第二章　随机变量及其分布

随机变量

离散型随机变量

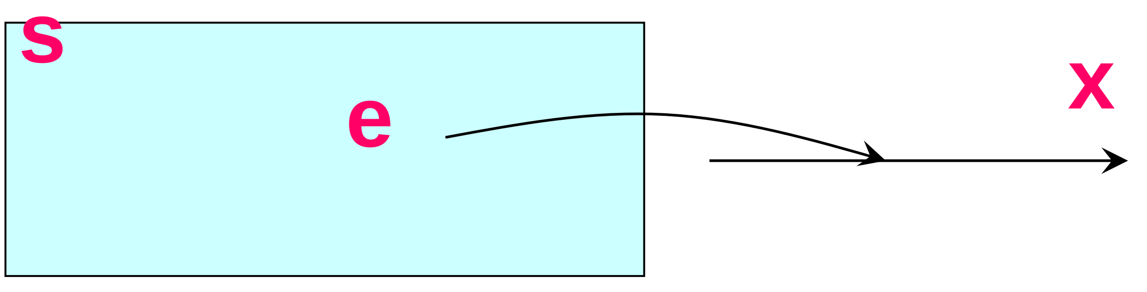
分布函数

连续型随机变量

随机变量的函数

2.1 随机变量

将随机试验结果数量化



$X=X(e)$——为S上的实值单值函数

定义：设随机试验的样本空间 $S=\{e\}$，若 $X = X(e)$ 为定义在样本空间 S 上的实值单值函数，则称 $X = X(e)$ 为随机变量。

一般地，若 I 是一个实数集合，则

$\{X \in I\}$ 为事件 $\{e: X(e) \in I\}$

- 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 $\begin{cases} \text{离散型的} \\ \text{连续型的} \end{cases}$

例1.1 掷硬币3次，出现正面的次数记为*X*.

样本点	TTT	TTH	THT	HTT	HHT	HTH	THH	HHH
*X*的值	0	1	1	1	2	2	2	3

$$P\{X=0\}=P\{TTT\}=1/8$$

$$P\{X=1\}=P\{TTH,THT,HTT\}=3/8$$

$$P\{X\le 1\}=P\{X=0\}+P\{X=1\}=1/2$$

X	0	1	2	3
p	1/8	3/8	3/8	1/8

2.2　离散型随机变量及其分布

定义：取值至多可数的随机变量为**离散型的随机变量**。概率分布律为

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

概率分布律性质： $p_i \geq 0, \sum_{i=1}^{\infty} p_i = 1$

概率分布律

写出所有可能取值；

写出每个取值相应的概率.

例2.1 某人骑自行车从学校到火车站，一路上要经过3个独立的交通信号灯，设各灯工作独立，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信号灯数，求X的概率分布律。

解：设$\boldsymbol{A_i}$={第$\boldsymbol{i}$个灯为红灯}，则$\boldsymbol{P(A_i)=p}$，$\boldsymbol{i}$=1, 2, 3 且$\boldsymbol{A_1, A_2, A_3}$相互独立。

$$P(X=0)=P(A_1)=p\ ;$$

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

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

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

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

例2. 2 若随机变量***X***的概率分布律为

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

求常数***c***.

解：$$1=\sum_{k=0}^{+\infty}P\{X=k\}$$

$$=c\sum_{k=0}^{+\infty}\frac{\lambda^{k}}{k!}=ce^{\lambda}$$

$$\Rightarrow \quad c=e^{-\lambda}$$

几个重要的离散型随机变量分布

一、0—1分布

若X的分布律为：

X	0	1
P	q	p

随机变量只可能取0、1 两个值

$$(p+q=1, p>0, q>0)$$

则称X服从参数为p的0-1分布，或两点分布.

记为 $X \sim 0-1(p)$ 或 $B(1, p)$

$0\text{-}1(\boldsymbol{p})$分布的分布律还可以写为

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

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。

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

来描述这个随机试验的结果。

检查产品的质量是否合格，对新生婴儿的性别进行登记，检验种子是否发芽以及前面多次讨论过的“抛硬币”试验都可以用（0-1）分布的随机变量来描述。

一个随机试验,设A是一随机事件,且$P(A)=p,(0<p<1)$.若仅考虑事件A发生与否,定义一个服从参数为p的0-1分布的随机变量:

$$X=\begin{cases}1, & 若A发生, \\ 0, & 若A不发生(即\overline{A}发生).\end{cases}$$

来描述这个随机试验的结果。

只有两个可能结果的试验，称为贝努利试验。

二、二项分布

n重贝努利试验：设试验E只有两个可能的结果：A与$\bar{A}$，$P(A)=p$，$0<p<1$，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。

即每次试验结果互不影响

在相同条件下重复进行

- 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，

  $$P(\text{出现正面}) = 1/2.$$

- 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：$A, \bar{A}$,

  $$P(A) = 1/6.$$

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：$A, \bar{A}$,

$$P(A)=1/2.$$

如果是不放回抽样呢？

不是独立试验！

设***A***在***n***重贝努利试验中发生***X***次，则

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

并称***X***服从参数为***n,p***的二项分布，记

$$X \sim B(n,\ p)$$

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

推导：以n=3为例，设A_i={ 第i次A发生 }

$$P(X=0)=P(\bar{A}_1\bar{A}_2\bar{A}_3)=(1-p)^3$$

$$P(X=1)=P(A_1\bar{A}_2\bar{A}_3\cup\bar{A}_1A_2\bar{A}_3\cup\bar{A}_1\bar{A}_2A_3)=C_3^1p^1(1-p)^{3-1}$$

$$P(X=2)=P(A_1A_2\bar{A}_3\cup A_1\bar{A}_2A_3\cup\bar{A}_1A_2A_3)=C_3^2p^2(1-p)^{3-2}$$

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

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

例2.3 有一大批产品，其验收方案如下：先作第一次检验，从中任取10件，经检验无次品接收这批产品，次品数大于2拒收；否则作第二次检验，从中任取5件，仅当5件中无次品便接收这批产品，设产品的次品率为p．求这批产品能被接收的概率.

解：设A={接收该批产品}。 设X为第一次抽得的次品数，Y为第2次抽得的次品数.

则$X \sim B(10, p), Y \sim B(5, p)$, 且$\{X = i\}$与$\{Y = j\}$独立.

$$P(A) = P(X = 0) + P(1 \le X \le 2 \text{且} Y = 0)$$

$$= P(X = 0) + P(1 \le X \le 2) \cdot P(Y = 0)$$

$$= P(X = 0) + (P(X = 1) + P(X = 2)) \cdot P(Y = 0)$$

$$= (1-p)^{10} + [10p(1-p)^9 + 45p^2(1-p)^8] \cdot (1-p)^5$$

例2.4 设随机变量 $X \sim B(100, 0.05)$,

求$P(X \leq 10)$和$P(X = 10)$

解：$P(X \leq 10) = \sum_{k=0}^{10} P(X = k)$

$$= \sum_{k=0}^{10} C_{100}^{k} 0.05^{k} 0.95^{100-k}$$

使用*Excel*表单：
在任一单元格中输入
“=*BINOM.DIST*(10,100,0.05,*TRUE*)”，
点“确定”后，在单元格中出现“0.988528”.
这里“*TRUE*”可用“1”代替.

计算$P(X=10)$，
在任一单元格中输入
“=*BINOM.DIST*(10,100,0.05, *FALSE*)”，
点“确定”后，在单元格中出现“0.016715884”.
这里“*FALSE*”可用“0”代替.

三、泊松分布(Poisson分布)

若随机变量X的概率分布律为

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

称X服从参数为λ的泊松分布，记

$$X \sim P(\lambda)$$

例2.5 某公交站单位时间内候车人数

$$X \sim P(4.8),$$

求(1)随机观察1个单位时间，至少有3人候车的概率；

(2)随机独立观察5个单位时间，恰有4个单位时间至少有3人候车的概率.

