西安電子科技大學
XIDIAN UNIVERSITY

概率论与数理统计



概率论与数理统计课程组

2.2 离散型随机变量

1. 离散型随机变量的定义

设 X 是随机变量，如果其可能的取值为有限个或可列无限个，则称 X 为离散型随机变量。

要掌握一个离散型随机变量 X 的统计规律，必须且只需知道 X 所有可能的取值以及取每一个值的概率。

2. 离散型随机变量的分布律

设 X 是离散型随机变量，其可能的取值为 x_1，x_2，…，x_i，…，称 $P(X=x_i)=p_i$，$i=1,2,\cdots$

为 X 的分布律，或表示为：

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

3. 离散型随机变量分布律的性质

性质 1 $p_k \geq 0 \quad (k = 1, 2, 3, \ldots)$

性质 2 $\sum_{k=1}^{\infty} p_k = 1$

4.Bernoulli (伯努利)试验

伯努利

Jacob Bernoulli

Born: 27 Dec 1654 in Basel, Switzerland
Died: 16 Aug 1705 in Basel, Switzerland

定义： 若 n 次试验满足:

(1) 每次试验只有两种可能结果：A发生和A不发生；
(2) 各次试验是相互独立的；
(3) 每次试验是重复进行的，即A每次发生的概率一样。

则称此试验为n重Bernoulli试验。

5．几种重要的离散型随机变量

（1）离散均匀分布

定义： 若离散型随机变量 X 的分布律为

$$P(X=k)=\frac{1}{n},\qquad k=1,2,\ldots,n$$

则称 X 服从离散均匀分布

5．几种重要的离散型随机变量

（2）二项分布

事件A在一次试验中发生的概率为P，把试验重复独立做n次，A恰好发生k次的概率？

定义： 若离散型随机变量 X 的分布律为

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

则称 X 服从参数为 n 、 p 的二项分布，记为 $X\sim B(n, p)$。

特别地，当 $n=1$ 时，二项分布变为

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

这时的分布成为参数为 p 的 0-1 分布，可以记为 $X\sim B(1, p)$。

思考:二项分布的分布律 $p_k=P(X=k)$ 是否满足概率的基本性质？

$$1=(p+q)^n=\sum_{k=0}^{n} C_n^k p^k q^{n-k}=\sum_{k=0}^{n} P(X=k)$$

例1: 有一大批零部件，一级品率为20%，取20件检验，求取出的零件中至少有两个一级品的概率。

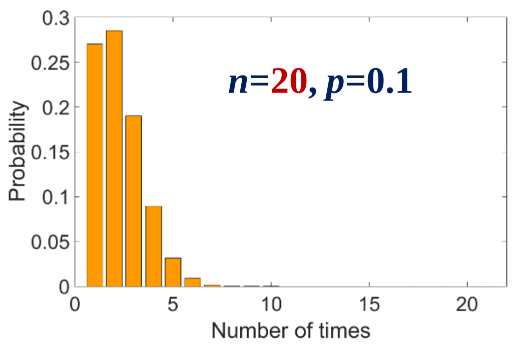
解：设X表示20件抽检产品中的一级品件数

由二项分布定义可知 $X \sim B(20, 0.2)$，求 $P(X \geq 2)$

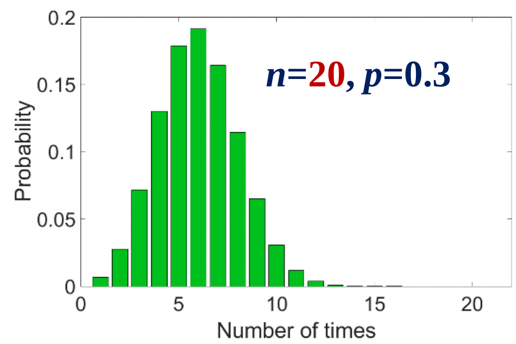
$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - P(X = 0) - P(X = 1)$$

