第二章　随机变量及其分布

关键词：

随机变量

概率分布函数

离散型随机变量

连续型随机变量

随机变量的函数

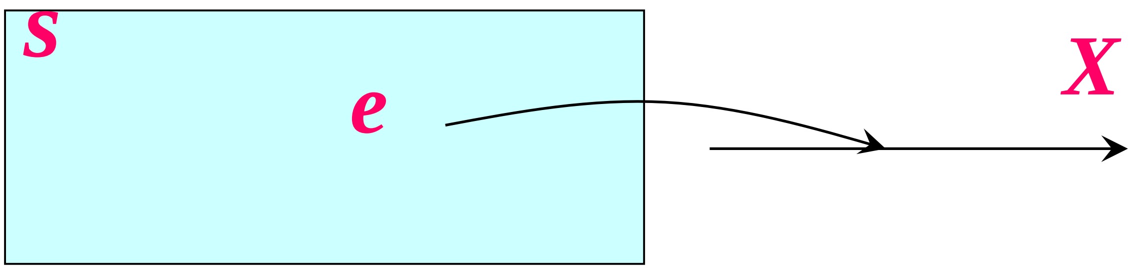
§1　随机变量

常见的两类试验结果：

示数的——降雨量；
候车人数；
发生交通事故的次数…

示性的——明天天气（晴，云…）；
化验结果（阳性，阴性）…

中心问题：将试验结果数量化



$X=X(e)$为S上的单值函数，X为实数.

- 例：任意投掷两骰子，研究两骰子之和。

则S={ (1,1), (1,2), …(1,6), (2,1), (2,2),…, (6,6) }共36个样本点

A={两骰子之和为7}={ (1,6), (2,5), (3,4), (4,3), (5,2),(6,1) }

若设两骰子之和为X，则{X=7}={ (1,6), (2,5), (3,4), (4,3), (5,2),(6,1) }，并且可方便列出其他值及概率。

X	2	3	4	5	6	7	8	9	10	11	12
P	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

定义：设随机试验的样本空间为$S=\{e\}$,若

$X=X(e)$为定义在样本空间上的实值

单值函数，则$X=X(e)$称为**随机变量**。

- 一般采用大写英文字母X,Y,Z来表示随机变量
- 引入随机变量的目的是用来描述随机现象

常见的两类随机变量 $\left\{\begin{array}{l}\text{离散型的}\\ \\ \text{连续型的}\end{array}\right.$

§2　离散型随机变量及其分布

定义：随机变量的取值是有限个或可列个，称为离散型随机变量或称为离散量。

离散量的概率分布律形如:

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

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

样本空间**S**={ **X**=**x**$_1$, **X**=**x**$_2$, …, **X**=**x**$_n$, … }

由于各样本点两两不相容，所以：

$$1 = P(S) = \sum_{i=1}^{\infty} P(X = x_i) = \sum_{i=1}^{\infty} p_i$$

$$\text{概率分布律}\begin{cases} \text{1、写出所有可能取值——样本点} \\ \text{2、写出相应的概率——每一个样本点出现的概率} \end{cases}$$

*excel*可以通过“随机数发生器”产生离散分布数据.

例1：盒装有5个球，其中3红2白，现不放回任取2个，用X表示取到红球的个数，求X的概率分布律。

解：由于X的取值只可能有0,1,2，所以它是一个离散随机变量。

$$P(X=0)=\frac{C_3^0C_2^2}{C_5^2}=0.1$$

$$P(X=1)=\frac{C_3^1C_2^1}{C_5^2}=0.6$$

$$P(X=2)=\frac{C_3^2C_2^0}{C_5^2}=0.3$$

罗列形式概率分布律

$$P(X=i)=\frac{C_3^iC_2^{2-i}}{C_5^2},\quad i=0,1,2$$

公式形式概率分布律

X	0	1	2
p_i	0.1	0.6	0.3

表格形式概率分布律

几个重要的离散型随机变量

一、0－1分布

若X的分布律为：

随机变量只可能取0、1两个值

X	0	1
p	q	p

（ $p+q=1, p>0, q>0$）

则称X服从参数为p的0-1分布，或两点分布.

0-1分布其他表示法

$$X \sim B(1, p) \quad 或$$

$$X \sim 0-1(p) \quad 或$$

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

*excel*产生0-1分布数据方法：

随机数发生器或函数*randbetween*(0,1).

用途：对于一个随机试验，如果它的样本空间只包含两个元素，即$S=\{e_1,e_2\}$，我们总能在S上定义一个服从（**0—1**）分布的随机变量。

$$X=X(e)=\begin{cases}\mathbf{0}, & \text{当}\ e=e_1,\\ \mathbf{1}, & \text{当}\ e=e_2.\end{cases}$$

来描述这个随机试验的结果。

例2：已知 X的可能取值0与1，并且

$$P(X=0)=9c^2-c,\ P(X=1)=3-8c$$

那么 $c=?$

解：由概率的性质：$(9c^2-c)+(3-8c)=1$

$$\Rightarrow (3c-1)(3c-2)=0 \qquad \Rightarrow c=\begin{cases}1/3\\ 2/3\end{cases}$$

$\because c=2/3$不合要求

$\therefore c=1/3$

二、二项分布（*Binomial*）

n重贝努利试验：设试验E只有两个可能的结果：A与$\overline{A}$，$P(A)=p$，$0<p<1$，将E独立地重复进行n次，则称这一串重复的独立试验为n重贝努利试验。

在相同条件下重复进行

即每次试验结果互不影响

例：独立重复地抛n次硬币，每次只有两个可能的结果：正面，反面，P(出现正面)$=1/2$.

例：将一颗骰子抛n次，设$A=\{$得到1点$\}$，则每次试验只有两个结果：$A, \bar{A}$,
$P(A)=1/6$.

例：从52张牌中有放回地取n次，设$A=\{$取到红牌$\}$，则每次只有两个结果：$A, \bar{A}$, 而且$P(A)=1/2$

现在要计算在以上的n次重复的独立试验中，事件A恰好发生了k次的概率。

设X表示n重贝努利试验中事件A发生次数，则

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

称X服从参数为$P(A)=p$的二项分布，记 $X\sim B(n,p)$

以n=3为例，设A_i=｛第i次A发生｝，i=0, 1, 2, 3

$$P(X=0)=P(\bar A_1\bar A_2\bar A_3)=P(\bar A_1)P(\bar A_2)P(\bar A_3)=(1-p)^3$$

$$P(X=1)=P(A_1\bar A_2\bar A_3\cup\bar A_1A_2\bar A_3\cup\bar A_1\bar A_2A_3)=C_3^1p^1(1-p)^{3-1}$$

$$P(X=2)=P(A_1A_2\bar A_3\cup A_1\bar A_2A_3\cup\bar A_1A_2A_3)=C_3^2p^2(1-p)^{3-2}$$

$$P(X=3)=P(A_1A_2A_3)=p^3$$

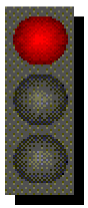
一般 $P(X=k)=C_n^k p^k(1-p)^{n-k},k=0,1,2,\cdots,n$

注：$1=(p+q)^n=\sum_{k=0}^{n}C_n^k p^k q^{n-k}$ **,其中**$q=1-p$

$$excel:P(X=k)=BINOMDIST(k,n,p,FALSE)$$

$$P(X\le k)=BINOMDIST(k,n,p,TRUE)$$

例3：某人骑了自行车从学校到火车站，一路上要经过8个独立的交通灯，且设各灯为红灯的概率为**0.6**，以***X***表示一路上遇到红灯的次数。求***X***的概率分布律.



解：设$A=$“遇到交通灯是红灯”， $P(A)=0.6$, $X\sim B(8,0.6)$

$$P(X=k)=C_8^k 0.6^k 0.4^{8-k},\ k=0,1,2,\cdots 8$$

X	0	1	2	3	4	5	6	7	8
P	0.007	0.0079	0.0413	0.1239	0.2322	0.2787	0.2090	0.0896	0.0168

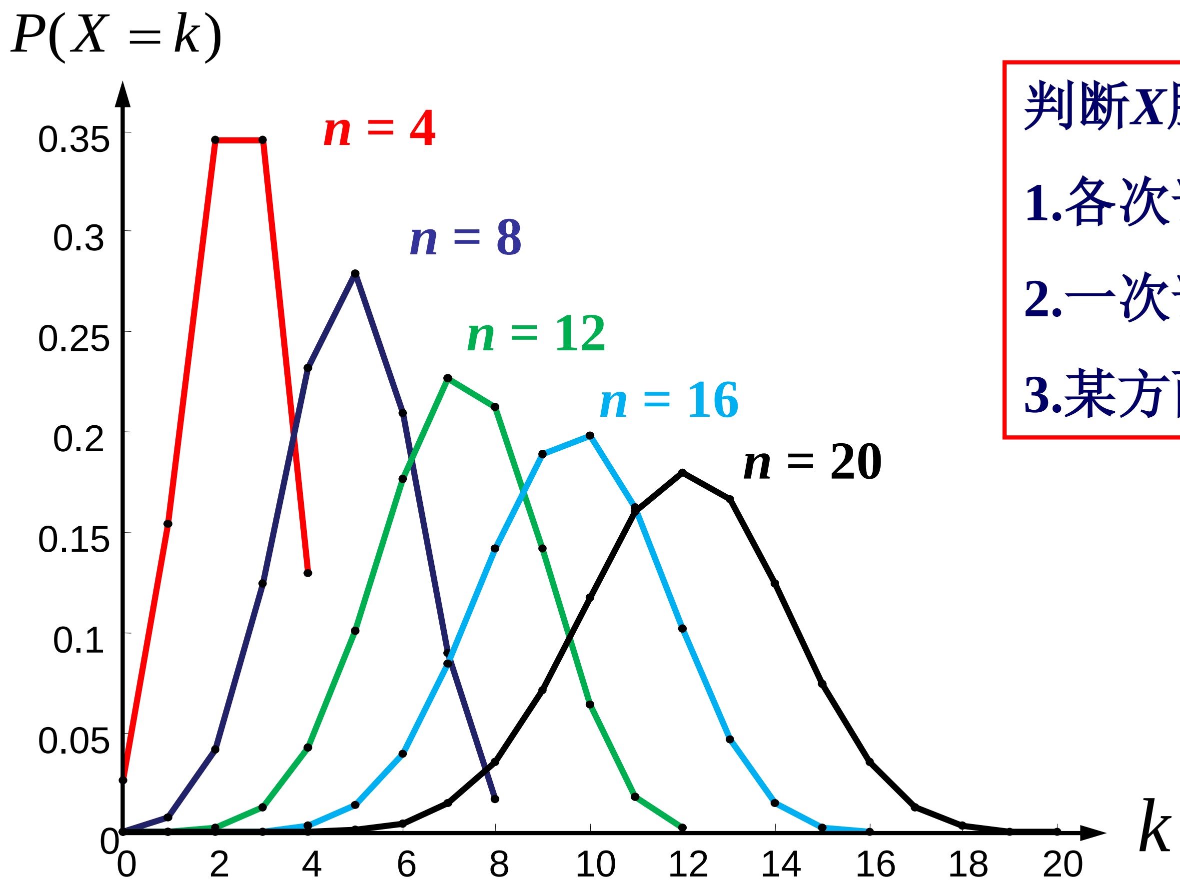
一般，当$(n+1)p$是整数时，X最可能取值是$(n+1)p$及$(n+1)p-1$.

当$(n+1)p$不是整数时，X最可能取值是$(n+1)p$的整数值.

本例中$(n+1)p$=**5.4**不是整数，X最可能取值是**5.4**的整数值**5**.

$$excel: P(X=5)=BINOMDIST(5,8,0.6,FALSE)=0.27869184$$

$B(n,0.6)$的概率分布折线图



判断X服从二项分布的方法：

1.各次试验间独立

2.一次试验有两方面

3.某方面发生次数X

$$P(X=k)=C_n^k 0.6^k 0.4^{n-k},\ k=0,1,2,\cdots,n$$

例4：某人独立射击n次，设每次命中率为p, $0<p<1$，设命中X 次，（1）求X的概率分布律；（2）求至少有一次命中的概率。

解：这是n重贝努利试验 $\Rightarrow X \sim b(n, p)$

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

(2) $P(X \geq 1) = 1 - P(X = 0) = 1 - (1-p)^n$

同时可知： $\lim\limits_{n \to \infty} P(X \geq 1) = 1$

上式的意义为：若p较小但$p \neq 0$, 只要n充分大，至少有一次命中的概率很大。 即“小概率事件”在大量试验中“至少有一次发生”几乎是必然的。

例5：某娱乐场提供玩客一项活动，玩客可以任选一种以下的玩法，如果要你选，你选哪种?

