概率统计B

Probability and Statistics

张思容
zhangsirong@buaa.edu.cn

数学与系统科学学院，北京航空航天大学
School of Mathematics and System Sciences, BUAA

September 22, 2014

1. 概率的计算
   - 集合运算
   - 概率计算
2. 条件概率
   - 条件概率
   - 全概率与贝叶斯公式
   - 计算与应用
3. 独立性与系统
   - 独立性
   - 系统与试验
   - 习题答疑

Review：回顾

RECALL:

- 概率模型=样本空间+概率律
- 公理化：有限到无穷 (数学分析工具!)
- 计数方法与生日问题。

TODAY

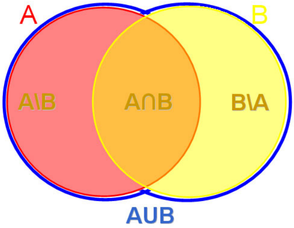
- 事件与集合，配对问题；
- 计算公式：加法和乘法；
- 条件概率 $\rightarrow$ 统计推断。

EXAMPLE (赌牌游戏)

有三张牌：一张两面都是红的，一张两面都是黑的，一张两面是一红一黑。随机取出一张，如果正面是红的，反面是黑还是红?
怎样赢钱?

事件与集合

- 集合Ω: $\omega \in \Omega$ 集合中的元素存在唯一判别法则;
  记子集$A = \{\omega | \omega \in A\}$. 记为$A \subset \Omega$;
- 集合运算: $\bar{A} = \Omega - A$, 乘法$A \cap B = AB$,"加法" $A \cup B$, 减法$A - B = A\bar{B}$
  对应的事件意义?
- 不相容事件: $A \cap B = \emptyset$.
  对立事件: $\bar{A}$
- 集合运算的规律: 交换律, 结合律;
  Venn韦恩图
- 有用公式:
  $A \cup B = A + \bar{A}B, A = AB + A\bar{B}$
  De-Morgan公式:
  $\overline{\bigcup_{j=1}^{j=n} A_i} = \bigcap_{j=1}^{j=n} \bar{A}_i.$



配对问题

EXAMPLE (配对问题)

任意n个同学交了n本作业，随机每人发回一本作业，试求有$k=0,1,2,...,n$个同学得到自己作业的概率？ 简单情形：$n=2$, $n=3$,

解答.

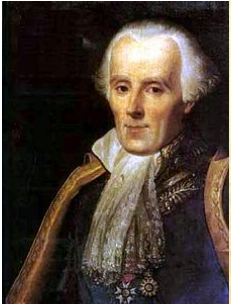
设$n=3$, 记三个人A,B,C,作业a,b,c, Ab表示A拿到b的作业。
样本空间：$\Omega=\{AaBbCc, AaBcCb, AbBaCc, AbBcCa, AcBbCa, AcBaCb\}$
事件：$X=k$个同学配对成功； $k=0$: $X=\{AbBcCa, AcBaCb\}$
$k=1$: $X=\{AaBcCb, AbBaCc, AcBbCa\}$
$k=2$: $X=\emptyset$; $k=3$: $X=\{AaBbCc\}$
古典概率计算有 $P(\{X=0\})=\frac{1}{3}, P(\{X=1\})=\frac{1}{2}$,
$P(\{X=2\})=0, P(\{X=3\})=\frac{1}{6}$,

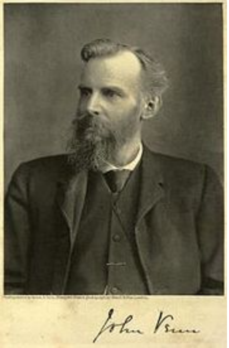
□

n变大时，情况迅速变复杂(样本空间大小$n!$)！见后续讲解。

概率律的指定：古典vs 频率学派

P.Laplace VS J.Venn





概率的真实大小?

- 投硬币：coin flipping 正面的概率?
- 拉普拉斯：1/2
- De Morgan 德摩根：1061/2048
- 蒲丰Buffon： 2048/4040
- John Kerrich: 5067/10000
- K Pearson: 12012/24000, Feller: 4979/10000;

古典模型: 集合中每个元素(基本事件或样本点)的概率一样;
频率学派：基本事件的概率由大量数据的决定。

计数方法 counting

Theorem (乘法原理)

完成一件事要k步，每一步分别有$n_1, n_2, n_3, \ldots, n_k$方法，则完成这件事的方法共有$n = n_1 \cdot n_2 \cdots n_k$.

