第二章　点估计

内容提要

§2.1 构造估计量的常用方法

一、矩估计

- 矩法是构造点估计的最古老的一种方法，直观、易行、适用面广。在复杂问题中，往往可以从矩估计出发寻找更好的估计量。

- 基本思路：用样本矩估计相应的总体矩；用样本矩的函数估计相应的总体矩的函数。（p312描述的是参数模型中的矩估计方法）

  (1) 设 $X_1, \cdots, X_n$ i.i.d. $P \in \mathscr{P}$，$\gamma_1, \cdots, \gamma_k$ 是k 个待估参数。若总体的前k 阶原点矩$\mu_j = E_P(X_1^j)$, $j = 1, \cdots, k$ 都存在，$\forall P \in \mathscr{P}$，且

$$\begin{cases} \mu_1 = \mu_1(\gamma_1, \cdots, \gamma_k), \\ \quad \vdots \\ \mu_k = \mu_k(\gamma_1, \cdots, \gamma_k), \end{cases}$$

  并由此可解出

$$\begin{cases} \gamma_1 = h_1(\mu_1, \cdots, \mu_k), \\ \quad \vdots \\ \gamma_k = h_k(\mu_1, \cdots, \mu_k), \end{cases}$$

  那么，将样本 j阶原点矩$m_j = \frac{1}{n}\sum_{i=1}^n X_i^j$ 替代 μ_j，$j = 1, \cdots, k$，即得 $\gamma_1, \cdots, \gamma_k$ 的矩估计

$$\begin{cases} \hat{\gamma}_1 = h_1(m_1, \cdots, m_k), \\ \quad \vdots \\ \hat{\gamma}_k = h_k(m_1, \cdots, m_k). \end{cases}$$

  (2) 设 $\eta = g(\gamma_1, \cdots, \gamma_k)$，$\hat{\gamma}_1, \cdots, \hat{\gamma}_k$是 $\gamma_1, \cdots, \gamma_k$ 的矩估计，则 $\hat{\eta} = g(\hat{\gamma}_1, \cdots, \hat{\gamma}_k)$ 是 η 的矩估计。

  (3) 若 $\gamma_1, \cdots, \gamma_k$ 能表示成若干原点矩与若干中心矩的函数，那么用样本原点矩替换同阶的总体原点矩、用样本中心矩 $\frac{1}{n}\sum_{i=1}^n (X_i - \mu_1)^j$ 替代同阶的总体中心矩，同样也能得到 $\gamma_1, \cdots, \gamma_k$ 的矩估计。

- 合理性：若$X_1, \cdots, X_n$ i.i.d. $P \in \mathscr{P}$，则矩估计往往有相合性。

- 例：

  (1) eg7.2.1 $N(\mu, \sigma^2)$ 两个参数的矩估计；

  (2) eg7.2.2 $B(k, p)$ 两个参数的矩估计；

(3) 若$X_1, \cdots, X_n$ i.i.d. $U(0, \theta)$, $\theta > 0$，θ 的矩估计$2\bar{X}$ 直观上有缺陷。

(4) 若$X_1, \cdots, X_n$ i.i.d. $Poi(\theta)$, $\theta > 0$，$\bar{X}, S^2$ 都是θ 的矩估计，哪个更好？

(5) eg7.2.3 Satterthwaite approximation.

- Generalized Method of Moments

二、极大似然估计(Maximum Likelihood Estimation, MLE)

极大似然估计方法主要适用于参数模型。设 X 的模型为$(\mathscr{X}, \mathscr{B}_X, \mathscr{P} = \{P_\theta : \theta \in \Theta\})$，其中$\Theta \subset \mathbb{R}^k$，$\mathscr{P}$ 关于$\sigma-$有限测度ν可控，且 $dP_\theta(x)/d\nu(x) = p(x; \theta)$。

- 似然函数：给定$x \in \mathscr{X}$，视 $p(x;\theta)$ 为θ 的函数，称此函数为观测值 x 的似然函数，记作
$$L(\theta; x) = p(x; \theta),\ \theta \in \Theta.$$
称
$$l(\theta; x) = \log p(x; \theta),\ \theta \in \Theta$$
为对数似然函数。

- MLE：记 $\hat{\theta}(x) = \arg\max_{\theta \in \Theta} L(\theta; x)$。若

  1) 对 $\forall x \in \mathscr{X} \setminus A$，$\hat{\theta}(x)$ 都存在，其中 $P_\theta(A) = 0,\ \forall \theta \in \Theta$；

  2) $\hat{\theta}(x)$ 是 $(\mathscr{X}, \mathscr{B}_X) \to (\overline{\Theta}, \mathscr{B}_{\overline{\Theta}})$ 的可测变换；

  则称 $\hat{\theta}(X)$ 是 θ 的极大似然估计(maximum likelihood estimator, MLE)。

- 不变原则：设 $\eta = \tau(\theta)$，称
$$L^*(\eta; x) = \sup_{\{\theta \in \overline{\Theta} : \tau(\theta) = \eta\}} L(\theta : x)$$
为导出似然函数；称
$$\hat{\eta}_{MLE} = \arg\max_{\eta} L^*(\eta; x)$$
为参数 η 的极大似然估计。显然 θ 的MLE
$$\hat{\theta}_{MLE} \in \{\theta : \tau(\theta) = \hat{\eta}_{MLE}\},$$
因此，$\tau(\hat{\theta}_{MLE})$ 是 η 的MLE。

- 计算方法：

  (1) 分析似然函数或者对数似然函数的单调性。

  (2) 在对数似然函数可微时，常通过求解对数似然方程（组）

  $$\frac{\partial l(\theta;x)}{\partial\theta}=0$$

  获得MLE。一般该方程（组）是非线性的，无显式解，需要采用Newton-Raphson 迭代法等数值方法求解。

- 例：eg7.2.5～eg7.2.13

三、构造点估计量的其他方法

U估计

最小二乘法

贝叶斯方法

惩罚似然方法

经验似然方法

等等

§2.2 一致最小方差无偏估计

§2.2.1 凸损失下点估计问题的一般结论

假定X 的模型为$(\mathscr{X},\mathscr{B}_X,\mathscr{P}=\{P_\theta,\theta\in\Theta\})$，$\gamma(\theta):\Theta\to\mathbb{R}^k$ 是一个感兴趣的待估参数(estimand)，x 是X 的观测值。点估计问题，就是要选择一个估计量$\delta(x):\mathscr{X}\to\mathbb{R}^k$ 去推断$\gamma(\theta)$。从统计推断角度看，希望该估计量的某种平均误差尽可能（关于参数θ）一致地小。从统计决策角度看，希望该估计量的风险函数尽可能一致地小。点估计的问题中，行动空间$\mathscr{A}$ 显然为$\mathbb{R}^k$ 中的子集，损失函数$L(\theta,a)$ 通常和$\gamma(\theta)$ 与a 之间的“距离”(即估计值的误差)有关：一般距离越大，损失越大；距离越小，损失越小。另外，往往还要求$L(\theta,a)$ 关于a 有连续性、可微性、凸性等良好的数学性质。

若行动空间$\mathscr{A}$ 为$\mathbb{R}^k$ 中的凸集，且对$\forall\theta\in\Theta$，当$a\in\mathscr{A}$ 时损失函数$L(\theta,a)$ 关于a 是凸函数，则称L 为凸损失函数。在凸损失情况下，点估计对应的统计决策问题的求解很大程度上可得以简化。

前一章谈到，对于欧氏样本空间，基于充分统计量的决策函数类构成全部决策函数类（包括随机化决策）的一个本质完全类。在凸损失情况下，对于点估计问题有更强的结论。

定理2.2.1 *[Shao J. p117/Th2.5]* 设 X 的模型为$(\mathscr{X},\mathscr{B}_X,\mathscr{P}=\{P_\theta,\theta\in\Theta\})$，欲估计某个 k 维参数 $\gamma(\theta)$，行动空间 $\mathscr{A}$ 为 $\mathbb{R}^k$ 中的凸集，L 为凸损失函数。

(i) δ 是一个随机化决策，满足 $\delta^*(x)\hat{=}\int_{\mathscr{A}}ad\delta(a|x)\in\mathscr{A}$, $\forall x\in\mathscr{X}$。则有

$$L(\theta,\delta^*(x))\le\int_{\mathscr{A}}L(\theta,a)d\delta(a|x).$$

若 L 严凸，且 $\exists\theta_0\in\Theta$，使得 $R(\theta_0,\delta^*)<\infty$ 且 $P_{\theta_0}\{x:\ \delta[\{\delta^*(x)\}|x]<1\}>0$，则 δ^* 优于 δ。

(ii)(Rao-Blackwell, 教材p342, Th7.3.17) 设样本空间是欧氏的，T 是分布族$\mathscr{P}$ 的充分统计量，δ 是一个非随机化估计量，且 $\delta(x)\in\mathscr{A}$, $\forall x\in\mathscr{X}$。记 $\eta(t)=E[\delta(X)|T=t]$，则估计量 $\eta(T)$ 的风险函数满足：

$$R(\theta,\eta)\le R(\theta,\delta),\quad\forall\theta\in\Theta.$$

若L 严凸，且$\exists\theta_0\in\Theta$ 使得$R(\theta_0,\delta)<\infty$，$P_{\theta_0}(\delta(X)\ne\eta(T))>0$，则$R(\theta_0,\eta)<R(\theta_0,\delta)$。

