第六章　数理统计的基本概念

关键词：

总 体

个 体

样 本

统 计 量

χ^2-分布

$t-$分布

$F-$分布

引言：

数理统计学是一门关于数据收集、整理、分析和推断的学科。

在概率论中已经知道，由于大量的随机试验中各种结果的出现必然呈现它的规律性，因而从理论上讲只要对随机现象进行足够多次观察，各种结果的规律性一定能清楚地呈现，但是实际上所允许的观察永远是有限的，甚至是少量的。

例如：若规定灯泡寿命低于1000小时者为次品，如何确定次品率？由于灯泡寿命试验是破坏性试验，不可能把整批灯泡逐一检测，只能抽取一部分灯泡作为样本进行检验，以样本的信息来推断总体的信息，这是数理统计学研究的问题之一。

§1 随机样本与统计量

- 总体：研究对象的全体。如一批灯泡。
- 个体：组成总体的每个元素。如某个灯泡。

总体是某一数量指标的全体，是具有确定分布的随机变量。

- 抽样：从总体X中抽取有限个个体, 进行观察的取值过程。
- 随机样本：随机抽取的n个个体的集合$(X_1,X_2,\ldots,X_n)$，n为样本容量
- 满足以下两个条件的随机样本$(X_1,X_2,\ldots,X_n)$称为简单随机样本：
  1. 代表性：每个X_i与X同分布
  2. 独立性：$X_1,X_2,\cdots,X_n$是相互独立的随机变量

[说明]：后面提到的样本均指简单随机样本。

由概率论知，若总体X具有概率密度$f(x)$，

则样本（$X_1, X_2, \cdots, X_n$）具有联合密度函数：

$$f_n(x_1,x_2,\cdots x_n) = f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_n}(x_n) = \prod_{i=1}^{n} f(x_i)$$

常用统计量

- 统计量：不含任何未知参数的样本的函数。

设（$X_1, X_2, \cdots, X_n$）为取自总体X的样本，

1. 样本均值 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

2. 样本方差 $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$，$S$为样本标准差

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

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

注：$A_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$ ，$B_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{n-1}{n} S^2$

样本与总体的各阶矩对比表

特征数据	样本(随机变量) $\xrightarrow{P}$	总体（常数）
均值	$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$	$E(X)=\sum_{i=1}^{+\infty}x_i p_i$
方差	$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$	$Var(X)=E\left(X-E(X)\right)^2$
均方差/标准差	$S=\sqrt{S^2}$	$\sigma=\sqrt{Var(X)}$
k 阶原点矩	$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$	$\mu_k=E(X^k)$
k 阶中心矩	$B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k$	$\nu_k=E(X-E(X))^k$

思考题：

设在总体$N\left(\mu,\sigma^2\right)$中抽取样本$\left(X_1,X_2,X_3\right)$,其中$\mu$已知，$\sigma^2$未知, 指出在

(1) $X_1+X_2+X_3$

(2) $X_2+2\mu$

(3) $\max\left(X_1,X_2,X_3\right)$

(4) $\dfrac{1}{\sigma^2}\sum_{i=1}^{3}X_i^2$

(5) $\left|X_3-X_1\right|$

中哪些是统计量，哪些不是统计量，为什么？

答：只有(4)不是统计量。

例1 设$X_1,X_2,\dots,X_n$是总体X的样本，若$E(X)=\mu, Var(X)=\sigma^2$，

则 $E(\bar{X})=$ $\underline{E(X)}$，$Var(\bar{X})=$ $\underline{Var(X)/n}$，$E(S^2)=$ $\underline{Var(X)}$.

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

$$Var(\bar{X})=Var(\frac{1}{n}\sum_{i=1}^{n}X_i)=\frac{1}{n^2}\sum_{i=1}^{n}Var(X_i)=\frac{1}{n^2}\sum_{i=1}^{n}\sigma^2=\frac{1}{n^2}n\sigma^2=\frac{\sigma^2}{n}$$

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

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

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

$$\therefore E[(n-1)S^2]=\sum_{i=1}^{n}E(X_i^2)-nE(\bar{X}^2)=\sum_{i=1}^{n}(Var(X_i)+E^2(X_i))-n(Var(\bar{X})+E^2(\bar{X}))$$

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

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

而$E(B_2)=E(\frac{n-1}{n}S^2)=\frac{n-1}{n}\sigma^2$

例2 设总体X的概率密度$f(x)=\begin{cases}2(1-x), & 0<x<1,\\ 0, & \text{其它}\end{cases}$

$X_1, X_2, \ldots, X_n$是样本，求$E(\bar{X}),\ E(S^2),\ E(\bar{X}^2)$，$P(\overline{X}=E(X))$

解：$E(X)=\int_0^1 2x(1-x)dx=\dfrac{1}{3},\qquad E(X^2)=\int_0^1 2x^2(1-x)dx=\dfrac{1}{6}$

$$Var(X)=E(X^2)-E^2(X)=\frac{1}{18}$$

$$E(\overline{X})=E(X)=\frac{1}{3}$$

$$E(S^2)=Var(X)=\frac{1}{18}$$

$$E(\overline{X}^2)=Var(\overline{X})+E^2(\overline{X})=\frac{Var(X)}{n}+E^2(X)=\frac{1}{18n}+\frac{1}{9}$$

$$P(\overline{X}=E(X))=?$$

例3 设$(X_1,X_2,\cdots,X_5)$是总体$X\sim N(12,4)$的一个样本，求

（1）样本均值与总体均值之差大于1的概率；

(2) $P(\min\{X_1,X_2,...,X_5\}\le 10)$.

