第二章 随机变量及其分布

北京化工大学数学系

苏贵福

在第一章我们看到有些随机试验的结果可以用数来表示, 有些试验的结果则不能用数来描述. 当样本空间U的元素不是一个数时, 人们对于U就难以 描述和研究.

在第二章我们讨论如何引入一个法则, 将随机试验的每一个结果, 即将U的每个元素e与实数x对应起来, 从而引入随机变量的概念, 在此基础上揭示随机现象的统计规律性.

2.1离散型随机变量及其分布

一. 随机变量

例1 将一枚硬币抛掷三次, 观察出现正面和反面的情况, 样本空间是

$$U = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$$

以X记三次投掷得到正面H的总数, 那么对于样本空间 $U = \{e\}$中的每一个样本点e, X都有一个数与之对应. X是定义在U上的一个实值单值函数, 它的 定义域是样本空间U, 值域是实数集合$\{0, 1, 2, 3\}$. 则

$$X = X(e) = \begin{cases} 3, & e = HHH \\ 2, & e = HHT, HTH, THH \\ 1, & e = HTT, THT, TTH \\ 0, & e = TTT \end{cases}$$

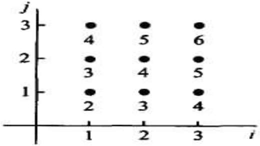
例2 在一袋中装有编号分别为1,2,3的3只球，在袋中任取一只球然后放回，再任取一只球，记录它们的号码，试验的样本空间为

$$U=\{e\}=\{(i,j)|i,j=1,2,3\}$$

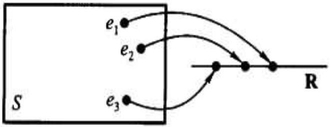
i,j分别为第1，第2次取到的球的号码.

以X记两球号码之和. 对于试验的每一个 结果$e=(i,j)\in U$，X都有一个指定的值$i+j$与之对应. 而X是定义在样本空间 U上的单值实值函数，它的定义域是样本空间U，值域是实数集合$\{2,3,4,5,6\}$. 则

$X=X(e)=X((i,j))=i+j \quad (i,j=1,2,3).$



定义1 设随机试验的样本空间为 $U=\{e\}$. $X=X(e)$是定义在样本空间U上的实值单值函数，称$X=X(e)$为随机变量.



随机变量的严格数学定义是测度论中的可测函数. 随机变量与微积分理论中的函数有着本质区别. 微积分中函数的取值是完全确定，而概率论中的函数是定义在样本空间上的函数，其取值依赖于随机试验的结果，因而其 取值具有一定的概率.

二. 离散型随机变量及其分布律

定义2 若随机变量的全部可能取值 是有限个或可列无限多个, 这种随机变量称为离散型随机变量.

因此要掌握一个离散型随机变量X的统计规律, 必须且只需知道X的所有可能取值以及每一个可能值的概率.

设离散型随机变量X所有可能的取值为x_k $(k=1,2,\cdots)$, X取各个可能值的概率, 即事件$\{X=x_k\}$的概率为

$$P\{X=x_k\}=p_k \quad (k=1,2,\cdots) \tag{1}$$

我们称(1)为离散型随机变量X的分布律.

离散型随机变量的分布律也可以用表格的形式表示

X	x_1	x_2	$\cdots$	x_n	$\cdots$
p_k	p_1	p_2	$\cdots$	p_n	$\cdots$

由概率的定义知, p_k满足如下两个条件:

- $p_k \geq 0, \quad k = 1, 2, \cdots.$
- $\sum_{k=1}^{\infty} p_k = 1.$

上述第二条成立是由于$\{X = x_1\} \cup \{X = x_2\} \cup \cdots$是必然事件, 且$\{X = x_i\} \cap \{X = x_j\} = \emptyset$. 故

$$1 = P[\bigcup_{k=1}^{\infty} \{X = x_k\}] = \sum_{k=1}^{\infty} P\{X = x_k\} = \sum_{k=1}^{\infty} p_k.$$

例3 设一汽车在开往目的地的道路上 需经过四组信号灯, 每组信号灯以1/2的概率允许或禁止汽车通过. 以X表示汽车首次 停下来时, 它已通过的信号灯的组数(设各组信号灯的工作相互独立). 求X的分布律.

