第二章　　随机变量及其分布

- 随机变量
- 离散型随机变量及其分布律
- 随机变量的分布函数
- 连续型随机变量及其密度函数
- 随机变量的函数的分布

2.1随机变量

Random Variable and Distribution

在前面的学习中，我们用字母**A、B、C...**表示事件，并视之为样本空间S的子集；针对等可能概型，主要研究了用排列组合手段计算事件的概率。

本章，将用随机变量表示随机事件，以便采用高等数学的方法描述、研究随机现象。

■基本思想

将样本空间数量化, 即用数值来表示试验的结果

- 有些随机试验的结果可直接用数值来表示.

例如: 在掷骰子试验中,结果可用**1,2,3,4,5,6**来表示

- 有些随机试验的结果不是用数量来表示，但可数量化

例如: 掷硬币试验,其结果是用汉字“正面”和“反面”来表示的

可规定： 用 **1**表示“正面朝上” 用 **0** 表示“反面朝上”

试验结果的数量化

例　设箱中有10个球，其中有2个红球，8个白球；从中任意抽取2个，观察抽球结果。

解、取球结果为：两个白球；两个红球；一红一白

如果用X表示取得的红球数，则X的取值可为0，1，2。

此时，“两只红球”＝“X取到值2”，可记为 $\{X=2\}$

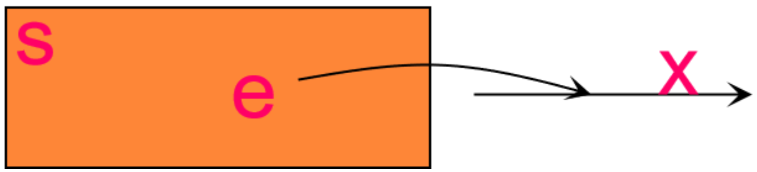
“一红一白”　记为 $\{X=1\}$，

“两只白球”　记为 $\{X=0\}$

特点：试验结果数量化了，试验结果与数建立了对应关系

随机变量的定义

定义. 设S={e}是试验的样本空间，如果量X是定义在S上的一个单值实值函数即对于每一个e∈S，有一实数X=X(e)与之对应，则称X为**随机变量**。



X=f(e)——为S上的单值函数，X为实数

简言之：随试验结果而变的量X为随机变量

随机变量常用大写字母X、Y、Z 等表示，而小写字母x,y,z等表示实数

引入随机变量的意义

随机变量的引入是概率论发展走向成熟的一个标志，它弥补了随机试验下的随机事件种类繁多，不易一一总结它们取值规律的缺陷，因为如果知道随机变量的分布，随机试验下任一随机事件的概率也随之可以得到；另外引入随机变量后，可以使用数学中微积分工具讨论随机变量的分布。

?

请举几个实际中随机变量的例子

EX. 引入适当的随机变量描述下列事件：

①将3个球随机地放入三个格子中，

事件A={有1个空格}，B={有2个空格}，

C={全有球}。

②进行5次试验，事件D={试验成功一次}，

F={试验至少成功一次}，G={至多成功3次}

随机变量的类型

■ 离散型

随机变量的所有取值是有限个或可列个

■ 连续型

随机变量的取值有无穷多个，且不可列

其中连续型随机变量是一种重要类型

连续型随机变量即取值于一个连续区间全部数值的随机变量。

2.2 离散随机变量及其分布律

设离散型随机变量 X **的所有可能取值是**

$x_1, x_2, \cdots, x_n, \cdots$ **，而取值** x_k **的概率为** p_k

即 $\boxed{P\{X = x_k\} = p_k}$

称此式为X的分布律或概率分布（**Probability distribution)**。

分布律也可以用表的形式来表示，如下

$\boldsymbol{X}$	x_1	x_2	$\cdots$	x_k	$\cdots$
$\boldsymbol{P}$	p_1	p_2	$\cdots$	p_k	$\cdots$

分布律的性质

- $p_k \geq 0,\ k = 1,2,\cdots$ ———— 非负性
- $\sum_{k=1}^{\infty} p_k = 1$ ———— 归一性

求分布律举例

例1 设有一批产品20件，其中有3件次品，从中任意抽取2件，如果用X表示取得的次品数，求随机变量X的分布律及事件“至少抽得一件次品”的概率。

解：**X**的可能取值为 **0，1，2**

$$P\{X=0\} = \frac{C_{17}^{2}}{C_{20}^{2}} = \frac{136}{190} = P\{\text{抽得的两件全为正品}\}$$

$$P\{X=1\} = \frac{C_{3}^{1}C_{17}^{1}}{C_{20}^{2}} = \frac{51}{190} = P\{\text{只有一件为次品}\}$$

$$P\{X=2\} = \frac{C_{3}^{2}}{C_{20}^{2}} = \frac{3}{190} = P\{\text{抽得的两件全为次品}\}$$

