06

统计量和抽样分布

《概率论与数理统计》数学科学学院统计系

目录/Contents

目录/Contents

6.1 总体与样本

一、总体

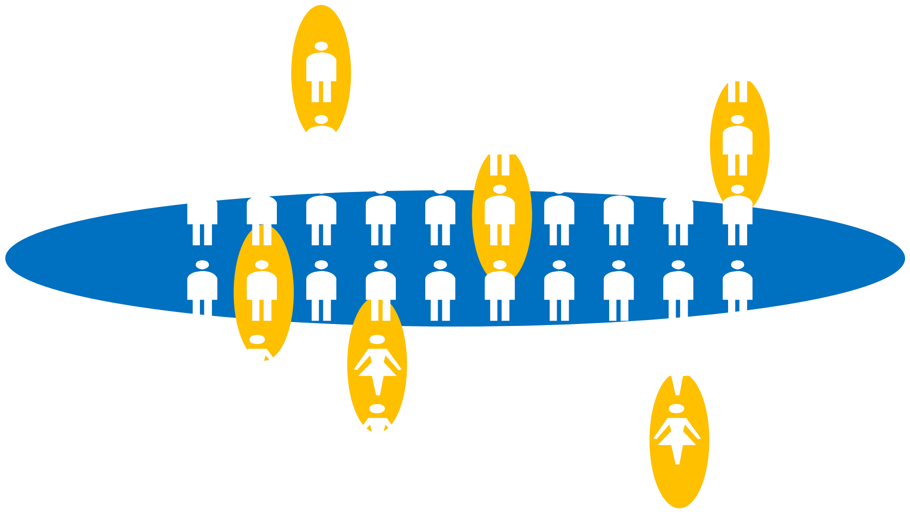
二、样本

一、总体

研究对象的全体称为**总体**，组成总体的每个成员称为**个体**.

特指

研究对象的某项数量**指标**的全体称为总体，组成总体的每个成员的该项数量**指标**称为个体.



一、总体

总体指标未知，可看作是一个随机变量，记为X .

(1)当总体X是离散型随机变量时，定义总体分布为

$$f(x;\theta) \triangleq P(X = x;\theta)$$ **，即为总体X的分布律.**

(2)当总体 X 是连续型随机变量时，定义总体分布为

$$f(x;\theta) \triangleq f_X(x;\theta)$$ **，即为总体X 的概率密度函数.**

例1 设总体$X\sim B(1, p)$，试写出总体分布律$f(x;p)$.

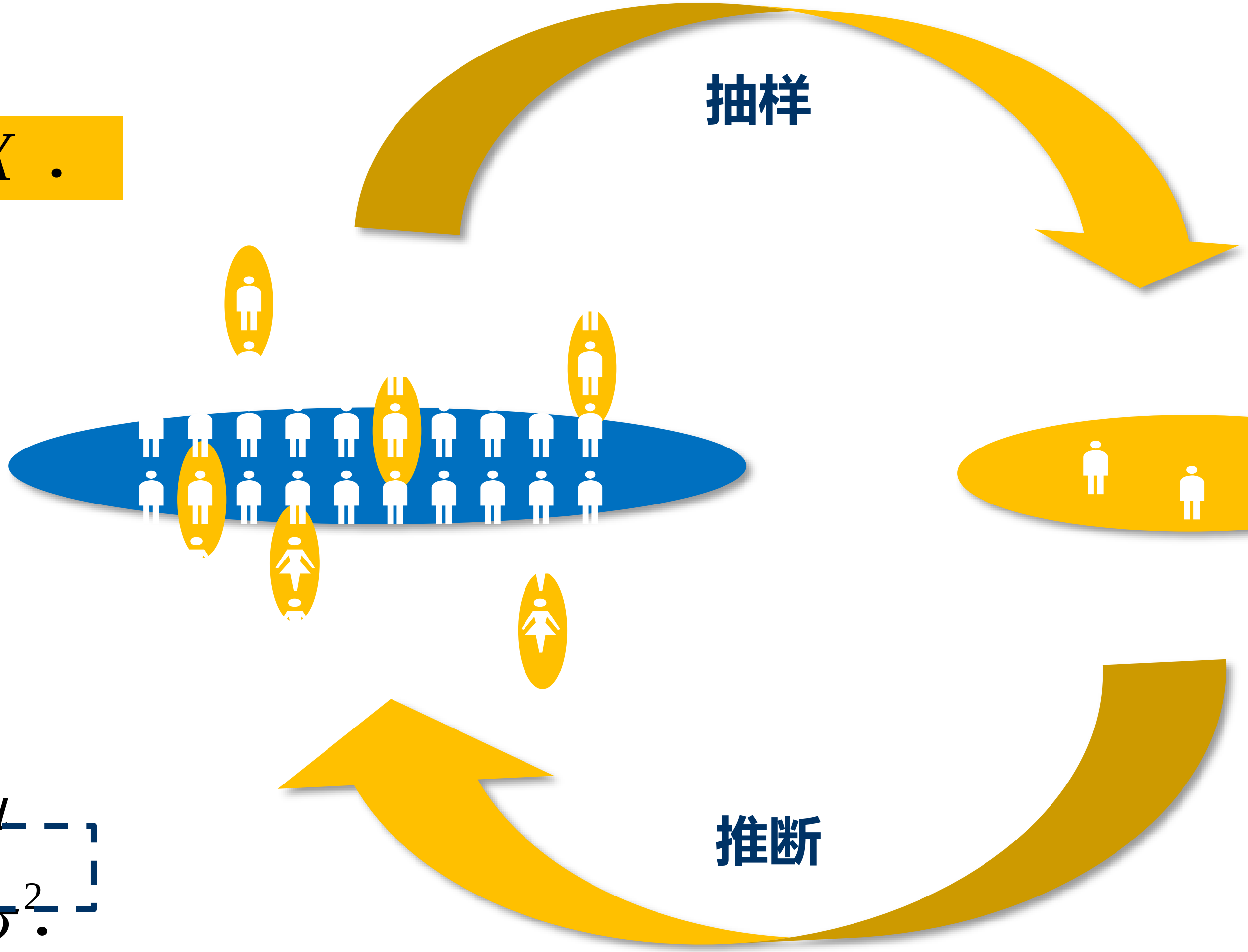
解 $f(x;p)=P(X=x)=(1-p)^{1-x}p^x, x=0,1,2,\cdots,n$

例2 设总体$X\sim N(\mu,\sigma^2)$，试写出总体分布$f(x;\mu,\sigma^2)$.

