概率论与数理统计
第一章 随机事件与概率

金玲飞

复旦大学计算机学院
Email: lfjin@fudan.edu.cn

2018.9.25

1.4 条件概率

Monty Hall problem

假设一个电视竞猜栏目，你面前有三扇门，其中一扇门后面有一台保时捷，猜到就送你了，剩下的什么也没有。假设你已经选了一扇门（不打开），此时主持人打开剩下的两扇门中什么都没有的那一扇，然后给你一次重选的机会，请问你会改变注意选择另一扇吗？

银行账户管理问题

- 某银行将客户分为信用好和坏两类，并通过分析历史数据得到：在每个月都会有近1%的好客户和10%坏客户透支银行账户。当一新客户来银行开办现金账户时，信用处通过基本检验后认为这位客户大概有70%的概率是好客户。问题是该客户第一个月内透支，问银行对这个客户的信用度有什么改变？如果他第二个月仍透支呢？

问题： 已知事件B发生，怎么求另一事件A发生的概率？

问题： 已知事件B发生，怎么求另一事件A发生的概率？

例子：任意抽取有两个小孩的家庭考察性别。$\Omega = \{bb, gb, bg, gg\}$。假定生男生女的概率是相同的，家中至少有一女孩的概率？已知家中至少有一男孩，有女孩的概率？

问题： 已知事件B发生，怎么求另一事件A发生的概率？

例子：任意抽取有两个小孩的家庭考察性别。$\Omega=\{bb,gb,bg,gg\}$。假定生男生女的概率是相同的，家中至少有一女孩的概率？已知家中至少有一男孩，有女孩的概率？

$P(A|B)$: 已知事件B发生，事件A发生的概率.

条件概率

定义 (条件概率 Conditional Probability)

设$(\Omega, \mathcal{F}, P)$为概率空间，$B \in \mathcal{F}$，$P(B) > 0$，则对任意$A \in \mathcal{F}$，令

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

并称它为在事件B发生的条件下事件A发生的条件概率。

条件概率

定义 (条件概率 Conditional Probability)

设$(\Omega, \mathcal{F}, P)$为概率空间，$B \in \mathcal{F}$，$P(B) > 0$，则对任意$A \in \mathcal{F}$，令

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

并称它为在事件B发生的条件下事件A发生的条件概率。

条件概率是概率，$P(\cdot|B)$具有概率所具有的一切性质。
$(i) P(A|B) \geq 0; (ii) P(\Omega|B) = 1;$
$(iii) A_i \in \mathcal{F}, A_iA_j = \emptyset, P(\cup_{i=1}^{\infty} A_i|B) = \sum_{i=1}^{\infty} P(A_i|B)$

条件概率的性质

- $P(\varnothing|B) = 0$.
- $P(A|B) = 1 - P(\overline{A}|B)$.
- $P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B) - P(A_1A_2|B)$.

条件概率的性质

- $P(\varnothing|B) = 0.$
- $P(A|B) = 1 - P(\overline{A}|B).$
- $P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B) - P(A_1 A_2|B).$

概率是一种特殊的条件概率。

乘法公式 Multiplication formula

乘法公式

- $P(AB) = P(A|B)P(B)$.
- $P(A_1A_2\cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1\cdots A_{n-1})$.

注：用于$P(A|B)$容易求得，$P(AB)$较难。

例子 (1.4.2)

袋中有a只白球和b只红球。现依次不放回地从袋中任取两球，试求两次均取到白球的概率。

例子 (1.4.2)

袋中有a只白球和b只红球。现依次不放回地从袋中任取两球，试求两次均取到白球的概率。

例子 (1.4.3)

袋中有一个白球和一个黑球，一次次地从袋中取球，如果取出白球，则除把白球放回外再加进一个白球，直至取出黑球为止。求取了n次都没有取到黑球的概率。

全概率公式

求事件A的概率难求怎么办？

定义 (1.4.2 完备事件组)

设$(\Omega, \mathcal{F}, P)$为概率空间，$B_i \in \mathscr{F}, i = 1, \ldots, n$，如果

$$\cup_{i=1}^{n} B_i = \Omega, B_iB_j = \phi(i \neq j), P(B_i) > 0, i = 1, \ldots, n,$$

