第二章 离散随机变量

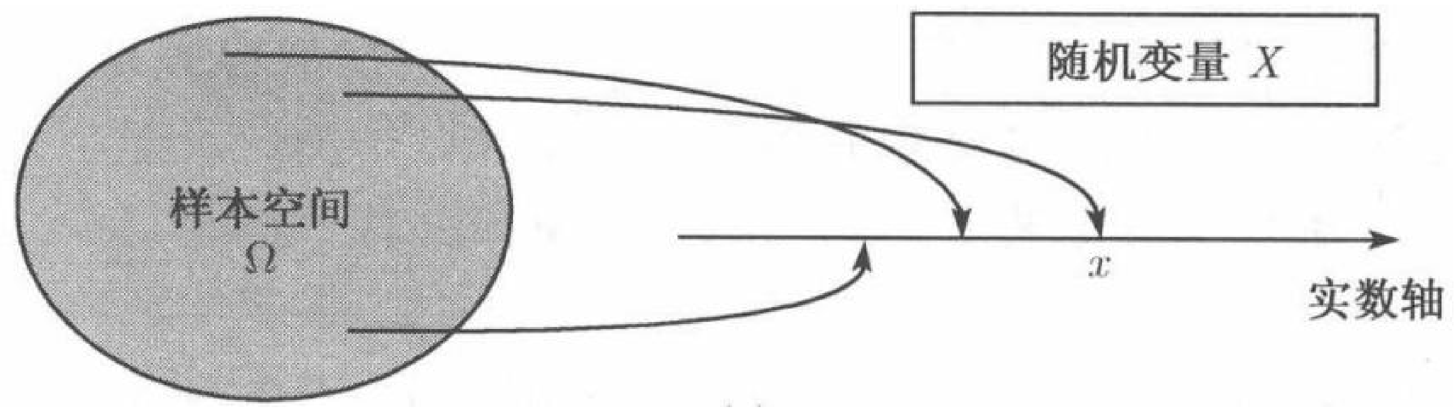
Probability and Statistics for Computer Scientists

第二章 离散随机变量

- 相关概念
- 分布列
- 随机变量的函数
- 期望、均值和方差
- 多个随机变量的联合分布
- 条件分布列
- 独立性

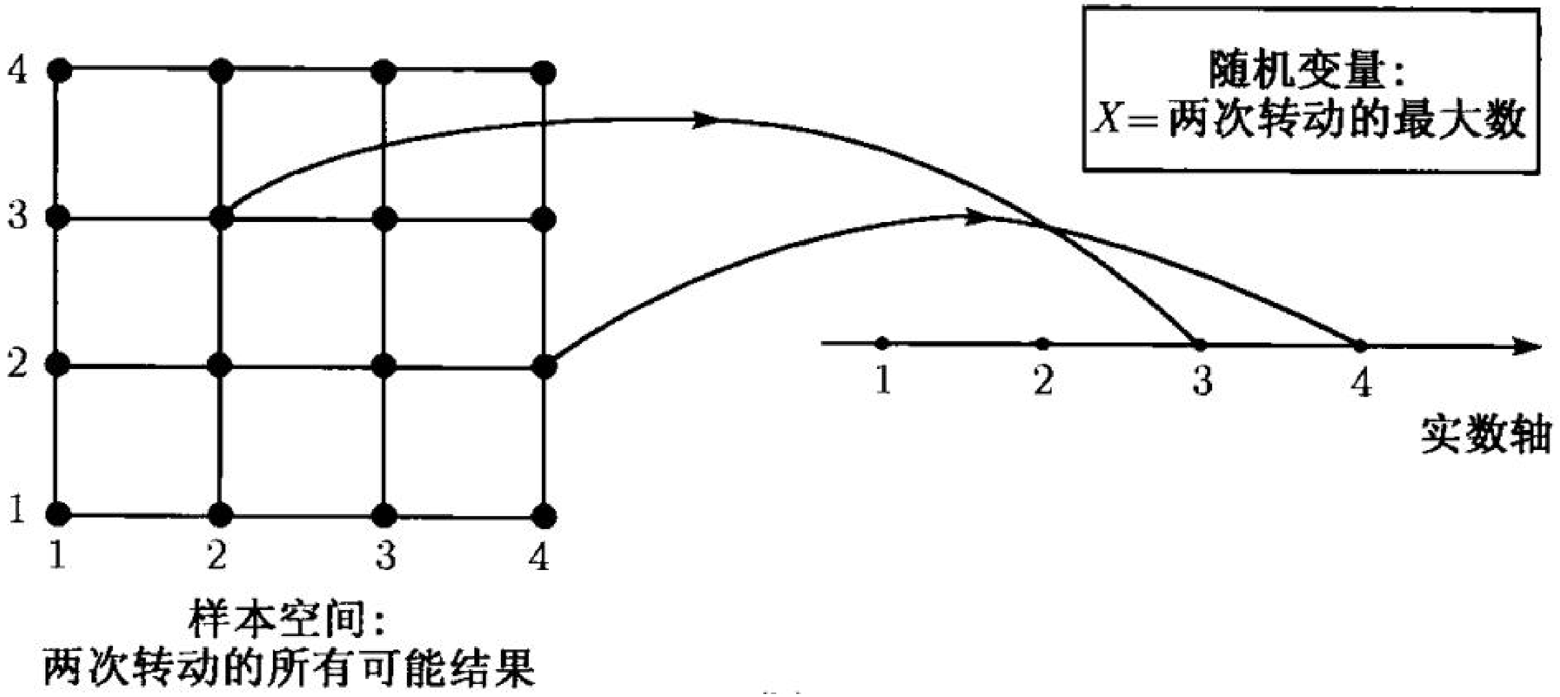
§1 相关概念

1.1 随机变量：通过随机变量，对实验结果确定概率。



随机变量是实验结果的一个实值函数。

§1 相关概念



§1 相关概念

1.2 与随机变量相关的概念

与随机变量相关的主要概念

在一个试验的概率模型之下:

- **随机变量**是试验结果的实值函数.
- **随机变量的函数**定义了另一个随机变量.
- 对于一个随机变量, 我们可以定义一些平均量, 例如**均值**和**方差**.
- 可以在某事件或某随机变量的**条件**之下定义一个随机变量.
- 存在一个随机变量与某事件或某随机变量相互**独立**的概念.

§1 相关概念

1.3 离散

若一个随机变量的值域 (随机变量的取值范围) 为一个有限集合或最多为可数无限集合, 则称这个随机变量为**离散的**.

