概率统计在实际问题中的应用举例

王　妍
(中国传媒大学理学院应用数学系,北京 100024)

摘　要:本文介绍了概率统计的某些知识在实际问题中的应用,主要围绕古典概型,全概率公式,正态分布,数学期望,极限定理等有关知识,探讨概率统计知识在实际生活中的广泛应用,进一步揭示概率统计与实际生活的密切联系,为应用概率知识解决实际问题,数学模型的建立,学科知识的迁移奠定一定的理论基础。

关键词:概率统计;古典概型;正态分布;数学期望;中心极限定理
中图分类号:O211.9　**文献标识码**:A　**文章编号**:1673-4793(2007)01-0015-05

The Application of Probability and Statistics in Reality

WANG Yan
(School of Science, Communication University of China, Beijing 100024, China)

Abstract: This paper introduced the appliance of probability and statistics in reality, including classical model, formula of total probability, normal distribute, mathematics expectation and the central limit theorem. While it also discusses the widely uses of probability and statistics and the close relationship with the real life. Therefore, it lays the theoretical foundation for the practical uses with probability and statistics, and the basis of mathematics model

Key words: probability and statistics; classical mordel; nrmal distribute; mathematics expectation; central limit theorem

1 引言

概率论在一定的社会条件下,通过人类的社会实践和生产活动发展起来,被广泛应用于各个领域,在国民经济的生产和生活中起着重要的作用。正如英国逻辑学家和经济学家杰文斯 (Jevons, 1835 - 1882)所说:概率论是“生活真正的领路人,如果没有对概率的某种估计,我们就寸步难行,无所作为。

在日常生活中,同样不难发现,周围的许多事物都和概率有着千丝万缕的联系,下面从几个方面具体阐述。

2 古典概型在博彩领域的应用

纵观概率发展的历史长河,可窥见概率和博彩已经鱼水相融。早在 15世纪上半叶,就已有数学家试图理论上思考赌博问题。从最初的意大利数学家帕乔利 (L. pacioli) 1494年出版的《算术》一书中提出赌注分配问题,到后来的卡丹 (Cardan Jerome, 1501 - 1576)重新就帕乔利赌注分配问题进行系列的理论探讨;从自然科学创始人之一的 ——伽利略 (Galileo, 1564 - 1642)解决掷骰子问题,到帕斯卡和费马用各自不同的方法解决 1654年 7月 29日法国

收稿日期:2006 - 06 - 23
作者简介:王妍 (1979 -),女 (汉族),辽宁葫芦岛人,中国传媒大学讲师. E - mail: wy@cuc. edu. cn

骑士梅累向帕斯卡提出的赌博问题,再到 1657年荷兰数学家惠更斯(G. Huygens, 1629 - 1695)一书《论赌博中的计算》的问世,都在探索赌博中的概率问题,并且也相应的使得概率论概念和定理得到延拓和发展。

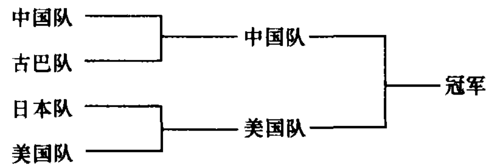
如今,博彩业雨后春笋般涌起,巨额奖金的诱惑,使得一些"有识之士"为实现自己的家庭梦想,不得不借助概率这个工具审时度势。下面一道例题作为对博采理论分析具有很好的指导作用:在考察时间跨度内,引起人们注意的偏号码或偏和值共有 10个,体彩"排列三"的"和 14"相邻两个开出期间隔甚至长达 96期[1],理论计算这些情况是否合理,在研究最初用到的就是古典概型和概率的有关性质。首先考虑各个位置号码,在 $k(k \geq 10)$ 期中,至少有某一位置的某一个数字没有被开出的概率为 $p_1 = 1 - (1 - (1 - 0.9^k)^{10})^3$。此问题抽象为概率问题,其实质是求"由 0～9十个数字组成的 k个位置的排列中,其中至少有一个数字在 k个位置都不出现的概率"。首先我们可以考虑:k个位置中某一个位置有一个数字不出现的概率为 $9/10$, k个位置都不出现该数字的概率则为 $(9/10)^k$(数字可以重复排列),而 k个位置至少有一个位置出现该数字的概率为 $(1 - 0.9^k)$,数字是 0～9中的任意一个,每个数字在该位置出现又是等可能的,所以 10个数字在此位置全出现过的概率为 $(1 - 0.9^k)^{10}$,根据概率性质,至少有一个数字在这个位置从未出现的概率为 $1 - (1 - 0.9^k)^{10}$,这样的位置有三个,所以 $(1 - (1 - 0.9^k)^{10})^3$。此问题的探讨反复利用概率的性质,最终使问题得到解决。

