5.5 充分统计量

1. 设 $x_1, x_2, \cdots, x_n$ 是来自几何分布

$$P(X=x)=\theta(1-\theta)^x, \quad x=0,1,2,\cdots$$

的样本，证明 $T=\sum_{i=1}^{n} x_i$ 是充分统计量。

证明：

$$p(x_1,x_2,\cdots,x_n;\theta)=\prod_{i=1}^{n}\theta(1-\theta)^{x_i}=\theta^n(1-\theta)^{\sum_{i=1}^{n}x_i}=\theta^n(1-\theta)^T$$

取 $g(T,\theta)=\theta^n(1-\theta)^T, h(X)=1$，由因子分解定理可知 $T=\sum_{i=1}^{n} x_i$ 是 θ 的充分统计量。 □

3. 设总体为如下离散分布：

x	a_1	a_2	$\cdots$	a_k
p	p_1	p_2	$\cdots$	p_k

。$x_1, x_2, \cdots, x_n$ 是来自该总体的样本，

(1) 证明次序统计量 $(x_{(1)}, x_{(2)}, \cdots, x_{(n)})$ 是充分统计量；

证明：

设 $T=(x_{(1)}, x_{(2)}, \cdots, x_{(n)})$，$X$ 表示一次取样。则

$$\begin{aligned}P(X=(x_1,x_2,\cdots,x_n)|T=t)&=\frac{P(X=(x_1,x_2,\cdots,x_n),T=t)}{P(T=t)}\\&=\frac{\prod_{i=1}^{n}p_i}{\mathrm{P}_n^n\prod_{i=1}^{n}p_i}=\frac{1}{\mathrm{P}_n^n}=\frac{1}{n!}\end{aligned}$$

可见与 T 无关，所以次序统计量 $(x_{(1)}, x_{(2)}, \cdots, x_{(n)})$ 是充分统计量。 □

(2) 以 n_j 表示 $x_1, x_2, \cdots, x_n$ 中等于 a_j 的个数，证明 $(n_1, n_2, \cdots, n_k)$ 是充分统计量。

设 $T=(n_1, n_2, \cdots, n_k)$，$X$ 表示一次取样。则

$$\begin{aligned}P(X=(x_1,x_2,\cdots,x_n)|T=t)&=\frac{P(X=(x_1,x_2,\cdots,x_n),T=t)}{P(T=t)}\\&=\frac{\prod_{j=1}^{n}p_j}{\mathrm{P}_n^n\prod_{j=1}^{n}p_j^{n_j}}\end{aligned}$$

应该与 T 无关，所以 $(n_1, n_2, \cdots, n_k)$ 是充分统计量。

8. 设 $x_1, x_2, \cdots, x_n$ 是来自拉普拉斯（Laplace）分布

$$p(x;\theta) = \frac{1}{2\theta} e^{-\frac{|x|}{\theta}}, \theta > 0$$

的样本，试给出一个充分统计量。

解：

设 X 表示一次取样，则

$$P(X = (x_1, x_2, \cdots, x_n); \theta) = \prod_{i=1}^{n} p(x_i;\theta) = \prod_{i=1}^{n} \frac{1}{2\theta} e^{-\frac{|x|}{\theta}} = \left(\frac{1}{2\theta}\right)^n e^{-\frac{1}{\theta}\sum_{i=1}^{n}|x_i|}$$

令 $T = \sum_{i=1}^{n} |x_i|$，则上式 $= \left(\frac{1}{2\theta}\right)^n \left(e^{-\frac{T}{\theta}}\right)$。则可以令 $g(T,\theta) = \left(\frac{1}{2\theta}\right)^n \left(e^{-\frac{T}{\theta}}\right)$, $h(X) = 1$，由因子分解定理可知 $T = \sum_{i=1}^{n} |x_i|$ 是 θ 的充分统计量。 □

10. 设 $x_1, x_2, \cdots, x_n$ 是来自正态分布 $N(\mu, \sigma^2)$ 的样本。

(1) 在 μ 已知时给出 σ^2 的一个充分统计量。

解：

$$p(x_1, x_2, \cdots, x_n; \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right\}$$