(1). 投1只骰子 4次，若能得一次“6”点就算赢

(2). 投2只骰子24次，若能得一次双“6”点就算赢

---

(1)设*A*=“任投一次得6点”，*X*表示4次投掷中得“6点”的次数

则 $P(A)=1/6$ $\qquad X\sim B(4,1/6)$

$$P(X\geq 1)=1-P(X=0)=1-C_4^0(1/6)^0(1-1/6)^4=0.5177$$

(2)设*A*=“任投一次得双6点”， 则 $P(A)=1/36$

*Y*表示24次投掷中得双“6点”的次数， $Y\sim B(24,1/36)$

$$P(Y\geq 1)=1-P(Y=0)=1-C_{24}^0(1/36)^0(35/36)^{24}=0.4991$$

因此应选（1）

例6：一袋有10个球，其中4红6白，若从中取三次，每次取一只，问恰有一只红球的概率**P(B)**多少？

解：不放回取球时，各次取球不独立

用分步法：$P(B)=\frac{4}{10}\times\frac{6}{9}\times\frac{5}{8}+\frac{6}{10}\times\frac{4}{9}\times\frac{5}{8}+\frac{6}{10}\times\frac{5}{9}\times\frac{4}{8}=\frac{1}{2}$

红 白 白 白 红 白 白 白 红

用超几何分布概率公式：$P(B)=\frac{C_4^1C_6^2}{C_{10}^3}=\frac{1}{2}$

放回取球时，各次取球独立，可用二项分布

设*A*=“任取一球为红球”，**$P(A)$=0.4**

*X*表示所取三只球中红球的只数

则 $X\sim B(3,0.4)$

$P(B)=P(X=1)=C_3^1 0.4\times0.6^2=0.432$

- 例7：设有80台同类型设备，各台工作是相互独立的，发生故障的概率都是0.01，且一台设备的故障只能由一个人来处理。

现考虑两种配备维修工人的方案，

其一是由3个人共同维护80台；

其二是由4个人维护，每人负责20台。

试比较这两种方案在设备发生故障时不能及时维修的概率的大小。

解：第一种方案(由3个人共同维护80台)

以X记80台中同一时刻发生故障的台数，

则 $X \sim B(80, 0.01)$

故80台中发生故障而不能及时维修的概率为：

$$P\{X \geq 4\} = 1 - \sum_{k=0}^{3} C_{80}^{k}(0.01)^{k}(0.99)^{80-k} = 0.0087$$

Excel: 1-BINOM.DIST(3,80,0.01,1)

第二种方案(由4个人维护，每人负责20台)

以A_i记“第i个人所维护的20台中发生了故障，不能及时维修”

则80台机器中发生故障不能及时维修的概率为:

$$P(A_1 \cup A_2 \cup A_3 \cup A_4) = 1 - P(\bar{A}_1 \bar{A}_2 \bar{A}_3 \bar{A}_4) = 1 - [P(\bar{A}_1)]^4,$$

以Y记“第一个人所维护的20台机器中同时发生故障的台数”

则$Y \sim B(20, 0.01)$,故有:

$$P(A_1) = P\{Y \geq 2\} = 1 - P\{Y = 0\} - P\{Y = 1\}$$

$$= 1 - C_{20}^{0} * 0.01^{0} * 0.99^{20} - C_{20}^{1} * 0.01^{1} * 0.99^{19} = 0.0169$$

即有：$P(A_1 \cup A_2 \cup A_3 \cup A_4) = 1 - (1 - 0.0169)^4 = 0.0659$

方案结果比较

方案一：由3个人共同维护80台

方案二：由4个人维护，每人负责20台

第一种方案下不能及时维修的概率（0.0087）小于于第二种方案的概率（0.0659）.

甚至第二种方案中$P\left(A_1\right)=0.0169$都大于第一种方案的概率（0.0087）.

例8：　某单位在进货时，一般从一大批产品中任取10件，若其中次品数不多于一件，则接受该批产品．现有一大批产品其次品率为***0.05*** ，则在以上验收方案下产品能被接受的概率***P*(*B*)**是多少？

分析：题中没有说明总的产品数、抽检时是否放回。

许多情况下的抽检时是不放回的，如果产品总数是100，若以次品率为***0.05***计算，那么次品数为5，正品为95.

$$P(B) \approx \frac{C_5^0 C_{95}^{10}}{C_{100}^{10}} + \frac{C_5^1 C_{95}^{9}}{C_{100}^{10}} = 92.31\%$$

Excel: =HYPGEOM.DIST(1,5,10,100,1)

但是，现在产品总数并不知道！

例8： 某单位在进货时，一般从一大批产品中任取10件，若其中次品数不多于一件，则接受该批产品．现有一大批产品其次品率为***0.05*** ，则在以上验收方案下产品能被接受的概率***P*(*B*)**是多少？

分析：注意到本题的产品数很多，所以可作近似处理：

1、可以认为抽检时，产品是一件一件抽取的

2、每取出一件检验后，又放回

结论：经以上近似处理后,可以把抽取10件产品检验看作是做了10次独立试验。为此，设 $A=$"任取一件产品为次品"

X表示10件品中的次品数， 则 $X\sim B(10,0.05)$

$$P(B)=P(X=0)+P(X=1)=C_{10}^{0}0.05^{0}0.95^{10}+C_{10}^{1}0.05^{1}0.95^{9}\quad=91.39\%$$

	A	B	C
1	100	0.92314328	=HYPGEOM.DIST(1,A1*0.05,10,A1,1)
2	1000	0.91469243	=HYPGEOM.DIST(1,A2*0.05,10,A2,1)
3	10000	0.91394384	=HYPGEOM.DIST(1,A3*0.05,10,A3,1)
4	100000	0.91386985	=HYPGEOM.DIST(1,A4*0.05,10,A4,1)
5	二项分布	0.91386164	=BINOM.DIST(1,10,0.05,1)

设有m只产品，其中a只正品，b只次品$(m=a+b)$，则不放回取n只产品中恰有k只元件正品的概率为：

$$P(X=k)=\frac{C_a^k C_{m-a}^{n-k}}{C_m^n},$$

此式即为**超几何分布**的概率公式。

当$m\to+\infty$时，记 $p=\dfrac{a}{m}, q=\dfrac{b}{m}$，则可证：

X近似服从二项分布$B(n,p)$，即

$$P(X=k)=\frac{C_a^k C_{m-a}^{n-k}}{C_m^n}\approx C_n^k p^k q^{n-k}$$

二项分布中，设$\lambda = np$,当$n \to \infty$，$p \to 0$时，

$$P(X = k) = C_n^k p^k (1 - p)^{n-k}$$

$$= \frac{n(n-1)\cdots(n-k+1)}{k!} p^k (1-p)^{n-k}$$

$$= \frac{1 \times (1 - \frac{1}{n}) \cdots (1 - \frac{k-1}{n})}{k!} (np)^k (1-p)^{\frac{1}{-p}[-p(n-k)]}$$

$$\xrightarrow[p \to 0]{n \to \infty} \frac{\lambda^k e^{-\lambda}}{k!}$$

三、泊松分布(*Poisson*)

若随机变量 X 的概率分布律为

$$P(X=k)=\frac{e^{-\lambda}\lambda^k}{k!},\quad k=0,1,2,\cdots,\quad \lambda>0$$

称 X 服从参数为 λ 的泊松分布，

记为 $X\sim P(\lambda)$，或记为 $X\sim\pi(\lambda)$.

$$excel: P(X=k)=\mathrm{POISSON}(k,\lambda,false)$$
$$excel: P(X\le k)=\mathrm{POISSON}(k,\lambda,true)$$

产生泊松分布背景

若在不相重叠的“时间”区间内需要“服务”的“顾客”数相互独立，则单位“时间”内需要服务的“顾客”数往往可视为服从泊松分布的。

这里的“时间”、“服务”、“顾客”都是广义的概念。

泊松分布的应用

- 教科书每页的错别字的个数(稀有事件)
- 某路段单位时间内发生的交通事故数
- 车站某时段等车人数
- 每天进商场购物人数
- 电话交换台在一个时间间隔内收到的呼叫次数
- 单位时间内商品销量
- 单位时间内网站访问人数
- 单位时间内棉纱断头数
- 单位时间间隔内某放射物发出的、经过计数器的粒子数

例1：设某汽车停靠站候车人数$X\sim P(\lambda)$，$\lambda=4.5$。

(1)求至少有两人候车的概率；

(2)已知至少有两人候车，求恰有两人候车的概率。

解：$P(X=k)=\dfrac{e^{-4.5}4.5^k}{k!}$，$k=0,1,2,\cdots$

(1) $P(X\geq 2)=1-P(X=0)-P(X=1)=1-e^{-4.5}(1+4.5)=0.9389$

(2) $P(X=2\mid X\geq 2)=\dfrac{P(X=2)}{P(X\geq 2)}=\dfrac{e^{-4.5}*4.5^2/2!}{0.9389}=0.1198$

$excel: P(X=2)=\mathrm{POISSON}(2,\ 4.5,\ \text{false})=0.11247859$

$P(X\geq 2)=1-P(X\leq 1)=1-\mathrm{POISSON}(1,4.5,\text{true})$

$=0.938900519$

例2：设$P(X=1)=P(X=2)=0.5$，当$X=x$时，随机变量$Y\sim P(x)$，试求 (1)$P(Y\geq 1)$；(2)$P(X=1|Y\geq 1)$。

解：注意此处Y并非服从泊松分布，而是有条件服从泊松分布。

可写为：$Y|X=x\sim P(x)$，把“$Y|X=x$”作为随机变量Z_x。

$$P(Z_x=k)=\frac{x^k}{k!}e^{-x},\quad k=0,1,2,\cdots,\quad x=1,2$$

(1) Y的取值受到X取值的影响，由全概率公式

$$P(Y\geq 1)=P(X=1)P(Y\geq 1|X=1)+P(X=2)P(Y\geq 1|X=2)$$

$$=P(X=1)P(Z_1\geq 1)+P(X=2)P(Z_2\geq 1)$$

$$=P(X=1)(1-P(Z_1=0))+P(X=2)(1-P(Z_2=0))$$

$$=0.5(1-e^{-1})+0.5(1-e^{-2})$$

例2：设$P(X=1)=P(X=2)=0.5$，当$X=x$时，随机变量$Y\sim P(x)$，试求 (1)$P(Y\geq 1)$；(2)$P(X=1|Y\geq 1)$。

---

(2)由贝叶斯公式：

$$P(X=1|Y\geq 1)=\frac{P(X=1)P(Y\geq 1|X=1)}{P(Y\geq 1)}$$

$$=\frac{P(X=1)P(Z_1\geq 1)}{P(Y\geq 1)}$$

$$=\frac{0.5(1-e^{-1})}{0.5(1-e^{-1})+0.5(1-e^{-2})}$$

$$=\frac{1-e^{-1}}{2-e^{-1}-e^{-2}}$$

*例4：某商店每天的顾客数是随机变量$X\sim P(\lambda)$，设每个进店购物的概率为p，顾客之间是否购物相互独立，求该店每天购物的顾客数（Y）的概率分布律。

