概率论与数理统计习题课-3^{rd}

复习复习!

2024 年 12 月 21 日

目录

1 大数定律与中心极限定律

1.1 基础知识

(弱) 大数定律讨论的是, 在何种条件下, 独立同分布随机变量的均值**依概率收敛**于期望. 称随机变量序列 $\{Y_n\}_{n=1}^{\infty}$ 依概率收敛于 Y, 如果对任意的 $\epsilon > 0$, 都有

$$\lim_{n\to\infty} \mathbb{P}\{|Y_n - Y| \geq \epsilon\} = 0,$$

或者, 等价地,

$$\lim_{n\to\infty} \mathbb{P}\{|Y_n - Y| < \epsilon\} = 1.$$

1.1 定理 (伯努利大数定理). *独立同分布的 0-1 随机变量服从弱大数定律. 具体地说, 设 p 是每次 Bernoulli 试验中事件 A 发生的概率, n_A 为 n 次试验中事件 A 发生的次数, 那么对任意的 $\epsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\left\{\left|\frac{n_A}{n} - p\right| < \epsilon\right\} = 1.$$

1.2 定理 (辛钦大数定律). 期望存在的独立同分布的随机变量服从弱大数定律. 具体地说, 设独立同分布的随机变量 $X_1, X_2, \ldots$ 期望为 $\mathbb{E}(X_k) = \mu$, 那么对任意的 $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| < \epsilon\right\} = 1.$$

中心极限定理讨论的是, 在何种条件下, 独立随机变量**和的极限分布为正态分布**. 和大数定律不同的是, 通过中心极限定理我们可以获取关于估计量分布的信息——从而提供了一种计算独立随机变量和的近似概率的简单**方法**. 它还解释了一个引人注目的事实, 即如此多的自然“种群”的经验频率呈现钟形 (即正态) 曲线.

1.3 定理 (独立同分布的中心极限定理). 设独立同分布的随机变量 $X_1, X_2, \ldots$ 期望为 $\mathbb{E}(X_k) = \mu$, 方差为 $D(X_k) = \sigma^2 > 0$, 那么标准化的随机变量之和的极限分布服从标准正态分布: 对任意 t,

$$\lim_{n\to\infty} \mathbb{P}\left\{\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \le t\right\} = \Phi(t).$$

下面的中心极限定理给出了二项分布的正态近似.

1.4 定理 (棣莫弗–拉普拉斯中心极限定理). 设 p 是每次 *Bernoulli* 试验中事件 A 发生的概率, n_A 为 n 次试验中事件 A 发生的次数, 那么标准化的 n_A 的极限分布服从标准正态分布: 对任意 t,

$$\lim_{n\to\infty} \mathbb{P}\left\{\frac{n_A - np}{\sqrt{np(1-p)}} \le t\right\} = \Phi(t).$$

1.2 例题示范

1.5 示例. 掷一枚质地均匀的骰子 100 次, 记 X_i 为第 i 次掷出的点数, $\overline{X}$ 为 $X_1, \ldots, X_{100}$ 的均值, 试用中心极限定理估计概率 $P\{3 \le \overline{X} \le 4\}$.

解. 由题意可得 $\mathbb{E}(X_i) = 3.5$, $D(X_i) = \frac{35}{12}$, $\mathbb{E}\left(\overline{X}\right) = 3.5$, $D(\overline{X}) = \frac{7}{240}$, 于是由中心极限定理,

$$\mathbb{P}\{3 \le \overline{X} \le 4\} \approx \mathbb{P}\left\{\frac{3-3.5}{\sqrt{7/240}} \le \frac{\overline{X}-3.5}{\sqrt{7/240}} \le \frac{4-3.5}{\sqrt{7/240}}\right\} = 2\Phi(2.9277) - 1 = 0.9966.$$

