第二章 随机变量及其分布

• 中山大学 • 计算机学院 • 郑培嘉 • Email：zhpj@mail.sysu.edu.cn

目录

1. 随机变量

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

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

X是定义在样本空间S上的一个实值单值函数(以X为三次投掷得到正面H的总数)，定义域是样本空间S，值域是实数集合{0, 1, 2, 3}. 用函数记号可将X写为:

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

随机变量

例2： 在一袋中装有编号1，2，3的3只球，在袋中任取一只球，放回，再另取一只球。记录它们的号码，该实验的样本空间为$S = \{e\} = \{(i,j)|i,j = 1,2,3\}$, i, j分别为第1，2次取到球的号码。

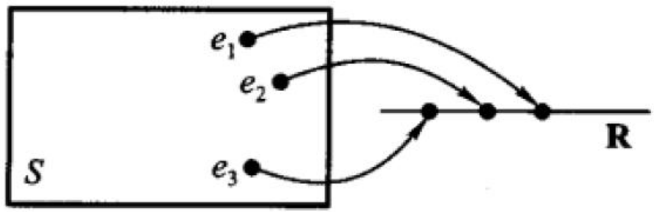
X是定义在样本空间S上的一个实值单值函数(以X为两球号码之和），定义域是样本空间S，值域是实数集合{2, 3, 4, 5, 6}.用函数记号可将X写为:

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

随机变量

◆ **定义：** 设随机实验的样本空间为S={e}. X=X(e)是定义在样本空间S上的实值单值函数，称X=X(e)为**随机变量**。

下图为样本点与实数X=X(e)对应的示意图



注： 有些随机实验的样本空间，其元素本身就是一个数，令X=X(e)=e，X就是一个随机变量。

以Y记录某车间一天的缺勤人数，以W记录某地区第一季度的降雨量，以Z记录某工厂一天的耗电量，以N记录某医院一天的挂号人数 ……

随机变量

- 一般用大写字母如 X，Y，Z，W… 表示随机变量，以小写字母 x，y，z，w… 表示实数。
- 随机变量的取证随实验的结果而定，而实验的各个结果出现有一定的概率，因而随机变量的取值有一定的概率。

如： 上例1中X的取值为2，记成{X=2}，对应的样本点集合为A={HHT，HTH，THH}，当且仅当事件A发生时有{X=2}。

我们称概率P(A)=P{HHT，HTH，THH}为{X=2}的概率，即P{X=2}=P(A)=3/8。

- 一般，L为实数集合，将X在L上取值写为$\{X \in L\}$，表示事件$B=\{e|X(e) \in L\}$，即B由S中使得$X(e) \in \mathrm{L}$的所有样本点e所组成的事件，有

$$P\{X \in L\} = P(B) = P\{e|X(e) \in L\}.$$

2. 离散型随机变量及其分布律

离散型随机变量及其分布律

- **定义**：如果随机变量的可能取值个数（注：非样本空间元素个数）是有限个或者可数无穷多个，这样的随机变量称为**离散型随机变量**。

如：上例1中X的取值只有0, 1, 2, 3四个值；
上例2中X的取值只有2, 3, 4, 5, 6五个值等等；

如：以T为某元件的寿命，它所可能取的值充满一个区间，无法按一定次序一一列举出来，故其为**非离散型随机变量**。

离散型随机变量及其分布律

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

$$P\{X = x_k\} = p_k,\ k = 1,2, \ldots$$

由概率的定义，p_k满足如下两个条件：

1) $p_k \geq 0,\ k = 1,2, \ldots$

2) $\sum_{k=1}^{\infty} p_k = 1$

我们称$P\{X = x_k\} = p_k$为离散型随机变量X的**分布律**，表格形式为：

X	X₁	X₂	…	Xₙ	…
p_k	p_1	p_2	…	p_n	…

离散型随机变量及其分布律

例：设一汽车在开往目的地的道路上需要经过四组信号灯，每组信号灯以0.5的概率允许或禁止汽车通行。以X表示汽车首次停下时，它已通过的信号灯的组数(设各组信号灯的工作是互相独立的)，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	$(1-p)p$	$(1-p)^2p$	$(1-p)^3p$	$(1-p)^4$

