9.4.5 泊松分布

泊松分布也是一种重要的离散型随机变量概率分布，它适于描述某些稀有事件的状态或出现机会非常小的一些事件（如特大洪水、火山爆发、民航飞机失事、核反应堆溢漏事件等），它是由普阿松于 1837 年提出的。

设随机变量 x 表示一实验的“成功”次数，即在一段时间或一定区域内，该实验中某一特定事件发生的次数，则普哇松实验具有以下性质：

（1）发生在一定时间或特定区域内的成功次数 x 的期望值 $E(x)=\mu$ 为已知，或 $E(x)=np$ 为已知。

（2）不管时间或区域的始点，某一特定事件在某一段时间或特定区域内发生的概率相同。

（3）在极短时间或极小区域内，某一特定事件发生超过一次的概率略而不计。

（4）某一特定事件在各段时间或特定区域上出现是相互独立的。

（5）特定事件成功次数的期望值 μ 与所选择的时间或区域的大小 t 成正比，其关系为 $\mu=\lambda t$。

泊松分布的定义为：若离散型随机变量 x 的分布具有下列概率函数，即

$$f(x)=\frac{e^{\lambda t}(\lambda t)^x}{x!}=\frac{e^{-\mu}(\mu)^x}{x!} \quad（x=0，1，2，\cdots）$$

称为泊松分布。其中 μ 为此分布的参数，$e=2.71828$。其分布的重要特征值如下。

① 期望值 $E(x)=\mu$

② 方差 $V(x)=\mu$

③ 偏态系数 $\beta_1=\dfrac{1}{\sqrt{\mu}}$

④ 峰态系数 $\beta_2=3+\dfrac{1}{\mu}$

期望值与方差均为 μ 是泊松分布的一大特性。当 $\beta_1>0$，$\beta_2>3$ 时，泊松分布为具有高狭峰的右偏分布；当 β_1 随 μ 增加而趋向于 0 时，其偏斜程度则随 μ 的增加而逐渐减小，最终成对称分布；β_2 随 μ 增加而趋向 3 时，则高狭程度的峰态会随 μ 的增加而逐渐减慢，最终成为常态峰。

【例 9.21】第二次世界大战后，对伦敦西面一片 144 平方公里的地区分为 576 个方块，每块 1/4 平方公里（t）进行德国炮弹投有数调查。整个地区一共投有炮弹 537 枚，每方块中弹的分布情况如表 9-8 和图 9-8 所示。试确定泊松概率分布。

表 9-8　每方块中弹分布表

每方块中弹数 x	$f(x)$	实际中弹方块数	理论中弹方块数
0	0.39365	229	226.74
1	0.36699	211	221.39
2	0.17107	93	98.54
3	0.05316	35	30.62
4	0.01239	7	7.14
5	0.00274	1	1.58
Σ	1.000	576	576.00

