课题：样本及抽样分布

讲师:Seven老师

2018-03-26（周一）上课时间8:30-10:00

Tanzhou EDU

Artificial Intelligence

AI学院 VIP

课程内容

- 大数定理
- 中心极限定理
- 随机样本
- 直方图和箱线图
- 抽样分布

大数定律的概念

概率论中用来阐明大量随机现象平均结果的稳定性的一系列定理，称为大数定律（law of large number)

本章将介绍三个大数定律：

(1) 切比雪夫大数定律、

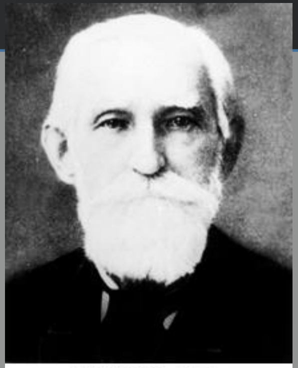
(2) 贝努里大数定律

(3) 辛钦大数定律。

它们之间既有区别也有联系。

三、大数定律

1、定理一（切比雪夫定理的特殊情况）



切比雪夫，П. Л.

切比雪夫

设随机变量$X_1, X_2, \cdots, X_n, \cdots$相互独立，且具有相同的数学期望和方差：

$$E(X_k)=\mu, D(X_k)=\sigma^2 (k=1,2,\cdots).$$

做前 n 个随机变量的算术平均 $\bar{X}=\frac{1}{n}\sum_{k=1}^{n} X_k$

则对任意的$\varepsilon>0$，有

$$\lim_{n\to\infty} P\{|\bar{X}-\mu|<\varepsilon\}$$

$$=\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{i=1}^{n} X_i-\mu|<\varepsilon\}=1$$

大数定理

证 由于

$$E\left(\frac{1}{n}\sum_{k=1}^{n}X_k\right)=\frac{1}{n}\sum_{k=1}^{n}E(X_k)=\frac{1}{n}\cdot n\mu=\mu$$

$$D\left(\frac{1}{n}\sum_{k=1}^{n}X_k\right)=\frac{1}{n^2}\sum_{k=1}^{n}D(X_k)=\frac{1}{n^2}\cdot n\sigma^2=\frac{\sigma^2}{n}$$

由切比雪夫不等式

$$P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\mu\right|<\varepsilon\right\}\geq 1-\frac{\sigma^2/n}{\varepsilon^2}$$

上式中令 $n\to\infty$ 得

$$\lim_{n\to\infty}P\{|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu|<\varepsilon\}=1$$

说明

1、定理中$\{|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu|<\varepsilon\}$是指一个随机事件，当$n\to\infty$时，这个事件的概率趋于1.

2、定理以数学形式证明了随机变量$X_1,\cdots X_n$的算术平均$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$接近数学期望E（$X_k$）$=\mu$（$k=1,2\cdots,n$），这种接近说明其具有的稳定性.

这种稳定性的含义说明算术平均值是依概率收敛的意义下逼近某一常数.

2、定理二（贝努里大数定律）

设 n_A 是n次独立重复试验中事件A发生的次数，p是事件A在每次试验中发生的概率，则对于任意正数$\varepsilon > 0$，有

$$\lim_{n\to\infty} P\{|\frac{n_A}{n} - p| < \varepsilon\} = 1$$

或

$$\lim_{n\to\infty} P\{|\frac{n_A}{n} - p| \geq \varepsilon\} = 0$$

此定理说明了频率的稳定性。

证：令 $X_k = \begin{cases} 0, & \text{在第}k\text{次试验中}A\text{不发生} \\ 1, & \text{在第}k\text{次试验中}A\text{发生} \end{cases}, \quad k = 1,2,\cdots,n$

则 $n_A = \sum_{k=1}^{n} X_k$，且 $X_1,\cdots,X_n$ 相互独立同服从于(0 – 1)分布

故 $EX_k = p$，$DX_k = p(1-p)$，$k = 1,2,\cdots,n,\cdots$

由定理一有

$$\lim_{n->\infty} P\{|\frac{1}{n}\sum_{i=1}^{n} X_i - p| < \varepsilon\} = 1,$$

即 $$\lim_{n->\infty} P\{|\frac{n_A}{n} - p| < \varepsilon\} = 1。$$

贝努里大数定律的重要意义：

(1) 从理论上证明了频率具有稳定性。

(2) 提供了通过试验来确定事件概率的方法：

$$\frac{n_A}{n} \approx p = P(A)$$

这种方法是参数估计的重要理论基础。

（3）是“小概率原理”的理论基础。

小概率原理：实际中概率很小的随机事件在个别试验中几乎是不可能发生的。

下面给出的独立同分布下的大数定律，
不要求随机变量的方差存在.

