概率论与数理统计（9）

清华大学

2020 年春季学期

充分统计量

- 假设总体分布具有未知参数 θ, $T(x_1, \ldots, x_n)$ 是一个样本的统计量，它被称为充分的如果没有其他其他同一样本的统计量能比它得到更多关于于未知参数的信息。
- 设 $F(\theta, x_1, \ldots, x_n)$ 是统计量 T 的抽样分布，我们考虑 $T=t$ 情形下的 $X=(x_1, \ldots, x_n)$ 的条件分布 $F(\theta, X|T=t)$, 则有以下两种可能
  - $F(\theta, X|T=t)$ 与未知参数 θ 有关;
  - $F(\theta, X|T=t)$ 与未知参数 θ 无关。
- 充分统计量的定义：设 $x_1, \ldots, x_n$ 是来自某总体的样本，总体的分布为 $F(x;\theta)$, 统计量 $T=T(x_1, \ldots, x_n)$ 称为 θ 的充分统计量如果在给定 T 的取值后，$x_1, \ldots, x_n$ 的条件分布于 θ 无关。

充分统计量

- 设总体的分布为 $b(1,\theta)$, $X_1,\ldots,X_n$ 为样本，令 $T=X_1+\cdots+X_n$，则其为 θ 的充分统计量。

$$
\begin{aligned}
& P(X_1=x_1,\ldots,X_n=x_n \mid T=t) \\
=& \frac{P(X_1=x_1,\ldots,X_{n-1}=x_n, X_n=t-\sum_{i=1}^{n-1} x_i)}{P(\sum_{i=1}^{n} X_i=t)} \\
=& \frac{\prod_{i=1}^{n-1} \theta^{x_i}(1-\theta)^{1-x_i}\theta^{t-\sum_{i=1}^{n-1} x_i}(1-\theta)^{1-t+\sum_{i=1}^{n-1} x_i}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} \\
=& \frac{\theta^t(1-\theta)^{n-t}}{\binom{n}{t}\theta^t(1-\theta)^{n-5}} = \frac{1}{\binom{n}{t}}
\end{aligned}
$$

充分统计量

- 若总体的分布为 $N(\mu,1)$, 则 $T=\bar{x}$ 为充分统计量。
- $$p(x_1,\ldots,x_{n-1},t;\mu)\sim \exp\{\frac{-1}{2}[\sum_{i=1}^{n-1}(x_i-\mu)^2+(nt-\sum_{i=1}^{n-1}x_i-\mu)^2]\}$$
  $$\sim \exp\{\frac{-1}{2}[n(t-\mu)^2+\sum_{i=1}^{n-1}x_i^2+(\sum_{i=1}^{n-1}x_i-nt)^2-nt^2]\}$$
- $p_{\bar{x}}(t)\sim \exp\{-\frac{n}{2}(t-\mu)^2\}$.

充分统计量

- 总体为均匀分布 $U(0,\theta)$，则 $x_{(n)}$ 是关于 θ 的充分统计量。

$$p(x_1;\theta)\ldots p(x_n;\theta)=\begin{cases}1/\theta^n, & 0<\min\{x_i\}\leqslant\max\{x_i\}<\theta,\\ 0, & else.\end{cases}$$

$g(t,\theta)=\frac{1}{\theta^n}I_{t<\theta}$，$h(X)=1$.

- 总体为 $N(\mu,\sigma^2)$，$\theta=(\mu,\sigma^2)$，则若 $t=\sum_{i=1}^n x_i$ 和 $T=\sum_{i=1}^n x_i^2$，则 (t,T) 为充分统计量。

$$p(X;\theta)\sim e^{-\frac{\sum_{i=1}^n(x_i-\mu)^2}{2\sigma^2}}\sim e^{-\frac{n\mu^2}{2\sigma^2}}e^{-\frac{\sum_{i=1}^n x_i^2-2\mu\sum_{i=1}^n x_i}{2\sigma^2}}$$

$$g(t,T,\theta)\sim e^{-\frac{n\mu^2}{2\sigma^2}}e^{-\frac{T-2\mu t}{2\sigma^2}},\quad h(X)=1.$$

