3.3 离散型随机变量的方差

数学期望反映了随机变量的平均值, 在很多实际应用中不仅需要知道随机变量的平均值, 还需要了解随机变量取值的偏离程度. 例如, 考虑三个随机变量 X, Y 和 Z, 其分布列分别为

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

容易得到 $E[X]=E[Y]=E[Z]=0$, 即三个随机变量期望相同, 但它们之间存在着明显的差异. 如何刻画它们的不同之处, 可以考虑三个随机变量的取值与期望的偏离程度, 即方差.

定义 3.5 设离散随机变量 X 的分布列为 $p_k=P(X=x_k)>0$, 若期望 $E[X]$ 和 $E[X-E[X]]^2$ 存在, 则称 $E[X-E[X]]^2$ 为随机变量 X 的 **方差** (variance), 记为 $\operatorname{Var}(X)$, 即

$$\operatorname{Var}(X)=E[X-E[X]]^2=\sum_k p_k(x_k-E[X])^2=\sum_k p_k\left(x_k-\sum_k x_k p_k\right)^2. \tag{3.2}$$

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

结合方差的定义和期望的性质有

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

由此给出方差的另一种定义

$$\operatorname{Var}(X)=E[X^2]-(E[X])^2. \tag{3.3}$$

尽管方差的两种定义等价, 然而在实际应用中却存在着不同的用处, (3.2) 给出了方差的物理含义, 而 (3.3) 更有利于方差的计算, 例如,

例 3.6 设随机变量 X 的分布列为 $P(X=x_k)=1/n$ $(k\in[n])$, 需要遍历数据 $x_1,x_2,\cdots,x_n$ 几遍才能计算方差 $\operatorname{Var}(X)$.

解 采用计算公式 $\operatorname{Var}(X)=E[X-E[X]]^2$, 则需要遍历数据 $x_1,x_2,\cdots,x_n$ 两遍, 第一遍计算期望 $E[X]$, 第二遍计算方差 $\operatorname{Var}(X)$.

采用计算公式 $\operatorname{Var}(X)=E[X^2]-(E[X])^2$, 则只需要遍历数据 $x_1,x_2,\cdots,x_n$ 一遍, 在遍历数据的过程中计算 $E[X^2]$ 和 $(E[X])^2$, 从而计算方差. 在此过程中也不需要将全部数据 $x_1,x_2,\cdots,x_n$ 存在内存中, 可以一个个轮流存取数据, 即机器学习中的在线学习.

下面给出方差的一些性质:

性质 3.6 设 $c\in\mathbb{R}$ 是常数, 若随机变量 $X\equiv c$, 则 $\operatorname{Var}(X)=0$.

性质 3.7 对随机变量 X 和常数 $a, b \in \mathbb{R}$, 有

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

证明 根据期望的性质有 $E[aX+b] = aE[X]+b$, 代入可得

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

值得注意的是, 方差通常不具有线性性, 即 $\operatorname{Var}(f(X)+g(X)) \neq \operatorname{Var}(f(X)) + \operatorname{Var}(g(X))$.

性质 3.8 对随机变量 X 和常数 $c \in \mathbb{R}$, 有

$$\operatorname{Var}(X) = E[X-E[X]]^2 \leqslant E[X-c]^2.$$

证明 根据期望的性质有

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

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

$$\operatorname{Var}(X) \leqslant (b-E[X])(E[X]-a) \leqslant (b-a)^2/4.$$

证明 对任意随机变量 $X \in [a,b]$, 有

$$(b-X)(X-a) \geqslant 0,$$

两边同时对随机变量取期望, 整理可得 $E[X^2] \leqslant (a+b)E[X]-ab$. 根据方差的定义有

$$\operatorname{Var}(X) = E[X^2] - (E[X])^2 \leqslant -(E[X])^2 + (a+b)E[X] - ab = (b-E[X])(E[X]-a).$$

利用二次函数 $f(t) = (b-t)(t-a) = -t^2 + (a+b)t - ab$ 的最大值可得

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

3.4 常用离散型随机变量

本节介绍几种常用的离散型随机变量, 并研究其性质.

