第 36 卷第 6 期
2020 年 12 月

大学数学
COLLEGE MATHEMATICS

Vol. 36, №. 6
Dec. 2020

基于数学期望的新冠肺炎核酸检测方法

赵小艳
(西安交通大学 数学与统计学院,西安 710049)

[摘 要] 以武汉市全民新冠肺炎核酸检测方法为例,阐述数学期望在实际问题中的应用,培养学生用概率统计课程所学的理论知识、统计方法解决实际问题的能力,诱导学生养成用数学理论、方法描述实际问题的思维习惯,以此激发学生学习数学的兴趣.

[关键词] 数学期望;混样检测;新冠肺炎核酸检测

[中图分类号] O211.9 [文献标识码] B [文章编号] 1672-1454(2020)06-0019-04

1 引 言

2020 年初,一场突如其来的新冠肺炎疫情[1]在全球迅速蔓延,不但打乱了相关人群正常的生活秩序,也给生命财产带来损失,是全球人民不得不面对的一场生存挑战. 如何迅速判断疫情的发展势态,并果断采取相应的措施进行疫情防控,是相关国家、省、市面临的重大问题. 为了给人民营造安全健康的生活环境,遏制经济继续衰退,尽快恢复正常社会生活秩序,5 月 11 日,武汉市新冠肺炎疫情防控指挥部涉疫大数据与流行病学调查组发布紧急通知,决定在武汉市进行全民新冠病毒核酸检测.

众所周知,武汉市是一个拥有千万以上人口的城市,逐人进行核酸检测不但工作量大,而且因为程序的繁杂还有可能导致漏检、误检、迟检等问题发生. 因此,设计一种科学、有效、全面、快速的检测方法至关重要.

本文通过分析此次疫情期间武汉市所采用的核酸混样检测方法,阐述概率论与数理统计课程中数学期望理论知识在混样检测中的应用,揭示解决实际问题时数学知识与方法的重要性.

2 混样检测方法

在医学疾病筛查检测中,常规的检测办法是逐人依次检测,也称作单样本检测. 对样本量比较小的情况,单样本检测可以保证检测结果及时有效,而对量大(如百万级以上样本),且患病率很低的疾病样本进行检测时,采用单样本检测进行筛查不仅需要大量的人力、物力和财力,而且比较费时,同时也可能因为耗时较多,影响检测速度,不利于相关单位及时制定措施. 如:2020 年 4 月底,武汉市常驻人口约 1100 万人,武汉市核酸检测机构有 55 家,核酸采样点 211 个,单日核酸检测量最多达 6.3 万人份. 按此检测速度,完成全市、全民核酸检测筛查需要 5 个月左右. 因此,选择科学、有效的检测方法,以便快速、准确地完成新冠病毒核酸检测筛查,就显得尤为重要.

混样检测是一种常用的提高检测效率的检测方法. 所谓混样检测就是把要检测的样本混合在一起做一次检测. 在具体实施时,由于样本量很大,通常会依据某个原则将检测样本分成若干组,对每组的 k 个样

[收稿日期] 2020-08-23; [修改日期] 2020-08-28
[基金项目] 西安交通大学本科教学改革研究基础课专项(1802Z-02);西安交通大学钱学森学院荣誉课程建设项目(C 类-2)
[作者简介] 赵小艳(1976—),女,博士,副教授,从事数学教学与研究. Email:zhaoswallow@xjtu.edu.cn

本进行混样检测,根据检测结果判断这组样本的检测结果. 如果混样检测结果显示正常,则说明这 k 个样本的检测结果都是正常的,即用混样检测结果代替混样个体样本检测结果;如果混样检测结果不正常,则说明这组样本中至少有一样本检测结果应属于不正常,此时,需对这 k 个样本进行单样本检测,方能确定本组样本个体结果. 这样,对该组样本实际上做了 $k+1$ 次检测. 事实上,在实际检测中,k 的选择不但与疾病的患病率 p 有关,而且还与检测方式的灵敏性等因素有关,需要综合考虑,才能制定科学的混样检测方案. 下面进行具体分析.

假设有一批需要检测的样本,每个样本可能被检测为呈阳性(即样本不正常)的概率为 p,自然检测为阴性(即样本正常)的概率是 $q=1-p$. 将这批样本分成若干组,每组 k 个.

首先假设:检测方式对要检测项的灵敏性比较高,也就是说,如果混合样本中有一个样本呈阳性,则混样检测结果也呈阳性;如果混样检测结果呈阴性,则认为该组人的检测结果都呈阴性.

