概率统计

习题輔导

《湖北电大学刊》编辑部

前 言

这份辅导材料着重于对一些概念、方法的解释，分析具体的问题，从而得出解题的方法。我们以典型的例题详细分析为主，而不着眼于收罗种种题目，给出提示和解答，希望对学生的解题能力真有帮助。

我们分四讲介绍有关的内容：

Ⅰ．随机变量（用随机变量表示随机现象，计算相应的概率）

Ⅱ．贝叶斯公式（反概率公式）及其应用

Ⅲ．最小二乘法（迴归分析）

Ⅳ．可靠性试验数据的统计分析

各讲均分小段，用1．2．……等分别表示各段的标号，在引用时，Ⅰ．2表示第Ⅰ讲中第2小段的内容；公式以(Ⅰ．3)（Ⅱ．14）等记号表示，即为第Ⅰ讲第3个公式，第Ⅱ讲第14个公式。其余的一些符号与教本相同，就不另作说明了。

希望读者在学习课本到一个阶段后，在复习教材，做题之后，再看本材料的有关内容，可以加深理解并掌握有关的方法。千万不要把本材料当作习题解答，做题有困难时，翻一下就可解决，这样是没有帮助的。

本材料的目的是帮助大家将所学的知识去解决实际问题。通常的毛病是学员往往只注意一些公式，以为记住了公式、会套用公式就可以解决问题了，这是非常片面的。高等

数学与中学的数学有一个比较明显的差别，就在于概念的理解。一些公式总是反映某些概念之间的内在联系的，如果概念不清楚，就无法理解公式所反映的真实内容，使用时往往容易用错，导致错误的结论。针对这一弱点，本材料着重于对公式中一些概念的解释，使读者真正能理解公式的内容，使用时就不会有较多的困难了。并且对于一些解题所得的结论，也往往进一步作一些说明，使读者能理解一些概率、统计的思想方法。

这四讲由于内容不同，我们的写法也是各有侧重的。原则是书上有的，讲得较清楚的，这里不再重复，书上没有的，说得较简单的，但又是学员容易发生问题的，这里再补充说详细一些，因此我们认为，这一材料对辅导教师可能也会有些帮助。

由于初次编写这类内容，是否能达到预期的目的，还是要看效果，欢迎读者广泛提出意见，帮助我们改进。

编 者　1985年8月

目　录

Ⅰ. 随机变量

概率论是研究随机现象的，要把随机现象用数学的语言来描述，才能用数学的工具去处理它。通常的概率论教材是由两步来完成的：先引入随机事件，讨论随机事件的运算与相应的概率关系；然后再引入随机变量，随机变量取值在某一个指定集合，它就是一个随机事件。这样一个过程产生了两个问题：开始随机事件的运算与随机现象的描述、概率的计算这几个难点都集中在一起，此时概率的计算不能用微积分，而大都用排列组合；在引入随机变量时，又必须与随机事件联上，数集合的运算与事件的运算联上，这也会造成一些困难。

本文的重点是希望通过随机变量来直接描述随机现象，把随机现象由随机变量直接过渡到数集合，然后利用微积分这个工具，求和号这个工具来计算相应的概率。

随机现象大致上可分两类：一是比较容易用一些量来描述的，例如测量误差的大小，电子管使用的时间，一平方米中昆虫的数目……等等；另一类是似乎与数量无关的，一台机床在八小时内连续工作是否会发生故障，这一次考试是否会全不及格，某人打靶一次能否打中……等等。前一类从随机变量来看，反而比较容易处理，而后一类看起来是要麻烦些，其实也是不困难的，下面我们就分别进行讨论。当然，这两类也不是能截然分开的，往往一个问题中含有这，也含有

那，这就造成了困难。

1．**指标的分类**。任何一种随机现象，从数学来看，都可以用量来反映它，这个量就是这一问题所涉及的指标。例如测量山的高度，山高就是指标；又如打靶时弹着点的环数，环数就是指标；又如一毫升水中的细菌的个数，个数就是指标；例如一个盒中有红、白、黑三种球，摸到一个球是白球这一现象如何用量来描述呢？这一问题的指标是什么呢？如果我们将这三种颜色编号，1表示红，2表示白，3表示黑，那么摸到白球就是出现编号为2的球，于是编号就是这个问题中的指标。显然最后一例中，1．2．3．仅仅是一个代号，它们的大小顺序没有任何意义，也可以用-1表示红，0表示白，11表示黑，它不会改变问题的实质。

