



离散数学

Discrete Mathematics

第十四讲：离散随机变量及其数学期望

吴楠

南京大学计算机学院



2025 年 4 月 8 日



前情提要



- 条件概率
- 条件概率的应用
- 全概率公式及其应用
- 贝叶斯定理及其应用
- 事件的独立性

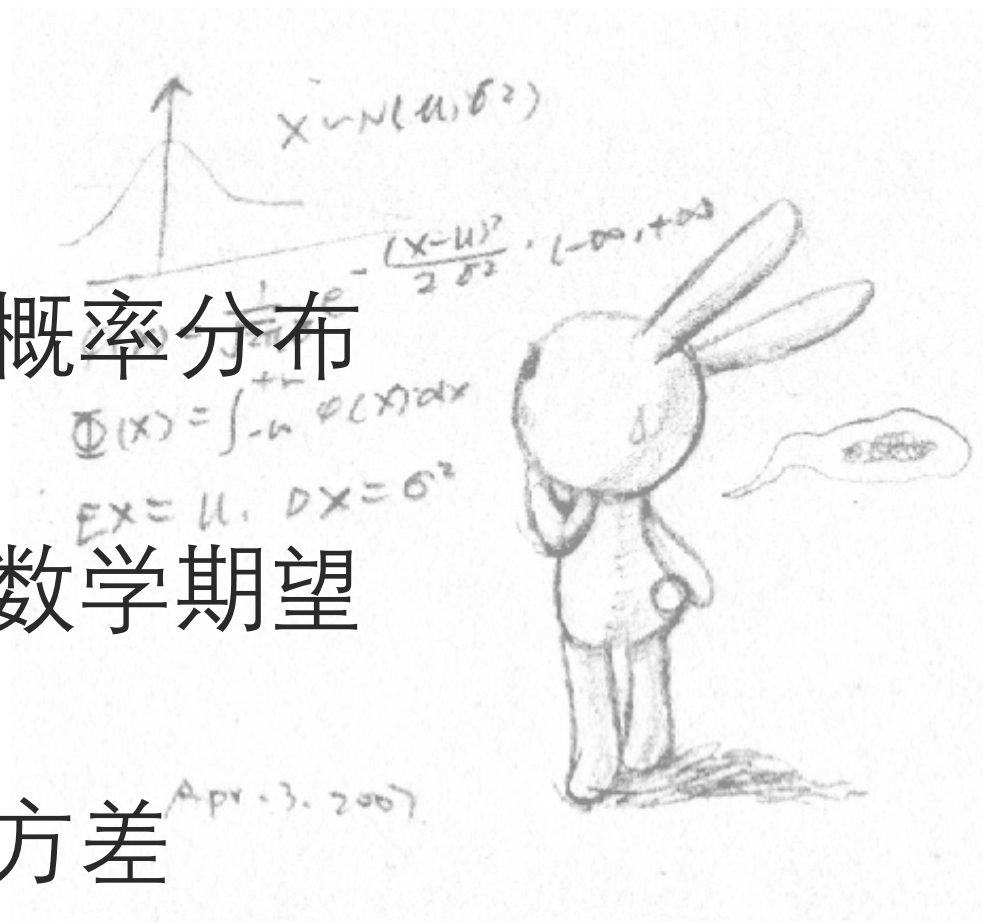




本讲主要内容



- 离散随机变量
- 离散随机变量的概率分布
- 离散随机变量的数学期望
- 离散随机变量的方差





为什么引入随机变量?



- 随机试验的结果多用文字描述，不利于数学处理
- 用实数表示随机试验 E 中的试验结果，则可建立由样本空间 Ω 到实数上的实值函数，这样该函数便可表示试验的结果并可作为变量参与计算
- 例如，在掷硬币试验中， $\Omega = \{\text{正面向上}, \text{反面向上}\}$

令函数 $X(\omega) = \begin{cases} 1 & (\omega = \text{正面向上}) \\ 0 & (\omega = \text{反面向上}) \end{cases}$ ，则 $X: \Omega \rightarrow \{0,1\}$

