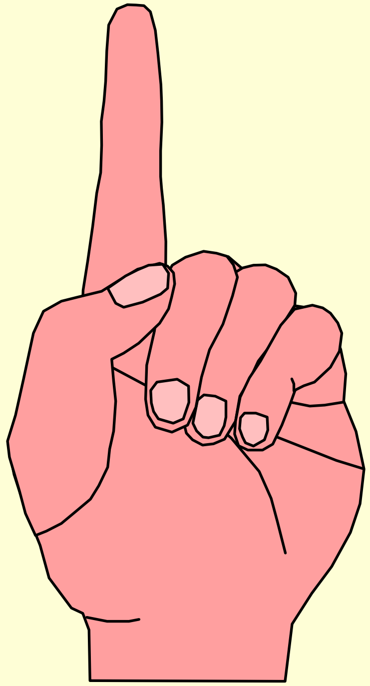
第二章 随机变量及其分布

- 随机变量
- 离散型随机变量的分布率
- 随机变量的分布函数
- 连续型随机变量及其概率密度
- 随机变量函数的分布



§1 随机变量

实例: 做试验抛一枚均匀硬币，其样本空间

$$S=\{\omega\}=\{H,T\}$$

可规定映射

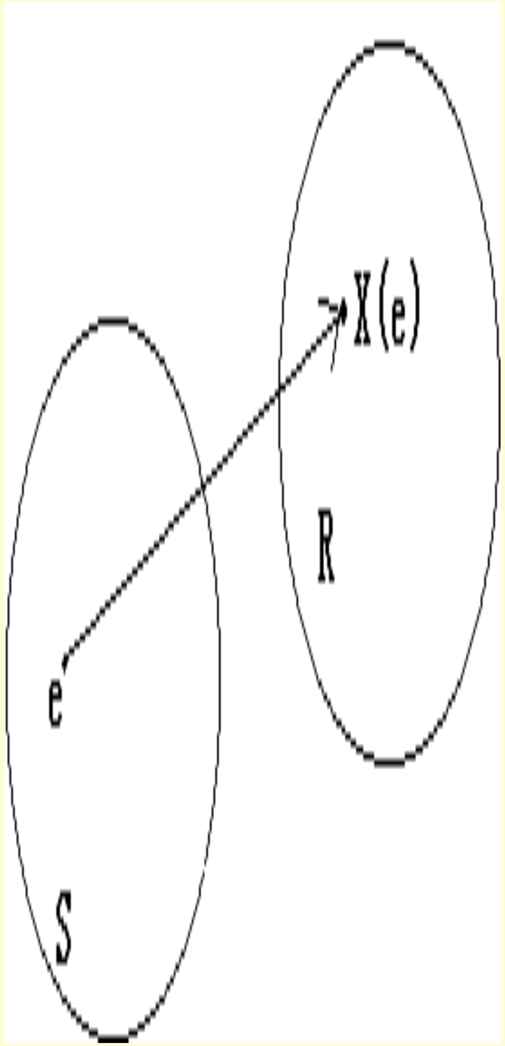
$$X=X(\omega)=\begin{cases}1, & \omega=H\\ 0, & \omega=T\end{cases}$$

随机变量实际上是定义在样本空间上的一个实函数。

$$X: S \to R$$

定义．设S={e}是试验的样本空间，如果量X是定义在S上的一个单值实值函数，即对于每一个e∈S，有唯一确定的实数X=X(e)与之对应，则称X为**随机变量**。

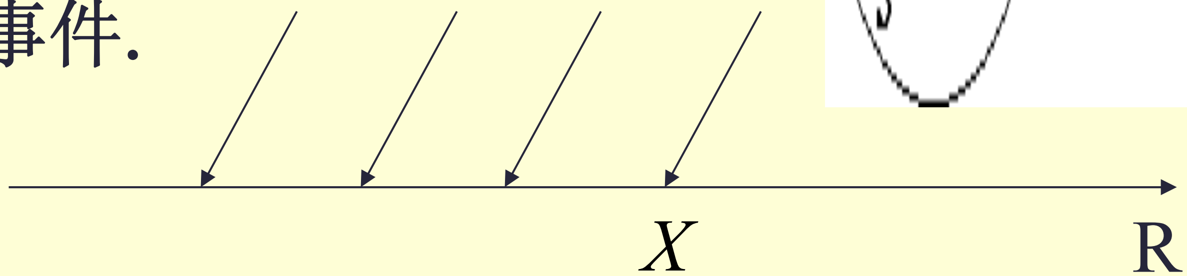
随机变量常用X、Y、Z 或 ξ、η、ζ等表示。记为r. v. X等。



引入随机变量的意义:

1. 将随机试验的结果数量化。
2. 描述随机事件.

几何意义:



例1：引入适当的随机变量描述下列事件：

①将3个球随机地放入三个格子中，事件 A={有1个空格}，B={有2个空格}，C={全有球}。

②进行5次试验，事件 D={试验成功一次}，F={试验至少成功一次}，G={至多成功3次}

解：① 设X为将3个球随机地放入三个格子后的空格数，则

A={X=1}，B={X=2}，C={X=0}

② 设Y为进行5次试验中成功的次数，则

D={Y=1}，F={Y≥1}，G={Y≤3}

随机变量的分类

随机变量
- 离散型随机变量
- 非离散型
  - 连续型
  - 奇异型（混合型）

§2 离散型随机变量的分布律

定义

若随机变量X取值$x_1, x_2, \ldots, x_n, \ldots$， 且取这些值的概率依次为$p_1, p_2, \ldots, p_n, \ldots$, 则称

$$P\{X=x_k\}=p_k,\ (k=1, 2, \ldots)$$

为X的分布律。

可表为

$$X\sim\ P\{X=x_k\}=p_k,\quad (k=1,\ 2,\ \cdots),$$

或…

$$X\sim\ \begin{array}{c|ccccc} X & x_1 & x_2 & \cdots & x_K & \cdots \\ \hline P_k & p_1 & p_2 & \cdots & p_k & \cdots \end{array}$$

● 分布律的性质

(1) $p_k \geq 0,\ k=1, 2, \ldots$;

