Ch 6. 数理统计的基本概念

钟友良

zhongyl@scut.edu.cn

Outline

6.0 前言

6.1 总体、样本和统计量

6.3 抽样分布

Outline

6.0 前言

6.1 总体、样本和统计量

6.3 抽样分布

Outline

6.0 前言

6.1 总体、样本和统计量

6.3 抽样分布

概率与统计

概率论和统计的区别:

- 概率论: 已知随机变量 X 的分布 (分布列或者分布密度函数), 进而根据分布进行事件概率和矩计算
- 统计: 总体的分布未知或者不完全已知的情形下, 通过抽样并研究抽样得到数据的统计特征, 并估计总体的分布。

问题提出

对某个学校的小学生身高 X 进行调查，确定其分布函数？ X 为待研究的 总体 ， 构成总体的每个成员被称为一个 个体 。

研究方法

对总体进行 简单随机抽样 ， 并根据抽样结果推断整体。

抽样 ：从 X 中随机的抽出 n 个个体，记为 n 维样本 。由于抽取的个体是随机的，即无法提前预知其数值，样本为 n 维随机向量 $(X_1, X_2, \cdots, X_n)$。

抛硬币 $X \sim B(1, p)$

ith . . . : $X_i \sim B(1, p)$

n次 . . . : $(X_1, \dots, X_n)$

简单随机抽样

简单随机抽样:

- 样本与总体 同分布
- $X_1, X_2, \cdots, X_n$ 相互独立

总体与样本

- 总体: 总体是一个分布待确定的随机变量 X。总体中包含很多个 个体 。
- 样本: 从总体中随机抽取的 n 个个体，为 n 维随机向量 $(X_1, X_2, \cdots, X_n)$。对样本进行一次观测，记 $(x_1, x_2, \cdots, x_n)$ 为样本的一个 观测值 。

分布函数

命题 6.1.1:
设总体 $X \sim F_X(\cdot, \theta)$， $(X_1, X_2, \cdots, X_n)$ 为其样本，
那么

$$
\begin{aligned}
& F_{X_1, X_2, \cdots, X_n}(x_1, x_2, \cdots, x_n; \theta) \\
= & F_X(x_1, \theta) F_X(x_2, \theta) \cdots F_X(x_n, \theta)
\end{aligned}
$$

经验分布函数

问题：由样本估计总体的分布函数。

经验分布函数：设总体的样本 $(X_1, X_2, \cdots, X_n)$ 的一次观测值为 $(x_1, x_2, \cdots, x_n)$，并将其进行升序排列 $x_{(1)} \le x_{(2)} \le \cdots \le x_{(\overset{n}{10})}$，并定义

$<$

$$F_n^X(x) = \begin{cases} 0, & x < x_{(1)} \\ \cdots \quad \frac{1}{n} & x_{(1)} \le x < x_{(2)} \\ \dfrac{k}{n}, & x \in [x_{(k)}, x_{(k+1)}) \\ \cdots & \\ 1, & x \ge x_{(n)} \end{cases}$$

称 F_n^X 为总体 X 的一个经验分布函数，或样本分布函数。

经验分布函数的例子

- 抛硬币
- 投骰子

正 $=0$，反 $=1$.

$X \sim B(1,p)$

$X_1,\dots,X_5 \sim B(1,p)$

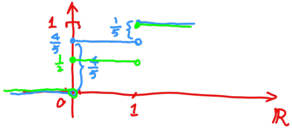
$(0,1,0,0,0)$

$X \sim U[0,1]$

$x_{(1)} \le x_{(2)} \le x_{(3)} \le x_{(4)} \le x_{(5)}$

$0, \quad 0, \quad 0, \quad 0, \quad 1$

$$F_5(x) = \begin{cases} 0, & x < 0 \\ \frac{4}{5}, & 0 \le x < 1 \\ 1, & 1 \le x \end{cases}$$



$$P(X_5=0) = \lim_{x\to 0^+} F_5(x) - \lim_{x\to 0^-} F_5(x) = \frac{4}{5} - 0 = \frac{4}{5}$$

$$P(X_5=1) = \lim_{x\to 1^+} F_5(x) - \lim_{x\to 1^-} F_5(x) = 1 - \frac{4}{5} = \frac{1}{5}$$

