数学聊斋连载

（连载四）

李尚志

千手观音有多少只手
——集合的元素个数

重庆大足石刻是世界文化遗产。大足石刻有一尊千手观音。导游带领游客参观的时候会告诉你：这尊千手观音一共有 1007 只手。然后问：1007 这个准确数字是怎样数出来的？

只要你看了这尊观音，就知道要数清有多少只手并不容易。观音的手如孔雀开屏般向各个方向伸出，长短各异，方向各异，千姿百态，无一雷同。各只手的位置没有规律，不可能按某种方式排列成先后顺序一只一只数清楚。

传说清代时有个工匠被请来对千手观音“贴金”，他为了数清千手观音的手，贴完一只手就朝桶里扔一支竹签。等所有的手臂都贴完了以后，再数一数竹签，一共有 1007 只。由此，千手观音 1007 只手的说法一直流传至今。

【注：据有关媒体报道，2009 年千手观音“大修”，专家用 248 张高精度照片纠正拼接成了千手观音的高清晰影像图，对每只手逐一编号，最后确认千手观音实际

千手观音舞蹈

上是 829 只手。也许专家确认的数字是正确的，但我仍然认为，传说中工匠的方法既简单又不容易出错。如果 1007 只手的数字有误，那也是因为从来也没有认真地按照这个方法操作一次。】

这个简单的办法包含了计数的基本思想：一一对应。将观音的手与金箔一一对应起来，又与竹签一一对应起来，这样，手就与竹签一样多。手不容易数清楚，竹签却容易一支一支数清楚。数出了竹签有多少支，就知道了手有多少只。

一支一支数竹签的时候，又建立了另外一个对应：将竹签依次与正整数 1, 2, 3, … 对应。数到最后一支竹签，发现它对应于正整数 1007, 这就将竹签的集合与从 1 到 1007 的正整数集合 {1, 2, …,1007} 建立了一一对应，竹签个数也是 1007。我们通常用尺子来量长度。类似地，集合元素的多少也是用正整数集合 N={1, 2, …} 这把“尺子”量出来的。用这把“尺子”去量竹签集合，从 1 开始到 1007 正好将竹签量完，也就是说从 N 这把“长尺子”截下“一小段”{1, …,1007} 正好与竹签集合建立一一对应，竹签集合的元素个数就是 1007。

如果用正整数集合 N 这把尺子去量某个集合 S 时，N 的正整数始终用不完，S 就是无穷集合。如果 S 是一个无穷序列 a_1, a_2,v …, a_n , … ，就可以依次将序列的各项对应于正整数 1, 2, …, n, …。这样就将这个序列各项组成的集合与全体正整数的集合 N 建立了一一对应。集合 N 中的元素有无穷多个，凡是与 N 能够建立一一对应的集合称为“可列集合”。可列集合是元素个数最少的无穷集合。

也许你会疑惑：怎么能说 N 是元素最少的无穷集合呢？假如将 N 中的奇数删去，剩下的全体正偶数组成一个集合 $2N$ ={2, 4, 6, …} ，仍然是无穷集合，元素个数是 N 的一半，岂不是比 N 的元素更少了吗？

你可以认为全体正偶数只占全体正整数的一半，但是如果用 N 这把“标准尺子”去“度量”正偶数集合 $2N$，将 2, 4, 6, … 依次与 1, 2, 3, … 对应，全体正偶数的集合 $2N$ 与全体正整数的集合 N 就建立了一一对应，这就说明集合 $2N$ 与 N 所含的元素一样多，正偶数与正整数一样多！明明正偶数只占正整数的一半，居然又说它们一样多，岂不是自相矛盾？

这种看起来矛盾的现象对于有限集合是不会发生的，只有对无穷集合才会发生。为了避免矛盾，我们规定：只要能够有一种方法在两个集合 A,B 之间建立一一对应，就称两个集合的势相等。这里，“势”就是“元素个数”的意思。只要在 A,B 之间曾经建立一一对应，即使还可以将 B 中去掉一些元素之后再将剩下的元素与 A 建立一一对应，也不能说 A 的元素比 B 少，而仍然坚持认为 A,B 的元素一样多。

