最后所求概率为

$$\begin{aligned} &P(X=n)+P(X=n-r) \\ =\ & \binom{2n-r}{n}(1/2)^n(1/2)^{n-r}+\binom{2n-r}{n-r}(1/2)^{n-r}(1/2)^n=\binom{2n-r}{n}\Big/2^{2n-r-1}\ , \end{aligned}$$

由此完成证明.

例 3.8 一个系统由 n 个独立的元件组成, 每个元件能正常工作的概率为 p, 若该系统中至少有一半的元件能正常工作则整个系统有效, 在什么情况下 5 个元件的系统比 3 个元件的系统更有效?

解 用 X 表示 n 个元件能正常工作的元件数, 则有 $X\sim B(n,p)$. 由此可知包含有 5 个元件的系统有效的概率为

$$\binom{5}{3}p^3(1-p)^2+\binom{5}{4}p^4(1-p)+\binom{5}{5}p^5=p^3(6p^2-15p+10)\ ,$$

而包含有 3 个元件的系统有效的概率为

$$\binom{3}{2}p^2(1-p)+\binom{3}{3}p^3=p^2(3-2p)\ .$$

当 $p^3(6p^2-15p+10)>p^2(3-2p)$ 时, 即当 $3(p-1)^2(2p-1)>0$ 时 5 个元件的系统比 3 个元件的系统更有效, 此时 $p>1/2$.

3.4.3 泊松分布

泊松分布是概率论中另一种重要的分布, 用于描述大量试验中稀有事件出现次数的概率模型. 例如, 一个月内网站的访问量, 一个小时内公共汽车站来到的乘客数, 书中一页出现错误的语法数, 一天中银行办理业务的顾客数, 一年内中国发生的地震次数等.

定义 3.8 给定常数 $\lambda>0$, 若随机变量 X 的分布列为

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

则称随机变量 X 服从 **参数为 λ 的泊松分布** (Possion distribution), 记为 $X\sim P(\lambda)$.

容易验证 $P(X=k)=\lambda^k e^{-\lambda}/k!\geqslant 0$, 并根据指数的泰勒展式 $e^x=\sum_{k=0}^{\infty}x^k/k!$ 有

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

关于泊松分布的数字特征有:

引理 3.3 若随机变量 $X\sim P(\lambda)$, 则有 $E(X)=\lambda$ 和 $\mathrm{Var}(X)=\lambda$.

因此泊松分布可由期望或方差唯一确定.

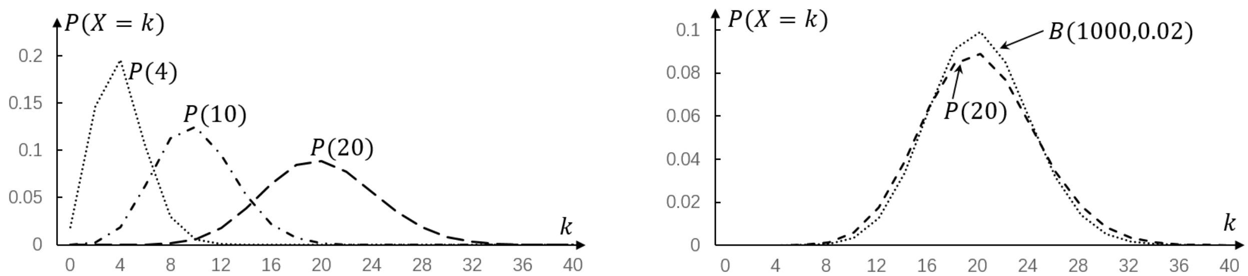
证明 根据指数的泰勒展开式有 $e^x = \sum_{k=0}^{\infty} x^k/k!$ 有

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

对于随机变量的方差, 首先计算

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

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



a) 泊松分布示意图　　**b)**二项分布$B(1000,0.02)$与泊松分布$P(20)$的逼近性

图 3.2 泊松分布示意图、以及泊松分布与二项分布的逼近图

从图 3.2(a) 中可以观察发现: 若随机变量 $X \sim P(\lambda)$, 则有 $P(X=k)$ 从一开始单调递增, 然而一致单调递减, 在期望 λ 附近取得最大值. 其次, 泊松分布与二项分布的分布图之间有一定的相似性, 如图 3.2(b) 所示, 下面的定理给出了二者之间的近似关系:

定理 3.4 (泊松定理) 设 $\lambda > 0$ 任意给定的常数, n 是一个正整数, 若 $np_n = \lambda$, 则对任意给定的非负整数 k, 有

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

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

泊松分布的应用: 若随机变量 $X \sim B(n,p)$, 当 n 比较大而 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. 针对彩票中奖、火山爆发、洪水泛滥、意外事故等小概率事件, 当试验的次数较多时, 可以将 n 重伯努利试验中小概率事件发生的次数近似服从泊松分布.