解:(1) $P(X \geq 3) = 1 - P(X = 0) - P(X = 1) - P(X = 2)$

$$= 1 - e^{-4.8}(1 + 4.8 + \frac{4.8^2}{2!}) = 0.8580$$

(2) 设5个单位时间内有Y个单位时间是“至少有3人候车”，

则$Y \sim B(5, p)$, 其中$p = P(X \geq 3) = 0.8580$,

于是 $P(Y = 4) = C_5^4 p^4 (1 - p) = 0.7696.$

泊松定理：

二项分布与泊松分布有下面的近似结果

当$n>10, p<0.1$时，

$$C_n^k p^k(1-p)^{n-k} \approx \frac{e^{-\lambda}\lambda^k}{k!}, \text{其中} \lambda = np.$$

事实上， $C_n^k p^k\left(1-p\right)^{n-k}=\dfrac{n!}{k!(n-k)!}\left(\dfrac{\lambda}{n}\right)^k\left(1-\dfrac{\lambda}{n}\right)^{n-k}$

$$=\frac{\lambda^k}{k!}\frac{n(n-1)\ldots(n-k+1)}{n^k}\left[\left(1-\frac{\lambda}{n}\right)^{-n/\lambda}\right]^{-\lambda}\Big/\left(1-\frac{\lambda}{n}\right)^k$$

$$\approx\frac{e^{-\lambda}\lambda^k}{k!}$$

因为当n充分大和适当的λ时，.

$$\frac{n(n-1)\ldots(n-k+1)}{n^k}\approx 1,\left(1-\frac{\lambda}{n}\right)^k\approx 1,\left[\left(1-\frac{\lambda}{n}\right)^{-n/\lambda}\right]^{-\lambda}\approx e^{-\lambda}$$

例2.6 某地区一个月内成年人患某种疾病的患病率为1/200，设各人是否患病相互独立。若该地区一社区有1000个成年人，求某月内该社区至少有3人患病的概率。

解：设该社区1000人中有X 个人患病，

则 $X \sim B(1000, \dfrac{1}{200})$,

$$P(X \geq 3) = 1 - P(X = 0) - P(X = 1) - P(X = 2)$$

$$= 0.8760;$$

利用泊松分布进行近似计算，取$\lambda = 5$,

$$P(X \geq 3) \approx 1 - \frac{e^{-5}}{0!} - \frac{5e^{-5}}{1!} - \frac{5^2 e^{-5}}{2!} = 0.8753.$$

泊松分布使用*Excel*表单：
在*Excel*的任一单元格输入“*=POISSON.DIST*(2,5,1)”，回车，就在单元格中出现“0.124652019”.

$$P(X \geq 3) = 1 - P(X \leq 2) = 0.875347981.$$

四、超几何分布

若随机变量*X*的概率分布律为

$$P(X=k)=\frac{C_a^k C_b^{n-k}}{C_N^n}, k=l_1, l_1+1, \ldots, l_2,$$

其中，$l_1=\max(0, n-b),\ l_2=\min(a, n).$

称*X*服从超几何分布.

例2.7　一袋中有a个白球，b个红球，$a+b=N$, 从中不放回地取n个球，设每次取到各球的概率相等，以X表示取到的白球数，则X服从超几何分布。

五、几何分布

若随机变量X的概率分布律为

$$P(X=k)=p(1-p)^{k-1}, k=1,2,3,\ldots, \quad 0<p<1.$$

称X服从参数p的几何分布.

例2.8 从生产线上随机抽产品进行检测，设产品的次品率为p，$0<p<1$，若查到一只次品就得停机检修，设停机时已检测到X只产品，则X服从参数p的几何分布。

六、巴斯卡分布

若随机变量***X***的概率分布律为

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

其中r为正整数，$0<p<1$.

称***X***服从参数为(***r,p***)的巴斯卡分布.

例2.9 独立重复地进行试验，每次试验的结果为成功或失败，每次试验中成功的概率均为p，$0<p<1$，试验进行到出现r次成功为止，以X表示试验次数，则X服从参数为(r,p)的巴斯卡分布。

思考题：一盒中有2个红球4个白球，

（1）从中取一球，X表示取到的红球数；

（2）采用不放回抽样取3球，Y表示取到的红球数；

（3）采用放回抽样取3球，Z表示取到的红球数；

（4）采用放回抽样取球，直到取到红球为止，U表示取球次数；

（5）采用放回抽样取球，直到取到3个红球为止，V表示取球次数。

上述随机变量X,Y,Z,U,V的分布律是什么呢？

解答：(1)*X*服从0－1分布，*P*(*X*=1)=1/3，*P*(*X*=0)=2/3；

(2)*Y*服从超几何分布，

$$P(Y=k)=\frac{C_2^k C_4^{3-k}}{C_6^3}, k=0,1,2;$$

(3)*Z*服从二项分布*B*(3, 1/3)，

$$P(Z=k)=C_3^k\frac{2^{3-k}}{3^3}, k=0,1,2,3;$$

(4)*U*服从几何分布，

$$P(U=k)=\frac{2^{k-1}}{3^k}, k=1,2,3,\dots$$

(5)*V*服从巴斯卡分布，

$$P(V=k)=C_{k-1}^2\frac{2^{k-3}}{3^k}, k=3,4,5,\dots$$

2.3　随机变量的分布函数

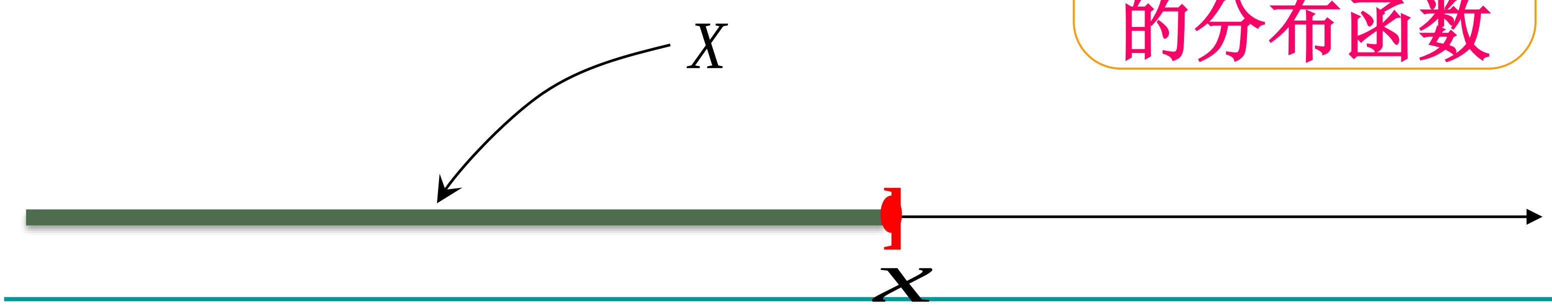
定义：随机变量X，若对任意实数x，函数

$$F(x) = P(X \le x)$$

称为X 的分布函数.

任何随机变量都有相应的分布函数

$F(x)$的几何意义：



分布函数的性质：

1) $0 \le F(x) \le 1$;

2) $F(x)$单调不减，且$F(-\infty)=0$，$F(+\infty)=1$

$\because \quad 0 \le P(x_1 < X \le x_2) = F(x_2) - F(x_1)$;

3) $F(x)$右连续,即$F(x+0)=F(x)$;

4) $F(x)-F(x-0)=P(X=x)$.

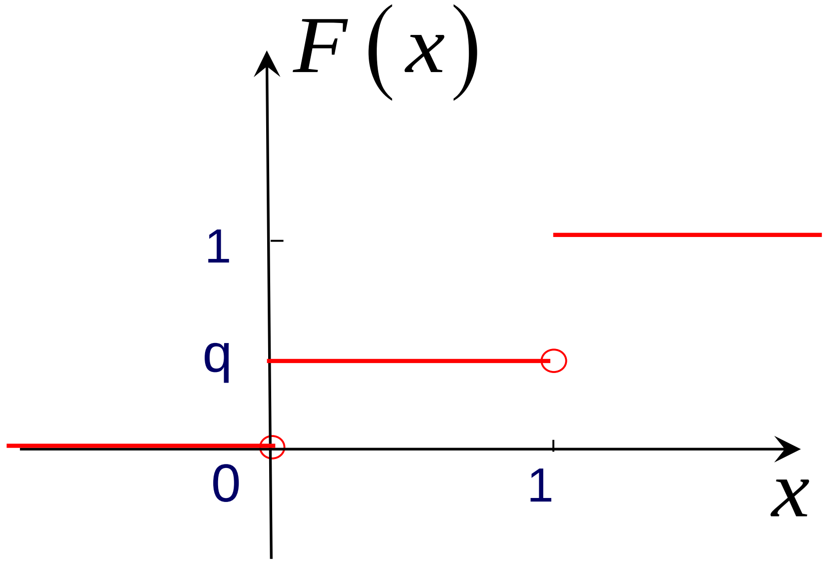
例3.1设 $X \sim B(1,p), 0<p<1, q=1-p,$

X	0	1
P	q	p

求 X 的分布函数 $F(x)$.

