

案例 1

大数定律与中心极限定理在彩票选号中的应用

1、背景知识

概率统计的应用十分广泛，它几乎遍及所有学科领域，买彩票中奖机会的大小，就是一个典型的概率统计问题。无论是福利彩票还是体育彩票，开奖结果中的数字都是随机产生的。每次随机确定一个数字（或称一次试验），其结果具有一定的偶然性，不过在大量重复试验中，中奖号码会呈现出某些必然规律。本文针对彩票奖号可重复出现数字情况，用概率统计方法寻找中奖号码规律。

2、案例分析

（1）先来求奖号中有重复数字的概率。目前，全国大多数地区体育彩票中奖号码是从0到9这十个数字中，可重复抽取七个数字依次排列组成，对于这种确定中奖号码的方式，可计算有重复数字的概率。由古典概率计算方法，中奖号码中七个数字全部不同的概率为

$$\frac{10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4}{10^7} = 0.06048$$

那么，七个数字中至少有两个数字相同的概率为 $1 - 0.06048 = 0.93952$ 。

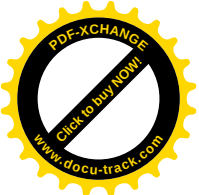
策略一：每注彩票七个数字中至少有两个相同。

（2）利用大数定律确定备选取数字。

大数定律：设 n_A 是 n 次独立重复试验中事件A发生的次数， p 是事件A在每次试验中发生的概率，则对于任意给定正数 $\varepsilon > 0$ ，有 $\lim_{n \rightarrow \infty} P(|\frac{n_A}{n} - p| < \varepsilon) = 1$ ，这个定律说明：当重复试验很多次时，随机事件A发生的频率 $\frac{n_A}{n}$ 在它的概率 p 附近摆动，若 $\frac{n_A}{n}$ 偏离 p ，只要继续试验下去， $\frac{n_A}{n}$ 就有向 p 靠近的趋势。

统计连续 n 期中奖号码，第 i 位上0到9个数字出现的次数，依次记为 $n_{i0}, n_{i1}, \dots, n_{i9}$ ，把较小的 K 个数之和记为 n_{Ai} ，即 n_{Ai} 是出现次数较少的 K 个数字一共出现的次数。它们出现的频率为 $\frac{n_{Ai}}{n}$ ，对应的概率大约为 $\frac{K}{10}$ ，则 $\frac{K}{10} - \frac{n_{Ai}}{n}$ 一般偏大，由大数定律，第 $n+1$ 期中奖号码第 i 位上频率 $\frac{(n+1)_{Ai}}{n+1}$ 应有向概率 $\frac{K}{10}$ 靠近的趋势，因而第 i 位上出现这 K 个数字之一的概率非常大。在第 $n+1$ 期投注时，第 i 位应首先考虑从这 K 个数字中挑选理论上讲，彩票投注范围共有 10^7 个不同号码，如果依上述方法，每个位置上确定 $K=5$ 个备选数字，备选号码范围就缩小到 5^7 个不同号码。当然，可能第 $n+1$ 期中奖号码的第 i 位上没有出现备选取的5个数字，由大数定律，往后的中奖号在第 i 位上出现这5个数字之一的可能性更大。

策略二：前 n 期中奖号在第 i 位上出现较少的数字，作为第 $n+1$ 期第 i 位上的备



选数字。

(3) 利用中心极限定理确定投注号码数字和的范围

中心极限定理：设随机变量序列 $\{X_i, i \geq 1\}$ 相互独立同分布，且具有数学期望

和方差，则当 n 充分大时，近似有 $\frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n\sigma}} : N(0,1)$ ，由标准正态分布的密度

函数的特征，易见中心极限定理的直观意义：当 n 充分大时， $\sum_{k=1}^n X_k$ 在其数学期

望 $n\mu$ 附近取值的概率非常大。

统计江苏省体育彩票中奖号数据（不包括2006年第46期以前的特别号码），得到0到9各数字出现的次数和频率，除数字9外，各数字出现的频率有向0.1靠近的趋势，为方便起见，不防设0到9各数字出现的概率均为0.1。记随机变量

$\{X_i, i=1, 2, \dots, L\}$ 为第 i 次确定的数字，易见 $\{X_i, i \geq 1\}$ 相互独立同分布， $\{X_i, i \geq 1\}$ 的

数学期望和方差为 $EX_i = 4.5, DX = 8.25$ ，令 $\xi_{7n} = X_1 + \dots + X_{7n}$ 是连续 n 期中奖号各

位数字总和，由和式和独立性，可得 $E(\xi_{7n}) = 31.5n, D(\xi_{7n}) = 57.75n$ ，由中心

极限定理，当 $7n$ 充分大时，有 $\frac{\xi_{7n} - 31.5n}{\sqrt{57.75n}} : N(0,1)$ ，那么 ξ_{7n} 的保证概率为0.6827

的估计区间是 $(31.5n - \sqrt{57.75n}, 31.5n + \sqrt{57.75n})$ 。

当 $n=1$ 时，区间为 $[24, 39]$ ，若 ξ_{7n} 的观察值靠近或大于上限 $31.5n + \sqrt{57.75n}$ ，

由中心极限定理， $\xi_{7(n+1)}$ 有靠近期望 $31.5(n+1)$ 的趋势，那么第 $n+1$ 期七个数字

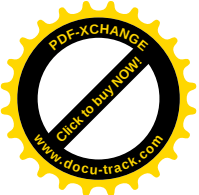
之和的估计区间的上下限应该分别小于区间 $[24, 39]$ 的上下限；若 ξ_{7n} 的观察值

靠近或小于下限 $31.5n - \sqrt{57.75n}$ ，在第 $n+1$ 期投注时，应考虑把区间 $[24, 39]$ 的上下限增大。

策略三：若连续 n 期中奖号的 $7n$ 个数字之和 ξ_{7n} 靠近 $31.5n + \sqrt{57.75n}$ 或 $31.5n - \sqrt{57.75n}$ ，就适当下调或上调区间 $[24, 39]$ 的上下限，所得区间作为第 $n+1$ 期投注号码的七个数字之和的范围。

3、结果说明

(1) 文中用极限定理观察中奖号码的运动趋势，要求观察次数足够多。在策略二中， n 的范围以 $30 \leq n \leq 50$ 为宜；在策略三中，最好 $5 \leq n \leq 7$ ，即连续观察



5至7期中奖号的数字。

(2)每期彩票特等奖号码只有一个，备选数字配置的所有号码有可能不包括特等奖号码，不过它覆盖部分中奖号码的概率非常大，对于仅期望能中奖（包括五等奖）的彩民，可以按文中介绍的三个策略有节制地购买彩票。