2.2 分布列

离散随机变量的取值概率是随机变量的最重要的特征. 我们用**分布列**表示这种特征,并且用 p_X 表示随机变量 X 的分布列. 设 x 是随机变量 X 的取值,则 X 取值为 x 的概率定义为事件 $\{X=x\}$ 的概率,即所有与 x 对应的试验结果所组成的事件的概率,用 $p_X(x)$ 表示,即

$$p_X(x)=\mathrm{P}(\{X=x\}).$$

例如, 在将一枚均匀的硬币独立地抛掷两次的试验中,令 X 为正面向上的次数. 则 X 的分布列由下式给出

$$p_X(x)=\begin{cases}1/4, & 若\ x=0\ 或\ x=2,\\ 1/2, & 若\ x=1,\\ 0, & 其他.\end{cases}$$

今后在不引起混淆的情况下, 我们将省去表示事件或集合的花括号.例如用 $\mathrm{P}(X=x)$ 表示事件 $\{X=x\}$ 的概率,尽管记号 $\mathrm{P}(\{X=x\})$ 比较确切一些.同时我们也会遵守下面的传统: **我们用大写字母表示随机变量, 用小写字母表示实数,例如随机变量的取值等.**

对于分布列, 我们有

$$\sum_x p_X(x)=1,$$

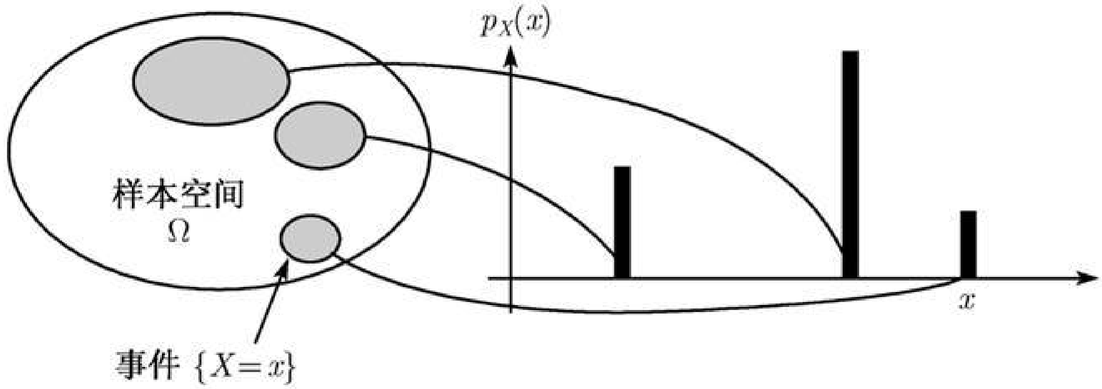
其中求和是对随机变量 X 的一切可能的取值而求的.上式之所以成立是由于概率的可加性和归一性公理. 对于不同的 x,事件 $\{X=x\}$ 是互不相容的, 并且对所有的 x,事件系列 $\{X=x\}$ 形成了样本空间的一个分割. 利用类似的原理可以证明,对于任意一个 X 的可能值的集合 S, 下式成立:

$$\mathrm{P}(X\in S)=\sum_{x\in S}p_X(x).$$

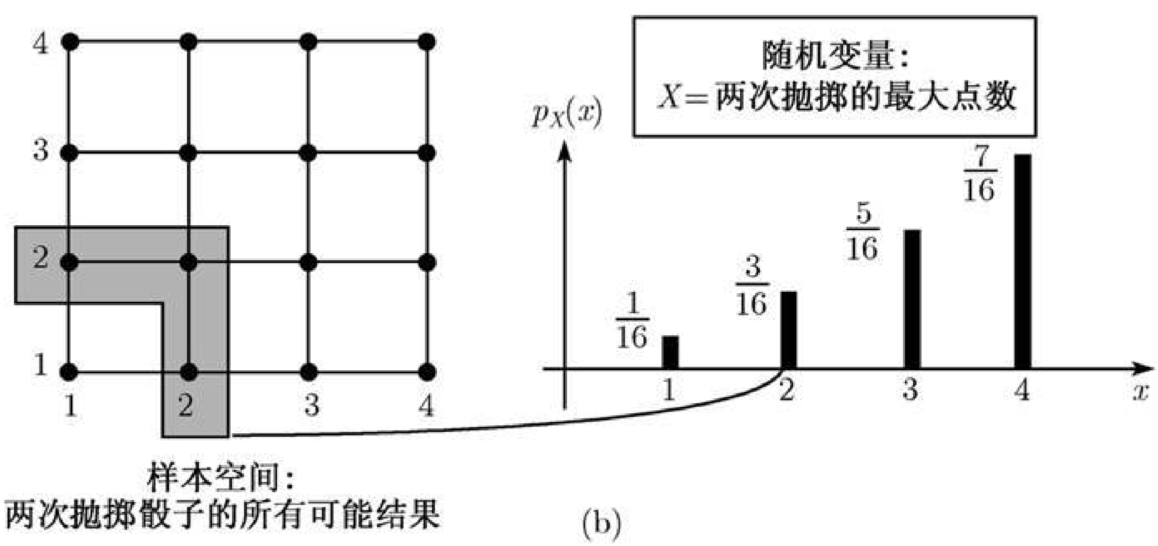
例如, 在将一枚均匀的硬币独立地抛掷两次的试验中,至少一次正面向上的概率为

$$\mathrm{P}(X>0)=\sum_{x=1}^{2}p_X(x)=\frac{1}{2}+\frac{1}{4}=\frac{3}{4}.$$

分布列的计算, 在概念上是很简单的, 图2.2给出了很直观的解释.



