3 离散型随机变量

3.1 离散型随机变量及分布列

对于很多实际的问题, 可以将试验结果用数值表示, 有些试验结果本身就是数值, 例如:

- 抛一枚骰子的点数$\{1, 2, 3, \cdots, 6\}$;
- 一年国家出生的婴儿数$\{1, 2, \cdots, n, \cdots\}$.

有些试验结果可能与数值无关, 但可以将结果数值化, 例如:

- 抛一枚硬币,正面朝上用0表示,正面朝下用1表示;
- 流星坠落地球的落脚点用坐标纬度表示.

当试验结果用数值表示, 可以引入一个变量来表示随机事件, 由此产生随机变量的概念.

将样本空间Ω中每个样本点ω与实数$X(\omega)$相互对应, $X(\omega)$是ω的实值函数. 称实值函数$X(\omega): \Omega \to \mathbb{R}$为随机变量(random variable), 简写为r.v., 一般用大写字母X, Y, Z表示. $X(\omega)$随样本点ω的不同而取不同的值, 例如:

- 抛一枚骰子, 令$X=$出现的点数, 则X是随机变量, X的取值为$1, 2, \ldots, 6$, 集合$\{X \leq 4\}$表示点数不超过4的事件; $\{X$为偶数$\}$ 表示出现偶数点的事件.
- 用X表示一盏电灯的寿命, 则X是随机变量, 取值为$[0, +\infty)$. 集合$\{X \leq 500\}$表示电灯寿命不超过500的事件.

引入随机变量, 可以通过数学形式化后的随机变量来描述随机现象或随机事件, 利用更多数学工具. 例如: $\{X \leq -\infty\}$表示不可能事件; $\{X \leq +\infty\}$表示必然事件.

随机变量的分类: 离散型随机变量和连续型随机变量,

- **离散型随机变量**: 随机变量取值是有限的, 或无限可列的.
- **非离散型随机变量**: 随机变量取值无限且不可列的.

定义3.1. *若随机变量X的取值是有限个或无限可列个, 称X为离散型随机变量. 设随机变量X的取值为$x_1, x_2, \cdots, x_n, \cdots$, 考虑$X$每个取值的概率为$p_k = P(X = x_k)$ $(k = 1, 2, \cdots)$, 称之为X的分布列, 常用表格表示为*

X	x_1	x_2	$\cdots$	x_n	$\cdots$
P	p_1	p_2	$\cdots$	p_n	$\cdots$

根据概率的定义可得

性质3.1. *对于随机变量X的分布列$p_k = P(X = x_k)$, 有$p_k \geq 0$ $(k \geq 1)$ 以及$\sum_k p_k = 1$.*

例3.1. *设随机变量X的分布列为$P(X = k) = c\left(\frac{1}{4}\right)^k (k = 0, 1, 2, \cdots)$, 求常数$c$.*

解. 根据分布列的性质有

$$1 = \sum_{k=0}^{\infty} P(X = k) = \sum_{k=0}^{\infty} c\left(\frac{1}{4}\right)^k = c\frac{1}{1 - 1/4} = \frac{4}{3}c$$

从而得到$c = 3/4$. □

例3.2. *从$\{1, 2, \ldots, 10\}$中任取5个, 令随机变量X表示取出5个数中的最大值, 求X的分布列.*

解. 由题意可知X的取值为5, 6, 7, 8, 9, 10, 且$P(X = k) = \binom{k-1}{4}/\binom{10}{5}$ $(5 \leq k \leq 10)$. □

X	5	6	7	8	9	10
P	$\frac{1}{252}$	$\frac{5}{252}$	$\frac{15}{252}$	$\frac{35}{252}$	$\frac{70}{252}$	$\frac{126}{252}$

例3.3. *给定常数$\lambda > 0$, 随机变量X的分布列$p_i = c\lambda^i/i!$ $(i \geq 0)$, 求$P(X > 2)$.*

解. 根据分布列的性质有

$$1 = \sum_{i=0}^{\infty} p_i = c\sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = c \cdot e^{\lambda}$$

从而得到$c = e^{-\lambda}$, 进一步得到

$$P(X > 2) = 1 - P(X \leq 2) = 1 - p_0 - p_1 - p_2 = 1 - e^{-\lambda}(1 + \lambda + \lambda^2/2).$$

□

3.2 离散型随机变量的期望和方差

给定一随机变量X, 由于X的取值具有一定的随机性, 在研究中希望研究X的一些本质特征, 即刻画随机变量特征的不变量, 两种最常见的特征是期望与方差.