我们已经看到无穷集合 N 的“一半”与 N 的元素一样多。反过来，容易想到，N 的“两倍”的元素也与 N 一样多。容易看出，全体正整数组成的集合 $N=\{1, 2, \cdots, n, \cdots\}$ 与全体负整数组成的集合 $-N=\{-1, -2, \cdots, -n, \cdots\}$ 可以建立一一对应，元素个数相等。因此，将 N 与 $-N$ 合并在一起得到的集合就是 N 的“两倍”，再添上 0 得到的就是全体整数组成的集合 Z，它的元素个数可以认为是 N 的两倍再加 1。然而，全体整数可以按如下方式排成一个无穷数列 $Z=\{0, 1, -1, 2, -2, \cdots, n, -n, \cdots\}$，数列中的各个数可以依次与正整数 1, 2, 3, ⋯对应，这说明了整数与正整数一样多！

每个实数可以用数轴上的一个点来表示。这样就建立了数轴与全体实数的一一对应。表示整数的点在数轴上稀稀拉拉，表示有理数的点在实数轴上密密麻麻，可见有理数比整数多得多。将实数轴分成很多长度为 1 的区间 $(n, n+1)$，则每个区间内只有一个表示整数的点，但它却有无穷多个有理数的点。可见有理数的个数是整数个数的无穷多倍。然而，全体有理数可以按如下方式排列成一个无穷数列：

$$\frac{0}{1}, \frac{1}{1}, -\frac{1}{1}, \frac{1}{2}, -\frac{1}{2}, \frac{2}{1}, -\frac{2}{1}, \cdots$$

排列的方法是：先将 0, 1, -1 排在前三项，再按如下原则排列其余各有理数：将 0, 1, -1 之外的每个非零有理数经过约分写成 q/p 或 $-q/p$ 或 p/q 或 $-p/q$ 的形式，其中 p, q 是互素的正整数并且 $p>q$。按 p 从小到大的顺序排列；p 相等的，按 q 从小到大的顺序排列；p 与 q 都相同的 4 个数，按 $q/p, -q/p, p/q, -p/q$ 顺序排列。

按以上方法，就将所有的有理数排成了一个无穷数列。将无穷数列的各项依次与正整数 1, 2, 3, ⋯ 对应，第 n 项与正整数 n 对应，这就建立了这个无穷数列与正整数集合 $N+$ 的一一对应，证明了有理数与正整数一样多！

能否将全体实数也排成数列 $a_1, a_2, \cdots, a_n, \cdots$，从而证明实数也与正整数一样多呢？

如果想出了一个办法将所有的实数排成无穷数列，就证明了实数与正整数一样多。反过来，即使全世界所有的人都没有找到这样的办法，也还不能就此断定这样的办法不存在。万一来了一个外星人想出这样的办法呢？

不过我们可以用反证法证明：无论何时，无论何地，谁也不能将全体实数排成一个数列！

数轴上的点与实数一一对应。假如全体实数被排成了一个数列 $a_1, a_2, \cdots, a_n, \cdots$，数轴上全体点也就对应地排成一个序列 $A_1, A_2, \cdots, A_n, \cdots$。预先任意取定一个小的正数 e。在数轴上取一个长度为 $e/2$ 的线段 E_1 将点 A_1 盖住，再取一个长度为 $e/4$ 的线段 E_2 将点 A_2 盖住。一般地，对序列中第 n 个点 A_n，取一个长度为 $e/(2^n)$ 的线段 E_n 将它盖住。所有这些线段 $E_1, E_2, \cdots, E_n, \cdots$ 合并起来得到的集合 E 将所有的点 $A_1, A_2, \cdots, A_n, \cdots$ 全都盖住了。我们知道，这些点就是数轴上所有的点，因此 E 将数轴全部覆盖了。但另一方面，E 由线段 $E_1, E_2, \cdots, E_n, \cdots$ 合并而成，E 的总长度不能超过这些线段的长度之和

$$\frac{e}{2}+\frac{e}{4}+\frac{e}{8}+\cdots\frac{e}{2^n}+\cdots<e$$

其中 e 可以取任意小的正数。E 的总长度小于任意正数，只能是 0，不可能覆盖整个数轴！

这个矛盾就证明了全体实数不可能排成一个数列，实数与正整数不一样多。

事实上，我们证明的结论是：如果数轴上某些点组成的集合能够排成一个序列，那么这些点的“总长度”是 0。在前面已经证明过数轴上的有理点（表示有理数的点）可以排成无穷数列，可见有理点的“总长度”是 0。除去有理点，剩下就是无理点（表示无理数的点）。数轴的总长度是无穷大，其中有理点的总长度是 0，剩下的无理点的总长度就应当是无穷大。如果要问在数轴上有理点和无理点各占百分之几，那就只能说有理点占百分之零，无理点占百分之百！