(2) $\sum_{k\geq 1} p_k = 1.$

例1 设袋中有5只球，其中有2只白3只黑。现从中任取3只球(不放回)，求抽得的白球数X的分布律。

解: X的可能取值为0，1，2

$$P\{X=k\}=\frac{C_2^k C_3^{3-k}}{C_5^3}.\quad k=0,1,2$$

对离散型随机变量来说, 概率分布律可以完全描述它的统计规律. 换句话说, 已知分布律, 就可以求出各种概率.

$$P(X \in (a,b)) = \sum_{x_i \in (a,b)} P(X = x_i)$$

例2　设随机变量X的分布律为

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 0.1 & 0.15 & 0.2 & 0.3 & 0.12 & 0.1 & 0.03 \end{pmatrix}$$

试求：$P(X \le 4), P(2 \le X \le 5), P(X \ne 3)$

解：0.87　0.72　0.7

几种常用的离散型随机变量

1. (0-1)分布

若X只能取0、1两个值，且 分布律为

$$P\{X=k\}=p^k(1-p)^{1-k},\ k=0,1。\ (0<p<1)$$

则称X服从参数为p的0—1分布或两点分布。

即

X	1	0
p_k	p	$1-p$

2. 二项分布

· 贝努利试验：若试验E只有两个结果，记为A、$\overline{A}$.

· n重贝努利试验：独立重复的进行n次贝努利试验。

a. 每次试验均为贝努利试验，只有两个结果。

b. 重复，指每次试验P(A)不变，为定值。

c. 独立，指某次试验事件A发生与否与其它次试验事件A发生与否互不影响。

例3. 某射手对目标独立射击5次，每次命中目标的概率为p，以X表示命中目标的次数，求X的分布律。

解：设Ai——第i次射击时命中目标，i=1, 2, 3, 4, 5

则$A_1, A_2, \ldots A_5$, 相互独立且 $P(A_i)=p, i=1, 2, \ldots 5.$

$S_X=\{0, 1, 2, 3, 4, 5\}$,

$$P\{X=0\}=P(\overline{A_1}\,\overline{A_2}\,\overline{A_3}\,\overline{A_4}\,\overline{A_5})=(1-p)^5$$

$$P\{X=1\}=P\{A_1\overline{A_2}\,\overline{A_3}\,\overline{A_4}\,\overline{A_5}\cup\overline{A_1}A_2\overline{A_3}\,\overline{A_4}\,\overline{A_5}\cup\ldots\}=5p(1-p)^4$$

$$P\{X=2\}=P\{A_1A_2\overline{A_3}\,\overline{A_4}\,\overline{A_5}\cup A_1\overline{A_2}A_3\overline{A_4}\,\overline{A_5}\cup\ldots\}=C_5^2P^2(1-P)^3$$

$$P\{X=k\}=C_5^k p^k(1-p)^{5-k}\qquad k=0,1,\ldots,5$$

若以X表示n重贝努里试验中事件A发生的次数，P(A)=p，则称X服从参数为n,p的二项分布。记作X~b(n,p)，其分布律为：

$$P\{X=k\}=C_n^k p^k(1-p)^{n-k},(k=0,1\ldots n)$$

例4 掷一颗骰子10次，求（1）双数点出现6次的概率？（2）“3”点出现两次的概率？

解：(1) 设X表出现双数点的次数，则X~b(10，1/2)

所求概率：$P(X=6)=C_{10}^6(\frac{1}{2})^6(\frac{1}{2})^{10-6}=C_{10}^6(\frac{1}{2})^{10}$

(2) 设Y表出现“3”点的次数，则Y~b(10，1/6)

