第六章：数理统计的基本概念

概率论和数理统计

概率论和数理统计的区别：

1. 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的计算.
2. 数理统计：研究对象的分布未知或者不完全已知的情形下，通过抽样对得到数据的进行分析、推断并作出一定的决策。

例子：灯泡寿命

生产厂家声称他们生产的灯泡平均寿命不低于6000小时，如何验证?

例子：灯泡寿命

生产厂家声称他们生产的灯泡平均寿命不低于6000小时，如何验证？

- 破坏性试验 $\rightarrow$ 不能全检
- 抽样检测
- 以部分数据信息来推断整体未知参数，就是数理统计研究问题的基本方式。

6.1 总体，样本和统计量

总体相关概念

- 总体：研究对象的全体
- 个体：总体中的成员
- 总体的容量：总体中包含的个体数
- 有限总体：容量有限的总体
- 无限总体：容量无限的总体，通常将容量非常大的有限总体也按无限总体处理

练习

(1) 了解某校大学生做过家教（包括正在做家教）的比例。
(2) 了解某城市的空气质量情况，关注该城市的PM2.5值。
(3) 药厂研究某种药物在人体中的吸收情况。

总体指标的特性

- 实际中人们通常只关注总体的某个（或几个）指标
- 总体的某个指标X可以看成一个随机变量
- 直接将X称为总体，其分布函数为$F(x)$

伯努利总体实例

了解某校学生"做过家教"的情况，对每个学生来说，以 $\{X = 1\}$ 表示"做过家教"，以 $\{X = 0\}$ 表示"未做过家教"，则总体

$$X \sim B(1, p),$$

p 是全校学生中做过家教所占的比例，未知。即

$$P(X = x) = p^x (1 - p)^{1-x}, \ x = 0, 1.$$

统计推断方法

- 如何推断总体分布的未知参数（或分布）？
- 方法：从总体中抽取样本，根据样本数据作出推断
- 样本：被抽取的部分个体

简单随机样本定义

Definition (简单随机样本)

若随机样本$(X_1, X_2, \cdots, X_n)$满足以下条件：

1. 代表性：每个X_i与总体X同分布
2. 独立性：$X_1, X_2, \cdots, X_n$相互独立

则称为容量为n的简单随机样本。

说明

后续讨论中提到的"样本"均指简单随机样本

简单随机抽样实施

- 获得简单随机样本的抽样称为简单随机抽样
- 实施方法：
  - 有限总体：有放回抽样
    - 大容量时近似采用不放回抽样
  - 无限总体：不放回抽样

样本与样本值的区分

1. 一个样本（容量为n）$X_1, X_2, \ldots, X_n$是指n个独立与总体分布相同的随机变量。
2. 对样本进行一次观测，得到实际数值（n个）$x_1, x_2, \ldots, x_n$称为样本观察值（或样本值）。

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

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

其中，$F_X(\cdot,\theta)$ 可以是分布函数，也可以是分布密度函数（连续型随机变量）或概率分布（离散型随机变量）.

习题

1. 设总体 $X \sim N(\mu, \sigma^2)$，$(X_1, X_2, \cdots, X_n)$ 为其样本，求$(X_1, X_2, \cdots, X_n)$ 的联合分布密度函数。
2. 设总体 $X \sim B(1, p)$，$(X_1, X_2, \cdots, X_n)$ 为其样本，求$(X_1, X_2, \cdots, X_n)$ 的联合概率分布函数。

统计量的引入

考虑某校展开英语水平考试，随机抽取学生的成绩分别为：$X_1, X_2, \ldots, X_n$

问题： 如何评价该校的英语整体学习情况?

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \max\{X_1, X_2, \cdots, X_n\} \quad \min\{X_1, X_2, \cdots, X_n\}$$

统计量的引入

考虑某校展开英语水平考试，随机抽取学生的成绩分别为：$X_1, X_2, \ldots, X_n$

问题：如何评价该校的英语整体学习情况?

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \max\{X_1, X_2, \cdots, X_n\} \quad \min\{X_1, X_2, \cdots, X_n\}$$

$$\text{样本} \xrightarrow[\text{研究的途径或手段}]{\text{统计量}} \text{总体}$$

统计量

统计量

设 $X_1, X_2, \ldots, X_n$ 是来自总体 $X \sim F(x)$ 的样本，$g(x_1, x_2, \ldots, x_n)$ 是 n 元实值连续函数，若函数 g 不含未知参数，则称随机变量 $g(X_1, X_2, \ldots, X_n)$ 为统计量。

练习

设 $X_1, X_2, \ldots, X_n$ 是来自总体$N(\mu, \sigma^2)$ 的样本，其中 μ, σ^2 均未知。则下列中哪些是统计量?