Glivenko 定理

设 $X_1, X_2, \cdots, X_n$ 是取自总体 X 的容量为 n 的样本，总体的分布函数为 $F_X(x)$，$F_n(x)$ 为其经验分布函数，那么

$$P\left(\lim_{n\to\infty} sup_{-\infty<x<\infty}\left|F_n^X(x) - F_X(x)\right| = 0\right) = 1$$

$(X_1, \ldots, X_n)$

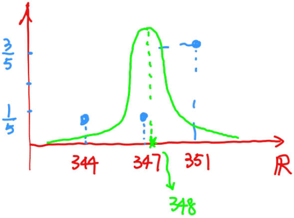
上面的结果说明，就是当样本容量足够大的时候，用经验分布函数来推断总体的分布函数是合理的。

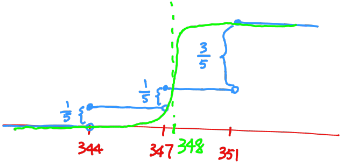
练习

某食品厂生产饮料，现在从生产线 (X) 上任意抽取 5 瓶，称得重量依次为 (g)

$$351 \quad 347 \quad 351 \quad 344 \quad 351$$

请写出总体 X 的经验分布函数





统计量

注意下面这个关系

$$\text{样本} \xrightarrow{\text{统计量}} \text{总体}$$

统计量是通过样本研究总体的手段
统计量是样本的函数，不含有任何未知参数，即
统计量由样本唯一决定。

常用统计量

- 样本均值:

$= E[X] = \sum_i X_i \cdot p_i$

$$\overline{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- 样本方差:

$\mathrm{Var}X = E[(X - EX)^2]$

$= \sum_i (X_i - EX)^2 \times p_i$

$p_i = \frac{1}{n}$

$EX = \overline{X}$

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

常用统计量

- 修正样本方差

$$S_n^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{n}{n-1}S_n^2$$

- 样本的 k 阶原点矩

$$\overline{X^k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

- 样本的 k 阶中心矩

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k$$

常用统计量

- 顺序统计量 $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$，其中 $X_{(1)} = \min\{X_1, X_2, \cdots, X_n\}$, $X_{(n)} = \max\{X_1, X_2, \cdots, X_n\}$，而 $X_{(k)}$ 是将 $X_1, X_2, \cdots, X_n$ 的取值从小到大排列的第 k 位的值。
- 样本中位数

$$\tilde{X} = \begin{cases} X_{\left(\frac{n+1}{2}\right)}, & \text{若}n\text{为奇数} \\ \frac{1}{2}\left(X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}\right), & \text{若}n\text{为偶数} \end{cases}$$

- 样本极差

$$R_n^X = X_{(n)} - X_{(1)}$$

Outline

6.0 前言

6.1 总体、样本和统计量

6.3 抽样分布

总体

$X \sim P$

个体

$X_1, \ldots, X_n \overset{i.i.d.}{\sim} P$

样本 $(X_1, \ldots, X_n)$

概率 已知 $\longrightarrow$ 未知

统计 未知 $\longleftarrow$ 已知

数字特征.

$E[X]$, $Var[X]$ ...

k阶原点矩

k阶中心矩.

统计量：由样本唯一确定

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

$S_n^2 := \frac{1}{n}\sum (X_i - \bar{X})^2$

$(E[S_n^2] = \frac{n-1}{n} Var[X])$

修正 $S_n^{*2} := \frac{n}{n-1} S_n^2$

$(E[S_n^{*2}] = Var[X])$

顺序 $X_{(1)} \le \ldots \le X_{(k)} \le \ldots \le X_{(n)}$

抽样分布

设计合适的统计量可以用来研究总体的数字特征和参数。

命题 6.3.1: 设 $(X_1, X_2, \cdots, X_n)$ 是来自总体 X 的样本，$E[X] = \mu$，$Var[X] = \sigma^2$，那么

- $$\boldsymbol{E[\overline{X}] = \mu, \quad Var[\overline{X}] = \frac{\sigma^2}{n}}$$

  $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

- $$\boldsymbol{E[S_n^2] = \frac{n-1}{n}\sigma^2, \quad E[S_n^{*2}] = \sigma^2}$$