或写成$P\{X=\mathrm{k}\}=(1-p)^kp, k=0,1,2,3, P\{X=4\}=(1-p)^4$

离散型随机变量及其分布律

◆ 0-1分布

随机变量X只可能取0-1两个值，它的分布律是 $P\{X=\mathrm{k}\}=p^k(1-p)^{1-k}$， $k=0,1$， 则称X服从以p为参数的(0-1)分布或两点分布。

表格形式：

X	0	1
p_k	$1-p$	p

离散型随机变量及其分布律

◆ 伯努利试验

设实验E只有两个可能结果 A，$\overline{\mathrm{A}}$，则称E为**伯努利试验**。

设$P(A)=p(0<p<1)$，此时$P(\overline{\mathrm{A}})=1-p$。

将E独立重复地进行n次，则称这一串重复地独立试验为**n重伯努利试验**。

“**重复**”指每次试验中$P(A)=p$保持不变。

“**独立**”指各次实验结果互不影响，设$P(C_1C_2\ldots C_n)=P(C_1)P(C_2)\ldots P(C_n)$，$C_{\mathrm{i}}$表示第i次实验的结果 i=1, 2…n， C_{i}取A或$\overline{\mathrm{A}}$。

离散型随机变量及其分布律

◆ 二项分布

以X表示n重伯努利试验中事件A发生地次数，则有

$$P\{X=\mathrm{k}\}=\mathrm{C}_n^k p^k(1-p)^{n-k},\ k=0,1,2,\ldots,n$$

并称随机变量X服从参数为n, p的二项分布，记为$X \sim b(n,p)$

特别地，当n=1时，二项分布化为：

$$P\{X=\mathrm{k}\}=p^k(1-p)^{1-k}$$

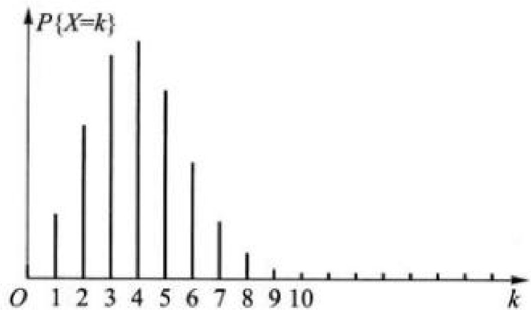
即退化为(0-1)分布。

离散型随机变量及其分布律

例：按规定，某种型号电子元件的使用寿命超过1500小时的为一等品，已知某一大批产品的一级品率为0.2， 现在从中随机地抽查20只，问20只元件中恰有k只（k=0, 1, …20）为一级品的概率是多少？

解：由于元件总数较大，而抽取的数量较小，近似看成放回抽样处理。以X记20只元件中一级品的只数，且有 $X \sim b(20,0.2)$, 则所求概率为

$$P\{X=\mathrm{k}\}=C_{20}^{k}0.2^{k}0.8^{20-k},\ k=0,1,2,\ldots,20$$



概率大小先增加，后减少。最大值出现在 k = 4。

我们指出，一般对于固定的n, p，二项分布b(n, p)都具有这一性质

离散型随机变量及其分布律

例：某人进行射击，设每次射击的命中率为0. 02，独立设计400次，试求至少击中两次的概率。

解：设击中次数为X，则$X\sim b(400,0.02)$, X的分布律为

$$P\{X=\mathrm{k}\}=\mathrm{C}_{400}^{k}0.02^{k}0.98^{400-k},\ k=0,1,2,\ldots,400$$

故所求概率为

$$P\{X\geq 2\}=1-P\{X=0\}-P\{X=1\}=1-0.98^{400}-400\times 0.02\times 0.98^{399}=0.9972$$

离散型随机变量及其分布律

上题的结果概率很接近100%，可从两方面讨论该结果：

- 虽然每次击中概率比较低(0.02)，但如果射击次数较多（400），击中至少2次是几乎可以肯定的。一事件尽管在一次实验中发生的概率很低，但只要试验次数多，且试验是独立进行的，则这一事件发生几乎是肯定的。
- 如果某一射手在400次中击中目标的次数不到两次，由于概率 $P\{X<2\} \approx 0.003$ 根据实际推断原理，我们将怀疑“每次射击的命中率为0.02”这一假设，认为命中率应该小于0.02。

离散型随机变量及其分布律

例：设有80台同类型设备，各台工作相互独立，发生故障的概率都为0.01，且一台设备的故障能由一个人处理，考虑两种配备维修个人的方法，其一是由4人维护，每人负责20台；其二是有3人维护共同维护80台，试比较这两种方法在设备发生故障时不能及时维修的概率。

