概率论与数理统计

Probability and

Mathematical Statistics

第六章 数理统计的基本概念

6.1 总体与样本

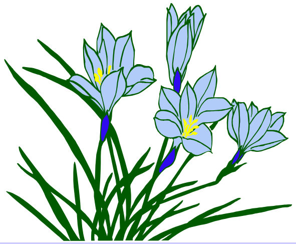
一 总体与个体的概念

二 简单随机样本

三 理论分布函数和经验分布函数

四 统计量

6.2 抽样分布



6.1 总体与个体

一 总体(母体)与个体

总体 ⟷ 研究对象全体组成的集合

R.V.X $\xleftrightarrow{\text{随机性}}$ 数量指标的所有取值的集合

个体 ⟷ 总体的每个元素

注　总体X $\begin{cases}\text{一维}\\ \text{多维}\end{cases}$ $\begin{cases}\text{有限}\\ \text{无限}\end{cases}$ $\begin{cases}\text{离散}\\ \text{连续}\end{cases}$

二 简单(随机)样本

抽样：从总体中抽取部分的手续。

(简单)(随机)抽样:

1. 随机性：每次抽取，每个个体都有相等的机会；

2. 独立性：每次抽取互不影响。

观测值:一次抽样的结果就称为总体X的一个观察测值。

简单(随机)样本的定义

设总体X的分布函数为$F(x)$.

若$X_1,X_2,\dots,X_n$ *i.i.d.*且与总体X同分布的R.V.，则称$(X_1,X_2,\dots,X_n)$是总体X(或$F(x)$)的容量为n的**(简单随机)样本**。

当$(X_1,X_2,\dots,X_n)$取定某组常数值$(x_1,x_2,\dots,x_n)$时，称这组常数$(x_1,x_2,\dots,x_n)$为样本$(X_1,X_2,\dots,X_n)$的一组**样本观测值**(或**样本实现**)。

注：试验前：X_1，X_2，…，X_n为R.V.

试验后：x_1，x_2，…，x_n为样本观察值（实数）

三 理论分布和经验分布

1.理论分布: 指总体X的分布。总体X的分布函数$F(x)$称为理论分布函数。

例 设总体$X\sim E(\lambda)$,则样本$(X_1,\cdots,X_n)$的联合概率密度为

$$\prod_{i=1}^{n} f(x_i)=\begin{cases}\lambda^n \exp\{-\lambda\sum_{i=1}^{n} x_i\}, & x_i>0, i=1,\cdots,n.\\ 0, & \text{其它}.\end{cases}$$

例 设总体$X\sim B(1,p)$,则样本$(X_1,\cdots,X_n)$的联合分布列为

$$P(X_1=x_1,\cdots,X_n=x_n)=\prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i},\quad x_i=0,1.$$

2 经验分布函数

设$(X_1,\ldots,X_n)$是总体X的样本，将其观察值$(x_1,\ldots,x_n)$由小到大排列为$x_1^* \le x_2^* \le \ldots \le x_n^*$，称

$$F_n(x)=\begin{cases}0, & x<x_1^*,\\ k/n, & x_k^* \le x < x_{k+1}^*,\ k=1,\cdots,n-1,\\ 1, & x \ge x_n^*.\end{cases}$$

为总体X的经验分布函数。

注1：经验分布函数用$\{X\le x\}$的频率估计总体的分布函数。

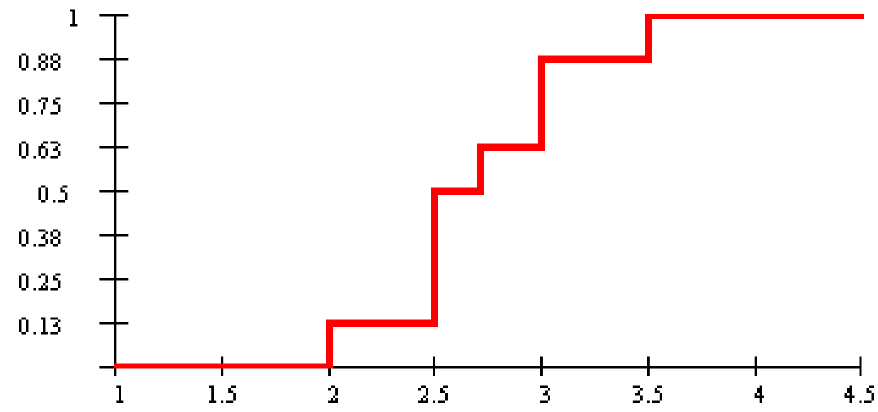
格列汶科定理 $P(\lim\limits_{n\to\infty}\sup\limits_{-\infty<x<\infty}\left|F_n(x)-F(x)\right|=0)=1$