解：（1）$E(\bar{X})=E(X)=12, Var(\bar{X})=\dfrac{Var(X)}{5}=\dfrac{4}{5}\quad \therefore \bar{X}\sim N(12,\dfrac{4}{5})$

$$P(\bar{X}-12>1)=P(\bar{X}>13)=1-\Phi\left(\frac{13-12}{\sqrt{4/5}}\right)=1-\Phi(1.12)=0.1314$$

（2）$P(X>10)=1-\Phi\left(\dfrac{10-12}{2}\right)=1-\Phi(-1)=\Phi(1)=0.8413$

$$P(\min\{X_1,X_2,...,X_5\}\le 10)=1-P(\min\{X_1,X_2,...,X_5\}>10)$$

$$\overset{独立}{=}1-P(X_1>10)P(X_2>10)\cdots P(X_5>10)$$

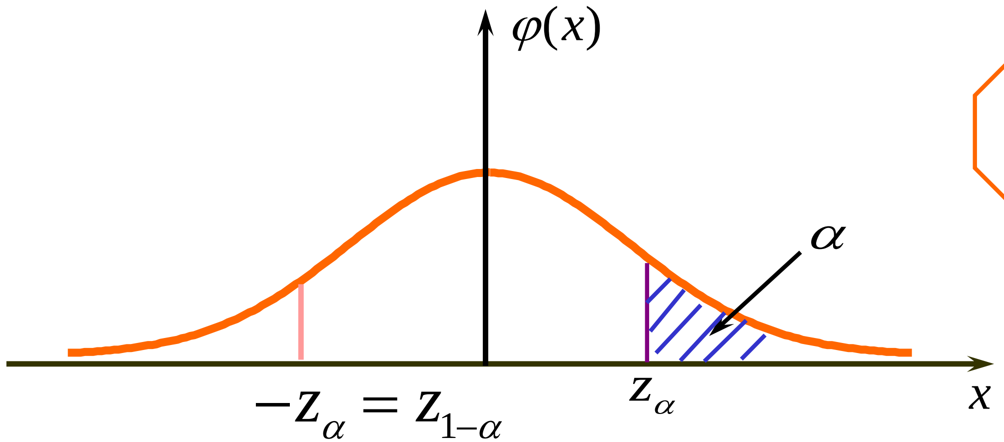
$$\overset{同分布}{=}1-[P(X>10)]^5$$

$$=1-[0.8413]^5=0.5785$$

标准正态分布的上侧α分位数(~~P127~~) P162

设$X \sim N(0,1)$,若z_α满足条件$P\{X > z_\alpha\} = \alpha$, $0<\alpha<1$

则称点 z_α 为标准正态分布的上侧 α 分位数。



$$\boxed{z_{1-\alpha} = -z_\alpha}$$

$$\Phi(z_\alpha) = P(X \le z_\alpha) = 1 - P(X > z_\alpha) = 1 - \alpha$$

例： 求 $z_{0.1} = ?$

$$\Phi(z_{0.1}) = 1 - 0.1 = 0.9 \quad \Rightarrow z_{0.1} \overset{\text{查表}}{=} 1.28$$

§ 2 常用的分布

χ^2分布(卡方分布)

定义：随机变量$X_1, X_2, \cdots, X_n$相互独立，而且

$$X_i \sim N(0,1) \qquad (i=1,2,\cdots,n)$$

则 $X = \sum_{i=1}^{n} X_i^2$

服从自由度为n的χ^2分布，记为 $X \sim \chi^2(n)$.

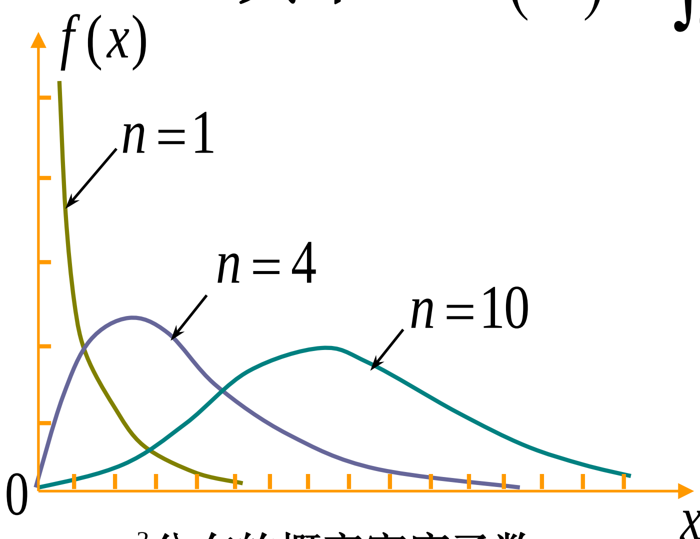
自由度指独立的标准正态分布的随机变量个数.

χ^2分布的概率密度

定理6.2.1: $\chi^2(n)$分布的概率密度为:

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

其中 $\Gamma(\alpha)=\int_0^{+\infty}t^{\alpha-1}e^{-t}dt$



χ^2分布的概率密度函数

$\Gamma(\alpha+1)=\alpha\,\Gamma(\alpha)$

$\Gamma(\alpha+1)=\alpha!$ (α为正整数)

$\Gamma(0.5)=\sqrt{\pi}$

χ^2分布的一些重要性质

(1) χ^2分布的可加性

设$Y_1 \sim \chi^2(n_1), Y_2 \sim \chi^2(n_2)$，且$Y_1, Y_2$相互独立，

则有 $Y_1 + Y_2 \sim \chi^2(n_1 + n_2)$

推广：

设$Y_i \sim \chi^2(n_i), i = 1, 2, \cdots m$，且$Y_1, Y_2, \cdots, Y_m$相互独立，

则 $Y_1 + Y_2 + \cdots + Y_m \sim \chi^2(n_1 + n_2 + \cdots + n_m)$

χ^2分布的一些重要性质

(2) χ^2分布的数学期望和方差

设$Y\sim\chi^2(n)$，则有 $E(Y)=n$, $Var(Y)=2n$

证：设$Y=X_1^2+X_2^2+\cdots+X_n^2$，$X_1,X_2,\cdots,X_n$均服从$N(0,1)$的独立随机变量

$$E(Y)=E(X_1^2+X_2^2+\cdots+X_n^2)=nE(X_1^2)=n[Var(X_1)+E^2(X_1)]=n$$

$$Var(Y)=Var(X_1^2+X_2^2+\cdots+X_n^2)=nVar(X_1^2)\ =n[E(X_1^4)-E^2(X_1^2)]$$

$$=n[3-1]=2n$$

$$E(X_1^4)=\int_{-\infty}^{+\infty}x^4\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx\ =\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}-x^3de^{-\frac{x^2}{2}}$$

$$=\frac{1}{\sqrt{2\pi}}[-x^3e^{-\frac{x^2}{2}}\Big|_{-\infty}^{+\infty}+\int_{-\infty}^{+\infty}e^{-\frac{x^2}{2}}3x^2dx]\ =\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}-3xde^{-\frac{x^2}{2}}$$

$$=\frac{1}{\sqrt{2\pi}}[-3xe^{-\frac{x^2}{2}}\Big|_{-\infty}^{+\infty}+3\int_{-\infty}^{+\infty}e^{-\frac{x^2}{2}}dx]\ =3\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=3$$

