§3 统计与信息不等式

1. **信息不等式(Rao-Cramer 不等式,或 Frechet 不等式)**:设随机向量 $\xi = (\xi_1,\cdots,\xi_n)$ 取值于 R^n,其概率分布由密度 $p(x \mid \theta)$ 决定,其中 $x = (x_1,\cdots,x_n)^T$,$\theta \in \Omega \subset R^1$,设统计量 $T = T(\xi)$ 满足条件 $E_\theta T = \theta + h(\theta)$.其中 $h(\theta)$ 为可微函数,称为 T 的偏倚,θ 为未知数值参数.Fisher 信息量定义为

$$I(\theta) = E\left[\frac{\partial \ln P(\xi \mid \theta)}{\partial \theta}\right]^2.$$

若 $I(\theta) \neq 0$,则 $$E_\theta \mid T - \theta \mid^2 \geqslant \frac{[1 + h'(\theta)]^2}{I(\theta)} + [h(\theta)]^2. \tag{3.1}$$

特别地,若 T 是θ 的无偏估计量(即 $E_\theta T = \theta$),则

$$DT = E_\theta \mid T - \theta \mid^2 \geqslant \frac{1}{I(\theta)}. \tag{3.2}$$

若(3.2) 式关于某个无偏估计量,T 变为等式,则在所有无偏估计类中在最小平方风

险意义下 T 是最优的,这样的估计量 T 称为**有效估计量**,例如,若 $\xi_1,\cdots,\xi_n$ 是独立随机变量,服从同一正态律 $N(\theta,1)$,则 $T=\frac{1}{n}\sum_{k=1}^{n}\xi_k$ 是未知均值 θ 的有效估计量.

在一般情形下,(3.2) 式中等号成立,仅当 $\{p(x\mid\theta)\}$ 是指数分布族,即随机向量 ξ 的概率密度可表示为

$$p(x\mid\theta)=C(x)\exp\{u(\theta)\varphi(x)-v(\theta)\}.$$

在向量参数情形下,(3.1) 式有不同的推广,并且可以推广到估计此参数的函数情形. 例如,设参数空间 $\Omega\subset R^1$,参数函数 $g(\theta)$ 的任意无偏估计量为 $\varphi(\xi)$,而 $\varphi_0(\xi)$ 是使方差在 θ_0 处为最小的估计量,$\theta_0,\theta\in\Omega$,以参数 θ 为附标的密度函数 $p(x\mid\theta)$ 的均值和方差分别记为 E_θ 和 D_θ,$\pi(x\mid\theta)=\frac{P(x\mid\theta)}{P(x\mid\theta_0)}$. 令

$$D_{\theta_0}(\varphi(\xi))=\sup\Big\{\sum_{k=1}^{n}\sum_{j=1}^{n}g(\theta_k)g(\theta_j)\lambda^{kj}\Big\},$$

式中上确界对 $\forall\, n\in N$ 和 $\theta_1,\cdots,\theta_n\in\Omega$ 而取的. λ^{kj} 是以 $\lambda_{kj}=E_{\theta_0}(\pi(\xi\mid\theta_k)\pi(\xi\mid\theta_j))$ 为元素的 $n\times n$ 矩阵 (λ_{kj}) 的逆 (k,j) 元素.

在某些正则条件下(例如 $E_{\theta_0}\{[\pi(\xi\mid\theta)]^2\}<\infty,\theta\in\Omega,p(x\mid\theta)$ 在 $\theta=\theta_0$ 处有偏导数 $p'(x\mid\theta_0)$(a.e. $P(x\mid\theta_0)$,) 及

$$\lim_{\Delta\theta\to0}E_{\theta_0}\Big[\frac{P(\xi\mid\theta_0+\Delta\theta)-P(\xi\mid\theta_0)}{P(\xi\mid\theta_0)\Delta\theta}-\frac{P'(\xi\mid\theta_0)}{p(\xi\mid\theta_0)}\Big]^2=0.$$

则成立**信息不等式**:

$$D_{\theta_0}(\varphi(\xi))\geqslant\frac{[g'(\theta_0)]^2}{E_{\theta_0}\left\{\left[\frac{\partial\ln P(\xi\mid\theta)}{\partial\theta}\Big|_{\theta=\theta_0}\right]^2\right\}}.$$