证明. 证(ii)，(i)的证明类似。因 T 是充分统计量，故 $\eta(T)$ 是一个统计量。记 $\phi_\theta(a)=L(\theta,a)$，在给定 $T=t$ 时 X 的条件分布记为$P^{X|t}$。由Jense不等式知，

$$L(\theta,\eta(t))=\phi_\theta(\eta(t))=\phi_\theta(E^{X|t}\delta(X))\le E^{X|t}\phi_\theta(\delta(X))=E^{X|t}L(\theta,\delta(X)).$$

对两边关于T 求期望，则有$R(\theta,\eta)\le R(\theta,\delta)$。这对所有$\theta\in\Theta$ 都成立。

若$P_{\theta_0}(\eta(T)\ne\delta(X))=P^T_{\theta_0}\{t:P^{X|t}[\eta(t)\ne\delta(X)]>0\}>0$，则$\exists B_T$，使得$P^T_{\theta_0}(B_T)>0$，当$t\in B_T$ 时，$P^{X|t}(\eta(t)\ne\delta(X))>0$。再由$R(\theta_0,\delta)<\infty$，可知$R(\theta_0,\eta)<R(\theta_0,\delta)$。 □

注：上述定理的两个结论分别说明，在凸损失情况下
(i) 非随机化估计类是一个本质完全类。若损失函数严凸，则随机化估计不可容许。
(ii) 由充分统计量构造的非随机化估计类是本质完全类。

推论2.2.1 (教材p343, Th7.3.19) *损失函数$L(\theta,a)$ 关于a 严凸，δ 是$\gamma(\theta)$ 的容许估计，若另一估计δ' 与δ 有相同的且处处有限的风险函数，即$R(\theta,\delta)=R(\theta,\delta')<\infty$, $\forall\theta\in\Theta$，则$P_\theta(\delta=\delta')=1,\forall\theta\in\Theta$。*

证明. (反证法)取$\delta^*=\frac{1}{2}(\delta+\delta')$，则由损失函数的凸性知

$$L(\theta,\delta^*)\le\frac{1}{2}[L(\theta,\delta)+L(\theta,\delta')],$$

因此，

$$R(\theta,\delta^*)\le\frac{1}{2}[R(\theta,\delta)+R(\theta,\delta')]=R(\theta,\delta),\quad\forall\theta\in\Theta.$$

若$\exists\theta_0\in\Theta$，使得$P_{\theta_0}(\delta\ne\delta')>0$，则由$R(\theta_0,\delta)<\infty$ 知

$$R(\theta_0,\delta^*)<R(\theta_0,\delta).$$

这与δ 的容许性矛盾。因此，$P_\theta(\delta=\delta')=1$, $\forall\theta\in\Theta$。 □

无偏估计

不加约束的一致最优估计一般不存在。无偏性是对估计量的一个直观而自然的要求。在无偏估计类中寻求一致最优估计，在不少情况下不仅理论上可行，而且应用上可取。

无偏估计的定义见定义第一章。设$\gamma(\theta)$ 是待估参数，δ是其无偏估计的充要条件是：
(1) 若 δ 是非随机化的，则

$$\int_{\mathscr{X}} \delta(x) dP_\theta(x) = \gamma(\theta), \quad \forall \theta \in \Theta.$$

(2) 若 δ 是随机化估计，则

$$\int_{\mathscr{X}} \big[\int_{\mathscr{A}} a d\delta(a|x)\big] dP_\theta(x) = \gamma(\theta), \quad \forall \theta \in \Theta.$$

无偏估计并不总是存在。若未知参数$\gamma(\theta)$ 存在无偏估计，则称$\gamma(\theta)$ 可估(estimable)。

例2.2.1 (Shao J. p154/§2.6, 84) 设$X \sim B(n,p)$，$0 < p < 1$ 未知，n 已知。记$\gamma(p) = \frac{1}{p}$，则$\gamma(p)$ 的U.E.不存在。

证明. 反证，假定$\delta(X)$ 是$\gamma(p)$ 的U.E.，则按U.E.的定义有

$$E_p\delta(X) = \sum_{x=0}^{n} \delta(x) \binom{n}{x} p^x (1-p)^{n-x} = \gamma(p) = \frac{1}{p}, \ \forall p \in (0,1).$$

显然，上式左边是p 的多项式，而右边不是，两者不可能在$p \in (0,1)$ 上处处相等。因此$\gamma(p)$ 的U.E.不存在。 □

但是，此例中 $\gamma(p)$ 的合适估计还是存在的，比如 $\delta(X) = \frac{n+1}{X+0.5}$。它是渐近无偏的、相合的。

$\gamma(\theta)$ 可估时，其U.E.未必优于有偏估计。

例2.2.2 (教材p331/Example 7.3.4) 设$X_1, \dots, X_n$ i.i.d. $N(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0$。记$\theta = (\mu, \sigma^2)$，考虑未知参数$\gamma(\theta) = \sigma^2$ 的估计。易知，$S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$ 是无偏估计，$\frac{n-1}{n}S^2$ 是有偏估计。但在平方损失$L(\theta, a) = (a-\sigma^2)^2$ 下，$\frac{n-1}{n}S^2$ 一致地优于S^2。

因为

$$T = \sum_{i=1}^{n} (X_i - \overline{X})^2/\sigma^2 \sim \chi^2(n-1),$$

故

$$E_\theta T = n-1, \quad \mathrm{Var}_\theta T = 2(n-1).$$

所以

$$E_\theta S^2 = \sigma^2,\ R(\theta, S^2) = E_\theta(S^2-\sigma^2)^2 = \mathrm{Var}_\theta(\frac{\sigma^2}{n-1}T) = \frac{2\sigma^4}{n-1}.$$

而

$$R(\theta, \frac{n-1}{n}S^2) = (E_\theta\frac{n-1}{n}S^2-\sigma^2)^2 + \mathrm{Var}_\theta\big(\frac{n-1}{n}S^2\big) = \frac{2n-1}{n^2}\sigma^4.$$

因此，当$n \geq 2$ 时，$R(\theta, \frac{n-1}{n}S^2) < R(\theta, S^2), \forall\theta$。

进一步可证，$S_*^2 = \frac{1}{n+1}\sum_{i=1}^n (X_i-\overline{X})^2$ 在估计类$\{c\sum_{i=1}^n (X_i-\overline{X})^2:\ c>0\}$ 中一致最优。

定义2.2.1 (UMVUE, 教材p334, Def7.3.7) 若δ^* 是可估参数$\gamma(\theta)$ 的U.E.，且对$\gamma(\theta)$ 的任一U.E. δ 满足

$$R(\theta, \delta^*) \leq R(\theta, \delta), \quad \forall\theta\in\Theta,$$

则称δ^* 是$\gamma(\theta)$ 的*一致最小风险无偏估计*。特别地，若$\gamma(\theta)$ 是一维的，$L(\theta, a) = [\gamma(\theta)-a]^2$，则一致最小风险无偏估计就是*一致最小方差风险无偏估计*(UMVUE)。

在凸损失情况下，无偏估计类中往往存在最优估计，而且借助于定理2.2.1等方法往往能找到一致最小风险无偏估计。但若损失函数非凸，那么即使限于无偏估计类，一般也不存在最优估计。

定理2.2.2 (Basu,1955) *设损失函数$L(\theta, a) \leq M < \infty$，且$L(\theta, \gamma(\theta)) = 0, \forall\theta\in\Theta$；$\gamma(\theta)$ 可估。则对$\forall\theta_0\in\Theta$，存在U.E.序列$\{\delta_n\}$，$R(\theta_0, \delta_n) \to 0 (n\to\infty)$。*

证明. 设$\delta(X)$ 是$\gamma(\theta)$ 的U.E.，$\pi\in(0,1)$。记

$$\delta_\pi(X) = \begin{cases} \gamma(\theta_0), & \text{以概率}1-\pi, \\ \frac{1}{\pi}[\delta(x)-\gamma(\theta_0)]+\gamma(\theta_0), & \text{以概率}\pi. \end{cases}$$

则$E_\theta[\delta_\pi(X)] = \gamma(\theta), \forall\theta\in\Theta$，即$\delta_\pi(X)$ 也是$\gamma(\theta)$ 的无偏估计。而

$$R(\theta_0, \delta_\pi) = (1-\pi)\cdot 0 + \pi E_{\theta_0} L\big(\theta_0, \frac{1}{\pi}[\delta(x)-\gamma(\theta_0)]+\gamma(\theta_0)\big) \leq \pi\cdot M,$$

因此当$\pi\to 0$ 时，$R(\theta_0, \delta_\pi) \to 0$。 □

注： 若损失函数有界，任何非常数参数都无一致最小风险无偏估计，甚至无局部最小风险无偏估计。

§2.2.2 存在充分完备统计量情况下的UMVUE

引理2.2.1 设X 的模型为$(\mathscr{X}, \mathscr{B}_X, \mathscr{P} = \{P_\theta, \theta \in \Theta\})$，样本空间是欧氏的，$T$ 为$\mathscr{P}$ 的充分完备统计量，则每一个可估参数 $\gamma(\theta)$ 有且仅有一个非随机化的无偏估计，它是 T 的函数。

证明. 存在性。设δ 是可估参数$\gamma(\theta)$ 的U.E.，取$\eta(t) = E(\delta|T = t)$，则$\eta(T)$ 是T 的函数，由T 的充分性知，它是$\gamma(\theta)$ 的U.E.。

