Basics of Probs and Stats

5. 数理统计学基本知识

梁雪吟

地球科学与资源学院
1001195109@cugb.edu.cn

Spring 2022

Statisticians, like artists, have the bad habit of falling in love with their models.

–George E. P. Box

Outline

1. Review
   - 切比雪夫不等式
   - LLM and CLT
2. 基本概念
3. 抽样分布
   - χ^2 分布
   - t 分布
   - F 分布
4. 三大分布之间的关系
   - 三大分布之间的关系
   - 三大分布有关证明
   - 上 α 分位点的应用

5.1 Review

切比雪夫不等式

Define （期望和方差之间的关系）

$$P\{|X - EX| \geq \varepsilon\} \leq \frac{DX}{\varepsilon^2}$$

- 离均值越远发生的概率越小
- 注意不等号的方向
- 两个及以上，注意题目条件**是否独立**（思考: 方差的求和）
- Hint: 看见期望、方差、不等关系，使用切比雪夫不等式

练习 1

设 X 与 Y 的期望分别为-2 和 2，方差分别为 1 和 4，$\rho_{XY} = -0.5$，则根据切比雪夫不等式，有 $P\{|X+Y| \geq 6\} \leq ?$

设 X 与 Y 的期望分别为-2 和 2，方差分别为 1 和 4，$\rho_{XY} = -0.5$，则根据切比雪夫不等式，有 $P\{|X + Y| \geq 6\} \leq ?$

Solution

由题可得：

$E(X + Y) = EX + EY = -2 + 2 = 0$

$D(X + Y) = DX + DY + 2Cov(X, Y) = 1 + 2 + 2\rho_{XY}\sqrt{DXDY} = 3$

所以：

$P\{|X + Y| \geq 6\} \leq \frac{3}{36} = \frac{1}{12}$

练习 2

设在每次试验中，事件 A 发生的概率为 $P(A) = 0.25$

(1) 进行 300 次独立重复实验，以 X 表示 A 发生的次数，用切比雪夫不等式估计 X 和 EX 的偏差不大于 50 的概率？

(2) 是否可用 0.925 的概率确信，在 1000 次试验中，A 发生次数在 200 和 300 之间？

设在每次试验中，事件 A 发生的概率为 $P(A) = 0.25$
(1) 进行 300 次独立重复实验，以 X 表示 A 发生的次数，用切比雪夫不等式估计 X 和 EX 的偏差不大于 50 的概率？
(2) 是否可用 0.925 的概率确信，在 1000 次试验中，A 发生次数在 200 和 300 之间？

Solution

(1) 由题可得：$X \sim B(300, 0.25)$

$EX = 300 \times 0.25 = 75, DX = 300 \times 0.25 \times 0.75 = \frac{225}{4}$

所以，由切比雪夫不等式可得

$P\{|X - EX| \leq 50\} \geq 1 - \frac{\frac{225}{4}}{50^2} \approx 0.9775$

(2) 由题可得：$X \sim B(1000, 0.25)$

故 $EX = 350, DX = \frac{375}{2}$

所以，由切比雪夫不等式可得

$P\{200 \leq X \leq 300\} = P\{|X - 250| \leq 50\} \geq 1 - \frac{\frac{375}{2}}{50^2} \approx 0.925$

总体和样本

- 总体：想要研究的那群人/物（统计学中样将总体视作无限大的，我们无法穷尽观察到总体中每个个体的情况）
- 简单随机抽样：获得独立同分布样本

  i.i.d. sample (Independently and Identically Distributed)
- 样本：从总体中随机抽取的独立的个体，每一个样本点与总体有相同的分布 $(X_1, X_2, ..., X_n)$
  - 在我们“具体确定样本观测值”之前，随机样本里面的每一个大写字母，都是随机变量.
- 观测值（观察值、样本值）$(x_1, x_2, ..., x_n)$

大数定律 LLM

- 大数定律：样本均值依概率收敛于总体期望
- 伯努利大数定律：频率依概率收敛于概率 (统计学的理论依据)

