概率论与数理统计

χ^2 分布

主讲人：郑旭玲

信息科学与技术学院

统计三大抽样分布

统计量的分布叫做“**抽样分布**”。

统计中的“**三大分布**”

- χ^2分布
- t 分布
- F 分布

χ^2分布

定义

设$X_1, X_2, \ldots, X_n$相互独立，都服从正态分布$N(0,1)$，

则称随机变量

$$\chi^2 = X_1^{\ 2} + X_2^{\ 2} + \cdots + X_n^{\ 2} = \sum_{i=1}^{n} X_i^{\ 2}$$

服从的分布为**自由度为 n 的 χ^2 分布**。记为 $\chi^2 \sim \chi^2(n)$

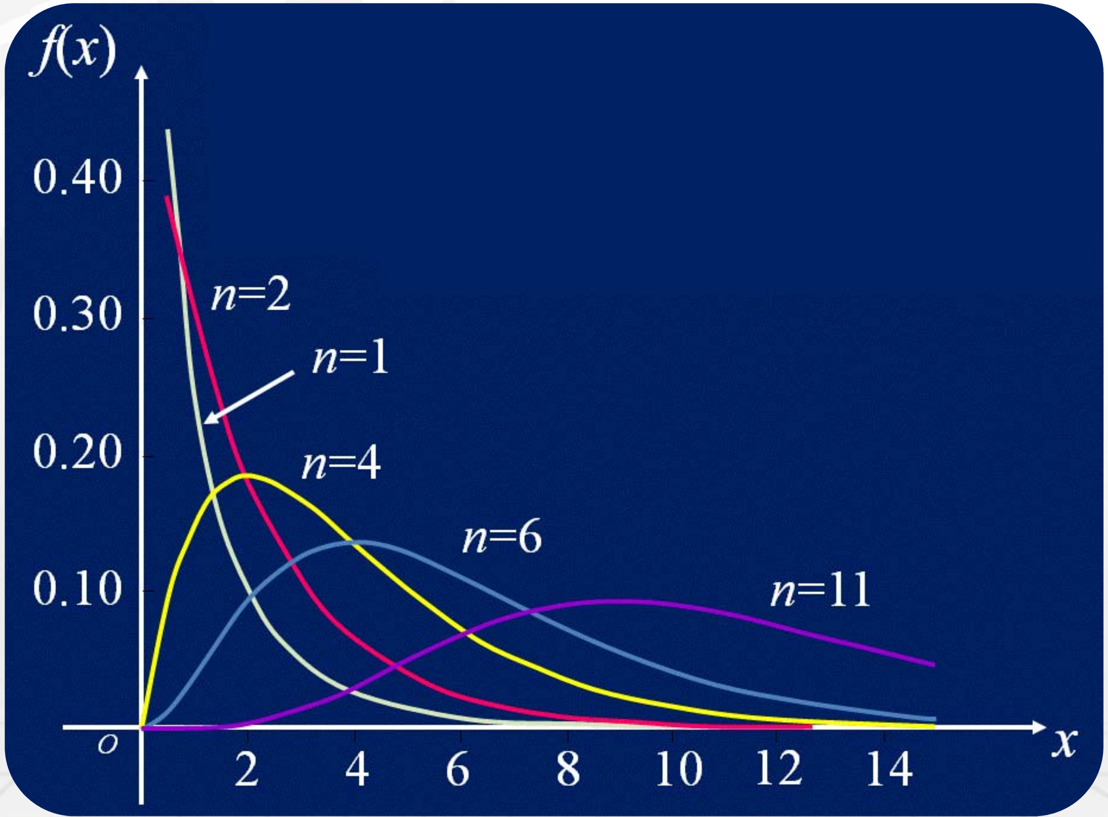
χ^2分布密度函数

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

其中，伽玛函数 $\Gamma(x)$的定义是

$$\Gamma(x)=\int_0^{\infty}e^{-t}t^{x-1}dt,\quad x>0$$

χ^2分布密度函数



χ^2分布的性质

1. 设 $X_1 \sim \chi^2(n_1)$, $X_2 \sim \chi^2(n_2)$ 且 X_1, X_2 相互独立，

则$X_1 + X_2 \sim \chi^2(n_1+n_2)$

χ^2 分布的可加性

$$X_1 = U_1^2 + U_2^2 + \cdots + U_{n_1}^2, \quad \forall U_i \sim N(0,1)$$

$$X_2 = V_1^2 + V_2^2 + \cdots + V_{n_2}^2, \quad \forall V_j \sim N(0,1)$$

$$X_1+X_2 = \left(U_1^2 + U_2^2 + \cdots + U_{n_1}^2\right) + \left(V_1^2 + V_2^2 + \cdots + V_{n_2}^2\right)$$

χ^2分布的性质

1. 设 $X_1 \sim \chi^2(n_1)$，$X_2 \sim \chi^2(n_2)$ 且 X_1, X_2 相互独立，

则$X_1 + X_2 \sim \chi^2(n_1+n_2)$

χ^2 分布的可加性

推广：设 $X_1, X_2, \ldots, X_n$相互独立，且 $X_i \sim \chi^2(n_i)$，则

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

χ^2分布的性质

2. 设 $X_1, X_2, \ldots, X_n$ 相互独立，都服从正态分布 $N \sim (\mu, \sigma^2)$，则

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

χ^2分布的性质

3. 若 $\chi^2 \sim \chi^2(n)$，χ^2 分布的数学期望与方差存在，

$E(\chi^2)=n$，$D(\chi^2)=2n$.