解：引起A=“Y=k”发生的原因有X=0,X=1,X=2，……等

由全概率公式：$$P(Y=k)=\sum_{n=k}^{+\infty}P(X=n)P(Y=k|X=n)$$

其中：$P(X=n)=\dfrac{\lambda^n e^{-\lambda}}{n!}$，　$n=0,1,2,\cdots$　　$Y|X=n\sim B(n,p)$

$$P(Y=k|X=n)=C_n^k p^k(1-p)^{n-k}=C_n^k p^k q^{n-k},k=0,1,2,\cdots n$$

$$\therefore P(Y=k)=\sum_{n=k}^{\infty}\frac{\lambda^n e^{-\lambda}}{n!}\times C_n^k p^k q^{n-k}=\sum_{n=k}^{\infty}\frac{\lambda^n e^{-\lambda}}{n!}\times\frac{n!}{k!(n-k)!}p^k q^{n-k}$$

$$=\frac{(\lambda p)^k e^{-\lambda}}{k!}\sum_{n=k}^{\infty}\frac{(\lambda q)^{n-k}}{(n-k)!}=\frac{(\lambda p)^k e^{-\lambda}}{k!}e^{\lambda q}=\frac{(\lambda p)^k e^{-\lambda p}}{k!},\ k=0,1,2,\cdots$$

$$\therefore\ Y\sim P(\lambda p)$$

『泊松定理』

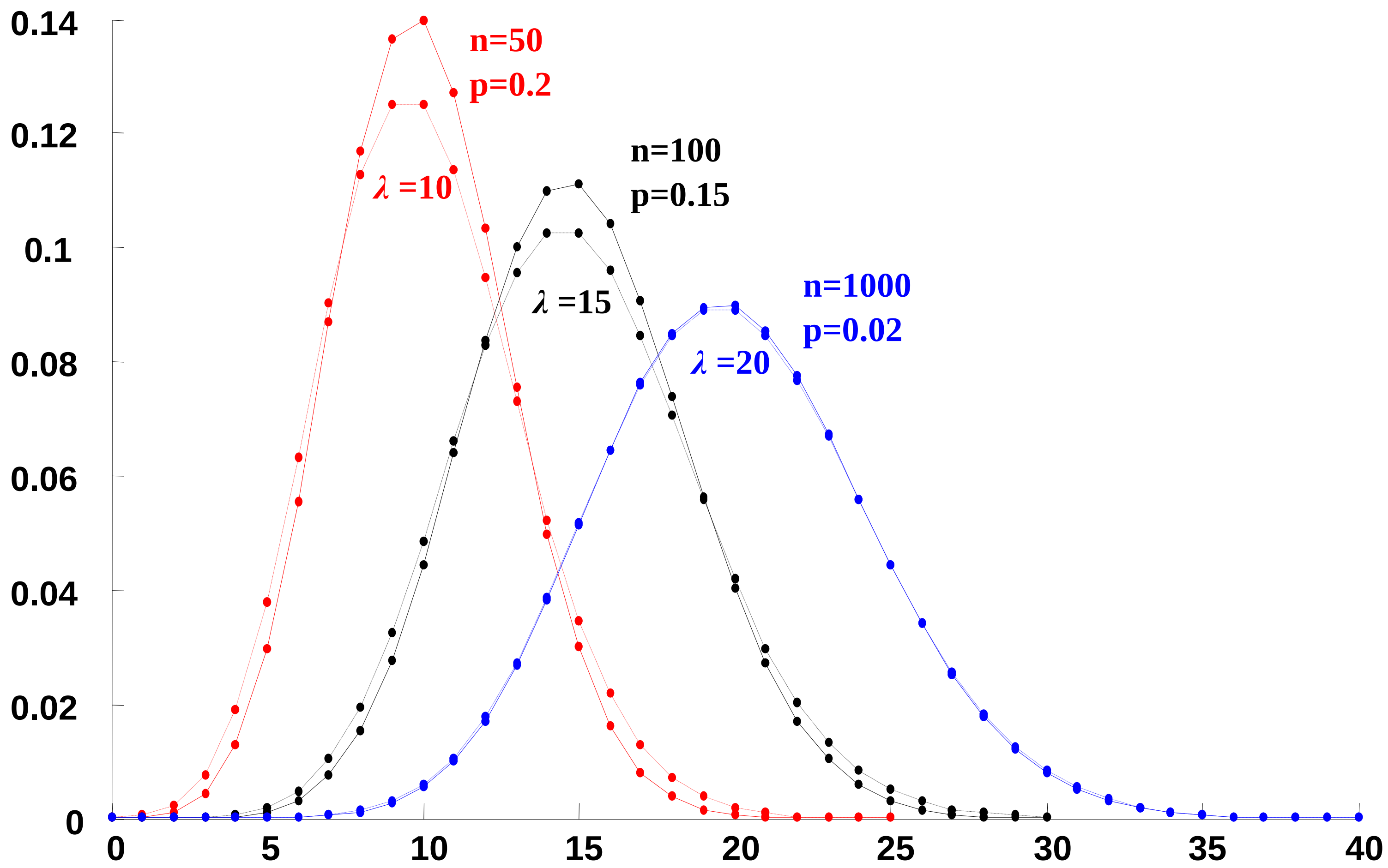
当n充分大, p足够小时，记$\lambda = np$, 有

$$C_n^k p^k (1-p)^{n-k} \approx \frac{\lambda^k e^{-\lambda}}{k!}$$

即认为可用泊松分布近似二项分布.

实际上当$n > 10$, $p < 0.1$时, 就可用泊松分布代替二项分布作近似计算。

二项分布与泊松分布分布律对比



注：

$$f(x)=\sum_{k=0}^{n}\frac{f^{(k)}(x_0)}{k!}(x-x_0)^k+\frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)^{n+1}$$

其中ξ在x和x_0之间

$$e^x=\frac{x^0}{0!}+\frac{x^1}{1!}+\frac{x^2}{2!}+\cdots \qquad (x_0=0)$$

$$e^\lambda=\sum_{k=0}^{+\infty}\frac{\lambda^k}{k!} \qquad (x=\lambda)$$

$$\sum_{k=0}^{+\infty}\frac{\lambda^k e^{-\lambda}}{k!}=1$$

四、几何分布（Geometric）

设独立重复试验中，每次试验有两个结果$A, \bar{A}$.

$P(A)=p$,随机变量X表示直到事件A发生为止所做的试验次数，则

$$P(X=k)=(1-p)^{k-1}p \quad , \quad k=1,2,3,\cdots$$

则称随机变量X服从参数为 p 的几何分布。

记为： $X \sim G(p)$

几何分布反映了直到事件A发生为止，所做的试验次数恰为 k 次的概率问题，也即反映了直到第 k 次才发生事件A的概率问题。

例1：某人进行独立射击，每次命中率为p，射击直到命中目标为止时的射击次数为X。求直到第3次才命中的概率及X的分布律。

解：设$B_i=\{$第i次命中目标$\}$，$i=1,2,3,\ldots\ldots$

则$B_1, B_2, \ldots$相互独立, $P(B_i)=p$

$$P(X=3)=P(\bar{B}_1\bar{B}_2B_3)\overset{\text{独立}}{=}P(\bar{B}_1)P(\bar{B}_2)P(B_3)=(1-p)^2p$$

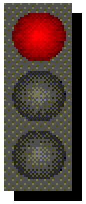
一般，第k次才命中的概率为：

$$P(X=k)=P(\bar{B}_1\bar{B}_2\cdots\bar{B}_{k-1}B_k)$$

$$\overset{\text{独立}}{=}P(\bar{B}_1)P(\bar{B}_2)\cdots P(\bar{B}_{k-1})P(B_k)$$

$$=(1-p)^{k-1}p\ ,\qquad k=1,2,3,\cdots$$

*例2：某人骑车从学校到火车站，一路上要经过3个独立的交通灯，且设各灯为红灯的概率为p，$0<p<1$，以X表示首次停车时所通过的交通灯数，求X的概率分布律。



解：设A_i={第i个灯为红灯}，则$P(A_i)= p$，i=1, 2, 3

且A_1, A_2, A_3相互独立。

$$P(X=0)=P(A_1)=p\ ;\ P(X=1)=P(\bar{A}_1A_2)=(1-p)p\ ;$$

$$P(X=2)=P(\bar{A}_1\bar{A}_2A_3)=(1-p)^2p\ ;$$

$$P(X=3)=P(\bar{A}_1\bar{A}_2\bar{A}_3)=(1-p)^3\ ;$$

注意：$(X=0),(X=1),(X=2)$
$(X=3)$为S的一个划分

X	0	1	2	3
p	p	$(1-p)p$	$(1-p)^2p$	$(1-p)^3$

*五、帕斯卡(负二项)分布（Pascal）

设独立重复试验中，每次试验有两个结果$A, \bar{A}$.
$P(A)=p$,随机变量X表示直到事件A发生了r次为止所作的试验次数, 则

$$P(X=k)=C_{k-1}^{r-1}p^r(1-p)^{k-r}\ ,\quad k=r,r+1,r+2,\ldots,$$

其中r为正整数,$0<p<1$.

称X服从参数为(r,p)的帕斯卡分布，$X\sim NB(r,p)$

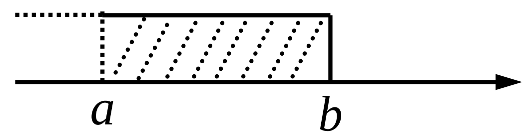
帕斯卡分布反映了独立试验中直到第 k 次发生了r次事件A的概率问题。

$$P(X=k)=C_{k-1}^{r-1}p^{r-1}(1-p)^{k-r}\cdot p$$

§ 3 随机变量的分布函数

对于非离散型变量，由于其不可列，故考虑区间上的概率问题. 注意到:

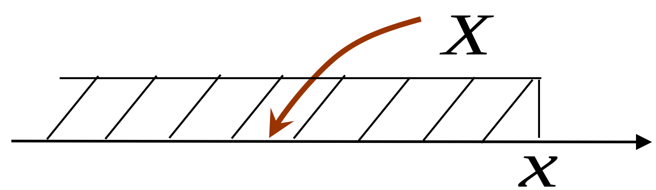
$$\{a < X \le b\} = \{X \le b\} - \{X \le a\}$$



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。

$F(x)$的几何意义：



$F(x)$的性质：

1) $F(x)$是单调不减函数

2) $0 \le F(x) \le 1$，且$F(-\infty) = 0$，$F(+\infty) = 1$

3) $F(x)$右连续,即$F(x+0) = F(x)$

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

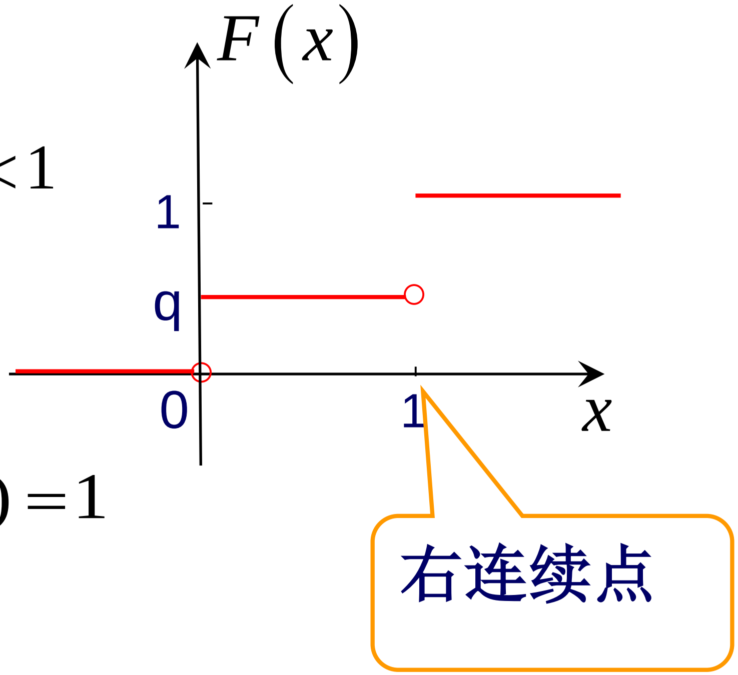
例1：

X	0	1
p	q	p

求X的概率分布函数$F(x)$及$P(X\geq 1)$的值。

解：

$$F(x)=P\{X\leq x\}=\begin{cases}0 & ,x<0\\ q & ,0\leq x<1\\ 1 & ,x\geq 1\end{cases}$$



$$P(X\geq 1)=P(\{X=1\}\cup\{X>1\})=p$$

比较 $P(X\geq 1)=p$ 与 当$x\geq 1$时$F(x)=1$

注意： $P(0<X<1)=0$

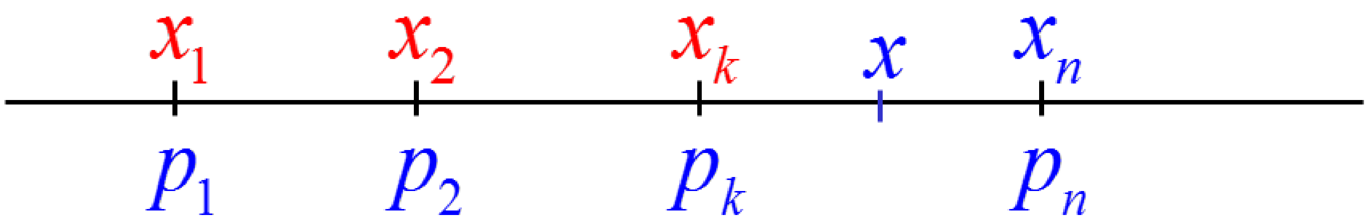
$$P(0<X\leq 1)=P(\{0<X<1\}\cup\{X=1\})\overset{\text{不相容}}{=}P(0<X<1)+P(X=1)=0+p=p$$

或 $P(0<X\leq 1)=F(1)-F(0)=1-q=p$

离散型随机变量的分布函数

一般，离散型随机变量X的分布律为：

$$P\{X = x_k\} = p_k, \qquad k = 1,2,\cdots$$



由概率的可列可加性得X的分布函数为：

$$F(x) = P(X \le x) = \sum_{x_k \le x} P\{X = x_k\} = \sum_{x_k \le x} p_k$$

分布函数$F(x)$在$X = x_k\ (k = 1,2,\cdots)$处

有跳跃，其跳跃值为 $p_k = P\{X = x_k\}$

例2：设$X \sim F(x) = \begin{cases} 0, & x < -1 \\ 0.2, & -1 \le x < 2 \\ 0.7, & 2 \le x < 4 \\ 1, & x \ge 4 \end{cases}$,