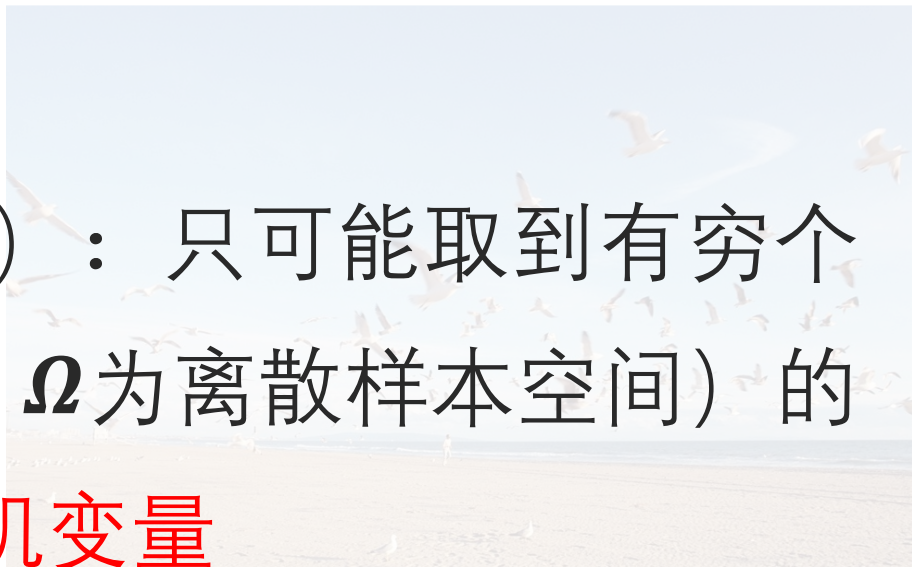
构成一个可表征掷硬币试验结果的变量



离散随机变量



- **定义 (随机变量)** : 设随机试验的样本空间为 Ω , 称定义在 Ω 上的实值函数 $X: \Omega \rightarrow \mathbb{R}$ 为 **随机变量** (random variable)。通常把函数 $X(\omega)$ 简记作 X
- **定义 (离散随机变量)** : 只可能取到有穷个或可数无穷个值 (*i.e.* Ω 为离散样本空间) 的随机变量称作 **离散随机变量**





离散随机变量的概率分布



- 设离散随机变量 X 可能取到的值为 a_1, a_2, \dots (有穷个或可数无穷个), 称 X 在所有取值处的概率:

$$P\{X = a_k\} = p_k, k = 1, 2, \dots$$

为 X 的 (离散) 概率分布 (probability distribution)

- 离散概率分布的性质:

- (1) $\forall k, 0 \leq p_k \leq 1$

- (2) $\sum_k p_k = 1$

- 例: 掷硬币试验 $P\{X = 1\} = 0.5, P\{X = 0\} = 0.5$



离散随机变量的概率分布 (续)



- **例：** 在一个袋子中有10个球，其中6个白球，4个红球。从中任取3个，求抽到红球数量的概率分布
- **解：** 设离散随机变量 X 为抽到的红球数，则 X 所有可能的取值为0, 1, 2, 3。故 X 的概率分布为：

$$P\{X = 0\} = \frac{\binom{6}{3}}{\binom{10}{3}} = 1/6, \quad P\{X = 1\} = \frac{\binom{4}{1}\binom{6}{2}}{\binom{10}{3}} = 1/2, \quad P\{X = 2\} = \frac{\binom{4}{2}\binom{6}{1}}{\binom{10}{3}} = 3/10,$$
$$P\{X = 3\} = \frac{\binom{4}{3}}{\binom{10}{3}} = 1/30, \quad \text{还可表为: } P\{X = k\} = \frac{\binom{4}{k}\binom{6}{3-k}}{\binom{10}{3}}, k = 0, 1, 2, 3$$

X	0	1	2	3
p_k	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{3}{10}$	$\frac{1}{30}$



离散随机变量的独立性



- **定义** (**独立离散随机变量**) : 样本空间 Ω 上的2个离散随机变量 X, Y 若满足:

$$P(X = r_1 \wedge Y = r_2) = P(X = r_1) \cdot P(Y = r_2)$$

则称随机变量 X 与 Y 相互独立

- 抛两个匀质骰子，第一个骰子的点数与第二个骰子的点数二者是否相互独立？上述情况下，第一个骰子的点数与两个骰子的点数之和二者是否相互独立？

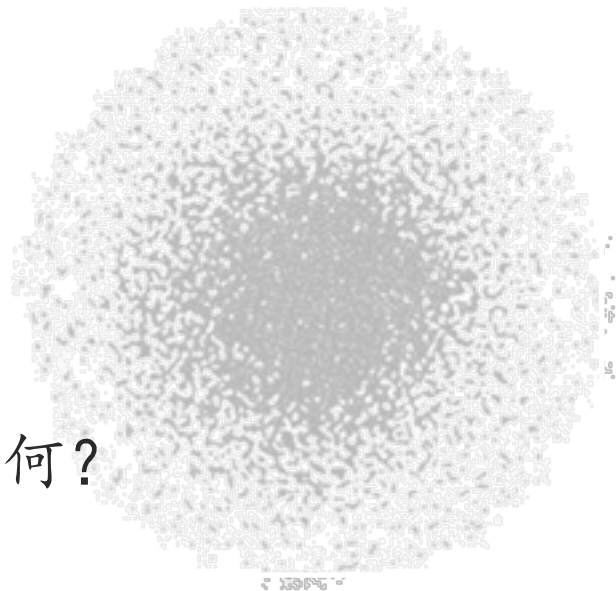


离散随机变量的分布特征



- 由于离散随机变量的分布是随机的，故其取值的规律可能非常复杂，如何刻画随机变量取值的整体特征呢？

- 一个班级学生的身高情况如何？
- 一个系微积分期中考试的整体情况如何？





离散随机变量的期望与方差



- **整体的取值情况：**常用**均值**来衡量一个离散变量整体取值的分布取向，即事件最**期望**产生的结果
 - 平均身高、平均成绩
- **个体之间的区别情况：**常用均值作为基准，考虑个体与基准的偏差程度



离散随机变量的期望



- 等概率分布的随机变量的均值一般用“平均”表示，而一般随机变量的“均值”是指**概率加权平均**，即**数学期望**（mathematical expectation），也称为**期望值**（expected value, EV）
- **定义（数学期望）**：样本空间 Ω 上的一个随机变量 X 的数学期望为

$$E(X) = \sum_{s \in \Omega} p(X = s)X(s)$$





离散随机变量的期望（续）

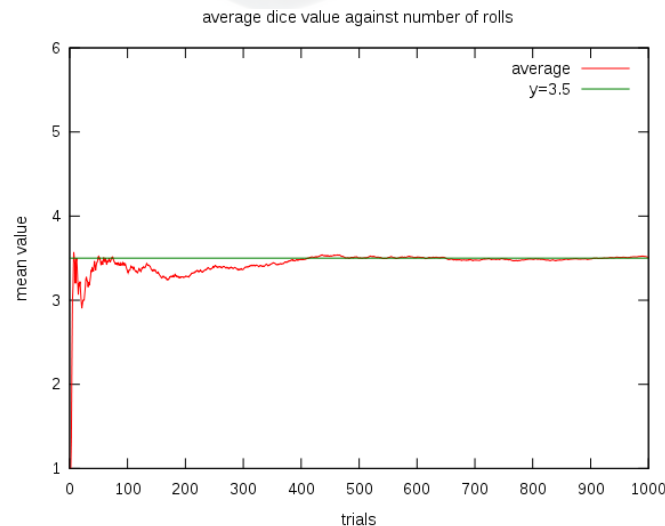


■ 例1：随机投1枚匀质骰子，求投出点数的数学期望

○ 解：设离散随机变量 X 为投出的点数，则

$$E(X) = \sum_{i=1}^6 \frac{1}{6} \cdot X(i) = \frac{1}{6} \cdot \sum_{i=1}^6 i = \frac{7}{2}$$

- 模拟实验结果（Matlab）：
- 本例的数学期望就是平均值





离散随机变量的期望 (续)



- **例2:** 随机掷3枚匀质硬币，求掷出头面向上硬币个数的数学期望

○ **解:** 设离散随机变量 X 为掷出的正面向上硬币的个数，则 $E(X)$

$$\begin{aligned} &= \frac{1}{2^3} [X(\text{HHH}) + X(\text{HHT}) + X(\text{HTH}) + X(\text{THH}) + X(\text{TTH}) + X(\text{THT}) \\ &\quad + X(\text{HTT}) + X(\text{TTT})] = \frac{1}{8} (3 + 2 + 2 + 2 + 1 + 1 + 1 + 0) = \frac{3}{2} \end{aligned}$$

- 以上2例为离散随机变量呈均匀分布之情况



离散随机变量的期望 (续)