(a)



(b)

图 2.2 (a)随机变量 X 的分布列计算方法的图像化表示.对每一个 X 的可能值 x ，找出使 $X=x$ 的所有试验结果,将它们的概率相加得到 $p_X(x)$.(b)设所涉及的试验是抛掷一个具有4个面的均匀骰子，独立地抛掷两次.所涉及的随机变量为 X=两次转动所得到的最大点数.X 的可能值为1,2,3,4. 对于给定的 x 的值，为计算 $p_X(x)$ 的值,将 X 取值为 x 的所有试验结果的概率相加，得到 $p_X(x)$ 的值. 例如,有三个试验结果((1,2),(2,2),(2,1))的 X 的值为2,而每一个试验结果的概率为1/16，故 $p_X(2)=3/16$

随机变量 X 的分布列的计算

对每一个随机变量 X 的值 x：.

(1) 找出与事件$\{X=x\}$相对应的所有试验结果;

(2) 将相应的试验结果的概率相加得到 $p_X(x)$.

2.2.1　伯努利随机变量

考虑抛掷一枚硬币，设正面向上的概率为 p，反面向上的概率为$1-p$. **伯努利随机变量**在试验结果为正面向上时取值为1,在试验结果为反面向上时取值为0，即

$$X = \begin{cases} 1, & \text{若正面向上}, \\ 0, & \text{若反面向上}. \end{cases}$$

它的分布列为

$$p_X(k) = \begin{cases} p, & \text{若 } k=1, \\ 1-p, & \text{若 } k=0. \end{cases}$$

由于伯努利随机变量非常简洁，因此它也是非常重要的随机变量.在实际中它用于刻画具有两个试验结果的概率模型. 例如：

(a) 在给定的时刻，一架电话机可处于待机状态或使用状态;

(b) 一个人可以处于健康状态或患有某种疾病状态;

(c) 作为一个人的政治态度，他可以赞成或反对某个候选人.

进一步,我们可以将多个伯努利随机变量综合成更加复杂的随机变量.下面我们要讨论的二项随机变量就是其中之一.

2.2.2　二项随机变量

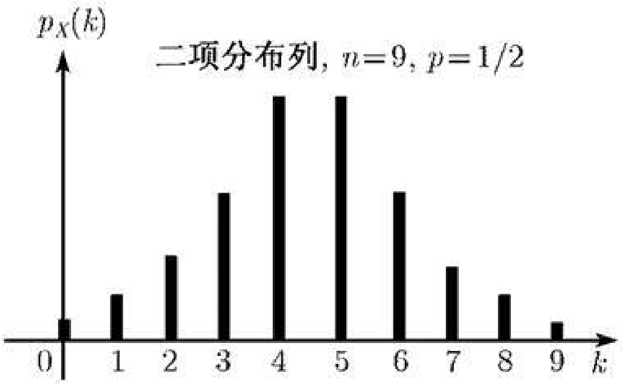
将一枚硬币抛掷 n 次，每次抛掷，正面出现的概率为 p,反面出现的概率为 $1-p$，而且各次抛掷是相互独立的.令 X 为 n 次抛掷得到正面的次数. 我们称 X 为**二项随机变量**，其参数为 n 和 p . X 的分布列就是在1.5节中讨论的二项概率：

$$p_X(k) = \mathrm{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \cdots, n.$$

(按照传统，我们用 k 代替x ，表示整数值随机变量 X 的取值.)对于二项随机变量,利用归一化公理可以得到

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

在图2.3中，用图像表示某些特殊情况的二项分布列.



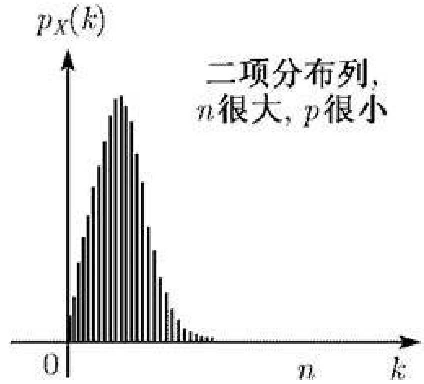


图 2.3　二项随机变量的分布列. 当 p=1/2时, 分布列是相对于 n/2对称的.当 $p < 1/2$ 时, 相应的分布偏向0, 当 $p > 1/2$ 时,相应的分布偏向 n

2.2.3　几何随机变量

在连续抛掷硬币的试验中, 每次抛掷, 正面出现的概率为 p ,反面出现的概率为 1-p , 而且各次抛掷是相互独立的.令 X 为连续地抛掷一枚硬币, 直到第一次出现正面所需要抛掷的次数. X 就称为**几何随机变量**.前 k-1 次抛掷的结果为反面向上, 第 k 次抛掷的结果为正面向上的概率为 $(1-p)^{k-1}p$. 因此 X 的分布列为

$$p_X(k) = (1-p)^{k-1}p, \quad k = 1, 2, \cdots .$$

几何随机变量的分布列的图像可见图2.4. 从

$$\sum_{k=1}^{\infty} p_X(k) = \sum_{k=1}^{\infty}(1-p)^{k-1}p = p\sum_{k=0}^{\infty}(1-p)^k = p\frac{1}{1-(1-p)} = 1$$

可知这是合格的分布列.