$$= 1 - (1 - 0.2)^{20} - C_{20}^{1} 0.2^{1} (1 - 0.2)^{19}$$

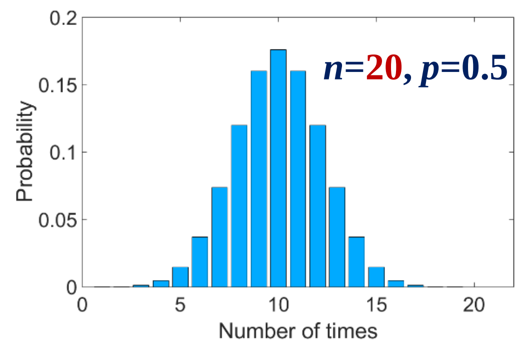
概率计算值



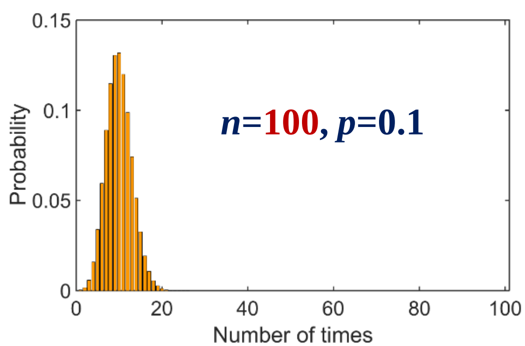
n=20, p=0.1
Probability
Number of times
0 5 10 15 20
0 0.05 0.1 0.15 0.2 0.25 0.3



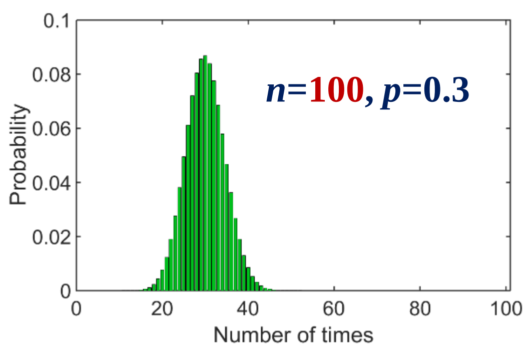
n=20, p=0.3
Probability
Number of times
0 5 10 15 20
0 0.05 0.1 0.15 0.2



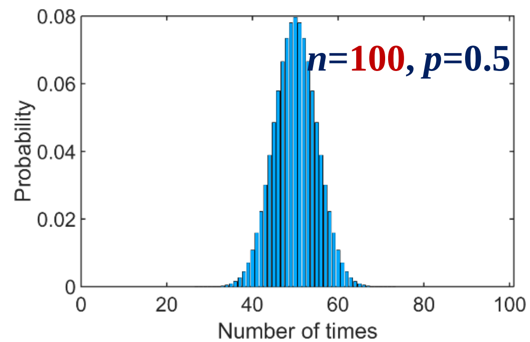
n=20, p=0.5
Probability
Number of times
0 5 10 15 20
0 0.05 0.1 0.15 0.2



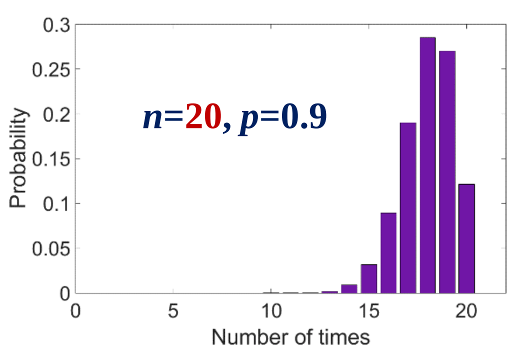
n=100, p=0.1
Probability
Number of times
0 20 40 60 80 100
0 0.05 0.1 0.15