离散随机变量如:

$$\operatorname{sgn}(a)=\begin{cases}1, & \text{若 } a>0,\\ 0, & \text{若 } a=0,\\ -1, & \text{若 } a<0.\end{cases}$$

§1 相关概念

1.4 与离散随机变量相关的概念

与离散随机变量相关的概念

在一个试验的概率模型之下:

- **离散随机变量**是试验结果的一个实值函数, 但是它的取值范围只能是有限多个值或可数无限多个值.
- 一个离散随机变量有一个**分布列**, 它对于随机变量的每一个取值, 给出一个概率.
- **离散随机变量的函数**也是一个离散随机变量, 它的分布列可以从原随机变量的分布列得到.

§2 分布列

2.1 分布列的概念

设 x 是随机变量 X 的取值，$p_X(x) = \mathrm{P}(\{X = x\})$. $p_X(x)$为分布列

对于分布列，我们有 $\sum_x p_X(x) = 1$

对于任意一个 X 的可能值的集合 S，下式成立：

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

§2 分布列

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 分布列

2.2 伯努利随机变量

考虑抛硬币，正面向上的概率

$$X=\begin{cases}1, & \text{若正面向上},\\ 0, & \text{若反面向上}.\end{cases} \qquad p_X(k)=\begin{cases}p, & \text{若} \quad k=1\\ 1-p, & \text{若} \quad k=0\end{cases}$$

用于刻画具有两个试验结果的概率模型.

§2 分布列

2.3 二项随机变量

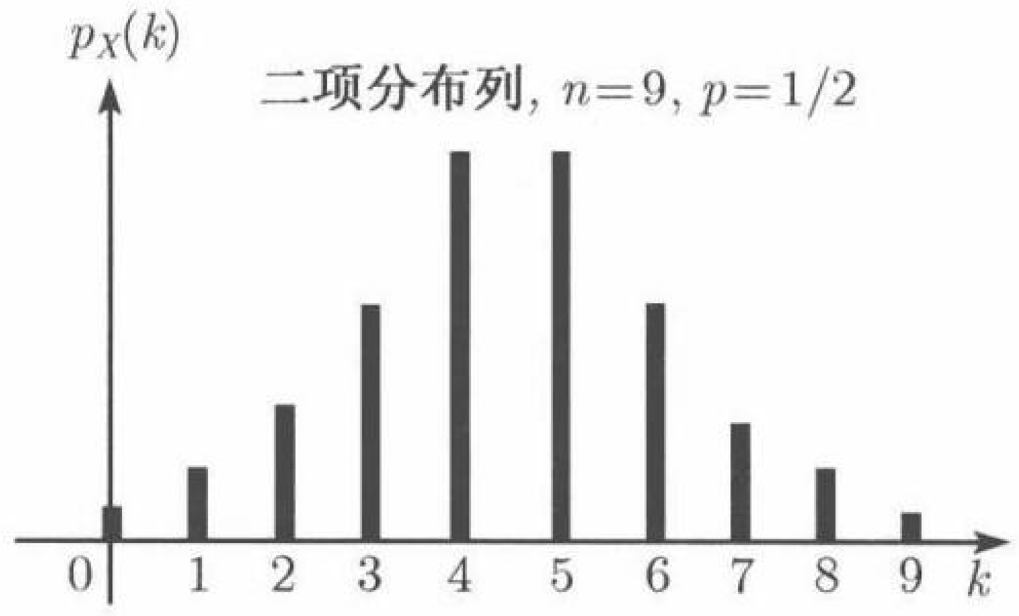
考虑抛硬币n次，正面向上的次数

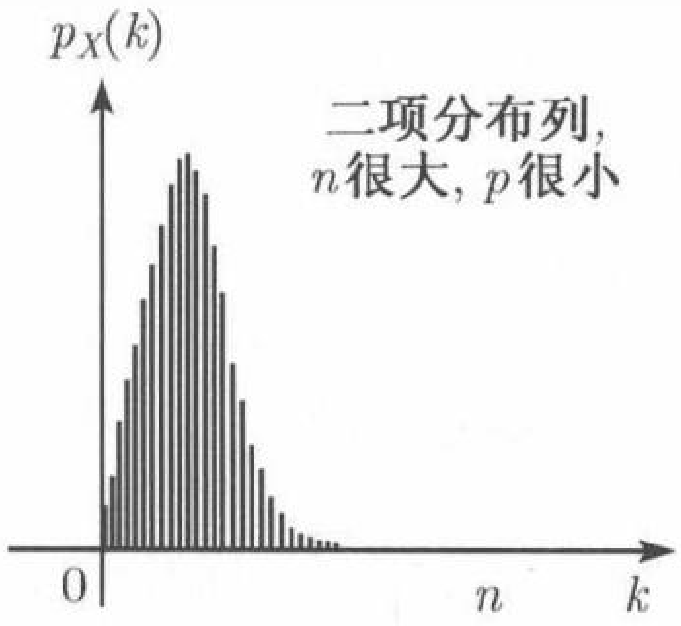
$$p_X(k) = \mathrm{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \cdots, n$$

利用归一化公理可以得到

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

§2 分布列





p=1/2时，分布相对于n/2对称

§2 分布列

证明：

$$\frac{P(X=k)}{P(X=k-1)}=\frac{C_n^k p^k q^{n-k}}{C_n^{k-1}p^{k-1}q^{n-k+1}}=\frac{(n-k+1)p}{kq}$$

$$=\frac{(n+1)p-kp}{kq}=\frac{(n+1)p-k(1-q)}{kq}$$

$$=1+\frac{(n+1)p-k}{kq}\quad(k=1,2,\cdots,n)$$

§2 分布列

（小于）（等于）　（大于）（等于）

当 $k<(n+1)p$ 时. $p(X=k)>p(X=k-1)$.

此时 $p(X=k)$ 随 k 的增加而上升.

当 $k>(n+1)p$ 时. $p(X=k)<p(X=k-1)$.

此时 $p(X=k)$ 随 k 的增加而下降.

√ 当 $k=(n+1)p$ 为正整数时. $p(X=k)=p(X=k-1)$, 此时 $p(X=k)$ 在 $k=(n+1)p$ 及 $(n+1)p-1$ 处都达到最大值.

