AI 中的数学

第十八讲

方聪，概率统计部分参考章复熹和张原老师课件

2024 年秋季

1. 无偏估计的优良性
2. 充分统计量
3. UMVUE

1. 无偏估计的优良性
2. 充分统计量
3. UMVUE

设 $X_1, \cdots, X_n \sim$iid $F_\theta(x)(\theta \in \Theta)$ 为统计模型，$g(\theta)$ 为待估量，$g(\theta)$ 的估计量 $T(X_1, \cdots, X_n)$ 的均方误差定义为

$$R(\theta, T) = E_\theta[T(X_1, \cdots, X_n) - g(\theta)]^2.$$

估计的均方误差有时也称为风险函数。

例：设 $X_1, \cdots, X_n \sim$iid $N(\mu, \sigma^2)$，$\mu \in (+\infty, -\infty)$，$\sigma^2 > 0$，在这个统计模型中，$\mu$ 的 ML 估计为 $\hat{\mu} = \bar{X}$，其均方误差或风险函数为

$$E(\bar{X} - \mu)^2 = \mathrm{var}(\bar{X}) = \mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$
$$= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{var}(X_i) = \frac{\sigma^2}{n}.$$

它是不随 μ 的变化而变化的。

例：设 $X_1, \cdots, X_n \sim \text{iid} B(1, p)$，即 $X_i (i = 1, \cdots, n)$ 以概率 p 取 1，以概率 $1 - p$ 取 0。该统计模型在实际应用中是很常见的，我们已经求得参数 p 的 ML 估计为 $\hat{p} = \bar{X}$，不难验证 $\hat{p}$ 也是 p 的无偏估计，其均方误差为

$$E(\bar{X} - p)^2 = \mathrm{var}(\bar{X}) = \frac{1}{n}\mathrm{var}(X_1) = \frac{p(1-p)}{n}.$$

由此可知，估计 $\bar{X}$ 的风险函数为 $R(p, \bar{X}) = \frac{p(1-p)}{n}$。

注意：对于任何一个待估参数 $g(\theta)$，可以定义估计 $g(\theta_0)$，它将保证在 θ_0 处，风险函数最小，这意味着一个估计如果要成为最优估计，它的风险函数必须处处为 0。

但是这是不可能的，为此我们需要考虑限制估计类。考虑无偏估计类：设 $T(X)$ 是 $g(\theta)$ 的无偏估计，其均方误差变成方差：

$$R(\theta, T) = E_\theta[T - g(\theta)]^2 = E_\theta[T - E_\theta(T)]^2 = \mathrm{var}_\theta(T).$$

定义：设 $X_1, \cdots, X_n \sim$iid $F_\theta(x)(\theta \in \Theta)$ 为统计模型，$g(\theta)$ 为待估量，$g(\theta)$ 的一个估计量为 $T(X_1, \cdots, X_n)$ 如果
(1) T 是 $g(\theta)$ 的无偏估计,
(2) 对于 $g(\theta)$ 的任意无偏估计 $\tilde{T} = \tilde{T}(X_1, \cdots, X_n)$, 都有

$$\mathrm{var}_\theta(T) \leqslant \mathrm{var}_\theta(\tilde{T}), \quad \forall \theta \in \Theta.$$

则称 T 为 $g(\theta)$ 的 (一致) 最小方差无偏估计 (Uniformly Minimum Variance Unbiased, UMVU).

寻求（优化）一个映射 T，要求对每一个 θ 方差最小

记

$$F_\theta(t) = P_\theta(T \leqslant t)$$

为统计量 T 的分布。

定义：假设统计量 $T = T(X_1, \cdots, X_n)$ 满足:
对任意统计量 $\tilde{T} = \tilde{T}(X_1, \cdots, X_n)$ 都有,

$$\text{在}\, T = t \text{ 的条件下，} \tilde{T} \text{ 的分布与参数} \theta \text{ 无关,}$$

那么，称 T 为充分统计量。

例：（对应郑书例 4.3）设总体：$X \sim B(1, p)$，样本量：n. 考虑 $T = X_1 + \cdots + X_n$. 易知 T 的分布为二项分布，现在讨论 T 的充分性。

例：（对应郑书例 4.3）设总体：$X \sim B(1, p)$，样本量：n. 考虑 $T = X_1 + \cdots + X_n$. 易知 T 的分布为二项分布，现在讨论 T 的充分性。

