计量经济学预备知识一

概率论和数理统计基础

孔伟杰

第一部分 概率和概率分布

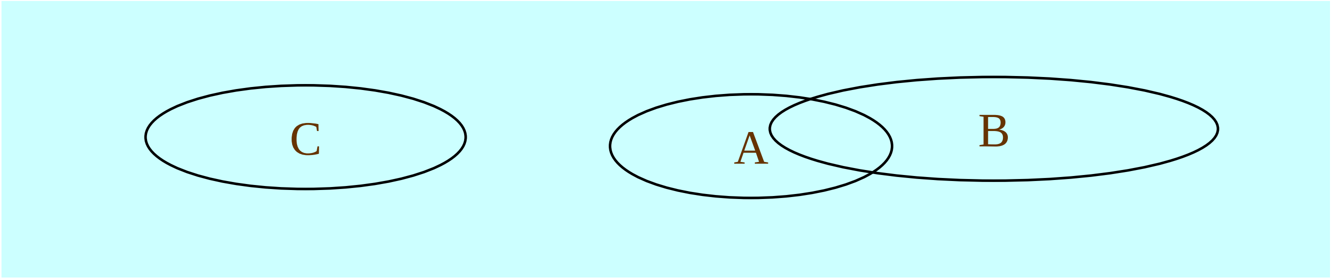
样本空间和随机事件

对客观现象的一次观察或者一次真正的科学实验统称为一个实验
随机试验满足下面的三个条件：
(1) 在相同的条件下，试验可以重复地进行；
(2) 试验的结果不止一种，而且事先可以确知试验的所有结果；
(3) 在进行试验前不能确定出现哪一个结果.
我们要研究随机实验的实验结果
在随机试验中，每一个可能出现的不可分解的最简单的结果称为随机试验的基本事件或样本点，用 ω 表示
由全体基本事件构成的集合称为基本事件空间或样本空间，记为 Ω

一个具体的人、一个具体的过程不能叫随机事件。
凡是能够预测的及其准确的不是随机事件。

概率和条件概率

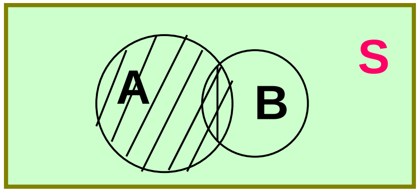
- 有关概率的概念
  样本空间、样本点、事件、样本、互斥事件、完备事件、概率等。



图：样本空间与事件

- 条件概率

$$P(B|A)=\frac{P(AB)}{P(A)} \qquad P(A)\neq 0$$



- 概率的性质
  （1） 0≤P(A)≤1，P(A)=0表明A为不可能事件，
  P(A)=1表明A为完备事件
  （2） 若A,B,C,...是完备事件集，则 P(A+B+C+...)=1
  （3） 若A,B,C,...是互斥事件，则 P(A+B+C+...)=
  P(A)+P(B)+P(C)+...

随机变量和概率分布的特征（1）

随机变量：随试验结果而变的量X为随机变量。

定义：随机变量X,对任意实数x,称函数

$F(x)=P(X\le x)$为X的概率分布函数，简称分布函数。

○ 定义：取值可数的随机变量为离散量

离散量的概率分布(分布律)

X	x_1	x_2	$\cdots$	x_i	$\cdots$
P	p_1	p_2	$\cdots$	p_i	$\cdots$

$$p_i\ge 0,\sum_{i=1}^{\infty}p_i=1$$

○ 定义：对于随机变量X的分布函数$F(x)$，若存在 非负的函数$f(x)$,使对于任意实数 x，有：

$$F(x)=\int_{\infty}^{x}f(t)dt$$

则称X为连续型随机变量，

其中 $f(x)$ 称为X的概率密度函数，简称**概率密度**。

随机变量和概率分布的特征（2）

3．期望值

- 离散随机变量X的期望值E(X)

$$E(X)=\sum_{x} xf(x)$$

- 连续随机变量X的期望值E(X)

$$E(X)=\int_{-\infty}^{+\infty} xf(x)dx$$

- 期望值E(X)的性质

(1) $E(C)=C$

(2) $E(CX)=CE(X)$

(3) $E(X+Y)=E(X)+E(Y)$ $E(X-Y)=E(X)-E(Y)$

$$E\left(\sum_{i=1}^{n} C_i x_i\right)=\sum_{i=1}^{n} C_i E(x_i)$$

(4) $E(xy)=E(x)E(y)$ X,Y独立

随机变量和概率分布的特征（3）

4．方差：若E(X)=μ，则X的方差定义为Var(X)=E[（X−μ）^{2}]