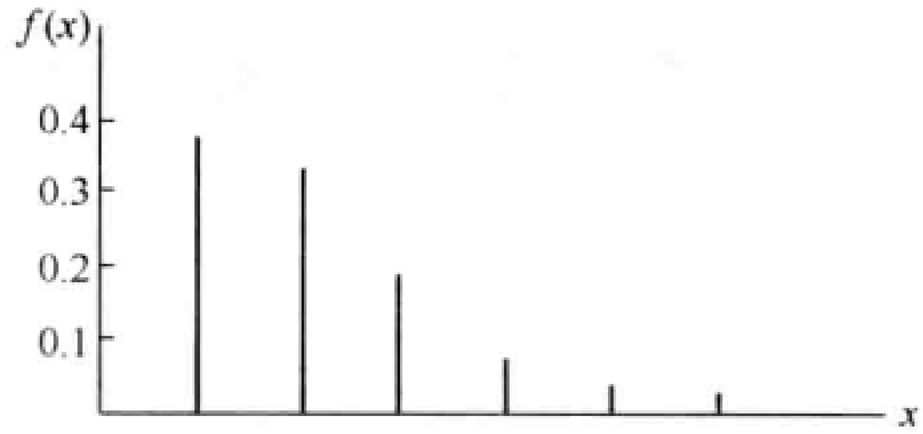

图 9-8　泊松概率分布图

解：

$$t=\frac{1}{4}\quad \lambda=\frac{537}{144}=3.7291666\text{（枚）}$$

$$\mu=\lambda t=0.93229166\text{（或}\frac{537}{576}\text{）}$$

则泊松分布的概率函数为

$$f(x)=\frac{\mathrm{e}^{-0.93229166}(0.93229166)^x}{x!}$$

① 期望值 $E(x)=\mu=0.9323$

② 方差 $V(x)=\mu=0.9323$

③ 偏态系数 $\beta_1=\frac{1}{\sqrt{\mu}}=1.04$

④ 峰态系数 $\beta_2=3+\frac{1}{\mu}=4.07>3$

此泊松分布为具有高狭峰的右偏分布。

【例 9.22】 某批产品的不良率为 2%，现从中重复抽取 10 个，试分别用二项分布和泊松分布确定 10 个中出现 1 个不良品的概率。

解：（1）

$$f(x)=C_n^x p^x(1-p)^{n-x}$$

$$f(x=1)=C_{10}^1 0.2^1(1-0.2)^{10-1}=0.1667$$

（2）

$$f(x)=\frac{e^{-\mu}(\mu)^x}{x!}$$

$$f(x=1)=\frac{e^{-0.2}0.2^1}{1!}=0.1667$$

结论：当二项分布中的 n 相当大，p 又较小时，二项分布可用 $u=np$ 的泊松分布来近似。

【例 9.23】 某市平均每月有 5 人因车祸死亡，试问每月没有人死亡的概率为多少？

解：$\mu=5$，故没有人死亡的概率为

$$f(x=0)=\frac{e^{-5}\cdot 5^0}{0!}=0.0067$$

【例 9.24】 设每 10 人中平均有 1 人吸烟，现随机调查 50 人，其吸烟人数最多有 3 人的概率为多少？

解：$\mu=\frac{1}{50}\times 50=5$，则吸烟人数最多有 3 人的概率为

$$f(x\leqslant 3)=\sum_{x=0}^{3}x\frac{e^{-5}\cdot 5^0}{3!}$$

$$=0.0067+0.0337+0.0842+0.1404=0.2650$$

9.5 连续型随机变量概率分布

9.5.1 正态分布

正态分布又称常态分布或高斯分布，是一种非常重要的连续型随机变量的概率分布。其定义为：若连续型随机变量 x 的分布具有下列概率密度函数，即

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

则称为正态分布。式中 μ 和 σ 为此分布的参数。（μ 为总体均值，σ 为总体标准差），$e=2.71828$，$\pi=3.1416$。

正态分布的重要特征值如下。

（1）期望值 $E(x)=\mu$，且 $\mu=M_e=M_0$

（2）方差 $V(x)=\sigma^2$

（3）偏态系数 $\beta_1=0$

（4）峰态系数 $\beta_2=3$

正态分布具有下列重要性质。

（1）正态分布具有常态峰，即以 μ 为中心的左右对称分布，左右二者面积相等，均为 $\frac{1}{2}$。

（2）正态分布曲线左右两尾与横轴渐近，但不与横轴相交，即 $-\infty<x<\infty$。

（3）当 $x=\mu$ 值时，正态分布的概率密度函数值最大，当 $x\neq\mu$ 时，$f(x)$ 的值随 $|x|$ 的值递增而递减，如图 9-9 所示。

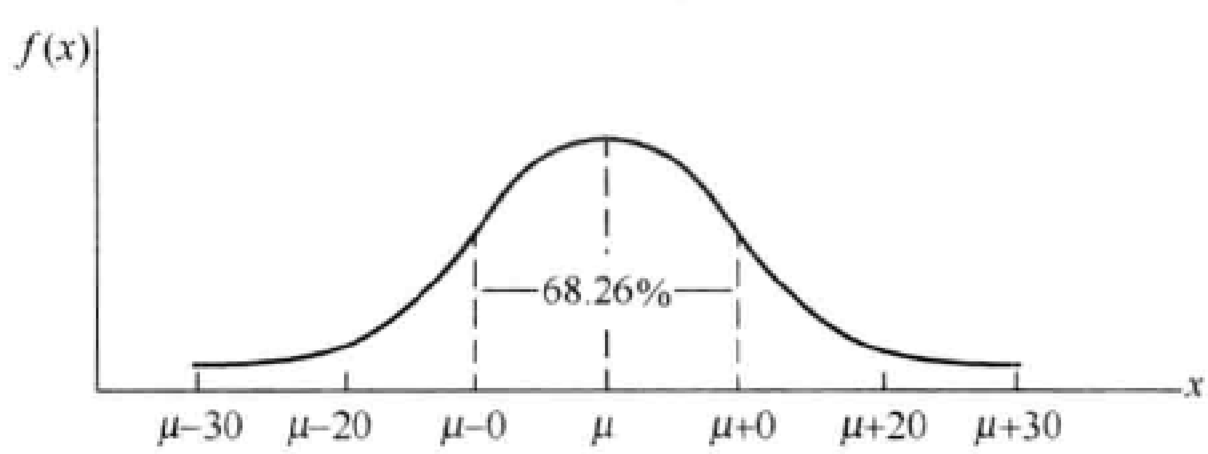

图 9-9 正态分布 x 的取值区间及概率

（4）正态分布曲线有两个拐点，分别在横轴 $\mu-\sigma$ 与 $\mu+\sigma$ 所对应的曲线上。

（5）正态分布曲线下的面积（区间概率）是固定的，如表 9-9 所示。

表 9-9 正态分布的概率与概率密度

x 的取值区间	概率 $p(x)$%	概率密度(x)
$\mu\pm\sigma$	68.26	1.000
$\mu\pm1.645\sigma$	90.00	1.645
$\mu\pm1.96\sigma$	95.00	1.960
$\mu\pm2\sigma$	95.45	2.000
$\mu\pm2.326\sigma$	98.00	2.326
$\mu\pm2.576\sigma$	99.00	2.576
$\mu\pm3\sigma$	99.74	3.000