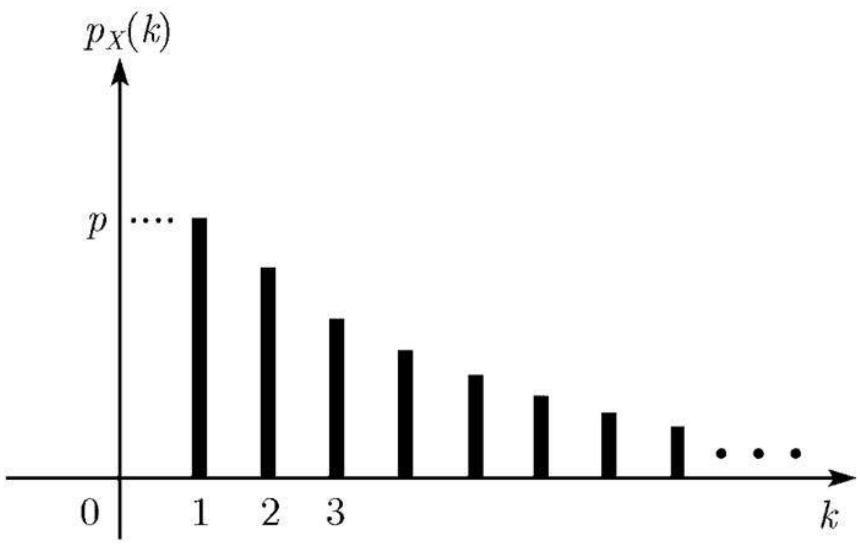


图 2.4　几何随机变量的分布列. $p_X(k)=(1-p)^{k-1}p(k=1,2\cdots)$ 是几何级数，递减的因子为1-p

此处，利用抛掷硬币的试验恰巧是抓住了事物的本质. 更一般地,连续抛掷硬币的试验序列中出现正面可以解释为独立试验序列中的一次试验“成功”,这样几何随机变量可以解释为独立试验序列中直到试验第一次“成功”所需的试验次数.而试验“成功”的意义是随着所讨论的问题的实际背景而变化的.例如可以是在某次测验中通过了考试,在某次搜索中发现目标,或成功地进入计算机系统等.

2.2.4　泊松随机变量

设随机变量 X 的分布列由下式给出

$$p_X(k)=\mathrm{e}^{-\lambda}\frac{\lambda^k}{k!},\quad k=0,1,2,\cdots,$$

其中 λ 是刻画分布列的取正值的参数,则称 X 是泊松随机变量(见图2.5). 由于

$$\sum_{k=0}^{\infty}\mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}=\mathrm{e}^{-\lambda}\left(1+\lambda+\frac{\lambda^2}{2!}+\cdots\right)=\mathrm{e}^{-\lambda}\mathrm{e}^{\lambda}=1,$$

这个数列符合分布列的定义.[1]

[1]若这个总和不等于1,就与概率的归一化定律相冲突. ——译者注

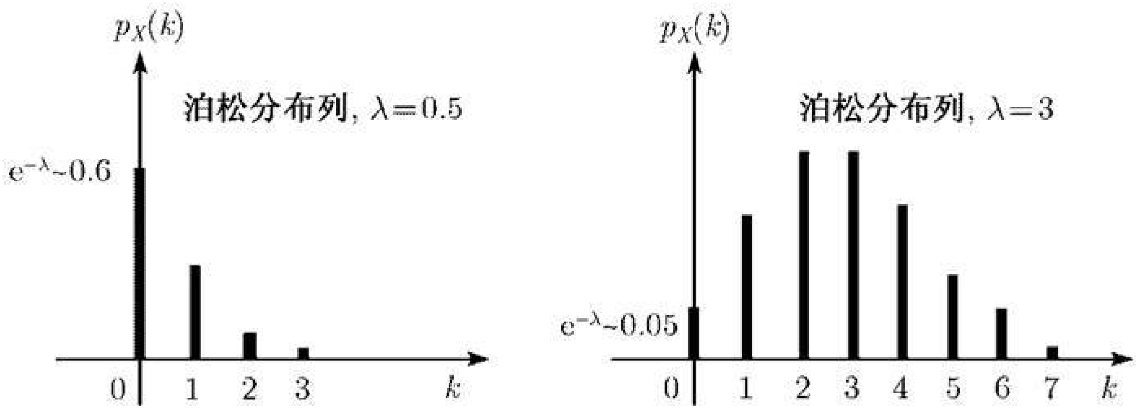


图 2.5　对应于不同的 λ 的泊松随机变量的分布列. 当 $\lambda\le 1$ 时,分布列是单调递减的. 当 $\lambda>1$ 时，分布列随着 k 的递增,先递增后递减(可参考本章末尾的习题)

为了给出泊松随机变量的直观印象，考虑当二项随机变量的参数 n 很大，p 很小的情况. 例如，令 X 为字数为 n 的一本书中含有打印错误的字数. 这样，X 是二项随机变量. 但是,由于一个字被打印错误的概率 p 非常小，X 也可以用泊松分布列刻画(打错一 个字相当于抛掷一枚硬币出现正面向上,但正面向上的概率 p 很小). 类似的例子很多,例如在一个城市中一天中发生车祸的事故数.[2]

[2]普遍认为,第一个关于二项随机变量和泊松随机变量之间联系的实证例子,是在19世纪后半叶用泊松分布列去逼近波兰骑兵被马踢伤的人数.

用泊松随机变量刻画这样的现象十分恰当. 更确切地说,参数为 λ 的泊松随机变量的分布列是二项随机变量分布列的很好的逼近：

$$\mathrm{e}^{-\lambda}\frac{\lambda^k}{k!} \approx \frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}, \quad k=0,1,\cdots,n,$$

其中 $\lambda = np$, n 很大, p 很小. 在这种情况下,泊松分布列使得模型简单, 计算方便. 例如, $n=100$, $p=0.01$,用二项随机变量计算成功次数 k=5的概率为

$$\frac{100!}{95!5!}\cdot 0.01^5(1-0.01)^{95} = 0.002\ 90.$$

利用泊松随机变量计算这个概率得到近似值

$$\mathrm{e}^{-1}\frac{1}{5!} = 0.003\ 06,$$

其中$\lambda = np = 100\cdot 0.01 = 1$.