3.2.1 期望

定义3.2. *设离散型随机变量X的分布律为$P(X = x_k) = p_k$ $(k = 1, 2, \cdots)$. 若级数$\sum_{k=1}^{\infty} x_k \cdot p_k$绝对收敛, 称级数和为随机变量$X$的期望(expectation), 记为$E(X)$, 即*

$$E(X) = \sum_{k=1}^{\infty} x_k \cdot p_k.$$

期望又被称为均值(mean), 或加权平均(weighted average).

期望反映随机变量X所取的平均值, 值得注意的是:

- 期望$E(X)$完全由X的概率分布决定, 不是变量, 其本质是随机变量的值x_i根据概率p_i加权所得.
- 级数的绝对收敛保证了级数和不随级数各项次序的改变而改变, 期望$E(X)$反映X可能值的平均值, 不应随次序改变而改变.

根据定义求期望的步骤: i) 求随机变量X的分布律; ii) 根据定义求期望.

例3.4. *随机掷一枚骰子, X表示点数, 其取值为$\{1,2,3,4,5,6\}$, 且每点等可能发生, 其分布列为$P(X=i)=1/6$ $(i\in[6])$. 因此随机变量X的期望$E(X)=(1+2+3+4+5+6)/6=3.5$.*

例3.5. *有4个盒子编号分别为$1,2,3,4$. 将3个不同的球随机放入4个盒子中, 用X表示有球盒子的最小号码, 求$E(X)$.*

解. 先求X的分布律

$$\begin{aligned}
P(X=1)&=\frac{\binom{3}{1}3^2+\binom{3}{2}3+1}{4^3}=\frac{37}{64}, \quad & P(X=2)&=\frac{\binom{3}{1}2^2+\binom{3}{2}2+1}{4^3}=\frac{19}{64},\\
P(X=3)&=\frac{\binom{3}{1}+\binom{3}{2}+1}{4^3}=\frac{7}{64}, & P(X=4)&=\frac{1}{64}.
\end{aligned}$$

进一步可得

$$E(X)=\frac{37}{64}+2\cdot\frac{19}{64}+3\cdot\frac{7}{64}+4\cdot\frac{1}{64}=\frac{25}{16}.$$

□

例3.6. *一串钥匙有n把, 只有一把能打开门, 随机选取一把试开门, 若打不开则除去, 求打开门次数的平均数.*

解. 设随机变量X表示试开门的次数, 则可得其分布列

$$P(X=k)=\frac{\binom{n-1}{k-1}}{\binom{n}{k-1}}\cdot\frac{1}{n-k+1}=\frac{1}{n}$$

于是有打开门次数的平均数$E(X)=\sum_{k=1}^{n}k/n=(1+n)n/(2n)=(n+1)/2$. □

根据期望的定义, 可以得到如下性质:

性质3.2. *若随机变量$X\equiv c$, 则$E(c)=c$; 对常数a,b和随机变量X, 有$E(aX+b)=aE(X)+b$.*

下面给出随机变量函数的期望计算定理:

定理3.1. *设X为离散型随机变量, $g:\mathbb{R}\to\mathbb{R}$是Þ 续函数, 随机变量$Y$是$X$的函数: $Y=g(X)$. 若X的分布列为$P(X=x_k)=p_k$ $(k=1,2,\ldots)$, 且$\sum_k g(x_k)p_k$绝对收敛, 则有*

$$E[Y]=E[g(X)]=\sum_{k=1}^{\infty}g(x_k)p_k.$$

此定理意义在于计算期望$E(Y)$时, 不必计算Y的分布律, 只须利用X的分布律即可.

Proof. 证明的本质是利用绝对收敛保证无穷级数任意重排的级数仍然收敛于原无穷级数的和. 首先给出X和Y的分布列.

