定理8.6. 若随机向量 $X=(X_1,X_2,\cdots,X_k)\sim Dir(\alpha_1,\alpha_2,\cdots,\alpha_k)$, 设 $\tilde{\alpha}=\alpha_1+\alpha_2+\cdots+\alpha_k$ 和 $\tilde{\alpha}_i=\alpha_i/\tilde{\alpha}$, 则

$$E[X_i]=\tilde{\alpha}_i \qquad \text{和} \qquad Cov(X_i,X_j)=\begin{cases}\frac{\tilde{\alpha}_i(1-\tilde{\alpha}_i)}{\tilde{\alpha}+1} & i=j,\\ -\frac{\tilde{\alpha}_i\tilde{\alpha}_j}{\tilde{\alpha}+1} & i\neq j.\end{cases}$$

Proof. 根据期望的定义有

$$\begin{aligned} E[X_i] &= \frac{\int\int_{\sum_i x_i=1,x_i\geq 0} x_1^{\alpha_1-1}x_2^{\alpha_2-1}\cdots x_k^{\alpha_k-1}\cdot x_i dx_1\cdots dx_k}{\mathrm{Beta}(\alpha_1,\alpha_2,\cdots,\alpha_k)} \\ &= \frac{\mathrm{Beta}(\alpha_1,\cdots,\alpha_i+1,\cdots,\alpha_k)}{\mathrm{Beta}(\alpha_1,\cdots,\alpha_i,\cdots,\alpha_k)}=\frac{\alpha_i}{\alpha_1+\alpha_2+\cdots+\alpha_k}=\tilde{\alpha}_i. \end{aligned}$$

若 $i=j$, 则有

$$\mathrm{Cov}(X_i,X_i)=E[X_i^2]-(E[X_i])^2=\frac{Beta(\alpha_1,\cdots,\alpha_i+2,\cdots,\alpha_k)}{Beta(\alpha_1,\cdots,\alpha_i,\cdots,\alpha_k)}-(\tilde{\alpha}_i)^2=\frac{\tilde{\alpha}_i(1-\tilde{\alpha}_i)}{\tilde{\alpha}+1}.$$

若 $i\neq j$, 则有

$$\begin{aligned} Cov(X_i,X_j) &= E[X_iX_j]-E[X_i]E[X_j]=\frac{\mathrm{Beta}(\alpha_1,\cdots,\alpha_i+1,\cdots,\alpha_j+1,\cdots,\alpha_k)}{\mathrm{Beta}(\alpha_1,\cdots,\alpha_i,\cdots,\alpha_j,\cdots,\alpha_k)}-\tilde{\alpha}_i\tilde{\alpha}_j \\ &= \frac{\alpha_i\alpha_j}{\tilde{\alpha}(\tilde{\alpha}+1)}-\tilde{\alpha}_i\tilde{\alpha}_j=-\frac{\tilde{\alpha}_i\tilde{\alpha}_j}{\tilde{\alpha}+1}. \end{aligned}$$

□

8.4 正态总体抽象分布定理

8.4.1 χ^2 分布

定义8.9. 若 $X_1,X_2,\cdots,X_n$ 是来自总体 $X\sim\mathcal{N}(0,1)$ 的一个样本, 称 $Y=X_1^2+X_2^2+\cdots+X_n^2$ 为服从自由度为 n 的 χ^2 分布, 记 $Y\sim\chi^2(n)$.

根据 $X_1^2\sim\Gamma(1/2,1/2)$ 和 Γ函数的可加性可得 $Y\sim\Gamma(n/2,1/2)$. 于是有随机变量 Y 的概率密度为

$$f_Y(y)=\begin{cases}\frac{(\frac{1}{2})^{\frac{n}{2}}}{\Gamma(n/2)}y^{\frac{n}{2}-1}e^{-\frac{y}{2}} & y>0\\ 0 & y\leq 0\end{cases}$$

下面研究 χ^2 分布的性质:

定理8.7. 若随机变量 $X\sim\chi^2(n)$, 则 $E(X)=n$ 和 $Var(X)=2n$; 若随机变量 $X\sim\chi^2(m)$ 和 $Y\sim\chi^2(n)$ 相互独立, 则 $X+Y\sim\chi^2(m+n)$;

Proof. 若随机变量 $X\sim\chi^2(n)$, 则有 $X=X_1^2+X_2^2+\cdots+X_n^2$, 其中 $X_1,X_2,\cdots,X_n$ 是总体为 $X'\sim\mathcal{N}(0,1)$ 的一个样本. 我们有

$$\begin{aligned} E[X] &= E[X_1^2+X_2^2+\cdots+X_n^2]=nE[X_1^2]=n, \\ \mathrm{Var}(X) &= n\mathrm{Var}(X_1^2)=n[E(X_1^4)-(E(X_1^2))^2]=n(E(X_1^4)-1). \end{aligned}$$

计算

$$E(X_1^4)=\int_{-\infty}^{+\infty}\frac{x^4}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=-\int_{-\infty}^{+\infty}\frac{x^3}{\sqrt{2\pi}}de^{-\frac{x^2}{2}}=3\int_{-\infty}^{+\infty}\frac{x^2}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=3$$

可得 $\mathrm{Var}(X)=2n$. □

若随机变量 $X\sim\mathcal{N}(0,1)$, 则

$$E(X^k)=\begin{cases}(k-1)!! & k\text{ 为偶数}\\ 0 & k\text{ 为奇数}\end{cases}$$

其中 $(2k)!!=2k\cdot(2k-2)\cdot\dots\cdot 2$ 和 $(2k+1)!!=(2k+1)\cdot(2k-1)\cdot\dots\cdot 1$.

例8.4. 设 X_1,X_2,X_3,X_4 是来自于总体 $\mathcal{N}(0,4)$ 的样本, 以及 $Y=a(X_1-2X_2)^2+b(3X_3-4X_4)^2$. 求 a,b 取何值时, Y 服从 χ^2 分布, 并求其自由度.

解. 根据正态分布的性质有 $X_1-2X_2\sim\mathcal{N}(0,20)$ 和 $3X_3-4X_4\sim\mathcal{N}(0,100)$, 因此

$$\frac{X_1-2X_2}{2\sqrt{5}}\sim\mathcal{N}(0,1),\qquad\frac{3X_3-4X_4}{10}\sim\mathcal{N}(0,1),$$

所以当$a=\frac{1}{2\sqrt{5}},b=\frac{1}{10}$时有$Y\sim\chi^2(2)$成立. □

分布可加性:

- 如果 $X\sim\mathcal{N}(\mu_1,a_1^2)$ 和 $Y\sim\mathcal{N}(\mu_2,a_2^2)$, 且 X 与 Y 独立, 那么 $X\pm Y\sim\mathcal{N}(\mu_1\pm\mu_2,a_1^2+a_2^2)$;
- 如果 $X\sim B(n_1,p)$ 和 $Y\sim B(n_2,p)$, 且 X 与 Y 独立, 那么 $X+Y\sim B(n_1+n_2,p)$;
- 如果 $X\sim P(\lambda_1)$ 和 $Y\sim P(\lambda_2)$, 且 X 与 Y 独立, 那么 $X+Y\sim P(\lambda_1+\lambda_1)$;
- 如果 $X\sim\Gamma(\alpha_1,\lambda)$ 和 $Y\sim\Gamma(\alpha_2,\lambda)$, 且 X 与 Y 独立, 那么 $X+Y\sim\Gamma(\alpha_1+\alpha_2,\lambda)$.
- 如果 $X\sim\chi(m)$ 和 $Y\sim\chi(n)$, 且 X 与 Y 独立, 那么 $X+Y\sim\chi(m+n)$.

8.4.2 t 分布 (student distribution)

定义8.10. 随机变量 $X\sim\mathcal{N}(0,1)$ 和 $Y\sim\chi^2(n)$ 相互独立, 则随机变量

$$T=\frac{X}{\sqrt{Y/n}}$$

服从自由度为 n 的 t-分布, 记 $T\sim t(n)$.