故 X的分布律为

X	0	1	2
p_k	$\frac{136}{190}$	$\frac{51}{190}$	$\frac{3}{190}$

而“至少抽得一件次品”$=\{X\geqslant 1\}$ $=\{X=1\}\cup\{X=2\}$

注意：$\{X=1\}$与$\{X=2\}$是互不相容的！

故 $P\{X\geqslant 1\}=P\{X=1\}+P\{X=2\}$ $=\frac{51}{190}+\frac{3}{190}=\frac{54}{190}=\frac{27}{95}$

实际上，这仍是古典概型的计算题，只是表达事件的方式变了

例设一汽车开往目的地的道路上需要经过4组信号灯，每组信号灯以0.5的概率允许或禁止汽车通过。以X表示汽车首次停下时，他通过的信号灯的组数（设各组信号灯的工作是相互独立的），求X的分布律。

解　用 p 表示每组信号灯禁止汽车通过的概率，可知

X的分布律如下

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

或写成　$p_k=(1-p)^k p, k=0,1,2,3;\ p_4=(1-p)^4$

当p=0.5时其结果如下

X	0	1	2	3	4
p_k	0.5	0.25	0.125	0.0625	0.0625

例6 某射手射击某一目标，直到射中为止。某每次射击命中率为p，求射击次数X的分布

解：X＝1表示第一次射击命中, $P\{X=1\}=p$

$X=2$表示第二次命中，第一次未中，

$$P\{X=2\}=(1-p)p$$

$X=i$表示第i次命中，前$i-1$次未中，

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

X的概率函数为

$$P\{X=i\}=p(1-p)^{i-1} \qquad i=1,2,\ldots$$

亦称X为服从参数p的几何分布。

三、几种常见的离散型分布

1、0-1分布(二点分布)

样本空间中只有两个样本点

△定义： 若随机变量X的分布律为:

X	0	1
P	1−p	p

则称X服从参数为p 的二点分布或(0-1)分布,

应用场合

凡试验只有两个结果, 常用**0－1**分布描述,

如产品是否合格、人口性别统计、系统是否正常、电力消耗是否超标等等。

例 设一个袋中装有3个红球和7个白球，现在从中随机抽取一球，如果每个球抽取的机会相等，

并且用数“1”代表取得红球，“0”代表取得白球，则随机抽取一球所得的值是一个离散型随机变量

$$X=\begin{cases}1 & （取得红球）\\ 0 & （取得白球）\end{cases}$$

其概率分布为 $P\{X=1\}=\dfrac{3}{10}$ $P\{X=0\}=\dfrac{7}{10}$

即X服从两点分布。

2、二项分布 Binomial distribution

伯努利试验概型

现实生活中我们接触到的随机现象，有不少恰好是有两个可能的结果，例如，抛硬币出现正面或反面；新生婴儿是男孩或女孩；射击命中还是不命中；种子发芽或不发芽；考试是否及格；上课迟到或不迟到；身体有病还是没病；买彩票是否中奖等等。这些“非此即彼”的结果都可以简化为“成功”和“失败”，即把我们感兴趣的结果称作“成功”，与之对立的结果称为失败。

假设随机试验是可以重复的，只有两种可能结果的试验就称为伯努利试验。它最先被瑞士数学家雅各布.伯努利研究。

n 重伯努利 (Bernoulli) 试验要满足以下三个条件:

- 试验可重复 n 次，“重复”是指在每次实验中 P(A)=p
- 每次试验只有两个可能的结果：A , $\overline{A}$
- 每次试验的结果与其他次试验无关——称为这 n 次试验是相互独立的