X	x_1	x_2	$\cdots$	x_n	$\cdots$
P	p_1	p_2	$\cdots$	p_n	$\cdots$
Y	y_1	y_2	$\cdots$	y_n	$\cdots$
P	$P(Y=y_1)$	$P(Y=y_2)$	$\cdots$	$P(Y=y_n)$	$\cdots$

将 $x_1, x_2, \ldots, x_n, \ldots$ 进行重新分组

$$\underbrace{x_{1,1}, x_{1,2}, \cdots, x_{1,k_1}}_{y_1=g(x_{1,j})\ (j\in[k_1])}, \quad \underbrace{x_{2,1}, x_{2,2}, \cdots, x_{2,k_2}}_{y_2=g(x_{2,j})\ (j\in[k_2])}, \quad \cdots, \quad \underbrace{x_{n,1}, x_{n,2}, \cdots, x_{n,k_n}}_{y_n=g(x_{n,j})\ (j\in[k_n])}, \quad \cdots$$

由此可得Y的分布列为

$$P[Y=y_i] = \sum_{j=1}^{k_i} p_{i,j}$$

从而得到

$$E[Y] = \sum_{i=1}^{\infty} y_i P[Y=y_i] = \sum_{i=1}^{\infty} y_i \sum_{j=1}^{k_i} p_{i,j} = \sum_{i=1}^{\infty}\sum_{j=1}^{k_i} g(x_{i.j})p_{i.j} = \sum_{j=1}^{\infty} g(x_j)p_j,$$

最后一个等式成立是因为绝对收敛级数的重排. □

推论3.1. *设X为离散型随机变量, 其分布列为$p_k = P(X = x_k)$. $g_i : \mathbb{R} \to \mathbb{R}$是连续函数$(i \in [n])$, 且$E(g_i(X))$存在. 对任意常数$c_1, c_2, \ldots, c_n$, 有*

$$E\left(c_1g_1(X) + c_2g_2(X) + \cdots + c_ng_n(X)\right) = \sum_{i=1}^{n} c_iE(g_n(X)).$$

Proof. 根据定理 3.1有

$$\begin{aligned} & E\left(c_1g_1(X) + c_2g_2(X) + \cdots + c_ng_n(X)\right) \\ = & \sum_{k=1}^{\infty} p_k(c_1g_1(x_k) + c_2g_2(x_k) + \cdots + c_ng_n(x_k)) \\ = & \sum_{i=1}^{n} c_i \sum_{k=1}^{\infty} p_kg_i(x_k) = \sum_{i=1}^{n} c_iE(g_i(X)). \end{aligned}$$

□

给定随机变量X和连续函数$g : \mathbb{R} \to \mathbb{R}$, $E[g(X)]$与$g(E[X])$有完全不同的涵义, 没有直接的大小关系, 但对于特殊的凸函数或凹函数, 有一个重要的结论.

定义3.3. *函数$g : (a, b) \to \mathbb{R}$满足:*

- 对$\forall x_1, x_2 \in (a,b)$, 有$g(x_1/2+x_2/2) \le (g(x_1)+g(x_2))/2$, 称函数$g$是定义在$(a,b)$上的凸函数;
- 对$\forall x_1, x_2 \in (a,b)$, 有$g(x_1/2+x_2/2) \ge (g(x_1)+g(x_2))/2$, 称函数$g$是定义在$(a,b)$上的凹函数.

定理3.2 (Jensen不等式)**.** *设X为随机变量, 若$g:\mathbb{R}\to\mathbb{R}$是连续的凸函数, 则有$g(E(X)) \le E(g(X))$; 若$g:\mathbb{R}\to\mathbb{R}$是连续的凹函数, 则有$g(E(X)) \ge E(g(X))$.*

此定理对人工智能或计算机学科的分析非常重要, 其证明略去. 根据此定理, 可以直接得到:

- 对任意随机变量X, 有$(E(X))^2 \le E(X^2)$;
- 对任意随机变量X, 有$e^{E(X)} \le E(e^X)$.

