1.3 统计量

样本 p 分位数 $X_{(m)}$, $m=[(n+1)p]$.

经验分布函数.

$$F_n(x) = \begin{cases} 0 & x \le X_{(1)} \\ \frac{k}{n} & X_{(k)} < x < X_{(k+1)}. \\ 1 & X_{(n)} < x. \end{cases}$$

单调非降, 左连续. $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty, x)}(X_i)$

$F_n(x)$ 可看作 n 个 iid $b(1, F(x))$ 变量 $Y_1 \cdots Y_n$ 的均值. 从而有

(i) $\dfrac{\sqrt{n}\,(F_n(x) - F(x))}{\sqrt{F(x)(1-F(x))}} \xrightarrow{d} N(0,1)$;

(ii) $F_n(x) \xrightarrow{P} F(x)$;

(iii) $F_n(x) \xrightarrow{a.s.} F(x)$.

(iv) Glivenko - Cantelli Theorem:

记 $D_n = \sup_x |F_n(x) - F(x)|$, 则 $D \to 0$, a.s.

常用不等式

Markov: X 取非负值 $P(X\geq a)\leq \frac{E[X]}{a}$.

Chebyshev: $P(|X|\geq a)\leq \frac{Eg(|X|)}{g(a)}$. g为$[0,+\infty)$非降非负值函数

特别. $P(|X-\mu|\geq k)\leq \frac{\sigma^2}{k^2}$.

Chebyshev (单边): $P(X>a)\leq \frac{\sigma^2}{\sigma^2+a^2}$

2.2 正态总体

Thm 2.2.2. 设 $X_1,\cdots X_n$ iid $\sim N(a,\sigma^2)$. 作正交变换 $Y=AX$. 则 Y 各个分量独立

Thm 2.2.3. 设 $X_1 \cdots X_n$ iid $\sim N(a,\sigma^2)$ 则

(i) $\bar{X}\sim N(a.\sigma^2/n)$

(ii) $(n-1)S^2/\sigma^2\sim \chi^2_{n-1}$

(iii) $\bar{X}$ 与 S^2 独立.

2.4. χ^2分布, t分布与F分布

χ^2分布: $g_n(x)=\frac{(\frac{x}{2})^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2})}\cdot\frac{1}{2}e^{-\frac{x}{2}}$ $(x>0)$

均值 n, 方差 $2n$.

T分布: $X\sim N(0,1)$. $Y\sim\chi^2_n$, X,Y独立.

$T=\frac{X}{\sqrt{Y/n}}$ 定义为自由度为 n 的t分布. $n=1$时为Cauchy分布. $n\to\infty$时 趋于$N(0,1)$

F分布: $X\sim\chi^2_m$, $Y\sim\chi^2_n$, X,Y独立

$F=\frac{X/m}{Y/n}$ 定义为 自由度是m和n的F分布.

2.6 指数族.

正态分布,二项分布,poisson分布,负二项分布,指数分布和Γ分布均属于指数族.

标准形式

$$f(x,\theta) = C(\theta)\cdot\exp\left\{\sum_{i=1}^{k} Q_i(\theta)T_i(x)\right\}h(x).$$

指数族支撑集与θ无关.

自然形式:

在标准形式下令$\varphi_i = Q_i(\theta)$.

$$f(x,\varphi) = C^*(\varphi)\exp\left\{\sum_{i=1}^{k}\varphi_i T_i(x)\right\}h(x).$$

自然参数空间:

$$\Theta^* = \{\varphi = (\varphi_1 \cdots \varphi_n)\}.$$

Thm 2.6.1 指数族在自然形式下,自然参数空间为凸集

2.7 充分统计量.

Def 2.7.1 设样本X分布族为$\{f(x,\theta), \theta\in\Theta\}$, $T(x)$为统计量.若样本X的条件分布与T无关则称T是充分的.

Thm 2.7.1 (因子分解定理). 设样本$X = (x_1, \dots x_n)$为概率密度$f(x,\theta)$. $T = T(x)$为统计量,则T为充分统计量$\Longleftrightarrow f(x,\theta) = g(t(x),\theta)\,h(x)$, $t(x)$为$T(x)$的观察值.

推论: 设$T = T(\vec{X})$为θ的充分统计量, $S = \varphi(T)$单值,可逆.则S也为θ的充分统计量.

Def 2.7.2. T是分布族下的极小充分统计量,若对所有充分统计量$S(X)$.有函数g_s,使得$T = g_s(S)$.

注:充分完全统计量必是极小充分的.

2.8 完全统计量

Def 2.8.1. 设 $\{f(x,\theta), \theta\in\Theta\}$ 为分布族. $T(x)$ 为统计量. 若对任何满足 $E_\theta[\varphi(T(x))]=0$ 的 $\varphi(T(x))$, 都有 $\varphi(T(x))=0$, a.s. 那么称 $T(x)$ 是完全统计量.

推论: T为完全统计量, S为任一可测函数, 则 $S(T)$ 也为完全统计量.

$T(x)=\sum_{i=1}^{n}x_i$ 对 $b(1,\theta)$, $N(\theta,1)$ 都是完全的; $X_{(n)}$ 对 $U(0,\theta)$ 也是完全的.