唯一性。若δ_1, δ_2 都是T 的函数且都是$\gamma(\theta)$ 的非随机化的U.E.，则$f(T) = \delta_1 - \delta_2$ 满足

$$E_\theta f(T) = 0, \quad \forall \theta \in \Theta.$$

由T 的完备性可知$f(T) = 0$ a.s.$\mathscr{P}$，即$\delta_1 = \delta_2$ a.s. $\mathscr{P}$。 □

由定理2.2.1 及引理2.2.1 即得下列定理。

定理2.2.3 (Lehmann-Scheffé, 教材p347, Th7.3.23) 设X 的模型为$(\mathscr{X}, \mathscr{B}_X, \mathscr{P} = \{P_\theta, \theta \in \Theta\})$，样本空间是欧氏的，$T$ 为$\mathscr{P}$ 的充分完备统计量。

(1) 若损失函数$L(\theta, a)$ 关于a 凸，则任一可估函数$\gamma(\theta)$，都存在唯一一个非随机化的*U.E.*，它是 T 的函数，且是一致最小风险无偏估计。

(2) 若损失函数严凸，且上述估计量的风险有限，则此估计是唯一一个一致最小风险无偏估计。特别地，若$\gamma(\theta)$为一维参数，对于平方损失，该估计是唯一的*UMVUE*。

注： 在较宽松的条件下，充分完备统计量的存在性是非常数可估函数存在UMVUE的充要条件。

推论2.2.2 若$\mathscr{P}$ 为s 维满秩的指数族，$T = (T_1, \ldots, T_s)$ 是充分完备统计量，则定理 *2.2.3*的结论成立。

求解UMVUE的方法。设T 为充分完备统计量。

- 若δ 是T 的函数，又是$\gamma(\theta)$ 的U.E.，则δ 是$\gamma(\theta)$ 的UMVUE;
- 若δ 是$\gamma(\theta)$ 的U.E.，则$E(\delta|T)$ 是$\gamma(\theta)$ 的UMVUE。

例2.2.3 (直接求UMVUE) 设$X_1,\cdots,X_n$ i.i.d. $B(1,p)$，$p\in(0,1)$。求$g(p)=p(1-p)$ 的UMVUE。显然，$T=\sum_{i=1}^n X_i$ 充分完备，且$T\sim B(n,p)$。设$\delta(T)$ 是$g(p)$ 的U.E.，则$\delta(T)$ 应满足

$$\sum_{t=0}^{n}\delta(t)\binom{n}{t}p^t(1-p)^{n-t}=p(1-p),\quad \forall p\in(0,1) \tag{2.2.1}$$

取$\rho=p/(1-p)$，则$p=\frac{\rho}{1+\rho}$，$1-p=\frac{1}{1+\rho}$，(2.2.1) 可改写为

$$\sum_{t=0}^{n}\delta(t)\binom{n}{t}\rho^t=\rho(1+\rho)^{n-2}=\sum_{t=1}^{n-1}\binom{n-2}{t-1}\rho^t.$$

比较等式两边可知$\delta(t)=\frac{t(n-t)}{n(n-1)},t=0,\ldots,n$，即$\delta(T)$ 是$g(p)$ 的UMVUE。

例2.2.4 (由求$E(\delta|T)$得UMVUE。) 设$X_1,\ldots,X_n$ i.i.d. $U(0,\theta)$, $\theta>0$，求$\gamma(\theta)=\theta/2$ 的UMVUE。前面我们已经证明过，$T=X_{(n)}$ 是此分布族的充分完备统计量。易知，X_1 是$\gamma(\theta)$的U.E.。给定$T=t$ 条件下，X_1 的条件分布为

$$P(X_1\le x|T=t)=\frac{n-1}{n}\frac{x}{t}I_{(0,t)}(x)+I_{[t,\infty)}(x),$$

因此，$E(X_1|T=t)=\frac{t}{n}+\frac{n-1}{n}\cdot\frac{t}{2}=\frac{n+1}{2n}t$。所以$\frac{n+1}{2n}X_{(n)}$ 是$\gamma(\theta)$ 的UMVUE，$\frac{n+1}{n}X_{(n)}$ 是θ 的UMVUE。

UMVUE数例

例2.2.5 (正态单样本问题) 设$X_1,\ldots,X_n$ i.i.d. $N(\xi,\sigma^2)$，其联合密度函数为

$$\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\xi)^2\right\},\quad (\xi,\sigma)\in\mathbb{R}\times\mathbb{R}^+.$$

一、估计ξ,σ 的函数

单参数情形

1. σ 已知，ξ 未知，估计ξ 的函数。可证$\bar{X}$ 是充分完备统计量，因此
   - $\overline{X}$ 是ξ 的UMVUE;
   - 一般地，若$\gamma(\xi)$ 是ξ 的可估函数，则在$\gamma(\xi)$ 的U.E.中有且仅有一个$\delta(\overline{X})$ 是$\overline{X}$ 的函数，且它是$\gamma(\xi)$ 的UMVUE;（是否是唯一的UMVUE？）

- 若$\gamma(\xi)$ 是ξ 的r 次多项式，则$\delta(\overline{X})$ 也是$\overline{X}$ 的r 次多项式。（可用归纳法证明）

2. ξ已知，σ未知，估计σ 的函数。可证$B^2 = \sum_{i=1}^n (X_i - \xi)^2$ 是充分完备统计量，$B^2/\sigma^2 \sim \chi^2(n)$。

   记$\kappa_{n,r} = [E(B/\sigma)^r]^{-1}$，它是$\chi^2(n)$ 分布的$\frac{r}{2}$ 阶矩的倒数，即

$$\frac{1}{\kappa_{n,r}} = \int_0^\infty x^{\frac{r}{2}} \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} dx = 2^{\frac{r}{2}} \Gamma(\frac{n+r}{2})/\Gamma(\frac{n}{2}).$$

   当$r > -n$ 时，$\kappa_{n,r}$存在。

   显然$\kappa_{n,r}B^r$ 是σ^r 的U.E.，因此它是σ^r 的UMVUE。具体地，若$r = 2$，$\kappa_{n,2} = \frac{1}{n}$，$B^2/n$ 是σ^2 的UMVUE。

双参数情形，即ξ, σ 都未知。可证 $(\overline{X}, Q^2)$ 是充分完备统计量，且$\overline{X}$ 与Q^2 独立，$Q^2/\sigma^2 \sim \chi^2(n-1)$，其中$Q^2 = \sum_{i=1}^n (X_i - \overline{X})^2$。因此，

- $\overline{X}$ 仍是ξ 的UMVUE；
- 当$r > -n+1$时，$\kappa_{n-1,r}Q^r$ 是σ^r 的UMVUE（$r = 2$ 时给出$\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$ 的U.E.）；
- 因为$\overline{X}$ 是ξ 的U.E.，$\kappa_{n-1,-1}Q^{-1}$ 是$\frac{1}{\sigma}$ 的U.E.($n > 2$ 时)，且$\overline{X}$ 与Q^2 相互独立，故$\kappa_{n-1,-1}\overline{X}/Q$ 是总体信噪比 $\gamma(\xi, \sigma) = \xi/\sigma$ 的UMVUE。

二、估计总体分位点、分布函数值

记$N(\xi, \sigma^2)$ 的分布函数为$F(\cdot)$，$N(0,1)$ 的分布函数为$\Phi(\cdot)$，两者的关系为

$$F(u) = P(X_1 \le u) = \Phi(\frac{u-\xi}{\sigma}).$$

1. 给定p，估计总体的p 分位点 $u_p = F^{-1}(p)$。

   显然，$u_p = \Phi^{-1}(p)\sigma + \xi$。因此双参数情形下，$u_p$ 的UMVUE为

$$\Phi^{-1}(p)\kappa_{n-1,1}Q + \overline{X}.$$

2. 给定u，估计总体的分布函数值 $F(u)$。

分两种情形讨论。

(1) σ 已知，不妨设$\sigma=1$。取$\delta(X)=I(X_1\le u)$，它显然是$F(u)$ 的U.E.，求它关于充分完备统计量$\overline{X}$ 的条件期望

$$\begin{aligned}E(\delta(X)|\overline{X}=\overline{x})&=P(X_1\le u|\overline{X}=\overline{x})\\&=P(X_1-\overline{X}\le u-\overline{x}|\overline{X}=\overline{x}).\end{aligned}$$

因为 $X_1-\overline{X}\sim N(0,\frac{n-1}{n})$，它的分布与$\xi$ 无关，是附属统计量，故由Basu定理知，$X_1-\overline{X}$ 与$\overline{X}$ 独立，$\forall\xi$。于是有

$$E(\delta(X)|\overline{X}=\overline{x})=P(X_1-\overline{X}\le u-\overline{x})=\Phi\left[\sqrt{\frac{n}{n-1}}(u-\overline{x})\right].$$

因此，$\hat{F}(u)=\Phi\left[\sqrt{\frac{n}{n-1}}(u-\overline{X})\right]$ 是$F(u)$ 的UMVUE。

(2) (ξ,σ) 都未知的情形。

此时，$\delta(X)=I(X_1\le u)$ 仍然是$F(u)$ 的U.E.。$T=(\overline{X},Q^2)$ 是充分完备统计量。求δ 关于T 的条件期望

$$\begin{aligned}E(\delta(X)|\overline{X}=\overline{x},Q^2=q^2)&=P(X_1\le u|\overline{X}=\overline{x},Q^2=q^2)\\&=P\left(\frac{X_1-\overline{X}}{Q}\le\frac{u-\overline{x}}{q}\Big|\overline{X}=\overline{x},Q^2=q^2\right).\end{aligned}$$