解 $f(x;\mu,\sigma^2)\triangleq f_X(x)=\dfrac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$， $-\infty<x<+\infty$

二、样本

总体，记为X.



$(X_1, X_2, \cdots, X_n)$

总体

样本

总体均值μ

总体方差σ^2.

二、样本

从总体中抽取样本的方法有很多， 我们主要采用简单随机抽样的方法，即有放回地重复独立抽取，这样得到的样本称为**简单随机样本**（简称样本）．记作 $(X_1,X_2,\cdots,X_n)$.

在试验前，样本的观测值是不确定的，为了体现随机性,在数理统计中样本记作 $(X_1,X_2,\cdots,X_n)$，事实上是一个n 维随机向量.

通过实验或观测得到的数值称为**样本观测值**，记作 $(x_1,x_2,\cdots,x_n)$，其中n称为**样本容量**(样本大小).

也就是说样本是一组随机变量，而样本观测值是抽样完成以后所得到的这组随机变量的一次具体取值.

二、样本

简单随机样本具有两个特点:

（1）**独立性**: $X_1, X_2, \cdots, X_n$**是相互独立的;**

（2）**代表性**: **每个个体** X_i **的分布都和总体分布相同.**

即$X_i \sim f(x_i, \theta), i = 1,2,\cdots,n.$

二、样本

(1) 设 X 为离散型随机变量，则$X \sim f(x;\theta) \triangleq P(X=x)$

而样本 $(X_1, X_2, \cdots, X_n)$ 的联合分布律为:

$$f(x_1, x_2, \cdots, x_n; \theta) \triangleq P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n; \theta)$$

$$= P(X_1 = x_1) P(X_2 = x_2) \cdots P(X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i; \theta)$$

(2) 设 X 为连续型随机变量，概率密度函数为 $f(x;\theta)$ ，

则样本 $(X_1,X_2,\cdots,X_n)$ 的联合概率密度函数为:

$$f(x_1,x_2,\cdots,x_n;\theta)\triangleq f_{(X_1,X_2\cdots,X_n)}(x_1,x_2,\cdots,x_n)$$

$$=f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_n}(x_n)$$

$$=f(x_1;\theta)\ f(x_2;\theta)\cdots f(x_n;\theta)=\prod_{i=1}^{n}f(x_i;\theta).$$

例3 设总体 $X \sim B(1, p)$, $(X_1, X_2, \cdots, X_n)$ 是取自该总体的一个样本,

试写出样本 $(X_1, X_2, \cdots, X_n)$ 的联合分布律.

解

$$f(x_1, x_2, \cdots, x_n; p) = \prod_{i=1}^{n} f(x_i; p)$$

$$= \prod_{i=1}^{n} P(X_i = x_i) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

$$= p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}, \quad x_i = 0,1, \quad i = 1, 2, \cdots, n.$$

例4

设总体 $X \sim P(\lambda)$,$(X_1, X_2, \cdots, X_6)$为取自该总体的一个样本,

求样本$(X_1, X_2, \cdots, X_6)$的联合分布律 $f(x_1, x_2, \cdots, x_6; \lambda)$?

解

$$f\left(x_1, x_2, \cdots, x_6; \lambda\right) = e^{-\lambda}\frac{\lambda^{x_1}}{x_1!}\cdot e^{-\lambda}\frac{\lambda^{x_2}}{x_2!}\cdot\cdots\cdot e^{-\lambda}\frac{\lambda^{x_6}}{x_6!} = e^{-6\lambda}\frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{6} x_i!}, \quad x_1, x_2, \cdots, x_6 = 0,1,2,\cdots$$

例5

设总体$X\sim U(0,\theta)$,$(X_1,X_2,\cdots,X_n)$是取自上均匀分布总体X的一个样本，$\theta>0$未知，求样本$(X_1,X_2,\cdots,X_6)$的联合密度函数?

解

$$f(x_1,x_2,\cdots,x_n;\theta)=f_{X_1}(x_1;\theta)\ f_{X_2}(x_2;\theta)\ldots\ldots\ f_{X_n}(x_n;\theta)$$

$$=\begin{cases}\theta^{-n} & 0<x_1,x_2,\cdots,x_n<\theta\\ 0 & \text{其它}\end{cases}$$

例6

设总体 $X \sim N(\mu,\sigma^2)$,$(X_1,X_2,\cdots,X_n)$为取自该总体的一个样本，求样本$(X_1,X_2,\cdots,X_n)$的联合密度函数?

解

$$f(x_1,x_2,\cdots,x_n;\ \mu,\sigma^2)=\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}}e^{-\frac{\sum(x_i-\mu)^2}{2\sigma^2}},-\infty<x_i<+\infty$$

例7 设 $(X_1, X_2, \cdots, X_n)$ 是取自总体 X 的一个样本, X 的 概率密度函数为

$$f(x,\theta)=\begin{cases}\dfrac{2x}{\theta^2} & 0<x<\theta\\ 0 & \text{其余}\end{cases}$$

试写出 $(X_1, X_2, \cdots, X_n)$ 的联合密度函数.

解 联合密度函数为

$$f^*(x_1,x_2,\cdots,x_n,\theta)=\begin{cases}\dfrac{2^n x_1\cdots x_n}{\theta^{2n}} & 0<x_i<\theta,\ i=1,2,\cdots,n\\ 0 & \text{其余}\end{cases}$$

例8 设 $(X_1, X_2, \cdots, X_n)$ 是取自总体 X 的一个样本,X 的分布律为

X	-1	0	1
Pr.	$\frac{\theta}{2}$	$1-\theta$	$\frac{\theta}{2}$

其中参数$\theta(0 < \theta < 1)$,

试写出$(X_1, X_2, \cdots, X_n)$的联合 分布律?

