第二章随机变量及其分布

2.1 随机变量的概念

随机变量是其值随机会而定的变量。

以 X 表示掷一次骰子得到的点数, X 是一个随机变量. 它可以取 $\{1, 2, 3, 4, 5, 6\}$ 中的一个值, 但到底取那个值, 要等掷了骰子才知道.

↑Example

↓Example

一张奖券的中奖金额是一个随机变量. 它的值要等开奖以后才知道.

↑Example

↓Example

在一批产品中随机地抽出 100 个产品, 其中所含的废品数是一个随机变量. 它的值要等检查了所有抽出的产品后才知道.

↑Example ↓Example

在另外的例子中, 随机试验的结果虽然不是一个数, 但仍可用数来描述.

掷一枚硬币出现正面或反面.

↑Example ↓Example

$$X(\omega) = \begin{cases} 2, & \omega = \omega_1 \\ 1, & \omega = \omega_2, \omega_3 \\ 0, & \omega = \omega_4 \end{cases}$$

↑Example

产品被分为正品或废品.

↓Example

上面两例中的结果均可用一个取值 0,1 的随机变量来描述, 其中可以 1 代表正面或正品, 以 0 代表反面或废品.

事实上, 对任意一个事件 A, 定义

$$I_A(\omega) = \begin{cases} 1 & \omega \in A\ , \\ 0 & \text{反之}\ , \end{cases}$$

则事件 A 由随机变量 I_A 表示出来. I_A 称为事件 A 的示性函数.

随机变量是把随机试验的结果，也就是样本空间，与一组实数联系起来. 这样的处理简化了原来的概率结构. 例如某机构调查民众对一提案的态度是支持 (1) 还是反对 (0). 如果随机访问 50 人，按照古典概型，所有可能的结果有 2^{50} 个. 但是如果我们用 X 记 1 的个数来表示赞成者的人数，则 X 为一个随机变量. 它的取值范围只在

$\{0, 1, \cdots, 50\}$. 所以随机变量的引进有利于我们对所研究的问题进行准确, 简练的描述. 又由于随机变量取实值, 随机变量之间的运算就变得容易了.

令 Ω 为一个样本空间. 令 X 是定义在 Ω 上的一个实函数, 如果对 Ω 中的任意点 ω, 总存在一个实数 $X(\omega)$ 与之对应, 则称 X 为一个 (一维) 随机变量.

Definition

常见的随机变量可以分为两大类. 只取有限个或可数个值的随机变量称为**离散型随机变量**; 取连续的值且密度存在的随机变量称为**连续型随机变量**. 当然, 存在既非离散型也非连续型的随机变量. 但它们在实际中并不常见, 也不是我们这里研究的对象.

2.2 离散型随机变量

设 X 为一随机变量. 如果 X 只取有限个或可数个值, 则称 X 为一个 (一维) 离散型随机变量.

Definition

由于一个随机变量的值是由试验结果决定的, 因而是以一定的概率取值. 这个概率分布称为离散型随机变量的概率函数.

设 X 为一离散型随机变量, 其全部可能值为 $\{a_1, a_2, \ldots\}$. 则

$$p_i = P(X = a_i), \quad i = 1, 2, \ldots \tag{2.1}$$

称为 X 的概率质量函数 (probability mass function, *pmf*) 或分布律.

Definition

概率质量函数 $\{p_i,\ i=1,2,..\}$ 必须满足下列条件:

$$p_i \geq 0,\ i=1,2,....$$
$$\sum_i p_i = 1.$$

$$F(x)=P(X\leq x)=\sum_{i:p_i\leq x} P(X=a_i)=\sum_{i:p_i\leq x} p_i$$

$$P(X=a_i)=P(a_{i-1}<X\leq a_i)=F(a_i)-F(a_{i-1})$$

概率质量函数 (2.1) 指出了全部概率 1 是如何在 X 的所有可能值之间分配的. 它可以列表的形式给出:

可能值	a_1	a_2	...	a_i	...
概率	p_1	p_2	...	p_i	...

