

Chapter 5 Random Variables on a Countable Space

Random variable

In this Chapter, we again assume Ω is countable and $\mathcal{A} = 2^\Omega$

A random variable X is defined to be a function from Ω into a set T

- A random variable represents an unknown quantity (hence the term variable) that varies not as a variable in an algebraic relation, but rather varies with the outcome of a random event. 随机变量表示未知量（因此称为术语变量），其不随代数关系的变量而变化，而是随随机结果的变化而变化的事件。
- Before the random event, we know which values X could possibly assume, but we do not know which one it will take until the random event happens. 在随机事件发生之前，我们知道 X 可能假定哪个值，但是在随机事件发生之前我们不知道它将取哪个值。

Remark

Note that even if the state space (or range space) T is not countable, the image T' of Ω under X (that is, all points $\{i\}$ in T for which there exists an $\omega \in \Omega$ such that $X(\omega) = i$) is either finite or countably infinite.

Distribution 分布

We define distribution of X (also called the law of X) on the range space T' of X by

$$P^X(A) = P(\{\omega : X(\omega) \in A\}) = P(X^{-1}(A)) = P(X \in A)$$

That this formula defined a Probability measure on T' (with the σ -algebra $2^{T'}$ of all subsets of T') is evident.

Remark

Since T' is at most countable, this probability is completely determined by the following numbers:

$$P_j^X = P(X = j) = \sum_{\{\omega: X(\omega)=j\}} p_\omega$$

Sometimes, the family $(p_j^X : j \in T')$ is also called the distribution (or the law) of X .

We have of course

$$P^X(A) = \sum_{j \in A} p_j^X$$

Define \mathcal{L}^1 to be the space of real valued random variables on (Ω, \mathcal{A}, P) which have a finite expectation, i.e.

$$\mathcal{L}^1 = \{X : X \in (\Omega, \mathcal{A}, P) \text{ and } E\{X\} < \infty\}$$

定义的 \mathcal{L}^1 是所有在概率空间 (Ω, \mathcal{A}, P) 上期望有限的随机变量的集合

- \mathcal{L}^1 is a vector space, and the expectation operator E is linear.
- the expectation operator E is positive:
If $X \in \mathcal{L}^1$ and $X \geq 0$, then $E(X) \geq 0$
If $X, Y \in \mathcal{L}^1$ and $X \leq Y$, then $E(X) \leq E(Y)$
- \mathcal{L}^1 contains all bounded random variables.
If $X \equiv a$, then $E(X) = a$
- If $X \in \mathcal{L}^1$, its expectation depends only on its distribution
If T' is the range of X ,

$$E\{X\} = \sum_{j \in T'} jP(X = j)$$

- If $X = 1_A$ is the indicator function of an event, then $E\{X\} = P(A)$

Theorem 5.1 更一般的马尔可夫 \rightarrow 找上界

Let $H : \mathbb{R} \rightarrow [0, \infty)$ be a nonnegative function 非负的函数

Let X be a real valued random variable. 实值随机变量

Then.

$$P\{\omega : h(X(\omega)) \geq a\} \leq \frac{E\{h(X)\}}{a} \quad \forall a > 0$$

Proof.

Since X is an *r.v.* so also is $Y = h(X)$; Let

$$A = Y^{-1}([a, \infty)) = \{\omega : h(X(\omega)) \geq a\} = \{h(X) \geq a\}$$

Then $h(X) \geq a1_A$, hence

$$E\{h(X)\} \geq E\{a1_A\} = aE\{1_A\} = aP(A)$$

And we have the result.

Proof_2 好一点的证明

$$\begin{aligned}
& P(\{\omega \in \Omega : h(X(\omega)) \geq a\}) \\
&= E1(\{\omega \in \Omega : h(X(\omega)) \geq a\}) \\
&= \int_{\omega \in \Omega} 1(\{\omega \in \Omega : h(X(\omega)) \geq a\}) dP \\
&= \int_{\omega \in \Omega: h(X(\omega)) \geq a} 1 dP \\
&\leq \int_{\omega \in \Omega: h(X(\omega)) \geq a} \frac{h(X(\omega))}{a} dP \\
&\leq \int_{\omega \in \Omega} \frac{h(X(\omega))}{a} dP \\
&= \frac{E\{h(X(\omega))\}}{a}
\end{aligned}$$

概率可以用示性函数的期望来表示

Corollary 5.1 Markov's Inequality

$$P\{|X| \geq a\} \leq \frac{E\{|X|\}}{a}$$

Proof.

