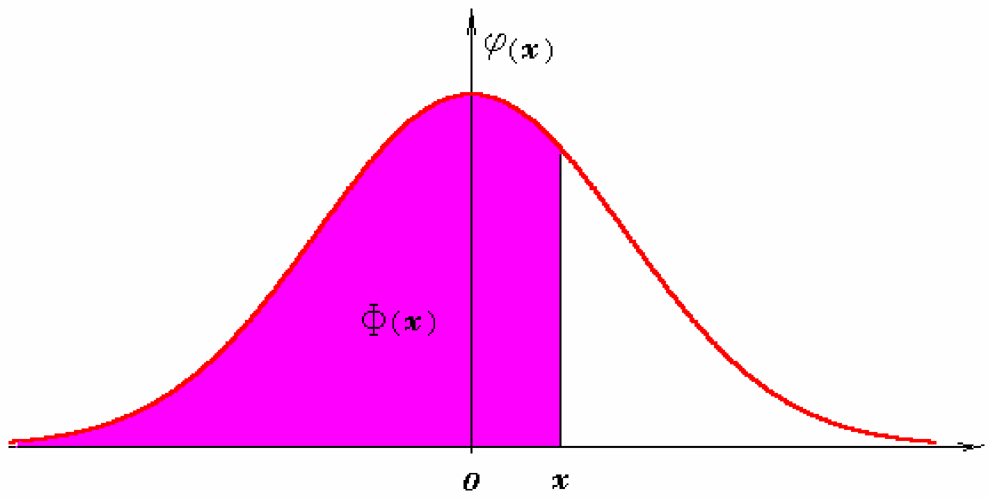
湖北省精品课程

概率论与数理统计



华中科技大学 概率统计系

叶 鹰 副教授

第六章　数理统计的基本概念

引例（习题4.9P$_{84}$截尾试验）

概率问题：　p已知，X为检查件数，则

$$P(X=k)=\begin{cases}(1-p)^{k-1}p & k=1,2,\cdots,n_0-1\\(1-p)^{k-1} & k=n_0\end{cases}$$

$$E(X)=\sum_{k=1}^{n_0-1}k(1-p)^{k-1}p+n_0(1-p)^{n_0-1}=\frac{1-(1-p)^{n_0}}{p}$$

统计问题：　p未知，确定适当的n_0，使

若$X<n_0$，则认为$p>p_0$（不合格）；

若$X=n_0$，则认为$p\leqslant p_0$（合格）。

§6.1 总体与样本

6.1.1 总体

研究对象的全体 $\xrightarrow{\text{量化}}$ 指标集 $\xrightarrow{\text{规律}}$ R.V. X 或$F(x)$

6.1.2 样本

总体的部分个体： $X_1, X_2, \cdots, X_n$ 独立同分布于$F(x)$

试验前： $X_1, X_2, \cdots, X_n$为R.V.

试验后： $x_1, x_2, \cdots, x_n$为样本观察值（实数）

n：样本容量（样本大小）

基本思想： 由样本对总体的分布（特征）进行合理地**推断**。

6.1.3 理论分布与经验分布函数

理论分布函数$F(x)$

对总体$F(x)$：样本的联合分布函数 $F(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} F(x_i)$

离散型总体：$P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i)$

如：$X_1, X_2, \cdots, X_n$ 为取自总体 $B(1,p)$ 的样本，则其联合分布律

$$P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$$

连续型总体：$f(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f(x_i)$

如：$X_1, X_2, \cdots, X_n$为取自总体 $N(0,1)$ 的样本，则其联合密度函数

$$f(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} = (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right)$$

经验分布函数$F_n(x)$

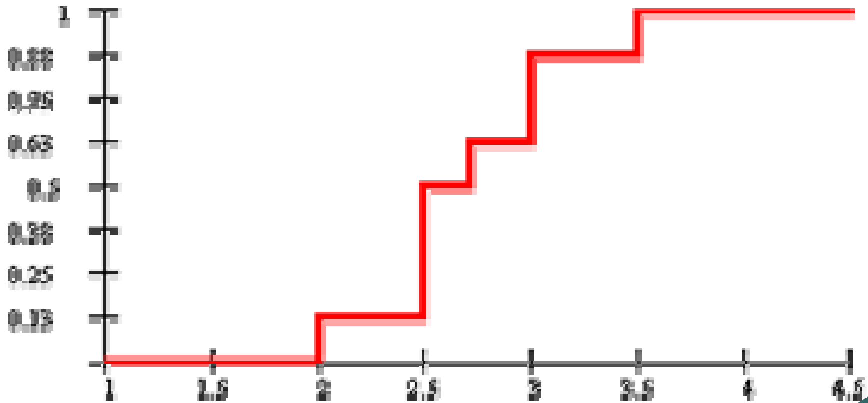
(1) $x_1, x_2, \cdots, x_n$ ➡ $x_1^* \leqslant x_2^* \leqslant \cdots \leqslant x_n^*$