(2.2)

有时也把 (2.2) 称为随机变量 X 的分布表.

设 Ω 为一样本空间. X 为定义于其上的一个离散型随机变量, 其取值为 $x_1, x_2, \ldots$.. 令 A 为 $\{x_1, x_2, \ldots\}$ 的任意一个子集. 事件 $\{X$ 取值于 A 中$\}$ 的概率可根据概率的可加性来计算:

$$P(A) = \sum_{x \in A} P(X = x).$$

这样知道了离散型随机变量 X 的概率函数, 我们就能给出关于 X 的任何概率问题的回答.

下面我们给出常见的离散型分布. 在描述离散概率模型时, Bernoulli 试验是最早被研究且应用及其广泛的概率模型.

设一个随机试验只有两个可能结果 A 和 $\bar{A}$, 则称此试验为一 Bernoulli 试验.

Definition

设将一个可能结果为 A 和 $\bar{A}$ 的 Bernoulli 试验独立地重复 n 次，使得事件 A 每次出现的概率相同，则称此试验为 n 重 Bernoulli 试验.

Definition

下面的 0-1 分布和二项分布都是以 Bernoulli 试验为基础的.

2.2.1 0-1 分布

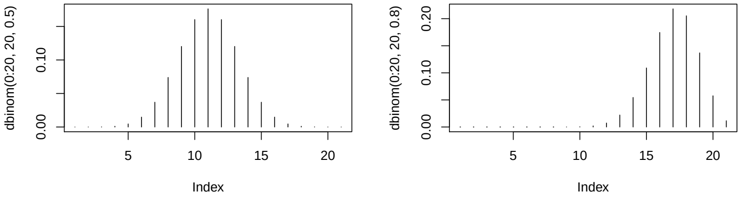
设随机变量 X 只取 0,1 两值，$P(X=1)=p$，$P(X=0)=1-p$，则称 X 服从 0-1 分布或 Bernoulli 分布. 0-1 分布是很多古典概率模型的基础.

2.2.2 二项分布

设某事件 A 在一次试验中发生的概率为 p. 现把试验独立地重复 n 次. 以 X 记 A 在这 n 次试验中发生的次数，则 X 取值 $0, 1, ..., n$，且有

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \cdots, n. \tag{2.3}$$

称 X 服从二项分布，记为 $X \sim B(n, p)$.



从

$$\sum_{i=1}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p+1-p)^n = 1,$$

我们知道 (2.3) 确实是一个概率函数.

in R

```
dbinom, rbinom, pbinom, qbinom
```

↑Code
↓Code

为了考察这个分布是如何产生的，考虑事件 $\{X = i\}$. 要使这个事件发生，必须在这 n 次试验的原始记录

$$AA\bar{A}A...\bar{A}A\bar{A}$$

中，有 i 个 A, $n-i$ 个 $\bar{A}$, 每个 A 有概率 p 而每个 $\bar{A}$ 有概率 $1-p$. 又由于每次试验独立，所以每次出现 A 与否与其它次试验的结果独立. 因此由概率乘法定理得出每个这样的原始结果序列发生的概率为

$p^i(1-p)^{n-i}$. 但是 i 个 A 和 $n-i$ 个 $\bar{A}$ 的排列总数是 $\binom{n}{k}$，所以有 i 个 A 的概率是:

$$P(X=i)=\binom{n}{i}p^i(1-p)^{n-i}, \quad i=0,1,\cdots,n.$$

由 $np_n \to \lambda > 0$，因此 $p_n \to 0$，从而

$$\begin{aligned}\binom{n}{k}p_n^k(1-p_n)^{n-k} &= \frac{1}{k!}\frac{n(n-1)\cdots(n-k+1)}{n^k}(np_n)^k(1-p_n)^n(1-p_n)^{-k} \\ &\to \frac{1}{k!}\lambda^k e^{-\lambda}\end{aligned}$$

