离散概率 (I)

离散数学课程组

南京大学计算机科学与技术系

内容提要

- 引论
- 概率论

引论

- 起源于17世纪的赌博游戏
- 奠基人之一：法国数学家拉普拉斯
  - Pierre-Simon Laplace (1749-1827)
- 军事、经济、政治、社会
  - 运筹帷幄
- 计算机科学：算法设计与分析
  - 设计概率算法
  - 分析算法的平均复杂度

有限概率

- 试验
  - 从一组可能的结果中得出一个结果的过程
- 试验的样本空间
  - 可能结果的集合
- 一个事件
  - 样本空间的一个子集

有限概率

- 事件的概率
  - 如果S是结果具有相等可能性的有限样本空间，E是其中的一个事件，即是S的一个子集，则事件E的概率是

$$p(E) = |E|/|S|$$

- 易见 $0 \le p(E) \le 1$

举例

- 掷两个骰子，点数之和等于7的概率？
  - $|S|=6*6=36$
  - $|E|=?$
    - $(x, y), x+y=7, x\geq 1, y\geq 1$
    - $(x', y'), x'+y'=5, x'\geq 0, y'\geq 0$
    - $C(5+2-1, 5)=C(6,1)=6$
  - $p(E) = |E|/|S|=6/36=1/6$

举例

- 标准扑克牌52张，一手牌（5张牌组成）出现满堂红(AAAKK)，即3张同一类且其余2张在另一类，概率是多少？
  - P(13, 2)C(4, 3)C(4, 2)/C(52, 5)≈0.0014
- 标准扑克牌52张，一手牌（5张牌组成）正好含有4种相同面值(AAAAK)的概率是多少？
  - C(13, 1)C(4, 4)C(48, 1)/C(52, 5)≈0.00024

事件组合的概率

定理 1: 设E是样本空间S中的一个事件。事件 $\overline{E}$ (事件E的补事件)的概率是

$$p(\overline{E}) = 1 - p(E).$$

证明: $|\overline{E}| = |S| - |E|,$

$$p(\overline{E}) = \frac{|S| - |E|}{|S|} = 1 - \frac{|E|}{|S|} = 1 - p(E).$$

事件组合的概率（举例）

随机生成10位0-1串，其中至少一位为0的概率?

解: 设E是10位中至少一位是0的事件。事件E的补事件是所有位都是1的事件。

$$p(E) = 1 - p(\overline{E}) = 1 - \frac{|\overline{E}|}{|S|} = 1 - \frac{1}{2^{10}} = 1 - \frac{1}{1024} = \frac{1023}{1024}.$$

事件组合的概率（续）

定理2: 设E_1和E_2是样本空间S中的事件。那么

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$$

证明： $p(E_1 \cup E_2) = \frac{|E_1 \cup E_2|}{|S|} = \frac{|E_1|+|E_2|-|E_1 \cap E_2|}{|S|}$

$$= \frac{|E_1|}{|S|} + \frac{|E_2|}{|S|} - \frac{|E_1 \cap E_2|}{|S|}$$

$$= p(E_1) + p(E_2) - p(E_1 \cap E_2). \blacktriangleleft$$

事件组合的概率（举例）

从不超过100的正整数中随机选一个，它能被2或5整除的概率?

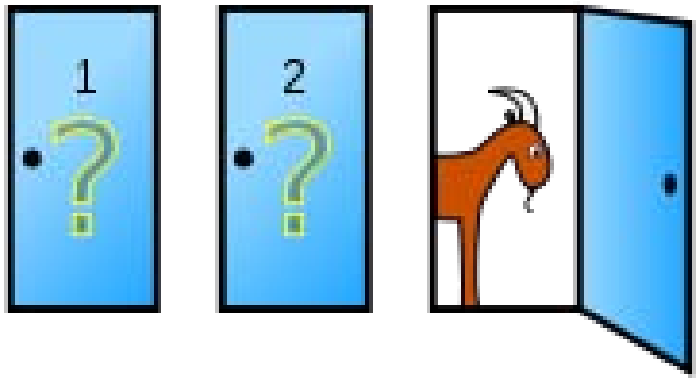
解: 设E_1是选出一个被2整除的事件，E_2是选出一个被5整除的事件。则$E_1 \cap E_2$是选出一个被10整除的事件。

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$$

$$= 50/100 + 20/100 - 10/100 = 3/5.$$

概率推理

- 常见问题：两个事件中哪个更有可能发生？
- 蒙蒂·霍尔游戏(电视节目, Monty Hall Puzzle)