若在下述正则条件下,$E_{\theta_0}\{[\pi(\xi\mid\theta)]^2\}<\infty,\theta\in\Omega,p(x\mid\theta)$ 在 $\theta=\theta_0$ 处关于 θ 为 k 次可偏微(a.e. $P(x\mid\theta_0)$);第 j 阶($1\leqslant j\leqslant k$) 偏导数 $p^{(j)}(x\mid\theta_0)$ 满足

$$\lim_{\Delta\theta\to0}E_{\theta_0}\left[\left(\frac{\Delta^jP(\xi\mid\theta)_{\theta=\theta_0}}{P(\xi\mid\theta_0)(\Delta\theta)^j}-\frac{P^{(j)}(\xi\mid\theta_0)}{P(\xi\mid\theta_0)}\right)^2\right]=0,$$

则成立 Bhattacharya 不等式:

$$D_{\theta_0}(\varphi(\xi))\geqslant\sum_{i=1}^{k}\sum_{j=1}^{k}g^{(i)}(\theta_0)g^{(j)}(\theta_0)K^{ij},$$

式中 $g^{(i)}(\theta_0)$ 是 $g(\theta)$ 在 θ_0 的 i 阶导数,K^{ij} 是以

$K_{ij}=E_{\theta_0}\left(\frac{P^{(i)}(\xi\mid\theta_0)}{P(\xi\mid\theta_0)}\cdot\frac{P^{(j)}(\xi\mid\theta_0)}{P(\xi\mid\theta_0)}\right)$　为元素的矩阵 (K_{ij}) 的逆的 (i,j) 元素,$i,j=1,2\cdots,k$. 若参数 $\theta=(\theta_1,\cdots,\theta_m)$ 是 m 维的,则在与一维相同的条件下,成立

$$D_{\theta_0}(\varphi(\xi))\geqslant\sum_{j=1}^{m}\sum_{k=1}^{m}g_j'(\theta_0)g_k'(\theta_0)A^{jk},$$

式中 $g_j'(\theta_0)$ 是 $g(\theta)$ 在 θ_0 关于 θ_j 的偏导数,A^{jk} 是以

$$A^{jk} = E_{\theta_0}\left\{\frac{\partial \ln P(\mid \xi \mid \theta)}{\partial \theta_j}\bigg|_{\theta=\theta_0} \cdot \frac{\partial \ln P(\xi \mid \theta)}{\partial \theta_k}\bigg|_{\theta=\theta_0}\right\}$$

为元素的矩阵 $A = (A_{jk})$ 的逆的 (j,k) 元素.

若 $\varphi(\xi)$ 是序贯无偏估计量,则成立 **Wolfowitz 不等式**:

$$D_{\theta_0}(\varphi(\xi)) \geqslant \frac{[g'(\theta)]^2}{E_{\theta_0}(N)I(\theta_0)},$$

式中 $I(\theta_0)$ 为信息量,N 是序贯抽样的样本大小.

见[144]P.1196－1199.

2.　**估计量的容许性不等式**:若参数函数 $g(\theta)$ 的任意估计量 $\varphi(\xi)$ 满足:

从 $r(\theta,\varphi) \leqslant r(\theta,\varphi_0)$ 可推出 $r(\theta,\varphi) = r(\theta,\varphi_0), \theta \in \Omega$,则称 $\varphi_0(\xi)$ 是容许的,式中 $r(\theta,\varphi) = E_\theta(W(\theta,\varphi(\xi)))$ 是 $\varphi(\xi)$ 的**风险函数**,而 $W(\theta,\varphi(\xi))$ 表示当参数的真值为 θ 时,由 $\varphi(\xi)$ 带来的损失,形如 $W(\theta,\varphi(\xi)) = \lambda(\theta)[\varphi(\xi) - g(\theta)]^2 (\lambda(\theta) > 0)$ 的损失函数 W 称为平方损失函数.特别当 $\lambda(\theta) = 1$ 时,$r(\theta,\varphi) = E_\theta([\varphi(\xi) - g(\theta)]^2)$ 称为 $\varphi(\xi)$ 的均方误差.若 $\varphi(\xi)$ 是 $g(\theta)$ 的无偏估计量,则均方误差与方差 $D_\theta(\varphi(\xi))$ 一致.若对实数 c,形如 $(\varphi(\xi))$ 的统计量作为 $g(\theta)$ 的估计量是容许的,则

$$\inf_\theta\left\{\frac{g(\theta)E_\theta(\varphi)}{E_\theta(\varphi^2)}\right\} \leqslant c \leqslant \sup_\theta\left\{\frac{g(\theta)E_\theta(\varphi)}{E_\theta(\varphi^2)}\right\}.$$