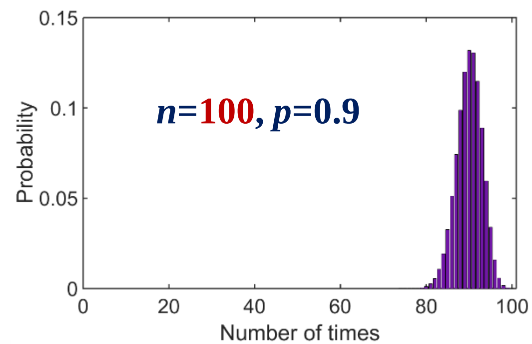
n=100, p=0.3
Probability
Number of times
0 20 40 60 80 100
0 0.02 0.04 0.06 0.08 0.1



n=100, p=0.5
Probability
Number of times
0 20 40 60 80 100
0 0.02 0.04 0.06 0.08



n=20, p=0.9
Probability
Number of times
0 5 10 15 20
0 0.05 0.1 0.15 0.2 0.25 0.3



n=100, p=0.9
Probability
Number of times
0 20 40 60 80 100
0 0.05 0.1 0.15

一个变量服从二项分布有两个条件：

1. 各次试验的独立性；2. 各次试验的条件是稳定的

例2: 双胞胎分别位于两个屋子里，取8张牌，4张红色，4张黑色。哥哥看一张牌，弟弟猜测哥哥看到的牌是什么颜色，一张一张轮流看，直至看完8张。问如果弟弟猜对的数目≥6张，那么二人之间是否有心灵感应？