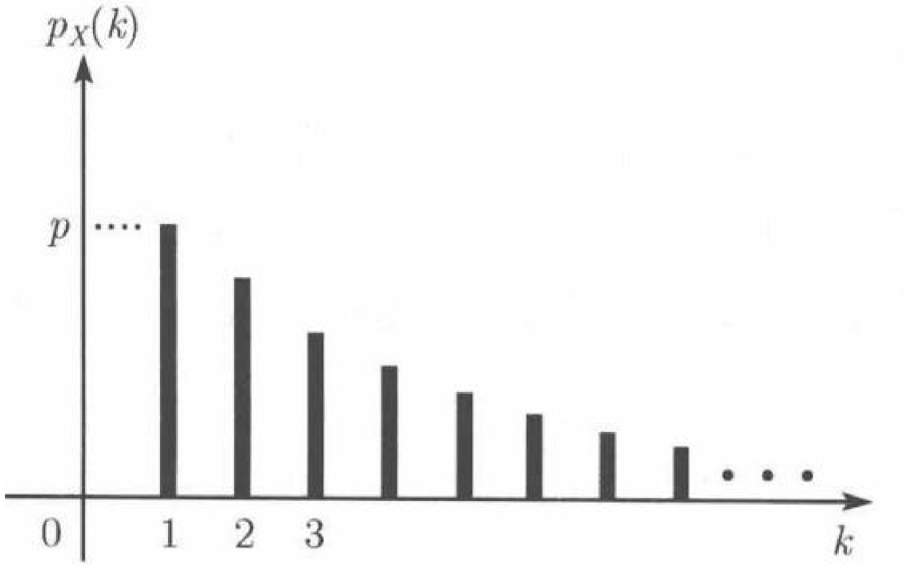
√ 若 $(n+1)p$ 不是整数, 则 $k=[(n+1)p]$ 时达到最大值.

§2 分布列

2.4 几何随机变量

独立试验序列中，直到试验第一次成功所需的试验次数为X

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



几何随机变量的分布列.

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

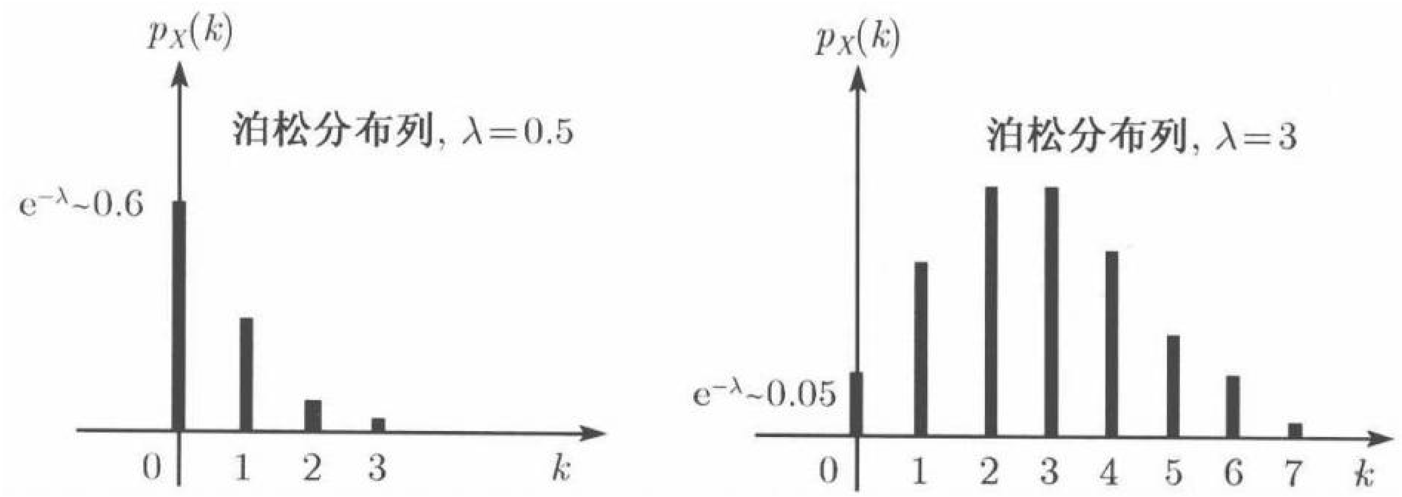
§2 分布列

2.5 泊松随机变量

随机变量X的分布由下式给出：

$$p_X(k) = \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \cdots, \quad \lambda \text{ 是分布列中取正值的参数.}$$

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



§2 分布列

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:

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

其中 $\lambda = np$, n 很大, p 很小.

例如, $n = 100$, $p = 0.01$, $k = 5$

$$\frac{100!}{95!5!} \cdot 0.01^5(1-0.01)^{95} = 0.002\ 90. \qquad e^{-1}\frac{1}{5!} = 0.003\ 06, \lambda = np = 100 \cdot 0.01 = 1.$$

§ 2 分布列

泊松逼近二项分布的证明.

证: 令 $\overset{\text{固定常数}}{\lambda}=np$.

$$P_X(k)=\frac{n!}{(n-k)!\,k!}p^k\cdot(1-p)^{n-k}=\frac{\overset{n(n-1)\cdots(n-k+1)}{\cancel{n(n-1)}\cdots 1}}{\cancel{(n-k)(n-k-1)\cdots 1}}\cdot\frac{p^k}{k!}\cdot(1-p)^{n-k}.$$

$$\overset{\text{将}\lambda=np\text{代入}}{=}\frac{n(n-1)\cdots(n-k+1)}{n^k}\cdot\frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^{n-k}.$$

固定k, 令 $n\to\infty$, 则

$$\frac{n-k+j}{n}\to 1\quad (j=1,2\cdots k).\qquad \left(1-\frac{\lambda}{n}\right)^{-k}\to 1,\quad \left(1-\frac{\lambda}{n}\right)^{n}\to e^{-\lambda}$$

利用: $\lim\limits_{n\to\infty}\left(1+\frac{1}{n}\right)^n=e$

~~则~~ $\dfrac{n(n-1)\cdots(n-k+1)}{n^k}=1$

则 $P_X(k)=e^{-\lambda}\cdot\dfrac{\lambda^k}{k!}$ (当 $n\to\infty$ 时)