则称$B_1, \ldots, B_n$为完备事件组。
类似地，称$\mathscr{F}$中可列无穷个事件$B_1, \ldots, B_n, \ldots$为完备事件组，如果

$$\cup_{i=1}^{\infty} B_i = \Omega, B_iB_j = \phi(i \neq j), P(B_i) > 0, i = 1, \ldots, n, \ldots$$

全概率公式

定理 (1.4.2 Total Probability Formula)

设$(\Omega, \mathcal{F}, P)$为概率空间，$B_1, \ldots, B_n$为完备事件组，则对任何事件$A \in \mathcal{F}$，

$$P(A) = \sum_{i=1}^{n} P(B_i)P(A|B_i)$$

类似地，若$B_1, \ldots, B_n, \ldots$为完备事件组，则对任何事件$A \in \mathscr{F}$，

$$P(A) = \sum_{i=1}^{\infty} P(B_i)P(A|B_i)$$

- 全概率公式用于求复杂事件的概率。
- 使用关键：找出完备事件组$B_i(i \geq 1)$ 使得$P(B_i)$和$P(A|B_i)$比较好计算。
- 全概率公式最简单的形式：

$$P(A) = P(B)P(A|B) + P(\overline{B})P(A|\overline{B})$$

- 全概率公式用于求复杂事件的概率。
- 使用关键：找出完备事件组$B_i (i \geq 1)$ 使得$P(B_i)$和$P(A|B_i)$比较好计算。
- 全概率公式最简单的形式：

$$P(A) = P(B)P(A|B) + P(\overline{B})P(A|\overline{B})$$

- 条件$B_1, \ldots, B_n$为样本空间的一个分割，可以改成$B_1, \ldots, B_n$互不相容且$A \subset \cup_{i=1}^n B_i$，定理依然成立。

例子

摸彩模型 设n张彩票中有一张奖券，求第二个人摸到奖券的概率？

例子

摸彩模型 设n张彩票中有一张奖券，求第二个人摸到奖券的概率？

解：设A_i =第i人摸到奖券

$$P(A_1) = 1/n, P(\overline{A_1}) = \frac{n-1}{n}$$

$$P(A_2|A_1) = 0, P(A_2|\overline{A_1}) = \frac{1}{n-1}$$

$$P(A_2) = P(A_1)P(A_2|A_1) + P(\overline{A_1})P(A_2|\overline{A_1}) = 1/n$$

注： 与先后顺序无关。

例子 (1.4.4)

某工厂有四条流水线生产同一种产品，该四条流水线的产量分别占到总产量的*15%,20%,30%,35%*。又这四条流水线的不合格品率依次为*0.05,0.04,0.03,0.02*。现在从出厂产品中任取一件，问恰好抽到不合格品的概率是多少？

例子 (1.4.4)

某工厂有四条流水线生产同一种产品，该四条流水线的产量分别占到总产量的*15%,20%,30%,35%*。又这四条流水线的不合格品率依次为*0.05,0.04,0.03,0.02*。现在从出厂产品中任取一件，问恰好抽到不合格品的概率是多少？

解：A：抽到不合格品， B：抽到第i条流水线的产品

$$P(A) = \sum_{i=1}^{4} P(B_i)P(A|B_i) = 0.0315$$

问： 已知抽到不合格品，问来自第一条流水线的概率？

Bayes formula

已知“结果”求“原因”

定理 (1.4.3（贝叶斯公式）)

设$B_1, \ldots, B_n$是概率空间$(\Omega, \mathscr{F}, P)$上的完备事件组，$P(A) > 0$，则

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{n} P(B_j)P(A|B_j)}, i = 1, \ldots, n$$

类似地，若$B_1, \ldots, B_n, \ldots$是完备事件组，$P(A) > 0$，则

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{\infty} P(B_j)P(A|B_j)}, i = 1, \ldots, n, \ldots$$

两种概率

- $B_1, B_2, \ldots, B_n$可以看成是导致A发生的原因

两种概率

- $B_1, B_2, \ldots, B_n$可以看成是导致A发生的原因
- 先验概率 **Prior probability**：$P(B_i)$是试验前就知道的
- 后验概率 **Posterior probability**：$P(B_i|A)$是观察到A后的B_i的概率
- 贝叶斯公式又称“逆概率公式”或“后验概率公式”