因为 $\frac{X_1-\overline{X}}{Q}$ 是附属统计量，由Basu定理知，它与$(\overline{X},Q^2)$ 独立。因此，

$$E(\delta(X)|\overline{X}=\overline{x},Q^2=q^2)=P\left(\frac{X_1-\overline{X}}{Q}\le\frac{u-\overline{x}}{q}\right).$$

可计算得$Z=\frac{X_1-\overline{X}}{Q}$ 的pdf为

$$f(z)=\frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{n-2}{2})}\sqrt{\frac{n}{n-1}}\left(1-\frac{n}{n-1}z^2\right)^{\frac{n}{2}-2}I_{[0,\sqrt{\frac{n-1}{n}}]}(|z|).$$

于是，

$$E(\delta(X)|\overline{X}=\overline{x},Q^2=q^2)=\int_{-\infty}^{\frac{u-\overline{x}}{q}}f(z)dz.$$

因此，

$$\hat{F}(u)=\int_{-\infty}^{\frac{u-\overline{X}}{Q}}f(z)dz$$

是$F(u)$ 的UMVUE。

三、估计u 处总体的密度函数值

1. $\sigma=1$ 的情形。总体的密度函数在u 处的值为 $p_{X_1}(u;\xi)=\phi(u-\xi)$，其中$\phi(\cdot)$ 是$N(0,1)$ 的密度函数。那么 $\frac{d\hat{F}(u)}{du}$ 是否是它的UMVUE？

   显然，$\frac{d\hat{F}(u)}{du}$ 是充分完备统计量$\overline{X}$ 的函数，只要

   $$E_\xi\Big(\frac{d\hat{F}(u)}{du}\Big)=\frac{dF(u)}{du}=p_{X_1}(u;\xi),\ \forall\xi,$$

   那么它就是总体密度函数的UMVUE。

   这可以直接验证。因为$\hat{F}(u)$ 是$X_1|\overline{X}$ 的条件分布函数，故$\frac{d\hat{F}(u)}{du}=p_{X_1|\overline{X}}(u)$ 为$X_1|\overline{X}$ 的条件密度函数。而X_1 的边际密度与该条件密度有如下关系

   $$\begin{aligned}p_{X_1}(u;\xi)&=\int_{-\infty}^{\infty}p_{X_1|\overline{X}}(u)\cdot p_{\overline{X}}(\overline{x};\xi)d\overline{x}\\&=E_\xi(p_{X_1|\overline{X}}(u))=E_\xi\Big(\frac{d\hat{F}(u)}{du}\Big),\quad\forall\xi,\end{aligned}$$

   因此，$\frac{d\hat{F}(u)}{du}=\sqrt{\frac{n}{n-1}}\phi\Big(\sqrt{\frac{n}{n-1}}(u-\overline{X})\Big)$ 是$p_{X_1}(u;\xi)$ 的UMVUE。

2. (ξ,σ) 均未知的情形。同理可证

   $$\frac{d\hat{F}(u)}{du}=\frac{d}{du}\int_{-\infty}^{\frac{u-\overline{X}}{Q}}f(z)dz=\frac{1}{Q}f\Big(\frac{u-\overline{X}}{Q}\Big)$$

   是 $p_{X_1}(u;\xi,\sigma)=\frac{1}{\sigma}\phi\Big(\frac{u-\xi}{\sigma}\Big)$ 的UMVUE。

例2.2.6 (正态两样本问题) 设$X_1,\ldots,X_m$ i.i.d. $N(\xi_1,\sigma_1^2)$，$Y_1,\ldots,Y_n$ i.i.d. $N(\xi_2,\sigma_2^2)$，且全样本独立。

一、$(\xi_1,\sigma_1^2,\xi_2,\sigma_2^2)$**均未知**。

易知，样本的联合分布族为指数族，$T=(\overline{X},\overline{Y},Q_X^2=\sum_{i=1}^m(X_i-\overline{X})^2,Q_Y^2)$ 是充分完备统计量。可证：

- ξ_1,σ_1^r 的UMVUE：$\overline{X},\kappa_{m-1,r}Q_X^r$；
- ξ_2,σ_2^r 的UMVUE：$\overline{Y},\kappa_{n-1,r}Q_Y^r$；
- $\xi_1-\xi_2$ 的UMVUE：$\overline{X}-\overline{Y}$；

- σ_1^r/σ_2^r 的UMVUE：$\kappa_{m-1,r}Q_X^r \cdot \kappa_{n-1,-r}/Q_Y^r \quad (1-m<r<n-1)$。

二、$(\xi_1, \xi_2, \sigma \hat{=} \sigma_1 = \sigma_2)$ 未知。

易知，$T = (\overline{X}, \overline{Y}, Q^2 = Q_X^2 + Q_Y^2)$ 是充分完备统计量。因此，$\xi_1, \xi_2, \sigma^r, \xi_1 - \xi_2, \frac{\xi_1-\xi_2}{\sigma}$ 的U.E.若是T 的函数，那么都是UMVUE。

三、$\xi \hat{=} \xi_1 = \xi_2$，但$\sigma_1, \sigma_2$ 不一定相同，欲估计ξ。

可证，统计量$T = (\overline{X}, \overline{Y}, Q_X^2, Q_Y^2)$ 最小充分，但不完备。

- 若已知$\sigma_1^2/\sigma_2^2 = \gamma$，则$(\sum_{i=1}^m X_i^2 + \gamma \sum_{i=1}^m Y_i^2, \sum_{i=1}^m X_i + \gamma \sum_{i=1}^m Y_i)$ 是充分完备统计量，常用$\overline{X}, \overline{Y}$ 的最佳线性组合δ_γ 估计共同的均值ξ
$$\delta_\gamma = \alpha \overline{X} + (1-\alpha)\overline{Y}, \quad 其中\alpha = \frac{\sigma_2^2}{n} / (\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}).$$
显然，此时δ_γ 是ξ 的UMVUE。
- 若 σ_1^2/σ_2^2 未知，则 ξ 的UMVUE 不存在，而且ξ 的任何可估函数都没有UMVUE。可先估计α，再由上述线性组合估计 ξ。

四、$p = P(X_1 < Y_1)$ 的估计。

可由
$$P(X_1 < Y_1 | \overline{X}, \overline{Y}, Q_X^2, Q_Y^2), \quad P(X_1 < Y_1 | \overline{X}, \overline{Y}, Q^2)$$
分别构造一、二两种情形下 p 的UMVUE。

例2.2.7 (多元正态单样本问题) 。

一、二元正态

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix}, \ i = 1, \ldots, n \ \text{ i.i.d. } N\left(\begin{pmatrix} \xi \\ \eta \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma\tau \\ \rho\sigma\tau & \tau^2 \end{pmatrix} \right).$$

在$|\rho| < 1$时，(X_i, Y_i)，$i = 1, \ldots, n$ 的jpdf 形成一个五参数的满秩指数族，$T = (\overline{X}, \overline{Y}, Q_X^2, Q_Y^2, Q_{XY} = \sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y}))$ 是充分完备统计量。

- ξ, η, σ^r, τ^r 的UMVUE： $\overline{X}$, $\overline{Y}$, $\kappa_{n-1,r}Q_X^r$, $\kappa_{n-1,r}Q_Y^r$；

- $\rho\sigma\tau$ 的UMVUE： $Q_{XY}/(n-1)$；

- ρ 的UMVUE。$R = Q_{XY}/\sqrt{Q_X^2 Q_Y^2}$ 是ρ 的常用估计，但有偏。可证

$$E(R) = \rho\Big[1 - \frac{1-\rho^2}{2n} + O(\frac{1}{n^2})\Big].$$

Olkin & Pratt(1958)直接用T 的函数构造ρ 的U.E.，同时利用Laplace 变换的一些结果，导出了ρ 的UMVUE

$$\begin{aligned} G(R) &= R \cdot F(\frac{1}{2}, \frac{1}{2}; \frac{n-1}{2}; 1-R^2) \\ &\approx R\big[1 + \frac{1-R^2}{2(n-1)} + O(\frac{1}{n^2})\big], \end{aligned}$$

其中

$$F(a,b;c;x) = \sum_{k=0}^{\infty} \frac{\Gamma(a+k)\Gamma(b+k)\Gamma(c)x^k}{\Gamma(a)\Gamma(b)\Gamma(c+k)k!} = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 \frac{t^{b-1}(1-t)^{c-b-1}}{(1-tx)^a} dt.$$

二、多元正态

设 $X_\nu = (X_{\nu 1}, \ldots, X_{\nu p})', \nu = 1, \ldots, n$ i.i.d. $N_p(\xi, \Sigma)$。在$\Sigma > 0$ 时，$X_\nu, \nu = 1, \ldots, n$ 的jpdf 形成一个$p + \frac{(p+1)p}{2}$ 个参数的满秩指数族。易知，$T = (\overline{X}, Q = \sum_{i=1}^n (X_\nu - \overline{X})(X_\nu - \overline{X})')$ 是充分完备统计量。可以考虑 $\xi, \Sigma, \rho_{ij} = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$ 等参数的UMVUE。