所求概率为：$P(Y=2)=C_{10}^2(\frac{1}{6})^2(\frac{5}{6})^8$

例5. 某人射击的命中率为0.02，他独立射击400次，试求其命中次数不少于2的概率。

解　　设X表示400次独立射击中命中的次数，

则X～B(400, 0.02)，故

$P\{X\geq 2\} = 1 - P\{X=0\} - P\{X=1\}$

$= 1 - 0.98^{400} - (400)(0.02)(0.98^{399}) = \ldots$

泊松定理　设随机变量X~B(n, p), (n＝0, 1, 2,…), 且n很大，p很小，记λ=np，则

$$P\{X=k\} \approx \frac{\lambda^k}{k!}e^{-\lambda}, \qquad k=0,1,2,\ldots$$

上题用泊松定理 取λ =np＝(400)(0.02)＝8，故

近似地有

$$P\{X\ge 2\}=1-P\{X=0\}-P\{X=1\}$$

$$=1-(1+8)e^{-8}=0.996981.$$

（二.） 泊松(Poisson)分布

$$X\sim P\{X=k\}=\frac{\lambda^k}{k!}e^{-\lambda},\quad k=0,1,2,\ldots$$

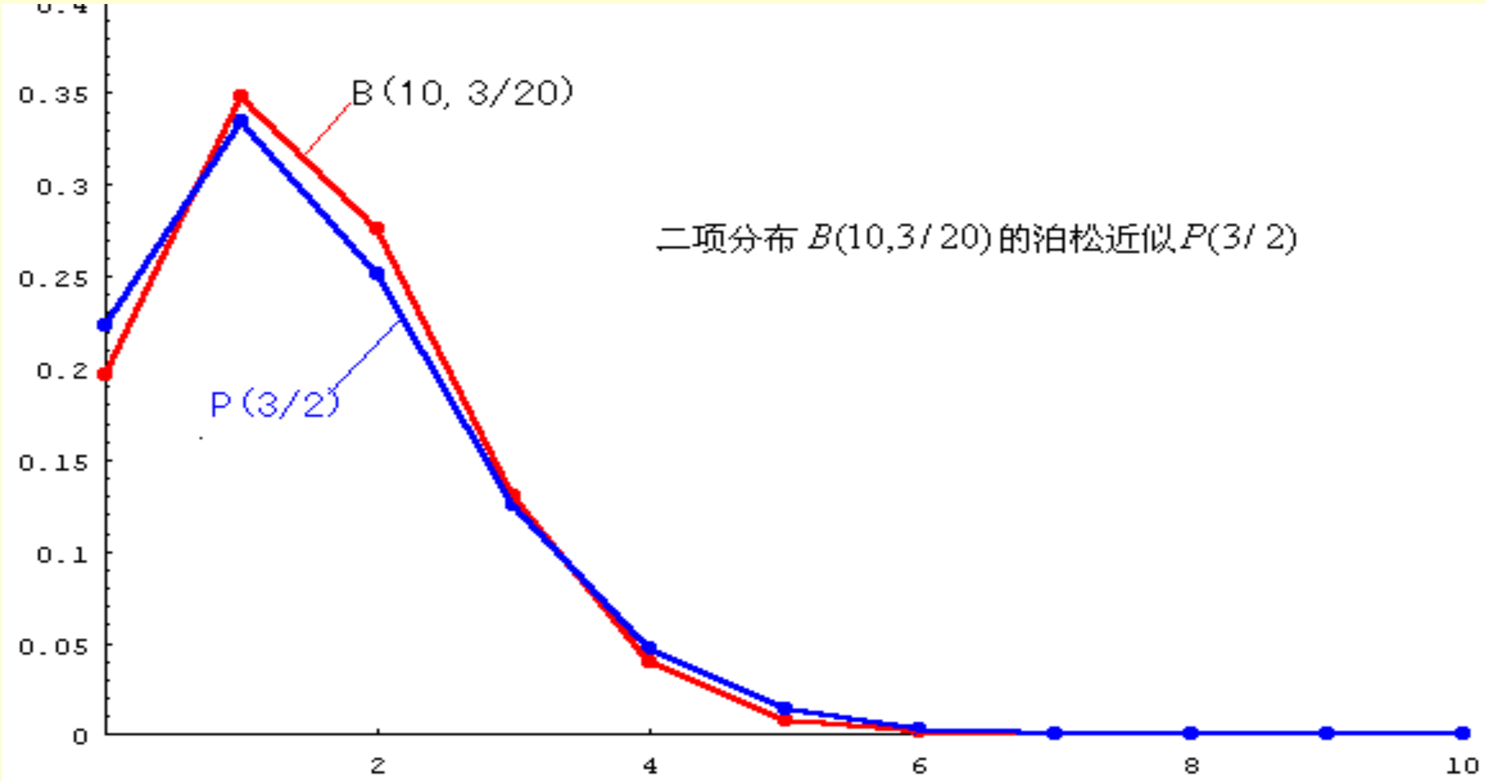
$$(\lambda>0)$$

则称r.v.X服从参数为λ的泊松分布。记为：$X\sim\pi(\lambda)$

泊松定理表明，泊松分布是二项分布的极限分布，

当**n**很大，**p**很小时，二项分布就可近似地

看成是参数**λ=np**的泊松分布



例4：某信息服务台在一分钟内接到的问讯次数X服从参数为λ的泊松分布，已知任一分钟内无问讯的概率为e^{-6}，求在指定的一分钟内至少有2次问讯的概率。

解：$\because X \sim \pi(\lambda)$，且$P\left(X=0\right)=e^{-6}$

即 $e^{-\lambda}=e^{-6} \Rightarrow \lambda=6$

$$\therefore \begin{aligned} &P\{X \geq 2\}=1-P\{X<2\} \\ &=1-P\{X=0\}-P\{X=1\} \end{aligned}$$

$$=1-e^{-6}-6e^{-6} \approx 0.9826$$

例5：设书中每一页上印刷错误个数服从参数为λ=1/2的泊松分布，求（1）一页上至少有一处印错的概率？(2) 10页中至多有一页有错的概率？

解：(1) 设X为一页上印刷错误的个数，则 $X \sim \pi(1/2)$

所求概率为：

$$P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-\frac{1}{2}} \approx 0.395$$

(2) 设Y为10页中有错的页数，则

$$Y \sim b(10, 0.395)$$

所求概率为：

$$P(Y \leq 1) = P(Y = 0) + P(Y = 1) \approx 0.049$$

想一想：离散型随机变量的统计特征可以用分布律描述，非离散型的该如何描述？

如：熊猫彩电的寿命X是一个随机变量，事件{X=5年}的概率为多少呢？

描述非离散随机变量统计特征，我们讨论它落在某区间的概率。

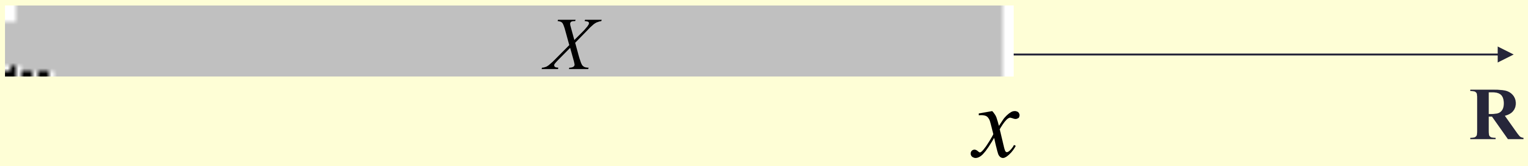
这相当于，只要知道，对任意实数x，事件{X≤x}的概率.

$$\because \{a < X \le b\} = \{X \le b\} - \{X \le a\}$$

§3 随机变量的分布函数

定义

（P31） 设X是随机变量，x是任意实数，函数 $F(x)=P\{X\leq x\}$ 称为随机变量X的**分布函数**。



易知，对任意实数a, b (a<b),

$$P\{a<X\leq b\}=P\{X\leq b\}-P\{X\leq a\}=F(b)-F(a).$$

分布函数的性质

1、单调不减性：若$x_1<x_2$，则$F(x_1)\le F(x_2)$;