厦门和美丽的鼓浪屿

九成厦门人感到幸福遭九成以上质疑？

——统计

网上看见一篇文章《当九成厦门人感到幸福遭九成以上质疑》。现摘引一部分内容供大家欣赏：

国家统计局厦门调查队的调查报告称九成以上厦门人感到幸福，此报告刚一发布，就引来众多网友和市民的争议，其中，受争议最大的是“九成以上”这个数据的真实性。有多少人质疑这个调查结论呢？居然也是“九成以上的厦门人”。（2008 年 6 月 26 日浙江在线）

但遗憾的是，有人对于“九成以上厦门人感到幸福”的调查结论却提出质疑，最奇怪的是，质疑的人居然也在九成之多。在厦门某社区网站上，截至 25 日 19 时，关于该话题的帖子点击率为 690 次，其中 63 人对此回复，仅 9 个人表示相比其他人觉得幸福。而记者随机调查的十几位市民中，肯定回答“幸福”的人仅一两个。

既然“九成以上厦门人感到幸福”的调查结论遭遇九成以上厦门人的质疑，我们就可以断言，这是一个虚假的调查结论。如果这个数据确实是经过调查的，也是一个靠不住的“伪调查”。“伪调查”是怎样进行的？据国家统计局厦门调查队城镇住户处处长林芙贞介绍，这份调查是在厦门某个社区做的入户调查，“我们是随机调查统计的”。该局的工作人员笑称，之所以有那么多人说自己不幸福，“只能说他们没有在我们的抽样调查之内”。是的，你们怎么都抽到了说好话的调查户，而大量提出质疑的市民就没有被抽样呢？如果你们抽样到一个民营企业家、一个国企老总、一个厅级干部，就能算出厦门市民平均年收入是 100 万吧？

这篇文章有很多跟贴，其中一份跟贴如下：

我也是调查队的，我知道这样的调查很难很难搞准，因为被问卷调查的人当面都说好、好、好，背后谁都说不好。

现在的人真的都是两面人，问到的人没有人说实话，一方面大骂统计数据不准，另一方面调查到自己时却编

晨运的人们

瞎话，从不说真实情况。

国家为了搞准调查数据真是煞费苦心，投入巨大人力、物力在全国开展住户调查，…

这篇文章和跟贴提出的问题很精彩，很有代表性，可以作为今后关于如何调查统计的所有的课本（包括大学和中学的课本）的经典案例。

首先，我们来看厦门调查队的工作人员的说法。

说法 1：该局的工作人员笑称，之所以有那么多人说自己不幸福，“只能说他们没有在我们的抽样调查之内”。调查队当然不可能调查全部人员。但一定要有代表性。在设计调查对象的时候，就应当尽可能让被调查的人的观点能够代表没有被调查的人的观点。不能只调查说幸福的人，也不能只调查说不幸福的人。如果真的有那么多说自己不幸福的人在抽样调查之外，只能说明调查组的抽样并没有代表性。工作人员居然能够如此理直气壮地说这句话，只能说明这个工作人员根本不懂调查的基本常识。

说法 2：我也是调查队的，我知道这样的调查很难很难搞准，因为被问卷调查的人当面都说好、好、好，背后谁都说不好。问到的人没有人说实话，一方面大骂统计数据不准，另一方面调查到自己时却编瞎话，从不说真实情况。如果被调查的人真的有幸福感，为什么不愿意说实话，为什么要编瞎话？如果这个调查队员的话是真的，如果被问卷调查的人“没有人说实话”，“背后谁都说不好”，那就不用调查了，就凭这个调查队员的结论就足以说明厦门 100% 的人都没有幸福感了。岂止是没有幸福感，这样的厦门简直就是一个恐怖世界！这位调查队员以为他在捍卫“九成以上厦门人感到幸福”这个结论，实际上是在否定这个结论。其实，如果被问卷调查的人不肯说实话，并不能怪被调查的人，只能怪这个调查本身有问题。任何调查都应适当地设计问卷和调查的方式，使得被调查的人说实话没有顾虑。比如不采用当面询问的方式，而采用类似于无记名投票的方式将自己的答案折好投入“票箱”。如果主持调查的人员连这一点常识都没有，那只能说明他们没有资格承担这项调查任务，国家向他们投入大量经费是投错了。