例：

1. 独立重复地抛n次硬币，每次只有两个可能的结果：

正面，反面，

$$P(\text{出现正面}) = 1/2$$

2. 将一颗骰子抛n次，设A={得到1点}，则每次试验只有两个结果：

$$A, \bar{A}, \qquad P(A) = 1/6$$

3. 从52张牌中<u>有放回</u>地取n次，设A={取到红牌}，则每次只有两个结果：

$$A, \bar{A}, \qquad P(A) = 1/2$$

如果是不放回抽样呢？

27

◆ 设A在n重贝努利试验中发生X次，则

$$P\{X=k\}=C_n^k p^k(1-p)^{n-k},\ k=0,1,\cdots,\ n$$

并称X服从参数为p的**二项分布**，记 $X\sim b(n,\ p)$

注：$1=(p+q)^n=\sum_{k=0}^{n} C_n^k p^k q^{n-k}$ 其中$q=1-p$

推导：设A_i={ 第i次A发生 }，先设n=3

$$P\{X=0\}=P(\bar{A}_1\bar{A}_2\bar{A}_3)=(1-p)^3$$

$$P\{X=1\}=P(A_1\bar{A}_2\bar{A}_3\cup\bar{A}_1A_2\bar{A}_3\cup\bar{A}_1\bar{A}_2A_3)=C_3^1p^1(1-p)^{3-1}$$

$$P\{X=2\}=P(A_1A_2\bar{A}_3\cup A_1\bar{A}_2A_3\cup\bar{A}_1A_2A_3)=C_3^2p^2(1-p)^{3-2}$$

$$P\{X=3\}=P(A_1A_2A_3)=p^3$$

一般 $P\{X=k\}=C_n^k p^k(1-p)^{n-k},k=0,1,2,\cdots,n$

27

二项分布　　Binomial distribution

- 在n重贝努利试验中，若以X表示事件A发生的次数，

则X可能的取值为0，1，2，3，…，n.

- 随机变量X的分布律

$$P\{X=k\}=C_n^k p^k q^{n-k},\ k=0,1,2\ldots,n;$$

其中**0< p <1**，则称**X**服从参数为 **n, p** 的二项分布(也称**Bernoulli** 分布），记为

$$X \sim b(n, p)$$

根据二项分布的定义，显然有

$$(1)\ P\{X=k\}\geq 0,\ k=0,1,2,\cdots,n;$$

$$(2)\sum_{k=0}^{n} P\{X=k\}=\sum_{k=0}^{n} C_n^k p^k q^{n-k}=(p+q)^n=1$$

当n=1时，二项分布退化为0-1分布，记为

X～b(1, p)

例 某人进行射击，设他每次命中目标的概率均为**0.6.** 现对目标射击**5**次,求（**1**）目标恰好被击中**3**次的概率;（**2**）目标被击中的概率.

解 将一次射击看作是一次伯努利试验，令X表示射击中命中次数,则 $X \sim b(5,0.6)$

（**1**）**5**次射击中目标恰好被击中**3**次的概率

$$P\{X=3\}=C_5^3(0.6)^3(0.4)^2=0.3456$$

（**2**）**5**次射击中目标被击中的概率

$$P\{X\geq 1\}=\sum_{k=1}^{5}P\{X=k\}=1-P\{X=0\}=1-C_5^0(0.6)^0(0.4)^5=0.9898$$

例 设在一次实验中A发生的概率为**P(A)=ε(ε**是很小的数**)**，把实验独立地重复做**n**次，求事件**A**至少发生一次的概率。

解 令X表示**n**次试验中**A**发生的次数**,**

则 $X \sim b(n, \varepsilon)$

于是所求概率为

$$P\{X \ge 1\} = 1 - P\{X < 1\} = 1 - P\{X = 0\}$$

$$= 1 - C_n^0 \varepsilon^0 (1-\varepsilon)^n = 1 - (1-\varepsilon)^n$$

易知，当$n \to \infty$时$, 1-(1-\varepsilon)^n \to 1.$

这就说明，当**n**充分大时，事件**A**迟早要发生。

从而，得到一个重要结论：“小概率事件在大量重复试验中迟早要发生”

由此可见日常生活中“提高警惕,防火防盗”的重要性.

由于时间无限,自然界发生地震、海啸、空难、泥石流等都是必然的，早晚的事，不用奇怪，不用惊慌.

同样,人生中发生车祸、失恋、患绝症、考试不及格、炒股大亏损等都是正常现象,大可不必怨天尤人,更不要想不开而跳楼自杀.

例 设同类型设备80台，每台工作相互独立，每台设备发生故障的概率都是0.01. 在通常情况下，一台设备发生故障可由一个人独立维修，每人同时也只能维修一台设备.考虑两种配备维修工人的方法，其一是由4人分别维护，每人负责20台；其二 3个人共同负责80台。试比较这两种方案设备发生故障时不能及时维修的概率的大小。