这样我们可以把指标分成三类：

a）连续型指标。如山的高度，它可以取实数轴上某一区间内的任何一个值。其他如灯泡的点燃寿命，某一质点运动的速度，矿石中某种化合物的含量等等都是连续型的。这一类指标就与连续型随机变量直接相联系，很自然地可以用连续型随机变量来处理。

b）计数型指标。如一毫升水中细菌的个数，它只能取整数值，不可能取到某一实数区间$[a,b]$内的任何值；又如打了n次靶，打中的次数k只能是从0到n这$n+1$个整数值，这也是计数型的。从数学上说，只要是取离散的值，它总是与离散型随机变量有关，在实际问题中，离散型随机变量所反映的，大都是计数型指标与下一种名义指标。

c）名义指标。如最后的例中用1．2．3．分别代表

红、白、黑，这种代码性的指标称为名义指标，它的数量大小，顺序等均无实际意义。名义指标原则上可以随便规定（但规定后在问题中不能再作改变），实际上也还有一个如何规定使得处理起来更为方便的考虑，这在下面的一些例题中就可以看到. 名义指标也与离散型随机变量有关.

2. 离散型随机变量的例题.

由于连续型随机变量与连续型指标有关，处理时反而很容易，往往就用连续型指标作为反映随机现象的随机变量，因此只要设法弄清楚它的分布密度，或利用一个基本公式，求概率的问题就解决了. 对于离散型随机变量，情况就有些不同，我们通过一些简单的例子，逐步过渡到复杂的例子.

例1. 一个人打靶，一发子弹打中的概率是p，打不中的概率是$1-p$.

这一现象应如何用随机变量来描述呢？你可以规定一个随机变量ξ如下：

$$(\text{I}.\ 1)\quad \begin{cases} \xi=5 & \text{子弹中靶} \\ \xi=7 & \text{子弹不中靶} \end{cases}$$

5与7仅仅是代号，也可以改成别的例如11与105，但是谁也不会这样去规定，因为这会造成不必要的麻烦。通常是规定随机变量ξ如下：

$$(\text{I}.\ 2)\quad \begin{cases} \xi=1 & \text{子弹中靶} \\ \xi=0 & \text{子弹未中靶} \end{cases}$$

这时ξ的值还是有意义的，它就是打中的次数（一发子弹只可能打中一次或零次，零次就是未打中）。如果一个人打靶打了n次，而每一次打中不打中用（I. 1）的ξ来描述，第

i 次记为 ξ_i， $i=1,2,\cdots n$， 这样 n 次的结果就无法用 $\xi_1+\cdots+\xi_n$ 来直接描述，但是用（Ⅰ.2）来描述时，ξ_i 就表示第 i 次打中的次数，$\xi_1+\cdots+\xi_n$ 就表示 n 次中一共打中了多少次，$\sum\limits_{i=1}^{n}\xi_i$ 有明确的实际意义。

（Ⅰ.2）对应的随机变量 ξ 只取 0、1 两个值，ξ 的分布就是两点分布：$P(\xi=1)=p$，$P(\xi=0)=1-p$。打了 n 次靶，每次的条件一样，于是用 ξ_i 表示第 i 次的结果，按（Ⅰ.2）的定义，$\xi_*=\sum\limits_{i=1}^{n}\xi_i$ 就是 n 次打靶中打中的次数。当 ξ_1，…，ξ_n 相互独立时，ξ_* 就是二项分布：

$$P(\xi_*=k)=\binom{n}{k}p^k(1-p)^{n-k},\quad k=0,1,\cdots,n.$$

这就弄清了两点分布与二项分布的关系。如果 $\xi_1\cdots$，ξ_n 之间是不独立的，但分布仍然是相同的两点分布，则 $\xi_*=\sum\limits_{i=1}^{n}\xi_i$ 就不再是二项分布了，这可用超几何分来说明。设有一批产品共 N 件，次品率为 p，从 N 件中抽了 n 件进行检验，n 件中恰有 k 件是次品的概率是

$$\binom{Np}{k}\binom{N(1-p)}{n-k}\Big/\binom{N}{n}$$

我们令

$$(\text{Ⅰ}.3)\qquad \xi_i=\begin{cases}1 & \text{第 } i \text{ 件抽到的是次品}\\ 0 & \text{第 } i \text{ 件抽到的不是次品}\end{cases}$$