3.4.1 0-1分布

0-1分布是概率统计中最经典、最简单的分布, 是很多概率模型的基础.

定义 3.6 设随机变量 X 的分布列 $P(X=1)=p$, $P(X=0)=1-p$, 等价于

$$P(X=k)=p^k(1-p)^{1-k} \qquad k=0,1,$$

则称随机变量 X 服从 **参数为 p 的 0-1 分布**, 又称 **两点分布**, 或 **伯努利分布** (Bernoulli distribution), 记 $X\sim \mathrm{Ber}(p)$. 0-1 分布也可以用表格表示为

X	0	1
P	$1-p$	p

根据上述定义可得

性质 3.9 若随机变量 $X\sim \mathrm{Ber}(p)$, 则有 $E[X]=p$ 和 $\mathrm{Var}(X)=p(1-p)$.

由此可知 0-1 分布也可由它的数学期望唯一确定.

若一次试验只考虑事件 A 发生或不发生两种情况, 称这样的试验为 **伯努利试验**, 可以通过 0-1 分布来描述伯努利试验:

$$X=\begin{cases}1 & \text{若事件 } A \text{ 发生},\\ 0 & \text{否则}.\end{cases}$$

此时容易得到 $E[X]=P(A)$, 即随机变量 X 的期望等于事件 A 发生的概率.

3.4.2 二项分布

伯努利试验考虑事件 A 发生或不发生, 设事件 A 发生的概率 $P(A)=p\in(0,1)$. 将伯努利试验独立重复地进行 n 次, 称这一系列独立重复的试验为 n **重伯努利试验**.

在 n 重伯努利试验中, 用 X 表示事件 A 发生了多少次, 其可能的取值为 $0,1,2,\cdots,n$. 事件 $\{X=k\}$ 表示在 n 重伯努利试验中事件 A 发生了 k 次, 到底是哪 k 次发生, 共有 $\binom{n}{k}$ 种不同的情况. 针对一种具体的情况, 不妨设前 k 次事件 A 发生, 后 $n-k$ 次事件 A 不发生, 此时发生的概率为

$$\underbrace{p\times p\times\cdots\times p}_{k\text{ 个}}\times\underbrace{(1-p)\times(1-p)\times\cdots\times(1-p)}_{n-k\text{ 个}}=p^k(1-p)^{n-p}.$$

由此可知在 n 重伯努利试验中事件 A 发生了 k 次的概率为 $P(X=k)=\binom{n}{k}p^k(1-p)^{n-p}$.

定义 3.7 若随机变量 X 的分布列为

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}\qquad k=0,1,2,\ldots,n, \tag{3.4}$$

则称随机变量 X 服从 **参数为 n 和 p 的二项分布** (binomial distribution), 记 $X\sim B(n,p)$.

容易发现 (3.4) 中 $P(X=k)$ 是二项式 $(1-p+xp)^n$ 展开式中 x^k 项的系数, 该分布被称为二项分布. 进一步可检验

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

若 $n=1$, 则二项分布退化为 0-1 分布, 即 $B(1,p)=\mathrm{Ber}(p)$. 关于二项分布的数字特征有

性质 3.10 若随机变量 $X\sim B(n,p)$, 则有 $E[X]=np$ 和 $\mathrm{Var}(X)=np(1-p)$.

若知道二项分布的期望和方差, 可反解出参数 n 和 p, 因此二项分布可由它的期望和方差唯一确定.