(2) $$F_n(x)=\begin{cases} 0 & x < x_1^* \\ k/n & x_k^* \le x < x_{k+1}^* \quad k=1,2,\cdots,n-1 \\ 1 & x \ge x_n^* \end{cases}$$

例1 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 $\mathbf{2 < 2.5 = 2.5 = 2.5 < 2.7 < 3 = 3 < 3.5}$

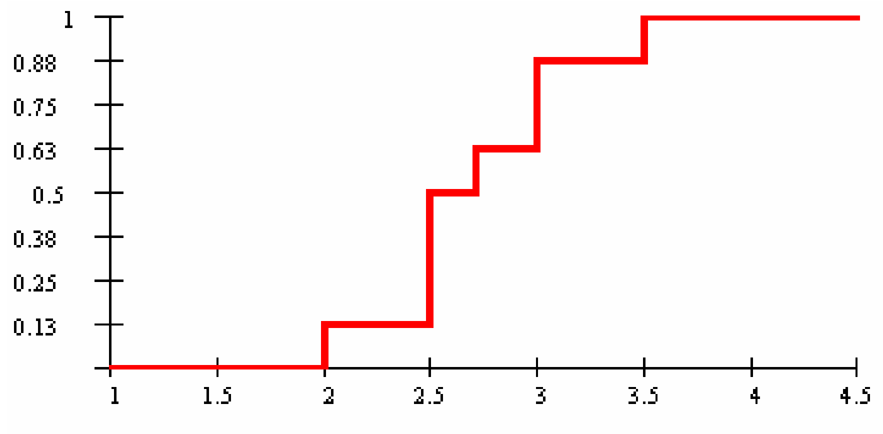
$$F_8(x)=\begin{cases} 0 & x<2 \\ 1/8 & 2\le x<2.5 \\ 4/8 & 2.5\le x<2.7 \\ 5/8 & 2.7\le x<3 \\ 7/8 & 3\le x<3.5 \\ 1 & x\ge 3.5 \end{cases}$$



例1 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：

$P(X=x_i)=1/n$，$i=1,2,\dots,n$

格列汶科定理

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

随机模拟演示

6.1.4 统计量

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

（1）$T = g(x_1, x_2, \cdots, x_n)$ 连续;

（2）$T = g(x_1, x_2, \cdots, x_n)$ 中不含有关总体的未知参数。

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

例如 设$X_1, X_2,\cdots, X_n$为来自总体$N(\mu, \sigma^2)$的样本，其中μ已知，σ^2未知，则

(1) $X_1 + X_n$ ✓

(2) $\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$ ✓

(3) $\sum_{i=1}^{n}\frac{|X_i|}{\sigma}$ ✗

(4) $\min_{1\leqslant i\leqslant n}\{X_i\}$ ✓

常用统计量

1、样本均值 $\overline{X}=\dfrac{1}{n}\sum_{i=1}^{n}X_i$ 提炼**E(*X*)**的信息

2、样本方差 $S^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 提炼**D(*X*)**的信息

样本标准差 $S=\sqrt{S^2}$

3、样本*k*阶原点矩 $A_k=\dfrac{1}{n}\sum_{i=1}^{n}X_i^k$ 提炼**E(*X*^{k})**的信息

$$A_1=\overline{X}$$

4、样本*k*阶中心矩 $B_k=\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k$ 提炼**E(*X*–E*X*)^k**的信息

$$B_2=\frac{n-1}{n}S^2=\tilde{S}^2$$

常用统计量

5、顺序统计量 $X_1, X_2, \cdots, X_n \Longrightarrow X_1^* \leqslant X_2^* \leqslant \cdots \leqslant X_n^*$

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

样本极差 $R = X_n^* - X_1^* \quad \to \quad D(X)$

注意：X_i^* 并不是 $X_1, X_2, \cdots, X_n$ 中的一个，如

$$X_1, X_2 \overset{iid}{\sim} B(1, p) \quad \Rightarrow \quad X_1^* \sim \mathrm{B}(1, \boldsymbol{p^2}), \ X_2^* \sim \mathrm{B}(1, \boldsymbol{1-(1-p)^2}).$$