解：按第一种方法。以X记为“第1人维护的20台中同一时刻发生故障的台数”，以$A_i(i=1,2,3,4)$表示事件“第i个人维护的20台中发生故障不能及时维修”，

离散型随机变量及其分布律

则80台中发生故障且不能及时维修的概率为

$$P(A_1 \cup A_2 \cup A_3 \cup A_4) \geq P(A_1) = P\{X \geq 2\}$$

而$X\sim b(20,0.01)$, 故有

$$P\{X \geq 2\} = 1 - \sum_{k=0}^{1} P\{X = k\} = 1 - \sum_{k=0}^{1} C_{20}^{k} 0.01^{k} 0.99^{20-k} = 0.0169$$

即有$P(A_1 \cup A_2 \cup A_3 \cup A_4) \geq 0.0169$

按第二种方法。设Y为80台中同一同一时刻发生故障的台数，此时Y$\sim b(80,0.01)$, 则80台中发生故障且不能及时维修的概率为

$$P\{Y \geq 4\} = 1 - \sum_{k=0}^{3} C_{80}^{k} 0.01^{k} 0.99^{80-k} = 0.0087$$

◆ 泊松分布

设随机变量X所有可能取值为0, 1, 2, ⋯, 而取各个值得概率为

$$P\{X=\mathrm{k}\}=\frac{\lambda^{\mathrm{k}}e^{-\lambda}}{k!},\ k=0,1,2,\ldots,$$

其中$\lambda>0$是常数，则称X服从参数为λ的**泊松分布**，记为$X\sim\pi(\lambda)$。

分布律满足两个条件：

- $P\{X=\mathrm{k}\}=\frac{\lambda^{\mathrm{k}}e^{-\lambda}}{k!}\geq 0$
- $\sum_{k=0}^{\infty}P\{X=\mathrm{k}\}=\sum_{k=0}^{\infty}\frac{\lambda^{\mathrm{k}}e^{-\lambda}}{k!}=e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^{k}}{k!}=e^{-\lambda}\cdot e^{\lambda}=1$

离散型随机变量及其分布律

◆ 泊松定理

设λ>0是一个常数，n是任意正整数，设$np_n = \lambda$, 则对于任一固定非负整数k，有

$$\lim_{n\to\infty} C_n^k {p_n}^k (1-p_n)^{n-k} = \frac{\lambda^k e^{-\lambda}}{k!}$$

证： 由$p_n = \frac{\lambda}{n}$,有

$$C_n^k {p_n}^k (1-p_n)^{n-k} = \frac{n(n-1)\dots(n-k+1)}{k!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{\lambda^k}{k!}\left[1\cdot\left(1-\frac{1}{n}\right)\dots\left(1-\frac{k-1}{n}\right)\right]\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k}$$

对于任意固定的k, 当$n\to\infty$时，

$$1\cdot\left(1-\frac{1}{n}\right)\dots\left(1-\frac{k-1}{n}\right)\to 1,\quad \left(1-\frac{\lambda}{n}\right)^n \to e^{-\lambda},\quad \left(1-\frac{\lambda}{n}\right)^{-k}\to 1$$

故有$\lim_{n\to\infty} C_n^k {p_n}^k (1-p_n)^{n-k} = \frac{\lambda^k e^{-\lambda}}{k!}$

离散型随机变量及其分布律

例：计算机硬件公司制造某种特殊型号的微型芯片，次品率达0.1%，各芯片称为次品相互独立，求1000只产品中至少有2只次品的概率。以X表示产品中的次品数，则

$X \sim b(1000,\ 0.001)$

解：所求概率为

$$P\{X \geq 2\} = 1 - P\{X = 0\} - P\{X = 1\}$$

$$= 1 - 0.999^{1000} - \mathrm{C}_{1000}^{1} 0.999^{999} \times 0.001 \approx 0.2642411$$

利用泊松分布对二项分布进行近似得到

$$\lambda = 1000 \times 0.001 = 1$$

$$P\{X \geq 2\} = 1 - P\{X = 0\} - P\{X = 1\} \approx 1 - \mathrm{e}^{-1} - \mathrm{e}^{-1} \approx 0.2642411$$