§3 随机变量的函数

随机变量的函数

$$Y = g(X)$$

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

例 2.1 计算 $Y = |X|$ 的分布列, 其中 X 的分布列由下式给出,

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

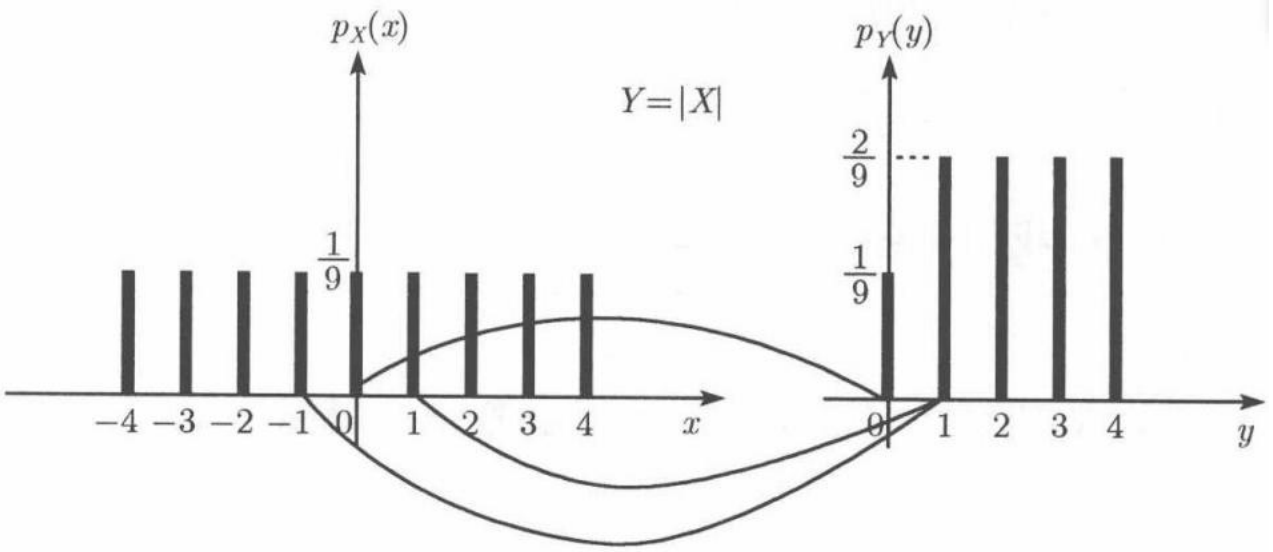
§3 随机变量的函数

Y 的值域为 $y=0,1,2,3,4$, 只需将满足 $|x|=y$ 的所有 $p_X(x)$ 的值相加.

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

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

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



§4 期望、均值、方差

4.1 期望、均值

X所有取值相对于它概率的加权平均

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:

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

§4 期望、均值、方差

4.2 方差、矩和随机变量的函数的期望规则

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

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

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*, $\sigma_X = \sqrt{\mathrm{var}(X)}$.

§4 期望、均值、方差

例 2.3 考虑例 2.1 中的随机变量 X, 它的分布列为

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

X 的方差为？

$\mathrm{E}[X] = 0$

令 $Z = (X - \mathrm{E}[X])^2 = X^2$. $p_Z(z) = \begin{cases} 2/9, & \text{若 } z = 1, 4.9, 16 \\ 1/9, & \text{若 } z = 0, \\ 0, & \text{其他}, \end{cases}$

$$\mathrm{var}(X) = \mathrm{E}[Z] = \sum_z z p_Z(z)$$

§4 期望、均值、方差

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

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

§4 期望、均值、方差

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

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

§4 期望、均值、方差

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

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

§4 期望、均值、方差

4.3 均值和方差的性质

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

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

用矩表达的方差公式

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

§4 期望、均值、方差

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

§ 4 期望、均值、方差

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?

$$p_T(t) = \begin{cases} 0.6, & t = 2/5\text{小时}, \\ 0.4, & t = 2/30\text{小时}, \end{cases} \longrightarrow \mathrm{E}[T] = 0.6 \cdot \frac{2}{5} + 0.4 \cdot \frac{2}{30} = \frac{4}{15}$$

$$\boxed{\mathrm{E}[V] = 0.6 \cdot 5 + 0.4 \cdot 30 = 15\text{英里/小时}, \longrightarrow \frac{2}{\mathrm{E}[V]} = \frac{2}{15}}$$

除非 $g(X)$ 是一个线性函数, 一般情况下 $\mathrm{E}[g(X)]$ 不等于 $g(\mathrm{E}[X])$.

§4 期望、均值、方差

4.4 常用随机变量的均值和方差

伯努利随机变量

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

离散均匀随机变量

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

泊松随机变量

$$\mathrm{E}[X] = \lambda \qquad \mathrm{var}(X) = \lambda$$

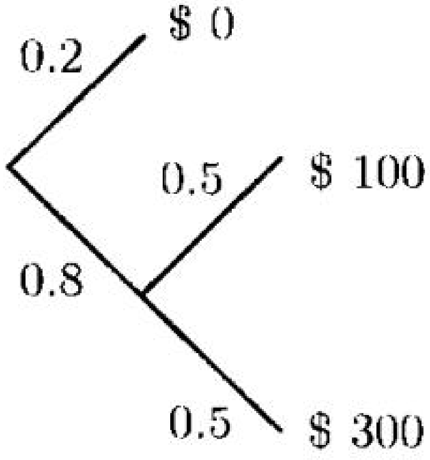
§4 期望、均值、方差

4.4 利用期望进行决策

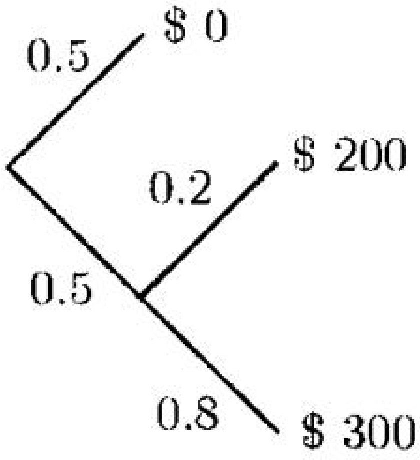
例 2.8 (智力测验) 这是一个具有随机回报的实施方案最优选择的典型例子.

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?