对 $t = 0, 1, \cdots, n$，令

$$S_t := \left\{(x_1, \cdots, x_n) : x_i \in \{0, 1\}, \forall i;\ 且 x_1 + \cdots + x_n = t\right\}.$$

那么，$\forall\,(x_1, \cdots, x_n) \in S_t$ ，

$$\begin{aligned}
& P_p\left(X_1 = x_1, \cdots, X_n = x_n \mid T = t\right) \\
= & \frac{P_p\left(X_1 = x_1, \cdots, X_n = x_n\right)}{P_p(T = t)} = \frac{p^t(1-p)^{n-t}}{C_n^t p^t (1-p)^{n-t}} = \frac{1}{C_n^t}.
\end{aligned}$$

因此 $\forall t$，在 $T = t$ 的条件下，$(X_1, \cdots, X_n)$ 服从 S_t 上的均匀分布. 该分布与 p 无关. 因此，T 是充分统计量.

定理：（因子分解定理）设 $X_1, \cdots, X_n \sim \text{iid} p(x, \theta)$，其中 $p(x, \theta)$ 为分布密度或分布列。若 $T = T(X_1, \cdots, X_n)$ 满足:

$$\prod_{i=1}^{n} p(x_i, \theta) = q_\theta(T(x_1, \cdots, x_n)) h(x_1, \cdots, x_n)$$

则 T 是充分统计量。

例：设总体：$X \sim N\left(\mu, \sigma^2\right)$，样本量：$n$.
联合密度为

$$
\begin{aligned}
\prod_{i=1}^{n} p\left(x_i, \theta\right) &= \frac{1}{\sqrt{2\pi\sigma^2}^{n}} \exp\left\{-\frac{\sum_{i=1}^{n}\left(x_i-\mu\right)^2}{2\sigma^2}\right\} \\
&= \frac{1}{\sqrt{2\pi\sigma^2}^{n}} \exp\left\{-\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2+n(\bar{x}-\mu)^2}{2\sigma^2}\right\}.
\end{aligned}
$$

指数上的部分可以写为 $\left(\bar{X}, \sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\right)$ 的函数。因此，$\left(\bar{X}, \sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\right)$ 是充分统计量.

若总体改为 $X \sim N(\mu, 1)$.

若总体改为 $X \sim N(\mu, 1)$.

$$\prod_{i=1}^{n} p\left(x_i, \theta\right) = \frac{1}{\sqrt{2\pi}^{n}} \exp\left\{-\sum_{i=1}^{n} \frac{\left(x_i-\mu\right)^2}{2}\right\}$$

$$= \frac{1}{\sqrt{2\pi}^{n}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right\} \exp\left\{\mu \sum_{i=1}^{n} x_i\right\} \exp\left\{-\frac{n}{2}\mu^2\right\}.$$

因此, $\bar{X}$ 是充分统计量.

例：假设二维随机向量 (X, Y) 服从二元正态分布，参数 $\theta=(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，从总体中抽取一个样本 $((X_1,Y_1),\cdots,(X_n,Y_n))$，样本量为 n，其联合密度为

$$\prod_{i=1}^{n} p\left(x_i, y_i; \theta\right)=\left(\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\right)^n \cdot \exp$$

$$\left\{-\frac{1}{2\left(1-\rho^2\right)}\sum_{i=1}^{n}\left[\left(\frac{x_i-\mu_1}{\sigma_1}\right)^2-\frac{2\rho(x_i-\mu_1)(y_i-\mu_2)}{\sigma_1\sigma_2}+\left(\frac{y_i-\mu_2}{\sigma_2}\right)^2\right]\right\},$$

$x_i-\mu_1=x_i-\bar{x}+\bar{x}-\mu_1$ 和 $y_i-\mu_2=y_i-\bar{y}+\bar{y}-\mu_2$ 代入上式 e 指数上的表达式中，化简，表达式可以写为

$$\left(\bar{x},\bar{y},\sum_{i=1}^{n}(x_i-\bar{x})^2,\sum_{i=1}^{n}(y_i-\bar{y})^2,\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right)$$

的函数.

