第一章　事件与概率

第1节　随机现象与统计规律性

李 娟

Email: juanlisdu@163.com

山东大学(威海)　数学与统计学院

授课要求：1. 理解随机现象；

2. 理解频率与概率的内在联系。

一、随机现象：在一定条件下，有各种可能发生的结果，而事先不能确定哪一种结果一定发生的现象。

为了说明什么是随机现象，书上讲了一个例子：**飞机票超售**。

一、随机现象

航空公司电脑订座系统的普遍采用给旅客和公司都带来极大的方便，但是也对管理工作提出了更高的要求。例如一架200座的飞机到底应出售多少座位？

简单而常用的方法是出售200座。不过，这并不是一个很好的答案，因为常有订了座位的旅客临时不来上机，出现空位，造成浪费。于是就实行超售，即在飞机起飞前出售的座位超过实有的座位。

一、随机现象

据统计，国内航班中订座而到时不来上机的旅客超过5%，因此若照实有座位数售座，则不可避免会出现大量空座。这些空座位的浪费，不仅使有些想搭乘此航班的客人失去了乘机的机会，而且也给航空公司造成经济损失，最后也被航空公司用提价的方式转嫁给旅客。

因此，超售是正确的选择。但是超售会造成拒登机，即有些持票者上不了机。虽然航空公司可以通过给自愿推迟者某种补偿(譬如提供一张免票或者免费安排食宿等)来化解矛盾，但还是会带来种种负面影响，使公司蒙受损失。

一、随机现象

从理论上讲，超售越多，空位损失越小，但拒登机的可能性越大；反之，超售越少，拒登机的可能性越小，但空位损失会越大，因此这是一个优化问题。

航空公司要确定准确的超售数额，这就要求确定该航班订座旅客不来上机的人数，但是这个量在登机前是无法准确确定的。订座的旅客为什么不来上机呢？原因各别，但大体上都是受一些偶然因素的影响，例如计划变动，行程更改，交通延误以及改乘其他航班等等。因此这里我们要处理的是一个受许多偶然因素影响的量，这正是概率论研究的对象。

超售问题是很典型的概率论问题，用概率论方法可以给这个问题以相当完满的解决。这里略述思路：假定每个订座旅客准时上机的可能性为95%，则采用适当的概率模型可以算出在不同的出售额N下，发生拒登机的可能性P 列于下表：

N	201	202	203	204	205	206	207
P	0.000	0.002	0.007	0.015	0.032	0.062	0.109

航空公司可以根据这些数据制定自己的超售和补偿方案。实践证明，超售带来巨大的经济效益，而且以超售为起点，当代航空业已发展出一套很先进的管理方法—收益管理。

一、随机现象

类似的例子在许多实际问题中出现，解决这类问题当然具有重要意义。它们都牵涉一类现象—随机现象，要求处理一类变量—随机变量，它的数值受许多偶然因素的影响，事先无法确知。

在自然界和人类社会中都存在着两类不同的现象。当我们多次观察自然现象和社会现象后，会发现许多事情在一定的条件下必然会发生。例如：在生活中，水加热到$100°C$必然沸腾......

一、随机现象

1. 必然事件：在一定条件下，必然会发生的事件成为必然事件。

2. 不可能事件：在一定条件下，必然不会发生的事件称为不可能事件。

注：从所举例子中看出，必然事件和不可能事件，虽然形式相反，但是两者的实质是相同的。必然事件的反面就是不可能事件，而不可能事件的反面就是必然事件。所有这种现象称之为**决定性现象**，或**确定性现象**，$i.e.$，在一定条件下，有唯一结果的现象。

但是在自然现象和社会现象中还广泛存在着与决定性现象有着本质区别的另一类现象，$i.e.$，**随机现象**。例如：上述机票超售问题。

类似的例子还可以举出很多，例如用同一仪器多次测量同一物体的重量，所得结果彼此总是略有差异，这是由于诸如测量仪器受大气影响，观察者生理上或心理上的变化等等偶然因素引起的。同样的，同一门大炮向同一目标发射多发同种炮弹，弹落点也不一样，因为炮弹制造时种种偶然因素对炮弹质量有影响，此外，炮筒位置的误差，天气条件的微小变化等等都影响弹落点。再如从某生产线上用同一种工艺生产出来的灯泡的寿命也有差异等等。

