定义 2.7　泊松分布

若随机变量 X 的分布律为

$$P(X=k)=\mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}, k=0,1,\cdots,\lambda>0, \tag{2.13}$$

则称 X 服从参数为 λ 的泊松分布，记为 $X\sim P(\lambda)$.

$$\sum_{k=0}^{\infty}P(X=k)=1. \qquad \text{(级数展开式：} \mathrm{e}^{\lambda}=1+\lambda+\frac{\lambda^2}{2!}+\cdots+\frac{\lambda^k}{k!}+\cdots\text{)}$$

- 泊松分布可看成二项分布的一种极限情形.

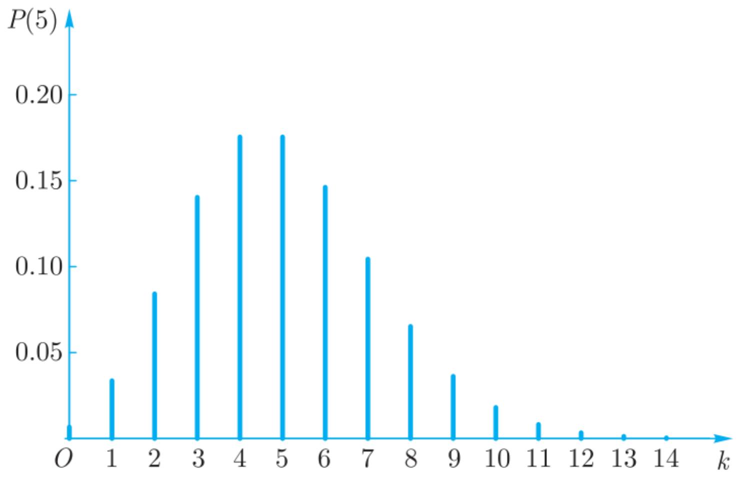


图 2.4　泊松分布

二项分布与泊松分布之间的关系

定理 2.2 泊松逼近定理

设一族随机变量 $X_n \sim B(n, p_n)$, 若当 $n \to \infty$ 时, $np_n \to \lambda > 0$, 则

$$\lim_{n\to\infty} P(X_n = k) = \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}, \quad k = 0, 1, 2, \cdots . \tag{2.14}$$

♡

- 实际应用中灵活掌握，泊松分布可近似试验次数n较大，p较小，np适中的二项分布.（如：$n \geq 30$, $np \leq 5$.）

⇓

简化计算

现在需要 100 个符合规格的元件. 从市场上买的该元件有废品率 0.01. 考虑到有废品存在, 我们准备买 $100 + a$ 个元件使得从中可以挑出 100 个符合规格的元件. 我们要求在这 $100 + a$ 个元件中至少有 100 个符合规格的元件的概率不小于 0.95. 问 a 至少要多大?

- 泊松分布适合于描述某段时间内随机事件发生的次数。
  - ✓如电话台接到的用户呼叫次数；
  - ✓某一服务设施在一定时间内到达的人数；
  - ✓交通路口的交通事故数；
  - ✓机器出现的故障数；
  - ✓自然灾害发生的次数；
  - ✓显微镜下单位分区内的细菌分布数；
  - ✓放射性物质放射的射线到达计数器的α 粒子数；
  - ✓……

例子

- 假设一块放射性物质在单位时间内发射出的α 粒子数X 服从参数为 λ 的泊松分布. 而每个发射出来的α粒子被记录下来的概率是 p, 就是说有$q = 1 - p$ 的概率被计数器漏记. 如果各粒子是否被计数器记录是相互独立的, 试求记录下来的α粒子数 Y 的分布.

补充：负二项分布与泊松分布

与二项分布类似, 负二项分布也可以由泊松分布来近似. 如果 $\lim_{r\to\infty} r(1-p)=\lambda$, 其中 $\lambda>0$ 为正常数, 那么

$$p \approx 1-\frac{\lambda}{r}.$$

由负二项分布的分布律 (2.8) 式, 我们有

$$\begin{aligned}\binom{k-1}{r-1}p^r(1-p)^{k-r} &= \frac{(k-1)\cdots r}{(k-r)!}\left(1-\frac{\lambda}{r}\right)^r\left(\frac{\lambda}{r}\right)^{k-r}\\ &= \frac{\lambda^{k-r}}{(k-r)!}\left(1+\frac{1}{r}\right)\cdots\left(1+\frac{k-r-1}{r}\right)\left(1-\frac{\lambda}{r}\right)^r.\end{aligned}$$