解:

$$F(x)=P\{X\le x\}=\begin{cases}0 & x<0\\ q & 0\le x<1\\ 1 & x\ge 1\end{cases}$$



一般地，设离散型随机变量**X**的分布律为

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

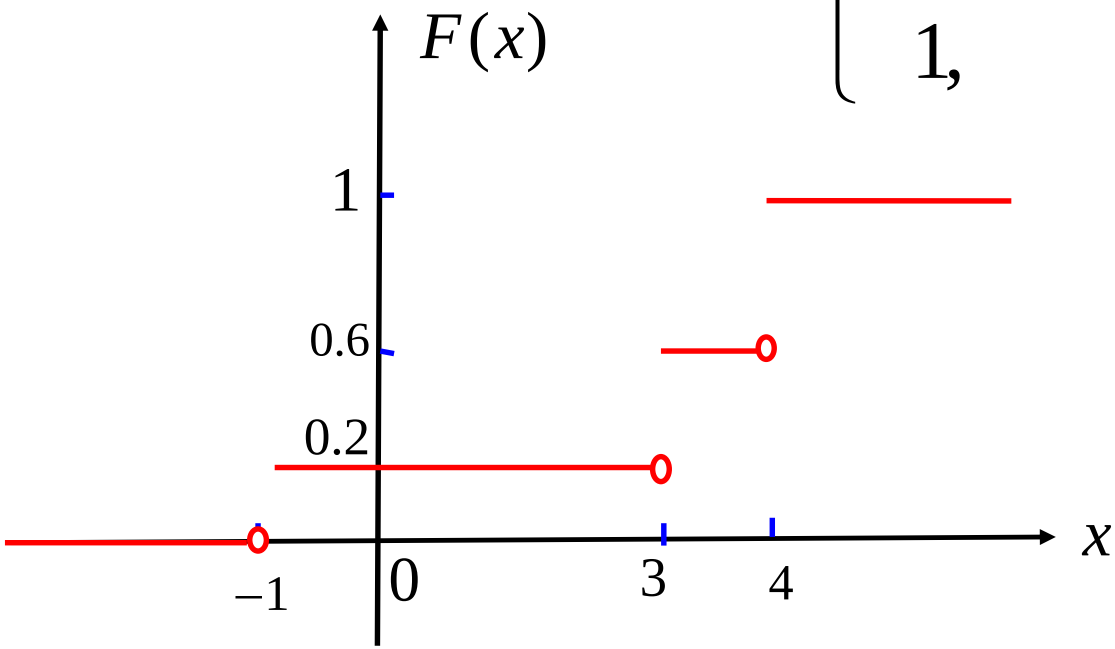
由概率的可列可加性得**X**的分布函数为

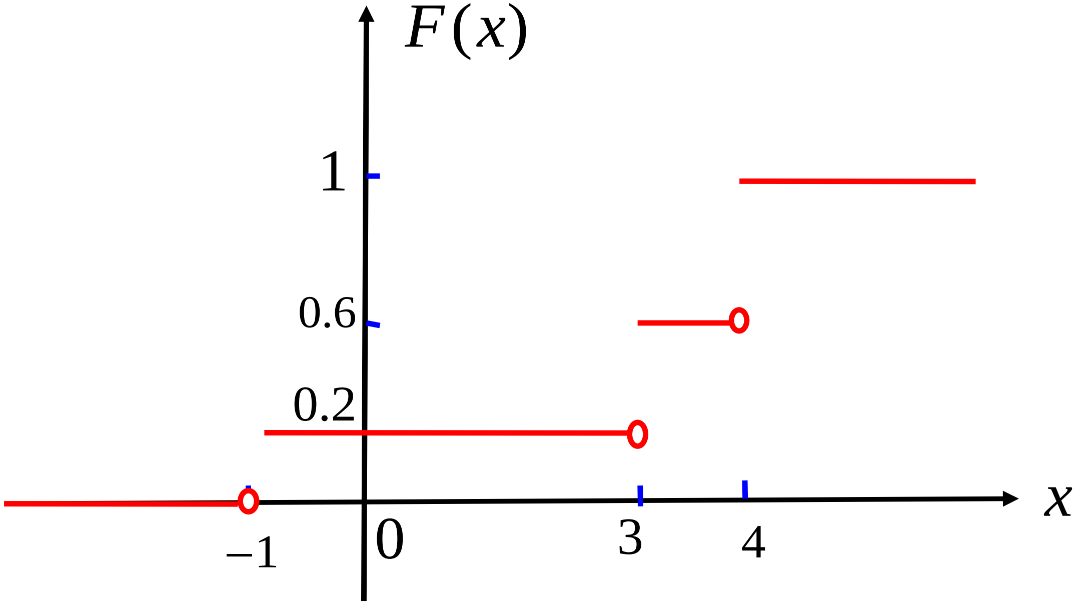
$$F(x) = \sum_{x_k \le x} p_k$$

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3.2 设随机变量X的分布函数如下，求X的分布律.

$$F(x)=\begin{cases}0, & x<-1,\\ 0.2, & -1\le x<3,\\ 0.6, & 3\le x<4,\\ 1, & x\ge 4.\end{cases}$$





解：$F(x)$只在$-1,3,4$有跳，跳的幅度分别是

0.2，0.4，0.4. $\therefore$分布律为

X	-1	3	4
p	0.2	0.4	0.4

例3. 3 设一物体在A,B两点间移动，A,B之间距离3个单位。该物体落在A,B间任一子区间的概率与区间长度成正比。设它离A点的距离为X，求X的分布函数。

解：根据题意，$P(0 \le X \le 3) = 1,$

$$P(0 \le X \le 3) = 3k = 1, \quad \Rightarrow k = \frac{1}{3},$$

当$x < 0$时， $F(x) = P(X \le x) = 0,$

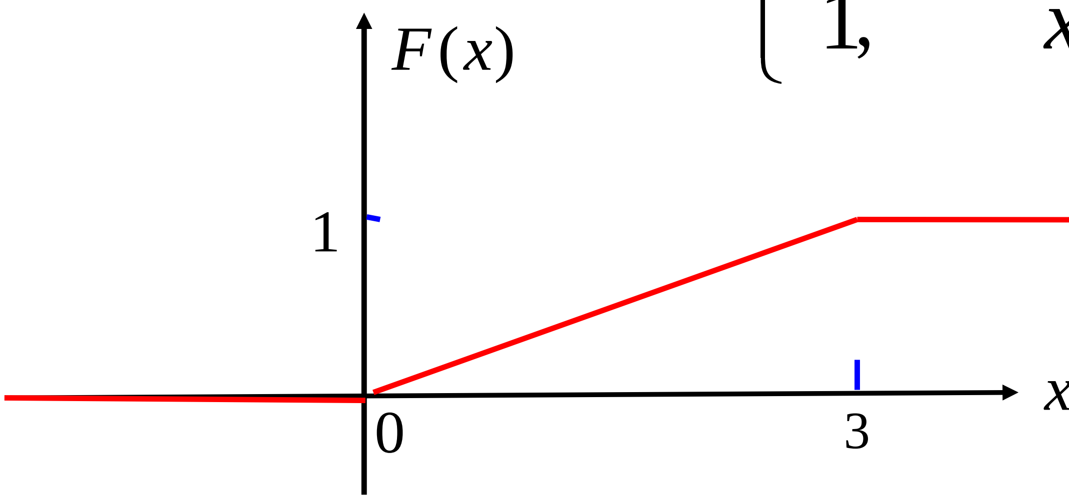
当$x \ge 3$时， $F(x) = P(X \le x) = 1,$

当$0 \le x < 3$时， $F(x) = P(X \le x)$

$$= P(X < 0) + P(0 \le X \le x) = \frac{x}{3}.$$

X的分布函数为

$$F(x)=P(X\le x)=\begin{cases}0, & x<0\\ \dfrac{x}{3}, & 0\le x<3.\\ 1, & x\ge 3\end{cases}$$



与离散型随机变量的分布函数不同！

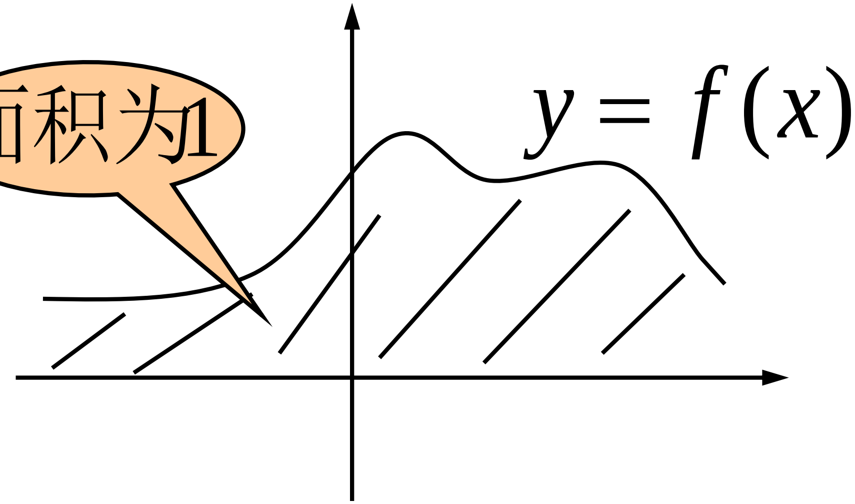
2.4 连续型随机变量及其概率密度函数

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**密度函数**.