§4 期望、均值、方差



首先回答问题 1



首先回答问题 2

§4 期望、均值、方差

推广

p_1 和 p_2 分别表示正确回答问题 1 和问题 2 的概率

v_1 和 v_2 分别表示正确回答问题后所得到的奖金

最优策略为先行回答问题 1 的充要条件是

$$p_1 v_1 + p_1 p_2 v_2 \geqslant p_2 v_2 + p_1 p_2 v_1$$

或

$$\frac{p_1 v_1}{1 - p_1} \geqslant \frac{p_2 v_2}{1 - p_2}$$

§5 多个随机变量的联合分布

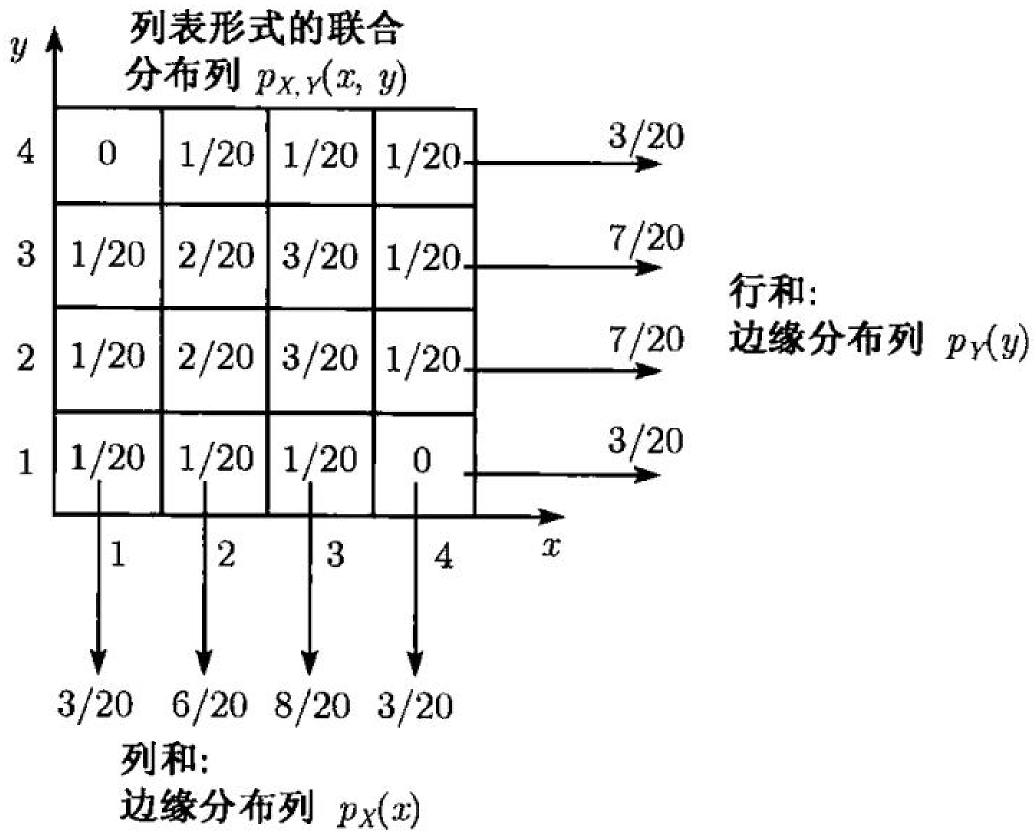
设在同一个试验中有两个随机变量 X 和 Y.

联合分布列用 $p_{X,Y}$ 表示

$$p_{X,Y}(x,y) = \mathrm{P}(X = x, Y = y).$$

边缘分布列

$$\begin{aligned} p_X(x) &= \mathrm{P}(X = x) \\ &= \sum_y \mathrm{P}(X = x, Y = y) \\ &= \sum_y P_{X,Y}(x,y), \end{aligned}$$



§5 多个随机变量的联合分布

5.1 多个随机变量的函数

二元函数 $Z = g(X, Y)$

$$p_Z(z) = \sum_{\{(x,y)|g(x,y)=z\}} p_{X,Y}(x, y).$$

$$\mathrm{E}\left[g(X, Y)\right] = \sum_x \sum_y g(x, y) p_{X,Y}(x, y)$$

当 g 是形如 $aX + bY + c$ 的时候，$\mathrm{E}[aX + bY + c] = a\mathrm{E}[X] + b\mathrm{E}[Y] + c$

§ 5 多个随机变量的联合分布

5.2 多于两个随机变量的情况

$$p_{X,Y,Z}(x,y,z) = \mathrm{P}(X=x, Y=y, Z=z)$$

$$p_{X,Y}(x,y) = \sum_{z} p_{X,Y,Z}(x,y,z),$$

$$p_X(x) = \sum_{y}\sum_{z} p_{X,Y,Z}(x,y,z).$$

边缘分布列

$$\mathrm{E}[g(X,Y,Z)] = \sum_{x}\sum_{y}\sum_{z} g(x,y,z)p_{X,Y,Z}(x,y,z)$$

形如$g = a_1X_1 + a_2X_2 + \ldots + a_nX_n$，则$E[a_1X_1 + a_2X_2 + \ldots + a_nX_n] = a_1E[X_1] + a_2E[X_2] + \ldots + a_nE[X_n]$

§5 多个随机变量的联合分布

例 2.10 (二项随机变量的均值) 你的概率班上有 300 个学生, 每个学生有 1/3 的概率可得到成绩 A, 并且相互独立. 记 X 为班上取得 A 的学生数. X 的平均数为多少?

$$\mathrm{E}[X]=\sum_{i=1}^{300}\mathrm{E}[X_i]=\sum_{i=1}^{300}\frac{1}{3}=300\cdot\frac{1}{3}=100.$$

推广：设班上有 n 个学生, 每个学生得 A 的概率为 p,

$$\mathrm{E}[X]=\sum_{i=1}^{n}\mathrm{E}[X_i]=\sum_{i=1}^{n}p=np.$$

§6 条件分布列

6.1 某个事件发生的条件下的随机变量

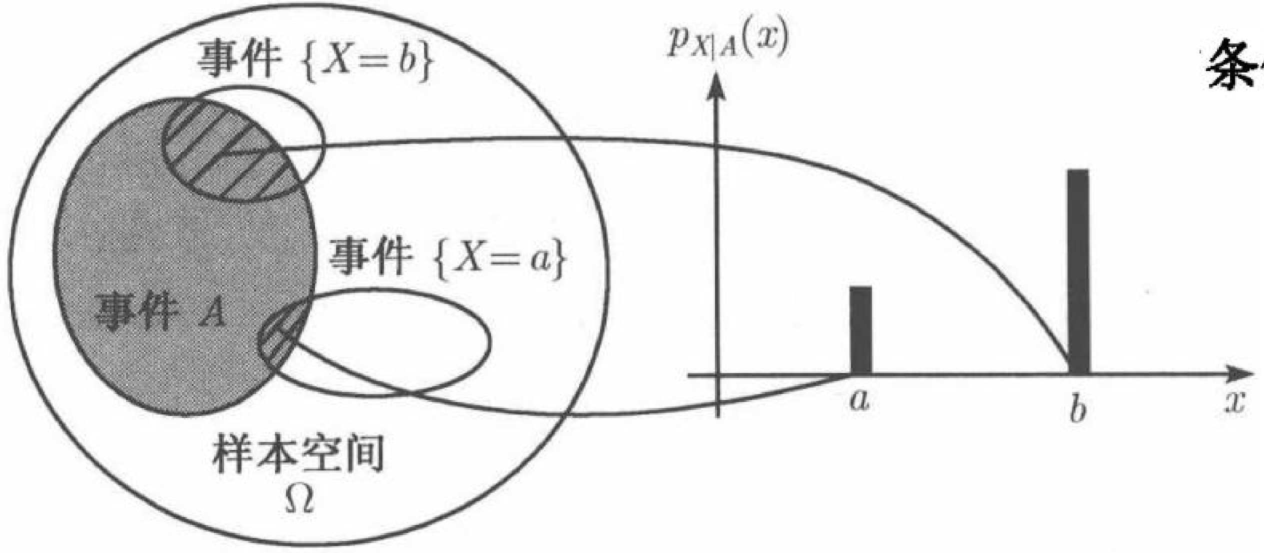
随机变量 X 的条件分布列

$$p_{X|A}(x) = \mathrm{P}(X = x|A) = \frac{\mathrm{P}\left(\{X = x\} \cap A\right)}{\mathrm{P}(A)}.$$

$\sum_{x} p_{X|A}(x) = 1$ 故 $p_{X|A}$ 符合分布列的要求.