3、定理三（辛钦大数定律）

设随机变量序列$X_1, X_2, \cdots$ 相互独立，服从同一分布，具有数学期$E(X_i)=\mu$, i=1,2,⋯， 则对于任意正数ε，有

$$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| < \varepsilon\} = 1$$

注 1、辛钦大数定律为寻找随机变量的期望值提供了一条实际可行的途径.

当n充分大时有 $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i\approx\mu$

2、伯努利大数定律是辛钦定理的特殊情况.

3、辛钦定理具有广泛的适用性.

要估计某地区的平均亩产量，要收割某些有代表性块，例如n块地．计算其平均亩产量，则当n较大时，可用它作为整个地区平均亩产量的一个估计.

三、小结

大数定律

伯努利大数定律 $\lim\limits_{n\to\infty} P\{|\frac{n_A}{n} - p| < \varepsilon\} = 1$ $n_A \sim b(n, p)$

切比雪夫大数定律 $\lim\limits_{n\to\infty} P\{|\frac{1}{n}\sum\limits_{i=1}^{n} X_i - \mu| < \varepsilon\} = 1$ $\begin{cases} E(X_k) = \mu \\ D(X_k) = \sigma^2 \end{cases}$

辛钦大数定律 $\lim\limits_{n\to\infty} P\{|\frac{1}{n}\sum\limits_{i=1}^{n} X_i - \mu| < \varepsilon\} = 1$ $E(X_k) = \mu$

大数定律以严格的数学形式表达了随机现象最根本的性质之一：

平均结果的稳定性

第二节　中心极限定理

- 中心极限定理的背景
- 中心极限定理的定义
- 中心极限定理
- 小结

一、中心极限定理的客观背景

在实际问题中许多随机变量是由相互独立随机因素的综合（或和)影响所形成的.

例如：炮弹射击的落点与目标的偏差，就受着许多随机因素（如瞄准，空气阻力，炮弹或炮身结构等）综合影响的.每个随机因素的对弹着点（随机变量和）所起的作用都是很小的.那么弹着点服从怎样分布 ？

如果一个随机变量是由大量相互独立的随机因素的综合影响所造成，而每一个别因素对这种综合影响中所起的作用不大，则这种随机变量一般都服从或近似服从正态分布.

二、中心极限定理定义

概率论中有关论证独立随机变量的和的极限分布是正态分布的一系列定理称为中心极限定理。

由于无穷个随机变量之和可能趋于∞，故我们不研究n个随机变量之和本身而考虑它的标准化的随机变量，即：

$$Y_n=\frac{\sum_{k=1}^{n}X_k-E(\sum_{k=1}^{n}X_k)}{\sqrt{D(\sum_{k=1}^{n}X_k)}}$$

讨论Y_n的极限分布是否为标准正态分布

1、定理四（独立同分布下的中心极限定理）

设随机变量$X_1, X_2, \cdots X_n, \cdots$相互独立，服从同一分布，且具有数学期望和方差：$E(X_k)=\mu, D(X_k)=\sigma^2$ $(k=1,2,\cdots)$，则随机变量之和$\sum_{k=1}^{n} X_k$的标准化变量

$$Y_n = \frac{\sum_{k=1}^{n} X_k - E(\sum_{k=1}^{n} X_k)}{\sqrt{D(\sum_{k=1}^{n} X_k)}} = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma}$$

的分布函数$F_n(x)$对于任意x满足

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P\left\{\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \le x\right\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = \Phi(x)$$

注 1、定理表明，独立同分布的随机变量之和$\sum_{k=1}^{n} X_k$，当n充分大时，随机变量之和与其标准化变量分别有

$$\sum_{k=1}^{n} X_k \overset{\text{近似地}}{\sim} N(n\mu, n\sigma^2)\ ;\quad \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \overset{\text{近似地}}{\sim} N(0,1).$$

2、独立同分布中心极限定理的另一种形式可写为

$$\overline{X} \overset{\text{近似地}}{\sim} N(\mu, \sigma^2/n)\ \text{或}\ \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \overset{\text{近似地}}{\sim} N(0,1)$$

其中$\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$

3、虽然在一般情况下，我们很难求出$\sum_{k=1}^{n} X_k$的分布的确切形式，但当n很大时，可以求出近似分布.