随机变量 $T\sim t(n)$ 的概率密度为

$$f(x)=\frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}}\left(1+\frac{x^2}{n}\right)^{\frac{n+1}{2}}\qquad x\in(-\infty,+\infty).$$

由此可知 t-分布的密度函数 $f(x)$ 是偶函数. 当 $n>1$ 为偶数时有

$$\frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}}=\frac{(n-1)(n-3)\cdots5\cdot3}{2\sqrt{n}(n-2)(n-4)\cdots4\cdot2};$$

当 $n>1$ 为奇数时有

$$\frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}}=\frac{(n-1)(n-3)\cdots4\cdot2}{\pi\sqrt{n}(n-2)(n-4)\cdots5\cdot3}.$$

当 $n\to\infty$ 时, 随机变量 $T\sim t(n)$ 的概率密度

$$f(x)\to\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}.$$

因此当 n 足够大时, $f(x)$ 可被近似为 $\mathcal{N}(0,1)$ 的密度函数.

8.4.3 F 分布

定义8.11. 设随机变量 $X\sim\chi^2(m)$ 和 $Y\sim\chi^2(n)$ 相互独立, 称随机变量

$$F=\frac{X/m}{Y/n}$$

服从自由度为 (m,n) 的 F-分布, 记 $F\sim F(m,n)$.

随机变量 $F\sim F(m,n)$ 的概率密度为

$$f(x)=\begin{cases}\frac{\Gamma(\frac{m+n}{2})(\frac{m}{n})^{\frac{m}{2}}x^{\frac{m}{2}-1}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})(1+\frac{mx}{n})^{\frac{m+n}{2}}} & x>0\\ 0 & x\le 0\end{cases}$$

若随机变量 $F\sim F(m,n)$, 则 $\frac{1}{F}=F(n,m)$.

课题练习:

- 独立同分布随机变量 $X_1,X_2,\cdots,X_n$ 满足 $X_i\sim\mathcal{N}(\mu_i,\sigma_i^2)$, 求 $\sum_{i=1}^n(X_i-\mu_i)^2/\sigma_i^2$ 的分布.
- 设 $X_1,X_2,\cdots,X_9$ 和 $Y_1,Y_2,\cdots,Y_9$ 是分别来自总体 $\mathcal{N}(0,9)$ 的两个独立样本, 求 $(X_1+X_2+\cdots+X_9)/\sqrt{Y_1^2+Y_1^2+\cdots+Y_9^2}$ 的分布.
- 设 $X_1,X_2,\dots,X_{2n}$ 来自总体 $\mathcal{N}(0,\sigma_2)$ 的样本, 求 $(X_1^2+X_3^2+\cdots+X_{2n-1}^2)/(X_2^2+X_4^2+\cdots+X_{2n}^2)$ 的分布.

8.4.4 正态分布的抽样分布定理

定理8.8. 设 $X_1,X_2,\cdots,X_n$ 是来自总体 $\mathcal{N}(\mu,\sigma^2)$ 的样本, 则有

$$\bar{X}=\frac{1}{n}\sum_{i=1}^n X_i\sim\mathcal{N}(\mu,\frac{\sigma^2}{n}),\qquad\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim\mathcal{N}(0,1).$$

定理8.9. 设 $X_1, X_2, \cdots, X_n$ 是来自总体 $\mathcal{N}(\mu, \sigma^2)$ 的样本, 其样本均值和修正样本方差分别为

$$\bar{X} = \sum_{i=1}^{n} X_i/n \quad \text{和} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2,$$

则有 $\bar{X}$ 和 S^2 相互独立, 且

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

此定理证明参考书的附件.