古典概型是概率里边最早的概型,也是应用较为广泛的概型。同时,古典概型与全概率公式相结合,对于实际问题中考虑整个系统的概率问题,或者得知整个系统的概率查找原因等问题,其作用也是不可磨灭的。下面例述全概率公式的作用。

3 全概率公式在实际问题中的应用

全概率公式是概率论中一个重要的公式,在实际中同样有广泛的应用。先引进定义:

设 $B_1, B_2, \ldots, B_n$为样本空间 Ω的一个划分,即 $B_1, B_2, \ldots, B_n$互不相容 $(B_i \cap B_j = \phi, i \neq j, i\ j = 1, 2, \ldots, n)$,且 $\bigcup_{i=1}^{n} B_i = \Omega$,如果 $P(B_i) > 0, i = 1, 2, \ldots, n$,则对任一事件 A有 $P(A) = \sum_{i=1}^{n} P(B_i) P(A \mid B_i)$。在某次世界女排赛中,中,日,美,古巴四对取得半决赛权,形式如下:



现根据以往的战绩,假定中国队战胜日本队,美国队的概率分别为 0.9与 0.4,而日本队战胜美国队的概率为 0.5,试问中国队取得冠军的可能性有多大?

根据上述形势,未完成的日美半决赛对中国冠军的影响很大,若日本队胜利,则中国队可有 90%的希望夺冠,若美国队胜利,则中国队夺冠的希望只有 40%。在日本队和美国队未比赛前,他们谁能取得半决赛权,两种情况都必须考虑到。

记"中国队得冠军"为事件 B,日本队胜美国队为事件 A_1,有 $P(A_1) = 0.5 = 50\%$.

美国队胜日本队为事件 A_2, $P(A_2) = 50\%$ 显然有,要么日本队胜,要么美国队胜,二者必居其一,所以 A_1, A_2 为一个划分,由全概率公式:这里 $(n = 2)$

$P(B) = P(A_1) P(B \mid A_1) + P(A_2) P(B \mid A_2)$ 其中 $P(B \mid A_1)$, $P(B \mid A_2)$,是两个条件概率。$P(B \mid A_1)$ 表示在日本队胜美国队的条件下中国队取得冠军的概率,由题意可知,$P(B \mid A_1) = 90\%$, $P(B \mid A_2)$ 表示在美国队胜日本队的条件下,中国队取得冠军的概率,由题意可知,$P(B \mid A_2) = 40\%$。

综上所述,在日、美未决赛前,估计中国队取得冠军的概率为:

$$P(B) = P(A_1) P(B \mid A_1) + P(A_2) P(B \mid A_2) = 50\% \times 90\% + 50\% \times 40\% = 50\% (90\% + 40\%) = 65\%$$ 。[5]

类似的利用全概率公式求解的案例有许多,比如工厂有多条流水线,求故障发生概率就是利用全概率公式求解,或者已知故障发生概率,追究不同流水线应承担的责任,利用的是全概率公式的反向——贝叶斯(逆概率)公式。在利用全概率公式求解实际问题中,关键是对问题的合理划分,考虑所有可能导致问题发生的情况。