目录/Contents

目录/Contents

6.2 统计量

数理统计的基本任务之一是利用样本所提供的信息来对总体分布中未知的量进行推断，但是样本观测值常常表现为一大堆数字，很难直接用来解决我们所要研究的具体问题．人们常常把数据加工成若干个简单明了的数字特征,由数据加工后的数字特征就是**统计量的观测值**.

不含有未知参数的样本的函数 $g(X_1,\cdots,X_n)$ 称为统计量.

例1 **假设总体 $X \sim U(0,\theta),(X_1,X_2,\cdots,X_n)$ 为取自该总体的一个样本,**

下列表达式是否为统计量:

1. $T_1=\dfrac{X_1+\cdots+X_6}{6}$ 是统计量，因为不含总体中的唯一未知参数 θ；
2. $T_2=X_6-\theta$ 不是统计量，因为含有未知参数 θ；
3. $T_3=\max(X_1,\cdots,X_6)$ 是统计量，因为不含总体中的唯一未知参数 θ .

补例 假设总体 $X \sim N(\mu,\sigma^2),(X_1,X_2,\cdots,X_n)$ 为取自该总体的一个样本,

其中 μ 未知而 σ^2 已知. 设 $n=4$，试判断下列表达式是否为统计量:

(1) $\sum_{i=1}^{4} X_i^2$; (2) $\frac{1}{4}\sum_{i=1}^{4}(X_i-\mu)^2$; (3) $\frac{1}{\sigma^2}\sum_{i=1}^{4}(X_i-\bar{X})^2$.

解 由定义即知(2)不是统计量，而(1)(3)是.

一、样本均值和样本方差

设 $(X_1, X_2, \cdots, X_n)$ 为取自总体的一个样本，称

(1)**样本均值** $\bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ 样本均值的观测值 $\bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$

(2)**样本方差** $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$ 样本方差的观测值 $s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$

(3)**样本标准差** $S = \sqrt{\dfrac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$ 样本标准差的观测值 $s = \sqrt{\dfrac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$

(4)样本的k 阶**原点矩**

$$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \qquad \left(A_1 = \bar{X}\right)$$

(5)样本的k 阶**中心矩**

$$M_k = \frac{1}{n}\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^k$$

其中k是正整数，而每个统计量也都有相应的观测值.

当 $k=2$ 时，样本的2阶中心矩

$$M^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2 \triangleq S_n^2,$$

由此可得：

$$S_n=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}$$

相应的观测值

$$s_n^2=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2,\quad s_n=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}$$

注 S^2, S_n^2 **在计算时的另一表达形式:**

$$\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2=\sum_{i=1}^{n}\left(X_i^2-2X_i\bar{X}+\bar{X}^2\right)=\sum_{i=1}^{n}X_i^2-2\bar{X}\sum_{i=1}^{n}X_i+n\bar{X}^2$$

$$=\sum_{i=1}^{n}X_i^2-2\bar{X}\cdot n\bar{X}+n\bar{X}^2=\sum_{i=1}^{n}X_i^2-n\bar{X}^2$$

由此知

$$S^2=\frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2-n\bar{X}^2\right)\qquad S_n^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}^2$$

显然有

$$(n-1)S^2=nS_n^2=\sum_{i=1}^{n}(X_i-\bar{X})^2$$

一、样本均值和样本方差

定理

设总体 X 的均值 $E(X)=\mu$ ，方差 $D(X)=\sigma^2$ ，$(X_1,X_2,\cdots,X_n)$ 为取自该总体的一个样本，则

（1） $E(\bar{X})=\mu, D(\bar{X})=\dfrac{\sigma^2}{n}$

（2） $E(S^2)=\sigma^2, E(S_n^2)=\dfrac{n-1}{n}\sigma^2, n\geq 2$

（3） $\bar{X}\xrightarrow{P}\mu$， $S^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2\xrightarrow{P}\sigma^2$， $S_n^2=\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\xrightarrow{P}\sigma^2$.

证（1）

$$E\left(\bar{X}\right)=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n}\sum_{i=1}^{n}E\left(X_i\right)=\mu,$$

$$D\left(\bar{X}\right)=D\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}D\left(X_i\right)=\frac{1}{n}\sigma^2,$$

证（2）

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

$$=n\left(D\left(X_1\right)+E^2\left(X_1\right)-D\left(\bar{X}\right)-E^2\left(\bar{X}\right)\right)$$

$$=n\left(\sigma^2-\frac{\sigma^2}{n}\right)=(n-1)\sigma^2$$

$$\Rightarrow E\left(S^2\right)=\sigma^2,\quad E\left(S_n^2\right)=\frac{n-1}{n}\sigma^2$$

证（3） 独立同分布的大数定律，即得 $\bar{X} \xrightarrow{P} \mu$

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{P} \sigma^2+\mu^2,$$

所以 $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2 = \frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}^2 \xrightarrow{P} \sigma^2+\mu^2-\mu^2=\sigma^2$

$$S^2=\frac{n}{n-1}S_n^2 \xrightarrow{P} \sigma^2$$

例2 设$(X_1,X_2,\cdots,X_n)$是取自总体 X 的一个样本,在下列三种情况下,

分别求 $E(\bar{X}),D(\bar{X}),E(S^2),E\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right)$.

（1） $X\sim B(1,p)$；（2） $X\sim E(\lambda)$；（3） $X\sim U(0,2\theta)$,其中$\theta>0$.

解 **由定理可得**

$$E(\bar{X})=E(X)\triangleq\mu,D(\bar{X})=\frac{D(X)}{n}\triangleq\frac{\sigma^2}{n},$$

$$E(S^2)=D(X)=\sigma^2,$$

$$E\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2\right)=D(X)+(E(X))^2=\sigma^2+\mu^2$$

根据题意可知

(1) $X \sim B(1,p)$，所以$E(X)=p, D(X)=p(1-p)$，

故$E(\bar{X})=E(X)=p$， $D(\bar{X})=\dfrac{D(X)}{n}=\dfrac{p(1-p)}{n}$，

$E(S^2)=D(X)=p(1-p)$，

$E(\frac{1}{n}\sum_{i=1}^{n}X_i^2)=\frac{1}{n}\sum_{i=1}^{n}E(X_i^2)=E(X^2)=p^2$.