- 离散随机变量X的方差Var(X)

$$Var(X) = \sum_{x} (X-\mu)^2 f(x)$$

- 连续随机变量X的方差Var(X)

$$Var(X) = \int_{-\infty}^{+\infty} (X-\mu)^2 f(x)dx$$

- 方差Var(X)的性质

（1）$D(C) = 0$

（2）$D(aX) = a^2 D(X)$

（3）$D(aX+b) = a^2 D(X)$

（4）$D(X) = E(x^2) - E^2(X) = E(X - E(X))^2$

（5）$D(X \pm Y) = D(X) + D(Y)$ 若X, Y为独立随机变量

$D(X \pm Y) = D(X) + D(Y) \pm 2E[(X - E(X))(Y - E(Y))]$ 无条件成立

随机变量和概率分布的特征（4）

5．协方差：两随机变量X、Y的期望值分别为μ_x和μ_y，则X、Y的协方差定义为Cov(X, Y)=E[（X-μ_x）（Y-μ_y）]，或者表示为：

$$\delta_{xy} = \mu_{11} = E[(X - E(X))(Y - E(Y))]$$

$$D(X) = \delta_{XX}$$

$$D(Y) = \delta_{YY}$$

6．相关系数

对于随机变量X与Y，如果D（X）>0，D(Y)>0，则称

$$\frac{\sigma_{XY}}{\sqrt{D(X)}\sqrt{D(Y)}}$$

为X与Y的相关系数，记作ρ_{XY}（有时可简记为 ρ ）。

$|\rho| < 1$，当$|\rho| = 1$时，称X与Y完全相关：$P(X = aY + b) = 1$

完全相关 $\begin{cases} \text{正相关，当} \rho = 1 \text{时} (a = 1), \\ \text{负相关，当} \rho = -1 \text{时} (a = -1), \end{cases}$ 而当$\rho = 0$时，称X与Y不相关。

随机变量和概率分布的特征（5）

相关系数

以下五个命题是等价的：

① $\rho_{XY}=0$；

②cov（X，Y）=0；

③E（XY）=E（X）E（Y）；

④D（X+Y）=D（X）+D（Y）；

⑤D（X−Y）=D（X）+D（Y）.

7. 矩：k阶矩：$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k \quad (k=1,2,\cdots)$

k阶中心矩：$B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k \quad (k=1,2,\cdots)$

四个重要的概率分布之一：正态分布

1． 正态分布

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad -\infty<x<\infty$$

其中μ、σ为分布参数，分别是随机变量X的均值和方差。该分布記为：

$X \sim N(\mu, \sigma^2)$

将正态变量X变换成 $Z=(X-\mu)/\sigma$，则Z称为均值μ是0、方差σ为1的标准正态分布。该分布記为： $Z \sim N(0, 1)$

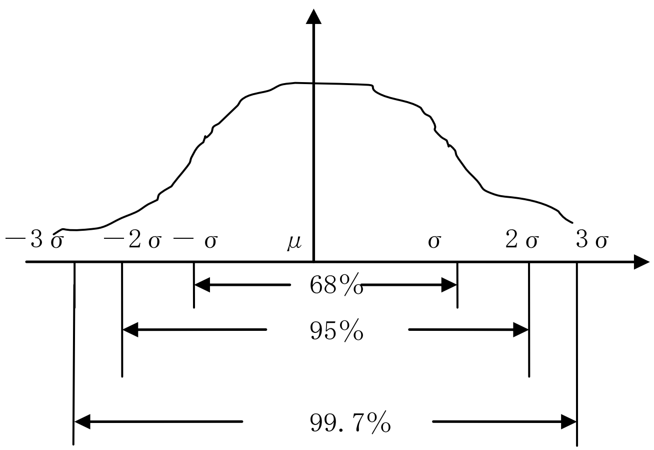


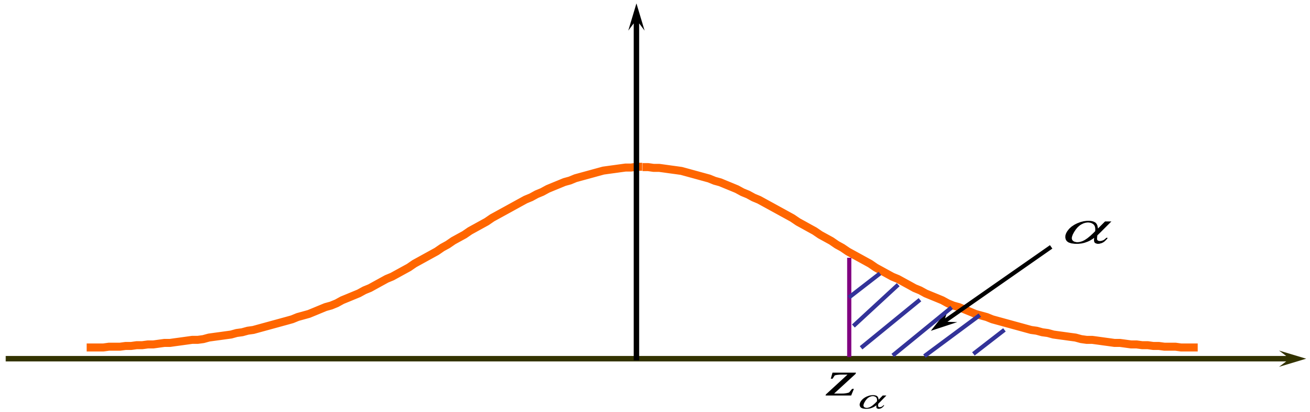
图2-1 正态分布的特征

☆ 正态分布关于其均值对称.