基本结果:

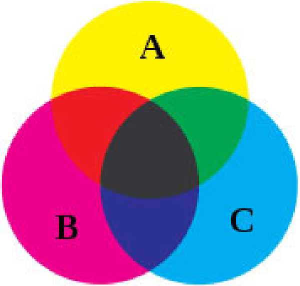
- (m次试验): 从n个元素中有放回的每次取一个；取出m个元素,排成一列；共有 n^m种可能；不同排列是等可能的；
- (m元排列) 从n个元素中无放回的每次取一个；取出m个元素,排成一列；共有 $A_n^m = \frac{n!}{(n-m)!}$种可能；不同排列是等可能的；
- (m元组合) 从n个元素中无放回的每次取一个；取出m个元素,放在一组；共有 $C_n^m = \frac{n!}{m!(n-m)!}$种可能；不同组合是等可能的；

例子：(集合子集的个数) 集合有n个元素，则所有子集的个数为2^n。 利用：组合数即二项式系数, 由二项式展开$(x+y)^n = \sum C_i^n x^i y^{n-i}$可得。

加法公式：从基本事件到复杂事件

Proposition (加法公式)

1. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;
2. $P(A_1 \cup A_2 \cup A_3) = \sum P(A_i) - \sum P(A_i A_j) + P(A_1 A_2 A_3)$;
3. *Jordan* 约旦公式:
   $P(\bigcup_{i=1}^n A_i) = \sum_{k=1}^n (-1)^{k-1} p_k$;
   其中 $p_k = \sum P(A_{i_1} A_{i_2} \ldots A_{i_k})$.



证明(1).

$$P(A \cup B) = P(A + \bar{A} \cap B) = P(A) + P(\bar{A}B)$$
$$= P(A) + P(\bar{A}B) + P(AB) - P(AB) =$$
$$P(A) + P(B) - P(AB) \quad \square$$

再谈：配对问题

EXAMPLE (配对问题)

任意n个同学交了n本作业，随机每人发回一本作业，问没有一个同学得到自己作业的概率？ n无穷大时，概率是多少？

解答.

(对立事件$\bar{A}$)计算至少有一个人得到自己的作业的概率.设E_i是第i个人得到自己作业的事件；则 $\bar{A} = \bigcup_{i=1}^{i=n} E_i$.
Jordan公式： $P(\bar{A}) = P(\bigcup_{i=1}^{i=n} E_i) = \sum_{k=1}^{n}(-1)^{k-1}p_k$,
$p_k = \sum P(E_{i_1}E_{i_2}\dots E_{i_k})$.
记结果为$(1,2,3,\dots,n)$,$(E_{i_1}E_{i_2}\dots E_{i_k})$为其中$k$个人拿到自己作业；剩下$n-k$个任意排列；样本空间是$n$个任意排列，断定:$P(E_{i_1}E_{i_2}\dots E_{i_k}) = \frac{(n-k)!}{n!}$
另外: n中选k个人共有C_n^k可能; Jordan公式中每一项求和有$p_k = \frac{1}{k!}$
$P(\bar{A}) = 1 - \frac{1}{2!} + \frac{1}{3!} - \dots + (-1)^{n-1}\frac{1}{n!}$, $P(A) = 1 - P(\bar{A})$.
特别： n充分大时， $P(A) \approx e^{-1} \approx 0.3678$.

产品检验

EXAMPLE (超几何分布)

假设N个产品中有M个次品,抽取n件产品检验，其中恰有m件次品的概率。

Proof.

A_m是恰有m件产品的事件$(m \leq n)$.

$P(A_m) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}$ □

排列组合续：分组问题

- (m次试验): 从n个元素中有放回的每次取一个；取出m个元素,排成一列；共有 n^m种可能；不同排列是等可能的；
- (m元排列) 从n个元素中无放回的每次取一个；取出m个元素,排成一列；共有 $A_n^m = \frac{n!}{(n-m)!}$种可能；不同排列是等可能的；
- (m元组合) 从n个元素中无放回的每次取一个；取出m个元素,放在一组；共有 $C_n^m = \frac{n!}{m!(n-m)!}$种可能；不同组合是等可能的；
- (k重分组) 将n个元素分成不同的k组，不考虑每组中的元素次序；第i个组恰有n_i个元素的可能分组为 $\frac{n!}{n_1!n_2!\ldots n_k!}$种可能；不同分组是等可能的；

例子：导师有4个研究生12个本科生，分配到4个项目中，随机分配，问每个项目正好有一个研究生的概率?