(见[144]P.1198)

3.　**极小极大估计量不等式**:若估计量 $\varphi^*(\xi)$ 满足

$$\sup_\theta r(\theta,\varphi^*) = \inf_\varphi \sup_\theta r(\theta,\varphi),$$

则称 φ^* 为极小极大的估计量,若存在先验分布列 $\{\xi_n\}$,使得 $\forall \theta \in \Omega$,成立

$$\liminf_{n\to\infty}\left[r(\theta,\varphi^*) - \inf_\varphi \int_\Omega r(\theta,\varphi)\mathrm{d}\xi_n(\theta)\right] \leqslant 0.$$

则 φ^* 是极小极大估计量(Wald).见[144]P.1199.

4.　**U-统计量列不等式**:设 ξ_1, ξ_n 是定义于同一完备概率空间 $(\Omega, \sum, P)$ 且取值于 R^1 的随机变量列,$S_0 = 0, S_n = \sum_{k=1}^n \xi_k$. $T_n = \sum_{1\leqslant k<j\leqslant n} \xi_k \xi_j, n = 1,2,\cdots, T_n^* = \max_{1\leqslant k\leqslant n} |T_k|$.令

$$S = \sum_{n=1}^\infty n^{\alpha p-2} P\{T_n^* \geqslant n^{2\alpha}\}.$$

1991 年周元燊证明:设 ξ, ξ_n 为独立同分布随机变量列,$E(\xi) = 0, 1 < r < 2, \alpha r > 1, \alpha p > 1$,则存在常数 $c = c(\alpha,p,r) > 0$,使得

$$S \leqslant c[E|\xi|^p + (E|\xi|^r)^q],\text{式中 } q = (\alpha p - 1)/(\alpha r - 1).$$

1993 年王岳宝证明:

(1)　设 ξ, ξ_n 为任意同分布的随机变量列,$0 < p < 1$,

若 $\alpha p > 1$,则 $S \leqslant cE(|\xi|^p)$;

若 $\alpha p = 1$,则 $\sum_{n=1}^\infty n^{-1} P\{T_n^* \geqslant n^{2\alpha}\} \leqslant cE(|\xi|^p)$.

(2) 设 ξ,ξ_n 是独立同分布随机变量列，则当 $p=1$ 时成立

$$S\leqslant c\max\{E(|\xi|),(E|\xi|)^2\};$$

当 $\alpha p=1,E(\xi)=0$ 时，成立

$$S\leqslant c\max\{E|\xi|,(E|\xi|)^2,(E|\xi|)\ln(|\xi|+1)(E|\xi|)^3\};$$

当 $\alpha p=1,1<p<2,E(\xi)=0$ 时成立

$$S\leqslant c\max\{E|\xi|^p,(E|\xi|)^2,[(E|\xi|^p)(E|\xi|)]^2\}.$$

见[333]1993,38(2):189－190.

5. **随机抽样不等式**：设 $\xi_1,\cdots,\xi_n$ 是取自分布函数 F 的随机样本. $\overline{\xi}=\dfrac{1}{n}\sum\limits_{k=1}^{n}\xi_k$ 称为样本均值. $S^2=\dfrac{1}{n-1}\sum\limits_{K=1}^{N}(\xi_k-\overline{\xi})^2$ 称为样本方差. 应注意样本统计量 $\overline{\xi},S^2$ 为随机变量，而本章 §2 中总体参数 $E(\xi)$ 与 σ^2 等是固定的常数，它们可以是未知的.

$$\max_{1\leqslant k\leqslant n}|\xi_k-\overline{\xi}|<\frac{(n-1)S}{\sqrt{n}}.$$

但要除去所有 n 个观察都相等或所有 ξ_k 中正好 $n-1$ 个相等的情形.(Samuelson).(见[143]P.364－365)

6. 设 F_n^* 是一个取自分布函数 F 的随机样本的经验分布，则 $\forall t>0$，成立

$$P\left\{|F_n^*(x)-F(x)|\geqslant\frac{t}{2\sqrt{n}}\right\}\leqslant\frac{1}{t^2}.\quad(见[143]P378)$$

7. **K-S 距离不等式**：设 $\xi_1,\cdots,\xi_n$ 是来自总体分布 $F(x)$ 的简单随机样本，$\varphi(A)$ 为随机事件 A 的示性函数. $F_n(x)=\dfrac{1}{n}\sum\limits_{k=1}^{n}\varphi\{\xi_k\leqslant x\}$ 称为经验分布函数.