根据题意可知

（2） $X \sim E(\lambda)$，所以$E(X)=\frac{1}{\lambda}, D(X)=\frac{1}{\lambda^2}$；

故$E(\bar{X})=\frac{1}{\lambda}, D(\bar{X})=\frac{1}{n\lambda^2}, E(S^2)=\frac{1}{\lambda^2}$， $E(\frac{1}{n}\sum_{i=1}^{n}X_i^2)=\frac{2}{\lambda^2}$.

（3） $X \sim U(0,2\theta)$，其中$\theta>0$，所以$E(X)=\theta, D(X)=\frac{\theta^2}{3}$，

故$E(\bar{X})=\theta, D(\bar{X})=\frac{\theta^2}{3n}, E(S^2)=\frac{\theta^2}{3}$， $E(\frac{1}{n}\sum_{i=1}^{n}X_i^2)=\frac{4}{3}\theta^2$.

一、样本均值和样本方差

***例3**

设$(X_1,X_2,\cdots,X_n)$是取自总体 X 的一个样本，总体$X \sim P(\lambda)$,求$\bar{X}$与$\frac{1}{n}\sum_{i=1}^{n}X_i^2$分别依概率收敛于什么值?

解

$$\bar{X}\xrightarrow{P}\lambda,\ \frac{1}{n}\sum_{i=1}^{n}X_i^2\xrightarrow{P}E\left(X^2\right)=\lambda+\lambda^2$$

二、次序统计量

设 $(X_1, X_2, \cdots, X_n)$ 为取自总体 X 的一个样本，其中取值最大的记为 $X_{(n)}$ ，称为最大次序统计量；而取值最小的记为 $X_{(1)}$ 称为最小次序统计量；

$X_{(1)}$ 与 $X_{(n)}$ 统称为次序统计量.

二、次序统计量

设总体 X 具有分布函数$F(x)$ 及概率密度函数 $f(x)$ ，

则最小次序统计量具有概率密度函数:

$$f_{X_{(1)}}(y)=n\left[1-F(y)\right]^{n-1}f(y)$$

最大次序统计量具有概率密度函数:

$$f_{X_{(n)}}(y)=nF(y)^{n-1}f(y).$$

二、次序统计量

由第三章中关于独立同分布情形下求最大或最小随机变量的分布函数之结论知

$$F_n(y)=F^n(y), F_1(y)=1-(1-F(y))^n$$

各自求导即得.

例4

设总体 $X \sim E(\lambda)$，$(X_1, X_2, \cdots, X_n)$是取自总体 X 的一个样本,

分别求次序统计量 $X_{(1)}$，$X_{(n)}$ 的分布.

解

$X \sim E(\lambda)$，所以密度函数为 $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & \text{其它}, \end{cases} \quad \lambda > 0$

分布函数为 $F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \ge 0. \end{cases}$

因此， 根据公式 $f_{X_{(1)}}(v)=n(1-F_X(v))^{n-1}f_X(v)$，

可得 $f_{X_{(1)}}(v)=\begin{cases} n\lambda e^{-n\lambda v}, & v>0 \\ 0, & \text{其余} \end{cases}$， 即 $X_{(1)}\sim E(n\lambda)$

根据公式 $f_{X_{(n)}}(u)=n\left(F_X(u)\right)^{n-1}f_X(u)$，

可得 $f_{X_{(n)}}(u)=\begin{cases} n\lambda e^{-\lambda u}(1-e^{-\lambda u})^{n-1}, & u>0 \\ 0, & \text{其余} \end{cases}$

例5 设$(X_1,X_2,\cdots,X_n)$是取自总体X的一个样本，总体$X\sim U(0,1)$.求最小次序统计量$X_{(1)}$的均值和方差.

解 $X_{(1)}$有密度函数$f_{X_{(1)}}(y)=\begin{cases} n(1-y)^{n-1}, 0<y<1 \\ 0,\text{其他} \end{cases}$，所以

$$E(X_{(1)})=n\int_0^1 y(1-y)^{n-1}\mathrm{d}y=n\int_0^1 y(1-y)^{n-1}\mathrm{d}y=n\times\frac{1}{(n+1)n}=\frac{1}{(n+1)};$$

又$E(X_{(1)}^2)=n\int_0^1 y^2(1-y)^{n-1}\mathrm{d}y=n\int_0^1 y^2(1-y)^{n-1}\mathrm{d}y=n\times\frac{2}{(n+2)(n+1)n}=\frac{2}{(n+2)(n+1)}$,

因此 $\mathrm{D}(X_{(1)})=\frac{2}{(n+2)(n+1)}-\frac{1}{(n+1)^2}=\frac{n}{(n+2)(n+1)^2}$.

目录/Contents

目录/Contents

6.3 总体与样本

一、χ^2分布

二、t 分布

三、F 分布

一、χ^2分布

(1) 定义

设 $X_1,\cdots,X_n$ 是独立同分布的随机变量，且都服从标准正态分布$N(0,1)$ ，

则称随机变量 $Y \triangleq \sum_{i=1}^{n} X_i^2$

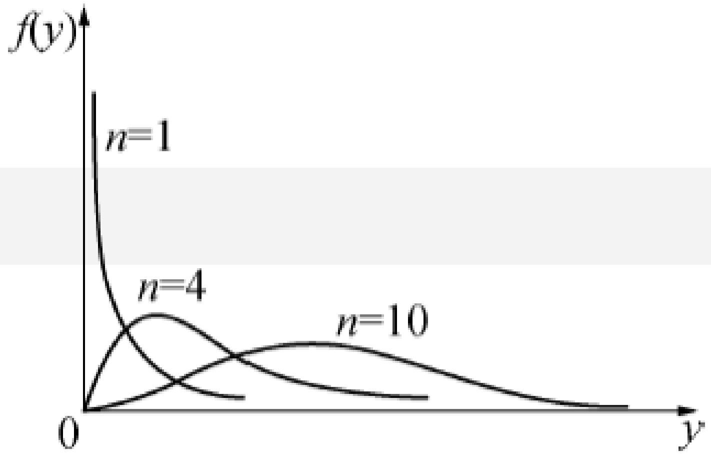
所服从的分布为自由度为n的 χ^2 **分布**，记为$Y \sim \chi^2(n)$.