  $< \sigma^2$

  $S_n^2 = \frac{1}{n}\sum_i (X_i - \overline{X})^2$

  $E[(X - EX)^2] = VarX = \sigma^2$

  $S_n^{*2} = \frac{n}{n-1} S_n^2$

  $= \frac{1}{n-1}\sum (X_i - \overline{X})^2$

抽样分布

证明:

$$E[\overline{X}] = E[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu$$

再由 X_i 之间的独立性有

$$Var[\overline{X}] = Var[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n^2}\sum_{i=1}^{n} Var[X_i]$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 = \frac{\sigma^2}{n}$$

抽样分布

对于 S_n^2

$$
\begin{aligned}
S_n^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i^2-2\overline{X}X_i+\overline{X}^2) \\
&= \frac{1}{n}\left(\sum_{i=1}^{n}X_i^2-2\overline{X}\sum_{i=1}^{n}X_i+n\overline{X}^2\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^2
\end{aligned}
$$

抽样分布

所以

$$E[S_n^2] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i^2] - E[\overline{X}^2]$$

$E[X_i^2] - \underline{E[X_i]^2 + E[X_i]^2}$

$$= \frac{1}{n}\sum_{i=1}^{n}(Var[X_i] + (E[X_i])^2) - \underline{(Var[\overline{X}] + (E[\overline{X}])^2)}$$

$Var\left[\frac{\Sigma X_i}{n}\right] = \frac{1}{n^2}\Sigma Var[X_i]$

$= \frac{1}{n^2}\cdot n \cdot \sigma^2 = \frac{1}{n}\sigma^2$

$$= \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2$$

$$= \frac{n-1}{n}\sigma^2$$

$$E[S_n^{*2}] = E\left[\frac{n}{n-1}S_n^2\right] = \frac{n}{n-1}E[S_n^2] = \sigma^2$$

χ^2 分布

chi-square

[tʃai]

命题 6.3.2:
设总体 $X \sim N(0,1)$，$(X_1, X_2, \cdots, X_n)$ 为其简单随机样本，则

$$\boldsymbol{X_1^2 + X_2^2 + \cdots + X_n^2 \sim \chi^2(n)}$$

其中

$$\chi^2(n) = \Gamma(\frac{n}{2}, \frac{1}{2}).$$

例 2.4.2

Recall $X \sim N(0,1)$

$\Rightarrow Y = X^2 \sim \Gamma(\frac{1}{2}, \frac{1}{2})$

$$\Rightarrow f_Y(y) = \begin{cases} \dfrac{(\frac{1}{2})^{\frac{1}{2}}}{\Gamma(\frac{1}{2})} \cdot y^{\frac{1}{2}-1} \cdot e^{-\frac{1}{2}y}, & y>0 \\ 0 & y \le 0 \end{cases}$$

一般地 $\Gamma(\alpha, \lambda)$ 的密度为

$$\begin{cases} \dfrac{\lambda^\alpha}{\Gamma(\alpha)} \cdot y^{\alpha-1} \cdot e^{-\lambda y}, & y>0 \\ 0 & y \le 0 \end{cases} \xrightarrow{(\alpha,\lambda)=(\frac{n}{2},\frac{1}{2})} \begin{cases} \dfrac{(\frac{1}{2})^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} \cdot y^{\frac{n}{2}-1} \cdot e^{-\frac{1}{2}y} & \cdots \\ \cdots & \cdots \end{cases}$$

$\dfrac{(\frac{1}{2})^n}{\cdots} \cdots\cdots$

χ^2 分布 chi-square distribution 卡方分布.

若 $X \sim \Gamma(\frac{n}{2}, \frac{1}{2})$，则称 X 服从自由度为 n 的 χ^2 分布，记为 $X \sim \chi^2(n)$，其分布密度函数为

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

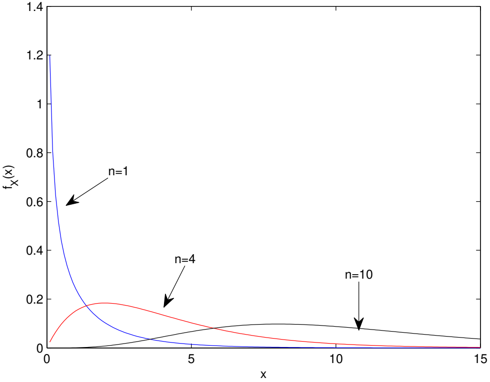
$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt$$

