概率统计第二讲：条件概率

史灵生　清华数学系

1、本讲提要

2、条件概率定义

- 在一随机试验中，我们用概率来评价某种不确定现象（事件A）发生的可能性。如果我们得到关于这个随机试验的进一步信息（比如，被告知事件B发生了），那么我们有必要对这种不确定现象（事件A）发生的可能性作出新的判断，条件概率（$P(A|B)$）就是我们经重新判断得到的概率值。我们称$P(A|B)$为在已知B发生的条件下，A的条件概率。
- 我们从频率的观点看看应该如何定义条件概率$P(A|B)$。我们将一个随机试验重复n次，其中事件B发生了$N_n(B)$次，在B发生的这$N_n(B)$次试验中，A发生的相对频率为

$$\frac{N_n(AB)}{N_n(B)} = \frac{N_n(AB)/n}{N_n(B)/n} \to \frac{P(AB)}{P(B)}, \quad 当n \to \infty.$$

3、条件概率定义

定义

设$(\Omega, \mathcal{F}, P)$是概率空间，$B \in \mathcal{F}$满足$P(B) > 0$。则对任何$A \in \mathcal{F}$，定义

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

为在B发生下A的条件概率，简称条件概率。

性质1

设$(\Omega, \mathcal{F}, P)$是概率空间，$B \in \mathcal{F}$满足$P(B) > 0$。令

$$P_B : \mathcal{F} \to \mathbb{R}, \quad P_B(A) = P(A|B) = P(AB)/P(B)。$$

则(a) $(\Omega, \mathcal{F}, P_B)$也是概率空间，即$P_B$满足非负性、正则性和可列可加性；(b) 如果$P(BC) > 0$，则$P_B(A|C) = P(A|BC)$。

4、条件概率定义

注：

- 定理表明，新的信息的获得使得我们对随机试验的结果必须重新进行评价。
- 在概念上没有必要定义“多重条件概率”。
- 当$P(B)=0$时，我们暂未给出条件概率的定义；这种情况下的条件概率我们将在连续型随机变量中遇到。

性质1

设$(\Omega,\mathcal{F},P)$是概率空间，$B\in\mathcal{F}$满足$P(B)>0$。令

$$P_B:\mathcal{F}\to\mathbb{R},\quad P_B(A)=P(A|B)=P(AB)/P(B)。$$

则(a) $(\Omega,\mathcal{F},P_B)$也是概率空间。即P_B满足非负性、正则性和可列可加性；(b) 如果$P(BC)>0$，则$P_B(A|C)=P(A|BC)$。

5、例

例

新搬来的邻居家有一对双胞胎小孩，你听说其中有男孩。问这对双胞胎是龙凤胎的可能性有多大？

解

- 样本空间：{(男,男),(男,女),(女,男),(女,女)}，古典概型。
- 事件A表示这是一对龙凤胎，事件B表示至少有一个男孩，$A=\{(男,女),(女,男)\}$，$B=\{(男,男),(男,女),(女,男)\}$。
- 于是，$P(A|B)=P(AB)/P(B)=2/3$。

- 为何样本空间不是{两个男孩,两个女孩,一男孩和一女孩}？
- 上述概率空间为什么是古典概型？
- 如你亲眼见到了双胞胎中的一个，而且是男孩，那么另一个是女孩的概率是几？参见[1]。

6、乘法公式

我们在处理复杂事件往往将它们表示成其他一些易于分析的事件的运算结果。

- 对求余运算，我们可以利用概率的简单性质$P(\bar{A})=1-P(A)$。
- 对极限运算，我们用概率的连续性。
- 而对交运算我们经常利用下述结论：

性质2（乘法公式）

1. 设A,B是事件，$P(B)>0$。则
$$P(AB)=P(B)P(A|B)。$$
2. 设$A_1,\ldots,A_n$是事件，$P(A_1A_2\cdots A_{n-1})>0$。则
$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1}).$$

7、可靠性

例

一设备的失效率函数为$\lambda(t)$。即
$P(设备在(t,t+\Delta t]时段中失效|设备在t时刻正常)=\lambda(t)\Delta t+o(\Delta t),\quad \Delta t\to 0$. 求设备在$(0,t)$时段正常的概率。

