MTH102 Engineering Mathematics II

Lesson 6: Commonly used discrete random variables

Term: 2024

Outline

Discrete random variables

Let X be a discrete random variable which takes the values $x_1, x_2, \ldots$, and the pmf is $p(x)$. Then

X	x_1	x_2	$\cdots$	x_i	$\cdots$
$P(X=x)$	$p(x_1)$	$p(x_2)$	$\cdots$	$p(x_i)$	$\cdots$

The cdf $F(x)$ is defined as

$$F(x) = P(X \le x) = \sum_{x_i : x_i \le x} p(x_i).$$

The mean and variance of X are defined as

$$E(X) = \sum_{i=1}^{\infty} x_i p(x_i), \;\; Var(X) = E[(X - E(X))^2].$$

Example 1

Let X have a uniform distribution on the first m positive integers. Then the pmf of X is

$$p(x) = \frac{1}{m}, \; x = 1, 2, \ldots, m.$$

The mean of X is

$$\mu = E[X] = \sum_{x=1}^{m} x \cdot \frac{1}{m} = \frac{1}{m} \cdot \frac{m(m+1)}{2} = \frac{m+1}{2}.$$

To compute the variance of X, we first compute

$$E[X^2] = \sum_{x=1}^{m} x^2 \cdot \frac{1}{m} = \frac{1}{m} \cdot \frac{m(m+1)(2m+1)}{6} = \frac{(m+1)(2m+1)}{6}.$$

Thus, the variance of X is

$$\sigma^2 = E[X^2] - \mu^2 = \frac{(m+1)(2m+1)}{6} - \left(\frac{m+1}{2}\right)^2 = \frac{m^2-1}{12}.$$

Outline

Bernoulli distribution

- A **Bernoulli experiment** is a random experiment with two outcomes, modeled with the sample space

$$S = \{0, 1\}.$$

- Let X be a random variable associated with a Bernoulli experiment with

$$P(X = 1) = p, \; P(X = 0) = 1 - p,$$

for some $0 \le p \le 1$. The pmf of X can be written as

$$p(x) = p^x (1 - p)^{1-x}, \; x = 0, 1.$$

- We say that X has a **Bernoulli distribution**, and

$$\mu = E[X] = (0)(1 - p) + (1)(p) = p,$$

$$\sigma^2 = Var(X) = (0 - p)^2 (1 - p) + (1 - p)^2 p = p(1 - p).$$

Example 2

Consider an experiment consists of tossing a coin, and let the random variable X be defined as

$$X = \begin{cases} 1, & \text{if it is a head,} \\ 0, & \text{if it is a tail.} \end{cases}$$

If it is a fair coin, then a head is as likely to appear as a tail. Therefore

$$P(X=1) = \frac{1}{2}, \; P(X=0) = \frac{1}{2}.$$

On the other hand, if the coin is biased and we feel that a head is twice as likely to appear as a tail, then we have

$$P(X=1) = \frac{2}{3}, \; P(X=0) = \frac{1}{3}.$$

In both cases, X has a Bernoulli distribution with different values of p.

Binomial distribution

- A Bernoulli experiment is performed n times independently, and let the random variable X be the number of times when the outcome is 1 in the n trials.
- The support (range) of X is $\{0, 1, \ldots, n\}$.
- The pmf of X is

$$p(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \ \ k = 0, 1, \ldots, n.$$

- X is said to have a **binomial distribution**, which is denoted by the symbol $b(n, p)$. The constants n and p are called the **parameters** of the binomial distribution.
- A Bernoulli distribution is just a binomial distribution with parameters $(1, p)$.

Binomial distribution

If a random variable X has a binomial distribution with parameters (n, p), then

$$E(X) = np, \ \ Var(X) = np(1-p).$$

The mean and variance can be computed in two ways.