现在来看 ξ_i 的分布是否相同。$P(\xi_1=1)$表示第1件是次品，今次品率为 p，因此$P(\xi_1=1)=p$，于是$P(\xi_1=0)=1-p$。考虑$\xi_2=1$，即第2件是次品。由于第二件抽取的结果依赖于第1件，于是用条件概率公式及加法公式

$$P(\xi_2=1)=P(\xi_2=1,\xi_1=1)+P(\xi_2=1,\xi_1=0)$$

而$P(\xi_1=1,\xi_2=1)=P(\xi_1=1)P(\xi_2=1\mid\xi_1=1)=pP(\xi_2=1\mid\xi_1=1)$，当第1件抽得的是次品时，第2件抽到是次品的条件概率为（因为次品少了一件）

$$\binom{Np-1}{1}\Big/\binom{N-1}{1}\quad\left(=\frac{Np-1}{N-1}\right)$$

因此$P(\xi_1=1,\xi_2=1)=p\cdot\dfrac{Np-1}{N-1}$。又可类似地求出$P(\xi_1=0,\xi_2=1)$，因为

$$\begin{aligned}P(\xi_1=0,\xi_2=1)&=P(\xi_1=0)P(\xi_2=1\mid\xi_1=0)\\&=(1-p)P(\xi_2=1\mid\xi_1=0)\end{aligned}$$

在第1件不是次品条件下，第2件抽到次品的概率是

$\binom{Np}{1}\Big/\binom{N-1}{1}$，就是$Np/(N-1)$，于是

$$P(\xi_1=0,\xi_2=1)=(1-p)Np/(N-1)$$

这样 $P(\xi_2=1)=p\dfrac{Np-1}{N-1}+(1-p)\dfrac{Np}{N-1}=\dfrac{p(N-1)}{N-1}=p$，因此$P(\xi_2=0)=1-p$。

只消将上述证明中ξ_1改为$\sum\limits_{j=1}^{i}\xi_j$，$\xi_2$改为$\xi_{j+1}$，

所有的推理全相仿（注意并不完全一样），就可以证明
$P(\xi_i=1)=p, P(\xi_i=0)=1-p, i=1,2,\cdots,n$。
很明显，在现在讨论的问题中，$\xi_1, \xi_2, \cdots, \xi_n$是不独立的。

用1与0这两个值来描述打中与未打中是有点普遍意义的，也就是说，更一般的现象有时也可以用0与1来描述它。例如某一特定的现象发生（相当于打中）与不发生（相当于未打中），也可以用1与0描述。$\xi=1$表示发生，$\xi=0$表示未发生，ξ实际上就是发生的次数。从这里可以看到，只取1与0两个值的、这么简单的随机变量，其实是很重要的，它与不少重要的分布有密切的联系。因此可以说，取1或0来描述某些随机现象是很重要的手段。（以后我们在Ⅲ最小二乘法中，还要回来提到这一点，现在先声明一下）

例2．设有三个信封均已写好地址，相应的信笺也已写好，现将信随便放入封内，问这三封信恰巧都放对的概率是多少？

解：这时三个信封用三个号码代表，选用那三个都可以，一般就用1．2．3这三个号码来表示。三封信笺放入那个信封，每一信笺可能放对，也可能放错，应怎样来描述呢？号码1信封相应的信笺放的结果用ξ_1表示，$\xi_1=1$表示放入第1号信封，$\xi_1=2$表示放入第2号信封，$\xi_1=3$表示放入第3号信封。同样地，号码2信封相应的信笺放的结果用ξ_2表示，ξ_2也可以取1．2．3这三个数中的一个；ξ_3表示3号信封相应的信笺所放的结果。三封信恰巧都放对