解、第一种方案：**X=**“一个人维护20台中同一时刻发生故障的台数”，A_i=“第**i**人维护的20台中发生故障时不能及时维修”，则80台中发生故障时不能及时维修的概率为

$$P(A_1 \cup A_2 \cup A_3 \cup A_4) = 1 - P(\overline{A_1})P(\overline{A_2})P(\overline{A_3})P(\overline{A_4})$$

$$P(A_i)=P\{X\geq 2\}$$

由于$X\sim b(20,0.01)$,则有

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

$$=0.0169$$

所以$P(A_1\cup A_2\cup A_3\cup A_4)=1-P(\overline{A_1})P(\overline{A_2})P(\overline{A_3})P(\overline{A_4})$

$$=1-(1-0.0169)^4=0.0659$$

第二种方案：**Y=“80**台中同一时刻发生故障的台数”

则$Y\sim b(80,0.01)$,于是，**80**台发生故障不能及时维修的概率为

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

$$=0.0087<0.0659$$

例 某地有**5000**人参加某种人寿保险，规定：参加保险者在一年的第一天交付**100**元保险金。如果在**1**年内投保人死亡，则家属可以从保险公司领取**3**万元赔偿费。设在一年内被保险者的死亡率为**0.001**，求该保险公司在这一年中至少盈利**20**万元的概率。

解、设X为5000个投保者在1年内死亡的人数，则

$$X \sim b(5000\ ,\ 0.001)$$

如果投保人在1年内的有X人死亡，则这一年内保险公司

的收入为 5000×100-30000X=500000-30000X=50-3X

（万元）。所求概率为

P{50-3X⩾20}=P{X⩽10}

$$= \sum_{k=0}^{10} C_{5000}^{k} 0.001^{k} 0.999^{5000-k}$$

P{50-3X⩾20}=P{X⩽10}

$$=\sum_{k=0}^{10} C_{5000}^{k} 0.001^{k} 0.999^{5000-k}$$

显然要准确计算上式不是一件很容易的事。下面给

出其近似值的泊松定理

则当n 较大，p 较小（一般$n \geq 20, p \leq 0.05$ ），有

$$C_n^k p^k (1-p)^{n-k} \approx \frac{e^{-\lambda}\lambda^k}{k!}, \qquad k=0,1,2,\cdots$$