一、χ^2分布

由定义可知，若 X_1, X_2 相互独立且都服从 $N(0,1)$ ，则

1 $X_1^2 \sim \chi^2(1)$, 2 $X_2^2 \sim \chi^2(1)$, 3 $X_1^2 + X_2^2 \sim \chi^2(2)$.

下图分别是当 $n=1,4,10$ 时的概率密度函数图形. 是偏峰的倒钟形.



一、χ^2分布

(2) χ^2 分布性质

①当 $Y\sim\chi^2(n)$ 时，$E(Y)=n, D(Y)=2n$ ；

②χ^2 分布的可加性:

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

一、χ^2分布

证①：$E(Y)=n, D(Y)=2n$

由χ^2分布定义知

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

$$=n\left(D\left(X_1\right)+E^2\left(X_1\right)\right)=n$$

$$D(Y)=D\left(\sum_{i=1}^{n}X_i^2\right)=\sum_{i=1}^{n}D\left(X_i^2\right)$$

$$=n\left(E\left(X_1^4\right)-E^2\left(X_1^2\right)\right)=n(3-1)=2n$$

一、χ^2分布

证②：设$X \sim \chi^2(m), Y \sim \chi^2(n)$，**且** X **与**Y**相互独立**

由χ^2**分布定义知** $X = \sum_{i=1}^{m} X_i^2; Y = \sum_{i=1}^{n} Y_i^2$ **，其中**

$X_1, \cdots, X_m, Y_1, \cdots, Y_n$ **都是相互独立的标准正态分布**

则：

$$X + Y = \sum_{i=1}^{m} X_i^2 + \sum_{i=1}^{n} Y_i^2 \sim \chi^2(m+n).$$

一、χ^2分布

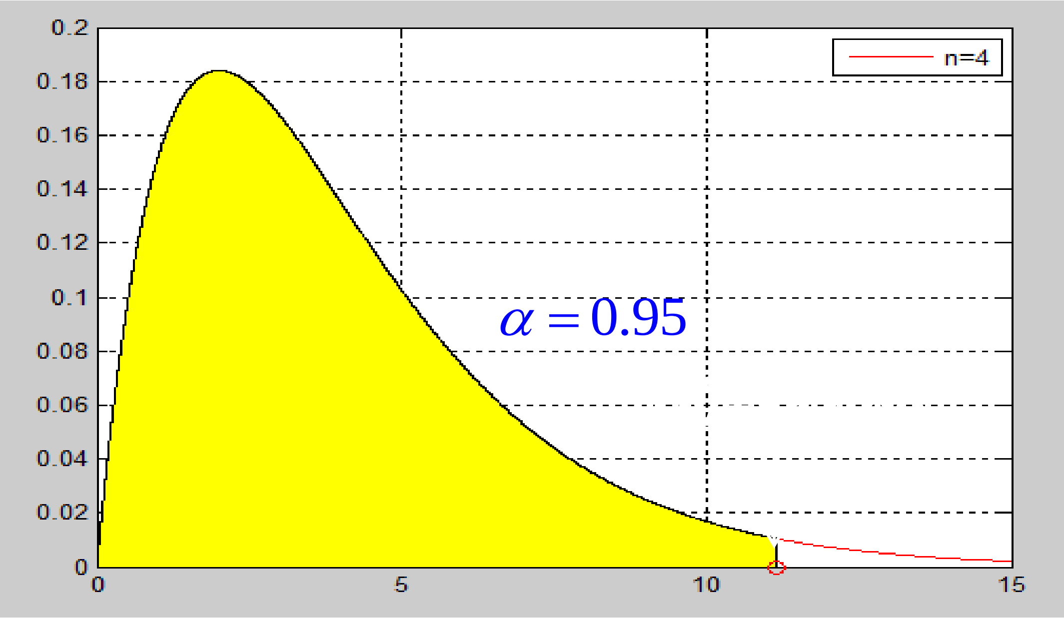
(3)χ^2分布的分位数

设 $X \sim \chi^2(n)$ ，记它的α 分位数为$\chi^2_\alpha(n)$，即$\chi^2_\alpha(n)$ 满足

$$P\left(X \le \chi^2_\alpha(n)\right) = \alpha.$$ 见图示.

分位数值可查表得到，比如

$$\chi^2_{0.95}(4) = 9.488$$



一、χ^2分布

例1 设 $(X_1,X_2,\cdots,X_6)$ 是取自总体$N(0,1)$ 的简单随机样本,

求下列三个统计量的分布

01 OPTION （1） $X_1^2+X_2^2$；

02 OPTION （2） X_1^2；

03 OPTION $(3)Q=X_1^2+a\left(X_2+X_3\right)^2+b\left(X_4-X_5+X_6\right)^2$

并求非零常数 a,b 的值.

一、χ^2分布

解 （1）由样本的定义可知，$X_1,X_2,\cdots,X_6$相互独立，且都服从$N(0,1)$，所以根据χ^2分布的定义可知$X_1^2+X_2^2\sim\chi^2(2)$；

（2）同上，$X_1^2\sim\chi^2(1)$；

一、χ^2分布

(3) $X_2+X_3 \sim N(0,2) \Rightarrow \dfrac{X_2+X_3}{\sqrt{2}} \sim N(0,1),$

$X_4-X_5+X_6 \sim N(0,3) \Rightarrow \dfrac{X_4-X_5+X_6}{\sqrt{3}} \sim N(0,1),$

且 $X_1, \dfrac{1}{\sqrt{2}}(X_2+X_3), \dfrac{1}{\sqrt{3}}(X_4-X_5+X_6)$ **相互独立,**

再由χ^2 分布的定义

$$X_1^2+\left(\frac{X_2+X_3}{\sqrt{2}}\right)^2+\left(\frac{X_4-X_5+X_6}{\sqrt{3}}\right)^2 \sim \chi^2(3).$$

可得 $a=\dfrac{1}{2}, b=\dfrac{1}{3}$

一、χ^2分布

例 2 设 $X_1,\cdots,X_n$ 是来自正态总体 $N(0,\sigma^2)$ 的样本，试证：

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

证明 (1) $\frac{X_i}{\sigma}, i=1,\cdots,n$ 独立同分布于 $N(0,1)$，由 χ^2 分布的定义，$\sum_{i=1}^{n}\left(\frac{X_i}{\sigma}\right)^2 \sim \chi^2(n)$，

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

---

(2) 易见，$\sum_{i=1}^{n} X_i \sim N(0, n\sigma^2)$，即 $\frac{\sum_{i=1}^{n} X_i}{\sqrt{n\sigma^2}} \sim N(0,1)$，由 χ^2 分布的定义，

$\left(\frac{\sum_{i=1}^{n} X_i}{\sqrt{n\sigma^2}}\right)^2 \sim \chi^2(1)$，即 $\frac{1}{n\sigma^2}\left(\sum_{i=1}^{n} X_i\right)^2 \sim \chi^2(1)$.