定理8.10. 设 $X_1, X_2, \cdots, X_n$ 是来自总体 $\mathcal{N}(\mu, \sigma^2)$ 的样本, 其样本均值和修正样本方差分别为

$$\bar{X} = \sum_{i=1}^{n} X_i/n \quad \text{和} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2,$$

则有

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1).$$

Proof. 根据前面两个定理可知 $(\bar{X} - \mu)/\sigma\sqrt{n} \sim \mathcal{N}(0, 1)$ 和 $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$, 于是有

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \Bigg/ \sqrt{\frac{(n-1)S^2}{\sigma^2(n-1)}} \sim t(n-1).$$

□

定理8.11. 设 $X_1, X_2, \cdots, X_m$ 和 $Y_1, Y_2, \cdots, Y_n$ 分别来自总体 $\mathcal{N}(\mu_X, \sigma^2)$ 和 $\mathcal{N}(\mu_Y, \sigma^2)$ 的两个独立样本, 令其样本均值分别 $\bar{X}$ 和 $\bar{Y}$, 修正样本方差分别为 S_X^2 和 S_Y^2, 则

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}}\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t(n_1 + n_2 - 2).$$

Proof. 根据正太分布的性质有 $\bar{X} \sim \mathcal{N}(\mu_X, \sigma^2/m)$ 和 $\bar{Y} \sim \mathcal{N}(\mu_Y, \sigma^2/n)$, 以及

$$\bar{X} - \bar{Y} \sim \mathcal{N}\left(\mu_X - \mu_Y, \left(\frac{1}{m} + \frac{1}{n}\right)\sigma^2\right),$$

进一步有

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sigma\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim \mathcal{N}(0, 1).$$

根据定理 8.9 有 $\frac{(m-1)S_X^2}{\sigma^2} \sim \chi^2(m-1)$ 和 $\frac{(n-1)S_Y^2}{\sigma^2} \sim \chi^2(n-1)$, 由此得到

$$\frac{(m-1)S_X^2 + (n-1)S_Y^2}{\sigma^2} \sim \chi^2(m+n-2).$$

从而完成证明. □

定理8.12. 设 $X_1, X_2, \cdots, X_m$ 和 $Y_1, Y_2, \cdots, Y_n$ 分别来自总体 $\mathcal{N}(\mu_X, \sigma_X^2)$ 和 $\mathcal{N}(\mu_Y, \sigma_Y^2)$ 的两个独立样本, 令其修正样本方差分别为 S_X^2 和 S_Y^2, 则有

$$\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F(m-1, n-1).$$

Proof. 根据定理 8.9 有 $\frac{(m-1)S_X^2}{\sigma_X^2} \sim \chi^2(m-1)$ 和 $\frac{(n-1)S_Y^2}{\sigma_Y^2} \sim \chi^2(n-1)$, 由此得到

$$\frac{\frac{(m-1)S_X^2}{\sigma_X^2}/(m-1)}{\frac{(n-1)S_Y^2}{\sigma_Y^2}/(n-1)} \sim F(m-1, n-1).$$

□

课堂习题:

- 若随机变量 $X \sim t(n)$, 求 $Y = X^2$ 的分布.
- 设 $X_1, X_2, \cdots, X_5$ 是来自总体 $\mathcal{N}(0,1)$ 的样本, 令 $Y = c_1(X_1 + X_3)^2 + c_2(X_2 + X_4 + X_5)^2$. 求常数 c_1, c_2 使 Y 服从 χ^2 分布.
- 设 X_1, X_2 是来自总体 $\mathcal{N}(0, \sigma^2)$ 的样本, 求 $\frac{(X_1+X_2)^2}{(X_1-X_2)^2}$ 的分布.

8.4.5 分位数(点)

定义8.12. 对给定 $\alpha \in (0,1)$ 和随机变量 X, 称满足 $\Pr(X > \lambda_\alpha) = \alpha$ 的实数 λ_α 为上侧 α 分位数*(点)*.

对正态分布 $X \sim \mathcal{N}(0,1)$, 给定 $\alpha \in (0,1)$, 满足 $\Pr(X > \mu_\alpha) = \int_{\mu_\alpha}^{\infty} f(x)dx = \alpha$ 的点 μ_α 称为正态分布上侧 α 分位点, 由对称性可知 $\mu_{1-\alpha} = -\mu_\alpha$.