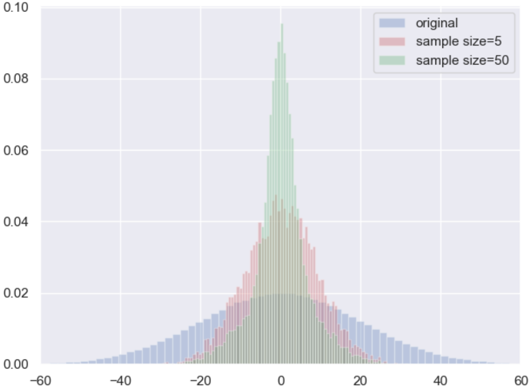


Figure 1: LLN 的模拟（Source：知乎）

中心极限定理 CLT

- 中心极限定理：样本点的样本均值依分布收敛到正态分布
  - 求解方式：因为我们有标准正态分布的表，所以根据 CLT 可以将所求的随机变量化成标准正态分布的形式，从而查表求解。

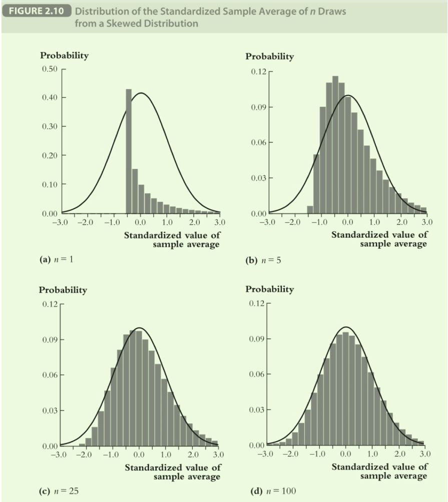


Figure 2: CLT 的模拟

CLT 的应用

某种产品由 20 个相同部件连接而成，每个部件的长度是均值为 2mm、标准差为 0.02mm 的随机变量。假如这 20 个部件的长度相互独立同分布，且规定产品总长为 $(40 \pm 0.2)mm$ 时为合格品，求该产品的不合格率。

某种产品由 20 个相同部件连接而成，每个部件的长度是均值为 2mm、标准差为 0.02mm 的随机变量。假如这 20 个部件的长度相互独立同分布，且规定产品总长为 $(40 \pm 0.2)mm$ 时为合格品，求该产品的不合格率。

Solution

记 X_i 为第 i 个部件的长度，则 $Y = X_1 + X_2 + ... + X_{20}$ 为总长度，且 $E(X_i) = 2$，$D(X_i) = 0.0004$，所以，$E(Y) = 40$，$D(Y) = 0.008$

根据独立同分布的中心极限定理可得：

$$P(39.8 \le Y \le 40.2) \approx \Phi\left(\frac{0.2}{\sqrt{0.008}}\right) - \Phi\left(-\frac{0.2}{\sqrt{0.008}}\right) \tag{1}$$

$$= 2\Phi(2.236) - 1 = 0.9746 \tag{2}$$

5.2 基本概念

基本概念

- 样本分布函数（书 P117）
- 经验分布函数（书 P117）
- 统计量：不含任何未知参数的样本函数称为统计量，使该函数值能反映总体某方面的信息。

重要的几个统计量

- 样本均值：$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
- 样本方差：$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i^2 - n\bar{X}^2)$
  - 自由度的调整是为了后续得到无偏估计，统计学中才调整自由度

Hint: basic properties of the summation operator

$$\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^{n} X_i(Y_i - \bar{Y}) = \sum_{i=1}^{n}(X_i - \bar{X})\,Y_i$$

- 样本标准差：$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$
- 样本 k 阶（原点）矩：$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \quad (k = 1, 2, 3...)$
- 样本 k 阶中心矩：$B_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k \quad (k = 2, 3...)$