在本章最后的习题中, 我们将给出泊松逼近的严格证明.第6章将作进一步解释和推广,并且将结果用 到泊松过程中去.

2.3 随机变量的函数

设 X 是一个随机变量. 对 X 施行不同的变换，可以得到其他的随机变量.作为例子，用 X 表示今天的气温(单位为摄氏度，°C).作变换 $Y=1.8X+32$，得到华氏温度的读数(°F).在这个例子中 Y 是 X 的**线性**函数

$$Y=g(X)=aX+b,$$

其中 a 和 b 是数值. 我们也可以考虑 X 的非线性函数

$$Y=g(X).$$

例如可以考虑对数度量，此时可用变换 $g(X)=\log(X)$.

设 $Y=g(X)$ 是随机变量 X 的函数，由于对每一个试验结果,也对应一个(Y 的)数值，故 Y 本身也是一个随机变量.如果 X 是离散的随机变量，其对应的分布列为 p_X，则 Y 也是离散随机变量，其分布列可通过 X 的分布列进行计算. 实际上,对固定的 y 值，$p_Y(y)$ 的值可以通过下式计算

$$p_Y(y)=\sum_{\{x|g(x)=y\}}p_X(x).$$

例 2.1 可以利用上述公式计算 $Y=|X|$ 的分布列，其中 X 的分布列由下式给出，

$$p_X(x)=\begin{cases}1/9, & \text{若 } x \text{ 是 } [-4,4] \text{ 中的整数},\\ 0, & \text{其他}.\end{cases}$$

由于 Y 的值域为 $y=0,1,2,3,4$，对于值域中的任意 y,只需将满足 $|x|=y$ 的所有 $p_X(x)$ 的值相加，就可以得到 $p_Y(y)$ 的值.当 y=0的时候，只有 x=0能够满足条件 $y=|0|=0$. 这样

$$p_Y(0)=p_X(0)=\frac{1}{9}.$$

对于 $y=1,2,3,4$，有两个 x 值满足条件 $y=|x|$.例如(见图2.6的图示说明)

$$p_Y(1)=p_X(-1)+p_X(1)=\frac{2}{9}.$$

这样，Y 的分布列为

$$p_Y(y) = \begin{cases} 2/9, & \text{若 } y = 1,2,3,4, \\ 1/9, & \text{若 } y = 0, \\ 0, & \text{其他.} \end{cases}$$

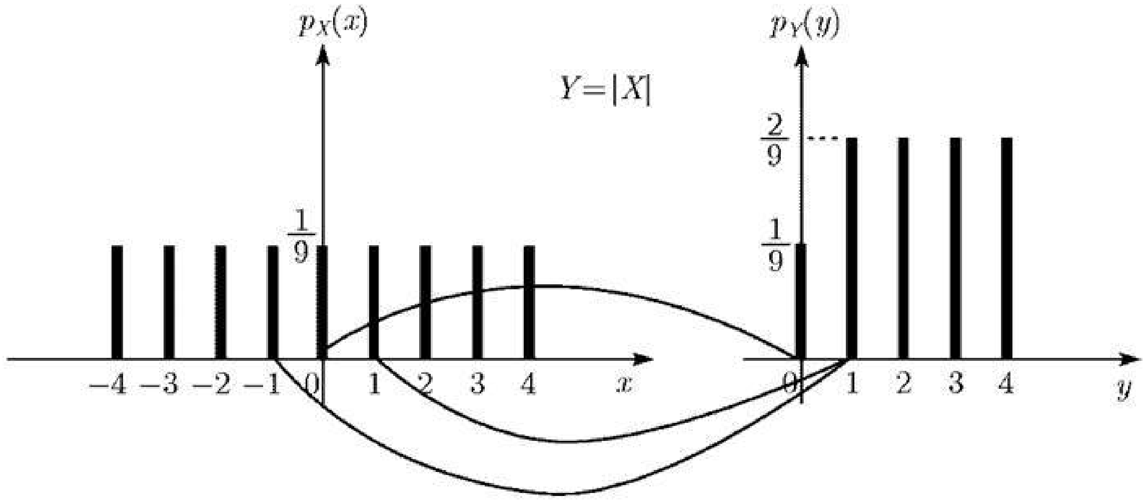


图 2.6 例2.1中 X 和 $Y = |X|$ 的分布列

现在看另一个随机变量 $Z = X^2$. 为了求得 Z 的分布列,我们既可以将它看成 X 的平方，也可以看成 $Y = |X|$ 的平方. 利用公式 $p_Z(z) = \sum_{\{x|x^2=z\}} p_X(x)$ 或 $p_Z(z) = \sum_{\{y|y^2=z\}} p_Y(y)$ ，得到

$$p_Z(z) = \begin{cases} 2/9, & \text{若 } z = 1,4,9,16, \\ 1/9, & \text{若 } z = 0, \\ 0, & \text{其他.} \end{cases}$$

2.4 期望、均值和方差

X 的分布列给出了 X 所有可能取值的概率. 通常,我们希望将这些信息综合成一个能够代表这个随机变量的数. X 的**期望**可以实现这个目的. X 的期望就是 X 的所有取值相对于它的概率的加权平均.

为了更好地理解期望的意义, 假定你有机会转动一个幸运轮许多次. 每次转动, 幸运轮会出现一个数, 不妨设为 $m_1, m_2, \cdots, m_n$ 中的一个. 这些数出现的概率分别为 $p_1, p_2, \cdots, p_n$. 而出现的数就是你所得到的钱数(给你的奖励). "每次"转动, 你所"期望"得到的钱数是多少? 此处"每次"和"期望"都是一些不确定的词汇. 但是下面的解 释可以把这些词汇的含义确定下来.