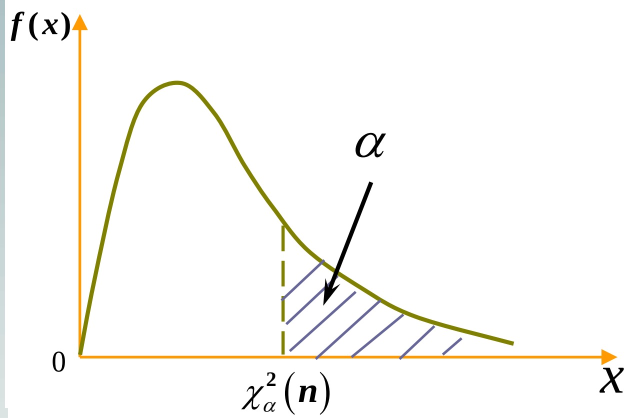
χ^2分布的上侧α分位数

$\chi^2(n)$－－分布记号

$\chi^2_\alpha(n)$－－分位数

对给定的概率α，$0<\alpha<1$，称满足条件$\int_{\chi^2_\alpha(n)}^{\infty} f(x)dx=\alpha$

的点$\chi^2_\alpha(n)$为$\chi^2(n)$分布的上侧α分位数。



χ^2分布的上α分位数

若$X\sim\chi^2(n)$, $P\left(X>\chi^2_\alpha(n)\right)=\alpha$,

上α分位数$\chi^2_\alpha(n)$的值可查χ^2分布表

~~P297~~ P366

$\chi^2_{0.1}(40)=$ 51.805

*当$n>40$时，有近似公式：

$$\chi^2_\alpha(n)\approx\frac{1}{2}\left(z_\alpha+\sqrt{2n-1}\right)^2$$

其中z_α为标准正态分布的上侧α分位数

$z_{0.1}=1.28$

当$n=40$时上式右边$=51.696$

若n足够大，$X\sim\chi^2(n)$,

由中心极限定理，

$$X\overset{\text{近似}}{\sim}N(n,2n)$$

例3：设总体$X\sim N(\mu,\sigma^2)$, μ,σ^2已知,$(X_1,X_2,\cdots X_n)$是取自总体X的样本，求

(1)统计量$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$的分布；

(2)设$n\ge 5$,若$a(X_1-X_2)^2+b(2X_3-X_4-X_5)^2\sim\chi^2(k)$,则$a,b,k$各位多少？

解：(1)作变换 $Y_i=\frac{X_i-\mu}{\sigma}\quad i=1,2,\cdots,n$

显然$Y_1,Y_2,\cdots,Y_n$相互独立，且$Y_i\sim N\left(0,1\right)\quad i=1,2,\cdots,n$

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

(2) $X_1-X_2\sim N(0,2\sigma^2)$, $\quad\left(\frac{X_1-X_2}{\sqrt{2}\sigma}\right)^2\sim\chi^2(1)$

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

X_1-X_2与$2X_3-X_4-X_5$相互独立，

故$\frac{(X_1-X_2)^2}{2\sigma^2}+\frac{(2X_3-X_4-X_5)^2}{6\sigma^2}\sim\chi^2(2)$

$a=\frac{1}{2\sigma^2}$,

$b=\frac{1}{6\sigma^2}$,

$k=2.$

*设总体$X \sim N(0,\sigma^2), X_1, X_2, \cdots X_n$为取自总体的一个样本。

(1) 问 $\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i}{\sigma}\right)^2$依概率收敛到什么?

(2) 记 $\dfrac{\sum_{i=1}^{n}\left(\frac{X_i}{\sigma}\right)^2 - n}{\sqrt{2n}}$ 的分布函数为$F_n(x)$,求$\lim\limits_{n\to\infty} F_n(1)$.

解:(1)收敛到$E\left(\frac{X_i}{\sigma}\right)^2 = 1$ $\quad \because \frac{X_i}{\sigma} \sim N(0,1)$

$$(2)\sum_{i=1}^{n}\left(\frac{X_i}{\sigma}\right)^2 \overset{\text{近似}}{\sim} N(n,2n), \quad \frac{\sum_{i=1}^{n}\left(\frac{X_i}{\sigma}\right)^2 - n}{\sqrt{2n}} \overset{\text{近似}}{\sim} N(0,1)$$

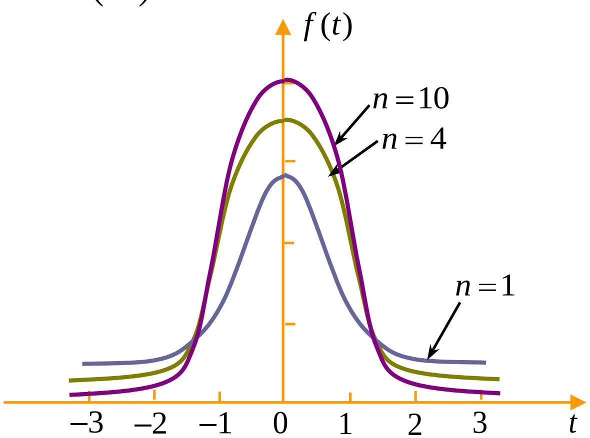
$$\lim_{n\to\infty} F_n(1) = \Phi(1) = 0.8413$$