□

1.6 示例. *进行独立重复试验, 每次试验中事件 A 发生的概率为 0.25. 试问能以 95% 的把握保证 1000 次试验中事件 A 发生的频率与概率相差多少? 此时 A 发生的次数在什么范围内?*

解. 记 Y_n 为 1000 次试验中事件 A 发生的次数, 则 $Y_n \sim b(1000, 0.25)$, 且 $\mathbb{E}\left(Y_n\right) = 250, D\left(Y_n\right) = 187.5$. 设事件 A 发生的频率 $Y_n/1000$ 与概率 0.25 的差为 k, 我们有不等式

$$\begin{aligned} 0.95 \leq \mathbb{P}\left\{\left|\frac{Y_n}{1000} - 0.25\right| \leq k\right\} &= \mathbb{P}\left\{\left|\frac{Y_n - 250}{\sqrt{187.5}}\right| \leq \frac{1000k + 0.5}{\sqrt{187.5}}\right\} \\ &\approx 2\Phi\left(\frac{1000k + 0.5}{\sqrt{187.5}}\right) - 1. \end{aligned}$$

即寻求 k, 使得

$$\Phi\left(\frac{1000k + 0.5}{\sqrt{187.5}}\right) \geq 0.975,$$

查表得 $\frac{1000k+0.5}{\sqrt{187.5}} \geq 1.96$, 解得 $k \geq 0.027$. 这表明能以 95% 的把握保证在 1000 次试验中事件 A 发生的频率与概率相差不大于 0.027. 或者说, 有 95% 的把握说, 在 1000 次试验中事件 A 发生的次数在即在 223 次到 277 次间. □

2 样本及抽样分布

2.1 基础知识

数理统计的核心问题是利用样本推断总体——我们在获取数据后, 需要使用有效的方法去集中和提取数据中的有关信息, 对所研究的问题做出一定的结论, 这在统计上被称为“推断”. 著名统计学家 C.R.Rao 在其著作《统计与真理——怎样运用偶然性》中概括道:

- 在终极的分析中, 一切知识都是历史;
- 在抽象的意义下, 一切科学都是数学;
- 在理性的基础上, 所有的判断都是统计学.

2.1.1 总体与样本

在一个统计问题中, 我们把研究对象的全体称为**总体**, 构成总体的每个成员称为**个体**. 而在实际问题中, 我们关心个体的数量指标值, 于是总体就是一堆数——

有大有小、有的出现机会大, 有的出现机会小. 因此用一个概率分布去描述和归纳总体是恰当的, 从这个意义看, 总体就是一个随机变量 X.

对总体的观察称为**抽样**: 我们从总体 X 中进行 n 次独立地、重复地观察, 得到 n 个结果 $X_1, \ldots, X_n$, 称为来自总体 X 的、容量为 n 的**简单随机样本**, 简称**样本**. 于是 $X_1, \ldots, X_n$ **独立**且与总体 X 有**相同分布** F, 记做 $X_1, \ldots, X_n$ i.i.d. $\sim F$.

样本具有**二重性**: 一方面, 由于样本是从总体中随机抽取的, 抽取前无法预知它们的数值, 因此样本是随机变量, 用大写字母 $X_1, \ldots, X_n$ 表示; 另一方面, 样本在抽取以后经观测就有确定的观测值, 因此样本又是一组数值, 此时用小写字母 $x_1, \ldots, x_n$ 表示是恰当的.

2.1.2 统计量

由于样本本身是一些杂乱无章的数字, 我们需要对数字加工处理, 计算出一些有用的量.

2.1 定义 (统计量). *由样本算出来的量称为统计量, 或者说, 统计量是样本的函数. (注意: 统计量中不能含有未知参数.)*

下面列出一些常见的统计量.

