性质3.5. 对随机变量 X 和常数 $a, b \in \mathbb{R}$, 有

$$Var(aX+b) = a^2\, Var(X).$$

Proof. 根据期望的性质有 $E(aX+b) = aE(X)+b$, 代入可得

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

□

一般情况下方差不具有线性性, 即 $\mathrm{Var}[f(X)+g(X)] \neq \mathrm{Var}[f(X)] + \mathrm{Var}[g(X)]$.

性质3.6. 对随机变量 X 和常数 $a \in \mathbb{R}$, 有

$$Var(X) = E(X-E(X))^2 \leq E(X-a)^2.$$

Proof. 我们有

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

从而完成证明. □

定理3.3 (Bhatia-Davis不等式)**.** 对随机变量 $X \in [a, b]$, 有

$$Var[X] \leq (b-E(X))(E(X)-a) \leq (b-a)^2/4.$$

Proof. 对任意随机变量 $X \in [a, b]$, 有 $(b-X)(X-a) \geq 0$, 两边同时对随机变量取期望, 整理可得

$$E(X^2) \leq (a+b)E(X) - ab.$$

根据方差的定义有

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

进一步对二次函数 $f(t) = -t^2 + (a+b)t - ab$ 求最大值, 可得 $(b-E(X))(E(X)-a) \leq (b-a)^2/4$. □

3.3 常用的离散型随机变量

本节介绍几种常用的离散型随机变量, 并研究其数字特征.

3.3.1 离散型均匀分布

定义3.4. 设随机变量 X 的取值为 $x_1, x_2, \cdots, x_n$, 且 $P(X = x_i) = 1/n$, 称 X 服从离散型均匀分布.

由定义可知

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

下面来看一个离散型均匀分布的例子.

例3.8 (德国坦克问题)**.** 假设德国生产了 N 辆坦克, 编号分别为 $1, 2, \cdots, N$, 盟军战斗中随机击毁了 k 辆, 被随机击毁坦克编号分别为 $x_1, x_2, \ldots, x_k$, 如何估计 N 的大小.

解. 对题意进行分析, 坦克被随机击毁可看作坦克被击毁的服从离散型均匀分布, 即 $P[X = i] = 1/N$ $(i \in [N])$, 因此有 $E(X) = (1 + N)/2$, 可用被击毁坦克编号的平均值去近似, 即

$$E(X) = \frac{1+N}{2} \approx \frac{1}{k}\sum_{i=1}^{k} x_i \qquad \Longrightarrow \qquad N \approx \frac{2}{k}\sum_{i=1}^{k} x_i - 1.$$

若 k 较大时, 上述表达式对 N 有一个较好的估计; 但 k 较小时, 往往更关注于被毁坦克的最大编号 m.

方法二 问题转化为从 $\{1, 2, \cdots, N\}$ 中以不放回随机抽取 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. 假设随机变量 X 表示抽取 k 个数中的最大数, 其分布列为

$$Pr[X = i] = \frac{\binom{i-1}{k-1}}{\binom{N}{k}} \quad (k \leq i \leq N).$$

由此得到

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

为计算期望 $E(X)$, 考虑从 $N + 1$ 个元素中随机取 $k + 1$ 个元素, 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

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

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

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

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 $N = 212 \times (1 + 1/6) - 1 \approx 246$. 针对二战德国坦克数量的实际估计情况可参见下表, 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3.5. *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

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

$0 - 1$分布是很多概率模型的基础.

3.3.3 二项分布

Bernoulli 试验只有两个结果: A 和 $\bar{A}$, 设 $P(A) = p$ $(p \in [0, 1])$, 因此 $P(\bar{A}) = 1 - p$. 将 Bernoulli 试验独立重复地进行 n 次, 称 n 重 Bernoulli 试验, 是一种非常重要的概率模型, 具有广泛的应用.

定义3.6. *用随机变量 X 表示 n 重 Bernoulli 试验试验中事件 A 发生的次数, 则 X 的分布列为*

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad k \in \{0, 1, 2, \ldots, n\}$$

称随机变量 X 服从二项分布 (binomial distribution), 记 $X \sim B(n, p)$.

我们称之为二项分布是因为与二项展开式 $(a + b)^n = \sum_{k=1}^n \binom{n}{k} a^k b^{n-k}$ 相似, 容易验证

$$P(X = k) \geq 0 \qquad \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = (p + 1 - p)^n = 1.$$

对二项分布, 有

性质3.7. *对随机变量 $X \sim B(n, p)$ 有*

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

Proof. 由定义可知

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

为计算 $E(X)$, 考虑二项展开式

$$(1+x)^n=\sum_{k=0}^{n}\binom{n}{k}x^k \;\Rightarrow\; n(1+x)^{n-1}=\sum_{k=1}^{n}\binom{n}{k}kx^{k-1} \;\Rightarrow\; nx(1+x)^{n-1}=\sum_{k=1}^{n}\binom{n}{k}kx^k,$$