2、定理五（李雅普诺夫(Liapounov)定理）

设随机变量$X_1, X_2, \cdots, X_n \cdots$相互独立，它们具有数学期望和方差：

$$E(X_k) = \mu_k,\quad D(X_k) = \sigma_k^2, (k = 1,2,\cdots)$$

记 $$B_n^2 = \sum_{k=1}^{n} \sigma_k^2$$

若存在正数δ，使得当$n \to \infty$时，

李雅普诺夫条件

$$\frac{1}{B_n^{2+\delta}} \sum_{k=1}^{n} E\left\{\left|X_k - \mu_k\right|^{2+\delta}\right\} \to 0$$

则随机变量之和$\sum_{k=1}^{n} X_k$的标准化变量：

$$Z_n = \frac{\sum_{k=1}^{n} X_k - E(\sum_{k=1}^{n} X_k)}{\sqrt{D(\sum_{k=1}^{n} X_k)}} = \frac{\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} \mu_k}{B_n}$$

的分布函数$F_n(x)$对于任意x，满足

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P\left\{ \frac{\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} \mu_k}{B_n} \le x \right\}$$

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

请注意：

1、定理中随机变量之和$\sum_{k=1}^{n} X_k$及其标准化变量Z_n在n很大时,分别近似服从

$$\sum_{k=1}^{n} X_k \overset{近似地}{\sim} N(\sum_{k=1}^{n}\mu_k, B_n^2); \quad Z_n \overset{近似地}{\sim} N(0,1)$$

2、随机变量X_k无论服从什么分布，只要满足定理条件，随即变量之和$\sum_{k=1}^{n} X_k$，当n很大时，就近似服从正态分布，这就是为什么正态分布在概率论中所占的重要地位的一个基本原因.

3、定理六(棣莫佛－拉普拉斯 (De Laplace定理)

设随机变量η_n(n=1,2,……)服从参数$n,p(0<p<1)$的二项分布，则对任意x，有

$$\lim_{n\to\infty} P\{\frac{\eta_n - np}{\sqrt{np(1-p)}} \le x\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(x)$$

证 由第四章知识知可将η_n分解成为n个相互独立、服从同一$(0-1)$分布的诸随机变量$X_1, X_2, \cdots X_n$之和，即有

$$\eta_n = \sum_{k=1}^{n} X_k$$

其中$X_k (k=1,2,\cdots,n)$的分布律为

$$P\{X_k = i\} = p^i (1-p)^{1-i}, i = 0,1$$

由于$E(X_k)=p, D(X_k)=p(1-p) \quad k=1,2,\cdots,n),$

由定理4得

$$\lim_{n\to\infty} P\{\frac{\eta_n - np}{\sqrt{np(1-p)}} \le x\} = \lim_{n\to\infty} P\{\frac{\sum_{k=1}^{n} X_k - np}{\sqrt{np(1-p)}} \le x\}$$

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

定理表明，当n很大，$0<p<1$是一个定值时（或者说，$np(1-p)$也不太小时），二项变量η_n的分布近似正态分布$N(np,np(1-p))$.

即 $\eta_n \overset{近似地}{\sim} N(np,np(1-p))$

四、小结

中心极限定理

- 独立同分布中心极限定理：$\begin{cases} E(X_k)=\mu,\ D(X_k)=\sigma^2 \\ \Rightarrow \sum\limits_{k=1}^{n} X_k \overset{\text{近似地}}{\sim} N(n\mu, n\sigma^2) \end{cases}$
- 棣莫弗－拉普拉斯中心极限定理：$\begin{cases} \eta_n \sim N(n,p) \\ \Rightarrow \eta_n \overset{\text{近似地}}{\sim} N(np, np(1-p)) \end{cases}$
- 李雅普诺夫中心极限定理：$\begin{cases} E(X_k)=\mu_k, D(x_k)=\sigma_k{}^2 \\ \Rightarrow \sum\limits_{k=1}^{n} X_k \overset{\text{近似地}}{\sim} N(\sum\limits_{k=1}^{n}\mu_k, B_n{}^2) \end{cases}$

注 随机变量$X_1, X_2, \cdots$是相互独立的

一、总体与样本

1、总体与个体

一个统计问题总有它明确的研究对象.

研究对象的全体称为总体，总体中每个成员称为个体

总体中所包含的个体的个数称为总体的容量.