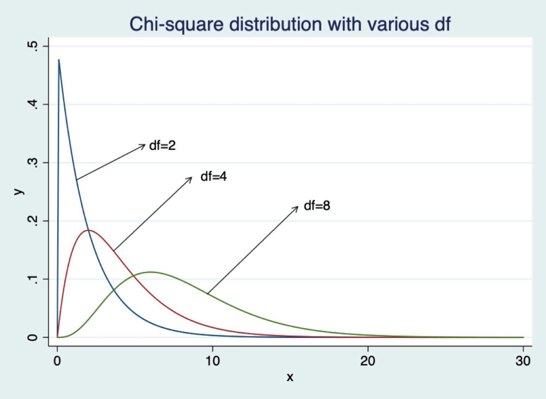
5.3 抽样分布

χ^2 分布

设 $X_1, X_2, X_3 ... X_n$ 是正态总体 $X \sim N(0,1)$ 的样本，则称统计量

$$\chi^2 = X_1^2 + X_2^2 + ... + X_n^2 \tag{3}$$

服从自由度（degrees of freedom）n 的 χ^2 分布记为 $\chi^2 = \chi^2(n)$



性质 1

性质 1: **设随机变量** $\chi_1^2 \sim \chi^2(n_1)$, $\chi_2^2 \sim \chi^2(n_2)$, **且** χ_1^2 **与** χ_2^2 **相互独立**, **则**:

$$\chi_1^2 + \chi_2^2 \sim \chi^2(n_1 + n_2) \tag{4}$$

推广:

$$\sum_{i=1}^{n} \chi_1^2 \sim \chi^2(n_1 + n_2 + ... + n_m) \tag{5}$$

性质 2

性质 2: 设 $X_1, X_2, X_3 ... X_n$ 是来自正态总体 $X \sim N(0,1)$ 的样本，则

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi^2(n) \tag{6}$$

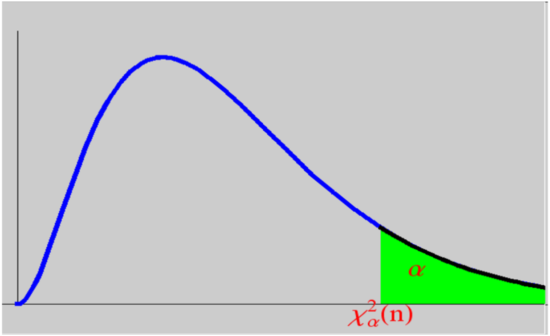
性质 3: 若随机变量量 $\chi^2 \sim \chi^2(n)$，则

$$E(\chi^2)=n, D(\chi^2)=2n \tag{7}$$

上 α 分位点

若随机变量量 $\chi^2 \sim \chi^2(n)$，对给定的正数 $\alpha(0 < \alpha < 1)$，称满足条件

$$P\{\chi^2 > \chi^2_\alpha(n)\} = \int_{i=\chi^2_\alpha(n)}^{\infty} f(y)\,dy = \alpha \tag{8}$$



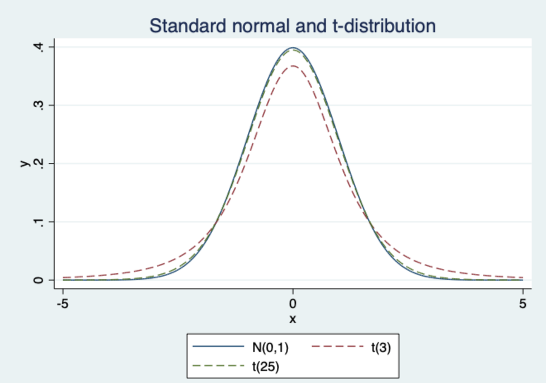
- 学会看表

t 分布

设随机变量 $X \sim N(0,1)$，$Y \sim \chi^2(n)$，且 X 和 Y 相互独立，则称统计量：

$$T = \frac{X}{\sqrt{\frac{Y}{n}}} \tag{9}$$

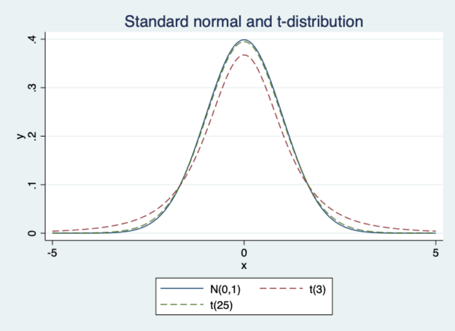
服从自由度（degrees of freedom）n 的 t 分布记为 $T \sim t(n)$