$f(x)$的性质：



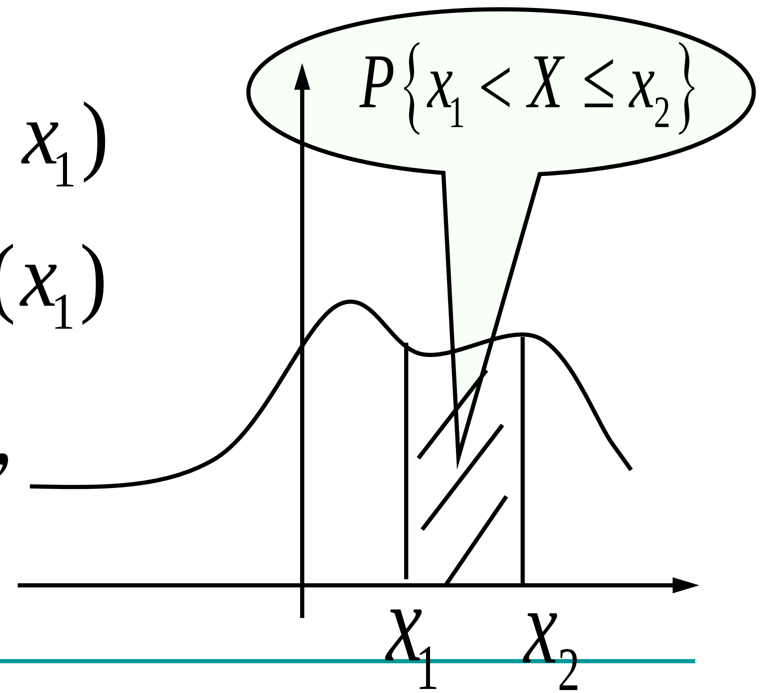
1) $f(x) \geq 0,$

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

3) 对于任意的实数x_1，$x_2(x_2 > x_1)$

$$P\{x_1 < X \leq x_2\} = F(x_2) - F(x_1)$$

$$= \int_{x_1}^{x_2} f(t)\, dt,$$



$\Rightarrow P(X = a) = 0.$

$f(x)$的性质：

4) 在$f(x)$连续点x, $F'(x) = f(x)$

即在$f(x)$的连续点

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

与物理学中的质量线密度的定义相类似

$$P(x < X \le x + \Delta x) \approx f(x) \cdot \Delta x$$

这表示X落在点x附近（$x, x+\Delta x$]的概率近似等于$f(x)\Delta x$

思考题：

设A，B为随机事件，

若$P(A)=1$，则A为必然事件吗？

若$P(B)=0$，则B为不可能事件吗？

若$P(AB)=0$，则A与B不相容吗？

答：都不一定。例如：

$$X \in [0,1],\ f(x)=\begin{cases} 1, x\in[0,1], \\ 0, \text{其他}. \end{cases} A=\{0<X<1\},$$

$B=\{X=0.5\}$, 则$P(A)=1$，$P(B)=0$，$P(AB)=0$.

例4.1 设X的密度函数为$f(x)=\begin{cases} c, & 0<x<1, \\ 2/9, & 3<x<6, \\ 0, & \text{其他}. \end{cases}$

(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。

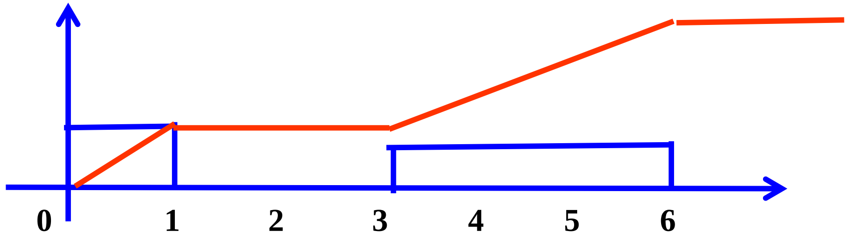
解：(1) $1=\int_{-\infty}^{+\infty} f(t)dt$

$$=c\int_0^1 dt+\frac{2}{9}\int_3^6 dt=\frac{2}{3}+c \Rightarrow c=\frac{1}{3}$$

$$(2)\quad F(x)=P\{X\le x\}=\int_{-\infty}^{x} f(t)dt$$

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

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



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

$$(3)\quad \text{使 } P(X<k)=\frac{2}{3},\ \text{即} F(k)=\frac{2}{3},$$

$$\Rightarrow 3<k<6,\ \therefore \frac{2k-3}{9}=\frac{2}{3},\ k=4.5.$$

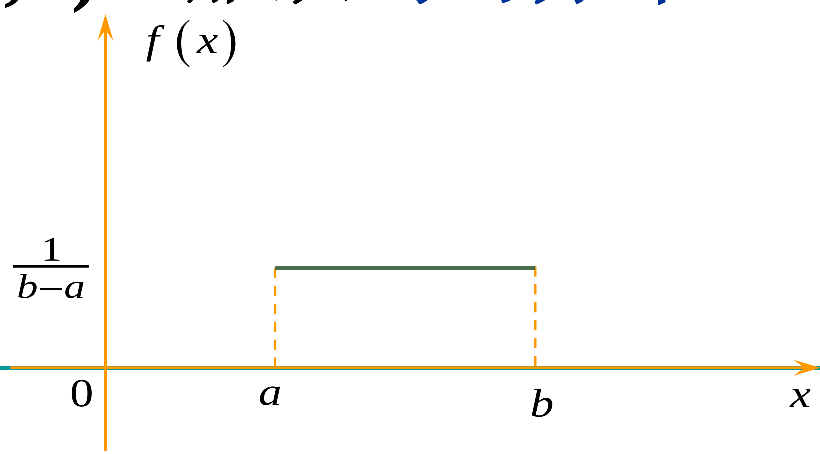
几个重要的连续型随机变量分布

一、均匀分布

定义：设随机变量*X*具有概率密度函数

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

称*X*在区间(*a*,*b*)上服从均匀分布，记为*X*～*U*(*a*,*b*).

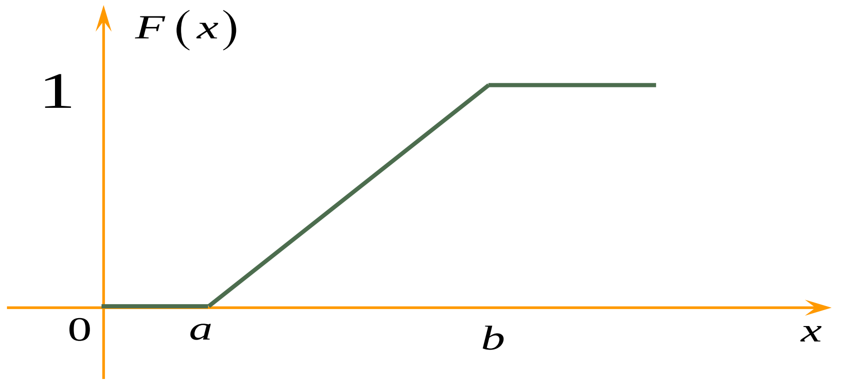


性质：设 $a \le c < c+l \le b$

$$\Rightarrow P(c < X < c+l) = \int_c^{c+l} \frac{1}{b-a} dt = \frac{l}{b-a} \quad \text{--与}c\text{无关}$$

分布函数

$$F(x) = \begin{cases} 0, & x \le a, \\ \dfrac{x-a}{b-a}, & a < x < b, \\ 1, & x \ge b. \end{cases}$$



例4.2 (1)在区间(-1,2)上随机取一数X，试写出X的概率密度函数。并求$P(X > 0)$的值；

(2)若在该区间上独立随机取10个数，求10个数中恰有两个数大于0的概率。

解：(1) 根据题意，X在区间$(-1,2)$上均匀分布

$$\Rightarrow f(x)=\begin{cases}\frac{1}{3}, & -1<x<2\\ 0, & \text{其他}\end{cases} \longrightarrow P(X>0)=\frac{2}{3},$$

（2）设10个数中有Y个数大于0， 则：

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

例4.3 杭州某长途汽车站每天从早上6点（第一班车）开始，每隔30分钟有一班车开往上海。王先生在早上6:20过*X*分钟到达车站，设*X*服从（0，50）上的均匀分布，