若用 X 表示每组中每个人需要的检测次数,则 X 是离散型随机变量,且

$$X=\begin{cases}\dfrac{1}{k}, & \text{该组混合样本检测结果呈阴性},\\ 1+\dfrac{1}{k}, & \text{该组混合样本检测结果呈阳性}.\end{cases} \tag{1}$$

易得 X 的分布律为

$$P\left\{X=\frac{1}{k}\right\}=q^k,\quad P\left\{X=1+\frac{1}{k}\right\}=1-q^k.$$

这样,随机变量 X 的数学期望 $\mathrm{E}(X)$ 表示每组中每个人的平均检测次数[2],且

$$\mathrm{E}(X)=\frac{1}{k}\cdot q^k+\left(1+\frac{1}{k}\right)\cdot(1-q^k)=1+\frac{1}{k}-q^k.$$

为了有效降低检测工作量,对给定的 $q=1-p$,只要选取每组人数 k,使得

$$\mathrm{E}(X)=1+\frac{1}{k}-q^k<1 \tag{2}$$

即可,而且 $\mathrm{E}(X)$ 的取值越小越好.

这样,问题就转化为:在条件(2)的约束下,求 $1+\dfrac{1}{k}-q^k$ 的极小值. 为了处理方便,可设 $f(x)=1+\dfrac{1}{x}-q^x$,$x>0$,借助数学软件,寻找使函数 $f(x)$ 取得近似极小值的正整数 x.

此处,分别选取 $p=0.3,0.2,0.1,0.01,0.001,0.0001$,用 MATLAB 软件编程[3],分别绘制函数 $f(x)$ 的曲线如图 1.

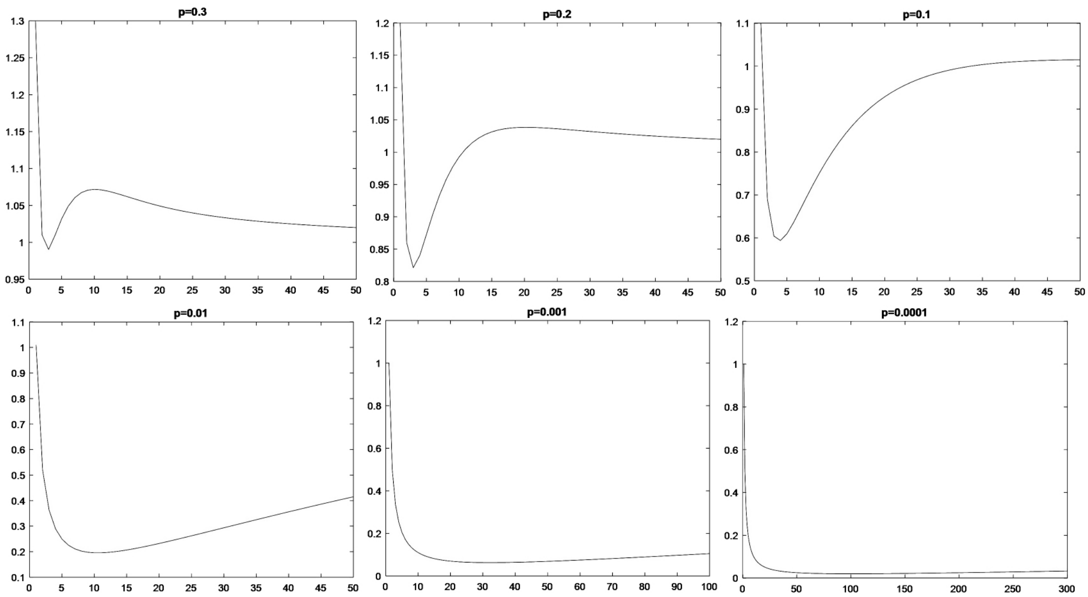

图 1 不同 p 值对应的函数 $f(x)$ 的曲线

同时，对不同的 p 值，计算得最优 k 值及对应的数学期望值，见表 1.

表 1　不同 p 值对应的最优 k 值

p	最优 k 值	对应 $E(X)$	p	最优 k 值	对应 $E(X)$
0.3	3	0.9903	0.01	11	0.1956
0.2	3	0.8213	0.001	32	0.06276
0.1	4	0.5939	0.0001	101	0.01995

综合图 1 和表 1 可以看出，若 p 较大，每组的样本数 k 应选小. 如 $p=0.3$，每组样本数应为 $k=3$，此时每个人需要的平均检测次数为 0.9903，非常接近于 1. 这说明即便进行混样检测也不会减少检测工作量. 当 p 较小，每组的样本数 k 应选大. 如 $p=0.01$，每组样本数 $k=11$，此时每个人需要的平均检测次数为 0.1956. 也就是说：如果每组 11 人（左右），大约可减少 80%的检测工作量，这时混样检测可以很大程度地减少检测工作量.

还需特别说明的是：在实际应用混样检测方法时，要注意以下几点：

（i）一般情况下，在某个确定的疾病筛查区域，患病率 p 是随时间变化的，在做全体筛查之前 p 不是常数，此时，利用统计方法，可以根据已有数据估计 p 的近似值，或者在小范围内通过单样本检测估计出 p 的值.

