小测验 1

1. 如何对随机现象的不确定性进行研究？

2. 频数资料的算术平均数如何计算？

3. 概率的统计定义？

4. 为什么要研究概率分布?

5. 离散型随机变量的特点？采用那些函数描述它们的概率分布？

6. 连续型随机变量的特点?采用那些函数描述它们的概率分布？

7. 标准差的含义及计算式是什么？

8. 变异系数的应用条件及其计算式是什么？

9. 数学期望的含义是什么?

10. 用哪两种方法可以计算连续型随机变量在任意区间的概率？

课前提问

- 研究概率分布的目的
- 概率分布
- 随机变量，类型及各自的特点
- 离散型随机变量的概率分布的研究方法?
- 连续型随机变量的概率分布的研究方法?
- 累积分布函数
- 随机变量的数学期望和方差

第三章　几种常见的概率分布律

张爱莲

新疆大学生命科学与技术学院　新疆生物资源基因工程重点实验室

zalxju@gmail. com

20100909

3.1 二项分布(*binomial distribution*)

3.1.1 贝努利试验及其概率公式

3.1.2 二项分布的定义及性质

3.1.3 二项分布的应用

3.1.4 二项分布的概率计算

3.1.5 二项分布的特征数

3.1.1 贝努利试验及其概率公式

在生物学研究中，我们经常碰到的一类**离散型随机变量:**

- 如入孵n枚种蛋的出雏数
- n头病畜治疗后的治愈数
- 某种遗传性状在子二代的出现数

实际问题中，有许多实验它们只包含两个结果。这些例子所具有的共同性质可总结如下：

- 试验包含了n个相同的试验；
- 每一次试验只有两个可能结果：“成功”或“失败”；
- 出现“成功”的概率 φ 对每一次试验是相同的，“失败”的概率 $1-\varphi$ 也不变，且 $\varphi+1-\varphi=1$
- 试验是相互独立的；

通常称具有上述特征的n次重复的独立试验为*n*重贝努利试验，简称贝努利试验。

摸棋子实验：

设把相当多的围棋子放在一个坛子里，黑白子之比1：2。完全混合，随机瞎摸. 棋子经混合后被摸到的概率相同，则：

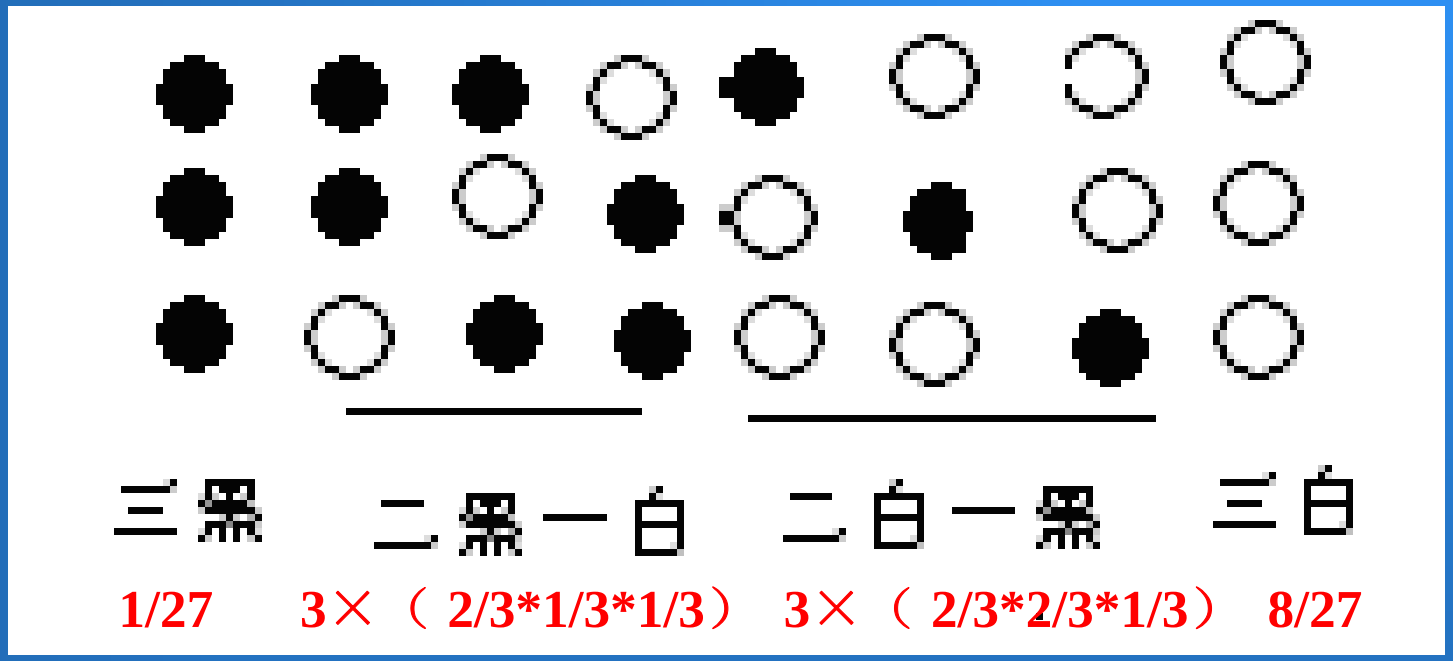
摸出一只黑子的概率是:$\varphi = 1/3$，

摸出一只白子的概率是（$1-\varphi$）$= 2/3$，

摸到一只“不是黑子就是白子”的概率是1。

因此，$\varphi + (1-\varphi) = 1/3 + 2/3 = 1$

先后从坛子里随机摸出一子，共摸三次排列起来，它们共有$2^3=8$种不同的排列方式, 4种不同的组合方式如图：



（若摸10次、20次，则各种情况的概率应是多少？）

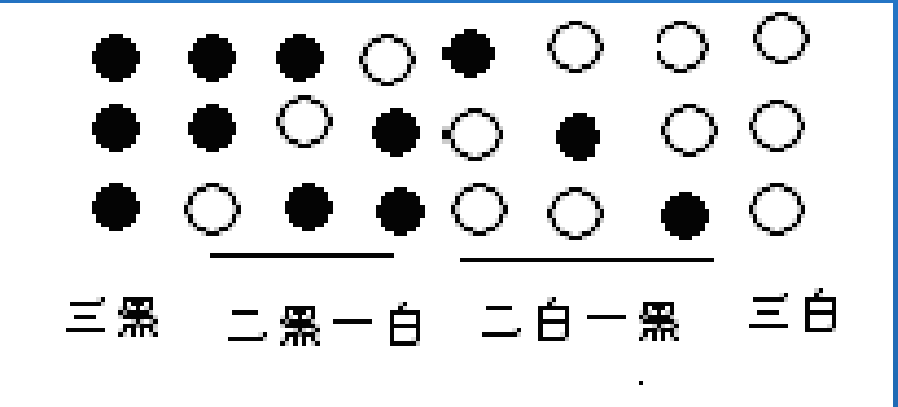
因为摸到一只黑子的概率是1/3，白子是2/3，它们相互独立，则各种组合方式的概率是：

“0白三黑”的概率是：$\left(\frac{2}{3}\right)^0\left(\frac{1}{3}\right)^3=\frac{1}{27}$

“一白二黑”的概率是：$3\left(\frac{2}{3}\right)^1\left(\frac{1}{3}\right)^2=\frac{6}{27}$

“二白一黑”的概率是：$3\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^1=\frac{12}{27}$

“三白0黑”的概率是：$\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^0=\frac{8}{27}$



这四种组合方式的概率相加应等于1。

$$\frac{1}{27}+\frac{6}{27}+\frac{12}{27}+\frac{8}{27}=1$$

$$\left(\frac{2}{3}\right)^0\left(\frac{1}{3}\right)^3+3\left(\frac{2}{3}\right)^1\left(\frac{1}{3}\right)^2+3\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^1+\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^0=\left(\frac{1}{3}+\frac{2}{3}\right)^3$$

容易看出，黑白子组合的各种方式的概率恰好是二项式展开式相应的各项。计算概率 $\left(\frac{1}{3}+\frac{2}{3}\right)^3$ 各项的系数.

也就是：1、3、3、1

c_3^0、c_3^1、c_3^2、c_3^3。 $C_n^r=\dfrac{n!}{r!(n-r)!}$

现引进一随机变量X，X表示摸白子，

设：X=0表示“0只白子，3只黑子”，概率为：

$$p(0)=C_3^0\left(\frac{2}{3}\right)^0\left(\frac{1}{3}\right)^3=\frac{1}{27}$$

X=1表示“1只白子，2只黑子”，概率为：

$$p(1)=C_3^1\left(\frac{2}{3}\right)^1\left(\frac{1}{3}\right)^2=\frac{6}{27}$$

X=2表示“2只白子，1只黑子”，概率为：

$$p(2)=C_3^2\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^1=\frac{12}{27}$$

X=3表示“3只白子，0只黑子”，概率为

$$p(3)=C_3^3\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^0=\frac{8}{27}$$

在n重贝努利试验中，事件 A 可能发生0，1，2，…，n次，现在我们来求事件 A 恰好发生x($0\leqslant x\leqslant n$)次的概率$P(x)$，

设 n 为试验次数（或样本含量），

X 为在n次试验中事件A出现的次数（白子出现次数）

φ为事件A发生x次的概率（白子出现的概率）。

$1-\varphi=\overline{A}$ 发生的概率（黑子的概率）

则对于任意的n和X有通式：

$$P(x)=C_n^x\varphi^x(1-\varphi)^{(n-x)},x=0,1,2,\ldots\ldots$$

此式正是二项式展开式 $(\varphi+1-\varphi)^n$ 的第X+1项，因此称为二项分布.

