全概率公式的两点注记

李 佳[a],邓有莲[b]

(江西师范大学 a.计算机信息工程学院;b.数学与信息科学学院,南昌 330022)*

摘 要:全概率公式和贝叶斯公式是《概率统计》中两个重要的概率计算公式,它们的使用条件不一定需要在样本空间划分的前提条件下,可以放宽到两两互不相容正概率事件的前提下;全概率公式的解题思路——分解法在《概率统计》离散型随机变量期望和方差的求解过程中也有着广泛的应用.文章介绍的这两点注记不仅是对现有《概率统计》教材内容的有益补充,也有利于激发学生的学习兴趣、加深学生对全概率公式和贝叶斯公式的全面理解和掌握。

关键词:全概率公式;分解法;期望;方差

doi:10.3969/**j.issn.**2095-5642.2018.**11**.097

中图分类号:G642;O211.9 **文献标志码:**A **文章编号:**2095-5642(2018)11-0097-05

1 全概率公式和贝叶斯公式的新定义

在文献[1]中,样本空间划分的定义是:设 S 为试验 E 的样本空间,B_1,B_2,L,B_n 为 E 的一组事件,若(1)$B_iB_j=\varnothing,i\neq j,i,j=1,2,L,n$;(2)$B_1\cup B_2\cup L\cup B_n=S$,则称 B_1,B_2L,B_n 为样本空间 S 的一个划分,B_i 为分块;并且文献[1]以定理的形式给出了全概率公式和贝叶斯公式的定义:设试验 E 的样本空间为 S,A 为 E 的事件,B_1,B_2,L,B_n 为 S 的一个划分,且 $P(B_i)>0(i=1,2,L,n)$,则称 $P(A)=P(A|B_1)P(B_1)+P(A|B_2)P(B_2)+L+P(A|B_n)P(B_n)$ 为全概率公式。当 $P(A)>0$ 时,称 $P(B_i|A)=\dfrac{P(A|B_i)P(B_i)}{\sum\limits_{j=1}^{n}P(A|B_j)P(B_j)},i=1,2,L,n$ 为贝叶斯公式。我们从下面这个例子可以看出,全概率公式和贝叶斯公式在使用时,并不需要在样本空间划分的这个前提条件下,只要正确写出事件等式就可以正确解题。

例 有朋自远方来,他乘坐火车、轮船、汽车和飞机的概率分别为0.3、0.2、0.1和0.4,并且迟到的概率分别为0.25、0.3、0.1和0,(1)求朋友迟到的概率;(2)如果这位朋友聚会迟到了,请问他乘坐哪种交通工具的可能性最大?

解:设 B_1:“朋友选的交通工具是火车”, B_2:“朋友选的交通工具是轮船”,B_3:“朋友选的交通工具是汽车”,B_4:“朋友选的交通工具是飞机”,

由已知 $P(B_1)=0.3,P(B_2)=0.2,P(B_3)=0.1,P(B_4)=0.4$,易知 B_1,B_2,B_3,B_4 构成样本空间的一个划分。

B_2 又设事件 A:“朋友聚会迟到了”,由于朋友迟到的原因只可能是乘坐火车,轮船和汽车,但朋友乘坐

* **收稿日期:**2018-05-11
基金项目:江西省教育厅科学技术研究项目“多维计算机化自适应测验新选题算法研究”(GJJ170212);江西师范大学教改课题资助项目“探究式教学方法在《离散数学》中的应用研究”(JXSDJG1767)
作者简介:李 佳(1979—),女,江西南昌人,讲师,硕士,研究方向:计算机辅助教学、心理测量;
邓有莲(1979—),女,江西泰和人,讲师,硕士,研究方向:计算机辅助教学。

飞机不会迟到,所以,事件 A 的发生与分块 B_4 无关。其示意图如图下所示：

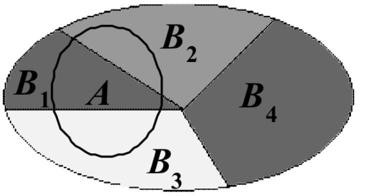

图1　全概率公式分解图

由图1,可以得到事件等式 $A=AB_1\cup AB_2\cup AB_3$,且事件 AB_1,AB_2,AB_3 两两互斥,由概率的可加性得 $P(A)=P(AB_1)+P(AB_2)+P(AB_3)$