求 $P(X \le 3), P(0.5 < X \le 3), P(X \ge 2)$, X的分布律

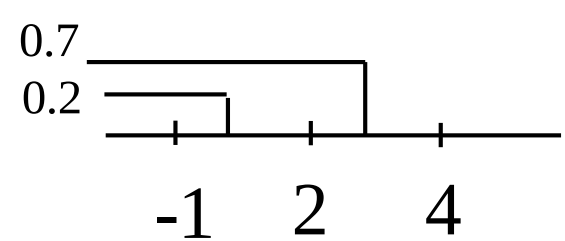
解：$P(X \le 3) = F(3) \ = 0.7$

$$P(0.5 < X \le 3) = F(3) - F(0.5) = 0.7 - 0.2 = 0.5$$

$$P(X \ge 2) = 1 - P(X < 2) = 1 - P(X \le 2) + P(X = 2)$$

$$= 1 - F(2) + P(X = 2)$$

$$= 1 - 0.7 + 0.5 = 0.8$$



例2：设$X \sim F(x)=\begin{cases}0, & x<-1\\ 0.2, & -1\le x<2\\ 0.7, & 2\le x<4\\ 1, & x\ge 4\end{cases}$,

求 $P(X\le 3), P(0.5<X\le 3), P(X\ge 2), X$的分布律

其实，可以证明$P(X=x_0)=F(x_0+0)-F(x_0-0)$

注意到$F(x)$中取值为常数及区间中-1、2、4上有等号

$$\therefore P(X=-1)=F(-1+0)-F(-1-0)=0.2-0=0.2$$

$$P(X=2)=0.7-0.2=0.5$$

$$P(X=4)=1-0.7=0.3$$

整理得X的分布律为：

X	-1	2	4
P_k	0.2	0.5	0.3

例3：设$X \sim F(x)=\begin{cases}\dfrac{Ax}{1+x} & ,x>0 \\ 0 & ,x\le 0\end{cases}$，求 A 和 $P(1<X\le 2)$

解：由分布函数的性质：

$$1=\lim_{x\to+\infty}F(x)=\lim_{x\to+\infty}\frac{Ax}{1+x}=A$$

$$P(1<X\le 2)=F(2)-F(1)$$

$$=\frac{2}{1+2}-\frac{1}{1+1}=\frac{1}{6}$$

例4：设$X \sim F(x)=\begin{cases} A+Be^{-x}, & x>0 \\ 0, & x\le 0 \end{cases}$，求 A, B=?

解：由分布函数的性质：

$$1=\lim_{x\to+\infty}F(x)=A+B\times 0=A$$

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

$$=A+Be^{0}$$

$$\Rightarrow B=-1$$

§4 连续型随机变量及其概率密度

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：

$$\boxed{F(x)=\int_{-\infty}^{x} f(t)dt} \quad \text{或} \quad \boxed{P(X \in D)=\int_{D} f(t)dt}$$

则称X为连续型随机变量.

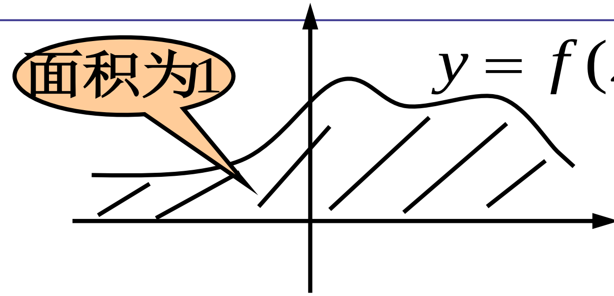
其中$f(x)$称为X的概率密度函数，

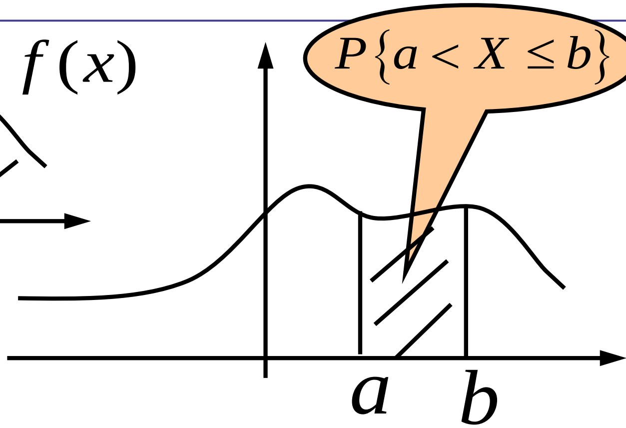
简称**概率密度**，也称X的密度函数。

概率密度函数 $f(x)$ 的性质

1) $f(x) \geq 0$

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





3) 对于任意的实数 $a,b \quad (b > a)$

$$P\{a < X \leq b\} = \int_a^b f(t)\,dt \qquad \Rightarrow P(X = a) = 0$$

4) 在$f(x)$的连续点 x，$F'(x) = f(x)$

即在$f(x)$的连续点x, 有：

$$f(x) = F'(x) = \lim_{\Delta x \to 0} \frac{F(x+\Delta x) - F(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{P(x < X \leq x + \Delta x)}{\Delta x}$$

$\therefore P(x < X \leq x + \Delta x) \approx f(x) \cdot \Delta x$ ，这表示X落在点 x 附近 $(x, x+\Delta x]$ 的概率近似等于$f(x)\Delta x$.

例1：设随机变量X的密度函数为

$$f(x)=\begin{cases}2x, & 0\le x<0.5\\ 6-6x, & 0.5\le x\le 1\\ 0, & \text{其他}\end{cases}$$，求分布函数$F(x)$。

解：参考$f(x)$的分段情况

$$F(x)=P\{X\le x\}=\int_{-\infty}^{x}f(t)dt=$$

$$\begin{cases}\int_{-\infty}^{x}0dt=0 & ,x<0\\ \int_{-\infty}^{0}0dt+\int_{0}^{x}2tdt=x^2 & ,0\le x<0.5\\ \int_{-\infty}^{0}0dt+\int_{0}^{0.5}2tdt+\int_{0.5}^{x}(6-6t)dt=6x-3x^2-2 & ,0.5\le x\le 1\\ \int_{-\infty}^{0}0dt+\int_{0}^{0.5}2tdt+\int_{0.5}^{1}(6-6t)dt+\int_{1}^{x}0dt=1 & ,\quad x>1\end{cases}$$

例2：设X的概率密度为

$$f(x)=\begin{cases} c & ,0<x<1 \\ 2/9 & ,3<x<6 \\ 0 & ,\text{其他} \end{cases}$$

(1) 求常数c的值；

(2) 写出X的概率分布函数；

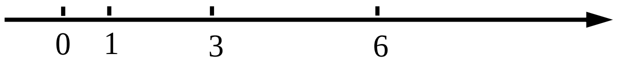
(3) 要使$P(X<k)=\dfrac{2}{3}$，求k的值。

解:(1) $1=\int_{-\infty}^{+\infty} f(t)dt$

$$=\int_{-\infty}^{0} 0dt+\int_{0}^{1} cdt+\int_{1}^{3} 0dt+\int_{3}^{6} \frac{2}{9}dt+\int_{6}^{\infty} 0dt$$

$$=c+\frac{2}{3} \qquad \Rightarrow c=\frac{1}{3}$$

(2) 参考 $f(x)$的分段情况，求分布函数 $F(x)=P\{X\le x\}=\int_{-\infty}^{x} f(t)dt$



$$f(x)=\begin{cases}1/3 & 0<x<1\\ 2/9 & 3<x<6\\ 0 & 其他\end{cases}$$

$$F(x)=\begin{cases}0 & ,x\le 0\\ \int_{-\infty}^{0}0dt+\int_{0}^{x}\frac{1}{3}dt & ,0<x<1\\ \int_{-\infty}^{0}0dt+\int_{0}^{1}\frac{1}{3}dt+\int_{1}^{x}0dt & ,1\le x\le 3\\ \int_{-\infty}^{0}0dt+\int_{0}^{1}\frac{1}{3}dt+\int_{1}^{3}0dt+\int_{3}^{x}\frac{2}{9}dt & ,3<x<6\\ 1 & ,x\ge 6\end{cases}$$

$$=\begin{cases}0 & ,x\le 0\\ x/3 & ,0<x<1\\ 1/3 & ,1\le x\le 3\\ (2x-3)/9 & ,3<x<6\\ 1 & ,x\ge 6\end{cases}$$

$(3)\ P(X<k)=\dfrac{2}{3} \quad =F(k)$

$$\Rightarrow \frac{2k-3}{9}=\frac{2}{3} \Rightarrow k=4.5$$

例3：设$X \sim F(x)=\begin{cases} A+Be^{-x}, x>0 \\ 0 \qquad\quad, x\le 0 \end{cases}$，求$A, B$=?

解：由分布函数的性质：$1=\lim\limits_{x\to+\infty}F(x)=A+B\times 0=A$

$$0=F(0)=\lim_{x\to 0+}F(x)=A+Be^{0}\Rightarrow B=-1$$

或 $f(x)=F^{'}(x)=\begin{cases} -Be^{-x}, x>0 \\ 0 \qquad\quad, x\le 0 \end{cases}$

$$1=\int_{-\infty}^{+\infty}f(x)dx=\int_{0}^{+\infty}-Be^{-x}dx=B\int_{0}^{+\infty}de^{-x}$$

$$=Be^{-x}\Big|_{0}^{+\infty}=-B$$

例4：设$F_1(x), F_2(x)$为两个分布函数，其相应的密度函数$f_1(x), f_2(x)$是连续函数，则必为概率密度的是

(A) $f_1(x)f_2(x)$ (B) $2f_2(x)F_1(x)$

(C) $f_1(x)F_2(x)$ (D) $f_1(x)F_2(x)+f_2(x)F_1(x)$

解：任何密度函数均应有 $\int_{-\infty}^{+\infty} f(x)dx = 1$

$$\int_{-\infty}^{+\infty}\left[f_1(x)F_2(x)+f_2(x)F_1(x)\right]dx$$

$$=\int_{-\infty}^{+\infty} d(F_1(x)F_2(x)) = F_1(x)F_2(x)\Big|_{-\infty}^{+\infty} = 1$$ 选D

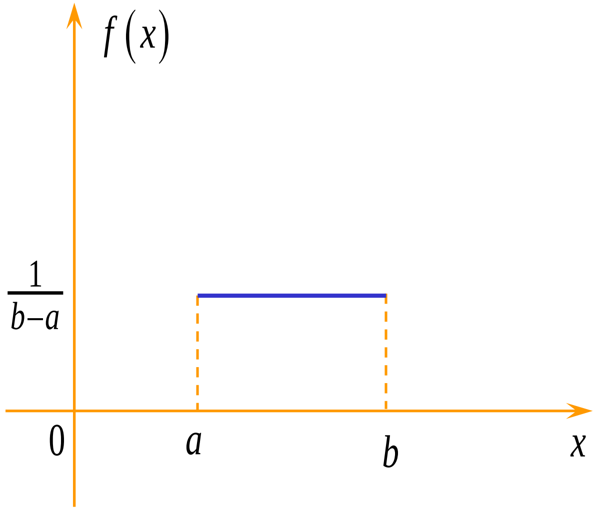
$$F_1(x)F_2(x) = \int_{-\infty}^{x}\left[f_1(x)F_2(x)+f_2(x)F_1(x)\right]dx$$

几个重要的连续随机变量

一、均匀分布（Uniform）

定义：X具有概率密度

$$f(x)=\begin{cases}\dfrac{1}{b-a} & x\in(a,b)\\ 0 & 其他\end{cases}$$



称X在(a,b)上服从均匀分布，

记为$X\sim U(a,b)$

而在 $[a,b]$上服从均匀分布，记为$X\sim U[a,b]$

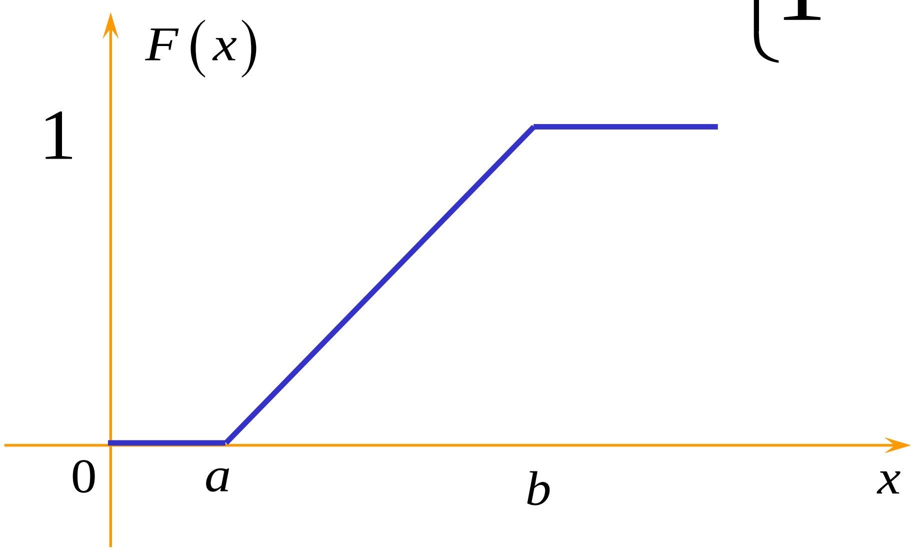
*excel*产生均匀分布数据方法:随机数发生器.

