数理统计笔记

作者：徐知南而北游
时间：2025 年 7 月 7 日

目录

Chapter 1

统计量及其分布

§1.1 总体和样本

- **总体**: 所要研究对象的全体被称为总体，总体中的元素被称为个体
- **样本**: 从总体中随机抽取 n 个个体，记其指标为 $X_1,\cdots,X_n$ ，则称 $X_1,\cdots,X_n$ 是来自总体的一个样本，其中 n 称为样本量。

【命题 1.1】 设总体 X 的概率（密度）函数为 $f(x)$（分布函数为 $F(x)$ ），则

1. $X_1,\cdots,X_n$ 独立同分布于 X ，记为 $X_1,\cdots,X_n \sim f(x)$ 。
2. $X_1,\cdots,X_n$ 的联合概率函数为
$$f\left(x_1,\cdots,x_n\right)=f_{X_1}\left(x_1\right)\cdots f_{X_n}\left(x_n\right)=f\left(x_1\right)\cdots f\left(x_n\right)=\prod_{i=1}^{n} f\left(x_i\right).$$
3. $X_1,\cdots,X_n$ 的联合分布函数
$$F\left(x_1,\cdots,x_n\right)=F_{X_1}\left(x_1\right)\cdots F_{X_n}\left(x_n\right)=F\left(x_1\right)\cdots F\left(x_n\right)=\prod_{i=1}^{n} F\left(x_i\right).$$

【例题 1.2】 设总体 $X\sim N\left(\mu,\sigma^2\right)$, $X_1,\cdots,X_n$ 是来自总体 X 的一组样本，求 $X_1,\cdots,X_n$ 的联合概率函数。

解. 由于
$$f(x)\sim N\left(\mu,\sigma^2\right) \implies f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}$$
$$\implies f\left(x_1,\cdots,x_n\right)=\prod_{i=1}^{n} f\left(x_i\right)=\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

□

§1.2 统计量

- **统计量**: 是样本的函数，且不含任何未知参数. 是可以由样本算出的量（完全由样本决定的量），通常用 $T=T\left(X_1,\cdots,X_n\right)$ 表示.

【例题 1.3】设 $X_1, \cdots, X_n$ 是来自 $N\left(\mu, \sigma^2\right)$ 的样本,其中 μ 已知,σ^2 未知.令 $X_{(1)} = \min\left[X_1, \cdots, X_n\right], X_{(n)} = \max\left[X_1, \cdots, X_n\right]$，判断

1. $X_{(n)} - X_{(1)}$ 和 $\frac{1}{2}\left(X_{(n)} + X_{(1)}\right)$ 是否为统计量
2. $\frac{1}{2}\left(X_{(n)} + X_{(1)}\right) - \mu$ 和 $\frac{1}{\sigma}\left(X_{(n)} - X_{(1)}\right)$ 是否为统计量.

1.2.1 样本均值

【定义 1.4】(样本均值) 设 $X_1, \cdots, X_n$ 是来自 $f(x;\theta)$ 的样本，称

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

为样本均值.

【定理 1.5】 设总体 X 的均值 $\mathrm{E}(X) = \mu$，方差 $\mathrm{var}(X) = \sigma^2 < +\infty$，令 $X_1, \cdots, X_n$ 是来自 X 的随机样本，样本均值为

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} x_i,$$

则

$$\mathrm{E}\left(\bar{X}_n\right) = \mu, \quad \mathrm{var}\left(\bar{X}_n\right) = \sigma^2/n.$$

证明. 直接计算得

$$\mathrm{E}\left(\bar{X}_n\right) = \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\mathrm{E}\left(\sum_{i=1}^{n} X_i\right) = \mu$$

方差同理 □

【注 1.6】 1. 根据辛钦大数定律

$$\lim_{n\to\infty} P\left(\left|\bar{X}_n - \mu\right| < \epsilon\right) = 1, \text{即} \bar{X}_n \xrightarrow{p} \mu$$

2. 根据 Chebyshev 不等式 (假设二阶矩存在),

$$P\left(\left|\bar{X}_n - \mu\right| < \epsilon\right) \geq 1 - \frac{\sigma^2}{n\epsilon^2}.$$

【例题 1.7】 假设某一分布的均值未知，方差为 1，则为了使得样本均值 $\bar{X}_n$ 和总体均值 μ 的差异以至少 0.95 的概率在 0.5 之内，则至少需要多少样本？

解. 使用切比雪夫不等式

$$P(|\bar{X}_n - \mu| \geqslant \varepsilon) \leqslant \frac{\sigma^2/n}{\varepsilon^2}$$

代入 $\sigma^2 = 1, \varepsilon = 0.5$:

$$P(|\bar{X}_n - \mu| \geqslant 0.5) \leqslant \frac{1/n}{(0.5)^2} = \frac{4}{n}$$

$\Longrightarrow$

$$\frac{4}{n} \leqslant 0.05 \implies n \geqslant \frac{4}{0.05} = 80$$

故最小样本量为 80.

□

对于一些具有可加性的分布，可以精确计算其样本均值的分布，见下例

【例题 1.8】 分别计算下列分布对应的样本均值 $\bar{X}$ 的精确分布

1. 伯努利分布 $B(1, p)$

2. 泊松分布 $P(\lambda)$
3. 指数分布 $\exp(\theta)$
4. 柯西分布 $\mathrm{Cauchy}(a,b)$
5. 正态分布 $\mathrm{N}(\mu,\sigma^2)$

解.

1. 注意到
$$X_k \sim B(1,p) \quad iid \implies \sum X_k \sim B(n,p)$$
$\implies$
$$P\left(\sum_{i=1}^n X_i = k\right) = \binom{n}{k} p^k (1-p)^{n-k} \quad k=0,\ldots,n.$$
$\implies$
$$P\left(\bar{X}_n = \frac{k}{n}\right) = \binom{n}{k} p^k (1-p)^{n-k} \quad k=0,\cdots,n.$$

2. 注意到
$$X_k \sim P(\lambda) \quad iid \implies \sum X_k \sim P(n\lambda)$$
$\implies$
$$P\left(\sum_{i=1}^n X_i = k\right) = \frac{(n\lambda)^k}{k!} e^{-n\lambda}, \quad \forall k \in \mathbb{Z} > 0$$
$\implies$
$$P\left(\bar{X}_n = \frac{k}{n}\right) = P\left(\sum_{i=1}^n X_i = k\right) = \frac{(n\lambda)^k}{k!} e^{-n\lambda}$$

3. 注意到
$$X_k \sim \exp(\theta) \quad iid \implies Y := \sum X_k \sim Ga(n,\theta)$$
$\implies$
$$f_Y(y) = \frac{\theta^n}{\Gamma(n)} y^{n-1} \exp(-\theta y) \quad y > 0$$
$\implies$
$$f_{\bar{X}}(x) = f_Y(nx)n \sim Ga(n, n\theta)$$

4.

5. 注意到
$$X_k \sim N(\mu,\sigma^2) \quad iid \implies \sum X_k \sim N(n\mu, n\sigma^2) \implies \bar{X}_n = \frac{1}{n}\sum X_k \sim N\left(\mu, \sigma^2/n\right).$$

□

对于一般的分布，仍然可以考虑其样本均值的渐进分布 (CLT)

【定理 1.9】(中心极限定理) 设 $f(x)$ 是一个具有均值 μ 和方差 σ^2 的概率密度函数。设 $\bar{X}_n$ 为来自 $f(x)$ 的大小为 n 的随机样本的样本均值。定义 Z_n
$$Z_n = \frac{\bar{X}_n - \mu}{\sqrt{\mathrm{var}(\bar{X}_n)}} = \frac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}}$$
则有
$$Z_n \xrightarrow{d} N(0,1) \iff \bar{X}_n \xrightarrow{d} N\left(\mu, \frac{\sigma^2}{n}\right)$$

【例题 1.10】 下图对比了 $n=1,5,30,100$ 时，总体 $X\sim\exp(1)$ 的渐进分布和精确分布的差异

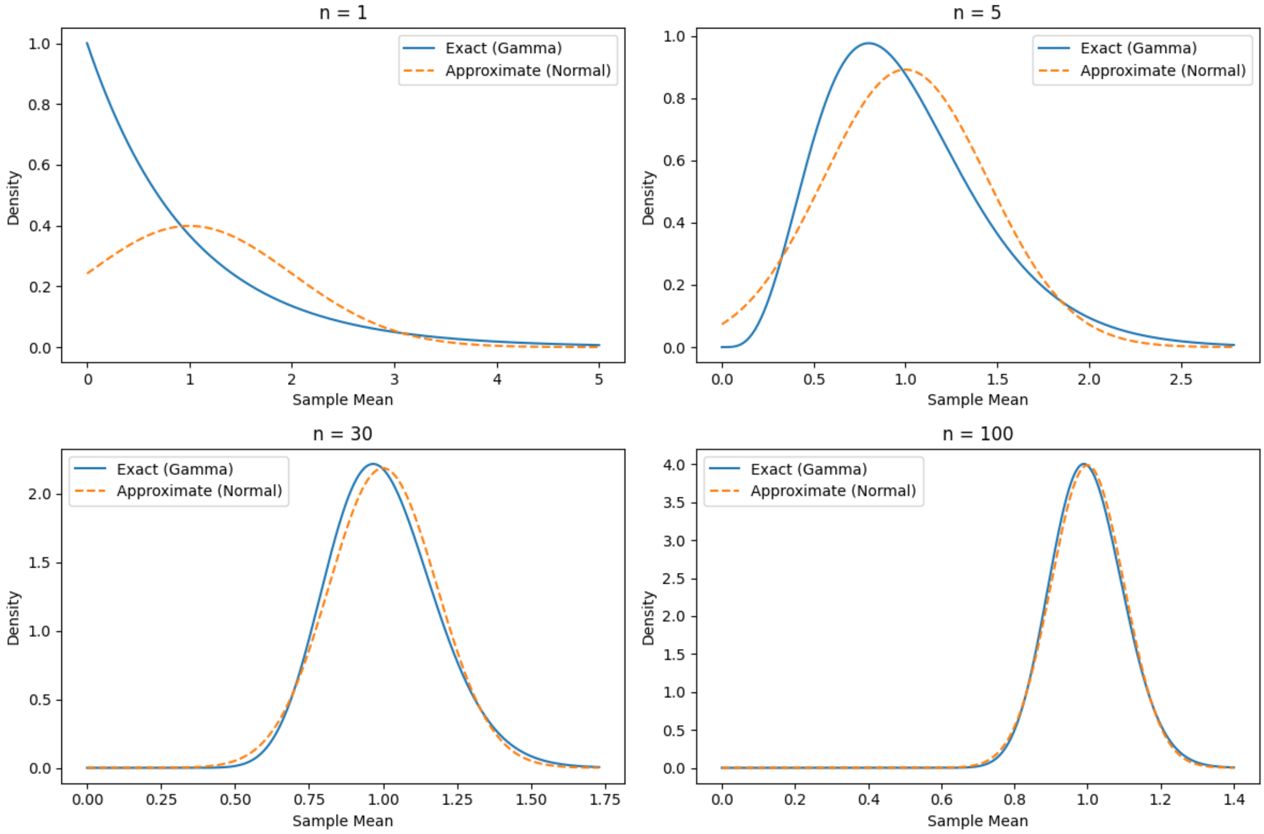

图 1.1: 指数分布

1.2.2 样本方差与样本矩

【定义 1.11】(样本方差) 设 $X_1,\cdots,X_n$ 是来自 $f(x)$ 的一个样本，且 $\bar{X}_n$ 为样本均值。则称

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2\quad(n>1)$$

为样本方差。

【定理 1.12】 设 $X_1,\cdots,X_n$ 是来自 $f(x)$ 的样本，且

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2.$$

则

$$E\left(S^2\right)=\sigma^2,\quad \operatorname{var}\left(S^2\right)=\frac{1}{n}\left(\mu_4-\frac{n-3}{n-1}\sigma^4\right)\quad n>1$$

其中 $\sigma^2=\operatorname{var}(X)$ 为总体方差，$\mu_4=E\left[(X-E(X))^4\right]$ 为总体四阶中心矩。

【注 1.13】 关于样本方差的计算大多要用到恒等变形

$$\sum_{i=1}^{n}(X_i-\bar{X}_n)^2=\sum_{i=1}^{n}X_i^2-n\bar{X}_n{}^2$$

以及

$$\sum_{i=1}^{n}(X_i-\bar{X}_n)^2=\sum_{i=1}^{n}(X_i-\mu)^2-n\left(\bar{X}_n-\mu\right)^2$$

证明.