再利用全概率公式得

$$P(A)=P(A|B_1)P(B_1)+P(A|B_2)P(B_2)+P(A|B_3)P(B_3)=0.3\times0.25+0.2\times0.3+0.1\times0.1=0.145$$

由朋友迟到的结果来分析他迟到的原因,我们可以使用贝叶斯公式进行计算：

$P(B_1|A)=\dfrac{P(A|B_1)P(B_1)}{P(A)}=\dfrac{0.3\times0.25}{0.145}=\dfrac{15}{29}$,同理 $P(B_2|A)=\dfrac{12}{29}$,$P(B_3|A)=\dfrac{2}{29}$,

因为从题目的已知就知道朋友乘坐飞机聚会是不会迟到的,所以在这里根本不需要计算 $P(B_4|A)$。

最后,答:朋友迟到的概率为0.145,并且如果朋友聚会迟到,则他乘坐火车的可能性最大。

由本例,笔者认为全概率公式和贝叶斯公式的使用条件可以不是样本空间的划分,只要正确写出事件等式,由分解法就可以得到正确的解答,所以我们得到全概率公式和贝叶斯公式的新定义：

设 B_1,B_2,L,B_n 是两两互不相容的正概率事件,且 $A\subseteq\bigcup\limits_{i=1}^{n}B_i$,

则全概率公式 $P(A)=P(A|B_1)P(B_1)+P(A|B_2)P(B_2)+L+P(A|B_n)P(B_n)$；

贝叶斯公式为 $P(B_i|A)=\dfrac{P(A|B_i)P(B_i)}{\sum\limits_{j=1}^{n}P(A|B_j)P(B_j)}$,$i=1,2,L,n$。

2 利用分解法求解离散型随机变量和卡方分布的数学期望和方差

概率论的重要研究课题之一是希望从已知简事单件的概率推算出未知复杂事件的概率。为了达到这个目的,经常把一个复杂事件分解为若干个互不相容的简单事件之和,再通过分别计算这些简单事件的概率,最后利用概率的可加性得到最终结果。全概率公式就采用了数学中的这种分解法,把复杂事件的概率计算问题,分解为若干个简单事件的概率计算问题,再通过概率的可加性求出最终结果。事实上,这一解题思路有广泛的意义,如《高等数学》[2]中定积分、二重积分、曲线积分和曲面积分的定义就采用了分解法,其基本思路都是:分割求近似、求和取极限。本文将介绍分解法在《概率统计》离散型随机变量和卡方分布的数学期望和方差求解过程中的应用研究。

定义 设离散型随机变量的分布律为 $P\{X=x_k\}=p_k,k=1,2,L$,若级数 $\sum\limits_{k=1}^{\infty}x_kp_k$ 绝对收敛,则称级数 $\sum\limits_{k=1}^{\infty}x_kp_k$ 的和为随机变量 X 的数学期望,记为 $E(X)$[1]。

性质:设 X,Y 是两个随机变量,则有 $E(X+Y)=E(X)+E(Y)$,这一性质可以推广到任意有限个随机变量之和的情况[3]。

定义设 X 是一个随机变量,若 $E\{[X-E(X)]^2\}$ 存在,则称 $E\{[X-E(X)]^2\}$ 为 X 的方差,记为 $D(X)$[1]。

性质:(1)随机变量的方差可以按下列公式计算 $D(X)=E(X^2)-E^2(X)$;

(2)设 X,Y 是两个随机变量且相互独立,则有 $D(X+Y)=D(X)+D(Y)$,这一性质可以推广到任意有限个随机变量之和的情况[4]。

2.1 计算二项分布的数学期望和方差

(1)0－1 分布的数学期望和方差

设随机变量 X 只可能取 0 与 1 两个值,它的分布律是 $P\{X=k\}=p^k(1-p)^{1-k},k=0,1(0<p<1)$,则称 X 服从以 p 为参数的(0－1)分布或两点分布[5]。

容易按定义求得数学期望 $E(X)=0\times(1-p)+1\times p=p$,$E(X^2)=0^2\times(1-p)+1^2\times p=p$,