对 $\chi^2(n)$ 分布 $X \sim \chi^2(n)$, 给定$\alpha \in (0,1)$, 满足 $\Pr(X \geq \chi_\alpha^2(n)) = \alpha$ 的点 $\chi_\alpha^2(n)$ 称为 $\chi^2(n)$ 分布上侧 α 分位点. 当 $n \to \infty$ 时有 $\chi_\alpha^2(n) \approx \frac{1}{2}(\mu_\alpha + \sqrt{2n-1})^2$, 其中 μ_α 表示正态分布上侧 α 分位点.

对 t-分布 $X \sim t(n)$, 给定 $\alpha \in (0,1)$, 满足 $\Pr(X > t_\alpha(n)) = \alpha$ 的点 $t_\alpha(n)$ 称为 $t(n)$-分布上侧 α 分位点. 由对称性可知 $t_{(1-\alpha)}(n) = -t_\alpha(n)$.

对 F-分布 $X \sim F(m,n)$, 给定 $\alpha \in (0,1)$, 满足 $\Pr[X > F_\alpha(m,n)] = \alpha$ 的点 $F_\alpha(m,n)$ 称为 $F(m,n)$ 分布上侧 α 分位点.

对于 F-分布, 有如下性质:

引理8.4. 对 F 分布的分位点有

$$F_{(1-\alpha)}(m,n) = \frac{1}{F_\alpha(n,m)}.$$

Proof. 设 $X \sim F(m,n)$, 根据定义有

$$1 - \alpha = \Pr(X > F_{1-\alpha}(m,n)) = \Pr\left(\frac{1}{X} < \frac{1}{F_{1-\alpha}(m,n)}\right) = 1 - \Pr\left(\frac{1}{X} \geq \frac{1}{F_{1-\alpha}(m,n)}\right).$$

再根据 $1/X \sim F(n,m)$, 结合上式有

$$\alpha = \Pr\left(\frac{1}{X} \geq \frac{1}{F_{1-\alpha}(m,n)}\right) = \Pr\left(\frac{1}{X} > \frac{1}{F_{1-\alpha}(m,n)}\right)$$

于是有 $F_\alpha(n,m) = 1/F_{1-\alpha}(m,n)$. □

课堂习题:

- 设 $X_1, X_2, \cdots, X_{10}$ 是总体 $\mathcal{N}(\mu, 1/4)$ 的样本, i) 若 $\mu = 0$, 求 $\Pr(\sum_{i=1}^{10} X_i^2 \geq 4)$; ii) 若 μ 未知, 求 $\Pr(\sum_{i=1}^{10}(X_i - \bar{X})^2 \geq 2.85)$.
- 设 $X_1, X_2, \cdots, X_{25}$ 是总体 $\mathcal{N}(12, \sigma^2)$ 的样本, i) 若 $\sigma = 2$, 求 $\Pr(\sum_{i=1}^{25} X_i/25 \geq 12.5)$; ii) 若 σ 未知但知道修正样本方差为 $S^2 = 5.57$, 求 $\Pr(\sum_{i=1}^{25} X_i/25 \geq 12.5)$.

9 参数估计

设总体 X 的分布函数为 $F(X,\theta)$, 其中 θ 为未知参数(也可为向量). 现从总体中抽取一样本 $X_1,X_2,\cdots,X_n$, 如何依据样本估计参数 θ, 或 θ 的函数 $g(\theta)$, 此类问题称为参数估计问题. 内容包括: 点估计, 估计量标准, 区间估计.

9.1 点估计

9.1.1 矩估计法

总体 X 的 k 阶矩: $a_k=E[X^k]$ 样本 k 阶矩: $A_k=\frac{1}{n}\sum_{i=1}^n X_i^k$

用相应的样本矩去估计总体矩, 求解参数 θ 的方法称为 **矩估计法**. 矩估计法的理论基础是大数定理: $X_1,X_2,\cdots,X_n$ 为 i.i.d. 的随机变量, 若 $E(X)=\mu$, 则当 $n\to\infty$ 时有

$$\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{P} \mu.$$

推论: 若 $E[X^k]=a_k$ 存在, 则当 $n\to\infty$ 时有

$$\frac{1}{n}\sum_{i=1}^n X_i^k \xrightarrow{P} a_k=E[X^k].$$

还可利用中心距进行估计:
总体 X 的 k 阶中心矩: $b_k=E[(X-E(X))^k]$ 样本 k 阶中心矩: $B_k=\frac{1}{n}\sum_{i=1}^n(X_i-\bar{X})^k$