χ^2 分布图像

根据例题 2.4.2，当 $X \sim N(0,1)$，

$$X^2 \sim \chi^2(1)$$

即 X^2 服从自由度为 1 的 χ^2 分布。



χ^2 分布的性质

$X=\sum_{i=1}^{n} X_i^2 \quad \bar{X}=\frac{1}{n}\sum_{i=1}^{n} X_i^2$

$E[\bar{X}]=1, \quad Var[\bar{X}]=\frac{2}{n}$

- 若 $X \sim \chi^2(n)$，那么 $E[X]=n$， $Var[X]=2n$。
- 若 $X_1 \sim \chi^2(n_1)$， $X_2 \sim \chi^2(n_2)$， 且 X_1 与 X_2 独立，那么 $X_1+X_2 \sim \chi^2(n_1+n_2)$。
- 若 $X \sim \chi^2(n)$，那么 $\frac{X-n}{\sqrt{2n}}$ 近似服从 $N(0,1)$，当 $n \to \infty$

① 记 $X_i \sim N(0,1)$

$X=\sum X_i^2$

则 $E[X]=E[\sum X_i^2]$

$=n\cdot E[X_i^2]$

$=n\cdot(Var X_i+(\overset{0}{E[X_i]})^2]$

$=n\cdot\sigma^2=n.$

$Var[X]=Var[\sum X_i^2]$

$=\sum Var[X_i^2]=n\cdot Var[X_i^2]$

$=n\left(\underbrace{E[X_i^4]}_{峰度}-(E[X_i^2])^2\right)$

$=n\cdot(3\cdot VarX_i^2-VarX_i^2)=n\cdot 2\cdot VarX_i^2=2n$

② $X_1=X_1^2+\dots+X_{n_1}^2$

$X_2=(X_1')^2+\dots+(X_{n_2}')^2$

$X_1+X_2=X_1^2+\dots+X_{n_1}^2+(X_1')^2+\dots+(X_{n_2}')^2$

$=(X_1'')^2+\dots+(X_{n_1+n_2}'')^2$

...

③ Recall CLT:

$\frac{\sum_i Y_i-E[\sum_i Y_i]}{(Var(\sum Y_i))^{\frac{1}{2}}} \to N(0,1)$

取 $Y_i=X_i^2$ with $EY_i=EX_i^2=1$

$Var(\sum Y_i)=VarX=\sqrt{2n}$

$\Rightarrow \frac{\sum_i Y_i-n}{\sqrt{2n}} \to N(0,1).$

习题 6.2

设 $(X_1, X_2, \cdots, X_5)$ 是来自总体 $N(0,4)$ 的一个样本，且

$$Y = \underbrace{aX_1^2}_{4a\cdot\left(\frac{X_1}{2}\right)^2} + \underbrace{b(2X_2+3X_3)^2}_{52b\cdot\left(\frac{2X_2+3X_3}{\sqrt{52}}\right)^2} + c(4X_4 - X_5)^2$$

$68c\cdot\left(\frac{4X_4-X_5}{\sqrt{68}}\right)^2$

请问 $a, b, c > 0$ 取什么值的时候，随机变量 Y 服从 $\chi^2(n)$ 分布，n 为多少？

$$\boxed{\begin{array}{l} X_1 \sim N(0,4) \\ \frac{X_1}{2} \sim N(0,1) \end{array}}$$

$\Rightarrow a = \frac{1}{4}$.

$2X_2+3X_3 \sim N(0,52)$

$\left(\begin{array}{l} Var[2X_2+3X_3] \\ = 2^2\cdot Var X_2 + 3^2\cdot Var X_3 \\ = 4\times4+9\times4 = 52 \end{array}\right)$

$\frac{2X_2+3X_3}{\sqrt{52}} \sim N(0,1)$

$4X_4-X_5 \sim N(0,68)$

$\Rightarrow \frac{4X_4-X_5}{\sqrt{68}} \sim N(0,1)$.