证明 根据定义有

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

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

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

将 $x=p/(1-p)$ 带入上式可得

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

对于方差, 首先计算

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

对二项展开式 $(1+x)^n=\sum_{k=0}^n\binom{n}{k}x^k$ 两边同时求导两次后乘 x^2 可得

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

将 $x = p/(1-p)$ 带入上式有

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

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

下面给出几个二项分布的概率分布示意图. 若随机变量 $X \sim B(n,p)$, 则概率 $P(X=k)$ 开始会随 k 的增加而增大, 一般在期望 np 附近的整数点取得最大值, 然后会随 k 的增加而减小.

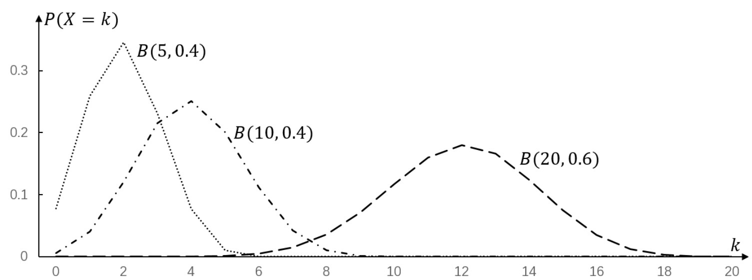


图 3.1 二项分布的概率分布示意图

例 3.7 两个盒子中分别放了 n 个球, 每次任选一个盒子并拿走一球, 重复这一过程, 求一个盒子中的球拿光且另一个盒子还剩下 r 个球的概率.

解 两盒子表示为第一个盒子和第二个盒子, 考虑伯努利试验: 事件 A 表示从第一个盒子拿走一球, 根据题意有 $P(A)=1/2$, 共进行了 $2n-r$ 重伯努利试验, 用 X 表示事件 A 发生的次数, 则有

$$X \sim B(2n-r, 1/2).$$

所求概率为

$$P(X=n) + P(X=n-r) = \binom{2n-r}{n}\frac{1}{2^{n-r}} + \binom{2n-r}{n-r}\frac{1}{2^{n-r}} = \binom{2n-r}{n}\frac{2}{2^{n-r}}\ ,$$

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 表示由 n 个元件构成的系统中能正常工作的元件数, 则有 $X \sim B(n,p)$. 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:

性质 3.11 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.

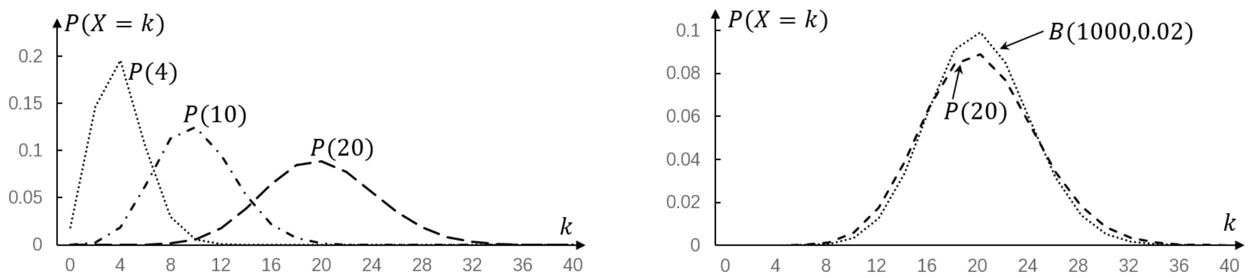
证明 根据期望的定义有

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

这里利用了指数的泰勒展开式 $e^x=\sum_{k=0}^{\infty}x^k/k!$. 对于随机变量的方差, 首先计算

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

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

图 3.2(a) 给出了几个泊松分布的概率分布示意图. 若随机变量 $X \sim P(\lambda)$, 则概率 $P(X=k)$ 开始会随 k 的增加而增大, 一般在期望 λ 附近的整数点取得最大值, 然后会随 k 的增加而减小.

如图 3.2(b) 所示, 泊松分布与二项分布的分布图之间有一定的相似性, 有如下定理:

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?