所以可以令 $T = \sum_{i=1}^{n}(x_i - \mu)^2$，则 T 是 σ^2 的一个充分统计量。 □

(2) 在 σ^2 已知时给出 μ 的一个充分统计量。

解：

$$\begin{aligned} p(x_1, x_2, \cdots, x_n; \sigma^2) &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right\} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{n\mu^2}{2\sigma^2}\right\} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2\right\} \exp\left\{\frac{\mu}{\sigma^2}\sum_{i=1}^{n}x_i\right\} \end{aligned}$$

令 $T = \sum_{i=1}^{n} x_i$, 则 $g(\mu, T) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{n\mu^2}{2\sigma^2}\right\} \exp\left\{\frac{\mu}{\sigma^2}T\right\}, h(\overrightarrow{x}) = \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2\right\}$。所以 T 是 μ 的一个充分统计量。

□

11. 设 $x_1, x_2, \cdots, x_n$ 是来自均匀分布 $U(\theta_1, \theta_2)$ 的样本，试给出一个充分统计量。

解：

$$p(x_1, x_2, \cdots, x_n; \theta_1, \theta_2) = \prod_{i=1}^{n} \frac{1}{\theta_2 - \theta_1} 1_{[\theta_1, \theta_2]}(x_i) = \left(\frac{1}{\theta_2 - \theta_1} \right)^n 1_{[\theta_1, \theta_2]}(x_{(1)}, x_{(n)})$$

所以 $(x_{(1)}, x_{(n)})$ 是一个充分统计量。 □

12. 设 $x_1, x_2, \cdots, x_n$ 是来自均匀分布 $U(\theta, 2\theta), \theta > 0$ 的样本，试给出充分统计量。

解：

$$p(x_1, x_2, \cdots, x_n; \theta) = \prod_{i=1}^{n} \frac{1}{\theta} 1_{[\theta, 2\theta]}(x_i) = \frac{1}{\theta^n} 1_{[\theta, 2\theta]}(x_{(1)}, x_{(n)})$$

所以 $(x_{(1)}, x_{(n)})$ 是一个充分统计量。 □

17. 设 $\begin{pmatrix} x_i \\ y_i \end{pmatrix}, i = 1, 2, \cdots, n$ 是来自正态分布族

$$\left\{ N\left(\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right) ; -\infty < \theta_1, \theta_2 < \infty, \sigma_1, \sigma_2 > 0, |\rho| \leqslant 1 \right\}$$

的一个二维样本，寻求 $(\theta_1, \sigma_1, \theta_2, \sigma_2, \rho)$ 的充分统计量。

解：

$$p\left(\begin{pmatrix} x_i \\ y_i \end{pmatrix}; (\theta_1, \sigma_1, \theta_2, \sigma_2, \rho) \right) = \prod_{i=1}^{n} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)} (a_i^2 + b_i^2 - 2\rho a_i b_i) \right\}$$

$$= \left(\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \right)^n \exp\left\{ -\frac{1}{2(1-\rho^2)} \left(\sum_{i=1}^{n} a_i^2 + \sum_{i=1}^{n} b_i^2 - 2\rho \sum_{i=1}^{n} a_i b_i \right) \right\}$$

其中

$$\sum_{i=1}^{n} a_i^2 = \sum_{i=1}^{n} \left(\frac{x_i - \theta_1}{\sigma_1} \right)^2 = \frac{1}{\sigma_1^2} \sum_{i=1}^{n} (x_i^2 - 2\theta_1 x_i + \theta_1^2) = \frac{1}{\sigma_1^2} \sum_{i=1}^{n} x_i^2 - \frac{2\theta_1}{\sigma_1^2} \sum_{i=1}^{n} x_i + \frac{\theta_1^2}{\sigma_1^2}$$