全概率公式是考虑整个系统,渗透到生活各个方面的正态分布,对于许多实际问题的解决更是举足轻重。下面从几方面阐述一下正态分布在实际问题中的应用。

4 正态分布在实际问题中的应用

正态分布也称"高斯分布",是概率论中最重要的一个分布,中心极限定理也告诉我们许多其它分布的极限为正态分布。许多实际问题,我们都可以将其转化为正态分布加以解决。

首先,正态分布可以比较乘车时间长短,从而选择出行路线,看下例:

从南郊某地乘车前往北区火车站搭火车,有两条路可走,第一条路线穿过市区,路程较短,但交通拥挤,所需时间(单位为分)服从正态分布 $N(50,100)$,第二条路线沿环城公路走,路线较长,但意外阻塞较多,所时间服从正态分布 $N(60,16)$。

(1) 假如有 70分钟可用,问应走哪条路线?

(2) 若只有 65分钟可用,又应走哪条路线?

分析:从概率角度先考虑(1)的情况,有 70分钟可用时,根据正态分布的性质,分别求两种情况下的概率,又由于所有的正态分布都可以通过标准化化成标准正态分布,利用标准正态的性质或查找正态分布表,可以比较两条路线按时到达的概率大小,哪条大就走哪条路线。情况(2)与情况(1)同。具体解法如下:

(1) 有 70分钟可用,走路线一到达的概率:$P(\zeta \le 70) = \Phi(\frac{70-60}{4}) = \Phi(2) = 0.9772$ (2)走路线二到达的概率:$P(\zeta \le 70) = \Phi(\frac{70-60}{4}) = \Phi(2.5) = 0.9938$。所以应走路线二。

(2) 有 65分钟可用,走路线一到达的概率:$P(\zeta \le 65) = \Phi(\frac{65-50}{10}) = \Phi(1.5) = 0.9332$,走路线二到达的概率:$P(\zeta \le 65) = \Phi(\frac{65-50}{4}) = \Phi(1.25) = 0.8944$,所以应该走路线一[3]。

其次,正态分布可以帮助应聘者分析形式,对应聘状况做正确估计:具体案例如下:

某企业准备通招聘考试招收 300名职工,其中正式工 280人,临时工 20人;报考的人数是 1657人,考试满分是 400分,考试后得知,报名者的成绩 X近似服从正态分布 $N(166,\sigma^2)$,360分以上的高分考生 31人,某考生 B得 256分,问:他能否被录取?能否被聘用为正式工?

这类问题求解大致分为这样三个步骤:首先根据问题中所给信息:高于 360分的有 31人,利用分数服从正态分布,遇到正态分布一般多联想到标准正态分布,求出 σ^2;然后根据招收 300名职工这个信息,再次利用分数服从正态分布求出最低分数线,将 B的成绩与最低分数线比较,从而确定是否被录取;最后,如果根据比较结果确定 B被录取,再求出第 280人的分数与 B的成绩比较,或者根据 B的成绩求出高于 B成绩的人数,再与 280比较,进而确定 B是否被录取为正事员工。

具体求解如下:

第一步,预测最低分数线,设最低分数为 x_1,考生成绩为 X,则对一次成功的考试来说,$X \sim N(166,\sigma^2)$

因为高于 360分的考生的频率是$\frac{31}{1657}$,故

$$P\{X > 360\} = \{\frac{X-166}{\sigma} > \frac{360-166}{\sigma}\} = 1 - \Phi(\frac{31}{1657}) \approx \frac{31}{1657}$$

因此 $\Phi(\frac{194}{\sigma}) \approx 1 - \frac{31}{1657} \approx 0.981$,

查表可知$\frac{194}{\sigma} \approx 2.08$,解得 $\sigma \approx 93$,故 $X \sim N(166,93^2)$。

因为最低分数线的确定应使录取考生的频率等于$\frac{300}{1657}$,即

$$P\{X > x_1\} = P\{\frac{X-166}{93} > \frac{X_1-166}{93}\} = 1 - \Phi(\frac{X_1-166}{93}) \approx \frac{300}{1657}$$