性质

性质 1: 设随机变量量 $T \sim t(n)$，则

$$E(T) = 0, D(T) = \frac{n}{n-2} \tag{10}$$



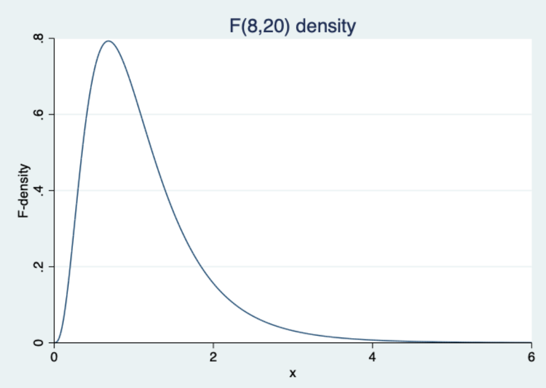
- 尾部比正态分布概率更大
- 因为 t 分布的对称性：$t_{1-\alpha}(n) = -t_\alpha(n)$

F 分布

设随机变量 $U \sim \chi^2(n_1)$， $V \sim \chi^2(n_2)$，且 U 和 V 相互独立，则称统计量：

$$F = \frac{U/n_1}{V/n_2} \tag{11}$$

服从自由度（degrees of freedom） $(n_1.n_2)$ 的 F 分布记为 $F \sim F(n_1, n_2)$



性质

性质 1: 若随机变量量 $F \sim F(n_1, n_2)$，则 $\frac{1}{F} \sim F(n_2, n_1)$
性质 2：F 的上 α 分位点有如下性质

$$F_{1-\alpha}(n_1, n_2) = \frac{1}{F_\alpha(n_2, n_1)} \tag{12}$$

可加性

7. 可加性

(1) 设 $X \sim B(m,p)$, $Y \sim B(n,p)$, 且相互独立, 则 $Z = (X+Y) \sim B(m+n,p)$.

(2) 设 $X \sim P(\lambda_1)$, $Y \sim P(\lambda_2)$ $\left(\text{泊松分布}: P(X=k) = \dfrac{\lambda^k e^{-\lambda}}{k!} (k=0,1,\cdots), \lambda > 0\right)$, 且相互独立, 则 $Z = (X+Y) \sim P(\lambda_1 + \lambda_2)$.

(3) 设 $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, 且相互独立, 则

$$Z = (X+Y) \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

(4) 设 $X \sim \chi^2(n)$, $Y \sim \chi^2(m)$, 且 X, Y 独立, 则 $Z = X+Y \sim \chi^2(m+n)$.

- 可加性的前提: 独立!!

练习 3

设 X 和 Y 都服从 $N(0,1)$，则（　）

A. X+Y 服从正态分布

B. $X^2 + Y^2$ 服从 χ^2 分布

C. X^2 和 Y^2 都服从 χ^2 分布

D. $\frac{X^2}{Y^2}$ 服从 F 分布

练习 3

设 X 和 Y 都服从 $N(0,1)$，则（C）

A. X+Y 服从正态分布

B.X^2+Y^2 服从 χ^2 分布

C.X^2 和 Y^2 都服从 χ^2 分布

D.$\frac{X^2}{Y^2}$ 服从 F 分布

- 题目中没有说 X 和 Y**独立!!** 。ABD 直接排除。

5.4 三大分布之间的关系

性质

设 X 的数学期望位 μ，方差为 σ^2，$X_1, X_2, X_3 ... X_n$ 是来自总体 X 的样本，$\bar{X}$ 为均值，则

$$E(\bar{X}) = \mu, D(\bar{X}) = \frac{\sigma^2}{n} \tag{13}$$

- Proof...

定理

定理 1: 设 X 的数学期望为 μ，方差为 σ^2，$X_1, X_2, X_3 ... X_n$ 是来自正态总体 $X \sim N(\mu, \sigma^2)$ 的样本，$\bar{X}$ 为均值，则

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \tag{14}$$

定理 2: 设 X 的数学期望为 μ，方差为 σ^2，$X_1, X_2, X_3 ... X_n$ 是来自正态总体 $X \sim N(\mu, \sigma^2)$ 的样本，$\bar{X}$ 为均值，S^2 为样本方差，则

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1) \tag{15}$$

且 $\bar{X}$ 与 S^2 相互独立

- Proof...