1. 直接计算得

$$
\begin{aligned}
(n-1)\mathbb{E}\left(S^2\right) &= \mathbb{E}\left(\sum_{i=1}^{n} x_i^2 - n\bar{X}_n^2\right) = n\mathbb{E}\left(\sum_{i=1}^{n} x_i^2\right) - n\mathbb{E}\left(\bar{X}_n^2\right) \\
&= n(\mu+\sigma^2) - n\left(\mu + \frac{\sigma^2}{n}\right) = (n-1)\sigma^2
\end{aligned}
$$

两边消去 $n-1$ 即可

2. 我们考虑中心化变换 $Y_i = X_i - \mu$ 则由注解知

$$
S^2 = \frac{1}{n-1}\left(\sum y_i^2 - n\bar{y}^2\right)
$$

则

$$
\begin{aligned}
(n-1)^2 E\left[\left(s^2\right)^2\right] &= E\left[\left(\sum y_i^2 - n\bar{y}^2\right)^2\right] \\
&= E\left(\sum y_i^2\right)^2 - 2nE\left(\sum y_i^2\bar{y}^2\right) + n^2E\left(\bar{y}^4\right) \\
&= E\left(\sum y_i^2\right)^2 - 2n^2E\left(y_1^2\bar{y}^2\right) + n^2E\left(y_1\bar{y}^3\right) \\
&= E\left(\sum y_i^2\right)^2 - 2E\left(y_1^2\left(\sum y_i\right)^2\right) + \frac{1}{n}E\left[y_1\left(\sum y_i\right)^3\right]
\end{aligned}
$$

分别计算得

$$
\begin{aligned}
E\left(\sum y_i^2\right)^2 &= E\left(\sum y_i^4 + 2\sum_{i<j} y_i^2 y_j^2\right) = n\mu_4 + n(n-1)\sigma^4 \\
E\left[y_1^2\left(\sum y_i\right)^2\right] &= E\left(y_1^4 + \sum_{i\neq 1} y_1^2 y_i^2\right) = \mu_4 + (n-1)\sigma^4 \quad \text{独立} \implies \text{交叉项期望为0} \\
E\left[y_1\left(\sum y_i\right)^3\right] &= E\left(y_1^4\right) + \binom{3}{1}E\left(y_1^2\sum_{i\neq 1} y_i^2\right) \quad \text{独立} \implies \text{其他交叉项期望为0} \\
&= \mu_4 + 3(n-1)\sigma^4
\end{aligned}
$$

相加并比较系数得

$$
\begin{aligned}
(n-1)^2E\left[\left(s^2\right)^2\right] &= \left(n-2+\frac{1}{n}\right)\mu_4 + \left(n^2-n-2n+2+3-\frac{3}{n}\right)\sigma^4 \\
&= \frac{(n-1)^2}{n}\mu_4 + \left[(n-1)^2 - \frac{n^2-4n+3}{n}\right]\sigma^4 \\
\implies \operatorname{Var}\left(s^2\right) &= E\left[\left(s^2\right)^2\right] - \sigma^4 \\
&= \frac{\mu_4}{n} - \frac{(n-1)(n-3)}{n(n-1)^2}\sigma^4 \\
&= \frac{\mu_4}{n} - \frac{n-3}{n(n-1)}\sigma^4
\end{aligned}
$$

□

1.2.3 次序统计量

【定义 1.14】(次序统计量) 设 $X_1, \cdots, X_n$ 是来自累积分布函数 $F(x)$ 的容量为 n 的随机样本，则

$$
Y_1 \leq \cdots \leq Y_n
$$

或

$$X_{(1)} \leq \cdots \leq X_{(n)}$$

【注 1.15】 $Y_1, \cdots, Y_n$ 是统计量，且为随机样本 $X_1, \cdots, X_n$ 的函数，但 $Y_1, \cdots, Y_n$ 并不独立。若 $Y_i > y$，则 $Y_j > y$ 对所有 $j = i+1, \cdots, n$ 成立。

次序统计量的边缘分布

【例题 1.16】 最小次序统计量 $Y_1 = \min\{X_1, \cdots, X_n\}$ 的分布函数和密度函数

解.

1. 分布函数

$$\begin{aligned} F_{Y_1}(y) &= P(Y_1 \leq y) = 1 - P(Y_1 > y) = 1 - P(\min\{X_1, \cdots, X_n\} > y) \\ &= 1 - P(X_1 > y, \cdots, X_n > y) = 1 - \prod_{i=1}^{n} P(X_i > y) \\ &= 1 - [1 - F_X(y)]^n . \end{aligned}$$

2. 密度函数

$$f_{Y_1}(y) = n[1 - F_X(y)]^{n-1} f_X(y).$$

□

【例题 1.17】 最大次序统计量 $Y_n = \max\{X_1, \cdots, X_n\}$ 的分布函数和密度函数

解.

1. 分布函数

$$\begin{aligned} F_{Y_n}(y) &= P(Y_n \leq y) = P(\max\{X_1, \cdots, X_n\} \leq y) \\ &= \prod_{i=1}^{n} P(X_i \leq y) = [F_X(y)]^n \end{aligned}$$

2. 密度函数

$$f_{Y_n}(y) = nF_X(y)^{n-1} f_X(y).$$

□

次序统计量的联合分布

1.2.4 充分统计量

【定义 1.18】(充分统计量) 设 $(x_1, x_2, \cdots, x_n)$ 是总体 X 的样本，其分布函数为 $F(x;\theta)$ 如果有

$$P_\theta((x_1 \cdots x_n)|T = t) = P((x_1 \cdots x_n)|T = t)$$

则称 $T(x_1 \cdots x_n)$ 为 θ 的充分统计量

【例题 1.19】 设 $X = (X_1, X_2, \cdots, X_n)$ 为从 0-1 分布中抽取的简单样本，则 $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$ 为充分统计量.

解. 直接计算

$$\begin{aligned} P(X_1 = x_1, \cdots, X_n = x_n \mid T = t_0) &= \frac{P(X_1 = x_1, \cdots, X_n = x_n, T = t_0)}{P(T = t_0)} \\ = \frac{P\left(X_1 = x_1, \cdots, X_n = t_0 - \sum_{i=1}^{n-1} x_i\right)}{P(T = t_0)} &= \frac{\theta^{t_0}(1-\theta)^{n-t_0}}{\binom{n}{t_0}\theta^{t_0}(1-\theta)^{n-t_0}} = \frac{1}{\binom{n}{t_0}}, \end{aligned}$$

因此有

$$P\left(X_1=x_1,\cdots,X_n=x_n \mid T=t_0\right)=\begin{cases}\dfrac{1}{\binom{n}{t_0}}, & \sum\limits_{i=1}^n x_i=t_0\\ 0, & \sum\limits_{i=1}^n x_i\neq t_0\end{cases}$$

上述条件概率与 θ 无关，因此 $T(\boldsymbol{X})=\sum\limits_{i=1}^n X_i$ 为 θ 的充分统计量. □

【例题 1.20】 设 $X=(X_1,X_2,\cdots,X_n)$ 为从正态总体 $N(\theta,1)$ 中抽取的简单样本，则 $T(\boldsymbol{X})=\sum\limits_{i=1}^n X_i/n=\bar{X}$ 为 θ 的充分统计量.

解. 作正交变换

$$(Y_1,\cdots,Y_n)'=\boldsymbol{A}\left(X_1,\cdots,X_n\right)'$$

其中正交阵

$$\boldsymbol{A}=\begin{pmatrix}\dfrac{1}{\sqrt{n}} & \dfrac{1}{\sqrt{n}} & \cdots & \dfrac{1}{\sqrt{n}}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & & \vdots\\ a_{n1} & a_{n2} & \cdots & a_{nn}\end{pmatrix}$$

由 Schmidt 正交化方法可知上述正交阵 $\boldsymbol{A}$ 是存在的. 变换的 Jacobi 行列式的绝对值为 $|J|=1$ 。

$$\begin{cases}Y_1=\dfrac{1}{\sqrt{n}}\sum\limits_{i=1}^n X_i=\sqrt{n}\bar{X}\\ Y_j=\sum\limits_{k=1}^n a_{jk}X_k, \quad j=2,\cdots,n\end{cases}$$

由于 $\sum\limits_{i=1}^n Y_i^2=\sum\limits_{i=1}^n X_i^2$ ，且 $Y_1,Y_2,\cdots,Y_n$ 是相互独立的，

$$Y_1\sim N(\sqrt{n}\theta,1), Y_i\sim N(0,1), i=2,\cdots,n$$

$\bar{X}$ 对原样本 $\boldsymbol{X}$ 的充分性等价于 Y_1 对 $(Y_1,Y_2,\cdots,Y_n)$ 的充分性. 因此只要证明给定 $Y_1=y_1$ 时，$(Y_1,Y_2,\cdots,Y_n)$ 的条件密度与 θ 无关，易见 $Y_1,\cdots,Y_n$ 的联合密度为

$$f\left(y_1,\cdots,y_n\right)=(2\pi)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2}\sum_{i=2}^n y_i^2-\frac{1}{2}\left(y_1-\sqrt{n}\theta\right)^2\right\}$$

Y_1 的边缘密度函数为

$$f_{Y_1}\left(y_1\right)=\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(y_1-\sqrt{n}\theta\right)^2\right\}$$

给定 $Y_1=y_1$ 时，$(Y_1,\cdots,Y_n)$ 的条件密度是

$$f\left(y_1,\cdots,y_n \mid y_1\right)=\frac{f\left(y_1,\cdots,y_n\right)}{f_{Y_1}\left(y_1\right)}=(2\pi)^{-\frac{n-1}{2}}\exp\left\{-\frac{1}{2}\sum_{i=2}^n y_i^2\right\},$$

与 θ 无关 □

【例题 1.21】 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为从指数分布 $\mathrm{Exp}(\theta)$ 中抽取的简单样本，其密度函数为 $f(x,\theta)=\theta e^{-\theta x}I_{[x>0]}$ ，则 $T(\boldsymbol{X})=\sum\limits_{i=1}^n X_i$ 为 θ 的充分统计量.

解. $\boldsymbol{X}$ 的联合密度为

$$f(\boldsymbol{x},\theta)=\theta^n\exp\left\{-\theta\sum_{i=1}^n x_i\right\}I_{[x_i>0,i=1,2,\cdots,n]},$$

作变换

$$\left\{\begin{array}{l} Y_1 = X_1 \\ \quad \cdots\cdots \\ Y_{n-1} = X_{n-1} \\ Y_n = X_1 + X_2 + \cdots + X_n = T \end{array}\right.$$

变换的Jacobi行列式的绝对值为 $|J| = 1$ 显然, $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$ 对原样本 $\boldsymbol{X}$ 的充分性等价于 $Y_n = T$ 对 $(Y_1, Y_2, \cdots, Y_n)$ 的充分性.因此只要证明给定 $Y_n = y_n$(即 $T = t$)时, $(Y_1, Y_2, \cdots, Y_n)$ 的条件密度与 θ 无关. 易见 $(Y_1, \cdots, Y_{n-1}, Y_n)$ 的联合密度为

$$f(y_1, y_2, \cdots, y_{n-1}, t, \theta) = \theta^n e^{-\theta t} I_{[y_i>0, i=1,2,\cdots,n-1, y_1+\cdots+y_{n-1}<t]}$$

由于 $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i \sim \Gamma(n, \theta)$ ，因此 $T = T(\boldsymbol{X})$ 有密度函数

$$f_T(t) = \frac{\theta^n}{\Gamma(n)} t^{n-1} e^{-\theta t} I_{[t>0]} = \frac{\theta^n}{(n-1)!} t^{n-1} e^{-\theta t} I_{[t>0]}$$

给定 $T = t$ 时，$(Y_1, \cdots, Y_n)$ 的条件密度为

$$\begin{aligned} f(y_1, \cdots, y_n \mid t) &= \frac{f(y_1, \cdots, y_{n-1}, t)}{f_T(t)} \\ &= \frac{(n-1)!\theta^n e^{-\theta t} \cdot I_{[y_i>0, i=1,\cdots,n-1, y_1+\cdots+y_{n-1}<t]}}{\theta^n t^{n-1} e^{-\theta t} I_{[t>0]}} \\ &= \frac{(n-1)! I_{[y_i>0, i=1,\cdots,n-1, y_1+\cdots+y_{n-1}<t]}}{t^{n-1} I_{[t>0]}} \end{aligned}$$

□

【定理 1.22】(因子分解) 设样本 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 的概率函数 $f(\boldsymbol{x}, \theta)$ 依赖于参数 $\theta, T = T(X)$ 是一个统计量，则 T 为充分统计量的充要条件是 $f(x, \theta)$ 可以分解为

$$f(\boldsymbol{x}, \theta) = g(t(\boldsymbol{x}), \theta) h(\boldsymbol{x})$$

的形状. 注意此处函数 $h(\boldsymbol{x}) = h(x_1, \cdots, x_n)$ 不依赖于 $\theta, t(\boldsymbol{x})$ 为 $T(\boldsymbol{X})$ 的观察值.

【推论 1.23】 设 $\boldsymbol{T} = T(\boldsymbol{X})$ 为 θ 的充分统计量，$S = \varphi(\boldsymbol{T})$ 是单值可逆函数，则 $S = \varphi(\boldsymbol{T})$ 也是 θ 的充分统计量.

【注 1.24】 若 $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$ 是充分统计量，则 $\bar{X}$ 也是充分统计量.

【例题 1.25】 设 $\boldsymbol{X} = (X_1, \cdots, X_n)$ 为从正态总体 $N\left(a, \sigma^2\right)$ 中抽取的简单样本，令 $\theta = \left(a, \sigma^2\right)$ ，则 $\left(\bar{X}, S^2\right)$ 为 θ 的充分统计量，此处 $\bar{X}, S^2$ 分别为样本均值和样本方差.

解. 样本 X 的联合密度为

$$\begin{aligned} f(\boldsymbol{x}, \theta) &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - a)^2\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2} \left(\sum_{i=1}^{n} x_i^2 - 2a \sum_{i=1}^{n} x_i + na^2\right)\right\} \\ &= g(t(\boldsymbol{x}), \theta) \cdot h(\boldsymbol{x}) \end{aligned}$$

此处 $h(\boldsymbol{x}) \equiv 1$ ，由因子分解定理可知 $T(\boldsymbol{X}) = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ 为 θ 的充分统计量. 由于 $\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ 与 $\left(\bar{X}, S^2\right)$ 为一一对应的变换，由推论可知 $\left(\bar{X}, S^2\right)$ 也是 θ 的充分统计量. □

解. 将平方项展开并分解:

$$\sum_{i=1}^{n}(x_i-\mu)^2=\sum_{i=1}^{n}(x_i-\bar{X})^2+n(\bar{X}-\mu)^2,$$

代入联合密度函数得:

$$f(x,\mu,\sigma^2)=(2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{n(\bar{X}-\mu)^2+(n-1)S^2}{2\sigma^2}\right\}.$$

进一步分解为:

$$f(x,\mu,\sigma^2)=\underbrace{(2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{n(\bar{X}-\mu)^2}{2\sigma^2}\right\}\exp\left\{-\frac{(n-1)S^2}{2\sigma^2}\right\}}_{g(\bar{X},S^2;\mu,\sigma^2)}\cdot\underbrace{1}_{h(x)}.$$

□

【例题 1.26】 设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为从总体 $b(1,\theta)$ 中抽取的简单样本，则 $T(\boldsymbol{X})=\sum_{i=1}^{n}X_i$ 是 θ 的充分统计量.