- **样本均值** $\overline{X} = \dfrac{1}{n} \sum_{i=1}^{n} X_i$;
- **样本方差** $S^2 = \dfrac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2 = \dfrac{1}{n-1} \left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2 \right)$;
- **样本标准差** $S = \sqrt{S^2}$;
- **样本 k 阶 (原点) 矩** $A_k = \dfrac{1}{n} \sum_{i=1}^{n} X_n^k$, $k = 1, 2, \ldots$;
- **样本 k 阶中心矩** $B_k = \dfrac{1}{n} \sum_{i=1}^{n} (X_n - \overline{X})^k$, $k = 1, 2, \ldots$.

2.2 命题. *设 $X_1, \ldots, X_n$ 为来自总体 X 的样本, $\mathbb{E}(X) = \mu$, $D(X) = \sigma^2$, 那*

么

$$\mathbb{E}(\overline{X}) = \mu,\ D(\overline{X}) = \frac{\sigma^2}{n},\ \mathbb{E}(S^2) = \sigma^2.$$

证明. 我们只证明样本方差的期望是 σ^2. 注意到 X_k 两两独立意味着

$$\mathbb{E}(\overline{X}^2) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i \cdot \overline{X}) = \mathbb{E}(X_k \cdot \overline{X}),$$

其中

$$\mathbb{E}(X_k \cdot \overline{X}) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_k \cdot X_i) = \frac{n-1}{n}\mathbb{E}(X)^2 + \frac{1}{n}\mathbb{E}(X^2).$$

于是

$$\begin{aligned}\mathbb{E}(S^2) &= \frac{1}{n-1}\sum_{i=1}^{n}\mathbb{E}(X_i - \overline{X})^2 = \frac{n}{n-1}\left[\mathbb{E}(X_k^2) - 2\mathbb{E}(X_k \cdot \overline{X}) + \mathbb{E}(\overline{X}^2)\right] \\ &= \frac{n}{n-1}\left[\mathbb{E}(X^2) - \frac{n-1}{n}\mathbb{E}(X)^2 - \frac{1}{n}\mathbb{E}(X^2)\right] = \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\ &= D(X) = \sigma^2.\end{aligned}$$

□

2.3 注. *在统计学中, **自由度**是指当以样本的统计量来估计总体的参数时, 样本中独立或能自由变化的数据的个数——这里的"度"指的是"维度". 若样本 X_1, X_2, X_3 的均值 $\overline{X}$ 是事先指定的, 那么此时真正取值"自由"的随机变量只有 2 个: 例如, 当 X_1, X_2 已知时, X_3 不得不取 $3\overline{X} - X_1 - X_2$. 而在计算 n 个样本的方差时, 样本均值已经给定, 此时自由度是 $n-1$.*

2.1.3 经验分布函数

若我们有总体 X 的 n 个观察值 $x_1, \ldots, x_n$, 我们可以写出分布的经验分布函数

$$F_n(x) = \frac{\#\{x_i : x_i \le x\}}{n},$$

即样本观察值小于等于 x 的频率. 可以看到这样的函数满足分布函数的非负、单增、右连续三个条件.

Glivenko–Cantelli 定理指出, 样本容量极大时, 样本的经验分布趋近于总体分布 (a.e.), 这被称为推断统计学的基石.

2.1.4 三大抽样分布

作为样本的函数, 统计量也是随机变量, 它的分布称为**抽样分布**. 以**标准正态变量**为基石而构造的三个著名统计量在实际中有广泛的应用, 它们被称作数理统计中的**三大抽样分布**.

2.4 定义 *(卡方分布)***.** 设 $X_1, \dots, X_n$ *i.i.d.* $\sim N(0,1)$, 则称随机变量

$$\chi^2 = \sum_{i=1}^{n} X_n^2$$

服从自由度为 n 的 χ^2 分布, 记做 $\chi \sim \chi^2(n)$.