一、随机现象

　　总之，所举这些现象的一个共同的特点是：在基本条件不变的情况下，一系列试验或观察会得到不同的结果，换句话说，就个别的试验或观察而言，它会时而出现这种结果，时而出现那种结果，呈现出一种偶然性，这种现象称为**随机现象**(random phenonmenon)
对于随机现象通常关心的是在试验或观察中某个结果是否出现，这些结果称为**随机事件**，简称**事件**(event)。以后我们一般都用A, B, C等大写拉丁字母表示随机事件。

3. 随机事件： 在一定条件下可能发生也可能不发生的事件称为随机事件。一般用$A, B, C, \cdots$ 表示。

二、频率稳定性

人们经过长期的实践发现，虽然个别随机事件在某次试验或观察中可以出现也可以不出现，但在大量试验中它却呈现出明显的规律性—频率稳定性。

1. 频率：对于随机事件A，若在N次试验中出现了n 次，则称

$F_N(A) = \frac{n}{N}$为随机事件A在N次试验中出现的频率。

二、频率稳定性

下面是关于频率稳定性的几个有名例子，援引这类例子是因为它们不但具有一定的权威性，而且都是可以反复验证的。

在掷一枚硬币时，既可能出现正面，也可能出现反面，预先作出确定的判断是不可能的，但是假如硬币均匀，直观上出现正面与出现反面的机会应该相等，即在大量试验中出现正面的频率，应接近于50%，为了验证这点，历史上曾有不少人做过这个试验，其结果如下表所示：

实验者	掷硬币次数	出现正面次数	频率
蒲　丰	4040	2048	0.5069
皮尔逊	12000	6019	0.5016
皮尔逊	24000	12012	0.5005

二、频率稳定性

又如，在英语中某些字母出现的频率远远高于另外一些字母。在进行了更深入的研究之后，人们还发现各个字母被使用的频率相当稳定。例如：下面就是英文字母使用频率的一份统计表。其他各种文字也都有着类似的规律。

字母	空格	E	T	O	A	N	I	R	S
频率	0.2	0.105	0.072	0.065	0.063	0.059	0.055	0.054	0.052
字母	H	D	L	C	F	U	M	P	Y
频率	0.047	0.035	0.029	0.023	0.0225	0.0225	0.021	0.0175	0.012
字母	W	G	B	V	K	X	J	Q	Z
频率	0.012	0.011	0.0105	0.008	0.003	0.002	0.001	0.001	0.001

二、频率稳定性

近年来对汉语的统计研究有了很大的发展。关于汉字的使用频率已有初步统计资料，对汉语常用词也作了一些统计研究。特别是结合汉字输入方案等的研制，正在对汉字的结构作深入的统计分析。这些研究对实现汉字信息处理自动化无疑具有重要的意义。

另一个验证频率稳定性的著名试验是由英国生物统计学家高尔顿(Galton)设计的。

自上端放入一小球，任其自由下落，在下落过程中当小球碰到钉子时，从左边落下与从右边落下的机会相等。碰到下一排钉子时又是如此。最后落入底板中的某一格子。因此，任意放入一球，则此球落入哪一个格子，预先难以确定。但是实验证明，如放入大量小球，则其最后所呈现的曲线，几乎总是一样的。也就是说，小球落入各个格子的频率十分稳定。这个试验模型称为**高尔顿板**。试验中呈现出来的规律性，在学习第五章极限定理之后，就会有更深刻的理解。

二、频率稳定性

另一个频率稳定性的有名例子是：在人类的生育中，男婴的出生率约为$\dfrac{22}{43}\left(>\dfrac{1}{2}\right)$.

同样，如果多次测量同一物体，其结果虽略有差异，但当测量次数增加时，就会越来越清楚地呈现出一些规律性：测量值的平均值在某固定常数附近波动，诸测量值在此常数两旁的分布呈现某种对称性。又如在射击的例子中，当射击次数不多时，炮弹的弹落点似乎是前后左右杂乱无章，看不出什么明显的规律；但当射击次数增加时，弹落点的分布就会呈现出一定的规律性：如弹落点关于目标的分布呈对称性。偏离目标远的弹落点比偏离目标近的弹落点少等等。其他如灯泡寿命等，在进行多次观察或试验后，也都可以发现类似的规律性。

