概率统计 (A) 课程作业: 数理统计的基本概念

周书予
2000013060@stu.pku.edu.cn

April 30, 2022

1

1. 对于 $x \in \{0,1\}^n$, 记 $k = \sum_{i=1}^{n} x_i$, 有

$$\mathbb{P}(X_1 = x_1, \cdots, X_n = x_n) = p^k (1-p)^{n-k}$$

2.

$$\begin{aligned}
\mathbb{E}\left[\overline{X}\right] &= \mathbb{E}[X] = p \\
\mathrm{Var}\left[\overline{X}\right] &= \mathbb{E}\left[\overline{X}^2\right] - \mathbb{E}\left[\overline{X}\right]^2 = \frac{n\mathbb{E}\left[X^2\right] - n(n-1)\mathbb{E}[X]^2}{n^2} - \mathbb{E}[X]^2 = \frac{\mathrm{Var}[X]}{n} = \frac{p(1-p)}{n} \\
\mathbb{E}\left[S^2\right] &= \mathrm{Var}[X] = p(1-p)
\end{aligned}$$

2

$$\begin{aligned}
\mathbb{E}\left[\overline{X}\right] &= \mathbb{E}[X] = m \\
\mathrm{Var}\left[\overline{X}\right] &= \mathbb{E}\left[\overline{X}^2\right] - \mathbb{E}\left[\overline{X}\right]^2 = \frac{n\mathbb{E}\left[X^2\right] - n(n-1)\mathbb{E}[X]^2}{n^2} - \mathbb{E}[X]^2 = \frac{\mathrm{Var}[X]}{n} = \frac{2m}{n} \\
\mathbb{E}\left[S^2\right] &= \mathrm{Var}[X] = 2m
\end{aligned}$$

3

$T = \frac{X}{\sqrt{Y/n}}$, 其中 $X \sim \mathcal{N}(0,1), Y \sim \chi^2(n)$ 且 X, Y 独立, 则 $T^2 = \frac{X^2}{Y/n}$, 注意到 $X^2 \sim \chi^2(1)$, 故 $T^2 \sim F(1,n)$.

4

$F = \frac{X/n_1}{Y/n_2}$, 其中 $X \sim \chi^2(n_1), Y \sim \chi^2(n_2)$ 且 X, Y 独立, 则 $1/F = \frac{Y/n_2}{X/n_1}$, 故 $1/F \sim F(n_2, n_1)$.

5

引理 1. 考虑多元正态分布 $\mathbf{X} \sim \mathcal{N}(\mathbf{a}, B)$, 其中 $\mathbf{a}$ 是分布的期望 (均值), B 是分布的协方差矩阵. 对于任意可逆矩阵 $A \in \mathbb{R}^{n\times n}$, 有 $A\mathbf{X} \sim \mathcal{N}(A\mathbf{a}, ABA^{\mathrm{T}})$.

证明. 记 $\mathbf{Y}=A\mathbf{X}$, 根据密度变换, 可得 $\mathbf{Y}$ 的概率密度函数为

$$
\begin{aligned}
f_{\mathbf{Y}}(\mathbf{y}) &= \left|\frac{\partial \mathbf{x}}{\partial \mathbf{y}}\right| f_{\mathbf{X}}(\mathbf{x}) = \left|\frac{\partial A^{-1}\mathbf{y}}{\partial \mathbf{y}}\right| f_{\mathbf{X}}(A^{-1}\mathbf{y}) \\
&= \frac{1}{\det(A)} \cdot \frac{1}{(2\pi)^{n/2}\sqrt{\det(B)}} \exp\left(-\frac{1}{2}(A^{-1}\mathbf{y}-\mathbf{a})^{\mathrm{T}} B^{-1}(A^{-1}\mathbf{y}-\mathbf{a})\right) \\
&= \frac{1}{(2\pi)^{n/2}\sqrt{\det(ABA^{\mathrm{T}})}} \exp\left(-\frac{1}{2}(\mathbf{y}-A\mathbf{a})^{\mathrm{T}}(ABA^{\mathrm{T}})^{-1}(\mathbf{y}-A\mathbf{a})\right)
\end{aligned}
$$