取 $Y_1 = \frac{X_1}{2}$, $Y_2 = \frac{\cdots}{\sqrt{52}}$, $Y_3 = \frac{\cdots}{\sqrt{68}}$

得 $Y = 4a\cdot Y_1^2 + 52b\cdot Y_2^2 + 68c\cdot Y_3^2$ where $Y_i \sim N(0,1)$

要使 $Y \sim \chi^2(n)$

则 $a = \frac{1}{4}$, $b = \frac{1}{52}$, $c = \frac{1}{68}$, $n = 3$

习题 6.6

设 $X_1, X_2, \cdots, X_{20}$ 是来自总体 $X \sim N(0,1)$ 的一个样本，记

$$Y = \frac{1}{10}\left(\sum_{i=1}^{10} X_i\right)^2 + \frac{1}{10}\left(\sum_{i=11}^{20} X_i\right)^2$$

请问 Y 服从什么分布？

$$\left.\begin{array}{l} Y_1 = \sum_{i=1}^{10} X_i \sim N(0,10) \\ Y_2 = \sum_{i=11}^{20} X_i \sim N(0,10) \end{array}\right\} \Rightarrow \frac{Y_i}{\sqrt{10}} \sim N(0,1)$$

$$Y = \left(\frac{Y_1}{\sqrt{10}}\right)^2 + \left(\frac{Y_2}{\sqrt{10}}\right)^2$$

$$\sim \chi^2(2)$$

t 分布

Student's t-distribution

↳ pseudonym of William Gosset

命题 6.3.3: 设 $X \sim N(0,1)$, $Y \sim \chi^2(n)$, 且 X 与 Y 相互独立, 令

$\frac{X}{(\frac{1}{n}\sum X_i^2)^{\frac{1}{2}}}$

$$\boldsymbol{T = \frac{X}{\sqrt{Y/n}} \sim t(n)}$$

则 T 的分布密度函数 f_T 为

$$f_T(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

(2)

$X_i \sim P$

$\bar{X}_n = \frac{1}{n}\sum X_i$

方差 $\overline{Var} = \frac{1}{n}\sum_i (X_i - EX)^2$

↳ 标准差 $(\overline{Var})^{\frac{1}{2}} =$

标准化 $\bar{X}_n$ 为 $\frac{\bar{X} - EX}{\sqrt{\overline{Var}}}$

if $X_i \sim N(0,1)$

then $EX = 0$

$\overline{Var} = \frac{1}{n}\sum (X_i - 0)^2$

$Y = \sum X_i^2 \sim \chi^2(n)$

t 分布图像

$X \sim t(n)$ 表示 X 服从自由度为 n 的 t 分布

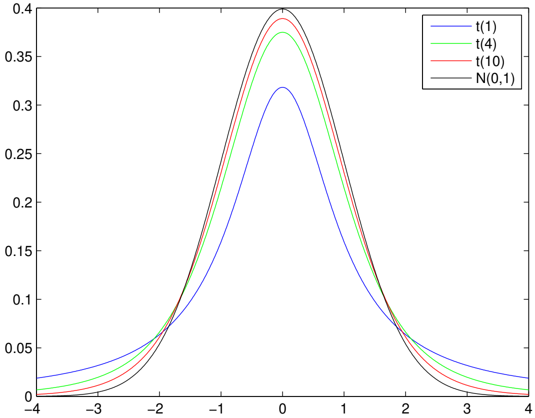


图: $t(n)$ 分布密度函数

t 分布的性质

- t 分布关于原点对称 $\dfrac{X}{\sqrt{\frac{Y}{n}}}$.
- 当 $n \to \infty$， $t(n)$ 趋于标准正态分布
- $e^x = (1+\frac{x}{n})^n$

$$E[X] = 0(n > 1), Var[X] = \frac{n}{n-2}(n > 2)$$

$X \sim N(0,1)$

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

$f_T(n) = \underbrace{\dfrac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\,\Gamma(\frac{n}{2})}}_{\mathrm{I}} \times \underbrace{(1+\frac{x^2}{n})^{-\frac{n+1}{2}}}_{\mathrm{II}}$

$\mathrm{II} = \left((1+\frac{x^2}{n})^{n+1}\right)^{-\frac{1}{2}}$

$\to (e^{x^2})^{-\frac{1}{2}} = e^{-\frac{x^2}{2}}$

$E[X] = 0$ （$\because$ 对称.）

$Var[X] = \frac{n}{n-2}$ (use Beta func).

