2023 Spring《数理统计》平时作业 1

@ 做作的 Morpheus

2023 年 5 月 6 日

题目 1. 设 $x_1, x_2, \cdots, x_n, x_{n+1}$ 是来自 $\mathcal{N}(\mu, \sigma^2)$ 的样本，又设 $\overline{x}_n = \frac{1}{n}\sum_{i=1}^n x_i$, $s_n^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x}_n)^2$，试求常数 c，使得 $t_c = c(x_{n+1} - \overline{x}_n)/s_n$ 服从 t 分布，并指出其自由度.

解答.

1) $x_{n+1} \sim \mathcal{N}(\mu, \sigma^2), \overline{x}_n \sim \mathcal{N}(\mu, \sigma^2/n) \Rightarrow x_{n+1} - \overline{x}_n \sim \mathcal{N}\left(0, \frac{n+1}{n}\sigma^2\right) \Rightarrow$

$$y_n = \frac{x_{n+1} - \overline{x}_n}{\sqrt{\frac{n+1}{n}}\sigma} \sim \mathcal{N}(0, 1)$$

2) 由样本方差的抽样分布定理（定理 1.3.3）：

   ◇ s_n^2 与 $\overline{x}_n$ 相互独立，又与 x_{n+1} 相互独立，故 s_n^2 与 y_n 相互独立；

   ◇ $\frac{(n-1)s_n^2}{\sigma^2} \sim \chi^2(n-1)$

   由 t 分布的定义（定理 1.3.4）可知

$$t_n = \frac{y_n}{s_n/\sigma} = \sqrt{\frac{n}{n+1}} \cdot \frac{x_{n+1} - \overline{x}_n}{s_n} \sim t(n-1),$$

   所以 $c = \sqrt{\frac{n}{n+1}}$，且自由度为 $n - 1$. □

题目 2. 设从两个方差相等，且相互独立的正态总体分别抽取容量为 15 与 20 的样本，若其样本方差分别为 s_1^2 与 s_2^2，试求 $P(s_1^2/s_2^2 > 2)$.

解答. 由 F 分布的定义（定理 1.3.6）可知

$$F = \frac{s_1^2}{s_2^2} \sim F(14, 19).$$

由 Mathematica 计算得：$P(F > 2) = 0.0798168$. □

题目 3. 设 $x_1, x_2, \cdots, x_{17}$ 是来自正态总体 $\mathcal{N}(\mu, \sigma^2)$ 的样本，$\overline{x}$ 与 s^2 分别为其样本均值与样本方差，求 k，使得 $P(\overline{x} > \mu + ks) = 0.95$.

解答. 由样本均值与样本标准差之比的抽样分布定理（定理 1.3.5）可知

$$\begin{aligned} t = \frac{\sqrt{17}(\overline{x} - \mu)}{s} \sim t(16) \Rightarrow P(\overline{x} > \mu + ks) &= P(t > \sqrt{17}k) = 0.95 \\ &\Rightarrow P(t \le \sqrt{17}k) = 0.05. \end{aligned}$$

查表（附表 4）得 $t_{0.05}(16) = -1.746 = \sqrt{17}k$，从而 $k = -0.4235$. □

题目 4. 设总体 X 的概率密度函数为：

$$p(x) = 3x^2, \quad 0 \le x \le 1$$

从中获得一个容量为 5 的样本 $x_1, x_2, \cdots, x_5$，试分别求 $x_{(1)}, x_{(5)}$ 的概率密度函数.

解答. 总体 X 的分布函数为 $F(x) = \int_0^x p(t)\mathrm{d}t = x^3$，根据次序统计量的密度函数计算公式（定理 1.4.1）：

1）$x_{(1)}$ 的概率密度函数为

$$p_1(x) = \frac{5!}{(1-1)!(5-1)!}(x^3)^{1-1}(1-x^3)^{5-1} \cdot 3x^2 = 15x^2(1-x^3)^4, \quad 0 \le x \le 1$$

1) $x_{(5)}$ 的概率密度函数为

$$p_5(x) = \frac{5!}{(5-1)!(5-5)!}(x^3)^{5-1}(1-x^3)^{5-5} \cdot 3x^2 = 15x^{14}, \quad 0 \le x \le 1$$

□

题目 5. 设某电子元件的寿命服从参数 $\lambda = 0.0015$ 的指数分布，其分布函数为:

$$F(x) = 1 - e^{-\lambda x}, \quad x > 0$$

今从中随机抽取 6 个元件，测得其寿命 $x_1, x_2, \cdots, x_6$，试求下列事件的概率:

1) 到 800 小时没有一个元件失效;

2) 到 3000 小时所有元件都失效.