令 $r\to\infty$, 以及 $k-r=x$ 为固定数, 得到

$$\binom{k-1}{r-1}p^r(1-p)^{k-r}\to\frac{\lambda^x}{x!}\mathrm{e}^{-\lambda},\quad x=0,1,\cdots.$$

定义 2.8　分布函数

设 X 为随机变量, x 为任一实数, 称

$$F(x) = P(X \leqslant x) \tag{2.15}$$

为随机变量 X 的 (累积) 分布函数.

$F(x)$ 的值等于随机变量不超过 x 所取值的概率之和, 故又称为累积分布函数 (cumulative distribution function, 简称 cdf).

- $F(x)$ 为一元实函数, 其定义域为 R, 而值域为 $[0,1]$. X 值落在区间 $(a,b]$ 内的概率为

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

- 分布函数的定义适用于一切类型的随机变量, 包括离散型随机变量.

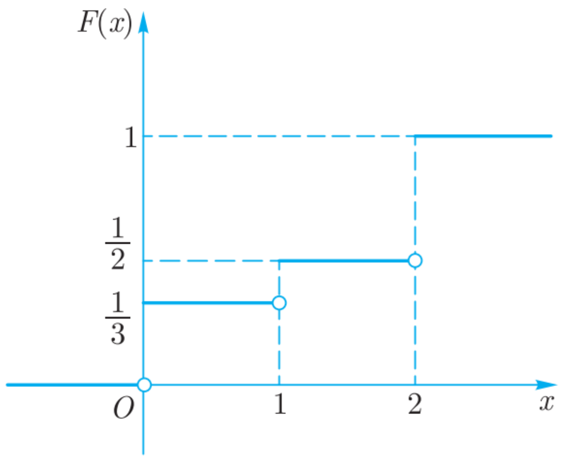
定理 2.3

设离散型随机变量 X 的可能值为 $x_1, x_2, \cdots, x_n, \cdots$, 则其分布律 $\{p_k\}$ 和分布函数 $F(x)$ 互相确定, 即它们等价地描述了 X 的概率分布情况.

- 例子：设离散型随机变量 X 的分布律如下，求它的分布函数.

X	0	1	2
P	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{2}$

- 离散型随机变量的分布函数为阶梯函数，不连续点为所有取值点，每个跳跃的高度即为取该点值的概率.



离散型随机变量 X 的分布函数图形

- 分布函数$F(x)$的性质：
  - **单调性**： $F(x)$是单调非减的函数，即对于任意的$x_1 \le x_2$，都有$F(x_1) \le F(x_2)$
  - **有界性**：对于任意的x，都有$0 \le F(x) \le 1$，且

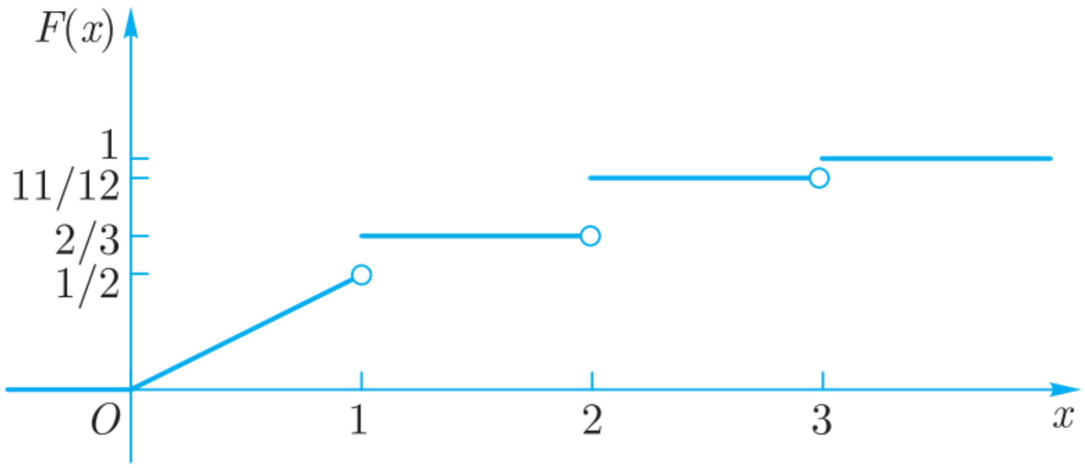
$$F(-\infty) = \lim_{x \to -\infty} F(x) = 0, \qquad F(+\infty) = \lim_{x \to +\infty} F(x) = 1$$