由定义, 不难看出卡方分布关于**自由度**具有**可加性**, 即若 χ_1 与 χ_2 相互独立, $\chi_1 \sim \chi^2(n)$, $\chi_2 \sim \chi^2(m)$, 那么

$$\chi_1 + \chi_2 \sim \chi^2(n+m).$$

此外我们还需熟记卡方分布的期望与方差: 若 $\chi \sim \chi^2(n)$, 那么

$$\mathbb{E}(\chi) = n,\ D(\chi) = 2n.$$

2.5 定义 *(t 分布)***.** 设 $X \sim N(0,1)$, $Y \sim \chi^2(n)$ 且 X 与 Y 相互独立, 则称随机变量

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

服从自由度为 n 的 $\boldsymbol{t}$ 分布, 记做 $t \sim t(n)$.

注意 t 分布的密度函数和标准正态曲线类似, 呈现"中间高, 两边低, 左右对称", 但是两侧的尾部概率比 $N(0,1)$ 大. 于是 t 分布的上分位数满足 $t_\alpha(n) + t_{1-\alpha}(n) = 0$(互为相反数).

2.6 定义 *(F 分布)***.** 设 $U \sim \chi^2(m)$, $V \sim \chi^2(n)$ 且 U, V 相互独立, 则称随机变量

$$F = \frac{U/m}{V/n}$$

服从自由度为 m, n 的 $\boldsymbol{F}$ 分布, 记做 $F \sim F(m,n)$.

由定义不难看出 $t^2(n) = F(1,n)$. 此外 $\frac{1}{F} \sim F(n,m)$, 于是 F 分布的上分位数满足 $F_\alpha(m,n) \cdot F_{1-\alpha}(n,m) = 1$(互为倒数):

$$1 - \alpha = \mathbb{P}(F \le F_\alpha(m,n)) = \mathbb{P}\left(\frac{1}{F} \ge \frac{1}{F_\alpha}(m,n)\right).$$

在命题2.2的基础上，若总体 X 服从正态分布，我们可以构造出一些常用统计量，在假设检验中有着很广泛的应用.

2.7 定理 *(单正态总体的抽样分布)*. *设样本* $X_1, X_2, \dots X_n$ *i.i.d.* $\sim N(\mu, \sigma^2)$，$\overline{X}$，S^2 *分别为样本均值与方差，那么*

1. $\overline{X} \sim N(\mu, \sigma^2/n)$，*于是* $U = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$；
2. $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$；*最关键的一点！请务必记住.*
3. $\overline{X}$ *与* S^2 *独立，于是*
$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1).$$

我们可以看到，最终构造的统计量 $\frac{\overline{X}-\mu}{S/\sqrt{n}}$ **不含总体方差这一未知参数**，并且分布信息可以通过查表得到——从而可以在 σ 未知的情况下，给出 μ 的区间估计.

有的时候，我们需要比较两个正态总体的参数，下面的统计量允许我们在参数 μ_1, μ_2 未知的情况下，对 σ_1/σ_2 做出推断.

2.8 定理 *(双正态总体的抽样分布)*. *设样本* $X_1, X_2, \dots X_m$ *i.i.d.* $\sim N(\mu_1, \sigma_1^2)$，$Y_1, Y_2, \dots Y_n$ *i.i.d.* $\sim N(\mu_2, \sigma_2^2)$，*且两样本之间相互独立.* $\overline{X}, \overline{Y}, S_1^2, S_2^2$ *分别为两个样本的均值与方差，那么*
$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(m-1, n-1).$$

证明. 在上一定理的基础上，我们有 $\frac{(m-1)S_1^2}{\sigma_1^2} \sim \chi^2(m-1)$，$\frac{(n-1)S_1^2}{\sigma_1^2} \sim \chi^2(n-1)$，于是由 F 分布的定义可知定理成立. □

2.2 例题示范

2.9 示例. 设 $x_1, \dots x_{16}$ 为来自 $N(\mu, \sigma^2)$ 的样本，经计算 $\bar{x} = 9$，$s^2 = 5.32$，试求 $\mathbb{P}\{|\bar{x}-\mu| < 0.6\}$.

