第三节 抽样分布

统计量的分布称为抽样分布.

统计中有四个常见分布：

正态分布 $N(\mu,\sigma^2)$（在概率论中已经介绍），

χ^2分布，t 分布，F 分布（即将介绍）.

一、χ^2 分布

定义 1 设$(X_1, X_2, \cdots, X_n)$为来自总体$X \sim N(0,1)$的一个样本，就称统计量

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

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

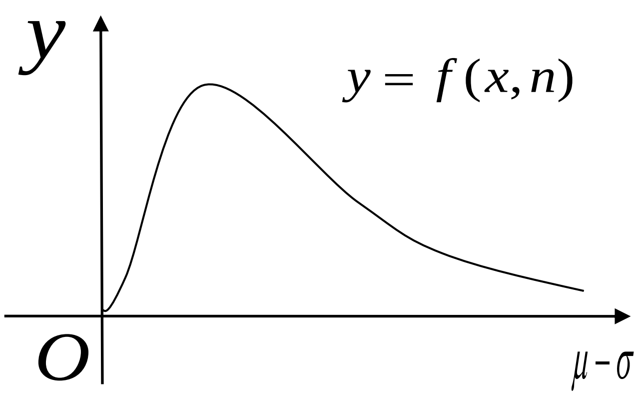
通俗地讲: 自由度为n的χ^2分布随机变量是由n个相互独立的标准正态随机变量的平方和组成.

定理 1 设 $\chi^2 \sim \chi^2(n)$，则 χ^2 的密度函数为

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

（淡化，不用记）

其中 $\Gamma(p)=\int_0^{+\infty}x^{p-1}e^{-x}dx\ (p>0)$ 称为 $\Gamma-$ 函数.



（记住图形）

性质 1 设 $\chi^2 \sim \chi^2(n)$，则 $E(\chi^2) = n$， $D(\chi^2) = 2n$．

证 $EX_i = EX = 0,\quad DX_i = DX = 1$， $E(X_i^2) = DX_i + (EX_i)^2 = 1$，

$$E(X_i^4) = E(X^4) = \int_{-\infty}^{+\infty} x^4 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = -\int_{-\infty}^{+\infty} x^3 d\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\right)$$

$$= -x^3 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \Bigg|_{-\infty}^{+\infty} + 3\int_{-\infty}^{+\infty} x^2 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 3E(X^2) = 3,$$

$$D(X_i^2) = E(X_i^4) - [E(X_i^2)]^2 = 3 - 1^2 = 2\ ,\quad i = 1, 2, \cdots, n,$$

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

（记住结论，了解证明过程）

$$D(\chi^2) = D\left(\sum_{i=1}^{n} X_i^2\right) \overset{\text{独立性}}{=} \sum_{i=1}^{n} D(X_i^2) = \sum_{i=1}^{n} 2 = 2n\ .$$

例 1 设(X_1,X_2,X_3)为来自总体$X\sim N(0,1)$的一个简单随机样本，求常数a和b，使得$aX_1^2+b(X_2-2X_3)^2\sim\chi^2(2)$.

解 由于$X_1\sim N(0,1)$, $X_1^2\sim\chi^2(1)$；又

$$X_2-2X_3\sim N(0,5),\qquad \frac{X_2-2X_3}{\sqrt{5}}\sim N(0,1),$$

得$\dfrac{(X_2-2X_3)^2}{5}\sim\chi^2(1)$，并且$X_1^2$与$\dfrac{\left(X_2-2X_3\right)^2}{5}$独立，

故$X_1^2+\dfrac{\left(X_2-2X_3\right)^2}{5}\sim\chi^2\left(2\right)$，所以$a=1$，$b=\dfrac{1}{5}$.

性质 2 设 $X \sim N(0,1)$，则 $X^2 \sim \chi^2(1)$.

证 在 χ^2 分布的定义中令 $n=1$，即可证明性质 2.

