统计量

信息减少 Def

- 给定 随机变量 $X_1, \cdots, X_n$, $X_1 \cdots X_n \to R^{k(n)}$ 为 k维统计量.

  $$t_n(x_1, \cdots, x_n) = (t_1, \cdots, t_n).$$

  为了减少数据.

  $k(n) \le n$ 对信息做总结.

  $k(n) = k$ 独立于 n.

Eg

- $t_n = \frac{1}{n}\{x_1 + \cdots + x_n\}$ $k(n)=1$

  $t_n = [n, (x_1 + \cdots + x_n), x_1^2 + \cdots + x_n^2\}$ $k(n)=3$.

  $t_n = [n, \text{mid}\{x_1, \cdots, x_n\}]$ $k(n)=2$

让信息无损失.

充分统计量

def. $t_1, t_2, \cdots,$ 是**充分统计量**,

$$p(x_1, \cdots, x_n \mid \theta) = h_n(\vec{t}_n, \theta)\, g(x_1, \cdots, x_n) \quad \text{分解.}$$

只与x有关.

for some function $h_n \geq 0$ $\quad g > 0$.

Thm. 如果 $\vec{t}_1, \vec{t}_2, \cdots$ 是充分的, 无需再考虑 $x_1, x_2, \cdots$ 计.

$p(x_1, \cdots, x_n \mid \theta, t_n)$ 是与θ独立的.

Prof: $\vec{t}_n = t_n(x_1, \cdots, x_n)$

$$\begin{aligned} p(x_1, \cdots, x_n \mid \theta) &= P(x_1, \cdots, x_n \mid \theta, t_n)\, P(t_n \mid \theta) \\ &= P(x_1, \cdots, x_n \mid t_n)\, P(t_n \mid \theta) \end{aligned}$$

例子

eg1. • 判断 $\sum x_i$ 是否是指数分布的充分统计量；

$P(X=x_i)=\lambda e^{-\lambda x_i}$.

$f(\vec{x})=\lambda^n e^{-\lambda\sum x_i}$　　$T(\vec{x})=\sum x_i=t$

$f(\vec{x},\lambda)=\lambda^n\cdot e^{-\lambda t}$　此时 $T(\vec{x})=\sum x_i$　$h(\vec{x})=1$.

$g(T(\vec{x}),\lambda)=\lambda^n e^{-\lambda t}$

$p(x_1,\cdots,x_n|\theta)=h_n(\bar{t}_n,\theta)\,g(x_1,\cdots,x_n)$ 分解

eg2. • 找出正态分布的充分统计量.

$f(\vec{x})=(2\pi\delta^2)^{-\frac{n}{2}}\exp\{-\frac{1}{2\delta^2}\sum(x_i-\mu)^2\}$

$=(2\pi\delta^2)^{-\frac{n}{2}}\exp\{-\frac{1}{2\delta^2}[\sum(x_i^2+\mu^2-2x_i\mu)]\}$

$=(2\pi\delta^2)^{-\frac{n}{2}}\exp\{-\frac{1}{2\delta^2}[\sum x_i^2+n\mu-2\mu\sum x_i]\}$

充分统计量设为 $\sum x_i^2$　$\sum x_i$