由$X_i \sim N(0,1)$，有$E(X_i^2)=D(X_i)=1$

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

又$D(X_i^2)=E(X_i^4)-[E(X_i^2)]^2=3-1=2$

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

χ^2分布的性质

4．若$X \sim \chi^2(n)$, 则当n充分大时，

$\dfrac{X-n}{\sqrt{2n}}$的分布近似服从标准正态分布$N(0,1)$。

由卡方分布的定义可知，$X=\chi^2=\sum_{i=1}^{n} X_i^2$, 其中$X_i \sim N(0,1)$

且$X_1, X_2, \cdots, X_n$相互独立 $\Rightarrow X_1^2, X_2^2, \cdots, X_n^2$相互独立

又$E(X_i^2)=1$, $D(X_i^2)=2$,

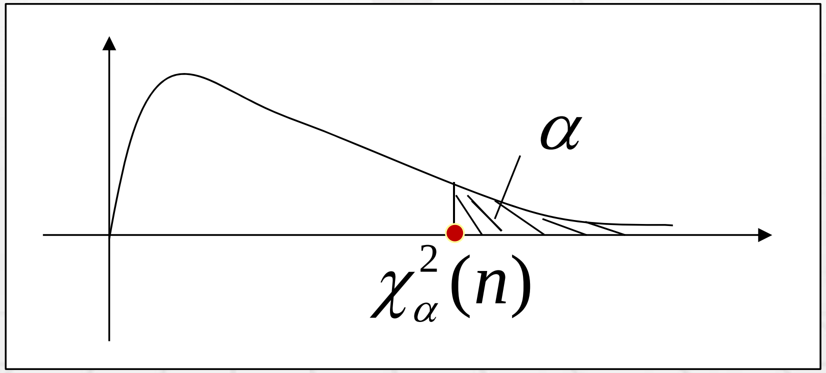
由中心极限定理可得：$\dfrac{\sum_{k=1}^{n} X_i^2 - n}{\sqrt{2n}} \overset{\text{近似地}}{\sim} N(0,1)$

χ^2分布的性质

5. 对于给定的正数$\alpha, 0<\alpha<1$,称满足条件

$$P\left\{\chi^2 > \chi_\alpha^2(n)\right\} = \int_{\chi_\alpha^2(n)}^{\infty} f(y)dy = \alpha$$

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



$\chi_\alpha^2(n)$可通过查表χ^2分布表来求得。

χ^2分布的性质

费希尔（R. A. Fisher）曾证明：

当n充分大时，近似地有

$$\chi_\alpha^2(n) \approx \frac{1}{2}(z_\alpha + \sqrt{2n-1})^2$$

其中，Z_α是标准正态分布的上α 分位点。

利用上式可求得当 $n>40$ 时，$\chi^2(n)$分布的上 α 分位点的近似值。

如：

$$\chi_{0.05}^2(50) \approx \frac{1}{2}(z_{0.05} + \sqrt{99})^2 = \frac{1}{2}(1.645 + \sqrt{99})^2 = 67.221$$

χ^2分布的例题

例 设总体$X \sim N\left(\mu, \sigma^2\right)$，$\mu, \sigma^2$ 已知。

$X_1, X_2, \cdots, X_5$是取自总体X的样本。

若$a(X_1 - X_2)^2 + b(2X_3 - X_4 - X_5)^2 \sim \chi^2(k)$，

问a, b, k各为多少？

解： 由正态分布的可加性，$X_1 - X_2 \sim N(0, 2\sigma^2)$，$\dfrac{(X_1 - X_2)^2}{2\sigma^2} \sim \chi^2(1)$

$2X_3 - X_4 - X_5 \sim N(0, 6\sigma^2)$，$\dfrac{(2X_3 - X_4 - X_5)^2}{6\sigma^2} \sim \chi^2(1)$

$X_1 - X_2$与$2X_3 - X_4 - X_5$相互独立，

故$\dfrac{(X_1 - X_2)^2}{2\sigma^2} + \dfrac{(2X_3 - X_4 - X_5)^2}{6\sigma^2} \sim \chi^2(2)$，有$a = \dfrac{1}{2\sigma^2}, b = \dfrac{1}{6\sigma^2}, k = 2.$

小结

- χ^2分布的定义
- χ^2分布的性质
- χ^2分布的上α分位点

谢 谢 大 家