解答. 代入 $\lambda = 0.0015$ 得分布函数 $F(x) = 1 - e^{-0.0015x}$，所求概率分别为

1)
$$\begin{aligned} P(x_{(1)} > 800) &= \prod_{i=1}^{6} P(x_i > 800) = \prod_{i=1}^{6} [1 - F(800)] \\ &= (e^{-1.2})^6 = e^{-7.2} \approx 7.4659 \times 10^{-4} \end{aligned}$$

2)
$$\begin{aligned} P(x_{(6)} < 3000) &= \prod_{i=1}^{6} P(x_i < 3000) = \prod_{i=1}^{6} F(3000) \\ &= (1 - e^{-4.5})^6 \approx 0.9352 \end{aligned}$$

□

题目 6. 设 $x_1, x_2, \cdots, x_n$ 是来自泊松分布 $P(\lambda)$ 的一个样本，证明：

1) $T = \sum_{i=1}^n x_i$ 是 λ 的充分统计量；

2) 依据条件分布 $P(\boldsymbol{X} = \boldsymbol{x} | T = t)$ 设计一个随机试验，使其产生的样本与原样本同分布；

3) 在 $n = 2$ 时，$x_1 + 2x_2$ 是统计量，但不是 λ 的充分统计量.

解答.

1) 由泊松分布的可加性知 $T \sim P(n\lambda)$，即

$$P(T = t) = \frac{(n\lambda)^t}{t!} e^{-n\lambda}, \quad t = 0, 1, \cdots$$

当 $T = t$ 时，样本的条件分布为

$$\begin{aligned}
& P_\lambda(X_1 = x_1, \cdots, X_n = x_n | T = t) \\
= & \frac{P(X_1 = x_1) \cdots P(X_{n-1} = x_{n-1}) P\left(X_n = t - \sum_{i=1}^{n-1} x_i\right)}{P(T = t)} \\
= & \left(\prod_{i=1}^{n-1} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}\right) \frac{\lambda^{t - \sum_{i=1}^{n-1} x_i} e^{-\lambda}}{\left(t - \sum_{i=1}^{n-1} x_i\right)!} \cdot \frac{t!}{(n\lambda)^t e^{-n\lambda}} \\
= & \frac{t!}{x_1! \cdots x_{n-1}! \left(t - \sum_{i=1}^{n-1} x_i\right)!} \cdot \frac{1}{n^t}
\end{aligned}$$

由于该条件分布在给定 $T = t$ 的情况下即完全确定，与参数 λ 无关，因此 T 是 λ 的充分统计量.

2) 根据上述条件分布可以设计如下的随机投球试验：把 t 个相同的球随机放入 n 个不同盒子，记 x_i' 为第 i 个盒子中球的个数 $(i = 1, \cdots, n)$，则所得样本 $x_1', \cdots, x_n'$ 的分布

$$\begin{aligned}
P(X_1' = x_1', \cdots, X_n' = x_n') &= P(X_1' = x_1', \cdots, X_n' = x_n' | T = t) P(T = t) \\
&= \frac{t!}{x_1'! \cdots x_n'!} \cdot \frac{1}{n^t} \cdot P(T = t)
\end{aligned}$$

与原样本 $x_1, \cdots, x_n$ 的分布

$$\begin{aligned} P(X_1 = x_1, \cdots, X_n = x_n) &= P(X_1 = x_1, \cdots, X_n = x_n | T = t) P(T = t) \\ &= \frac{t!}{x_1! \cdots x_n!} \cdot \frac{1}{n^t} \cdot P(T = t) \end{aligned}$$

具有相同的表达式，即该随机试验所产生的样本与原样本同分布.

3) 不妨设 $T' = x_1 + 2x_2 = 2$，则 $(x_1, x_2) = (2, 0)$ 或 $(0, 1)$，此时样本 $(0, 1)$ 的条件分布

$$\begin{aligned} P_\lambda(X_1 = 0, X_2 = 1 | T' = 2) &= \frac{P(X_1 = 0, X_2 = 1)}{P(T' = 2)} \\ &= \frac{P(X_1 = 0, X_2 = 1)}{P(X_1 = 2, X_2 = 0) + P(X_1 = 0, X_2 = 1)} \\ &= \frac{\lambda e^{-2\lambda}}{\frac{1}{2}\lambda^2 e^{-2\lambda} + \lambda e^{-2\lambda}} = \frac{2}{\lambda + 2} \end{aligned}$$

依赖于参数 λ，所以 $T' = x_1 + 2x_2$ 不是 λ 的充分统计量. □

题目 7. 给定 r，寻求如下负二项分布参数 p 的充分统计量:

$$P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \quad x = r, r+1, \cdots$$