Proof.

样本空间大小： $\frac{16!}{4!4!4!4!}$.
事件大小: $\frac{12!}{3!3!3!3!} \cdot 4!$. 概率$P(A) = \frac{64}{455}$ □

Review：回顾

- 计算公式：加法公式$P(A \cup B) =$
- 乘法公式 $P(A \cap B) =?$
- 条件概率 $\rightarrow$ 统计推断。不断通过新的信息获得正确的概率推断!

EXAMPLE (赌牌游戏)

有三张牌：一张两面都是红的，一张两面都是黑的，一张两面是一红一黑。 随机取出一张，如果正面是红的，反面是黑还是红？
怎样赢钱？

赢钱策略：永远赌反面和正面一样!

条件概率的定义

EXAMPLE (概率的变化)

掷骰子一次，结果为6的概率为$\frac{1}{6}$. 如果已知结果为偶数，结果为6的概率为$\frac{1}{3}$. 如果已知结果比4小呢?

解释：概率空间发生了变化。

Definition (条件概率)

设A, B是事件，已知A发生，B发生的概率记为$P(B|A)$. 其公式为$P(B|A) = \frac{P(AB)}{P(A)}$, 其中$P(A) > 0$。

- 古典模型解释: $P(A \cap B) = \frac{\sharp(A \cap B)}{\sharp\Omega} = \frac{\sharp(A \cap B)}{\sharp A} \cdot \frac{\sharp(A)}{\sharp\Omega} = P(A)P(B|A)$
- 公理化模型 $\rightarrow$ 定理：记$P_A(B) = P(B|A)$,P_A是一个概率(满足概率公理);

条件概率是一个概率律! ***

Theorem (条件概率)

条件概率满足概率公理化的三个条件。

- 非负： $P_A(B) \geq 0$
- 归一化： $P_A(\Omega) = 1$
- (可列) 可加性 $P_A(B \cup C) = P_A(B) + P_A(C)$.

Corollary (条件概率下的条件概率)

设有事件A, B, C, 已知A发生，有条件概率P_A, 又已知B发生，有条件概率P_{AB}, 则$P_A(C|B) = P_{AB}(C)$.

乘法公式

Proposition (乘法公式)

$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$,可以推广到$n$个事件。$P(A_1A_2\dots A_n) = P(A_1)P(A_2|A_1)\dots P(A_n|A_1A_2\dots A_{n-1})$

EXAMPLE (扑克)

一副牌*(52张)*连续抽三张都不是红桃的概率?

Proof.

设A_1, A_2, A_3表示第i张牌不是红桃。
$P(A_1) = \frac{39}{52}$, $P(A_2) = ?$
$P(A_2|A_1) = \frac{38}{51}$, $P(A_3|A_1A_2) = \frac{37}{50}$
则$P(A_1A_2A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)$. $\square$

第二次作业

北航教材:

习题一. 11,16,18，24,26,28

Monty Hall 三门问题

Monty Hall: 美国电视节目主持人.

问题： 有三扇门，其中一个后面有大奖(汽车). 幸运观众选好一个门后，主持人打开一扇门，显示其中没有奖，问 该观众应该坚持原来的选择还是改变选择?

- 答案：永远应该改变选择!
- 概率分析：坚持(一种可能) $\rightarrow$ 赢奖概率 1/3, 改变(两种可能) $\rightarrow$赢奖概率 2/3
- 理解性解释：假设有一万扇门，选定一个门后，主持人打开所有9998扇门，只剩一扇门，是否改变选择?
- 条件概率分析：假设$C = 1, 2, 3$表示汽车在那个门,$S = 1, 2, 3$表示观众的选择, $H = 1, 2, 3$表示主持人打开的门。设$S = 1, H = 3$, 计算概率$P(C = 2|H = 3, S = 1) = 2/3$。

Review：回顾

Remark (作业事宜)