- Direct computation with series:

$$E(X) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = np \sum_{k=1}^{n-1} \binom{n-1}{k-1} p^{k-1}(1-p)^{n-k} = np,$$

$$Var(X) = E(X^2) - [E(X)]^2 = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} - (np)^2 = np(1-p).$$

- For $i = 1, 2, \ldots, n$, let X_i be the outcome of the i-th trial. Then X_i has a Bernoulli distribution ($E(X_i) = p, \ \ Var(X_i) = p(1-p)$) and

$$X = X_1 + X_2 + \cdots + X_n.$$

Therefore, $\mathbb{E}(X) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n) = np$, $Var(X) = Var(X_1) + \cdots + Var(X_n)$
$= np(1-p)$

$$E(X) = E(X_1) + \cdots + E(X_n) = np, \ \ Var(X) = Var(X_1) + \cdots + Var(X_n) = np(1-p)$$

Example 3

Suppose that 1 out of 10 items produced by a process is defective. Select 5 items independently from the production line and test them. Let X denote the number of defective items among the 5 items. Then X follows $b(5, 0.1)$. Furthermore,

$$E[X] = 5 \cdot 0.1 = 0.5, \; Var(X) = 5 \cdot 0.1 \cdot (1 - 0.1) = 0.45.$$

The probability of observing at most one defective item is

$$\begin{aligned} P(X \leq 1) &= P(X = 0) + P(X = 1) \\ &= \binom{5}{0}(0.1)^0(0.9)^5 + \binom{5}{1}(0.1)^1(0.9)^4 \\ &= 0.9185. \end{aligned}$$

Example 4

Suppose that a particular trait of a person is classified on the basis of one pair of genes, and suppose also that d represents a dominant gene and r a recessive gene. Thus, a person with dd genes is purely dominant, one with rr is purely recessive, and one with dr is hybrid. The purely dominant and the hybrid individuals are alike in appearance. Children receive 1 gene from each parent. If, with respect to a particular trait, 2 hybrid parents have a total of 4 children, what is the probability that 3 of the 4 children have the outward appearance of the dominant gene?

mother, father
dr, dr $\longrightarrow$ dominant gene alike child
dd or dr

$\because$ "success rate" $= p = \mathbb{P}(\{\text{dominant gene alike for each child}\})$

$$= \frac{1}{2}\times\frac{1}{2} + \frac{1}{2}\times\frac{1}{2}\times 2 = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$$

Sol: Let X be the number of children having outward appearance of the dominant gene among 4 children. Then $X \in \{0, 1, 2, 3, 4\}$.

Since each child may inherit the appearance of the dominant gene independently with probability $p = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} \times 2 = \frac{3}{4}$ from their parents.

Therefore X is a Binomial r.v. with parameter $(4, 3/4)$. So we have

$\mathbb{P}$(3 of the 4 children have the outward appearance of the dominant gene) $= \mathbb{P}(X = 3)$

$$= \binom{4}{3} \left(\frac{3}{4}\right)^3 \left(1 - \frac{3}{4}\right)^{4-3}$$

$$= 4 \times \left(\frac{3}{4}\right)^3 \times \frac{1}{4}$$

$$= \frac{27}{64}$$

Exercise

(Monte Carlo method)

Suppose that 2000 points are selected independently and at random from the unit square

$$\{(x, y) : 0 < x < 1, 0 < y < 1\}.$$

Let X equal the number of points that fall into the unit circle

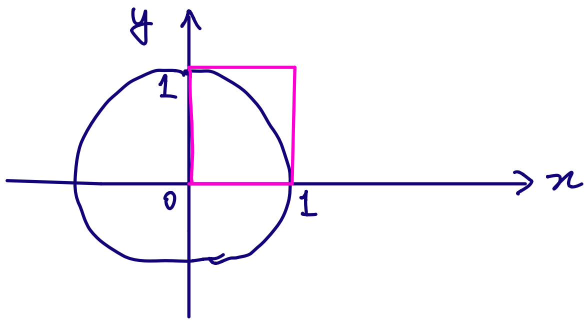
$$\{(x, y) : x^2 + y^2 < 1\}.$$

(a) Give the mean, variance and standard deviation of X.

(b) The built-in function "rand" in Matlab can generate a number in $[0, 1]$ at random. Can we use the "rand" function to find an estimate for the number π?