所以方差 $D(X)=E(X^2)-E^2(X)=p-p^2=p(1-p)$

(2)求二项分布的数学期望和方差

设试验 E 只有两个可能结果:A 和 $\bar{A}$,则称 E 为伯努利试验,设 $P(A)=p(0<p<1)$,此时 $P(\bar{A})=1-p$,将 E 独立重复地进行 n 次,则称这一串重复的独立试验为 n 重伯努利试验。以 X 表示 n 重伯努利试验中事件 A 发生的次数,则称随机变量 X 服从参数为 n,p 的二项分布[5],并记为 $X:b(n,p)$,其分布律为 $P\{X=k\}=C_n^k p^k(1-p)^{n-k},k=0,1,L,n$

解:如果按定义计算二项分布的数学期望和方差很麻烦,我们可以引入随机变量$X_k=\begin{cases}1, A\text{ 在第 }k\text{ 次试验发生,}\\0, A\text{ 在第 }k\text{ 次试验不发生,}\end{cases}$ $k=1,2,L,n$,易知 $X=X_1+X_2+L+X_n$,由于 X_k 只依赖于第 k 次试验,且各次试验相互独立,于是 X_1,X_2,L,X_n 相互独立,

因为 X_1,X_2,L,X_n 都服从 0—1 分布,所以由数学期望的性质得 X 的数学期望为 $E(X)=E(\sum\limits_{k=1}^{n}X_k)=\sum\limits_{k=1}^{n}E(X_k)=np$,

又由于 X_1,X_2,L,X_n 相互独立,由方差的性质得 X 的方差 $D(X)=D(\sum\limits_{k=1}^{n}X_k)=\sum\limits_{k=1}^{n}D(X_k)=np(1-p)$。

注:本例将以 n,p 为参数的二项分布分解为 n 个相互独立且都服从以 p 为参数的 0－1 分布的随机变量之和。将 X 分解成数个随机变量之和,然后利用随机变量和的数学期望等于随机变量数学期望之和来求数学期望,独立随机变量和的方差等于随机变量方差的和来求方差,这种处理方法具有一定的普遍意义。

2.2 计算负二项分布(巴斯卡分布)的数学期望和方差

(1)几何分布的数学期望和方差

进行重复独立试验,设每次试验的成功概率为 p,则失败概率为 $1-p(0<p<1)$,将试验进行到出现一次成功为止,以 X 表示所需的试验次数,称随机变量 X 服从以 p 为参数的几何分布[5],其分布律为 $P\{X=k\}=p(1-p)^{k-1},k=1,2,L$。

由定义计算 X 的数学期望

$$E(X)=\sum_{k=1}^{\infty}kp(1-p)^{k-1}=\sum_{k=1}^{\infty}(k-1+1)p(1-p)^{k-1}=\sum_{k=1}^{\infty}(k-1)p(1-p)^{k-1}+\sum_{k=1}^{\infty}p(1-p)^{k-1}$$

$$=\sum_{k=0}^{\infty}jp(1-p)^j+p\,\frac{1}{1-(1-p)}=(1-p)\sum_{k=1}^{\infty}kp(1-p)^{k-1}+1=(1-p)E(X)+1,$$

所以,$E(X)=\dfrac{1}{p}$,而

$$E(X^2)=\sum_{k=1}^{\infty}k^2p(1-p)^{k-1}=\sum_{k=1}^{\infty}(k-1+1)^2p(1-p)^{k-1}=\sum_{k=1}^{\infty}(k-1)^2p(1-p)^{k-1}+2\sum_{k=1}^{\infty}(k-1)p(1-p)^{k-1}+\sum_{k=1}^{\infty}p(1-p)^{k-1}$$

$$=\sum_{j=0}^{\infty}j^2p(1-p)^j+2\sum_{j=0}^{\infty}jp(1-p)^j+1=(1-p)\sum_{k=1}^{\infty}k^2p(1-p)^{k-1}+2\sum_{k=1}^{\infty}kp(1-p)^k+1=(1-p)E(X^2)+2(1-p)E(X)+1$$