§6 条件分布列

6.2 给定另一个随机变量的值的条件下的随机变量



条件分布列 $p_{X|Y}$

$$p_{X|Y}(x|y) = \mathrm{P}(X = x|Y = y).$$

$$= \frac{\mathrm{P}(X = x, Y = y)}{\mathrm{P}(Y = y)} = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

§ 6 条件分布列

条件分布列也可以用于计算边缘分布列，即有

$$p_X(x)=\sum_y p_{X,Y}(x,y)=\sum_y p_Y(y)p_{X|Y}(x|y).$$

应用:

例 2.15 考虑计算机网络中的一个信息传送器. 下面是有关的随机变量.

X : 给定消息的传送时间　　Y : 给定消息的长度.

我们知道给定消息长度的条件下传送时间的分布列和消息长度的分布列. 我们希望找到传送一个消息的时间的 (无条件) 分布列.

§ 6 条件分布列

假定一个消息的长度可以取两个可能值：$y = 10^2$ 和 $y = 10^4$(单位：比特)

$$p_Y(y) = \begin{cases} 5/6, & 若\ y = 10^2, \\ 1/6, & 若\ y = 10^4. \end{cases}$$

传送时间依赖于消息的长度和当时网络的拥塞程度, 特别传送时间为 $10^{-4}Y$ 的概率为 1/2, 传送时间为 $10^{-3}Y$ 的概率为 1/3, 传送时间为 $10^{-2}Y$ 的概率为 1/6.

$$p_{X|Y}(x|10^2) = \begin{cases} 1/2, & 若\ x = 10^{-2}, \\ 1/3, & 若\ x = 10^{-1}, \\ 1/6, & 若\ x = 1; \end{cases} \qquad p_{X|Y}(x|10^4) = \begin{cases} 1/2, & 若\ x = 1, \\ 1/3, & 若\ x = 10, \\ 1/6, & 若\ x = 100. \end{cases}$$

§6 条件分布列

$$p_X(x) = \sum_y p_Y(y) p_{X|Y}(x|y),$$

得到

$$p_X(10^{-2}) = \frac{5}{6} \cdot \frac{1}{2}, \quad p_X(10^{-1}) = \frac{5}{6} \cdot \frac{1}{3}, \quad p_X(1) = \frac{5}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{2},$$

$$p_X(10) = \frac{1}{6} \cdot \frac{1}{3}, \quad p_X(100) = \frac{1}{6} \cdot \frac{1}{6}.$$

§6 条件分布列

6.3 条件期望

关于条件期望的小结

设X 和 Y 为某一试验中的两个随机变量.

- 设A 为某事件, $\mathrm{P}(A)>0$. 随机变量 X 在给定 A 发生的条件下的条件期望为

$$\mathrm{E}[X|A]=\sum_x x p_{X|A}(x).$$

  对于函数 $g(X)$, 我们有

$$\mathrm{E}\left[g(X)|A\right]=\sum_x g(x) p_{X|A}(x).$$

- 给定$Y=y$ 的条件下 X 的条件期望由下式定义

$$\mathrm{E}[X|Y=y]=\sum_x x p_{X|Y}(x|y).$$

§ 6 条件分布列

- 设 $A_1, \cdots, A_n$ 是互不相容的事件并且形成样本空间的一个分割，假定 $\mathrm{P}(A_i) > 0$ 对一切 i 成立. 则

$$\mathrm{E}[X] = \sum_{i=1}^{n} \mathrm{P}(A_i)\mathrm{E}[X|A_i].$$

进一步假定事件 B 满足对一切 i, $\mathrm{P}(A_i \cap B) > 0$, 则

$$\mathrm{E}[X|B] = \sum_{i=1}^{n} \mathrm{P}(A_i|B)\mathrm{E}[X|A_i \cap B].$$

- 我们有

$$\mathrm{E}[X] = \sum_{y} p_Y(y)\mathrm{E}[X|Y = y].$$

全期望定理

§6 条件分布列

验证第一个公式

$$
\begin{aligned}
\mathrm{E}[X] &= \sum_x x p_X(x) \\
&= \sum_x x \sum_{i=1}^n \mathrm{P}(A_i) p_{x|A_i}(x|A_i) \\
&= \sum_{i=1}^n \mathrm{P}(A_i) \sum_x x p_{x|A_i}(x|A_i) \\
&= \sum_{i=1}^n \mathrm{P}(A_i) \mathrm{E}[X|A_i].
\end{aligned}
$$

§ 6 条件分布列

例 2.17 (几何随机变量的均值和方差) 你一次又一次地写一个计算机软件, 每写一次都有一个成功的概率 p. 假定每次成功与否与以前的历史记录相互独立. 令 X 是你一直到成功为止所写的次数 (最后一次你成功了！). X 的期望和方差是多少？

由于 X 是一个几何随机变量, 其分布列为

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

X 的均值和方差的公式是

$$\mathrm{E}[X] = \sum_{k=1}^{\infty} k(1-p)^{k-1}p, \quad \mathrm{var}(X) = \sum_{k=1}^{\infty}(k - \mathrm{E}[X])^2(1-p)^{k-1}p.$$

§ 6 条件分布列

利用全期望定理进行计算

记 $A_1 = \{X = 1\} = \{$第一次就写成功$\}$, $A_2 = \{X > 1\} = \{$第一次没有成功$\}$.