- 记A_t为设备在$(0,t)$时段正常，$p(t)=P(A_t)$。
- 则$P(\overline{A_{t+\Delta t}}|A_t)=\lambda(t)\Delta t+o(\Delta t)$。
- $$\begin{aligned}p(t+\Delta t)&=P(A_{t+\Delta t})=P(A_tA_{t+\Delta t})=P(A_t)P(A_{t+\Delta t}|A_t)\\&=p(t)\big[1-\lambda(t)\Delta t+o(\Delta t)\big],\quad \Delta t\to 0。\end{aligned}$$
- $\frac{dp(t)}{dt}=-\lambda(t)p(t)$，
- 分离变量后得，$d\ln p(t)=\frac{dp(t)}{p(t)}=-\lambda(t)dt$，
- 在区间$[0,t]$上积分上式，并注意到$p(0)=1$，得到$p(t)=\exp\left(-\int_0^t\lambda(s)ds\right)$。

8、全概率公式

性质3

设$(\Omega,\mathcal{F},P)$是概率空间，事件列$B_1,\ldots,B_n,\ldots$是样本空间Ω的一个分割，即$B_iB_j=\emptyset$（$\forall i\neq j$），且$\bigcup_{n=1}^{\infty}B_n=\Omega$。
如果$P(B_n)>0$（$\forall n$），则对任一事件$A\in\mathcal{F}$有，
$P(A)=\sum_{n\geq 1}P(B_n)P(A|B_n)$。
这里事件列$\{B_n\}_n$由有限多个或可列无穷多个事件组成。

注：

- 全概率公式使得我们研究复杂问题时，可以将所有情况按一定原则分成若干情形；
- 在每个情形下，原问题的解决变得相对简单了。
- 而综合考虑各种情形下的结论得到整体的答案。
- 这朴素而基本的思想在概率论中起着极其重要的作用。

9、全概率公式

性质3

设$(\Omega,\mathcal{F},P)$是概率空间，事件列$B_1,\ldots,B_n,\ldots$是样本空间Ω的一个分割，即$B_iB_j=\emptyset$（$\forall i\neq j$），且$\bigcup_{n=1}^{\infty}B_n=\Omega$。
如$P(B_n)>0$（$\forall n$），则对任一事件$A\in\mathcal{F}$有，
$P(A)=\sum_{n\geq 1}P(B_n)P(A|B_n)$。
这里事件列$\{B_n\}_n$由有限多个或可列无穷多个事件组成。

证明：

- 因为$A=A\Omega=\cup_{n\geq 1}AB_n$且$AB_n$彼此互不相容。
- 由可加性得：$P(A)=P\left(\bigcup_{n\geq 1}AB_n\right)=\sum_{n\geq 1}P(AB_n)$。
- 再由乘法公式得：$P(A)=\sum_{n\geq 1}P(B_n)P(A|B_n)$。

10、赌徒输光问题（首步分析法）

例

甲乙二人赌钱，各有赌资i元、$a-i$元。每赌一次必须分出胜负，胜方从负方的赌资中赢得1元，到有人输光时结束。已知每次赌，甲胜的概率为$0<p<1$，乙胜的概率为$q=1-p$。求甲最终破产的概率。

- 记A_i表示“甲有赌资i元但最终破产”，$p_i=P(A_i)$。
- 首步分析法就是用第一步所有可能结果对样本空间作分割。
- 记B表示“第一次甲胜”。则$B,\bar{B}$是Ω的一个正分割，$P(B)=p>0$，$P(\bar{B})=q>0$。
- 而当B发生(即第一次甲胜)时,以后的游戏相当于从甲、乙的初始赌资分别为$i+1$元和$a-i-1$元的情况下重新开始，
- 因此$P(A_i|B)=P(A_{i+1})$，同理$P(A_i|\bar{B})=P(A_{i-1})$。
- 因此当$1\leq i<a$时，由全概率公式得：

$$P(A_i)=P(B)P(A_i|B)+P(\bar{B})P(A_i|\bar{B})=pP(A_{i+1})+qP(A_{i-1}).$$

11、赌徒输光问题

- 这样我们建立了一个二阶常系数线性差分方程

$$p_i=pp_{i+1}+qp_{i-1}, \tag{1}$$

- 将$p_i=z^i$代入(1)，得到特征方程

$$pz^2-z+q=0, \tag{2}$$

- 解得$z=1$或$z=q/p$.
- 当$p\neq q$时，上述差分方程(1)的通解为

$$p_i=A\cdot 1^i+B\left(\frac{q}{p}\right)^i,$$

- 对边值条件$p_0=1,p_a=0$可解得

$$A=1-B,\qquad B=1/[1-(q/p)^a].$$

12、赌徒输光问题

- 当$p = q = 1/2$时，$z = 1$是特征方程(2)的二重根，由它得到差分方程(1)的常数解，而$p_i = i$也是差分方程(1)的解，且与常数解线性无关，这时差分方程的通解是