所以,$E(X^2)=\dfrac{2-p}{p^2}$,故 X 的方差 $D(X)=E(X^2)-E^2(X)=\dfrac{1-p}{p^2}$。

(2)求负二项分布(巴斯卡分布)的数学期望和方差

进行重复独立试验,设每次试验的成功概率为 p,则失败概率为 $1-p(0<p<1)$,将试验进行到出现 r

次成功为止。以 X 表示所需的试验次数，称随机变量 X 服从以 r,p 为参数的负二项分布或巴斯卡分布[5]，其分布律为 $P\{X=n\}=C_{n-1}^{r-1}p^r(1-p)^{n-r},n=r,r+1,L$。

解：如果按照定义计算负二项分布的数学期望和方差很麻烦，我们将得到 r 次成功所需的试验次数 X 分解为 $X=X_1+X_2+L+X_r$，其中 X_1 表示第一次成功时试验的次数，X_2 表示第一次成功之后，直到第二次成功时所需的试验次数，X_3 表示第二次成功之后，直到第三次成功所需的试验次数。因为试验是相互独立的，且每次成功的概率为 p，所以，X_1,X_2,LX_r 都服从几何分布且相互独立，

所以由数学期望的性质得 X 的数学期望 $E(X)=E(\sum\limits_{k=1}^{r}X_k)=\sum\limits_{k=1}^{r}E(X_k)=r\,\dfrac{1}{p}=\dfrac{r}{p}$，

由方差的性质得 X 的方差 $D(X)=D(\sum\limits_{k=1}^{r}X_k)=\sum\limits_{k=1}^{r}D(X_k)=r\,\dfrac{1-p}{p^2}=\dfrac{r(1-p)}{p^2}$。

注：故本例是将以 r,p 为参数的负二项分布（巴斯卡分布）分解为 n 个相互独立且都服从以 p 为参数的几何分布的随机变量之和。

2.3 计算超几何分布的数学期望和方差

设袋子里有 N 个球，其中有 m 个白球，$N-m$ 个黑球，从中随机地（无放回）取出 n 个球，令随机变量 X 表示取出来的白球数，称 X 服从超几何分布[5]。其分布律为 $P\{X=k\}=\dfrac{C_m^kC_{N-m}^{n-k}}{C_N^n},k=0,1,\ldots,min\{n,m\}$

解：如果按定义计算超几何分布的数学期望很繁琐，所以我们引入随机变量 $X_i=\begin{cases}1,\text{第 } i \text{ 次抽得白球,}\\0,\text{第 } i \text{ 次抽得黑球,}\end{cases}$

$i=1,2,L,min\{m,n\}$，易知 X_i 服从 0−1 分布且 $P\{X_i=1\}=\dfrac{m}{N}$，所以 $E(X_i)=\dfrac{m}{N}$，$E(X_i^2)=\dfrac{m}{N}$。而 $X=X_1+X_2+L+X_n$ 表示 n 次抽球中抽出的白球数，它服从超几何分布，则由数学期望的性质得 X 的数学期望 $E(X)=E(\sum\limits_{k=1}^{n}X_k)=\sum\limits_{k=1}^{n}E(X_k)=n\,\dfrac{m}{N}=\dfrac{mn}{N}$，

因为取球是采用不放回抽样，所以 X_1,X_2,L,X_n 并不独立，不可以利用方差的性质直接求超几何分布 X 的方差。

但是，因为 $X=\sum\limits_{i=1}^{n}X_i$，所以 $E(X^2)=E[(\sum\limits_{i=1}^{n}X_i)(\sum\limits_{i=1}^{n}X_i)]=E[\sum\limits_{i=1}^{n}X_i(X_i+\sum\limits_{i\neq j}X_j)]=E[\sum\limits_{i=1}^{n}X_i^2+\sum\limits_{i=1}^{n}\sum\limits_{i\neq j}X_iX_j)]$

$=\sum\limits_{i=1}^{n}E[X_i^2]+\sum\limits_{i=1}^{n}\sum\limits_{j\neq i}E[X_iX_j]=n\,\dfrac{m}{N}+\sum\limits_{i=1}^{n}\sum\limits_{j\neq i}E[X_iX_j]$