解. 样本 X 的联合分布是

$$\begin{aligned}f(\boldsymbol{x},\theta)&=P_\theta\left(X_1=x_1,\cdots,X_n=x_n\right)\\&=\theta^{\sum_{i=1}^{n}x_i}(1-\theta)^{n-\sum_{i=1}^{n}x_i}=g(t(\boldsymbol{x}),\theta)h(\boldsymbol{x})\end{aligned}$$

其中 $h(\boldsymbol{x})\equiv 1$，由因子分解定理可知 $T(\boldsymbol{X})=\sum_{i=1}^{n}X_i$ 为 θ 的充分统计量. □

【例题 1.27】(次序统计量的充分性) 设 $\mathscr{F}=\{F\}$ 为一维分布族，这里对分布函数 F 没有任何限制. 设 $X=(X_1,\cdots,X_n)$ 是从某个 F 中抽出的简单样本，$T(\boldsymbol{X})=\left(X_{(1)},X_{(2)},\cdots,X_{(n)}\right)$ 为次序统计量，则次序统计量 $T(\boldsymbol{X})$ 是充分的.

解. 特别地，若 F 有密度 f，则

$$f\left(x_1,\cdots,x_n\right)=f\left(x_1\right)\cdots f\left(x_n\right)=f\left(x_{(1)}\right)\cdots f\left(x_{(n)}\right)h(\boldsymbol{x}),$$

其中 $h(\boldsymbol{x})\equiv 1,f$ 起着参数 θ 的作用，则由因子分解定理可知 $T(\boldsymbol{X})=\left(X_{(1)},\cdots,X_{(n)}\right)$ 为充分统计量. □

【例题 1.28】设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为从均匀分布 $U(0,\theta)$ 中抽取的简单样本，则 $T(\boldsymbol{X})=X_{(n)}=\max\{X_1,\cdots,X_n\}$ 为 θ 的充分统计量.

解. 样本 $\boldsymbol{X}$ 的联合密度为

$$f(\boldsymbol{x},\theta)=\frac{1}{\theta^n}I_{(0,\theta)}\left(x_{(n)}\right)=g(t(\boldsymbol{x}),\theta)h(\boldsymbol{x}),$$

其中 $h(\boldsymbol{x})\equiv 1$. 由因子分解定理可知 $T(\boldsymbol{X})=X_{(n)}$ 为 θ 的充分统计量. □

【例题 1.29】 若 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为从指数族中抽取的简单样本，则 $T(\boldsymbol{X})=(T_1(\boldsymbol{X}),\cdots,T_k(\boldsymbol{X}))$ 为充分统计量.

解. 样本 $\boldsymbol{X}$ 的联合密度为

$$f(\boldsymbol{x},\theta)=C(\theta)\exp\left\{\sum_{i=1}^{k}Q_i(\theta)t_i(\boldsymbol{x})\right\}h(\boldsymbol{x})=g(t(\boldsymbol{x}),\theta)h(\boldsymbol{x}),$$

其中 $g(t(\boldsymbol{x}),\theta)=C(\theta)\exp\left\{\sum_{i=1}^{k}Q_i(\theta)t_i(\boldsymbol{x})\right\}$,$t(\boldsymbol{x})$ 为 $T(\boldsymbol{X})$ 的观察值.由因子分解定理立得 $T(\boldsymbol{X})=(T_1(\boldsymbol{X}),\cdots,T_k(\boldsymbol{X}))$ 为 θ 的充分统计量. □

【定义 1.30】 设 T 是分布族 $\mathscr{F}$ 的充分统计量，若对 $\mathscr{F}$ 的任一充分统计量 $S(\boldsymbol{X})$，存在一个函数 $q_S(\cdot)$ 使得 $T(\boldsymbol{X})=q_S(S(\boldsymbol{X}))$，则称 $T(\boldsymbol{X})$ 是此分布族的极小充分统计量.

§1.3 常见分布族

1.3.1 Gamma 分布族

【定义 1.31】(Gamma 分布) 如果 X 的密度是

$$f(x;\alpha,\lambda)=\frac{\lambda^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}\mathrm{e}^{-\lambda x}\chi_{(0,\infty)}(x),$$

就称 X 服从参数 α,λ 的 Gamma 分布，记做 $X\sim \mathrm{Ga}(\alpha;\lambda)$，其中 $\alpha>0$ 为形状参数，$\lambda>0$ 为尺度参数。

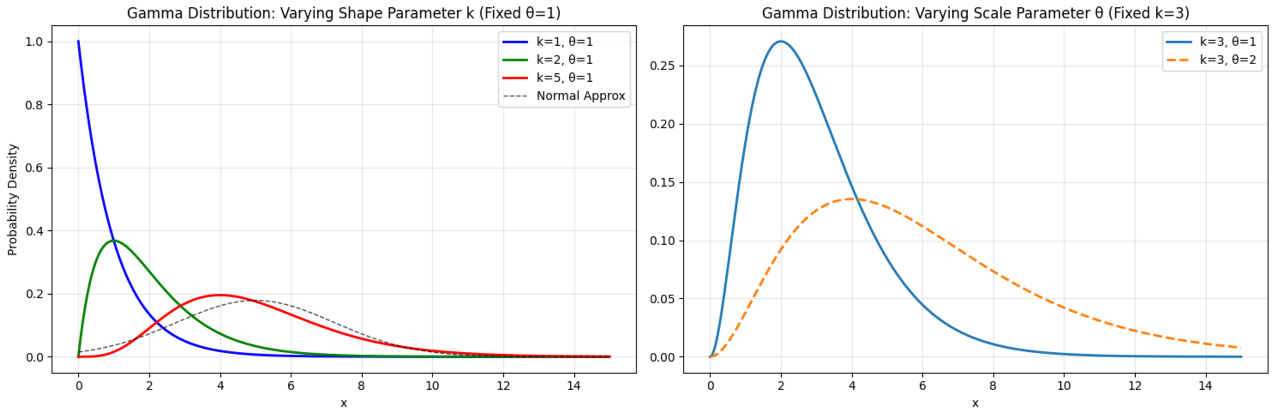

图 1.2: Caption

1.3.2 正态分布族

【定义 1.32】(多元正态分布) 设 $X\in\mathbb{R}^n$ 若 X 密度为

$$f_X(x)=\frac{1}{(2\pi)^{n/2}|\Sigma|^{-1}}\exp\left(\frac{1}{2}(X-\boldsymbol{\mu})^T\Sigma^{-1}(X-\boldsymbol{\mu})\right)$$

则称其为多元正态分布，记为 $X\sim N(\boldsymbol{\mu},\Sigma)$

【注 1.33】 考虑积分

$$\int_{\mathbb{R}n}f_X(x)\mathrm{d}x=\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\int_{\mathbb{R}n}\exp\left(\frac{1}{2}(X-\boldsymbol{\mu})^T\Sigma^{-1}(X-\boldsymbol{\mu})\right)\mathrm{d}x$$

由于 Σ 正定，做特征分解 $\Sigma=U^T\Lambda U$ 其中, $U^TU=UU^T=I,\Lambda=diag(\sigma_1^2,\cdots\sigma_n^2)$, 故

$$\int_{\mathbb{R}n}\exp\left(\frac{1}{2}(X-\boldsymbol{\mu})^T\Sigma^{-1}(X-\boldsymbol{\mu})\right)\mathrm{d}x=\int_{\mathbb{R}n}\exp\left(\frac{1}{2}Y^T\Lambda^{-1}Y\right)\mathrm{d}y$$

其中 $Y=U(X-\mu)$, 且有对应 Jacobi 行列式为 1，最后注意到 $|\Sigma|^{1/2}=\prod_{k=1}^{n}\sigma_k$

$$\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\int_{\mathbb{R}n}\exp\left(\frac{1}{2}Y^T\Lambda^{-1}Y\right)\mathrm{d}y=1$$

即原表达式的确是一个密度函数

【定理 1.34】 设 $X\sim N(\mu,\Sigma)$ 则对应特征函数 $\phi_X(\omega)=\exp(j\omega^T\mu-\frac{1}{2}\omega^T\Sigma\omega)$

证明. 直接计算得

$$\phi_X(\omega)=\mathbb{E}\left[\exp\left(j\omega^T X\right)\right]=\int_{\mathbb{R}n}\exp(j\omega^T x)f_X(x)\mathrm{d}x$$

考虑配方

$$-\frac{1}{2}(x-\mu-j\Sigma\omega)^T\Sigma^{-1}(x-\mu-j\Sigma\mu)+j\omega^T\mu-\frac{1}{2}\omega^T\Sigma\omega.$$

则原积分

$$I=\exp\left(j\omega^T\mu-\frac{1}{2}\omega^T\Sigma\omega\right)\cdot\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\int_{\mathbb{R}^n}\exp(*)dx.$$

其中 $(*)$ 是配方之后的二次型，则后面积分值必为 1，故得到特征函数

$$\boxed{\phi_X(\omega)=\exp(j\omega^T\mu-\frac{1}{2}\omega^T\Sigma\omega)}$$

□

【定理 1.35】(线性性) 设 $X\sim N_n(\mu,\Sigma)$, A 为任一 $r\times n$ 阶阵, 令 $Y=AX$，则

$$Y\sim N_r\left(A\mu,A\Sigma A^T\right)$$

证明. 直接计算 Y 的特征函数

$$\begin{aligned}\phi_Y(\omega)&=\mathbb{E}[\exp(j\omega^T AX)]=\mathbb{E}[\exp(j(A^T\omega)^T X)]=\phi_X(A^T\omega)\\&=\exp\left(j\left(A^T\omega\right)^T\mu-\frac{1}{2}\left(A^T\omega\right)^T\Sigma\left(A^T\omega\right)\right).\\&=\exp\left(j\omega^T A\mu-\frac{1}{2}\omega^T A\Sigma A^T\omega\right)\end{aligned}$$

由特征函数唯一性，显然有 $Y\sim N_r\left(A\mu,A\Sigma A^T\right)$ □

【推论 1.36】(边缘分布)

证明.

□

【注 1.37】 这个命题的逆是错的

将上面几个定理总结如下：

- 正态变量的线性变换仍然是正态变量.
- 正态分布的边缘分布仍然是正态分布
- 正态分布变量独立等价于不相关
- 正态分布的条件分布仍然是正态分布

1.3.3 指数族

【定义 1.38】 设 $\mathscr{F}=\{f(x,\theta):\theta\in\Theta\}$ 是定义在样本空间 Ω 上的分布族，其中 Θ 为参数空间. 若其概率函数 $f(x,\theta)$ 可表示成如下形式：

$$f(x,\theta)=C(\theta)\exp\left\{\sum_{i=1}^{k}Q_i(\theta)T_i(x)\right\}h(x)$$

则称此分布族为指数型分布族，其中 k 为正整数，$C(\theta)>0$ 和 $Q_i(\theta)(i=1,2,\cdots,k)$ 都是定义在参数空间 Θ 上的函数，$h(x)>0$ 和 $T_i(x)(i=1,2,\cdots,k)$ 都是定义在样本空间 Ω 上的函数.

【注 1.39】 指数族的一个重要性质是族中的所有分布具有共同的支撑集 ($G(x)$ 称为概率函数 $f(x,\theta)$ 的支撑集，若 $G(x)=\{x: f(x,\theta)>0\}$). 由定义可见指数族的支撑集 $\{x: f(x,\theta)>0\}=\{x: h(x)>0\}$ 与 θ 无关. 任一分布族若其支撑集与 θ 有关，则族中分布不再具有共同支撑集，因而必不是指数族.

【命题 1.40】 容易验证下列分布族是指数族

1. $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为从正态分布 $N\left(\mu,\sigma^2\right)$ 中抽取的简单样本
2. $\boldsymbol{X}=(X_1,\cdots,X_n)$ 为从 Gamma 分布 $\Gamma(\gamma,\lambda)$ 中抽取的简单样本
3. 二项分布族 $\{b(n,\theta): 0<\theta<1\}$
4. Poisson 分布族 $\{P(\theta): \theta>0\}$

证明.

1.
2.
3.
4.

□

解. 样本 X 的联合密度为

$$f\left(\boldsymbol{x};\mu,\sigma^2\right)=(\sqrt{2\pi}\sigma)^{-n}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\mu)^2\right\}.$$

记 $\theta=\left(\mu,\sigma^2\right)$，则参数空间为 $\theta=\left\{\theta=\left(\mu,\sigma^2\right): -\infty<\mu<+\infty,\sigma^2>0\right\}$.

$$\begin{aligned}f(\boldsymbol{x},\theta)&=(\sqrt{2\pi}\sigma)^{-n}e^{-\frac{n\mu^2}{2\sigma^2}}\exp\left\{\frac{\mu}{\sigma^2}\sum_{i=1}^n x_i-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2\right\}\\&=C(\theta)\exp\left\{Q_1(\theta)T_1(\boldsymbol{x})+Q_2(\theta)T_2(\boldsymbol{x})\right\}h(\boldsymbol{x}),\end{aligned}$$

其中 $C(\theta)=(\sqrt{2\pi}\sigma)^{-n}\exp\left\{-n\mu^2/\left(2\sigma^2\right)\right\}, Q_1(\theta)=\mu/\sigma^2, Q_2(\theta)=-1/\left(2\sigma^2\right)$，$T_1(\boldsymbol{x})=\sum_{i=1}^n x_i, T_2(\boldsymbol{x})=\sum_{i=1}^n x_i^2, h(\boldsymbol{x})\equiv 1$. 因此，由定义可知上述样本分布族是指数族. □

【例题 1.41】 均匀分布族 $\{U(0,\theta),\theta>0\}$ 和双参数指数分布不是指数族.