注2：经验分布函数是分布函数，依赖样本观测值。

例 随机地观测总体X 得8个数据：2.5，3，2.5，3.5，3，2.7，2.5，2，试求X 的一个经验分布函数。

•解　　2 < 2.5 = 2.5 = 2.5 < 2.7 < 3 = 3 < 3.5

$$F_8(x)=\begin{cases} 0 & x<2 \\ 1/8 & 2\le x<2..5 \\ 4/8 & 2.5\le x<2.7 \\ 5/8 & 2.7\le x<3 \\ 7/8 & 3\le x<3.5 \\ 1 & x\ge 3.5 \end{cases}$$



X	2	2.5	2.7	3	3.5
P	1/8	3/8	1/8	2/8	1/8

一般$F_n(x)$对应分布列：

$P(X=x_i)=1/n$，i=1,2,...,n

•格列汶科定理

$$P(\lim_{n\to\infty}\sup_{-\infty<x<\infty}\left|F_n(x)-F(x)\right|=0)=1$$

四 统计量

1 定义 设$X_1, X_2,\ldots,X_n$为来自总体X的一个样本, 若

(1) $T = g(x_1, x_2,\ldots,x_n)$ 连续;

(2) $T = g(x_1, x_2,\ldots,x_n)$ 中不含未知参数,

则称 $T = g(X_1, X_2,\ldots,X_n)$ 为统计量，称 $t= g(x_1, x_2,\ldots,x_n)$ 为统计量观察值。

例 设X_1，X_2，…，X_n为来自总体$N(\mu,\sigma^2)$的样本，其中μ已知，σ^2未知，则 ( ) 不是统计量。

(1) $X_1 + X_n$； (2) $\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$；

(3) $\sum_{i=1}^{n}|X_i|/\sigma$； (4) $\min\limits_{1\le i\le n}\{X_i\}$.

2 常用统计量

- 样本均值 $\bar{X}=\dfrac{1}{n}\sum_{i=1}^{n}X_i \quad\to\quad E(X)$
- 样本方差 $S^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 \quad\to\quad D(X)$

  样本标准差 $S=\sqrt{S^2}$
- 样本k阶原点矩 $A_k=\dfrac{1}{n}\sum_{i=1}^{n}X_i^k \quad\to\quad E(X^k)$
- 样本k阶中心矩 $B_k=\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k \quad\to\quad E(X-EX)^k$

注 $B_2=\dfrac{n-1}{n}S^2=\tilde{S}^2$, $A_1=\bar{X}$, $S^2=\dfrac{1}{n-1}(\sum_{i=1}^{n}X_i^2-n\bar{X}^2)$

顺序统计量 $X_1, X_2, \cdots, X_n \to X_1^* \le X_2^* \le \cdots \le X_n^*$

中位数 $\tilde{X} = \begin{cases} X_{m+1}^*, & n = 2m+1, \\ \frac{1}{2}(X_m^* + X_{m+1}^*), & n = 2m. \end{cases} \to E(X)$

极差 $R = X_n^* - X_1^* \to D(X)$

注 X_i^* 并不是 $X_1, X_2, \dots, X_n$ 中的一个，例如

$$X_1, X_2 \overset{i.i.d}{\sim} B(1, p) \Rightarrow X_1^* \sim B(1, p^2),\ X_2^* \sim B(1,\ 1-(1-p)^2)$$