$$p_i = A + Bi,$$

- 对边值条件$p_0 = 1, p_a = 0$可解得

$$A = 1, \qquad B = -\frac{1}{a}.$$

- 从而

$$p_i = \begin{cases} 1 - \dfrac{1-r^i}{1-r^a}, & \text{若} p \neq 1/2; \\ 1 - \dfrac{i}{a}, & \text{若} p = 1/2. \end{cases}, \quad r = \frac{q}{p}, \quad \forall 1 \leq i < a.$$

13、赌徒输光问题

不难发现：

1. 当a有限时，最终甲输光或乙输光的概率为1，也就是赌博几乎注定（以概率1）要在有限时间内结束。
2. 当$a \to \infty$时（乙具有无限财富，比如赌场），甲最终输光的概率为

$$P(A_i) = \begin{cases} \left(\dfrac{q}{p}\right)^i, & \text{若} p > q; \\ 1, & \text{若} p \leq q. \end{cases} = \min\left\{1, \left(\frac{q}{p}\right)^i\right\}, \quad \forall i \geq 1.$$

只要$p > q$，即使甲只有1元，也有$1 - \frac{q}{p}$的机会永不破产。

3. 这个例子也可被描述为一醉汉在一条两端有陷阱的街道上进行的随机徘徊（*random walk*），每次以一定的概率前进或后退一步，且每次选择前进还是后退与其他次的行为无关。
4. 另外可以将这里求解常系数线性差分方程的办法与常系数线性常微分方程的解法做个对比。

14、末步分析法

例

连续地抛掷一个不均匀的硬币n次。第一次出现正面的概率为a，第二次后每次出现与前一次相同的面的概率为b。求第n次时出现正面的概率。

解：

- 令$A_n =$“第n次出现正面”；$p_n = P(A_n)$。
- A_{n+1}发生与A_n发生与否是相关的：$P(A_{n+1}|A_n) = b$, $P(A_{n+1}|\overline{A_n}) = 1 - b$。
- 显然，$A_n, \overline{A_n}$是样本空间的一个正分割，
- 由全概率公式得：

$$\begin{aligned} p_{n+1} &= P(A_{n+1}) = P(A_{n+1}|A_n)P(A_n) + P(A_{n+1}|\overline{A_n})P(\overline{A_n}) \\ &= bp_n + (1-b)(1-p_n) = (2b-1)p_n + 1 - b. \end{aligned}$$

15、末步分析法

由$p_1 = a$，递推计算得：

$$\begin{aligned} p_n &= (2b-1)p_{n-1} + 1 - b \\ &= (2b-1)[(2b-1)p_{n-2} + 1 - b] + 1 - b \\ &= \cdots \\ &= (2b-1)^{n-1}p_1 + (1-b)[1 + (2b-1) + \cdots + (2b-1)^{n-2}] \\ &= \begin{cases} a(2b-1)^{n-1} + \frac{1}{2} - \frac{1}{2}(2b-1)^{n-1}, & \text{若} b \neq 1, \\ a, & \text{若} b = 1. \end{cases} \end{aligned}$$

16、全概率公式注记

注：

全概率公式：$P(A)=\sum_n P(B_n)P(A|B_n)$。

通常在应用全概率公式时，样本空间的正分割是需要我们特意构造的，其原则是：

1. 每个B_n的情况清楚（$P(B_n)$容易被确定）；
2. 每个B_n对所研究的对象A的影响清楚（$P(A|B_n)$容易被确定，这往往是要在B_n发生的情况下选取经适当改造的模型）。这正是我们在例子中选择首或末步的不同情况进行分类的原因。

17、Bayes公式

性质4（Bayes公式）

1. 如果事件A,B满足$P(A)>0,P(B)>0$，则
$$P(B|A)=\frac{P(A|B)}{P(A)}P(B)。$$
2. 设事件列$B_1,\ldots,B_n,\ldots$是样本空间Ω的一个正分割，事件A满足$P(A)>0$。则
$$P(B_i|A)=\frac{P(B_i)P(A|B_i)}{\sum_n P(B_n)P(A|B_n)}.$$