例 3.9 设有 80 台同类型设备独立工作, 每台发生故障的概率为 0.01, 一台设备发生故障时只能由一人处理, 考虑两种方案: I) 由四人维护, 每人单独负责 20 台; II) 由三人共同维护 80 台. 哪种方案更为合理?

解 首先讨论方案 I), 用事件 A_i 表示第 i 人负责的设备发生故障不能及时维修, 用 X_i 为第 i 人负责的 20 台设备同一时刻发生故障的台数, 则有 $X \sim B(20,0.01)$, 根据泊松定理有近似有 $X \sim P(0.2)$, 进一步有

$$P(A_i)=P(X_i\geqslant 2)=1-P(X=0)-P(X=1)\approx 1-\sum_{k=0}^{2}\frac{(0.2)^k}{k!}e^{-0.2}\approx 0.0175.$$

因四人独立维修, 有设备发生故障时而不能及时的概率

$$P(A_1\cup A_2\cup A_3\cup A_4)\geqslant P(A_1)\approx 0.0175.$$

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 $Y \sim B(80,0.01)$, 根据泊松定理有近似有 $Y \sim P(0.8)$, 则有设备发生故障不能及时维修的概率为

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

由此比较可知方案 II) 更优.

例 3.10 一个公共汽车站有很多路公交车, 若一个时间段内到站的乘客数 $X \sim P(\lambda)$ $(\lambda > 0)$, 所有到站的乘客是相互独立的、且选择 D1 路公交车的概率为 p $(p > 0)$, 求乘坐 D1 路公交车的乘客数 Y 的分布.

解 设一个时间段内到站的乘客数为 k, 该事件发生的概率

$$P(X=k)=\lambda^k e^{-\lambda}/k!\ .$$

根据题意可知到达公交站的 k 个人中乘坐 D1 的人数服从参数为 k 和 p 的二项分布 $B(k,p)$, 即

$$P(Y=i|X=k)=\binom{k}{i}p^i(1-p)^{k-i}\ .$$

根据全概率公式和指数函数 e^x 的泰勒展开式有

$$\begin{aligned}
P(Y=i) &= \sum_{k=i}^{+\infty} P(X=k)P(Y=i|X=k) = p^i e^{-\lambda} \sum_{k=i}^{+\infty} \binom{k}{i} \frac{\lambda^k}{k!}(1-p)^{k-i} \\
&= \frac{(p\lambda)^i e^{-\lambda}}{i!} \sum_{k=i}^{+\infty} \frac{((1-p)\lambda)^{k-i}}{(k-i)!} = \frac{(p\lambda)^i e^{-\lambda}}{i!} \sum_{k=0}^{+\infty} \frac{((1-p)\lambda)^k}{(k)!} \\
&= \frac{(p\lambda)^i e^{-\lambda}}{i!} e^{(1-p)\lambda} = \frac{(p\lambda)^i e^{-p\lambda}}{i!} \ ,
\end{aligned}$$

由此可知乘坐 D1 路公交车的乘客数 $Y \sim P(p\lambda)$.

3.4.4 几何分布

在多重 Bernoulli 试验中, 设事件 A 发生的概率为 p. 用随机变量 X 表示事件 A 首次发生需要的试验次数, 事件 $\{X=k\}$ 发生当且仅当事件 A 在前 $k-1$ 次不发生而第 k 次发生, 根据多重 Bernoulli 试验的独立性可知概率 $P(X=k)=(1-p)^{k-1}p$.

定义 3.9 设 $p \in (0,1)$ 是一个常数, 若随机变量 X 的分布列为

$$P(X=k) = (1-p)^{k-1}p \qquad (k \geqslant 1) \ , \tag{3.5}$$

称 X 服从 **参数为 p 的几何分布** (geometric distribution), 记 $X \sim G(p)$.

容易得到 $P(X=k) \geqslant 0$ 以及

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

从而验证了 (3.5) 构成概率分布列. 几何分布有一个重要的性质: **无记忆性** (memoryless property).

