单样本方差
统计量的分布

设计与制作：刘琼荪

问题 设样本$X_1,\ldots,X_n$来自总体X

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \xrightarrow{\text{估计}} EX$$

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 \xrightarrow{\text{估计}} DX \quad ?$$

如何使得这种估计既精确又可靠？这就需要我们研究样本方差S^2的分布。

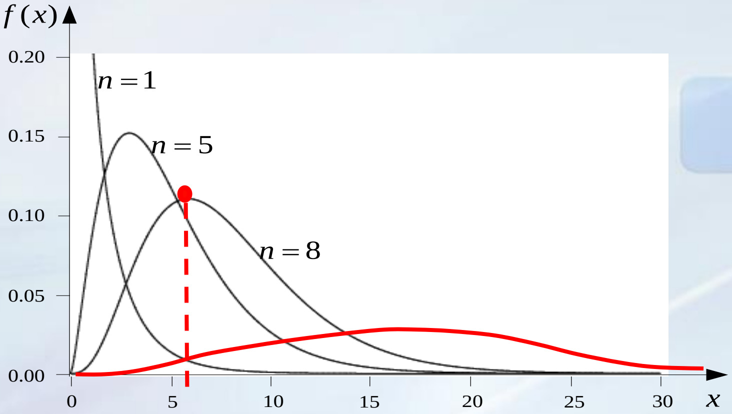
定义 如果$X_1, X_2, \ldots, X_n \sim N(0,1)$，则

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

参数为n的卡方分布密度为：

$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\dfrac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x > 0 \\ 0, & \text{其他} \end{cases}$$

密度曲线如图所示



性质

（1） 设$X \sim \chi^2(m)$， $Y \sim \chi^2(n)$， X与Y相互独立，则

$$X+Y \sim \chi^2(m+n)$$

（2） 设$X \sim \chi^2(n)$，则$EX = n$， $DX = 2n$

例1

如果$X_1,\ldots,X_{20}$来自均值为μ和方差为16的正态总体X，试分析下面两个统计量的分布及其参数，并确定它们的期望和方差。

（1） $Y=\dfrac{1}{32}\sum\limits_{i=1}^{10}(X_{2i-1}-X_{2i})^2$

（2） $\bar{X}_1=\dfrac{1}{10}\sum\limits_{i=1}^{10}X_i,\quad \bar{X}_2=\dfrac{1}{10}\sum\limits_{j=1}^{10}X_{10+j},\quad Z=\dfrac{10}{32}(\bar{X}_1-\bar{X}_2)^2$

分析

（1） $X_{2i-1}-X_{2i}\sim N(0,32),\quad i=1,2,\ldots,10$

标准化 $Y_i=\dfrac{X_{2i-1}-X_{2i}}{\sqrt{32}}\sim N(0,1),\quad i=1,2,\ldots,10$

并且$Y_1,Y_2,\ldots,Y_{10}$相互独立，

由卡方分布的定义和性质知，则

$$Y=\sum_{i=1}^{10}Y_i^2=\sum_{i=1}^{10}\left(\frac{X_{2i-1}-X_{2i}}{\sqrt{32}}\right)^2\sim\chi^2(10).\qquad EY=10,DY=20.$$

分析

（2） 因为$\bar{X}_1$与$\bar{X}_2$独立，且服从正态分布$N(\mu, \frac{16}{10})$

$$\bar{X}_1 - \bar{X}_2 \sim N(0, \frac{32}{10})$$

$$\frac{\sqrt{10}(\bar{X}_1 - \bar{X}_2)}{\sqrt{32}} \sim N(0,1)$$

由卡方分布的定义知，

$$Z = \frac{10}{32}(\bar{X}_1 - \bar{X}_2)^2 \sim \chi^2(1), \qquad EZ = 1, \quad DZ = 2$$

定理 如果样本 $X_1, X_2, \ldots, X_n$ 来自期望为μ和方差为σ的正态总体X, 则

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(n) \quad (1)$$

$$\Leftrightarrow \quad \chi^2 = \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{\sigma^2} \sim \chi^2(n) \quad (2)$$

注意： 理论上可以证明：

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

$$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n) \quad (2)$$

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

?

$$\boxed{\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)} \quad (4)$$

例2一个搬运公司对每袋货物的装载重量有规定，不允许有太大的出入，以免造成诸如搬运负担、计量不准确等情况。假设每袋货物的重量服从正态分布。从整批货物中随机抽查了15袋，测得重量（kg）数据如下：

11.5	12.3	11.0	12.0	12.2
12.1	10.9	12.3	12.5	10.6
11.7	11.2	12.4	11.0	10.0

试以90%的可靠度估计这15个样本数据的波动大小。

分析

假设货物的袋装重量为总体X，且 $X \sim N(\mu, \sigma^2)$

$$S^2 \xrightarrow{\text{估计}} \sigma^2$$

估计的可靠度为90%。

考虑设置两个待定常数c1，c2，满足

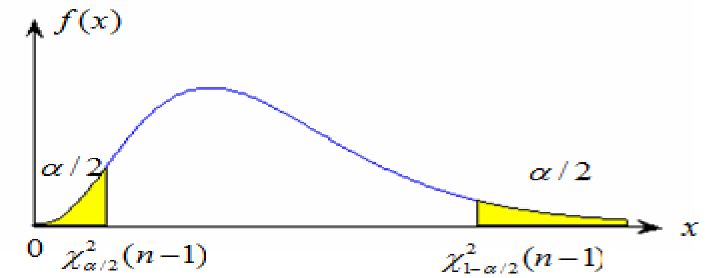
$$P\{c_1 < \frac{S^2}{\sigma^2} < c_2\} = 0.9 \qquad (1)$$

$$\Leftrightarrow \quad P\{\frac{S^2}{c_2} < \sigma^2 < \frac{S^2}{c_1}\} = 0.9 \qquad (2)$$

由于 $P\{c_1 < \frac{S^2}{\sigma^2} < c_2\} = 0.9$

$$\Rightarrow \quad P\{14c_1 < \frac{14S^2}{\sigma^2} < 14c_2\} = 0.9 \qquad \frac{14S^2}{\sigma^2} \sim \chi^2(14)$$

由分位数的定义知，如图所示，



$$14c_1 = \chi^2_{0.05}(14)$$
$$14c_2 = \chi^2_{0.95}(14)$$

经查卡方分布表，

$$\left.\begin{aligned} \chi^2_{0.05}(14) &= 6.5708 \\ \chi^2_{0.95}(14) &= 23.685 \end{aligned}\right\} \quad c_1 = 0.469, \quad c_2 = 1.692$$

$s^2 = 0.5817$

由此表达式 $P\{\frac{S^2}{c_2} < \sigma^2 < \frac{S^2}{c_1}\} = 0.9$ （2）

σ^2 的取值区间为 $[0.3438, 1.2403]$.

区间长度为 0.8965.

小结

（1）卡方分布的定义和特征

（2）样本方差的分布和应用