解. 对于单正态总体，我们考虑构造服从 t 分布的统计量. 由于 $\bar{x} - \mu \sim$

$N(0,\sigma^2/n)$, $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$, 那么

$$t := \frac{\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)s^2}{\sigma^2}/(n-1)}} = \frac{\sqrt{n}(\bar{x}-\mu)}{s} = \frac{4}{\sqrt{5.32}}(\bar{x}-\mu) \sim t(15).$$

于是

$$\begin{aligned}\mathbb{P}\{|\bar{x}-\mu| < 0.6\} = \mathbb{P}\{|\bar{x}-\mu| < 0.6\} = \mathbb{P}\left\{|t| < \frac{4\times 0.6}{\sqrt{5.32}}\right\} \\ = 0.6854.\end{aligned}$$

$\square$

2.10 示例. 若随机变量 $X \sim F(n,n)$, 证明 $\mathbb{P}\{X<1\} = 0.5$.

证明. 由 F 分布的定义有 $Y := 1/X \sim F(n,n)$ 与 X 同分布, 于是

$$\mathbb{P}\{X<1\} = \mathbb{P}\{Y<1\} = \mathbb{P}\{1/X<1\} = \mathbb{P}\{X>1\}.$$

另一方面, 由于 X 是连续型随机变量, 我们有 $\mathbb{P}\{X<1\} + \mathbb{P}\{X>1\} = 1$, 于是 $\mathbb{P}\{X<1\} = 0.5$. $\square$

2.11 示例. 设 $x_1,\dots,x_{15}$ 是总体 $N(0,\sigma^2)$ 的一个样本, 求

$$\frac{x_1^2+\cdots+x_{10}^2}{2(x_{11}^2+\cdots+x_{15}^2)}$$

的分布.

解. 由于 $x_1,\dots x_n$ i.i.d. $N(0,1)$, 那么

$$\frac{1}{\sigma^2}(x_1^2+\cdots+x_{10}^2) \sim \chi^2(10),\ \frac{1}{\sigma^2}(x_{11}^2+\cdots+x_{15}^2) \sim \chi^2(5),$$

于是

$$\frac{x_1^2+\cdots+x_{10}^2}{2(x_{11}^2+\cdots+x_{15}^2)} = \frac{\frac{1}{\sigma^2}(x_1^2+\cdots+x_{10}^2)/10}{\frac{1}{\sigma^2}(2(x_{11}^2+\cdots+x_{15}^2)/5} \sim F(10,5).$$

$\square$

3 参数估计

3.1 基础知识

3.1.1 点估计

对于未知参数 θ, 我们选用统计量 $\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$ 的取值作为 θ 的一个估计量, $\hat{\theta}(x_1, x_2, \cdots, x_n)$ 作为估计值. 常用的点估计有**矩估计**和**最大似然估计**.

矩估计

矩估计的基本想法是"替换原理": 用相应的样本矩 (的函数) **替换**总体矩 (的函数), 并且尽量选取低阶矩. 矩估计的理论依据是辛钦大数定律

$$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{\mathbb{P}} \mathbb{E}(X^k).$$

例如若总体 X 的期望 μ 与方差 σ^2 未知, 我们通常选用一阶样本原点矩 (即样本均值) 和二阶样本中心距分别作为二者的矩估计:

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

最大似然估计

最大似然估计法是求点估计用得最多的方法, 它的基本想法是, 未知参数 θ 的估计应该是使得当前结果出现可能性最大的那个 θ.