（1）求王先生候车时间不超过15分钟的概率；

（2）如果王先生一月中有两次按此方式独立地去候车，求他一次候车不超过15分钟，另一次候车大于10分钟的概率。

解： (1)P(候车时间不超过15钟)=25/50=0.5,

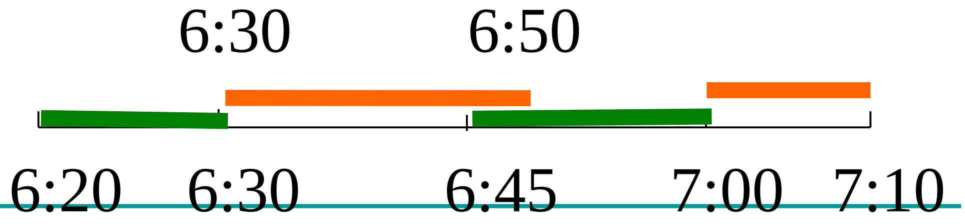
(2) P(候车时间大于10分钟)=30/50=3/5,

P(一次候车时间不超过15分钟，另一次大于10分钟) $= P((X_1 < 15, X_2 > 10) \cup (X_1 > 10, X_2 < 15))$

$$= P(X_1 < 15, X_2 > 10) + P(X_1 > 10, X_2 < 15)$$

$$-P(10 < X_1 < 15, 10 < X_2 < 15)$$

$$= 0.5 \times 3/5 + 3/5 \times 0.5 - 0.1 \times 0.1 = 0.59.$$



二、正态分布

定义：设X的概率密度函数为

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

其中 $-\infty<\mu<\infty,\sigma>0$ 为常数，称X服从参数为 μ,σ 的正态分布(Gauss分布)，

记为 $X\sim N(\mu,\sigma^2)$

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

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

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

记 $I=\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt$，只需证明 $I=\sqrt{2\pi}$.

$$\because I^2=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-\frac{(x^2+y^2)}{2}} dxdy$$

$$=\int_0^{2\pi} d\theta \int_0^{+\infty} re^{-\frac{r^2}{2}} dr = 2\pi \times 1 = 2\pi$$

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

性质：若$X \sim N(\mu,\sigma^2)$

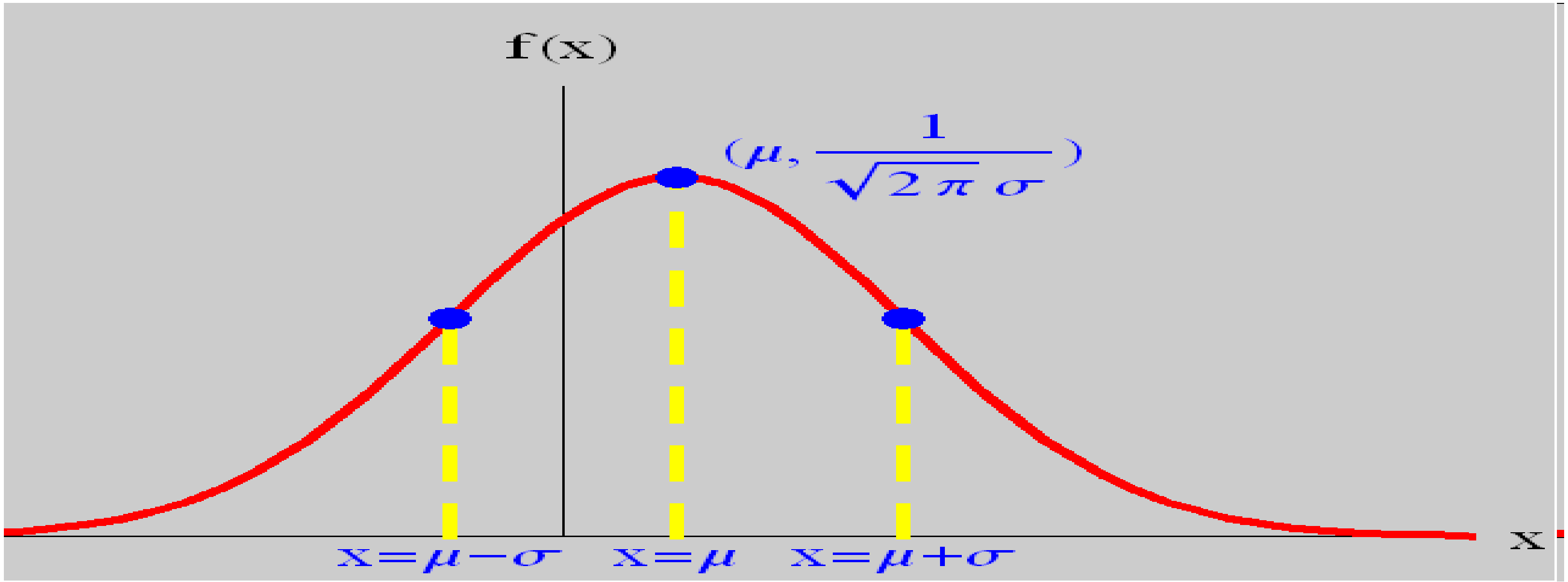
$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}，\quad -\infty < x < +\infty$$

则 1° $f(x)$关于$x=\mu$对称

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

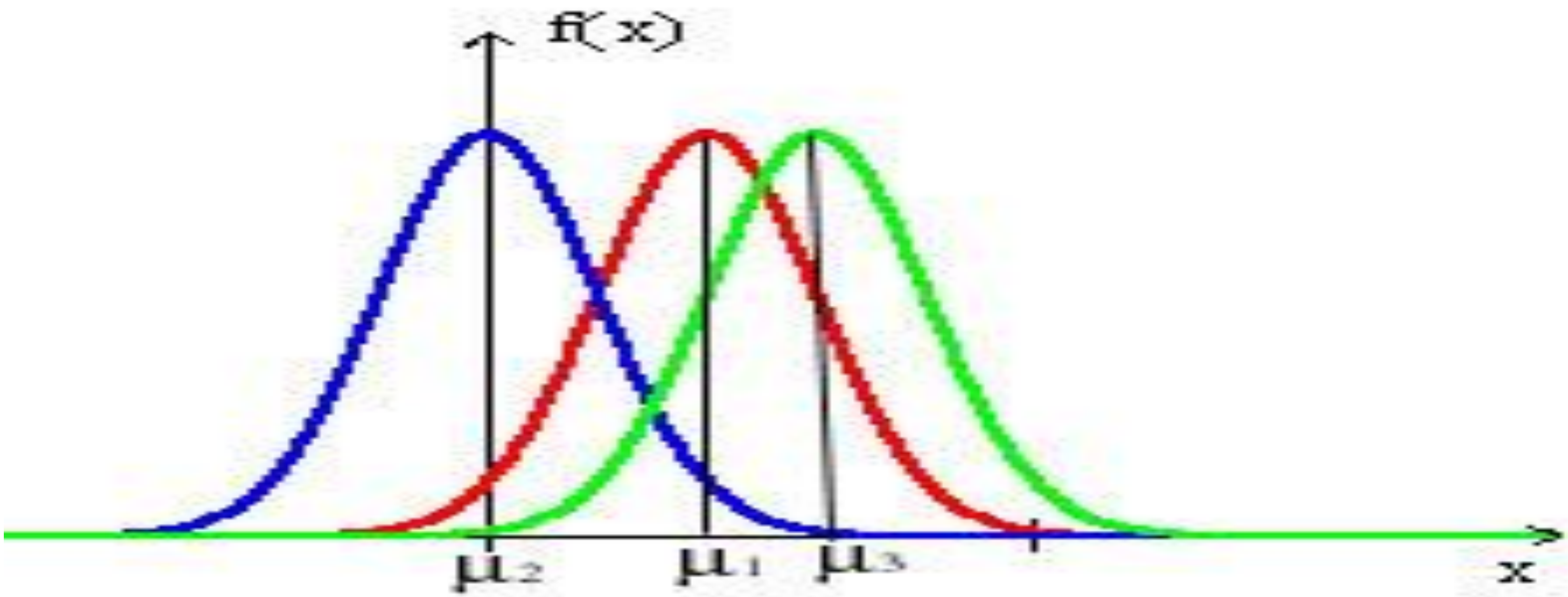
3° $\lim\limits_{|x-\mu|\to\infty} f(x) = 0$