思路：假定无心灵感应，计算弟弟猜中6张以上的可能性，看是否是小概率事件。

追问：若试验独立重复做10次，猜中6张以上算成功，那么成功多少次才能算有心灵感应？

（2）Poisson分布

定义： 若离散型随机变量 X 的分布律为

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

则称 X 服从参数为 λ 的 Poisson 分布，记为 $X\sim P(\lambda)$。

Poisson 分布是一种应用非常广泛的概率分布，用于描述稀有事件发生的概率。

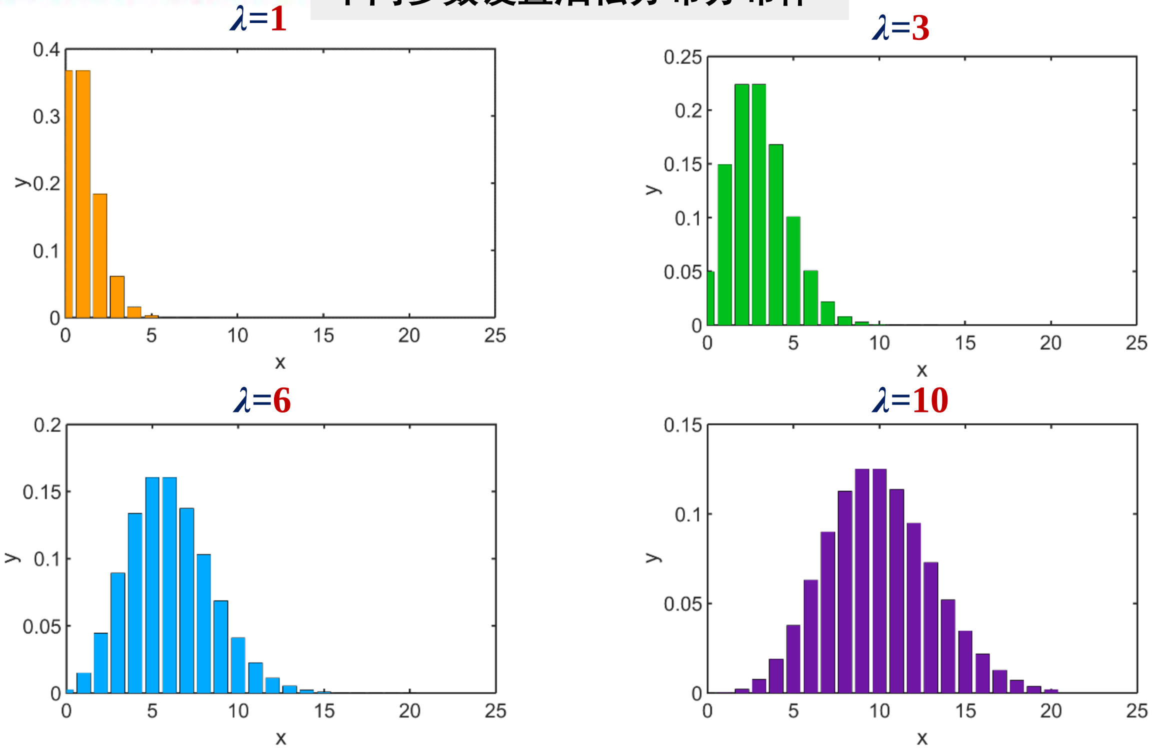
例如，一批产品中出现的次品数，

单位时间内商店售出的某种特殊商品的件数，

一本书一页中的印刷错误数，

某地区某段时间内交通事故次数等。

不同参数设置泊松分布分布律



- **Poisson 分布是二项分布的一种近似。**

泊松(*Poisson*)定理

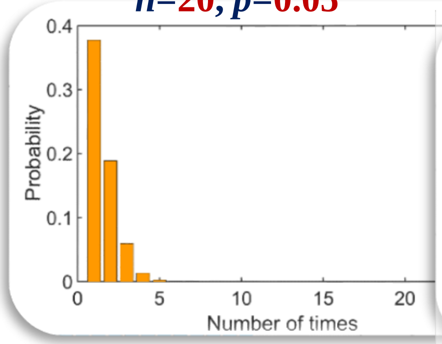
设$\lambda>0$是一常数，n是任意正整数，设$np_n=\lambda$，则对于任一固定的非负整数k，有

$$\lim_{n\to\infty} C_n^k p_n^{\ k}(1-p_n)^{n-k}=\frac{\lambda^k e^{-\lambda}}{k!}$$