解. 均匀分布族 $\{U(0,\theta),\theta>0\}$ 的支撑集为 $\{x: f(x,\theta)>0\}=(0,\theta)$ 与 θ 有关，因此它不是指数族。

双参数指数分布族的密度函数如下：

$$p(x;\mu,\sigma)=\frac{1}{\sigma}\exp\left\{-\frac{x-\mu}{\sigma}\right\}I_{[x>\mu]},\quad -\infty<\mu<+\infty,\quad \sigma>0$$

其中 μ 和 σ 为两个参数，它的支撑集为 $\{x: p(x;\mu,\sigma)>0\}=(\mu,\infty)$ 与未知参数 μ 有关，故它也不是指数族. 但若 μ 已知，如 $\mu=0$，则单参数指数分布族 $\operatorname{Exp}(1/\sigma)$ 属于指数族. □

§1.4 三大抽样分布

1.4.1 卡方分布

【定义 1.42】(卡方分布) 如果随机变量 X 的密度函数为

$$f(x)=\frac{(1/2)^{k/2}}{\Gamma(k/2)}x^{\frac{k}{2}-1}e^{-\frac{x}{2}}\quad x>0,$$

则称 X 服从自由度为 k 的卡方分布，记为 $X \sim \chi^2(k)$

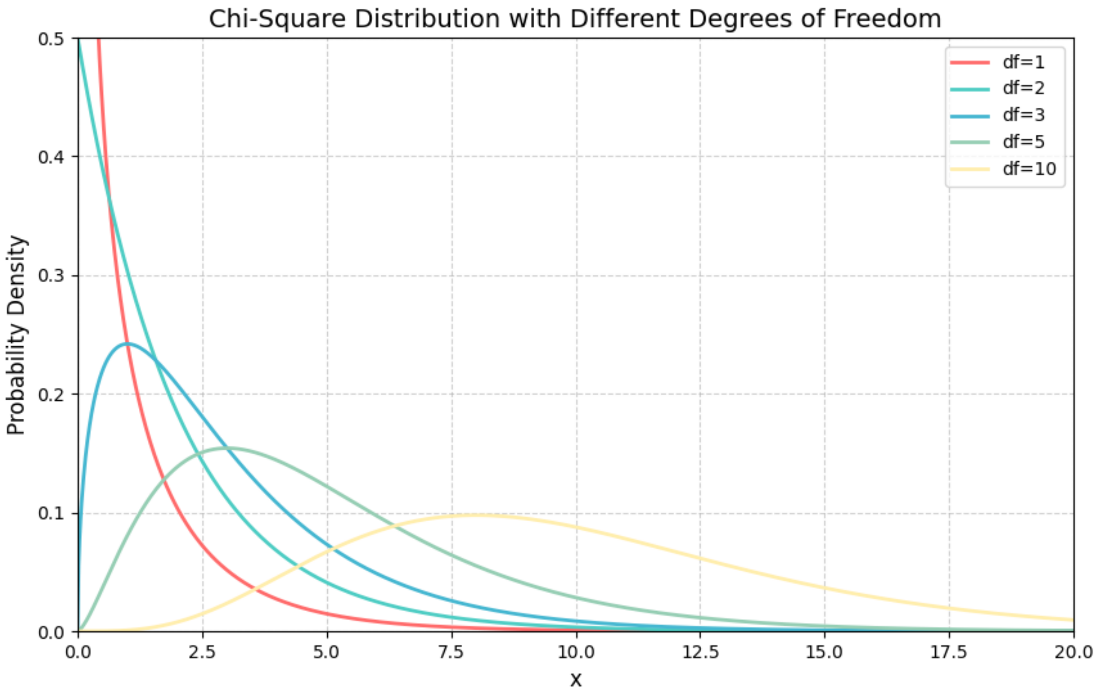

图 1.3: 不同自由度卡方分布对比

【命题 1.43】 由 Gamma 分布族的知识立刻推出卡方分布具有以下事实：

1. 卡方分布是参数为 $k/2$ 和 $1/2$ 的伽马分布 $Ga\left(\frac{k}{2}, \frac{1}{2}\right)$
2. 随机变量 $X \sim \chi^2(k)$，则

$$\mathrm{E}(X) = \frac{k/2}{1/2} = k, \quad \mathrm{var}(X) = \frac{k/2}{(1/2)^2} = 2k$$
$$m_X(t) = (1-2t)^{-k/2}, \quad \psi_X(t) = (1-2it)^{-k/2}$$

3. 设 $X \sim \chi^2(k_1), Y \sim \chi^2(k_2)$，且 X 和 Y 独立，则 $X+Y \sim \chi^2(k_1+k_2)$.

【定理 1.44】 设随机变量 $X_1, \cdots, X_k \quad iid \sim N(0,1)$ 则 $X = \sum_{j=1}^{k} X_j^2 \sim \chi(k)$

<u>**证明.**</u> 概率法：

$$X_i \sim N(0,1) \implies f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \implies f_Y(y) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y}{2}\right) \frac{1}{\sqrt{y}}$$

显然有 $Y \sim Ga(\frac{1}{2}, \frac{1}{2})$ 则 $\sum X_j^2 = \sum Y_j \sim Ga(\frac{k}{2}, \frac{1}{2}) = \chi(k)$ □

【推论 1.45】 如果随机变量 $X_1, \cdots, X_k$ 相互独立，且 $X_i \sim N\left(\mu_i, \sigma_i^2\right)$，则

$$\sum_{i=1}^{k} \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2$$

服从自由度为 k 的卡方分布.

【推论 1.46】 如果 $X_1, \cdots, X_k$ 是来自正态总体 $N\left(\mu, \sigma^2\right)$ 的随机样本，则

$$\sum_{i=1}^{k}\left(\frac{X_i-\mu}{\sigma}\right)^2$$

服从自由度为 k 的卡方分布.

1.4.2 F 分布

【定义 1.47】(F 分布) 如果随机变量 X 有如下密度函数

$$f(x)=\frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma(m/2)\Gamma(n/2)}\left(\frac{m}{n}\right)^{m/2}\frac{x^{\frac{(m-2)}{2}}}{[1+(m/n)x]^{(m+n)/2}},\quad x>0$$

则称 X 为具有自由度 m 和 n 的 F 分布，记为 $X \sim F(m,n)$.

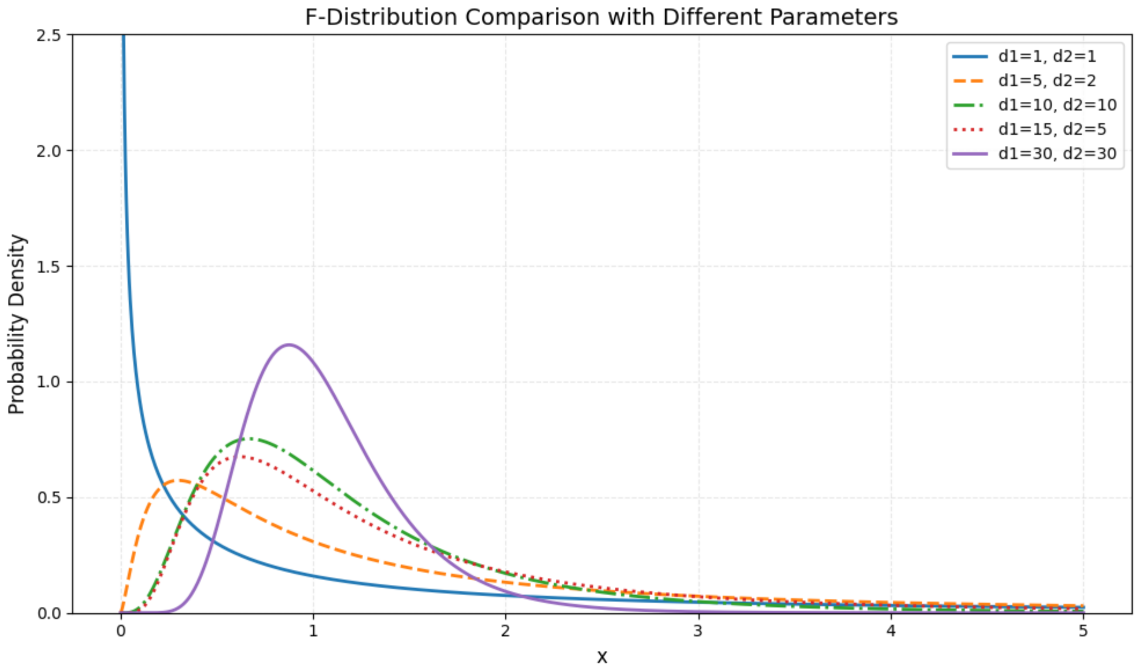

图 1.4: 不同自由度 F 分布对比

【定理 1.48】 令 $U \sim \chi^2(m)$，$V \sim \chi^2(n)$，且 U 和 V 独立，则随机变量

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

服从自由度为 m 和 n 的 F 分布

证明. 令

$$\begin{cases} X=\dfrac{U/m}{V/n} \\ Y=V \end{cases} \Longrightarrow \begin{cases} U=\dfrac{m}{n}XY \\ V=Y \end{cases}$$

变换矩阵的行列式:

$$J=\begin{vmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \\ \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \end{vmatrix}=\begin{vmatrix} \dfrac{m}{n}y & \dfrac{m}{n}x \\ 0 & 1 \end{vmatrix}=\frac{m}{n}y$$

□

【定理 1.49】

1.4.3 t 分布

【定义 1.50】 设 $X \sim N(0,1), Y \sim \chi_n^2$ 且 X 和 Y 独立，则称

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

是自由度为 n 的 t 变量，其分布称为自由度为 n 的 t 分布，记为 $T \sim t_n$.

【定理 1.51】 设随机变量 $T \sim t_n$ ，则其密度函数为

$$t_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}}\left(1+\frac{x^2}{n}\right)^{-(n+1)/2}, \quad -\infty < x < \infty$$

证明.

□

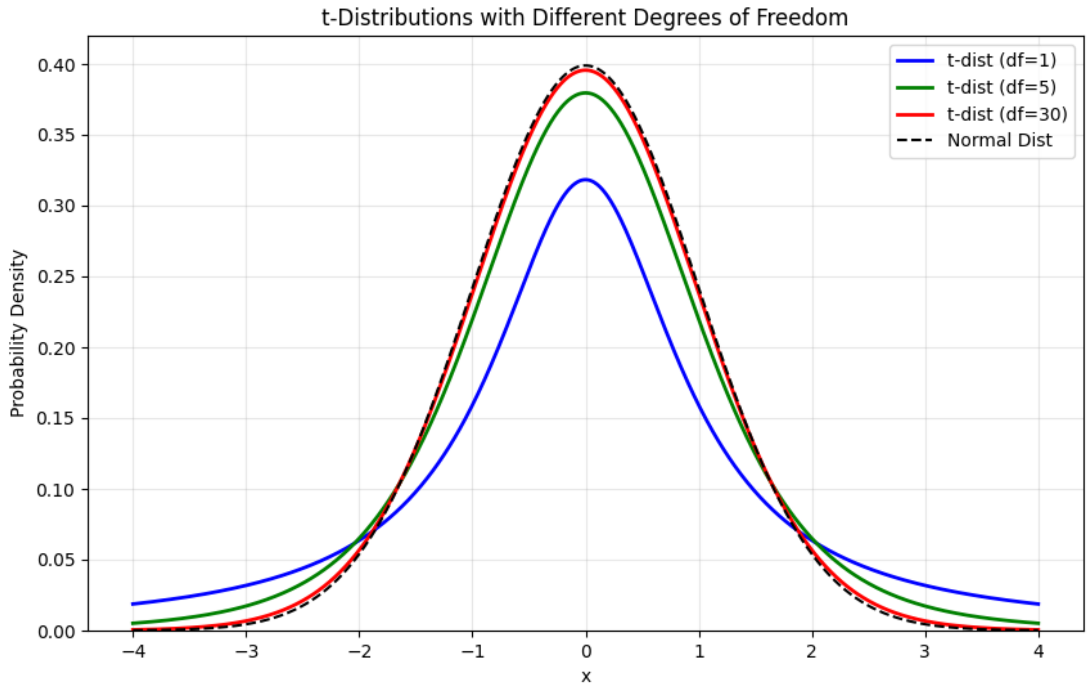

图 1.5: 不同参数 t 分布对比

【注 1.52】 如果 $X \sim t(k)$ ，则

1. 当 $k=1$ 时，$f(x) = \dfrac{1}{\pi(1+x^2)}$ ，即 $t(1)$ 是标准的 Cauchy 分布.
2. $\lim\limits_{k\to\infty} f(x;k) = \phi(x)$.
3. 当 $k>1, \mathrm{E}(X)=0$ ；当 $k>2$ 时，$\mathrm{var}(X) = \dfrac{k}{k-2}$.
4. 密度函数关于 $x=0$ 对称，且在 $x=0$ 处达到最大.

【推论 1.53】 如果 $X_1, \cdots, X_n$ 是来自正态分布 $N\left(\mu, \sigma^2\right)$ 的随机样本，则

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

因此

$$\frac{(\bar{X}-\mu)/(\sigma/\sqrt{n})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2/(n-1)\sigma^2}}=\frac{\bar{X}-\mu}{S_n/\sqrt{n}}\sim t_{n-1}$$

其中 S_n 是样本标准差.

