§6.3 抽样分布

一、分位数

对于给定的正数α, $0<\alpha<1$，称满足条件

$$1-\alpha=F(b), \quad \text{即}\ P(X>b)=\alpha$$

的点b为分布函数$F(x)$的上α分位点（数）。

注意：如果上述的b不存在，则取满足条件

$F(b-0)\le 1-\alpha$和$F(b)\ge 1-\alpha$的b.

这就是所谓的“理论分位数”，与此相对应有“样本分位数”。

二、统计量

设$X_1, X_2, \cdots, X_n$ 是来自总体X的一个样本，$g(X_1, X_2, \cdots, X_n)$ 是$X_1, X_2, \cdots, X_n$ 的函数。若g 中不含未知参数，则称$g(X_1, X_2, \cdots, X_n)$ 为统计量。

统计量——样本的已知函数

设$x_1, x_2, \cdots, x_n$是相应于样本$X_1, X_2, \cdots, X_n$的样本值,则称$g(x_1, x_2, \cdots, x_n)$是$g(X_1, X_2, \cdots, X_n)$的观察值.

样本平均值 $$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

样本方差 $$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right)$$

样本标准差 $S=\sqrt{S^2}=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}$

样本k阶(原点)矩 $A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k, \quad k=1,2,\cdots$

样本k阶中心矩 $B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k, \quad k=2,3,\cdots$

若总体X的k阶矩 $\mu_k = E(X^k)$存在，则

$A_k \xrightarrow{P} \mu_k (\mathrm{n} \to \infty)$。即对任意的$\varepsilon > 0$有

$$\lim_{n \to +\infty} \mathrm{P}(|A_k - \mu_k| > \varepsilon) = 0$$

样本的"矩"是怎么来的呢？

三、经验分布函数

设$X_1, X_2, \cdots, X_n$是来自X的样本，经验分布函数为

$$F_n(x) = \frac{\text{小于等于}x\text{的}X_j\text{的个数}}{n} = f_n(X \le x), -\infty < x < \infty$$

设$x_1, x_2, \cdots, x_n$ 是总体F 的一个容量为n 的样本值。先将$x_1, x_2, \cdots, x_n$ 按自小到大的次序排列,并重新编号。设为 $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$,

则经验分布函数$F_n(x)$ 为

$$F_n(x) = \begin{cases} 0, & 若 x < x_{(1)} \\ \dfrac{k}{n}, & 若 x_{(k)} \le x < x_{(k+1)} \\ 1, & 若 x \ge x_{(n)} \end{cases}$$

经验分布函数是一个离散型随机变量的分布函数,对应的概率分布是:

X	$x_{(1)}$	$x_{(2)}$	…	$x_{(k)}$	…	$x_{(n)}$
p_k	1/n	1/n	…	1/n	…	1/n

（$x_1, x_2, \cdots, x_n$各不相同时）

所以，它的k阶矩就是k阶样本矩

用经验分布函数代替总体的分布函数，总体的k次方的期望就变成样本的k阶矩

用经验分布函数代替总体的分布函数，总体的方差就变成样本的2阶中心矩。不是样本方差。

说明:

（1）给定$x, F_n(x)$是样本的函数，是统计量。

（2）给定样本值，作为 x 的函数，它是一个分布函数.

