第1.2节 充分统计量与完备统计量

- 一、充分统计量
- 二、因子分解定理
- 三、完备统计量
- 四、指数型分布族









一、充分统计量

1. 问题的引出

由于样本来自总体,抽取出来的样本包含有总体 的信息。数理统计主要是利用样本信息推断总体的信 息,如何将样本中包含总体的信息提取出来?以及是 否将样本中包含总体的信息完全提取出来? 这些都是 数理统计需要解决的问题。

例如,设总体服从 $N(\mu,\sigma^2)$,在上一节中,用 \overline{X},S^2_n , 去估计总体的 μ 和 σ^2 , \overline{X} , S_n^2 是否将 μ 和 σ^2 的信息完全提 炼出来呢? 这就涉及到充分统计量的概念。





2. 定义

1922年英国统计学家Fisher提出了描述总体信息 是否被完全提炼的概念——充分统计量.

定义 1.4 设 X_1, X_2, \dots, X_n 是来自总体X具有分布函数 $F(x,\theta)$ 的一个样本, $T = T(X_1, X_2, \dots, X_n)$ 为一个(一维或多 维)统计量,当给定T=t时,若样本 $(X_1,X_2,\dots,X_n)^T$ 的 条件分布(离散总体为条件概率,连续总体为条件密度) 与参数 θ 无关,则称T为 θ 的充分统计量.





3. 充分统计量的意义

如果知道了统计量T的观察值以后,样本的条件分布与θ无关,这就表明样本的剩余部分不再包含关于θ的信息,换言之,在T中包含了关于θ的全部信息,因此要做关于θ的统计推断,只需用统计量T就足够了.

例1(p6 例1.3) 设总体X服从两点分布B(1,p),即

$$P{X = x} = p^{x}(1-p)^{1-x}, x = 1,0,$$

 $(X_1, X_2, \dots, X_n)^T$ 是来自总体X的一个样本,试证

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_{i}$$
是参数 p 的充分统计量.







证 利用定义证明其是充分统计量

$$P\{X_{1} = x_{1}, X_{2} = x_{2}, \dots, X_{n} = x_{n} \mid \overline{X} = \frac{k}{n}\}$$

$$= \frac{P\{X_{1} = x_{1}, X_{2} = x_{2}, \dots, X_{n} = x_{n}, \overline{X} = \frac{k}{n}\}}{P\{\overline{X} = \frac{k}{n}\}}$$

$$= \frac{P\{X_{1} = x_{1}, X_{2} = x_{2}, \dots, X_{n} = x_{n}, n\overline{X} = k\}}{P\{n\overline{X} = k\}}$$

$$= \frac{P\{X_{1} = x_{1}, X_{2} = x_{2}, \dots, X_{n} = x_{n}, \sum_{i=1}^{n} X_{i} = k\}}{P\{X_{1} = x_{1}, X_{2} = x_{2}, \dots, X_{n} = x_{n}, \sum_{i=1}^{n} X_{i} = k\}}$$

$$= \frac{P\{X_{1} = x_{1}, X_{2} = x_{2}, \dots, X_{n} = x_{n}, \sum_{i=1}^{n} X_{i} = k\}}{P\{\sum_{i=1}^{n} X_{i} = k\}}$$





$$= \begin{cases} \frac{P\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\}}{P\{\sum_{i=1}^n X_i = k\}}, & \sum_{i=1}^n x_i = k, \\ 0, & \sum_{i=1}^n x_i \neq k. \end{cases}$$

$$= \begin{cases} \frac{P\{X_1 = x_1\}P\{X_2 = x_2\}\cdots P\{X_n = x_n\}}{P\{\sum_{i=1}^n X_i = k\}}, & \sum_{i=1}^n x_i = k, \\ P\{\sum_{i=1}^n X_i = k\} & \sum_{i=1}^n x_i \neq k. \end{cases}$$







$$= \begin{cases} \frac{\sum_{i=1}^{n} x_{i}}{(1-p)^{n-\sum_{i=1}^{n} x_{i}}}, & \sum_{i=1}^{n} x_{i} = k, \\ 0, & \sum_{i=1}^{n} x_{i} \neq k. \end{cases}$$

$$= \begin{cases} \frac{1}{C_{n}^{k}}, & \sum_{i=1}^{n} x_{i} = k, \\ 0, & \sum_{i=1}^{n} x_{i} \neq k. \end{cases}$$

$$= \begin{cases} 0, & \sum_{i=1}^{n} x_{i} \neq k. \end{cases}$$
显然该条件分布与 p 无关,因而 \overline{X} 是 p 的充分统计量.