其中：$p(x)=C_n^x\varphi^x(1-\varphi)^{n-x}$

称为二项分布的概率函数.

系数：$C_n^x=\dfrac{n!}{x!(n-x)!}$

3.1.2 二项分布的定义及性质

1. 二项分布定义如下：

设随机变量 x 所有可能取的值为零和正整数：0,1,2,…，n，且有

$$p(x)=C_n^x\varphi^x(1-\varphi)^{n-x} \quad x=0,1,2\ldots,n$$

其中$\varphi>0$，$1-\varphi>0$，则称随机变量 x 服从参数为 n 和 φ 的二项分布，记为：

$$x\sim B(n,\varphi)$$

二项分布是一种离散型随机变量的概率分布。

$$p(x) = C_n^x \varphi^x (1-\varphi)^{n-x}$$

参数n称为离散参数， 只能取正整数；

φ是连续参数，它能取0与1之间的任何数值。

2. 二项分布的性质：

$P(X=x)= P(x) \quad (x=0,1,\ldots, n)$

二项分布的概率之和等于1，即

$$\sum_{x=0}^{n} C_n^x \varphi^x (1-\varphi)^{n-x} = 1$$

$$P(X \le m) = P_n(x \le m) = \sum_{x=0}^{m} C_n^x \varphi^x (1-\varphi)^{n-x}$$

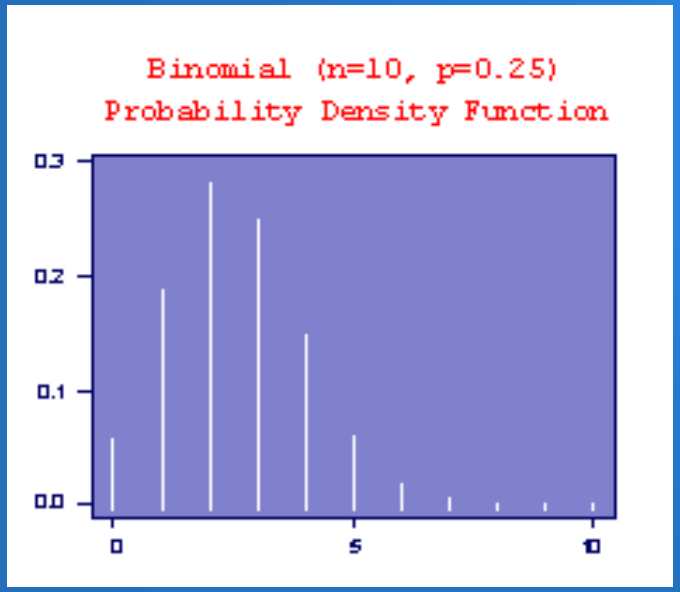
$$P(X \ge m) = P_n(x \ge m) = \sum_{x=m}^{n} C_n^x \varphi^x (1-\varphi)^{n-x}$$

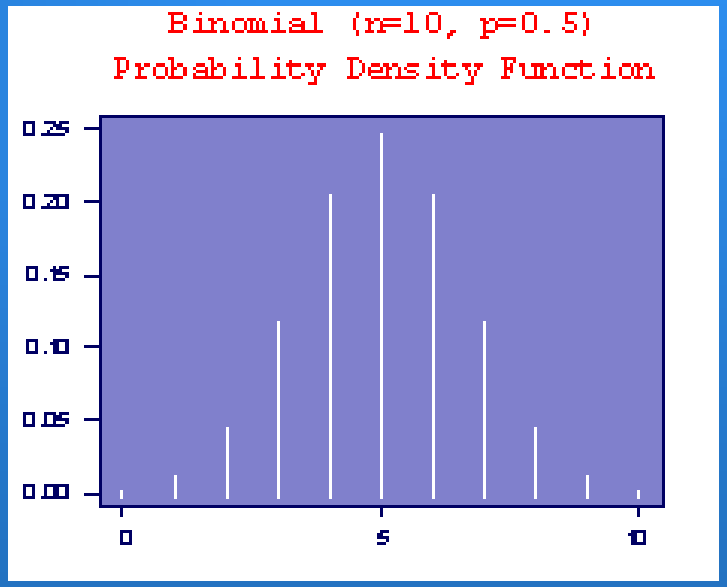
$$P(m_1 \le X \le m_2) = p_n(m_1 \le x \le m_2) = \sum_{x=m_1}^{m_2} C_n^x \varphi^x (1-\varphi)^{n-x}$$

（$m1<m2$）

3. 二项分布的形状由*n* 和*φ*两个参数决定：

当*φ*值较小且*n*不大时，分 布 是偏倚的。但随着*n*的增大，分布逐渐趋于对称.