四、几个主要的分布

(一) χ^2分布

设$X_1, X_2, \cdots, X_n$ 独立同分布于$N(0,1)$, 则称

$$\chi^2 := X_1^2 + X_2^2 + \cdots + X_n^2$$

的分布是自由度为 n 的χ^2分布。或说它服从自由度为 n 的χ^2分布，记为

$$\chi^2 \sim \chi^2(n).$$

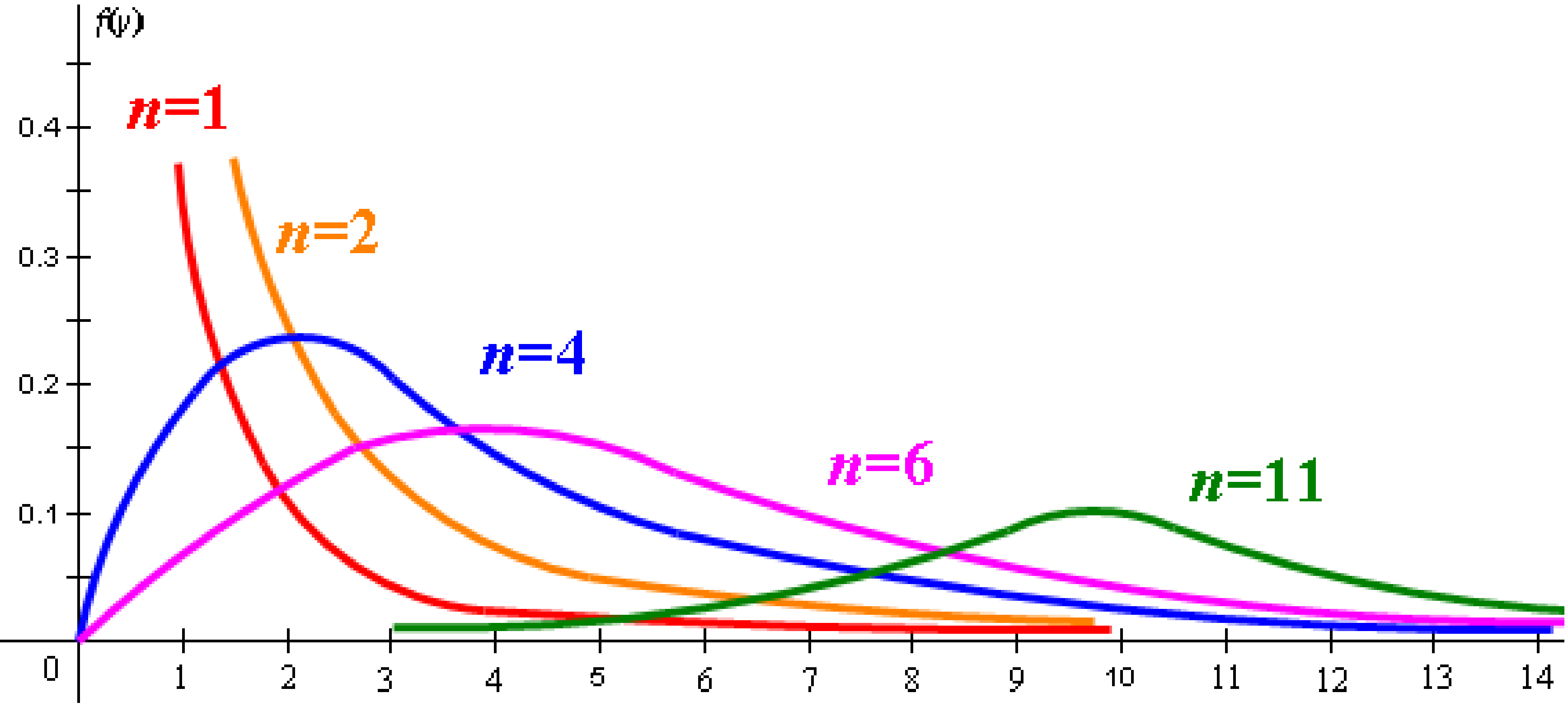
$\chi^2(n)$分布的概率密度为

$$f(y)=\begin{cases}\dfrac{1}{2^{n/2}\Gamma(n/2)}y^{n/2-1}e^{-y/2}, y>0 \\ 0, \qquad\qquad\qquad\qquad\qquad\qquad\quad \text{其它}\end{cases} \qquad (3.2)$$