说明 利用定义判别充分统计量比较麻烦,因而需 要更好的判别准则。

二、因子分解定理

1. 充分统计量的判别准则

定理1.3(因子分解定理)(Fisher-Nerman准则)

(1) 连续型情况

设总体X具有分布密度 $f(x,\theta),(X_1,X_2,\dots,X_n)^T$ 是一个样本, $T(X_1,X_2,\dots,X_n)$ 是一个统计量,则T是 θ 的充分统计量的充要条件是:样本的联合分布密度可以分解为





$$L(\theta) = \prod_{i=1}^{n} f(x_i, \theta) = h(x_1, x_2, \dots, x_n) g(T(x_1, x_2, \dots, x_n), \theta)$$

其中h是 x_1, x_2, \dots, x_n 的非负函数且与 θ 无关,g仅通过T依赖于 x_1, x_2, \dots, x_n .

(2) 离散型情况

设总体X的分布律 $P\{X = x^{(i)}\} = p(x^{(i)}, \theta), (i = 1, 2, \cdots), (X_1, X_2, \cdots, X_n)^T$ 是一个样本, $T(X_1, X_2, \cdots, X_n)$ 是一个统计量,则T是 θ 的充分统计量的充要条件是:样本的联合分布律可以分解为





$$\prod_{i=1}^{n} P(x^{(i)}, \theta) = h(x_1, x_2, \dots, x_n) g(T(x_1, x_2, \dots, x_n), \theta)$$

其中h是 x_1, x_2, \dots, x_n 的非负函数且与 θ 无关,g仅通过T依赖于 x_1, x_2, \dots, x_n .

证明涉及测度论, 从略

说明:

如果参数 θ 为向量时,统计量T也是随机向量,例如 $\theta = (\mu, \sigma^2)$,则相应的统计向量可以为 $T = (\bar{X}, S_n^2)$. 以下将通过几个例子来说明判别法则的应用





例2(p8 例1.4) 根据因子分解定理证明例1.3

解
$$P\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\}$$

$$= p^{\sum_{i=1}^{n} x_i} (1-p)^{n-\sum_{i=1}^{n} x_i} = (1-p)^n (\frac{p}{1-p})^{\sum_{i=1}^{n} x_i}$$

$$= 1\Box (1-p)^n (\frac{p}{1-p})^{n\frac{\sum_{i=1}^{n} x_i}{n}} = 1\Box (1-p)^n (\frac{p}{1-p})^{n\overline{X}}$$
其中 $T(x_1, x_2, \dots, x_n) = \overline{X}, h(x_1, x_2, \dots, x_n) = 1, g(T(x_1, x_2, \dots, x_n), \theta) = (1-p)^n (\frac{p}{1-p})^{nT},$ 因而, \overline{X} 是充分统计量





例3(p8 例1.5设 $(X_1, X_2, \dots, X_n)^T$ 是来自泊松分

$$\pi P(\lambda)$$
的一个样本,试证 $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ 是参数 λ 的充

分统计量.

其中
$$T(x_1, x_2, \dots, x_n) = \overline{X}, h(x_1, x_2, \dots, x_n) = \frac{1}{\prod_{i=1}^{n} x_i!}, g(T(x_1, x_2, \dots, x_n))$$

 $(x_1, \dots, x_n), \theta) = \lambda^{nT} e^{-n\lambda}, 因而, \overline{X}$ 是充分统计量







例4(p9 例1.6)设 $(X_1, X_2, \dots, X_n)^T$ 是来自正态总体

 $N(\mu, 1)$ 的一个样本,试证 $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ 是参数 μ 的充

分统计量.

解 $L(\mu) = \frac{1}{(\sqrt{2\pi})^n} e^{-\left\{\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}}$ $= \frac{1}{(\sqrt{2\pi})^n} \exp\{-\frac{1}{2} \sum_{i=1}^n (x_i - \overline{x} + \overline{x} - \mu)^2\}$ $= \frac{1}{(\sqrt{2\pi})^n} \exp\{-\frac{1}{2} \sum_{i=1}^n (x_i - \overline{x})^2 - \frac{n}{2} (\mu - \overline{x})^2\}$ $= \frac{1}{(\sqrt{2\pi})^n} \exp\{-\frac{1}{2} \sum_{i=1}^n (x_i - \overline{x})^2\} \exp\{-\frac{n}{2} (\mu - \overline{x})^2\}$







其中
$$T(x_1, x_2, \dots, x_n) = \overline{x}, h(x_1, x_2, \dots, x_n) = \exp\{-\frac{1}{2}\}$$

$$\sum_{i=1}^{n} (x_i - \overline{x})^2 \}, g(T(x_1, x_2, \dots, x_n), \theta) = \frac{1}{(\sqrt{2\pi})^n} \exp\{$$

 $-\frac{n}{2}(\mu-T)^2$ },因而, \bar{X} 是充分统计量

例5(p9 例1.7) 设 $(X_1, X_2, \dots, X_n)^T$ 是来自正态总体 N (μ, σ^2) 的一个样本,试证 $T(X_1, X_2, \dots, X_n) = (\overline{X}, T(X_1, X_2, \dots, X_n))$

 $\sum_{i=1}^{n} X_i^2)^T$ 是参数 $\theta = (\mu, \sigma^2)^T$ 的联合充分统计量.





