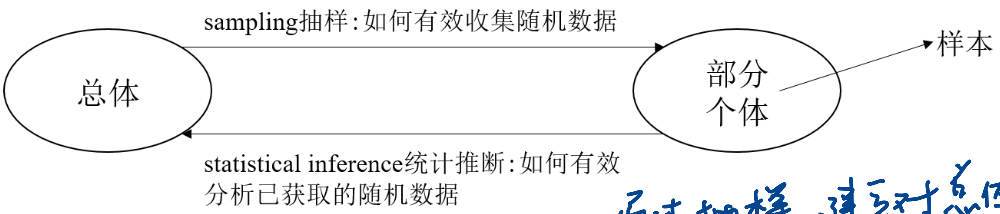
8 统计的基本概念

到19世纪末20世纪初, 随着近代数学和概率论的发展, 诞生了《数理统计》这门学科.

数理统计: 以概率论为基础, 研究如何有效收集研究对象的随机数据资料, 以及如何运用所获得的数据去揭示研究问题统计规律的一个学科.



通过抽样，建立对总体的估计，从而形成假设建模评估.

数理统计研究内容: i) 抽样; ii) 参数估计; iii) 假设检验.

8.1 总体(population) VS 样本(sample)

总体: 研究问题所涉及的对象全体;

个体: 总体中每个元素称为个体.

总体分为有限或无限总体. 例如: 全国人民的收入是总体, 一个人的收入是个体.

在研究总体时, 通常关心总体的某项或某些数量指标X, 而总体的数量指标X常是一随机变量. 因此对总体的研究归纳为对随机变量X的分布或其数字特征的研究. 故总体分布与随机变量X的分布不再区分, 常称总体X.

总体: 研究对象的全体 $\Rightarrow$ 数据 $\Rightarrow$ 分布(一般是未知的).

样本: 从总体中随机抽取的一些个体, 一般表示为$X_1, X_2, \cdots, X_n$, 称$X_1, X_2, \cdots, X_n$抽取自总体X的随机样本, 其样本容量为n

抽样: 抽取样本的过程.

样本值: 对样本观察得样本的数值, 例如: $X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n$为样本观察值或样本值.

样本的二重性: i) 就一次具体观察而言, 样本值是确定的数; ii)不同的抽样下, 样本值会发生变化, 可看作随机变量.

1)随机性: 多次采样, 值随机 ; 2)确定性: 一次采样, 值固定.

定义8.1 (简单随机样本)**.** *称样本$X_1, X_2, \cdots, X_n$是总体X的简单随机样本(简称样本), 是指样本满足: 1) 代表性, 即X_i与X同分布; 2) 独立性, 即$X_1, X_2, \cdots, X_n$之间相互独立.*

后面我们所考虑的样本均为简单随机样本.

离散型、连续型.

设总体X的联合分布函数为$F(x)$, 则$X_1, X_2, \cdots, X_n$的联合分布函数为

$$F(x_1, x_2, \cdots, x_n) = \underline{\prod_{i=1}^{n} F(x_i)}.$$

独立性.

连续型. 若总体X的概率密度为$f(x)$, 则样本$X_1, X_2, \cdots, X_n$的联合概率密度为

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

→ 独立性.

离散型. 若总体X的分布列$\Pr(X = x_i)$, 则样本$X_1, X_2, \cdots, X_n$的联合分布式为

$$\Pr(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n) = \prod_{i=1}^{n} \Pr(X_i = x_i).$$

N个高斯分布的加权和能够逼近任意分布.

8.2 常用统计量

定义8.2. 设$X_1, X_2, \cdots, X_n$是总体X的一个样本, $g(X_1, X_2, \cdots, X_n)$是$X_1, X_2, \cdots, X_n$的一个函数. 若g连续且不含任意参数, 称$g(X_1, X_2, \cdots, X_n)$是一个统计量. 统计量也是随机变量.

统计量是随机变量. $g(x_1, x_2, \cdots, x_n)$为$g(X_1, X_2, \cdots, X_n)$的观察值. 我们研究常用统计量: 假设$X_1, X_2, \cdots, X_n$是总体X 的一个样本, 则样本均值定义为:

样本均值.

$$\boxed{\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,}$$

公式总结:
样本均值 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
样本方差 $S_0^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2$
样本均值的期望:
$E(\bar{X}) = \mu$
样本均值的方差:
$Var(\bar{X}) = \frac{\sigma^2}{n}$
样本方差的均值:
$E[S_0^2] = \frac{n-1}{n}\sigma^2$

根据独立同分布可得 每一个样本都与总体同分布.