性质 3 设 $\chi_i^2 \sim \chi^2(n_i)$，$i=1,2$，且 χ_1^2, χ_2^2 相互独立，则

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

推论 1 设 $\chi_i^2 \sim \chi^2(n_i)$，$i=1,2,\cdots,k$，且 $\chi_1^2, \chi_2^2, \cdots, \chi_k^2$ 相互独立，则

$$\sum_{i=1}^{k} \chi_i^2 \sim \chi^2\left(\sum_{i=1}^{k} n_i\right).$$

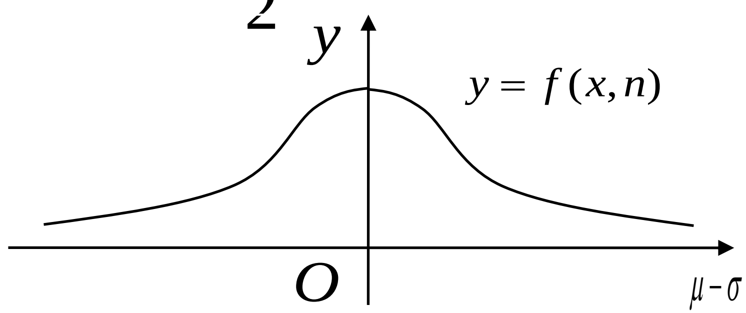
二、t 分布

定义 2 设随机变量 $X \sim N(0,1)$，$Y \sim \chi^2(n)$，且 X 与 Y 相互独立，就称 $T=\dfrac{X}{\sqrt{Y/n}}$ 为服从自由度为 n 的 t 分布，记作 $T \sim t(n)$.

定理 2 设 $T \sim t(n)$，则 T 的密度函数为

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

（淡化，不用记）



（记住图形）

性质 4 设 $T \sim t(n)$，则 $ET = 0\ (n > 1),\ DT = \dfrac{n}{n-2}\ (n > 2)$.

性质 5 设 $T \sim t(n)$，则当 n 充分大时，$T \overset{\text{近似}}{\sim} N(0,1)$.

（以上两个性质只作了解, 不必记）

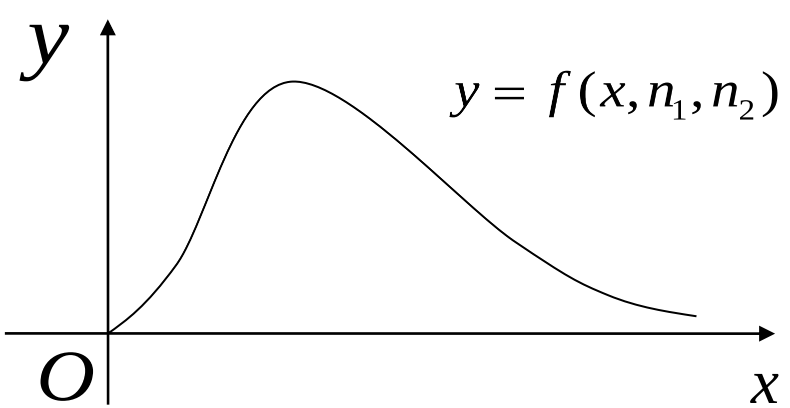
三、F 分布

定义 3 设随机变量 $X \sim \chi^2(n_1)$，$Y \sim \chi^2(n_2)$，且 X 与 Y 相互独立，就称 $F = \dfrac{X/n_1}{Y/n_2}$ 为服从第一自由度为 n_1，第二自由度为 n_2 的 F 分布，记作 $F \sim F(n_1, n_2)$.

定理 3 设 $F \sim F(n_1, n_2)$，则 F 的密度函数为

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

（淡化，不用记）



（记住图形）

性质 6 如果$F \sim F(n_1, n_2)$，则$\dfrac{1}{F} \sim F(n_2, n_1)$.

性质 7 如果$T \sim t(n)$，则$T^2 \sim F(1, n)$.

例 2 设$T \sim t(n)$，则$Y = \dfrac{1}{T^2} \sim$（ **C** ）.

(A) $\chi^2(n)$ **(B)** $\chi^2(n-1)$ **(C)** $F(n,1)$ **(D)** $F(1,n)$

例 3 设(X_1, X_2, X_3, X_4)为来自总体$X \sim N(0, \sigma^2)$的一个样本，求$\dfrac{X_1 - X_2}{|X_3 + X_4|}$所服从的分布.