解
$$L(\mu, \sigma^2) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\}$$

$$= \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\{-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 + \frac{n\mu}{\sigma^2} \bar{x} - \frac{n\mu^2}{2\sigma^2}\}\$$

其中
$$T(x_1, x_2, \dots, x_n) = (\bar{x}, \sum_{i=1}^n x_i^2)^T, h(x_1, x_2, \dots, x_n) = 1,$$

$$g(T(x_1, x_2, \dots, x_n), \theta) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\{-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 + \frac{n\mu}{\sigma^2} \overline{x}$$

$$-\frac{n\mu^2}{2\sigma^2}$$
},因而, $T(x_1, x_2, \dots, x_n) = (\bar{x}, \sum_{i=1}^n x_i^2)^T$ 是充分统计向量。





2. 充分统计量的函数特性

定理1.4 设 $T = T(X_1, X_2, \dots, X_n)$ 是 θ 的一个充分统计量,f(t)是一个单值可逆函数,则f(T)也是 θ 的一个充分统计量.

证 以连续型为例,由因子分解定理可知 $L(\theta) = \prod_{i=1}^{i=1} f(x_i, \theta) = h(x_1, x_2, \dots, x_n) g(T(x_1, x_2, \dots, x_n), \theta)$ $= h(x_1, x_2, \dots, x_n) g(f^{-1}(f(T(x_1, x_2, \dots, x_n))), \theta)$ $= h(x_1, x_2, \dots, x_n) q(f(x_1, x_2, \dots, x_n), \theta)$

由因子分解定理可知, $f(x_1, x_2, \dots, x_n)$ 是 θ 的充分统计量,因而充分统计量不唯一.





三、完备统计量

统计量的充分性与完备性在寻找参数的优良估计中将起到重要的作用.

定义1.5 设总体的分布函数族为 $F(x,\theta)(\theta \in \Theta)$

,若对于任意一个满足

$$E_{\theta}[g(X)] = 0$$
,对一切 $\theta \in \Theta$,

的随机变量g(X),总有

$$P_{\theta}\{g(X)=0\}=1,$$
对一切 $\theta\in\Theta$,

则称 $\{F(x,\theta),\theta\in\Theta\}$ 为完备的分布函数族.







定义 1.6 设 X_1, X_2, \dots, X_n 是来自总体X具有分布函数 $F(x,\theta)$ 的一个样本, $T = T(X_1, X_2, \dots, X_n)$ 的分布函数族 $\{F_T(x,\theta), \theta \in \Theta\}$ 是完备的分布函数族,则称 $T = T(X_1, X_2, \dots, X_n)$ 为完备统计量.

说明 完备性的含义不是很显然. 但它具有下列性质

一方面,
$$P_{\theta}\{g_1(T) = g_2(T)\} = 1, \forall \theta \in \Theta,$$

$$\Leftrightarrow E_{\theta}(g_1(T)) = E_{\theta}(g_2(T)), \forall \theta \in \Theta$$







例6(p11 例1.8) 设总体X服从两点分布B(1,p),即 $P\{X=x\}=p^x(1-p)^{1-x},x=1,0,$

 $(X_1, X_2, \dots, X_n)^T$ 是来自总体X的一个样本,试证

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$
是参数p的完备统计量.

证由于
$$P\{\overline{X} = \frac{k}{n}\} = P\{n\overline{X} = k\} = C_n^k p^k (1-p)^{n-k}$$
,因而如果 $E_p(g(\overline{X})) = \sum_{k=0}^n g(\frac{k}{n}) C_n^k p^k (1-p)^{n-k} = 0$,则
$$(1-p)^n \sum_{k=0}^n g(\frac{k}{n}) C_n^k (\frac{p}{1-p})^k = 0,$$







即对任意的
$$0 , $\sum_{k=0}^{n} g(\frac{k}{n}) C_n^k (\frac{p}{1-p})^k = 0$, 而此式$$

是关于 $\frac{p}{1-p}$ 的多项式,因而每项系数只能为0,则 $g(\frac{k}{n})=0$,因而满足 $P_p\{g(\frac{k}{n})=0\}=1$,所以 \overline{X} 是完备 统计量.