两种概率

- $B_1, B_2, \ldots, B_n$可以看成是导致A发生的原因
- 先验概率 **Prior probability**：$P(B_i)$是试验前就知道的
- 后验概率 **Posterior probability**：$P(B_i|A)$是观察到A后的B_i的概率
- 贝叶斯公式又称“逆概率公式” 或“后验概率公式”

最常用的贝叶斯公式

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

例子 (1.4.5)

用甲胎蛋白法普查肝癌。令

$$A = \text{甲胎蛋白检验结果为阳性}, B = \text{被检验者患肝癌}$$

$$\overline{A} = \text{甲胎蛋白检验结果为阴性}, \overline{B} = \text{被检验者未患肝癌}$$

由过去的资料知，

$$P(A|B) = 0.95, P(\overline{A}|\overline{B}) = 0.90$$

又知某地居民的肝癌发病率为$P(B) = 0.0004$。在一次普查中查出一批甲胎蛋白检验结果为阳性的人，求这批人中真的患有肝癌的概率$P(B|A)$。

例子 (1.4.5)

用甲胎蛋白法普查肝癌。令

$$A = \text{甲胎蛋白检验结果为阳性}, B = \text{被检验者患肝癌}$$

$$\overline{A} = \text{甲胎蛋白检验结果为阴性}, \overline{B} = \text{被检验者未患肝癌}$$

由过去的资料知，

$$P(A|B) = 0.95, P(\overline{A}|\overline{B}) = 0.90$$

又知某地居民的肝癌发病率为$P(B) = 0.0004$。在一次普查中查出一批甲胎蛋白检验结果为阳性的人，求这批人中真的患有肝癌的概率$P(B|A)$。

后验概率的大小受先验概率的影响。

解. 由贝叶斯公式 (1.4.9), 可得

$$P(B|A) = \frac{P(B)\,P(A|B)}{P(B)\,P(A|B) + P(\overline{B})\,P(A|\overline{B})} \tag{1.4.10}$$

$$= \frac{0.0004 \times 0.95}{0.0004 \times 0.95 + (1 - 0.0004) \times (1 - 0.90)}$$

$$= 0.0038$$

例子 (1.4.6 “狼来了”的贝叶斯分析)

设$A=$ “孩子说谎”，$B=$ “孩子可信”。设村民过去对这个孩子的印象（先验概率）为

$$P(B)=0.8, P(\overline{B})=0.2$$

假设可信的孩子说谎的概率为*10%*，不可信的孩子说谎的概率为*50%*，即

$$P(A|B)=0.1, P(A|\overline{B})=0.5$$

计算第一次和第二次“狼来了”说谎发生后，村民对孩子印象（可信度）。

例子：罐子模型

设罐中有b个黑球，r个红球，每次随机取出一球，取出后将原球放回，再加入c个同色球和d个异色球。记B_i为第i次取出的是黑球，R_j为第j次取出的是红球。若连续从罐中取出三个球，其中两个红球，一个黑球的概率是

例子：罐子模型

设罐中有b个黑球，r个红球，每次随机取出一球，取出后将原球放回，再加入c个同色球和d个异色球。记B_i为第i次取出的是黑球，R_j为第j次取出的是红球。若连续从罐中取出三个球，其中两个红球，一个黑球的概率是

$$\begin{aligned} P(B_1R_2R_3) &= P(B_1)P(R_2|B_1)P(R_3|B_1R_2) \\ &= \frac{b}{b+r}\frac{r+d}{b+r+c+d}\frac{r+d+c}{b+r+2c+2d} \end{aligned}$$

$$
\begin{aligned}
P(R_1B_2R_3) &= P(R_1)P(B_2|R_1)P(R_3|R_1B_2) \\
&= \frac{r}{b+r}\frac{b+d}{b+r+c+d}\frac{r+d+c}{b+r+2c+2d}
\end{aligned}
$$

$$
\begin{aligned}
P(R_1R_2B_3) &= P(R_1)P(R_2|R_1)P(B_3|R_1R_2) \\
&= \frac{r}{b+r}\frac{r+c}{b+r+c+d}\frac{b+2d}{b+r+2c+2d}
\end{aligned}
$$