解 以p表示每组信号灯禁止汽车通过的概率, 易知X的分布律为

X	0	1	2	3	4
p_k	p	$(1-p)p$	$(1-p)^2p$	$(1-p)^3p$	$(1-p)^4$

或可以写成

$$P\{X=k\}=(1-p)^kp,\quad k=0,1,2,3.$$

定义3 设随机变量X只可能取0与1两个值，它的分布律是

$$P\{X=k\}=p^k(1-p)^{1-k},\quad (k=0,1\text{且}0<p<1)$$

则称X服从以p为参数的(0-1)-分布或两点分布.

(0-1)-分布的分布律也可以写成

X	0	1
p_k	$1-p$	p

对一个随机试验，如果它的样本空间只包含两个元素，即$U=\{e_1,e_2\}$. 我们总能在U上定义一个服从(0-1)-分布的随机变量

$$X=X(e)=\begin{cases}0, & \text{当}e=e_1\\ 1, & \text{当}e=e_2\end{cases}$$

定义4 若随机变量X的分布律为

$$P\{X=k\}=\binom{n}{k}p^k(1-p)^{n-k}=\frac{n!}{k!(n-k)!}p^kq^{n-k}$$

其中$k=0,1,2,\cdots,n$, $0<p<1$, $q=1-p$. 则称随机变量X服从参数为n,p的二项分布, 记作$X\sim\mathcal{B}(n,p)$.

定义5 若随机变量X的分布律为

$$P\{X=k\}=\frac{e^{-\lambda}\lambda^k}{k!}$$

其中$k=0,1,2,\cdots$且$\lambda>0$. 则称随机变量X服从参数为λ的泊松分布, 记作$X\sim\mathcal{P}(\lambda)$.

定理1 设随机变量$X_n(n=1,2,\cdots)$服从 二项分布, 其分布律为

$$P\{X_n=k\}=\binom{n}{k}p_n^k(1-p_n)^{n-k}\quad(k=0,1,2,\cdots,n)$$

其中$n=1,2,\cdots$, p_n与n有关, 且$np_n=\lambda>0$, λ为常数. 则有

$$\lim_{n\to\infty}P\{X_n=k\}=\frac{e^{-\lambda}\lambda^k}{k!}$$

证明 由$p_n=\frac{\lambda}{n}$可知

$$\begin{aligned}P\{X_n=k\}&=\frac{n(n-1)\cdots(n-k+1)}{k!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}\\&=\frac{\lambda^k}{k!}\left[\underbrace{1\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)}_{s_1}\right]\underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{s_2}\underbrace{\left(1-\frac{\lambda}{n}\right)^{-k}}_{s_3}.\end{aligned}$$

而对于任意固定的$k(k=0,1,\cdots,n)$，有

$$s_1=\lim_{n\to\infty}\left[1\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\right]=1$$

$$s_2=\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n=\lim_{n\to\infty}\left[\left(1-\frac{\lambda}{n}\right)^{-\frac{n}{\lambda}}\right]^{-\lambda}=e^{-\lambda}$$

$$s_3=\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k}=1$$

于是得到

$$\lim_{n\to\infty}P\{X_n=k\}=\frac{e^{-\lambda}\lambda^k}{k!}.$$

上述定理表明：既可以用二项分布逼近泊松分布，也可以用泊松分布来近似二项分布。若给定泊松分布$\mathcal{P}(\lambda)$，可以用参数n比较大的二项分布$\mathcal{B}(n,\frac{\lambda}{n})$来逼近。若给定 参数$n$比较大的二项分布$\mathcal{B}(n,\frac{\lambda}{n})$，可以用泊松分布$\mathcal{P}(\lambda)$ 来近似。

例4 某人进行射击，设每次射击的命中率为0.02，独立射击400次，试求至少击中两次的概率.

解 将一次射击看成是一次试验. 设击中的次数为X，则$X \sim b(400, 0.02)$. 于是X的分布律为

$$P\{X=k\} = \binom{400}{k}(0.02)^k(0.98)^{400-k}, \quad k=0,1,\cdots,400.$$

于是所求概率为

$$\begin{aligned} P\{X \geq 2\} =& 1 - P\{X=0\} - P\{X=1\} \\ =& 1 - (0.98)^{400} - 400(0.02)(0.98)^{399} = 0.9972. \end{aligned}$$

♠ 虽然每次射击的概率都是0.02，但如果射击400次，则击中目标至少两次是几乎可以肯定的. 由此告诉人们绝不能轻视小概率事件.