就是$\xi_1=1$，$\xi_2=2$，$\xi_3=3$同时发生，因此所求的概率是$P(\xi_1=1, \xi_2=2, \xi_3=3)$。

求这个概率可以用两种方法来做：一是用排列组合的方法，三封信的放法共有3！这么多，全部都放对，只有一种情形，而3！种情形都是等概的，因此所求的概率是$\frac{1}{6}$(3！=6)。另一种用条件概率来做，即

$$
\begin{aligned}
&P(\xi_i=i,\ i=1,2,3)\\
&=P(\xi_1=1)P(\xi_2=2, \xi_3=3 \mid \xi_1=1)\\
&=P(\xi_1=1)P(\xi_2=2 \mid \xi_1=1)P(\xi_3=3 \mid \xi_1=1, \xi_2=2)
\end{aligned}
$$

显然$P(\xi_1=1)=\frac{1}{3}$，因为第一号信笺，在三个信封内可以随便放，放对的概率是$^1/_3$。第一个放对的条件下，第二号信笺只能在二个信封中随便放，放对的概率是$^1/_2$。第一、二号信笺都放对的条件下，第三号信笺只能放到第三号信封，因此$P(\xi_3=3 \mid \xi_1=1, \xi_2=2)=1$，于是

$P(\xi_1=1, \xi_2=2, \xi_3=3)=\frac{1}{3}\cdot\frac{1}{2}\cdot 1=\frac{1}{6}$这两种算法的结果是一样的。

从上面的讨论可以看到，我们讨论的是三个信封，三封信笺，实际上，这种方法对任意给定的n个信封，n封信笺也是完全一样的。

如果把问题稍换一下，随机变量的取法就有些不同。条

件还和原来一样，现在要求“至少有一封信放对的概率”。此时，如果随机变量的取法还和原来一样，描述“至少有一封信放对”会有困难，这时要引入新的随机变量，能反映出放对、放不对这一现象，注意到例1中告诉我们的1与0的表示法，取

$$(\text{I}.4)\qquad \eta_i=\begin{cases}1 & \text{当}\ \xi_i=i\\ 0 & \text{当}\ \xi_i\neq i\end{cases},\quad i=1,2,3$$

也即第 i 封信笺放对了，就为1，如第 i 封信没放对，就取0。于是 $\sum\limits_{i=1}^{3}\eta_i\geqslant 1$ 就是至少有一封放对，它的“逆事件”——即 $\sum\limits_{i=1}^{3}\eta_i<1$，但又因为 $\eta_i\geqslant 0$，因此不是 $\sum\limits_{i=1}^{3}\eta_i\geqslant 1$ 就是 $\sum\limits_{i=1}^{3}\eta_i=0$，并且当 $\sum\limits_{i=1}^{3}\eta_i=0$ 时，只可能 $\eta_1=\eta_2=\eta_3=0$，也即全都没有放对。可见 $\sum\limits_{i=1}^{3}\eta_i$ 是一个重要的随机变量，它反映了放对的信封数。可以看出，用随机变量描述问题时，事件的一些概念可以不出现，记 $\eta=\sum\limits_{i=1}^{3}\eta_i$，由于 $0\leq\eta_i\leq 1$，因此 $\eta\geqslant 0$，$\eta\leq 3$，实际上 η 只取整数值0，1，2，3这四个值。这就告诉我们 $\eta\geqslant 1$ 或 $\eta=0$ 两者必须出现其中之一，而且也只出现其中之一，$\eta\geqslant 1$ 可以用 $\eta\in[1,\infty)$ 表示，$\eta=0$ 可以用 $\eta\in\{0\}$ 表示，$[1,\infty)$ 与 $\{0\}$ 是两个互不相交的数集合，$\eta\in[1,\infty)$ 与 $\eta\in\{0\}$ 是两个互