矩估计方法: 总体 X 的分布函数 F 包含 m 个未知参数 $\theta_1,\theta_2,\cdots,\theta_m$,

1) 求总体 X 的 k 阶矩: $a_k=a_k(\theta_1,\theta_2,\cdots,\theta_m)=E[X^k],\ k\in[m]$ (a_k 一般为 $\theta_1,\theta_2,\cdots,\theta_m$ 的函数).
2) 计算样本的 k 阶矩: $A_k=\frac{1}{n}\sum_{i=1}^n X_i^k$.
3) 令样本矩等于总体矩 $A^k=a^k=a^k(\theta_1,\theta_2,\cdots,\theta_m)$ $(k=1,2,\cdots,m)$, 得到 m 个关于 $\theta_1,\theta_2,\cdots,\theta_m$ 的方程组.
4) 求解方程组得到估计量 $\widehat{\theta}_1,\widehat{\theta}_2,\cdots,\widehat{\theta}_n$.

例9.1. 设总体 X 的概率密度函数

$$f(x)=\begin{cases}(\alpha+1)x^\alpha & x\in(0,1)\\ 0 & \text{其它},\end{cases}$$

设 $X_1,X_2,\cdots,X_n$ 是来自总体 X 的样本, 求参数 α 的矩估计.

解. 首先计算总体 X 的期望

$$E[X]=\int_{-\infty}^{+\infty} xf(x)dx=\int_0^1 x(\alpha+1)x^{\alpha+1}dx=\frac{\alpha+1}{\alpha+2}.$$

样本 X 的均值 $\bar{X}=\sum_{i=1}^n X_i/n$. 样本矩等于总体矩有

$$E(X)=\frac{\alpha+1}{\alpha+2}=\bar{X},$$

求解可得 $\alpha=(2\bar{X}-1)/(1-\bar{X})$. □

例9.2. 设 $X_1,X_2,\cdots,X_n$ 是来自总体 X 的样本, 且总体 X 的密度函数为

$$f(x)=\begin{cases}\theta e^{-(x-\mu)\theta} & x\geq\mu\\ 0 & \text{其它},\end{cases}$$

其中 $\theta>0$, 求 μ,θ 的矩估计.

解. 设随机变量 $Y=X-\mu$, 则 Y 服从参数为 θ 的指数分布, 有

$$E(Y)=1/\theta \quad \text{和} \quad \mathrm{Var}(Y)=1/\theta^2.$$

由此可得 $E(X)=\mu+1/\theta$ 和 $\mathrm{Var}(X)=1/\theta^2$. 计算对应的样本矩

$$A_1=\bar{X}=\frac{1}{n}\sum_{i=1}^n X_i, \qquad B_2=\frac{1}{n}\sum_{i=1}^n (X_i-\bar{X})^2.$$

求解方程组

$$\mu+\frac{1}{\theta}=A_1, \qquad \frac{1}{\theta^2}=B_2,$$

解得 $\mu=\bar{X}-\sqrt{\sum_{i=1}^n (X_i-\bar{X})^2/n}$ 和 $\theta=1/\sqrt{\sum_{i=1}^n (X_i-\bar{X})^2/n}$. □

课堂习题:

- 求正态总体 $\mathcal{N}(\mu,\sigma^2)$ 的 μ,σ^2 的矩估计法.
- 求总体 $X\sim\mathcal{U}(a,b)$ 中 a,b 的矩估计法.

9.1.2 最大似然估计法

设 $X_1,X_2,\cdots,X_n$ 是来自总体 X 的一个样本. 若 X 为离散型随机变量, 其分布列为 $\Pr(X=x)=\Pr(X=x;\theta)$, 则样本 $X_1,X_2,\cdots,X_n$ 的分布列为

$$L(\theta)=L(x_1,x_2,\cdots,x_n;\theta)=\prod_{i=1}^n \Pr(x_i;\theta).$$

这里 $L(\theta)$ 表示样本 $X_1=x_1,X_2=x_2,\cdots,X_n=x_n$ 发生的概率.