$t-$分布

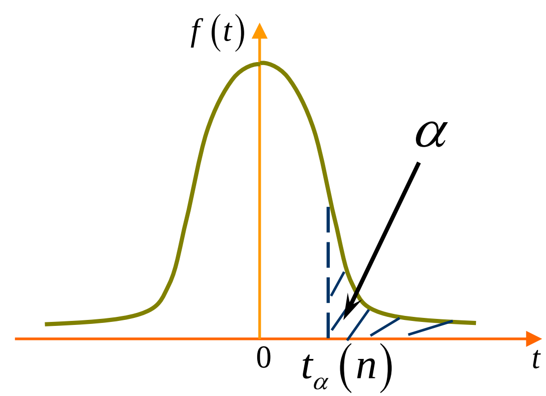
- 定义：设$X\sim N(0,1),Y\sim\chi^2(n)$,并且$X,Y$相互独立，则称随机变量$t=\dfrac{X}{\sqrt{Y/n}}$服从自由度为$n$的$t$分布，记为$t\sim t(n)$.

$\boxed{\text{定理6.2.2:}}$ $t(n)$分布的概率密度为：$f(t)=\dfrac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\dfrac{t^2}{n}\right)^{-\frac{n+1}{2}},-\infty<t<+\infty$

对给定的α，$0<\alpha<1$,称满足条件$P(t>t_\alpha(n))=\int_{t_\alpha(n)}^{\infty}f(t)dt=\alpha$的点$t_\alpha(n)$为$t(n)$分布的上侧$\alpha$分位数。$t$分布的上$\alpha$分位数可查$t$分布表（~~P296~~ P366）.

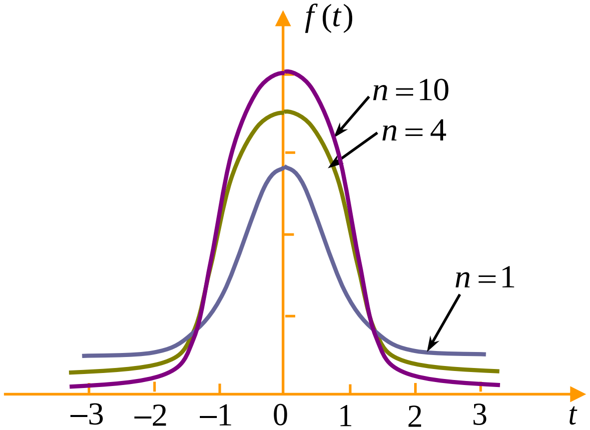


t 分布的密度函数

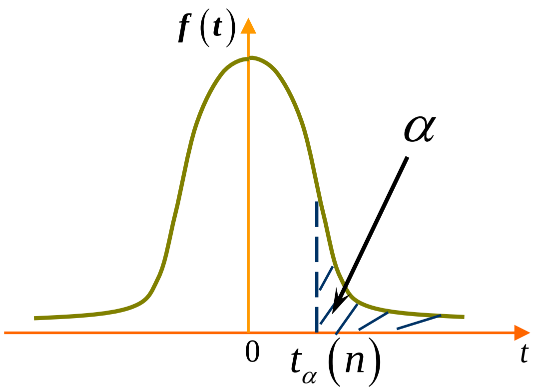


t 分布的上α分位数

t 分布的性质



t分布的密度函数



t分布的上α分位数

(1) $\lim\limits_{n\to\infty} f(t) = \dfrac{1}{\sqrt{2\pi}} e^{-t^2/2} = \varphi(t)$

(2) 若$X \sim t(n)$，则$E(X)=0$, $Var(X) = \dfrac{n}{n-2} > 1$ ， $n>2$

(3) $t_{1-\alpha}(n) = -t_{\alpha}(n)$

(4) 当$n>45$时，$t_{\alpha}(n) \approx z_{\alpha}$

若n充分大时， $X \sim t(n)$，也可认为： $X \overset{\text{近似}}{\sim} N(0,1)$

F 分布

- 定义：设$X\sim\chi^2(n_1),Y\sim\chi^2(n_2)$，且$X,Y$独立，则称随机变量

$$F=\frac{X/n_1}{Y/n_2}$$服从自由度(n_1,n_2)的F分布，记为$F\sim F(n_1,n_2)$

其中n_1称为第一自由度，n_2称为第二自由度.

$\boxed{\text{定理6.2.3:}}$ $F(n_1,n_2)$分布的概率密度为：

$$f(x)=\begin{cases}\dfrac{1}{B\left(\frac{n_1}{2},\frac{n_2}{2}\right)}n_1^{\frac{n_1}{2}}n_2^{\frac{n_2}{2}}x^{\frac{n_1}{2}-1}(n_2+n_1x)^{-\frac{n_1+n_2}{2}} & x>0\\ 0 & x\le 0\end{cases}$$

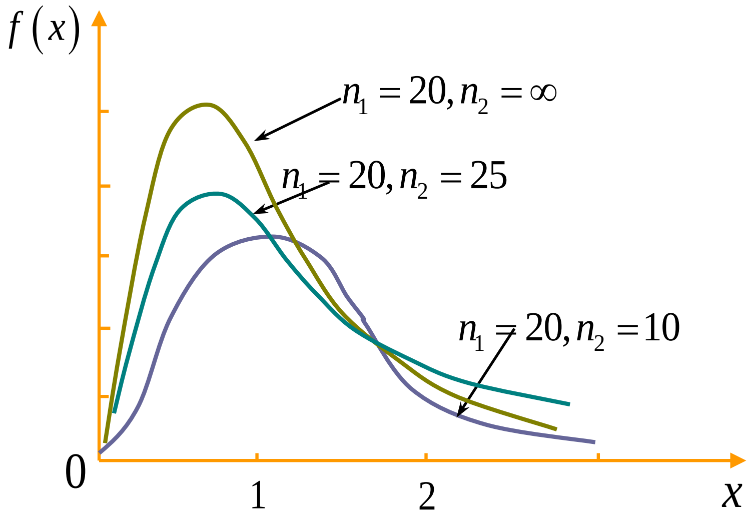
其中$B(a,b)=\int_0^1 x^{a-1}(1-x)^{b-1}dx=\dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