引理8.1. 设总体X的期望为$\mu = E[X]$, 方差$\sigma^2 = Var(X)$, 则有

$$E[\bar{X}] = \mu, \qquad Var(\bar{X}) = \sigma^2/n, \qquad \bar{X} \xrightarrow{d} \mathcal{N}(\mu, \sigma^2/n).$$

定义样本方差为

$$S_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2.$$

根据

$$E(\bar{X}^2) = E\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2\right] = \frac{\sigma^2}{n} + \mu^2,$$

我们得到

$$E(S_0^2) = E(X_i^2) - E(\bar{X}^2) = \underbrace{\sigma^2 + \mu^2}_{\text{iid}} - \frac{\sigma^2}{n} - \mu^2 = \frac{n-1}{n}\sigma^2.$$

样本方差与总体方差σ^2有偏差. 进一步定义样本标准差为:

$$S_0 = \sqrt{S_0^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

$$E\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2\right]$$
$$= E\left[\frac{1}{n^2}\left(\sum_{i=1}^{n} X_i^2 + \sum_{i\neq j} X_i X_j\right)\right]$$
$$= \frac{1}{n^2}E\left[\sum_{i=1}^{n} X_i^2\right] + \frac{1}{n^2}E\left[\sum_{i\neq j} X_i X_j\right]$$
$$= \frac{\sigma^2+\mu^2}{n} + \frac{n-1}{n}\mu^2 = \frac{\sigma^2}{n} + \mu^2.$$

S：修正后样本方差、S_0：修正前有偏样本方差.

修正后的样本方差为:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \Longrightarrow S^2 = \frac{n}{n-1}S_0^2,$$

所以$E(S^2) = \sigma^2$.

样本k阶原点矩:

$$A_k = \frac{1}{n}\sum_{i=1}^{n}X_i^k, \qquad k = 1, 2, \cdots.$$

维度更高

样本期望即为1阶原点矩

样本方差即为2阶中心矩

样本k阶中心矩为:

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

例8.1. 设总体$X \sim \mathcal{N}(20, 3)$, 从中抽取两独立样本, 容量分别为10和15. 求这两个样本均值之差的绝对值大于0.3的概率.

总体 → 样本分布 → 中心极限定理.

样本均值 → 实际上和中心极限那里的题目区别仅在于Σ…换成了样本均值

解. 根据中心极限定理近似有

$$\bar{X}_1 = \frac{1}{10}\sum_{i=1}^{10}X_i \sim \mathcal{N}(20, 3/10), \qquad \bar{X}_2 = \frac{1}{15}\sum_{i=1}^{15}X_i' \sim \mathcal{N}(20, 3/15).$$

所以$\bar{X}_1 - \bar{X}_2 \sim \mathcal{N}(0, \frac{1}{2})$, 进一步得到

$$\Pr(|\bar{X}_1 - \bar{X}_2| > 0.3) = 2 - 2\Phi(0.3/\sqrt{0.5}).$$

□

最小最大次序统计量分别定义为:

$$X_{(1)} = \min\{X_1, X_2, \cdots, X_n\}, \qquad X_{(n)} = \max\{X_1, X_2, \cdots, X_n\},$$

且定义样本极差为 可用于归一化.

$$R_n = X_{(n)} - X_{(1)}.$$

设总体X的分布函数为$F(x)$, 则

$$F_{X_{(1)}}(x) = \Pr(X_{(1)} \le x) = 1 - \Pr(X_{(1)} > x) = 1 - (1 - F(x))^n, \qquad F_{X_{(n)}}(x) = F^n(x).$$

定理8.1. 设总体X的密度函数为$f(x)$, 分布函数为$F(x)$, $X_1, X_2, \cdots, X_n$是样本, 则第k个次序统计量$X_{(k)}$的密度函数为

表示样本中第k小元素

$$f_k(x) = \frac{n!}{(k-1)!(n-k)!}(F(x))^{k-1}f(x)(1 - F(x))^{n-k}.$$

补充知识点:

定义8.3 (Γ分布). 如果随机变量X的概率密度

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x} & x > 0 \\ 0 & x \le 0 \end{cases},$$

其中α和λ为正常数, 则称随机变量X服从参数为α和λ的Γ分布, 记为$X \sim \Gamma(\alpha, \lambda)$.