再来看这篇文章对“九成以上的人质疑”提供的依据：

依据：关于该话题的帖子点击率为 690 次，其中 63 人对此回复，仅 9 个人表示相比其他人觉得幸福。

幸福家庭

63 人回复，仅 9 人感到幸福，其余 54 人不幸福，占 63 人的 54/63 = 85.7%，四舍五入可以算作九成，将它说成“九成”还可以。当然，小学生都知道 85.7% 不是“九成以上”，但考虑到媒体写文章习惯于夸张以吸引眼球，我们也不必苛求他们。不过，我们想知道，厦门有多少万人口？这 63 人能否代表他们？而且，在点击“该话题的帖子”的 690 人中，除了这 63 人，其余没有回复的人是什么态度，他们是感到幸福还是不幸福？假如“该话题的帖子”就是那份宣布“九成以上厦门人感到幸福”的调查报告，我估计凡是感到幸福的人一般不会回复，而强烈反对这个结论的人回复的可能性就比较大。因此，这 63 人更多地代表了反对意见，而不能代表这 690 人。更何况，这 690 个上网的人的态度恐怕也不能代表厦门成千上万不上网的人。写这篇文章的人批评调查组的抽样没有代表没有被抽样的人，这样的批评原则上是对的。可是，他仅凭 63 人中的 54 份没有表示幸福的回复、或者仅凭采访的十几个人的态度，就斩钉截铁地说“我们可以断言”，岂不是犯了与调查组同样的错误甚至更严重的错误吗？

我不是厦门人，但去过厦门几次，平心而论我对厦门的印象是很好的，我所接触的厦门人多数是有幸福感的，因此我也相信这篇文章所说的“九成以上的人反对”是错误的。但我没有进行科学的调查，因此没有充分的理由断定我的这个感觉是否正确，没有充分的理由判定调查组的结论或者文章的结论是否正确。但可以斩钉截铁地断言的是：以上所提到的两位调查人员，以及文章的作者，他们对调查的观点和方法都是完全错误的。这种错误，来源于他们调查知识的欠缺，也来源于他们的偏见。

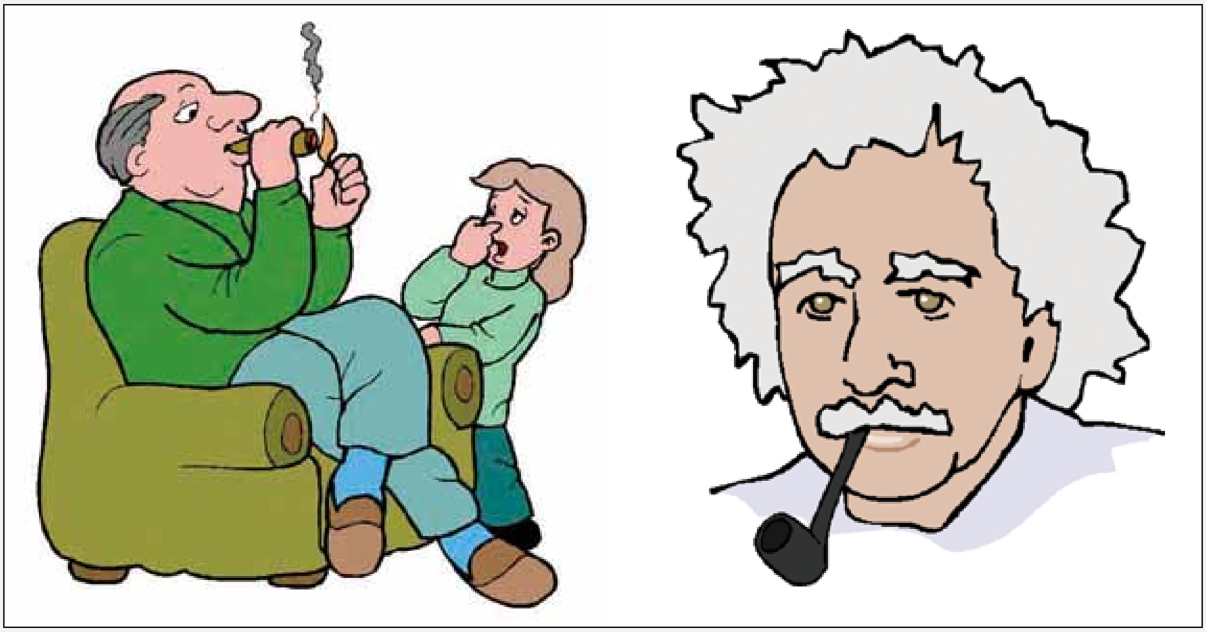
调查组为了显示政绩，希望说幸福的人越多越好。媒体写文章时为了吸引眼球，希望说幸福的人越少越好。这种希望也无可指责。但是，只要进行调查，一条基本的原则就是：不论自己喜欢与否，都要努力保证调查结果的客观真实性。在任何一本有关调查统计的书籍里都有许多相关的知识教你怎样实现这一点。我建议以后写书时将这件事情作为典型案例写进去，使该书更加通俗易懂。不要说让全国人民都懂一点统计常识，至少应当让主持统计工作的负责人懂一点统计常识吧？

