\begin{cases} f(x,y) \geq 0 & \text{非负性} \\ \iint f(x,y)\,dy\,dx = 1 \end{cases}$

$$\int_0^{\infty}\int_x^{\infty} cx(y-x)e^{-y}\,dy\,dx = \int_0^{\infty}\int_0^{y} cx(y-x)e^{-y}\,dx\,dy$$

$$= \int_0^{\infty} ce^{-y}\cdot\left[\int_0^{y} yx-x^2\,dx\right]dy$$

$$= c\int_0^{\infty} e^{-y}\cdot\left(\tfrac{1}{6}y^3\right)dy$$

$$= c\cdot\frac{\Gamma(4)}{6} = c = 1$$

(2) 先求边际密度，再根据条件密度的定义.

$$f_X(x) = \int_x^{+\infty} x(y-x)e^{-y}\,dy = x\cdot\int_x^{+\infty} ye^{-y}\,dy - x^2\int_x^{+\infty} e^{-y}\,dy = x\cdot(-(y+1))e^{-y}\Big|_x^{\infty} + x^2e^{-y}\Big|_x^{\infty}$$

$$= xe^{-x}, \quad 0 < x < \infty. \qquad X \sim Ga(2,1)$$

$$f_Y(y) = \int_0^{y} x(y-x)e^{-y}\,dx = \frac{1}{6}y^3e^{-y}, \quad 0 < y < \infty \qquad Y \sim Ga(4,1)$$

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = 6x(y-x)y^{-3} \quad 0 \le x \le y$$

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = (y-x)e^{x-y} \quad x \le y < \infty$$

(3)

① $E[XY] = \int_0^{\infty}\int_0^{y} xy\,f(x,y)\,dx\,dy = \int_0^{\infty} ye^{-y}\int_0^{y} x^2(y-x)\,dx$

$$= \int_0^{\infty} ye^{-y}\cdot\tfrac{1}{12}y^4\,dy = \tfrac{1}{12}\Gamma(6) = 10$$

② 如答案所给.

(4) $X \sim Ga(\alpha,\lambda)$ 则 $E(X) = \frac{\alpha}{\lambda}$. $Var(X) = \frac{\alpha}{\lambda^2}$ 相关系数 Corr 的计算.

$X \sim Ga(2,1)$ $\quad E(X) = 2$. $\quad Var(X) = 2$

$Y \sim Ga(4,1)$ $\quad E(Y) = 4$ $\quad Var(Y) = 4$

$$Corr(X,Y) = \frac{E(XY) - E(X)E(Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

$$= \frac{10 - 2\times 4}{\sqrt{2}\cdot\sqrt{4}} = \frac{\sqrt{2}}{2}$$

均值为2

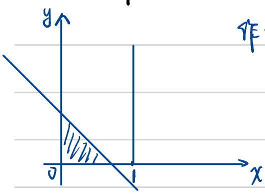
三. $X\sim U(0,1)$　　$Y\sim Exp(1/2)$

(1) X, Y相互独立　$f(x,y)=f_X(x)\,f_Y(y)$

$f(x,y)=\frac{1}{2}e^{-\frac{1}{2}y}$,　$0<x<1$, $y>0$

(2)

法一. 分布函数法.　$Z=X+Y$. 先求Z的分布函数 $F_Z(z)$.



作曲线簇 $x+y=z$. 得z的分段点为0, 1.

当 $z<0$ 时. $F_Z(z)=P(X+Y\le z)=0$

当 $0\le z<1$ 时. $F_Z(z)=\int_0^z\int_0^{z-x}\frac{1}{2}e^{-\frac{1}{2}y}\,dy\,dx$
$=\int_0^z -e^{-\frac{1}{2}y}\Big|_0^{z-x}dx=\int_0^z 1-e^{-\frac{1}{2}(z-x)}dx$
$=z-2+2e^{-\frac{z}{2}}$