3.2.2 方差

数学期望反映了X取值的平均值, 而方差则反映了X偏离平均值$E(X)$的程度. 例如三个随机变量X, Y, Z, 其分布列分别为

$$P(X=0)=1;\qquad P(Y=1)=1/2, P(Y=-1)=1/2;\qquad P(Z=2)=1/5, P(Z=-1/2)=4/5.$$

尽管$EX=EY=EZ=0$, 即三随机变量的均值相同, 但这三个随机变量与均值0的偏离程度有很大的差异, 即本节所研究的方差.

定义3.4. *设X为离散性随机变量, 其分布列为$p_k=P(X=x_k)$ $k=1,2,\ldots$, 若$E(X)=\sum_k x_k p_k$存在, 以及$E(X-E(X))^2=\sum_k p_k(x_k-E(X))^2$存在, 称$E(X-E(X))^2$为随机变量$X$的方差(variance), 记为$Var(X)$或$D(X)$, 即*

$$Var(X)=D(X)=E(X-E(X))^2=\sum_k p_k(x_k-E(X))^2=\sum_k p_k\left(x_k-\sum_k x_k p_k\right)^2.$$

称$\sqrt{Var(X)}$为标准差(standard deviation), 记为$\sigma(X)$.

方差有另外一种简化计算的定义:

定理3.3. *对随机变量X, 有*

$$Var(X)=E(X-E(X))^2=E(X^2)-(E(X))^2.$$

Proof. 根据定义和推论 3.1有

$$\begin{aligned}\mathrm{Var}(X)&=E(X-E(X))^2=E(X^2-2XE(X)+E^2(X))\\&=E(X^2)-2E(X)E(X)+(E(X))^2=E(X^2)-(E(X))^2.\end{aligned}$$

□

例3.7. 随机变量X的取值为$x_1, x_2, \cdots, x_n$, 其分布列为$P(X = x_i) = 1/n$. 求计算随机变量X方差需要遍历$x_1, x_2, \cdots, x_n$几遍.

解. 若利用定义$\mathrm{Var}(X) = E(X - E(X))^2$需要遍历$x_1, x_2, \cdots, x_n$两遍, 若利用$\mathrm{Var}(X) = E(X^2) - (E(X))^2$则只需要遍历$x_1, x_2, \cdots, x_n$一遍. □

下面看看方差的性质:

性质3.3. *i) 如果随机变量$X \equiv c$, 则有$Var(X) = 0$;*
ii) 对任意随机变量X和常数a, b, 有$Var(aX) = a^2 Var(X)$以及$Var(aX + b) = a^2 VarD(X)$;
iii) 对任意随机变量X和常数c, 有$Var(X) = E(X - E(X))^2 \le E(X - c)^2$.

Proof. 如果随机变量$X \equiv c$, 则有$E(X) = c$, 从而得到$\mathrm{Var}(X) = E(X - E(X))^2 = E(c - c)^2 = 0$.
对于ii), 有$E(aX) = aE(X)$和$E(aX + b) = aE(X) + b$, 以及

$$\mathrm{Var}(aX) = E(aX - E(aX))^2 = a^2 E(X - E(X))^2 = a^2 \mathrm{Var}(X),$$
$$\mathrm{Var}(aX + b) = E(aX + b - E(aX + b))^2 = a^2 E(X - E(X))^2 = a^2 \mathrm{Var}(X).$$

对于iii), 有

$$\begin{aligned} E(X - c)^2 &= E(X - E(X) + E(X) - c)^2 \\ &= E(X - E(X))^2 + E[(X - E(X))(E(X) - c)] + (E(X) - c)^2 \\ &= E(X - E(X))^2 + (E(X) - c)^2 \ge E(X - E(X))^2. \end{aligned}$$

□

3.3 常用的离散型随机变量

本节介绍几种常见且重要的离散型随机变量

3.3.1 离散型均匀分布

定义3.5. 设随机变量X的取值$a_1, a_2, \cdots, a_n$, 且$P(X=a_i)=1/n$, 称X服从离散型均匀分布.

由定义可知

$$\begin{aligned} E(X) &= \frac{1}{n}\sum_{i=1}^{n} a_i, \\ \mathrm{Var}(X) &= \frac{1}{n}\sum_{i=1}^{n} a_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} a_i\right)^2. \end{aligned}$$

例3.8 (德国坦克数量问题)**.** 二战期间假德国坦克编号为 $1, 2, \cdots, N$, 共生产 N 辆. 盟军战斗中随机击毁了 k 辆, 且所击毁的最大编号为 m. 如何估计 N 的大小.