Take $h(x) = |x|$ in Theorem 5.1

Proof 2 尽量不用Th5.1来证明，而是用Th5.1的证明方法来证

考试的时候用这个方法

$$\begin{aligned}
P\{|X| \geq a\} &= E1(\{|X| \geq a\}) \\
&= \int_{|X| \geq a} 1 dP \\
&\leq \int_{|X| \geq a} \frac{|X|}{a} dP \\
&\leq \int \frac{|X|}{a} dP \\
&= \frac{E\{|X|\}}{a}
\end{aligned}$$

Definition 5.2 方差

Let X be a real valued random variable with X^2 in \mathcal{L}^1

The **Variance** of X is defined to be

$$\sigma^2 = \sigma_X^2 \equiv E \left\{ (X - E(X))^2 \right\}$$

The **standard deviation** of X , is the nonnegative square root of the variance.

- The primary use of the standard deviation is to report statistics in the correct (and meaningful) units.

条件: 二阶矩存在 $X^2 \in \mathcal{L}^1$, 即 $E(X^2)$ 存在

也可以写成 $X \in \mathcal{L}^2$

Remark

- $E\{X\}$ represents the expected, or average, value of X (often called the mean)
- $E\{|X - E(X)|\} = E\{|X - \mu|\}$ where $\mu = E\{X\}$, represents the average difference from the mean, and is a measure of how "spread out" the values of X are. It measures how the values vary from the mean

一阶绝对中心矩

- The variance is the average squared distance from the mean. This has the effect of diminishing small deviations from the mean and enlarging big ones.

方差放大了大偏差, 缩小了小偏差 (马太效应)

- The variance is usually easier to compute than $E\{|X - \mu|\}$, and often it has a simpler expression

方差比一阶绝对中心矩更好算, 但是方差有其缺点

统计四大天王

- **jrssb**: Journal of the Royal Statistical Society, Series B
- **aos**: Annals of Statistics
- **jasa**: Journal of the American Statistical Association
- **biometrika**

Chebyshev's Inequality

If X^2 is in \mathcal{L}^1 , then we have

$$(a). \quad P\{|X| \geq a\} \leq \frac{E\{X^2\}}{a^2} \quad \text{for } a > 0$$

$$(b). \quad P\{|X - E\{X\}| \geq a\} \leq \frac{\sigma_X^2}{a^2} \quad \text{for } a > 0$$

还是用证明定理5.1的方法来证明（略）

Examples

- X is Poisson with parameter λ . Then $E\{X\} = \lambda$
- X has the Bernoulli distribution if X takes on only two values: 0 and 1.
 X corresponds to an experiment with only two outcomes, usually called "success" and "failure"

Usually $\{X = 1\}$ corresponds to "success"

Also it is customary to call $P(\{X = 1\}) = p$ and $P(\{X = 0\}) = q = 1 - p$

Note $E\{X\} = p$

伯努利分布

- X has the Binomial distribution if P^X is the Binomial probability.

That is, for given and fixed n , X can take on the values $\{0, 1, \dots, n\}$

$$P(\{X = k\}) = C_n^k p^k (1-p)^{n-k} \quad \text{where } 0 \leq p \leq 1 \text{ is fixed}$$

Suppose we perform a success/failure experiment n times independently. Let

$$Y_i = \begin{cases} 1 & \text{if success on the } i^{\text{th}} \text{ trial} \\ 0 & \text{if failure on the } i^{\text{th}} \text{ trial} \end{cases}$$

Then $X = Y_1 + \dots + Y_n$ has the Binomial distribution.

That is, a Binomial random variable is the sum of n Bernoulli random variables.