一、χ^2分布

补例 设$(X_1, X_2, \cdots, X_n)$ 是取自总体 $X \sim \chi^2(n)$ 的一个样本，

定义 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$，试求$E(\bar{X}), D(\bar{X})$.

解 由 χ^2分布性质知 $E(X) = n, D(X) = 2n,$ 故

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = n.$$

一、χ^2分布

补例 设 $(X_1, X_2, \cdots, X_n)$ 是取自总体 $X \sim \chi^2(n)$ 的一个样本，

定义 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ ，试求 $E(\bar{X}), D(\bar{X})$.

解 由 χ^2 分布性质知 $E(X) = n, D(X) = 2n,$ 故

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

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

二、t分布

（1）定义

设随机变量 $X \sim N(0,1), Y \sim x^2(n)$ ，且X 与Y 相互独立,

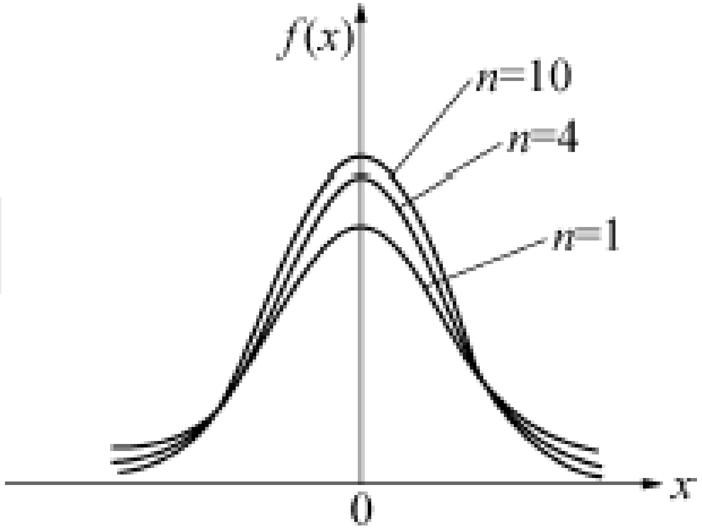
则称 $T \triangleq \dfrac{X}{\sqrt{Y/n}}$ 所服从的分布为自由度为n 的 t分布或者学生分布,

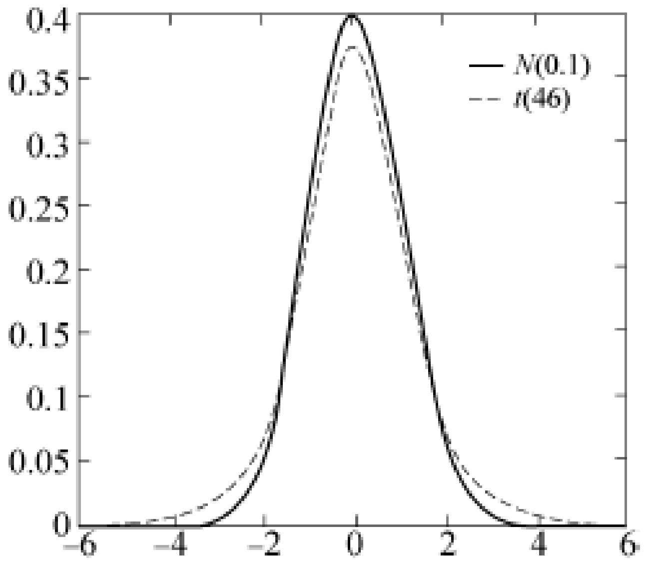
记为 $T \sim t(n)$.

二、t分布

t分布的概率密度函数图形，

当 n 充分大时，$t(n)$分布近似于 $N(0,1)$分布





二、t分布

（2）分位数

设 $X \sim t(n)$ ，记它的α 分位数为$t_\alpha(n)$ ，即$t_\alpha(n)$ 满足

$$P\left(X \le t_\alpha(n)\right) = \alpha.$$

根据 t分布密度函数的对称性，有性质 $t_\alpha(n) = -t_{1-\alpha}(n)$.

该性质类似于正态分布的分位数性质．分位数值查表可得.

比如 $t_{0.95}(10) = -t_{0.05}(10)=1.812.$

例3

设$T \sim t(10)$，求常数c使$P(T > c) = 0.95$；

解

由$P(T > c) = 0.95$可知，$P(T \le c) = 0.05$，所以$c = t_{0.05}(10)$，由 t 分布的密度函数图像对称性，所以$c = t_{0.05}(10) = -t_{0.95}(10)$，查附录 6 得$c = -1.8125$.

补充例题

设 $X_1, X_2, \cdots, X_5$ 为取自标准正态总体的一个样本，

试给出常数 d ，使得

$$d\frac{X_1+X_2}{\sqrt{X_3^2+X_4^2+X_5^2}}$$

服从 t 分布，并指出它的自由度.

解

因 $X_i \sim N(0,1),\ i=1,2,\cdots,5$ ，且相互独立，故有

$$X_3^2+X_4^2+X_5^2 \sim \chi^2(3),\ X_1+X_2 \sim N(0,2)$$

且两者相互独立，由定义知

$$\frac{\dfrac{X_1+X_2}{\sqrt{2}}}{\sqrt{\dfrac{X_3^2+X_4^2+X_5^2}{3}}} \sim t(3) \Rightarrow \frac{\sqrt{3}}{\sqrt{2}} \frac{X_1+X_2}{\sqrt{X_3^2+X_4^2+X_5^2}} \sim t(3)$$

所以，取 $d=\dfrac{\sqrt{3}}{\sqrt{2}}=\dfrac{\sqrt{6}}{2}$ 即可，且自由度为3.