所以 $\Phi(\frac{X_1-166}{93}) \approx 1 - \frac{300}{1657} \approx 0.819$

查表得$\frac{X_1-166}{93} \approx 0.91$,解得 $X_1 \approx 251$,也就是说,最低分数线是 251分。

第二步:预测 B的考试名次,这样就可以确定他是否能被录取。在 $x = 256$分时,由查表可知:

$$P\{X > 256\} = 1 - \Phi\left(\frac{256 - 166}{93}\right) = 1 - \Phi(0.9677) \approx 1 - 0.8315 = 0.1685$$

这表明,考试成绩高于 256分的频率是 0.1685,也就是成绩高于考生 B的人数大约占总考生 16.8%,所以名次排在考生 B之前的考生人数约有;1657 ×16.85% ≈ 280即考生 B大约排在 281名[2],即 B只能作为一名临时工被录用。

正态分布应用的广泛性有目共睹,但是,作为数字特征的期望,在探讨利润最优化问题时,其作用也是独树一帜的,下面探讨利用期望获取利润问题。

5 数学期望在求解最大利润问题中的应用

如何获取最大利润不但成为商界追求的目标,同时也为越来越多的人所关注,许多数学模型也从概率角度利用期望求解最大利润问题,为问题的解决提供新的思路。下面就是一道应用期望探讨利润的问题。

设某产品每周需求量 Q取 1, 2, 3, 4, 5为值,是等可能的,生产每件产品的成本为 $C_1 = 3$元;每件产品的售价为 $C_2 = 9$元;设售出的产品以每件 $C_3 = 1$元的费用存入仓库,问生产者每周生产多少件产品能使所期望的利润最大。

此问题的解决先是建立利润与销售量的函数;然后求利润的期望;即求关于销量 Q的函数的期望;得到关于 T关于生产量 N的函数,再求函数的导数,根据原函数和导函数的关系以及极值与导数的性质得出结果。

此外,期望的思想用于某项活动中,可以减少工作量。血液检查的案例很好的说明了期望的这方面作用。

例:通过血检对某地区的 N个人进行某种疾病普查,有两套方案:方案一是逐一检查,方案是分组检查. 那么哪一种方案好 ?如果用方案二应怎样分组可以减少工作量 ?显然方案一需要检查 N次。下面我们讨论方案二:假设检验结果阴性为“正常”、阳性为“患者”,把受检查分为 k个人一组,把这 k个人的血混和在一起进行检查,如果检验结果为阴性,这说明 k个人的血液全为阴性,因而这 k个人总共只要检验一次就够了;如果检验结果为阳性,要确定 k个人的血液哪些是阳性就需要逐一再检验,因而方案二在实施时有两种可能性,要和方案一比较,就要求出它的平均值 (即平均检验次数).

具体做法如下:假设这一地区患病率 (即检验结果为阳性的概率)为 p,那么检验结果为阴性的为阴性的概率为 $q = 1 - p$, 这时 k个人一组的混合血液是阴性的概率为 q^k,是阳性的概率为 $1 - q^k$,则每一组所需要的检验次数是一个服从二点分布的一个随机变量,即

ζ	1	$1 + k$
p	q^k	$1 - q^k$

由此可求得每组所需的平均检验次数为:$E\zeta = 1 \times q^k + (1 + k) \times (1 - q^k) = 1 + k - kq^k$由以上计算结果可以得出:当 $1 + k - kq^k < k$,即 $kq^k > 1$, $q^k > 1/k$时,方案二就比方案一好,总的检验次数为 $(1 + k - kq^k) \times N/k$。某医疗机构在一次普查中,由于采用了上述分组方法,结果每 100个人的平均检验次数为 21,减少工作量达 79%。[3]

上述例子足见期望在解决实际问题中的重要性,而作为概率核心定理的中心极限定理也在某些实际问题求解中起到指导性的作用,下面阐述一下中心极限定理的实际应用。

6 中心极限定理在实际问题中的应用

中心极限定理指出:如果一个随机变量由众多的随机因素所引起,每个年十的变化起着不大作用,就可以推断描述这个随机现象的随机变量近似的服从正态分布 ,所以要求随机变量之和落在某个区间上的概率,只要把它标准化. 用正态分布作近似计算即可.