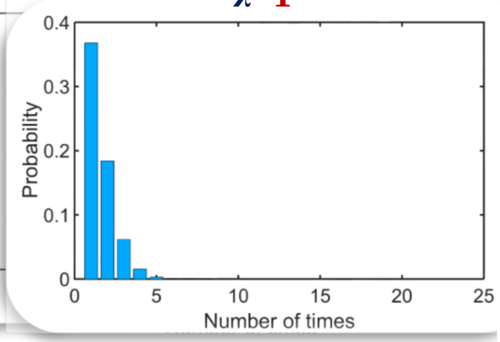
二项分布与泊松分布之间的数值关系

二项分布　分布律比较　泊松分布

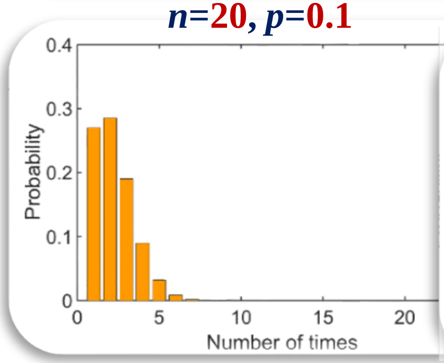
$n=20,\ p=0.05$



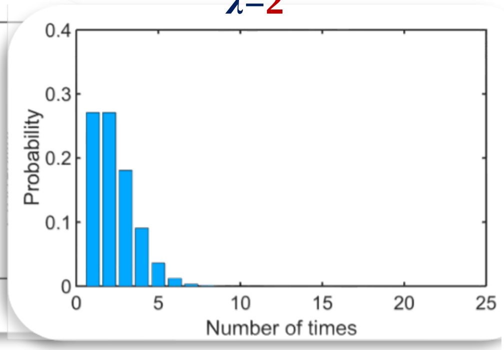
$\lambda=1$



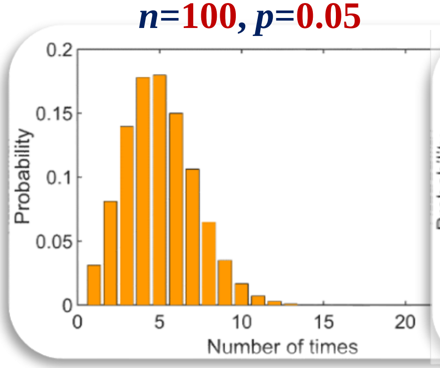
$n=20,\ p=0.1$



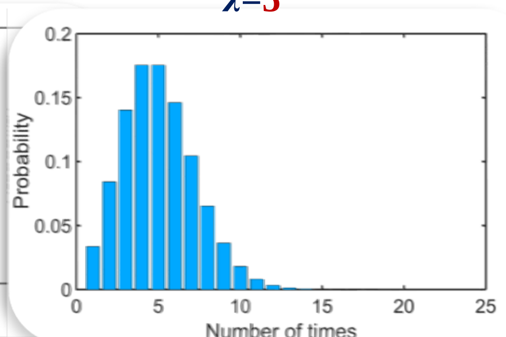
$\lambda=2$



$n=100,\ p=0.05$

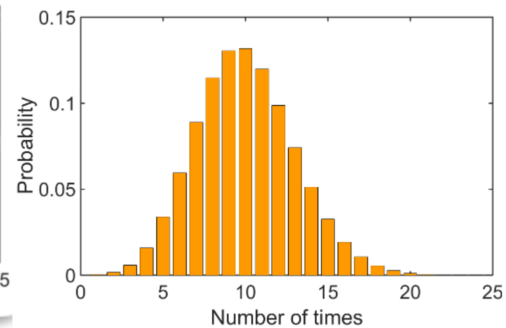


$\lambda=5$

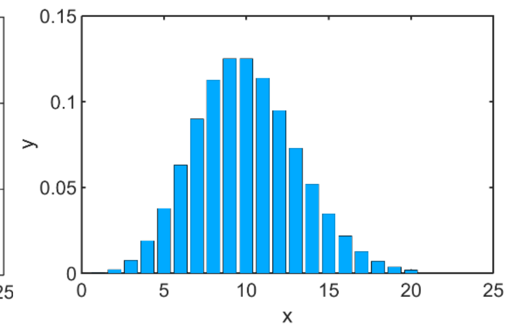


二项分布　　泊松分布

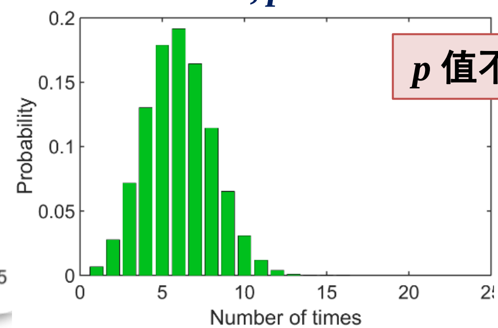
$n=100,\ p=0.1$



$\lambda=10$

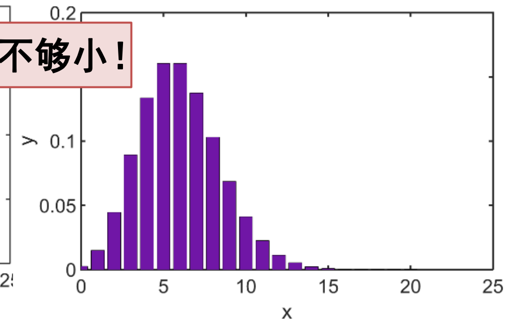


$n=20,\ p=0.3$



p 值不够小！

$\lambda=6$



Poisson定理

二项分布在n很大，p很小时($np=\lambda$) 可以用泊松分布来近似

当 $n>10$，$p<0.1$ 时，

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

在 $n>20$，$p<0.05$ 时近似效果颇佳

例 设某工厂有400台同类机器，各台机器发生故障的概率都是0.02，各台机器工作是相互独立的，试求机器出故障的台数不小于2的概率。

解 设X为机器故障台数，$X \sim B(400, 0.02)$，两种方法求解

(i) 二项分布

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

$$= 1 - 0.98^{400} - 400 \times 0.02 \times 0.98^{399} = 0.9972$$

(ii) 泊松分布近似

$$\lambda = np = 400 \times 0.02 = 8$$

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1) \approx 1 - 0.000335 - 0.002684 \approx 0.9969$$