三、F 分布

（1）定义

设随机变量 $X \sim x^2(m), Y \sim x^2(n)$ ，且X 与Y 相互独立,

则称 $F \triangleq \dfrac{X/m}{Y/n}$ 服从第一自由度为 m，第二自 由度为 n 的**F分布**,

记为 $F \sim \mathrm{F}(m,n)$.

三、 F 分布

补充例题

设随机变量 $X_1, X_2, \cdots, X_5$ 相互独立且都服从相同 分布 $X_i \sim N(0,1)$ ，$i = 1, \cdots, 5$.

试给出常数 a ，使得

$$a\frac{X_1^2 + X_2^2}{X_3^2 + X_4^2 + X_5^2}$$

服从 F 分布，并指出它的自由度.

三、F 分布

解 因 $X_i \sim N(0,1),\ i=1,2,\cdots,5$，且相互独立，故有

$$X_1^2+X_2^2 \sim \chi^2(2),\ X_3^2+X_4^2+X_5^2 \sim \chi^2(3)$$

且两者相互独立，由定义知

$$\frac{(X_1^2+X_2^2)/2}{(X_3^2+X_4^2+X_5^2)/3} \sim \mathrm{F}(2,3)$$

所以，取 $a=\dfrac{3}{2}$ 即可，且自由度为2和3.

三、F 分布

（2）分位数

设 $X \sim F(m,n)$ ，记它的α 分位数为 $F_\alpha(m,n)$ ，即$F_\alpha(m,n)$ 满足

$$P\left(X \le F_\alpha(m,n)\right) = \alpha.$$

根据F分布的定义，有性质

$$F_\alpha(m,n) = \frac{1}{F_{1-\alpha}(n,m)}.$$

三、F 分布

证明

设 $F \sim \mathrm{F}(m,n)$. 则可以定义 $F \triangleq \dfrac{X/m}{Y/n}$，其中 $X \sim \chi^2(m), Y \sim \chi^2(n)$

且 X与 Y相互独立。则 $\dfrac{1}{F} \triangleq \dfrac{Y/n}{X/m} \sim \mathrm{F}(n,m)$.

所以有 $P\left(F \le F_\alpha(m,n)\right) = P\left(\dfrac{1}{F} \ge \dfrac{1}{F_\alpha(m,n)}\right) = \alpha$ 即 $P\left(\dfrac{1}{F} \le \dfrac{1}{F_\alpha(m,n)}\right) = 1-\alpha$

根据前面 $\dfrac{1}{F} \sim \mathrm{F}(n,m)$ 及分位数的定义，显然有

$$\frac{1}{F_\alpha(m,n)} = F_{1-\alpha}(n,m)$$

例4

设随机变量$T \sim t(n)$，求 $F = \frac{1}{T^2}$ 的分布?

解

由于 $T \sim t(n)$，不妨设 $T = \frac{X}{\sqrt{Y/n}}$，其中随机变量 X 与 Y 独立，$X \sim N(0,1)$，$Y \sim \chi^2(n)$.

则 $F = \frac{1}{T^2} = \frac{Y/n}{X^2}$，因为 $X^2 \sim \chi^2(1)$，且 X^2 与 Y 独立，故由 F 分布定义，即知 $F \sim F(n,1)$。

三、F 分布

补例 由F定义可知，若 X_1, X_2, X_3 相互独立且都服从 $N(0,1)$，

则有 $X_1^2 \sim \chi^2(1), X_2^2 + X_3^2 \sim \chi^2(2)$

且 X_1^2 与 $X_2^2 + X_3^2$ 相互独立，故

$$\frac{X_1^2/1}{(X_2^2+X_3^2)/2} = \frac{2X_1^2}{X_2^2+X_3^2} = \left(\frac{\sqrt{2}X_1}{\sqrt{X_2^2+X_3^2}}\right)^2 \sim F(1,2).$$

目录/Contents

定理 1 设 $(X_1, X_2, \cdots, X_n)$ 是取自总体 $N(\mu, \sigma^2)$ 的一个简单随机样本,

则有

（1） $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ ，即 $\sqrt{n}\frac{\bar{X}-\mu}{\sigma} \sim N(0,1)$;

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

（3） $\bar{X}$ 与 S^2 相互独立.

定理2

设 $(X_1, X_2, \cdots, X_n)$ 是取自总体 $N(\mu, \sigma^2)$ 的一个简单随机样本，

则有

$$\sqrt{n}\,\frac{\bar{X}-\mu}{S}=\sqrt{n-1}\,\frac{\bar{X}-\mu}{S_n}\sim t(n-1)$$

证明 由定理1可知

由（1）知 $\sqrt{n}\dfrac{\bar{X}-\mu}{\sigma}\sim N(0,1)$，

由（2）知 $\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$

由（3）知 $\bar{X}$ 与 S^2 相互独立，

证明

根据t分布的定义，可得

$$\frac{\sqrt{n}\,\dfrac{\bar{X}-\mu}{\sigma}}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2}/(n-1)}}=\sqrt{n}\,\frac{\bar{X}-\mu}{S}\sim t(n-1)$$

补例 设总体 $X \sim N\left(40,5^2\right)$ ，试求解下列各题:

1 抽取容量为36的样本，求 $P\left(38 \le \bar{X} \le 43\right)$ ；

2 问样本容量n 多大时，才能使 $P\left(\left|\bar{X}-40\right|<1\right)=0.95$ ？

解(1) 由抽样分布及 $n=36$ 可知

$$\bar{X} \sim N\left(40, \frac{25}{36}\right)$$

所以

$$P\left(38 \le \bar{X} \le 43\right)=\Phi\left(\frac{43-40}{5/6}\right)-\Phi\left(\frac{38-40}{5/6}\right)$$

$$=\Phi(3.6)-1+\Phi(2.4) \approx 0.9918.$$

(2) 由 $X \sim N(40,25) \Rightarrow \bar{X} \sim N\left(40,\dfrac{25}{n}\right)$， 也即 $\dfrac{\bar{X}-40}{\dfrac{5}{\sqrt{n}}} \sim N(0,1)$，