定理

定理 3: **设** X **的数学期望为** μ，**方差为** σ^2，$X_1, X_2, X_3 ... X_n$ **是来自正态总体** $X \sim N(\mu, \sigma^2)$ **的样本**，$\bar{X}$ **为均值**，S^2 **为样本方差**，**则**

$$\frac{(\bar{X} - \mu)}{S/\sqrt{n}} \sim t(n-1) \tag{16}$$

- Proof...
- **有** S^2 **考虑构造** χ^2 **分布（应用定理** 2**）**

定理 4

定理 4 设 $X_1, X_2, \cdots, X_{n_1}$ 与 $Y_1, Y_2, \cdots, Y_{n_2}$ 分别是来自正态总体 $X \sim N(\mu_1, \sigma_1^2)$ 和 $Y \sim N(\mu_2, \sigma_2^2)$ 的样本，且这两个样本互相独立，$\overline{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$，$\overline{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$ 分别是它们的样本均值，$S_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_i - \overline{X})^2$，$S_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(Y_i - \overline{Y})^2$ 分别是它们的样本方差，则

(1) $\dfrac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$； ⑤

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

$$\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{S_w\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim t(n_1 + n_2 - 2), \tag{6}$$

其中 $$S_w^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}, \quad S_w = \sqrt{S_w^2}.$$

练习

（2021 年期末最后一题）设总体 $X \sim N(0,1)$，$X_1, X_2, X_3 ... X_n$ 是来自总体的简单随机样本，$\bar{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$，$S^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2$，求 $\frac{\sqrt{n}\bar{X}}{S}$ 的分布

（2021 年期末最后一题）设总体 $X \sim N(0,1)$，$X_1, X_2, X_3 \ldots X_n$ 是来自总体的简单随机样本，$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$，$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$，求 $\frac{\sqrt{n}\bar{X}}{S}$ 的分布

Solution

由定理 1 可得：$\bar{X} \sim N(0, \frac{\sigma^2}{n})$

由定理 2 可得：$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

且 $\bar{X}$ 与 S^2 相互独立

构造 t 分布：$\dfrac{\frac{\bar{X}}{\sqrt{\frac{\sigma^2}{n}}}}{\sqrt{\frac{\frac{(n-1)S^2}{\sigma^2}}{n-1}}}$

所以化简可得，$\frac{\sqrt{n}\bar{X}}{S} \sim t(n-1)$

- 注意自由度和独立

练习

设总体 $X \sim N(\mu, \sigma^2)$，$X_1, X_2, X_3 \ldots X_n$ 是来自总体的简单随机样本，$\bar{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$，$S^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2$，求 $\sqrt{\frac{n}{n+1}}\frac{X_{n+1}-\bar{X}}{S}$ 的分布

设总体 $X \sim N(\mu, \sigma^2)$，$X_1, X_2, X_3 ... X_n$ 是来自总体的简单随机样本，

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$，$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$，求 $\sqrt{\frac{n}{n+1}}\frac{X_{n+1}-\bar{X}}{S}$ 的分布

Solution

由于简单随机抽样，样本与总体具有同分布，且根据定理 1 可知:

$X_{n+1} \sim N(\mu, \sigma^2)$，$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

随机样本及均值彼此独立，所以: $X_{n+1} - \bar{X} \sim N(0, \frac{n+1}{n}\sigma^2)$

由定理 2 可得: $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

构造 t 分布: $\dfrac{\frac{X_{n+1}-\bar{X}}{\sqrt{\frac{n+1}{n}\sigma^2}}}{\sqrt{\frac{\frac{(n-1)S^2}{\sigma^2}}{n-1}}}$