将 $x=p/(1-p)$ 带入可得

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

对于方差, 首先计算

$$\begin{aligned} E(X^2) &= \sum_{k=0}^{n}k^2\binom{n}{k}p^k(1-p)^{n-k} \\ &= \sum_{k=1}^{n}k(k-1)\binom{n}{k}p^k(1-p)^{n-k}+np \\ &= (1-p)^n\sum_{k=2}^{n}k(k-1)\binom{n}{k}\left(\frac{p}{1-p}\right)^k+np \end{aligned}$$

对二项展开式 $(1+x)^n=\sum_{k=0}^{n}\binom{n}{k}x^k$ 两边同时求导两次可得

$$n(n-1)(1+x)^{n-2}=\sum_{k=2}^{n}\binom{n}{k}k(k-1)x^{k-2} \;\Rightarrow\; n(n-1)x^2(1+x)^{n-2}=\sum_{k=2}^{n}\binom{n}{k}k(k-1)x^k,$$

将 $x=p/(1-p)$ 带入有

$$E(X^2)=n(n-1)p^2+np=n^2p^2+np(1-p),$$

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

例3.9. *有 5 个选择题, 每个选择题有 4 种答案, 只有一种正确, 求一学生随机猜对 4 个选择题的概率?*

解. 将每一个选择题看作一次 Bernoulli 试验, 事件 A 表示猜正确, 则有 $P(A)=1/4$. 整个问题等价于 5 重 Bernoulli 试验, 用 X 表示学生猜对题的个数, 则 $X\sim B(5,1/4)$, 从而得到

$$P(X=4)=\binom{5}{4}\frac{1}{4^4}\frac{3}{4}=\frac{15}{4^5}.$$

□

3.3.4 几何分布

定义3.7. *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, 则 X 的取值为 $1,2,\cdots$, 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, 记为 $X\sim G(p)$.*

首先可知 $P(X=k)=(1-p)^{k-1}p\geq 0$ 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. 对几何分布, 我们有

性质3.8. 若随机变量 $X\sim G(p)$ $(0<p<1)$, 则有

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

Proof. 根据期望的定义有

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

对级数展开式 $(1-x)^{-1}=\sum_{k=0}^{\infty}x^k$ 两边求导有

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

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

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

对级数展开式 $(1-x)^{-1}=\sum_{k=0}^{\infty}x^k$ 两边求二阶导有

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

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

下面给出几何分布的一个重要性质: 无记忆性 (memoryless property).

定理3.4. 设随机变量 $X\sim G(p)$, 对任意正整数 m,n, 有

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

直观理解几何分布的无记忆性: 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. 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 $\{X>m+n\}\cap\{X>m\}=\{X>m+n\}$. □

例3.10. *古人非常重视生男孩且资源有限, 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. 多年后男女是否失衡?*

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 $X = 1, 2, \cdots$, 以及

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

这里 $p = 1/2$ 表示生男孩的概率. 根据几何分布的期望有一个家庭小孩数的期望为 $E[X] = 1/p = 2$, 由此可得一个家庭的小孩男女比例 $1 : 1$. □

3.3.5 泊松分布

定义3.8. *若随机变量 X 的分布列为*

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

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: 在一段时间内电话收到的呼叫次数, 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等.

性质3.9. *对任意给定的 $\lambda > 0$, 若 $X \sim P(\lambda)$, 则*

$$E(X) = \lambda \qquad \text{和} \qquad Var(X) = \lambda.$$

Proof. 根据期望的定义和幂级数 $e^\lambda = \sum_{k=0}^{\infty} \lambda^k/k!$ 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

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

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 $P(X=k)=\lambda^k e^{-\lambda}/k!$ 和 $P(X=1)=P(X=2)$ 可得

$$\lambda e^{-\lambda}=\lambda^2 e^{-\lambda}/2 \quad \Longrightarrow \quad \lambda=2,$$

进一步得到

$$\begin{aligned} P\{X\geq 4\} &= 1-P(X=0)-P(X=1)-P(X=2)-P(X=3) \\ &= 1-e^{-2}-2e^{-2}-\frac{4}{3}e^{-2}=1-\frac{13}{3e^2}. \end{aligned}$$

□

下面研究二项分布和泊松分布的关系, 即泊松定理:

定理3.5. 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 $n\to\infty$ 时有 $(1-\frac{\lambda}{n})^{\frac{n}{\lambda}}\to e$ 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.

例3.12. 射击训练每次命中目标的概率为 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 射击训练中射中目标的次数, 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由一人处理, 考虑方案: I) 由四人维护, 每人单独负责 20 台; II) 由三人共同维护 80 台. 哪种方案更为可取?*

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

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

因四人独立维修, 有设备发生故障时而不能及时的概率

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

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 $X \sim P(0.8)$, 则有设备发生故障不能及时维修的概率为

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

由此比较可知方案 II) 更优. □