以上概率与黑球是第几次抽到有关。

(1) 不返回抽样：$c=-1, d=0$。只要抽取的黑球红球个数确定，概率不依赖与先后次序。

$$P(B_1R_2R_3)=P(R_1B_2R_3)=P(R_1R_2B_3)$$

(1) 不返回抽样：$c=-1, d=0$。只要抽取的黑球红球个数确定，概率不依赖与先后次序。

$$P(B_1R_2R_3)=P(R_1B_2R_3)=P(R_1R_2B_3)$$

(2) 返回抽样：$c=0, d=0$。此时抽取结果互不影响。

(1) 不返回抽样：$c = -1, d = 0$。只要抽取的黑球红球个数确定，概率不依赖与先后次序。

$$P(B_1R_2R_3) = P(R_1B_2R_3) = P(R_1R_2B_3)$$

(2) 返回抽样：$c = 0, d = 0$。此时抽取结果互不影响。

(3) 传染病模型：$c > 0, d = 0$。每发现一个传染病患者，都会增加再传染概率。

$$P(B_1R_2R_3) = P(R_1B_2R_3) = P(R_1R_2B_3)$$

(1) 不返回抽样：$c = -1, d = 0$。只要抽取的黑球红球个数确定，概率不依赖与先后次序。

$$P(B_1R_2R_3) = P(R_1B_2R_3) = P(R_1R_2B_3)$$

(2) 返回抽样：$c = 0, d = 0$。此时抽取结果互不影响。

(3) 传染病模型：$c > 0, d = 0$。每发现一个传染病患者，都会增加再传染概率。

$$P(B_1R_2R_3) = P(R_1B_2R_3) = P(R_1R_2B_3)$$

(4) 安全模型：$c = 0, d > 0$。

1.5 事件的独立性

事件的独立性

直观的说：若A的发生与否与B的发生与否互不影响，则这两个事件是独立的。即

$$P(A|B) = P(A), P(B|A) = P(B)$$

事件的独立性

直观的说：若A的发生与否与B的发生与否互不影响，则这两个事件是独立的。即

$$P(A|B) = P(A), P(B|A) = P(B)$$

定义 (1.5.1（两个事件的独立性）)

设$(\Omega, \mathcal{F}, P)$为概率空间，$A, B \in \mathcal{F}$。如果

$$P(AB) = P(A)P(B)$$

则称事件A与B相互独立(独立)。否则称A与B不独立或相依。

- 概率为0的事件与概率为1的事件与任何事件独立。

- 概率为0的事件与概率为1的事件与任何事件独立。

例子 (1.5.1)

袋中有a只黑球和b只白球，采取有放回方式摸球两次，记

$$A = \text{第一次摸得黑球}, B = \text{第二次摸得黑球}$$

计算$P(B|A)$，判断A, B是否独立。

- 概率为0的事件与概率为1的事件与任何事件独立。

例子 (1.5.1)

袋中有a只黑球和b只白球，采取有放回方式摸球两次，记

$$A = \text{第一次摸得黑球}, B = \text{第二次摸得黑球}$$

计算$P(B|A)$，判断A, B是否独立。

注: 若摸球改为不放回方式，判断此时A, B是否独立。

例子 (1.5.3)

若事件A, B独立，证明下列各对事件也相互独立：

$$\{\overline{A}, B\}, \{A, \overline{B}\}, \{\overline{A}, \overline{B}\}$$

例子 (1.5.3)

若事件A, B独立，证明下列各对事件也相互独立：

$$\{\overline{A}, B\}, \{A, \overline{B}\}, \{\overline{A}, \overline{B}\}$$

$$P(\overline{A}B) = P(B - AB) = P(B) - P(AB) = P(\overline{A})P(B)$$

例子 (1.5.3)

若事件A, B独立，证明下列各对事件也相互独立：

$$\{\overline{A}, B\}, \{A, \overline{B}\}, \{\overline{A}, \overline{B}\}$$

$$P(\overline{A}B) = P(B - AB) = P(B) - P(AB) = P(\overline{A})P(B)$$

- 事件A, B独立 VS 事件A, B互不相容

例子 (1.5.3)

若事件A, B独立，证明下列各对事件也相互独立：

$$\{\overline{A}, B\}, \{A, \overline{B}\}, \{\overline{A}, \overline{B}\}$$

$$P(\overline{A}B) = P(B - AB) = P(B) - P(AB) = P(\overline{A})P(B)$$

- 事件A, B独立 VS 事件A, B互不相容
- $P(A) > 0, P(B) > 0$，独立与互不相容不能同时成立。