$$\Gamma(a)=\int_0^{+\infty}x^{a-1}e^{-x}dx$$

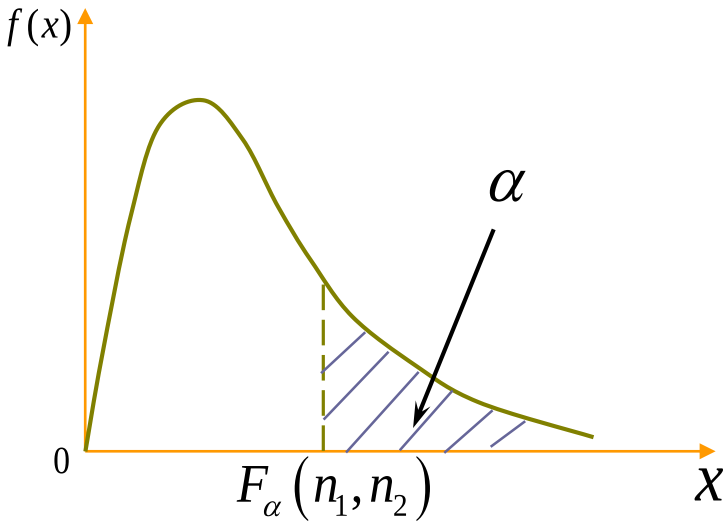
F分布的上α分位数

对于给定的α，$0<\alpha<1$，称满足条件$\int_{F_\alpha(n_1,n_2)}^{\infty} f(x)dx=\alpha$的点$F_\alpha(n_1,n_2)$为$F(n_1,n_2)$分布的上$\alpha$分位数。$F_\alpha(n_1,n_2)$的值可查$F$分布表(~~P298-302~~).

P367 - 371



F分布的密度函数



F分布的上α分位数

F分布的性质

（1）若$F \sim F(n_1, n_2)$，则 $\dfrac{1}{F} \sim F(n_2, n_1)$

（2）若$t \sim t(n)$，则 $t^2 \sim F(1, n)$

*（3）$\left(t_{\alpha/2}(n)\right)^2 = F_\alpha(1, n)$

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

F分布分位数表中，下标（概率值）只有

0.1，0.05，0.025，0.01，0.005

证：（2）设$t=\dfrac{X}{\sqrt{Y/n}}\sim t(n)$

其中$X\sim N(0,1)$，$Y\sim\chi^2(n)$，并且两者独立

$$\therefore\quad t^2=\frac{X^2}{Y/n}=\frac{X^2/1}{Y/n}\sim F(1,n)$$

（3）设$t\sim t(n)$，则$t^2\sim F(1,n)$

$$P(\{t^2>(t_{\alpha/2}(n))^2\})$$

$$=P(\{t>t_{\alpha/2}(n)\}\cup\{t<-t_{\alpha/2}(n)\})=\frac{\alpha}{2}+\frac{\alpha}{2}=\alpha$$

$$\therefore\left(t_{\alpha/2}(n)\right)^2=F_\alpha(1,n)$$

(4)设$F \sim F(n_1,n_2)$, 则$\dfrac{1}{F} \sim F(n_2,n_1)$

$$1-\alpha = P(F > F_{1-\alpha}(n_1,n_2)) = P\left(\frac{1}{F} < \frac{1}{F_{1-\alpha}(n_1,n_2)}\right)$$

$$= 1 - P\left(\frac{1}{F} \ge \frac{1}{F_{1-\alpha}(n_1,n_2)}\right) \quad \text{即} P\left(\frac{1}{F} \ge \frac{1}{F_{1-\alpha}(n_1,n_2)}\right) = \alpha$$

$$\therefore F_\alpha(n_2,n_1) = \frac{1}{F_{1-\alpha}(n_1,n_2)} \quad \text{也即} F_{1-\alpha}(n_1,n_2) = \frac{1}{F_\alpha(n_2,n_1)}$$

例：$F_{0.99}(20,10) = \dfrac{1}{F_{0.01}(10,20)} = \dfrac{1}{3.37}$

例4：若$F \sim F(5,10)$，求λ值使 $P(F<\lambda)=0.05$

解：注意到F分布的分位数表中，只有

$$\alpha = 0.1,\quad 0.05,\quad 0.025,\quad 0.01,\quad 0.005$$

所以用 $P(F \geq \lambda)=1-P(F<\lambda)=0.95$ 不能求得λ。

$$\because\ P(F<\lambda)=P\left(\frac{1}{F}>\frac{1}{\lambda}\right)=0.05$$

注意到 $\dfrac{1}{F} \sim F(10,5)$

$$\therefore\ \frac{1}{\lambda}=F_{0.05}(10,5)=4.74$$

$$\Rightarrow\ \lambda=0.21$$

§ 3　正态总体下的抽样分布

定理6.3.1: 设$(X_1, X_2, \cdots, X_n)$是总体$N\left(\mu, \sigma^2\right)$的样本，$\bar{X}, S^2$分别是样本均值和样本方差，则有：

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \quad ，也即 \quad \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

证明　$\because \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$，

即$\bar{X}$是n个独立的服从正态分布随机变量的线性组合,

$\therefore \bar{X}$也是服从正态分布的。两个参数是:

$$E(\bar{X}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \mu \qquad Var(\bar{X}) = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{\sigma^2}{n}$$

$\boxed{\text{定理6.3.2:}}$ 设$(X_1,\cdots,X_n)$是总体$N\left(\mu,\sigma^2\right)$的样本，$\bar{X}$和$S^2$分别是样本均值和样本方差，则有：

(1) $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$　（2）$\bar{X}$和S^2相互独立

［分析］ $$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2 \sim \chi^2(n-1)$$

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

$\because \sum_{i=1}^{n}\left(\dfrac{X_i-\bar{X}}{\sigma}\right)=0$ 或 $\bar{X}=\dfrac{1}{n}\sum_{i=1}^{n}X_i$，有一个约束条件。