例2.2.8 (指数分布单样本问题) 设$X_1, \ldots, X_n$ i.i.d. $E(a,b)$，总体的pdf 为

$$f(x;a,b) = \frac{1}{b} e^{-\frac{x-a}{b}} I_{(a,\infty)}(x).$$

$a = 0$时，也即通常的指数分布，期望为b。

1. b 已知，$X_{(1)}$充分、完备。故$X_{(1)} - b/n$ 为a 的UMVUE。

2. a已知，b未知，$\sum_{i=1}^n (X_i - a)$ 充分、完备。则$\sum_{i=1}^n (X_i - a)/n$ 是b 的UMVUE。

3. a,b 均未知，则$(X_{(1)},\sum_{i=1}^n(X_i-\overline{X}))$ 充分完备，相互独立，且

$$n[X_{(1)}-a]/b\sim E(0,1),\quad 2\sum_{i=1}^n(X_i-X_{(1)})/b\sim\chi^2(2(n-1)).$$

因此，

$$\frac{1}{n-1}\sum_{i=1}^n(X_i-X_{(1)}),\quad X_{(1)}-\frac{1}{n(n-1)}\sum_{i=1}^n(X_i-X_{(1)})$$

分别是b,a 的UMVUE。

4. 还可讨论a/b, $P(X_1\le u)$, u_p 等参数的UMVUE，两样本问题等。

例2.2.9 (二项分布) 设$X_1,\dots,X_n$ i.i.d. $B(1,p)$, $p\in(0,1)$，则统计量$T=\sum_{i=1}^n X_i$ 充分、完备。

- p的UMVUE，T/n；
- pq 的UMVUE：$\frac{T(n-T)}{n(n-1)}$；
- p^m 的UMVUE：$\frac{T(T-1)\cdots(T-m+1)}{n(n-1)\cdots(n-m+1)}$, $(m\le n)$；
- 若$\gamma(p)$ 是p 的$m(\le n)$ 次多项式，则可估，UMVUE可由各p^r 的UMVUE获得，$r=0,1,\dots,m$；
- 但$\frac{1}{p}$，$\frac{1}{1+p^2}$ 等参数不可估。

例2.2.10 (负二项分布) 二项分布中，$\frac{1}{p}$ 不可估。采用逆抽样方法能构造$\frac{1}{p}$ 的无偏估计。逆抽样：独立地依次进行Bernoulli试验，直到成功m 次为止。记总的试验次数为$Y+m$，每次试验的成功概率为p。则Y 服从负二项分布(NB(m,p))，分布列为

$$P(Y=y)=\binom{y+m-1}{m-1}p^m(1-p)^y,\quad y=0,1,2,\dots$$

可证

$$EY=\frac{m(1-p)}{p},\quad \mathrm{Var}Y=\frac{m(1-p)}{p^2},$$

$\delta(Y)=\frac{Y+m}{m}$ 是$1/p$ 的U.E.。又Y 是此分布族的充分完备统计量，因此$\delta(Y)$ 是$\frac{1}{p}$ 的UMVUE。

例2.2.11 (幂级数分布) 由离散型单参数指数族

$$\{P(Y=y)=A(\eta)e^{\eta\cdot T(y)}\cdot h(y):\ \eta\in\tilde{\Theta}\}$$

可获得一大类离散型分布族。记$\theta=e^{\eta}$，$x=T(y)$，$c(\theta)=1/A(\eta)$，则统计量T的分布为：

$$P(T=x)=a(x)\cdot\theta^{x}/c(\theta),\ x=0,1,2,\ldots,$$

其中$\theta>0,\ \theta\in\Theta\subset\mathbb{R}$。只要 $\{a(x),x=0,1,\ldots\}$ 非负，且满足 $\sum_{x=0}^{\infty}a(x)\theta^{x}<\infty,\ \forall\theta\in\Theta$，则 $\{P_{\theta}(T=x):\theta\in\Theta\}$ 就形成一族幂级数分布。

特例：二项分布族$B(n,p)$；负二项分布族$NB(m,p)$；Poisson分布族等。

若$a(x)>0,x=0,1,\ldots$，则对$\forall r\in\mathbb{N}$，θ^{r} 可估，其UMVUE可由解下列方程获得

$$\sum_{x=0}^{\infty}\delta(x)a(x)\cdot\theta^{x}=\theta^{r}c(\theta),\ \forall\theta\in\Theta.$$

因为 $c(\theta)=\sum_{x=0}^{\infty}a(x)\theta^{x}$，比较上式左右两边$\theta^{x}$ 的系数可得

$$\delta(x)=\begin{cases}0, & \text{if } x=0,1,\ldots,r-1,\\ \frac{a(x-r)}{a(x)}, & \text{if } x\geq r.\end{cases}$$

设$X_1,\ldots,X_n$ i.i.d.，且

$$P(X_1=x)=a(x)\cdot\theta^{x}/c(\theta),\ x=0,1,2,\ldots,$$

其中$\theta>0,\theta\in\Theta$。则$T=\sum_{i=1}^{n}X_i$ 充分完备，且分布为

$$P(T=t)=\frac{A(t,n)\theta^{t}}{c^{n}(\theta)},\ t=0,1,\ldots,$$

其中$A(t,n)$ 是$c^{n}(\theta)$ 的幂级数展式中θ^{t} 项的系数。此时，θ^{r} 的UMVUE为

$$\delta(T)=\begin{cases}0, & T=0,1,\ldots,r-1,\\ \frac{A(T-r,n)}{A(T,n)}, & T\geq r.\end{cases}$$

对于给定的x，$\gamma(\theta)=P_{\theta}(X_1=x)$ 的UMVUE为

$$\delta(T)=\begin{cases}\frac{a(x)A(T-x,n-1)}{A(T,n)}, & 0\leq x\leq T,\\ 0, & x>T.\end{cases}$$

§2.2.3 UMVUE的充分必要条件

当充分完备统计量不存在的时候，求解UMVUE 往往是比较困难的。有些情况下，可根据这样一条直观思路去求：若已知δ_0 是$\gamma(\theta)$ 的某一个U.E.，那么$\gamma(\theta)$ 的任意一个U.E.可表为$\delta = \delta_0 - U$，其中 U 是0 的U.E.，即$E_\theta U(X) = 0,\ \forall\theta \in \Theta$。如果$\delta_0$ 的方差有限，那么从δ_0 出发找UMVUE，相当于在所有0 的U.E. 中找一个U 最小化$E_\theta(\delta_0 - U)^2$。这一思路的可行性跟模型假定下所有0 的U.E. 的结构有关。

例2.2.12 随机变量X 的分布列为

$$P(X=-1)=p, P(X=k)=q^2p^k, k=0,1,2,\ldots, p\in(0,1), q=1-p,$$

考虑 p 与q^2 的无偏估计。

显然，

$$\delta_1(X)=\begin{cases}1, & 若X=-1,\\ 0, & 其他.\end{cases} \qquad \delta_2(X)=\begin{cases}1, & 若X=0,\\ 0, & 其他.\end{cases}$$

分别为 p、q^2 的U.E.。若U 是0 的U.E.，即$E_pU(X)=0,\ \forall p\in(0,1)$，那么有

$$U(k)=-kU(-1)\hat{=}a\cdot k,\quad k=-1,0,1,\ldots$$

其中$a=-U(-1)$。找 p 与q^2 的U.E.使之在p_0 处方差最小，即简化为解

$$\min_a E_{p_0}[\delta_i(X)-aX]^2,\quad i=1,2.$$

因为

$$E_{p_0}[\delta_i(X)-aX]^2=\sum_{k=-1}^{\infty}P_{p_0}(X=k)[\delta_i(k)-ak]^2,$$

因此这两个问题的解分别为 $a_1^*=-\frac{1-p_0}{2}$, $a_2^*=0$。这就求得在p_0 处使方差最小的U.E.分别为

$$\begin{array}{ll}\delta_1^*=\delta_1-a_1^*X, & (依赖于p_0, 说明p不存在\text{UMVUE.})\\ \delta_2^*=\delta_2, & (不依赖于p_0, 说明这是q^2的\text{UMVUE.})\end{array}$$

虽然p 无UMVUE，但是处处存在局部最小方差无偏估计(LMVUE)，δ_1^* 即为p 在p_0 处的LMVUE。

定理2.2.4 (教材p344, Th7.3.20) 设X 的模型为$(\mathscr{X},\mathscr{B}_X,\mathscr{P}=\{P_\theta,\theta\in\Theta\})$。$\gamma(\theta)$ 是一维实函数，δ 是$\gamma(\theta)$ 的*U.E.*，且$Var_\theta\delta(X)<\infty$, $\forall\theta\in\Theta$。记$\mathcal{U}=\big\{U: E_\theta U(X)=0,\forall\theta\in\Theta\big\}$.

(i) δ 为$\gamma(\theta)$ 的*UMVUE*的充要条件是：对于$\theta\in\Theta$，若$U\in\mathcal{U}$ 满足$Var_\theta U(X)<\infty$，则有$E_\theta[\delta(X)U(X)]=0$.

(ii) 设T 是充分统计量，$\delta=h(T)$，$h(\cdot)$ 是一个*Borel* 函数。记$\mathcal{U}_T$ 是$\mathcal{U}$ 的子集，其中的元素都是T 的*Borel* 函数。那么δ 为$\gamma(\theta)$ 的*UMVUE*的充要条件是：对于$\theta\in\Theta$，若$U\in\mathcal{U}_T$ 满足$Var_\theta U(X)<\infty$，则有$E_\theta[\delta(X)U(X)]=0$.