☆ 正态分布变量的线性函数也服从正态分布。

设$X \sim N(0,1)$,若Z_α满足条件$P\{X > Z_\alpha\} = \alpha, 0 < \alpha < 1$

则称点Z_α为标准正态分布的上α分位数。



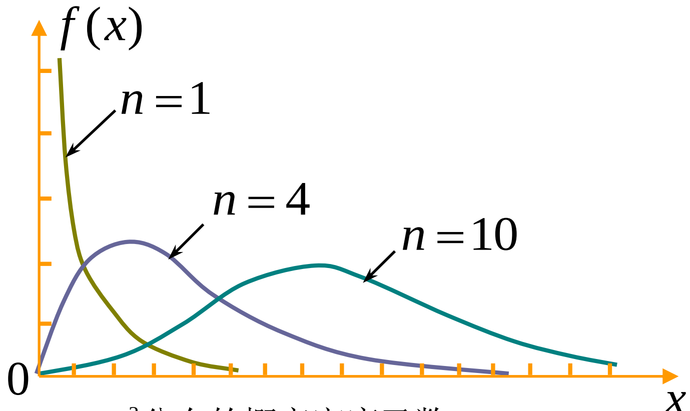
四个重要的概率分布之二：x^2分布

2. x^2分布

若Z_1、Z_2、…、Z_n是独立的标准正态变量$Z_i \sim N(0, 1)$，则其平方和

$$Z = \sum_{i=1}^{n} Z_i^2$$

服从自由度为n的X^2分布。記为$X^2(n)$。（卡方发布）



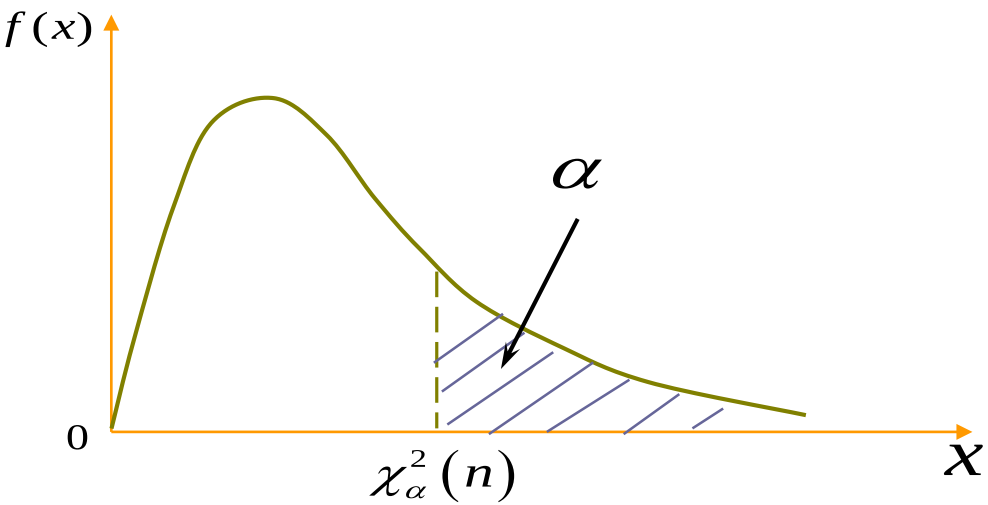
χ^2分布的概率密度函数

☆ x^2分布是一个偏斜分布.

☆ x^2分布的均值是n，方差是2n。

☆ 若Z_1、Z_2是自由度分别为n_1和n_2的x^2变量，则Z_1+Z_2为n_1+n_2的x^2变量。

对给定的概率$\alpha, 0<\alpha<1$,称满足条件$\int_{\chi_\alpha^2(n)}^{\infty} f_n(y)dy=\alpha$的点$\chi_\alpha^2(n)$为$\chi^2(n)$分布的上$\alpha$分位数,上$\alpha$分位数$\chi_\alpha^2(n)$的值可查$\chi^2$分布表