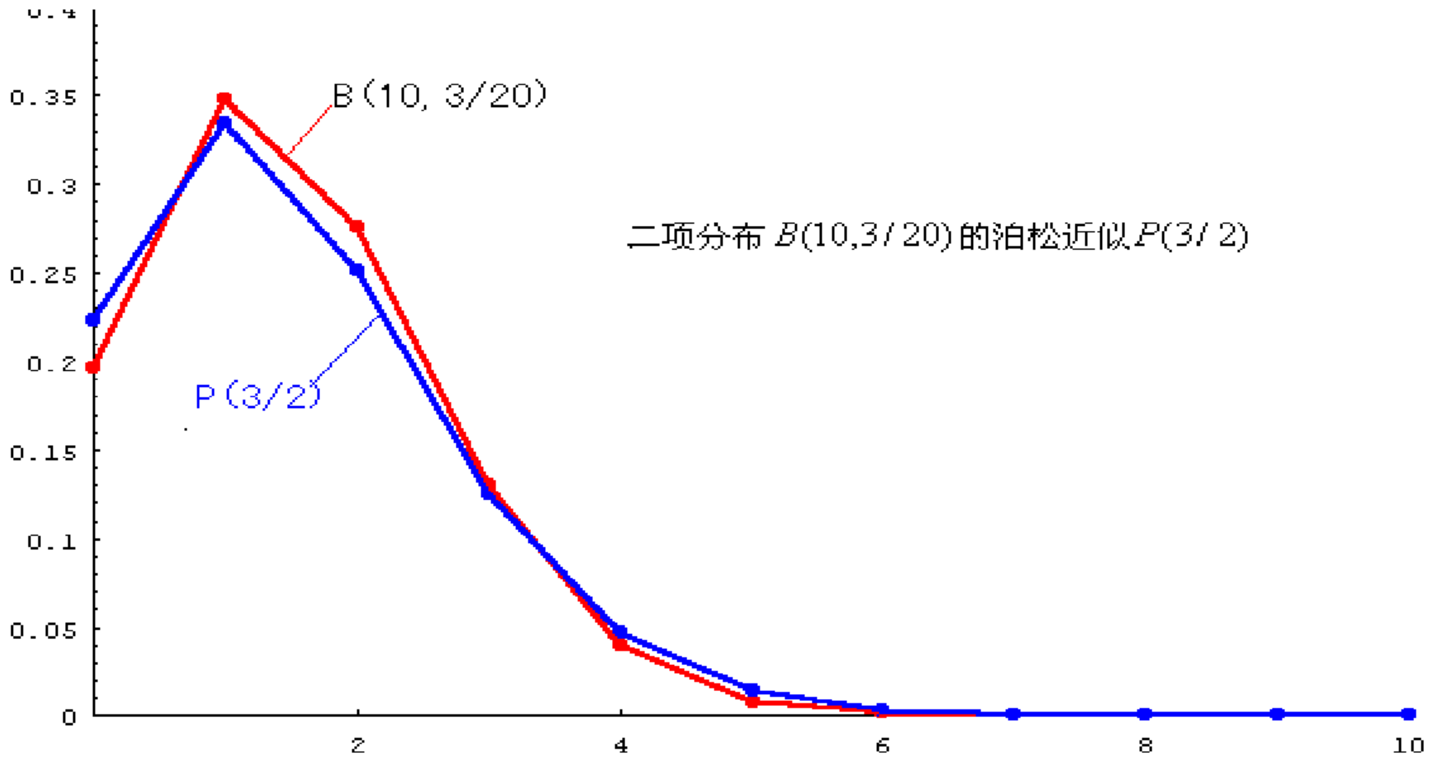
式中$\lambda=np$。

而$\dfrac{e^{-\lambda}\lambda^k}{k!}$的值可通过查泊松分布表得到。

泊松定理表明，泊松分布是二项分布的极限分布，

当n很大，p很小时，二项分布就可近似地

看成是参数λ=np的泊松分布



P{50-3X⩾20}=P{X⩽10}

$$=\sum_{k=0}^{10} C_{5000}^{k} 0.001^{k} 0.999^{5000-k}$$

而在本例中，$n=5000, p=0.001,$ 所以$\lambda=np=5$

P{50-3X⩾20}=P{X⩽10} $\approx \sum_{k=0}^{10} \frac{e^{-5}5^{k}}{k!}$

$$=1-\sum_{k=11}^{+\infty} \frac{e^{-5}5^{k}}{k!} \overset{\text{查表}}{=} 1-0.013695 \approx 0.986305$$

(3) Poisson 分布

若 $P\{X=k\}=e^{-\lambda}\dfrac{\lambda^k}{k!},\quad k=0,1,2,\cdots$

其中 $\lambda>0$ 是常数，则称 **X** 服从参数为 λ 的**Poisson 分布**. 记作 $X\sim\pi(\lambda)$

$$e^{-\lambda}\frac{\lambda^k}{k!}>0$$

$$\sum_{k=0}^{\infty}e^{-\lambda}\frac{\lambda^k}{k!}=e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=e^{-\lambda}\left(1+\lambda+\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}+\cdots\right)$$

$$=e^{-\lambda}\cdot e^{\lambda}=1$$

在某个时段内：

应用场合

① 某一医院在一天内的急诊病人数;

② 某地区一个时间间隔内拨错号的电话呼唤次数;

③ 某地区一个时间间隔内发生的交通事故的次数.

④ 一匹布上的疵点个数;

⑤ 某地区一天内邮递遗失的信件数;

⑥ 在一段时间内，某操作系统发生的故障次数

⑦ 一本书一页中的印刷错误数;

…………

例7 在某一个繁忙的汽车站有大量的汽车通过，设每辆汽车在一天的某段时间内发生事故的概率为**0.0001**，在某天的该段时间内有**1000**辆汽车通过，求某天该段时间内发生事故的次数不小于**2**的概率是多少？

解、该问题可看做**n**重伯努利试验，设**X**为**n**重伯努利试验中发生事故的次数，则**X~B(1000,0.0001).**有两种方法解决。

方法一：按二项分布计算

$$
\begin{aligned}
P\{X \geq 2\} &= 1 - P\{X = 0\} - P\{X = 1\} \\
&= 1 - C_{1000}^{0} \cdot 0.0001^{0} \cdot (1 - 0.0001)^{1000} \\
&\quad - C_{1000}^{1} \cdot 0.0001^{1} \cdot (1 - 0.0001)^{999} \\
&= 0.0046748
\end{aligned}
$$

方法二：应用泊松定理

由**p=0.0001,n=1000,λ=np=0.1,**所以可以用泊松定理来近似计算。根据泊松定理，发生事故的次数不小于**2**的概率为

$$P\{X \geq 2\} \approx \sum_{k=2}^{+\infty} \frac{0.1^k}{k!} e^{-0.1} = 0.0046788$$

想一想：离散型随机变量的统计特征可以用分布律描述，非离散型的该如何描述？

如：海信彩电的寿命X是一个随机变量，对消费者来说，你是否在意

{X>5年}还是{X>5年零1分钟}