$\therefore f_T(n) = \mathrm{I} \cdot \mathrm{II}$

$\to \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

$\Rightarrow t(n) \to N(0,1)$

I: Stirling's approximation : for large α

$\Gamma(\alpha) = \sqrt{\frac{2\pi}{\alpha}} \cdot (\frac{\alpha}{e})^\alpha (1 + O(\frac{1}{\alpha}))$

$\mathrm{I} = \frac{1}{\sqrt{n\pi}} \cdot \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \to \frac{1}{\sqrt{n\pi}} \times (\frac{n+1}{2})^{\frac{1}{2}}$

$= \frac{1}{\sqrt{2\pi}}$

$\Rightarrow \dfrac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} = \dfrac{\sqrt{\frac{2\pi}{(n+1)/2}} \cdot (\frac{(n+1)/2}{e})^{\frac{n+1}{2}} \cdot (1+O(\frac{2}{n+1}))}{\sqrt{\frac{2\pi}{n/2}} \cdot (\frac{n/2}{e})^{n/2} (1+O(\frac{2}{n}))} = (\frac{n+1}{2e})^{\frac{1}{2}} \cdot \underbrace{(\frac{n+1}{n})^{\frac{n}{2}}}_{\to ((1+\frac{1}{n})^n)^{\frac{1}{2}} = e^{\frac{1}{2}}} = (\frac{n+1}{2})^{\frac{1}{2}}$

习题 6.4

设 $(X_1, X_2, \cdots, X_9)$ 是来自总体 $N(0,1)$ 的简单随机样本，请确定正数 C，使得

$$\frac{C(X_1+X_2+X_3)}{\sqrt{(X_4+X_5)^2+(X_6+X_7)^2+(X_8+X_9)^2}}$$

服从 t 分布，并指出其自由度。

$Y_1 = X_1+X_2+X_3 \sim N(0,3)$

$\underbrace{X_4+X_5}_{Y_2},\ \underbrace{X_6+X_7}_{Y_3},\ \underbrace{X_8+X_9}_{Y_4} \sim N(0,2)$

$$\frac{CY_1}{\sqrt{Y_2^2+Y_3^2+Y_4^2}} = \frac{\sqrt{3}C\cdot Z_1}{\sqrt{2(Z_2^2+Z_3^2+Z_4^2)}} \quad \left(\because\ Z_1=\frac{Y_1}{\sqrt{3}},\ Z_i=\frac{Y_i}{\sqrt{2}}\right)$$

$$= \frac{\sqrt{3}C}{\sqrt{2}}\cdot\frac{Z_1}{\sqrt{Z_2^2+Z_3^2+Z_4^2}}$$

$$= \frac{C}{\sqrt{2}}\ \frac{Z_1}{\sqrt{\frac{Z_2^2+Z_3^2+Z_4^2}{3}}}$$

要使 $\cdots \sim t(n)$.

则 $\frac{C}{\sqrt{2}} = 1$ i.e. $C=\sqrt{2}$.

$n=3$

F 分布

命题 6.3.4：若 $X \sim \chi^2(m)$，$Y \sim \chi^2(n)$，且 X 与 Y 相互独立，令

$$\frac{\frac{1}{m}\sum_{i=1}^{m} X_i^2}{\frac{1}{n}\sum_{i=1}^{n} X_i^2}$$

$$Z = \frac{X/m}{Y/n} \sim F(m, n),$$

$$\frac{1}{Z} = \frac{Y/n}{X/m} \sim F(n, m)$$

则 Z 的分布密度函数为

$$f_Z(x) = \begin{cases} \dfrac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} m^{\frac{n}{2}} n^{\frac{m}{2}} \dfrac{x^{m/2-1}}{(mx+n)^{(m+n)/2}}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

称随机变量 Z 服从第一自由度为 m，第二自由度为 n 的 F 分布，记作 $Z \sim F(m, n)$。

F 分布图像

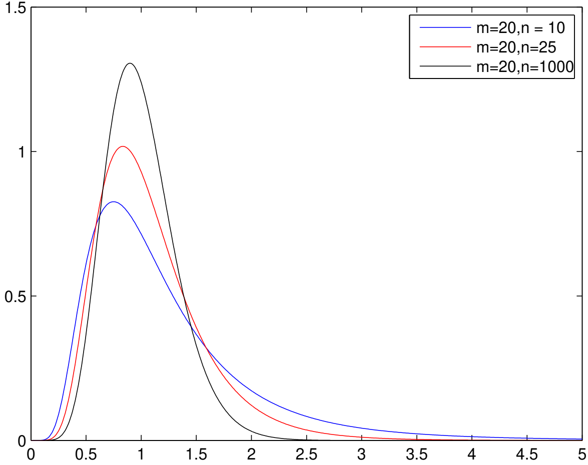


图: $F(m, n)$ 分布密度函数

F 分布

$Z \sim F(m,n)$，那么 $\boldsymbol{\frac{1}{Z} \sim F(n,m)}$。
证明:

$$\begin{aligned} F_{\frac{1}{Z}}(t) &= P(\frac{1}{Z} \le t) \\ &= P(Z \ge \frac{1}{t}) \\ &= 1 - P(Z < \frac{1}{t}) \\ &= 1 - \int_0^{\boxed{\frac{1}{t}}} f_Z(x)dx \end{aligned}$$