当 $z\ge 1$ 时. $F_Z(z)=\int_0^1\int_0^{z-x}\frac{1}{2}e^{-\frac{1}{2}y}\,dy\,dx=\int_0^1 1-e^{-\frac{1}{2}(z-x)}dx$
$=1-2e^{-\frac{1}{2}(z-1)}+2e^{-\frac{z}{2}}$

得Z的密度:
$$f_Z(z)=\begin{cases}0 & z<0\\ 1-e^{-z/2} & 0\le z<1\\ e^{-\frac{1}{2}(z-1)}-e^{-\frac{1}{2}z}=(e^{1/2}-1)e^{-\frac{1}{2}z}, & z\ge 1\end{cases}$$

法二: 卷积公式　$f_Z(z)=\int_{-\infty}^{\infty}f_X(z-y)\,f_Y(y)\,dy$

当 $z<0$ 时. $f_Z(z)=0$

当 $0<z<1$ 时. $f_Z(z)=\int_0^z\frac{1}{2}e^{-1/2y}dy=-e^{-\frac{1}{2}y}\Big|_0^z=1-e^{-\frac{z}{2}}$

当 $z\ge 1$ 时. $f_Z(z)=\int_{z-1}^{z}\frac{1}{2}e^{-1/2y}dy=e^{-\frac{1}{2}(z-1)}-e^{-\frac{1}{2}z}$

结果一致.

法三. 增补变量法.

$\begin{cases}Z=X+Y\\ W=X\end{cases}\Rightarrow\begin{cases}X=W\\ Y=Z-W\end{cases}$　　雅可比行列式 $J=\begin{vmatrix}0 & 1\\ 1 & -1\end{vmatrix}=-1$

$f(z,w)=\frac{1}{2}e^{-\frac{1}{2}(z-w)}$　　$w<z<\infty$

当 $z<0$,　$f_Z(z)=0$

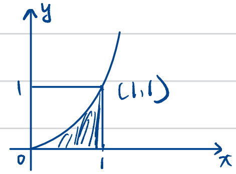
当 $0\le z<1$,　$f_Z(z)=\int_0^z\frac{1}{2}e^{-\frac{1}{2}(z-w)}dw=1-e^{-\frac{z}{2}}$

当 $z\ge 1$　$f_Z(z)=\int_0^1\frac{1}{2}e^{-\frac{1}{2}(z-w)}dw=e^{-\frac{1}{2}(z-1)}-e^{\frac{1}{2}z}$

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

则直接与法二一致.

(3) $t^2+2Xt+Y=0$ 有实根 $\Leftrightarrow$ $4X^2-4Y\geq 0$ $\Rightarrow$ $X^2\geq Y$



$$P(Y\leq X^2)=\iint_G \frac{1}{2}e^{-\frac{1}{2}y}\,dy\,dx$$

$$=\int_0^1\int_0^{x^2}\frac{1}{2}e^{-\frac{1}{2}y}\,dy\,dx$$

标准正态pdf $\approx \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$

$$=\int_0^1\left(1-e^{-\frac{x^2}{2}}\right)dx$$

$$=1-\sqrt{2\pi}\,[\Phi(1)-\Phi(0)]$$ $\nearrow \frac{1}{2}$

附表：$\Phi(1)=0.8413$

四. $X\sim \text{Pareto}(\alpha, c)$. $\quad f(x)=\frac{\alpha c^\alpha}{x^{\alpha+1}}$, $x\geq c$

(1) 矩估计. (不唯一)

$$E(X)=\int_c^{\infty}x\cdot\frac{\alpha c^\alpha}{x^{\alpha+1}}\,dx=\int_c^{\infty}\alpha c^\alpha\cdot x^{-\alpha}dx=\alpha c^\alpha\cdot\frac{1}{-\alpha+1}\alpha^{-\alpha+1}\Big|_c^{\infty}=\frac{c\alpha}{1-\alpha}$$