2、归一 性：对任意实数x，$0\le F(x)\le 1$，且

$$F(-\infty)=\lim_{x\to-\infty}F(x)=0,$$

$$F(+\infty)=\lim_{x\to+\infty}F(x)=1;$$

3、右连续性：对任意实数x_0，

$$F(x_0+0)=\lim_{x\to x_0^+}F(x)=F(x_0).$$

反之，具有上述三个性质的实函数，必是某个随机变量的分布函数。故该三个性质是分布函数的充分必要性质。

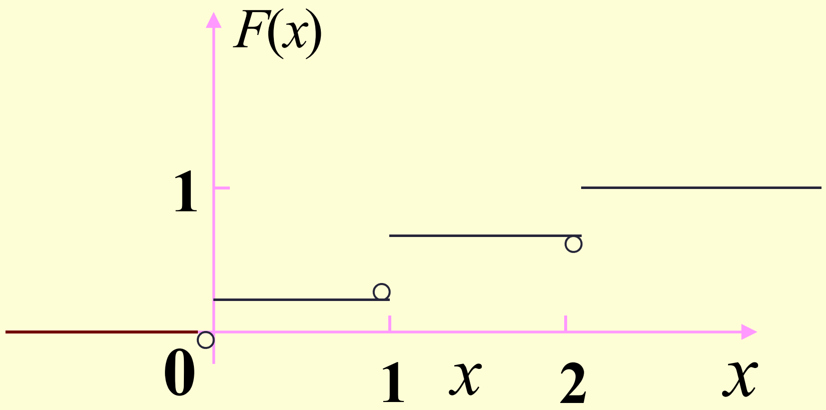
例1: 设随机变量X分布律如右表

试求出X的分布函数。

X	0	1	2
P	0.1	0.6	0.3

解: $F(x)=P\{X\le x\}$

当 $x<0$ 时, $F(x)=0$



当$0\leqslant x<1$ 时,

$F(x)=P\{X\leqslant x\}=P\{X=0\}=0.1$

当$1\leqslant x<2$ 时,

$F(x)=P\{X\leqslant x\}=P\{X=0\}+P\{X=1\}=0.1+0.6=0.7$

当$x\geqslant 2$ 时, $F(x)=P\{X\leqslant x\}=P\{X=0\}+P\{X=1\}+P\{X=2\}=1$

一般地，对离散型随机变量

$$X \sim P\{X = x_k\} = p_k, \quad k = 1, 2, \ldots$$

其分布函数为 $F(x) = P\{X \le x\} = \sum_{k: x_k \le x} p_k$

离散型随机变量的分布函数是阶梯函数，分布函数的跳跃点对应离散型随机变量的可能取值点，跳跃高度对应随机变量取对应值的概率；

反之，如果某随机变量的分布函数是阶梯函数，则该随机变量必为离散型.

利用分布函数计算概率的一些公式

$$(1)\quad P(a < X \le b) = F(b) - F(a)$$

$$(2)\quad P(X > a) = 1 - F(a)$$

$$(3)\quad P(X = a) = \lim_{x \to a+} F(x) - \lim_{x \to a-} F(x)$$

$$= F(a) - F(a-0)$$

$$(4)\quad P(X < a) = F(a-0)$$

例2：设离散r.v. X的分布函数为：

$$F(x)=\begin{cases} A & x<0 \\ \dfrac{1}{2} & 0\le x<1 \\ \dfrac{2}{3} & 1\le x<2 \\ \dfrac{11}{12} & 2\le x<3 \\ B & x\ge 3 \end{cases}$$

求 r.v.X的分布律，并求 $P\{X\le 3\}$，$P\{X>0.5\}$，$P\{2\le X<4\}$

解：$F(-\infty)=A=0,\ F(+\infty)=B=1$ $\quad P\{X\le 3\}=1,\ P\{X>0.5\}=\dfrac{1}{2},$