上述定义中Γ函数为:

$$\Gamma(\alpha)=\int_0^{+\infty} x^{\alpha-1}e^{-x}dx,\quad (\alpha>0).$$

根据上式有$\Gamma(1)=1$, $\Gamma(1/2)=\sqrt{\pi}$.
τ分布的可加性:

定理8.2. *若$X\sim\Gamma(\alpha_1,\lambda)$, $Y\sim\Gamma(\alpha_2,\lambda)$, 且$X$与$Y$独立, 则$X+Y\sim\Gamma(\alpha_1+\alpha_2,\lambda)$.*

上述定理的证明留作习题. 特别地, 当$\alpha=1/2$和$\lambda=1/2$时, 有

$$f(x)=\begin{cases}\frac{1}{\sqrt{2\pi}}x^{-\frac{1}{2}}e^{-\frac{1}{2}} & x>0\\ 0 & x\le 0\end{cases}$$

例8.2. *若$X\sim\mathcal{N}(0,1)$, 则$X^2\sim\Gamma(1/2,1/2)$.*

解. 首先求解随机变量函数$Y=X^2$的分布函数: 当$y>0$时,

$$F_Y(y)=\Pr(X^2\le y)=\Pr(-\sqrt{y}\le X\le\sqrt{y})=\int_{-\sqrt{y}}^{\sqrt{y}}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},$$

由此得到概率密度为$f_Y(y)=\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{y}}e^{-\frac{y}{2}}$. 当$y\le 0$时有$f_Y(y)=0$. 从而得到$X^2\sim\Gamma(1/2,1/2)$. □

8.3 正态总体抽象分布定理

8.3.1 χ^2分布

定义8.4. *若$X_1,X_2,\cdots,X_n$是总体为$\mathcal{N}(0,1)$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$\chi^2$分布, 记为$Y\sim\chi^2(n)$.*

根据$X_1^2\sim\Gamma(1/2,1/2)$, 以及Γ函数的可加性, 可得$Y\sim\Gamma(n/2,1/2)$. 因此Y的概率密度为

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

下面研究χ^2分布的性质:

定理8.3. *若$X\sim\chi^2(n)$, 则$E(X)=n$, $Var(X)=2n$; 若$X\sim\chi^2(m)$, $Y\sim\chi^2(n)$且独立, 则$X+Y\sim\chi^2(m+n)$;*

Proof. 若$X\sim\chi^2(n)$, 设$X=X_1+X_2+\cdots+X_n$, 其中$X_1,X_2,\cdots,X_n$独立同分布于$\mathcal{N}(0,1)$, 则有

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

计算

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

可得$\mathrm{Var}(X)=2n$. □

更一般的结论: 若$X\sim\mathcal{N}(0,1)$, 则

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

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

例8.3. 设X_1,X_2,X_3,X_4是来自于$\mathcal{N}(0,4)$的样本, $Y=a(X_1-2X_2)^2+b(3X_3-4X_4)^2$. 求$a,b$ 取何值时, Y服从χ^2分布, 并求其自由度.

解. 根据正太分布的性质有$X_1-2X_2\sim\mathcal{N}(0,20)$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$X+Y\sim\Gamma(\alpha_1+\alpha_2,\lambda)$.

8.3.2 t分布

定义8.5. 随机变量$X\sim\mathcal{N}(0,1)$, $Y\sim\chi^2(n)$, X与Y独立, 则$T=X/\sqrt{Y/n}$服从自由度为n的t-分布, 记$T\sim t(n)$.

随机变量$T\sim t(n)$, 其概率密度为

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

由此可知t分布的密度函数$f(x)$是偶函数. 当$n\to\infty$时,

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

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

8.3.3 F分布

定义8.6. 设随机变量$X \sim \chi^2(m)$, $Y \sim \chi^2(n)$, 且X与Y独立, 称$F = \frac{X/m}{Y/n}$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

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

注: 若$F \sim F(m,n)$, 则$\frac{1}{F} = F(n,m)$.