中心极限定理对保险业具有指导性的意义,一个保险公司的亏盈,是否破产,和通过学习中心极限定理的只是都可以做到估算和预测,大数定律是近代保险业赖以建立的基础. 下面例题阐述了大数定律和中心极限定理在保险业中的重要作用和具体应用。

已知在某人寿保险公司有 10000个同一年龄段的人参加保险,在同一年里这些人死亡率为 0.1%,死亡的家属在一年的头一天交付保险费 10元,死亡时家属可以从保险公司领取 2000元的抚恤金,求保

险公司一年中获利不少于 40000元的概率;保险公司亏本的概率是多少?

解:设一年中死亡的人数为 x人,死亡率为 $p = 0.001$,把考虑 10000人在一年里是否死亡看成10000重贝努里试验,保险公司每年收入为 $10000 \times 10 = 100000$元,付出 2000元。

P(保险公司获利不少于 40000元) $= P^{6}\{(10000 - 2000) > 40000\} = P(0 < x < 30)$ $np = \sqrt{D(x)} = \sqrt{np \times (1-p)} = \sqrt{10 \times 0.999} = 3.161$

$$P\{0 \le x \le 30\} = P\left\{\frac{-10}{3.161} \le \frac{x-10}{3.161} \le \frac{50-10}{3.161}\right\} = 1 - \Phi(1.6542) + \Phi(-3.1641) = 0.0008$$[4]

上面只是列举了概率在实际问题中应用的几个小片段,然而,作为一门独立的学科,概率的足迹可以说已经深入到每一个领域,在实际问题中的应用随处可见。尤其随着科技飞速发展,知识产业化的今天。许多基础学科从幕后走到台前,概率的许多其他方面也正在或将要发挥它应有的作用。诸如方差分析、回归分析等内容在医学,军事等领域都正在发挥它的最大作用。相信人类能够更好的"挖掘概率的潜能",使之最大限度地为人类服务。

参考文献:

[1] 李刚我国彩票业现状实证分析与未来发展对策的研究[J].复旦大学博士论文, 2006.

[2] 陈文灯,黄开先等. 概率论与数理统计复习指导——思路、方法与技巧[M]. 北京:清华大学出版社, 2003.

[3] 魏宗舒等. 概率论与数理统计[M]. 北京:高等教育出版社, 2004.

[4] 菜海涛等编著. 概率论与数理统计典型例题与解法[M]. 长沙:国防科技大学出版社, 2003.

[5] 杜镇中. 全概率公式及其应用[J]. 遵义师范学院学报, 2005, 7(5): 76—77.

(责任编辑:龙学锋)

(上接第 43页)

参考文献:

[1] 尹长川,等,多载波宽带无线通信技术[M]. 北京:邮电大学出版社, 2004. 7.

[2] 李栋. 数字声音广播[M]北京:北京广播学院出版社, 2001.

[3] 佟学俭,罗涛. 移动通信技术原理与应用[M]北京:人民邮电出版社, 2003.

[4] Xianbin Wang Reduction of Peak - to - Average Power Ratio of OFDM System Using A Companding Technique [J]. IEEE Transactlons on Broadcasting 1999, 45(3).

[5] Yuanbin Guo, Joseph R. Cavallaro. Reducing peak - to - average power patio in OFDM systems by adartive dynamic range companding Rice University, 2004. http://cml rile. edu/docs/Guo2002may5OntheReduc. pdf

[6] Tao Jiang, Guangxi Zhu. Nonlinear Companding Transform for Reducing Peak - to - Average Power Ratio of OFDM Signal [J]. IEEE Transactions on Broadcasting, 2004, 50(3).

[7] Mary Ann Ingram. OFDM Simulation Using Matlab [M]. Smart Antenna Research Laboratory, August 2000. http://users ece. getech. edu/～mai/totorial/OFDM/Tutorial/_web. pdf

[8] 张乃谦. 李栋. OFDM峰均值功率比研究[A]. 第十届海峡两岸无线电技术研讨会论文集[C]. 北京:中国传媒大学出版社 2005. 37 - 41.

(责任编辑:王　谦)