Sol :

(a) $p = \dfrac{\text{area of the quater circle}}{\text{area of the unit square}}$



$= \dfrac{\frac{1}{4}\pi \times 1^2}{1^2} = \frac{1}{4}\pi$

X follows Binomial distribution with parameter $(2000, \frac{\pi}{4})$.

Therefore,

$\mathbb{E}(X) = np = 2000 \times \frac{\pi}{4} = 500\pi$

$\mathrm{Var}(X) = np(1-p) = 2000 \times \frac{\pi}{4} \times (1 - \frac{\pi}{4}) = 125\pi(4-\pi)$

(b) Yes, from part (a),

$\pi = \dfrac{\mathbb{E}(X)}{500} \approx \dfrac{\text{Sample mean}}{500}$

Outline

1 Binomial distribution

2 Geometric distribution

3 Poisson distribution

Geometric distribution

- Motivation: we are interested in X the number of independent trials needed in order to get the first success. The probability of success in each trial is constantly $p \in (0, 1)$.
- The support (range) of X is $\{1, 2, 3, \dots\}$.
- A random variable X is said to have a **geometric distribution** if the pmf of X is defined by

$$p(k) = P(X = k) = q^{k-1}p, \;\; k = 1, 2, 3, \dots,$$

where $0 < p < 1$, $q = 1 - p$.

Geometric distribution

- The cdf of a geometric random variable X is

$$P(X \le k) = \sum_{i=1}^{k} q^{i-1} p = 1 - q^k, \ \ k = 1, 2, \ldots.$$

- The mean of X is

$$E(X) = \frac{1}{p}.$$

- The variance of X is

$$Var(X) = \frac{1-p}{p^2}.$$

Example 5

A box contains N white and M black balls. Balls are selected at random, one at a time and replaced, until a black one is obtained. What is the probability that

1. exactly k draws are needed?
2. at least k draws are needed?

Sol: Let X be the number of draws need. Then X is a Geometric r.v. with parameter $p = \frac{M}{N+M}$.

① $\mathbb{P}(X=k) = \left(1-\frac{M}{N+M}\right)^{k-1} \frac{M}{N+M} = \left(\frac{N}{N+M}\right)^{k-1} \frac{M}{N+M}$

② $\mathbb{P}(X \geq k) = 1 - \mathbb{P}(X \leq k-1) = 1 - \sum_{i=1}^{k-1} \mathbb{P}(X=i)$

$= 1 - \sum_{i=1}^{k-1} \left(\frac{N}{N+M}\right)^{i-1} \frac{M}{N+M} = 1 - \frac{M}{N+M} \sum_{i=1}^{k-1} \left(\frac{N}{N+M}\right)^{i-1} = 1 - \frac{M}{N+M} \cdot \frac{1-\left(\frac{N}{M+N}\right)^{k-1}}{1-\frac{N}{M+N}}$

$$= 1 - \frac{M\left[1-\left(\frac{N}{M+N}\right)^{k-1}\right]}{M}$$

$$= 1 - \left[1-\left(\frac{N}{M+N}\right)^{k-1}\right]$$

$$= \left(\frac{N}{N+M}\right)^{k-1}$$

Example 6

Problem. Finding a four-leaf clover is one of the luckiest things that can happen to a person. Statistically, only 1 in 10000 classic clovers has the coveted 4 leaves. How many clovers in average we need to check to find a four-leaf one?

Solution. Let X be the number of clovers we have checked until one four-leaf clover is discovered. Then for $k = 1, 2, \ldots,$

$$P(X = k) = (0.9999)^{k-1}(0.0001).$$

Therefore, X follows a geometric distribution and thus

$$E[X] = \frac{1}{0.0001} = 10000.$$

Outline

Motivation

The **Poisson distribution** is a discrete probability distribution that gives the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known average rate.

Some examples of random variables that generally obey the Poisson distribution are as follows:

1. The number of telephone calls received in a customer service center each day.
2. The number of packages of biscuits sold in a supermarket each day.
3. The number of occurrences of the DNA sequence "ACGT" in a gene.
4. The number of meteors greater than 1 meter diameter that strike the earth per year.
5. The number of students visiting MB523A during the office hour each Thursday (not before the exams).