Therefore

$$E\{X\} = E\left\{\sum_{i=1}^n Y_i\right\} = \sum_{i=1}^n E\{Y_i\} = np$$

Note that we could also have computed $E\{X\}$ combinatorially by using the definition, but this would have been an unpleasant calculation.

二项分布, 随机变量 X 是成功次数

- - We are performing repeated independent Bernoulli trials.
 If instead of having a fixed number n of trials to be chosen in advance, suppose we keep performing trials until we have achieved a given number of success.
 注意: 这里的随机变量 X 代表第一次成功时, 失败的次数, 这样设置的好处可以从负二项分布看出来
 - Let X denote the number of failures before we reach a success
 X has a Geometric distribution, with parameter $1 - p$:

$$P(X = k) = (1-p)^k p, \quad k = 0, 1, 2, \dots \quad \text{注意这里的 } k \text{ 从 } 0 \text{ 开始计数}$$

Where p is the probability of success.

$$\text{We then have } E\{X\} = \frac{1-p}{p} = \frac{1}{p} - 1$$

含义: 若成功的概率为 0.1, 则一般做 10 次试验才成功, 即做 9 次失败的试验后会成功一次。

- - In the same framework as Geometric distribution, if we continue independent Bernoulli trials until we achieve the r^{th} success, then we have Pascal's distribution, also known as the Negative Binomial distribution.
 - We say X has the Negative Binomial distribution with parameters r and p if

$$P(X = j) = C_{j+r-1}^{r-1} p^r (1-p)^j \quad \text{for } j = 0, 1, \dots$$

- X represents the number of failures that must be observed before r success are observed

随机变量 X 代表 r 次成功时, 失败的总次数, 这样的设随机变量的好处是可以将其分解为两次相邻的成功之间的失败次数。

- If one is interested in the total number of trials required, call that $r.v.$ Y , then $Y = X + r$
- Note that if X is Negative Binomial, then

$$X = \sum_{i=1}^r Z_i$$

Where Z_i are geometric random variables with parameter $1 - p$. Therefore

$$E\{X\} = \sum_{i=1}^r E\{Z_i\} = \frac{r(1-p)}{p} = r \times \frac{1-p}{p}$$

i. i. d : independently identically distribution

- ◦ A distribution common in the social sciences is the Pareto distribution, also known as the Zeta distribution.
- X takes its values in \mathbb{Z}^+ , where

$$P(X = j) = c \frac{1}{j^{\alpha+1}}, \quad j = 1, 2, 3, \dots \quad \text{for a fixed parameter } \alpha > 0$$

- The constant c is such that $c \sum_{j=1}^{\infty} \frac{1}{j^{\alpha+1}} = 1$
- The function

$$\zeta(s) = \sum_{k=1}^{\infty} \frac{1}{k^s}, \quad s > 1$$

is known as the Riemann zeta function, and it is extensively tabulated.

- $c = \frac{1}{\zeta(\alpha+1)}$
- $P(X = j) = \frac{1}{\zeta(\alpha+1)} \frac{1}{j^{\alpha+1}}$
- If $\alpha > 1$, then

$$E\{X\} = \sum_{j=1}^{\infty} j P(X = j) = \frac{1}{\zeta(\alpha+1)} \sum_{j=1}^{\infty} \frac{j}{j^{\alpha+1}} = \frac{1}{\zeta(\alpha+1)} \sum_{j=1}^{\infty} \frac{1}{j^{\alpha}} = \frac{\zeta(\alpha)}{\zeta(\alpha+1)}$$

- If the state space E of a random variable X has only a finite number of points, say n , and each point is equally likely, then X is said to have a uniform distribution.

In this case

$$P(X = j) = \frac{1}{n}, \quad j = 1, 2, \dots, n$$

then X has the Discrete Uniform distribution with parameter n

$$E\{X\} = \frac{n+1}{2}$$

Exercise 5.2 尽量不用定理来证明

Let $h : \mathbb{R} \rightarrow [0, a]$ be a nonnegative (bounded) function.