证明：

利用条件概率定义、乘法公式和全概率公式。

18、Bayes公式

注： $P(B_i|A)=\frac{P(B_i)P(A|B_i)}{\sum_n P(B_n)P(A|B_n)}$

- Bayes公式告诉我们应该如何根据新的观察结果A修正我们对事物B的认识，因此我们分别称$P(B)$和$P(B|A)$为B的先验概率和后验概率。Bayes学派认为概率反映人对事物的信任程度，人们根据经验或直观判断确定先验概率，然后再通过观测结果将先验概率修正成后验概率。
- Bayes公式常被用来进行概率意义上的因果分析。比如，B_i是某种疾病，$P(B_i)$是这种疾病的发病率，它的数值来自卫生部门的统计结果。而A是一个病人的症状，$P(A|B_i)$是不同疾病导致这种症状的可能性，医生的任务就是根据症状来确诊患者的病因（比如使$P(B_i|A)$明显大于其他条件概率值的B_i）。
- 又如B_k是关于某个历史事件的各种假说，根据史书的记载或已有的考古发现，我们确定各种假说的可信程度$P(B_k)$。而A是一个新的考古发现，我们需要更新我们对历史的认识，决定那种假说更接近真实情况。

19、Monty Hall问题

Monty Hall问题

[a] 在一个猜奖游戏中，主持人事先在A、B、C三个箱子中的一个箱子里藏了奖品。当观众猜奖品在A时，主持人在B、C两个箱子中选一个（比如B）打开，示意其中无奖品。请问这时最有可能在哪一个箱子中藏有奖品？

[a] Monty Hall是美国电视游戏节目“让我们做个交易”的主持人，关于Monty Hall问题可参见http://en.wikipedia.org/wiki/Monty_Hall_problem。

解：

- 记A表示事件“奖品在A箱中”，类似定义事件B,C。
- 记B^*表示事件“主持人打开B箱示意其中无奖”。
- 则A,B,C是样本空间的一个分割，并且
$$P(A)=P(B)=P(C)=1/3。$$

20、Monty Hall问题

- $P(A|B^*) = \dfrac{P(A)P(B^*|A)}{P(A)P(B^*|A) + P(B)P(B^*|B) + P(C)P(B^*|C)}$.
- 因为主持人事先知道奖品藏在哪个箱子里，因此如果观众第一次猜的箱子里确有奖品，那么他可以打开剩下两个箱子中的任何一个示意其中无奖，即$P(B^*|A) = 1/2$；
- 如果观众第一次猜的箱子里没有奖品，那么他**必然**打开剩下两个箱子中无奖的一个示意其中无奖。
- 所以$P(B^*|B) = 0, \quad P(B^*|C) = 1$。
- 因此$P(A|B^*) = 1/3$。

问题：

若主持人事先也**不知**奖品藏在哪个箱子里，情况又如何呢？

本讲提要 条件概率 公式 | 乘法公式 全概率公式 Bayes公式

21、法庭审讯

- 在法庭审讯时，陪审员也在（可能是无意识地）使用Bayes公式。他根据一些事实判断被告有罪A和无罪$\bar{A}$的概率，并根据控辩双方提供的进一步证据B修正这些概率值：

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

- 我们知道后验概率$P(A|B)$明显依赖先验概率$P(A)$和$P(\bar{A})$，先验概率的不同选择往往导致后验概率值的天壤之别。
- 陪审员对案件往往做先验假定$P(A) = P(\bar{A}) = 1/2$（就像我们上面假设哪个箱内有奖那样）。这种基于无知的先验假定很可能造成无辜者被确定有罪。
- 为保证人的权利不受侵害，法庭应该首先对被告人作无罪假定，然后等待控方提供强有力的有罪证明。

史灵生 清华数学系 | 概率统计第二讲：条件概率

本讲提要 条件概率 公式 | 乘法公式 全概率公式 Bayes公式

22、参考文献

1、Kai Lai Chung and Farid AitSahlia，Elementary probability theory：with stochastic processes and an introduction to mathematical finance，4th ed. New York：Springer，2003.

史灵生 清华数学系 | 概率统计第二讲：条件概率