例:工厂有300台机器，每台机器发生故障的概率为1%，每台机器发生故障时需要一人维修，那么需要配套多少人才能使机器发生故障单得不到维修的概率≤0.01？

（3）几何分布

定义：　若离散型随机变量 X 的分布律为

$$P(X=k)=q^{k-1}p,\quad k=1,2,\cdots,0<p<1,q=1-p$$

则称 X 服从参数为 p 的几何分布。

进行重复独立试验，每次试验成功的概率为P，直到第r次才成功的概率。

若将试验进行到r次成功为止，X将服从Pascal（巴斯卡）分布。

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

（4）超几何分布

定义： 设有 N 件产品，其中有 M 件次品，从中任取 n 件，则取出的次品数 X 的分布律为

$$P(X=k)=\frac{C_M^k C_{N-M}^{n-k}}{C_N^n},\quad k=0,1,2,\cdots,\min\{M,n\}$$

称 X 服从参数为 N 、 M 、 n 的超几何分布。

注：二项分布与超几何分布分别对应于不同的抽取方式：

设有 N 件产品，其中有 M 件次品，从中任取 n 件：
① 有放回：X服从二项分布
② 无放回：X服从超几何分布

2.3 连续型随机变量的分布

离散型随机变量可能的取值是有限个或可列无限多个，它的概率分布可以用分布律来刻画，如果随机变量可能的取值充满某个区间，那么它的概率分布需要如何来刻画呢?

1、随机变量的分布函数

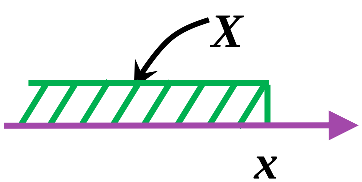
定义 设 X 是一个随机变量，称函数

$$F(x)=P(X\le x),\quad x\in \mathrm{R}$$

为随机变量 X 的分布函数。

对于随机变量$F(x)$的几何意义

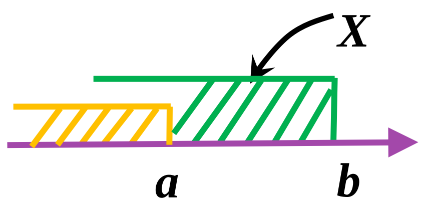
$F(x)=P(X\le x), -\infty<x<+\infty$



随机变量取值落在小于等于x一侧的概率

对于任意实数$a, b(a<b)$，有

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



注意开闭区间!

已知X分布函数，就可知X落在任意区间 $(a, b]$ 的概率

性质

1° **单调不减函数**：对于任意实数$x_1<x_2$，有 $F(x_2)-F(x_1)=P(x_1<X\le x_2)\ge 0$

2° $0\le F(x)\le 1$且 $\begin{cases} F(-\infty)=\lim\limits_{x\to-\infty}F(x)=0 \\ F(\infty)=\lim\limits_{x\to+\infty}F(x)=1 \end{cases}$

3° $F(x+0)=F(x)$且即$F(x)$为**右连续函数**

假想将分布函数定义改为$G(x)=P(X<x)$，则为左连续

$P\{X\le x\}$关于x右连续

$P\{X<x\}$关于x左连续

满足其上三点的$F(x)$必为某随机变量的分布函数

性质1–3是鉴别一个函数是否是某个随机变量的分布函数的充分必要条件。

离散型随机变量的分布函数

$$P(X=x_k)=p_k\ (k=1,2,3,\dots)$$

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

$$p_i=P(X=x_i)=F(x_i+0)-F(x_i-0)=F(x_i)-F(x_i-0)\ (i=1,2,3,\dots)$$

例 2-10 设离散型随机变量 X 的分布律为

X	-1	2	3
P	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

求 X 的分布函数 $F(x)$ 及概率 $P\left(X\leqslant\frac{1}{2}\right)$、$P\left(\frac{3}{2}<X\leqslant\frac{5}{2}\right)$和 $P(2\leqslant X\leqslant 3)$。