【推论 1.54】 如果 $X_1,\cdots,X_m$ 是来自正态总体 $X\sim N\left(\mu_X,\sigma^2\right)$ 的随机样本，$Y_1,\cdots,Y_n$ 是来自正态总体 $Y\sim N\left(\mu_y,\sigma^2\right)$，并且 X 和 Y 独立，因此

$$\frac{\bar{X}-\bar{Y}-(\mu_x-\mu_y)}{S_w\sqrt{\frac{1}{m}+\frac{1}{n}}}\sim t(n+m-2)$$

其中

$$S_w^2=\frac{(m-1)S_x^2+(n-1)S_y^2}{m+n-2}$$

这里 S_x^2 和 S_y^2 分别是随机样本 $X_1,\cdots,X_m$ 和 $Y_1,\cdots,Y_n$ 的样本方差.

1.4.4 重要推论

【定理 1.55】 如果 $X_1,\cdots,X_n$ 是来自 $N\left(\mu,\sigma^2\right)$ 的随机样本，则

1. $\bar{X}$ 服从均值为 μ 方差为 σ^2/n 的正态分布
2. $\bar{X}$ 和 S^2 独立
3. $\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$

证明.

1. 略
2.
3.

□

【推论 1.56】 如果 $S^2=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\bar{X})^2$ 是来自正态总体 $N\left(\mu,\sigma^2\right)$ 的随机样本的样本方差，则

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

【推论 1.57】 S^2 的密度函数可以根据 U 的密度函数导出：

$$f_{S^2}(s)=f_U\left(\frac{(n-1)s}{\sigma^2}\right)\cdot\frac{(n-1)}{\sigma^2}=\left(\frac{n-1}{2\sigma^2}\right)^{\frac{n-1}{2}}\frac{1}{\Gamma\left(\frac{n-1}{2}\right)}s^{\frac{n-3}{2}}e^{-\frac{(n-1)s}{2\sigma^2}},\quad s>0.$$

【例题 1.58】 设 X_1,X_2,X_3,X_4 为来自正态总体 $N\left(\mu,\sigma^2\right)$ 的样本，求统计量

$$\frac{X_1-X_2}{|X_3+X_4-2\mu|}$$

的分布

【例题 1.59】 设 $X_1,\cdots,X_n$ 为来自正态总体 $N\left(0,\sigma^2\right)$ 的样本，$\bar{X}_n$ 和 S_n^2 分别是样本均值和样本方差. 统计量 $T_1=\bar{X}_n^2-S_n^2/n,T_2=\bar{X}_n^2/S_n^2$，计算 T_1 和 T_2 的方差.

§1.5 统计量的极限分布

Chapter 2

点估计

§2.1 参数估计方法

2.1.1 矩估计

- 设 $X_1, \cdots, X_n$ 是来自总体 $f(x;\theta)$ 的随机样本，其中 $\theta = (\theta_1, \theta_2, \cdots, \theta_k)^{\mathrm{T}}$. 令 M_r' 是样本 r 阶原点矩，即:

$$M_r' = \frac{1}{n}\sum_{i=1}^{n} X_i^r,$$

  并且 $\mathrm{E}\left(M_r'\right) = \mathrm{E}\left(X^r\right) = \mu_r' = \mu_r'\left(\theta_1, \theta_2, \cdots, \theta_k\right), M_r' \xrightarrow{p} \mu_r'$.

- 利用样本矩来代替总体矩，k 个参数 $\theta_1, \theta_2, \cdots, \theta_k$ 列 k 个方程:

$$\begin{aligned}
M_1' &= \mu_1' = \mu_1'\left(\theta_1, \theta_2, \cdots, \theta_k\right) \\
M_2' &= \mu_2' = \mu_2'\left(\theta_1, \theta_2, \cdots, \theta_k\right) \\
&\vdots \\
M_k' &= \mu_k' = \mu_k'\left(\theta_1, \theta_2, \cdots, \theta_k\right).
\end{aligned}$$

- 令 $\hat\theta_1, \cdots, \hat\theta_k$ 是上述方程的解(设上述方程有唯一解), 称 $\left(\hat\theta_1, \cdots, \hat\theta_k\right)$ 是利用矩方法得到的参数 $\left(\theta_1, \cdots, \theta_k\right)$ 的估计量，其中 $\hat\theta_j$ 是对应的 θ_j 的估计.

【例题 2.1】 设总体 X 的概率密度为

$$f(x;\theta) = \begin{cases} \dfrac{2x}{\theta} e^{-x^2/\theta}, & x \geq 0, \\ 0 & \text{其它}, \end{cases}$$

其中 $\theta > 0$ 是未知参数，$X_1, \cdots, X_n$ 为来自总体 X 的样本，求 θ 的矩估计.

解. 因为

$$\mathrm{E}(X) = \int_0^\infty x f(x;\theta) dx = \int_0^\infty \frac{2x^2}{\theta} e^{-x^2/\theta} dx = \frac{\sqrt{\theta\pi}}{2},$$

令

$$\mathrm{E}(X) = \frac{\sqrt{\theta\pi}}{2} = \bar{X}_n,$$

故 θ 的矩估计为

$$\hat\theta = \frac{\left(2\bar{X}_n\right)^2}{\pi}.$$

□

【例题 2.2】 令 $X_1,\cdots,X_n$ 是来自正态总体 $N\left(\mu,\sigma^2\right)$ 的随机样本，求参数 μ 和 σ^2 的矩估计.

解. 我们有

$$\mathrm{E}(X)=\mu=\frac{1}{n}\sum_{i=1}^{n}X_i,\quad E\left(X^2\right)=\mu^2+\sigma^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2.$$

则

$$\Longrightarrow\begin{cases}\bar{X}=\mu,\\ \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}X_i^2=\mu^2+\sigma^2\end{cases}$$

解得

$$\Longrightarrow\begin{cases}\hat{\mu}=\bar{X},\\ \hat{\sigma}^2=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2.\end{cases}$$

□

【例题 2.3】 $X_1,\cdots,X_n$ 是来自 $(\mu-\sqrt{3}\sigma,\mu+\sqrt{3}\sigma)$ 的均匀分布的随机样本．求 μ 和 σ 的矩估计.

解.

$$\mathrm{E}(X)=\mu=\frac{1}{n}\sum_{i=1}^{n}X_i,\quad \mathrm{E}\left(X^2\right)=\mu^2+\sigma^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2.$$

$$\Longrightarrow\begin{cases}\hat{\mu}=\bar{X},\\ \hat{\sigma}=\sqrt{\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}\end{cases}$$

□

【例题 2.4】 令 $X_1,\cdots,X_n$ 是来自参数为 λ 的 Poisson 分布的随机样本，利用矩方法估计 λ .

解.

1. 利用一阶样本原点矩代替一阶原点矩，得到
$$\frac{1}{n}\sum_{i=1}^{n}x_i=\mathrm{E}(X)=\lambda.\Longrightarrow\hat{\lambda}=\bar{X}$$
2. 利用二阶样本原点矩代替二阶原点矩，得到
$$\frac{1}{n}\sum_{i=1}^{n}X_i^2=\mathrm{E}\left(X^2\right)=\lambda^2+\lambda.$$
$$\Rightarrow\tilde{\lambda}=\frac{-1+\sqrt{1+\frac{4}{n}\sum_{i=1}^{n}X_i^2}}{2}.$$
3. 利用样本方差代替总体方差，得到
$$\lambda=\mathrm{var}(X)=S_n^2\Longrightarrow\check{\lambda}=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2.$$

□

【例题 2.5】 令 $X_1,\cdots,X_n$ 是来自总体 X 的随机样本，X 的密度函数为

$$f\left(x;\theta_1,\theta_2\right)=\frac{\theta_2}{\Gamma\left(\left(1+\theta_1\right)/\theta_2\right)}x^{\theta_1}\exp\left(-x^{\theta_2}\right),x>0$$

其中 $-1<\theta_1<+\infty,\theta_2>0$. 求 θ_1 和 θ_2 的矩估计.

【例题 2.6】 令 $X_1,\cdots,X_n$ 是来自正态总体 $X\sim N\left(\mu,\sigma^2\right)$ 的随机样本，求 $P(X>1)$ 的矩估计.

【例题 2.7】 设 $\left(X_1,Y_1\right),\cdots,\left(X_n,Y_n\right)$ 是来自二元总体 (X,Y) 中的样本，设 $\mathrm{E}\left(X^2\right)$ 和 $\mathrm{E}\left(X^2\right)$ 存在，求 $\sigma_{XY}=\mathrm{Cov}(X,Y)$ 和 $\mathrm{Corr}(X,Y)=\rho$ 的矩估计.

2.1.2 极大似然估计

【定义 2.8】(似然函数) 令 $X_1,\cdots,X_n$ 是来自总体 $f(x;\theta)$ 的随机样本，则称 $X_1,\cdots,X_n$ 的联合概率函数为其似然函数，即 $f_{X_1,\cdots,X_n}(x_1,\cdots,x_n;\theta)$ 被看作参数 θ 的函数．特别地，随机样本 $X_1,\cdots,X_n$ 的似然函数等于

$$f_{X_1,\cdots,X_n}(x_1,\cdots,x_n;\theta)=f(x_1;\theta)\,f(x_2;\theta)\cdots f(x_n;\theta)\,.$$

似然函数通常用 $L(\theta;x_1,\cdots,x_n)$ 或 $L(\theta)$ 表示．

- 似然函数是固定 $x_1,\cdots,x_n$ ，将其看作定义在参数空间 Θ 上的参数 θ 的函数.
- 如果 $X_1,\cdots,X_n$ 是来自 $f(x;\theta)$ 的随机样本，则

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

- 似然函数 $L(\theta)$ 度量了随机变量 $X_1,\cdots,X_n$ 取 $x_1,\cdots,x_n$ 的可能性.

【定义 2.9】(MLE) 令 $L(\theta)=L(\theta;x_1,\cdots,x_n)$ 是随机变量 $X_1,\cdots,X_n$ 的似然函数．如果

$$\hat{\theta}=\operatorname{argmax}_{\theta\in\Theta}L(\theta)$$

称 $\hat{\theta}=\hat{\theta}(X_1,\cdots,X_n)$ 是参数 θ 的极大似然估计量，$\hat{\theta}=\hat{\theta}(x_1,\cdots,x_n)$ 是参数 θ 的极大似然估计值，其中 $x_1,\cdots,x_n$ 是观测值.

- 如果 $L(\theta)$ 是可导的，则 MLE 是方程 $\dfrac{dL(\theta)}{d\theta}=0$ 的解.
- 因为 $\ln L(\theta)$ 是 $L(\theta)$ 在同一 θ 处达到最大值，有时 $\ln L(\theta)$ 的最值更容易计算，故可以通过最大化 $\ln L(\theta)$ 求参数的 MLE.
- 如果 $L(\theta)$ 关于 θ 不可导，甚至不连续，则要从定义出发，寻找参数的极大似然估计.

【定理 2.10】(极大似然估计的不变性) 设总体的概率函数 $f(x;\theta)$ ，假设参数 θ 是一维的，令 $\hat{\theta}(X_1,\cdots,X_n)$ 是参数 θ 的极大似然估计．如果 $\tau(\theta)$ 是 θ 的单值可逆函数，则 $\tau(\theta)$ 的极大似然估计是 $\tau(\hat{\theta})$ 。

【注 2.11】 定理可以从两个方向推广，极大似然估计的不变性依然成立：

1. θ 可以是 k 维向量；
2. $\tau(\theta)$ 是单值可逆函数的假设可以去掉.

§2.2 点估计的性质

2.2.1 无偏性

【定义 2.12】 设 $X_1,\cdots,X_n$ 是来自总体 $f(x;\theta)$ 的随机样本，其中 $\theta\in\Theta.T=T(X_1,\cdots,X_n)$ 是 $\tau(\theta)$ 的一个估计量，如果

$$\mathrm{E}_\theta(T)=\mathrm{E}_\theta\left[T(X_1,\cdots,X_n)\right]=\tau(\theta),\quad \theta\in\Theta,$$

则称 $T=T(X_1,\cdots,X_n)$ 是 $\tau(\theta)$ 的无偏估计.

- 如果 $\mathrm{E}_\theta(T)\neq\tau(\theta)$ ，则估计量 $T=T(X_1,\cdots,X_n)$ 是 $\tau(\theta)$ 的有偏估计；称 $\tau(\theta)-\mathrm{E}_\theta(T)$ 是估计量 T 的偏
- 如果 $\lim\limits_{n\to\infty}\mathrm{E}_\theta(T)=\tau(\theta)$ ，则称估计量 T 是 $\tau(\theta)$ 的渐近无偏估计

【例题 2.13】 令 $X_1,\cdots,X_n$ 是来自正态总体 $N\left(\mu,\sigma^2\right)$ 的随机样本．判断 μ 和 σ^2 的极大似然估计是否是无偏估计．

解. μ 和 σ^2 的 MLE 分别为：

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

因为 $\mathrm{E}(\hat{\mu})=\mathrm{E}(\bar{X})=\mu$ 故 $\bar{X}$ 是无偏的．
因为

$$\mathrm{E}\left(\hat{\sigma}^2\right)=\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\right)=\frac{n-1}{n}\sigma^2\to\sigma^2,$$

故 $\hat{\sigma}^2$ 是渐近无偏的．$\hat{\sigma}^2$ 的偏为 $=\sigma^2-E\left(\hat{\sigma}^2\right)=\dfrac{1}{n}\sigma^2\to 0$． □

【注 2.14】 说明 MLE 和矩估计可能有偏也可能无偏

【例题 2.15】 令 $X_1,\cdots,X_n$ 是来自指数总体 $\mathrm{Exp}(\lambda)$ 的随机样本．

1. 求 λ 的 MLE $\hat{\lambda}$
2. 证明 $\hat{\lambda}$ 是有偏的．
3. 根据（2）中的结果，构造 λ 的一个无偏估计．

解.