充分统计量

- 寻找充分统计量：总体为 $p(x;\theta) = mx^{m-1}\theta^{-m}e^{-(\frac{x}{\theta})^m}$, $x > 0$, $\theta > 0$, $m > 0$. m 是已知量，$\theta > 0$ 为未知量。

$$p(x_1, \dots, x_n; \theta) = m^n (x_1 \dots x_n)^{m-1} \theta^{-mn} e^{-\sum_{i=1}^n x_i^m / \theta^m}.$$

若令 $T = \sum_{i=1}^n x_i^m$, 取 $g(t,\theta) = \theta^{-nm} e^{-t/\theta^m}$,
$h(x_1, \dots, x_n) = m^n (\prod_{i=1}^n x_i)^{m-1}$, 则有因子分解定理得知，
$T = \sum_{i=1}^n x_i^m$ 是 θ 的充分统计量。

无偏差估计

- 对于任一总体，样本均值都是总体的均值（数学期望）的无偏差估计。

$$E(\bar{x}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = E(X).$$

- 样本方差 $s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ 不是总体方差的无偏差估计：

$$E(s_n^2) = \frac{n-1}{n}\sigma^2.$$

不过，$E(s_n^2) \to \sigma^2$, $n \to \infty$。这情形，我们称之为渐进无偏差估计。

- 对样本方差稍微修整以下 $s^2 = \frac{n}{n-1}s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$. s^2 为总体方差的无偏差估计。

- 统计量的无偏差性不具有函数不变性。即，如果 $\hat{\theta}$ 是参数 θ 的无偏差估计，我们不能直接认为 $g(\hat{\theta})$ 是 $g(\theta)$ 的无偏差估计。如 s^2 是方差的无偏差估计，$s=\sqrt{s^2}$ 一般不是标准差的无偏差估计。
- 假设总体是 $N(\mu,\sigma^2)$，则 $Y=\frac{(n-1)s^2}{\sigma^2}\sim\chi^2(n-1)$，则

$$E(s)=\frac{\sigma}{\sqrt{n-1}}E(Y^{1/2})=\sqrt{\frac{2}{n-1}}\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}\sigma.$$

在这种情形下，s 是标准差的渐进无偏差估计。也可以修正为无偏差估计

$$s/\left(\sqrt{\frac{2}{n-1}}\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}\right).$$

缩小估计中偏差的方法：刀切法

- 设 $T(x)$ 是基于样本 $x=(x_1,\ldots,x_n)$ 的关于参数 $g(\theta)$ 的统计量，且 $E_\theta T(x)=g(\theta)+O(\frac{1}{n})$. 用 x_{-i} 表示在 x 中删除 x_i 后得到的样本向量，即 $(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n)$。定义刀切统计量为：

$$T_J(x)=nT(x)-\frac{n-1}{n}\sum_{i=1}^{n}T(x_{-i}).$$

- $E_\theta T_J(x)=g(\theta)+O(\frac{1}{n^2})$. 而且其方差也不会增大。
- 大概的原因：假设 $E_\theta T(x)=g(\theta)+\frac{a_1}{n}+\frac{a_2}{n^2}+\cdots$. 则

$$E_\theta T_J=(ng(\theta)+\frac{na_1}{n}+\frac{na_2}{n^2})-\frac{n-1}{n}(ng(\theta)+\frac{na_1}{n-1}+\frac{na_2}{(n-1)^2})$$
$$=g(\theta)+\frac{a_2}{n}-\frac{a_2}{n-1}+O(\frac{1}{n^3}).$$

刀切法

- 有时候能得到无偏差估计。如果总体为 $b(1,\theta)$，而 $g(\theta)=\theta^2$. 考虑 $T(x)=\bar{x}^2$.

$$E(T(x))=\theta^2+\frac{\theta(1-\theta)}{n}.$$

应用刀切法，

$$T(x_{-i})=(\frac{n\bar{x}-x_i}{n-1})^2=\frac{n^2\bar{x}^2+x_i^2-2nx_i\bar{x}}{(n-1)^2}.$$