- 1/3 （总是不变）
- 2/3 （总是改变）

概率论

概率指派

事件的组合

条件概率

独立性

伯努利（**Bernoulli** ）试验与二项分布

随机变量

蒙特卡洛（**Monte Carlo**）算法

概率指派

假定：各种结果的可能性都是相等的. 取消这个假定。

- 设S是某个具有可数个结果的试验的样本空间。我们赋给每个结果s一个概率 $p(s)$，满足下列条件：

  i. $0 \le p(s) \le 1 \ (\forall s \in S)$

  ii. $\sum_{s \in S} p(s) = 1$

  这个函数p:S→[0,1]称为概率分布.

概率指派（举例）

掷一枚均匀的硬币，结果H(heads) 和 T (tails) 应该赋予什么概率?对于一枚不均匀的硬币，如果如果头像向上常常是头像向下的两倍，相应的概率又如何?

解: 我们有$p(H) = 2p(T)$.

因为 $p(H) + p(T) = 1$,

所以 $2p(T) + p(T) = 3p(T) = 1$.

从而, $p(T) = 1/3$， $p(H) = 2/3$.

均匀分布

定义: 假设*S* 是一个含n个元素的集合. 均匀分布(*uniform distribution*)赋给S中每个元素1/*n* 的概率.

举例: 对于均匀的硬币，$p(H) = p(T) = 1/2$.

一个事件的概率

定义:一个事件E的概率是E中各结果的概率之和.

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

举例：掷一个不均匀的骰子，3这一面出现的次数是其他面的两倍，其它五个面的出现是均等的，出现奇数点的概率？

解: 令E = {1,3,5}表示出现奇数点的事件. 我们有 $p(3) = 2/7$，$p(1) = p(2) = p(4) = p(5) = p(6) = 1/7$.

因此, $p(E) = p(1) + p(3) + p(5) = 4/7$.

事件的组合

- **补事件:** $p(\overline{E}) = 1 - p(E)$
- **事件的并:** $p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$
- **定理: 如果E_1, E_2, … 是样本空间S中两两互不相交的事件序列，那么**

$$p\left(\bigcup_i E_i\right) = \sum_i p(E_i)$$

条件概率

定义: 设E和F是事件，且$p(F) > 0$. E在给定F条件下的概率, 记作$p(E|F)$, 定义为:

$$p(E|F) = \frac{p(E \cap F)}{p(F)}$$

条件概率（举例）

随机产生的4位二进制串，16个串是均匀产生的，那么在第一位是0的条件下，串中含有2个连续0的概率？

解: 设E是串中含有2个连续0的事件, F是第一位为0的事件.

- $E \cap F$ = {0000, 0001, 0010, 0011, 0100}.
- $p(E \cap F)=5/16$.
- $p(F) = 8/16 = ½$.
- 所以, $$p(E|F) = \frac{p(E \cap F)}{p(F)} = \frac{5/16}{1/2} = \frac{5}{8}.$$

条件概率（举例）

在至少有一个男孩的条件下，有两个孩子的家庭正好均是男孩的条件概率？假设BB, BG, GB, 和GG是等可能的。

解: 令E是家庭有两个男孩的事件，F是家庭至少有一个男孩的事件。我们有E = {BB}, F = {BB, BG, GB}, $E \cap F$ = {BB}.

- $p(F) = 3/4$， $p(E \cap F) = 1/4$.
- 因此， $$p(E|F) = \frac{p(E \cap F)}{p(F)} = \frac{1/4}{3/4} = \frac{1}{3}.$$

独立性

定义: 事件*E*和*F*是独立的，当且仅当

$$p(E \cap F) = p(E)p(F).$$

举例. 一个有两个孩子的家庭有四种情形 (**BB, GG, BG,GB**)，假设是等可能的。事件*E*是两个孩子的家庭有两个男孩，事件*F*是两个孩子的家庭至少有一个男孩。事件*E*和*F*是否独立？

解: $p(E) = 1/4$, $p(F) = 3/4$， $p(E \cap F)=1/4$,

$p(E)\, p(F)= 3/16 \neq 1/4= p(E \cap F)$. *E*和*F*不是独立的.

伯努利（Bernoulli ）试验

Jacob Bernoulli (**James** or **Jacques**)

瑞士数学家 (1654 – 1705)

定义: 假设一个试验只有两种可能的结果.