小结

- 理解理论分布和经验分布函数，统计量的概念
- 深刻理解简单随机样本的概念
- 熟练掌握常见统计量

6.2 抽样分布

——统计量的分布

一 χ^2分布

二 t 分布

三 F分布

四 正态总体的样本均值和样本方差的分布

五 顺序统计量的分布

二 χ^2分布

1 定义 设$(X_1,\ldots,X_n)$是正态总体$N(0,1)$的样本, 则称

$$\boxed{\chi^2 = X_1^2 + X_2^2 + \cdots + X_n^2}$$

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

2 性质

(1)$\chi^2 \sim \chi^2(n) \Rightarrow E\chi^2 = n,\ D\chi^2 = 2n.$

(2)可加性: $\chi_1^2 \sim \chi^2(n_1),\ \chi_2^2 \sim \chi^2(n_2)$,且相互独立,

$$\Rightarrow \chi_1^2 + \chi_2^2 \sim \chi^2(n_1 + n_2).$$

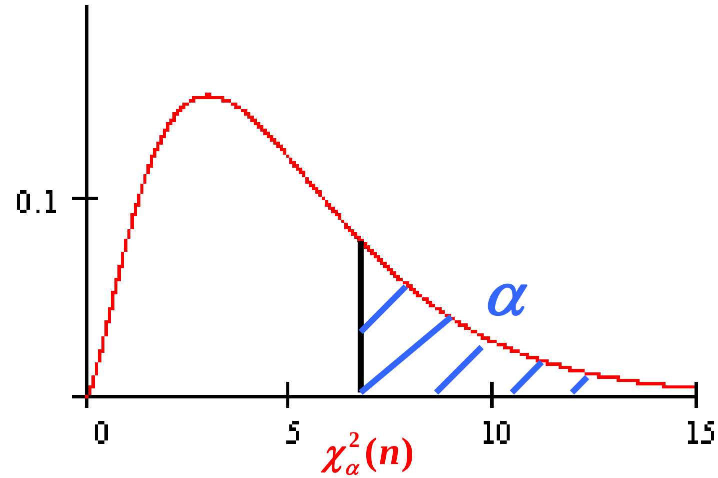
(3)$\chi^2 \sim \chi^2(2) \Rightarrow \chi^2 \sim E(\frac{1}{2}).$

3 上α分位点

对给定$\alpha\,(0<\alpha<1)$，称满足条件

$$P(\chi^2 > \qquad\qquad) = \alpha$$

的点$\chi_\alpha^2(n)$为$\chi^2(n)$的上α分位点。



注 对$n\le 45$，直接查表；例如： $\chi^2_{0.01}(20) = 37.566$

对$n>45$， $\chi^2_\alpha(n) \approx \dfrac{1}{2}(u_\alpha + \sqrt{2n-1})^2.$

例(填空) $X \sim N(0,\sigma^2), \dfrac{X^2}{\sigma^2} \sim$ $\underline{\chi^2(1)}$

例 设总体$X \sim N(0,0.3^2),(X_1,\cdots,X_{10})$是$X$的样本，

求$P(\sum_{i=1}^{10} X_i^2 > 1.44)$.

解 $\because \sum_{i=1}^{10}(\dfrac{X_i}{0.3})^2 \sim \chi^2(10),$

$$\therefore P(\sum_{i=1}^{10} X_i^2 > 1.44) = P(\sum_{i=1}^{10}\frac{X_i^2}{0.3^2} > \frac{1.44}{0.3^2})$$

$$= P(\chi^2(10) > 16) = 0.10$$

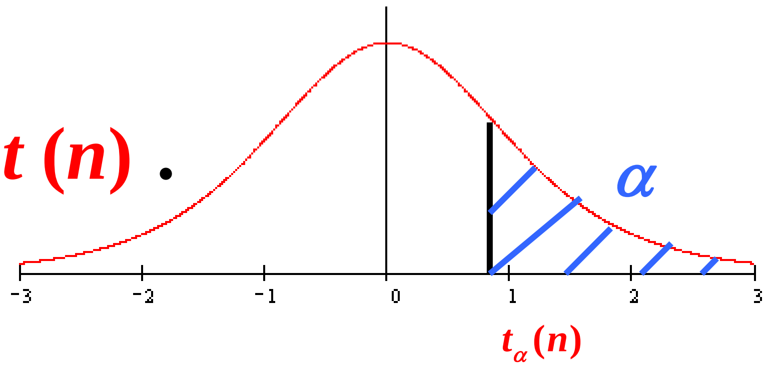
查表

二 t 分布

1 定义 设$X\sim N(0,1)$, $Y\sim \chi^2(n)$, 且X,Y独立,则称

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

服从自由度为 n 的t 分布,记$T\sim t(n)$.