  - **右连续型**：F(x)是x的右连续函数，$F(x+0) = F(x)$.

$$(\text{对任意的}x_0\text{，有}\lim_{x \to x_0+} F(x) = F(x_0))$$

- 设随机变量 X 的分布函数为

$$F(x)=\begin{cases}0, & x<0,\\ x/2, & 0\leqslant x<1,\\ 2/3, & 1\leqslant x<2,\\ 11/12, & 2\leqslant x<3,\\ 1, & 3\leqslant x.\end{cases}$$



X 的分布函数图形

求 (1) $P(X<3)$; (2) $P(X=1)$; (3) $P\left(X>\dfrac{1}{2}\right)$; (4) $P(2<X\leqslant 4)$.

事件的概率用分布函数表示

- $P(a < X \leq b) = F(b) - F(a)$
- $P(X > a) = 1 - F(a)$
- $P(X = a) = F(a) - F(a-0)$
- 还有更多的…
- $P(a < X < b) = F(b-0) - F(a)$
- $P(a \leq X < b) = F(b-0) - F(a-0)$
- $P(a \leq X \leq b) = F(b) - F(a-0)$
- $P(X \geq a) = 1 - F(a-0)$

- 引例：假定步枪射手瞄准靶子, 在固定的位置进行一系列的射击. 令 X是命中点与过靶心垂线的水平偏离值 (单位: cm), 设 X 取值 $[-5,5]$. 为了计算 X值落在某区间的概率,将 $[-5,5]$ 等分为长为 1 cm 的小区间. 对于每个小区间, 以落在这个小区间的弹孔数除以弹孔总数得到落在这个区间的弹孔位点相对频数. 设弹孔总数为 100. 我们得到表 2.2. 将其用直方图表示得到图 2.8 .

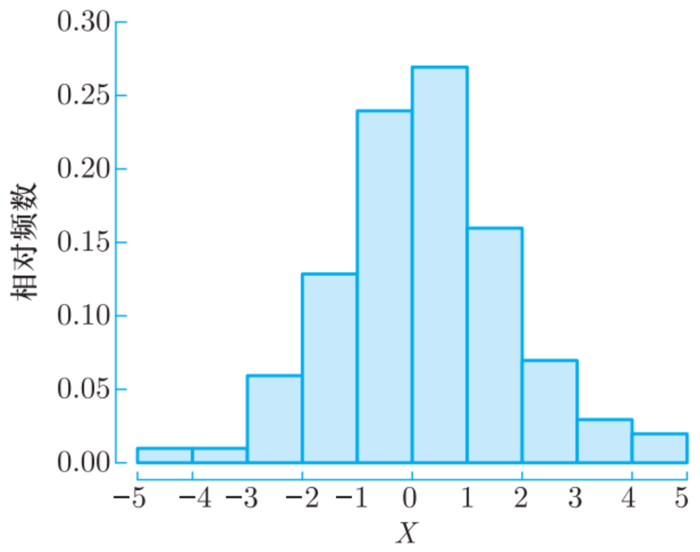


图 2.8 弹孔位点分布图

✓每个矩形的面积为相对频数.

✓全部矩形的面积是 1.

✓相对频数≈落在此区间的概率(更多实验次数，更密区间).

✓对于$[-5,5]$ 中任一子区间，区间面积≈落在此区间的概率.

概率的统计定义：频率 $\frac{n_A}{n} \approx$ 概率 p

表 2.2 弹孔位点相对频数

区间	弹孔数	相对频数
$[-5,-4)$	1	0.01
$[-4,-3)$	1	0.01
$[-3,-2)$	6	0.06
$[-2,-1)$	13	0.13
$[-1,0)$	24	0.24
$[0,1)$	27	0.27
$[1,2)$	16	0.16
$[2,3)$	7	0.07
$[3,4)$	3	0.03
$[4,5]$	2	0.02

定义 2.9 连续型随机变量和概率密度函数

设随机变量 X 的分布函数为 $F(x)$，若存在非负函数 $f(x) \geqslant 0$，使得 $\forall x \in \mathbb{R}$

$$F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t, \tag{2.17}$$

则称 X 为连续型随机变量，$f(x)$ 称为分布函数的概率密度函数 (pdf)，简称密度函数，记为 $X \sim f(x)$.