所以化简可得，$t(n-1) \sim \sqrt{\frac{n}{n+1}}\frac{X_{n+1}-\bar{X}}{S}$

练习

设总体 $X \sim N(\mu, \sigma^2)$，$X_1, X_2, ..., X_{2n}$ 是来自总体的简单随机样本，其样本均值为 $\bar{X} = \frac{1}{2n}\sum_{i=1}^{2n} X_i$，求 $Y = \sum_{i=1}^{n}(X_i + X_{i+1} - 2\bar{X})^2$ 的数学期望

设总体 $X \sim N(\mu, \sigma^2)$，$X_1, X_2, ..., X_{2n}$ 是来自总体的简单随机样本，其样本均值为 $\bar{X} = \frac{1}{2n}\sum\limits_{i=1}^{2n} X_i$，求 $Y = \sum\limits_{i=1}^{n}(X_i + X_{i+1} - 2\bar{X})^2$ 的数学期望

Solution

令 $y_i = X_i + X_{i+1}$，则 y_i 可以看做来自 $N(2\mu, 2\sigma^2)$ 的样本，所以可得

$Y = \sum\limits_{i=1}^{n}(X_i + X_{i+1} - 2\bar{X})^2 = \sum\limits_{i=1}^{n}(y_i - \bar{y})^2$ 服从正态分布 $N(0, 2\sigma^2)$

$\frac{Y}{2\sigma^2} \sim \chi^2(n-1)$

所以，$E(Y) = 2(n-1)\sigma^2$

- 注意自由度（样本均值已知，多了一个限制条件，$df = n - 1$）

练习

设总体 $X \sim N(0,1)$，$(X_1, X_2, .., X_{2n})$ 为 X 的简单随机样本，求下列统计量服从的分布：

(1) $T_1 = \frac{1}{2}\sum\limits_{i=1}^{2n} X_i^2 + \sum\limits_{i=1}^{n} X_{2i-1}X_{2i}$

(2) $T_2 = \frac{\sqrt{2n-1}X_1}{\sqrt{\sum\limits_{i=2}^{2n} X_i^2}}$

(3) $T_3 = \frac{(2n-3)\sum\limits_{i=1}^{3} X_i^2}{3\sum\limits_{i=4}^{2n} X_i^2}$

设总体 $X \sim N(0,1)$，$(X_1, X_2, .., X_{2n})$ 为 X 的简单随机样本，求下列统计量服从的分布：

(1) $T_1 = \frac{1}{2}\sum\limits_{i=1}^{2n} X_i^2 + \sum\limits_{i=1}^{n} X_{2i-1}X_{2i}$

Solution

$$\begin{aligned} T_1 &= \frac{1}{2}(X_1^2 + X_2^2 + ... + X_{2n-1}^2 + X_{2n}^2) + X_1X_2 + X_3X_4 + ... + X_{2n-1}X_{2n} \\ &= \frac{1}{2}(X_1 + X_2)^2 + \frac{1}{2}(X_3 + X_4)^2 + ... + \frac{1}{2}(X_{2n-1} + X_{2n})^2 \\ &= \sum_{i=1}^{n}(\frac{X_{2i-1} + X_{2i}}{\sqrt{2}})^2 \end{aligned}$$

而 $Y_i = \frac{X_{2i-1}+X_{2i}}{\sqrt{2}} \sim N(0,1)$，且相互独立，故 $T_1 = \sum\limits_{i=1}^{n} Y_1^2 \sim \chi^2(n)$

练习

设总体 $X \sim N(0,1)$，$(X_1, X_2, .., X_{2n})$ 为 X 的简单随机样本，求下列统计量服从的分布：

（2） $T_2 = \frac{\sqrt{2n-1}X_1}{\sqrt{\sum\limits_{i=2}^{2n} X_i^2}}$

Solution

$$T_2 = \frac{X_1}{\sqrt{\frac{\sum\limits_{i=1}^{2n} X_i^2}{2n-1}}} \sim t(2n-1)$$

练习

设总体 $X \sim N(0,1)$，$(X_1, X_2, .., X_{2n})$ 为 X 的简单随机样本，求下列统计量服从的分布：

(3) $T_3 = \dfrac{(2n-3)\sum\limits_{i=1}^{3} X_i^2}{3\sum\limits_{i=4}^{2n} X_i^2}$