设 $X_1, \ldots, X_n$ 为来自总体 X 的样本, X 的分布已知而参数 θ 未知. 最大似然估计的方法步骤为:

1. 求出当前事件出现的概率/密度, 它是样本和未知参数的函数 $L(X_1, \ldots, X_n; \theta)$, 我们称之为**似然函数**;

2. 寻求使得似然函数达到最大的点 $\hat{\theta}$ 作为 θ 的估计

$$\hat{\theta} = \arg\max_{\theta \in \Theta} L(X_1, \ldots, X_n; \theta).$$

   其中 Θ 为参数 θ 的取值范围, 例如方差的参数空间是 $\mathbb{R}_{\geq 0}$.

若 X 为离散型随机变量, 它的分布律为 $\mathbb{P}\{X = x\} = p(x; \theta)$, 由样本的独立性,

$$L(x_1, \ldots, x_n; \theta) := \mathbb{P}\{X_1 = x_1, \cdots, X_n = x_n\} = \prod_{i=1}^{n} p(x_i; \theta).$$

若 X 为连续型随机变量, 它的密度为 $f(x;\theta)$, 此时似然函数为

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

3.1 注. *大多情况下, 似然函数 L 关于 θ 可微, 于是极值点可由似然方程*

$$\frac{\mathrm{d}}{\mathrm{d}\theta}L(X_1,\ldots,X_n;\theta)=0$$

解出驻点, 再得到最大值点. 特别当驻点唯一时, 最大值点只能是此驻点. 进一步地, 由于似然函数 L 和它的对数函数 $\ln L$ 同时取得最值, 实际计算中我们也常用对数似然方程

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\ln L(X_1,\ldots,X_n;\theta)=0$$

解出 (对数可以把乘积转换为求和的形式, 便于求导, 特别对正态分布、指数分布这种密度包含指数函数的情况, 能够极大的方便我们的计算).

若有多个未知参数 $\theta_1,\ldots,\theta_m$, 则从似然方程组 $\frac{\partial}{\partial\theta_i}L$, $i=1,\ldots,m$ 或者对数似然方程组解出驻点, 再判断驻点是否为最大值点.

估计量的评选标准

由于不同方法可能得到不同的点估计, 我们需要对不同的估计量做出评价. 常用的标准有**无偏性**、**有效性**及**相合性**等, 这需要点估计量分别满足如下要求:

- **无偏性**: 称 $\hat\theta$ 是 θ 的无偏估计量, 如果 $\mathbb{E}(\hat\theta)=\theta$;
- **有效性**: 称无偏估计量 $\hat\theta_1$ 比无偏估计量 $\hat\theta_2$ 更有效, 如果对任意 $\theta\in\Theta$, 都有 $D(\hat\theta_1)\le D(\hat\theta_2)$ 且存在某个 θ 使得不等号严格成立;
- **相合性**: 称 $\hat\theta=\hat\theta(X_1,\ldots,X_n)$ 为参数 θ 的相合估计量, 如果 $\hat\theta\xrightarrow{\mathbb{P}}\theta$.

3.2 注. *由辛钦大数定律, 样本的二阶中心矩是方差的相合估计, 但不是无偏估计——由命题2.2, 样本方差才是总体方差的无偏估计, 此外样本方差也是总体方差的相合估计.*

3.1.2 区间估计

由于时间缘故, 本节未能来得及准备, 私密马赛.

3.2 例题示范

3.3 示例. 设 x_1, x_2, x_3 是来自总体的样本, 是证明下列统计量都是总体均值 μ 的无偏估计, 并比较方差存在时的有效性.

(1) $\hat{\mu}_1 = \frac{1}{2}x_1 + \frac{1}{3}x_2 + \frac{1}{6}x_3$;

(2) $\hat{\mu}_2 = \frac{1}{3}x_1 + \frac{1}{3}x_2 + \frac{1}{3}x_3$;

(3) $\hat{\mu}_3 = \frac{1}{6}x_1 + \frac{1}{6}x_2 + \frac{2}{3}x_3$.