解. 上述问题本质上是从 $1, 2, \cdots, N$ 中以'不放回'方式抽取 k 个数, 观察到 k 个数中最大数为 m, 如何利用 m 和 k 估计 N.

随机变量$X=${抽取 k 个数中的最大数} $(X=k, k+1, \cdots, N)$, 有

$$Pr[X=i] = \binom{i-1}{k-1}\Big/\binom{N}{k}, \quad \text{从而得到} \quad E(X) = \sum_{i=k}^{N} i\binom{i-1}{k-1}\Big/\binom{N}{k}.$$

现考虑从 $N+1$ 个元素中取 $k+1$ 个元素, 共有多少种不同的取法, 此问题可等价于按所抽取 $k+1$ 个元素中最大元进行分类, 即

$$\binom{N+1}{k+1} = \sum_{i=k}^{N}\binom{i}{k} = \sum_{i=k}^{N}\frac{i}{k}\binom{i-1}{k-1}$$

代入 $E(X)$ 可得

$$E(X) = \sum_{i=k}^{N} i\binom{i-1}{k-1}\Big/\binom{N}{k} = k\binom{N+1}{k+1}\Big/\binom{N}{k} = \frac{k}{k+1}(N+1).$$

由于仅做了一次观察, 可以将一次观察的最大值 m 看作为 $E[X]$ 的近似, 即

$$m = E(X) = \frac{k}{k+1}(N+1) \quad \Rightarrow \quad N = m(1+k^{-1}) - 1$$

从而完成 N 的估计. □

例如, 如果观察到被击毁坦克编号分别为 $17, 68, 94, 127, 135, 212$, 根据上面的推到可估计出 $N = 212(1+\frac{1}{6}) - 1 \approx 246$. 下表给出这种方法的估计效果:

由此可见: 统计估计比情报估计准确得多, 接近德国的实际产量.

时间	统计估计	英国情报估计	德国实际产量
1940-06	169	1000	122
1941-06	244	1550	271
1942-08	327	1550	342

3.3.2 0-1分布

定义3.6. *随机变量 X 的取值为 $\{0,1\}$, 其分布列 $P(X=1)=p$, $P(X=0)=1-p$, 称 X 服从参数为 p 的 $0-1$ 分布, 又称两点分布, 或Bernoulli分布, 记 $X \sim Ber(p)$.*

由定义可知

$$E(X)=p, \qquad Var(X)=p-p^2=p(1-p).$$

$0-1$分布是很多概率模型的基础.

3.3.3 二项分布

Bernoulli试验E只有两个结果: A 和 $\bar{A}$, 将试验E独立重复地进行 n 次, 称为 n 重Bernoulli试验. n 重Bernoulli试验是一种很重要的数学模型, 具有广泛的应用.

定义3.7. *在 n 重Bernoulli试验, 设事件A发生的概率为p, 用随机变量 X 表示 n 次独立试验中事件 A 发生的次数, X 的取值为$0,1,2,\ldots,n$, 其分布列为 $P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$ $(k=0,1,2,\cdots,n)$, 称 X 服从二项分布(binomial distribution), 记$X \sim B(n,p)$.*

由于分布列类似于二项展开式$(a+b)^n=\sum_{k=1}^{n}\binom{n}{k}a^k b^{n-k}$, 因此称为二项分布. 容易验证

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k} \geq 0, \quad \sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k}=1,$$

从而得到二项分布是一个分布列.

