AI 中的数学

第十九讲

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. 估计的相合性
2. 估计的渐近分布

1. 估计的相合性
2. 估计的渐近分布

定义：设 $T_n = T_n(X_1, \cdots, X_n)$ 满足: $\forall \varepsilon > 0$,

$$P_\theta\left(|T_n - g(\theta)| \geqslant \varepsilon\right) \xrightarrow{n\to\infty} 0, \quad \forall \theta \in \Theta.$$

则称 T_n 为 $g(\theta)$ 的相合估计，或估计 T_n 具有相合性.

要求对所有 θ，依概率收敛

定理：设 $X_1, \cdots, X_n \sim$iid $F_\theta(x)(\theta \in \Theta)$，$E_\theta(X_i)$ 存在且有限，则

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} E_\theta(X_1) \quad (n \to \infty),$$

定理说明，在简单随机抽样的情况下，样本均值是总体均值的相合估计。

推论：设 $X_1, \cdots, X_n \sim$iid $F_\theta(x)(\theta \in \Theta)$，则 $\alpha_l = E_\theta(X_1^l)$ 的矩估计 $a_l = \frac{1}{n}\sum_{j=1}^{n} X_j^l$ 为 α_l 的相合估计。

定理：设 $X_1, \cdots, X_n \sim$iid $F_\theta(x)(\theta \in \Theta)$，则 θ 的函数 $g(\theta)$ 的矩估计具有相合性. (注: $g(\theta) = \phi(\alpha_1, \cdots, \alpha_k)$, 其中 ϕ 为连续函数).

隐藏使用了定理：

定理：若 $T_n \xrightarrow{P} \theta$，$f$ 为连续函数，则 $f(T_n) \xrightarrow{P} f(\theta)$.

由连续性：对于任意 $\epsilon > 0$, 存在 δ, 有 $|T_n - \theta| \leq \delta$, 则 $|f(T_n) - f(\theta)| \leq \epsilon$。故后者事件是前者的子集，有 $P(|T_n - \theta| \leq \delta) \geq P(|f(T_n) - f(\theta)| \leq \epsilon)$。

例： 设总体: $X \sim U(0,\theta)$, 样本量: n.

例：设总体: $X \sim U(0,\theta)$, 样本量: n.

由定理可以直接得到，参数 θ 的矩估计 $2\bar{X}$ 具有相合性.
最大似然估计 $T_n = \max_{1\leqslant i\leqslant n} X_i$ 也具有相合性:
$\forall 0<\varepsilon<\theta$,

$$\begin{aligned} P_\theta\left(|T_n-\theta| \geqslant \varepsilon\right) &= P_\theta\left(T_n \leqslant \theta-\varepsilon\right) \\ &= P_\theta(X \leqslant \theta-\varepsilon)^n = \left(\frac{\theta-\varepsilon}{\theta}\right)^n \to 0. \end{aligned}$$

例：（对应郑书例 5.3）考虑某物种三种类型的个体，以 $1,2,3$ 表示个体的三种类型，设此三种类型出现的概率分别为：

$$p(1,\theta)=\theta^2,\quad p(2,\theta)=2\theta(1-\theta),\quad p(3,\theta)=(1-\theta)^2,$$

其中 $0<\theta<1$, 现有 n 个个体的类型 $\mathbf{x}=(x_1,\cdots,x_n)$，设其中共有 n_1 个 1,n_2 个 2，n_3 个 3，求 ML 估计并探究相合性。

例：（对应郑书例 5.3）考虑某物种三种类型的个体，以 $1,2,3$ 表示个体的三种类型，设此三种类型出现的概率分别为：

$$p(1,\theta)=\theta^2,\quad p(2,\theta)=2\theta(1-\theta),\quad p(3,\theta)=(1-\theta)^2,$$

其中 $0<\theta<1$, 现有 n 个个体的类型 $\mathbf{x}=(x_1,\cdots,x_n)$，设其中共有 n_1 个 $1,n_2$ 个 2，n_3 个 3，求 ML 估计并探究相合性。