X	0	1	2	3
P_k	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{12}$

$$P\{2\le X<4\}$$
$$=F(4-0)-F(2-0)=1/3$$

例3：向[0,1]区间随机抛一质点，以X表示质点坐标.假定质点落在[0,1]区间内任一子区间内的概率与区间长成正比，求X的分布函数

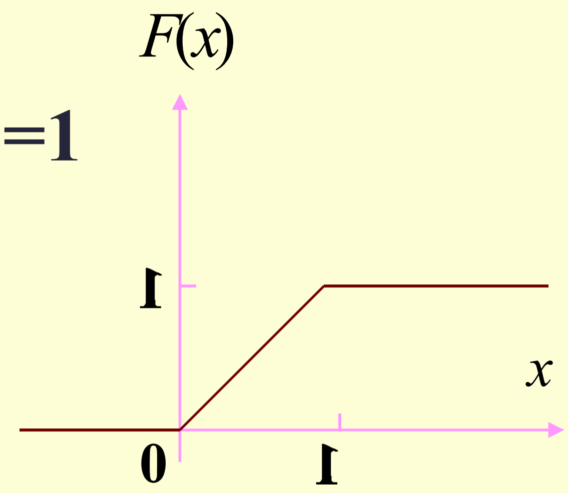
解：　F(x)=P{X≤x}

当x<0时,F(x)=0;　　当x>1时,F(x)=1

当0≤x≤1时,

$$F(x) = P\{0 \le X \le x\} = kx$$

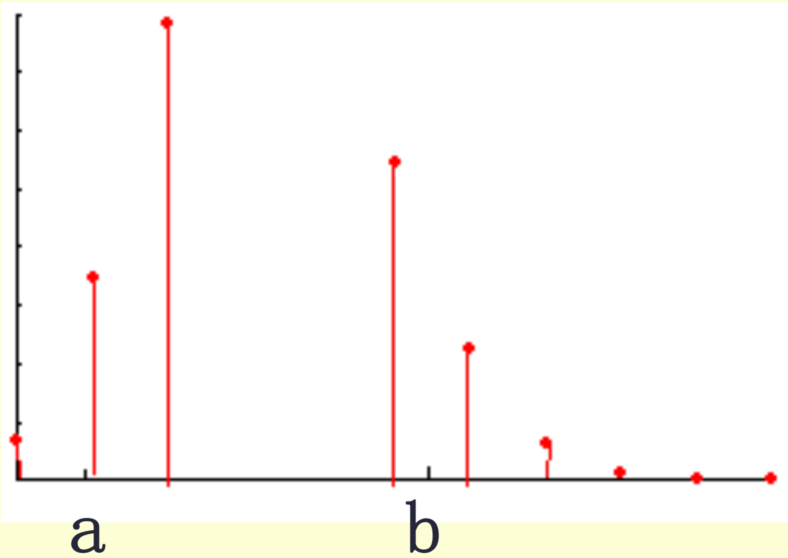
特别,F(1)=P{0≤X≤1}=k=1

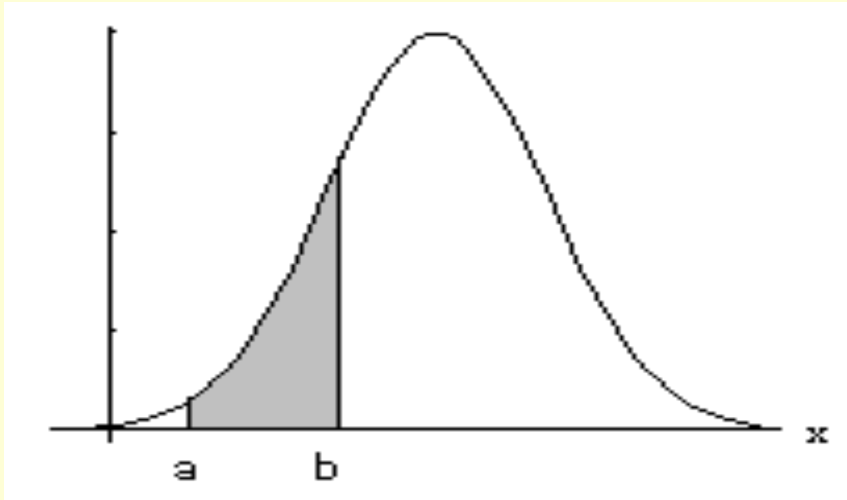


$$\therefore F(x)=P(X \le x)=\begin{cases} 0, & x<0 \\ x, & 0 \le x \le 1 \\ 1, & x>1 \end{cases}$$

用分布函数描述随机变量不如分布律直观，

对非离散型随机变量，是否有更直观的描述方法？





$$p\{a < X \le b\} = ?$$

§ 4 连续型随机变量及其概率密度

一、概率密度

1. 定义　对于随机变量X，若存在非负函 数f(x)，(-∞<x<+∞)，使对任意实数x，都有

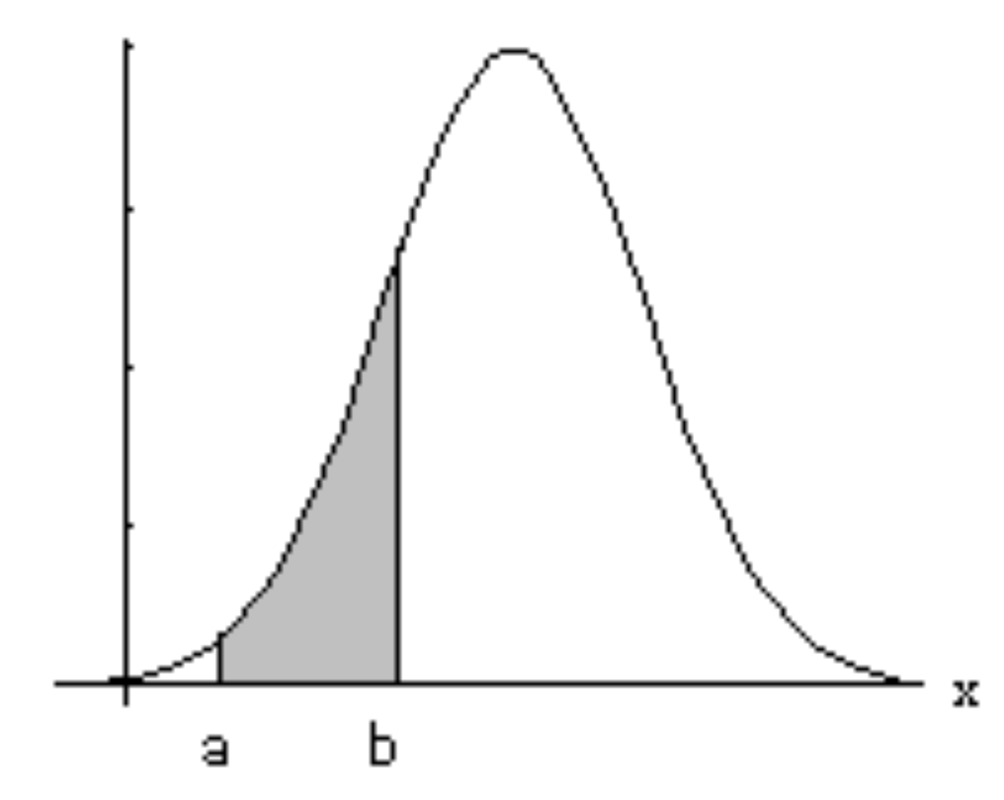
$$F(x)=P(X\le x)=\int_{-\infty}^{x} f(u)\,du$$

则称X为连续型随机变量， f(x)为X的**概率密度函数**，简称概率密度或密度函数.

常记为:

X～ f(x) ，(-∞<x<+∞)

密度函数的几何意义为



$$P(a < X \leq b) = \int_a^b f(u)\,du$$

随机变量**X**取值于区间**(a, b]**的概率等于**(a, b]**区间上曲线**f(x)**覆盖下的曲边梯形的面积

2. 密度函数的性质

(1) 非负性　f(x)≥0，(-∞<x<∞)；

(2) 归一性　$\int_{-\infty}^{+\infty} f(x)dx = 1.$

性质(1)、(2)是密度函数的充要性质；

设随机变量X的概率密度为

$$f(x) = ae^{-|x|}$$

求常数a.

答:

$$a = \frac{1}{2}$$

(3) 若x是f(x)的连续点，则

$$\frac{dF(x)}{dx}=f(x)$$

EX

设随机变量X的分布函数为

求$f(x)$

$$F(x)=\begin{cases}\frac{1}{2}e^{x} & x<0\\ 1-\frac{1}{2}e^{-x} & x\geq 0\end{cases}$$

（4） 对任意实数b，若X～ f(x)，

(-∞<x<∞)，则P{X=b}＝0。

于是

$$P\{a<X<b\}=P\{a\le X<b\}$$

$$=P\{a\le X\le b\}=\int_a^b f(x)dx$$

例1：已知随机变量X的概率密度为

$$f(x)=\begin{cases}2Ax & 0<x<1\\ 0 & 其他\end{cases}$$

（1）求参数A. (2) P{0.5<X<3}. (3) 求分布函数F(X).