注：$\Gamma(p)=\int_0^{\infty}x^{p-1}e^{-x}dx \quad (p>0)$。

$$\chi$$

卡方分布密度的图像——分布曲线



卡方分布的可加性

设$\chi_1^2 \sim \chi^2(n_1)$与$\chi_2^2 \sim \chi^2(n_2)$相互独立，则

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

卡方分布的期望与方差

若$\chi^2 \sim \chi^2(n)$,则有

$$E(\chi^2) = n, \ \ D(\chi^2) = 2n \quad (3.5)$$

设$X_1, X_2, \cdots, X_n$相互独立且同分布于$N(0,1)$,

按照卡方分布的定义，不妨设

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

由正态分布的定义可以计算得到DX_1^2 =2。

从而由数学期望与方差的性质得

$$E\chi^2 = nEX_1^2 = n,\ D\chi^2 = nDX_1^2$$

10月28日

卡方分布的上分位点

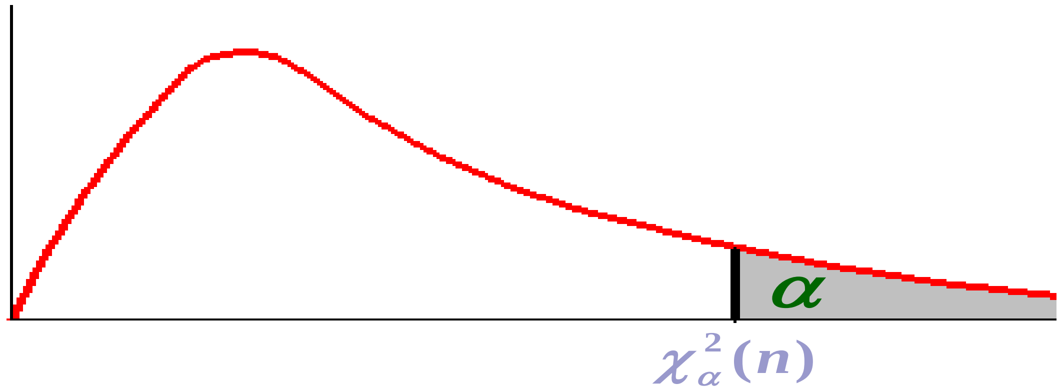
对于给定的正数$\alpha, 0<\alpha<1$,称满足条件

$$P\{\chi^2>b\}=\int_b^{\infty} f(y)dy=\alpha$$

的点b为$\chi^2(n)$分布的上α分位点, 记为$\chi^2_\alpha(n)$

$\chi^2_\alpha(n)$ 的数值可以查表

$\chi^2_{0.1}(25)=?$



34.382

10月28日

(二)　t 分布

设$X \sim N(0,1)$与$Y \sim \chi^2(n)$相互独立,则称

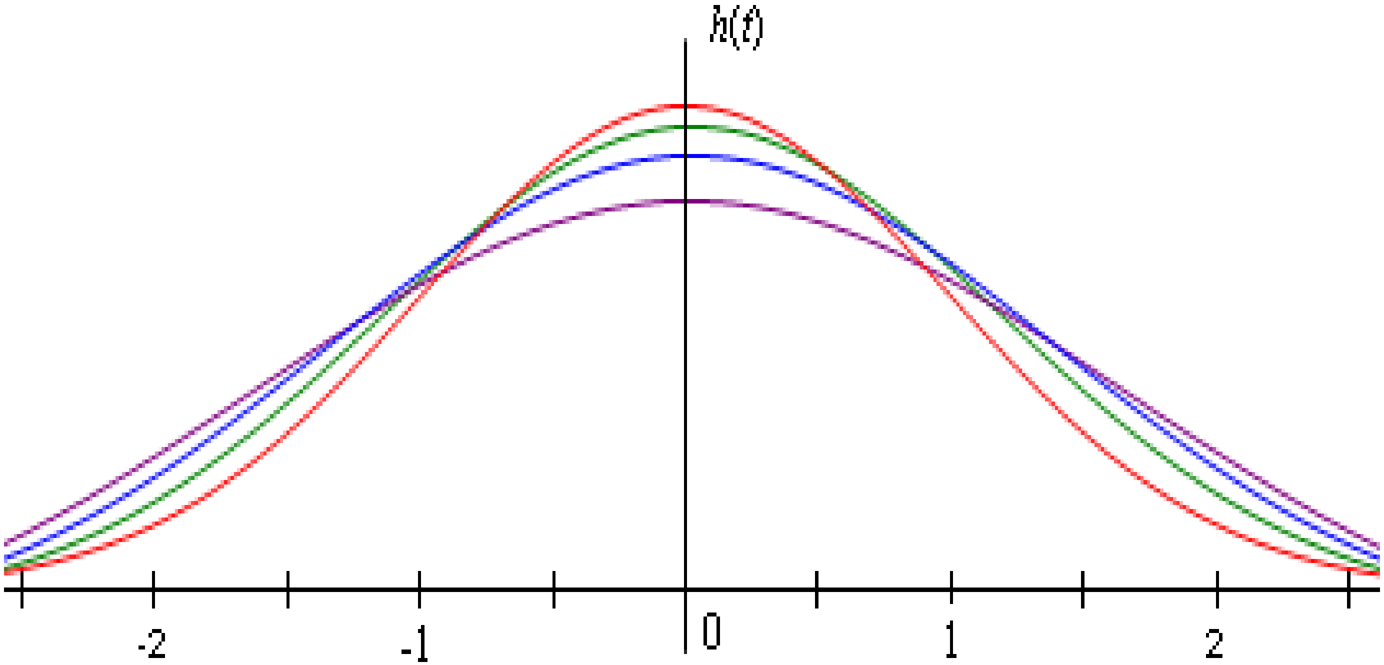
$$t = \frac{X}{\sqrt{Y/n}}$$

的分布是自由度为n的t分布。或说它服从自由度为n的t分布，记为 $t \sim t(n)$.