解答. 设 $\boldsymbol{x} = (x_1, \cdots, x_n)$ 是来自负二项分布的一个样本，则该样本的联合分布为

$$\begin{aligned} P_p(\boldsymbol{X} = \boldsymbol{x}) &= \prod_{i=1}^n P_p(X_i = x_i) \\ &= \prod_{i=1}^n \binom{x_i - 1}{r-1} p^r (1-p)^{x_i - r} \\ &= (1-p)^{\sum_{i=1}^n x_i} \left(\frac{p}{1-p}\right)^r \prod_{i=1}^n \binom{x_i - 1}{r-1}. \end{aligned}$$

由因子分解定理（定理 1.5.2）知，$T(\boldsymbol{x}) = \sum_{i=1}^n x_i$ 是参数 p 的充分统计量. □

题目 8. 设 $x_1, x_2, \cdots, x_n$ 是来自如下分布的一个样本，寻求各自的充分统计量：

1) $p_\theta(x) = \theta x^{\theta-1}$, $0 < x < 1$, $\theta > 0$;

2) $p_\theta(x) = \theta a^\theta x^{-(\theta+1)}$, $x > a$, $\theta > 0$, a 已知；

3) $p_\theta(x) = \frac{1}{2\theta} e^{-|x|/\theta}$, $-\infty < x < +\infty$, $\theta > 0$.

解答. 设 $\boldsymbol{x} = (x_1, \cdots, x_n)$ 是来自上述分布的一个样本，则该样本的联合概率密度函数分别为

1) $p_\theta(\boldsymbol{x}) = \prod_{i=1}^n p_\theta(x_i) = \prod_{i=1}^n \theta x_i^{\theta-1} = \theta \left(\prod_{i=1}^n x_i\right)^{\theta-1}$，由因子分解定理知 $T(\boldsymbol{x}) = \prod_{i=1}^n x_i$ 是参数 θ 的充分统计量；

2) $p_\theta(\boldsymbol{x}) = \prod_{i=1}^n p_\theta(x_i) = \prod_{i=1}^n \theta a^\theta x_i^{-(\theta+1)} = \theta a^\theta \left(\prod_{i=1}^n x_i\right)^{-(\theta+1)}$，由因子分解定理知 $T(\boldsymbol{x}) = \prod_{i=1}^n x_i$ 是参数 θ 的充分统计量；

3) $p_\theta(\boldsymbol{x}) = \prod_{i=1}^n p_\theta(x_i) = \prod_{i=1}^n \frac{1}{2\theta} e^{-|x_i|/\theta} = \frac{1}{2\theta} e^{-\sum_{i=1}^n |x_i|/\theta}$，由因子分解定理知 $T(\boldsymbol{x}) = \sum_{i=1}^n |x_i|$ 是参数 θ 的充分统计量. □

题目 9. 寻求如下三种不同均匀分布的充分统计量：

1) $U(0, \theta)$;

2) $U(\theta_1, \theta_2)$, $\theta_1 < \theta_2$;

3) $U(\theta, 2\theta)$.

解答. 设 $\boldsymbol{x} = (x_1, \cdots, x_n)$ 是来自上述均匀分布的一个样本，则该样本的联合概率密度函数分别为

1)

$$\begin{aligned} p_\theta(\boldsymbol{x}) &= \prod_{i=1}^n p_\theta(x_i) = \prod_{i=1}^n \frac{1}{\theta} I_{(0,+\infty)}(x_i) I_{(-\infty,\theta)}(x_i) \\ &= \frac{1}{\theta} I_{(0,+\infty)}(x_{(1)}) I_{(-\infty,\theta)}(x_{(n)}), \end{aligned}$$

由因子分解定理知 $T(\boldsymbol{x}) = x_{(n)}$ 是参数 θ 的充分统计量；

2)

$$\begin{aligned} p_{\theta_1,\theta_2}(\boldsymbol{x}) &= \prod_{i=1}^n p_{\theta_1,\theta_2}(x_i) = \prod_{i=1}^n \frac{1}{\theta_2-\theta_1} I_{(\theta_1,+\infty)}(x_i) I_{(-\infty,\theta_2)}(x_i) \\ &= \frac{1}{\theta_2-\theta_1} I_{(\theta_1,+\infty)}(x_{(1)}) I_{(-\infty,\theta_2)}(x_{(n)}), \end{aligned}$$

由因子分解定理知 $T(\boldsymbol{x}) = (x_{(1)}, x_{(n)})$ 是参数 (θ_1, θ_2) 的充分统计量；

3)

$$\begin{aligned} p_\theta(\boldsymbol{x}) &= \prod_{i=1}^n p_\theta(x_i) = \prod_{i=1}^n \frac{1}{\theta} I_{(\theta,+\infty)}(x_i) I_{(-\infty,2\theta)}(x_i) \\ &= \frac{1}{\theta} I_{(\theta,+\infty)}(x_{(1)}) I_{(-\infty,2\theta)}(x_{(n)}), \end{aligned}$$

由因子分解定理知 $T(\boldsymbol{x}) = (x_{(1)}, x_{(n)})$ 是参数 θ 的充分统计量.