解：

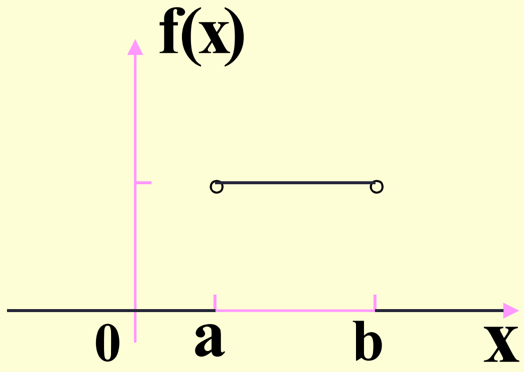
(1) $$\int_{-\infty}^{\infty} f(x)dx = \int_0^1 2Axdx = 1 \quad \Rightarrow \quad A=1$$

(2) $$P\{0.5<X<3\} = \int_{0.5}^{3} f(x)dx = \int_{0.5}^{1} 2xdx = 0.75$$

(3) $$F(x) = P\{X \le x\} = \int_{-\infty}^{x} f(t)dt = \begin{cases}0 & x\le 0\\ x^2 & 0<x\le 1\\ 1 & x>1\end{cases}$$

二、几个常用的连续型分布

1. 均匀分布



若X～f(x)＝$\begin{cases} \dfrac{1}{b-a}, a<x<b \\ 0 \quad ，其它 \end{cases}$

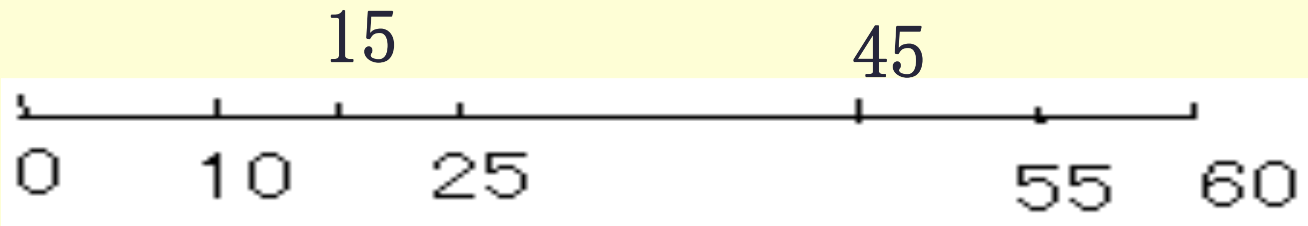
则称X在(a, b)内服从均匀分布。记作 **X~U(a, b)**

对任意实数c, d (a<c<d<b)，都有

$$P\{c<X<d\}=\int_c^d f(x)dx=\int_c^d \frac{1}{b-a}dx=\frac{d-c}{b-a}$$

这说明X落在(a, b)中任一区间的概率只与该区间的长度成正比，而与该区间的位置无关，这就是均匀分布的概率意义。

例2. 长途汽车起点站于每时的10分、25分、55分发车，设乘客不知发车时间，于每小时的任意时刻随机地到达车站，求乘客候车时间超过10分钟的概率.



解：设A—乘客候车时间超过10分钟

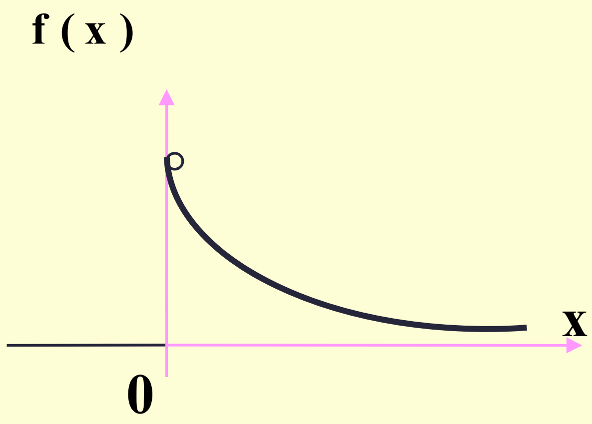
X—乘客于某时X分钟到达，则X~U(0,60)

$$P(A)=P\{10<X\le 15\}+P\{25<X\le 45\}+P\{55<X\le 60\}$$

$$=\frac{5+20+5}{60}=\frac{1}{2}$$

2. 指数分布

若 X～ $f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0\\ 0, & x\le 0\end{cases}$



则称X服从参数为λ>0 的指数分布。
其分布函数为

$$F(x)=\begin{cases}1-e^{-\lambda x}, & x>0\\ 0, & x\le 0\end{cases}$$

例3. 电子元件的寿命X(年）服从参数为3的指数分布

(1) 求该电子元件寿命超过2年的概率。

(2) 已知该电子元件已使用了1.5年，求它还能使用两年的概率为多少?

解:

$$f(x)=\begin{cases} 3e^{-3x} & x>0 \\ 0 & x\le 0, \end{cases}$$

$$(1)\ P\{X>2\}=\int_2^{\infty} 3e^{-3x}dx = e^{-6}$$

$$(2)P\{X > 3.5 \mid X > 1.5\}$$

$$= \frac{P\{X > 3.5, X > 1.5\}}{P\{X > 1.5\}}$$

$$= \frac{\int_{3.5}^{\infty} 3e^{-3x}dx}{\int_{1.5}^{\infty} 3e^{-3x}dx} = e^{-6}$$

