

单样本均值 统计量的分布

设计与制作：刘琼荔



引例

现在城市里买车的人越来越多了，导致交通拥堵的问题越来越严重。有没有办法能合理地估计明年城市里小轿车增长的比例 p 呢？



从该城市的 n 人中可以统计今年新买车的人数 v ，计算频率 v/n ，这个数值就能估计明年车辆数增长的比例！

大数定律： $\frac{v}{n} \xrightarrow[n \rightarrow \infty]{P} p$

等价于

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{v}{n} - p \right| < c \right\} = 1$$

可靠度

估计精度



引例

首先，进行如下的数学描述：

$$X_i = \begin{cases} 1, & \text{今年抽取的}n\text{人中，第}i\text{人新买了车} \\ 0, & \text{今年抽取的}n\text{人中，第}i\text{人没有买车} \end{cases} \quad i = 1, 2, \dots, n$$

$$X_1, X_2, \dots, X_n \sim B(1, p)$$

其中 p 表示今年车辆数增长的比例。

显然
$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{v}{n}$$

表达式 (1) 转换为：
$$\lim_{n \rightarrow \infty} P \left\{ \left| \bar{X} - p \right| < c \right\} = 100\% \quad (2)$$



引例

已知95%和样本容量 n ，可以确定精度 c ，使得表达式 (3) 满足。

$$P\{|\bar{X} - p| < c\} = 95\% \quad (3)$$

下面，我们就来学习样本均值 \bar{X} 的分布。



定理1 当已知总体方差 σ^2 时，如果样本 X_1, X_2, \dots, X_n 来自期望为 μ 和方差为 σ^2 的总体 X ，且 n 充分大，则

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{or} \quad \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0, 1) \quad (\text{近似})$$

注意：假设样本 X_1, X_2, \dots, X_n 来自期望为 μ 和方差为 σ^2 的正态总体 X ，则样本均值 \bar{X} 精确地服从正态分布。



例 假设样本 X_1, X_2, \dots, X_n 来自参数为 μ 和 0.25 的正态总体 X , 如果要以至少 99.7% 的概率使得 $|\bar{X} - \mu| < 0.1$, 问样本容量 n 应取多大?

根据题意,

$$\begin{aligned} P\{|\bar{X} - \mu| < 0.1\} &= P\left\{\frac{\sqrt{n}|\bar{X} - \mu|}{0.5} < \frac{0.1}{0.5}\sqrt{n}\right\} \\ &= 2\Phi(0.2\sqrt{n}) - 1 \geq 0.997 \end{aligned}$$

即 $\Phi(0.2\sqrt{n}) \geq 0.9985$

由查表知, $0.2\sqrt{n} \geq 2.96 \Rightarrow n \geq 220$



进一步地问，如果方差 σ^2 是未知的，定理1还成立吗？

请看下面的定理2.



定理2 如果样本 X_1, X_2, \dots, X_n 来自均值为 μ 和方差为 σ 的正态总体 X , \bar{X} 和 S 是样本均值和样本方差, 则

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

称 T 为 t 统计量。

t 分布具有什么特征呢?



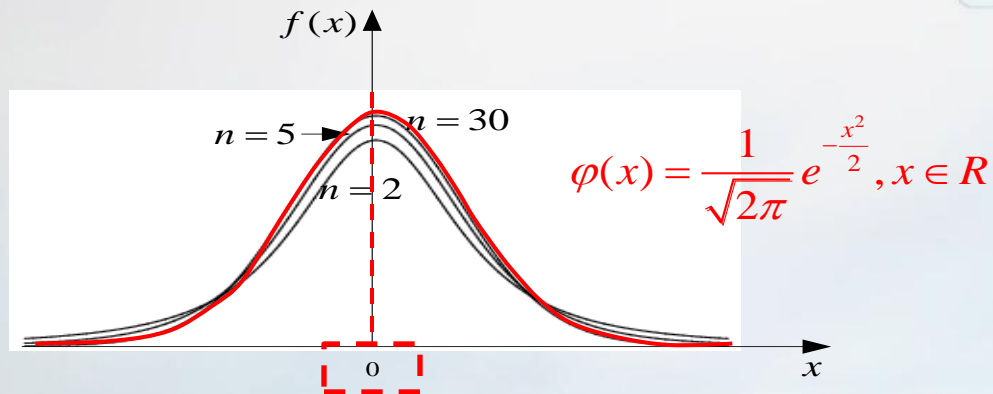
参数为 n 的 t 分布的密度函数为

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

记为 $T \sim t(n)$.