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

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

$$\min\{X_1, X_2, \ldots, X_n\}$$

常用的统计量 I

1. 样本均值：

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

2. 样本方差及样本标准差：

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

修正样本方差及样本修正标准差：

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

常用的统计量 II

3 样本矩

- 样本的 k 阶原点矩：

$$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k.$$

- 样本的 k 阶中心矩：

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

常用的统计量 III

4 顺序统计量： 将样本 $X_1, X_2, \ldots, X_n$ 按由小到大的顺序排成

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$$

则称统计量 $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ 为顺序统计量

- 样本极小值 $X_{(1)} = \min\{X_1, X_2, \cdots, X_n\}$
- 样本极大值 $X_{(n)} = \max\{X_1, X_2, \cdots, X_n\}$
- 样本中位数：

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

- 样本极差：

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

6.2 经验分布函数

经验分布函数——从局部推断整体

问题

如何由样本估计总体的分布函数?

问题

如何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$x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$，定义

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

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——从局部推断整体

利用伯努利大数定律，可以证明对于任意的$\epsilon > 0$，有

$$\lim_{n\to\infty} P\left(\left|F_n^X(x) - F_X(x)\right| \geq \epsilon\right) = 0, \quad \forall x \in (-\infty, \infty).$$

上面的结果说明，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

某产品40个装为一盒，为判断其每盒中次品数的情况，抽取7盒产品，检查其次品数。得样本观察值(0,3,2,1,1,0,1)。试指出总体和样本，并写出经验函数。

6.3 抽样分布

样本均值与样本方差的数字特征

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

1. $$E[\overline{X}] = \mu, \quad Var[\overline{X}] = \frac{\sigma^2}{n}.$$
2. $$E[S_n^2] = \frac{n-1}{n}\sigma^2, \quad E[S_n^{*2}] = \sigma^2.$$

抽样分布 II

证明：

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

再由 X_i 之间的独立性有

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

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

抽样分布 III

所以

$$
\begin{aligned}
E[S_n^2] &= E\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i^2] - E[\overline{X}^2] \\
&= \frac{1}{n}\sum_{i=1}^{n}(Var[X_i] + (E[X_i])^2) - (Var[\overline{X}] + (E[\overline{X}])^2) \\
&= \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 \\
&= \frac{n-1}{n}\sigma^2.
\end{aligned}
$$

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

练习

设 $(X_1, X_2, \cdots, X_n)$ 是取自总体 X 的一个样本，在 (1) $X \sim B(1, p)$, (2) $X \sim Exp[\lambda]$ 两种情况下，分别求 $E[\bar{X}]$, $Var[\bar{X}]$, $E[S_n^2]$.

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

$$Y = X_1^2 + X_2^2 + \cdots + X_n^2 \sim \chi^2(n).$$

称随机变量Y服从自由度为 n 的 χ^2 分布.

特别地，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 II

若 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

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

其中 $\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t}\, \mathrm{d}t.$

若 $X \sim \Gamma(\alpha, \lambda)$ （Γ 分布）,其的分布密度函数为

$$f_X(x) = \begin{cases} \dfrac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x > 0, \\ 0, & x \le 0. \end{cases}$$

$$X \sim \chi^2(n) \Leftrightarrow X \sim \Gamma(\tfrac{n}{2}, \tfrac{1}{2})$$

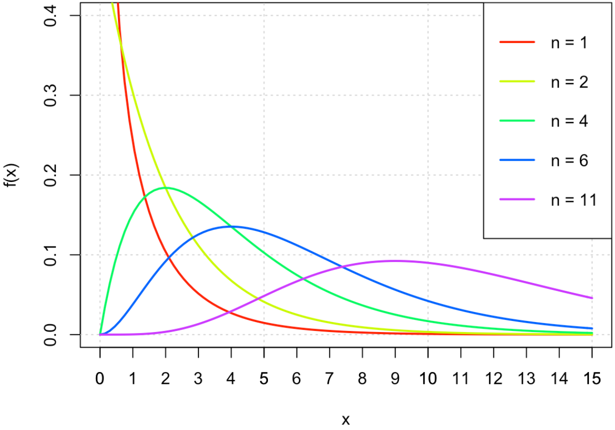


Figure: $\chi^2(n)$的分布密度

性质

Γ 分布的性质：

1. 若 $X \sim \Gamma(\alpha, \lambda)$，那么 $E[X] = \frac{\alpha}{\lambda}$，$Var[X] = \frac{\alpha}{\lambda^2}$.
2. 若 $X_1 \sim \Gamma(\alpha_1, \lambda)$，$X_2 \sim \Gamma(\alpha_2, \lambda)$，且 X_1 与 X_2 独立，那么 $X_1 + X_2 \sim \Gamma(\alpha_1 + \alpha_2, \lambda)$.（$\lambda$ 要相同！）