例4. 某公路桥每天从零时刻开始到第一辆汽车过桥经过的时间为T，设每t时段内过桥的汽车数X_t服从参数为λt的泊松分布，求T的概率密度。

解: $F(t)=P\{T\le t\}$

当$t\le 0$时， $F(t)=0$

当$t>0$时， $F(t)=P\{T\le t\}=1-P\{T>t\}$

=1-P\{在[0，t]时段内无汽车过桥\}

$$=1-P\{X_t=0\}=1-e^{-\lambda\cdot t}$$

于是

$$f(t)=F'(t)=\begin{cases}\lambda e^{-\lambda t} & t>0\\ 0 & t\le 0\end{cases}$$

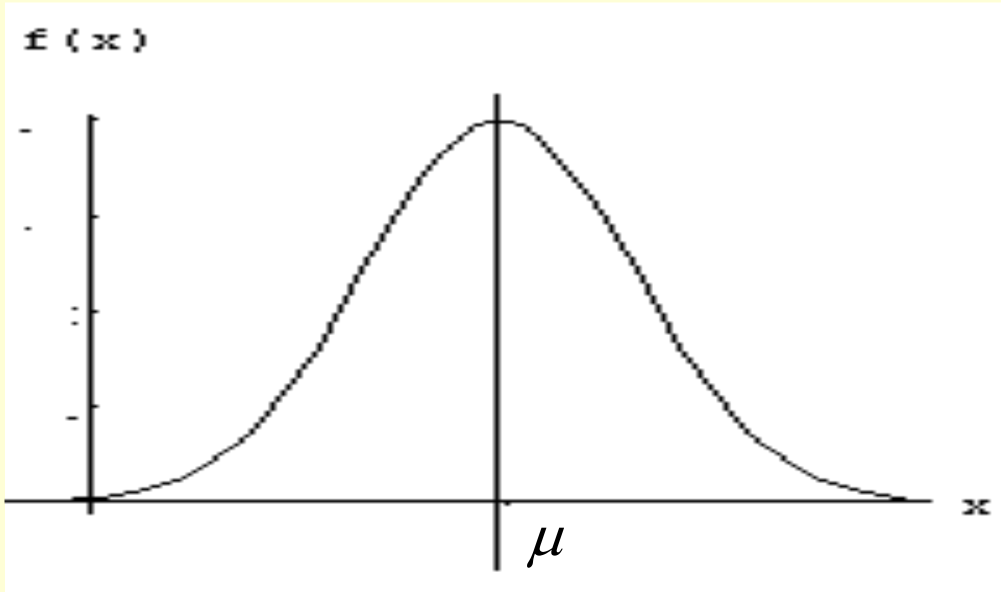
3. 正态分布

正态分布是实践中应用最为广泛，在理论上研究最多的分布之一，故它在概率统计中占有特别重要的地位。

A，B间真实距离为μ，测量值为**X**。

X的概率密度应该是什么形态？

若随机变量



$$X \sim f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in R$$

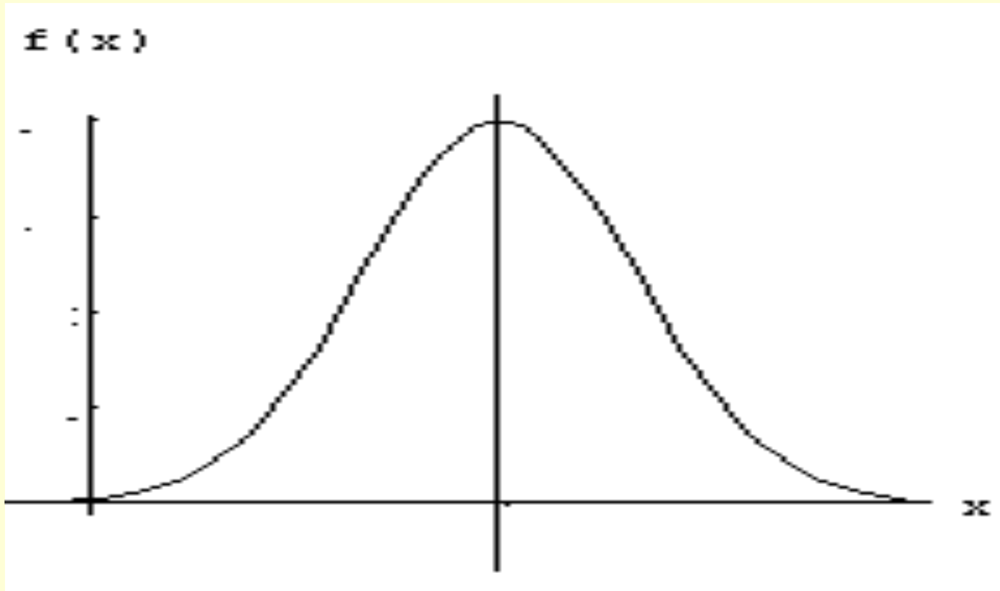
其中 μ为实数， σ>0 ，则称**X**服从参数为μ ,σ^2的正态分布,记为**N(μ, σ^2)**，可表为**X～N(μ, σ^2).**