性质3.4. *若 $X \sim B(n,p)$, 则有*

$$E(X)=np, \qquad Var(X)=np(1-p).$$

Proof. 由定义可知:

$$E(X)=\sum_{k=0}^{n}k\binom{n}{k}p^k(1-p)^{n-k}=(1-p)^n\sum_{k=0}^{n}\binom{n}{k}k\left(\frac{p}{1-p}\right)^k.$$

为计算 $\sum_{k=0}^{n}\binom{n}{k}k\left(\frac{p}{1-p}\right)^k$, 考虑

$$(1+x)^n=\sum_{k=0}^{n}\binom{n}{k}x^k \Longrightarrow n(1+x)^{n-1}=\sum_{k=0}^{n}\binom{n}{k}x^{k-1}k \Longrightarrow nx(1+x)^{n-1}=\sum_{k=0}^{n}\binom{n}{k}x^k k$$

将 $x = p/(1-p)$带入上式可得

$$E(X) = (1-p)^n \sum_{k=0}^{n} \binom{n}{k} k \left(\frac{p}{1-p}\right)^k = (1-p)^n n \frac{p}{1-p} \frac{1}{(1-p)^{n-1}} = np.$$

对于方差, 首先计算

$$E(X^2) = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} nk \binom{n-1}{k-1} p \cdot p^{k-1} (1-p)^{n-k}$$

令$i = k-1$, 带入上式可得

$$\begin{aligned} E(X^2) &= np \sum_{i=0}^{n-1} (i+1) \binom{n-1}{i} \cdot p^i (1-p)^{n-1-i} \\ &= np \sum_{i=0}^{n-1} i \binom{n-1}{i} \cdot p^i (1-p)^{n-1-i} + np \sum_{i=0}^{n-1} \binom{n-1}{i} \cdot p^i (1-p)^{n-1-i} \\ &= np(n-1)p + np. \end{aligned}$$

从而得到$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = np(1-p)$. □

例3.9. *一张试卷上有5道选择题, 每道选择题有4个答案, 只有一个是正确的, 某学生靠随机猜测至少能做对4个题的概率是多少?*

解. 每答一道题相当于一次Bernoulli试验, 事件 A ={答对一题}, 有 $P(A) = 1/4$. 5道题相当于5重Bernoulli试验. 用 X 表示学生猜对的题数, 则 $X \sim B(5, 1/4)$, 从而得到

$$P(X \geq 4) = P(X=4) + P(X=5) = \binom{5}{4} \left(\frac{1}{4}\right)^4 \cdot \frac{3}{4} + \left(\frac{1}{4}\right)^5 = \frac{1}{64}.$$

□

3.3.4 几何分布

定义3.8. *在多重Bernoulli试验, 设事件A发生的概率为p, 用随机变量 X 表示事件 A 首次发生时的试验次数, X 的取值为$1, 2, \ldots$, 其分布列为 $P(X = k) = (1-p)^{k-1} p$ $(k = 1, 2, \ldots)$, 称 X 服从参数为 p 的几何分布, 记$X \sim G(p)$.*

容易得到 $P(X = k) = (1-p)^{k-1} p \geq 0$, 以及

$$\sum_{k=1}^{\infty} P(X = k) = p \sum_{k=1}^{\infty} (1-p)^{k-1} = p \frac{1}{1-(1-p)} = 1,$$

从而验证了几何分布构成一个分布列.

性质3.5. *若随机变量 $X \sim G(P)$, 则有*

$$E(X) = \frac{1}{p}, \quad Var(X) = \frac{1-p}{p^2}.$$

Proof. 根据期望的定义有

$$E(X)=\sum_{k=1}^{\infty}kP(X=k)=\sum_{k=1}^{\infty}kp(1-p)^{k-1}=p\sum_{k=1}^{\infty}k(1-p)^{k-1}.$$

根据公式

$$\sum_{k=0}^{\infty}x^k=\frac{1}{1-x}\quad\Rightarrow\quad\sum_{k=1}^{\infty}kx^{k-1}=\frac{1}{(1-x)^2}.$$

令$x=1-p$可得$E(X)=1/p$. 对于随机变量 X 的方差, 首先计算

$$E(X^2)=\sum_{k\geq 1}k^2p(1-p)^{k-1}=p\sum_{k\geq 1}k^2(1-p)^{k-1}=p\cdot\frac{2-p}{p^3}=\frac{2-p}{p^2}.$$

根据公式

$$\sum_{k=0}^{\infty}x^k=\frac{1}{1-x}\quad\Rightarrow\quad\sum_{k=1}^{\infty}kx^{k-1}=\frac{1}{(1-x)^2}\quad\Rightarrow\quad\sum_{k=1}^{\infty}kx^k=\frac{x}{(1-x)^2}\quad\Rightarrow\quad\sum_{k=1}^{\infty}k^2x^{k-1}=\frac{1+x}{(1-x)^3}.$$

令 $x=1-p$ 可得 $E(X^2)=(2-p)/p^2$, 于是有 $\mathrm{Var}(X)=E(X^2)-(EX)^2=(1-p)/p^2$. □

下面研究几何分布的特性: 无记忆性(memoryless property).