$\boxed{\text{定理6.3.3:}}$ 设$(X_1,\cdots,X_n)$是总体$N(\mu,\sigma^2)$的样本，$\bar{X}$和S^2

分别是样本均值和样本方差，则有：

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

证明：
$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \triangleq U \sim N(0,1), \quad \frac{(n-1)S^2}{\sigma^2} \triangleq V \sim \chi^2(n-1),$$

且两者独立，由t分布定义得：

$$\frac{U}{\sqrt{V/(n-1)}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \Bigg/ \sqrt{\frac{(n-1)S^2}{\sigma^2}\Big/(n-1)}$$

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

定理6.3.4: 设样本$\left(X_1,\cdots,X_{n_1}\right)$和$\left(Y_1,\cdots,Y_{n_2}\right)$分别来自总体$N\left(\mu_1,\sigma_1{}^2\right)$和$N\left(\mu_2,\sigma_2{}^2\right)$并且它们相互独立，其样本方差分别为$S_1^2,S_2^2$,则：

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

$$(2)\ \frac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim N(0,1),$$

(3)当$\sigma_1{}^2=\sigma_2{}^2=\sigma^2$时，$\dfrac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_1-\mu_2\right)}{S_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim t\left(n_1+n_2-2\right)$

其中$S_w^2=\dfrac{\left(n_1-1\right)S_1^2+\left(n_2-1\right)S_2^2}{n_1+n_2-2},S_w=\sqrt{S_w^2}$

证明：(1) $\dfrac{(n_1-1)S_1^2}{\sigma_1^{\ 2}} \sim \chi^2(n_1-1)$，$\dfrac{(n_2-1)S_2^2}{\sigma_2^{\ 2}} \sim \chi^2(n_2-1)$

且两者独立，由F分布的定义，有：

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

(2) $\bar{X} \sim N(\mu_1, \dfrac{\sigma_1^2}{n_1}), \bar{Y} \sim N(\mu_2, \dfrac{\sigma_2^2}{n_2})$，且$\bar{X}$与$\bar{Y}$相互独立，

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

即 $$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}} \sim N(0,1)$$

(3) 当$\sigma_1^2=\sigma_2^2=\sigma^2$时，由（2）得

$$U=\frac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_1-\mu_2\right)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim N\left(0,1\right)$$

又由前定理知：

$$\frac{\left(n_1-1\right)S_1^2}{\sigma^2}\sim\chi^2\left(n_1-1\right),\ \frac{\left(n_2-1\right)S_2^2}{\sigma^2}\sim\chi^2\left(n_2-1\right)$$

且它们相互独立,故有χ^2分布的可加性：

$$V=\frac{\left(n_1-1\right)S_1^2}{\sigma^2}+\frac{\left(n_2-1\right)S_2^2}{\sigma^2}\sim\chi^2\left(n_1+n_2-2\right)$$

且U与V相互独立，于是根据t分布：

$$\frac{U}{\sqrt{V/\left(n_1+n_2-2\right)}}=\frac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_1-\mu_2\right)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t\left(n_1+n_2-2\right)$$

例6：设$X_1, X_2, \ldots X_{10}$是总体$N(0,0.3^2)$的样本，求$P(\sum_{i=1}^{10} X_i^2 > 1.44)$

解：
$$P(\sum_{i=1}^{10} X_i^2 > 1.44) = P(\sum_{i=1}^{10}\left(\frac{X_i - 0}{0.3}\right)^2 > \frac{1.44}{0.3^2})$$

$$\overset{Y\sim\chi^2(10)}{====} P(Y > 16) \quad \approx 0.10$$

$$\chi^2_{0.05}(9) = 16.919 \qquad \chi^2_{0.10}(9) = 14.684$$

$$\chi^2_{0.05}(10) = 18.307 \qquad \chi^2_{0.10}(10) = 15.987$$

例7：设$X \sim N(\mu,\sigma^2),(X_1,X_2,\cdots,X_n)$是一样本，求$E(S^2),Var(S^2),Var(\overline{X}S^2)$

解：$\because \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

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

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

$$=\left(\frac{\sigma^2}{n-1}\right)^2 2(n-1) \ =\frac{2}{n-1}\sigma^4$$

$$Var(\overline{X}S^2)=E(\overline{X}^2S^4)-\left[E(\overline{X}S^2)\right]^2=E(\overline{X}^2)E(S^4)-\left[E(\overline{X})E(S^2)\right]^2$$

$$=\left[E^2(\overline{X})+Var(\overline{X})\right]\left[E^2(S^2)+Var(S^2)\right]-E^2(\overline{X})E^2(S^2)$$

$$=\left[\mu^2+\frac{\sigma^2}{n}\right]\left[\sigma^4+\frac{2}{n-1}\sigma^4\right]-\mu^2\sigma^4$$

思考：$Var\left(n(\overline{X}-\mu)^2-\sum_{i=1}^{n}(X_i-\overline{X})^2\right)=?$ 答案:$2n\sigma^4$

例8：设总体$X \sim N\left(\mu,\sigma^2\right)$，$(X_1,\cdots,X_4)$与$(Y_1,\cdots,Y_9)$是取自总体$X$的两个独立样本，$\bar{X},S_1^2$和$\bar{Y},S_2^2$分别为样本均值和样本方差；

(1) 证明：$\dfrac{\bar{X}-\mu}{S_2/\sqrt{4}} \sim t(8)$　　(2) $a\dfrac{\bar{X}-\bar{Y}}{S_1} \sim t(k)$，则$a,k$各为多少？

(3) $b\sum_{i=1}^{4}(X_i-\mu)^2 \Big/ S_2^2 \sim F(n_1,n_2)$，则$b,n_1,n_2$各为多少？

---

(1) 证：令$U=\dfrac{\bar{X}-\mu}{\sigma/\sqrt{4}} \sim N(0,1)$，　$V=\dfrac{(9-1)S_2^2}{\sigma^2} \sim \chi^2(8)$

