Advances in Applied Mathematics 应用数学进展, 2022, 11(8), 5951-5955
Published Online August 2022 in Hans. http://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2022.118626

Hans汉斯

基于概率统计角度分析核酸检测分组问题

李　俏

辽宁师范大学数学学院，辽宁　大连

收稿日期：2022年7月22日；录用日期：2022年8月15日；发布日期：2022年8月25日

摘　要

进入2022年，新型冠状病毒仍在我国肆虐，各地区核酸检测已经常态化。本文基于我国人口众多的基本国情，将从概率统计的角度分析在大规模核酸检测中应该如何确定混检人数问题。对各地区实际情况进行数据整理，统计分析，建立数学模型，提出了核酸检测分组问题的最佳解决方案。

关键词

核酸检测，最佳分组，概率统计

The Grouping Problem of Nucleic Acid Detection Was Analyzed Based on Probability Statistics

Qiao Li

School of Mathematics, Liaoning Normal University, Dalian Liaoning

Received: Jul. 22nd, 2022; accepted: Aug. 15th, 2022; published: Aug. 25th, 2022

Abstract

Entering 2022, the new coronavirus is still raging in my country, and nucleic acid testing in various regions has been normalized. Based on the basic national conditions of my country's large population, this article will analyze how to determine the number of mixed tests in large-scale nucleic acid testing from the perspective of probability and statistics. Data sorting, statistical

文章引用: 李俏. 基于概率统计角度分析核酸检测分组问题[J]. 应用数学进展, 2022, 11(8): 5951-5955.
DOI: 10.12677/aam.2022.118626

analysis, and mathematical model establishment were carried out on the actual situation of each region, and the best solution to the problem of nucleic acid detection grouping was proposed.

Keywords

Nucleic Acid Detection, Optimal Grouping, Probability Statistics

1. 引言

就目前国家采取的核酸检测方法来看，相关医疗小组开发了五合一、十合一和二十合一的混检方式，其中十合一的混检方式被大部分城市地区应用。国务院应对疫情采取联防联控机制，医疗救治组曾在2020年发出《关于印发新冠病毒核酸 10 合 1 混采检测技术规范的通知》[1]，其中强调了针对新冠病毒核酸 10 合 1 混采检测(10-in-1 test)技术，从而进一步提升核酸检测能力和效率[2]。但是就此前全国疫情形势来看，存在疫情高暴发区和低暴发区，从节约资源，保证效率的角度考虑，各个地区因确诊率不同或确诊人数不同不应该采用相同人数的检测方案，并且随着居家隔离，全域静态等措施的施行各轮核酸检测人数也不应相同。目前，曹振民[3]从数学组合角度进行了分析，但是从概率统计角度讨论分组问题的文献较少，因此具有一定的研究意义。

2. 基本假设

1) 排除偶然事件，假设混合检测的准确率为 100%，即不存在检测错误的情况；
2) 检测出的阳性试管将在下一轮核酸前单独进行单人单管检测；
3) 检测出阳性的病例将不再参加下一轮的核酸检测；
4) 假设每轮核酸检测出的阳性病例为 60%；
5) 假设出现阳性病例后进行封城管控，人员流动性较小，传染率为 5%。

3. 检测模型说明

3.1. 单轮核酸检测

利用概率论与数理统计的相关知识，将某地区中每人所做核酸检测次数看作一个随机变量，不同的随机变量可能拥有不同的分布，包括分布列、密度函数或者分布函数，这些分布全面描述了随机变量取值的统计规律性，用分布可以算出有关随机变量时间的概率。除此之外，我们可以算得相应的期望，用来描述分布的特征。下面我们来定义数学期望[4]。