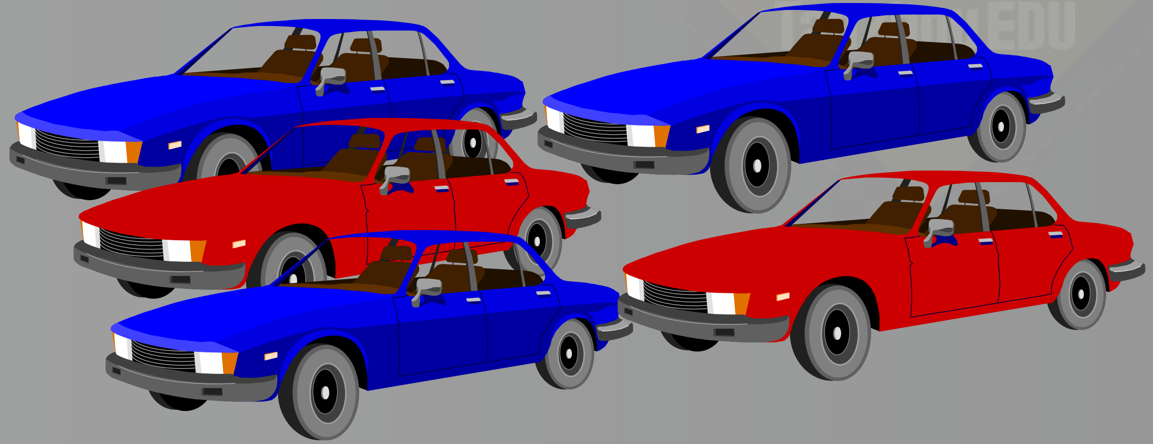
研究某批灯泡的质量

总体 $\begin{cases} \text{有限总体} \\ \text{无限总体} \end{cases}$

2、样本

总体中抽出若干个体而成的集体,称为样本。

样本中所含个体的个数，称为样本容量。



从国产轿车中抽5辆
进行耗油量试验
样本容量为5

抽到哪5辆是随机的

随机样本

注1：所谓样本就是n个与总体同分布的随机变量。

对总体X在相同的条件下，进行n次重复、独立观察，其结果依次记为X_1，X_2,···，X_n.

这样得到的随机变量$X_1,X_2,\cdots X_n$是来自总来自的一个随机样一个随机体随机变随机变量具有分布.

一旦取定一组样本X_1，··· ,X_n ,得到n个具体的数$(x_1,x_2,\cdots,x_n)$，称为样本的一次观察值，简称样本值 .

注2 最常用的一种抽样叫作"简单随机抽样"，其特点：

1. 代表性： $X_1,X_2,\cdots,X_n$中每一个与所考察的总体有相同的分布.

2. 独立性： $X_1,X_2,\cdots,X_n$是相互独立的随机变量.

随机样本

定义：

设X是具有分布函数F的随机变量，若$X_1, X_2, \cdots, X_n$是具有同一分布函数F的、相互独立的随机变量，则称$X_1, X_2, \cdots, X_n$为从分布函数F（或总体F、或总体X）得到的容量为n的简单随机样本，简称样本，它们的观察值$x_1, x_2, \cdots, x_n$称为样本值，又称为X的n个独立的观察值.

由简单随机抽样得到的样本称为简单随机样本，它可以用与总体独立同分布的n个相互独立的随机变量$X_1, X_2, \cdots, X_n$表示.

随机样本

若总体的分布函数为$F(x)$、概率密度函数为$f(x)$,则其简单随机样本的联合分布函数为

$$F^*(x_1,x_2,\cdots,x_n) = F(x_1)\,F(x_2)\cdots F(x_n)$$

其简单随机样本的联合概率密度函数为

$$f^*(x_1,x_2,\cdots,x_n) = f(x_1)\,f(x_2)\cdots f(x_n)$$

简单随机样本是应用中最常见的情形，今后，当说到“$X_1,X_2,\cdots,X_n$是取自某总体的样本”时，若不特别说明，就指简单随机样本.

一、统计量与经验分布函数

1. 统计量

不含任何未知参数的样本的函数称为统计量.
它是完全由样本决定的量.

设$X_1, X_2, \cdots, X_n$是来自总体X的一个样本，$g(X_1, X_2, \cdots, X_n)$是$X_1, X_2, \cdots, X_n$的函数，若g中不含未知参数，则$g(X_1, X_2, \cdots, X_n)$称是一个统计量.