$$\sum_{i=1}^{n} b_i^2 = \sum_{i=1}^{n} \left(\frac{y_i - \theta_2}{\sigma_2} \right)^2 = \frac{1}{\sigma_2^2} \sum_{i=1}^{n} (y_i^2 - 2\theta_2 y_i + \theta_2^2) = \frac{1}{\sigma_2^2} \sum_{i=1}^{n} y_i^2 - \frac{2\theta_2}{\sigma_2^2} \sum_{i=1}^{n} y_i + \frac{\theta_2^2}{\sigma_2^2}$$

$$\sum_{i=1}^{n} a_i b_i = \sum_{i=1}^{n} \left(\frac{x_i - \theta_1}{\sigma_1} \right) \left(\frac{y_i - \theta_2}{\sigma_2} \right) = \frac{1}{\sigma_1\sigma_2} \sum_{i=1}^{n} (x_i y_i - \theta_1 y_i - \theta_2 x_i + \theta_1\theta_2)$$

$$= \frac{1}{\sigma_1\sigma_2} \sum_{i=1}^{n} x_i y_i - \frac{\theta_1}{\sigma_1\sigma_2} \sum_{i=1}^{n} y_i - \frac{\theta_2}{\sigma_1\sigma_2} \sum_{i=1}^{n} x_i + \frac{n\theta_1\theta_2}{\sigma_1\sigma_2}$$

仔细观察即可发现

$$\left(\sum_{i=1}^n x_i,\ \sum_{i=1}^n x_i^2,\ \sum_{i=1}^n y_i,\ \sum_{i=1}^n y_i^2,\ \sum_{i=1}^n x_i y_i\right)$$

是此二维正态分布的充分统计量。 □

19. 设 $x_1, x_2, \cdots, x_n$ 是来自两参数指数分布

$$p(x;\theta,\mu) = \frac{1}{\theta} e^{-\frac{x-\mu}{\theta}}, \quad x > \mu, \theta > 0$$

的样本，证明 $(\bar{x}, x_{(1)})$ 是充分统计量。

证明：

$$p(x_1, x_2, \cdots, x_n; \theta, \mu) = \prod_{i=1}^n \frac{1}{\theta} e^{-\frac{x_i-\mu}{\theta}} = \frac{1}{\theta^n} \exp\left\{-\frac{1}{\theta}\sum_{i=1}^n (x_i - \mu)\right\}$$
$$= \frac{1}{\theta^n} \exp\left\{-\frac{1}{\theta}\sum_{i=1}^n x_i\right\} \exp\left\{\frac{n\mu}{\theta}\right\}, \quad x_1, x_2, \cdots, x_n > \mu$$

其中 $x_1, x_2, \cdots, x_n > \mu \iff x_{(1)} > \mu$，并且 $\sum_{i=1}^n x_i = n\bar{x}$，所以 $(\bar{x}, x_{(1)})$ 是充分统计量。 □

20. 设随机变量 $Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2), i = 1, 2, \cdots, n$，诸 Y_i 独立，$x_1, x_2, \cdots, x_n$ 是已知常数，证明 $\left(\sum_{i=1}^n Y_i,\ \sum_{i=1}^n x_i Y_i,\ \sum_{i=1}^n Y_i^2\right)$ 是充分统计量。

解：

$$p(Y_1, Y_2, \cdots, Y_n; \beta_0, \beta_1, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{Y_i - (\beta_0 + \beta_1 x)}{\sigma}\right)^2\right\}$$
$$= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_i)^2\right\}$$

其中

$$\sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_i)^2 = \sum_{i=1}^n Y_i^2 + n\beta_0^2 + n\beta_1^2 \sum_{i=1}^n x_i^2 - 2\beta_0 \sum_{i=1}^n Y_i - 2\beta_1 \sum_{i=1}^n x_i Y_i + \beta_0\beta_1 \sum_{i=1}^n x_i$$

其中 β_0, β_1, σ 为参数，$x_1, x_2, \cdots, x_n$ 已知，所以 $\left(\sum_{i=1}^n Y_i,\ \sum_{i=1}^n x_i Y_i,\ \sum_{i=1}^n Y_i^2\right)$ 是充分统计量。
□