正态概率密度函数

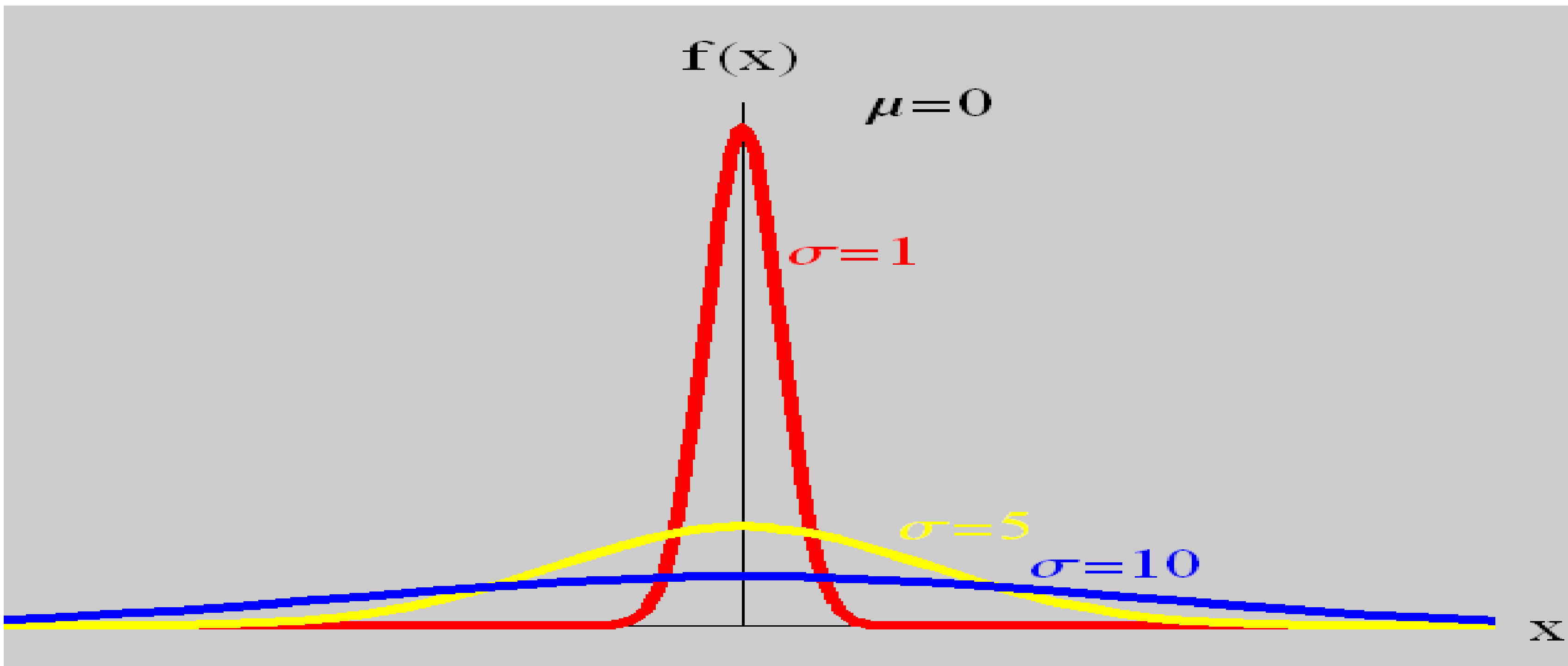


$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\quad -\infty<x<+\infty$$

当固定σ，改变μ的大小时，密度函数图形形状不变，只是沿着x轴作平移变换；



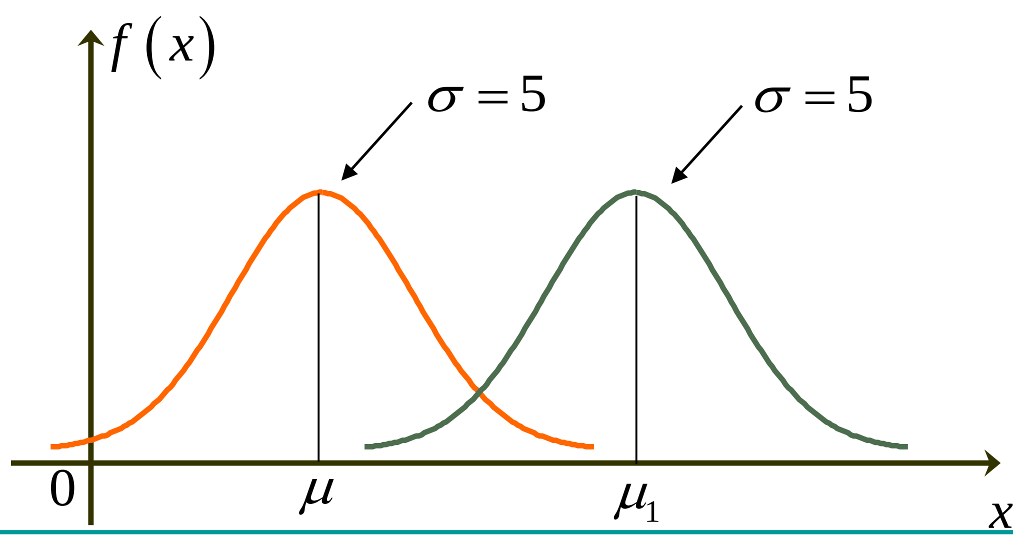
当固定μ，改变σ的大小时，密度函数图形的对称轴不变，形状发生改变，σ越小，图形越高瘦，σ越大，图形越矮胖.



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

称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)



- 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，即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。

正态分布下的概率计算

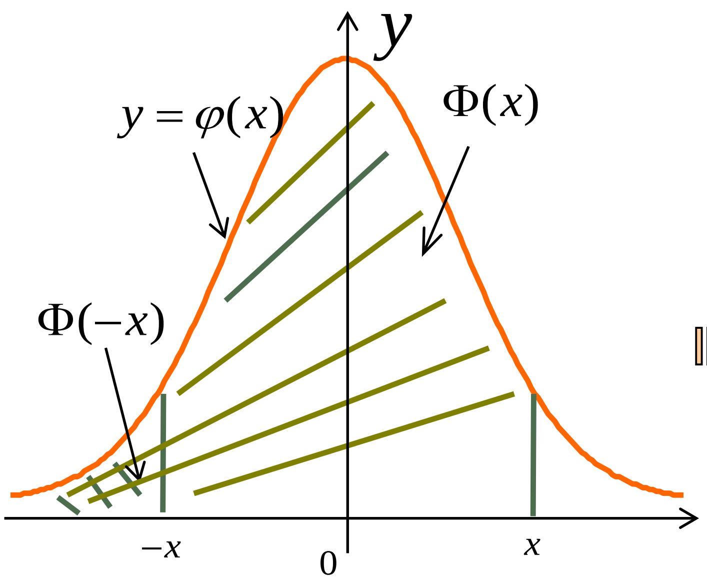
$$P\{X \le x\} = F(x) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,\mathrm{d}t$$

$$= ?$$

若$Z \sim N(0, 1)$，称 Z服从标准正态分布．

Z 的密度函数为：$\varphi(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$，

Z的分布函数为：$\Phi(x) = \displaystyle\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$



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

✦ 当 $X \sim N(\mu,\sigma^2)$ 时

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

（作变换：$\frac{x-\mu}{\sigma} = t$）

$$= \int_{-\infty}^{\frac{b-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

$$\boxed{\Rightarrow P(X \le b) = \Phi(\frac{b-\mu}{\sigma})}$$

当 $X \sim N(\mu,\sigma^2)$ 时 $\frac{X-\mu}{\sigma} \sim N(0,1)$

例4.4 $X \sim N(\mu, \sigma^2)$

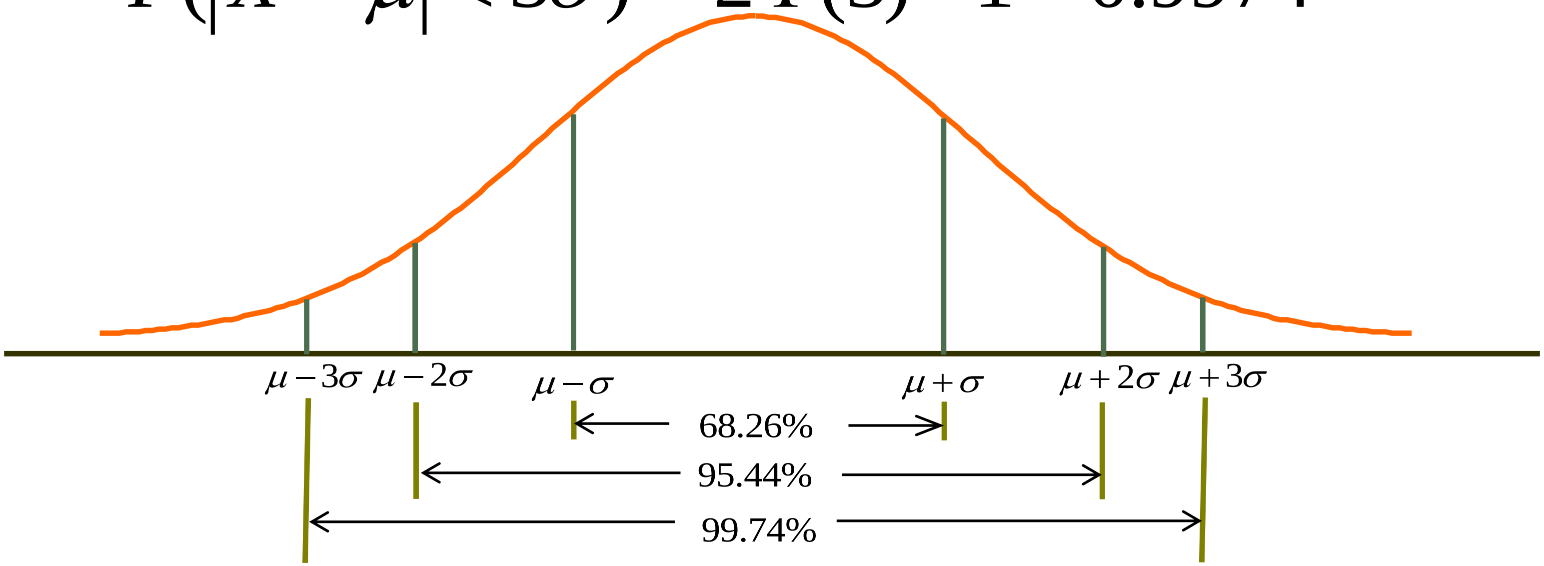
$$P(|X-\mu|<\sigma) = P(\mu-\sigma < X < \mu+\sigma)$$