证明. (i) 充分性。设δ' 也是$\gamma(\theta)$ 的U.E.，则$\delta'-\delta\in\mathcal{U}$。对于$\theta\in\Theta$，若$\mathrm{Var}_\theta\delta'=\infty$，则$\mathrm{Var}_\theta\delta'>\mathrm{Var}_\theta\delta$。若$\mathrm{Var}_\theta\delta'<\infty$，则$\mathrm{Var}_\theta(\delta'-\delta)<\infty$。由条件知

$$\mathrm{Var}_\theta\delta'=\mathrm{Var}_\theta(\delta'-\delta)+\mathrm{Var}_\theta(\delta)+2\mathrm{Cov}_\theta(\delta,\delta'-\delta)\geq\mathrm{Var}_\theta(\delta).$$

因此，δ 是$\gamma(\theta)$的UMVUE。

必要性。设$U\in\mathcal{U}$，且在θ 处 $\mathrm{Var}_\theta U(X)<\infty$。则对$\forall c\in\mathbb{R},\delta_c=\delta+cU$ 都是$\gamma(\theta)$ 的U.E.，且$\mathrm{Var}_\theta\delta_c<\infty$。因$\delta$ 为$\gamma(\theta)$ 的UMVUE，故$\mathrm{Var}_\theta\delta\leq\mathrm{Var}_\theta\delta_c$, $\forall c\in\mathbb{R}$。这等价于

$$\mathrm{Var}_\theta\delta\leq\mathrm{Var}_\theta\delta+c^2\mathrm{Var}_\theta U+2c\mathrm{Cov}_\theta(\delta,U),\ \forall c\in\mathbb{R}.$$

这就意味着$\mathrm{Cov}^2_\theta(\delta,U)\leq 0$，即$E_\theta[\delta(X)U(X)]=0$。

(ii) 只须证$E_\theta[\delta(X)U(X)]=0,\forall U\in\mathcal{U}_T$ 意味着$E_\theta[\delta(X)U(X)]=0,\forall U\in\mathcal{U}$。

□

注：(1) 若U.E. δ 满足对θ_0，$\mathrm{Var}_{\theta_0}(\delta)<\infty$，且对$U\in\mathcal{U}$，若$\mathrm{Var}_{\theta_0}U(X)<\infty$，则有$\mathrm{Cov}_{\theta_0}(\delta,U)=0$。那么$\delta$ 是$\gamma(\theta)$ 在θ_0 处的LMVUE。

(2) 在一定条件下，定理 2.2.4可推广至一般的凸损失。

例2.2.12 （续）用定理 2.2.4给出全部的UMVUE。由定理 2.2.4知，若δ 是其期望的UMVUE，则δ 满足$E_p(\delta\cdot U)=0,\forall p\in(0,1)$，即$\delta\cdot U$ 还是0的无偏估计。由此可知δ 应满足$k\cdot\delta(k)=k\cdot\delta(-1),k=0,1,\ldots$，即$\delta(k)=\delta(-1),\quad k=1,2,\ldots$，其中$\delta(0),\delta(-1)$ 不定。不妨设$a=\delta(-1),b=\delta(0)$，那么估计量δ 的一、二阶矩为

$$\gamma(p)=E_p(\delta)=bq^2+a(1-q^2),\quad E_p(\delta^2)=b^2q^2+a^2(1-q^2).$$

只要$a, b \neq \pm\infty$，总有$\mathrm{Var}_p\delta < \infty$。因此，所有存在方差有限UMVUE的$\gamma(p)$ 必具有形式$a + c \cdot q^2, a, c \neq \pm\infty$。

例2.2.13 (p167/Example 3.7) 设$X_1, \cdots, X_n$ i.i.d. $U(0, \theta), \theta \in [1, \infty)$。此时，$X_{(n)}$ 充分但不完备。我们可以根据定理 2.2.4找到θ 的UMVUE。

注：(1) UMVUE存在与否的四种情形：

Case 1 $\gamma(\theta) = const$，UMVUE 必存在且方差恒为0。

Case 2 任意非常数函数$\gamma(\theta)$ 都无UMVUE。

例2.2.14 (p168/Example3.8) 设$X_1, \ldots, X_n$ i.i.d.，若其共同的分布为

$$(i) P(X_1 = \theta - 1) = P(X_1 = \theta) = P(X_1 = \theta + 1) = \frac{1}{3},\ \theta \in \mathbb{R} \text{ 或}$$
$$(ii) X_1 \sim U(\theta - \frac{1}{2}, \theta + \frac{1}{2}),\ \theta \in \mathbb{R},$$

即属这种情况。

Case 3 有些非常数函数$\gamma(\theta)$ 有UMVUE，有些没有。

Case 4 所有可估的$\gamma(\theta)$ 都有UMVUE。

(2) MLE与UMVUE的比较。设$X_1, \ldots, X_n$ i.i.d. $N(\xi, \sigma^2)$。

1. 若ξ, σ^2 未知，$\gamma(\xi, \sigma) = \xi$，则MLE与UMVUE一致，都为$\overline{X}$。
2. 若$\sigma = 1$，$\gamma(\xi) = \Phi(u - \xi)$，$u$ 给定。则
   MLE：$\Phi(u - \overline{X})$，UMVUE：$\Phi\left[\sqrt{\frac{n}{n-1}}(u - \overline{X})\right]$。
   从风险角度比较，若$R(\xi, \delta) = E(\gamma(\xi) - \delta)^2$，Zack & Even(1996)发现两者各有优势。例：$n = 4$，$|u - \xi| > 1.3$时，UMVUE好；反之，MLE好。
3. 若ξ, σ^2 未知，$\gamma(\xi, \sigma^2) = \sigma^2$，则
   MLE：$\sum_{i=1}^n (X_i - \overline{X})^2/n$，UMVUE：$\sum_{i=1}^n (X_i - \overline{X})^2/(n-1)$。MLE 一致优于UMVUE。
4. 若σ^2已知，$\gamma(\xi) = \xi^2$，则UMVUE $\overline{X}^2 - \frac{\sigma^2}{n}$ 一致优于MLE $\overline{X}^2$。

(3) UMVUE的不良表现

设$X\sim$ Poisson(θ)，$\gamma(\theta)=e^{-a\theta}$，其中$a$ 给定。则可证 $\gamma(\theta)$ 的UMVUE为$\delta(X)=(1-a)^X$。若$a=3$，则δ 正负交替，表现怪异。但当样本量增大时，可消除这种不良表现。设$X_1,\ldots,X_n$ i.i.d. Poisson(θ)，则$\delta(T)=(1-\frac{a}{n})^T$ 为$\gamma(\theta)$ 的UMVUE，其中$T=\sum_{i=1}^n X_i$。当$n>a$ 时，$\delta(T)$ 合理。

推论2.2.3 *(i)* 若δ_j 是未知参数$\gamma_j(\theta)$ 的*UMVUE*，且$Var_\theta(\delta_j)<\infty$, $\forall\theta\in\Theta$，$j=1,\cdots,k$，其中k 是一个给定的正整数。则对任意常数$c_1,\cdots,c_k\in\mathbb{R}$，$\sum_{j=1}^k c_j\delta_j$ 是$\sum_{j=1}^k c_j\gamma_j(\theta)$ 的*UMVUE*。

(ii) 设δ_1,δ_2 是$\gamma(\theta)$ 的两个*UMVUE*，且$Var_\theta(\delta_j)<\infty$, $\forall\theta\in\Theta$，$j=1,2$，则$\delta_1=\delta_2$ $a.s.\mathscr{P}$。

§2.2.4 U-统计量

前面我们讨论了UMVUE 的一般理论，考察了诸如正态、指数、均匀、二项等一些常用参数分布族下UMVUE 的例子。本小节讨论采用U-统计量解决一些非参数分布族下的UMVUE 问题。U-统计量是Hoeffding (1948) 提出的，他证明了U-统计量的渐近性质。这类统计量在参数估计、假设检验方面都有应用。

一、次序统计量的充分完备性

由Lehmann-Scheffé 定理（定理2.2.3）知，若分布族$\mathscr{P}$ 有充分完备统计量T，则由T 的函数构造的可估参数的无偏估计即为该参数的UMVUE。Shao (2003) p111/Example 2.17 与p166/Example 3.6 说明，对于有些非参数分布族，次序统计量是充分完备统计量；若估计量δ 是次序统计量的函数、且方差处处有限，则它即是其自身期望的UMVUE。这就涉及两个问题：1)对于哪些非参数分布族，次序统计量是充分完备的？2)怎样的统计量才是次序统计量的函数？

先讨论问题1)。设$\mathscr{P} = \{P\}$ 是任一给定的定义在$(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ 上的分布族，$X_1, \cdots, X_n$ i.i.d. $P \in \mathscr{P}$，$(X_{(1)}, \cdots, X_{(n)})$ 是次序统计量。设$\varphi(x_1, \cdots, x_n)$ 是任一$P^n (\forall P \in \mathscr{P})$ 可积的函数，那么（参见成平等(1985)p36/例5.1 或讲义第二章例2.5），

$$\begin{aligned}&E_P\left[\varphi(X_1,\cdots,X_n)|X_{(1)}=x_{(1)},\cdots,X_{(n)}=x_{(n)}\right]\\=&\frac{1}{n!}\sum_{(i_1,\cdots,i_n)\in\mathscr{S}_n}\varphi(x_{i_1},\cdots,x_{i_n}),\end{aligned}$$