故证明了 $\mathbf{Y}\sim\mathcal{N}(A\mathbf{a}, ABA^{\mathrm{T}})$. □

当 $\mathbf{X}\sim\mathcal{N}(\mathbf{0}, I_n)$ 时, $\mathbf{Y}=A\mathbf{X}\sim\mathcal{N}(A\mathbf{0}, AI_nA^{\mathrm{T}})=\mathcal{N}(\mathbf{0}, AA^{\mathrm{T}})$, 故 $\mathbf{Y}\sim\mathcal{N}(\mathbf{0}, I_n)$ 是 n 维标准正态分布的充要条件是 $AA^{\mathrm{T}}=I_n$, 也即 A 是正交矩阵.

6

取 $\mathbb{R}^n$ 中向量

$$
\begin{aligned}
\alpha &= \left(\frac{(t_1-\bar{t})}{\sqrt{\sum_{k=1}^n (t_k-\bar{t})^2}}, \cdots, \frac{(t_n-\bar{t})}{\sqrt{\sum_{k=1}^n (t_k-\bar{t})^2}}\right)^{\mathrm{T}} \\
\beta &= \left(\frac{1}{\sqrt{n}}, \cdots, \frac{1}{\sqrt{n}}\right)^{\mathrm{T}}
\end{aligned}
$$

不难验证 $|\alpha|=|\beta|=1$, 且 $\alpha\cdot\beta=0$. 故存在正交矩阵 $A\in\mathbb{R}^{n\times n}$ 以 α^{T} 和 β^{T} 作为其前两行. 由于 $\mathbf{X}=(X_1,\cdots,X_n)\sim\mathcal{N}(\mathbf{0},\sigma^2 I_n)$, 根据上一题的结论, 有 $\mathbf{Y}=(Y_1,\cdots,Y_n)=A\mathbf{X}\sim\mathcal{N}(\mathbf{0},\sigma^2 I_n)$.

注意到

$$
\begin{aligned}
\sum_{i=1}^n Y_i^2 &= |\mathbf{Y}|^2 = |A\mathbf{X}|^2 = |\mathbf{X}|^2 = \sum_{i=1}^n X_i^2 \\
Y_1^2 &= \left(\sum_{j=1}^n \frac{(t_j-\bar{t})X_j}{\sqrt{\sum_{k=1}^n (t_k-\bar{t})^2}}\right)^2 \\
Y_2^2 &= \frac{1}{n}\left(\sum_{j=1}^n X_j\right)^2
\end{aligned}
$$

以及 $\sum\limits_{i=1}^n (X_i-\overline{X})^2 = \sum\limits_{i=1}^n X_i^2 - \frac{1}{n}\left(\sum\limits_{j=1}^n X_j\right)^2$, 故 $Q=\sum\limits_{i=1}^n Y_i^2 - Y_2^2 - Y_1^2 = \sum\limits_{i=3}^n Y_i^2$. 由于 $Y_i\sim$ i.i.d. $\mathcal{N}(0,\sigma^2), Y_i/\sigma\sim$ i.i.d. $\mathcal{N}(0,1)$, 所以 $Q/\sigma^2=\sum\limits_{i=3}^n (Y_i/\sigma)^2\sim\chi^2(n-2)$.

同理, 对于 $F=\frac{(Y_1/\sigma)^2}{(Q/\sigma^2)/(n-2)}$, 由于 $(Y_1/\sigma)^2\sim\chi^2(1)$ 且与 Q 独立, 所以 $F\sim F(1,n-2)$.

7

1. 注意到 $Z,W\sim$ i.i.d. $\mathcal{N}(0,1)$, 故 $U=Z^2+W^2\sim\chi^2(2)$, 概率密度函数为 $f_U(u)=\frac{1}{2}\mathrm{e}^{-u/2}\mathbb{1}[u>0]$, 从而也有 $U\sim\mathrm{Exp}\left(\frac{1}{2}\right)$.