（ii）疾病筛查检测通常会有多种检测方式与多种检测试剂等[4]. 为了保证混样检测结果准确有效，应尽可能地选择灵敏性高、特异性强的检测方式与检测试剂，避免可能出现的假阴性、假阳性等情况.

（iii）对有些特别检测结果，不但应采取多样本、多试剂的平行检测，还应与临床诊断结果相结合.

3　混样检测在新冠肺炎检测中的应用

将上述混样检测方法应用到武汉市新冠肺炎核酸检测中，需考虑新冠肺炎的患病率、核酸检测方式的灵敏性和特异性等问题.

根据武汉市卫生健康委员会资料显示，截止 2020 年 4 月 30 日武汉市累计确诊新冠肺炎病例 50333 例[1]，占武汉市常驻人口的 0.35%. 4 月 8 日至 4 月 15 日，武汉市对重点人群、复工复产人员等完成了 27.54 万人次的核酸检测，检出新冠肺炎无症状感染者 182 人，占比约 0.066%. 据此可以近似地认为新冠肺炎的患病率 p 约为 0.001.

科研人员经过多次试验、测试发现，带有新冠肺炎病毒的样本中，每毫升通常含有百万（以上）个病毒. 与其他多份正常样本混合之后，新冠肺炎病毒仅仅是浓度被稀释了，病毒依然存在，而且病毒均匀分布在样本保存液中. 通常情况下，只要一毫升样本中有 500 个新冠肺炎病毒就可以用核酸检测出来. 因此，混合样本后虽然病毒浓度被稀释了，但对检测结果影响很小. 科研人员经过进一步试验认为，只要混合的样本数不超过 30 个，核酸检测结果的准确性在概率意义上是可以保证的.

由于新冠病毒传播快、传播途径多、潜伏时间长等特点，为了及时、准确地掌握新冠病毒的感染情况，尽可能避免检测结果假阴性和假阳性等情况，武汉市最终采用了 10 人一组的混样检测，这样，每人平均检测次数约为 $1+\frac{1}{10}-(1-0.001)^{10}=0.11$，可减少 89%的检测工作量，提高核酸检测效率. 如果按照武汉市核酸检测机构每日最高检测人数 6.3 万来计算，这种分组的混样检测方法可以使每天检测人数达 57 万左右，这样，大约 18 天就可以完成约一千万人的核酸检测工作.

根据武汉市卫生健康委员会通报数据表明[5]，从 5 月 14 日 0 时至 6 月 1 日 24 时的 19 天时间内，武汉市对 9899828 人进行了新冠肺炎核酸检测，检出无症状感染者 300 人，没有发现确诊病例. 检测所用的时间和上面预测的 18 天基本吻合.

4 结　论

基于随机变量数学期望的混样检测方法尽管简单、方便、易操作，在一定条件下可以提高检测效率.

但是,在实际应用时,该方法又依赖于检测技术、检测试剂、甚至检测标的物的浓度等其它因素,所以,在使用时需根据实际情况,科学修正检测方法,制定与实际问题相适应的检测方案.

本文目的仅是为了说明随机变量的数学期望,以及统计学中的一些方法在实际问题的应用,以此来引导学生在学习相关课程时,重视理论知识联系实际问题,学会从实际问题中凝练数学问题,养成用数学方法解决实际问题的思维习惯.

致谢 作者非常感谢审稿专家提出的宝贵意见,感谢相关文献对本文的启发.

[参 考 文 献]

[1] 武汉市卫生健康委员会. 疫情通报[EB/OL].(2020-01-31)[2020-08-16],http://wjw.wuhan.gov.cn/ztzl_28/fk/yqtb/index_28.shtml.

[2] 复旦大学. 概率论[M]. 北京:高等教育出版社,1979.

[3] 李继成,赵小艳,李萍. 数学实验[M]. 3版. 北京:高等教育出版社,2020.

[4] 颜新生,杨荟荟,蒿叶霞,等. 4种SARS-CoV-2核酸检测试剂一致性评价[J]. 检验医学,2020,35(7):706-709.

[5] 湖北省人民政府. "新型冠状病毒感染的肺炎疫情防控工作"新闻发布会第104场[EB/OL].(2020-06-02)[2020-08-18],http://www.hubei.gov.cn/hbfb/xwfbh/202006/t20200602_2376181.shtml.

COVID-19 Testing For Nucleic Acid Based on Mathematical Expectation

ZHAO Xiao-yan

(School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: Taking the nucleic acid testing method for COVID-19 testing in Wuhan as an example, this article illustrates the application of mathematical expectations in practical problems, and cultivates students' ability to solve practical problems with the theoretical knowledge and statistical methods learned in the probability and mathematical statistics courses, and induces students to develop the thinking habits describe practical problems by mathematical theories and methods, so as to stimulate students' interest in learning mathematics.

Key words: mathematical expectation; mixed sample testing; COVID-19 testing for nucleic acid