

数学无处不在 – 血液检测的数学模型和理论（一）

胡晓东

中国科学院数学与系统科学研究院

联合国教科文组织在 2019 年 11 月 26 日第四十届大会批准宣布, 3 月 14 日为“国际数学日(International Day of Mathematics, 简称 IDM)”。这是一个由国际数学联盟(International Mathematical Union, 简称 IMU)发起和领导的项目。此项目得到了来自世界各地众多国际和地区组织的支持。第一次庆祝活动于 2020 年 3 月 14 日举行。非常遗憾, 原计划于 2020 年 3 月 13 日(星期五)举行的联合国教科文组织国际官方发布会因新冠肺炎疫情而被取消。



今年国际日的主题是“数学无处不在”。这是因为数学在我们日常生活的几乎每个领域都发挥着至关重要的作用: 从自然模式到气候科学, 从医学成像到搜索引擎, 从运输网络到 AI 的优化, 从建模到流行病的控制。它扩大了“圆周率日(以前在 3 月 14 日庆祝)”的范围, 涵盖了整个数学领域乃至整个世界的各个方面。

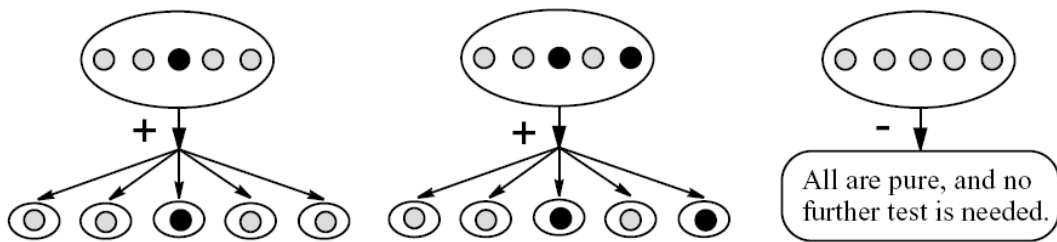
今天是 3 月 14 日, 全国上下都还处在抗击疫情的阶段。这两个多月以来, 大家持续在关注新冠肺炎确诊病例数, 也知道确诊病例时需要进行核酸检测, 而且每一次检测既需要花费时间也需要花很多钱。

下面我就给大家介绍一下与血液检测相关的数学模型和理论: 如何应用数学的方法, 用最少的成本(检验次数和时间), 在大量的血液样本中准确地检验哪些是阴性的(好的), 哪些是阳性的(坏的)。



在第二次世界大战期间，美国政府因战事需要征集了大量的年轻人参军。在征兵过程中需要体检，其中一个环节就是验血，用来筛查出应征参军的报名者中携带淋病病毒的年青人。通常的筛查方法是采集血样以后，对它们一个一个地进行检测，我们不妨称为**逐一检测法**。1943年 R. Dorfman [1]为美国公共健康服务和筛选服务系统设计了一个**组合检测法**：

- 首先，将所要检测的 n 个样本中的每一个样本都分成两部分（类似于现在兴奋剂检测采用的 A 瓶和 B 瓶方法）；
- 然后，将 n 个 A 瓶样本分成 $n/5$ 组，每组 5 个血液样本充分混合，得到了 $n/5$ 个样本；
- 最后，将 $n/5$ 个样本组，每一个组每一个组进行检测。这样检测一个样本组可能出现的结果：
 - （1）某一个样本组显示是阴性的，那么说明该组样本中的每一个样本都是好的。这样我们通过检测 1 次就确认了 5 个样本都是阴性的，从而比逐一检测法少用了 4 次。
 - （2）某一个样本组显示是阳性的，那么说明该组样本中至少有一个样本是不好的，但是不清楚哪一个或者哪几个是阳性的；因而需要对该组中 5 个样本的 B 瓶，进行逐一检测以便确认每个样本是阴性还是阳性。这样我们通过检测 6 次可以确认 5 个样本是阴性还是阳性的，比逐一检测法多用了 1 次。



不难看出，如果在所有样本中阳性的比较少，那么组合检测法会减少大量的检测次数；而这正是当年征兵进行血液检测时的情形（现在每年高考学生进行体检时也类似）。可是如果在所有样本中阳性的比较多，那么组合检测法很有可能不仅不会减少检测的次数，还会增加检测的次数。

由此就自然地产生了一个数学问题：什么时候组合检测法可以减少检测的次数，什么时候不可以呢？1960年 P. Ungar [2]对这个问题进行了研究，称其为**Group Testing Problem**。他假设所需检测的 n 个样本有 pn 个是阳性的，其中 $0 \leq p < 1$ ；每组可以包含任意多的血液样本（不再限定最多 5 个样本），而且每一个样本可以重复使用（一个血液样本可以分装在很多瓶子里）。在这些假设下，他证明了如下定理（大意是：当少于 38.2% 的样本是阳性的时候，组合检测法平均需要的检测次数比逐一检测法需要的检测次数少）：

THEOREM. *The range of p for which there are group test plans such that the expected number of tests is less than n is*

$$(1) \quad 0 \leq p < \frac{1}{2}(3 - \sqrt{5}).$$

R. Dorfman 和 P. Ungar 等人对血液检测问题的研究激发了很多后续的相关研究，很多有意思的数学模型和方法也先后被提出来，并逐渐形成了一个研究方向：组合检测/搜索。下一次我将给大家介绍这个方向中的一个非常有趣的数学猜想。

参考文献

[1] Robert Dorfman, The Detection of Defective Members of Large Populations, *The Annals of Mathematical Statistics*, 14 (4)(1943), 436-440.

[2] Peter Ungar, The Cutoff Point for Group Testing, *Communications on Pure and Applied Mathematics*, XIII (1960), 49-54.