- 由定义可知，连续型随机变量的分布函数$F(x)$一定是一个连续函数(绝对连续函数).

性质 设随机变量 X 的概率密度函数为 $f(x)$, 分布函数为 $F(x)$, 则有

(1) $f(x) \geqslant 0, \forall\, x \in \mathbb{R}$;

(2) $\displaystyle\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$;

(3) $\forall x_1 < x_2$, 有 $P(x_1 < X \leqslant x_2) = F(x_2) - F(x_1) = \displaystyle\int_{x_1}^{x_2} f(x)\mathrm{d}x$. 特别地, 对任意 (可测) 集合 $A \subseteq \mathbb{R}$ 有

$$P(X \in A) = \int_A f(x)\mathrm{d}x;$$

(4) 若 $f(x)$ 在 x_0 点连续, 则 $F'(x_0) = f(x_0)$.

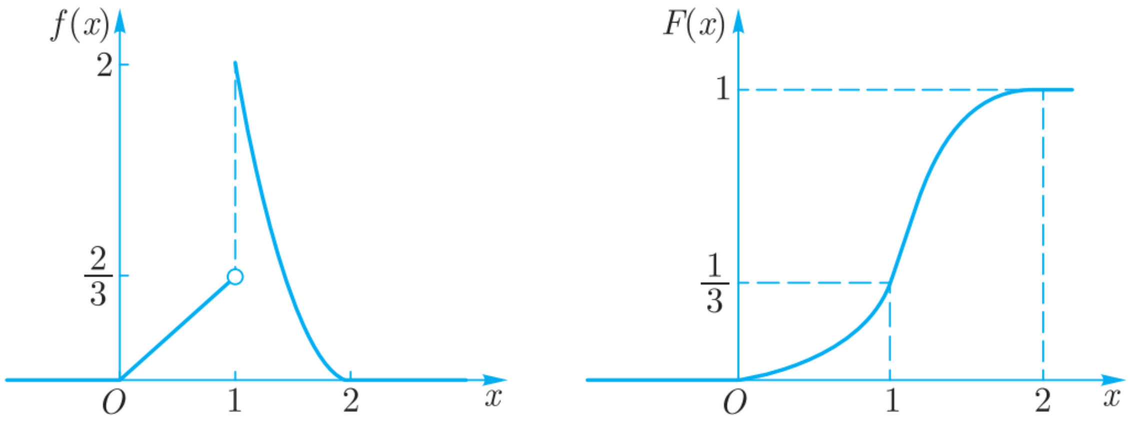
(5) $\forall x \in \mathbb{R}, P(X = x) = 0$.

- 连续型随机变量的分布函数是连续函数, 但是其概率密度函数未必都是连续函数.

- 例子：设随机变量 X 的概率密度函数如左下，由 $F(x)=\int_{-\infty}^{x} f(u)\mathrm{d}u$ 可求得其分布函数$F(x)$如右下.

$$f(x)=\begin{cases}\frac{2}{3}x, & 0\leqslant x<1,\\ 2(x-2)^2, & 1\leqslant x\leqslant 2,\\ 0, & \text{其他}.\end{cases}$$

$$F(x)=\begin{cases}0, & x\leqslant 0,\\ \frac{x^2}{3}, & 0<x\leqslant 1,\\ 1+\frac{2}{3}(x-2)^3, & 1<x\leqslant 2,\\ 1, & x>2.\end{cases}$$