其中 $I_A(\boldsymbol{x})$ 表示集合 A 的示性函数. □

题目 10. 设 $x_1, x_2, \cdots, x_n$ 是来自双参数指数分布

$$p_{\mu,\theta}(x) = \frac{1}{\theta} \exp\left\{-\frac{x-\mu}{\sigma}\right\}, \quad x > \mu,\ \theta > 0$$

的一个样本，证明 $(\overline{x}, x_{(1)})$ 是该分布的充分统计量.

解答. 记 $\boldsymbol{x} = (x_1, \cdots, x_n)$ 为上述样本，则该样本的联合概率密度函数为

$$\begin{aligned}
p_{\mu,\theta}(\boldsymbol{x}) &= \prod_{i=1}^{n} p_{\mu,\theta}(x_i) \\
&= \prod_{i=1}^{n} \frac{1}{\theta} \exp\left\{-\frac{x_i - \mu}{\sigma}\right\} I_{(\mu,+\infty)}(x_i) \\
&= \frac{1}{\theta} \exp\left\{-\frac{1}{\sigma}\left(\sum_{i=1}^{n} x_i - n\mu\right)\right\} I_{(\mu,+\infty)}(x_{(1)}) \\
&= \frac{1}{\theta} \exp\left\{-\frac{n}{\sigma}(\overline{x} - \mu)\right\} I_{(\mu,+\infty)}(x_{(1)}),
\end{aligned}$$

其中 $I_A(\boldsymbol{x})$ 表示集合 A 的示性函数. 因此由因子分解定理可知，$T(\boldsymbol{x}) = (\overline{x}, x_{(1)})$ 是该分布的充分统计量. □

题目 11. 设 $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ 是来自二维正态分布 $\mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ 的一个二维样本，寻求该二维正态分布的充分统计量.

解答. 记 $\boldsymbol{x} = (x_1, \cdots, x_n), \boldsymbol{y} = (y_1, \cdots, y_n)$，则该二维样本的联合概率密度函数为

$$\begin{aligned}
p(\boldsymbol{x}, \boldsymbol{y}) &= \prod_{i=1}^{n} p(x_i, y_i) = \prod_{i=1}^{n} \mathcal{N}(x_i, y_i; \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho) \\
&= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)} \sum_{i=1}^{n} \left[\frac{(x_i - \mu_1)^2}{\sigma_1^2}\right.\right. \\
&\qquad \left.\left. - \frac{2\rho(x_i - \mu_1)(y_i - \mu_2)}{\sigma_1\sigma_2} + \frac{(y_i - \mu_2)^2}{\sigma_2^2}\right]\right\},
\end{aligned}$$

其中

- $(x_i - \mu_1)^2 = (x_i - \overline{x})^2 + (\overline{x} - \mu_1)^2 + 2(x_i - \overline{x})(\overline{x} - \mu_1) \Rightarrow$
$$\sum_{i=1}^{n}(x_i - \mu_1)^2 = \sum_{i=1}^{n}(x_i - \overline{x})^2 + n(\overline{x} - \mu_1)^2$$
- $(x_i - \mu_1)(y_i - \mu_2) = (x_i - \overline{x})(y_i - \overline{y}) + x_i(\overline{y} - \mu_2) + y_i(\overline{x} - \mu_1) \Rightarrow$
$$\sum_{i=1}^{n}(x_i - \mu_1)(y_i - \mu_2) = \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) + n\overline{x}(\overline{y} - \mu_2) + n\overline{y}(\overline{x} - \mu_1)$$

- $(y_i-\mu_2)^2=(y_i-\overline{y})^2+(\overline{y}-\mu_2)^2+2(y_i-\overline{y})(\overline{y}-\mu_2)\Rightarrow$

$$\sum_{i=1}^{n}(y_i-\mu_2)^2=\sum_{i=1}^{n}(y_i-\overline{y})^2+n(\overline{y}-\mu_2)^2$$

由因子分解定理可知,$T(\boldsymbol{x},\boldsymbol{y})=\big(\overline{x},\overline{y},\sum_{i=1}^{n}(x_i-\overline{x})^2,\sum_{i=1}^{n}(y_i-\overline{y})^2,\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})\big)$ 是该二维正态分布的充分统计量. $\square$