正态分布有两个特性：

(1) 单峰对称

密度曲线关于直线**x=μ**对称；

$$f(\mu)=\max f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}$$



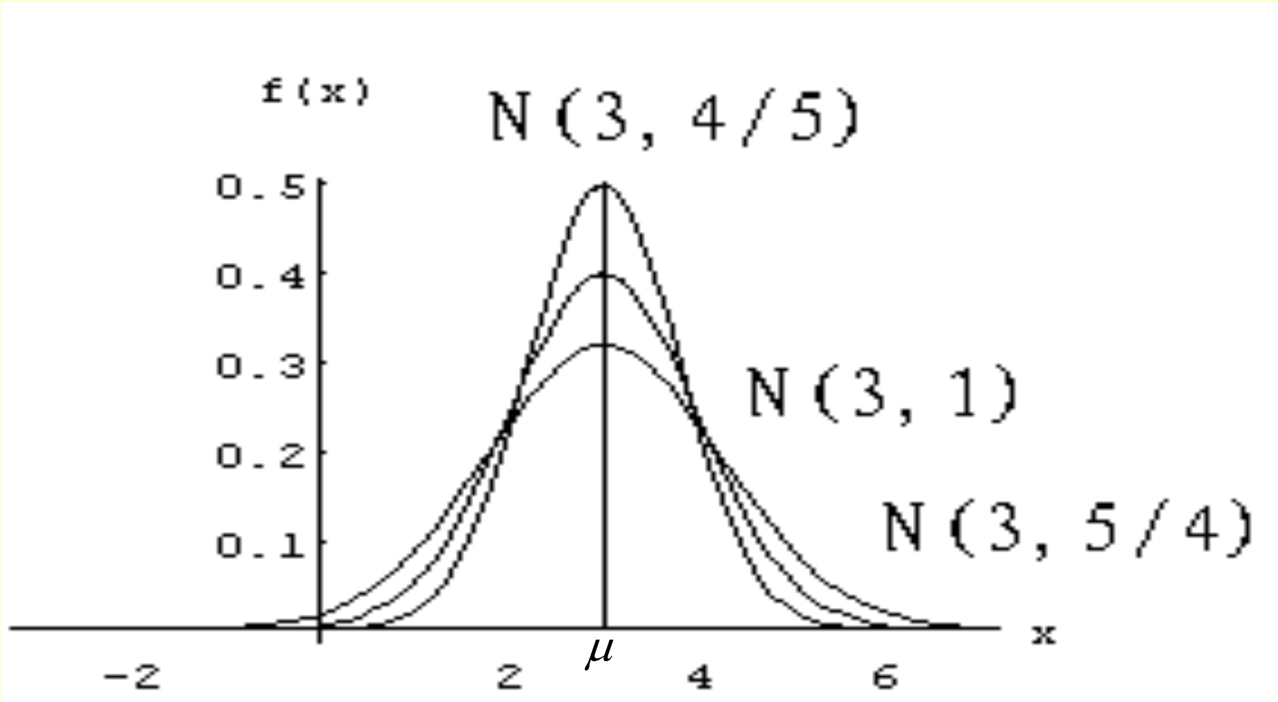
μ

(2) σ的大小直接影响概率的分布

σ越大，曲线越平坦，

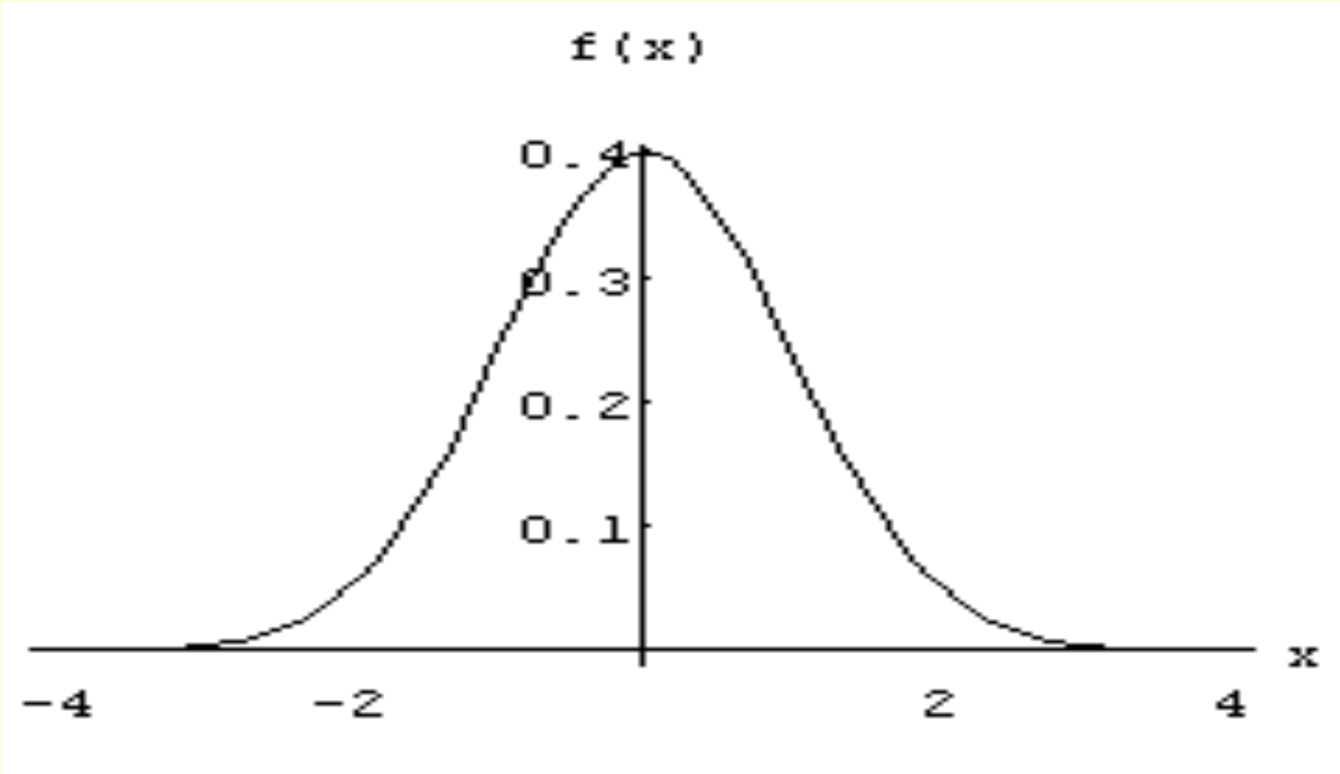
σ越小，曲线越陡峭。

正态分布也称为高斯(**Gauss**)分布



4. 标准正态分布

参数$\mu=0$，$\sigma^2=1$的正态分布称为标准正态分布，记作**X~N(0, 1)**。



其密度函数表示为

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, -\infty < x < +\infty.$$

分布函数表示为

$$\Phi(x) = P\{X \le x\}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt, -\infty < x < +\infty$$

一般的概率统计教科书均附有标准正态分布表供读者查阅$\Phi(x)$的值。如，若

$X \sim N(0,1)$，$\Phi(0.55)=0.7088$,

$P\{1.32<X<2.43\}=\Phi(2.43)-\Phi(1.32)$ 正态分布表

$=0.9925-0.9066$

注：(1) $\Phi(x)=1-\Phi(-x)$；

(2) 若$X \sim N(\mu, \sigma^2)$，则

$$F(x)=P\{X \le x\}=\Phi\left(\frac{x-\mu}{\sigma}\right).$$

正态分布表

例5　设 $X \sim N(\mu, \sigma^2)$，求 $P\{\mu-3\sigma<X<\mu+3\sigma\}$.

本题结果称为3σ原则.在工程应用中，通常认为 $P\{|(X-\mu)/\sigma|\leqslant 3\}\approx 1$，忽略 $\{|(X-\mu)/\sigma|>3\}$ 的值.

如在质量控制中，常用标准指标值 $\pm 3\sigma$ 作两条线，当生产过程的指标观察值落在两线之外时发出警报.

表明生产出现异常.

例6 一种电子元件的使用寿命X（小时）服从正态分布$N(100,15^2)$,某仪器上装有3个这种元件，三个元件损坏与否是相互独立的.求：使用的最初90小时内无一元件损坏的概率.

解: 设Y为使用的最初90小时内损坏的元件数, 则$Y\sim b(3,p)$

其中

$$p = P\{X < 90\} = \Phi(\frac{90-100}{15}) \approx \Phi(-0.67) = 0.2514$$

故

$$P\{Y = 0\} = (1-p)^3 \approx 0.4195$$

正态分布表