- **例3:** 随机投2枚匀质骰子, 求2个骰子所投出点数之和的数学期望

○ **解:** 设离散随机变量 X 为2个骰子投出的点数之和, 则

$$p(X = 2) = p(X = 12) = 1/36,$$

$$p(X = 3) = p(X = 11) = 2/36 = 1/18,$$

$$p(X = 4) = p(X = 10) = 3/36 = 1/12,$$

$$p(X = 5) = p(X = 9) = 4/36 = 1/9,$$

$$p(X = 6) = p(X = 8) = 5/36,$$

$$p(X = 7) = 6/36 = 1/6.$$





离散随机变量的期望 (续)



- 例3：随机投2枚匀质骰子，求2个骰子所投出点数之和的数学期望

○ 解（续）：故有

$$\begin{aligned} E(X) &= 2 \cdot \frac{1}{36} + 3 \cdot \frac{1}{18} + 4 \cdot \frac{1}{12} + 5 \cdot \frac{1}{9} + 6 \cdot \frac{5}{36} + 7 \cdot \frac{1}{6} \\ &\quad + 8 \cdot \frac{5}{36} + 9 \cdot \frac{1}{9} + 10 \cdot \frac{1}{12} + 11 \cdot \frac{1}{18} + 12 \cdot \frac{1}{36} \\ &= 7. \end{aligned}$$





多个离散随机变量期望值的性质



- **定理**（**数学期望的线性性质**）：对于**任意**离散随机变量 X_i ($i = 1, 2, \dots, n$)及任意 $a, b \in \mathbb{R}$ ，有

- $E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n)$;

- $E(aX_i + b) = aE(X_i) + b$

由随机变量的线性特征和数学期望的定义易证

- 注意，这里**并不**要求各随机变量之间相互独立
- 试用此定理考虑随机抛双骰子的点数和的期望

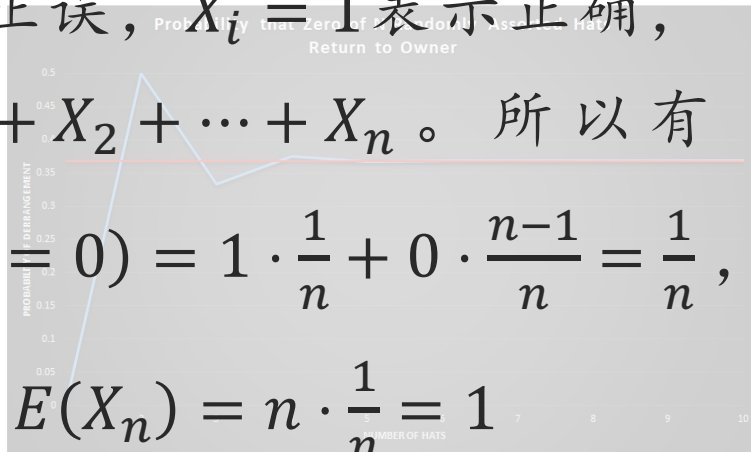


多个离散随机变量期望值的性质 (续)



- **例1 (hat-check problem)** : 负责寄存帽子的服务生把客人的帽子弄乱了, 只能随机发还。求其发还正确的客人数的数学期望

- **解:** 令离散随机变量 X 表示客人拿到正确的帽子的数量, X_i 表示第 i 位客人拿到帽子的正误, $X_i = 1$ 表示正确, $X_i = 0$ 表示错误, 故 $X = X_1 + X_2 + \cdots + X_n$ 。所以有
$$E(X_i) = 1 \cdot p(X_i = 1) + 0 \cdot p(X_i = 0) = 1 \cdot \frac{1}{n} + 0 \cdot \frac{n-1}{n} = \frac{1}{n},$$
故
$$E(X) = E(X_1) + E(X_2) + \cdots + E(X_n) = n \cdot \frac{1}{n} = 1$$





多个离散随机变量期望值的性质 (续)



- **例2** (线性查表算法的平均查找次数) : 给定1个元素 x 及1个长度为 n 且元素互异的线性表, 线性搜索算法通过比较依次检查表中的元素, 看是否为 x , 直至找到或者所有元素都查完但未找到 x 为止。求平均查找 (比较) 的次数

```
int lsearch( int x, int ltable[] )
{
    int n = sizeof(ltable) / sizeof(int);
    int i = 1;
    while ( i <= n && x != ltable[i] )
        i ++;
    if ( i <= n ) return i
    else return 0;
}
```



多个离散随机变量期望值的性质 (续)