U,V两者独立！

$$\therefore \frac{U}{\sqrt{V/8}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{4}} \Bigg/ \sqrt{\frac{(9-1)S_2^2}{\sigma^2}/8} = \frac{\bar{X}-\mu}{S_2/\sqrt{4}} \sim t(8)$$

注意 $\dfrac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$ 成立的条件！

例8：设总体$X \sim N\left(\mu,\sigma^2\right)$，$\left(X_1,\cdots,X_4\right)$与$\left(Y_1,\cdots,Y_9\right)$是取自总体$X$的两个独立样本，$\bar{X},S_1^2$和$\bar{Y},S_2^2$分别为样本均值和样本方差；

(1) 证明：$\dfrac{\bar{X}-\mu}{S_2/\sqrt{4}} \sim t(8)$　　(2) $a\dfrac{\bar{X}-\bar{Y}}{S_1} \sim t(k)$，则$a,k$各为多少？

(3) $b\sum_{i=1}^{4}(X_i-\mu)^2 \Big/ S_2^2 \sim F(n_1,n_2)$，则$b,n_1,n_2$各为多少？

---

(2) $\because \bar{X} \sim N(\mu,\dfrac{\sigma^2}{4}),\ \bar{Y} \sim N(\mu,\dfrac{\sigma^2}{9})$，且$\bar{X}$与$\bar{Y}$相互独立，

$$\therefore \bar{X}-\bar{Y} \sim N(0,\frac{13\sigma^2}{36}),\qquad U=\frac{\bar{X}-\bar{Y}}{\sigma\sqrt{13/6}} \sim N(0,1)$$

$V=\dfrac{3S_1^2}{\sigma^2} \sim \chi^2(3)$，　$\because \bar{X}-\bar{Y}$与S_1^2相互独立，$\therefore U$与V独立！

$$\frac{U}{\sqrt{V/3}}=\frac{\dfrac{\bar{X}-\bar{Y}}{\sigma\sqrt{13/6}}}{\sqrt{\dfrac{3S_1^2}{\sigma^2}/3}}=\frac{6}{\sqrt{13}}\frac{\bar{X}-\bar{Y}}{S_1} \sim t(3) \Rightarrow a=\frac{6}{\sqrt{13}},\quad k=3$$

例8：设总体$X \sim N\left(\mu, \sigma^2\right)$，$(X_1, \cdots, X_4)$与$(Y_1, \cdots, Y_9)$是取自总体$X$的两个独立样本，$\bar{X}, S_1^2$和$\bar{Y}, S_2^2$分别为样本均值和样本方差；

(1) 证明：$\dfrac{\bar{X}-\mu}{S_2/\sqrt{4}} \sim t(8)$　　(2) $a\dfrac{\bar{X}-\bar{Y}}{S_1} \sim t(k)$，则$a, k$各为多少？

(3) $b\sum_{i=1}^{4}(X_i-\mu)^2 \Big/ S_2^2 \sim F(n_1, n_2)$，则$b, n_1, n_2$各为多少？

---

(3) $\sum_{i=1}^{4}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(4)$，$\frac{8S_2^2}{\sigma^2} \sim \chi^2(8)$，且$\sum_{i=1}^{4}(X_i-\mu)^2$与$S_2^2$独立

$$\Rightarrow \sum_{i=1}^{4}\left(\frac{X_i-\mu}{\sigma}\right)^2/4 \Bigg/ \frac{8S_2^2}{\sigma^2}/8 = \frac{1}{4}\sum_{i=1}^{4}(X_i-\mu)^2 \Big/ S_2^2 \sim F(4,8),$$

$$\Rightarrow b = \frac{1}{4}, (n_1, n_2) = (4,8).$$

思考：$b\sum_{i=1}^{4}(X_i-\bar{X})^2 \Big/ S_2^2 \sim F(n_1, n_2)$，则$b, n_1, n_2$各为多少？

答案:$(b, n_1, n_2) = (1/3, 3, 8)$

例9：从总体$X \sim N(3,\sigma^2)$中取$n=10$的一个样本，已求得样本方差为4，求样本均值落在2.1253到3.8747间的概率。

解：虽然由正态分布的性质知道，样本均值也服从正态分布，但由于总体方差未知，故不能用下面正态分布方法求：

$$\overline{X} \sim N(3,\sigma^2/10)$$

$$P(2.1253<\overline{X}<3.8747)=\Phi\left(\frac{3.8747-3}{\sigma/\sqrt{10}}\right)-\Phi\left(\frac{2.1253-3}{\sigma/\sqrt{10}}\right)=?$$

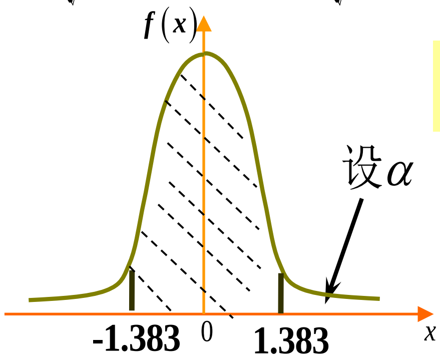
但注意到：$\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$，　此处即：$\dfrac{\overline{X}-3}{S/\sqrt{10}} \overset{\text{记为}}{=} t \sim t(9)$

$$P(2.1253<\overline{X}<3.8747)=P\left(\frac{2.1253-3}{2/\sqrt{10}}<\frac{\overline{X}-3}{S/\sqrt{10}}<\frac{3.8747-3}{2/\sqrt{10}}\right)$$

$$=P(-1.383<t\quad<1.383)$$

$$=1-2\alpha=1-2*0.1=0.8$$

查$t_\alpha(9)=1.383$



若$\sigma=2$，此概率$=2\Phi(1.38)-1=0.8324$

例10： 总体$X \sim N(\mu,4)$, μ未知，$X_1, X_2, \cdots, X_n$是一样本，问当至少多大时才能使$P(|\overline{X}-\mu| \le 0.1) \ge 0.95$ ？