产生均匀分布的动态数据用函数RAND().

利用RAND()也可产生动态的正态分布、二项分布的随机数据.

均匀分布的分布函数

$$F(x)=\int_{-\infty}^{x} f(t)dt=\begin{cases} 0 & x\le a \\ \dfrac{x-a}{b-a} & a<x<b \\ 1 & x\ge b \end{cases}$$



均匀分布的性质

$$f(x)=\begin{cases}\dfrac{1}{b-a} & x\in(a,b)\\ 0 & \text{其他}\end{cases}$$

设 $a<s<s+L<b$

$$\Rightarrow P(s<X<s+L)=\int_s^{s+L}\frac{1}{b-a}dt=\frac{L}{b-a}$$

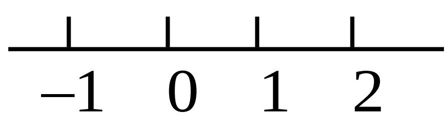
与 **s** 无关

例1：在区间(-1, 2)上随机取一实数*X*，

(1)试写出*X*的概率密度；　(2)求 $P(X>0)$ 的值；

(3)若在该区间上随机取10个数，求10个数中恰有两个数大于0的概率。

解：(1) *X*在区间(-1, 2)上均匀分布 $\Rightarrow f(x)=\begin{cases}\dfrac{1}{3}, & -1<x<2\\ 0, & \text{其他}\end{cases}$

(2) 

$$P(X>0)=\frac{2-0}{2-(-1)}=\frac{2}{3}$$

或 $$P(X>0)=\int_{0}^{+\infty}f(x)dx=\int_{0}^{2}\frac{1}{3}dx=\frac{2}{3}$$

(3) 设*A*=“*X*>0”，*Y*=“10个数中大于0的个数”，则：

$$Y\sim B(10,\frac{2}{3})\qquad \Rightarrow P(Y=2)=C_{10}^{2}\left(\frac{2}{3}\right)^{2}\left(\frac{1}{3}\right)^{8}$$

例2：设X的分布律为：

X	1	2	3
P_k	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$

$Y \sim U(0, X)$，求 $P(Y<0.5)$，$P(X=1|Y<0.5)$

解：Y的取值与X有关，把"$Y<0.5$"作为一个事件，该事件与X=1，2，3事件有关。由全概率公式：

$$P(Y<0.5)=\sum_{i=1}^{3}P(X=i)P(Y<0.5|X=i)$$

$$=\sum_{i=1}^{3}P(X=i)P(Z_i<0.5)$$

$$=\frac{1}{6}\times\frac{1}{2}+\frac{1}{3}\times\frac{1}{4}+\frac{1}{2}\times\frac{1}{6}=\frac{1}{4}$$

由贝叶斯公式：

$$P(X=1|Y<0.5)=\frac{P(X=1,Y<0.5)}{P(Y<0.5)}=\frac{P(X=1)P(Y<0.5|X=1)}{P(Y<0.5)}=\frac{1}{3}$$

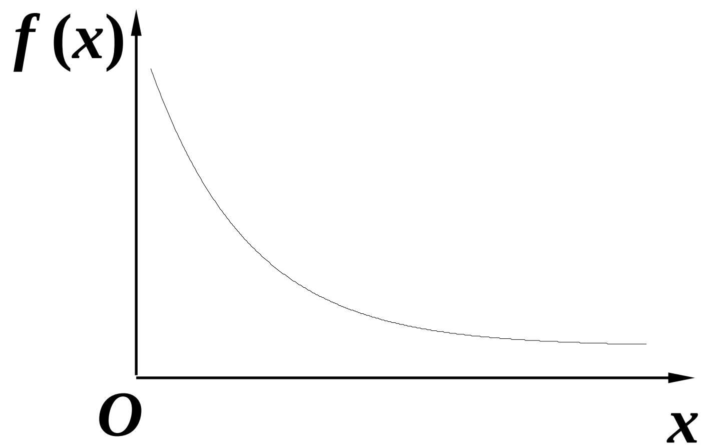
把$Y \sim U(0,X)$理解成：$Y|X=x \sim U(0,x)$ 或 $Z_x \sim U(0,x)$

二、指数分布（Exponential）

定义：设X的概率密度为 $f(x)=\begin{cases}\lambda e^{-\lambda x} & x>0\\ 0 & x\le 0\end{cases}$，

其中$\lambda>0$为常数，则称X服从参数为λ的指数分布。

分布可记为：$X\sim E(\lambda)$



指数分布密度函数图

指数分布密度函数、分布函数、应用

注意:有的资料用θ表示参数!
$$f(x)=\begin{cases}\frac{1}{\theta}e^{-\frac{x}{\theta}} & x>0\\ 0 & x\le 0\end{cases}$$

应用：动植物的寿命、电器的寿命分布等。

$$F(x)=\int_{-\infty}^{x}f(t)dt=\begin{cases}1-e^{-\lambda x} & x>0\\ 0 & x\le 0\end{cases}$$

$excel$：$F(x)=EXPONDIST(x,\lambda,true)$

指数分布X的具有无记忆性

设 $t_0 > 0, t > 0,$

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

$$P(X > t_0 + t \mid X > t_0) = \frac{P(X > t_0 + t)}{P(X > t_0)}$$

$$= \frac{1 - F(t_0 + t)}{1 - F(t_0)} = \frac{1 - (1 - e^{-\lambda(t_0 + t)})}{1 - (1 - e^{-\lambda t_0})} = e^{-\lambda t}$$

$$= P(X > t)$$

无记忆性可写为:

$$P(X > t_0 + t) = P(X > t_0)P(X > t)$$

也可写为：$P(X > s \mid X > t_0) = P(X > s - t_0)$

例3：设某种电子元件的寿命X(以小时计)的密度函数为：$f(x)=\begin{cases}0.1e^{-0.1x}, & x>0\\ 0, & x\le 0\end{cases}$

（1）求任取一只元件其寿命大于10小时的概率

（2）求任取三只元件中正好有一只寿命大于10小时的概率

（3）已知一只元件使用了15小时未坏，求该元件还能用10小时的概率

解：(1) $P(X>10)=\int_{10}^{+\infty}0.1e^{-0.1x}dx=-\int_{10}^{+\infty}de^{-0.1x}=(-e^{-0.1x})_{10}^{+\infty}=e^{-1}$

(2) 设A=“任取一只元件其寿命大于10小时”

$$P(A)=P(X>10)=e^{-1}\triangleq p$$

Y表示任取三只元件中寿命大于10小时的元件只数

$$Y\sim B(3,e^{-1}),\qquad P(Y=1)=C_3^1e^{-1}(1-e^{-1})^2$$

(3) $P(X>15+10\mid X>15)=P(X>10)=e^{-1}$

例4：已知$X \sim f(x)=\begin{cases} \frac{1}{\theta}e^{-\frac{x}{\theta}}, x>0 \\ 0, \quad\quad x\le 0 \end{cases}, \theta>0$未知，

$P(5<X<10)=\frac{1}{4}$，求$P(X>20)$

解：$\frac{1}{4}=P(5<X<10)=\int_5^{10}\frac{1}{\theta}e^{-\frac{x}{\theta}}dx=e^{-\frac{5}{\theta}}-e^{-\frac{10}{\theta}}$

即：$\frac{1}{4}=e^{-\frac{5}{\theta}}-(e^{-\frac{5}{\theta}})^2 \quad \Rightarrow e^{-\frac{5}{\theta}}=\frac{1}{2}$

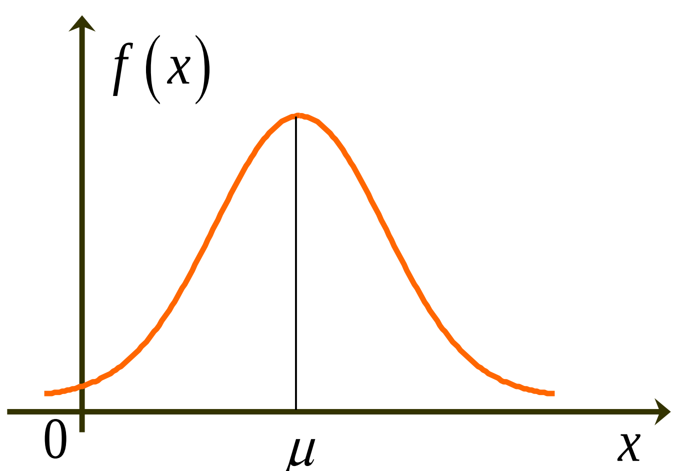
$$P(X>20)=\int_{20}^{+\infty}\frac{1}{\theta}e^{-\frac{x}{\theta}}dx=e^{-\frac{20}{\theta}}=(e^{-\frac{5}{\theta}})^4=\frac{1}{16}$$

三、正态分布（*Normal*）

定义：设X的概率密度为：

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

其中μ,σ^2为常数，称X服从参数为μ,σ^2的正态分布(Gauss分布)，记为$X\sim N(\mu,\sigma^2)$



验证：$\int_{-\infty}^{+\infty} f(x)dx = 1$

$$\because \int_{-\infty}^{+\infty} f(x)dx \overset{令 t=\frac{x-\mu}{\sigma}}{=\!=\!=} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{t^2}{2}} \sigma dt = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

记 $I = \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt,\quad \begin{cases} x = r\cos\theta \\ y = r\sin\theta \end{cases},\quad J = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r$

$$\Rightarrow I^2 = \iint e^{-\frac{(x^2+y^2)}{2}} dxdy = \int_0^{2\pi} d\theta \int_0^{+\infty} r e^{-\frac{r^2}{2}} dr$$

$$= (\theta)_0^{2\pi} \left(-e^{-\frac{r^2}{2}} \right)_0^{\infty} = 2\pi$$

$\Rightarrow I = \sqrt{2\pi} \qquad \Rightarrow \int_{-\infty}^{+\infty} f(x)dx = 1$

正态分布密度函数的性质

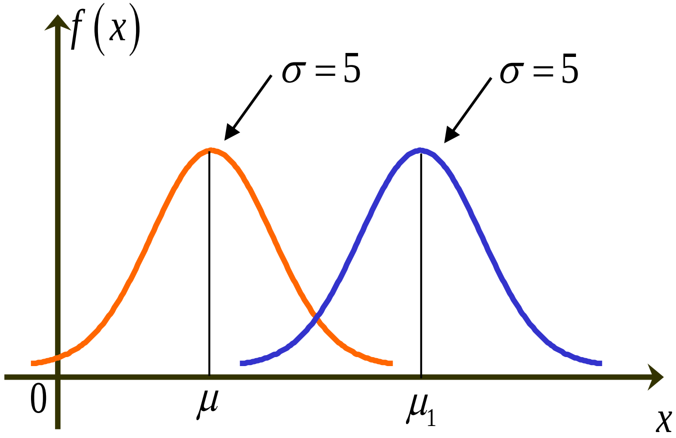
$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

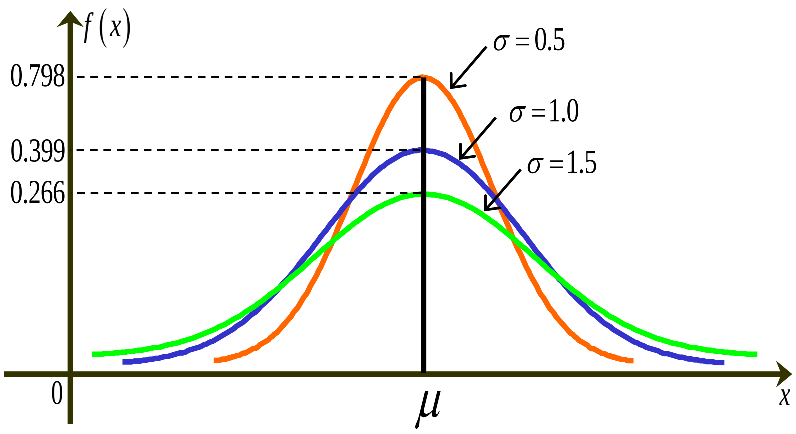
1° $f(x)$关于$x=\mu$对称

2° $f_{\max}(x)=f(\mu)=\dfrac{1}{\sqrt{2\pi}\sigma}$

3° $\lim\limits_{|x-\mu|\to\infty} f(x)=0$

μ为位置参数(决定对称轴位置)　　σ为尺度参数(决定曲线分散性)