1. 考虑似然函数

$$L(\lambda)=\prod_{i=1}^{n}f\left(x_i;\lambda\right)=\lambda^n\exp\left(-\lambda\sum_{i=1}^{n}x_i\right)\implies \ln L(\lambda)=n\ln\lambda-\lambda\sum_{i=1}^{n}x_i.$$

估计方程：

$$\frac{d\ln L(\lambda)}{d\lambda}=\frac{n}{\lambda}-\sum_{i=1}^{n}x_i=0\quad\implies\hat{\lambda}=\frac{1}{\bar{X}}$$

2. 因为 $Z=\sum_{i=1}^{n}X_i\sim\Gamma(n,\lambda)$，

$$\mathrm{E}(\hat{\lambda})=\mathrm{E}\left(\frac{1}{\bar{X}}\right)=\mathrm{E}\left(\frac{n}{\sum_{i=1}^{n}X_i}\right)=n\int_0^\infty\frac{1}{z}\frac{\lambda^n}{\Gamma(n)}z^{n-1}e^{-\lambda z}dz=\frac{n}{n-1}\lambda$$

因此 $\hat{\lambda}$ 是有偏的．

3. 根据 (2) 中的结果，我们可得

$$\mathrm{E}\left(\frac{n-1}{n}\cdot\frac{1}{\bar{X}}\right)=\lambda$$

是所求无偏估计

□

【例题 2.16】 令 $X_1,\cdots,X_n$ 是来自总体 $N\left(\mu,\sigma^2\right)$ 的随机样本．证明 $\hat{\sigma}=\sqrt{S_n^2}$ 不是 σ 的无偏估计，其中

$$S_n=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2.$$

解. 因为 $(n-1)S_n^2/\sigma^2\sim\chi^2(n-1)$，故

$$\begin{aligned}\mathrm{E}(\hat{\sigma})=\mathrm{E}\left(\sqrt{S_n^2}\right)&=\frac{\sigma}{\sqrt{n-1}}\int_0^\infty\sqrt{t}\frac{2^{-\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)}t^{\frac{n-1}{2}-1}e^{-\frac{t}{2}}dt\\&=\frac{\sqrt{2}}{\sqrt{n-1}}\cdot\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\sigma\neq\sigma,\end{aligned}$$

从而 $\hat{\sigma}$ 不是 σ 的无偏估计． □

【注 2.17】 无偏估计不具有不变性

【例题 2.18】 设 X_1 是来自二项分布总体 $B(n,\theta)$ 的一个随机样本，$1/\theta$ 是否存在无偏估计？

解. 假设 $T(X_1)$ 是 $1/\theta$ 的无偏估计．根据无偏估计的定义，我们有：

$$E\left(T\left(X_1\right)\right)=\frac{1}{\theta},\quad \forall\theta\in(0,1).$$

则对任意的 $\theta\in(0,1)$，下面等式恒成立：

$$\sum_{x=0}^{n}T(x)C_n^x\theta^x(1-\theta)^{n-x}=\frac{1}{\theta},\Longrightarrow\sum_{x=0}^{n}T(x)C_n^x\theta^{x+1}(1-\theta)^{n-x}-1=0.$$

但注意到这是一个关于 θ 的多项式，不可能满足对 $\forall\theta\in(0,1)$ 满足方程，故 $1/\theta$ 不存在无偏估计。 □

【注 2.19】 无偏估计可能不存在

【例题 2.20】 设 $X_1,\cdots,X_n$ 是来自总体 $f(x;\theta)=\theta^{-1}I_{[0,\theta]}(x)$ 的样本，其中 $\theta>0$．

1. 求参数 θ 的矩估计 $\hat{\theta}$ 和 MLE $\tilde{\theta}$．
2. 判断 $\hat{\theta}$ 和 $\tilde{\theta}$ 是否为无偏估计．如果不是无偏的，做适当的修正得到 θ 的无偏估计．
3. 计算矩估计 $\hat{\theta}$ 和 MLE $\tilde{\theta}$ 的方差．

解.

1. 由

$$\mathrm{E}(X)=\frac{\theta}{2}=\bar{X}_n$$

得 θ 的矩估计为 $\hat{\theta}=2\bar{X}_n$
似然函数

$$L(\theta)=\frac{1}{\theta^n}\prod_{i=1}^{n}I_{[0,\theta]}\left(x_i\right)=\frac{1}{\theta^n}I_{[0,\theta]}\left(y_n\right),$$

其中 y_n 是 $x_1,\cdots,x_n$ 的最大值．当 $\theta=y_n$ 时，$L(\theta)$ 达到最大值，从而 θ 的极大似然估计为 $\tilde{\theta}=Y_n$．

2. 因为

$$\mathrm{E}(\hat{\theta})=\mathrm{E}\left(2\bar{X}_n\right)=2\cdot\frac{\theta}{2}=\theta,$$

从而 $\hat{\theta}$ 是 θ 的无偏估计．总体的分布函数为

$$F(x)=\frac{x}{\theta}l_{[0,\theta]}(x),$$

最大次序统计量 Y_n 的密度函数

$$f_{Y_n}(y)=n[F(y)]^{n-1}f(y)=\frac{ny^{n-1}}{\theta^n}l_{[0,\theta]}(y),$$

从而

$$\mathrm{E}(\tilde{\theta})=\mathrm{E}\left(Y_n\right)=\int_0^\theta\frac{ny^n}{\theta^n}dy=\frac{n\theta}{n+1},$$

因此 $\tilde{\theta}$ 不是 θ 的无偏估计，从而

$$\check{\theta}=\frac{n+1}{n}Y_n$$

是 θ 的无偏估计．

3. 矩估计 $\hat{\theta}$ 的方差
$$\operatorname{var}(\hat{\theta})=\operatorname{var}\left(2\bar{X}_n\right)=4\cdot\frac{\theta^2}{12n}=\frac{\theta^2}{3n}.$$
极大似然估计量 $\tilde{\theta}$ 的方差
$$\operatorname{var}(\tilde{\theta})=\mathrm{E}\left(Y_n^2\right)-\left[\mathrm{E}\left(Y_n\right)\right]^2=\int_0^\theta \frac{ny^{n+1}}{\theta^n}dy-\left[\mathrm{E}\left(Y_n\right)\right]^2=\frac{n\theta^2}{(n+2)(n+1)^2}.$$
改进无偏估计 $\check{\theta}$ 的方差
$$\operatorname{var}(\check{\theta})=\frac{\theta^2}{n(n+2)}.$$

□

【注 2.21】 不难证明 $\operatorname{var}(\tilde{\theta})\leq\operatorname{var}(\check{\theta})\leq\operatorname{var}(\hat{\theta})$ ．但是矩估计 $\hat{\theta}$ 和改进的极大似然估计 $\check{\theta}$ 是无偏的，极大似然估计量 $\tilde{\theta}$ 是有偏的。

【定义 2.22】(有效性) 令 T_1 和 T_2 是 $\tau(\theta)$ 的两个无偏估计，如果对任意的 $\theta\in\Theta$，有 $\operatorname{var}\left(T_1\right)\leq\operatorname{var}\left(T_2\right)$，并且至少存在一个 $\theta\in\Theta$，使得不等号严格成立，则称 T_1 比 T_2 更有效.

2.2.2 均方误差

【定义 2.23】(均方误差) $T=T\left(X_1,\cdots,X_n\right)$ 是 $\tau(\theta)$ 的估计，称
$$\mathrm{E}_\theta(T-\tau(\theta))^2$$
为 $T=T\left(X_1,\cdots,X_n\right)$ 的均方误差.

【注 2.24】
$$\mathrm{MSE}_T(\theta)=\int\cdots\int\left(T\left(x_1,\cdots,x_n\right)-\tau(\theta)\right)^2\prod_{i=1}^n f\left(x_i;\theta\right)dx_1\cdots dx_n.$$

【注 2.25】
$$\mathrm{MSE}_T(\theta)=\mathrm{E}_\theta(T-\mathrm{E}(T)+\mathrm{E}(T)-\tau(\theta))^2=\operatorname{var}(T)+[\operatorname{Bias}(T)]^2$$
故如果 T 是 $\tau(\theta)$ 的无偏估计，则 $\mathrm{MSE}_T(\theta)=\operatorname{var}(T)$

【注 2.26】 如果 T 是 r 维向量，则其均方误差定义为
$$\mathrm{MSE}_T(\theta)=\mathrm{E}\left(\|T-\tau(\theta)\|^2\right)=\sum_{j=1}^r\mathrm{E}\left(T_j-\tau_j(\theta)\right)^2$$
其中 $\|\cdot\|$ 表示二范数．因此，估计向量的均方误差等于各个分量的均方误差之和。

【例题 2.27】(例2.20续) 设 $X_1,\cdots,X_n$ 是来自总体 $f(x;\theta)=\theta^{-1}l_{[0,\theta]}(x)$ 的样本，其中 $\theta>0$．求参数 θ 的矩估计 $\hat{\theta}$，极大似然估计 $\tilde{\theta}$ 和其修正后的无偏估计 $\check{\theta}$ 的均方误差.

解.

1. $\hat{\theta}$ 的均方误差
$$\mathrm{MSE}_{\hat{\theta}}(\theta)=\operatorname{var}(\hat{\theta})=\frac{\theta^2}{3n}.$$

2. $\tilde{\theta}$ 均方误差
$$\begin{aligned}\mathrm{MSE}_{\tilde{\theta}}(\theta)&=\operatorname{var}(\tilde{\theta})+\operatorname{Bias}(\tilde{\theta})\\&=\frac{n\theta^2}{(n+2)(n+1)^2}+\left(\frac{\theta}{n+1}\right)^2\\&=\frac{2\theta^2}{(n+2)(n+1)}.\end{aligned}$$

3. $\check{\theta}$ 的均方误差

$$\mathrm{MSE}_{\check{\theta}}(\theta)=\operatorname{var}(\check{\theta})=\frac{\theta^2}{n(n+2)}.$$

不难证明：$\mathrm{MSE}_{\check{\theta}}(\theta)\le\mathrm{MSE}_{\tilde{\theta}}(\theta)\le\mathrm{MSE}_{\hat{\theta}}(\theta)$

□

2.2.3 相合性

【定义 2.28】(均方相合性) 令 $T_n=T_n(X_1,\cdots,X_n)$ 是 $\tau(\theta)$ 的一个估计量，其中 n 是样本量. 如果对任意的 $\theta\in\Theta$，有

$$\lim_{n\to\infty}E_\theta\left[T_n-\tau(\theta)\right]^2=0,$$

则称 T_n 是 $\tau(\theta)$ 的均方相合估计.

【注 2.29】 T_n 是均方相合的，意味着 T_n 的偏和方差在 $n\to\infty$ 时均趋于零。

【定义 2.30】(弱相合性) 令 $T_n=T_n(X_1,\cdots,X_n)$ 是 $\tau(\theta)$ 的一个估计量，其中 n 是样本量. 如果对任意的 $\varepsilon>0$，有

$$\lim_{n\to\infty}P_\theta\left[|T_n-\tau(\theta)|<\varepsilon\right]=1,\theta\in\Theta,$$

则称 T_n 是 $\tau(\theta)$ 的弱相合估计.

【定理 2.31】 设 $\hat{\theta}_n$ 是 θ 的估计，且满足

$$\lim_{n\to\infty}\mathrm{E}\left(\hat{\theta}_n\right)=\theta,\quad\lim_{n\to\infty}\operatorname{var}\left(\hat{\theta}_n\right)=0$$

则 $\hat{\theta}_n$ 是 θ 的弱相合估计.

【定理 2.32】 如果 $\hat{\theta}_n=\left(\hat{\theta}_{n1},\hat{\theta}_{n2},\cdots,\hat{\theta}_{nk}\right)$ 是 $\theta=(\theta_1,\cdots,\theta_k)$ 的弱相合估计，且 $g(\theta)$ 是参数 θ 的连续函数，则 $g\left(\hat{\theta}_n\right)$ 是 $g(\theta)$ 的弱相合估计.

【例题 2.33】 设总体 X 的均值为 μ，方差为 $\sigma^2<+\infty,X_1,\cdots,X_n$ 是来自总体 X 的随机样本，证明 $\bar{X}_n$ 和 S_n^2 分别为 μ 和 σ^2 的相合估计.

2.2.4 渐进正态性

【定义 2.34】 T_n 是 $\tau(\theta)$ 的估计量，如果 T_n 满足如下条件：

- 当 $n\to\infty$，

$$\sqrt{n}\left(T_n-\tau(\theta)\right)\to N\left(0,\sigma^2(\theta)\right).$$

- $\forall\varepsilon>0$，

$$\lim_{n\to\infty}P_\theta\left[|T_n-\tau(\theta)|>\varepsilon\right]=0,\theta\in\Theta.$$

则称 T_n 是 $\tau(\theta)$ 的相合渐近正态估计，称 $\sigma^2(\theta)$ 为 $\sqrt{n}T_n$ 的渐近方差. 如果 T_n 还同时满足下面条件

- 令 $\left\{\widetilde{T}_n\right\}$ 是 $\tau(\theta)$ 的另外一个相合估计，并且

$$\sqrt{n}\left(\widetilde{T}_n-\tau(\theta)\right)\xrightarrow{d}N\left(0,\tilde{\sigma}^2(\theta)\right).$$

- 对任意的 $\theta\in\Theta$（参数空间是开区间），$\sigma^2(\theta)\le\tilde{\sigma}^2(\theta)$，则称 T_n 是 $\tau(\theta)$ 的最优渐近正态估计

§2.3 一致最小方差无偏估计

- UMVUE 就是在寻找均方误差最小，我们只研究无偏的情况
- $\tau(\theta)$ 的无偏估计 T 的均方误差可以写为
$$\mathrm{E}(T-\tau(\theta))^2=\mathrm{var}(T)+[\tau(\theta)-\mathrm{E}(T)]^2.$$
如果 T 是无偏估计，则 $\mathrm{E}(T-\tau(\theta))^2=\mathrm{var}(T)$.