从而有 $0.95 = P\left(\left|\bar{X}-40\right|<1\right) = P\left(\left|\dfrac{\bar{X}-40}{5/\sqrt{n}}\right| < \dfrac{1}{5/\sqrt{n}}\right)$

$$= P\left(\left|\frac{\bar{X}-40}{5/\sqrt{n}}\right| < \frac{\sqrt{n}}{5}\right) = 2\Phi\left(\frac{\sqrt{n}}{5}\right) - 1,$$

$$\Rightarrow \Phi\left(\frac{\sqrt{n}}{5}\right) = 0.975 \Rightarrow \frac{\sqrt{n}}{5} = u_{0.975} = 1.96 \Rightarrow n = 96.04.$$

所以，取 $n = 97.$

正态总体的抽样分布

补例

设 $(X_1, X_2, \cdots, X_n)$ 是独立同分布的随机变量,且 $X \sim N(0,1)$

请证明以下结论

01 OPTION

$$n\bar{X}^2 \sim \chi^2(n-1);$$

02 OPTION

$$D\left(\bar{X}^2\right) = \frac{2}{n^2};$$

03 OPTION

$$D\left(\bar{X}^2 + S^2\right) = \frac{2}{n^2} + \frac{2}{n-1}.$$

定理3

设 $\left(X_1, X_2, \cdots, X_m\right)$ 是取自总体 $X \sim N\left(\mu_1, \sigma_1^2\right)$ 的一个简单随机样本，

$\left(Y_1, Y_2, \cdots, Y_n\right)$ 是取自总体 $Y \sim N\left(\mu_2, \sigma_2^2\right)$ 的一 个简单随机样本，

两个总体相互 独立。

定义：

$$\bar{X}=\frac{1}{m}\sum_{i=1}^{m}X_i,\bar{Y}=\frac{1}{n}\sum_{i=1}^{n}Y_i,$$

$$S_X^2=\frac{1}{m-1}\sum_{i=1}^{m}\left(X_i-\bar{X}\right)^2,S_Y^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2,$$

$$S_w^2=\frac{(m-1)S_X^2+(n-1)S_Y^2}{m+n-2},$$

正态总体的抽样分布

则有

01 OPTION

$$\bar{X}-\bar{Y}\sim N\left(\mu_1-\mu_2,\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}\right);\quad \text{即}\ \frac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}\sim N(0,1).$$

02 OPTION

$$\frac{(m-1)S_X^2}{\sigma_1^2}+\frac{(n-1)S_Y^2}{\sigma_2^2}\sim\chi^2(m+n-2)$$

03 OPTION

$$\frac{S_X^2/\sigma_1^2}{S_Y^2/\sigma_2^2}\sim F(m-1,n-1)$$

04 OPTION

$$\sigma_1^2=\sigma_2^2=\sigma^2\text{时，}\ \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{m}+\frac{1}{n}}}\sim t(m+n-2)$$

正态总体的抽样分布

证明（1）

1 $\bar{X} \sim N\left(\mu_1, \dfrac{\sigma_1^2}{m}\right)$　　**2** $\bar{Y} \sim N\left(\mu_2, \dfrac{\sigma_2^2}{n}\right)$

则 $\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \dfrac{\sigma_1^2}{m} + \dfrac{\sigma_2^2}{n}\right)$ **,即**

$$\frac{\left(\bar{X} - \bar{Y}\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\dfrac{\sigma_1^2}{m} + \dfrac{\sigma_2^2}{n}}} \sim N(0,1).$$

证明（2） 由定理1可知

$$\frac{\sum_{i=1}^{m}(X_i-\bar{X})^2}{\sigma_1^2}=\frac{(m-1)S_X^2}{\sigma_1^2}\sim\chi^2(m-1)$$

$$\frac{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}{\sigma_2^2}=\frac{(n-1)S_Y^2}{\sigma_2^2}\sim\chi^2(n-1)$$

则由χ^2分布的可加性可得

$$\frac{(m-1)S_X^2}{\sigma_1^2}+\frac{(n-1)S_Y^2}{\sigma_2^2}\sim\chi^2(m+n-2)$$

证明（3）则由 F 分布的定义可得

$$\frac{(m-1)S_X^2}{\sigma_1^2(m-1)}\bigg/\frac{(n-1)S_Y^2}{\sigma_2^2(n-1)}\sim F(m-1,n-1)，即 \frac{S_X^2/\sigma_1^2}{S_Y^2/\sigma_2^2}=\frac{S_X^2/S_Y^2}{\sigma_1^2/\sigma_2^2}\sim F(m-1,n-1)。$$

当 $\sigma_1^2=\sigma_2^2=\sigma^2$ 时，

即有 $\dfrac{S_X^2}{S_Y^2}\sim F(m-1,n-1)$.

证明（4）

当两个正态总体有相同的方差，即 $\sigma_1^2=\sigma_2^2=\sigma^2$ 时，

结论（1）和（2）分别转变为下面形式：

1 $$\frac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_1-\mu_2\right)}{\sigma\sqrt{\frac{1}{m}+\frac{1}{n}}} \sim N(0,1);$$

2 $$\frac{(m-1)S_X^2+(n-1)S_Y^2}{\sigma^2} \sim \chi^2(m+n-2)$$

知 $\bar{X}$、$\bar{Y}$ 、S_X^2、S_Y^2 相互独立，根据 t 分布的定义，可得

$$\frac{[(\bar{X}-\bar{Y})-(\mu_1-\mu_2)] \Big/ \sigma\sqrt{\frac{1}{m}+\frac{1}{n}}}{\sqrt{[(m-1)S_X^2+(n-1)S_Y^2] \big/ \sigma^2(m+n-2)}} \sim t(m+n-2)$$

即

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{m}+\frac{1}{n}}} \sim t(m+n-2)$$

总结/summary

统计量和抽样分布	总体与样本	
	常用统计量	样本均值，样本方差，次序统计量等
	三大分布	χ^2分布，t分布，F分布
	正态总体的抽样分布	