正态分布密度函数的性质

X的取值呈中间多，两头少，对称的特性。

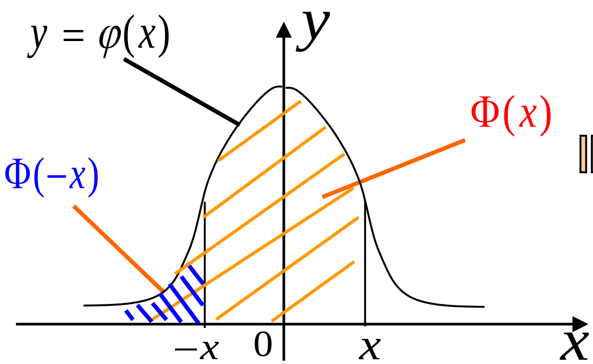
当固定μ时，σ越大，曲线的峰越低，落在μ附近的概率越小，取值就越分散，所以，σ是反映X的取值分散性的一个指标。

在自然现象和社会现象中，大量随机变量服从或近似服从正态分布。

记 $X \sim N(0,1)$，称X服从**标准正态分布**

X的概率密度：$\varphi(x)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$

X的分布函数：$\Phi(x)=\int_{-\infty}^{x}\varphi(t)dt=\int_{-\infty}^{x}\dfrac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$



$$\varphi(x)=\varphi(-x)$$
$$\Phi(x)+\Phi(-x)=1$$

查~~P295~~表： P364

	$\Phi(0)$	$\Phi(1)$	$\Phi(1.23)$	$\Phi(3.49)$
	0.5	0.8413	0.8907	0.9998

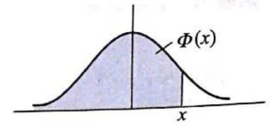
反查表：若$\Phi(c)=0.975$，则$c=?$　　$c=1.96$

反查表：若$\Phi(d)=0.025$，则$d=?$

$\Phi(-d)=1-\Phi(d)=1-0.025=0.975$，　$d=-c=-1.96$

附表 2　标准正态分布表

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



x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500 0	0.504 0	0.508 0	0.512 0	0.516 0	0.519 9	0.523 9	0.527 9	0.531 9	0.535 9
0.1	0.539 8	0.543 8	0.547 8	0.551 7	0.555 7	0.559 6	0.563 6	0.567 5	0.571 4	0.575 3
0.2	0.579 3	0.583 2	0.587 1	0.591 0	0.594 8	0.598 7	0.602 6	0.606 4	0.610 3	0.614 1
0.3	0.617 9	0.621 7	0.625 5	0.629 3	0.633 1	0.636 8	0.640 6	0.644 3	0.648 0	0.651 7
0.4	0.655 4	0.659 1	0.662 8	0.666 4	0.670 0	0.673 6	0.677 2	0.680 8	0.684 4	0.687 9
0.5	0.691 5	0.695 0	0.698 5	0.701 9	0.705 4	0.708 8	0.712 3	0.715 7	0.719 0	0.722 4
0.6	0.725 7	0.729 1	0.732 4	0.735 7	0.738 9	0.742 2	0.745 4	0.748 6	0.751 7	0.754 9
0.7	0.758 0	0.761 1	0.764 2	0.767 3	0.770 4	0.773 4	0.776 4	0.779 4	0.782 3	0.785 2
0.8	0.788 1	0.791 0	0.793 9	0.796 7	0.799 5	0.802 3	0.805 1	0.807 8	0.810 6	0.813 3
0.9	0.815 9	0.818 6	0.821 2	0.823 8	0.826 4	0.828 9	0.831 5	0.834 0	0.836 5	0.838 9
1.0	0.841 3	0.843 8	0.846 1	0.848 5	0.850 8	0.853 1	0.855 4	0.857 7	0.859 9	0.862 1
1.1	0.864 3	0.866 5	0.868 6	0.870 8	0.872 9	0.874 9	0.877 0	0.879 0	0.881 0	0.883 0
1.2	0.884 9	0.886 9	0.888 8	0.890 7	0.892 5	0.894 4	0.896 2	0.898 0	0.899 7	0.901 5
1.3	0.903 2	0.904 9	0.906 6	0.908 2	0.909 9	0.911 5	0.913 1	0.914 7	0.916 2	0.917 7
1.4	0.919 2	0.920 7	0.922 2	0.923 6	0.925 1	0.926 5	0.927 8	0.929 2	0.930 6	0.931 9
1.5	0.933 2	0.934 5	0.935 7	0.937 0	0.938 2	0.939 4	0.940 6	0.941 8	0.942 9	0.944 1
1.6	0.945 2	0.946 3	0.947 4	0.948 4	0.949 5	0.950 5	0.951 5	0.952 5	0.953 5	0.954 5
1.7	0.955 4	0.956 4	0.957 3	0.958 2	0.959 1	0.959 9	0.960 8	0.961 6	0.962 5	0.963 3
1.8	0.964 1	0.964 9	0.965 6	0.966 4	0.967 1	0.967 8	0.968 6	0.969 3	0.969 9	0.970 6
1.9	0.971 3	0.971 9	0.972 6	0.973 2	0.973 8	0.974 4	0.975 0	0.975 6	0.976 1	0.976 7
2.0	0.977 2	0.977 8	0.978 3	0.978 8	0.979 3	0.979 8	0.980 3	0.980 8	0.981 2	0.981 7
2.1	0.982 1	0.982 6	0.983 0	0.983 4	0.983 8	0.984 2	0.984 6	0.985 0	0.985 4	0.985 7
2.2	0.986 1	0.986 4	0.986 8	0.987 1	0.987 5	0.987 8	0.988 1	0.988 4	0.988 7	0.989 0
2.3	0.989 3	0.989 6	0.989 8	0.990 1	0.990 4	0.990 6	0.990 9	0.991 1	0.991 3	0.991 6
2.4	0.991 8	0.992 0	0.992 2	0.992 5	0.992 7	0.992 9	0.993 1	0.993 2	0.993 4	0.993 6
2.5	0.993 8	0.994 0	0.994 1	0.994 3	0.994 5	0.994 6	0.994 8	0.994 9	0.995 1	0.995 2
2.6	0.995 3	0.995 5	0.995 6	0.995 7	0.995 9	0.996 0	0.996 1	0.996 2	0.996 3	0.996 4
2.7	0.996 5	0.996 6	0.996 7	0.996 8	0.996 9	0.997 0	0.997 1	0.997 2	0.997 3	0.997 4
2.8	0.997 4	0.997 5	0.997 6	0.997 7	0.997 7	0.997 8	0.997 9	0.997 9	0.998 0	0.998 1
2.9	0.998 1	0.998 2	0.998 2	0.998 3	0.998 4	0.998 4	0.998 5	0.998 5	0.998 6	0.998 6
3.0	0.998 7	0.998 7	0.998 7	0.998 8	0.998 8	0.998 9	0.998 9	0.998 9	0.999 0	0.999 0
3.1	0.999 0	0.999 1	0.999 1	0.999 1	0.999 2	0.999 2	0.999 2	0.999 2	0.999 3	0.999 3
3.2	0.999 3	0.999 3	0.999 4	0.999 4	0.999 4	0.999 4	0.999 4	0.999 5	0.999 5	0.999 5
3.3	0.999 5	0.999 5	0.999 5	0.999 6	0.999 6	0.999 6	0.999 6	0.999 6	0.999 6	0.999 7
3.4	0.999 7	0.999 7	0.999 7	0.999 7	0.999 7	0.999 7	0.999 7	0.999 7	0.999 7	0.999 8

✦ 设 $X \sim N(\mu, \sigma^2)$，即X为一般正态分布。

$$\because F(x) = P(X \le x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \qquad \text{作变换：} \frac{t-\mu}{\sigma} = z$$

$$\Rightarrow F(x) = \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{z^2}{2}} \sigma dz = \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

$$P(a < X < b) = F(b) - F(a) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$P(X > a) = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right), \quad P(X < b) = \Phi\left(\frac{b-\mu}{\sigma}\right)$$

实际上，若$X \sim N(\mu, \sigma^2)$，则 $\dfrac{X-\mu}{\sigma} \sim N(0,1)$

Excel 中的正态分布函数

$X \sim N(\mu,\sigma^2)$ $f(x) = NORM.DIST(x,\mu,\sigma,FALSE)$

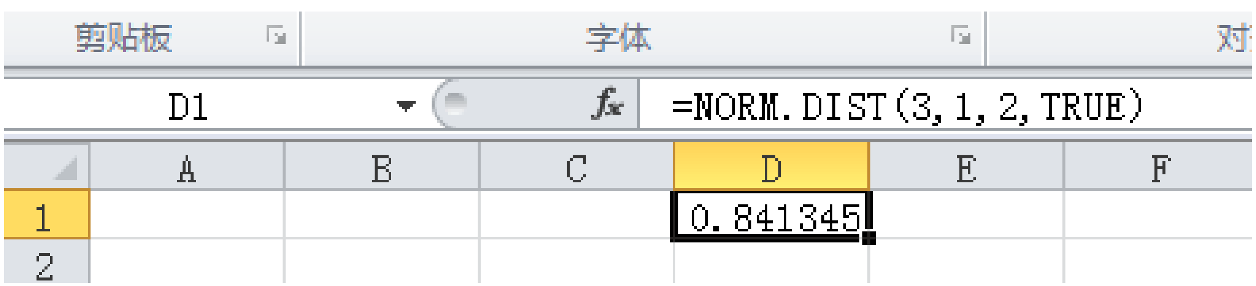
$F(x) = NORM.DIST(x,\mu,\sigma,TRUE)$

$X \sim N(0,1)$ $\varphi(x) = NORM.S.DIST(x,FALSE)$

$\Phi(x) = NORM.S.DIST(x,TRUE)$

例：$X \sim N(1,2^2)$，求 $F(3) = P(X \le 3)$

=NORM.DIST(3,1,2,TRUE)=0.841345

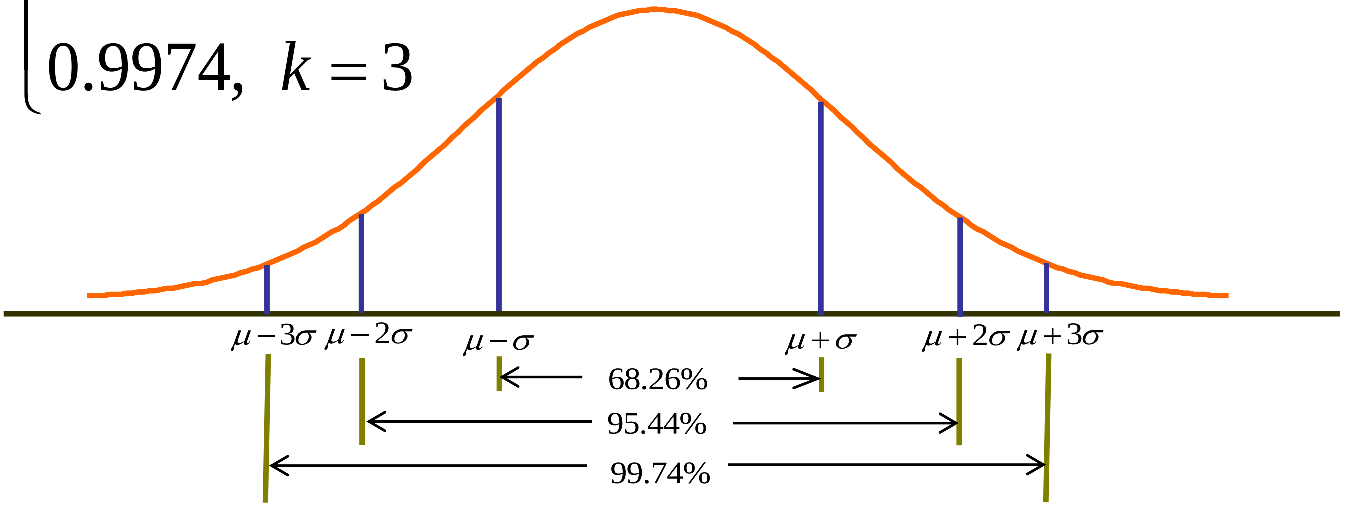


例5：设$X \sim N(\mu,\sigma^2)$，则

$$P(|X-\mu|<k\sigma)=P(\mu-k\sigma<X<\mu+k\sigma)$$

$$=\Phi\left(\frac{\mu+k\sigma-\mu}{\sigma}\right)-\Phi\left(\frac{\mu-k\sigma-\mu}{\sigma}\right)=2\Phi(k)-1$$

$$=\begin{cases}0.6826, & k=1\\ 0.9544, & k=2\\ 0.9974, & k=3\end{cases}$$



3σ法则：3σ之外的数据可认为异常数据。

例6：一批钢材(线材)长度 (cm) $X \sim N(\mu,\sigma^2)$

(1)若 μ=100， σ=2，求这批钢材长度小于97.8cm的概率；(2)若 μ=100，要使这批钢材的长度至少有90%落在区间(97,103)内，问 σ 至多取何值？