设离散随机变量 X 的分布列为

$$p(x_i)=P(X=x_i),i=1,2,\cdots.$$

如果

$$\sum_{i=1}^{\infty}|x_i|\,p(x_i)<\infty$$

则称

$$E(X)=\sum_{i=1}^{\infty}x_i p(x_i)$$

随机变量 X 的数学期望，或称为该分布的数学期望[3]。

数学期望 $E(X)$的物理解释是重心。若把概率 $p(x_i)=P(X=x_i)$ 看作点 x_i 上的质量，概率分布看作质量在 x 轴上的分布，则 X 的数学期望 $E(X)$就是该质量分布的重心所在位置。数学期望是消除随机性的主要手段，具有广泛的应用。

此问题中，我们将一个地区一轮疫情中核酸检测的总人数设为 N，并将这 N 人按 k 个人一组进行分组，把同组 k 个人的采样样本混合后混检，如果混合样本呈阴性反应，说明 k 个人的样本都呈阴性反应，这 k 个人都无感染病毒，因而这 k 个人只要检测 1 次，相当于每人检测 $\frac{1}{k}$ 次。如果混合样本呈阳性反应，说明 k 人中，至少一人样本呈阳性反应，则再对此 k 个人的样本分别进行检验，因而这 k 个人的样本要检验 $k+1$次，相当于每个人检验 $1+\frac{1}{k}$ 次，这时增加了检验次数。假设一个地区的确诊率为 p, X 为该地区中每人需要的核酸次数，则 X 的分布列为：

X	$\frac{1}{k}$	$1+\frac{1}{k}$
P	$(1-p)^k$	$1-(1-p)^k$

所以每人平均核酸次数为

$$E(X)=\frac{1}{k}(1-p)^k+\left(1+\frac{1}{k}\right)\left[1-(1-p)^k\right]=1-(1-p)^k+\frac{1}{k}$$

因此，只要知道一个地区的确诊率 p，就可以求出核酸检测次数，$E(X)$越小，所对应的分组混检人数最高效。需要指出的是，本论文在计算确诊率时，确诊人数为 3 月到 4 月每日新增确诊人数与无症状感染者之和，于是确诊率为：$\frac{\text{总确诊人数}}{\text{检测总人数}}$。

3.2. 多轮核酸检测

在进行多轮核酸检测中，需要排除上一轮中已经检测出的阳性病例，还需要考虑阴性检测者在核酸检测或者日常生活中被感染的概率。假设在每一轮核酸检测中检测出阳性的概率为 q，传染率为 r，则在进行多轮核酸检测中，确诊率的变化模型为：

$$X_n=p,n=1;$$

$$X_n=\frac{p(1-q+r)^{n-1}}{1-\sum_{i=0}^{n-2}pq(1+q-r)^i},n=2,3,4,\cdots.$$

具体实现过程为：根据第一轮得到的检测总人数和确诊人数求得第一轮的确诊率，根据检测出阳性概率 q 和传染率 r，算出第一轮检测出的阳性人数和传染人数[4]，那么，

第 n 轮的确诊人数 = 第 $(n-1)$ 轮确诊人数 − 第 $(n-1)$ 轮检测出阳性人数 + 第 $(n-1)$ 轮被传染人数，

而

$$第n轮检测总人数 = 第(n-1)轮检测总人数 - 第(n-1)轮检测出阳性人数,$$

故

$$第n轮确诊率 = \frac{第n轮阳性人数}{第n轮检测总人数}.$$

由上可知，对于多轮核酸检测情况，有必要调整混检方案。

4. 举例分析

在现实生活中，疫情一旦爆发，就不能只做一轮核酸检测，往往是多轮的。因此，这里只以多轮核酸检测举例。

我们以 2022 年 3 月上海疫情中具有确诊病例的某 2000 人小区为例，具体分析在 4 轮核酸检测中，应该如何调整混检方案。

假设该小区第 1 轮中检测总人数为 2000 人，确诊率为 $p = 0.02$，核酸检测出阳性的概率 $q = 0.6$，传染率为 $r = 0.05$，那么根据数学建模公式可以算出：

$$第2轮的确诊人数 = 2000\times0.02 - 2000\times0.02\times0.6 + 2000\times0.02\times0.05 = 18,$$
$$第2轮检测总人数 = 2000 - 2000\times0.02\times0.6 = 1976,$$
$$第2轮确诊率 = \frac{第2轮的确诊人数}{第2轮检测总人数} = \frac{18}{1976} \approx 0.008.$$