解： $\because \ \overline{X} \sim N(\mu, \frac{4}{n})$

$$\therefore P(\left|\overline{X}-\mu\right| \le 0.1) = P(\mu - 0.1 < \overline{X} < \mu + 0.1)$$

$$= \Phi\left(\frac{\mu+0.1-\mu}{2/\sqrt{n}}\right) - \Phi\left(\frac{\mu-0.1-\mu}{2/\sqrt{n}}\right)$$

$$= \Phi\left(\frac{0.1}{2/\sqrt{n}}\right) - \Phi\left(\frac{-0.1}{2/\sqrt{n}}\right) = 2\Phi\left(\frac{0.1}{2/\sqrt{n}}\right) - 1 \ge 0.95$$

与μ无关

$$\therefore \quad \Phi\left(\frac{0.1}{2/\sqrt{n}}\right) \ge 0.975 \qquad \Rightarrow \quad \frac{0.1}{2/\sqrt{n}} \ge 1.96$$

$$\Rightarrow \ n \ge 1536.6$$

$$n = 1537$$

例11：设总体$X\sim N(6,\sigma_1^2)$，$Y\sim N(5,\sigma_2^2)$，有$n_1=n_2=10$的独立样本，若

（1）已知$\sigma_1^2=\sigma_2^2=1$ （2）σ_1^2，σ_2^2未知，但两者相同，样本方差分别为0.9130，0.9816，求$P(\bar{X}-\bar{Y}<1.3)$。

解：(1)由定理得：$\bar{X}\sim N(6,\dfrac{1}{10})$，同理得：$\bar{Y}\sim N(5,\dfrac{1}{10})$

$\bar{X}-\bar{Y}$是两个独立的服从正态分布的变量线性组合，则为正态的。

$$E(\bar{X}-\bar{Y})=E(\bar{X})-E(\bar{Y})=1$$

$$Var(\bar{X}-\bar{Y})=Var(\bar{X})+Var(\bar{Y})=\frac{1}{5}$$

$$\therefore\ \bar{X}-\bar{Y}\sim N(1,\frac{1}{5})$$

$$P(\bar{X}-\bar{Y}<1.3)=\Phi\left(\frac{1.3-1}{\sqrt{1/5}}\right)=\Phi\left(0.671\right)=0.7486$$

例11： 设总体$X \sim N(6,\sigma_1^2)$, $Y \sim N(5,\sigma_2^2)$, 有$n_1 = n_2 = 10$的独立样本，若

（1）已知$\sigma_1^2 = \sigma_2^2 = 1$ （2）σ_1^2，σ_2^2未知，但两者相同，样本方差分别为0.9130，0.9816，求$P(\bar{X} - \bar{Y} < 1.3)$。

解：（2）若$\sigma_1^2 = \sigma_2^2 = \sigma^2$但未知，则$\bar{X} - \bar{Y} \sim N(1, \dfrac{\sigma^2}{5})$

$$P(\bar{X} - \bar{Y} < 1.3) = \Phi\left(\frac{1.3-1}{\sqrt{\sigma^2/5}}\right) = \Phi\left(\frac{0.3}{\sigma}\sqrt{5}\right)$$ 因σ未知而无法求得！

由定理，注意到$t = \dfrac{\bar{X} - \bar{Y} - (6-5)}{S_w\sqrt{1/10 + 1/10}} \sim t(10+10-2)$

$$s_w = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{(10-1)0.9130 + (10-1)0.9816}{10+10-2}} = 0.9733$$

$$P(\bar{X} - \bar{Y} < 1.3) = P(t < \frac{1.3-1}{0.9733\sqrt{1/10+1/10}}) = P(t < 0.6884)$$

$$= 1 - P(t \ge 0.6884) \overset{excel}{=} 1 - TDIST(0.6884, 18, 1) \approx 0.75$$

查表

$$0.6884 = t_\alpha(18) \Rightarrow \alpha > 0.2$$

例12：从总体$X \sim N(\mu,3), Y \sim N(\mu,5)$中分别抽取$n_1=10, n_2=15$的独立样本，求两个样本方差之比大于1.272的概率。

解：

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

本题中有，$$\frac{5S_1^2}{3S_2^2}\sim F(9,14)$$

$$P(\frac{S_1^2}{S_2^2}>1.272)=P(\frac{5S_1^2}{3S_2^2}>\frac{5}{3}*1.272)=P(\frac{5S_1^2}{3S_2^2}>2.12)\overset{\text{查表}}{=}0.1$$

$$\because F_{0.1}(9,14)=2.12$$

$$=F.DIST.RT(2.12,9,14)=0.100267382170487$$

设总体 $X \sim N(\mu,\sigma^2)$，$(X_1, X_2,\cdots X_n)$ 为样本，求下列分布

$$\overline{X} \sim N(\mu,\frac{\sigma^2}{n})$$

$$\overline{X}-\mu \sim N(0,\frac{\sigma^2}{n}) \qquad \overline{X}+\mu \sim N(2\mu,\frac{\sigma^2}{n})$$

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

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

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

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

$$\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2 \sim \chi^2(1)$$

$$\left(\frac{\overline{X}-\mu}{S/\sqrt{n}}\right)^2 \sim F(1,n-1)$$

$$\left(\frac{X_1-\mu}{X_2-\mu}\right)^2 \sim F(1,1)$$

$$\frac{S^2}{n(\overline{X}-\mu)^2} = \left(\frac{S/\sqrt{n}}{\overline{X}-\mu}\right)^2 \sim F(n-1,1)$$

复习思考题 6

1. 什么叫总体？什么叫简单随机样本？总体X的样本$X_1,X_2,\ldots,X_n$有哪两个主要性质？

2. 什么是统计量？什么是统计量的值？

3. 样本均值和样本方差如何计算？

4. $N(0,1)$分布，t分布，χ^2分布和F分布的双侧、下侧、上侧分位点是如何定义的？怎样利用附表查这些分位点的值？

5. 对一个正态总体的三个常用统计量及其分布是什么？

6. 对两个正态总体的三个常用统计量及其分布是什么？