解:(1) $P(X < 97.8) = \Phi(\frac{97.8-100}{2}) = \Phi(-1.1) = 1-\Phi(1.1)$

$$\overset{\text{查附表}}{=\!=\!=} 1-0.8643 = 0.1357$$

(2) 令：$P\{97 < X < 103\} \geq 90\%$

即 $\Phi(\frac{103-100}{\sigma}) - \Phi(\frac{97-100}{\sigma}) = 2\Phi(\frac{3}{\sigma}) - 1 \geq 90\%$

$\Rightarrow \Phi(\frac{3}{\sigma}) \geq 0.95$ $\Rightarrow \frac{3}{\sigma} \geq 1.645$

$\Rightarrow \sigma \leq 1.8237$

例7：设某地区男子身高 $X(cm) \sim N(169.7, 4.1^2)$

(1) 从该地区随机找一男子测身高，求他的身高大于175cm的概率；(2) 若从中随机找5个男子测身高,问至少有一人身高大于175cm的概率是多少？恰有一人身高大于175cm的概率为多少？

解：（1）$P(X > 175) = 1 - \Phi(\frac{175 - 169.7}{4.1}) = 1 - \Phi(1.293)$

$$\overset{\text{查表}}{=\!=} 1 - 0.9015 = 0.0985$$

（2）设5人中有 Y 人身高大于175cm，则 $Y \sim B(5, p)$,

其中 $p = 0.0985$

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - (1 - p)^5 = 0.4045$$

$$P(Y = 1) = C_5^1 p^1 (1 - p)^4 = 0.3253$$

例8: 设从甲地到乙地有两条路，所需时间（分钟）分别是 $X \sim N(50,10^2)$ 和 $Y \sim N(60,4^2)$，若某人有**70**分钟时间从甲地赶往乙地，为及时参加重要会议，应该选哪条路？如果时间只有**65**分钟又该选哪条路？

解：

$$P(X \le 70) = \Phi(\frac{70-50}{10}) = \Phi(2)$$

$$P(Y \le 70) = \Phi(\frac{70-60}{4}) = \Phi(2.5)$$

$$P(X \le 65) = \Phi(\frac{65-50}{10}) = \Phi(1.5)$$

$$P(Y \le 65) = \Phi(\frac{65-60}{4}) = \Phi(1.25)$$

所以，有70分钟时选第二条路，有65分钟时选第一条路

§5 随机变量的函数分布

问题： 已知随机变量*X*的概率分布，且已知*Y*=*g*(*X*)，求*Y*的概率分布。

例如，若要测量一个圆的面积*Y*，总是测量其半径，半径的测量值可看作随机变量*X*，若 $X \sim N(\mu, \sigma^2)$，则*Y*服从什么分布？

例1：已知*X*具有概率分布，且设*Y*=*X*²，求*Y*的概率分布。

X	-1	0	1
P_k	0.2	0.4	0.4
$Y=X^2$	1	0	1

解：*Y*的所有可能取值为0, 1

$$P(Y=0)=P(X^2=0)=P(X=0)=0.4$$

$$P(Y=1)=P\{(X=1)\cup(X=-1)\}=P(X=1)+P(X=-1)=0.6$$

$\Rightarrow$

Y	0	1
P_k	0.4	0.6

即找出(*Y*=0)的等价事件(*X*=0)；

(*Y*=1)的等价事件(*X*=1)或(*X*=-1)

或把*Y*值列在表的下面，合并同值项，对应概率相加。

例2：设

X	-1	0	1
p	$1/3$	$1/3$	$1/3$

$Y=2X$, $Z=X^2$, 求Y, Z的概率分布。

X	-1	0	1
p	$1/3$	$1/3$	$1/3$
$Y=2X$	-2	0	2
$Z=X^2$	1	0	1

解：Y的可能取值为$-2, 0, 2$

Z的可能取值为$0, 1$

$(Y=-2)$的等价事件为$(X=-1)$…

$(Z=1)$的等价事件为$(X=1)\cup(X=-1)$

故得：

Y	-2	0	2
p	$1/3$	$1/3$	$1/3$

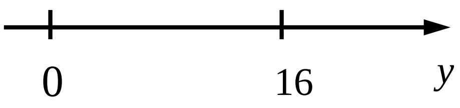
Z	0	1
p	$1/3$	$2/3$

例3：设随机变量*X*具有概率密度 $f_X(x)=\begin{cases}\dfrac{x}{8}, & 0<x<4\\ 0, & 其他\end{cases}$

求*Y*=*X*²的概率密度。

解： 分别记*X*,*Y*的分布函数为 $F_X(x)$，$F_Y(y)$

$$F_Y(y)=P\{Y\le y\}=P\left\{X^2\le y\right\}$$



当 $y\le 0$ 时，$F_Y(y)=0$;　　当 $y\ge 16$时，$F_Y(y)=1$;

当 $0<y<16$ 时，$F_Y(y)=P\{Y\le y\}=P\left\{-\sqrt{y}<X<\sqrt{y}\right\}$

$$=P\left(\left\{-\sqrt{y}<X\le 0\right\}\cup\left\{0<X<\sqrt{y}\right\}\right)=P\left(0<X<\sqrt{y}\right)$$

$$=\int_0^{\sqrt{y}}\frac{t}{8}dt=\frac{y}{16}$$

$$=\int_{-\sqrt{y}}^{\sqrt{y}}\frac{t}{8}dt=0?$$

$$\therefore f_Y(y)=F_Y'(y)=\begin{cases}\dfrac{1}{16}, & 0<y<16\\ 0, & 其他\end{cases}$$

*Y*在区间 $(0, 16)$ 上均匀分布。

$$f_X(x)=\begin{cases}\dfrac{x}{8}, & 0<x<4\\ 0, & \text{其他}\end{cases}$$

注意, 当 $0<y<16$ 时,

$$F_Y(y)=\cdots=P\left\{0<X<\sqrt{y}\right\}=\int_0^{\sqrt{y}}\frac{t}{8}dt=\frac{y}{16}$$

可用以下方法代替:

$$F_Y(y)=\cdots=P\left\{0<X<\sqrt{y}\right\}=F_X(\sqrt{y})-F_X(0)$$

$$f_Y(y)=F_Y'(y)=f_X(\sqrt{y})\left(\sqrt{y}\right)'-0$$

$$=f_X(\sqrt{y})\frac{1}{2\sqrt{y}}=\frac{\sqrt{y}}{8}\frac{1}{2\sqrt{y}}=\frac{1}{16}$$

*例3：设随机变量X具有概率密度 $f_X(x)=\begin{cases}\dfrac{x}{8}, & 0<x<4\\ 0, & 其他\end{cases}$

求$Y=X^2$的概率密度。

解： $F_Y(y)=P\{Y\le y\}=P\{X^2\le y\}$

当 $y\le 0$时，$F_Y(y)=0$，$f_Y(y)=F_Y'(y)=0$

当 $y>0$ 时，$F_Y(y)=P\{-\sqrt{y}<X<\sqrt{y}\}$ $=F_X(\sqrt{y})-F_X(-\sqrt{y})$

$$f_Y(y)=f_X(\sqrt{y})\frac{1}{2\sqrt{y}}+f_X(-\sqrt{y})\frac{1}{2\sqrt{y}}=\frac{1}{2\sqrt{y}}f_X(\sqrt{y})$$

当 $0<y<16$时，$f_Y(y)=\dfrac{1}{2\sqrt{y}}\dfrac{\sqrt{y}}{8}=\dfrac{1}{16}$，

当 $y\ge 16$时，$f_Y(y)=0$

综上，$f_Y(y)=\begin{cases}\dfrac{1}{16}, & 0<y<16\\ 0, & 其他\end{cases}$

例4：随机变量X的密度函数$f(x)=\begin{cases}\frac{3}{2}x^2, -1<x<1\\ 0, \quad 其他\end{cases}$，

$Y=1-|X|$，求Y的密度函数。

解:当X在$(-1,1)$取值时，$Y=1-|X|$在$(0,1]$内取值。

$\therefore$ 当$y\le 0, F_Y(y)=P(Y\le y)=0,\ \ f_Y(y)=F_Y'(y)=0$

当$y>1, F_Y(y)=P(Y\le y)=1,\ \ f_Y(y)=F_Y'(y)=0$

当$0<y\le 1$时，$F_Y(y)=P(Y\le y)=P(1-|X|\le y)$

$$=P(|X|\ge 1-y)=1-P(y-1<X<1-y)$$

$$=1-F_X(1-y)+F_X(y-1)$$

$$f_Y(y)=0-f_X(1-y)(1-y)'+f_X(y-1)(y-1)'$$

$$=f_X(1-y)+f_X(y-1)$$

例4：随机变量X的密度函数$f(x)=\begin{cases}\dfrac{3}{2}x^2,-1<x<1\\ 0,\quad\text{其他}\end{cases}$，

$Y=1-|X|$，求Y的密度函数。

$$f_Y(y)=f_X(1-y)+f_X(y-1)$$

$$=\frac{3}{2}(1-y)^2+\frac{3}{2}(y-1)^2=3(y-1)^2$$

$$\therefore f_Y(y)=\begin{cases}3(y-1)^2,0<y\le 1\\ 0,\qquad\text{其他}\end{cases}$$

*例5：设X的概率密度为$f(x)$，$|x|<\infty$，$Y=X^2$，

求Y的概率密度$f_Y(y)$.

解：设Y的概率分布函数为$F_Y(y)$

当 $y>0$时，$F_Y(y)=P(Y\le y)=P(X^2\le y)=\int_{-\sqrt{y}}^{\sqrt{y}}f(t)dt$

$f(x)$连续时，$\frac{d}{dx}\int_a^x f(t)dt=f(x)$

$\frac{d}{dx}\int_a^{u(x)} f(t)dt=f(u(x))u'(x)$

$=\int_0^{\sqrt{y}}f(t)dt-\int_0^{-\sqrt{y}}f(t)dt$

或$F_Y(y)=P(X^2\le y)=P(-\sqrt{y}<X<\sqrt{y})$

$=F_X(\sqrt{y})-F_X(-\sqrt{y})$

$$\Rightarrow f_Y(y)=F_Y'(y)=\begin{cases}[f(\sqrt{y})+f(-\sqrt{y})]\dfrac{1}{2\sqrt{y}}, & y>0\\ 0, & y\le 0\end{cases}$$

例6：若$X \sim U(-1,0)\cup(1,2)$， $Y = X^2$， 求 $f_X(x)$及 $f_Y(y)$

解： $f_X(x) = \begin{cases} \dfrac{1}{1+1} = \dfrac{1}{2}, x \in (-1,0)\cup(1,2) \\ 0, \quad\quad 其他 \end{cases}$ $F_Y(y) = P(Y \le y) = P(X^2 \le y)$

$y \le 0,\ F_Y(y) = 0$

$y > 0,\ F_Y(y) = P(Y \le y) = P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$

$$\therefore f_Y(y) = \begin{cases} \dfrac{1}{2\sqrt{y}}\left[f_X(\sqrt{y}) + f_X(-\sqrt{y}) \right], y > 0 \\ 0, \quad\quad y \le 0 \end{cases}$$

当$x = -1,0,1,2$时， $y = 0,1,4$ $\therefore$考虑区间（$-\infty$,0]$(0,1)[1,4)[4,+\infty)$

当$y \le 0$时，$f_Y(y) = 0$

当$0 < y < 1$时，$f_Y(y) = \dfrac{1}{2\sqrt{y}}\left[0 + \dfrac{1}{2}\right] = \dfrac{1}{4\sqrt{y}}$

当$1 \le y < 4$时，$f_Y(y) = \dfrac{1}{2\sqrt{y}}\left[\dfrac{1}{2} + 0\right] = \dfrac{1}{4\sqrt{y}}$

当$y \ge 4$时，$f_Y(y) = \dfrac{1}{2\sqrt{y}}[0+0] = 0$ 综上， $f_Y(y) = \begin{cases} \dfrac{1}{4\sqrt{y}}, 0 < y < 4 \\ 0, \quad 其他 \end{cases}$

例7：设$X \sim N(\mu,\sigma^2)$, $Y=aX+b(a\neq 0)$, 求Y的概率密度$f_Y(y)$.