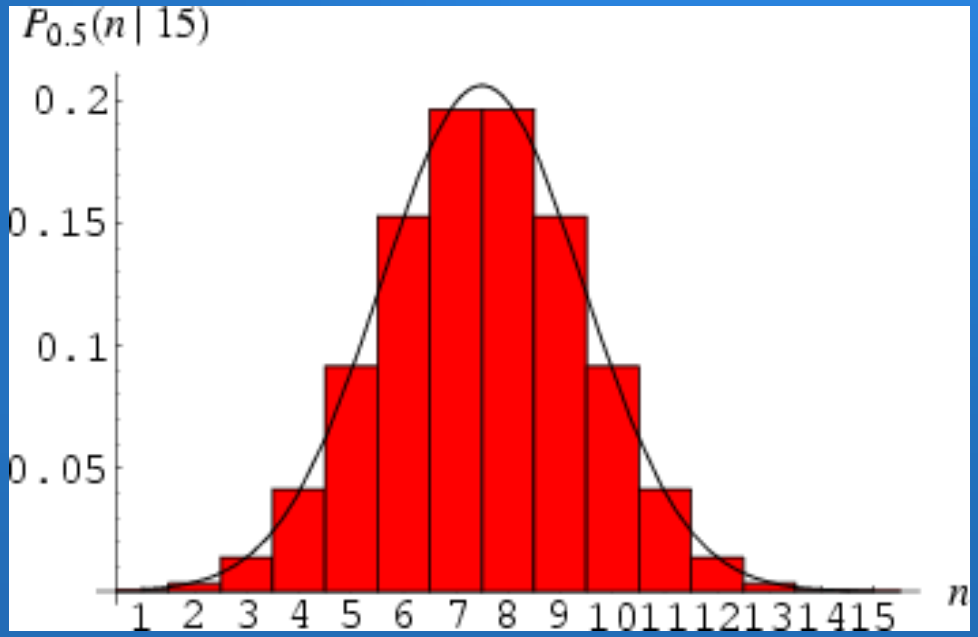


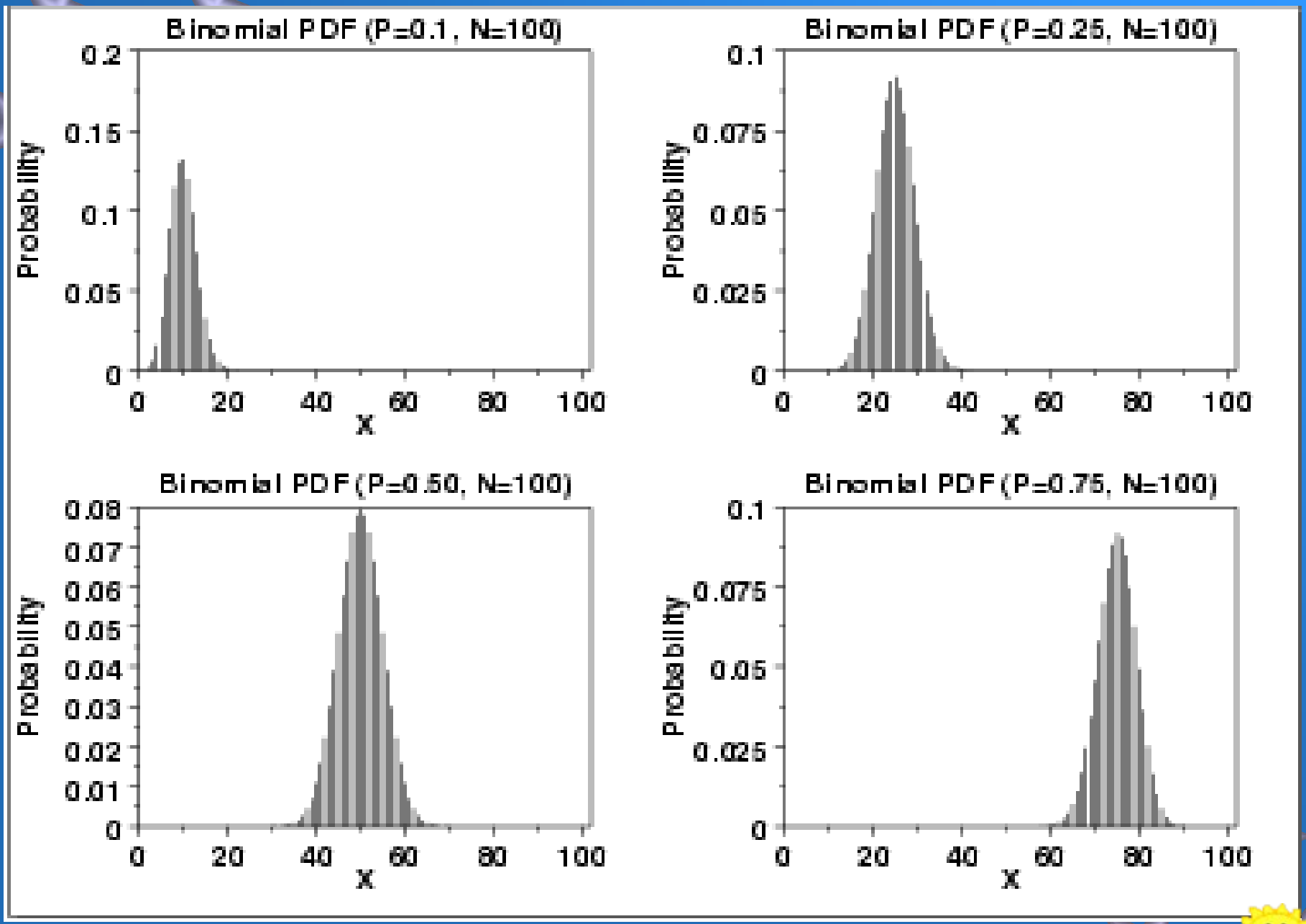


当 *φ* 值 趋 于 **0.5** 时，分 布 趋于对称；

对于固定的n及φ，当x增加时，$P(x)$先随之增加并达到其极大值，以后又下降。

此外，在n较大，$n\varphi$、$n(1-\varphi)$较接近时，二项分布接近于正态分布；当$n\to\infty$时，二项分布的极限分布是正态分布。




Binomial PDF (P=0.1, N=100)
Probability
0.2
0.15
0.1
0.05
0
0
20
40
60
80
100
X
Binomial PDF (P=0.25, N=100)
Probability
0.1
0.075
0.05
0.025
0
0
20
40
60
80
100
X
Binomial PDF (P=0.50, N=100)
Probability
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0
20
40
60
80
100
X
Binomial PDF (P=0.75, N=100)
Probability
0.1
0.075
0.05
0.025
0
0
20
40
60
80
100
X

3.1.3 二项分布的应用

- 判断一个现象是不是二项分布
- 求概率
- 在遗传学上又大量应用

二项分布的应用条件（三个）

- 各观察单位只具有互相对立的一种结果，如阳性或阴性，生存或死亡等，属于二项分类资料。
- 已知发生某一结果(如死亡) 的概率为p，其对立结果的概率则为1-P=q，实际中要求p 是从大量观察中获得的比较稳定的数值。
- n个观察单位的观察结果互相独立，即每个观察单位的观察结果不会影响到其它观察单位的结果。

3.1.4 二项分布的概率计算

【例1】牛的一种的病在某地区的死亡率为40%，现对该病有一种新药进行治疗试验，每次抽样10头病牛作为一组治疗。试问如新药无疗效则在10头病牛中死去3头、2头、1头，及全部愈好的概率是多少？

假设无效（零假设）利用已知条件，每次抽n头相当于一次抽一头，共抽了n次。

先判断是否属二项分布，φ=0.4（死亡的概率）

1-0.4=0.6（愈好的概率）

死去3头，愈好7头的概率：$p(3)=C_{10}^{3}(0.4)^{3}0.6^{7}=0.21499$

死去2头，愈好8头的概率：$p(2)=C_{10}^{2}(0.4)^{2}0.6^{8}=0.12093$

死去1头，愈好9头的概率：$p(1)=C_{10}^{1}(0.4)^{1}0.6^{9}=0.04031$

10头全部愈好的概率：$p(0)=C_{10}^{0}(0.4)^{0}0.6^{10}=0.00605$

死去2头以下包括2头的概率是多少？

$$F(x\le 2)=p(0)+p(1)+p(2)$$
$$=0.00605+0.04031+0.12093$$
$$=0.16729$$

【例2】遗传学中，两个纯合亲本杂交（RR×rr）F1代自交，F2代的基因型分离比（1RR：2Rr:1rr）是一个二项分布问题，其中

事件A：出现显性基因，φ=1/2

事件 $\overline{A}$：出现隐性基因，1-φ=1/2

*x*为显性基因的出现数

*n*为等位基因数（一对等位基因n为2，两对n为4）

$$[\varphi+(1-\varphi)]^2 = C_2^0\varphi^0(1-\varphi)^2 + C_2^1\varphi^1(1-\varphi)^1 + C_2^2\varphi^2(1-\varphi)^0$$

$$= \varphi^2 + 2\varphi(1-\varphi) + (1-\varphi)^2 = \frac{1}{4} + \frac{2}{4} + \frac{1}{4} = 1$$

3.1.5 二项分布的特征数

统计学证明，服从二项分布$B(n, \varphi)$的随机变量的平均数μ、标准差σ与参数n、φ有如下关系：

- 当试验结果以事件A发生次数表示时

平均数：$\mu = n\varphi$

意义：做n次独立试验，某事件平均出现的次数为$n\varphi$

方差：$\sigma^2 = n\Phi(1-\Phi)$

标准差：$\sigma = \sqrt{n\varphi(1-\varphi)}$

- 当试验结果以事件A发生的频率k/n表示时

平均数：$\mu_\varphi = \varphi$

方差：$$\sigma_\varphi^{\ 2} = \frac{\varphi(1-\varphi)}{n}$$

3.2 波松poisson分布

3.2.1 波松分布的概率函数

3.2.2 波松分布的含义

3.2.3 波松分布重要的特征

3.2.4 波松分布的应用

波松分布是一种可以用来描述和分析随机地发生在单位空间或时间里的稀有事件的概率分布。

在生物、医学研究中服从波松分布的随机变量是常见的：

- 一定畜群中某种患病率很低的非传染性疾病患病数 或死亡数；
- 畜群中遗传的畸形怪胎数；
- 每升饮水中大肠杆菌数；
- 计数器小方格中血球数；
- 单位空间中某些野生动物或昆虫数等，都是服从波松分布的。

- 在二项分布中，当某事件出现的概率特别小（$\varphi \to 0$），而样本含量又很大（$n \to \infty$）时，二项分布就变成泊松分布了。
- 波松分布是描述在一定空间、长度、面积、体积或一定时间间隔内，点子散布状况的理想化模型。

3.2.1 波松分布的概率函数

- 泊松分布的概率函数为：

$$P(x)=\frac{\mu^{x}}{x!}e^{-\mu}$$

X =0，1，……

3.2.2 波松分布的含义

若随机变量 x 只取零和正整数值0，1，2，…，且其概率分布为

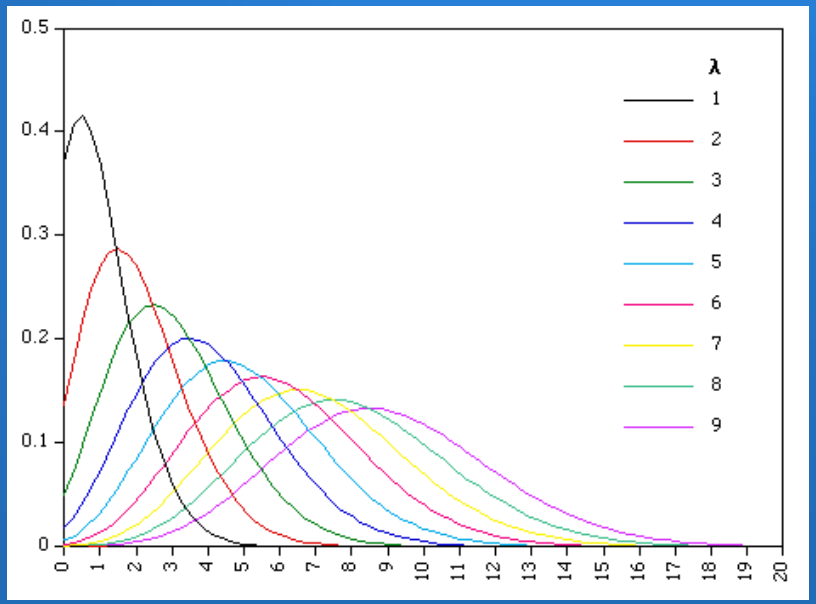
$$P(x)=\frac{\mu^{x}}{x!}e^{-\mu} \quad ,x=0，1，……$$

其中μ＞0；e=2.7182… 是自然对数的底数，

则称 x 服从参数为 μ 的波松分布，记为 $x\sim P(\mu)$。

Poisson分布的形状由参数 μ 确定:

- 当 μ 较小时，泊松分布是偏倚的，随μ增大，分布　逐渐对称.
- 当μ=20时，泊松分布已和正态分布非常接近.
- 当μ=50时，泊松分布和正态分布除一种是离散型一种是连续型外，已没有多大区别。