因为需要给出两个参数的估计，仅由 $E(X)$ 无法解出。

$$E(X^2)=\int_c^{\infty}x^2\cdot\frac{\alpha c^\alpha}{x^{\alpha+1}}\,dx=\frac{c^2\alpha}{\alpha-2}$$

$$Var(X)=E(X^2)-[E(X)]^2=\frac{c^2\alpha}{(\alpha-1)^2(\alpha-2)}$$

$$\begin{cases}E(X)=\frac{c\alpha}{1-\alpha} & ① \\ Var(X)=\frac{c^2\alpha}{(\alpha-1)^2(\alpha-2)} & ②\end{cases}$$

$$\frac{[E(X)]^2}{Var(X)}=\alpha(\alpha-2)\Rightarrow(\alpha-1)^2=1+\frac{[E(X)]^2}{Var(X)}\Rightarrow\hat{\alpha}_1=1+\sqrt{1+\frac{(\bar{X})^2}{S^2}}$$ 且 $\alpha>2$

代α

$$\frac{E(X)\cdot(1-\alpha)}{\alpha}=c\Rightarrow\hat{c}_1=\frac{\bar{X}\cdot(1-\hat{\alpha})}{\hat{\alpha}}$$

(2) 极大似然估计

似然函数. $\quad L(\alpha,c)=\prod_{i=1}^{n}f(x_i;\alpha,c)=\alpha^n c^{n\alpha}\prod_{i=1}^{n}x_i^{-(\alpha+1)}\mathbb{1}(x_i>c)$

$$=\alpha^n c^{n\alpha}\prod_{i=1}^{n}x_i^{-(\alpha+1)}\mathbb{1}(x_{(1)}>c)$$

$c\uparrow \quad L(\alpha,c)\uparrow \qquad \hat{c}_2=X_{(1)}$

$$\ln L(\alpha,c)=n\ln\alpha+n\alpha\ln c-(\alpha+1)\sum_{i=1}^{n}\ln x_i$$

$$\frac{\partial\ln L(\alpha,c)}{\partial\alpha}=\frac{n}{\alpha}+n\ln c-\sum_{i=1}^{n}\ln x_i=0$$

$$\alpha=\frac{n}{\sum_{i=1}^{n}\ln x_i-n\ln c}\Rightarrow\hat{\alpha}=\frac{n}{\sum_{i=1}^{n}\ln x_i-n\ln x_{(1)}}$$

(3) 最大值 $X_{(n)}$ 的分布.

$\hat{c}_2 = X_{(n)} := Y$ 其分布函数为 $F_Y(y) = 1-[1-F(y)]^n$

密度函数为 $f_Y(y) = n[1-F(y)]^{n-1} f(y)$

这里 $F(y) = \begin{cases} \int_c^y \frac{\alpha c^\alpha}{x^{\alpha+1}} dx = c^\alpha(-x^{-\alpha})\Big|_c^y = 1-c^\alpha y^{-\alpha} & y \geq c \\ 0 & y<c \end{cases}$

$f_Y(y) = n \cdot (c^\alpha y^{-\alpha})^{n-1} \cdot (\alpha c^\alpha \cdot y^{-(\alpha+1)})$

$= \frac{n\alpha \cdot c^{n\alpha}}{y^{n\alpha+1}}, \quad y \geq c \qquad Y \sim \text{Pareto}(n\alpha, c)$

$E(Y) = \frac{n\alpha \cdot c}{n\alpha-1} \neq c$

修正: $\frac{n\alpha-1}{n\alpha} \cdot X_{(n)} := \hat{c}_2^*$

五. 成对数据检验. (注意与两样本t检验区分) H_0 的提法

8个成人. 早晚身高. 数据. 属于同一个人. 所以用成对数据检验.

序号	1	2	3	……	8
早 x_i					
晚 y_i					

差 $z_i = x_i - y_i$

$Z = X-Y \sim N(\mu, \sigma^2)$ 检验早晨身高是否显著高于晚上身高

$H_0: \mu \leq 0$ vs $H_1: \mu>0$ 拒绝域 $W = \{t > t_\alpha(n-1)\}$

$t = \frac{\bar{z}-0}{S_z/\sqrt{n}} = 2.758 > t_{0.05}(7) = 1.895$

$\geq$