斥的事件，可见互斥事件由随机变量描述后用两个互不相交的数集合来反映。注意到$\eta\in\{0\}$与$\eta<1$描述的是同一个事件，而$[1,\infty)$与$(-\infty,1)$不仅互不相交，并且这两个集合之和是$(-\infty,\infty)$，即整个实轴，也即将互逆的事件变成了两个互补的数集合，由于随机变量η只取有限的实数，$P(\eta\in(-\infty,\infty))=1$，这样

$$P(\eta\in(-\infty,1))=1-P(\eta\in[1,\infty))$$

即 $$P(\eta=0)=1-P(\eta\geqslant 1)$$

从这里，可以明显地感到随机变量起的作用是把"抽象的"随机事件、随机现象变成具体的"数集合"，对于掌握了数学工具的人，区间、整数集，自然比事件这类概念要熟悉得多。

从这个例还可以看出，有时确实要考虑随机变量的函数，例如本例中η_i就是ξ_i的函数。当然我们也可以不通过ξ_i，直接对η_i给出相应的定义，但是由ξ_i来定义η_i是比较自然和易于理解的。另外如η是$\eta_1+\eta_2+\eta_3$，它更明显地表成函数了。

本例中，$\{\eta_i\}$不是相互独立的，$P(\eta_3=1\mid\eta_1=\eta_2=1)=1$。这就是$P(\xi_3=3\mid\xi_1=1,\xi_2=2)=1$，因此用$P(\eta\geqslant 1)=1-P(\eta=0)=1-P(\eta_1=0,\eta_2=0,\eta_3=0)$的公式，先求$P(\eta_1=\eta_2=\eta_3=0)$并不好求，可以直接去求$P(\eta\geqslant 1)$的值。

不难看出，上述讨论也可适用于n个信封、n封信笺的

问题，此时η应为$\sum\limits_{i=1}^{n}\eta_i$这里的$\eta$又是$n$个只取0，1值的随机变量之和，它们（指$\eta_1,\cdots,\eta_n$）是不是同分布的？是不是相互独立的？$\eta$的分布是什么？这些都是很有意思的问题。通常的教科书上常有本例的问题，用随机事件的方法讨论的，只消运用上述的随机变量的定义，就可以作一次“翻译”，把随机事件都翻成随机变量的说法，这对于我们理解这两者的关系是有好处的。一般的学员只要知道这一点就够了，将来在工作中遇到实际问题时，学会用随机变量去描述。

例3. 已知生男孩、女孩的概率是相同的，即各为$\frac{1}{2}$，问在全部有两个孩子的家庭中，已知一个是男孩，另一个也是男孩的概率是多少？

解：两个孩子要用两个随机变量来描述比较方便，令ξ_i表示第i个孩子的性别情况，规定

$$\xi_i=\begin{cases}1 & \text{第 } i \text{ 个是男孩}\\ 0 & \text{第 } i \text{ 个不是男孩}\end{cases}\quad i=1,2$$

于是$\xi_1+\xi_2$就是两个孩子中男孩的个数，已知一个是男孩，也即已知$\xi_1+\xi_2\geqslant1$，所谓另一个也是男孩，实际上是两个都是男孩，也即$\xi_1+\xi_2=2$，因此所求的概率是

$$P(\xi_1+\xi_2=2\mid\xi_1+\xi_2\geqslant1)$$

但是另一个也是男孩，也可以由$\xi_1\xi_2$的值来表示，注意到

$$\xi_1\xi_2=\begin{cases}1 & \text{两个都是男孩}\\ 0 & \text{有一个不是男孩}\end{cases}$$

因此，所求的概率也可以写成

$$P(\xi_1\xi_2=1\mid\xi_1+\xi_2\geqslant 1)$$

注意到ξ_1与ξ_2是相互独立的，同分布的，并且有$P(\xi_i=1)=P(\xi_i=0)=\frac{1}{2}$，$i=1,2$．由例1知道$\xi_1+\xi_2$是二项分布，因此

$$P(\xi_1+\xi_2=1)=\binom{2}{1}\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)=\frac{1}{2}$$

$$P(\xi_1+\xi_2=2)=\binom{2}{2}\left(\frac{1}{2}\right)^2=\frac{1}{4}$$

$$P(\xi_1+\xi_2\geqslant 1)=\frac{1}{2}+\frac{1}{4}=\frac{3}{4}$$

于是由条件概率公式知道

$$\begin{aligned}&P(\xi_1+\xi_2=2\mid\xi_1+\xi_2\geqslant 1)\\&=\frac{P(\xi_1+\xi_2=2,\ \xi_1+\xi_2\geqslant 1)}{P(\xi_1+\xi_2\geqslant 1)}\\&=\frac{P(\xi_1+\xi_2=2)}{P(\xi_1+\xi_2\geqslant 1)}\\&=\frac{\frac{1}{4}}{\frac{3}{4}}=\frac{1}{3}\end{aligned}$$