- 这种试验的每一次实行称为一次伯努利试验。
- 两种结果之一称为“成功”，另一种称为“失败”。
- 如果成功的概率为 p，失败的概率为q，那么$p + q = 1$.

伯努利试验（举例）

掷一个不均匀硬币，头像向上的概率为2/3。掷硬币7次，正好4次出现头像的概率是多少？

解：掷的7次中出现4次头像的方式共$C(7,4)$种.

每一种的概率为$(2/3)^4(1/3)^3$ ，因此，所求概率为：

$C(7,4)\,(2/3)^4(1/3)^3 = (35\cdot 16)/\,3^7 = 560/\,2187.$

在n次独立的伯努利试验中有k次成功的概率

定理2: 在成功概率为p、失败概率为$q = 1 - p$的n次独立的伯努利试验中，有k次成功的概率是

$$C(n,k)p^k q^{n-k}.$$

证明: n次伯努利试验的结果表示为n元组$(t_1,t_2,\ldots,t_n)$, $t_i = S$ (成功) 或 F (失败). 有$C(n,k)$个元组正好包含k个S，对应于一种k次成功的方式。每种的概率为$p^k q^{n-k}$. 所以，有k次成功的概率是

$$C(n,k)p^k q^{n-k}.$$ ◀

二项分布

- $b(k; n, p) = C(n,k)p^k q^{n-k}$
- 作为k的函数, $b(k; n, p)$ 称为二项分布

随机变量（random variable）

定义:一个随机变量是从样本空间到实数集的一个函数，也就是对每个可能结果指派一个实数。

- 一个随机变量是一个函数。它既不是一个变量，也不是随机的。
- In the late 1940s W. Feller and J.L. Doob flipped a coin to see whether both would use “random variable” or the more fitting “chance variable.” Unfortunately, Feller won and the term “random variable” has been used ever since.

随机变量

举例: 假设一个硬币被掷3次. 令 $X(t)$ 是头像在结果t中出现的次数。那么随机变量 $X(t)$ 取值如下:

$X(HHH) = 3,\ X(TTT) = 0,$

$X(HHT) = X(HTH) = X(THH) = 2,$

$X(TTH) = X(THT) = X(HTT) = 1.$

8种结果的每一个出现的概率为1/8. 因此, $X(t)$的(概率)分布

$p(X = 3) = 1/8,$

$p(X = 2) = 3/8,$

$p(X = 1) = 3/8,$

$p(X = 0) = 1/8.$

随机变量的分布

定义: X是样本空间S上的随机变量，X的分布是形如 $(r, p(X = r))$ 的二元组集合，其中

$r \in X(S)$,

$p(X = r)$是X取值为r的概率。

生日问题

- 367人中至少有2人生日相同。
- 人数？使得至少2人有相同生日的概率大于1/2

解: 假定生于某天是等可能的，相互独立的，一年有 366 天。p_{n}： n个人的生日互不相同的概率。

- $p_1 = 1$
- $p_n = (365/366)(364/366)\cdots(367-n)/366.$ n=2,..366
- $1-p_n = 1-(365/366)(364/366)\cdots(367-n)/366$
- $1-p_{22} \approx 0.457, 1-p_{23} \approx 0.506$。

$\boxed{1-p_{57} \approx 0.99}$

蒙特卡洛（Monte Carlo）算法

- 概率算法: 在一步或多步作随机选择的算法
- 蒙特卡洛算法：用来回答判定问题的概率算法
  - 一系列测试，每个测试结果为“真”或者“未知”
  - 如果有一个为“真”，算法结束，返回结果为“真”。
  - 指定的一系列测试完成，且每个测试结果都是“未知”，则算法输出“假”。

蒙特卡洛算法的一个例子

- 判定“n是合数吗？”（素数的概率测试）
  - 随机选取正整数$1 < b < n$，实施以b为基的米勒测试，如果测试失败（表明n必为合数），返回“真”。如果通过测试，返回“未知”。
  - 重复上述过程k次，若每次都是“未知”，则返回“假”。
  - //备注：合数通过米勒测试的概率小于1/4.
  - //一个合数通过k次米勒测试(b随机)的概率小于 $(1/4)^k$.
  - //k=10，误判的概率小于$1/10^6$.

作业

- 教材[6.1, 6.2]
  - p. 307： 8, 10, 12, 18, 38
  - p. 318： 8, 19, 20, 34, 35