与P 无关，其中$\mathscr{S}_n$ 为$(1, \cdots, n)$ 的所有排列的集合。这说明次序统计量为分布族$\{P^n : P \in \mathscr{P}\}$ 的充分统计量。

次序统计量的完备性对$\mathscr{P}$ 有一定的要求。Halmos (1946), Lehmann (1950), Fraser (1954), 陈希孺等人分别研究了次序统计量的完备性。主要结论有（参见Fraser (1954), Non-parametric theory: scale and location parameters, Canadian Journal of Mathematics, 6(1),46–68. 及陈希孺(1981) p83/定理1.6.2），若定义在$(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ 上的分布族$\mathscr{P}$

(1) (Lehmann) 是关于Lebesgue 测度具有如下形式pdf 的分布族，

$$p(x) = C(\theta_1, \cdots, \theta_n) \exp(-x^{2n} + \sum_{j=1}^{n} \theta_j x^j),$$

其中$(\theta_1, \cdots, \theta_n) \in \mathbb{R}^n, C(\theta_1, \cdots, \theta_n)$ 是正则化常数；

(2) 是关于Lebesgue 测度具有pdf 的分布族；

(3) (Halmos) 是在有限个点上离散型分布的分布族；

(4) (Fraser) 是在有限个区间上均匀分布的分布族；

(5) (陈希孺) 是满足条件(i) $\mathscr{P}$ 是凸集，即$\forall P, Q \in \mathscr{P}$，$\forall p \in [0,1]$，$pP+(1-p)Q \in \mathscr{P}$，(ii) 若$P \in \mathscr{P}, -\infty < a < b < \infty$，且$P([a,b)) > 0$，则$P_{[a,b)} \in \mathscr{P}$，其中$P_{[a,b)}$ 定义为$P_{[a,b)}(A) = P(A \cap [a,b))/P([a,b)), \forall A \in \mathscr{B}_{\mathbb{R}}$；

(6) 是(1)～(5) 之一的一个子族，其中各分布的$r(>0)$ 阶矩都存在。

则次序统计量对于分布族$\{P^n : P \in \mathscr{P}\}$ 是完备的。

但是对于有些分布族，次序统计量并不完备。如$\mathscr{P}$ 为具有Lebesgue pdf 及有限期望的对称分布族(Shao (2003),p166/Example 3.6)，或者$\mathscr{P}$ 为具有Lebesgue pdf 及满足$P((-\infty,0]) = p, \forall P \in \mathscr{P}$ 的分布族，其中$p \in (0,1)$为给定常数（Fraser (1954)），则次序统计量对于分布族$\{P^n : P \in \mathscr{P}\}$ 不完备。

二、U-统计量

问题2) 的答案比较直观。若统计量δ 是次序统计量$(X_{(1)},\cdots,X_{(n)})$ 的函数，则它必是n 元的对称函数，即满足

$$\delta(x_{i_1},\cdots,x_{i_n}) = \delta(x_1,\cdots,x_n),\ \forall(i_1,\cdots,i_n) \in \mathscr{S}_n, \quad \text{a.s.}P^n, \forall P \in \mathscr{P}.$$

设$\gamma(P)$ 是一个可估参数，$\varphi(X_1,\cdots,X_m)$ 是它的一个无偏估计，其中$m \le n$。易知

$$h(X_1,\cdots,X_m) = \frac{1}{m!} \sum_{(i_1,\cdots,i_m)\in\mathscr{S}_m} \varphi(X_{i_1},\cdots,X_{i_m}) \tag{2.2.2}$$

也是$\gamma(P)$ 的无偏估计，且是一个m 元的对称函数。

定义2.2.2 (U-统计量) 设$h(X_1,\cdots,X_m)$ 是一个统计量，且关于其$m(\le n)$ 个自变量对称，则称下列n 元的对称函数

$$U_n(X_1,\cdots,X_n) = \frac{1}{\binom{n}{m}} \sum_{\{i_1,\cdots,i_m\}\in\mathscr{C}_n^m} h(X_{i_1},\cdots,X_{i_m})$$

为具有对称核 h 的U-统计量，其中$\mathscr{C}_n^m$ 表示从$\{1,2,\cdots,n\}$ 中选m 个互不相同元素的所有$\binom{n}{m}$ 种组合形成的集合。

显然，U_n 是一个n 元的对称函数，因而是次序统计量的函数。同时，若h 是$\gamma(P)$ 的无偏估计则U_n 也是。如果次序统计量$(X_{(1)},\cdots,X_{(n)})$ 是分布族$\{P^n: P\in\mathscr{P}\}$ 的充分完备统计量，那么它就是$\gamma(P)$ 的UMVUE。从另一角度看，

$$U_n = E_P\left[h(X_1,\cdots,X_m)|X_{(1)},\cdots,X_{(n)}\right].$$

这就给出了一条由U-统计量构造可估参数UMVUE 的方法。

对于一般的未知参数$\gamma(P)$，在构造其UMVUE 前有两个问题需要考虑：(1) 是否可估？(2) 使其可估的最小样本容量$m=$？这两个问题的讨论参见Lehmann *et al* (1998)。使$\gamma(P)$ 可估的最小样本容量m 称为阶。下面考察一些U-统计量的例子。

例2.2.15 设$\mathscr{P}=\{P\}$ 是定义在$(\mathbb{R},\mathscr{B}_{\mathbb{R}})$ 上的分布族，$X_1,\cdots,X_n$ i.i.d. $P\in\mathscr{P}$。

(1) 估计分布函数。设$\gamma(P)=\int_{(-\infty,a]}dP(x)$，其中$a$ 给定。取$h(X_1)=I_{(-\infty,a]}(X_1)$，它是$\gamma(P)$ 的无偏估计，相应的U-统计量为$U_n=\frac{1}{n}\sum_{i=1}^n I_{(-\infty,a]}(X_i)$，正好是经验分布函数在$a$ 处的值。

(2) 估计r 阶原点矩。假定$\mathscr{P}$ 中每个分布的$r(>0)$ 阶矩都存在。设$\gamma(P)=\int_{\mathbb{R}}x^r dP(x)$。取$h(X_1)=X_1^r$，它是$\gamma(P)$ 的无偏估计，相应的U-统计量为$U_n=\frac{1}{n}\sum_{i=1}^n X_i^r$，就是$r$ 阶样本原点矩。

(3) 估计方差。假定$\mathscr{P}$ 中每个分布的二阶矩都存在。设$\gamma(P)=\mathrm{Var}_P(X_1)=\int_{\mathbb{R}}(x-EX_1)^2dP(x)$。易看出，该参数的阶为2，$\varphi(X_1,X_2)=X_1^2-X_1X_2$ 是其无偏估计，但不对称。由(2.2.2)可得对称核

$$h(X_1,X_2)=\frac{1}{2}\left[\varphi(X_1,X_2)+\varphi(X_2,X_1)\right]=\frac{1}{2}(X_1-X_2)^2.$$

相应的U-统计量为

$$U_n=\frac{1}{n(n-1)}\sum_{1\le i<j\le n}(X_i-X_j)^2=\frac{1}{n-1}\sum_{i=1}^n(X_i-\overline{X})^2,$$

即通常的样本方差。

(4) 估计Gini 均差。假定$\mathscr{P}$ 中每个分布的一阶矩都存在。设$\gamma(P)=E_P|X_1-X_2|$，该参数度量的是分布P 的离散度。该参数的阶也是2，$h(X_1,X_2)=|X_1-X_2|$ 是其无偏估计，且对称，故相应的U-统计量为

$$U_n=\frac{2}{n(n-1)}\sum_{1\le i<j\le n}|X_i-X_j|.$$

如果次序统计量$(X_{(1)},\cdots,X_{(n)})$ 是分布族$\{P^n: P\in\mathscr{P}\}$ 的充分完备统计量，那么这些例子中的U-统计量就是相应参数的UMVUE。

除了用于参数估计，U-统计量也可用于假设检验。设欲检验$H_0: P\in\mathscr{P}_0$ v.s. $H_1: P\in\mathscr{P}-\mathscr{P}_0$，其中$\mathscr{P}_0$ 是$\mathscr{P}$ 的一个子集。考虑找一个适当的参数$\gamma(P)$，使之满足

$$\gamma(P)\begin{cases}\equiv\gamma_0, & \text{if } P\in\mathscr{P}_0,\\ >(\text{or }<,\neq)\gamma_0, & \text{else.}\end{cases}$$

然后由其无偏估计为核构造U-统计量U_n，用$U_n-\gamma_0$ 作为检验统计量构造检验。

例2.2.16 (Wilcoxon统计量) 设$X_1,\cdots,X_n$ i.i.d. $P\in\mathscr{P}$，

$$\begin{aligned}\mathscr{P}=\{P: P \text{ 的分布函数} F(x)=G(x-\xi),\\ G\text{是关于原点对称的一元连续型分布函数}, \xi\in\mathbb{R}\}.\end{aligned}$$

欲检验$H_0:\xi=0$ v.s. $H_1:\xi\neq 0$。为此，令

$$\begin{aligned}\gamma(P) &= P(X_1+X_2\le 0)=\int_{\mathbb{R}}P(X_1+X_2\le 0|X_1=x)dP(x)\\ &= \int_{\mathbb{R}}F(-x)dF(x)=\int_{\mathbb{R}}G(-x-\xi)dG(x-\xi)\\ &= 1-\int_{\mathbb{R}}G(x+2\xi)dG(x)\quad\begin{cases}<1/2, & \text{if } \xi>0,\\ =1/2, & \text{if } \xi=0,\\ >1/2, & \text{if } \xi<0.\end{cases}\end{aligned}$$