因此联合密度具有形式

$$\prod_{i=1}^{n} p\left(x_i, y_i; \theta\right) = q_\theta\left(\bar{x}, \bar{y}, \sum_{i=1}^{n}(x_i - \bar{x})^2, \sum_{i=1}^{n}(y_i - \bar{y})^2, \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})\right)$$

由因子分解定理知

$$T = \left(\bar{X}, \bar{Y}, \sum_{i=1}^{n}(X_i - \bar{X})^2, \sum_{i=1}^{n}(Y_i - \bar{Y})^2, \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})\right)$$

例：设总体：$X \sim \mathrm{Exp}(\lambda)$，样本量：$n$.
联合密度具有形式

$$p_\lambda\left(x_1, \cdots, x_n\right) = \lambda^n \exp\left\{-\lambda \sum_{i=1}^n x_i\right\} \cdot 1_{\{x_1, \cdots, x_n > 0\}}.$$

其中 $1_{\{x_1, \cdots, x_n > 0\}}$ 是示性函数，，当所有 x_i 大于 0 时取 1，否则为 0。显然该函数与参数无关，因此由因子分解定理有，令 $T = T\left(X_1, \cdots, X_n\right) = X_1 + \cdots + X_n$. 则 T 是充分统计量.

注意充分统计量不唯一，没有起到数据压缩的作用。

定义：设 $X_1, \cdots, X_n \sim \text{iid} p(x, \theta)$，又设 $T = T(X_1, \cdots, X_n)$ 为充分统计量. 若对任意 ϕ,

$$E_\theta \phi(T) = 0, \forall \theta \in \Theta \text{ 可推出 } P_\theta(\phi(T) = 0) = 1, \forall \theta \in \Theta.$$

则称 T 为完全充分统计量.

例：设总体 $X \sim N(\theta, 1)$，样本量为 n。由因子分解定理知，$T_1 = \sum_{i=1}^{n} X_i$ 是一个充分统计量，$T_2 = (T_1, X_1 - X_2)$ 也是充分统计量。取 $\phi(T_2) = X_1 - X_2$，易知

$$E_\theta[\phi(T_2)] = E_\theta(X_1) - E_\theta(X_2) \equiv 0$$

但是 $P(\phi(T_2) = 0) = P(X_1 = X_2) \neq 1$，这说明 $T_2 = (T_1, X_1 - X_2)$ 不是完全充分统计量。

定理：设 $T = T(X_1, \cdots, X_n)$ 为完全充分统计量. 若

$$E_\theta(\phi(T)) = g(\theta), \forall\theta,$$

则 $\phi(T)$ 是 $g(\theta)$ 的 UMVU 估计. 该定理说明，对于待估量 $g(\theta)$，

只要找到依赖于完全充分统计量的函数 $\phi(T)$，使得 $\phi(T)$ 是 $g(\theta)$ 的无偏估计，则 $\phi(T)$ 就是 $g(\theta)$ 的 UMVU 估计。因此，要找到 $g(\theta)$ 的 UMVU 估计，只需在完全充分统计量中寻找即可。

例：设总体 $\xi \sim U(0,\theta)$，已经证明样本量为 n 的样本的最大值为 θ 的充分统计量，记为 $\xi_{(n)}$，证明 $\xi_{(n)}$ 也为 θ 的完全充分统计量。

证明：因为 $\xi_{(n)}$ 的密度为

$$f_{\xi_{(n)}}(x;\theta) = nf(x;\theta)[F(x;\theta)]^{n-1} = \begin{cases} \frac{nx^{n-1}}{\theta^n}, & 0 \leqslant x \leqslant \theta, \\ 0, & \text{其他}, \end{cases}$$

如果有函数 $g(x)$，使得对一切 $0 < \theta$，有

$$E_\theta[g(\xi_{(n)})] = 0,$$

即

$$0 \equiv \int_0^\theta g(x) f_{\xi_{(n)}}(x;\theta)dx = \frac{n}{\theta^n}\int_0^\theta g(x)x^{n-1}dx,$$

故

$$\int_0^\theta g(x)x^{n-1}dx \equiv 0.$$

对两侧求导得

$$g(\theta)\theta^{n-1} \equiv 0.$$

从而对一切 $\theta > 0$，$g(\theta) = 0$，因此 $\xi_{(n)}$ 是完全充分统计量。

1. 无偏估计的优良性
2. 充分统计量
3. UMVUE

定义：若密度或分布列 $p(x,\theta)$ 能进行如下分解:

$$p(x,\theta) = S(\theta)h(x)\exp\left\{\sum_{k=1}^{m} C_k(\theta)T_k(x)\right\}, \quad (\theta \in \Theta)$$

则称 $p(x,\theta), \theta \in \Theta$ 为指数族分布.