例 设总体X服从两点分布$b(1,p)$，其中p是未知参数，$X_1,\cdots,X_5$是来自X的简单随机样本.试指出X_1+X_2, $\max\limits_{1\le i\le 5} X_i$, X_5+2p, $(X_5-X_1)^2$之中哪些是统计量,哪些不是统计量，为什么？

解：X_1+X_2, $\max\limits_{1\le i\le 5} X_i$, $(X_5-X_1)^2$都是统计量，但X_5+2p不是统计量（因为p是未知数）.

请注意：

（1）统计量是一个随机变量。

(2) 设$X_1, X_2, \cdots X_n$是来自总体X的一个样本，$x_1, x_2, \cdots x_n$是一个样本的观察值，则$f(x_1, x_2, \cdots x_n)$也是统计量$f(X_1, X_2, \cdots X_n)$的观察值.

抽样分布

几个常见统计量

样本平均值

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

它反映了总体均值的信息

样本方差

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

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

它反映了总体方差的信息

样本标准差

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

样本k阶原点矩

$$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \quad k=1,2,\cdots$$

它反映了总体k阶矩的信息

样本k阶中心矩

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

它反映了总体k阶中心矩的信息

统计量的观察值

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i; \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}; \qquad \alpha_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k \qquad k = 1,2,\cdots$$

$$b_k = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^k \qquad k = 1,2,\cdots$$

2. 经验分布函数

设$X_1, X_2, \cdots, X_n$是总体F的一个样本，用$s(x)$ $|x| < \infty$

表示$x_1, x_2, \cdots, x_n$中不大于x的随机变量的个数.

定义　经验分布函数为

$$F_n(x) = \frac{1}{n} s(x) \quad -\infty < x < \infty$$

例 设总体F具有一个样本值1,1,2，则经验分布函数

$F_3(x)$的观察值为

$$F_3(x) = \begin{cases} 0, & 若 x < 1 \\ \frac{2}{3}, & 若 1 \le x < 2 \\ 1, & 若 x \ge 2 \end{cases}$$

一般，设$x_1,x_2,\cdots,x_n$是总体的一个容量为n的样本值.将它们按大小次序排列如下：$x_{(1)}\le x_{(2)}\le\cdots\le x_{(n)}$

则经验分布函数$F_n(x)$的观察值为

$$F_n(x)=\begin{cases}0, & 若x<x_{(1)}\\ \dfrac{k}{n}, & 若x_{(k)}\le x<x_{(k+1)},\ (k=1,2,\cdots,n-1)\\ 1, & 若x\ge x_{(n)}\end{cases}$$

三、几个重要的抽样分布定理

设总体X的均值为μ，方差为σ^2，$X_1, X_2, \cdots, X_n$是来自总体的一个样本，则样本均值$\bar{X}$和样本方差S^2有

$$E(\bar{X}) = \mu,$$

$$D(\bar{X}) = \sigma^2 / n,$$

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

事实上 $E(s^2)=E\left[\frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2-n\overline{X}^2\right)\right]$

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

$$=\frac{1}{n-1}\left[\sum_{i=1}^{n}(\sigma^2+\mu^2)-n(\sigma^2/n+\mu^2)\right]=\sigma^2$$

当总体为正态分布时，给出几个重要的抽样分布定理.

定理 1（样本均值的分布）

设 $X_1, X_2, \cdots, X_n$ 是来自正态总体 $N(\mu,\sigma^2)$ 的样本，$\overline{X}$是样本均值，则有

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

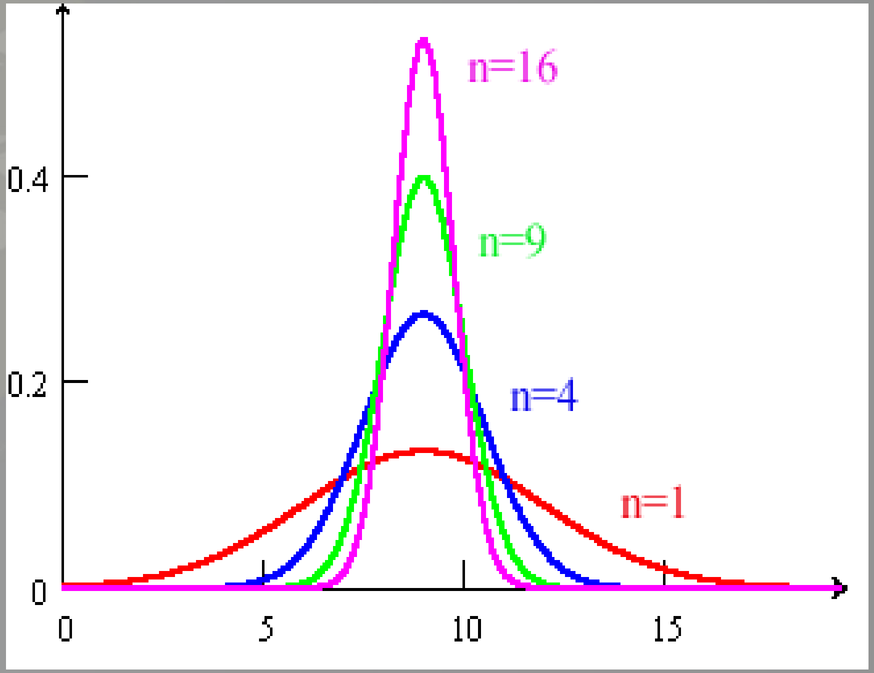
即 $$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