定理 3.5 设随机变量 $X \sim G(p)$, 对任意正整数 m, n, 有

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

证明 根据几何分布的定义, 对任何正整数 k 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

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

这里利用事件 $\{X > m+n\} \cap \{X > m\} = \{X > m+n\}$, 从而完成证明.

几何分布无记忆性的直观解释: 假设以前经历了 m 次失败, 从当前起至成功的次数与 m 无关. 例如, 一人赌博时前面总输, 觉得下一次应该会赢了, 然而无记忆性告诉大家: 下一次是否会赢与前面输了多少次没有任何关系.

关于几何分布的数字特征, 我们有

引理 3.4 若随机变量 $X \sim G(p)$ $(0 < p < 1)$, 则有

$$E(X) = \frac{1}{p} \qquad \text{和} \qquad \mathrm{Var}(X) = \frac{1-p}{p^2} \ .$$

证明 根据几何分布的定义有

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

对于非负整数的随机变量 X 有

$$E(X) = \sum_{i=1}^{+\infty} P(X \geqslant i) = \sum_{i=1}^{+\infty}(1-p)^{i-1} = 1/p \ .$$

对于随机变量 X 的方差, 首先计算

$$E(X^2) = \sum_{k=1}^{\infty} k^2 p(1-p)^{k-1} = p\sum_{k=2}^{\infty} k(k-1)(1-p)^{k-1} + 1/p \ .$$

对级数展开式 $(1-x)^{-1} = \sum_{k=0}^{\infty} x^k$ 两边先求二阶导后乘 x 有

$$\sum_{k=2}^{\infty} k(k-1)x^{k-1} = \frac{2x}{(1-x)^3} \ .$$

令 $x = 1-p$ 代入可得 $E(X^2) = (2-p)/p^2$. 最后有 $\mathrm{Var}(X) = E(X^2) - (EX)^2 = (1-p)/p^2$.

例 3.11 在古代非常重视生男孩但生存资源有限, 于是规定: 每个家庭可生一个男孩, 如果没男孩则可以继续生育直至有一个男孩; 若已有一个男孩, 则不再生育. 不妨假设每个家庭生男孩的概率为 $p = 1/2$, 问题: 1) 一个家庭恰好有 n 个小孩的概率; 2) 一个家庭至少有 n 个小孩的概率; 3) 男女比例是否会失衡?

解 用随机变量 X 表示一个家庭的小孩个数, 其取值为 $\{1, 2, \cdots\}$, 根据题意可知 X 服从参数为 $p = 1/2$ 的几何分布, 因此一个家庭恰好有 n 个小孩的概率为

$$P(X = n) = p(1-p)^{n-1} = 1/2^n \ .$$

一个家庭至少有 n 个小孩的概率为

$$P(X \geqslant n) = \sum_{k=n}^{+\infty} P(X=k) = 1/2^{n-1} \ .$$

至于男女比例是否会失衡, 考虑一个家庭平均的孩子个数为 $E[X]=1/p=2$, 由此可知在平均的情形下, 一个家庭的小孩男女比例 $1:1$, 因此不会造成男女失衡.

几何分布考虑在多重试验中事件 A 首次发生时所进行的试验次数, 可以进一步考虑事件 A 第 r 次发生时所进行的试验次数. 设随机事件 A 发生的概率为 $p \in (0,1)$, 用 X 表示事件 A 第 r 次成功时发生的试验次数, 则 X 取值 $r, r+1, r+2, \cdots$, 其分布列为

$$P(X=k) = \binom{k-1}{r-1} p^r (1-p)^{k-r} \qquad (k = r, r+1, r+2, \cdots) \ ,$$

称随机变量 X 服从 **参数为** r **和** p **的负二项分布** 或 **帕斯卡分布**. 可以验证上述概率构成一个分布列, 以及随机变量 X 的期望 $E(X)=r/p$ 和方差 $\mathrm{Var}(X) = r(1-p)/p^2$. 相关证明将作为练习题.

3.5 案例分析

3.5.1 德国坦克问题

在二战期间, 同盟国一直在努力确定德国坦克的生产数量, 有助于对德国战力的评估. 这个问题可描述为: 德国生产了 n 辆坦克, 编号分别为 $1, 2, \cdots, n$. 盟军在战斗中任意击毁了 k 辆坦克, 被击毁的坦克编号分别为 $x_1, x_2, \ldots, x_k$, 能否通过被击毁的坦克编号来估计 n 的大小, 即估计德国生产了多少辆坦克.