注: x 可为高维向量, 于是 $p(x,\theta)$ 为联合密度/联合分布列.

例：设总体：$X \sim N\left(\mu, \sigma^2\right)$，样本量：$n$，参数 $\theta = \left(\mu, \sigma^2\right)$，分布密度具有形式：

$$p(x,\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\mu^2}{2\sigma^2}} e^{\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2}.$$

显然，它具有指数分布的形式。

例：设总体：X 服从参数为 p 的二项分布，样本量：n.

例：设总体：X 服从参数为 p 的二项分布，样本量：n.

$$\begin{aligned} P(X=x) &= C_n^x p^x (1-p)^{n-x} \\ &= (1-p)^n C_n^x \exp\left\{ x \ln \frac{p}{1-p} \right\} \end{aligned}$$

$P(X=x)$ 所表示的三个因子的乘积符合指数族分布的要求。

例：设总体 $X \sim N(\mu, \sigma^2)$，可得 $(X_1, \cdots, X_n)$ 的联合分布密度为

$$p(x_1, \cdots, x_n; \mu, \sigma^2) = \prod_{i=1}^{n} p(x_i, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}^n} \exp\left\{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right\}$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}^n} \exp\left\{-\frac{n\mu^2}{2\sigma^2}\right\} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2 + \frac{\mu}{\sigma^2}\sum_{i=1}^{n} x_i\right\},.$$

故正态分布的一个样本的联合分布也是指数族分布。直接使用引理也可以得证。

定理：总体 X 具有指数族分布，$\Theta \in \mathbb{R}^m$ 且含内点；$(C_1, \cdots, C_m)$ 是在 Θ 上一对一、连续的函数；诸 C_i 之间（T_i 之间）无线性关系．则

$$\left(\sum_{i=1}^{n} T_1(X_i), \cdots, \sum_{i=1}^{n} T_k(X_i)\right)$$

是完全充分统计量.

例：设总体：$X \sim N\left(\mu, \sigma^2\right)$，样本量：$n$，参数 $\theta = \left(\mu, \sigma^2\right)$。
设 $T_1 = \sum_{i=1}^n X_i$, $T_2 = \sum_{i=1}^n X_i^2$，则由上面的定理知 (T_1, T_2) 是完全充分统计量.

$\bar{X}, S^2$ 是 μ, σ^2 的 UMVU 估计，其中：

$$\bar{X} = \frac{1}{n} T_1, \quad S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 = \frac{1}{n-1} \left(T_2 - \frac{1}{n} T_1^2\right)$$

是 (T_1, T_2) 的函数，并且是 μ, σ^2 的无偏估计.

改为已知 μ (例如, 已知 $\mu = 1$). 则 $\theta = \sigma^2, m = 1$:

$$p(x, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-1)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-1)^2}.$$

$T_1 = \sum_{i=1}^{n} (X_i - \mu)^2$ 是完全充分统计量.

$\hat{\sigma}^2$ 是 σ^2 的 UMVU 估计, 其中

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2.$$

例：某工人生产 20 件产品，其中 1 件为次品. 求：次品率的 UMVU 估计.

解：由题意，总体：$Y \sim B(1,p)$，参数 $p=\theta \in [0,1]$，样本量：$n=20$.
分布列：（记 $k=y_1+\cdots+y_n$）

$$P_p\left(Y_1=y_1,\cdots,Y_n=y_n\right)=\prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i}$$
$$=p^k(1-p)^{n-k}=e^{k\cdot\log p+(n-k)\log(1-p)}=e^{n\log(1-p)}e^{(\log p-\log(1-p))k}$$

可见 Y 的分布列具有指数族分布的形式. 因此，
$T_1=X=Y_1+\cdots+Y_{20}$ 是完全充分统计量.
又由于 $E(X)=nE(Y)=np$，即 $\frac{X}{n}$ 是 p 的无偏估计，因此，
$\hat{p}=X/20$ 是 UMVU 估计.