F 分布

$$= \frac{d\frac{1}{t}}{dt} \cdot \frac{dF_{\frac{1}{Z}}(t)}{d\frac{1}{t}} = -\frac{1}{t^2} \cdot \frac{d}{d\frac{1}{t}}\left(1-\int_0^{\frac{1}{t}} f_Z(x)dx\right)$$

$$= \frac{1}{t^2} \cdot \frac{d}{d\frac{1}{t}} \int_0^{\frac{1}{t}} f_Z(x)dx$$

$$f_{\frac{1}{Z}}(t) = \frac{d}{dt}F_{\frac{1}{Z}}(t) = t^{-2} f_Z(\frac{1}{t})$$

$$f_Z(x) = C_{m,n} \cdot \frac{x^{\frac{m}{2}-1}}{(mx+n)^{\frac{m+n}{2}}}$$

$$= t^{-2} C_{m,n} \frac{t^{1-\frac{m}{2}}}{(mt^{-1}+n)^{\frac{m+n}{2}}}$$

$$= t^{-2} C_{m,n} \frac{t^{\frac{m+n}{2}-\frac{m}{2}+1}}{(nt+m)^{\frac{m+n}{2}}}$$

$$= C_{m,n} \frac{t^{\frac{n}{2}-1}}{(nt+m)^{\frac{m+n}{2}}}$$

（相当于 m,n 交换

$\Rightarrow \frac{1}{Z} \sim F(n,m)$

其中

$$C_{m,n} = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} m^{\frac{n}{2}} n^{\frac{m}{2}}$$

习题 6.3

设 $X_1, X_2, \cdots, X_8$ 是来自总体 $N(0,1)$ 的简单随机样本，求常数 c，使得

$$\frac{c(X_1^2 + X_2^2)}{(X_3 + X_4 + X_5)^2 + (X_6 + X_7 + X_8)^2}$$

服从 F 分布，并指出其自由度

$Y_1 = \frac{X_3+X_4+X_5}{\sqrt{3}}$, $Y_2 = \frac{X_6+X_7+X_8}{\sqrt{3}} \sim N(0,1)$

原 r.v. $= \frac{c}{3} \cdot \frac{X_1^2+X_2^2}{Y_1^2+Y_2^2}$ $\rightarrow \chi^2(2)$ $\rightarrow \chi^2(2)$

服从 $F(2,2)$

当 $c=3$

分位数

分位数: 设 $X \sim \phi(n)$，$0 < \alpha < 1$，称满足

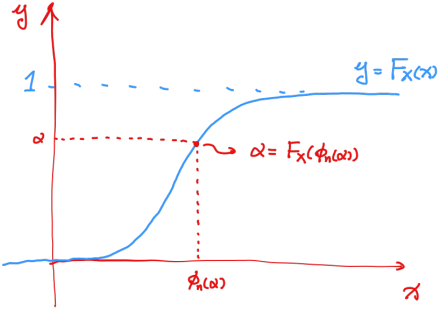
$$P(X \leq \phi_\alpha(n)) = \alpha$$

的实数 $\phi_\alpha(n)$ 为分布 $\phi(n)$ 的 α 分位数，或叫做 α 分位点。

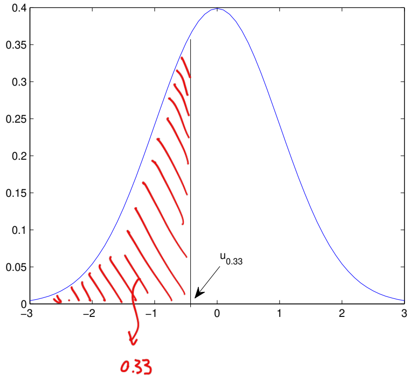
Fix n.