χ^2 分布的性质：

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

练习

1. 设 $X_1, X_2, \cdots, X_n$ 是来自总体 $X \sim N(\mu, \sigma^2)$ 的一个样本，请问 $\frac{1}{\sigma^2}\sum_{k=1}^{n}(X_k - \mu)^2$ 服从什么分布？
2. （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

$$Y = aX_1^2 + b(2X_2 + 3X_3)^2 + c(4X_4 - X_5)^2.$$

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？

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

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

则 称随机变量T 服从自由度为 n 的 t 分布.

T 的分布密度函数 f_T 为

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

t 分布 II

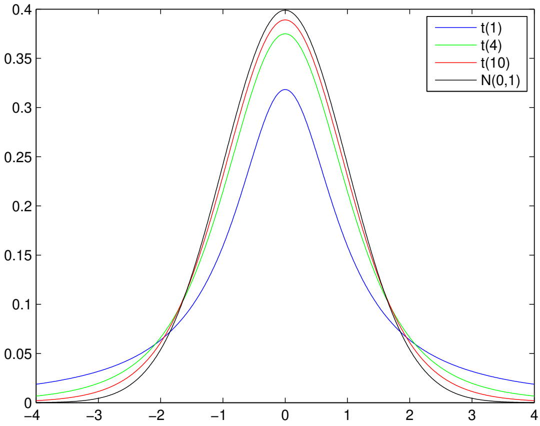


Figure: $t(n)$分布密度函数

t 分布的性质

1. t 分布关于原点对称.
2. $E[X] = 0\ (n > 1), \quad Var[X] = \frac{n}{n-2}\ (n > 2).$
3. 当 $n \to \infty$ 时，$t(n)$ 趋于标准正态分布.

练习

1. 设总体 $X \sim N(0,1)$，$(X_1, X_2, \cdots, X_n)$ 为简单随机样本。试问下列统计量各服从什么分布？
   (1) $\frac{X_1 - X_2}{\sqrt{X_3^2 + X_4^2}}$；　(2) $\frac{\sqrt{n-1}X_1}{\sqrt{\sum_{i=2}^n X_i^2}}$.

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.

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

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

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

Z 的分布密度函数为

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

F 分布 II

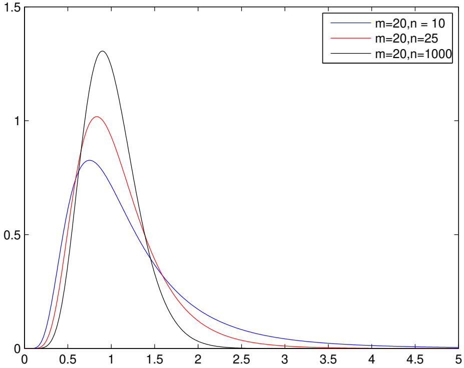


Figure: $F(m,n)$分布密度函数

F 分布 III

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

F 分布 IV

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

证明：

$$\begin{aligned}
F_{\frac{1}{Z}}(t) &= P(\frac{1}{Z} \leq t) \\
&= P(Z \geq \frac{1}{t}) \quad \text{（}Z\text{取值非负）} \\
&= 1 - P(Z < \frac{1}{t}) \\
&= 1 - \int_0^{\frac{1}{t}} f_Z(x)\,\mathrm{d}x
\end{aligned}$$

F 分布 V

故

$$
\begin{aligned}
f_{\frac{1}{Z}}(t) = \frac{\mathrm{d}F_{\frac{1}{Z}}(t)}{\mathrm{d}t} &= t^{-2} f_Z(\frac{1}{t}) \\
&= t^{-2} C_{m,n} \frac{t^{1-\frac{m}{2}}}{(mt^{-1}+n)^{\frac{m+n}{2}}} \\
&= t^{-2} C_{m,n} \frac{t^{\frac{m+n}{2}-\frac{m}{2}+1}}{(nt+m)^{\frac{m+n}{2}}} \\
&= C_{m,n} \frac{t^{\frac{n}{2}-1}}{(nt+m)^{\frac{m+n}{2}}},
\end{aligned}
$$

其中

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

练习

1. 设总体 $X \sim N(0,1)$，$(X_1, X_2, \cdots, X_n)$ 为简单随机样本。试问 $\frac{(\frac{n}{3}-1)\sum_{i=1}^{3} X_i^2}{\sum_{i=4}^{n} X_i^2}$ 服从什么分布？
2. 若$T \sim t(n)$，问 T^2服从什么分布？
3. （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.