2 性质

(1)t分布的概率密度为偶函数,有$E(T)=0$.

(2)渐近正态: $t(\infty)=N(0,1)$. $t(n)$比$N(0,1)$有较大尾事件概率。

3上α分位点 $P(T>\quad)=\alpha$

注 对$n\leq 45$，查表；$n>45$, $t_\alpha(n)\approx u_\alpha$.

三 F分布

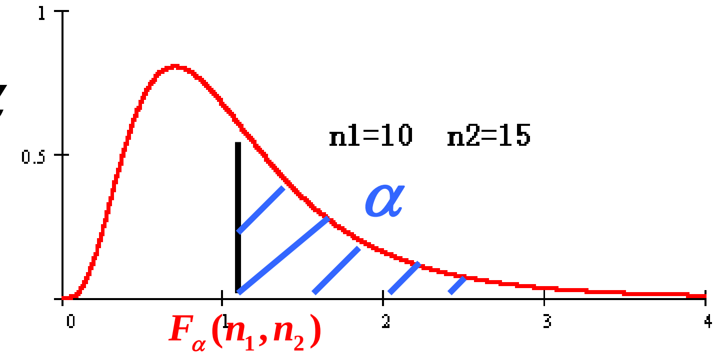
1 定义 设$X\sim\chi^2(n_1)$, $Y\sim\chi^2(n_2)$, 且X,Y独立,则称

$$\boxed{F=\frac{X/n_1}{Y/n_2}}$$

服从自由度为 (n_1,n_2) 的F 分布,记$F\sim F(n_1,n_2)$.

2 性质 $F\sim F(n_1,n_2) \Longrightarrow 1/F\sim F(n_2,n_1)$

3上α分位点 $P(F>\quad\quad)=\alpha$



$$F_{1-\alpha}(n_2,n_1)=\frac{1}{F_\alpha(n_1,n_2)}$$

四 正态总体的样本均值和样本方差的分布

定理 设$X_1, X_2, \ldots, X_n$是取自总体$N(\mu, \sigma^2)$的一个样本，$\bar{X}$和S^2分别为样本均值和样本方差，则

(1) $\bar{X} \sim N(\mu, \dfrac{\sigma^2}{n}) \Leftrightarrow \dfrac{\bar{X}-\mu}{\sigma}\sqrt{n} \sim N(0,1)$

(2) $\bar{X}$ 与 S^2 相互独立

(3) $\dfrac{n-1}{\sigma^2}S^2 \sim \chi^2(n-1)$ 即 $\dfrac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2 \sim \chi^2(n-1)$

推论1 $T = \dfrac{\bar{X}-\mu}{S}\sqrt{n} \sim t(n-1).$

推论2 $(X_1, X_2, \cdots, X_{n_1})$和$(Y_1, Y_2, \cdots, Y_{n_2})$分别取自两个相互独立的正态总体$N(\mu_1, \sigma_1^2)$，$N(\mu_2, \sigma_2^2)$的两个样本，其样本方差分别记为$S_1^2$，$S_2^2$，则：

$$(1) F = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$$

$$(2) \frac{\frac{1}{n_1}\sum_{i=1}^{n_1}\left(\frac{X_i - \mu_1}{\sigma_1}\right)^2}{\frac{1}{n_2}\sum_{i=1}^{n_2}\left(\frac{Y_i - \mu_2}{\sigma_2}\right)^2} \sim F(n_1, n_2)$$

推论3 条件同推论2，且 $\sigma_1 = \sigma_2 = \sigma$，则

$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_w\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2), \quad S_w^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}.$$