$$= P\left(\frac{(\mu-\sigma)-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{(\mu+\sigma)-\mu}{\sigma}\right)$$

$$= \Phi(1)-\Phi(-1) = 2\Phi(1)-1 = 0.6826$$

$$P(|X-\mu|<2\sigma) = 2\Phi(2)-1 = 0.9544$$

$$P(|X-\mu|<3\sigma) = 2\Phi(3)-1 = 0.9974$$



例4.5 用天平称一实际重量为a的物体，天平的读数为随机变量X，设$X \sim N(a,\ 0.01^2)$时，

（1）求读数与a的误差小于0.005的概率；

（2）求读数至少比a多0.0085的概率。

解:(1) $P(|X-a|<0.005)$

$$=\Phi(\frac{0.005}{0.01})-\Phi(-\frac{0.005}{0.01})$$

$$=2\Phi(0.5)-1$$

$$\overset{\text{查附表}}{=\!=\!=}2\times 0.6915-1=0.3830$$

(2) $P(X-a\geq 0.0085)=1-\Phi(0.85)$

$$=1-0.8023=0.1977.$$

注：计算$P(X-a<0.0085)$使用$Excel$表单：

在$Excel$表单的任一单元格输入

“$=NORM.DIST(0.0085, 0, 0.01, 1)$”

点击“确定”,即在单元格中出现"0.802337508".

例4.6 一批钢材(线材)长度$X(cm) \sim N(\mu, \sigma^2)$

(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) $P(X<97.8)=P(\frac{X-100}{2}<\frac{97.8-100}{2})$

$$=\Phi(\frac{97.8-100}{2})=1-\Phi(1.1)$$

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

(2) 需: $P\{97<X<103\}\geq 90\%$

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

$$\Rightarrow \Phi(\frac{3}{\sigma})\geq 0.95 \quad \Rightarrow \frac{3}{\sigma}\geq 1.645 \quad \Rightarrow \sigma\leq 1.8237$$

例4.7 设一天中经过一高速公路某一入口的重型车辆数*X*近似服从 $N(\mu,\sigma^2)$，已知有25%的天数超过400辆，有33%的天数不到350辆，求

$$\mu,\sigma.$$

解： 已知 $P(X > 400) = 0.25,\ P(X < 350) = 0.33$

而 $P(X > 400) = 1 - \Phi(\frac{400 - \mu}{\sigma}) = \Phi(\frac{\mu - 400}{\sigma}),$

$P(X < 350) = \Phi(\frac{350 - \mu}{\sigma}),$

$\Phi(-0.675) = 0.25, \quad \Phi(-0.440) = 0.33,$

于是 $\begin{cases} \dfrac{\mu - 400}{\sigma} = -0.675 \\ \dfrac{350 - \mu}{\sigma} = -0.440 \end{cases} \Rightarrow \begin{cases} \mu \approx 369.7 \\ \sigma \approx 44.8 \end{cases}$

三、指数分布

定义：设**X**的密度函数为 $f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0,\\ 0, & x\le 0.\end{cases}$

其中**λ>0**为常数，则称**X**服从参数为**λ**的指数分布。记为 $X\sim E(\lambda)$或$X\sim Exp(\lambda)$.

X的分布函数为：

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

X具有如下的无记忆性：$t_0 > 0, t > 0,$

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

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

如果X表示等待时间，那么无记忆性说明只要还没等到，那么剩余等待时间仍然服从参数为λ的指数分布.

如果X表示元件寿命，那么无记忆性说明只要还没坏掉，那么剩余寿命仍然服从参数为λ的指数分布.

例4.8 某大型设备在任何长度为t的时间区间内发生故障的次数$N(t)$服从参数为λt的泊松分布.记设备无故障运行的时间为T,

(1)求T的概率分布函数;

(2)已知设备无故障运行了10个小时，求该设备再无故障至少运行8个小时的概率.

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

$$F_T(t)=P\{T\le t\}=1-P\{T>t\}$$

当$t<0$时，$F_T(t)=0$

当$t\ge 0$时，$F_T(t)=1-P\{N(t)=0\}=1-e^{-\lambda t}$

(2) $P\{T\ge 18 \mid T>10\}=\dfrac{P\{T>18\}}{P\{T>10\}}=e^{-8\lambda}=P\{T>8\}$

例4.9　一银行服务需要等待，设等待时间*X*(分钟)的概率密度函数为

$$f(x)=\begin{cases}\dfrac{1}{10}e^{\frac{-x}{10}}, & x>0\\ 0, & x\le 0\end{cases}$$

某人进了银行，且打算过会儿去办另一件事，于是先等待，如果超过15分钟还没有等到服务就离开，设他实际的等待时间为*Y*，(1)求*Y*的分布函数;(2)问*Y*是离散型随机变量吗？连续型随机变量吗？

解： $(1) 0 \le Y \le 15, F(y) = P(Y \le y)$

当$y < 0$时, $F(y) = P(Y \le y) = 0,$

当$y \ge 15$时, $F(y) = P(Y \le y) = 1,$

当$0 \le y < 15$时, $F(y) = P(Y \le y)$

$$= P(X \le y) = 1 - e^{\frac{-y}{10}}.$$

(2)*Y*的取值范围为[0, 15],故不是离散型随机变量；

$$P\{Y = 15\} = P(Y \le 15) - P(Y < 15)$$

$$= F(15) - F(15-0) = e^{-1.5} > 0,$$

因此*Y*也不是连续型随机变量；*Y*既没有分布律，也没有概率密度函数.

2.5　随机变量函数的分布

例如，若要测量一个圆的面积Y，可以先测量其周长，如果周长的测量值X是随机变量，有一个概率分布，那么Y服从什么分布？

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5.1 已知X具有分布律

X	-1	0	1
p	0.2	0.5	0.3

且设$Y=X^2$，求Y的概率分布。

解：***Y***的所有可能取值为0, 1

$$P(Y=0)=P(X=0)=0.5$$

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

$$=P(X=1)+P(X=-1)=0.5$$

即找出(***Y***=0)的等价事件(***X***=0)；

(***Y***=1)的等价事件(***X***=1)与(***X***=−1)的和事件.

例5.2 设随机变量X具有概率密度函数

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

求 $Y=X^2$ 的概率密度函数。

解：记*Y*的分布函数为$F_Y(y)$

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

当 $y \le 0$时，$F_Y(y) = 0;$

当 $y \ge 16$时，$F_Y(y) = 1$

当 $0<y<16$ 时,

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

$$=P\left\{-\sqrt{y}\le X\le 0\right\}+P\left\{0<X\le\sqrt{y}\right\}$$

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

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

Y在区间(0, 16)上均匀分布。

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为离散型随机变量，则先写出Y的可能取值：$y_1, y_2, \cdots y_j, \cdots$, 再找出$\left(Y = y_j\right)$的等价事件$(X \in D_j)$, 得$P(Y = y_i) = P(X \in D_j)$;

2. 若Y为连续型随机变量，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$(X \in D_y)$, 得$F_Y(y) = P(X \in D_y)$；再求出Y的概率密度函数$f_Y(y)$.

关键是找出等价事件。

例5.3 设随机变量X的分布律如下表

X	-1	0	1	2
p	$1/3$	$1/6$	$1/6$	$1/3$

$\boldsymbol{Y=2X+1}$, $\boldsymbol{Z=X^2}$, 求$\boldsymbol{Y,Z}$的概率分布律.

解：Y的可能取值为−1, 1, 3, 5,

Z的可能取值为0, 1, 4,

$(Y=-1)$的等价事件为$(X=-1)$…

$(Z=1)$的等价事件为$(X=1)\cup(X=-1)$

故得：

Y	-1	1	3	5
p	$1/3$	$1/6$	$1/6$	$1/3$

Z	0	1	4
p	$1/6$	$1/2$	$1/3$

例5.4： 设随机变量X具有密度函数

$f_X(x), -\infty < x < +\infty,$

分别求$Y = |X|,\ Z = X^2$的

概率密度函数$f_Y(y),\ f_Z(z).$

解：分别记X, Y, Z的分布函数为$F_X(x), F_Y(y), F_Z(z)$.