课题练习:

1) 随机变量$X_1, X_2, \cdots, X_n$相互独立, 且$X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, 求$\sum_{i=1}^n (X_i - \mu_i)^2/\sigma_i^2$的分布.

2) 总体$X \sim \mathcal{N}(0,9)$与总体$Y \sim \mathcal{N}(0,9)$独立, $X_1, X_2, \cdots, X_9$和$Y_1, Y_2, \cdots, Y_9$分别为来自总体X和Y的两样本, 求$(X_1 + X_2 + \cdots + X_9)/\sqrt{Y_1^2 + Y_1^2 + \cdots + Y_n^2}$的分布.

3) 设X_1, X_2, 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.

8.3.4 正态分布的抽样分布定理

定理8.4. 设$X_1, X_2, \cdots, X_n$是取自总体$\mathcal{N}(\mu, \sigma^2)$的样本, 则有

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

无偏方差.

定理8.5. 随机变量$X_1, X_2, \cdots, X_n$是取自总体$\mathcal{N}(\mu, \sigma^2)$的样本, 设$\bar{X} = \sum_{i=1}^n X_i/n$和$S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$, 则有

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1), \qquad 且\ \bar{X}\ 和\ S^2\ 相互独立.$$

样本方差可转换为χ^2分布来计算,且不需要知道μ的具体大小

此定理证明参考书的附件.

结论是易于猜测的:样本方差并不依赖于总体的均值.

定理8.6. 设$X_1, X_2, \cdots, X_n$是取自总体$\mathcal{N}(\mu, \sigma^2)$的样本, 则$\bar{X} = \sum_{i=1}^n X_i/n$与$S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$相互独立, 且

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

在已知S和总体均值的情况下,样本均值的分布可转化为t分布来计算,且不需要知道总体均值的大小.

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

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

□

定理8.7. 设$X \sim \mathcal{N}(\mu_1, \sigma^2)$, $Y \sim \mathcal{N}(\mu_2, \sigma^2)$, 且$X$与$Y$独立. 设$X_1, X_2, \cdots, X_m$, $Y_1, Y_2, \cdots, Y_n$分别来自总体X和Y的两个样本, $\bar{X}, \bar{Y}$为样本均值, S_X^2和S_Y^2为修正样本方差, 则

应用、在X和Y的均值相同时, 可以由样本统计量$S_1^2 \cdot S_2^2$计算出分布.

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}}\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t(n_1 + n_2 - 2).$$

Proof. 对正太分布有$\bar{X} \sim \mathcal{N}(\mu_1, \sigma^2/m)$和$\bar{Y} \sim \mathcal{N}(\mu_2, \sigma^2/n)$, 并$\bar{X} - \bar{Y} \sim \mathcal{N}\left(\mu_1 - \mu_2, \left(\frac{1}{m} + \frac{1}{n}\right)\sigma^2\right)$, 进一步有

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sigma^2\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim \mathcal{N}(0, 1).$$

根据定理 8.5 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8.8. 设$X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$, 且$X$与$Y$独立. 设$X_1, X_2, \cdots, X_m$, $Y_1, Y_2, \cdots, Y_n$分别来自总体X和Y 的两个样本, S_X^2和S_Y^2为样本的修正样本方差, 则有

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

两总体采样后可能得到的分布关系。

Proof. 根据定理 8.5有$\frac{(m-1)S_X^2}{\sigma_X^2} \sim \chi^2(m-1)$和$\frac{(n-1)S_Y^2}{\sigma_Y^2} \sim \chi^2(n-1)$, 所以

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

□

课题练习题:

1) 若$X \sim t(n)$, 求$Y = X^2$的分布.

2) 若总体$X \sim \mathcal{N}(0, 1)$, $X_1, X_2, \cdots, X_5$为一样本. 设$Y = c_1(X_1 + X_3)^2 + c_2(X_2 + X_4 + X_5)^2$. 求常数$c_1, c_2$使$Y$服从$\chi^2$分布.

3) 设X_1, X_2是总体$\mathcal{N}(0, \sigma^2)$的样本, 求$\frac{(X_1+X_2)^2}{(X_1-X_2)^2}$的分布.