T分布具有以下特征



$$ET = 0 \quad (n > 1)$$

$$DT = \frac{n}{n-2} \quad (n > 2)$$

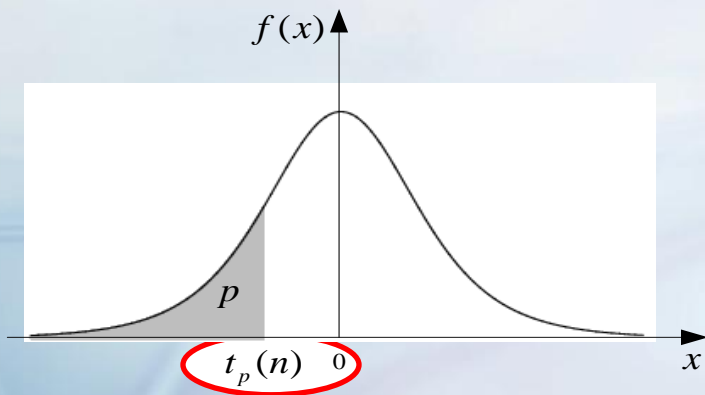
当 $n=1$ 时, $f_T(x) = \frac{1}{\pi} \frac{1}{1+x^2}, x \in R$ (柯西分布)



如何利用t分布计算概率呢？

注意：类似标准正态分布，计算概率也需要查t分布表。

$t_p(n)$ ——表示t分布的左侧 p 分位数，几何特征如图所示。



t 分布表

n	p						
	0.600	0.700	0.800	0.900	0.950	0.975	0.985
1	0.324 9	0.726 5	1.376 4	3.077 7	6.313 8	12.706 2	21.204 9
2	0.288 7	0.617 2	1.060 7	1.885 6	2.920 0	4.302 7	5.642 8
3	0.276 7	0.584 4	0.978 5	1.637 7	2.353 4	3.182 4	3.896 0
4	0.270 7	0.568 6	0.941 0	1.533 2	2.131 8	2.776 4	3.297 6
5	0.267 2	0.559 4	0.919 5	1.475 9	2.015 0	2.570 6	3.002 9
6	0.264 8	0.553 4	0.905 7	1.439 8	1.943 2	2.446 9	2.828 9
7	0.263 2	0.549 1	0.896 0	1.414 9	1.894 6	2.364 6	2.714 6
8	0.261 9	0.545 9	0.888 9	1.396 8	1.859 5	2.306 0	2.633 8
9	0.261 0	0.543 5	0.883 4	1.383 0	1.833 1	2.262 2	2.573 8
10	0.260 2	0.541 5	0.879 1	1.372 2	1.812 5	2.228 1	2.527 5
11	0.259 6	0.539 9	0.875 5	1.363 4	1.795 9	2.201 0	2.490 7
12	0.259 0	0.538 6	0.872 6	1.356 2	1.782 3	2.178 8	2.460 7
13	0.258 6	0.537 5	0.870 2	1.350 2	1.770 9	2.160 4	2.435 8

$$t_{0.95}(12)=1.7823$$

$$t_{0.05}(12)=?$$

$$=-t_{0.95}(12)$$

$$=-1.7823$$

$$t_{0.95}(45)=?$$

$$\approx 1.65$$

在后续的统计方法学习中，我们还会学习 t 分布的计算。



小结

- (1) 样本均值的分布
- (2) t 分布的密度函数形式
- (3) t 分布的特征