在实践中，由于不同现象的随机变量有不同的参数 μ 和 σ，且不同随机变量的计量单位也不同，因而有不同的正态分布形状，从而给正态分布的应用带来了不便之处。为此，可令正态分布概率密度中的 $\frac{x-u}{\sigma}=z$，则有

$$f(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{\frac{1}{2}x^2} \qquad (-\infty < x < \infty)$$

因此，新的随机变量 z 仍服从正态分布，且该正态分布的参数 $\mu=0$，$\sigma=1$。同时，无论 x 的计量单位如何，新变量以 σ 为计量单位，则称 z 为标准正态随机变量，称 z 的分布为标准正态分布，见图 9-10 所示。其重要的特征值为

（1）期望值　$E(z)=0$

（2）方差　$V(z)=1$

（3）偏态系数　$\beta_1=0$

（4）峰态系数　$\beta_2=3$

（5）最高纵轴　$f(z)=\frac{1}{\sqrt{2\pi}}$

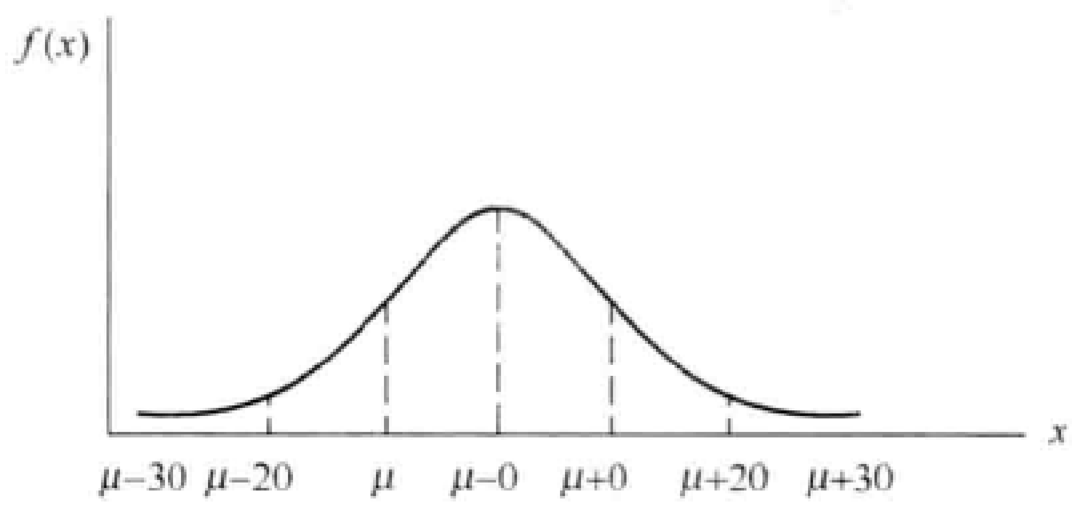

图 9-10　标准正态分布图

由于任何正态分布都可以通过 $\frac{x-\mu}{\sigma}$ 的变量转换化为标准正态分布（z 分布），因此，只要计算出正态随机变量 z 的取值区间$[-\infty, z]$，就可求出相应的区间概率 p（$z\leqslant z_i$），并可利用编成的 z 分布表可求出任何正态随机变量 x 的取值区间$[x_1, x_2]$的概率。即

$$\begin{aligned}p(x_1\leqslant x\leqslant x_2) &= p\left(\frac{x_1-\mu}{\sigma}\leqslant\frac{x-\mu}{\sigma}\leqslant\frac{x_2-\mu}{\sigma}\right)\\ &= p(z_1\leqslant z\leqslant z_2)\\ &= \int_{z_1}^{z_2} f(z)dz\\ &= \int_{-\infty}^{z_2} f(z)dz-\int_{-\infty}^{z_1} f(z)dz\\ &= p(z\leqslant z_2)-(z\leqslant z_1)\end{aligned}$$

【例 9.25】某机床制造某零件的长度 x 服从正态分布，其平均长度为 15 公分，标准差为 0.2 公分，求零件长度 x 在 14.7 公分至 15.3 公分的概率。

解：　$\mu=15$，$\sigma=0.2$

$z_1=\frac{14.7-15}{0.2}=-1.5$　查 z 分布表 $P(z\leqslant-1.5)=0.0668$

$z_2=\frac{15.3-15}{0.2}=+1.5$　查 z 分布表 $P(z\leqslant+1.5)=0.9332$