若 X 为连续型随机变量, 其概率密度为 $f(x;\theta)$, 则 $X_1=x_1,X_2=x_2,\cdots,X_n=x_n$ 的联合概率密度为 $\prod_{i=1}^n f(x_i;\theta)$, 记

$$L(\theta)=L(x_1,x_2,\cdots,x_n;\theta)=\prod_{i=1}^n f(x_i;\theta).$$

根据概率密度定义可知 $L(\theta)$ 越大, 样本 $(X_1, X_2, \cdots, X_n)$ 落入 $(x_1, x_2, \cdots, x_n)$ 的邻域内概率越大.

综合上述离散和连续两种随机变量, 统称 $L(\theta)$ 为样本 $X_1, X_2, \cdots, X_n$ 的似然函数, 可以发现 $L(\theta)$ 是 θ 的函数. 若

$$\hat{\theta} = \arg\max_{\theta} L(x_1, x_2, \cdots, x_m; \theta),$$

则称 $\hat{\theta}$ 为 θ 的最大似然估计量. 直觉而言: 最大似然估计量 $\hat{\theta}$ 是使观测值 $X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n$ 出现的概率最大.

求解最大化似然估计的步骤如下:

1) 写出对数似然函数 $\log(L(x_1, x_2, \cdots, x_m; \theta)$;

2) 对对数似然函数的参数 θ 求一阶导数或一阶偏导, 令其导数等于0;

4) 求解方程组得到最大似然估计量 $\hat{\theta}$.

例9.3. 设 $X_1, X_2, \cdots, X_n$ 是取自总体 $X \sim B(1, p)$ 的一个样本, 求参数 p 的极大似然估计.

解. 首先计算似然函数

$$L(p) = \prod_{i=1}^{n} p^{X_i}(1-p)^{1-X_i} = p^{\sum_{i=1}^{n} X_i}(1-p)^{n-\sum_{i=1}^{n} X_i},$$

以及对数似然函数

$$\ln L(p) = \sum_{i=1}^{n} X_i \ln p + \left(n - \sum_{i=1}^{n} X_i\right) \ln(1-p),$$

设

$$0 = \frac{d \ln L(p)}{dp} = \frac{1}{p}\sum_{i=1}^{n} X_i - \frac{1}{1-p}\left(n - \sum_{i=1}^{n} X_i\right).$$

由此可得 $p = \sum_{i=1}^{n} X_i/n = \bar{X}$. [验证矩估计法] □

例9.4. 设 $X_1, X_2, \cdots, X_n$ 为总体 $X \sim \mathcal{N}(\mu, \sigma^2)$ 的样本, 求 μ 和 σ^2 的最大似然估计.

解. 根据正太分布可知 X 的概率密度函数为 $f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. 样本 $X_1, X_2, \cdots, X_n$ 的似然函数为

$$L(\mu, \sigma^2) = \frac{1}{(\sqrt{2\pi}\sigma)^n}\exp\left(-\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{2\sigma^2}\right).$$

其对数似然函数为 $\ln L(\mu, \sigma^2) = -n\ln(2\pi)^{\frac{1}{2}} - n\ln\sigma - \sum_{i=1}^{n}(X_i-\mu)^2/2\sigma^2$. 对参数 μ 求导计算可得

$$\frac{d \ln L(\mu, \sigma^2)}{d\mu} = \sum_{i=1}^{n}(X_i - \mu) = 0 \quad \Longrightarrow \quad \mu = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X},$$

对 σ 求导计算可得

$$\frac{d \ln L(\mu, \sigma^2)}{d\sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i-\mu)^2 = 0 \quad \Longrightarrow \quad \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2.$$

□

前面两例中的矩估计和最大似然估计完全一样.

例9.5. 设总体 X 的密度函数为

$$f(x) = \begin{cases} (\alpha+1)x^{\alpha} & x \in (0,1) \\ 0 & \text{其它} \end{cases}$$

设 $X_1, X_2, \cdots, X_n$ 是总体 X 的样本, 求 α 的极大似然估计.