显然，$h(X_1,X_2)=I_{(-\infty,0]}(X_1+X_2)$ 是$\gamma(P)$ 的一个对称的无偏估计，以此为核所得的U-统计量为

$$U_n=\frac{2}{n(n-1)}\sum_{1\le i<j\le n}I_{(-\infty,0]}(X_i+X_j).$$

可以证明，

$$U_n=1-\binom{n}{2}^{-1}\left[\sum_{i=1}^n I_{(0,\infty)}(X_i)R_i^+-r\right],$$

其中$(R_1^+,\cdots,R_n^+)$ 是$|X_1|,\cdots,|X_n|$ 对应的秩向量，$W_n^+=\sum_{i=1}^n I_{(0,\infty)}(X_i)R_i^+$ 为Wilcoxon 符合秩统计量，$r=\sum_{i=1}^n I_{(0,\infty)}(X_i)$。相应地，可取拒绝域为

$$\left\{(x_1,\cdots,x_n):\binom{n}{2}^{-1}\left[\sum_{i=1}^n I_{(0,\infty)}(X_i)R_i^+-r\right]-\frac{1}{2}\le C_1 \text{ or } \ge C_2\right\},$$

其中C_1,C_2 须根据显著性水平及H_0 真时U_n 的抽样分布来定。

三、U-统计量的方差

定理2.2.5 (Hoeffding) 设$X_1, \cdots, X_n$ *i.i.d.* $P \in \mathscr{P}$，U_n 是由m 元对称核h 构造的U-统计量，$E_P[h(X_1, \cdots, X_m)]^2 < \infty$。则

$$Var_P(U_n) = \binom{n}{m}^{-1} \sum_{k=1}^{m} \binom{m}{k} \binom{n-m}{m-k} \zeta_k, \tag{2.2.3}$$

其中，对于$k = 1, \cdots, m$，$\zeta_k = Var_P[h_k(X_1, \cdots, X_k)]$，而

$$\begin{aligned} h_k(x_1, \cdots, x_k) &= E_P[h(X_1, \cdots, X_m) | X_1 = x_1, \cdots, X_k = x_k] \\ &= E_P[h(x_1, \cdots, x_k, X_{k+1}, \cdots, X_m)]. \end{aligned}$$

证明. 记$\gamma(P) = E_P h(X_1, \cdots, X_m)$。显然，$E_P(U_n) = \gamma(P)$。为求$U_n$ 的方差，先考察h_k 的期望与方差。由其定义易知，$h_m = h$，且

$$h_k(x_1, \cdots, x_k) = E_P[h_{k+1}(x_1, \cdots, x_k, X_{k+1})].$$

对$k = 1, \cdots, m$，根据定理条件有

$$\begin{aligned} &E_P h_k(X_1, \cdots, X_k) = \gamma(P), \\ &\zeta_k = \mathrm{Var}_P h_k(X_1, \cdots, X_k) \leq \zeta_m < \infty. \end{aligned}$$

由U_n 的定义知，

$$\mathrm{Var}_P(U_n) = \binom{n}{m}^{-2} \sum_{\{i_1, \cdots, i_m\} \in \mathscr{C}_n^m} \sum_{\{j_1, \cdots, j_m\} \in \mathscr{C}_n^m} \mathrm{Cov}_P[h(X_{i_1}, \cdots, X_{i_m}), h(X_{j_1}, \cdots, X_{j_m})].$$

若$\{i_1, \cdots, i_m\} \cap \{j_1, \cdots, j_m\} = \varnothing$，则

$$\mathrm{Cov}_P[h(X_{i_1}, \cdots, X_{i_m}), h(X_{j_1}, \cdots, X_{j_m})] = 0.$$

若$\{i_1, \cdots, i_m\} \cap \{j_1, \cdots, j_m\} = \{c_1, \cdots, c_k\}$，$k = 1, \cdots, m$，记$\{i_1, \cdots, i_m\} - \{c_1, \cdots, c_k\} = \{i'_{k+1}, \cdots, i'_m\}$，$\{j_1, \cdots, j_m\} - \{c_1, \cdots, c_k\} = \{j'_{k+1}, \cdots, j'_m\}$ 则

$$\begin{aligned} & \mathrm{Cov}_P[h(X_{i_1}, \cdots, X_{i_m}), h(X_{j_1}, \cdots, X_{j_m})] \\ = \ & E_P E_P \Big\{ [h(X_{c_1}, \cdots, X_{c_k}, X_{i'_{k+1}}, \cdots, X_{i'_m}) - \gamma(P)] \\ & \cdot [h(X_{c_1}, \cdots, X_{c_k}, X_{j'_{k+1}}, \cdots, X_{j'_m}) - \gamma(P)] \Big| X_{c_1}, \cdots, X_{c_k} \Big\} \\ = \ & E_P[h_k(X_{c_1}, \cdots, X_{c_k}) - \gamma(P)]^2 \\ = \ & \zeta_k. \end{aligned}$$

而$\{i_1, \cdots, i_m\}$ 与$\{j_1, \cdots, j_m\}$ 恰好有k 个共同元素的情况共$\binom{n}{m}\binom{m}{k}\binom{n-m}{m-k}$ 种。因此，(2.2.3) 得证。 □

推论2.2.4 在定理*2.2.5* 的条件下，有

(i) $\frac{m^2}{n}\zeta_1 \le Var_P(U_n) \le \frac{m}{n}\zeta_m$；

(ii) $(n+1)\,Var_P(U_{n+1}) \le n\,Var_P(U_n),\ \forall n > m$；

(iii) 对于$k = 1, \cdots, m$，若$\zeta_j = 0$，当$j < k$，而$\zeta_k > 0$，则当$n \to \infty$ 时

$$Var_P(U_n) = \frac{k!\binom{m}{k}^2}{n^k}\zeta_k + O(n^{-(k+1)}).$$

(iv) $\zeta_1 \le \zeta_2 \le \cdots \le \zeta_k$。

该推论说明当$n \to \infty$ 时，$U_n \to_{L_2} \gamma(P)$；若(iii)中的条件成立，则有

$$n^{k/2}[U_n - \gamma(P)] = O_p(1).$$

例2.2.17 设$\mathscr{P} = \{P\}$ 是定义在$(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ 上的分布族，$X_1, \cdots, X_n$ i.i.d. $P \in \mathscr{P}$。

(1) 假定$\mathscr{P}$ 中每个分布的二阶矩都存在。记$\mu = E_P X_1, \sigma^2 = \mathrm{Var}_P X_1$，考虑参数$\gamma(P) = \mu^2$ 的估计。显然$h(X_1, X_2) = X_1 X_2$ 是$\gamma(P)$ 的无偏估计，且是对称函数，由其构造的U-统计量为

$$U_n = \binom{n}{2}^{-1} \sum_{1\le i<j\le n} X_i X_j.$$

因为

$$\begin{aligned} h_1(x_1) = x_1 E_P X_2 = \mu x_1, \quad \zeta_1 = \mathrm{Var}_P h_1(X_1) = \mu^2\sigma^2, \\ \zeta_2 = \mathrm{Var}_P(X_1X_2) = E_P(X_1X_2)^2 - [E_P(X_1X_2)]^2 = \sigma^4 + 2\mu^2\sigma^2, \end{aligned}$$

故

$$\begin{aligned} \mathrm{Var}_P(U_n) &= \binom{n}{2}^{-1} \sum_{k=1}^{2} \binom{2}{k}\binom{n-2}{2-k}\zeta_k \\ &= \binom{n}{2}^{-1} [2(n-2)\zeta_1 + \zeta_2] = \frac{4\mu^2\sigma^2}{n} + \frac{2\sigma^4}{n(n-1)}. \end{aligned}$$

与$P \in \mathscr{P}' = \{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2\text{已知}\}$ 情况下$\gamma(P)$ 的UMVUE $\delta = \overline{X}^2 - \frac{\sigma^2}{n}$ 相比，

$$\mathrm{Var}_P(U_n) - \mathrm{Var}_P(\delta) = \frac{2\sigma^4}{n^2(n-1)}.$$

(2) 续例2.2.16。易知，

$$h_1(x_1) = E_P[h(X_1, X_2)|X_1 = x_1] = F(-x_1), \quad \zeta_2 = \gamma(P)[1 - \gamma(P)].$$

故

$$\mathrm{Var}_P(U_n) = \binom{n}{2}^{-1}[2(n-2)\zeta_1 + \gamma(P)(1 - \gamma(P))].$$

当$\xi = 0$ 时，因$F(-X_1) = 1 - F(X_1) \sim U(0, 1)$，故$\zeta_1 = \frac{1}{12}$.

(3) 续例2.2.15(4)。

$$h_1(x_1) = E_P|x_1 - X_2| = \int_{\mathbb{R}} |x_1 - y| dP(y),$$

$$\zeta_1 = \mathrm{Var}_P h_1(X_1) = \int_{\mathbb{R}} \left[\int_{\mathbb{R}} |x - y| dP(y)\right]^2 dP(x) - \gamma^2(P).$$

四、其他场合下的U-统计量

多个独立样本的情况。

多元非参数分布族。

样本独立不同分布、不独立等情况。