- **解：** 设元素 x 出现在线性表中的概率为 p ，且目标元素等可能地出现在每个位置，令 $q = 1 - p$ ，则有：

$$\begin{aligned} E &= \frac{3p}{n} + \frac{5p}{n} + \cdots + \frac{(2n+1)p}{n} + (2n+2)q \\ &= \frac{p}{n}(3 + 5 + \cdots + (2n+1)) + (2n+2)q \\ &= \frac{p}{n}((n+1)^2 - 1) + (2n+2)q \\ &= p(n+2) + (2n+2)q. \end{aligned}$$

```
int lsearch( int x, int ltable[] )
{
    int n = sizeof(ltable) / sizeof(int);
    int i = 1;
    while ( i <= n && x != ltable[i] )
        i ++;
    if ( i <= n ) return i
    else return 0;
}
```



独立离散随机变量的数学期望



- 定理（独立离散随机变量的数学期望）：离散样本空间 Ω 上相互独立的随机变量 X 与 Y ，其数学期望满足

$$E(XY) = E(X)E(Y)$$

由离散随机变量相互独立的定义和数学期望的定义易

证



离散随机变量的方差



- 方差是刻画离散随机变量的值随机分布范围的统计量，直观分析，方差是随机变量的各个取值与“基线标准”——数学期望的“差值”（平方运算统一了正负两个方向的差异）
- **定义（离散随机变量的方差）**：设 X 是离散样本空间 Ω 上的随机变量，则 X 的方差（variance）记为

$$V(X) = E \left[(X - E(X))^2 \right] = \sum_{s \in \Omega} (X(s) - E(s))^2 p(s)$$



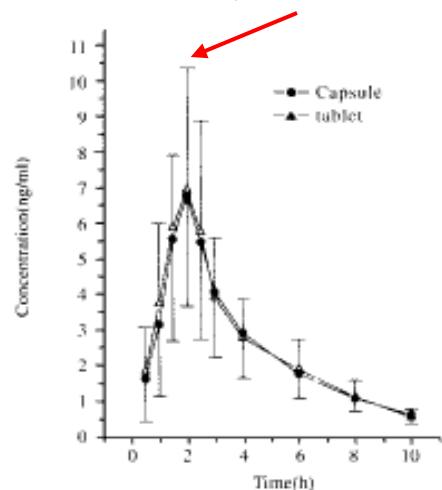
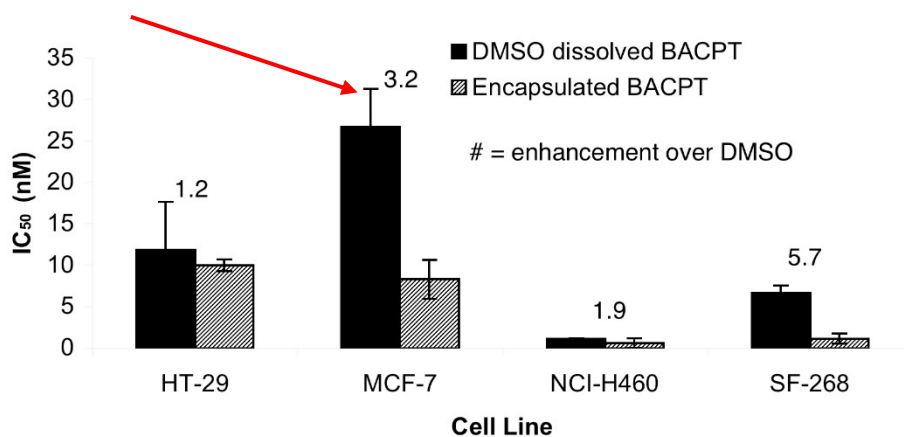
离散随机变量的方差



- **定义（标准差）**：设 X 是离散样本空间 Ω 上的随机变量，则 X 的标准差（standard deviation）记为

$$\sigma(X) = \sqrt{V(X)}$$

均值和标准差往往联合反映统计变量的基线和分布情况





离散随机变量的方差（续）



- 定理（离散随机变量的方差与数学期望的关系）：设

X 是离散样本空间 Ω 上的随机变量，则

$$V(X) = E(X^2) - E(X)^2$$

- 证明：

$$\begin{aligned} V(X) &= \sum_{s \in \Omega} (X(s) - E(X))^2 p(s) = \\ &= \sum_{s \in \Omega} X(s)^2 p(s) - 2E(X) \sum_{s \in \Omega} X(s) p(s) + E(X)^2 \sum_{s \in \Omega} p(s) = E(X^2) \\ &\quad - 2E(X)E(X) + E(X)^2 \cdot 1 = E(X^2) - E(X)^2 \end{aligned}$$