假定你一共转动幸运轮 k 次, 而其中有 k_i 次转动的结果为 m_i. 你所得到的总钱数为 $m_1k_1 + m_2k_2 + \cdots + m_nk_n$. 每次转动所得到的钱数为

$$M = \frac{m_1k_1 + m_2k_2 + \cdots + m_nk_n}{k}.$$

现在假定 k 是很大的一个数, 我们有理由假定概率与频率相互接近, 即

$$\frac{k_i}{k} \approx p_i, \quad i = 1, \cdots, n.$$

这样你每次转动幸运轮所期望得到的钱数是

$$M = \frac{m_1k_1 + m_2k_2 + \cdots + m_nk_n}{k} \approx m_1p_1 + m_2p_2 + \cdots + m_np_n.$$

由这个例子的启发, 我们引进下面的定义.[3]

[3]当随机变量的取值范围为可数无限集合的时候, 可能会遇到这样的情况: 和号 $\sum_x xp_X(x)$ 没有确切定义. 通常, 当 $\sum_x |x|p_X(x) < \infty$ 的时候, X 的期望值有确切定义, 它的值是一个有限数并且等于级数 $\sum_x xp_X(x)$ 的部分和的极限, 而这个极限值与求和号内各项的次序无关.
作为一个反例, 考虑随机变量 X 的取值范围为 $2^1, 2^2, \cdots$, 相应的概率分别为 $2^{-1}, 2^{-2}, \cdots$, 此时级数 $\sum_x xp_X(x) = \infty$, 并称 X 的期望无确切定义. 另一个反例是: X 取 2^k 和 -2^k 的概率为 $2^{-k}, k = 2, 3, \cdots$. 这个例子中 X 的期望也无确切定义, 其原因是 $\sum_x |x|p_X(x) = \infty$. 尽管这个随机变量是相对于0对称的, 其期望值似乎可以定义为0.
本书所涉及的随机变量的期望总是有定义的, 因此在论证中默认随机变量的期望是有定义的.

期望

设随机变量 X 的分布列为 p_X. X 的**期望值**(也称**期望**或**均值**)由下式给出:

$$\mathrm{E}[X] = \sum_x xp_X(x).$$

例 2.2 考虑两次抛掷一枚硬币的试验，而硬币的两面是不均匀的,正面向上的概率为 3/4. 令 X 是得到的正面数，这是一个二项随机变量,$n=2, p=3/4$. 它的分布列为

$$p_X(k)=\begin{cases}(1/4)^2, & \text{若 } k=0,\\ 2\cdot(1/4)\cdot(3/4), & \text{若 } k=1,\\ (3/4)^2, & \text{若 } k=2,\end{cases}$$

故其均值为

$$\mathrm{E}[X]=0\cdot\left(\frac{1}{4}\right)^2+1\cdot\left(2\cdot\frac{1}{4}\cdot\frac{3}{4}\right)+2\cdot\left(\frac{3}{4}\right)^2=\frac{24}{16}=\frac{3}{2}.$$

通常将 X 的均值解释为 X 的代表值，它位于 X 的值域中间的某一点.更确切地，可以将分布的均值看成分布列的“**重心**”(见图2.7的解释). 特别，当随机变量的分布列具有对称中心的时候,这个对称中心必定为这个对称随机变量的均值.

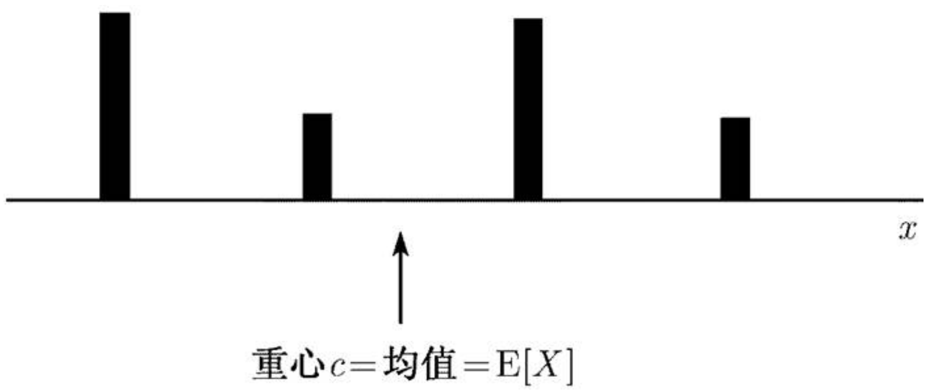


图 2.7 均值作为重心的解释，设在一根杆上在 x 处放上质量为 $p_X(x)$ 的物质，$p_X(x)>0$. 所谓重心是指杆上的平衡位置 c,使得 c 的右边的力矩等于 c 的左边的力矩. 即 满足

$$\sum_x (x-c)p_X(x)=0$$

的 c. 因此 $c=\sum_x xp_X(x)$, 即 $\mathrm{E}[X]$ 等于 X 的质量分布的重心

2.4.1 方差、矩和随机变量的函数的期望规则

期望是随机变量及其分布列的重要特征. 此外，还有其他重要的特征量.例如随机变量 X 的**二阶矩**定义为随机变量 X^2 的均值. 进一步 n **阶矩** $\mathrm{E}[X^n]$ 定义为 X^n 的期望值. 这样均值本身就刚好是一阶矩.