密度函数 $f(x)$ 和分布函数 $F(x)$ 的图形

- 均匀分布
- 指数分布
- 正态分布

定义 2.10　均匀分布

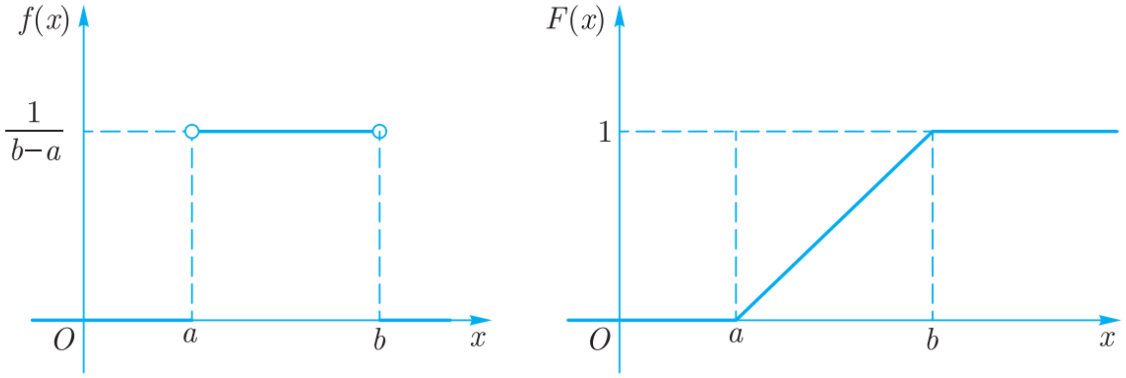
随机变量 X 在有限区间 (a,b) 内取值 $(-\infty < a < b < \infty)$，且概率密度函数为

$$f(x) = \frac{1}{b-a} I_{(a,b)}(x), \tag{2.19}$$

则称 X 服从区间 (a,b) 上的均匀分布 (uniform distribution)，记为 $X \sim U(a,b)$.

区间 (a,b) 上的均匀分布随机变量的分布函数为

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



(a)密度函数　(b)分布函数

- 对连续型随机变量 X 而言, 因为 $P(X = c) = 0$, 所以均匀分布中的区间可以是开区间, 也可以是闭区间, 或半开半闭区间, 这对分布没有影响.

- 均匀分布随机变量 X 落在 (a, b) 中任一子区间中的概率仅与区间长度成正比, 与起点无关.

定义 2.11 指数分布

若随机变量 X 的密度函数为

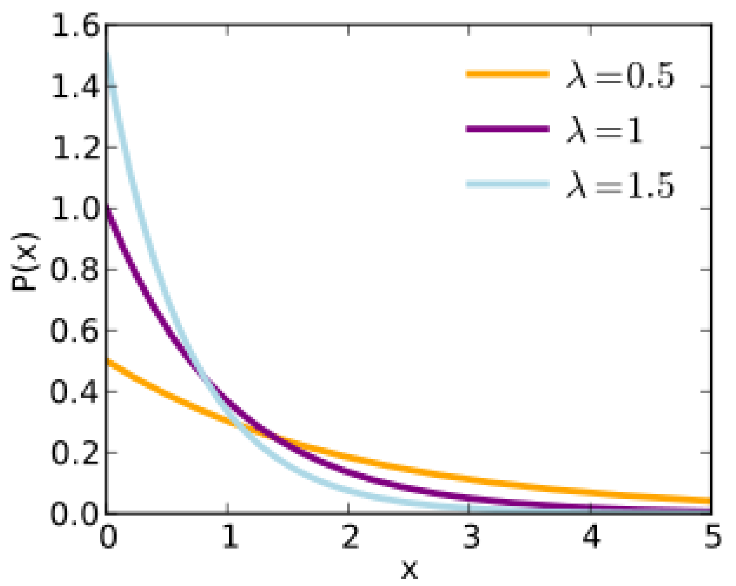
$$f(x) = \lambda \mathrm{e}^{-\lambda x} I_{(0,\infty)}(x), \tag{2.20}$$

其中 $\lambda > 0$ 为参数, 则称 X 服从参数为 λ 的指数分布 (exponential distribution), 记为 $X \sim Exp(\lambda)$.

指数分布的分布函数为

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

可以看出, 参数 λ 愈大, 密度函数下降得愈快.



指数分布的无记忆性

性质 设 $X \sim Exp(\lambda)$, 则对任意 $s, t > 0$ 有

$$P(X > s + t | X > t) = P(X > s). \tag{2.22}$$

这一性质称为无记忆性.

- 指数分布常用来作为各种产品寿命的分布的近似.
- 若把 X 视为一件仪器的寿命, 则上式就是说在仪器正常使用 t 小时后, 仪器能继续再正常使用 s 小时的概率等于仪器正常使用超过 s 小时的概率, 即寿命具有指数分布特性的仪器与之前已经用了多少小时无关.
- 指数分布是唯一的无记忆的连续型分布.(几何分布—离散型)