3.2.3 波松分布重要的特征

平均数和方差相等，都等于常数 μ，

即 $\mu = \sigma^2 = \mu$

3.2.4　波松分布的应用

- 求概率

　　在实际应用上，主要在医学科学、稀有动植物分布等领域

课前提问

- 二项分布的概率函数
- 二项分布的应用条件

3.3 正态分布（Normal Distribution）

3.3.1 正态分布的定义及其特征

3.3.2 标准正态分布

3.3.3 正态分布的概率计算

3.3.4 正态分布的临界值

本节要点

- 正态分布　密度函数　分布曲线　　累积分布函数
- 标准正态分布　密度函数　分布曲线
- 正态分布的临界值（分位数）

正态分布是一种很重要的连续型随机变量的概率分布。生物现象中有许多变量是服从或近似服从正态分布的。许多统计分析方法都是以正态分布为基础的。

正态分布的研究意义

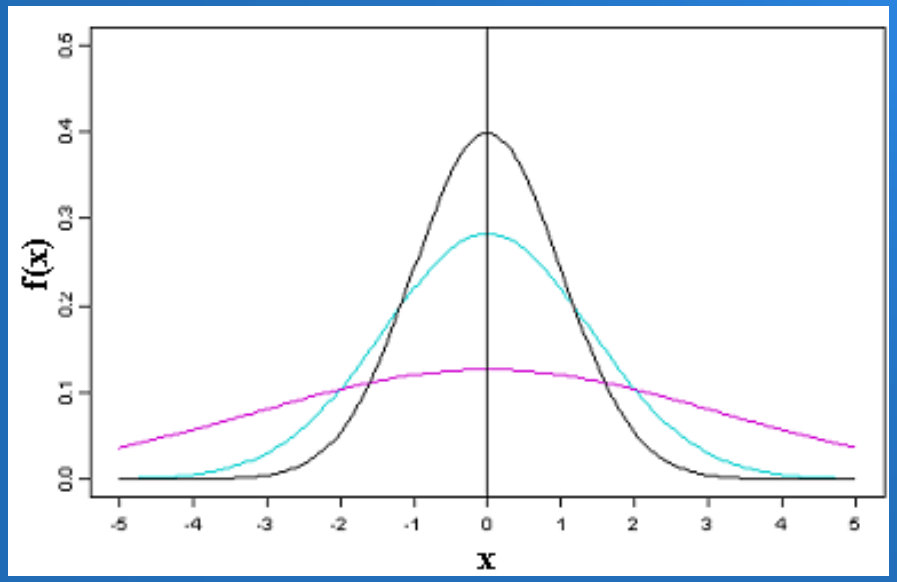
(1) 在统计推断中它扮演着中心角色

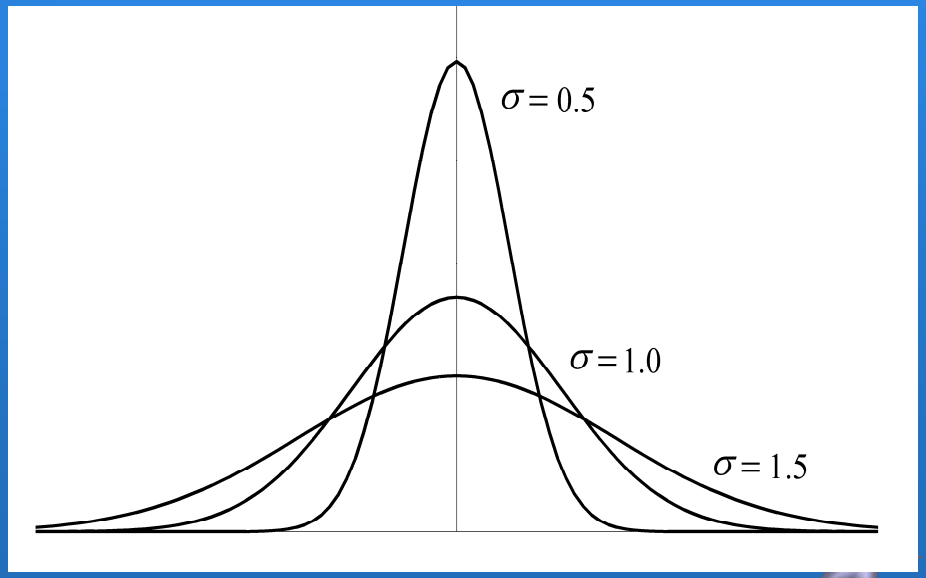
(2)在实际中的许多随机变量表现出来的频数分布非常接近正态分布

(3) 它能方便的用于近似许多其它的概率分布，如二项分布和Poisson分布。绝少部分资料的抽样分布，当n适当大时，接近正态分布

正态分布：

调查数据以平均数为原点，向左右两侧做对称分布，即两头少，中间多，两侧对称。数据的这种分布规律，称为正态分布。





当一个随机变量受很多小的随机因素影响时，其分布是正态分布

3.3.1 正态分布的定义及其特征

（一）正态分布的定义：

若连续型随机变量X 的概率分布密度函数为

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

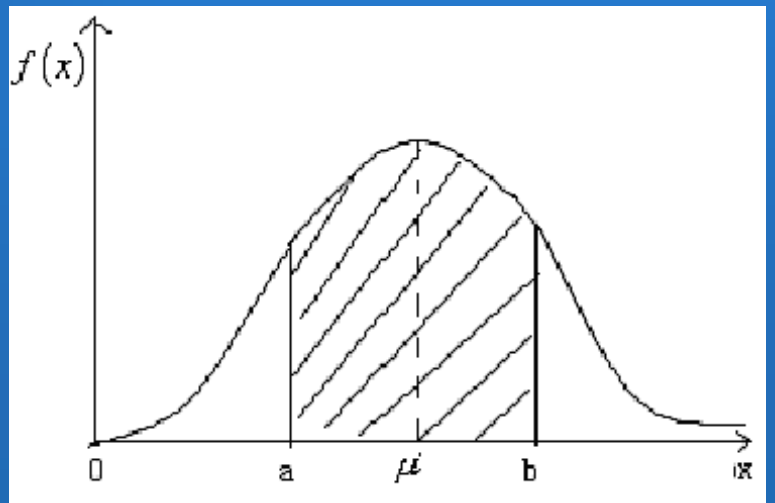
其中 μ 为平均数， σ^2为方差，则称随机变量X服从正态分布， 记为$X\sim N(\mu,\sigma^2)$。

随机变量的值落在任意区间（*a*,*b*）内的概率

$$P(a < X < b) = \int_a^b f(x)dx = \frac{1}{\sigma\sqrt{2\pi}}\int_a^b e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$

相应的概率分布函数为：

$$F(x) = P(X < x) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$



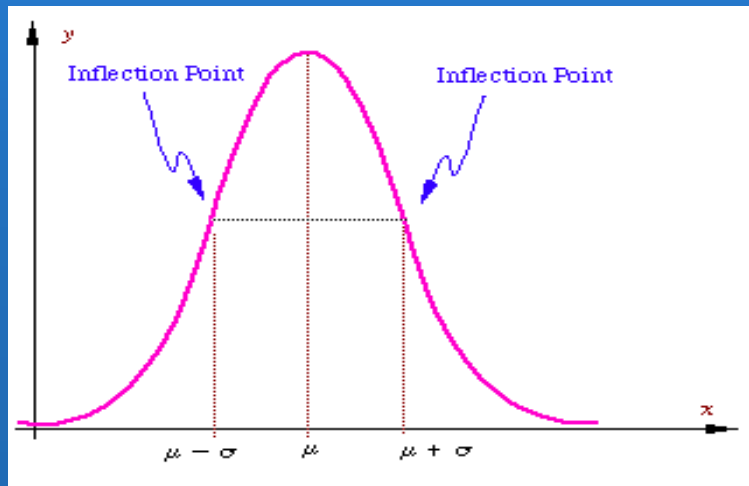
（二）正态分布密度曲线特征

1、正态分布密度曲线是单峰、对称的悬钟形曲线，对称轴为$x=\mu$；

2、$f(x)$ 在 $x=\mu$ 处达到极大，极大值 $f(\mu)=\dfrac{1}{\sigma\sqrt{2\pi}}$

3、$f(x)$是非负函数，以x轴为渐近线，分布从$-\infty$至$+\infty$；

4、曲线在$x=\mu\pm\sigma$处各有一个拐点，即曲线在$(-\infty,\mu-\sigma)$和$(\mu+\sigma,+\infty)$区间上是下凸的，在$[\mu-\sigma,\mu+\sigma]$区间内是上凸的；



5、正态分布有两个参数即平均数μ和标准差σ。

μ是位置参数，当σ恒定时，μ愈大，则曲线沿x轴愈向右移动；反之，μ愈小，曲线沿x轴愈向左移动。

σ是变异度参数，当μ恒定时，σ愈大，表示x的取值愈分散，曲线愈“胖”；σ愈小，x的取值愈集中在μ附近，曲线愈“瘦”。

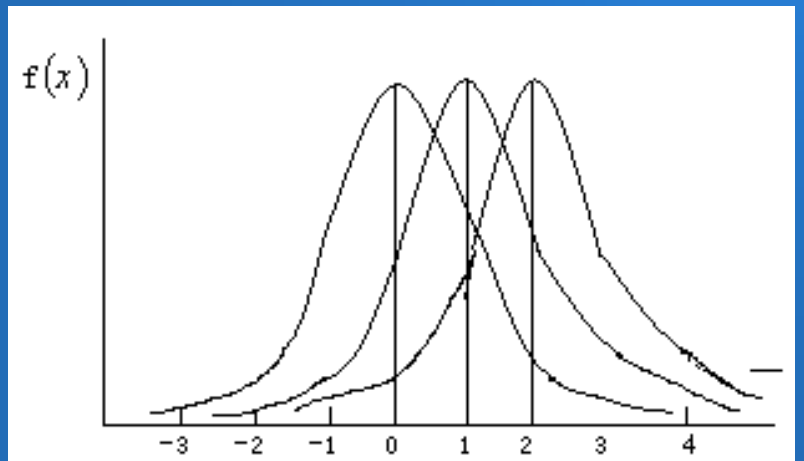
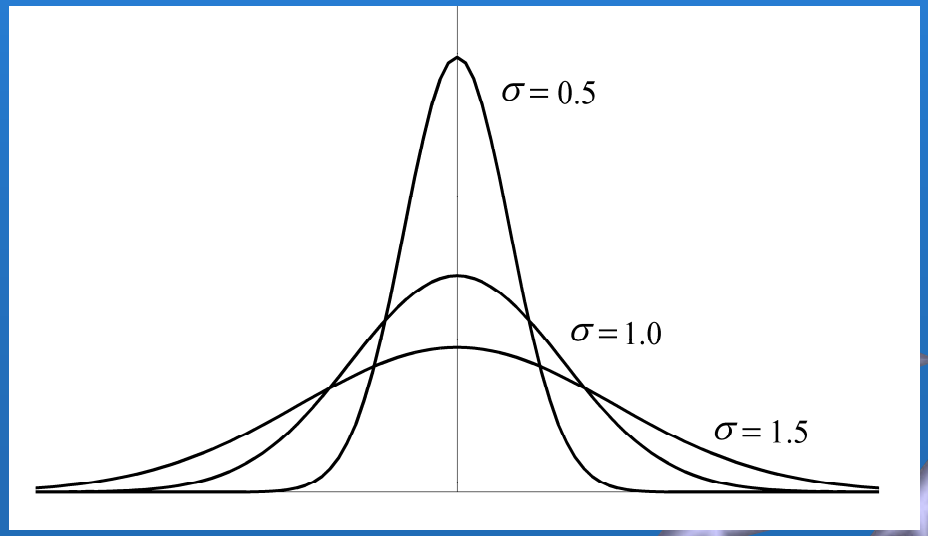