RECALL:

- 概率模型=样本空间+概率律
- 计算：加法公式，乘法公式，条件概率
- 例子: 配对问题，

TODAY

- 全概率，贝叶斯公式；
- 计算例子：抽样原理，保险评估，假阳性之谜；
- 条件概率与独立性。
- 真实世界的概率模型？
影响因素极多(独立性) → 简化模型
- 例子：系统与重复试验

全概率公式

Proposition (全概率公式)

$P(B) = P(A)P(B|A) + P(\bar{A})P(B|\bar{A})$,可以推广到$n$个互不相容事件(满足$\bigcup A_i = \Omega$)。

证明： $P(B) = P(B \cap \Omega) = P(B \cap (A \cup \bar{A}))$
$P(B) = P(BA) + P(B\bar{A}) = P(A)P(B|A) + P(\bar{A})P(B|\bar{A})$

EXAMPLE (风险评估)

保险公司认为开车者分为两类。一类容易出事故，一类为安全者。统计发现容易出事故在一年内发生事故的概率是0.4；安全者为0.2. 设第一类人口比例为30%. 现有一个新司机买保险，问他在一年内出事故的概率是多少？如果他一年内出了事故，他是容易出事故者的概率是多大？

解答：记$A =$容易出事故者；$B =$一年内发生事故；
(1)：$P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A}) = 0.4 \times 0.3 + 0.2 \times 0.7 = 0.26$
(2)：$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(A)P(B|A)}{P(B)} = 6/13.$

Bayes概率与因果推理 ***

Thomas Bayes(1702-1761)

Theorem (贝叶斯定理)

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} = P(A)\frac{P(B|A)}{P(B)}。$$

Proof: $P(AB) = P(A|B)P(B) = P(B)P(B|A)$

- 设A是原因，B是结果。已知B发生，推断A是否发生?
- 称$P(A)$是先验概率(prior)，$P(A|B)$是后验概率(posterior). $P(B|A)$是似然性(likeliwood), $P(B)$是边缘概率 (marginal).
- 贝叶斯推断：新的事件或信息B改变了A的概率(贝叶斯定理)。→ 通过更多信息得到更确切的原因判断.

贝叶斯公式

Proposition (贝叶斯公式(逆概率公式))

$P(A|B)=\frac{P(A)P(B|A)}{P(A)P(B|A)+P(\bar{A})P(B|\bar{A})}$,可以推广到$n$个互不相容事件。

[证明]: 全概率公式有 $P(B)=P(A)P(B|A)+P(\bar{A})P(B|\bar{A})$，代入贝叶斯定理可得。

EXAMPLE

赌牌游戏： 有三张牌。一张两面都是红的，一张两面都是黑的，一张两面是一红一黑。 随机取出一张，如果是红的，反面是黑的概率多大?

记牌为RR, RB, BB,取出的牌朝上为红事件为A,
$P(RB|A)=\frac{P(RB\cap A)}{P(A|RR)P(RR)+P(A|RB)P(RB)+P(A|BB)P(BB)}=1/3.$

抽签原理

EXAMPLE (抽签原理)

n个签中有m个为中奖。有放回的抽签：任一次中奖的概率为m/n。 无放回的随机抽签，任一次的中奖概率也是m/n. 即抽奖与次序无关!

Proof.

有放回情形： 显然任一次概率m/n.
无放回情形：记A_k表示第k次中奖的事件。则$P(A_1) = \frac{m}{n}$.
数学归纳法:设$P(A_{k-1}) = m/n$, 对所有m, n都成立，则用全概率公式
$P(A_k) = P(A_k|A_1)P(A_1) + P(A_k|\bar{A}_1)P(\bar{A}_1)$
$P(A_k) = \frac{m-1}{n-1}\frac{m}{n} + \frac{m}{n-1}\frac{m}{n-1} = m/n$.
直接证明：设第k次中奖，样本空间为n个取k个的排列共A_n^k. 事件A_k,第k次抽奖中奖的可能性为m种;其余$k-1$个任意排列共A_{n-1}^{k-1}.
$P(A_k) = m * A_{n-1}^{k-1}/A_n^k = m/n$.

□

假阳性之谜

医学里诊断疾病时对某种化验结果或试验结果，也有阴阳性的区分。阳性表示体内有某种病原体存在或者对某种药物有过敏反应.