充分完备统计量

如果一个统计量既是充分的,又是完备的,则称为充分完备统计量.







四、指数型分布族

1、指数型分布族的概念

判断一个统计量 $T(X_1, X_2, \dots, X_n)$ 是否为充分 完备统计量比较复杂,为此介绍一类分布族,其参 数的充分完备统计量容易发现。

定义1.7设总体X的分布密度为 $f(x,\theta)$,其中 $\theta = (\theta_1, \theta_2, \dots, \theta_m)^T, X_1, X_2, \dots, X_n$ 为其样本,若样本 的联合分布密度具有形式

$$\prod_{i=1}^{n} f(x_{i}, \theta) = C(\theta) \exp\{\sum_{j=1}^{m} b_{j}(\theta) T_{j}(x_{1}, x_{2}, \dots, x_{n})\} h(x_{1}, x_{2}, \dots, x_{n})\}$$







其中 $C(\theta)$, $b_i(\theta)$ 只与 θ 有关而与样本无关, T_i ,h 只与 样本有关,而与 θ 无关.则称 $f(x,\theta)$ 为指数型分布族.

说明 对于离散型分布律也有类似的定义.

2、指数型分布族参数的充分完备统计量的构造

定理1.5 设总体X的分布密度为 $f(x,\theta)$ 为指数型 分布族, 即样本的联合分布密度具有形式

$$\prod_{i=1}^{n} f(x_{i}, \theta) = C(\theta) \exp\{\sum_{j=1}^{m} b_{j}(\theta) T_{j}(x_{1}, x_{2}, \dots, x_{n})\} h(x_{1}, x_{2}, \dots, x_{n})$$

其中 $\theta = (\theta_1, \theta_2, \dots, \theta_m)^T$, $\theta \in \Theta$,如果 Θ 包含一个m维矩形,而且 $B = (b_1(\theta_1), b_2(\theta_2), \dots, b_n(\theta_n))^T$ 的值域





包含有一个m维的开集,则 $T = (T_1(X_1, X_2, \dots, X_n), T_2(X_1, X_2, \dots, X_n), \dots, T_m(X_1, X_2, \dots, X_n))^T$ 是参数 $\theta = (\theta_1, \theta_2, \dots, \theta_m)^T$ 的充分完备统计量. 证明(略)





解
$$P\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\} = \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} e^{-n\lambda}$$

$$= \frac{1}{\prod_{i=1}^{n} x_i!} \lambda^{n\frac{\sum_{i=1}^{n} x_i}{n}} e^{-n\lambda} = e^{-n\lambda} \exp\{\overline{X} \ln \ln \lambda\} \frac{1}{\prod_{i=1}^{n} x_i!}$$

其中
$$T(x_1, x_2, \dots, x_n) = \overline{X}, h(x_1, x_2, \dots, x_n) = \frac{1}{\prod_{i=1}^n x_i!}$$

 $C(\lambda) = e^{-n\lambda}, b(\lambda) = n \ln \lambda,$ 因而, \bar{X} 是充分完备统计量





例8(p12 例1.10) 设 $(X_1, X_2, \dots, X_n)^{\mathrm{T}}$ 是来自正态总体

$$N(\mu, \sigma^2)$$
的一个样本,试证 $T(X_1, X_2, \dots, X_n) = (\bar{X}, \frac{1}{n} \sum_{i=1}^n X_i^2)^T$

是参数 $\theta = (\mu, \sigma^2)$ 的联合充分完备统计量.

解
$$L(\mu, \sigma^2) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\}$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{n\mu^2}{2\sigma^2}} \exp\{-\frac{n}{2\sigma^2} \frac{1}{n} \sum_{i=1}^n x_i^2 + \frac{n\mu}{\sigma^2} \overline{x}\}$$





其中
$$T(x_1, x_2, \dots, x_n) = (\overline{x}, \frac{1}{n} \sum_{i=1}^n x_i^2)^T$$
,

$$h(x_1, x_2, \cdots, x_n) = 1,$$

$$C(\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{n\mu^2}{2\sigma^2}}, B = (\frac{n\mu}{\sigma^2}, -\frac{n}{2\sigma^2})^{\mathrm{T}},$$

因此,
$$T(X_1, X_2, \dots, X_n) = (\bar{X}, \frac{1}{n} \sum_{i=1}^n X_i^2)^T \mathbb{E}(\mu, \sigma^2)^T$$
的

充分完备统计量.





Thank You!