Fig 3-4 σ相同而μ不同的三个正态曲线

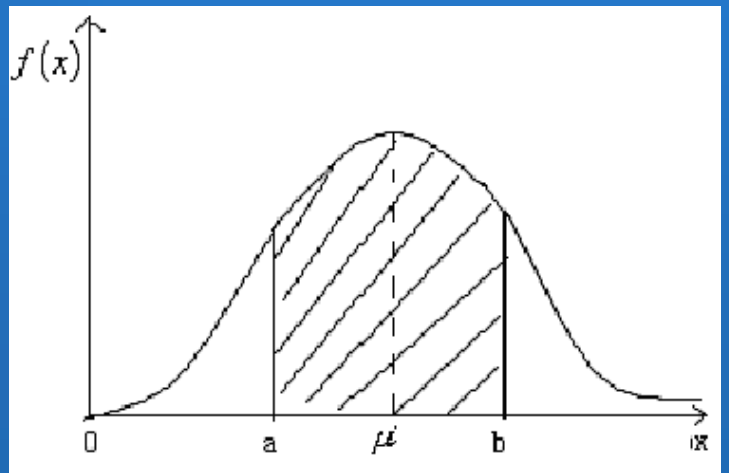


6、分布密度曲线与横轴所夹的面积为1，即：

$$P(-\infty < x < +\infty) = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$$

随机变量在区间（a，b）内取值的概率：即图中阴影部分的面积。

$$p(a \le x \langle b) = \int_a^b f(x)dx = F(b) - F(a)$$



(三)一些特殊的值:

X在μ左右一倍标准差范围内取值的概率为:

$$p(\mu-\sigma \le x \le \mu+\sigma)=\int_{\mu-\sigma}^{\mu+\sigma} f(x)dx = 0.6827$$

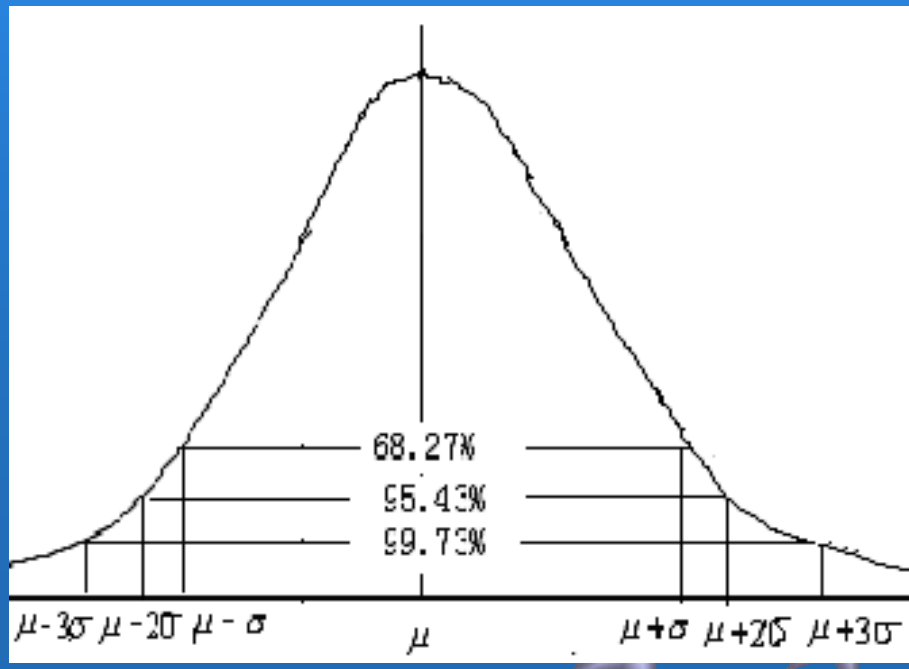
$$p(\mu-2\sigma \le x \le \mu+2\sigma)=\int_{\mu-2\sigma}^{\mu+2\sigma} f(x)dx = 0.9543,$$

$$p(\mu-3\sigma \le x \le \mu+3\sigma)=\int_{\mu-3\sigma}^{\mu+3\sigma} f(x)dx = 0.9973$$

此范围外仅0. 0027。

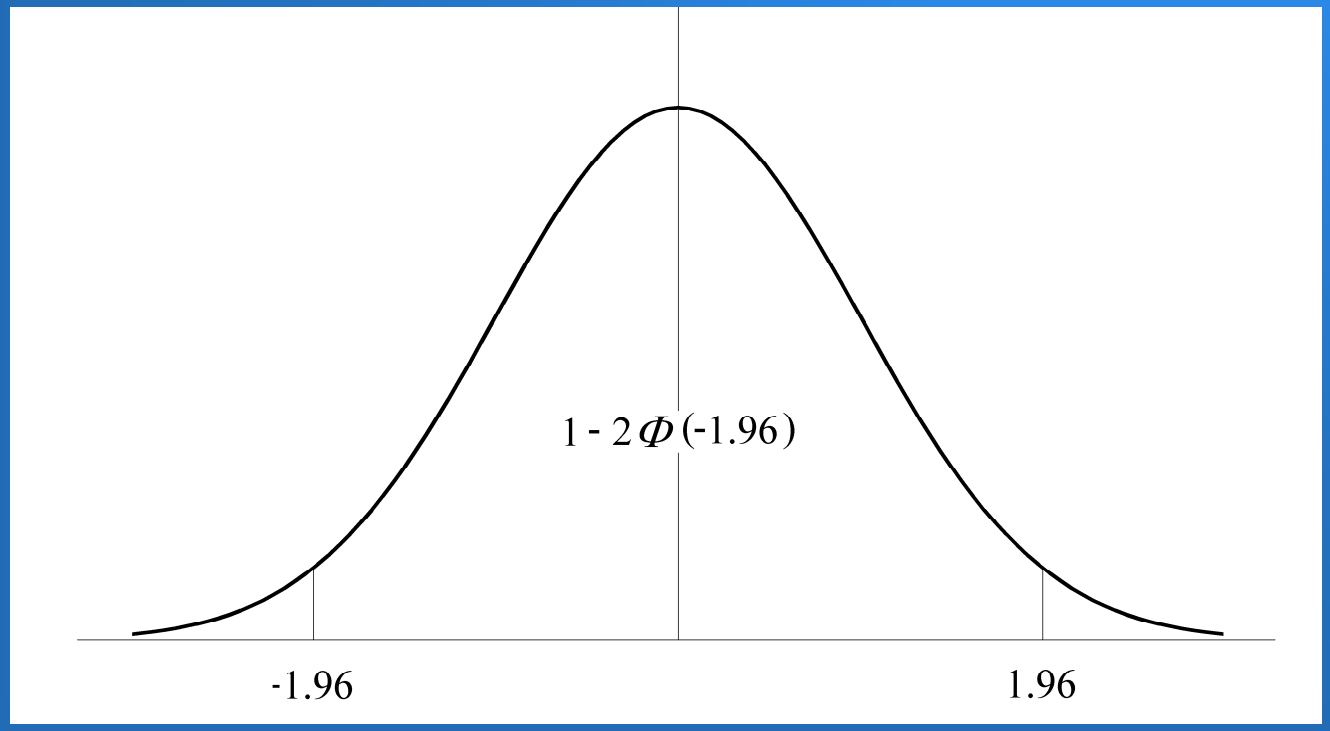
科研论文数据，以 $\bar{x} \pm S$ 表示。

若在平均数左右一倍标准差外，出现的概率为1-0.6827=0.3173，



$$p(\mu - 1.96\sigma \le x \le \mu + 1.96\sigma) = 0.95$$

$$p(\mu - 2.58\sigma \le x \le \mu + 2.58\sigma) = 0.99$$



3.3.2 标准正态分布

- 由上述正态分布的特征可知 ， 正态分布是依赖于参数 μ 和 σ^2 (或 σ) 的一簇 分布 ；
- 正态曲线之位置及形态随 μ 和 σ^2的不同而不同 ,这就给研究具体的正态总体带来困难;
- 需将一般的$N(\mu, \sigma^2)$ 转 换为 $\mu = 0, \sigma^2=1$ 的正态分布

（一）定义

- 我们称 $\mu=0, \sigma^2=1$ 的正分布为标准正态分布 (standard normal distribution)。
- 标准正态分布的概率密度函数及分布函数分别记作 $\varphi(u)$ 和 $\Phi(u)$，