分位数

分位数

设连续型随机变量 $X \sim f(x)$，对给定的 α，$0 < \alpha < 1$，存在一个实数 x_α，使得

$$F_X(x_\alpha) = P\{X \le x_\alpha\} = \int_{-\infty}^{x_\alpha} f(x)dx = \alpha$$

则称 x_α 为密度函数 $f(x)$ 的α分位数，或叫做α分位点.

计算分位数

查询的两种方式：

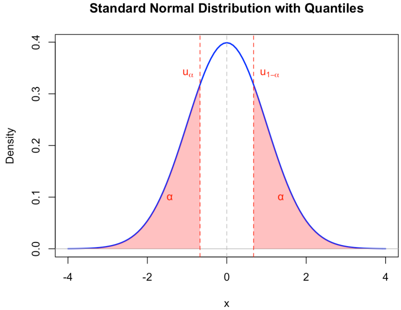
- R软件: $qnorm(\alpha), qchisq(\alpha, n), qt(\alpha, n), qf(\alpha, m, n)$
- 查表

标准正态分布的分位点

- 标准正态分布 $N(0,1)$ 的α分位点记为 u_α

$$\Phi(u_\alpha) = \int_{-\infty}^{u_\alpha} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \alpha$$

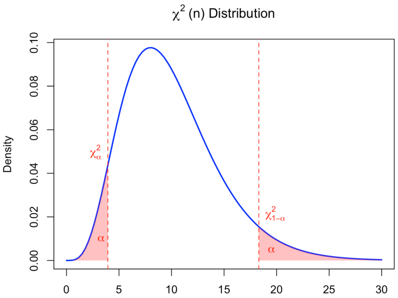
- 性质：$u_\alpha = -u_{1-\alpha}$
- 常用查表值：$u_{0.99} = 2.33$ $\quad u_{0.975} = 1.96$ $\quad u_{0.95} = 1.645$



χ^2分布的分位点 I

- $\chi^2(n)$ 分布的α分位点记为 $\chi^2_\alpha(n)$
- 若 $X \sim \chi^2(n)$ 分布，当 $n \to \infty$,

$$\chi^2_\alpha(n) \approx \sqrt{2n}u_\alpha + n.$$



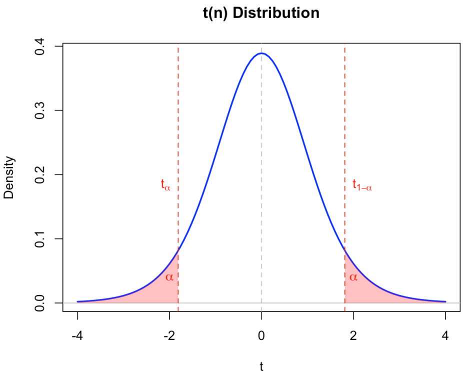
χ^2分布的分位点 II

习题：

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

t分布的分位点 I

- $t(n)$ 分布的α分位点记为 $t_\alpha(n)$
- 性质：$t_\alpha(n) = -t_{1-\alpha}(n)$
- 当 $n > 45$ 时，$t_\alpha(n) \approx u_\alpha$



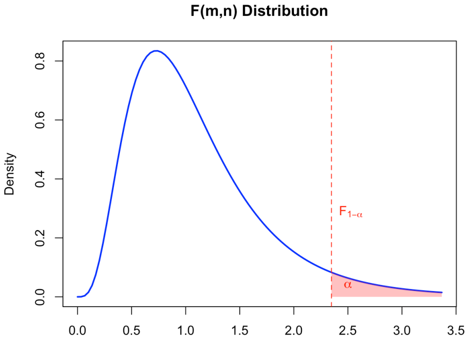
t分布的分位点 II

习题：

- $t_{0.95}(10)$
- $t_{0.1}(20)$
- $t_{0.99}(50)$

F分布的分位点 I

- $F(m,n)$ 分布的α分位点记为 $F_\alpha(m,n)$
- 性质： $F_\alpha(m,n) = \frac{1}{F_{1-\alpha}(n,m)}$



F分布的分位点 II

习题：

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

正态总体的抽样分布

抽样分布基本定理

设总体 $X \sim N(0,1)$，$(X_1, X_2, \cdots, X_n)$ 为其样本，则样本均值 $\overline{X} \sim N(0, \frac{1}{n})$，$nS_n^2 \sim \chi^2(n-1)$，并且 $\overline{X}$ 与 S_n^2 相互独立.