2. 由于 $X_1,\cdots,X_n\sim$ i.i.d. $\mathrm{Exp}(\lambda)$, 故 n 个指数分布随机变量的和 $S=n\overline{X}$ 的概率密度为

$$
f_S(s) = \frac{s^{n-1}\lambda^n \mathrm{e}^{-\lambda s}}{(n-1)!}\cdot\mathbb{1}[s>0]
$$

从而 $T = 2\lambda S$ 的概率密度为

$$f_T(t) = \frac{1}{2\lambda} f_S\left(\frac{t}{2\lambda}\right) = \frac{1}{2\Gamma(n)}\left(\frac{t}{2}\right)^{n-1} \mathrm{e}^{-t/2} \cdot \mathbb{1}[t > 0]$$

故 $T \sim \chi^2(2n)$.

3. 注意到 $\frac{Y}{\sqrt{\lambda \overline{X}}} = \frac{Y}{\sqrt{T/(2n)}}$, 其中 $Y \sim \mathcal{N}(0,1), T \sim \chi^2(2n)$ 且 Y 与 $(X_1, \cdots, X_n)$ 独立说明 Y 与 T 独立, 故根据定义 $\frac{Y}{\sqrt{\lambda \overline{X}}} \sim t(2n)$.

8

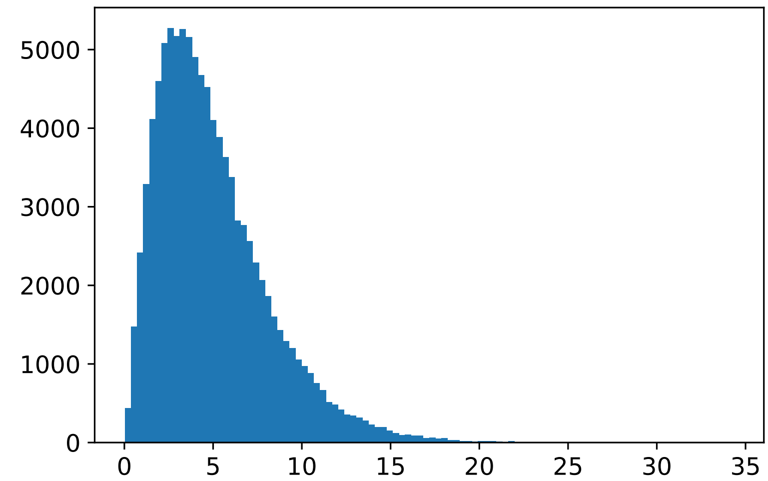

图 1: $\chi^2(5)$ 分布直方图

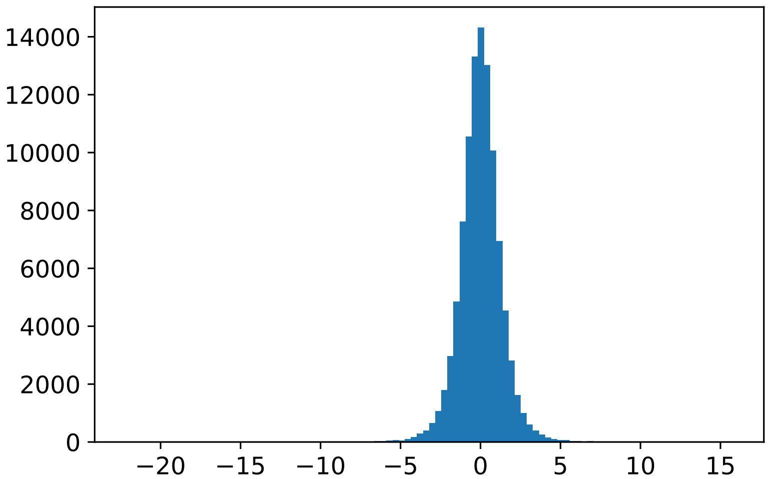

图 2: $t(5)$ 分布直方图

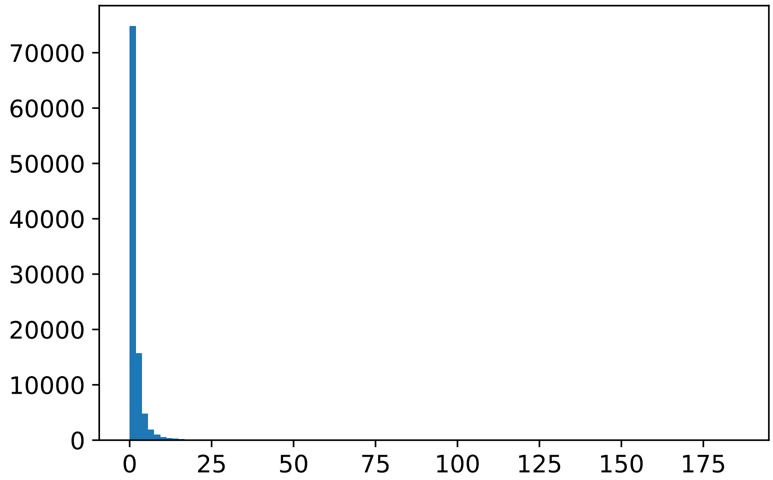

图 3: $F(3,5)$ 分布直方图

9

1. 注意到 $F_n(x;\omega)$ 与 $F(x)$ 都是单调增函数, 故

$$F_n(x;\omega)-F(x)\leqslant F_n(x_{M,k+1}-0;\omega)-F(x_{M,k})\leqslant F_n(x_{M,k+1}-0)-F(x_{M,k+1}-0)+\frac{1}{M}$$

2. 由于 $|F_n(x;\omega)-F(x)|\leqslant \max\limits_{1\leqslant k\leqslant M}\max\left\{|F_n(x_{M,k}-0;\omega)-F(x_{M,k}-0)|,|F_n(x_{M,k};\omega)-F(x_{M,k})|\right\}+\frac{1}{M}$ 对任意 $x\in\mathbb{R}$ 成立, 故