定理3.4. *设$X\sim G(P)$, 对任意正整数m,n, 有*

$$P(X>m+n|X>m)=P(X>n).$$

直观理解: 假设已经经历了 m 次失败, 从当前起直至成功的次数与 m 无关.

Proof. 对任何正整数 k, 根据几何分布的定义有

$$P(X>k)=\sum_{i=k+1}^{\infty}p(1-p)^{i-1}=p\sum_{i=k+1}^{\infty}(1-p)^{i-1}=p\frac{(1-p)^k}{1-(1-p)}=(1-p)^k.$$

根据条件概率的定义有

$$P(X>m+n|X>m)=\frac{P(X>m+n)}{P(X>m)}=\frac{(1-p)^{m+n}}{(1-p)^m}=(1-p)^n=P(X>n)$$

这里利用$\{X>m+n\}\cap\{X>m\}=\{X>m+n\}$. □

例3.10. *古代有一村落因为资源限制不能养太多的人, 每个家庭都很重视男性, 于是制定一条规则: 每个家庭可以生一个男孩, 如果没有男孩则可以继续生育直至有一个男孩为止. 若已有一个男孩, 则不再生育. 若生男孩的概率为p, 问多年后该村的男女比例是多少?*

解. 对一个家庭而言, 用随机变量 X 表示该家庭的小孩个数, 则 $X=1,2,\ldots$, 以及

$$P(X=i)=p(1-p)^{i-1},$$

即 $X\sim G(p)$. 于是一个家庭的小孩数的期望是 $E[X]=1/p$, 从而得到一个家庭中小孩的男/女比例为 $1:(1/p-1)$, 特别地, 当 $p=1/2$ 时男女比例 $1:1$. □

3.3.5 泊松分布

定义3.9. 如果随机变量X的分布列为

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

其中 $\lambda>0$ 是一个给定的常数, 称随机变量X服从参数为λ的泊松分布, 记为 $X\sim P(\lambda)$.

容易验证$P(X=k)=\lambda^k e^{-\lambda}/k!\geq 0$, 根据泰勒展式$e^\lambda=\sum_{k=0}^{\infty}\lambda^k/k!$有

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

泊松分布用于描述大量试验中稀有事件出现次数的概率模型, 例如:

- 电话在一段时间内收到的呼叫次数,
- 放射物在一段时间内放射的粒子数,
- 一段时间内通过某路口的出租车数,
- 一本书中一页出现的语法错误数,
- 一天内道一所银行办理业务的顾客数.

性质3.6. 对任意给定的 $\lambda>0$, 若 $X\sim P(\lambda)$, 则

$$E(X)=\lambda, \qquad Var(X)=\lambda.$$

Proof. 根据期望的定义有

$$E(X)=\sum_{i=0}^{\infty}i\cdot P(X=i)=\sum_{i=1}^{\infty}i\cdot\frac{\lambda^i}{i!}e^{-\lambda}=\lambda\sum_{i=1}^{\infty}\frac{\lambda^{i-1}}{(i-1)!}e^{-\lambda}=\lambda.$$

对于随机变量的方差, 首先计算

$$\begin{aligned}E[X^2] &= \sum_{i=0}^{\infty}i^2P(X=i)=\sum_{i=1}^{\infty}i^2\frac{\lambda^i}{i!}e^{-\lambda}=\lambda\sum_{i=1}^{\infty}i\cdot\frac{\lambda^{i-1}}{(i-1)!}e^{-\lambda}\\ &= \lambda\sum_{i=1}^{\infty}(i-1)\frac{\lambda^{i-1}}{(i-1)!}e^{-\lambda}+\lambda=\lambda\sum_{k=0}^{\infty}k\frac{\lambda^k}{k!}e^{-\lambda}+\lambda=\lambda^2+\lambda.\end{aligned}$$

从而得到$\mathrm{Var}(X)=E[X^2]-(E[X])^2=\lambda$. □

例3.11. 设随机变量X服从参数为λ的泊松分布, 且$P(X=1)=P(X=2)$, 求$P\{X\geq 4\}$.