例：（对应郑书例 4.15）总体: $X \sim N(\mu, 1)$, 样本量: n, 求 μ^2 的 UMVU 估计.

例：（对应郑书例 4.15）总体: $X \sim N(\mu, 1)$, 样本量: n, 求 μ^2 的 UMVU 估计.

解：参数 $\theta = \mu$ ，联合密度为:

$$p(x, \theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\mu^2} e^{-\frac{1}{2}x^2} e^{\mu x}.$$

$T_1 = \sum_{i=1}^{n} X_i$ 是完全充分统计量, 因此 $\bar{X}$ 也是完全充分统计量. 由 $\mathrm{var}(Y) = E(Y^2) - (E(Y))^2$ 知,

$$\mu^2 = (E_\mu \bar{X})^2 = E_\mu \bar{X}^2 - \mathrm{var}_\mu(\bar{X}) = E_\mu \bar{X}^2 - \frac{1}{n} = E_\mu \left(\bar{X}^2 - \frac{1}{n} \right)$$

因此, $\bar{X}^2 - \frac{1}{n}$ 是 μ^2 的 UMVU 估计.

例：设总体 $Y \sim N(X\beta, \sigma^2 I_n)$，其中 $Y \in \mathbb{R}^n, X \in \mathbb{R}^{n\times p}(p \leqslant n), \beta \in \mathbb{R}^{p\times 1}$。$X$ 已知。(β, σ^2) 是参数。设 $\hat{\beta} = (X^\top X)^{-1}X^\top Y$，$\hat{\sigma}^2 = \frac{1}{n-p}(Y - X\hat{\beta})^\top (Y - X\hat{\beta})$，证明 $(\hat{\beta}, \hat{\sigma}^2)$ 为 (β, σ^2) 的 UMVU 估计。

例：设总体 $Y \sim N(X\beta, \sigma^2 I_n)$，其中 $Y \in \mathbb{R}^n, X \in \mathbb{R}^{n\times p}(p \leqslant n), \beta \in \mathbb{R}^{p\times 1}$。$X$ 已知。(β, σ^2) 是参数。设 $\hat{\beta} = (X^\top X)^{-1}X^\top Y$，$\hat{\sigma}^2 = \frac{1}{n-p}(Y - X\hat{\beta})^\top(Y - X\hat{\beta})$，证明 $(\hat{\beta}, \hat{\sigma}^2)$ 为 (β, σ^2) 的 UMVU 估计。

证明：$Y_1, \cdots, Y_n$ 的联合密度函数为

$$p(y_1, \cdots, y_n; \beta, \sigma^2) = (\sqrt{2\pi}\sigma)^{-n} \exp\left\{-\frac{1}{2\sigma^2}||Y - X\beta||_2^2\right\}$$
$$= Q(\theta)\exp\{\theta_1 T_1(Y) + \theta_2 T_2(Y)\}.$$

其中 $\theta_1 = -\frac{1}{2\sigma^2}, \theta_2 = \frac{\beta}{\sigma^2}$，$T_1(Y) = Y^\top Y$，$T_2(Y) = X^\top Y$，$Q(\theta) = (\sqrt{2\pi}\sigma)^{-n} \cdot \exp\{-\frac{1}{2\sigma^2}(X\beta)^\top(X\beta)\}$。
由指数分布族的性质知，$T_1(Y)$ 和 $T_2(Y)$ 为完全充分统计量.

而

$$\hat{\beta} = (X^\top X)^{-1} X^\top Y = (X^\top X)^{-1} T_2(Y)$$

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{n-p}(Y - X\hat{\beta})^\top (Y - X\hat{\beta}) \\ &= \frac{1}{n-p}(Y^\top Y - 2Y^\top X\hat{\beta} + \hat{\beta}X^\top X\hat{\beta}) \\ &\frac{1}{n-p}(T_1(Y) - T_2^\top(Y)(X^\top X)^{-1} T_2(Y))\end{aligned}$$

这表明 $(\hat{\beta}, \hat{\sigma}^2)$ 都是完全充分统计量的函数，又由于它们分别是 β 和 σ^2 的无偏估计，可知 $\hat{\beta}$ 和 $\hat{\sigma}^2$ 分别是 β 和 σ^2 的 UMVU 估计。