$$\sup_{x\in\mathbb{R}}|F_n(x;\omega)-F(x)|\leqslant \max_{1\leqslant k\leqslant M}\max\left\{|F_n(x_{M,k}-0;\omega)-F(x_{M,k}-0)|,|F_n(x_{M,k};\omega)-F(x_{M,k})|\right\}+\frac{1}{M}$$

$$\begin{aligned}&\mathbb{P}\left(\sup_{x\in\mathbb{R}}|F_n(x;\omega)-F(x)|\geqslant\frac{2}{M}\right)\\ \leqslant&\sum_{k=1}^{M}\mathbb{P}\left(|F_n(x_{M,k}-0;\omega)-F(x_{M,k}-0)|\geqslant\frac{2}{M}\right)+\mathbb{P}\left(|F_n(x_{M,k};\omega)-F(x_{M,k})|\geqslant\frac{2}{M}\right)\end{aligned}$$

由于 $|F_n(x;\omega)-F(x)|\xrightarrow{P}0\ (n\to\infty)$, 对于任意 $\varepsilon>0$ 都有 $\lim\limits_{n\to\infty}\mathbb{P}\left(|F_n(x_{M,k};\omega)-F(x_{M,k})|\geqslant\varepsilon\right)=0$, 进一步也可以证明 $\lim\limits_{n\to\infty}\mathbb{P}\left(|F_n(x_{M,k};\omega)-F(x_{M,k}-0)|\geqslant\varepsilon\right)=0$, 从而

$$\lim_{n\to\infty}\mathbb{P}\left(\sup_{x\in\mathbb{R}}|F_n(x;\omega)-F(x)|\geqslant\frac{2}{M}\right)=0$$

3. 由于 M 的任意性, 对于任意 $\varepsilon>0$,

$$\lim_{n\to\infty}\mathbb{P}\left(D_n(\omega)\geqslant\varepsilon\right)=\lim_{n\to\infty}\mathbb{P}\left(\sup_{x\in\mathbb{R}}|F_n(x;\omega)-F(x)|\geqslant\varepsilon\right)=0$$

即说明 $D_n(\omega)\xrightarrow{P}0\ (n\to\infty)$.