$d_n=\sup\{|F_n(x)-F(x)|:x\in R^1\}$ 称为 K-S 距离(Kolmogorov-Smirnov 距离).

(1) **DKW 不等式(Dvoretzky-Kiefer-Wolfowitz 不等式)**：$\forall t>0$，存在不依赖于 F 的有限正常数 c，使得

$$P\{d_n>t\}\leqslant ce^{-2nt^2}.$$

(2) $\forall t>0$，存在 $c>0$，使得

$$P\{\sup_{k\geqslant n}d_k>t\}\leqslant\frac{c}{1-r}r^n,$$

式中 $r=\exp(-2t^2)$，(Serfling). 见[30]P.324.

注 事件 A 的示性函数 $\varphi(A)$，又称为 A 的特征函数，一般记为

$$\varphi_A(x)=\begin{cases}1,x\in A,\\0,x\notin A,\end{cases}$$

8. **Bogolyubov 不等式**：设 H,H_1,H_2 为 Hermit 算子. 自由能泛函定义为 $f(H)=-\dfrac{t}{n}\ln\mathrm{Tr}e^{-H/t}$，而 $\langle H_1-H_2\rangle_{H_1}=\dfrac{\mathrm{Tr}[(H_1-H_2)e^{-\frac{H_1}{t}}]}{\mathrm{Tr}(e^{-H_1/t})}$ 是 Hermit 算子 H_1 的热力学平均，则

$$\frac{1}{n}\langle H_1 - H_2\rangle_{H_1} \leqslant f(H_1) - f(H_2) \leqslant \frac{1}{n}\langle H_1 - H_2\rangle_{H_2}, \tag{3.3}$$

式中广延参数 n 是粒子数或容积,依系统而定,t 为能量单位的绝对温度.

(3.3) 式及统计力学中 Green 函数与关联函数的 Bogolyubov 不等式见[107]1:380 - 381.

9. **熵不等式**:设 $p_k \geqslant 0, \sum_{k=1}^{n} p_k = 1, S(x) = \sum_{k=1}^{n}(np_k)^x \ln(np_k), p = \frac{n}{n-1} - \frac{1}{\ln n}$, $q = \frac{1}{\ln n}\ln\left(\frac{n-1}{\ln n}\right)$,则 $S(p) \geqslant S(q) \geqslant 0$.

(Alzer, H., [372]1995, 38:13 - 18)

10. **Kraft 不等式**:给定两符号集 $X = \{x_1, \cdots, x_m\}$ 和 $A = \{a_1, \cdots, a_n\}$,用 A 中元素的序列作为元素构成一个新集 $B = \{b_1, \cdots, b_m\}$,并在 X 与 B 的元素间建立某种对应关系. 此过程称为**编码**,A 称为**码元素**,其元素称为**码元**. B 称为**码**,其元素称为**码字**,码元素 A 元素个数为 $m(m \geqslant 2)$ 的码称为 m **进元码**,码元素 A 各码元所占用的时间均相同的码称为**同阶码**,组成一个码字的码元个数称为**码长**,没有相同码字的码,称为**非奇异码**,在非奇异码中,若任何有限长码字序列都能被接收者惟一地分成单个码字,则该码称为**单义码**.

设码元素 A 含 d 个码元,所需编码含 m 个码字,各码字对应码长为 $n_k(k = 1, 2, \cdots, m)$,则单义码存在的充要条件是:

$$\sum_{k=1}^{m} d^{-n_k} \leqslant 1. \tag{3.4}$$

若一个码的任一个码字都不是其他码字的字首,则称该码为**非延长码**(或**即时码**),(3.4) 式对于非延长码也成立.

11. 约束的 Kantorovich 不等式及其在统计中的应用见[30]P273 - 295. 此外,线性规划所研究的领域,不过是以随机方程和不等式这个总题目出现的广泛得多的理论的一个小部分,许多物理现象都可以利用这种类型的数学方程来描述,这些数学方程中所包含的参数必须通过测量来确定,或利用计算技术得到,而计算又牵涉到舍入舍出误差等. 总之,参数实际上都是随机变量.