正态分布的特殊性

1. 对于服从正态分布的总体，样本均值 $\overline{X}$ 服从一维正态分布.
2. $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$，即 S_n^2 是 $\overline{X}$ 的函数，但是它们却相互独立.

几个推论 I

1. 推论 **6.3.1** 设总体 $X \sim N(\mu, \sigma^2)$，$(X_1, X_2, \cdots, X_n)$ 为其样本，则$\overline{X}$ 与 S_n^2 相互独立，且

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

$$\frac{(n-1)S_n^{*2}}{\sigma^2} = \frac{nS_n^2}{\sigma^2} \sim \chi^2(n-1).$$

2. 推论 **6.3.2** 设总体 $X \sim N(\mu, \sigma^2)$，$(X_1, X_2, \cdots, X_n)$ 为其样本，那么

$$\frac{\overline{X} - \mu}{S_n^*} \sqrt{n} = \frac{\overline{X} - \mu}{S_n} \sqrt{n-1} \sim t(n-1).$$

几个推论 II

3. 推论 **6.3.3** 设总体 $X \sim N(\mu_1, \sigma_1^2)$，$(X_1, X_2, \cdots, X_m)$ 为其样本，样本均值为 $\overline{X}$，样本方差为 S_{1m}^2；另有与 X 独立的总体 $Y \sim N(\mu_2, \sigma_2^2)$，$(Y_1, Y_2, \cdots, Y_n)$ 为其样本，样本均值为 $\overline{Y}$，样本方差为 S_{2n}^2，那么

$$\frac{mS_{1m}^2}{nS_{2n}^2} \frac{\sigma_2^2}{\sigma_1^2} \frac{n-1}{m-1} = \frac{S_{1m}^{*2}}{S_{2n}^{*2}} \frac{\sigma_2^2}{\sigma_1^2} \sim F(m-1, n-1).$$

4. 推论 **6.3.4** 在推论 6.3.3 的假定中增加一个条件 $\sigma_1 = \sigma_2$，那么

$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{mS_{1m}^2 + nS_{2n}^2}{m+n-2}} \sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t(m+n-2).$$

命题6.3.5　设总体 X 的分布函数为 F_X，分布密度函数为 f_X，那么

$$f_{X_{(k)}}(x) = \frac{n!}{(n-k)!(k-1)!}[F_X(x)]^{k-1}[1-F_X(x)]^{n-k}f_X(x),$$

$k = 1, 2, \cdots, n$.

特别地，

$$f_{X_{(1)}}(x) = n[1-F_X(x)]^{n-1}f_X(x),$$
$$f_{X_{(n)}}(x) = n[F_X(x)]^{n-1}f_X(x).$$

练习

1. 从总体 $X \sim N(20, 16)$ 中抽取容量为 $n = 25$ 的样本，求样本均值 $\bar{X}$ 落在区间 $(18, 22)$ 内的概率。

2. 设总体 $X \sim N(\mu, 4)$，$(X_1, X_2, \cdots, X_n)$ 为简单随机样本，样本均值为 $\bar{X}$。试问：样本容量 n 应取多大才能使：
   (1) $E[|\bar{X} - \mu|^2] \leq 0.1$，
   (2) $P(|\bar{X} - \mu| \leq 0.1) \geq 0.95$?

3. 设总体 $X \sim N(0, 1)$，(X_1, X_2) 为简单随机样本。试求常数 k，使

$$P\left(\frac{(X_1 + X_2)^2}{X_1^2 + X_2^2} > k\right) = 0.10.$$

4. （习题 6.5）设 $X_1, X_2, \cdots, X_{10}$ 是来自总体 $X \sim N(\mu, \sigma^2)$ 的一个样本，且

$$\overline{X} = \frac{1}{9}\sum_{i=1}^{9} X_i, \quad S_9^{*2} = \frac{1}{8}\sum_{i=1}^{9}(X_i - \overline{X})^2, \quad T = \frac{3(X_{10} - \overline{X})}{S_9^* \sqrt{10}},$$

请问 T 服从何种分布?