Poisson distribution

Roughly speaking, the Poisson distribution is appropriate if the following assumptions are true:

- X is the number of times an event occurs in an interval and X can take values $0, 1, 2, \ldots$.
- Events occur randomly and independently.
- The rate at which events occur is constant. The rate cannot be higher in some intervals and lower in other intervals.
- The probability of an event in an interval is proportional to the length of the interval.

Poisson distribution

We said that the random variable X has a **Poisson distribution** if its pmf is of the form

$$p(k) = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \; k = 0, 1, 2, \ldots,$$

where $\lambda > 0$.

The mean and variance of X are

$$E(X) = \lambda, \; Var(X) = \lambda.$$

Example 7

Suppose that earthquakes on an island occur at the rate of 2 per week. By modeling this as a Poisson distribution, find

(a) the probability of 3 earthquakes in the next week;

(b) the probability that at least 2 earthquakes occur during the next week.

Sol: Let X be the number of earthquakes on an island within a week. X is a Poisson r.v. with parameter $\lambda = 2$. And $X \in \{0, 1, \cdots\}$

(a) $\mathbb{P}(X=3) = p(3) = e^{-2} \frac{\lambda^3}{3!} = e^{-2} \frac{2^3}{3!} = \frac{4}{3} e^{-2}$

(b) $\mathbb{P}(X \geq 2) = 1 - \mathbb{P}(X \leq 1) = 1 - \mathbb{P}(X=0) - \mathbb{P}(X=1)$

$$= 1 - e^{-2} \frac{2^0}{0!} - e^{-2} \frac{2^1}{1!}$$

$$= 1 - e^{-2}(1+2)$$

$$= 1 - 3e^{-2}$$

Example 8

Suppose that the number of typographical errors on a single page of this book has a Poisson distribution with parameter $\lambda = \frac{1}{2}$.

(a) Find the probability that there is at least one error on one page.

(b) Find the probability that there are two errors on two pages.

Sol: Let X be the number of typos on a single page of this book. Then $X \in \{0, 1, 2, \cdots\}$
X is a Poisson r.v. with parameter $\lambda = 1/2$.

(a) $\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X=0) = 1 - p(0)$

$$= 1 - e^{-1/2} \frac{(1/2)^0}{0!}$$

$$= 1 - e^{-1/2}$$

(b) Now consider number of typos on 2 pages, the rate become $2\lambda = 1$. Let Y be the no. of typos on 2 pages, then Y is a Poisson r.v. with parameter $2\lambda = 1$.

Therefore,

$$\mathbb{P}(\text{2 errors on 2 pages}) = \mathbb{P}(Y = 2)$$

$$= e^{-2\lambda} \cdot \frac{(2\lambda)^2}{2!} = e^{-1} \cdot \frac{1^2}{2!} = \frac{1}{2} e^{-1}.$$

Exercise

The number of students visiting the library per day follows a Poisson distribution with mean λ. The probability that each student borrows books is p, and the students borrow books independently.

(a) If there are n students having visited the library on one day, find the conditional probability that there are k of them who have borrowed books.

(b) Determine the distribution of the number of students borrowing books per day.

Sol: Let N be the number of students visiting the library per day. Then N is a poisson r.v. with parameter λ and $N \in \{0, 1, \cdots\}$

(a) Given $N = n$, the number of students among n who borrow books is a r.v. following Binomial distribution with parameter (n, p).

So we have

$$\mathbb{P}(\text{no. of student borrowed books} = k \mid N = n)$$

$$= \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{for } k = 0, 1, \cdots, n$$

(b) Let M be the number of students who borrowed books, then from part (a) and law of total probability, for $k = 0, 1, \cdots n$

$$\mathbb{P}(M = k) = \sum_{n=0}^{+\infty} \mathbb{P}(M = k \mid N = n)\, \mathbb{P}(N = n)$$

$$= \sum_{n=0}^{+\infty} \binom{n}{k} p^k (1-p)^{n-k} \cdot e^{-\lambda} \frac{\lambda^n}{n!}$$