解 由于 X 仅在 $x=-1$， 2， 3 三点处概率不等于零，而 $F(x)$的值是 $X \le x$ 的累积概率值，由概率的有限可加性知，它即为小于或等于 x 的那些 x_i 处的概率 p_i 之和，因此当 $x<-1$ 时，

$$F(x)=0$$

当 $-1 \leqslant x < 2$ 时，

$$F(x)=P(X=-1)=\frac{1}{4}$$

当 $2 \leqslant x < 3$ 时，

$$F(x)=P(X=-1)+P(X=2)=\frac{3}{4}$$

当 x ≥3 时，

$$F(x)=P(X=-1)+P(X=2)+P(X=3)=1$$

即 *X* 的分布函数为

$$F(x)=\begin{cases}0, & x<-1\\ \frac{1}{4}, & -1\leqslant x<2\\ \frac{3}{4}, & 2\leqslant x<3\\ 1, & x\geqslant 3\end{cases}$$

分布函数*F* (*x*)的图像如图2-1所示 。

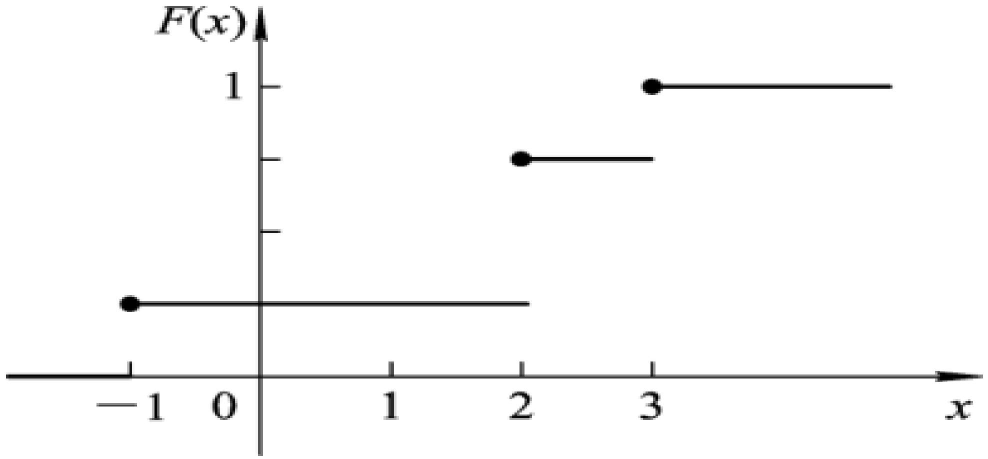


图 2-1　分布函数$F(x)$的图像

如图2-1所示，分布函数$F(x)$的图像是一条阶梯形的曲线，在 $x=-1$， 2， 3 处有跳跃，跳跃值分别为$\frac{1}{4}$，$\frac{1}{2}$，$\frac{1}{4}$ 。

如图2-1所示，分布函数$F(x)$的图像是一条阶梯形的曲线，在 $x=-1$， 2， 3 处有跳跃，跳跃值分别为$\frac{1}{4}$，$\frac{1}{2}$，$\frac{1}{4}$。

所求的概率分别为

$$P\left(X\leqslant\frac{1}{2}\right)=F\left(\frac{1}{2}\right)=\frac{1}{4}$$

$$P\left(\frac{3}{2}<X\leqslant\frac{5}{2}\right)=F\left(\frac{5}{2}\right)-F\left(\frac{3}{2}\right)=\frac{3}{4}-\frac{1}{4}=\frac{1}{2}$$

$$P(2\leqslant X\leqslant 3)=P(\{2<X\leqslant 3\}\cup\{X=2\})=P(2<X\leqslant 3)+P(X=2)$$
$$=F(3)-F(2)+P(X=2)=1-\frac{3}{4}+\frac{1}{2}=\frac{3}{4}$$

例 2-11 设离散型随机变量 X 的分布函数为

$$F(x)=\begin{cases}0, & x<-1\\ \dfrac{1}{4}, & -1\leqslant x<2\\ \dfrac{3}{4}, & 2\leqslant x<3\\ 1, & x\geqslant 3\end{cases}$$

试求随机变量 X 的分布律。