§6.2 抽样分布

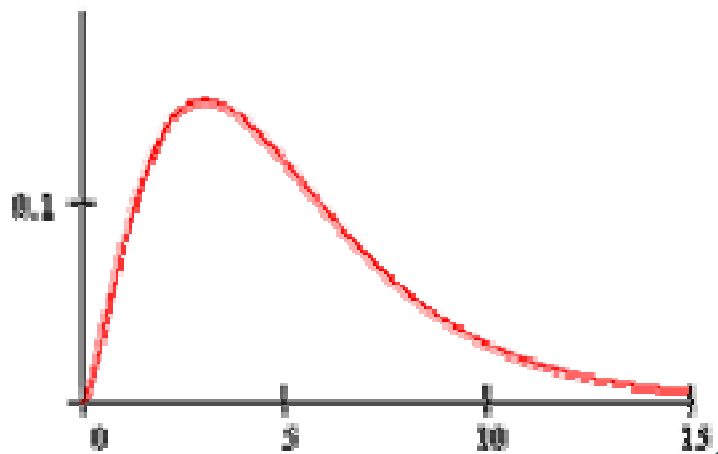
6.2.1 χ^2分布

设X_1，X_2，…，X_n独立同分布于N(0, 1)，则称

$$\chi^2 = X_1^2 + X_2^2 + \dots + X_n^2$$

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

$$f(x) = \begin{cases} \dfrac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}, & x \ge 0 \\ 0, & x < 0 \end{cases}$$



$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$$

• 数字特征:

$$E(\chi^2)=\sum_{i=1}^{n}E(X_i^2)=\sum_{i=1}^{n}D(X_i)=n$$

$$D(\chi^2)=\sum_{i=1}^{n}D(X_i^2)=\sum_{i=1}^{n}[E(X_i^4)-E^2(X_i^2)]=\sum_{i=1}^{n}(3-1)=2n$$

• 可加性:

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

• 上侧分位点:

$$P(\chi^2>\chi_\alpha^2(n))=\alpha \qquad \chi_{0.01}^2(20)=37.566$$

当$n>45$时，有 $\chi_\alpha^2(n)\approx\frac{1}{2}(u_\alpha+\sqrt{2n-1})^2$

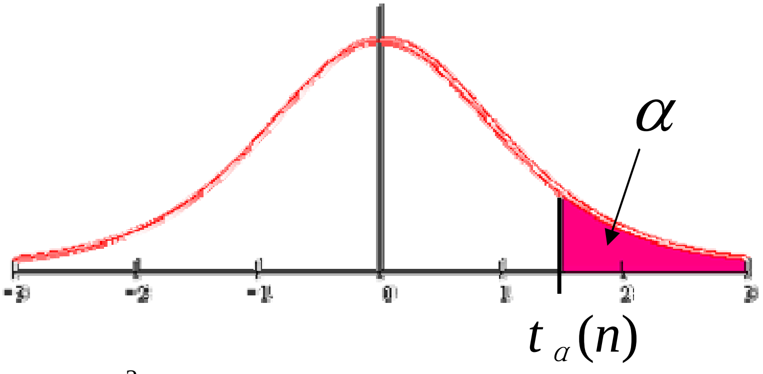
查表

6.2.2　t 分布

设 $X\sim N(0,1)$ 与 $Y\sim\chi^2(n)$ 独立，则称 $T=\dfrac{X}{\sqrt{Y/n}}$

服从自由度为 n 的 **t 分布**，记 $T\sim t(n)$。

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



- 数字特征　　$E(X)=0$
- 渐近正态　　$\lim\limits_{n\to\infty} f(x)=\varphi(x)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$
- 上侧分位点：　$P(T>t_\alpha(n))=\alpha$　　$t_{0.05}(8)=1.8595$

当$n>45$时，有 $t_\alpha(n)=u_\alpha$

查表

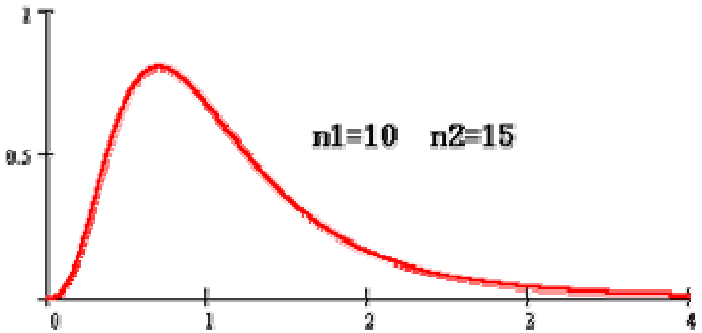
6.2.3 F 分布

设$X \sim \chi^2(n_1)$与$Y \sim \chi^2(n_2)$相互独立，则称

$$F = \frac{X/n_1}{Y/n_2} \qquad \frac{1}{F} \sim F(n_2, n_1)$$

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

$$f(x) = \begin{cases} \dfrac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})} \dfrac{(\frac{n_1}{n_2})(\frac{n_1}{n_2}x)^{\frac{n_1}{2}-1}}{(1+\frac{n_1}{n_2}x)^{\frac{n_1+n_2}{2}}}, & x \geq 0 \\ 0 \quad , & x < 0 \end{cases}$$



- 上侧分位点：$P(F > F_\alpha(n_1, n_2)) = \alpha$。

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

查表