最后是因为

$$|(1-p_n)^n-(1-\tfrac{\lambda}{n})^n| \le n|(1-p_n)-(1-\tfrac{\lambda}{n})| = |np_n-\lambda| \to 0$$

以及 $(1-\frac{\lambda}{n})^n \to e^{-\lambda}$。

一个变量服从二项分布有两个条件:

- 各次试验的条件是稳定的，这保证了事件 A 的概率 p 在各次试验中保持不变

- 各次试验的独立性

现实生活中有许多现象不同程度地满足这些条件. 例如工厂每天生产的产品. 假设每日生产 n 个产品. 若原材料质量, 机器设备, 工人操作水平等在一段时间内保持稳定, 且每件产品是否合格与其它产品合格与否并无显著性关联, 则每日的废品数服从二项分布.

2.2.3 几何分布 (Geometric distribution)

在 n 重贝努里实验中, 当试验次数 $n \to \infty$ 时, 称为可列重贝努里试验。

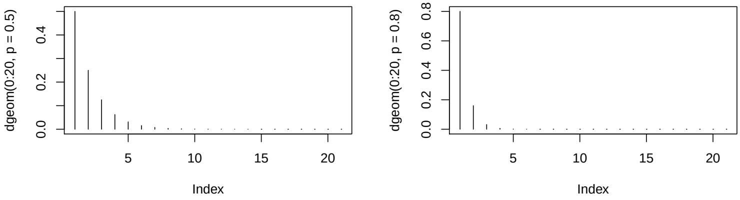
Definition

若以 X 表示在可列重贝努里试验中结果 A 出现时的试验次数, 即若以"成功"表示结果 A 发生, $p = P(A) = 1 - q$, 则 X 表示首

次成功时的试验次数，所以

$$P(X=k)=q^{k-1}p, \quad k=1,2,\cdots. \tag{2.4}$$

称此分布为几何分布. 记为 $X \sim G(p)$.



in R

```
dgeom, rgeom pgeom, qgeom
```

↑Code
↓Code

一个人要开门, 他共有 n 把钥匙。其中仅有一把可以打开门。现随机地有放回的从中选取一把开门，若不成功再放回去重新随机选取一把开门，问这人在第 S 次才首次试开成功的概率。

↑Example

↓Example

定理 1. *以所有正整数为取值集合的随机变量 ξ 服从几何分布 $G(p)$, 当且仅当对任何正整数 m 和 n, 都有*

$$P(\xi > m+n \mid \xi > m) = P(\xi > n). \tag{2.5}$$

这个性质称为几何分布的无记忆性 (memoryless property).

证:设随机变量 ξ 服从几何分布 $G(p)$, 写 $q=1-p$, 那么对任何非负整数 k, 都有

$$P(\xi > k) = \sum_{j=k+1} P(\xi = j) = p \sum_{j=k+1} q^{j-1} = q^k.$$

所以对任何正整数 m 和 n, 都有

$$P(\xi > m+n \mid \xi > m) = \frac{(\xi \geq m+n, \xi \geq m)}{P(\xi > m)}$$

$$= \frac{P(\xi \geq m+n)}{P(\xi > m)} = \frac{q^{m+n}}{q^{n}} = q^{n} = P(\xi > n).$$

知 (2.5) 式成立.

反之, 设对任何正整数 m 和 n, 都有 (2.5) 式成立. 对非负整数 k, 我们记 $p_k = P(\xi > k)$. 于是由 (2.5) 式知, 对任何正整数 k, 都有 $p_k > 0$, 并且对任何正整数 m 和 n, 都有 $p_{m+n} = p_m \cdot p_n$. 由此等式立知, 对任何正整数 m, 都有 $p_m = p_1$. 由于 $p_1 > 0$, 而若 $p_1 = 1$, 则必导致对一切正整数 m, 都有 $p_m = 1$, 此为不可能, 所以对某个小于 1 的正数 q, 有 $p_1 = q$. 由此不难得, 对任何正整数 m, 都有

$$P(\xi = m) = P(\xi > m-1) - P(\xi > m) = p_{m-1} - p_m = q^{m-1} - q^{m} = p\, q^{m-1},$$

其中 $p = 1 - q$, 所以 ξ 服从几何分布 $G(p)$.