Solution

由于 $\sum\limits_{i=1}^{3} X_i^2 \sim \chi^2(3)$，$\sum\limits_{i=4}^{2n} X_i^2 \sim \chi^2(2n-3)$

$T_3 = \dfrac{\frac{\sum_{i=1}^{3} X_i^2}{3}}{\frac{\sum_{i=4}^{2n} X_i^2}{2n-3}}$ 所以，$T_3 \sim F(3, 2n-3)$

Tips

- 有平方，优先考虑 χ^2 分布
- 有根号，形式较复杂。优先考虑 t 分布
- 有分式，有平方。优先考虑 F 分布
- 若所求统计量形式较复杂，先根据题目化简，得到每个部分属于什么分布
- 最后一定要检查一下自由度是否匹配

上 α 分位点的应用

设从两个方差相等的独立正态总体中分别抽取容量为 15，20 **的样本，样本方差分别为** S_1^2，S_2^2。**试求** $P(S_1^2/S_2^2 > 2)$

设从两个方差相等的独立正态总体中分别抽取容量为 15，20 的样本，样本方差分别为 S_1^2，S_2^2。试求 $P(S_1^2/S_2^2 > 2)$

Solution

假设正态总体样本方差为 σ^2，则有：

$\frac{14S_1^2}{\sigma^2} \sim \chi^2(14)$，$\frac{19S_2^2}{\sigma^2} \sim \chi^2(19)$

$F = \frac{S_1^2}{S_2^2} \sim F(14, 19)$

$P(S_1^2/S_2^2 > 2) = P(F > F_2(14, 19)) = 0.0798$

- 用 Matlab 求解：$1 - fcdf(2, 14, 19)$。理解题即可，求解不做要求。

练习

设 $X_1, X_2, ..., X_{17}$ 是来自总体 $N(\mu, \sigma^2)$ 的简单随机样本，$\bar{X}$ 与 S^2 分别是样本均值与样本方差。求 k，使得 $P(\bar{X} > \mu + ks) = 0.95$

设 $X_1, X_2, ..., X_{17}$ 是来自总体 $N(\mu, \sigma^2)$ 的简单随机样本，$\bar{X}$ 与 S^2 分别是样本均值与样本方差。求 k，使得 $P(\bar{X} > \mu + kS) = 0.95$

Solution

化简可得：$P(\bar{X} > \mu + kS) = P(\frac{\bar{X}-\mu}{S} > k)$

因为在正态总体下，有 $\frac{(\bar{X}-\mu)}{S/\sqrt{n}} \sim t(n-1)$（定理 3）

所以，$P(\bar{X} > \mu + kS) = P(\frac{\bar{X}-\mu}{S} > k) = P(\frac{(\bar{X}-\mu)}{S/\sqrt{n}} > k\sqrt{n}) = 0.95$

得：$P(t > k\sqrt{n}) = 0.95$

$k\sqrt{n} = t_{0.95}(n-1)$，已知 n=17，查表得

$t_{0.95}(16) = -t_{0.05}(16) = -1.7459$

所以 $k = \frac{-1.7459}{\sqrt{17}} = -0.4234$

Reference

- 王翠香, 褚宝增. “概率统计.”
- 茆诗松, 程依明, 濮晓龙. “概率论与数理统计 (第 3 版).”
- 茆诗松, 程依明, 濮晓龙. “概率论与数理统计习题与解答 (第 2 版).” (2012).
- 汤家凤. “考研接力题典 1800——数学一.” (2022).
- 陈强. “计量经济学及 Stata 应用.” (2015).
- 李林. “李林考研数学系列概率论与数理统计辅导讲义.” (2021).
- Jeffery M. Wooldridge. 2019. Introductory Econometrics: A Modern Approach. La Jolla, CA: South-Western College Publishing.
- Bruce Hansen. 2021. Probability and Statistics for Economists. Princeton, NJ: Princeton University Press.
- https://www.zhihu.com/question/276627271/answer/1693881405
- https://blog.csdn.net/xq151750111/article/details/120439438

Thank You

November 18, 2023