χ^2分布的分位数

四个重要的概率分布之三：t分布

3. t分布

若Z_1是一个标准正态变量，Z_2是n个自由度的x^2变量，则

$$t=\frac{Z_1}{\sqrt{Z_2/n}}$$

服从n个自由度的t变量，記为t(n)或t_n。

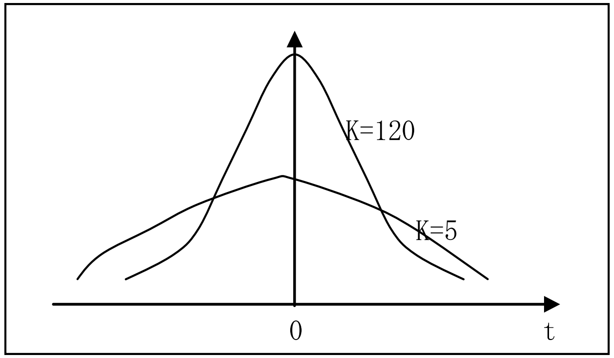
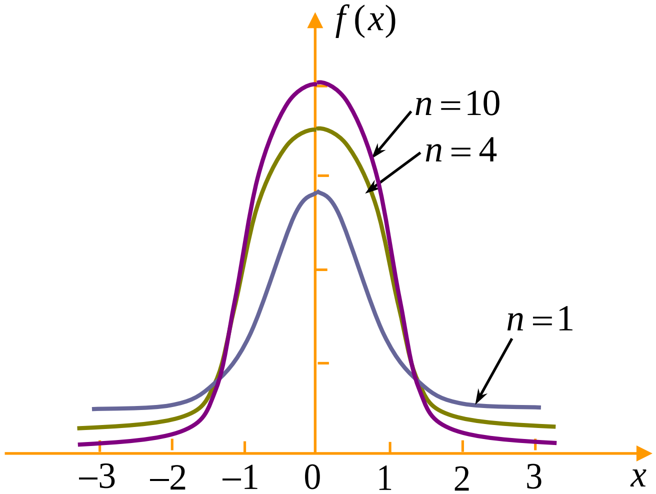


图2-3　t分布的特征

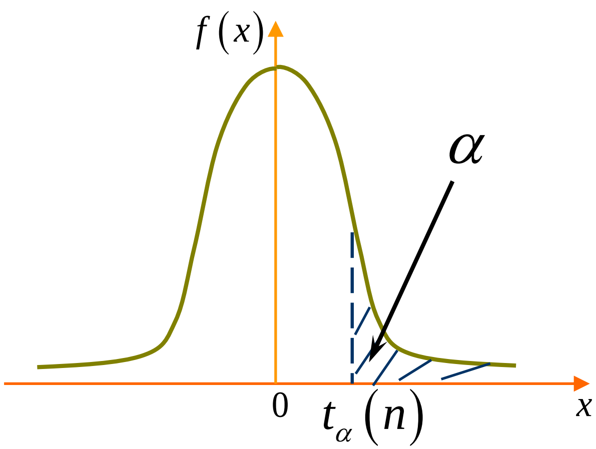
☆　t分布与正态分布一样是对称的，但要平一些。当自由度无限大时，t分布成为标准正态分布。

☆　t分布的均值是0，方差是n/（n-2）。

对给定的α，$0<\alpha<1$，称满足条件$\int_{t_\alpha(n)}^{\infty} f(t,n)\,dt=\alpha$的点$t_\alpha(n)$为$t(n)$分布的上$\alpha$分位数。$t$分布的上$\alpha$分位数可查$t$分布表



t分布的密度函数



t分布的分位数

四个重要的概率分布之四：F分布

4. F分布

若Z_1、Z_2是独立分布的自由度分别为K_1、K_2的x^2变量，则

$$F = \frac{Z_1 / K_1}{Z_2 / K_2}$$

服从自由度为K_1、K_2的F分布，記为$F(K_1, K_2)$。

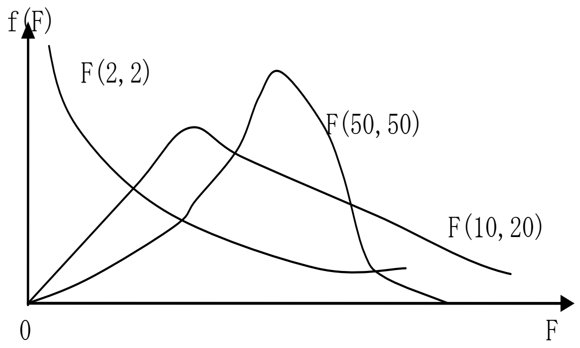


图2-4　F分布的特征

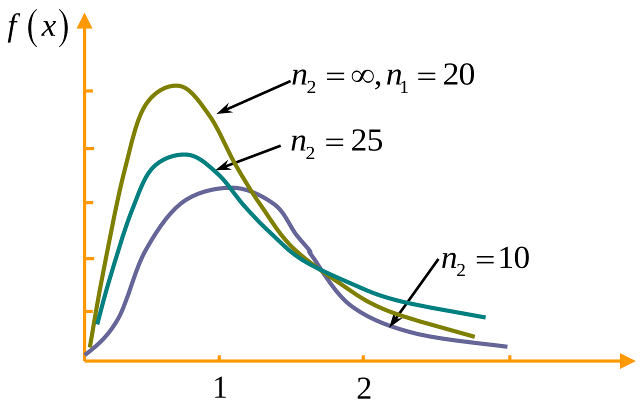
☆　F分布与x^2分布一样是向右偏斜的，当K_1、K_2的增大时，F分布近似正态分布。