2.2.4 Pascal 分布 (负二项分布)

在可列重贝努里试验中，若以 X_r 表示第 r 次成功发生时的试验次数，则 X_r 的分布律为

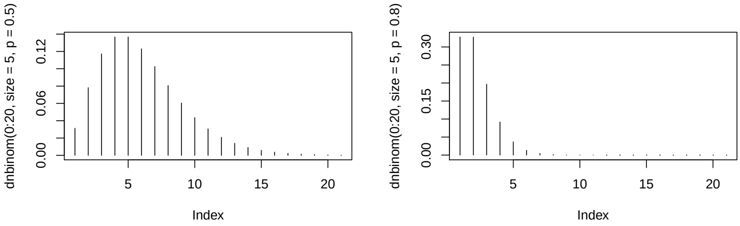
$$\begin{aligned}
P(X_r=k) &= P(\{前k-1次恰有r-1次成功且第k次成功\}) \\
&= P(\{前k-1次恰有r-1次成功\})P(\{第k次成功\}) \\
&= C_{k-1}^{r-1}p^{r-1}q^{k-r}\cdot p \\
&= C_{k-1}^{r-1}p^{r}q^{k-r}, \quad k=r,r+1,\cdots.
\end{aligned}$$

称此概率分布为 Pascal 分布。

in R

```
dnbinom, rnbinom, pnbinom, qnbinom
P(X_r = k)=dnbinom(k-r,size=r,prob=p)
```

↑Code ↓Code



如果记

$$p_k = C_{k-1}^{r-1} p^r q^{k-r}, \quad k = r, r+1, \cdots \tag{2.6}$$

那么显然有

$$\sum_{k=r}^{\infty} p_k = \sum_{k=r}^{\infty} C_{k-1}^{r-1} p^{\ r} q^{k-r} = p^{\ r} \sum_{k=0}^{\infty} C_{r+k-1}^{r-1} q^{\ k} = p^{\ r} (1-q)^{-r} = 1 \ ,$$

所以 (2.6) 式的确是一个离散型随机变量的分布律. 我们将其称为参

数为 p 和 r 的 Pascal 分布. 又因为上式表明, 它可以用负二项展开式中的各项表示, 所以又称为负二项分布.

(Banach 火柴问题) 某人口袋里放有两盒火柴, 每盒装有火柴 n 根. 他每次随机取出一盒, 并从中拿出一根火柴使用. 试求他取出一盒, 发现已空, 而此时另一盒中尚余 r 根火柴的概率.

↑Example

↓Example

解:

在可列重贝努里试验中，求事件 E =$\{n$ 次成功发生在 m 次失败之前$\}$ 的概率。

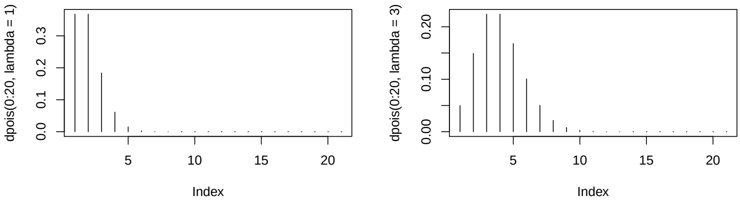
解:

2.2.5 Poisson 分布

设随机变量 X 的概率分布为

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}, \quad k=0,1,2,\cdots \quad \lambda>0, \tag{2.7}$$

则称 X 服从参数为 λ 的 Poisson 分布，并记 $X\sim P(\lambda)$.