解 由正态分布的性质，$X_1 - X_2 \sim N(0, 2\sigma^2)$，$\dfrac{X_1 - X_2}{\sqrt{2}\sigma} \sim N(0,1)$，

$X_3 + X_4 \sim N(0, 2\sigma^2)$，$\dfrac{X_3 + X_4}{\sqrt{2}\sigma} \sim N(0,1)$，$\dfrac{(X_3 + X_4)^2}{2\sigma^2} \sim \chi^2(1)$，

且$\dfrac{X_1 - X_2}{\sqrt{2}\sigma}$与$\dfrac{(X_3 + X_4)^2}{2\sigma^2}$独立，故

$$\frac{\dfrac{X_1 - X_2}{\sqrt{2}\sigma}}{\sqrt{\dfrac{(X_3 + X_4)^2}{2\sigma^2}\Big/1}} = \frac{X_1 - X_2}{|X_3 + X_4|} \sim t(1).$$

上侧分位点

定义 4 设 X 为随机变量， $0<\alpha<1$，如果点 x_α 满足

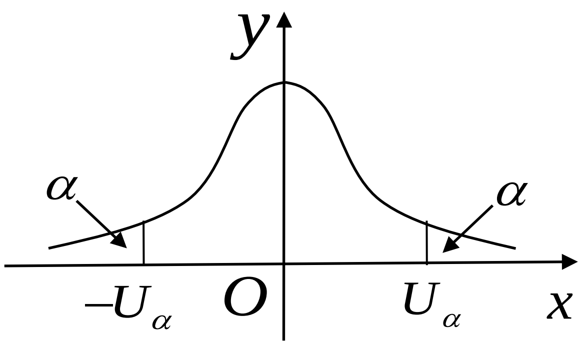
$$P\{X \geq x_\alpha\} = \alpha,$$

就称 x_α 为 X 的上 α 分位点.

1. 标准正态分布 $N(0,1)$ 的上侧分位点

设随机变量 $U \sim N(0,1)$，称满足

$$P\{U \geq U_\alpha\} = \alpha$$



的 U_α 为 $N(0,1)$ 分布的上侧 α 分位点.

注 1： U_α 可以通过标准正态分布表查得.

例如，$U_{0.05} = 1.645$；　$U_{0.025} = 1.96$.

例 4 设随机变量 $X \sim N(0,1)$，对给定的 $\alpha\ (0<\alpha<1)$，数 U_α 满足 $P\{X>U_\alpha\}=\alpha$．若 $P\{|X|<x\}=\alpha$，则 x 等于（　　）.

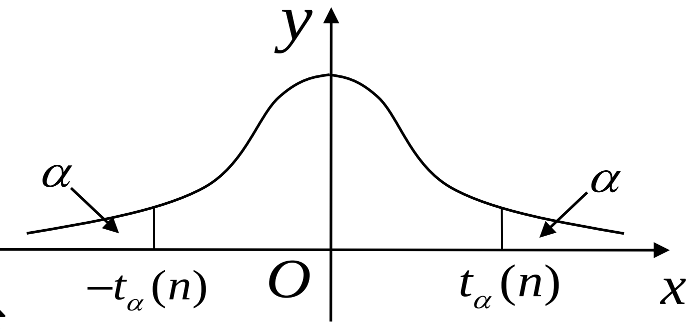
（A）$U_{\frac{\alpha}{2}}$　　（B）$U_{1-\frac{\alpha}{2}}$　　（C）$U_{\frac{1-\alpha}{2}}$　　（D）$U_{1-\alpha}$

解　由 $P\{|X|<x\}=\alpha$ 可得 $P\{X\geq x\}=\dfrac{1-\alpha}{2}$，所以 $x=U_{\frac{1-\alpha}{2}}$．选（C）.

2. t 分布 $t(n)$ 的上侧分位点

设随机变量 $T \sim t(n)$，称满足

$$P\{T \geq t_{\alpha}(n)\} = \alpha$$



的 $t_{\alpha}(n)$ 为 $t(n)$ 分布的上侧 α 分位点.