解. 根据泊松分布的定义可知 $P(X=k)=\lambda^k e^{-\lambda}/k!$, 由此可得

$$P(X=1)=P(X=2) \quad \Rightarrow \quad \frac{\lambda e^{-\lambda}}{1!}=\frac{\lambda^2}{2!} \quad \Rightarrow \quad \lambda=2,$$

进一步得到

$$\begin{aligned} P\{X\geq 4\} &= 1-P(X=0)-P(X=1)-P(X=2)-P(X=3) \\ &= 1-e^{-2}-4e^{-2}-\frac{4}{3}e^{-2}=1-\frac{19}{3}e^{-2}. \end{aligned}$$

□

下面研究二项分布于泊松分布的关系: 即可以用泊松分布逼近二项分布.

定理3.5 (泊松定理)**.** *对任意给定的常数$\lambda>0$, n为任意正整数, 设$np_n=\lambda$, 则对任意给定的非负整数k, 有*

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

Proof. 由 $p_n=\lambda/n$, 有

$$\begin{aligned} \binom{n}{k}p_n^k(1-p_n)^{n-k} &= \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!}(1-\frac{1}{n})\cdots(1-\frac{k-1}{n})\left(1-\frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!}(1-\frac{1}{n})\cdots(1-\frac{k-1}{n})\left(1-\frac{\lambda}{n}\right)^{\frac{n}{\lambda}\frac{n-k}{n}\lambda} \end{aligned}$$

当$n\to\infty$, 有$(1-\frac{\lambda}{n})^{\frac{n}{\lambda}}\to e$以及$\frac{n-k}{n}\to 1$, 从而完成证明. □

泊松分布的应用: 若 $X\sim B(n,p)$, 当 n 比较大, p 比较小时, 令 $\lambda=np$, 有

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}\approx\frac{\lambda^k}{k!}e^{-\lambda}.$$

即利用泊松分布近似计算二项分布.

例3.12. *设每次射击命中目标的概率为0.002, 现射击1000次, 求命中目标在500次与600次之间的概率. (用泊松近似计算)*

解. 将1000次射击可看作1000重Bernoulli试验, 设随机变量 X 表示1000次设计中射中目标的次数, 则 $X\sim B(1000,0.002)$, 利用泊松分布近似, 则可以看作 $X\sim P(2)$, 于是有

$$P(500\leq X\leq 600)=\sum_{k=500}^{600}\binom{1000}{k}(0.002)^k 0.998^{1000-k}\approx\sum_{k=500}^{600}\frac{2^k}{k!}e^{-2}.$$

□

例3.13. *有80台同类型设备, 各台独立工作, 发生故障的概率是0.01, 一台设备发生故障时只能1人处理, 考虑方案: I) 由4人维护, 每人单独负责20台; II) 由三人共同维护80台. 哪种方案更为可取?*

解. 对方案 I): 令 X 为第1人负责20台中同一时刻发生故障的台数, 则$X \sim B(20, 0.01)$. 设A_i表示第i人负责的设备发生故障不能及时维修的事件, 则

$$P(A_1 \cup A_2 \cup A_3 \cup A_4) \geq P(A_1) = P\{X \geq 2\} = 1 - P(X=0) - P(X=1) \approx 1 - \sum_{k=0}^{n} \frac{(0.2)^k}{k!} e^{-0.2} \approx 0.0175.$$

对方案 II): 设随机变量Y为80台设备中同一时刻发生故障的台数, 则$Y \sim (B, 0.01)$, 则有设备发生故障不能及时维修的概率为

$$P(Y \geq 4) = 1 - \sum_{k=1}^{3} \binom{80}{k} 0.01^k 0.99^{80-k} \approx 1 - \sum_{k=0}^{3} \frac{(0.8)^k}{k!} e^{-0.8} \approx 0.0091.$$

由此可知: 方案 II)更优. □