在没有其它信息的情况下, 不妨假设被随机击毁的坦克是等可能事件, 即第 i 辆坦克被击毁的概率为 $1/n$. 可以将问题看作从集合 $\{1, 2, \cdots, n\}$ 中不放回随机抽取 k 个数, 用 X 表示抽到的 k 个数中的最大数. 则 X 的取值为 $\{k, k+1, \cdots, n\}$ 以及概率

$$P(X=i) = \binom{i-1}{k-1} \Big/ \binom{n}{k} \qquad (i = k, k+1, \cdots, n) \ .$$

于是得到

$$E(X) = \binom{n}{k}^{-1} \sum_{i=k}^{n} \binom{i-1}{k-1} i \ .$$

针对上面的求和表达式, 可以考虑从 $n+1$ 个元素中选取 $k+1$ 个元素, 共有 $\binom{n+1}{k+1}$ 种不同的方法. 将这些不同的方法分情况讨论, 按照选取的 $k+1$ 个元素中最大元素 $i = k+1, k+2, \cdots, n+1$ 进行分类; 若最大元素为 i, 则有 $\binom{i-1}{k}$ 种不同的方法. 于是有

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

代入期望 $E(X)$ 可得

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

由于仅做了一次观察, 将观察中 k 个数的最大值近似期望 $E[X]$, 即 $E(X) \approx \max(x_1, x_2, \cdots, x_n)$, 由此估计

$$n \approx \max(x_1, x_2, \cdots, x_n)\left(1 + \frac{1}{k}\right) - 1,$$

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

$$n \approx 212 \times (1 + 1/6) - 1 = 246.$$

针对德国坦克数量的实际估计情况见下表, 可以发现利用上述所提的统计估计方法接近德国的实际产量, 比英国的情报估计准确得多.

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

3.5.2 集卡活动

很多小朋友喜欢参加各种集卡活动, 如奥特曼卡和叶罗丽卡等. 事实上很多成年人也对集卡游戏并不陌生, 例如 80 年代的葫芦娃洋画、或 90 年代的小虎队旋风卡等. 问题可以描述为: 市场上有 n 种不同类型的卡片, 假设一个小朋友每次都能以等可能概率、独立地收集一张卡片, 问一个小朋友在平均情况下至少要收集多少张卡才能收集齐 n 种不同类型的卡片.

这里先补充一个需要用到的引理, 后面将给出详细的证明:

引理 3.5 对任意的随机变量 $X_1, X_2, \cdots, X_n$ 有

$$E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n) \ .$$

用 X 表示收集齐 n 种不同类型的卡片所需要的收集次数, 用 X_k 表示收集齐 $k-1$ 种和 k 种不同类型卡片之间所需要的收集次数 ($k \in [n]$), 于是有 $X = X_1 + X_2 + \cdots + X_n$. 我们的问题是计算期望 $E(X)$.

很容易发现随机变量 X_k 服从参数为 p_k 的几何分布. 当已经收集到 $k-1$ 种不同类型的卡片时, 再获得一张新卡的概率

$$p_k = 1 - (k-1)/n \ .$$

根据几何分布的性质有 $E[X_k] = 1/p_k = n/(n-k+1)$. 利用引理 3.5 有

$$E(X) = E\left(\sum_{k=1} X_k\right) = \sum_{k=1} E(X_k) = \sum_{k=1}^{n} \frac{n}{n-k+1} = n\sum_{k=1}^{n} \frac{1}{k} = nH(n)\ ,$$

这里 $H(n)$ 表示参数为 n 的调和数, 即 $H(n) = \sum_{k=1}^{n} 1/k$. 关于调和数有

引理 3.6 针对调和数 $H(n) = \sum_{k=1}^{n} 1/k$ 有 $H(n) \in [\ln(n+1), 1+\ln(n)]$.

证明 因为函数 $1/x$ 在 $x \in (0, +\infty)$ 单调递减, 有

$$\ln(n+1) = \int_{x=1}^{n+1} \frac{1}{x} dx \leqslant \sum_{k=1}^{n} \frac{1}{k} = 1 + \sum_{k=2}^{n} \frac{1}{k} \leqslant 1 + \int_{x=1}^{n} \frac{1}{x} dx = 1 + \ln(n).$$

最后得到 $n\ln(n+1) \leqslant E(X) \leqslant n + n\ln n$.