3. 随机变量的分布函数

随机变量的分布函数

◆ 设X是一个随机变量（离散或连续）， x是任意实数，函数

$$F(x) = P\{X \leq x\},\ -\infty < x < +\infty$$

称为X的**分布函数**。

对于任意实数$x_1, x_2 (x_1 < x_2)$, 有

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

若已知X的分布函数，我们就可以计算X落在任一区间（x_1, x_2]上的概率；从这个意义上看，分布函数完整地描述了随机变量的统计规律。

随机变量的分布函数

◆ 分布函数的**基本性质**：

□ $F(x)$是一个不减函数。

对于任意实数$x_1, x_2(x_1 < x_2)$, 有

$$F(x_2) - F(x_1) = P\{x_1 < X \le x_2\} \ge 0$$

□ $0 \le F(x) \le 1$，且

$$F(-\infty) = \lim_{x\to-\infty} F(x) = 0, \; F(\infty) = \lim_{x\to\infty} F(x) = 1$$

□ $F(x+0) = F(x)$, 即$F(x)$是右连续的。

如某一函数同时满足上述三条基本性质，则该函数是某随机变量的分布函数。

随机变量的分布函数

例：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$P\{X \le 0.5\}, P\{1.5 \le X \le 2.5\}, P\{2 \le X \le 3\}$.

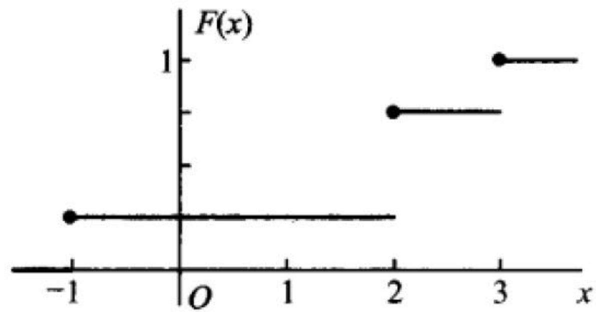
解：有

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

随机变量的分布函数

即有

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



$F(x)$的图形如上图所示：

$P\{X\leq 0.5\}=F(0.5)=\frac{1}{4}$

$P\{1.5\leq X\leq 2.5\}=F(2.5)-F(1.5)=\frac{1}{2}$

$P\{2\leq X\leq 3\}=F(3)-F(2)+P\{X=2\}=\frac{3}{4}$

随机变量的分布函数

例：一个靶子是半径为2m的圆盘，击中靶上任一同心圆盘上的点的概率与该圆盘的面积成正比，并假设射击都能中靶，以X表示弹着点与圆心的距离，试求随机变量X的分布函数。

解：若$x < 0$, 则$\{X \le x\}$是不可能事件，$F(x) = P\{X \le x\} = 0$

若$0 \le x \le 2$，则$P\{0 \le X \le x\} = \frac{x^2}{4}$,

$$F(x) = P\{X \le x\} = F(x) = P\{X < 0\} + P\{0 \le X \le x\} = \frac{x^2}{4}$$

若$x \ge 2$，则$\{X \le x\}$是必然事件，$F(x) = P\{X \le x\} = 1$

综上，X的分布函数为

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

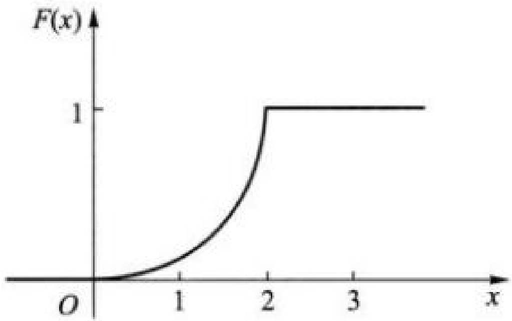
随机变量的分布函数

它的图像如图所示：

本题的分布函数，对于任意x可以写成

$$F(x) = \int_{-\infty}^{x} f(t)dt,$$

其中$f(t) = \begin{cases} \frac{t}{2}, & 0 < t < 2 \\ 0, & \text{其他} \end{cases}$



即$F(x)$恰为非负函数$f(t)$在区间$(-\infty, x]$上的积分，此时我们称X为**连续型随机变量**。

4. 连续型随机变量及其概率密度

连续型随机变量及其概率密度

◆ 如果对于随机变量X的分布函数$F(x)$, 存在非负可积函数$f(x)$, 使对于任意实数x有

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

则称X为**连续型随机变量**，$f(x)$称为X的**概率密度函数**，简称**概率密度。**