T04: Random Variables

MATH 2411 Applied Statistics

WANG Zhiwei

MATH, HKUST

2024-09-30

There are two common types of r.v.'s.

We first talk about the discrete random variable.

Discrete Random Variable

Let X be a discrete random variable:

- $p(x)$, the Probability Mass Function (PMF) of X, is the probability that event $X = x$ will occur for each x in the range of X, i.e., $p(x) = P(X = x)$.
- $F(x)$, the Cumulative Distribution Function (CDF) of X, is defined as $F(x) = P(X \leq x)$.
- $E(X)$, the Expectation of X, is defined as

$$E(X) = \sum_{x \in \mathrm{Range}(X)} [x \cdot P(x)]$$

- $\mathrm{Var}(X)$, the Variance of X, is defined as

$$\mathrm{Var}(X) = \sum_{x \in \mathrm{Range}(X)} [(x - E(X))^2 \cdot P(x)] = E((X - E(X))^2)$$

Properties of population mean and variance

When it exists, the mathematical expectation E satisfies the following properties: Suppose X, Y are random variables and a and b are two constants. Then

- $E(b) = b$
- $E(aX) = aE(X)$
- $E(aX + b) = aE(X) + b$
- $E(X + Y) = E(X) + E(Y)$

When it exists, the population variance satisfies the following properties: Suppose X and Y are random variables and a and b are two constants. Then

- $\mathrm{Var}(b) = 0$
- $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$
- $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$
- If X and Y are independent random variables, then $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

Problem 1.1

Let W be a random variable giving the number of heads minus the number of tails in four tosses of a coin. List the elements of the sample space S for the four tosses of the coin and to each sample point assign a value w of W.

Solution Let 'H' denote head and 'T' denote tail. The sample space S and the random variable W are as follows:

S	W
HHHH	4
HHHT, HHTH, HTHH, THHH	2
HHTT, HTHT, HTTH, THHT, THTH, TTHH	0
HTTT, THTT, TTHT, TTTH	-2
TTTT	-4

Problem 1.2

A coin is flipped until 2 heads occur in succession. List only those elements of the sample space that require 6 or fewer tosses. Is this a discrete sample space? Explain.

Solution Let 'H' denote head and 'T' denote tail. The possible elements are:

- HH
- THH
- TTHH, HTHH
- HTTHH, TTTTH, THTHH
- TTTTHH, HTTTHH, THTTHH, TTHTHH, HTHTHH

The sample space is discrete since it contains finite elements.

Problem 1.3

Let X be a random variable with the following probability distribution:

x	**2**	**3**	**4**
$p(x)$	$\frac{4}{15}$	$\frac{4}{15}$	$\frac{7}{15}$

a. Find Expected Value $E(X)$

b. Find Variance $\mathrm{Var}(X)$

c. Find Standard Deviation $\sigma(Y)$, where $Y = 2X - 1$ (Hint: $\sigma(X) = \sqrt{\mathrm{Var}(X)}$)

d. Find the CDF of X and plot the graph of the CDF.

Problem 1.4

Given that the cdf (cumulative distribution function) of a discrete random variable X is

$$F(x) = \begin{cases} 0, & x < 1 \\ \frac{1}{4}, & 1 \leq x < 3 \\ \frac{3}{4}, & 3 \leq x < 5 \\ 1, & x \geq 5 \end{cases}$$

i) Find the pmf (probability mass function) of X.

ii) Draw the graphs of cdf and pmf of X.

iii) Evaluate $P(X > 3)$ and $P(1.5 < X < 5)$.

iv) Let $Y = 3X - 1$, find the population mean and population variance of Y.

Some probability distributions of discrete random variables.

Bernoulli distribution

A Bernoulli random variable takes on only two values: 0 and 1, with probabilities $1-p$ and p, respectively. Its PMF is thus

$$\begin{aligned} p(1) &= p \\ p(0) &= 1-p \\ p(x) &= 0, \quad \text{if } x \neq 0 \text{ and } x \neq 1. \end{aligned}$$

An alternative and sometimes useful representation of this function is

$$p(x) = \begin{cases} p^x(1-p)^{1-x}, & \text{if } x = 0 \text{ or } x = 1 \\ 0, & \text{otherwise} \end{cases}$$

Typical examples: (If an experiment/trial has only **two possible outcomes**, it can be described by a Bernoulli r.v.)

- Tossing a coin: “head” or “tail”
- The grade of a P/F course: “Pass” or “Fail”

Questions: CDF? $E(X)$ and $Var(X)$?

Binomial distribution

- If X is the discrete r.v. of the **number of successes in n trials**, then we can use a **Binomial distribution** to characterize its random behavior.
- Here are some typical examples:
  - The number of heads in n tosses of a coin.
  - The number of defective products in a sample of n products.
  - The number of students who pass a test in a class of n students.
  - ......
- A Binomial distribution is related to a random experiment with the following features:
  - Fixed **finite** number of identical trials, say $n < \infty$.
  - Trials are **independent**.
  - Trials result in **two possible outcomes** denoted by success and failure.