于是

$$T_J(x)=n\bar{x}^2-\frac{n-1}{n}\sum_{i=1}^{n}\frac{n^2\bar{x}^2+x_i^2-2nx_i\bar{x}}{n-1}=\frac{n\bar{x}^2}{n-1}-\frac{\sum_{i=1}^{n}x_i^2}{n(n-1)}.$$

$$E(T_J)=\frac{n}{n-1}(\theta^2+\frac{\theta(1-\theta)}{n})-\frac{n(\theta^2+\theta(1-\theta))}{n(n-1)}=\theta^2.$$

不可估参数

- 当总体中的某未知参数存在无偏差估计，则称该参数为可估的，否则称其为不可估的。
- 总体为 $b(1,p)$，则参数 $\theta = \frac{1}{p}$ 是不可估的。
- $T = x_1 + \cdots + x_n$ 是充分统计量。若 $\hat{\theta}$ 是 θ 的无偏差估计，则对于任意 $p > 0$，

$$\frac{1}{p} = E_\theta(\hat{\theta}) = E(E(\hat{\theta}|T=t)) = \sum_{i=1}^{n} E(\hat{\theta}|T=i)\binom{n}{i}p^i(1-p)^{n-i}.$$

显然 $E(\hat{\theta}|T=t)$ 不可能都等于 0，则上式为 p 的 $n+1$ 次多项式，最多 $n+1$ 个 p 是等号成立。也就是不可能对所有可能的 p 都成立。矛盾。

无偏差估计的有效性

- 假设 θ_1 和 θ_2 均是参数 θ 的无偏差估计，若有

$$Var_\theta(\theta_1) \leqslant Var_\theta(\theta_2)$$

且至少有一个 θ 使得不等号严格成立，则称 θ_1 比 θ_2 有效。

- 如总体的均值为 μ，方差为 σ^2。$x_1, \dots, x_n$ 是一个样本，则 $\mu_3 = \frac{1}{3}(x_1 + x_2 + x_3)$ 和样本均值 $\bar{x}$ 都是 μ 的无偏差估计，

$$Var(\mu_3) = \frac{\sigma^2}{3}, \quad Var(\bar{x}) = \frac{\sigma^2}{n}.$$

当样本容量大于 3 时，$\bar{x}$ 比 μ_3 有效。

无偏差估计的有效性

- 若 $x_1, \ldots, x_n$ 为总体 $U(0,\theta)$ 的样本，则 $\theta_1 = \frac{n+1}{n} x_{(n)}$ 是 θ 的无偏差估计。

$$E(\theta_1) = \frac{n+1}{n} E(x_{(n)}) = \frac{n+1}{n} \frac{n}{n+1} \theta = \theta.$$

$$Var(\theta_1) = (\frac{n+1}{n})^2 Var(x_{(n)}) = \frac{\theta^2}{n(n+2)}.$$

- $\theta_2 = 2\bar{x}$ 也是 θ 的无偏差估计。

$$E(\theta_2) = 2E(\bar{x}) = 2\frac{\theta}{2} = \theta$$

$$Var(\theta_2) = \frac{4}{n} \frac{\theta^2}{12} = \frac{\theta}{3n}.$$

- 样本容量大于 1 时，θ_1 比 θ_2 有效。

矩估计

- 用样本矩来估计总体的矩。
- 用样本的矩的函数来估计总体矩的函数。
- 随机变量的矩：$E(X^k)$(k-阶原点矩), $E((X-E(X))^k)$(k 阶中心矩).
- 样本的矩：k-阶样本原点矩：

$$\frac{1}{n}\sum_{i=1}^{n} x_i^k.$$

k-阶样本中心矩：

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^k.$$

- 假设总体具有已知的概率函数 $p(x;\theta_1,\dots,\theta_k)$. 且总体的 k-阶原点矩 μ_k 存在。又假设 $\theta_j=\theta_j(\mu_1,\dots,\mu_k)$, 则 θ_j 的矩估计为

$$\hat{\theta}_j=\theta_j(a_1,\dots,a_k),\quad a_i=\frac{1}{n}\sum_{k=1}^{n}x_k^i.$$

如果要估计 $\eta=g(\theta_1,\dots,\theta_k)$, 用矩估计方法得到的统计量为

$$\hat{\eta}=g(\hat{\theta}_1,\dots,\hat{\theta}_k).$$

- $\mu=E(X)$, $\hat{\mu}=\bar{x}$.
- $\sigma^2=Var(X)=E(X^2)-(E(X))^2$, $\hat{\sigma}^2=\frac{1}{n}\sum_{k=1}^{n}x_i^2-(\bar{x})^2$.