解：$F_Y(y)=P(Y\leq y)=P(aX+b\leq y)=P(aX\leq y-b)$

当$a>0$时, $F_Y(y)=P(X\leq \frac{y-b}{a})=F_X(\frac{y-b}{a})$

$$f_Y(y)=F_Y'(y)=f_X(\frac{y-b}{a})\frac{1}{a}$$

当$a<0$时, $F_Y(y)=P(X\geq \frac{y-b}{a})=1-F_X(\frac{y-b}{a})$

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

$$f_Y(y)=F_Y'(y)=f_X(\frac{y-b}{a})\frac{1}{-a}$$

$$\therefore f_Y(y)=f_X(\frac{y-b}{a})\frac{1}{|a|}=\frac{1}{\sqrt{2\pi}|a|\sigma}e^{-\frac{[y-(a\mu+b)]^2}{2a^2\sigma^2}}$$

即：$Y=aX+b\sim N(a\mu+b,a^2\sigma^2)$

正态分布重要结论

若$X \sim N(\mu,\sigma^2)$，则$Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$

即，正态分布的随机变量线性变换后正态性不变

证明前面已述结论：若$X \sim N(\mu,\sigma^2)$, 则$\dfrac{X-\mu}{\sigma} \sim N(0,1)$

$$\because \frac{X-\mu}{\sigma} = \frac{1}{\sigma}X + (-\frac{\mu}{\sigma}) \qquad \text{其中 } a = \frac{1}{\sigma}, \quad b = -\frac{\mu}{\sigma}$$

$$a\mu + b = \frac{1}{\sigma}\mu - \frac{\mu}{\sigma} = 0 \qquad a^2\sigma^2 = \frac{1}{\sigma^2}\sigma^2 = 1$$

$$\therefore \frac{X-\mu}{\sigma} = \frac{1}{\sigma}X + (-\frac{\mu}{\sigma}) \sim N(0,1)$$

例8：设某纺纱机在任一时间间隔t（分钟）内出现的断头次数 $N(t) \sim P(\lambda t)$，求

（1）首次断头在10分钟以后出现的概率；

（2）两次断头之间的间隔时间Y（分钟）的概率密度。

解：泊松分布分布律：$P(N(t)=k)=\dfrac{(\lambda t)^k e^{-\lambda t}}{k!}$，$k=0,1,2,\cdots$

$$(1)P(N(10)=0)=\frac{(10\lambda)^0 e^{-10\lambda}}{0!}=e^{-10\lambda}$$

$$(2)y>0,\ F_Y(y)=P(Y\le y)\ =1-P(Y>y)$$

$$=1-P(N(y)=0)\ =1-\frac{(\lambda y)^0 e^{-\lambda y}}{0!}=1-e^{-\lambda y}$$

$$\therefore F_Y(y)=\begin{cases}1-e^{-\lambda y}, & y>0\\ 0, & y\le 0\end{cases} \Rightarrow f_Y(y)=F_Y'(y)=\begin{cases}\lambda e^{-\lambda y}, & y>0\\ 0, & y\le 0\end{cases}$$

Y服从指数分布！

*例9：某大型设备在任何长度为t的区间内发生故障的次数$N(t) \sim P(\lambda t)$,即服从参数为λt 的Poisson分布，记设备无故障运行时间为T.

(1) 求T的概率分布函数；

(2) 已知设备无故障运行t_0个小时，求再无故障运行t个小时的概率。

解：(1) $P\{N(t)=k\}=\dfrac{(\lambda t)^k e^{-\lambda t}}{k!},\ k=0,1,2,\cdots$ 　　指数分布

当$t \le 0$时，$F_T(t)=P\{T \le t\}=0$

当$t>0$时,

$$F_T(t)=P\{T \le t\}=1-P\{T>t\}=1-P\{N(t)=0\}=1-e^{-\lambda t}$$

$$\boxed{\therefore F_T(t)=\begin{cases}1-e^{-\lambda t}, & t>0\\ 0, & t \le 0\end{cases}}$$

$$(2)\ P\{T \ge t_0+t \mid T>t_0\}=\frac{P\{T>t_0+t\}}{P\{T>t_0\}}=e^{-\lambda t}=P\{T>t\}$$

其中：$P\{T>t_0+t\}=1-F(t_0+t)=1-(1-e^{-\lambda(t_0+t)})=e^{-\lambda(t_0+t)}$

$P\{T>t_0\}=1-F(t_0)=1-(1-e^{-\lambda t_0})=e^{-\lambda t_0}$

例10：设$X \sim f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0\\ 0, & x\le 0\end{cases}$，$F(x)$为$X$的分布函数。

设$Y=F(X)$,试证$Y\sim U(0,1)$ (即均匀分布)。

解：由前知, $F(x)=\int_{-\infty}^{x} f(x)dx=\begin{cases}1-e^{-\lambda x}, & x>0\\ 0, & x\le 0\end{cases}$

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

$\Rightarrow 0\le Y\le 1$

注意：$P(X>0)=1$

记$F_Y(y)$为Y的概率分布函数,

当$y\le 0$时，$F_Y(y)=P\{Y\le y\}=0$

当$y\ge 1$时，$F_Y(y)=P\{Y\le y\}=1$

当$0<y<1$时，$F_Y(y)=P\{1-e^{-\lambda X}\le y\}=P\{e^{-\lambda X}\ge 1-y\}$

$$=P\left\{X\le -\frac{1}{\lambda}ln(1-y)\right\}=F(-\frac{1}{\lambda}ln(1-y))=1-e^{-\lambda\left[-\frac{1}{\lambda}ln(1-y)\right]}=y$$

即 $F_Y(y)=\begin{cases}0, & y\le 0\\ y, & 0<y<1\\ 1, & y\ge 1\end{cases}$, $\therefore\ Y\sim U(0,1)$.

*随机数的产生

用均匀分布随机数产生λ=3的指数分布，即

$$f_X(x)=\begin{cases}3e^{-3x}, x\geq 0\\ 0 \quad\quad, x<0\end{cases}，\quad F_X(x)=\begin{cases}1-e^{-3x}, x\geq 0\\ 0 \quad\quad\quad, x<0\end{cases}$$

$$\therefore 当x\geq 0, y=F_X(x)=1-e^{-3x}, x=-\frac{1}{3}\ln(1-y)$$

先产生100个均匀分布随机数$Y\sim U(0,1)$,再得到100个

$X=-\dfrac{1}{3}\ln(1-Y)$就是服从指数分布的随机数。

一般，若已知***X***的概率分布，***Y=g*（*X*）**，求***Y***的概率分布的过程为：

1. 若Y为离散量，则先写出Y的可能取值：$y_1, y_2, \cdots y_i, \cdots,$
再找出$\left(Y = y_i\right)$的等价事件$(X \in D)$，得$P(Y = y_i) = P(X \in D)$;

或把***Y***值列在表的下面，合并同值项，对应概率相加。

2. 若Y为连续量，则先写出Y的概率分布函数：
$F_Y(y) = P(Y \le y)$，找出$\left(Y \le y\right)$的等价事件$(X \in D)$，
得$F_Y(y) = P(X \in D)$；再求出Y的概率密度函数$f_Y(y)$；

✦ 不管哪种类型，关键是找出等价事件。

定理：设$X \sim f_X(x), -\infty < x < +\infty$，$Y = g(X)$，$g'(x) > 0$

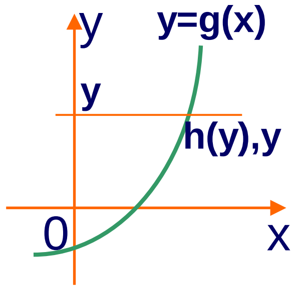
(或$g'(x) < 0$)，反函数为$X = h(Y)$, 则：

$$f_Y(y) = \begin{cases} f_X(h(y)) \cdot |h'(y)|, & \alpha < y < \beta \\ 0, & \text{其他} \end{cases}$$

其中$\alpha = g(-\infty)$, $\beta = g(\infty)$\{当$g'(x) < 0$时$\alpha = g(\infty)$, $\beta = g(-\infty)$\}

证明：不妨设$g'(x) > 0$,　　则$g(x)$为单调增函数,

而且有 $h'(y) > 0$



显然，当 $y \le \alpha$ 时, $F_Y(y) = 0$; 当 $y \ge \beta$ 时，$F_Y(y) = 1$;

当 $\alpha < y < \beta$ 时，$F_Y(y) = P(Y \le y) = P(g(X) \le y)$

$$= P(X \le h(y)) = \int_{-\infty}^{h(y)} f_X(t)dt \overset{\text{或}}{=} F_X(h(y))$$

$$\Rightarrow f_Y(y) = f_X(h(y))h'(y) = f_X(h(y)) \cdot |h'(y)|$$

同理可证：当$g'(x) < 0$ 时，定理为真.

例11：若$X \sim f(x)=\begin{cases}\dfrac{x}{8}, & 0<x<4 \\ 0, & \text{其他}\end{cases}$ $Y=X^3$，求 $f_Y(y)$。

解：$y=g(x)=x^3$, $\quad g'(x)=3x^2>0$,

$$x=h(y)=y^{\frac{1}{3}}, \quad x'=h'(y)=\frac{1}{3}y^{-\frac{2}{3}}$$

$\therefore$ 当$0<x<4$ 即 $0<y=x^3<64$时，

$$f_Y(y)=f_X(h(y))\cdot|h'(y)|=f_X(y^{\frac{1}{3}})\frac{1}{3}y^{-\frac{2}{3}}$$

$$f_Y(y)=\begin{cases}\dfrac{1}{24}y^{-\frac{1}{3}}, & 0<y<64 \\ 0, & \text{其他}\end{cases}$$

例12：设$X \sim N(\mu, \sigma^2)$, $Y = aX + b(a \neq 0)$, 求Y的概率密度$f_Y(y)$.

解：$y = g(x) = ax + b$, $\quad g'(x) = a \neq 0$,

$$x = h(y) = \frac{y-b}{a}, \quad x' = h'(y) = \frac{1}{a}$$

$$f_Y(y) = f_X\left(\frac{y-b}{a}\right)\frac{1}{|a|}$$

$$= \frac{1}{\sqrt{2\pi}|a|\sigma} e^{-\frac{[y-(a\mu+b)]^2}{2a^2\sigma^2}}$$

$$\Rightarrow Y \sim N(a\mu + b, a^2\sigma^2)$$

推论：设$Y=g(X)$的反函数为$X=h(Y)$, X的密度函数为$f_X(x)$, 在$f_X(x)$的非零区间(a,b)上有$g'(x)>0$（或$g'(x)<0$）。则：

$$f_Y(y)=\begin{cases} f_X(h(y))\cdot\left|h'(y)\right|, & \alpha<y<\beta \\ 0, & \text{其他} \end{cases}$$

其中$\alpha=\min(g(a),g(b))$, $\beta=\max(g(a),g(b))$。

说明：可以对定理放宽条件，即$y=g(x)$只要在$f(x)$的非零区域上单调即可。

当然，函数关系不是单调，就需从定义出发求解。

例13： $X \sim f(x)=\begin{cases}\dfrac{x}{8}, & 0<x<4 \\ 0, & \text{其他}\end{cases}$ $\quad Y=X^2$, 求 $f_Y(y)$。

$\because \; Y=g(X)=X^2$， $y'=g'(x)=2x$ 虽然x在（$-\infty,\infty$）上$y=g(x)$不是单调的，但是在$f(x)$的非零区间$(0,4)$上是单调递增的，故可应用定理的推论。

在$0<x<4$上，$x=h(y)=\sqrt{y}$， $x'=h'(y)=\dfrac{1}{2\sqrt{y}}$

$\therefore$ 当$0<y=x^2<16$时，$f_Y(y)=f_X(h(y))\left|h'(y)\right|$

$$=f_X(\sqrt{y})\frac{1}{2\sqrt{y}}=\frac{\sqrt{y}}{8}\frac{1}{2\sqrt{y}}=\frac{1}{16}$$

$$\therefore \quad f_Y(y)=\begin{cases}\dfrac{1}{16} & 0<y<16 \\ 0 & \text{其他}\end{cases}$$

复习思考题 2

1. 什么量被称为随机变量？它与样本空间的关系如何？
2. 满足什么条件的试验称为“n重贝努里试验”？
3. 事件A在一次试验中发生的概率为p, $0<p<1$。若在n次独立重复的试验中，A发生的总次数为X, 则X服从什么分布？并请导出：

   $$P\{X=k\}=C_n^k p^k(1-k)^{n-k},\ k=0,1,\cdots,n$$
4. 什么条件下使用泊松近似公式等式较为合适？
5. 什么样的随机变量称为连续型的？
6. 若事件A为不可能事件，则$P(A)=0$, 反之成立吗？又若A为必然事件，则$P(A)=1$，反之成立吗？
7. 若连续型随机变量X在某一区间上的概率密度为0，则X落在该区间的概率为0，对吗？
8. 若随机变量X在区间(a,b)上均匀分布，则X落入(a,b)的任意一子区间 (a_1,b_1)上的概率为$(b_1-a_1)/(b-a)$, 对吗？
9. 若$X\sim N(\mu,\sigma^2)$，则X的概率密度函数$f(x)$在$x=\mu$处值最大，因此X落在μ附近的概率最大，对吗？