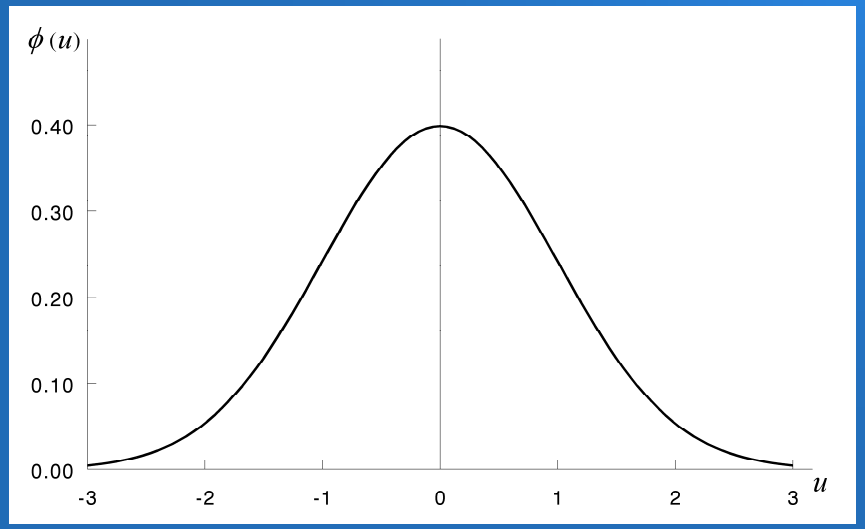
概率密度函数：$\varphi(u)=\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}$

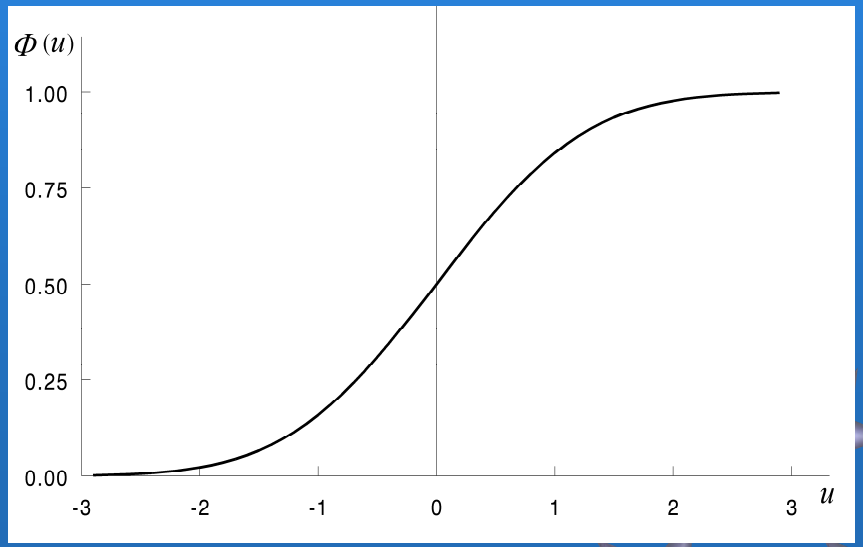
分布函数：$\phi(u)=P(U<u)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u}e^{-\frac{v^2}{2}}dv$

服从标准正态分布的随机变量，记作

$$U \sim N(0,1)$$

分布密度曲线如下图所示。





标准正态分布有以下特征：

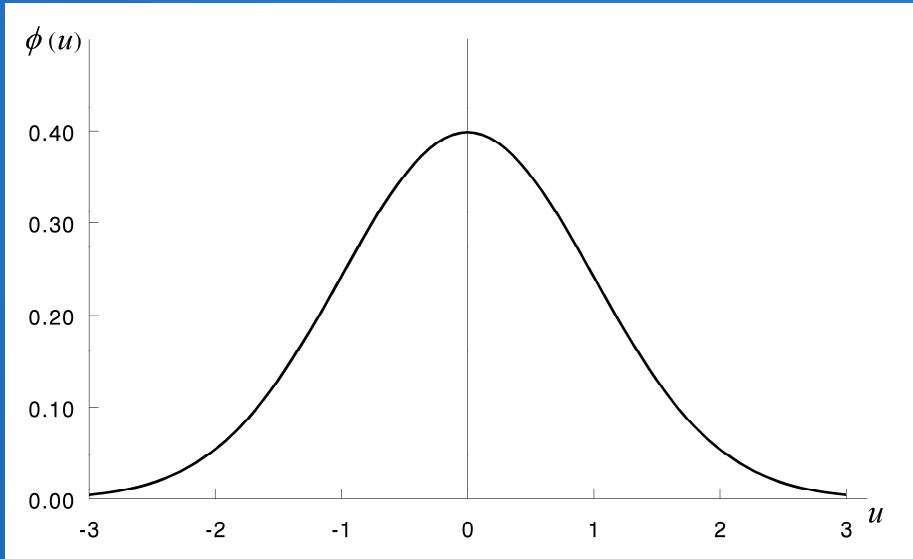
（1）u=0时，φ(u)达最大值；

（2）曲线纵坐标两侧对称，即φ(-u) = φ(u)；

（3）曲线在u±1处有两个拐点；

（4）曲线与横坐标轴所夹面积是1；

$$\sum p(u) = \int_{-\infty}^{+\infty} \varphi(u)du = 1 = \phi(+\infty) = 1$$



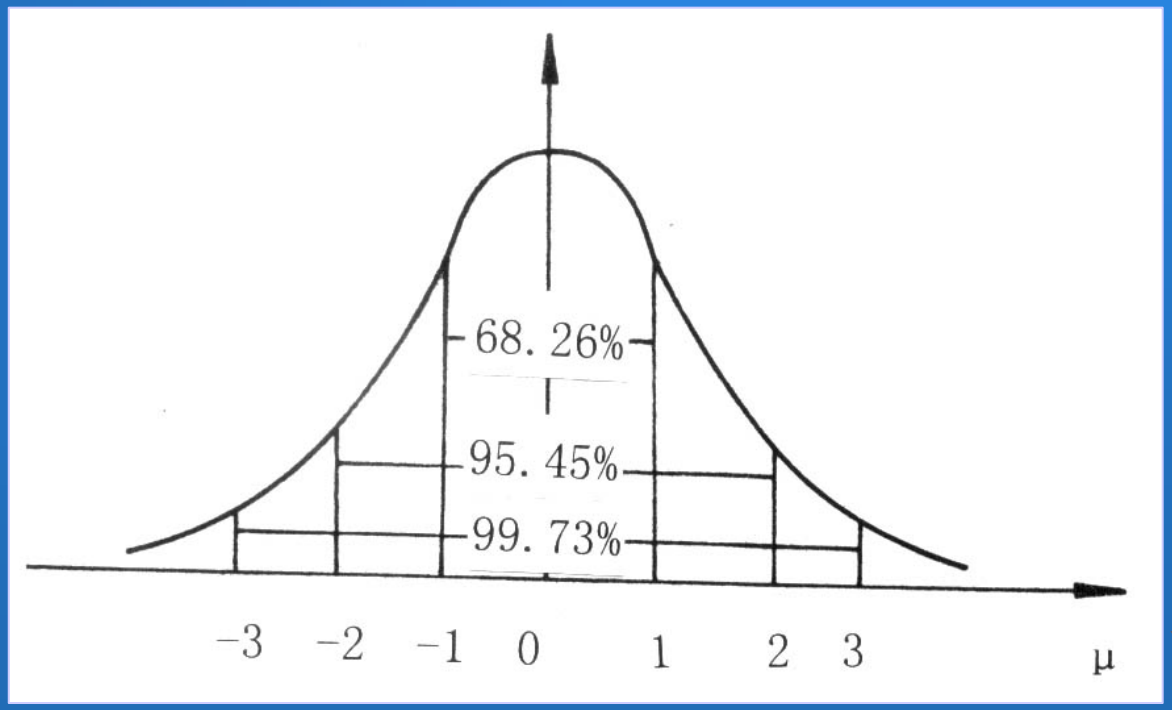
5.u=-1到1， 面积或概率=0.6827

u=-2到2， 面积或概率=0.9543

u=-3到3， 面积或概率=0.9973

u=-1.96到1.96， 面积或概率=0.95

u=-2.58到2.58， 面积或概率=0.99



（二）标准化的方法

对于任何一个服从正态分布$N(\mu, \sigma^2)$的随机变量x，都可以通过标准化变换：

$$u = \frac{x - \mu}{\sigma}$$

将 其变换为服从标准正态分布的随机变量u。

u 称 为 标 准 正 态变量或标准正态离差 **(standard normal deviate)**。

$$P(x_1 < X < x_2)$$

$$= P\left(\frac{x_1-\mu}{\sigma} < U < \frac{x_2-\mu}{\sigma}\right)$$

$$= P(u_1 < U < u_2)$$

$$= \phi(u_2) - \phi(u_1)$$

这表明服从正态分布 $N(\mu, \sigma^2)$的随机变量x 在 $[x_1, x_2)$ 内取值的概率， 等于服从标准正态分布的随 机变量u在$[(x_1-\mu)/\sigma, (x_2-\mu)/\sigma)$ 内取值的概率 。