被动吸烟的人高达几亿？
——数学模型的检验

中央电视台的新闻节目中曾经宣布了一个统计数据，说是我国被动吸烟的人口高达多少多少亿。为了核实具体数据，我登陆百度，输入关键词“被动吸烟”，查到这样一篇文章：

“我国吸烟人数为 3.5 亿。根据研究推算，目前我国人群中遭受被动吸烟危害的人数可高达 5.4 亿，其中 15 岁以下儿童有 1.8 亿。”

我不知道是经过怎样的“研究推算”得到这样的结果的。我国人口总数超过 13 亿，除去主动吸烟人数 3.5 亿和被动吸烟人数 5.4 亿，至少还应当有 4 亿幸运者既没有主动吸烟也没有被动吸烟。这 4 亿幸运者占全国总人口的 30% 以上，这个比例应当还不算小，至少不像大熊猫那样稀少，因此我们应当有机会经常见到这样的幸运者。不过，我想问：要达到怎样的条件才能够成为



这样的一位幸运者？当然，不主动吸烟容易做到，比如我就做到了。但我经常吸入别人抽烟产生的烟雾，应当属于那 5.4 亿被动吸烟者的行列。然而怎样才算是没有被动吸烟？是从来没有遇见别人在他们面前抽烟吗？或者，别人抽烟的烟雾从来没有被吸入他的鼻孔。不过我很怀疑，这样的幸运者不但达不到 30%，而且比大熊猫更稀少，甚至比华南虎更稀少，有可能根本不存在。我们周围哪一位人士敢于说他从来没有吸入过别人抽烟产生的烟雾？也许在某一个与世隔绝的世外桃园里生活的人群全都是这样的幸运者，不过，这样的世外桃园可能还不为我们所知，因此那里的人们还没有进入我们能够统计到的 13 亿人口之列，可以不予考虑。坦白地说，我猜想：遭受被动吸烟的人数不是“高达 5.4 亿”，而是更高达 13 亿 − 3.5 亿 = 9.5 亿。我强烈希望我的这个猜想是错的，希望进行“研究推算”的专家能够拿出证据推翻我的这个猜想，也就是说找出一些这样的幸运者来。不要求他们找出 4 亿幸运者，先找出哪怕一个幸运者让我们高兴一下吧！

如果觉得“从来没有吸入烟雾”这个标准太高，那么可以放宽标准，改为平均每天吸入的烟雾量在多少毫克以下。不知道那些进行“研究推算”的专家们研究过这样的标准没有？也许按照他们的新标准，我也是那 4 亿没有被动吸烟的幸运者之一呢。

我猜想，专家的推算方法大约是：先估计每个主动吸烟者造成几个被动吸烟者，比如假设每个主动吸烟者造成 1.54 个被动吸烟者。主动吸烟的人数比较明确，可以大致统计出来，比如确实是 3.5 亿。再通过简单的乘法算式 $1.54\times3.5=5.39$ 就研究推算出了被动吸烟的人数高达 5.4 亿。我们不好说他们的算法不对。也许专家们不仅念过小学，还念过中学和大学，根据我们都不懂的高级理论采用了我们都不懂的更复杂的算法。但是，不论他们的理论和算法是怎样的高不可攀，所得的结论必须符合普通草民们能够理解的常识，经受实际的检验。也就是说：数学模型得出的结论必须经过实际效果的检验。具体地说，既然宣布有 4 亿人口没有被动吸烟，至少应当很容易从这 4 亿人口找到其中一部分人向我们宣讲一下他们能够避免成为被动吸烟人的宝贵经验供我们学习和效仿吧？