【定义 2.35】 令 $X_1,\cdots,X_n$ 是来自 $f(x;\theta)$ 的样本。$\tau(\theta)$ 的估计 $T^*=T^*\left(X_1,\cdots,X_n\right)$ 被称为 $\tau(\theta)$ 的一致最小方差无偏估计量当且仅当

1. $\mathrm{E}_\theta\left(T^*\right)=\tau(\theta)$ ，即 T^* 是无偏的；
2. $\mathrm{var}\left(T^*\right)\leq\mathrm{var}(T)$ ，其中 $T=T\left(X_1,\cdots,X_n\right)$ 是 $\tau(\theta)$ 的任意一个无偏估计.

Cramer-Rao 不等式

令 $X_1,\cdots,X_n$ 是来自 $f(x;\theta)$ 的随机样本，其中 $\theta\in\Theta\subseteq R$. 令 $T=T\left(X_1,\cdots,X_n\right)$ 是 $\tau(\theta)$ 的无偏估计，我们假设下列正则条件成立

1. 对所有的 $x\in\{x:f(x;\theta)>0\}$ 和 $\theta\in\Theta$, $\frac{\partial}{\partial\theta}\ln f(x;\theta)$ 存在.
2. $\frac{\partial}{\partial\theta}\int\prod_{i=1}^n f\left(x_i;\theta\right)dx_1\cdots dx_n=\int\frac{\partial}{\partial\theta}\prod_{i=1}^n f\left(x_i;\theta\right)dx_1\cdots dx_n$.
3. $\frac{\partial}{\partial\theta}\int T\left(x_1,\cdots,x_n\right)\prod_{i=1}^n f\left(x_i;\theta\right)dx_1\cdots dx_n=\int T\left(x_1,\cdots,x_n\right)\frac{\partial}{\partial\theta}\prod_{i=1}^n f\left(x_i;\theta\right)dx_1\cdots dx_n$.
4. $\forall\theta\in\Theta,\quad 0<\mathrm{E}\left[\frac{\partial}{\partial\theta}\ln f(X;\theta)\right]^2<\infty$.

【定理 2.36】(C-R 不等式) 假设上述正则条件 1-4 成立，$T=T\left(X_1,\cdots,X_n\right)$ 是 $\tau(\theta)$ 的任一无偏估计，则有下面不等式成立：
$$\mathrm{var}[T]\geq\frac{\left[\tau'(\theta)\right]^2}{n\mathrm{E}\left[\frac{\partial}{\partial\theta}\ln f(X;\theta)\right]^2}$$
等号成立当且仅当存在函数 $K(\theta,n)$ ，使得
$$\frac{\partial}{\partial\theta}\sum_{i=1}^n\ln f\left(x_i;\theta\right)=\frac{\partial\ell(\theta)}{\partial\theta}=K(\theta,n)\left[T\left(x_1,\cdots,x_n\right)-\tau(\theta)\right].$$

证明. 根据假设 1 可得：
$$\begin{aligned}\tau'(\theta)=\frac{\partial\tau(\theta)}{\partial\theta}&=\frac{\partial}{\partial\theta}\int\cdots\int T\left(x_1,\cdots,x_n\right)\prod_{i=1}^n f\left(x_i;\theta\right)dx_1\cdots dx_n\\&=\int\cdots\int T\left(x_1,\cdots,x_n\right)\frac{\partial}{\partial\theta}\left[\prod_{i=1}^n f\left(x_i;\theta\right)\right]dx_1\cdots dx_n\end{aligned}$$
根据假设 2 可得：
$$\begin{aligned}0&=\frac{\partial}{\partial\theta}\int\cdots\int\prod_{i=1}^n f\left(x_i;\theta\right)dx_1\cdots dx_n\\&=\int\cdots\int\frac{\partial}{\partial\theta}\left[\prod_{i=1}^n f\left(x_i;\theta\right)\right]dx_1\cdots dx_n\\&=\int\cdots\int\tau(\theta)\frac{\partial}{\partial\theta}\left[\prod_{i=1}^n f\left(x_i;\theta\right)\right]dx_1\cdots dx_n\end{aligned}$$

因此

$$\begin{aligned}\tau'(\theta) &= \int\cdots\int \left[T\left(x_1,\cdots,x_n\right)-\tau(\theta)\right]\frac{\partial}{\partial\theta}\left[\prod_{i=1}^{n} f\left(x_i;\theta\right)\right] dx_1\cdots dx_n \\ &= \int\cdots\int \left[T\left(x_1,\cdots,x_n\right)-\tau(\theta)\right]\frac{\partial}{\partial\theta}\ln\left[\prod_{i=1}^{n} f\left(x_i;\theta\right)\right]\times\prod_{i=1}^{n} f\left(x_i;\theta\right) dx_1\cdots dx_n \\ &= \mathrm{E}\left\{\left[T\left(X_1,\cdots,X_n\right)-\tau(\theta)\right]\left[\frac{\partial}{\partial\theta}\ln\prod_{i=1}^{n} f\left(X_i;\theta\right)\right]\right\}\end{aligned}$$

根据 Cauchy-Schwarz 不等式

$$\left[\tau'(\theta)\right]^2 \le \mathrm{E}\left[T\left(X_1,\cdots,X_n\right)-\tau(\theta)\right]^2 \mathrm{E}\left[\frac{\partial}{\partial\theta}\ln\prod_{i=1}^{n} f\left(X_i;\theta\right)\right]^2.$$

$\Longrightarrow$

$$\left[\tau'(\theta)\right]^2 \le \operatorname{var}[T] E\left[\frac{\partial}{\partial\theta}\ln\prod_{i=1}^{n} f\left(X_i;\theta\right)\right]^2$$

$\Longrightarrow$

$$\operatorname{var}[T] \ge \frac{\left[\tau'(\theta)\right]^2}{\mathrm{E}\left[\frac{\partial}{\partial\theta}\ln\prod_{i=1}^{n} f\left(X_i;\theta\right)\right]^2}$$

由于

$$\begin{aligned}&\mathrm{E}\left[\frac{\partial}{\partial\theta}\ln\prod_{i=1}^{n} f\left(X_i;\theta\right)\right]^2 = \mathrm{E}\left[\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln f\left(X_i;\theta\right)\right]^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{E}\left\{\left[\frac{\partial}{\partial\theta}\ln f\left(X_i;\theta\right)\right]\left[\frac{\partial}{\partial\theta}\ln f\left(X_j;\theta\right)\right]\right\} \\ &= n\mathrm{E}\left[\frac{\partial}{\partial\theta}\ln f(X;\theta)\right]^2 + \sum_{i=1}^{n}\sum_{j\ne i}^{n}\mathrm{E}\left\{\left[\frac{\partial}{\partial\theta}\ln f\left(X_i;\theta\right)\right]\left[\frac{\partial}{\partial\theta}\ln f\left(X_j;\theta\right)\right]\right\}\end{aligned}$$

注意到

$$\mathrm{E}\left[\frac{\partial}{\partial\theta}\ln f(X;\theta)\right] = \int\left[\frac{\partial}{\partial\theta}\ln f(x;\theta)\right] f(x;\theta)dx = \int\frac{\partial}{\partial\theta} f(x;\theta)dx = \frac{\partial}{\partial\theta}\int f(x;\theta)dx = 0$$

当 $i\ne j$ 时，X_i 和 X_j 独立

$$\mathrm{E}\left\{\left[\frac{\partial}{\partial\theta}\ln f\left(X_i;\theta\right)\right]\left[\frac{\partial}{\partial\theta}\ln f\left(X_j;\theta\right)\right]\right\} = 0$$

$\Longrightarrow$

$$\mathrm{E}\left[\frac{\partial}{\partial\theta}\ln\prod_{i=1}^{n} f\left(X_i;\theta\right)\right]^2 = n\mathrm{E}\left[\frac{\partial}{\partial\theta}\ln f(X;\theta)\right]^2$$

定理成立，并且等号成立当且仅当存在常数 $K=K(\theta,n)$ 满足

$$\frac{\partial}{\partial\theta}\ln\left[\prod_{i=1}^{n} f\left(x_i;\theta\right)\right] = K(\theta,n)\left[T\left(X_1,\cdots,X_n\right)-\tau(\theta)\right]$$

□

【注 2.37】 如果 θ 的极大似然估计 $\hat\theta=\hat\theta\left(X_1,\cdots,X_n\right)$ 是下面方程的解

$$\frac{\partial}{\partial\theta}\ln L(\theta) \equiv \frac{\partial}{\partial\theta}\ln\prod_{i=1}^{n} f\left(x_i;\theta\right) = 0$$

如果 $T^*=T^*\left(X_1,\cdots,X_n\right)$ 是 $\tau^*(\theta)$ 的无偏估计,且其方差可以达到C-R下界,则 $T^*\left(x_1,\cdots,x_n\right)=\tau^*\left(\hat\theta\left(x_1,\cdots,x_n\right)\right)$

【注 2.38】 在二阶导数存在，微分和积分可以交换顺序的假设下，下列等式成立

$$\mathrm{E}\left\{\left[\frac{\partial}{\partial\theta}\ln f(X;\theta)\right]^2\right\} = -\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\ln f(X;\theta)\right].$$

证明. 注意到

$$\mathrm{E}\left[\frac{\partial}{\partial\theta}\ln f(X;\theta)\right]=0$$

两端对参数求导：

$$0=\frac{\partial}{\partial\theta}\mathrm{E}\left[\frac{\partial}{\partial\theta}\ln f(X;\theta)\right]=\frac{\partial}{\partial\theta}\int\left[\frac{\partial}{\partial\theta}\ln f(x;\theta)\right]f(x;\theta)dx=\int\frac{\partial^2\ln f(x;\theta)}{\partial\theta^2}f(x;\theta)dx+\int\left[\frac{\partial}{\partial\theta}\ln f(x;\theta)\right]\frac{\partial f(x;\theta)}{\partial\theta}dx$$
$$=\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\ln f(X;\theta)\right]+\int\left[\frac{\partial}{\partial\theta}\ln f(x;\theta)\right]\frac{\partial\ln f(x;\theta)}{\partial\theta}f(x;\theta)dx=\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\ln f(X;\theta)\right]+\mathrm{E}\left\{\left[\frac{\partial}{\partial\theta}\ln f(X;\theta)\right]^2\right\}.$$

□

【定义 2.39】 C-R 不等式中的

$$I(\theta)=\mathrm{E}\left[\frac{\partial}{\partial\theta}\ln f(X;\theta)\right]^2$$

称为 Fisher 信息量.

【定理 2.40】(Lehmann-Scheffé) 令 $X_1,\cdots,X_n$ 是来自 $f(x;\theta)$ 的随机样本. 如果 $S=s\left(X_1,\cdots,X_n\right)$ 是完备充分统计量，令 $T^*=T^*(S)$ 是 S 的函数且是 $\tau(\theta)$ 的无偏估计，则 T^* 是 $\tau(\theta)$ 唯一的 UMVUE.

证明. 唯一性：
令 $\tilde{T}$ 是 $\tau(\theta)$ 的任意一个无偏估计，且 $\tilde{T}=\tilde{T}(S)$.

$$\forall\theta\in\Theta,\mathrm{E}\left(T^*-\tilde{T}\right)\equiv 0$$

即 $T^*-\tilde{T}$ 是 0 的无偏估计. 又因为 $T^*-\tilde{T}$ 是完备充分统计量 S 的函数，由完备统计量的定义，$\forall\theta\in\Theta$,

$$P\left(T^*(S)=\tilde{T}(S)\right)=1$$

因此，由 S 的函数构成的 $\tau(\theta)$ 的无偏估计是唯一的

□

练习

【例题 2.41】 Let $X_1,\ldots,X_n$ be a random sample from $N(\theta,1)$.

1. Find the Cramér-Rao lower bound for the variance of unbiased estimators of θ,θ^2, and $P[X>0]$.
2. Is there an unbiased estimator of θ^2 for $n=1$? If so, find it.
3. Is there an unbiased estimator of $P[X>0]$? If so, find it.
4. What is the maximum-likelihood estimator of $P[X>0]$?
5. Is there an UMVUE of θ^2 ? If so, find it.
6. Is there an UMVUE of $P[X>0]$? If so, find it.

Chapter 3

区间估计

§3.1 基础概念

【定义 3.1】 设 $X_1,\cdots,X_n$ 是来自 $f(x;\theta)$ 的样本. 令 $T_1=T_1(X_1,\cdots,X_n)$ 和 $T_2=T_2(X_1,\cdots,X_n)$ 是两个统计量，且满足条件 $T_1\leq T_2$ 和

$$P(T_1<\tau(\theta)<T_2)\equiv\gamma,$$

其中 γ 不依赖于参数 θ，称 γ 为置信系数；称随机区间 (T_1,T_2) 是 $\tau(\theta)$ 的置信系数为 γ 的置信区间；T_1 和 T_2 分别为 $\tau(\theta)$ 的置信下限和置信上限. 随机区间 (T_1,T_2) 取值 (t_1,t_2) 也称为 $\tau(\theta)$ 的置信系数为 γ 的置信区间.

【注 3.2】 T_1,T_2 满足概率 $P(T_1\leq\tau(\theta)\leq T_2)\equiv\gamma$，称 $[T_1,T_2]$ 是 $\tau(\theta)$ 的置信区间.