4) 设$X_1, X_2, \cdots, X_{10}$是总体$\mathcal{N}(\mu, \frac{1}{4})$的样本, i) 若$\mu = 0$, 求$\Pr(\sum_{i=1}^{10} X_i^2 \geq 4)$; ii) 若$\mu$未知, 求$\Pr(\sum_{i=1}^{10}(X_i - \bar{X})^2 \geq 2.85)$.

5) 设$X_1, X_2, \cdots, X_{25}$是总体$\mathcal{N}(12, \sigma^2)$的样本, i) 若$\sigma = 2$, 求$\Pr(\frac{1}{25}\sum_{i=1}^{25} X_i \geq 12.5)$; ii) 若$\sigma$未知, 求$\Pr(\frac{1}{25}\sum_{i=1}^{25} X_i \geq 12.5)$(此时样本方差$S^2 = 5.57$).

8.3.5 分位数(点)

定义8.7. 设X为一随机变量, 给定$\alpha \in (0,1)$, 称满足$\Pr(X > \lambda_\alpha) = \alpha$的实数$\lambda_\alpha$为上侧$\alpha$分位数*(点)*.

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

$\chi^2(n)$分位点: $X \sim \chi^2(n)$, 对给定$\alpha \in (0,1)$, 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$\alpha$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$\mu_\alpha$表示正态分布上侧$\alpha$分位点.

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

F分布的分位点: $X \sim F(m,n)$, 对给定$\alpha \in (0,1)$, 满足$\Pr[X > F_\alpha(m,n)] = \alpha$的点$F_\alpha(m,n)$称为$F(m,n)$分布上侧$\alpha$分位点. 对于$F$分布, 我们有如下性质:

引理8.2. 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$F_\alpha(n,m) = \frac{1}{F_{1-\alpha}(m,n)}$. □

习题: 1) $X_1 \cdots X_5$ 是总体 $X \sim N(0,1)$ 的一个样本，

$Y = C_1(X_1+X_3)^2 + C_2(X_2+X_4+X_5)^2 \sim \chi(2)$

则 $C_1 = ?$ $\quad C_2 = ?$

$\frac{1}{2}$ $\quad \frac{1}{3}$

2) X_1, X_2, i.i.d. $N(0,\sigma^2)$，求 $\frac{(X_1+X_2)^2}{(X_1-X_2)^2} \sim ?$

∵ $X_1+X_2 \sim N(0, 2\sigma^2)$，∴ $\frac{X_1+X_2}{\sqrt{2\sigma^2}} \sim N(0,1)$.

$\frac{X_1-X_2}{\sqrt{2\sigma^2}} \sim N(0,1)$

∴ $$\frac{\frac{(X_1+X_2)^2}{(\sqrt{2\sigma^2})^2}}{\frac{(X_1-X_2)^2}{(\sqrt{2\sigma^2})^2}} = \frac{(X_1+X_2)^2}{(X_1-X_2)^2} \sim F(1,1)$$

3) $X_1 \cdots X_{10}$ 是 $N(\mu, \frac{1}{4})$ 的样本.

i) $\mu=0$ 时，求 $Pr[\sum_{i=1}^{10} X_i^2 \geqslant 4]$

∵ $2X_i \sim N(0,1)$，∴ $\sum_{i=1}^{10} X_i^2 = \frac{1}{4}\sum_{i=1}^{10}(2X_i)^2$

∴ 原概率 $= Pr[\sum_{i=1}^{10}(2X_i)^2 \geqslant 16] = Pr[Y \geqslant 16]$

ii) μ 未知时，求 $Pr[\sum_{i=1}^{10}(X_i-\bar{X})^2 \geqslant 2.85]$

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

∴ $Pr[\frac{\sum_{i=1}^{10}(X_i-\bar{X})}{0.25} \geqslant 2.85 \times 4]$ 计算 χ^2 分布即可.

因此，μ 已知时可直接使用分布的性质计算.

μ 未知时可使用前述公理进行计算.

(5) $X_1 \cdots X_{25}$ 是 $N(12, \sigma^2)$ 样本

~~[illegible]~~

σ未知时，$S^2 = 5.57$，求 $Pr\left[\frac{1}{25}\sum_{i=1}^{25} X_i \geq 12.5\right]$