Show that for $0 \leq a \leq \alpha$

$$P\{h(X) \geq a\} \geq \frac{E\{h(X) - a\}}{\alpha - a}$$

(补)

Proof.

$$\begin{aligned} P\{h(X) \geq a\} &\geq \frac{E\{h(X) - a\}}{\alpha - a} \\ \Leftrightarrow P\{h(X) < a\} &\leq \frac{\alpha - E\{h(X)\}}{\alpha - a} \\ \Leftrightarrow P\{\alpha - h(X) > \alpha - a\} &\leq \frac{\alpha - E\{h(X)\}}{\alpha - a} \end{aligned}$$

Exercise 5.9-5.10

Let X be Poisson(λ),

- What value of j maximizes $P(X = j)$?

Solve

$$\frac{P(X = j)}{P(X = j + 1)} = \dots = \frac{j + 1}{\lambda}$$

由这个式子可以看出来，当 j 取 λ 前面的整数时可以取到最大值

- For fixed $j > 0$, what value of λ maximizes $P(X = j)$?

Solve

$$P\{X = j\} = \frac{\lambda^j}{j!} e^{-\lambda} \quad \text{看作是关于 } \lambda \text{ 的一个函数}$$

令其为 $f(\lambda) = \frac{\lambda^j}{j!} e^{-\lambda}$

$$\begin{aligned} f'(\lambda) = \dots = \frac{1}{j!} (\dots) &= \frac{e^{-\lambda} \lambda^{j-1}}{j!} (j - \lambda) = 0 \\ &\Downarrow \\ &\lambda = j \end{aligned}$$

分析: $\begin{matrix} \lambda > j & f'(\lambda) < 0 \\ \lambda < j & f'(\lambda) > 0 \end{matrix} \Rightarrow f(\lambda) \text{ 在 } \lambda = j \text{ 时取最大值}$

Exercise 5.11

Let X be Poisson(λ) with λ a positive integer.

Show

$$E\{|X - \lambda|\} = \frac{2\lambda^\lambda e^{-\lambda}}{(\lambda - 1)!}$$

Proof.

Obviously

$$E\{|X - \lambda|\} = \sum_{k=0}^{\lambda} (\lambda - k) \frac{\lambda^k}{k!} e^{-\lambda} + \sum_{k=\lambda+1}^{\infty} (k - \lambda) \frac{\lambda^k}{k!} e^{-\lambda} \triangleq l_1 + l_2$$

On the other hand

$$\begin{aligned} l_1 &= \lambda \sum_{k=0}^{\lambda} \frac{\lambda^k}{k!} e^{-\lambda} - \sum_{k=1}^{\lambda} \frac{\lambda^k}{(k-1)!} e^{-\lambda} = \lambda \sum_{k=0}^{\lambda} \frac{\lambda^k}{k!} e^{-\lambda} - \lambda \sum_{k=0}^{\lambda-1} \frac{\lambda^k}{k!} e^{-\lambda} \\ l_2 &= \sum_{k=\lambda+1}^{\infty} \frac{\lambda^k}{(k-1)!} e^{-\lambda} - \sum_{k=\lambda+1}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \sum_{k=\lambda}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} - \lambda \sum_{k=\lambda+1}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \end{aligned}$$

Combining l_1 and l_2 , we get the result.

对于 l_2 , 要保证其绝对收敛。

Exercise 5.13

Let X_n be Binomial $B(n, p)$ with $0 < p < 1$ fixed.

Show that for any fixed $b > 0$, $P(X_n \leq b)$ tends to 0

Proof.

One can have

$$\begin{aligned} P(X_n \leq b) &= \sum_{i=0}^{[b]} C_n^i p^i (1-p)^{n-i} \\ &\leq \sum_{i=1}^{[b]} n^i (1-p)^{n-[b]} \\ &\leq C_n^{[b]} (1-p)^n \\ &\rightarrow 0 \quad \text{as } n \rightarrow \infty \end{aligned}$$

Since $0 < p < 1$ and b are fixed real numbers.

补充:

$$\begin{aligned} C_n^i &= \frac{n(n-1)(n-2)\cdots(n-i+1)}{i!} < n^i \\ p^i &< 1 \\ (1-p)^{n-i} &< (1-p)^{n-[b]} \end{aligned}$$