因此，计算一般正态分布的概率时， 只要将区间的上下限作适当变换(标准化)， 就可用查标准正态分布的概率表的方法求得概率了。

3.3.3 正态分布的概率计算

（一）标准正态分布的概率计算

设$U \sim N(0,1)$ 分布，则U 在(u_1, u_2)内取值的概率为：

$$P(u_1 \le U \le u_2) = \Phi(u_2) - \Phi(u_1)$$

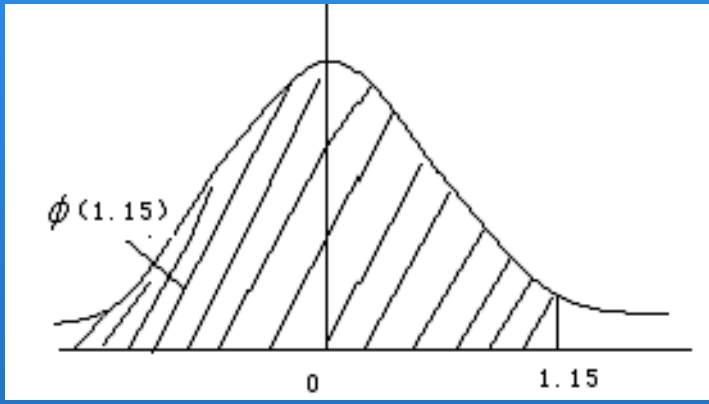
$\Phi(u_1)$与$\Phi(u_2)$可查表得到。

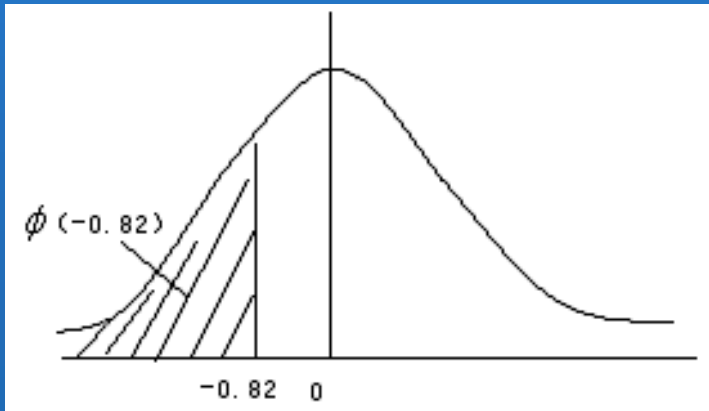
例如$u = 1.15$，查表

$\Phi(1.15) = p(U<1.15) = 0.87493$

$u = -0.82$，查表

$\Phi(-0.82) = p(U<-0.82) = 0.20611$





由正态分布的对称性可推出下列关系式：

$P(0 \leqslant U < u_1) = \Phi(u_1) - 0.5$

$P(U \geqslant u_1) = \Phi(-u_1)$

$P(|U| \geqslant u_1) = 2\Phi(-u_1)$ （两侧）

$P(|U| < u_1) = 1 - 2\Phi(-u_1)$ （中部）

$P(u_1 \leqslant U < u_2) = \Phi(u_2) - \Phi(u_1)$

例 随机变量U服从正态分布N（0，1），问随机变量的值落在0，1.21间的概率是多少？落在－1.96，1.96间的概率是多少？

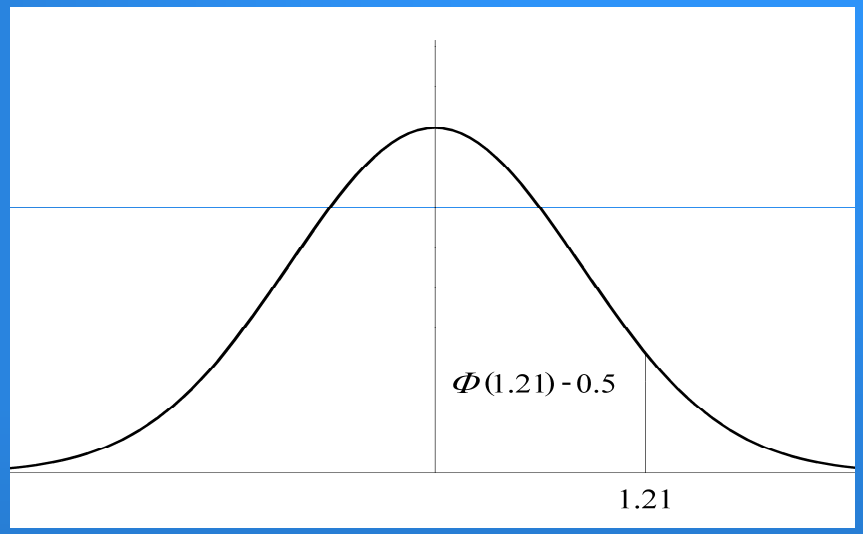
解： 1)

$P(0<U<u)$

$= \Phi(1.21)-0.5$

$=0.88686-0.5000$

$=0.38686$



2)

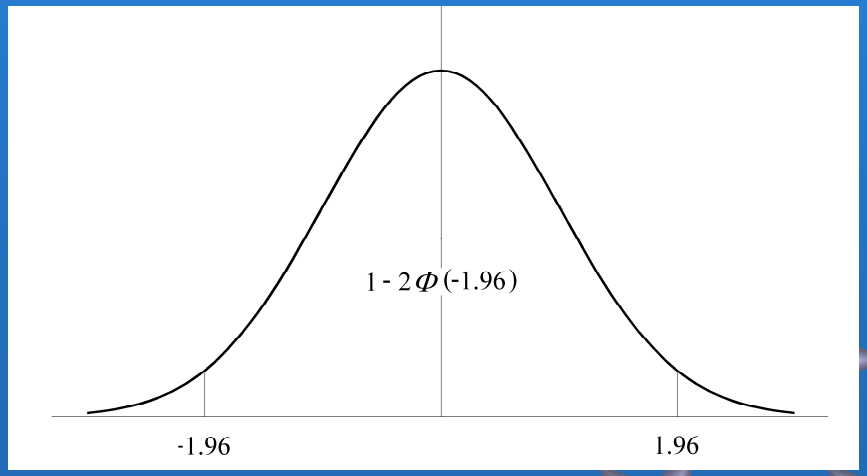
$P(|U|<u)$

$=1-2\Phi(-u)$

$=1-\Phi(-1.96)$

$=1-0.05000$

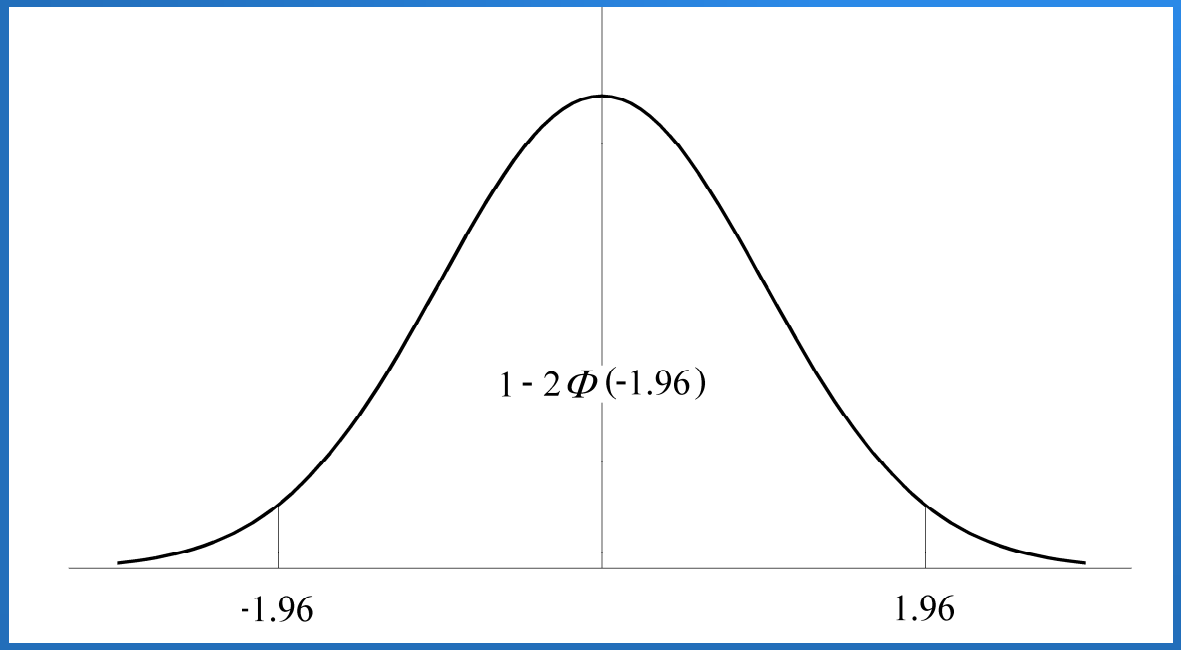
$=0.95000$



关于标准正态分布，以下几种概率应当熟记：

P（-1.96≤u＜1.96）=0.95

P (-2.58≤u＜2.58)=0.99



（二）一般正态分布的概率计算

❶将区间的上下限标准化，服从正态分布的随机变量x落在〔x1,x2〕内的概率，等于服从标准正态分布的随机变量u落在 $[(x_1-\mu)/\sigma,(x_2-\mu)/\sigma)$ 的概率。

❷然后查标准正态分布的概率表

[例]若x服从 $\mu = 30.26, \sigma^2 = 5.10^2$ 的正态分布，试求 $P(21.64 \le x < 32.98)$ 。

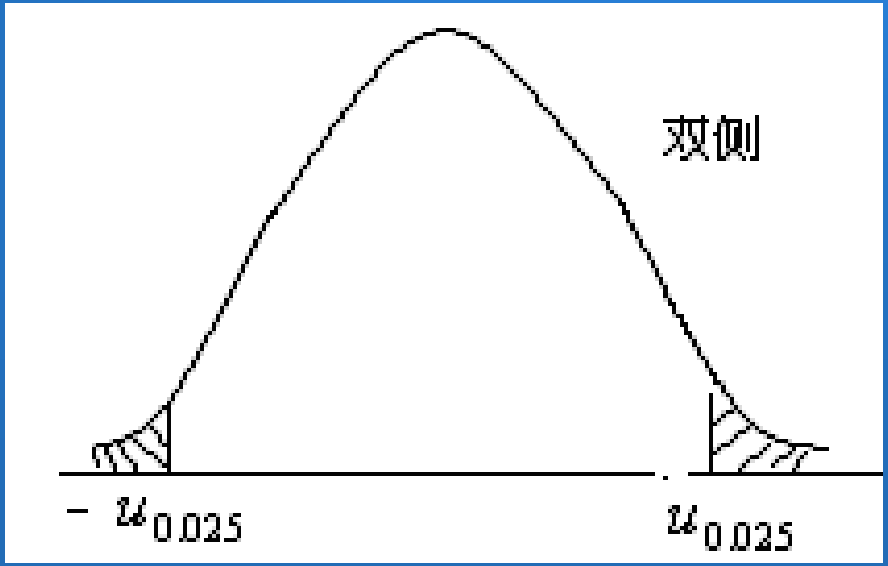
令u=(x-30.26)/5.10,则u服从标准正态分布，故

$$P(21.64 \le x < 32.98) = P\left(\frac{21.64-30.26}{5.10} \le \frac{x-30.26}{5.10} < \frac{32.98-30.26}{5.10}\right)$$