$F_X(\phi_\alpha(n)) = P(X \leq \phi_\alpha(n)) = \alpha$

$\Rightarrow \phi_\alpha(n) = F_X^{-1}(\alpha)$



分位数



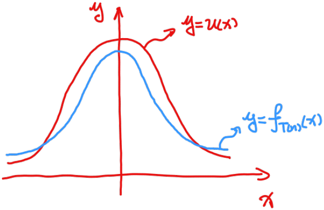
上图是标准正态分布的密度区线，其中 $u_{0.33}$ 是其 0.33 分位点，表示

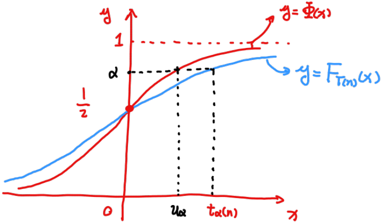
$$P(X \leq u_{0.33}) = 0.33$$

计算 t 分位数

若 $X \sim t(n)$，由于 $\lim_{n\to\infty} X \sim N(0,1)$，

$$\lim_{n\to\infty} t_\alpha(n) = u_\alpha$$

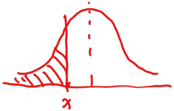




练习

$$t_{0.95}(10),\ t_{0.1}(20),\ t_{0.9}(50)$$

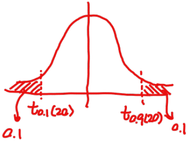
查表 (P.228)



$t_{0.95}(10) = 1.8125$

$t_{0.1}(20) = -t_{0.9}(20)$

$= -1.3253$



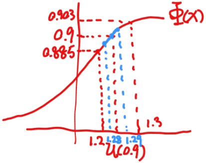
$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

$t_{0.9}(50) \sim u(0.9) \approx 1.28+1.29$

≈ 1.285

x	0.00	0.01	0.02	...	0.08
0.0	$x=0.0+0.00$ $\Phi(0)$				
0.1					
0.2					
0.3					$\Phi(x)=0.9$

$u_\alpha = \Phi^{-1}(\overset{0.9}{\alpha})$



计算 χ 分位数

$X \sim \chi^2(n)$

$\Rightarrow \begin{cases} EX = n \\ \mathrm{Var}X = 2n \end{cases} \Rightarrow$ 标准化 $\dfrac{X-n}{\sqrt{2n}} \xrightarrow{CLT} N(0,1)$

- 若 $X \sim \chi^2(n)$ 分布，当 $n \to \infty$，$(X-n)/\sqrt{2n} \sim N(0,1)$，所以

$$
\begin{aligned}
\alpha &= P(X \leq \chi^2_\alpha(n)) \\
&= P((X-n)/\sqrt{2n} \leq (\chi^2_\alpha(n) - n)/\sqrt{2n})
\end{aligned}
$$

即

n较大时.

$$u_\alpha \approx (\chi^2_\alpha(n) - n)/\sqrt{2n}$$

解出

$$\boldsymbol{\chi^2_\alpha(n) \approx n + \sqrt{2n} \times u_\alpha}, \quad n \to \infty$$

练习

$$\chi^2_{0.99}(10), \chi^2_{0.05}(20), \chi^2_{0.95}(60)$$

查表 P.229

$\chi^2_{0.99}(10) = 23.209$

$\chi^2_{0.05}(20) = 10.851$

$\chi^2_{0.95}(60) \approx n + \sqrt{2n} \cdot u_{0.95}$ where $n = 60$

$= 60 + \sqrt{2 \times 60} \times u_{0.95}$

计算 F 分位数

若 X 服从 $F(m, n)$，则有 $\overset{Y=}{\boxed{\frac{1}{X}}} \sim F(n, m)$，若已知 $F_\alpha(m, n)$，那么

$$\alpha = P(X < F_\alpha(m, n)) = P(\frac{1}{X} > \frac{1}{F_\alpha(m, n)})$$

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

即

$$P(\frac{1}{X} \le \frac{1}{F_\alpha(m, n)}) = 1 - \alpha \quad \checkmark$$

$$\text{"} P[Y \le \frac{1}{F_\alpha(m,n)}] \quad (Y \sim F(n,m))$$

也即

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

练习

$$F_{0.99}(5,4),\ F_{0.05}(3,7)$$

查表 P.238

$$P_{0.99}(5,4) = 15.5$$

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

得 $F_{0.05}(3,7) = \dfrac{1}{F_{1-0.05}(7,3)} = \dfrac{1}{F_{0.95}(7,3)} = \dfrac{1}{8.89}$