矩估计

- 总体为二点分布 $b(1,p)$, $\theta = \frac{1}{p}$, $\hat{\theta} = \frac{1}{\bar{x}}$.

矩估计

- 总体为二点分布 $b(1,p)$, $\theta = \frac{1}{p}$, $\hat{\theta} = \frac{1}{\bar{x}}$.
- 总体为指数分布 $Exp(\lambda)$: $p(x:\lambda) = \lambda e^{-\lambda x}$, $x > 0$. $E(X) = \frac{1}{\lambda}$. 所以 $\lambda = \frac{1}{E(X)}$.

$$\hat{\lambda} = \frac{1}{\bar{x}}.$$

$Var(X) = \frac{1}{\lambda}$, 则 $\lambda = \frac{1}{\sqrt{Var(X)}}$,

$$\lambda = \frac{1}{s}.$$

- 矩估计可以是多个形式。尽量采取使用低阶矩的矩估计。

矩估计

- 总体为均匀分布 $U(a,b)$, a 和 b 都是未知参数。

$$EX = \frac{a+b}{2}, \quad Var(X) = \frac{(b-a)^2}{12}.$$

$$a = EX - \sqrt{3\,Var(X)}, \quad b = E(X) + \sqrt{3\,Var(X)}.$$

a, b 的矩估计为

$$\hat{a} = \bar{x} - \sqrt{3}s, \hat{b} = \bar{x} + \sqrt{3}s.$$

- $\hat{a}_1 = x_{(1)}$, $\hat{b}_1 = x_{(n)}$.
- 都是渐进无偏估计。

统计估计量的相合性

- 若 $\theta_n = \hat{\theta}_n(x_1, \ldots, x_n)$ 是 θ 的一个估计量，n 是样本容量，若对任意的 $\epsilon > 0$, 都有

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| \geqslant \epsilon) = 0$$

  则称 θ_n 为参数 θ 的相合估计。
- 总体为正态分布 $N(\mu, \sigma^2)$，$\bar{x}$ 为 μ 的相合估计，s_n^2 与 s^2 均是 σ_2^2 的相合估计.
- 总体为 $U(a, b)$ 分布，则 $x_{(1)}$ 为 a 的相合估计，$x_{(n)}$ 是 b 的相合估计。

- 设 $\theta_n = \hat{\theta}_n(x_1, \ldots, x_n)$ 是 θ 的一个估计量，如果 $\lim_{n\to\infty} E(\hat{\theta}_n) = \theta, \quad \lim_{n\to\infty} Var(\hat{\theta}_n) = 0$，则 $\hat{\theta}_n$ 是 θ 的相合估计。

$$P(|\hat{\theta}_n - E(\hat{\theta}_n)| \geqslant \frac{\epsilon}{2}) \leqslant \frac{4}{\epsilon^2} Var(\hat{\theta}_n).$$

当 n 充分大时 $|E(\hat{\theta}_n) - \theta| < \frac{\epsilon}{2}$.

$$\{|\hat{\theta}_n - E\hat{\theta}_n| < \epsilon/2\} \subset \{|\hat{\theta}_n - \theta| < \epsilon\}.$$

- 若 $\hat{\theta}_{n1}, \ldots, \hat{\theta}_{nk}$ 是 $\theta_1, \ldots, \theta_k$ 的相合估计，$\eta = g(\theta_1, \ldots, \theta_k)$ 则 $\hat{\eta}_n = g(\hat{\theta}_{n1}, \ldots, \hat{\theta}_{nk})$ 是 η 的相合估计。也就是说，相合估计具有连续函数不变性。
- 矩估计一般都具有相合性：由大数定律，样本原点矩是原点矩的相合估计（在满足某阶矩的有界性的前提下）。？为何?