解.

$$\begin{aligned}
\mathbb{E}\left(\hat{\mu}_1\right) &= \frac{1}{2}\mathbb{E}\left(x_1\right) + \frac{1}{3}\mathbb{E}\left(x_2\right) + \frac{1}{6}\mathbb{E}\left(x_3\right) = \frac{1}{2}\mu + \frac{1}{3}\mu + \frac{1}{6}\mu = \mu, \\
\mathbb{E}\left(\hat{\mu}_2\right) &= \frac{1}{3}\mathbb{E}\left(x_1\right) + \frac{1}{3}\mathbb{E}\left(x_2\right) + \frac{1}{3}\mathbb{E}\left(x_3\right) = \frac{1}{3}\mu + \frac{1}{3}\mu + \frac{1}{3}\mu = \mu, \\
\mathbb{E}\left(\hat{\mu}_3\right) &= \frac{1}{6}\mathbb{E}\left(x_1\right) + \frac{1}{6}\mathbb{E}\left(x_2\right) + \frac{2}{3}\mathbb{E}\left(x_3\right) = \frac{1}{6}\mu + \frac{1}{6}\mu + \frac{2}{3}\mu = \mu.
\end{aligned}$$

于是它们都是 μ 的无偏估计. 不妨设总体方差为 σ^2, 那么

$$\begin{aligned}
D\left(\hat{\mu}_1\right) &= \frac{1}{4}D\left(x_1\right) + \frac{1}{9}D\left(x_2\right) + \frac{1}{36}D\left(x_1\right) = \frac{1}{4}\sigma^2 + \frac{1}{9}\sigma^2 + \frac{1}{36}\sigma^2 = \frac{7}{18}\sigma^2, \\
D\left(\hat{\mu}_3\right) &= \frac{1}{9}D\left(x_1\right) + \frac{1}{9}D\left(x_2\right) + \frac{1}{9}D\left(x_3\right) = \frac{1}{9}\sigma^2 + \frac{1}{9}\sigma^2 + \frac{1}{9}\sigma^2 = \frac{1}{3}\sigma^2, \\
D\left(\hat{\mu}_3\right) &= \frac{1}{36}D\left(x_1\right) + \frac{1}{36}D\left(x_2\right) + \frac{4}{9}D\left(x_3\right) = \frac{1}{36}\sigma^2 + \frac{1}{36}\sigma^2 + \frac{4}{9}\sigma^2 = \frac{1}{2}\sigma^2.
\end{aligned}$$

不难看出 $D(\hat{\mu}_2) < D(\hat{\mu}_1) < D(\hat{\mu}_3)$, 于是 $\hat{\mu}_2$ 最有效. □

3.4 示例. 设总体密度函数如下, $x_1, \ldots, x_n$ 为来自总体的样本, 试求未知参数的矩估计:

(1) $p(x;\theta) = (\theta+1)x^\theta$, $0 < x < 1$, $\theta > 0$;

(2) $p(x;\theta,\mu) = \frac{1}{\theta}e^{-\frac{x-\mu}{\theta}}$, $x > \mu$, $\theta > 0$.

解. (1) 由 $\mathbb{E}(X) = \int_0^1 x \cdot p(x;\theta)\,\mathrm{d}x = \frac{\theta+1}{\theta+2}$, 于是 $\theta = \frac{1-2\mathbb{E}(X)}{\mathbb{E}(X)-1}$, 从而由替换原理, 矩估计 $\hat{\theta} = \frac{1-2\bar{x}}{\bar{x}-1}$;

(2) 总体的期望和方差分别为

$$\mathbb{E}(X)=\int_{\mu}^{\infty} x\cdot p(x;\theta)\,\mathrm{d}x=\int_0^{\infty}\frac{t}{\theta}e^{-\frac{t+\mu}{\theta}}\,\mathrm{d}t=\theta+\mu,$$
$$\mathbb{E}(X^2)=\int_{\mu}^{\infty} x^2\cdot p(x;\theta)\,\mathrm{d}x=\int_0^{\infty}\frac{(t+\mu)^2}{\theta}e^{-\frac{t}{\theta}}\,\mathrm{d}t=2\theta^2+2\mu\theta+\mu^2,$$
$$D(X)=\mathbb{E}(X^2)-\mathbb{E}(X)^2=\theta^2.$$