解. 首先得到似然函数为

$$L(\alpha) = (\alpha+1)^n \prod_{i=1}^{n} X_i^{\alpha} = (\alpha+1)^n (X_1 X_2 \cdots X_n)^{\alpha},$$

以及其对数似然函数 $\ln L(\alpha) = n\ln(\alpha+1) + \alpha \ln(X_1 X_2 \cdots X_n)$. 求导并令导数为零有

$$\frac{d\ln L(\alpha)}{d\alpha} = \frac{n}{\alpha+1} + \ln(X_1 X_2 \cdots X_n) = 0,$$

求解得

$$\alpha = \frac{n}{\sum_{i=1}^{n} \ln(\frac{1}{X_i})} - 1 = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} \ln(\frac{1}{X_i})} - 1.$$

□

对上例, 矩估计值为 $\alpha = (2\bar{X}-1)/(1-\bar{X})$, 因此矩估计值与极大似然估计值可能不同.

例9.6. 总体$X \sim U(a,b)$, 求a,b的极大似然估计.

解. 首先得到X的密度函数为$f(x) = 1/(b-a)$ $(x \in [a,b])$, 其它为零, 因此似然函数为

$$L(a,b) = \begin{cases} \frac{1}{(b-a)^n} & a \le X_1, X_2, \cdots, X_n \le b \\ 0 & \text{其它} \end{cases}$$

直接求导无法解出a与b的估计, 此时可以从极大似然定义出发, 最大化$L(a,b)$, 应使得b尽可能小、a尽可能大, 但需满足$a \le X_1, X_2, \cdots, X_n \le b$, 因此极大似然估计量为:

$$b = \max\{X_1, X_2, \cdots, X_n\} \qquad a = \min\{X_1, X_2, \cdots, X_n\}.$$

□

极大似然估计不可变性: 设 $\hat{\theta}$ 为 θ 的极大似然估计, 设 $\mu(\theta)$ 为 θ 的函数, 且存在反函数 $\theta = \theta(\mu)$, 则 $\hat{\mu} = \mu(\hat{\theta})$ 是 μ 的最大似然估计.

例9.7. 设 $X_1, X_2, \cdots, X_n$ 是总体 X 的样本, 以及总体的概率密度为

$$f(x) = \begin{cases} \theta e^{-(x-\mu)\theta} & x \ge \mu \\ 0 & \text{其它}, \end{cases}$$

求 μ, θ 的极大似然估计.

解. 首先得到似然函数为

$$L(\theta,\mu)=\begin{cases}\theta^n e^{-\theta\sum_{i=1}^n(X_i-\mu)} & X_i\geq\mu\\ 0 & \text{其它}\end{cases}$$

进一步得到对数似然函数为

$$\ln L(\theta,\mu)=n\ln\theta-\theta\sum_{i=1}^n(X_i-\mu).$$

求偏导、并零偏导等于零有

$$\frac{\ln L(\theta,\mu)}{d\theta}=\frac{n}{\theta}-\sum_{i=1}^n(X_i-\mu)=0 \quad\Rightarrow\quad \theta=\frac{1}{\frac{1}{n}\sum_{i=1}^n(X_i-\mu)},$$

另一方面有

$$\frac{\ln L(\theta,\mu)}{d\mu}=n\theta=0 \quad\Rightarrow\quad \theta=0,$$

此时无法求解 θ,μ 的极大似然估计.

回到似然函数的定义

$$L(\theta,\mu)=\begin{cases}\theta^n e^{-\theta\sum_{i=1}^n(X_i-\mu)} & X_1,X_2,\cdots,X_n\geq\mu\\ 0 & \text{其它}\end{cases}$$

可以发现 μ 越大似然函数 $L(\theta,\mu)$ 越大, 但须满足 $X_i\leq\mu$ $(i\in[n])$. 由此可得极大似然估计

$$\hat{\mu}=\min\{X_1,X_2,\cdots,X_n\},$$

进一步求解可得

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

□