除了均值，随机变量 X 的最重要的特征量是**方差**，记作 $\mathrm{var}(X)$. 它由下式定义

$$\operatorname{var}(X) = \mathrm{E}\left[(X - \mathrm{E}[X])^2\right].$$

由于 $(X-\mathrm{E}[X])^2$ 只能取非负值，故方差只能取非负值. 方差提供了X 在期望周围分散程度的一个测度. 分散程度的另一个测度是**标准差**,它由下式定义

$$\sigma_X = \sqrt{\operatorname{var}(X)}.$$

标准差具有实用性，因为它的量纲与 X 的相同. 例如 X 是以米为单位的长度，方差的单位为平方米，而标准差的单位为米.

计算方差的一种方法是先行计算随机变量 $(X-\mathrm{E}[X])^2$ 的分布列,然后利用期望值的定义计算 X 的方差. $(X-\mathrm{E}[X])^2$ 是随机变量 X 的函数，可利 用前面提供的方法计算 $(X-\mathrm{E}[X])^2$ 的分布列.

例 2.3 考虑例2.1中的随机变量 X,它的分布列为

$$p_X(x) = \begin{cases} 1/9, & \text{若 } x \text{ 是 } [-4,4] \text{ 中的整数}, \\ 0, & \text{其他}. \end{cases}$$

此时，均值 $\mathrm{E}[X]=0$. 这可以从分布的对称性看出,也可以从期望的定义直接计算得到

$$\mathrm{E}[X] = \sum_x x p_X(x) = \frac{1}{9}\sum_{x=-4}^{4} x = 0.$$

令 $Z=(X-\mathrm{E}[X])^2=X^2$. 在例2.1中,已经得到

$$p_Z(z) = \begin{cases} 2/9, & \text{若 } z = 1,4,9,16, \\ 1/9, & \text{若 } z = 0, \\ 0, & \text{其他}, \end{cases}$$

这样，X 的方差为

$$\operatorname{var}(X) = \mathrm{E}[Z] = \sum_z z p_Z(z) = 0\cdot\frac{1}{9} + 1\cdot\frac{2}{9} + 4\cdot\frac{2}{9} + 9\cdot\frac{2}{9} + 16\cdot\frac{2}{9} = \frac{60}{9}.$$

计算 $\operatorname{var}(X)$ 时并不需要先行计算 $(X-\mathrm{E}[X])^2$ 的分布列,而另有更加便利的方法. 这种方法根据下面的规则得到.

随机变量的函数的期望规则

设随机变量 X 的分布列为 p_X, 又设$g(X)$ 是 X 的一个函数,则 $g(X)$ 的期望由下列公式得到

$$\mathrm{E}[g(X)] = \sum_x g(x)p_X(x).$$

为验证此公式, 令 $Y = g(X)$ 并利用2.3节导出的公式

$$p_Y(y) = \sum_{\{x|g(x)=y\}} p_X(x),$$

得到

$$\begin{aligned}
\mathrm{E}[g(X)] &= \mathrm{E}[Y] \\
&= \sum_y y p_Y(y) \\
&= \sum_y y \sum_{\{x|g(x)=y\}} p_X(x) \\
&= \sum_y \sum_{\{x|g(x)=y\}} y p_X(x) \\
&= \sum_y \sum_{\{x|g(x)=y\}} g(x) p_X(x) \\
&= \sum_x g(x) p_X(x).
\end{aligned}$$

将期望规则应用到 X 的方差, 我们得到

$$\mathrm{var}(X) = \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] = \sum_x (x - \mathrm{E}[X])^2 p_X(x).$$

相似地, 对于 X 的 n 阶矩, 我们有

$$\mathrm{E}[X^n] = \sum_x x^n p_X(x).$$

因此在计算 X 的 n 阶矩的时候, 我们不必先求 X^n 的分布列.

例 2.3(续) 设随机变量 X 的分布列由下式给出,

$$p_X(x) = \begin{cases} 1/9, & \text{若 } x \text{ 是 } [-4,4] \text{ 中的整数}, \\ 0, & \text{其他}. \end{cases}$$

利用期望规则得到

$$
\begin{aligned}
\operatorname{var}(X) &= \mathrm{E}\left[(X-\mathrm{E}[X])^2\right] \\
&= \sum_x (x-\mathrm{E}[X])^2 p_X(x) \\
&= \frac{1}{9}\sum_{x=-4}^{4} x^2 (\text{因为}\mathrm{E}[\mathrm{X}]=0) \\
&= \frac{1}{9}(16+9+4+1+0+1+4+9+16) \\
&= \frac{60}{9}.
\end{aligned}
$$

这个结果与早先得到的结果是一样的.

先前已经提到，方差是非负的. 那么是否可为0？由于在方差的公式 $\sum_x (x-\mathrm{E}[X])^2 p_X(x)$ 中, 每一项都是非负的. 为了使得这个和式为0，其充要条件是对每一个 x, $(x-\mathrm{E}[X])^2 p_X(x)=0$. 这个条件说明对每一个使得 $p_X(x)>0$的 x, 均有 $x=\mathrm{E}[X]$. 这说明 X 其实不是随机的, 随机变量 X 等于 $\mathrm{E}[X]$ 的概率为1.

方差

随机变量 X 的方差由下列公式所定义：

$$\operatorname{var}(X) = \mathrm{E}\left[(X-\mathrm{E}[X])^2\right].$$

并且可以用下式进行计算：

$$\operatorname{var}(X) = \sum_x (x-\mathrm{E}[X])^2 p_X(x).$$

它是非负的，其平方根称为**标准差**, 记为 σ_X.