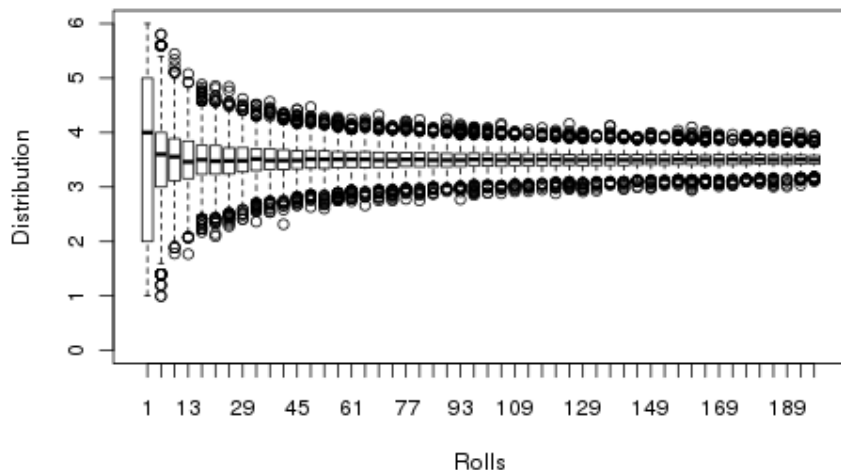


离散随机变量的方差（续）



- **例：**抛一枚匀质骰子，求抛出点数的方差
- **解：**设离散随机变量 X 为抛出的点数，则有

$$V(X) = E(X^2) - E(X)^2$$
$$= \frac{1}{6}(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$





期望与方差的关系*



请谈谈“准 (bias) ”与“确 (variance) ”的辩证关系



"准" (bias) 和"确" (variance) 是统计学和机器学习中两个重要的概念，它们描述了模型的预测性能和泛化能力之间的权衡关系。在机器学习中，理解和平衡准和确之间的关系对于构建高性能的模型至关重要。



首先，"准" (bias) 是模型对于真实数据生成过程的偏差，即模型的预测与真实值之间的差距。偏差较大的模型通常会在训练集上表现得较差，因为它们不能捕捉到数据生成过程的真实特征。这种情况称为欠拟合，模型在训练集上的性能较差，无法很好地适应训练数据。

其次，"确" (variance) 是模型对于输入数据的敏感性，即模型的预测在不同的训练集上可能会有较大的波动。方差较大的模型通常在训练集上表现得较好，因为它们能够较好地适应训练数据，但在未见过的数据上的性能可能较差。这种情况称为过拟合，模型在训练集上表现得很好，但在未见过的数据上的泛化性能较差。

准和确之间存在一种辩证关系。当模型的偏差较大时，它可能无法捕捉到数据生成过程的真实特征，从而导致欠拟合。而当模型的方差较大时，它对于输入数据非常敏感，可能在训练集上表现得很好，但在未见过的数据上性能较差，从而导致过拟合。

在实际应用中，我们需要在准和确之间取得平衡，以构建具有良好泛化性能的模型。一种方法是通过调整模型的复杂度，例如增加或减少模型的层数、节点数等，来平衡准和确。较简单的模型通常具有较高的偏差但较低的方差，而较复杂的模型则可能具有较低的偏差但较高的方差。此外，还可以通过使用正则化技术、增加训练数据量、进行交叉验证等方法来改善模型的泛化性能。

需要注意的是，准和确不是完全独立的，它们之间存在权衡和折衷。在实际模型选择和调优过程中，了解准和确

↻ Regenerate response





离散随机变量的方差（续）



- **定理（独立随机变量的方差）**：设 X, Y 是离散样本空间 Ω 上相互独立的随机变量，则

$$V(X + Y) = V(X) + V(Y)$$

- **定理（Bienaymé, 1853）**：设 X_i 是离散样本空间 Ω 上的一组两两相互独立的随机变量， $i = 1, 2, \dots, n$ ，则

$$V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i)$$

以上两定理由两变量方差和期望的性质及其推广易证



本次课后作业



- 教材内容：[Rosen] 7.4.1—7.4.4节，7.4.6 节
- 课后习题：
 - Problem Set 13 & 14
- 提交时间：4月15日 10:00前