仿照第 2 轮的求解方法，可以求出第 3，4 轮的确诊率，如表 1 所示。

Table 1. Diagnosis rate of multiple rounds of testing
表 1. 多轮检测确诊率

轮数	第 1 轮	第 2 轮	第 3 轮	第 4 轮
确诊率	0.02	0.008	0.003	0.001

之后，分别对以上 4 轮确诊率进行分组人数的求解，我们采用 mathematics 软件进行运算，具体运算公式为：

$$\text{Simplify}\left[\text{Minimize}\left[\left\{\frac{1}{k}(1-p)^k + \left(1+\frac{1}{k}\right)\left(1-(1-p)^k\right), 30 > k > 1\right\}, k\right]\right]$$

将每轮确诊率 p 代入求解结果如表 2 所示。

Table 2. Multiple rounds of detection of optimal group number
表 2. 多轮检测最佳分组人数

轮数	第 1 轮	第 2 轮	第 3 轮	第 4 轮
人数	7.59664	11.6945	18.7653	30

同样地，根据 MATLAB 软件编程可以得出同样的结论，即检测次数最少的为最佳混检人数，如图 1 所示。

第一轮核酸检测		第二轮核酸检测		第三轮核酸检测		第四轮核酸检测	
混检人数	检测次数	混检人数	检测次数	混检人数	检测次数	混检人数	检测次数
5	554	5	462	5	427	5	411
6	522	6	411	6	367	6	347
7	511	7	379	7	326	7	302
8	510	8	358	8	298	8	270
9	516	9	346	9	277	9	245
10	526	10	339	10	262	10	226
11	544	11	335	11	249	11	210
12	561	12	333	12	241	12	198
13	581	13	335	13	234	13	188
14	593	14	339	14	230	14	180
15	618	15	345	15	228	15	174
16	648	16	352	16	227	16	170
17	661	17	353	17	226	17	163
18	687	18	367	18	226	18	161
19	705	19	372	19	225	19	158
20	736	20	384	20	227	20	156
21	749	21	391	21	231	21	154
22	776	22	396	22	229	22	151
23	787	23	408	23	233	23	151
24	814	24	417	24	238	24	151
25	845	25	431	25	242	25	151
26	862	26	438	26	244	26	149
27	884	27	454	27	248	27	150
28	909	28	461	28	253	28	150
29	926	29	461	29	250	29	149
30	933	30	480	30	259	30	148

Figure 1. Number of multiple rounds of testing
图 1. 多轮核酸检测次数

5. 结论

本文利用概率论的知识给出了单轮和多轮核酸检测的数学模型，之后以上海某小区为例，对模型进行了实际应用，得出了在多轮核酸检测时，需要调整混检方案的结论，即随着确诊率的不断下降，混检中每组的检测人数应随之增加。该模型具有一定的实际应用意义。

参考文献

[1] 卫生健康委网站. 关于印发新冠病毒核酸 10 合 1 混采检测技术规范的通知[EB/OL]. http://www.gov.cn/xinwen/2020-08/19/content_5535756.htm, 2020-08-19.

[2] 新华社. 我国核酸检测能力达每天 5165 万管，检测能力和可及性不断提升[EB/OL]. http://www.news.cn/2022-04/17/c_1128568759.htm, 2022-04-17.

[3] 曹振民. 核酸检测最佳分组人数分析[J]. 延安职业技术学院学报, 2021, 35(6): 96-99.

[4] 茆诗松. 概率论与数理统计[M]. 第 2 版. 北京: 高等教育出版社, 2011.