目 录

第二章　随机变量及其分布

教学目的:

1) 掌握随机变量的概念。掌握离散型随机变量的概率函数, 连续型随机变量的概率密度, 及任意的随机变量的分布函数的概念.

2) 掌握二项分布、Poisson分布，以及相应的概率计算.

3) 掌握正态分布，指数分布和均匀分布，会进行相应的概率计算.

4) 掌握多维随机变量的概念。了解n维随机变量的联合分布函数的概念和性质.

5) 掌握二维离散型和连续型随机变量的边缘分布与联合分布之间的关系，会用这些关系式求边缘分布.

§2.1　随机变量的概念

随机变量是其值随机会而定的变量。

例 2.1.1. *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.*

例 2.1.2. *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.*

例 2.1.3. *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.*

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.

例 2.1.4. *掷一枚硬币出现正面或反面.*

例 2.1.5. *产品被分为正品或废品.*

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.

对于随机变量的研究，是概率论的中心内容. 因为对于一个随机试验，我们关心的通常是与所研究的问题有关的某个量或某些量. 而这些量就是随机变量.

定义 2.1.1. *令Ω为一个样本空间. 令X是定义在Ω上的一个实函数, 则称X为一个(一维)随机变量.*

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

定义 2.2.1. *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.*

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.

定义 2.2.2. *设X为一离散型随机变量，其全部可能值为$\{a_1, a_2, ...\}$. 则*

$$p_i = P(X = a_i),\ i = 1, 2, ... \tag{2.2.1}$$

称为X的概率函数.

概率函数$\{p_i,\ i=1,2,..\}$必须满足下列条件:

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

概率函数(2.2.1) 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	...

(2.2.2)

有时也把(2.2.2)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$x_1, x_2, ...$. 令A为$\{x_1, x_2, ...\}$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.

定义 2.2.3. *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.*

定义 2.2.4. *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.*

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

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}, \quad k=0,1,\cdots,n. \tag{2.2.3}$$

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(2.2.3) 确实是一个概率函数.

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：

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

一个变量服从二项分布有两个条件：一是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；二是各次试验的独立性. 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　Poisson分布

设随机变量X的概率分布为

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}, \quad k=0,1,2,\cdots,\ \lambda>0, \tag{2.2.4}$$

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

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

所以

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

穆德和格雷比尔著的《统计学导论》给出了Poisson分布的如下推导.

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？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

$$\binom{N}{x}\left(\frac{D}{V}\right)^{x}\left(1-\frac{D}{V}\right)^{N-x}. \tag{2.2.5}$$

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(2.2.5)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(2.2.5)式改写成如下形式：

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

当N变成无限时其极限为

$$e^{-Dd}(Dd)^{x}/x! \tag{2.2.6}$$

令$Dd=\lambda$，则(2.2.6)和(2.2.4)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 2.2.1. *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,*

$$\binom{n}{k}p_n^{k}(1-p_n)^{n-k}\to\frac{\lambda^{k}}{k!}e^{-\lambda}. \tag{2.2.7}$$

例 2.2.1. 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?

解: 令

$$A=\{\text{在}100+a\text{个元件中至少有}100\text{个符合规格的元件}\}.$$

假定各元件是否合格是独立的. 以X记在 $100+a$ 个元件中的废品数. 则X服从 $n=100+a$ 和 $p=0.01$ 的二项分布, 且

$$P(A)=\sum_{i=1}^{a}\binom{100+a}{i}(0.01)^i(0.99)^{100+a-i}.$$

上式中的概率很难计算. 由于 $100+a$ 较大而 0.01 较小, 且 $(100+a)(0.01)=1+0.01a\approx 1$, 我们以$\lambda=1$的Poisson分布来近似上述概率. 因而

$$P(A)=\sum_{i=1}^{a}e^{-1}/i!.$$

当$a=0,1,2,3$时, 上式右边分别为 $0.368, 0.736, 0.920$ 和 0.981. 故取 $a=3$ 已够了.

例 2.2.2. 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。

解: 以事件$\{\xi=n\}, n=0,1,2,\cdots$为划分，则由全概率公式有

$$\begin{aligned}P(\eta=k) &= \sum_{n=0}^{\infty}P(\eta=k|\xi=n)P(\xi=n)\\ &= \sum_{n=k}^{\infty}\binom{n}{k}p^kq^{n-k}\frac{\lambda^n}{n!}e^{-\lambda}\\ &= \sum_{n=k}^{\infty}\frac{(\lambda q)^{n-k}}{k!(n-k)!}e^{-\lambda}(\lambda p)^k=\frac{(\lambda p)^k}{k!}e^{-\lambda p},\quad k=0,1,2,\cdots.\#\end{aligned}$$

§2.2.4 几何分布(Geometric distribution)

定义 2.2.5. 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。

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，则X表示首次成功时的试验次数，所以

$$P(X=k)=q^{k-1}p, \quad k=1,2,\cdots. \tag{2.2.8}$$

称此分布为几何分布. 记为$X \sim G(p)$.

例 2.2.3. 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，问这人在第S次试开成功的概率。

定理 2.2.2. *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*

$$P(\xi > m+n \mid \xi > m) = P(\xi > n). \tag{2.2.9}$$

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

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

所以对任何正整数m和n,都有

$$\begin{aligned} P(\xi > m+n \mid \xi > m) &= \frac{P(\xi > m+n,\ \xi > m)}{P(\xi > m)} \\ &= \frac{P(\xi > m+n)}{P(\xi > m)} = \frac{q^{m+n}}{q^n} = q^n = P(\xi > n). \end{aligned}$$

故知(2.2.9)式成立.

反之,设对任何正整数m和n,都有(2.2.9)式成立.对非负整数k,我们记$p_k = P(\xi > k)$. 于是由(2.2.9)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$p_m = p_1^m$.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$G(p)$.

我们还可以证明几何分布是唯一的具有无记忆性的取值集合为正整数集的离散型分布.

§2.2.5 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

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

称此概率分布为Pascal分布。如果记

$$p_k=C_{k-1}^{r-1}p^rq^{k-r}, \quad k=r,r+1,\cdots \tag{2.2.10}$$

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(2.2.10)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.

例 2.2.4. **(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.

解:以A表示甲盒已空,而此时乙盒中尚余r根火柴的事件.由对称性知,所求的概率等于$2P(A)$.我们将每取出甲盒一次视为取得一次成功,以ξ表示取得第$n+1$次成功时的取盒次数,则ξ服从参数为0.5和$n+1$ 的Pascal 分布(因为每次取出甲盒的概率是0.5).易知,事件A发生,当且仅当ξ等于$2n-r+1$.所以所求的概率等于

$$2P(A)=2P(\xi=2n-r+1)=C_{2n-r}^{n}2^{\ r-2n}.$$

例 2.2.5. 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: 记F_k=$\{$第n次成功发生在第k次试验$\}$,则

$$E=\bigcup_{k=n}^{n+m-1}F_k$$

且诸F_k两两互斥，故

$$P(E)=\sum_{k=n}^{n+m-1}P(F_k)=\sum_{k=n}^{n+m-1}C_{k-1}^{n-1}p^nq^{n-k}.$$

§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

$$P(X=a_k)=\frac{1}{n},\ k=1,...,n. \tag{2.2.11}$$

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.