如果计算$P(\xi_1\xi_2=1\mid\xi_1+\xi_2\geqslant 1)$，其结果也是$\frac{1}{3}$。一个方法只计算$\xi_1+\xi_2$的分布，另一个就要计算两个函数$\xi_1+\xi_2$与$\xi_1\xi_2$，可见怎么去描述随机现象还是有值得注意的地方。

最后我们举一个比较简单的离散型随机变量的例子。

例4．假定从观察资料已查到某一种害虫的分布情况。

平均每平方米有10.5条，如果害虫数的分布是泊松分布，问每平方米最多可能会有多少害虫(以0.99的概率估计)？

解：害虫数ξ(每平方米内)是一个随机变量，它遵从泊松分布是已知的，相应的参数$\lambda=10.5$。用0.99概率来估计最多的害虫数，就是要找一个数k_0，使

$$P(\xi>k_0)=0.01$$

也即 $$P(\xi\leq k_0)=0.99$$

利用泊松分布的表示式，也就是求k使

$$0.01=\sum_{k_0\leq k}\frac{\lambda^k}{k!}e^{-\lambda}$$

由于$\lambda=10.5$是已知的，因此可以查有关的泊松分布表，求出相应的k_0。若粗略地计算，由于0.01是一个很小的数，$\lambda^k/k!$趋于0的速度较快，因此可以近似认为$0.01\doteq\lambda^k e^{-\lambda}/k!$，用$k!$的近似值(斯特林公式)$k!\doteq\sqrt{2\pi k}k^k e^{-k}$，代入上式，得

$$0.01\sqrt{2\pi k}k^k\cdot e^{\lambda}=(\lambda e)^k$$

两边取对数，

$$\lambda+ln(0.01)\sqrt{2\pi}+\frac{1}{2}ln k+k\,lnk=k\,ln\,\lambda e$$

解上述方程，就可以求得k_0的近似值。

3．连续型随机变量的例题。

连续型随机变量的问题在选择随机变量上比较简单，困难往往集中在如何计算积分以求出相应的概率。

例5．已知测量误差公分遵从正态分布$N(0,12)$，物体的实际长度是201公分，独立测量了九次，问九次平均值ξ

遵从什么分布？

解：设 ξ_1，…，ξ_9分别依次为第1次到第9次的测量值，e_1，…，e_9分别依次为相应的误差，易见有关系式

$$\xi_i = 201 + e_i, \quad i = 1, 2, \cdots, 9$$

$$\xi = \frac{1}{9}\sum_{i=1}^{9} \xi_i$$

已知 e_1，…，e_9是独立同分布$N(0, 12)$，因此ξ_1，…，ξ_9也是独立同分布的（为什么应好好想一下），从随机变量线性变换的公式知道$\xi_i \sim N(201, 12)$，ξ又是ξ_1，…，ξ_9的均值，因此ξ的分布还是正态，期望值$E\xi$与$E\xi_i$一样，方差$D(\xi)$是原来的$1/9$，也即$\xi \sim N(201, \frac{12}{9})$，即$\xi \sim N(201, \frac{4}{3})$。

这里用上了线性变换的一般公式，用了两次。如果$\xi = a_1\xi_1 + a_2\xi_2 + \cdots + a_k\xi_k$，$a_1$，…，$a_k$是常数，$\xi_1$，…，$\xi_k$是随机变量，则有

$$E\xi = a_1 E\xi_1 + \cdots + a_k E\xi_k$$

特别$k = 2$，ξ_2是常数1，则$\xi = a_1\xi_1 + a_2$，就是通常的最简单的线性变换，从$E\xi$的一般公式得到

$$E\xi = a_1 E\xi_1 + a_2$$

如果ξ_1，…，ξ_k相互独立（这是最常见的情形），就有方差的关系式

$$D(\xi) = a_1^2 D(\xi_1) + \cdots + a_k^2 D(\xi_k)$$