☆　F分布变量的均值、方差分别是：

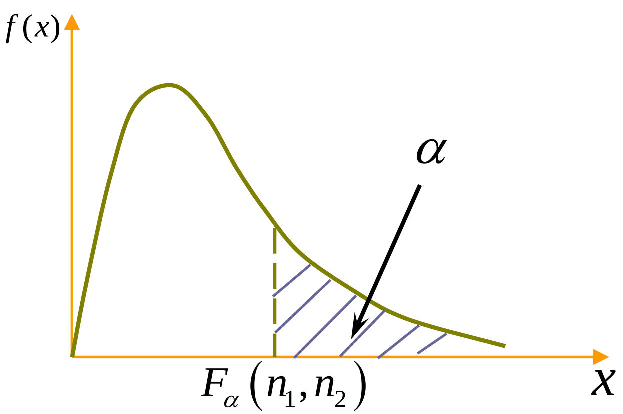
$$\frac{K_2}{K_2 - 2}, \quad \frac{2K_2^2(K_1 + K_2 - 2)}{K_1(K_2 - 2)^2(K_2 - 4)}$$

对于给定的α，$0<\alpha<1$，称满足条件$\int_{F_\alpha(n_1,n_2)}^{\infty} f(x;n_1,n_2)dx=\alpha$的点$F_\alpha(n_1,n_2)$为$F(n_1,n_2)$分布的上$\alpha$分位数。$F_\alpha(n_1,n_2)$的值可查$F$分布表

$$\boxed{F_{1-\alpha}(n_1,n_2)=[F_\alpha(n_2,n_1)]^{-1}}$$



F分布的密度函数



F分布的分位数

0

x

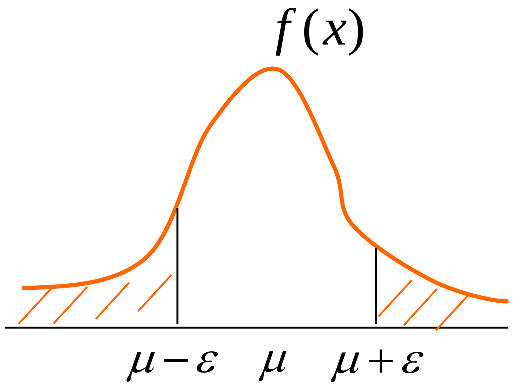
0

定理（契比雪夫不等式）:

设随机变量X具有数学期望$E(X)=\mu$,方差$D(X)=\sigma^2$

则对于任意$\varepsilon>0$,都有：$P\{|X-E(X)|\geq\varepsilon\}\leq\dfrac{\sigma^2}{\varepsilon^2}$

定理的**等价形式**为：$P\{|X-E(X)|<\varepsilon\}\geq 1-\dfrac{\sigma^2}{\varepsilon^2}$



大数定理

Chebyshev 大数定律

设随机变量序列$X_1, X_2, \cdots, X_n, \cdots$ 相互独立，

(指任意给定 $n > 1$, $X_1, X_2, \cdots, X_n$ 相互独立)，且具有相同的数学期望和方差

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

则 $\forall \varepsilon > 0$ 有 $$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| \geq \varepsilon\right) = 0$$

或 $$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| < \varepsilon\right) = 1$$

大数定理

定理(贝努里大数定理)

设事件A在每次试验中发生的概率为p，记n_A为n次独立重复试验中A发生的次数,则$\forall \varepsilon > 0$,有：$\lim\limits_{n\to+\infty} P\left\{\left|\frac{n_A}{n}-p\right|<\varepsilon\right\}=1$

贝努里大数定律建立了在大量重复独立试验中事件出现频率的稳定性，正因为这种稳定性，概率的概念才有客观意义，贝努里大数定律还提供了通过试验来确定事件概率的方法，既然频率n_A/n与概率p有较大偏差的可能性很小，我们便可以通过做试验确定某事件发生的频率并把它作为相应的概率估计。

大数定理

大数定理的意义:

具有相同数学期望和方差的独立随机变量序列的算术平均值依概率收敛于数学期望.

当 n 足够大时，算术平均值几乎就是一个常数,可以用算术平均值近似地代替数学期望.