由于 e^λ 有级数展开式

$$e^\lambda=1+\lambda+\frac{\lambda^2}{2!}+\ldots+\frac{\lambda^k}{k!}+\ldots$$

所以

$$\sum_{k=0}^{\infty} P(X=k)=1.$$

in R

```
dpois, rpois, ppois, qpois
```

↑Code
↓Code

假定体积为 V 的液体包含有一个大数目 N 的微生物. 再假定微生物没有群居的本能, 它们能够在液体的任何部分出现, 且在体积相等的部分出现的机会相同. 现在我们取体积为 D 的微量液体在显微镜下观察, 问在这微量液体中将发现 x 个微生物的概率是什么?

↑Example
↓Example

我们假定 V 远远大于 D. 由于假定了这些微生物是以一致的概率在液体中到处散布, 因此任何一个微生物在 D 中出现的概率都是

D/V. 再由于假定了微生物没有群居的本能, 所以一个微生物在 D 中的出现, 不会影响另一个微生物在 D 中的出现与否. 因此微生物中有 x 个在 D 中出现的概率就是

$$\binom{N}{x}\left(\frac{D}{V}\right)^x\left(1-\frac{D}{V}\right)^{N-x}. \tag{2.8}$$

在这里我们还假定微生物是如此之小, 拥挤的问题可以忽略不考虑, 即 N 个微生物所占据的部分对于体积 D 来说是微不足道.

在 (2.8) 中令 V 和 N 趋向于无穷, 且微生物的密度 $N/V = d$ 保持常数. 将 (2.8) 式改写成如下形式:

$$\begin{aligned}&\frac{N(N-1)(N-2)\dots(N-x+1)}{x!N^x}\left(\frac{ND}{V}\right)^x\left(1-\frac{ND}{NV}\right)^{N-x}\\ =\ &\frac{\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\dots\right)\left(1-\frac{x-1}{N}\right)(Dd)^x\left(1-\frac{Dd}{N}\right)^{N-x}}{x!}.\end{aligned}$$

当 N 变成无限时其极限为

$$e^{-Dd}(Dd)^x/x! \tag{2.9}$$

令 $Dd = \lambda$，则 (2.9) 和 (2.7) 的形式相同. 这一推导过程还证明了 λ 是 x 的平均数，因为所考察的一部分体积 D 乘以整个的密度 d 就给出了在 D 中所预计的平均数目.

当 N 很大，p 很小且 Np 趋于一个极限时，Poisson 分布是二项分布的一个很好的近似. 而在 N 未知时，Poisson 分布更显得有用. 我们有下面的定理.

定理 2. *在 n 重 Bernoulli 试验中，以 p_n 代表事件 A 在试验中出现的概率，它与试验总数 n 有关. 如果 $np_n \to \lambda$，则当 $n \to \infty$ 时，*

$$\binom{n}{k} p_n^k (1-p_n)^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda}. \tag{2.10}$$

↑Example

现在需要 100 个符合规格的元件. 从市场上买的该元件有废品率 0.01. 考虑到有废品存在, 我们准备买 $100+a$ 个元件使得从中可以挑出 100 个符合规格的元件. 我们要求在这 $100+a$ 个元件中至少有 100 个符合规格的元件的概率不小于 0.95. 问 a 至少要多大?

↓Example

解:

假设一块放射性物质在单位时间内发射出的 α 粒子数 ξ 服从参数为 λ 的 Poisson 分布。而每个发射出来的 α 粒子被记录下来的概率是 p，就是说有 $q = 1 - p$ 的概率被记数器漏记。如果各粒子是否被记数器记录是相互独立的，试求记录下来的 α 粒子数 η 的分布。

↑Example
↓Example

解:

2.2.6 离散的均匀分布

设随机变量 X 取值 $a_1, a_2, ..., a_n$, 且有

$$P(X = a_k) = \frac{1}{n}, \ k = 1, ..., n. \tag{2.11}$$

则称 X 服从离散的均匀分布.

可以看出, 离散的均匀分布正是古典概型的抽象.

in R

↑Code

```
x<-c(0,2,4,6,10)
prob<-rep(0.2,5)
plot(x,prob,type="h")
```

↓Code