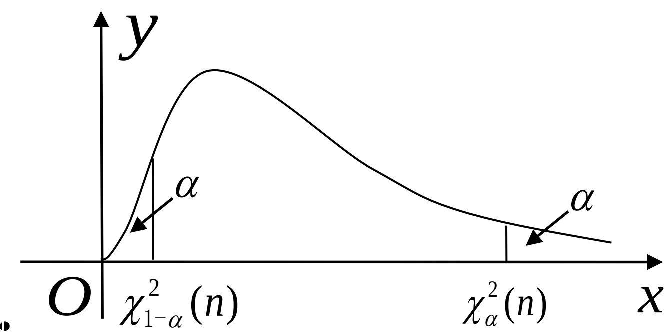
注 2： 利用 t 分布的对称性，$t_{1-\alpha}(n) = -t_{\alpha}(n)$.

3. χ^2 分布 $\chi^2(n)$ 的上侧分位点

设随机变量 $\chi^2 \sim \chi^2(n)$，称满足

$$P\{\chi^2 \geq \chi^2_\alpha(n)\} = \alpha$$

的 $\chi^2_\alpha(n)$ 为 $\chi^2(n)$ 分布的上侧 α 分位点.

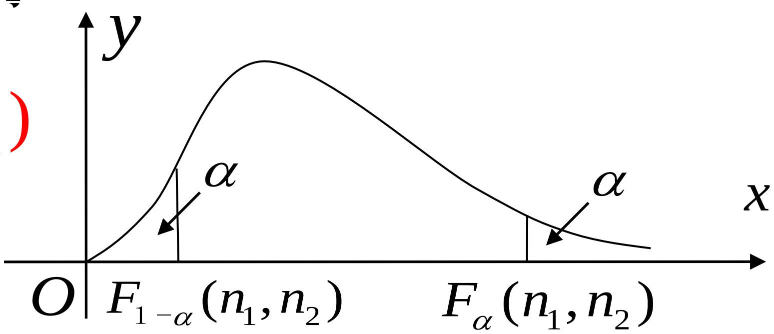


注 3： 由于 χ^2 分布为非对称的分布，因此，

$$P\{\chi^2 \geq \chi^2_{1-\alpha}(n)\} = 1-\alpha，即 P\{\chi^2 \leq \chi^2_{1-\alpha}(n)\} = \alpha .$$

4. F 分布 $F(n_1, n_2)$ 的上侧分位点

设随机变量 $F \sim F(n_1, n_2)$，称满足 $P\{F \geq F_\alpha(n_1, n_2)\} = \alpha$ 的 $F_\alpha(n_1, n_2)$ 为 $F(n_1, n_2)$ 分布的上 α 分位点.



注 4 $F_{1-\alpha}(n_1, n_2) = \dfrac{1}{F_\alpha(n_2, n_1)}$.

例如， $F_{0.99}(4,10) = \dfrac{1}{F_{0.01}(10,4)} = \dfrac{1}{14.55}$.

四、正态总体下几个重要统计量的分布

（本节为第七章和第八章的基础）

内容:

单正态总体样本均值和样本方差的分布（重点讲授）

双正态总体样本均值和样本方差的分布（简单介绍）

1. 单正态总体下样本均值和样本方差的分布

定理 4 设$(X_1,X_2,\cdots,X_n)$为来自总体$X\sim N(\mu,\sigma^2)$的一个样本，则

(1) $\bar{X}\sim N(\mu,\frac{\sigma^2}{n})$，或 $U=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$；

(2) $T=\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$；

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

(4) $\chi^2=\frac{(n-1)S^2}{\sigma^2}=\frac{\sum_{i=1}^{n}(X_i-\boxed{\bar{X}})^2}{\sigma^2}\sim \boxed{\chi^2(n-1)}$，

且 $\bar{X}$ 与 S^2 相互独立.

差别

不同

不同

差别

例 5 设 $X_1, X_2, \cdots, X_n$ 为来自总体 $X \sim N(\mu, \sigma^2)$ 的一个样本，求 $P\left\{\left(\overline{X}-\mu\right)^2 \le \dfrac{\sigma^2}{n}\right\}$.