$\mathrm{E}[X|X=1] = 1 \qquad \mathrm{E}[X|X>1] = 1 + \mathrm{E}[X]$

因此, 由全期望定理

$$\begin{aligned}\mathrm{E}[X] &= \mathrm{P}(X=1)\mathrm{E}[X|X=1] + \mathrm{P}(X>1)\mathrm{E}[X|X>1] \\ &= p + (1-p)(1+\mathrm{E}[X]).\end{aligned}$$

由此可得

$$\mathrm{E}[X] = \frac{1}{p}.$$

§6 条件分布列

相似地, 我们有

$$\mathrm{E}[X^2|X=1]=1, \quad \mathrm{E}[X^2|X>1]=\mathrm{E}[(1+X)^2]=1+2\mathrm{E}[X]+\mathrm{E}[X^2],$$

故

$$\mathrm{E}[X^2]=p\cdot 1+(1-p)(1+2\mathrm{E}[X]+\mathrm{E}[X^2]),$$

从而

$$\mathrm{E}[X^2]=\frac{1+2(1-p)\mathrm{E}[X]}{p},$$

再利用 $\mathrm{E}[X\ ]=1/p$, 得到

$$\mathrm{E}[X^2]=\frac{2}{p^2}-\frac{1}{p}.$$

最后我们得到

$$\operatorname{var}(X)=\mathrm{E}[X^2]-(\mathrm{E}[X])^2=\frac{2}{p^2}-\frac{1}{p}-\frac{1}{p^2}=\frac{1-p}{p^2}.$$

§6 条件分布列

例 2.18 (两个信封的悖论)

主持人给你两个信封, 并且告诉你两个信封里有现金, 其中一个信封里的钱是另一个信封里的 m 倍 ($m > 1$, 且是一个整数). 当你打开其中一个信封, 看到信封里面的钱数以后, 你可以收下这个信封里面的钱作为你的奖金, 也可以要另一个信封里的钱作奖金. 有什么好的策略可使你拿到较多的奖金?

§ 6 条件分布列

推理一:

令 A 是你打开的信封, B 是你可能换的信封. 令 x 和 y 分别为信封 A 和 B 中的钱数.

y 的期望值为

$$\frac{1}{2}\cdot\frac{x}{m}+\frac{1}{2}\cdot mx=\frac{1}{2}\left(\frac{1}{m}+m\right)x=\frac{1+m^2}{2m}x>x.$$

你应该总是转向信封 B. 当你转向 B 的时候, 由于同样的理由, 又得转回

§6 条件分布列

推理二:

X 和 Y 的联合分布率

$$p_{X,Y}(mz,z)=p_{X,Y}(z,mz)=\frac{1}{2}p_Z(z),$$

对于不具有 (mz,z) 或 (z,mz) 的形式的 (x,y),

$$p_{X,Y}(x,y)=0.$$

用以下换信封的规则: 换信封的充要条件为 $\mathrm{E}[Y|X=x]>x$.

§6 条件分布列

抛掷一枚均匀的硬币，直到出现正面为止. 记 N 为抛掷硬币的次数. 此时你将 m^N 元放进一个信封内，将 m^{N-1} 元放进另一个信封内.

$$\frac{P(Y=m^{m+1}|X=m^n)}{P(Y=m^{m-1}|X=m^n)}=\frac{P(Y=m^{m+1},X=m^n)}{P(Y=m^{m-1},X=m^n)}=\frac{P(N=n+1)}{P(N=n)}=\frac{1}{2}$$

$$P(Y=m^{m-1}|X=m^n)=\frac{2}{3},\quad P(Y=m^{m+1}|X=m^n)=\frac{1}{3},$$

$$E[\text{信封 } B \text{ 中的钱数}|X=m^n]=\frac{2}{3}m^{n-1}+\frac{1}{3}m^{n+1}=\frac{2+m^2}{3m}\cdot m^n.$$

$(2+m^2)/(3m)>1$ 的充要条件是 $m^2-3m+2>0$ 或 $(m-1)(m-2)>0$.

若 $m>2$, 为了获得最大的期望奖金，你应该转向信封 B.

§7 独立性

7.1 随机变量与事件的相互独立性

刻画条件的事件的发生与否不会对随机变量取值提供新的信息.

随机变量 X 独立于事件 A

$$\mathrm{P}(X=x \text{和} A)=\mathrm{P}(X=x)\mathrm{P}(A)=p_X(x)\mathrm{P}(A)$$

$$p_{X|A}(x)=p_X(x), \quad \forall x.$$

§7 独立性

7.2 随机变量之间的相互独立性

随机变量 X 和 Y 称为**相互独立**的随机变量,

$$p_{X,Y}(x,y) = p_X(x)p_Y(y), \quad \forall x, y.$$

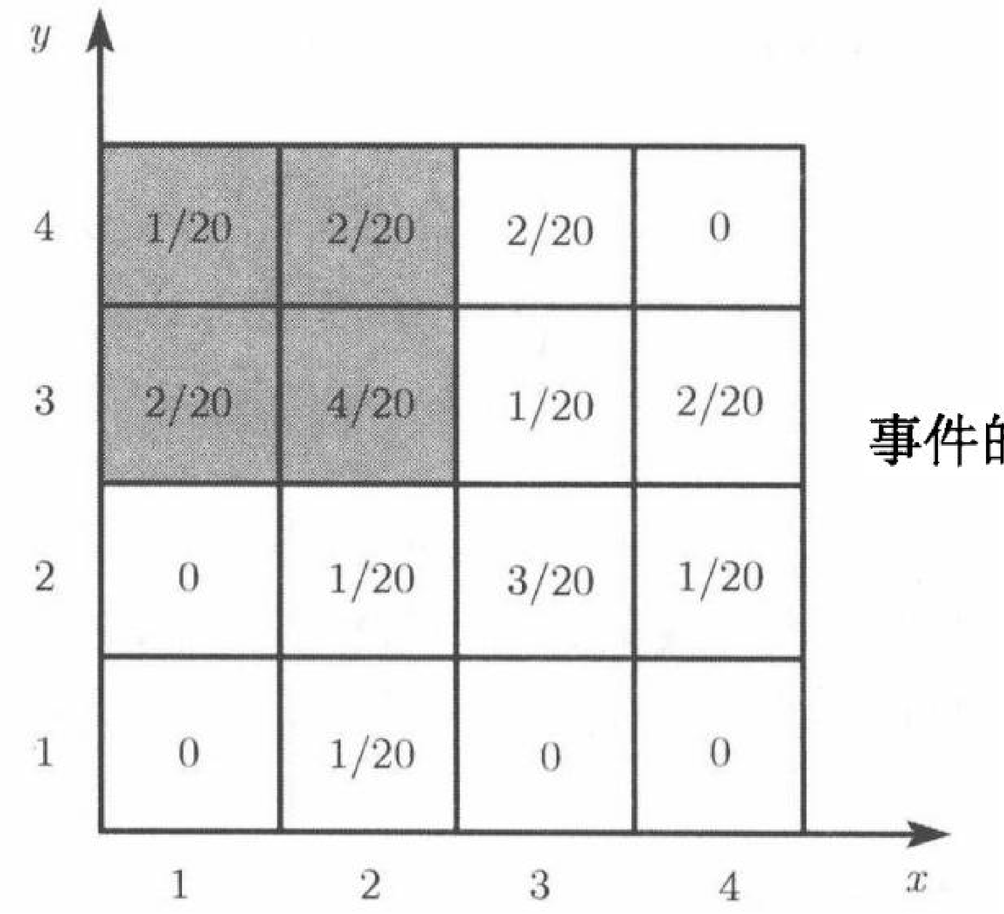
$$p_{X|Y}(x|y) = p_X(x), \quad \forall x \text{和一切满足} p_Y(y) > 0 \text{的} y.$$

