《概率论与数理统计》习题

第十七讲　充分统计量

1. 设 $x_1, x_2, \cdots, x_n$ 是来自几何分布

$$P(X=x)=\theta(1-\theta)^x$$

的样本，证明 $T=\sum_{i=1}^n x_i$ 是充分统计量。

证明:

$$p(T=t)=\begin{pmatrix} n+t-1 \\ t \end{pmatrix}\theta^n(1-\theta)^t, t=0,1,2,\ldots.$$

当给定 $T=t$ 时，$\sum_{i=1}^n x_i=t$.

$$\begin{aligned}
P(x_1,x_2,\ldots,x_n|T=t) &= \frac{P(x_1,x_2,\ldots,x_{n-1},t-\sum_{i=1}^{n-1}x_i)}{P(T=t)} \\
&= \frac{\prod_{i=1}^{n-1}P(x_i)P(t-\sum_{i=1}^{n-1}x_i)}{\begin{pmatrix} n+t-1 \\ t \end{pmatrix}\theta^n(1-\theta)^t} \\
&= \frac{\prod_{i=1}^{n-1}\Big(\theta(1-\theta)^{x_i}\Big)\theta(1-\theta)^{t-\sum_{i=1}^{n-1}x_i}}{\begin{pmatrix} n+t-1 \\ t \end{pmatrix}\theta^n(1-\theta)^t} \\
&= \frac{\theta^n(1-\theta)^t}{\begin{pmatrix} n+t-1 \\ t \end{pmatrix}\theta^n(1-\theta)^t} \\
&= \frac{1}{\begin{pmatrix} n+t-1 \\ t \end{pmatrix}}
\end{aligned}$$

该条件分布与 θ 无关，因而 $T=\sum_{i=1}^n x_i$ 是充分统计量。

(采用定义法或因子分解定理均可。)

2. 设 $x_1, x_2, \cdots, x_n$ 来自均匀分布 $U(\theta_1, \theta_2)$ 的样本，试给出充分统计量。
解:

总体的密度函数为:

$$p(x;\theta_1,\theta_1)=\begin{cases}\frac{1}{\theta_2-\theta_1}, & \theta_1<x<\theta_2\\ 0, & others\end{cases}$$

样本的联合密度函数为:

$$p(x_1,x_2,\ldots,x_n;\theta_1,\theta_1)=\begin{cases}\left(\frac{1}{\theta_2-\theta_1}\right)^n, & \theta_1<x_{(1)}<x_{(n)}<\theta_2\\ 0, & others\end{cases}$$

即可以写成:

$$p(x_1,x_2,\ldots,x_n;\theta_1,\theta_1)=\left(\frac{1}{\theta_2-\theta_1}\right)^n I_{\theta_1<x_{(1)}<x_{(n)}<\theta_2}$$

令 $t_1=x_{(1)},t_2=x_{(n)}$,

$$p(x_1,x_2,\ldots,x_n;\theta_1,\theta_1)=g(t;\theta_1,\theta_2)h(x)$$

$$h(x)=1$$

$$g(t;\theta_1,\theta_2)=\left(\frac{1}{\theta_2-\theta_1}\right)^n I_{\theta_1<x_{(1)}<x_{(n)}<\theta_2}$$

由因子分解定理得，$T=(x_{(1)},x_{(n)})$ 是参数 (θ_1,θ_2) 的充分统计量。

3. 设 $x_1,x_2,\cdots,x_n$ 是来自正态总体 $N(\mu,\sigma_1^2)$ 的样本，$y_1,y_2,\cdots,y_m$ 是来自于另一正态总体 $N(\mu,\sigma_2^2)$ 的样本,这两个样本相互独立,试给出 $(\mu,\sigma_1^2,\sigma_2^2)$ 的充分统计量。

解:

样本 $x_1,x_2,\ldots,x_n$□$y_1,y_2,\ldots,y_n$ 的联合密度函数为:

$$\begin{aligned}&p(x_1,x_2,\ldots,x_n,y_1,y_2,\ldots,y_m)\\&=\prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi\sigma_1^2}}e^{-\frac{(x_i-\mu)^2}{2\sigma_1^2}}\right)\prod_{i=1}^{m}\left(\frac{1}{\sqrt{2\pi\sigma_2^2}}e^{-\frac{(y_i-\mu)^2}{2\sigma_2^2}}\right)\\&=(2\pi)^{-\frac{m+n}{2}}\sigma_1^{-n}\sigma_2^{-m}e^{-\left(\frac{1}{2\sigma_1^2}\sum_{i=1}^n x_i^2+\frac{1}{2\sigma_2^2}\sum_{i=1}^n y_i^2-(\frac{n\bar{x}}{\sigma_1^2}+\frac{m\bar{y}}{\sigma_2^2})\mu+(\frac{n}{2\sigma_1^2}+\frac{m}{2\sigma_2^2})\mu^2\right)}\end{aligned}$$

其中，$\bar{x}=\frac{1}{n}\sum_{i=1}^n x_i$,$\bar{y}=\frac{1}{m}\sum_{i=1}^m y_i$.

令 $t=(t_1,t_2,t_3,t_4)=(\bar{x},\bar{y},\sum_{i=1}^n x_i^2,\sum_{i=1}^m y_i^2)$，取:

$$h(x_1,x_2,\ldots,x_n,y_1,y_2,\ldots,y_m)=1$$

$$g(t;\mu,\sigma_1,\sigma_2)=(2\pi)^{-\frac{m+n}{2}}\sigma_1^{-n}\sigma_2^{-m}e^{-\left(\frac{1}{2\sigma_1^2}t_3+\frac{1}{2\sigma_2^2}t_4-(\frac{nt_1}{\sigma_1^2}+\frac{mt_2}{\sigma_2^2})\mu+(\frac{n}{2\sigma_1^2}+\frac{m}{2\sigma_2^2})\mu^2\right)}$$

由因子分解定理得, $t=(t_1,t_2,t_3,t_4)=(\bar{x},\bar{y},\sum_{i=1}^{n}x_i^2,\sum_{i=1}^{m}y_i^2)$ 是 $(\mu,\sigma_1^2,\sigma_2^2)$ 的充分统计量。