赠券收集问题

温馨提示：读懂本文需要学习过浙大版《概率论与数理统计》或类似课程。

1. **问题定义**

假设有n种赠券，每种赠券获取机率相同，而且赠券亦无限供应。若取赠券t张，求能集齐n种赠券的机率。

1. **数学期望**

对于赠券收集问题，所有赠券都集齐的最小抽取次数的数学期望为

其中 为第n个调和数，即 。

证明：

设 为表示n种赠券都至少被抽中一次时的抽取次数的离散随机变量。

设 为表示k – 1种不同赠券被抽中后，为抽中第k种赠券所需的抽取次数的离散随机变量。显然 ，因为第一个被抽中的赠券与所有已抽中赠券相比，一定是独一无二。

由数学期望的线性可加性[[1]](#footnote-1)，得

当k - 1种不同的赠券被抽中后，还有n – k + 1种赠券还没被抽中。设 为表示在下一次抽奖中，抽中新品种赠券的随机事件。根据题设，每次抽取每种赠券获取机率相同，可得

且 服从几何分布，即 。

所以 。

所以 。

数学期望的估计公式：

调和数没有解析解，但是有估计公式。

当n特别大时，第n个调和数的估计值为

其中 为欧拉-马斯克若尼常数，

因此，赠券收集问题的数学期望的估计公式为

当 时，上述估计公式精确到小数点后一位。

1. **方差**

对于赠券收集问题，所有赠券都集齐的最小抽取次数的方差为

其中 为第n个调和数，即 。

证明：

设 为表示n种赠券都至少被抽中一次时的抽取次数的离散随机变量。

设 为表示k – 1种不同赠券被抽中后，为抽中第k种赠券所需的抽取次数的离散随机变量。

因为 ，所以

因为 之间相互独立，所以

得证。

方差的估计公式：

由巴塞尔问题可得 。

所以当n特别大时， 。

所以方差的估计公式为

1. **累积分布函数**

为了计算双侧置信区间和单侧置信区间，需要计算累积分布函数。

在计算累积分布函数前，先介绍第二类Stirling数。第二类Stirling数实际上是集合的一个拆分，表示将n个不同的元素拆分成m个集合的方案数，记为 或 。

第二类Stirling数的**递推公式**为

证明：

为求得递推公式，可从定义出发，考虑第n + 1个元素的情况，即假设要把n + 1个元素分成m个集合。

情况1：如果n个元素构成了m - 1个集合，那么第n + 1个元素需要单独构成一个集合，方案数为 。

情况2：如果n个元素已经构成了m个集合，则需将第n + 1个元素插入到任意一个集合，方案数 。

综合两种情况，可得 。

第二类Stirling数的**通项公式**为

证明：

第二类Stirling数为将n个不同的小球放入m个相同的盒子里（n≥m）且不允许有空盒的方案数。

现假设将n个不同的小球放入m个不同的盒子里，允许有空盒的方案总数为 。

设 表示第i个盒子为空时的方案总数。根据假设情形的定义，可得下列三个公式：

根据容斥原理[[2]](#footnote-2)，可得

所以没有空盒的方案数为

当盒子相同时，方案数为上式处以盒子的全排列数 。

所以 。

得证。

第二类Stirling数的性质：

上述性质均可用小球盒子模型证明，证略。

对于赠券收集问题而言，可将t次抽取看作t个小球，n种赠券看作n个不同盒子，则有 种方案使得每个盒子不为空，即在t次抽取后，每种赠券至少抽得一张。

又因为t次抽取的方案总数为 ，所以赠券收集问题的累积分布函数为

上式的具体含义为t次抽取后，n种赠券中的每种赠券至少已经抽得一张的概率。

1. **概率密度函数**

记抽取次数为T，赠券种类为N，则赠券收集问题的概率密度函数为

上式的具体含义为在第t次抽取时，恰好抽中第n种赠券的概率。

证明：

因为在第t次抽取时，恰好抽中第n种赠券，所以前面的t – 1次抽取一共抽中n - 1种赠券，方案总数为 。因为第t次抽取抽中的可能是n种赠券中的任意一种，所以在第t次抽取时，恰好抽中第n种赠券的方案总数为

又因为t次抽取的方案总数为 ，所以赠券收集问题的概率密度函数为

得证。

赠券收集问题的概率密度函数同样可由累积分布函数推导得到：

1. 见浙江大学《概率论与数理统计》（第四版） 高等教育出版社，第98页，数学期望的性质3 [↑](#footnote-ref-1)
2. 容斥原理为-。容斥原理可用数学归纳法证明，具体证明略。 [↑](#footnote-ref-2)