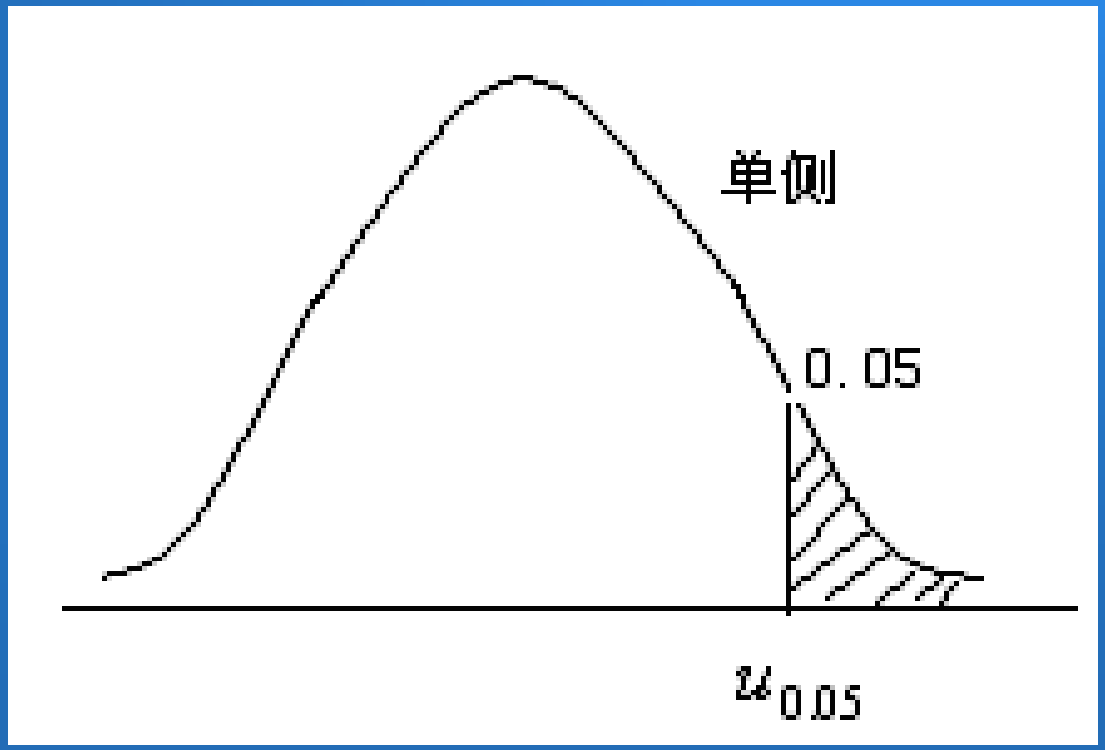
$$= P(-1.69 \le u < 0.53) = \Phi(0.53) - \Phi(-1.69) = 0.6564$$

3.3.4　正态分布的临界值

- 不仅注意x落在($\mu - k\sigma, \mu + k\sigma$)之内的概率而且 也很关心 x落在此区间之外的概率。
- 随机变量x落在平均数$\mu \pm K\sigma$区间之外的概率称为**双侧概率(两尾概率)**，记作α。



- 对应于双侧概率可以求得随机变量X小于 $\mu - k\sigma$或大于 $\mu + k\sigma$的概率，称为**单侧概率(一尾概率)**,记作 α 。

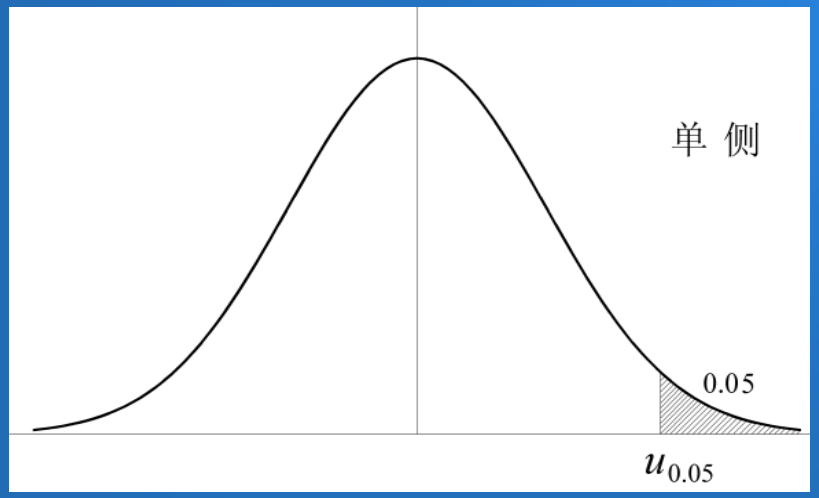


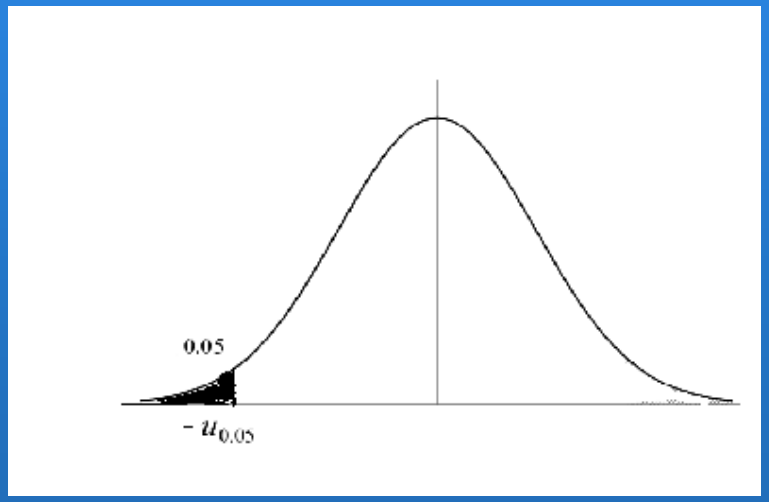
- 正态分布表给出的是U<u曲线下面积，反过来，如果求曲线右侧尾区一定面积（概率）下所对应的u的值， 查附表3。

例如，查右侧尾区曲线下面积P=0.05时的u值，得u=1.645。

以 α 表示尾区概率，则上述结果可表示为 $u_{0.05}=1.645$ 。

- 右侧尾区的$u_{0.05}$称为 α 的上侧分位数(临界值)，用 u_{α}表示，满足 $P(U>u_{\alpha})=\alpha$ 。
- 左侧尾区，满足$P(U<-u_{\alpha})=\alpha$， $-u_{\alpha}$ 称为 α 的下侧分位数。





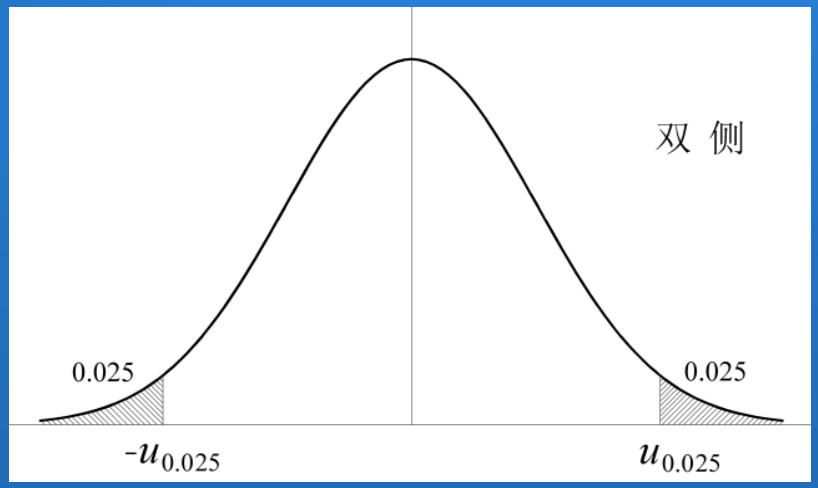
正态分布的上、下侧分位数

如果将α平分到两侧尾区，则每一尾区的曲线下面积，只有α /2。

- 满足 $p(|U|>u_{\frac{\alpha}{2}})=\alpha$ 时的 $u_{\frac{\alpha}{2}}$ 称为α的**双侧分位数**（临界值)。

例如查α=0.05的双侧分位数 $u_{\frac{0.05}{2}}$ ，得$u_{0.025}$=1.96

这与前面讲的P(|U|<1.96)=0.95一致的。



正态分布的双侧分为点

3.3.5 中心极限定理

设x_1，x_2,…… x_n从一个平均数为μ方差为σ2的正态总体得到的随机样本，则当n很大时，样本平均数的分布逼近正态分布N(μ，σ^2/n)

$$E(\bar{x}) = \mu \qquad Var(\bar{x}) = Var\left(\frac{1}{n}\sum x_i\right) = \frac{1}{n}\sigma^2$$

化为标准正态分布U~N(0,1)

$$U = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

作业：

Page 52　3.12, 3.13, 3.14