二、频率稳定性

上述种种事实表明，随机现象有其偶然性的一面，也有其必然性的一面。这种必然性表现为大量试验中随机事件出现的频率的稳定性，即一个随机事件出现的频率常在某个固定的常数附近摆动，这种规律性我们称之为**统计规律性**。频率的稳定性说明随机事件发生的可能性大小是随机事件本身固有的，不随人们意志而改变的一种客观属性，因此可以对它进行度量。

二、频率稳定性

对于一个随机事件A，用一个数$P(A)$来表示该事件发生的可能性大小，这个数$P(A)$就称为随机事件A的概率(probability)。因此概率度量了随机事件发生的可能性的大小。

对于随机现象，只讨论它可能出现什么结果，价值不大，而指出各种结果出现的可能性的大小则具有很大意义。有了概率的概念就使我们能对随机现象进行定量研究，由此建立了一个新的数学分支—概率论。

2. 概率：对于一个随机事件A，用一个数$P(A)$来表示该事件发生的可能性大小，这个数$P(A)$就称为随机事件A的概率。

(一)概率的直观含义:

定义1

随机事件A发生的可能性大小的度量(数值)称为A的概率,记为$P(A)$。并且规定$0 \le P(A) \le 1$。

注1: 概率由事件本身确定,是客观存在的。

注2: 概率是一种测度。

在概率论的发展历史上,人们曾针对不同的问题,从不从的角度给出了定义概率和计算概率的各种方法。然而所定义的概率都存在一定的缺陷,所以不如说它是一些计算概率的方法。例如:古典概率、几何概率,等等。

1. 定义：在N次重复试验中，事件A发生的次数n_A 称为事件A发生的频数。比值$\dfrac{n_A}{N}$称为事件A 发生的频率，并记成$F_N(A), i.e.,$

$$F_N(A) = \frac{n_A}{N} = \frac{\text{事件A发生的次数}}{\text{试验的总次数}}。$$

用$F_N(A)$作为事件A的概率的一个度量，这样计算的概率称为**统计概率**。我们可以验证，当试验次数N固定时，事件A发生的频率$F_N(A)$有如下性质：

2、频率的性质

(1) 非负性：$0 \le F_N(A) \le 1.$ (1.1.1)

(2) 规范性：对于必然发生的事件，在N次试验中应出现N 次。若以Ω 记必然事件，则应有：$F_N(\Omega) = 1.$ (1.1.2)

(3) 有限可加性：若A和B是两个不会同时发生的随机事件，以$A + B$表示A或B至少出现其一这个事件，则应有：

$$F_N(A + B) = F_N(A) + F_N(B). \tag{1.1.3}$$

推广至：若有限个事件$A_1, \cdots, A_k$两两不会同时发生，则

$$F_N(A_1 + \cdots + A_k) = \sum_{i=1}^{k} F_N(A_i)$$

2、频率的性质

统计概率同样满足上述3个性质，但它有许多缺点。因为我们没有理由认为，取试验次数为$N+1$来计算频率，总会比取试验次数为N来计算频率将会更准确、更逼近所求的概率。在实际应用上，我们也不知道N取多大才行。

3、概率与频率的联系

最后，根据上述频率稳定性的讨论似乎可以提出这样的猜想，即当N足够大时，$F_N(A)$与$P(A)$应充分接近。这一想法有很大的启发性，在历史上它一直是概率论研究的一个重大课题。以后我们将会看到，在很一般的条件下，这个结论的确成立，但同时还需对问题的提法进一步明确化。

3、概率与频率的联系

(1) 当试验次数n很大时，可以用频率近似地代替概率。

(2) 概率是频率的稳定中心。

(3) 设$P(A)=p$，则$\lim_{N\to\infty}\frac{n_A}{N}=\lim_{N\to\infty}F_N(A)=p$ 是错误的，而应是：$\forall\varepsilon>0,$

$$\lim_{N\to\infty}P\left\{\left|\frac{n_A}{N}-p\right|\geq\varepsilon\right\}=0。$$

大数定律：序列$\frac{n_A}{N}$依概率收敛于p，从直观上看，如果对于无论多么小的$\varepsilon > 0$，当N大到一定程度后，$\frac{n_A}{N}$与$P(A)$ 的距离小于ε的概率几乎是1，则"$\frac{n_A}{N}$靠近 p"这个随机事件就几乎是必然事件了。

四、概率论简史（略）。