而 $X_iX_j=\begin{cases}1 & X_i=1,X_j=1\\0 & \text{其它}\end{cases}$，

所以 $E(X_iX_j)=P\{X_i=1,X_j=1\}$

在给定第 i 次试验中取出的是白球的条件下，在剩余的 $N-1$ 个球中有 $m-1$ 个白球，它们被等可能地在第 $j(i\neq j)$ 次试验中抽取出来，因此当 $i\neq j$ 时，

$E(X_iX_j)=P\{X_i=1,X_j=1\}=P\{X_i=1\}P\{X_j=1|X_i=1\}=\dfrac{m}{N}\dfrac{m-1}{N-1}$

所以 $E(X^2)=n\,\dfrac{m}{N}+n(n-1)\dfrac{m}{N}\dfrac{m-1}{N-1}$，

故由方差的计算公式得 X 的方差 $D(X)=E(X^2)-E^2(X)=n\,\dfrac{m}{N}+n(n-1)\dfrac{m}{N}\dfrac{m-1}{N-1}-(n\,\dfrac{m}{N})^2=\dfrac{mn}{N}(1-\dfrac{m}{N})\dfrac{N-n}{N-1}$

2.4 计算卡方分布的数学期望和方差

设 X_1,X_2,L,X_n 是来自总体 $N(0,1)$ 的样本，称统计量 $\chi^2=X_1^2+X_2^2+L+X_n^2$ 服从自由度为 n 的卡方

分布[5],记为 $\chi^2 : \chi^2(n)$

解:因为 $X_i : N(0,1)$,以 $E(X^i)=D(X_i)+E^2(X_i)=1+0=1$,由定义可计算得到 $E(X_i^4)=3$,所以 $D(X_i^2)=E(X_i^4)-[E(X_i^2)]^2=3-1=2, i=1,2,L,n$

于是由数学期望的性质得到卡方分布的数学期望 $E(\chi^2)=E(\sum_{i=1}^{n}X_i^2)=\sum_{i=1}^{n}E(X_i^2)=n$,

由方差的性质得到卡方分布的方差 $D(\chi^2)=D(\sum_{i=1}^{n}X_i^2)=\sum_{i=1}^{n}D(X_i^2)=n$

3 结束语

化繁为简是我们日常生活中常用的一种思维方式,分解法就是该思维方式在数学中的应用研究。本文主要介绍了将《概率统计》中复杂的离散型随机变量分解为简单的随机变量之和,再利用数学期望和方差的性质求得复杂随机变量的数学期望和方差的方法。本文意在使广大读者了解数学中的分解法,理解数学中的分解法,并能在实践中灵活应用。

参考文献:

[1] 盛骤,谢式千,潘承毅.概率论与数理统计(第四版)[M].北京:高等教育出版社,2008:90-105.

[2] 同济大学数学系.高等数学(第七版)[M].北京:高等教育出版社,2014:224-236.

[3] 严士健,刘秀芳,徐承彝.概率论与数理统计[M].北京:高等教育出版社,2011:87-88.

[4] 魏宗舒.概率论与数理统计教程(第二版)[M].北京:高等教育出版社,2008:76-79.

[5] RONALD E.概率与统计[M].袁东学,译.北京:中国人民大学出版社,2016:125-131.

Two Notes on Total Probability Theorem

LI Jia[a], *DENG Youlian*[b]

(a. School of Computer and Information Engineering, b. School of Mathematics and Information Science, Jiangxi Normal University, Nanchang 330022, China)

Abstract: Total probability theorem and Bayes' rule are two important probability calculation theorems in *Probability Statistics*. The condition of using total probability theorem and Bayes' rule can be relaxed under the presupposition of mutually exclusive events with normal probability but not necessarily under the presupposition of sample space partition. The decomposition method, the idea of solving the problem in total probability theorem, can be widely applied in solving expected value and variance of discrete random variables in *Probability Statistics*. The two notes are beneficial to the supplement for the content of *Probability Statistics*, the stimulation of students' learning interest and the enhancement of students' comprehension and mastering of total probability theorem and Bayes' rule.

Keywords: total probability theorem; decomposition method; expected value; variance

(实习编辑:杨晓玲　责任校对:曲　比)