$\mathbf{x}$ 出现的概率为

$$\begin{aligned}P(X=\mathbf{x})&=p(1,\theta)^{n_1}p(2,\theta)^{n_2}p(3,\theta)^{n_3}\\&=\theta^{2n_1+n_2}(1-\theta)^{n_2+2n_3}2^{n_2}\\&=2^{n_2}\theta^{2n_1+n_2}(1-\theta)^{2n-(2n_1+n_2)}.\end{aligned}$$

这是对 θ 作 ML 估计的似然函数，与总样本量 $2n$，某事件出现 $2n_1 + n_2$ 次的二项分布的似然函数相同。利用二项分布的 ML 估计公式，可得 θ 的 ML 估计为

$$\hat{\theta}_n = \frac{2n_1 + n_2}{2n}.$$

为证明 $\hat{\theta}_n$ 的相合性，我们只需证明

$$\frac{n_1}{n} \xrightarrow{P} \theta^2, \quad \frac{n_2}{n} \xrightarrow{P} 2\theta(1-\theta)(n \to \infty).$$

事实上，利用大数定律可得 $\frac{n_1}{n} \xrightarrow{P} p(1,\theta) = \theta^2$，$\frac{n_2}{n} \xrightarrow{P} p(2,\theta) = 2\theta(1-\theta)$，将两个极限合并即可得到

$$\hat{\theta}_n = \frac{n_1}{n} + \frac{n_2}{2n} \xrightarrow{P} \theta^2 + \theta - \theta^2 = \theta$$

即 $\hat{\theta}_n$ 为 θ 的相合估计。

例：设总体 $\xi \sim U(0,\theta)$，$Y=(\prod_{i=1}^{n}\xi_i)^{\frac{1}{n}}$，证明 eY 是 θ 的相合估计。

例：设总体 $\xi \sim U(0,\theta)$，$Y=(\prod_{i=1}^n \xi_i)^{\frac{1}{n}}$，证明 eY 是 θ 的相合估计。

证明：需要证明 $eY \xrightarrow{P} \theta$，等价于 $Y \xrightarrow{P} \frac{\theta}{e}$。注意到

$$E(\ln\xi)=\int_0^\theta \ln x \frac{1}{\theta}dx=\frac{1}{\theta}x\ln x\Big|_0^\theta-\int_0^\theta \frac{1}{\theta}dx=\ln\theta-1.$$

因此，由独立同分布的大数定律得，

$$Y=\left(\prod_{i=1}^n \xi_i\right)^{\frac{1}{n}}=e^{\frac{1}{n}\sum_{i=1}^n \ln\xi_i} \xrightarrow{P} e^{E(\ln\xi)}=\frac{\theta}{e}$$

1 估计的相合性

2 估计的渐近分布

定义：设 $T_n = T_n(X_1, \cdots, X_n)$ 满足:

$$\sqrt{n}\left(T_n - g(\theta)\right) \xrightarrow{\omega} Z \sim N\left(0, \sigma^2\right), \quad \forall \theta \in \Theta,$$

则称 T_n 是渐近正态的，其中 $\sigma^2 = \sigma_\theta^2$ 称为渐近方差.

定理：若 $Y_n \xrightarrow{\omega} Y$，且随机变量序列 A_n 和 $B_n (n = 1, 2, \cdots)$ 分别依概率收敛于 a 和 b，则 $A_n + B_n Y_n \xrightarrow{\omega} a + bY$。

定义：设 $T_n = T_n(X_1, \cdots, X_n)$ 满足:

$$\sqrt{n}\left(T_n - g(\theta)\right) \xrightarrow{\omega} Z \sim N\left(0, \sigma^2\right), \quad \forall \theta \in \Theta,$$

则称 T_n 是渐近正态的，其中 $\sigma^2 = \sigma_\theta^2$ 称为渐近方差.