【定义 3.3】 设 $X_1,\cdots,X_n$ 是来自 $f(x;\theta)$ 的随机样本. 令 $T_1=T_1(X_1,\cdots,X_n)$ 是统计量，且满足

$$P(T_1<\tau(\theta))\equiv\gamma$$

则称 T_1 为 $\tau(\theta)$ 的单边置信下限. 同样地，令 $T_2=T_2(X_1,\cdots,X_n)$ 是统计量，且满足

$$P(\tau(\theta)<T_2)\equiv\gamma$$

则称 T_2 为 $\tau(\theta)$ 的单边置信上限 (γ 不依赖于 θ)

【注 3.4】 如果 $\tau(\theta)$ 是 θ 的严格单调增函数，θ 的置信系数为 γ 的置信区间是 (T_1,T_2)，则 $\tau(\theta)$ 的置信区间是 $(\tau(T_1),\tau(T_2))$，因为

$$P(\tau(T_1)<\tau(\theta)<\tau(T_2))=P(T_1<\theta<T_2)=\gamma$$

§3.2 枢轴量-正态总体参数的置信区间

【定义 3.5】(枢轴量) 我们称满足下列条件的统计量 $S(T,\theta)$ 为枢轴变量

1. $S(T,\theta)$ 包含优良点估计 T 并且是 θ 的函数

2. $S(T,\theta)$ 的分布与 θ 无关

3. $\forall a<b\quad a\leq S(T,\theta)\leq b\implies A\leq\theta\leq B$

3.2.1 单个正态总体的置信区间

假设 $\boldsymbol{X}=(X_1,\cdots,X_n)$ 是从正态总体 $N\left(\mu,\sigma^2\right)$ 抽取的简单随机样本. 记

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

即 $\bar{X}$ 和 S^2 分别为样本均值和样本方差.

1. σ^2 已知，求 μ 的置信区间

解. 显然，μ 的一个良好的点估计是 $\bar{X}=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}X_i$，其分布为 $\bar{X}\sim N\left(\mu,\sigma^2/n\right)$，将其标准化得

$$U=\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}\sim N(0,1)$$

其分布与 μ 无关. 由正态分布的对称性，可得

$$P_\mu\left(\left|\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}\right|\leqslant u_{\alpha/2}\right)=1-\alpha \iff P_\mu\left(\bar{X}-\frac{\sigma}{\sqrt{n}}u_{\alpha/2}<\mu<\bar{X}+\frac{\sigma}{\sqrt{n}}u_{\alpha/2}\right)=1-\alpha$$

此处 $u_{\alpha/2}$ 为标准正态分布的上侧 $\alpha/2$ 分位数. 因此

$$\left[\bar{X}-\frac{\sigma}{\sqrt{n}}u_{\alpha/2},\bar{X}+\frac{\sigma}{\sqrt{n}}u_{\alpha/2}\right]$$

为 μ 的置信系数 $1-\alpha$ 的置信区间. □

2. σ^2 未知，求 μ 的置信区间

解. μ 的良好的点估计仍为 $\bar{X}$，由于 σ^2 未知，随机变量 $U=\sqrt{n}(\bar{X}-\mu)/\sigma$ 在此不能作为枢轴变量，将其中的 σ 用 S（S^2 是样本方差）代替，得到

$$T=\frac{\sqrt{n}(\bar{X}-\mu)}{S}\sim t_{n-1}.$$

可见 T 的表达式与 μ 有关，而其分布与 μ 无关，故取 T 为枢轴变量. 由于 t 分布关于原点对称，令

$$P(|T|\leqslant c)=P\left(-c\leqslant\frac{\sqrt{n}(\bar{X}-\mu)}{S}\leqslant c\right)=1-\alpha$$

置信区间

$$\left[\bar{X}-\frac{S}{\sqrt{n}}t_{n-1}(\alpha/2),\bar{X}+\frac{S}{\sqrt{n}}t_{n-1}(\alpha/2)\right],$$

□

3. μ 已知，求 σ^2 的置信区间

解. 当 μ 已知时，σ^2 的一个良好的无偏估计为 $S_\mu^2=\displaystyle\sum_{i=1}^{n}(X_i-\mu)^2/n$，且 $nS_\mu^2/\sigma^2\sim\chi_n^2$，则取 $T=nS_\mu^2/\sigma^2$ 为枢轴变量，其表达式与 σ^2 有关，但其分布与 σ^2 无关，找 c_1 和 c_2 使得

$$P_{\sigma^2}\left(c_1\leqslant\frac{nS_\mu^2}{\sigma^2}\leqslant c_2\right)=1-\alpha$$

一般令 c_1 和 c_2 满足下列要求：

$$P_{\sigma^2}\left(\frac{nS_\mu^2}{\sigma^2}<c_1\right)=\frac{\alpha}{2},\quad P_{\sigma^2}\left(\frac{nS_\mu^2}{\sigma^2}>c_2\right)=\frac{\alpha}{2}$$

由 χ^2 分布的上侧分位数表可知 $c_1=\chi_n^2(1-\alpha/2),c_2=\chi_n^2(\alpha/2)$，即有

$$P_{\sigma^2}\left(\chi_n^2\left(1-\frac{\alpha}{2}\right)\leqslant\frac{nS_\mu^2}{\sigma^2}\leqslant\chi_n^2\left(\frac{\alpha}{2}\right)\right)=1-\alpha$$

置信区间

$$\left[\frac{nS_\mu^2}{\chi_n^2(\alpha/2)}, \frac{nS_\mu^2}{\chi_n^2(1-\alpha/2)}\right]$$

□

4. μ 未知，求 σ^2 的置信区间

解. 记 $\theta=\left(\mu,\sigma^2\right)$ 。此时 $S^2=\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2/(n-1)$ 是 σ^2 的良好估计，它是无偏的，且 $(n-1)S^2/\sigma^2\sim\chi_{n-1}^2$. 取 $T=(n-1)S^2/\sigma^2$ 为枢轴变量，其表达式与 σ^2 有关，而其分布与 σ^2 无关. 找 d_1 和 d_2 ，使得

$$P_\theta\left(d_1\leqslant\frac{(n-1)S^2}{\sigma^2}\leqslant d_2\right)=1-\alpha$$

类似于 3，取 $d_1=\chi_{n-1}^2(1-\alpha/2), d_2=\chi_{n-1}^2(\alpha/2)$ ，故有

$$P_\theta\left(\chi_{n-1}^2(1-\alpha/2)\leqslant\frac{(n-1)S^2}{\sigma^2}\leqslant\chi_{n-1}^2(\alpha/2)\right)=1-\alpha$$

置信区间

$$\left[\frac{(n-1)S^2}{\chi_{n-1}^2(\alpha/2)}, \frac{(n-1)S^2}{\chi_{n-1}^2(1-\alpha/2)}\right].$$

□

3.2.2 两个正态总体参数的置信区间

设 $X_1,\cdots,X_m$ 是自正态总体 $N\left(a,\sigma_1^2\right)$ 抽取的简单随机样本，$Y_1,\cdots,Y_n$ 是自正态总体 $N\left(b,\sigma_2^2\right)$ 抽取的简单随机样本，且 $X_1,\cdots,X_m$ 和 $Y_1,\cdots,Y_n$ 独立. 设 $\bar{X},\bar{Y}$ 和 S_1^2,S_2^2 分别为这两组样本的样本均值和样本方差.

A. 均值差 $b-a$ 的置信区间

1. 当 $m=n$ 时，令 $Z_i=Y_i-X_i, i=1,2,\cdots,n$ ，且记 $\tilde{\mu}=b-a, \tilde{\sigma}^2=\sigma_1^2+\sigma_2^2$ ，则有

   $$Z_i\sim N\left(\tilde{\mu},\tilde{\sigma}^2\right),\quad i=1,2,\cdots,n$$

   这就转化为单个正态总体当 $\tilde{\sigma}^2$ 未知，求其均值 $\tilde{\mu}$ 的置信区间问题. 显见 $\bar{Z}=\bar{Y}-\bar{X}$ 是 $\tilde{\mu}$ 的一个良好的无偏估计，枢轴变量

   $$T_Z=\frac{\sqrt{n}(\bar{Z}-\tilde{\mu})}{S_Z}\sim t_{n-1}$$

   此处 $S_Z^2=\sum_{i=1}^{n}\left(Z_i-\bar{Z}\right)^2/(n-1), T_Z$ 的表达式与 $\tilde{\mu}=b-a$ 有关，但其分布与 $\tilde{\mu}$ 无关，因此取 T_Z 为枢轴变量. 由前面已讨论过的情形的结果，可知 $\tilde{\mu}=b-a$ 的置信系数为 $1-\alpha$ 的置信区间为

   $$\left[\bar{Z}-\frac{S_Z}{\sqrt{n}}t_{n-1}(\alpha/2), \bar{Z}+\frac{S_Z}{\sqrt{n}}t_{n-1}(\alpha/2)\right]$$

2. 当 σ_1^2 和 σ_2^2 已知时，易知 $\bar{Y}-\bar{X}$ 为 $b-a$ 的一个良好的无偏估计，且

   $$\bar{Y}-\bar{X}\sim N\left(b-a,\sigma_1^2/m+\sigma_2^2/n\right)\implies U:=\frac{\bar{Y}-\bar{X}-(b-a)}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}\sim N(0,1)$$

   U 的表达式与 $b-a$ 有关，但其分布与 $b-a$ 无关，取 U 为枢轴变量，故有

   $$P_{a,b}\left(\left|\frac{\bar{Y}-\bar{X}-(b-a)}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}\right|\leqslant u_{\alpha/2}\right)=1-\alpha.$$

   等价变形得到 $b-a$ 的置信系数为 $1-\alpha$ 的置信区间为

   $$\left[\bar{Y}-\bar{X}-u_{\alpha/2}\sqrt{\sigma_1^2/m+\sigma_2^2/n}, \bar{Y}-\bar{X}+u_{\alpha/2}\sqrt{\sigma_1^2/m+\sigma_2^2/n}\right].$$

3. 当 $\sigma_1^2=\sigma_2^2=\sigma^2$ 未知时，令

$$S_\omega^2=\frac{1}{m+n-2}\left[(m-1)S_1^2+(n-1)S_2^2\right]=\frac{1}{m+n-2}\left[\sum_{i=1}^{m}(X_i-\bar{X})^2+\sum_{i=1}^{n}(Y_i-\bar{Y})^2\right]$$

显然 $\bar{Y}-\bar{X}$ 是 $b-a$ 的无偏估计，

$$T_\omega=\frac{\bar{Y}-\bar{X}-(b-a)}{S_\omega}\sqrt{\frac{mn}{m+n}}\sim t_{n+m-2}$$

T_ω 的表达式与 $b-a$ 有关，但其分布与 $b-a$ 无关，取 T_ω 为枢轴变量，故有

$$P\left(\left|\frac{\bar{Y}-\bar{X}-(b-a)}{S_\omega}\right|\sqrt{\frac{mn}{m+n}}\leqslant t_{m+n-2}(\alpha/2)\right)=1-\alpha$$

$b-a$ 的置信系数为 $1-\alpha$ 的置信区间为

$$\left[\bar{Y}-\bar{X}-S_\omega t_{m+n-2}\left(\frac{\alpha}{2}\right)\sqrt{\frac{1}{m}+\frac{1}{n}},\bar{Y}-\bar{X}+S_\omega t_{m+n-2}\left(\frac{\alpha}{2}\right)\sqrt{\frac{1}{m}+\frac{1}{n}}\right],$$

4. 当 $\sigma_1^2\neq\sigma_2^2$ 皆未知时，求 $b-a$ 的置信区间问题．这是著名的 Behrens-Fisher 问题

   - 当 m 与 n 都充分大时可用大样本方法，由于

   $$\frac{\bar{Y}-\bar{X}-(b-a)}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}\sim N(0,1),$$

   且当 $m\to\infty$ 时 $S_1^2\xrightarrow{P}\sigma_1^2$，当 $n\to\infty$ 时 $S_2^2\xrightarrow{P}\sigma_2^2$，将式（4．2．8）中的 σ_1^2 和 σ_2^2 分别用 S_1^2 和 S_2^2 代入，利用引理 2．5．1 可知，当 $m,n\to\infty$ 时，有

   $$\widetilde{U}=\frac{\bar{Y}-\bar{X}-(b-a)}{\sqrt{S_1^2/m+S_2^2/n}}=\frac{\bar{Y}-\bar{X}-(b-a)}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}\cdot\frac{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}{\sqrt{S_1^2/m+S_2^2/n}}\xrightarrow{\mathscr{L}}N(0,1).$$

   因此，取 $\widetilde{U}$ 为枢轴变量．当 m,n 充分大时，$b-a$ 的置信系数近似为 $1-\alpha$ 的置信区间是

   $$\left[\bar{Y}-\bar{X}-u_{\alpha/2}\sqrt{\frac{S_1^2}{m}+\frac{S_2^2}{n}},\bar{Y}-\bar{X}+u_{\alpha/2}\sqrt{\frac{S_1^2}{m}+\frac{S_2^2}{n}}\right]$$

   -

B. 方差比 σ_1^2/σ_2^2 的置信区间

1. 若 a 和 b 已知，记 $S_a^2=\sum_{i=1}^{m}(X_i-a)^2/m,S_b^2=\sum_{i=1}^{n}(Y_i-a)^2/n$，显见 $mS_a^2/\sigma_1^2\sim\chi_m^2,nS_b^2/\sigma_2^2\sim\chi_n^2$，且 S_a^2 为 σ_1^2 的无偏估计，S_b^2 为 σ_2^2 的无偏估计，且二者独立，故由推论 2．4．4 可知

$$F=\frac{S_a^2/\sigma_1^2}{S_b^2/\sigma_2^2}=\frac{S_a^2}{S_b^2}\cdot\frac{\sigma_2^2}{\sigma_1^2}\sim F_{m,n}$$