t分布的密度函数为

$$h(t) = \frac{\Gamma[(n+1)/2]}{\sqrt{\pi n}\,\Gamma(n/2)}\left(1+\frac{t^2}{n}\right)^{-(n+1)/2} \qquad -\infty < t < \infty$$

t分布密度的图像——分布曲线



$n=$ ∞ **25** **9** **2**

红色那一条就是标准正态分布曲线

t分布的性质

t分布的密度的图像关于y轴对称----偶函数

当n大于40时，t分布密度接近于标准正态分布的密度

所以，当n大于40时，t分布的分布函数值表可以用标准正态分布函数值表近似。

t分布的上分位点

对于给定的正数$\alpha, 0<\alpha<1$,称满足条件

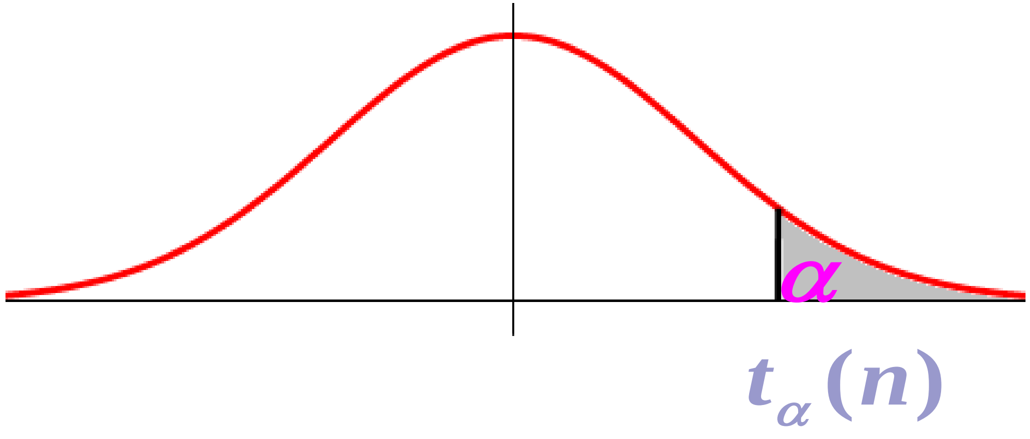
$$P\{t>b\}=\int_b^{\infty} h(y)dy=\alpha$$

的点b为t分布的上α分位点, 记为$t_\alpha(n)$

$t_\alpha(n)$ 的数值可以查表

$t_{0.1}(25)=?$

1.3163



样本均值与样本方差的期望和方差

设总体X的均值为μ和方差为σ^2存在，$X_1, X_2, \cdots, X_n$是来自X的一个样本，$\overline{X}$，S^2是样本均值和样本方差，则总有

$$E(\overline{X}) = \mu, D(\overline{X}) = \sigma^2 / n$$

$$E(S^2) = \sigma^2$$

$$D(\overline{X}) = D\left[\frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)\right] = \frac{1}{n^2}(\sum_{i=1}^{n} D(X_i))$$

$$= \frac{1}{n^2}\left[\sum_{i=1}^{n} \sigma^2\right] = \frac{1}{n}\sigma^2$$

$$E(S^2) = E\left[\frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right)\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n} E(X_i^2) - nE(\overline{X}^2)\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}(\sigma^2 + \mu^2) - n(\sigma^2/n + \mu^2)\right] = \sigma^2$$

五、正态总体的抽样分布

定理. 设$X_1, X_2, \cdots, X_n$是来自正态总体$N(\mu, \sigma^2)$的样本,$\overline{X}$是样本均值,则有

(1)$\overline{X} \sim N(\mu, \sigma^2/n)$;(定理2)

(2)$\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$　　(3) $\overline{X}$与S^2独立(定理3)

(4)$\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$(定理4)

作业：1，6，8，9

补充：设 $(2, 5, 3, 3, 3, 2)$ 是来自总体X 的容量为6 的样本值，求经验分布函数$F_6(x)$。