抽样分布

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

请注意：

在已知总体μ，σ^2时，可用本定理计算样本均值$\overline{X}$.



n取不同值时样本均值 $\overline{X}$ 的分布

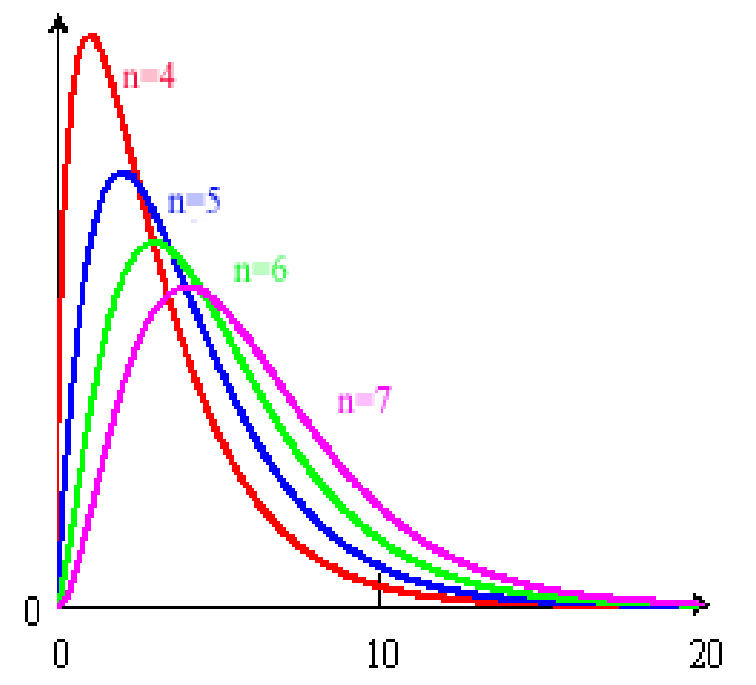
定理 2 (样本方差的分布)

设$X_1,X_2,\cdots,X_n$是来自正态总体 $N(\mu,\sigma^2)$的样本，$\overline{X}$和S^2分别为样本均值和样本方差，则有

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

(2) $\overline{X}$与S^2独立.

n取不同值时 $\dfrac{(n-1)S^2}{\sigma^2}$ 的分布



定理 3 (样本均值的分布)

设$X_1,X_2,\cdots,X_n$是取自正态总体 $N(\mu,\sigma^2)$

的样本, $\bar{X}$和S^2 分别为样本均值和样本方差,

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

证 由定理1、2,t分布的定义可得

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1), \quad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$ 且相互独立

则 $$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \Bigg/ \sqrt{\frac{(n-1)S^2}{\sigma^2}} \sim t(n-1)$$

在未知总体σ^2时,可用本定理计算样本均值$\bar{X}$.

抽样分布

定理 4 (两总体样本均值差、样本方差比的分布)

设$X \sim N(\mu_1,\sigma^2)$，$Y \sim N(\mu_2,\sigma^2)$，且X与Y独立，$X_1,X_2,\cdots,X_{n_1}$是来自X的样本，$Y_1,Y_2,\cdots,Y_{n_2}$是取自Y的样本，$\overline{X}$和$\overline{Y}$分别是这两个样本的样本均值，S_1^2和S_2^2分别是这两个样本的样本方差，则有

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

2、$$\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} \sim t(n_1+n_2-2)$$

六、小结

在这一节中我们学习了统计量的概念，几个重要的统计量及其分布，即抽样分布. 要求大家熟练地掌握它们.

课程内容

Tanzhou EDU

Thanks For Your Attention!

感谢观看!

Artificial Intelligence