2.4.2 均值和方差的性质

我们将用随机变量的函数的期望规则导出一些均值和方差的重要性质. 首先考虑随机变量 X 的函数

$$Y = aX + b,$$

其中 a 和 b 是已知常数. 关于线性函数 Y 的均值和方差, 我们有

$$\mathrm{E}[Y] = \sum_x (ax+b)p_X(x) = a\sum_x x p_X(x) + b\sum_x p_X(x) = a\mathrm{E}[X] + b.$$

进一步地

$$
\begin{aligned}
\operatorname{var}(Y) &= \sum_x \left(ax+b-\mathrm{E}[aX+b]\right)^2 p_X(x) \\
&= \sum_x \left(ax+b-a\mathrm{E}[X]-b\right)^2 p_X(x) \\
&= a^2 \sum_x (x-\mathrm{E}[X])^2 p_X(x) \\
&= a^2 \operatorname{var}(X).
\end{aligned}
$$

随机变量的线性函数的均值和方差

设 X 为随机变量，令

$$Y = aX + b,$$

其中 a 和 b 为给定的常数，则

$$\mathrm{E}[Y] = a\mathrm{E}[X] + b, \quad \operatorname{var}(Y) = a^2 \operatorname{var}(X).$$

此外，我们还将证明如下一个方差的重要公式.

用矩表达的方差公式

$$\operatorname{var}(X) = \mathrm{E}[X^2] - (\mathrm{E}[X])^2.$$

这个用矩表达的方差公式的证明可以通过下列等式完成：

$$
\begin{aligned}
\operatorname{var}(X) &= \sum_x \left(x-\mathrm{E}[X]\right)^2 p_X(x) \\
&= \sum_x \left(x^2 - 2x\mathrm{E}[X] + (\mathrm{E}[X])^2\right) p_X(x) \\
&= \sum_x x^2 p_X(x) - 2\mathrm{E}[X] \sum_x x p_X(x) + (\mathrm{E}[X])^2 \sum_x p_X(x) \\
&= \mathrm{E}[X^2] - 2(\mathrm{E}[X])^2 + (\mathrm{E}[X])^2 \\
&= \mathrm{E}[X^2] - (\mathrm{E}[X])^2.
\end{aligned}
$$

最后我们用例子说明一个陷阱：除非 $g(X)$ 是一个线性函数,一般情况下 $\mathrm{E}[g(X)]$ 不等于 $g(\mathrm{E}[X])$.

例 2.4(平均速度和平均时间) 如果遇到好天气(这种天气出现的概率为0.6),爱丽丝会步行2英里上学，步行速度为每小时5英里(V=5).天气不好的时候，她骑摩托车上学，时速30英里(V=30).她上学所用的平均时 间是多少？

正确的方法是先计算 时间 T 的分布列,

$$p_T(t) = \begin{cases} 0.6, & t = 2/5\text{小时}, \\ 0.4, & t = 2/30\text{小时}, \end{cases}$$

然后计算均值

$$\mathrm{E}[T] = 0.6 \cdot \frac{2}{5} + 0.4 \cdot \frac{2}{30} = \frac{4}{15} \text{ 小时}.$$

然而，下面的计算是错误的：先计算平均速度

$$\mathrm{E}[V] = 0.6 \cdot 5 + 0.4 \cdot 30 = 15 \text{ 英里/小时}.$$

然后声称平均时间为

$$\frac{2}{\mathrm{E}[V]} = \frac{2}{15} \text{ 小时}.$$

总之，在这个例子中

$$T = \frac{2}{V}, \quad \mathrm{E}[T] = \mathrm{E}\left[\frac{2}{V}\right] \neq \frac{2}{\mathrm{E}[V]}.$$

2.4.3　某些常用的随机变量的均值和方差

我们将推导出一些重要的随机变量的均值和方差,在本课程中经常会遇到这些公式.

例 2.5(伯努利随机变量的均值和方差)　考虑抛掷一枚硬币，设正面出现的概率为 p ,反面出现的概率为 $1-p$. 伯努利随机变量的分布列为

$$p_X(k) = \begin{cases} p, & \text{若 } k = 1, \\ 1-p, & \text{若 } k = 0. \end{cases}$$

下面给出了它的均值、二阶矩和方差的计算公式

$$\begin{aligned} \mathrm{E}[X] &= 1 \cdot p + 0 \cdot (1-p) = p, \\ \mathrm{E}[X^2] &= 1^2 \cdot p + 0^2 \cdot (1-p) = p, \\ \mathrm{var}(X) &= \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = p - p^2 = p(1-p). \end{aligned}$$

例 2.6(离散均匀随机变量)　设涉及的试验是抛掷一个均匀的具有6个面的骰子. 其平均点数和方差是多少？我们将试验结果看成一个随机变量,它的分布列为

$$p_X(k)=\begin{cases}1/6, & 若\ k=1,2,3,4,5,6,\\ 0, & 其他.\end{cases}$$

由于分布列相对于3.5是对称的，我们得到 $\mathrm{E}[X]=3.5$. 关于方差,我们有

$$\begin{aligned}\operatorname{var}(X)&=\mathrm{E}[X^2]-(\mathrm{E}[X])^2\\&=\frac{1}{6}(1^2+2^2+3^2+4^2+5^2+6^2)-(3.5)^2,\end{aligned}$$

这样，可得到 $\operatorname{var}(X)=35/12$.