EXAMPLE (疾病判断)

已知人群中某种疾病的发病率是0.1%. 有个抽血试验可以诊断该疾病，但准确率是90%(有病为阳性或无病为阴性的概率).有个人(甲)抽血试验为阳性， 问医生有多大把握(概率)判断这个人有该疾病?

解答.

记A =甲患病， B =甲的试验结果阳性；

$P(A) = 0.001, P(B|A) = 0.9, P(B|\bar{A}) = 0.1$

$P(A|B) = \frac{0.001\times 0.9}{0.001\times 0.9+0.999\times 0.1} \approx 0.0089$ □

罕见病的检验结果更有可能是假阳性!

独立性的定义

Definition (两个事件独立)

如果A, B满足$P(AB) = P(A)P(B)$, 称事件A, B互相独立。

- 条件概率解释：如果A, B独立$(P(A) > 0)$，则 $P(B|A) = P(B)$。
- 理解：不相容事件$(A \cap B = \emptyset)$ 不相互独立!
- 必然事件Ω和不可能事件$\emptyset$与任何事件相互独立。
- 如果A, B独立,则$\bar{A}, B$相互独立.
- 例子：掷骰子：$A =$结果为偶数；$B =$结果为三的倍数。问A, B是否独立?
  正四面体, 正八面体骰子呢?

独立性的判断

EXAMPLE

掷骰子两次，事件$A=$第一次为1, 事件$B=$第二次为6, A与B是相互独立。记事件$C=$两次的和为7,A与C是否相互独立?

Proof.

$$P(A=1,B=6)=\frac{1}{36}=P(A=1)P(B=6)$$
$$P(A=1,C=7)=\frac{1}{36}=P(A=1)P(C=7)$$
$\square$

问题：已知C发生，A,B独立吗?

多次事件的独立性

EXAMPLE (条件影响独立)

掷硬币两次。事件A =第一次为正面H, 事件B =第二次为正面H, 事件C =两次的结果不同,如果已知C发生，问A, B是否相互独立?

解答 $P(A|C) = P(A) = 1/2, P(B|C) = P(B) = 1/2,$
$P(AB|C) = 0 \neq P(A|C)P(B|C)$.
事实上A, B, C两两独立但三个不独立。

Definition (n个事件独立)

如果$A_1, A_2, \ldots, A_n$满足,任取其中$k(1 \leq k \leq n)$个事件,有
$P(A_{i_1}A_{i_2}\cdots A_{i_k}) = P(A_{i_1})P(A_{i_2})\cdots P(A_{i_k})$. 称$A_1, A_2, \ldots, A_n$相互独立。

- n次独立试验; n次无放回抽样?
- 所有事件相互独立$\rightarrow$ 任意两组事件相互独立。
- 理解: 两两相互独立不一定是所有事件相互独立。

系统的可靠性

EXAMPLE (系统)

系统有若干元件组成，系统的可靠性由元件的可靠性和系统的结构(网络)共同决定。

- 基本假定：所有元件是相互独立的。
- 连接方式：并联或串联。
  设每个元件正常工作的概率为$P(A_i)$.
  概率公式：串联系统：$P(A_1A_2\cdots A_n)=P(A_1)P(A_2)\cdots P(A_n)$.
  并联$P(A_1\cup A_2\cdots\cup A_n)=1-(1-P(A_1))(1-P(A_2)\cdots(1-P(A_n))$.

EXAMPLE (教材P22,例4)

$2n$个元件可以组成两个系统，比较两个系统的可靠性。

独立试验与二项概率模型

EXAMPLE (教务网站堵塞)

设教务网站有c个服务器，每个可以处理100个网页访问请求，用户帐号为n个。如果固定时间段内每个帐号访问的概率是p,问网站堵塞的概率？（即服务器不够用）。

- 基本假定：所有帐号访问是独立的。
- 固定时间段内访问帐号的个数记为X,则$P(X=k)=C_n^k p^k(1-p)^{n-k}$
- 网站堵塞的概率$P=\sum_{i=100c}^{\infty} P(X=i)$
- 若$c=15, n=20000, p=0.1$ 呢？

一般称n次重复试验结果出现k次的概率为二项概率。

作业

第三次作业： 北航教材： P35 习题一. 31， 35,38,39,40