这种方法即是参数估计法，参数估计的重要理论基础之一就是大数定理。

中心极限定理

设随机变量$X_1, X_2, \cdots, X_n, \cdots$相互独立同分布，

$E(X_i) = \mu, D(X_i) = \sigma^2, i = 1, 2, \cdots$

则前n个变量的和的标准化变量为：$Y_n = \dfrac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}$

$\forall x \in R$,有：$$\lim_{n\to+\infty} P(Y_n \le x) = \lim_{n\to+\infty} P\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \le x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

容易知道：对$\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$的近似分布为$N(\mu, \dfrac{\sigma^2}{n})$。

★许多随机变量是由大量的相互独立的随机变量的综合影响所形成的，而其中每个个别的因素作用都很小，这种随机变量往往服从或近似服从正态分布，中心极限定理正是从数学上论证了这一现象。

第二部分 数理统计

样本

我们把从总体中抽取的部分样品 $x_1, x_2, \cdots, x_n$ 称为样本。样本中所含的样品数称为样本容量，一般用n表示。在一般情况下，总是把样本看成是n个相互独立的且与总体有相同分布的随机变量，这样的样本称为简单随机样本。在泛指任一次抽取的结果时，$x_1, x_2, \cdots, x_n$ 表示n个随机变量（样本）；在具体的一次抽取之后，$x_1, x_2, \cdots, x_n$ 表示n个具体的数值（样本值）。我们称之为样本的两重性。

总体：研究对

个体：组成总

抽样：从总体X

随机样本：随

简单随机样本：满足以下两个条件的随机样本（X1, X2, …, Xn）称为简单随机样本。

1. 每个Xi与X同分布
2. X1, X2, …, Xn是相互独立的随机变量

[说明]：后面提到的样本均指简单随机样本，由概率论知，若总体X具有概率密度f(x)，

则样本（X1, X2, …, Xn）具有联合密度函数：

$$f_n(x_1, x_2, \cdots x_n) = \prod_{i=1}^{n} f(x_i)$$

- 统计量：样本的不含任何未知参数的函数。
- 常用统计量：设（$X_1, X_2, \cdots, X_n$）为取自总体X的样本

1\. 样本均值 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

2\. 样本方差 $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, S为样本标准差

3\. 样本矩　　k阶矩：$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \quad (k = 1, 2, \cdots)$

k阶中心矩：$B_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k \quad (k = 1, 2, \cdots)$

统计推断

● 均值的抽样分布

☆ 定理：若X ～ N（μ，σ^2），则对于容量为N的均值在重复抽样的状况下，有：

$$\overline{X} \sim \mathrm{N}(\mu, \sigma^2 / \mathrm{N})$$

也就是说，样本均值的抽样分布是均值为μ，方差为σ^2/N的正态分布。

☆ 中心极限定理： 对于从一均值为μ，方差为σ^2的非正态总体中抽取N个独立观测值的随机样本的均值，只要N充分大（如N＞30），该均值的抽样分布近似于均值为μ，方差为σ^2/N的正态分布。即：

$$\overline{X} \sim \mathrm{N}(\mu, \sigma^2 / \mathrm{N})$$

参数估计（１）

● 点估计

１．估计量和估计值：　随机变量X的概率密度函数为f(x, θ)，从样本推出 θ 的一个公式

$$\theta^* = g(x_1, x_2, \ldots, x_n)$$

称为总体参数 θ 的估计量，估计量所取的一个具体值称为 θ 的一个估计值。如算术平均和几何平均

$$\theta^* = \frac{1}{n}(x_1 + x_2 + \ldots + x_n) = \overline{X}$$

$$\theta^* = \sqrt[n]{x_1 x_2 \ldots x_n} = \overline{X}$$

就是总体均值 μ 的二个估计量，而

$$\overline{X} = 1.7$$

就是总体均值 μ 的一个估计值。此为点估计。

参数估计（2）

2．点估计量的统计性质

（1）小样本性质

- 无偏性或无偏估计量：$E(\theta^*)=\theta$
- 有效性：若 θ_1，θ_2 是 θ 的两个无偏估计量，且 $Var(\theta_1)\leq Var(\theta_2)$，则称 θ_1 相对 θ_2 为有效估计量。无偏估计量中具有最小方差的估计量称为最佳估计量。
- 最佳线性无偏性：若估计量是样本观测值的一个线性函数，则称为线性估计量。线性、无偏的最佳估计量称为最佳线性无偏估计量(BLUE:the best linear unbiased estimator)

（1）大样本性质

- 渐近无偏性或渐近无偏估计量：$\lim\limits_{n\to\infty} E(\theta^*)=\theta$
- 一致性或一致估计量：$P\lim\limits_{n\to\infty}\theta^*=\theta$

区间估计（1）

● 区间估计

1．置信区间、置信限： 由前面定理可知，若X ～ N（μ，σ²），则对于容量为N的均值在重复抽样的状况下，有：

$$\overline{X} \sim \mathrm{N}(\mu,\ \sigma/\sqrt{\mathrm{N}})$$

由正态分布的性质可知，在重复抽样的状况下，95%的样本均值将落在下面区间内

$$\mu - 1.96\sigma_{\bar{x}} \le \overline{X} \le \mu + 1.96\sigma_{\bar{x}}$$

经变换可得： $\overline{X} - 1.96\sigma_{\bar{x}} \le \mu \le \overline{X} + 1.96\sigma_{\bar{x}}$