上面的随机变量是**离散均匀随机变量**的特殊情况.按定义离散均匀随机变量的取值范围是由相邻的整数所组成的有限集,而取每个整数 的概率都是相等的. 这样它的分布列为

$$p_X(k)=\begin{cases}\dfrac{1}{b-a+1}, & 若\ k=a,a+1,\cdots,b,\\ 0, & 其他,\end{cases}$$

其中 a,b 是两个整数，作为随机变量的值域的两个端点,$a<b$（X 的分布列的图示见图2.8).由于它的分布列相对于 $(a+b)/2$ 是对称的，其均值为

$$\mathrm{E}[X]=\frac{a+b}{2}.$$

为计算 X 的方差，先考虑 a=1和 $b=n$ 的简单情况. 利用归纳法可以证明

$$\mathrm{E}[X^2]=\frac{1}{n}\sum_{k=1}^{n}k^2=\frac{1}{6}(n+1)(2n+1)$$

(具体证明过程留作习题). 这样利用一、二阶矩，可得到 X 的方差

$$\begin{aligned}\operatorname{var}(X)&=\mathrm{E}[X^2]-(\mathrm{E}[X])^2\\&=\frac{1}{6}(n+1)(2n+1)-\frac{1}{4}(n+1)^2\\&=\frac{n^2-1}{12}.\end{aligned}$$

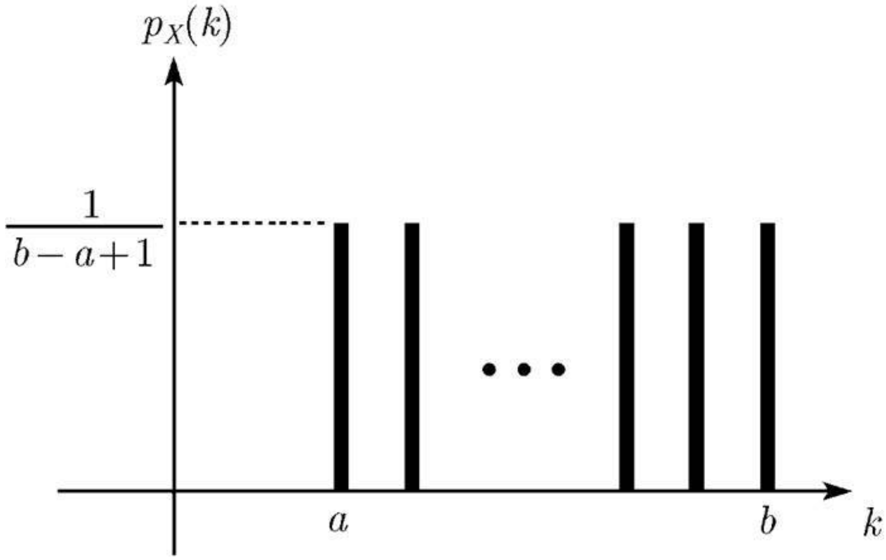


图 2.8　在 a 和 b 之间均匀分布的随机变量的分布列. 它的均值和方差为

$$\mathrm{E}[X]=\frac{a+b}{2},\quad \mathrm{var}(X)=\frac{(b-a)(b-a+2)}{12}$$

对于 a 和 b 的一般情况，实际上在区间$[a,b]$上的均匀分布与在区间$[1,b-a+1]$上的分布之间的差异,只是一个分布是另一个分布的平移,因此两者具有相同的方差(此处区间$[a,b]$是指处于 a 和 b 之间的整数的集合). 这样，在一般情况下,X 的方差只需将简单情况下公式中的 n 替换成$b-a+1$，即

$$\mathrm{var}(X)=\frac{(b-a+1)^2-1}{12}=\frac{(b-a)(b-a+2)}{12}.$$

例 2.7(泊松随机变量的均值)　设 X 的分布列为泊松分布列，即

$$p_X(k)=\mathrm{e}^{-\lambda}\frac{\lambda^k}{k!},\quad k=0,1,2,\cdots,$$

其中 $\lambda>0$ 为常数. 其均值可从下列等式得到

$$
\begin{aligned}
\mathrm{E}[X] &= \sum_{k=0}^{\infty} k\mathrm{e}^{-\lambda}\frac{\lambda^k}{k!} \\
&= \sum_{k=1}^{\infty} k\mathrm{e}^{-\lambda}\frac{\lambda^k}{k!} \qquad (k=0\text{这一项为}0) \\
&= \lambda\sum_{k=1}^{\infty} \mathrm{e}^{-\lambda}\frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda\sum_{m=0}^{\infty} \mathrm{e}^{-\lambda}\frac{\lambda^m}{m!} \qquad (\text{令}m=k-1) \\
&= \lambda.
\end{aligned}
$$

最后一个等式利用了泊松分布列的归一化性质.

相似的计算指出泊松随机变量的方差为 λ（见本章2.7节的例2.20).在以后的章节中将用不同的方法导出这个事实.

2.4.4　利用期望值进行决策

设想有一个项目，有几种处理方案. 而每种处理方案都有随机的回报,那么用什么样的准则去最优地选择处理方案呢？期望值是一个合理且方便的准则.如果把期望回报看成一个处理方案 长期重复执行的平均回报,那么选择具有最大期望回报的策略是合理的.下面是一个例子.

例 2.8(智力测验)　这是一个具有随机回报的实施方案最优选择的典型例子.

在一个智力游戏中一共有两个问题需要回答,但游戏规则要求你选择一个问题作为首先回答的问题. 问题1比较容易,你能够正确回答的概率为0.8. 回答正确就能够得到100美元的奖金.问题2比较难, 你能够正确回答的概率为0.5.回答正确就能够得到200美元的奖金.若你选定一个首先回答的问题却不能正确地回答, 你不但不能拿到奖金,而且也不容许回答第二个问题. 若你能够正确地回答第一个问题,就还有机会回答第二个问题. 为了使奖金总和的期望值最大,你应该选择哪一个问题作为首先回答的问题？

这个问题并不简单, 高回报必有高风险. 希望首先回答问题2, 奖金多,但是问题比较难, 并且要冒着不让回答问题1的风险.我们将所得到的奖金总额作为随机变量 X, 并且计算两种可能的回答问题的次序下的期望值 $\mathrm{E}[X]$（见图2.9).