五 顺序统计量的分布

设$X_1,X_2,\ldots,X_n$独立同分布于X，记$F(x), f(x)$分别为X的分布函数和概率密度。

$$F_{X_i^*}(x)=P(X_i^*\le x)=P(X_1,\cdots,X_n\text{中至少有}i\text{个}\le x)$$

$$=P(X_1,\cdots,X_n\text{中恰有}i\text{个}\le x)+\cdots+P(X_1,\cdots,X_n\text{中恰有}n\text{个}\le x)$$

$$=\sum_{k=i}^{n}C_n^k(F(x))^k(1-F(x))^{n-k}$$

$$f_{X_i^*}(x)=(F_{X_i^*}(x))'$$

$$=\sum_{k=i}^{n}C_n^k[k(F(x))^{k-1}(1-F(x))^{n-k}-(n-k)(F(x))^k(1-F(x))^{n-k-1}]f(x)$$

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

小结

- 熟记χ^2分布，t分布，F分布的定义及正态总体的样本均值与样本方差的定理

- 能灵活应用上述三种分布的定义及正态总体的样本均值与样本方差的定理判断统计量的分布

一. 填空题

练习

1. 设$(X_1,\cdots,X_n,X_{n+1})$是取自总体$N(\mu,\sigma^2)$的容量为$n+1$的样本,$\bar{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i$,$\tilde{S}_n^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2$,则

$$T=\frac{X_{n+1}-\bar{X}_n}{\tilde{S}_n}\sqrt{\frac{n-1}{n+1}}\sim \underline{\quad t(n-1)\quad}$$

2. 设$X\sim t(m)$,则$Y=X^2\sim$ $\underline{F(1,m)}$

一. 填空题

练习

3. (X_1, X_2, X_3, X_4)是取自总体$N(0,\sigma^2)$的样本，则当$C=$ $\underline{\sqrt{3}}$ 时，$U = CX_1/\sqrt{X_2^2+X_3^2+X_4^2}$服从自由度为 $\underline{3}$ 的 $\underline{t}$ 分布。

4. $(X_1,\cdots,X_n)$是来自总体$N(0,\sigma^2)$的样本，记$\bar{X}$为样本均值，$Q^2=\sum_{i=1}^{n}(X_i-\bar{X})^2$，则统计量$T=\dfrac{n(n-1)\bar{X}^2}{Q^2}$服从分布 $\underline{F(1,n-1)}$.

二. 选择题

练

习

1. $(X_1,\cdots,X_n)$为总体$N(1,2^2)$的一个样本, $\bar{X}$为样本均值,下面正确的(D)

A. $\dfrac{\bar{X}-1}{2/\sqrt{n}} \sim t(n)$　　B. $\dfrac{1}{4}\sum_{i=1}^{n}(X_i-1)^2 \sim F(n,1)$

C. $\dfrac{\bar{X}-1}{\sqrt{2}/\sqrt{n}} \sim N(0,1)$　　D. $\dfrac{1}{4}\sum_{i=1}^{n}(X_i-1)^2 \sim \chi^2(n)$

2. $X\sim F(n,m)$, $P(X\geq\lambda)=1-\alpha(0<\alpha<1)$,则$\lambda$的值为(D)

A. $F_\alpha(n,m)$　B. $F_\alpha(m,n)$　C. $(F_\alpha(n,m))^{-1}$　D. $(F_\alpha(m,n))^{-1}$

二. 选择题

练

习

3. 总体$X \sim B(1,p)$, p未知, $X_1,\cdots,X_n$是X的样本,

$\overline{X}$为样本均值，则$P(\overline{X}=\dfrac{k}{n})=$(C)

A.p B.$1-p$ C.$C_n^k p^k(1-p)^{n-k}$ D.$C_n^k p^{n-k}(1-p)^k$

4. R.V.X,Y独立均服从$N(0,\sigma^2)$, (X_1,X_2,X_3),$(Y_1,Y_2,$

$Y_3,Y_4)$是X,Y的样本, 则$\sum_{i=1}^{3}X_i^2/\sum_{i=1}^{4}(Y_i-\overline{Y})^2\sim$(C)

A.$N(0,1)$ B.$t(3)$ C.$F(3,3)$ D.$F(3,4)$

二. 选择题

练

习

5. 下列不等式中正确的是(B)

A. $t_{0.025}(10) > t_{0.025}(5) > u_{0.025}$

B. $t_{0.025}(5) > t_{0.025}(10) > u_{0.025}$

C. $u_{0.025} > t_{0.025}(10) > t_{0.025}(5)$

D. $u_{0.025} > t_{0.025}(5) > t_{0.025}(10)$