解 因 $\overline{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$，则 $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$，故

$$P\left\{(\bar{X}-\mu)^2 \le \frac{\sigma^2}{n}\right\} = P\left\{\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2 \le 1\right\} = P\left\{\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| \le 1\right\}$$

$$= P\left\{-1 \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le 1\right\} = \Phi(1)-\Phi(-1) = 2\Phi(1)-1$$

$$= 2\times 0.8413-1 = 0.6826.$$

例 6 设$(X_1, X_2, \cdots, X_9)$为来自总体$X \sim N(\mu, \sigma^2)$的一个样本，求$P\{-0.4656 < \dfrac{\overline{X}-\mu}{S} < 0.9655\}$.

解 因$\dfrac{\overline{X}-\mu}{S/\sqrt{9}} \sim t(8)$，所以

$$P\{-0.4656 < \frac{\overline{X}-\mu}{S} < 0.9655\} = P\{-1.3968 < \frac{\overline{X}-\mu}{S/\sqrt{9}} < 2.8965\}$$

$$= P\{-t_{0.10}(8) < \frac{\overline{X}-\mu}{S/\sqrt{9}} < t_{0.01}(8)\} = P\{t_{0.90}(8) < \frac{\overline{X}-\mu}{S/\sqrt{9}} < t_{0.01}(8)\}$$

$$= 0.90 - 0.01 = 0.89.$$

例 7 设 $(X_1, X_2, \cdots, X_n)$ 为来自总体 $X \sim N(\mu, \sigma^2)$ 的一个样本，其中 $n>1$．令 $S_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$，分别计算

$E(S_0^2)$，$D(S_0^2)$，$E(S^2)$ 和 $D(S^2)$．

解 由定理 4 知，

$$\frac{nS_0^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi^2(n), \quad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1),$$

所以 $E(\frac{nS_0^2}{\sigma^2}) = n,\quad D(\frac{nS_0^2}{\sigma^2}) = 2n$，

$$E[\frac{(n-1)S^2}{\sigma^2}] = n-1,\quad D[\frac{(n-1)S^2}{\sigma^2}] = 2(n-1),$$

故得 $E(S_0^2) = \sigma^2, D(S_0^2) = \frac{2\sigma^4}{n}$，$E(S^2) = \sigma^2, D(S^2) = \frac{2\sigma^4}{n-1}$．

2. 双正态总体下样本均值差和样本方差比的分布

定理 5 设 $(X_1, X_2, \cdots, X_{n_1})$ 为来自总体 $X \sim N(\mu_1, \sigma_1^2)$ 的样本，

样本均值为 $\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$，样本方差为 $S_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2$.

$(Y_1, Y_2, \cdots, Y_{n_2})$ 为来自总体 $Y \sim N(\mu_2, \sigma_2^2)$ 的一个样本，样本均

值为 $\bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$，样本方差为 $S_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2$，且

$X_1, X_2, \cdots, X_{n_1}$ 与 $Y_1, Y_2, \cdots, Y_{n_2}$ 相互独立. 则

(1) $\bar{X}-\bar{Y}\sim N(\mu_1-\mu_2,\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2})$；

(2) 当σ_1^2,σ_2^2未知，但$\sigma_1^2=\sigma_2^2=\sigma^2$时，

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

其中$S_\omega=\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$；

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

例 8 从总体 $X \sim N(1,3)$ 中分别抽取容量为 $20, 30$ 的两个独立样本，求其样本均值差的绝对值小于1的概率.

解 设两个样本均值分别为 $\overline{X}$ 和 $\overline{Y}$，由定理 5，可得

$\overline{X} - \overline{Y} \sim N(0, \frac{1}{4})$，所以

$$P\{\left|\overline{X} - \overline{Y}\right| < 1\} = P\{\left|\frac{\overline{X} - \overline{Y}}{1/2}\right| < 2\}$$

$$= 2\Phi(2) - 1 = 2 \times 0,9772 - 1 = 0.9544 .$$

注 5: $D(\overline{X} - \overline{Y}) = D(\overline{X}) + D(\overline{Y}) = \frac{3}{20} + \frac{3}{30} = \frac{1}{4}$.