定理：若 $Y_n \xrightarrow{\omega} Y$，且随机变量序列 A_n 和 $B_n(n = 1, 2, \cdots)$ 分别依概率收敛于 a 和 b，则 $A_n + B_n Y_n \xrightarrow{\omega} a + bY$。

定理：（中心极限定理）设 $X_i(i = 1, \cdots, n)$ 是独立同分布的，且 $E(X_i) = \mu$，$\mathrm{var}(X_i) = \sigma^2 < +\infty$，那么 $\sqrt{n}(\bar{X} - \mu)$ 弱收敛到 $N(\mu, \sigma^2)$，因此 $\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$ 弱收敛到标准正态分布 $N(0, 1)$。

定理：(Δ 方法). 设 T_n 为 θ 的估计，
$\sqrt{n}\left(T_n-\theta\right) \xrightarrow{\omega} Z \sim N\left(0, \tau^2\right)$, $h'(\theta)$ 存在且不为 0 , 则

$$\sqrt{n}\left(h\left(T_n\right)-h(\theta)\right) \xrightarrow{\omega} W \sim N\left(0, h'(\theta)^2 \tau^2\right).$$

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

UMVU 估计：$\hat{\mu} = \bar{X}, \widehat{\sigma^2} = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$.

例：设总体：$X \sim N\left(\mu, \sigma^2\right)$，样本量：$n$.
UMVU 估计：$\hat{\mu} = \bar{X}, \widehat{\sigma^2} = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$.

$\hat{\mu}$ 渐近正态：事实上，

$$\sqrt{n}(\bar{X} - \mu) \sim N\left(0, \sigma^2\right).$$

例：（对应郑书例 6.2）设 $X_1, \cdots, X_n$ 是来自总体 X 的一个样本，总体 X 服从泊松分布，分布列为

$$P(X_i = k) = \frac{\lambda^k}{k!} \exp\{-\lambda\} \quad (k = 0, 1, 2, \cdots; \lambda > 0).$$

已经求得 λ 的 ML 估计为 $\hat{\lambda} = \bar{X}$，不难验证 λ 的矩估计也是 $\bar{X}$，利用中心极限定理，有

$$\sqrt{n}(\bar{X} - \lambda) = \sqrt{n}[\bar{X} - E(X)] \xrightarrow{\omega} N(0, \mathrm{var}(X)),$$

其中 $\mathrm{var}(X) = \lambda$，故 $\bar{X}$ 是渐近正态的，并且渐近方差为 λ。

现在试求 λ^2 的 ML 估计，并讨论其渐近分布。

现在试求 λ^2 的 ML 估计，并讨论其渐近分布。

因 λ 的 ML 估计为 $\hat{\lambda} = \bar{X}$，故 λ^2 的 ML 估计为 $\bar{X}^2$，有

$$\sqrt{n}[h(\bar{X}) - h(\lambda)] \xrightarrow{\omega} N(0, (h'(\lambda))^2\lambda) = N(0, 4\lambda^3) \quad (n \to \infty),$$

式中 $h(\bar{X}) = \bar{X}^2$，$h(\lambda) = \lambda^2$，由此可知 λ^2 的 ML 估计 $\bar{X}^2$ 是渐近正态的，其渐近方差为 $4\lambda^2$。

例：总体: $X \sim N(\mu, 1)$, 待估量: $g(\mu) = P_\mu(X \leqslant x_0)$。(使用 ML 估计和经验估计方法)

例：总体: $X \sim N(\mu, 1)$, 待估量: $g(\mu) = P_\mu(X \leqslant x_0)$。(使用 ML 估计和经验估计方法)

方法一、设 $g(\mu) = P_\mu(X - \mu \leqslant x_0 - \mu) = \Phi(x_0 - \mu)$.
由 CLT, μ 的最大似然估计 $\hat{\mu} = \bar{X}$ 渐近正态, 渐近方差 $= 1$.
再由 Δ 方法, $g(\mu)$ 的最大似然估计 $g(\hat{\mu}) = \Phi\left(x_0 - \bar{X}\right)$ 渐近正态, 渐近方差为

$$\sigma_1^2 = g'(\mu)^2 \cdot 1 = \varphi(x_0 - \mu)^2.$$

方法二、注意到 $\frac{1}{n}\sum_{i=1}^n 1_{\{X_i \leqslant x_0\}} \xrightarrow{P_\mu} P_\mu(X_i \leqslant x_0) = g(\mu)$, 渐近正态, 渐近方差为

$$\sigma_2^2 = \mathrm{var}\left(1_{\{X \leqslant x_0\}}\right) = g(\mu)(1 - g(\mu)) = \Phi(x_0 - \mu)(1 - \Phi(x_0 - \mu))$$