上述区间称为 μ 的95%置信区间，该区间的上、下限称为 μ 的95%置信限。

一般的： $P(\overline{X} - Z_{\frac{\alpha}{2}}\sigma_{\bar{x}} \le \mu \le \overline{X} + Z_{\frac{\alpha}{2}}\sigma_{\bar{x}}) = 1 - \alpha$

其中概率1－α称为估计 μ 时的置信水平。若α＝5%时，$Z_{\alpha/2}$=1.96；若α＝1% 时，$Z_{\alpha/2}$=2.58， μ 的99%置信限为 $\overline{X} \pm 2.58\sigma_{\bar{x}}$

区间估计（2）

2．均值的置信限： 由于总体标准差σ一般未知，只能用样本标准差S代替，于是有：

$$t=\frac{\overline{X}-\mu}{S_{\bar{x}}}\sim t(N-1)$$

与正态分布类似有： $P(\overline{X}-\mathrm{t}_{\frac{\alpha}{2}}\sigma_{\bar{x}}\leq\mu\leq\overline{X}+\mathrm{t}_{\frac{\alpha}{2}}\sigma_{\bar{x}})=1-\alpha$

- 若取α=0. 05，则 $P(\overline{X}-Z_{0.025}\sigma_{\bar{x}}\leq\mu\leq\overline{X}+Z_{0.025}\sigma_{\bar{x}})=0.95$

  即μ的95%置信限为： $\overline{X}\pm t_{0.025}S_{\bar{x}}$

- 例：若置信水平是95%即α＝5%，自由度N-1=18时，查t表可得 $t_{0.025}$=2. 101。所以μ的95%置信限为：

$$\overline{X}\pm 2.101S_{\bar{x}}$$

假设检验的步骤(1)

- 假设检验的逻辑

  ○ 假设检验实际上是检验是否出现了小概率事件，如果出现，则拒绝原来关于总体参数的假设；如果检验表明得到的样本值并不属于小概率事件，则接受该假设。

  ○ 假设检验作为统计推断两个分支之一有两种检验方法：显著性检验方法和置信区间法。

  ○ 显著性水平：概率小到什么程度才算小概率事件。一般看该检验能承担多大风险。在实践中，一般取5%，称为5%的显著性水平。

假设检验的步骤

基本步骤

假设检验的基本步骤如下：

(i)提出零假设H_0；

(ii)选择统计量K；

(iii)对于检验水平α查表找分位数λ；

(iv)由样本值$x_1, x_2, \cdots, x_n$计算统计量之值K；

将$\hat{K}$与λ进行比较，作出判断：当$|\hat{K}| > \lambda$(或$\hat{K} > \lambda$)时否定H_0，否则认为H_0相容。

假设检验

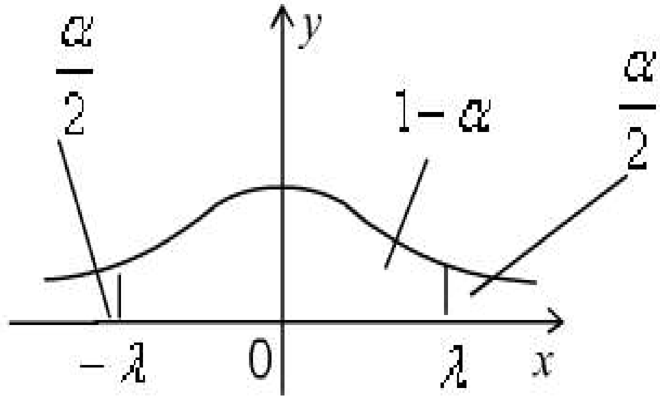
$$P\left(\left|\frac{\overline{X}-\mu}{\delta_0/\sqrt{n}}\right|\le\lambda\right)=\alpha$$

H_0：假定总体期望 μ

$$\left|\frac{\overline{X}-\mu}{\delta_0/\sqrt{n}}\right|>\lambda_y$$ 否定 H_0

$$\left|\frac{\overline{X}-\mu}{\delta_0/\sqrt{n}}\right|\le\lambda_y$$ 接受 H_0

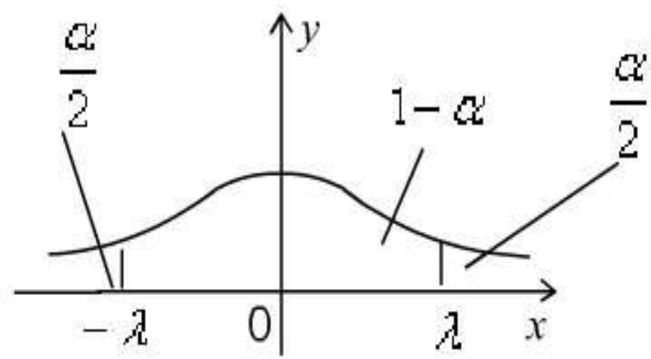
λ 如何确定



假设检验

$$P\left(\left|\frac{\overline{X}-\mu}{\delta_0/\sqrt{n}}\right|>\lambda\right)=\alpha \qquad \lambda=\lambda_{1-\frac{\alpha}{2}}$$