事实上, 在例4中的计算是相当复杂的. 如果用泊松分布近似计算将会更加便捷:

$$\lambda = np = 400 \times 0.02 = 8$$

$$P\{X = 0\} \approx \frac{8^0}{0!}e^{-8} = e^{-8}$$

$$P\{X = 1\} \approx \frac{8^1}{1!}e^{-8} = 8e^{-8}$$

因此所求的概率为

$$P\{X \geq 2\} \approx 1 - e^{-8} - 8e^{-8} = 1 - 9e^{-8} = 0.997.$$

三. 随机变量的分布函数

定义6 对任意实数x, 随机变量X取值不超过x的累积 概率$P\{X \leq x\}$是实数x的函数, 称为随机变量X的分布函数或累积概率, 记作$F(x)$:

$$F(x) = P\{X \leq x\}$$

若$F(x)$是随机变量X的分布函数, 对于任意实数$x_1, x_2 (x_1 < x_2)$, 有

$$P\{x_1 < X \leq x_2\} = P\{X \leq x_2\} - P\{X \leq x_1\} = F(x_2) - F(x_1).$$

♣ 分布函数$F(x)$可以表示随机变量X落在任意区间$(x_1, x_2]$上的概率.

♣ 分布函数可以完整地描述随机变量概率分布的规律性.

♣ 如果将X看成是数轴上的随机点的坐标, 那么分布函数$F(x)$在x处的函数 值则表示X落在区间$(-\infty, x]$上的概率.

随机变量的分布函数是定义在实数轴上, 以闭区间$[0,1]$ 为值域的普通函数, 而这函数就是微积分理论中定义的函数.

有了分布函数就可以利用微积分理论来研究随机变量.

分布函数具有下列性质:

- $0 \le F(x) \le 1 \ (-\infty < x < +\infty)$.
- 若$x_1 < x_2$, 则$F(x_1) \le F(x_2)$, 即任一分布函数都是单调不减的.
- $F(-\infty) = \lim_{x\to-\infty} F(x) = 0$, $F(+\infty) = \lim_{x\to+\infty} F(x) = 1$.
- $\lim_{x\to x_0^+} F(x) = F(x_0)$.

反之, 具有上述性质的实函数都是某个随机变量的分布函数.

例5 设随机变量X的分布律为

X	-1	2	3
p_k	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

求X的分布函数，并求$P\{X \le \frac{1}{2}\}$，$P\{\frac{3}{2} < X \le \frac{5}{2}\}$，$P\{2 \le X \le 3\}$.

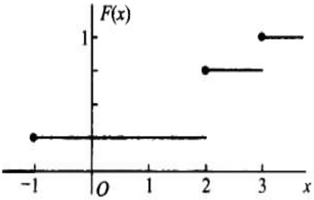
解 X仅在$x=-1,2,3$三点处其概率不为零，而$F(x)$的值是 $X \le x$的累积概率值. 由概率的有限可加性，它即为小于或等于x的那些 x_k处的概率p_k之和:

$$F(x)=\begin{cases} 0, & x<-1 \\ P\{X=-1\}, & -1 \le x < 2 \\ P\{X=-1\}+P\{X=2\}, & 2 \le x < 3 \\ 1, & x \ge 3 \end{cases}$$

也就是说

$$F(x)=\begin{cases}0, & x<-1\\ \frac{1}{4}, & -1\le x<2\\ \frac{3}{4}, & 2\le x<3\\ 1, & x\ge 3\end{cases}$$

易知$F(x)$的图像如下图, 它是一条阶梯形的曲线, 在$x=-1,2,3$处有跳跃点, 跳跃值分别为$\frac{1}{4},\frac{1}{2},\frac{1}{4}$.



因此所求的概率为:

$$P\left\{X \le \frac{1}{2}\right\} = F\left(\frac{1}{2}\right) = \frac{1}{4}$$
$$P\left\{\frac{3}{2} < X \le \frac{5}{2}\right\} = F\left(\frac{5}{2}\right) - F\left(\frac{3}{2}\right) = \frac{3}{4} - \frac{1}{4} = \frac{1}{2}$$
$$P\{2 \le X \le 3\} = F(3) - F(2) + P\{X = 2\} = 1 - \frac{3}{4} + \frac{1}{2} = \frac{3}{4}$$