随机变量 X 和 Y 是条件独立的,

$$\mathrm{P}(X = x, Y = y|A) = \mathrm{P}(X = x|A)\mathrm{P}(Y = y|A), \quad \forall x \text{ 和 } y,$$

$$p_{X|Y,A}(x|y) = p_{X|A}(x), \quad \forall x \text{ 和 } y, \text{ 但 } y \text{ 必须满足条件 } p_Y(y) > 0.$$

§7 独立性



事件的条件独立性并不包含独立性，反之亦然.

§7 独立性

讨论1:

若 X 和 Y 相互独立, 则对任意函数 g 和 h, $\mathrm{E}[g(X)h(Y)] = \mathrm{E}[g(X)]\mathrm{E}[h(Y)]$.

由 X 和 Y 的相互独立性可以蕴涵 $g(X)$ 和 $h(Y)$ 的相互独立性.

讨论2:

独立随机变量的和的方差等于它们的方差之和.

§ 7 独立性

$$
\begin{aligned}
\mathrm{E}[XY] &= \sum_x \sum_y xyp_{X,Y}(x,y) \\
&= \sum_x \sum_y xyp_X(x)p_Y(y) \quad \text{(根据独立性)} \\
&= \sum_x xp_X(x) \sum_y yp_Y(y) \\
&= \mathrm{E}[X]\mathrm{E}[Y].
\end{aligned}
$$

若 X 和 Y 相互独立, 则对任意函数 g 和 h, $\mathrm{E}[g(X)h(Y)] = \mathrm{E}[g(X)]\mathrm{E}[h(Y)]$.

§7 独立性

令 $\tilde{X} = X - \mathrm{E}[X], \tilde{Y} = Y - \mathrm{E}[Y]$,

$$\begin{aligned}
\operatorname{var}(X+Y) &= \operatorname{var}(\tilde{X}+\tilde{Y}) \\
&= \mathrm{E}[(\tilde{X}+\tilde{Y})^2] \\
&= \mathrm{E}[\tilde{X}^2 + 2\tilde{X}\tilde{Y} + \tilde{Y}^2] \\
&= \mathrm{E}[\tilde{X}^2] + 2\mathrm{E}[\tilde{X}\tilde{Y}] + \mathrm{E}[\tilde{Y}^2] \\
&= \operatorname{var}(\tilde{X}) + \operatorname{var}(\tilde{Y}) \\
&= \operatorname{var}(X) + \operatorname{var}(Y).
\end{aligned}$$

§7 独立性

关于独立随机变量的性质的小结

设在某一试验中, A 是一个事件, 满足条件 $\mathrm{P}(A)>0$, 又设 X 和 Y 是在同一个试验中的两个随机变量.

- 称 X 为相对于事件 A 独立, 如果满足

$$p_{X|A}(x)=p_X(x), \quad \text{对一切 } x \text{ 成立},$$

即对一切 x, 事件 $\{X=x\}$ 与 A 相互独立.

- 称 X 和 Y 为相互独立的随机变量, 如果对一切可能的数对 (x,y), 事件 $\{X=x\}$ 和 $\{Y=y\}$ 相互独立, 或等价地

$$p_{X,Y}(x,y)=p_X(x)p_Y(y), \quad \text{对一切 } x,y \text{ 成立}.$$

§7 独立性

- 若X 和 Y 相互独立, 则

$$\mathrm{E}[XY] = \mathrm{E}[X]\mathrm{E}[Y].$$

进一步地, 对于任意函数 g 和 h, 随机变量 $g(X)$ 和 $h(Y)$ 也是相互独立的, 并且

$$\mathrm{E}[g(X)h(Y)] = \mathrm{E}[g(X)]\mathrm{E}[h(Y)].$$

- 若 X 和 Y 相互独立, 则

$$\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y).$$

§7 独立性

7.3 几个随机变量的相互独立性

随机变量 X, Y 和 Z 是三个相互独立的随机变量,

$$p_{X,Y,Z}(x,y,z) = p_X(x)p_Y(y)p_Z(z), \quad \text{对一切 } x,y,z \text{ 成立}.$$

任何形如 $f(X), g(Y), h(Z)$ 的三个随机变量也是相互独立的.

§7 独立性

7.4 几个相互独立的随机变量的和的方差

设 $X_1, \cdots, X_n$ 为相互独立的随机变量序列, 则

$$\operatorname{var}(X_1 + \cdots + X_n) = \operatorname{var}(X_1) + \cdots + \operatorname{var}(X_n).$$

应用:

例 2.21 (样本均值的期望和方差) 我们希望估计总统的支持率. 为此, 我们随机地选取 n 个选民, 询问他们的看法. 令 X_i 表示第 i 个被问的选民的态度:

$$X_i = \begin{cases} 1, & \text{若第 } i \text{ 个被问的选民支持总统,} \\ 0, & \text{若第 } i \text{ 个被问的选民不支持总统.} \end{cases}$$

§7 独立性

样本均值

$$S_n = \frac{X_1 + \cdots + X_n}{n}$$

$$\mathrm{E}[S_n] = \sum_{i=1}^{n} \frac{1}{n}\mathrm{E}(X_i) = \frac{1}{n}\sum_{i=1}^{n} p = p,$$

$$\mathrm{var}(S_n) = \sum_{i=1}^{n} \frac{1}{n^2}\mathrm{var}(X_i) = \frac{p(1-p)}{n}.$$

作业

21. 圣彼得堡悖论. 抛掷一枚均匀的硬币, 直到出现反面向上为止. 假定每次抛掷是独立的. 若你抛掷了 n 次, 你可以获得 2^n 元. 你得到的钱数的期望值是多少? 你愿意付多少钱玩这个游戏呢?