$$\longrightarrow N(0,1)$$



假设检验的步骤（2）

- 假设检验的步骤和方法

下面用一个例子来说明假设检验：某厂生产一种直径为100mm的轴，随机抽取16根轴的样本，算出平均直径为110mm，方差为100，检验总体均值是否为100mm。

1．建立关于总体的原假设和备择假设：

原假设(null hypothesis)　　H_o： $\mu=100$

备择假设(alternative hypothesis)　　H_a： $\mu\neq100$

上述类型的检验称为双侧检验。若备择假设变为 $\mu>100$ 或 $\mu<100$，则称为单侧检验。

假设检验的步骤（3）

2. 计算检验统计量

构造下列**统计**量，若 σ 已知，很容易**计**算出其**值**。

$$Z=\frac{\overline{X}-\mu}{\sigma_{\bar{x}}}=\frac{\overline{X}-\mu}{\sigma/\sqrt{N}}\sim N(0,1)$$

但由于式中的**总体标**准差 σ 一般是未知的，所以**该统计**量不常用。而用**样**本数据代替，重新构造**统计**量。

$$S^2=\frac{\sum(X-\overline{X})^2}{N-1}$$

$$t=\frac{\overline{X}-\mu}{\hat{\sigma}_{\bar{x}}}=\frac{\overline{X}-\mu}{S_{\bar{x}}}=\frac{\overline{X}-\mu}{S/\sqrt{N}}\sim t(16)$$

t**统计**量与Z**统计**量不同，不再服从正**态**分布，而是服从自由度**为**N-1的 t 分布。

假设检验的步骤（4）

将样本均值、样本标准差代入t统计量中，计算 t 值。

$$\overline{X} = 110,\ \mu = 100$$

$$S = \sqrt{100} = 10,\ S_{\bar{x}} = \frac{S}{\sqrt{N}} = \frac{10}{\sqrt{16}} = 2.5$$

$$\Rightarrow t = \frac{\overline{X} - \mu}{S_{\bar{x}}} = \frac{110 - 100}{2.5} = 4$$

事实上，可以将均值的 t 检验推广至任何总体参数 β。只要 β 的样本估计量服从正态分布，构造的t统计量就服从 t 分布。

$$t = \frac{\hat{\beta} - \beta}{Se(\hat{\beta})}$$

假设检验的步骤（5）

3．检验原假设

若显著性水平5%确定以后，可以通过查 t 表查出 t 统计量的临界值，再和上一步算出的 t 值相比较，若

$$|t| > t_c$$

则拒绝原假设，否则接受原假设(参考图2-5)。本例中由于

$$|t| = 4 > t_{0.05} = 2.13$$

则拒绝原假设H_0，即样本数据不支持总体均值为100的假设。

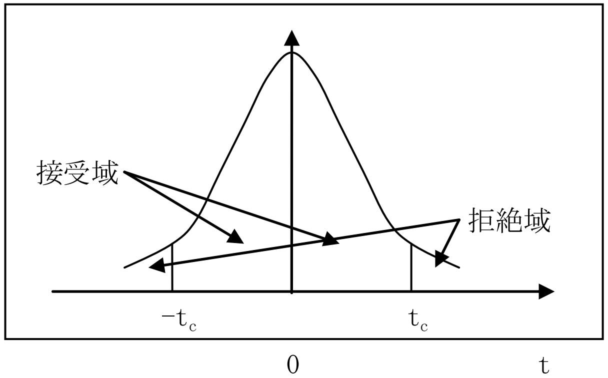


图2-5　接受域和拒绝域

二、两种类型的错误

● 第Ⅰ、Ⅱ类错误

○ 拒绝正确的原假设，称为第Ⅰ类错误；接受错误的原假设，称为第Ⅱ类错误。两类错误产生的概率如下表所示。

	H_0 正確	H_0 錯誤
接受 H_0	正確決策 （$1-\alpha$）	第二類錯誤 （β）
拒絕 H_0	第一類錯誤 （α）	正確決策 （$1-\beta$）

○ 犯第Ⅰ类错误的概率是 α，即**显著水平**；犯第Ⅱ类错误的概率是 β，（$1-\beta$）则给出正确地拒绝错误的原假设的概率，（$1-\beta$）称为统计检验的**功效**。

谢谢！