当$y \le 0$时，$F_Y(y) = 0$. 当$y > 0$时，

$$F_Y(y) = P\{Y \le y\} = P\{|X| \le y\} = F_X(y) - F_X(-y).$$

$$f_Y(y) = \begin{cases} f_X(y) + f_X(-y), & y > 0 \\ 0, & y \le 0 \end{cases}$$

同理，当 $z \le 0$ 时, $F_Z(z) = 0$. 当 $z > 0$ 时,

$$F_Z(z) = P\{Z \le z\} = P\{X^2 \le z\}$$

$$= P\{-\sqrt{z} \le X \le \sqrt{z}\} = F_X(\sqrt{z}) - F_X(-\sqrt{z})$$

$$f_Z(z) = \begin{cases} \dfrac{1}{2\sqrt{z}}[f_X(\sqrt{z}) + f_X(-\sqrt{z})], & z > 0 \\ 0, & z \le 0. \end{cases}$$

例5.5 设 $X \sim U(-1, 2)$，求

$Y = |X|$ 的概率密度函数 $f_Y(y)$.

解：X的密度函数为 $f_X(x)=\begin{cases}1/3, & -1<x<2\\ 0, & \text{其它}\end{cases}$

由例5.4得$Y=|X|$的概率密度函数为

$$f_Y(y)=\begin{cases}f_X(y)+f_X(-y), & y>0\\ 0, & y\le 0\end{cases}$$

$$=\begin{cases}\dfrac{1}{3}+\dfrac{1}{3}, & 0<y<1\\ \dfrac{1}{3}+0, & 1<y<2\\ 0, & \text{其它}\end{cases}=\begin{cases}\dfrac{2}{3}, & 0<y<1\\ \dfrac{1}{3}, & 1<y<2\\ 0, & \text{其它}\end{cases}$$

例5.6 设X~N(0, 1)，求

$Z = X^2$ 的概率密度函数 $f_Z(z)$.

解：X的密度函数为 $f_X(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},-\infty<x<+\infty,$

由例5.4得$Z=X^2$的概率密度函数为

$$f_Z(z)=\begin{cases}\frac{1}{2\sqrt{z}}[f_X(\sqrt{z})+f_X(-\sqrt{z})], & z>0\\ 0\qquad\qquad\qquad\qquad\qquad\qquad, & z\le 0.\end{cases}$$

$$=\begin{cases}\frac{1}{\sqrt{2\pi}}z^{-\frac{1}{2}}e^{-\frac{z}{2}}, & z>0\\ 0\qquad\qquad\qquad, & z\le 0.\end{cases}$$

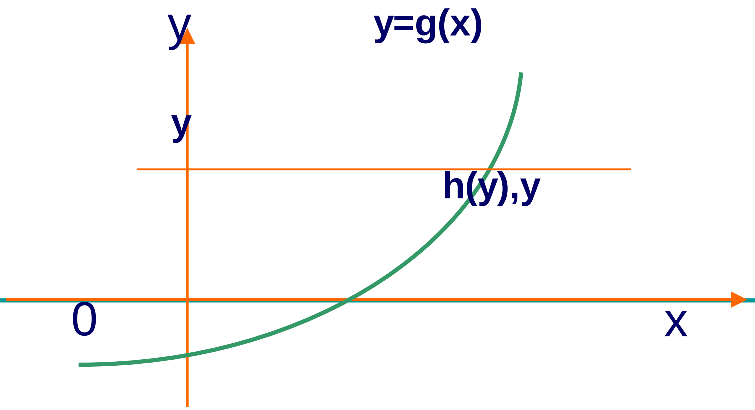
此 时 称 Z 服 从 自 由 度 为 1 的 χ^2分 布 。

定理2.5.1：设$X \sim f_X(x), -\infty < x < +\infty$，$y = g(x)$ 是严增(减)可微函数，即$g'(x) > 0$ (或$g'(x) < 0$)。$Y = g(X)$, 则Y具有概率密度函数为：

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

这里(α, β)是Y的取值范围，h是g的反函数

$$h(y) = x \Leftrightarrow y = g(x)$$



证明：不妨设$g'(x)>0$，则$g(x)$为单调增函数，

且：$h'(y)>0$

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

当 $\alpha < y < \beta$ 时，

$$F_Y(y)=P(Y \le y)=P(g(X) \le y)$$

$$=P(X \le h(y))=F_X(h(y))$$

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

同理可证：当$g'(x)<0$时，定理为真

例5.7 设$X \sim N(\mu,\sigma^2)$, $Y = aX + b(a \neq 0)$,

求Y的概率密度函数$f_Y(y)$.

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

$$x = h(y) = \frac{y-b}{a}$$

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

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

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

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

一般若$X \sim N(\mu, \sigma^2)$,

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

例5.8 设 $X \sim U(-\pi/2, \pi/2), Y = \sin X$,

求Y的概率密度函数$f_Y(y)$.

解：$Y=\sin X$对应的函数$y=g(x)$在$(-\pi/2,\pi/2)$

上恒有$g'(x)=\cos x>0$，且有反函数

$$x=h(y)=\arcsin y,\ h'(y)=1/\sqrt{1-y^2}$$

X的概率密度函数为

$$f_X(x)=\begin{cases}\dfrac{1}{\pi}, & -\dfrac{\pi}{2}<x<\dfrac{\pi}{2}\\ 0\ , & \text{其它}\end{cases}$$

由定理得$Y=\sin X$的密度函数为

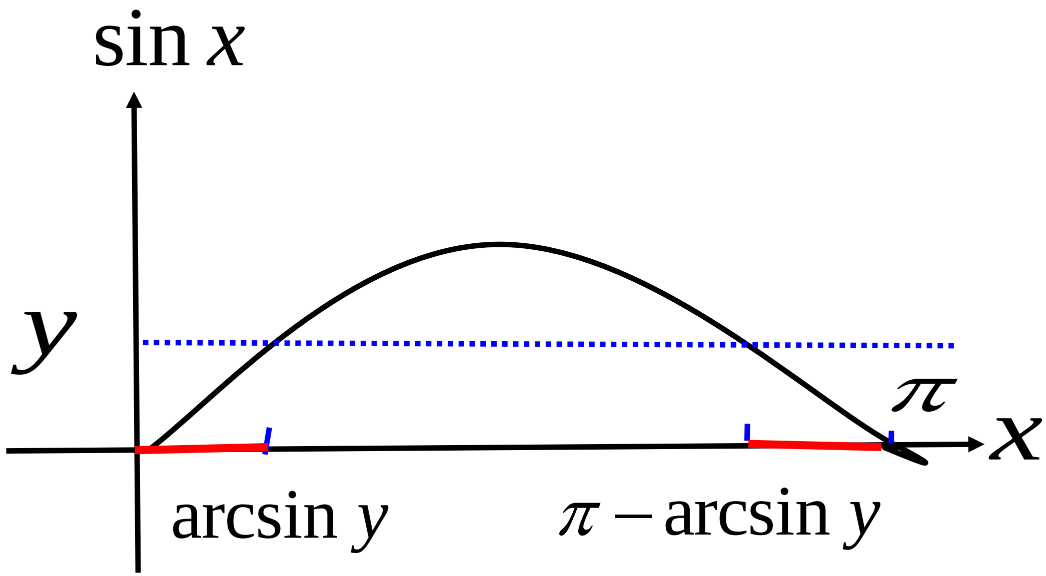
$$f_Y(y)=f_X(h(y))\cdot\left|h'(y)\right|=f_X(h(y))h'(y)$$

$$=\begin{cases}\dfrac{1}{\pi}\cdot\dfrac{1}{\sqrt{1-y^2}}, & -1<y<1\\ 0\ , & \text{其它}\end{cases}$$

例5.9　$Y=\sin X, X\sim U(0,\pi)$，求$f_Y(y)$.

解：$Y=\sin X$在（0，π）不单调，所以不能应用定理。对$0<y\le 1$,

$$F_Y(y)=P(Y\le y)=P\{\sin X\le y)=\frac{2\arcsin y}{\pi}$$



所以Y的概率密度函数为：

$$f_Y(y)=\begin{cases}\dfrac{2}{\pi\sqrt{1-y^2}}, & 0<y<1,\\ 0, & \text{其他}.\end{cases}$$

例5.10　设$X \sim E(\lambda)$，$F\left(x\right)$为X的分布函数,

记$Y = F\left(X\right)$,试证$Y \sim U\left(0,1\right)$(即均匀分布).

解： 由$X \sim E(\lambda)$,∴分布函数$F(x)=\begin{cases}1-e^{-\lambda x}, x>0, \\ 0, \quad x\le 0.\end{cases}$

当$x>0$时, $y=F(x)=1-e^{-\lambda x}\in(0,\ 1)$单调增,

反函数$x=F^{-1}(y)=-\dfrac{1}{\lambda}\ln(1-y)$,

$$\frac{\mathrm{d}}{dy}F^{-1}(y)=\frac{1}{\lambda(1-y)},$$

由定理2.5.1， 当$0<y<1$时,

$$f_Y(y)=\lambda\exp\{-\lambda[-\frac{1}{\lambda}\ln(1-y)]\}\frac{1}{\lambda(1-y)}=1,$$

∴ $Y\sim U(0,1)$. **更一般的结果见书中例2.5.6.**