于是 $\theta=\sqrt{D(X)}$, $\mu=\mathbb{E}(X)-\sqrt{D(X)}$, 从而由替换原理, 矩估计 $\hat{\theta}=\sqrt{\frac{1}{n}\sum_{i=1}^n(x_i-\bar{x})^2}$, $\hat{\mu}=\bar{x}-\sqrt{\frac{1}{n}\sum_{i=1}^n(x_i-\bar{x})^2}$.

□

3.5 示例 *(均匀分布的矩估计)***.** *设样本 $X_1,\ldots,X_n$ 来自均匀分布 $U(a,b)$, 给出未知参数 a,b 的矩估计.*

解. 由于 $\mathbb{E}(X)=\frac{a+b}{2}$, $D(X)=\frac{(b-a)^2}{12}$, 不难得出

$$a=\mathbb{E}(X)-\sqrt{3D(X)},\ b=\mathbb{E}(X)+\sqrt{3D(X)}.$$

我们用一阶样本原点矩和二阶样本中心矩分别替换 $\mathbb{E}(X)$ 和 $D(X)$ 即可:

$$\hat{a}=\overline{X}-\sqrt{\frac{3}{n}\sum_{i=1}^n(X_i-\overline{X})^2},\ \hat{b}=\overline{X}+\sqrt{\frac{3}{n}\sum_{i=1}^n(X_i-\overline{X})^2}.$$

□

3.6 示例 *(均匀分布的最大似然估计)***.** *参照教材例 6.*

3.7 示例. *设总体概率密度如下, $x_1,\ldots,x_n$ 为来自总体的样本, 试求未知参数的最大似然估计:*

(1) $p(x;\theta)=\sqrt{\theta}x^{\sqrt{\theta}-1}$, $0<x<1$, $\theta>0$;

(2) $p(x;\theta,c)=\frac{1}{\theta}e^{-\frac{x-c}{\theta}}$, $x>c$, $\theta>0$.

解. (1) 似然函数为 $L(\theta)=\theta^{\frac{n}{2}}(x_1x_2\cdots x_n)^{\sqrt{\theta}-1}$, 于是对数似然函数为

$$\ln L(\theta)=\frac{n}{2}\ln\theta+(\sqrt{\theta}-1)\sum_{i=1}^n\ln x_i.$$

于是似然方程为

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\ln L(\theta)=\frac{n}{2\theta}+\frac{\sum_{i=1}^n\ln x_i}{2\sqrt{\theta}}=0,$$

解之得唯一驻点 $\hat{\theta} = \left(\frac{1}{n}\sum_{i=1}^{n}\ln x_i\right)$.

(2) 记 $x_{(1)} = \min\{x_1, \ldots, x_n\}$, 此时似然函数为

$$L(\theta) = \theta^{-n}\exp\left\{-\frac{1}{\theta}\sum_{i=1}^{n}(x_i - c)\right\},\ c < x_{(1)},$$

注意到对数似然函数

$$\ln L(\theta) = -n\ln\theta - \frac{1}{\theta}\sum_{i=1}^{n}(x_i - c),\ c < x_{(1)}$$

关于 c 单调增, 要使似然函数达到最大, c 应在限制范围 $c < x_{(1)}$ 内尽可能取大, 于是最大似然估计 $\hat{c} = x_{(1)}$. 另一方面, 我们令 $\ln L(\theta, \hat{c})$ 关于 θ 的导数为 0 解出似然方程得

$$\hat{\theta} = \frac{\sum_{i=1}^{n}(x_i - \hat{c})}{n} = \bar{x} - x_{(1)}.$$

□