Thm 2.8.1 设样本 $X=(X_1,\cdots X_n)$ 概率函数

$$f(x,\theta)=C(\theta)\exp\{\sum_{i=1}^{k}\varphi_i T_i(x)\}h(x) \quad . \theta=(\varphi_1\dots\varphi_k)\in\Theta^*$$

为指数族的自然形式. 令 $T(x)=(T_1(x),\cdots,T_k(x))$. 若 Θ^* 有内点(作为 R^k 子集), 则 $T(x)$ 为完全统计量.

Thm 2.8.2 (Basu) $T(x)$ 为有界完全统计量, 也是充分统计量. 若随机变量 $V(x)$ 的分布与 θ 无关, 则对一切 $\theta\in\Theta$, $V(x)$ 与 $T(x)$ 独立.

3.2 矩估计.

样本k阶原点矩 $a_{nk} = \frac{1}{n}\sum_{i=1}^{n} x_i^k$;

样本k阶中心矩 $m_{nk} = \frac{1}{n}\sum_{i=1}^{n} (x_i-\bar{x})^k$

3.3. 极大似然估计

似然函数 $f(x,\theta) = P_\theta(X_1 = x_1, \cdots X_n = x_n)$.

求极大似然估计.

$\frac{\partial l(\theta,x)}{\partial \theta_i} = 0$ 称为似然方程组，若满足①似然函数的极大值在参数空间 Θ 内取到;

②似然方程只有唯一解. 对指数族来说，有解则必唯一.

注：当 $\hat{\theta}^* \notin \Theta$ 时若存在一列 $\{\hat{\theta}_n^*\} \in \Theta$. $\lim_{n\to\infty} \hat{\theta}_n^* = \hat{\theta}^*$，则 $\hat{\theta}^*$ 为 MLE.

Thm 3.3.2

若 $\{f(x,\theta), \theta\in\Theta\}$ 满足.

(1) $\frac{\partial \log f(x,\theta)}{\partial\theta}$, $\frac{\partial \log f(x,\theta)}{\partial\theta^2}$, $\frac{\partial \log f(x,\theta)}{\partial\theta^3}$ 存在.

(2) 存在定义于样本空间的函数 $F_1(x), F_2(x), H(x)$. 使得对一切 $\theta\in\Theta$.

$\left|\frac{\partial f(x,\theta)}{\partial\theta}\right| < F_1(x)$, $\left|\frac{\partial^2 f(x,\theta)}{\partial\theta^2}\right| < F_2(x)$, $\left|\frac{\partial^3 f(x,\theta)}{\partial\theta^3}\right| \le H(x)$

其中 $\int_{-\infty}^{+\infty} F_i(x)\,dx < \infty$; $\int_{-\infty}^{+\infty} H(x) f(x,\theta)\,dx < M$. $\theta\in\Theta$.

(3) 对一切 $\theta\in\Theta$,

$0 < I(\theta) < \infty$. $I(\theta) := E\left[\left(\frac{\partial \log f(x,\theta)}{\partial\theta}\right)^2\right]$.

(4) 对数似然方程有唯一根 $\hat{\theta}^*$.

则 $\sqrt{n}(\hat{\theta}^* - \theta) \xrightarrow{d} N\left(0, \frac{1}{I(\theta)}\right)$. 且 $\hat{\theta}^* \to \theta$.

3.4. 一致最小方差无偏估计.

Def 3.4.2 (UMVUE) $\hat{g}^*(X)$ 为 $g(\theta)$ 的 UMVUE, 若对 θ 的任意无偏估计 $\hat{g}(X)$

$Var(\hat{g}^*(X)) \le Var(\hat{g}(X))$.

lemma 3.4.1. $T=T(X)$ 为充分统计量, $\hat{g}(X)$ 为 $g(\theta)$ 的一个无偏估计. 则
$h(T)=E[\hat{g}(X)|T]$ 为 $g(\theta)$ 的无偏估计, 且 $Var(h(T)) \le Var(\hat{g}(X))$.

定理 (零无偏估计法): $\hat{g}(X)$ 为 $g(\theta)$ 的一个无偏估计. $Var_\theta(\hat{g}(X)) < \infty$.
若对所有满足 $E_\theta[l(X)]=0$ 对所有 θ 成立的 $l(X)$, 都有 $E_\theta[\hat{g}(X)l(X)]=0$
则 $\hat{g}(X)$ 为 $g(\theta)$ 的 UMVUE.

推论: T 为充分统计量. $h(T(X))$ 为无偏估计. 若对所有零无偏估计量 $\delta(T)$.
都有 $E_\theta[h(T)\cdot\delta(T)]=0$, 则 $h(T)$ 为 $g(\theta)$ 的 UMVUE.

Thm 3.4.2 (Lehmann-Scheff) $T(X)$ 为充分完全统计量. 若 $\hat{g}(T(X))$ 为无偏估计, 则 $\hat{g}(T(X))$ 是 $g(\theta)$ 的唯一 UMVUE

3.5. Cramer-Rao 不等式

Thm 3.5.1 $\{f(x,\theta), \theta\in\Theta\}$ 是 C-R 正则分布族 $g(\theta)$ 是 Θ 上的可微函数.
$\hat{g}(X)$ 是 $g(\theta)$ 的任一无偏估计. 则:

$$Var_\theta[\hat{g}(X)] \ge \frac{[g'(\theta)]^2}{nI(\theta)}, \quad I(\theta)=E\left[\left(\frac{\partial \log f(x,\theta)}{\partial\theta}\right)^2\right]$$

对于指数族, 唯有在 $\hat{g}(X)=aT(X)+b$ ($T(X)$ 为指数族参数) 时才能达到 C-R 不等式的下界.