Binomial distribution

- **Notation**

$$X \sim \text{Binomial}(n, p),$$

where n: number of trials; p: probability of success.

- **PMF**

$$p(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad \text{for } x = 0, 1, \ldots, n$$
$$p(x) = 0, \quad \text{otherwise.}$$

Remark: The total number of ways that x successes occur in n trials is $\binom{n}{x}$.

- **Mean and Variance**

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

Poisson distribution

The **Poisson distribution** can be used to model the **number of events occurring in a fixed interval of time or space**.

Here are some typical examples for Poisson distributions:

- The number of traffic accidents occurring on a highway in a day.
- Crashes of a computer network per week.
- The number of people joining a line in an hour.
- The number of customers arrived per day.
- The number of goals scored in a hockey game.
- The number of typos per page of an essay.
- The number of mutations on a DNA strand.
- ……

Poisson distribution

- **Notation** Let X be the number of occurrences of an event over a unit time. Then we say that X follows a Poisson distribution, denoted by

$$X \sim Poisson(\lambda),$$

where $\lambda \in (0, \infty)$ is the **rate of occurrences** of an event **per unit time (or space)** or the **average number of occurrences** of the event **per unit time (or space)**.

- **PMF**

$$p(x) = P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad \text{for } x = 0, 1, \ldots$$
$$p(x) = 0, \quad \text{otherwise.}$$

Remark $\sum_{x=0}^{\infty} P\{X = x\} = \sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \cdot \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} \cdot e^{\lambda} = 1$ (Using Maclaurin series of e^{λ}).

Problem 2.1

The probability that a patient recovers from a delicate heart operation is 0.8. What is the probability that

i) exactly 2 of the next 3 patients who have this operation survive?

Solution

Let $P(A_i) = 0.8$ for $i = 1, 2, 3$. The probability that a patient does not survive is $P(A_i^c) = 0.2$.

The probability that exactly 2 out of 3 patients survive is:

$$\binom{3}{2} \cdot (0.8)^2 \cdot (0.2) = 3 \cdot 0.64 \cdot 0.2 = 0.384$$

ii) all of the next 3 patients who have this operation survive?

Solution

The probability that all 3 patients survive is:

$$(0.8)^3 = 0.512$$

Problem 2.2

There are 8 female students and 24 male students in MATH2411 Tutorial. For the next 5 weeks, every week after class one student will be chosen randomly to invite everybody for tea. Let X be the total number of female students chosen.

i) Find the probability that the 'paying gender' sequence will be FMMFM.

Solution

$$\frac{1}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{1}{4} \cdot \frac{3}{4}$$

ii) Find the probability that the 'paying gender' sequence will be MMFFM.

Solution

$$\frac{3}{4} \cdot \frac{3}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{3}{4}$$

Problem 2.2

iii) In how many ways 2 out of the 5 places can be chosen?

Solution

$$\binom{5}{2} = 10$$

iv) What is the probability that 2 out of the 5 weeks will be paid by female students?

Solution

$$\binom{5}{2} \cdot \left(\frac{1}{4}\right)^2 \cdot \left(\frac{3}{4}\right)^3$$

Problem 2.2

v) What is the probability that no more than 1 week will be paid by female students?

Solution

$$P(X \leq 1) = P(X = 0) + P(X = 1) = \left(\frac{3}{4}\right)^5 + \binom{5}{1} \cdot \left(\frac{1}{4}\right) \cdot \left(\frac{3}{4}\right)^4$$

vi) What is the probability that at least 1 week will be paid by a male student?

Solution

$$P(X < 5) = 1 - P(X = 5) = 1 - \left(\frac{1}{4}\right)^5$$

vii) What is the mean of X?

Solution

$$E(X) = 5 \cdot \frac{1}{4}$$

Problem 2.3

In California, we are monitoring earthquakes that occur at a magnitude greater than 6.7 on the Richter scale. Using the Poisson distribution with the average rate $\lambda = 1.5$ to model the number of such earthquakes per year, we would like to find the following probability:

i) What is the probability of having 3 or more occurrences in the upcoming year?

Solution

$$P(X \geq 3) = 1 - P(X < 3) = 1 - \left(\frac{e^{-1.5} \cdot 1.5^0}{0!} + \frac{e^{-1.5} \cdot 1.5^1}{1!} + \frac{e^{-1.5} \cdot 1.5^2}{2!} \right)$$

Problem 2.3

ii) What is the probability of having 3 or more occurrences in the next 5 years?

Solution Since it covers 5 years, it is equivalent to $P(X \geq 3)$ when $X \sim \text{Poisson}(7.5)$, where the rate parameter comes from $1.5 \times 5 = 7.5$. Therefore,

$$P(X \geq 3) = 1 - P(X < 3) = 1 - \left(\frac{e^{-7.5} \cdot 7.5^0}{0!} + \frac{e^{-7.5} \cdot 7.5^1}{1!} + \frac{e^{-7.5} \cdot 7.5^2}{2!} \right)$$

iii) Given that there are 52 weeks in a year, what is the probability of going 1 week without an earthquake over 6.7?

Solution In one week, the number of such earthquakes follows $X \sim \text{Poisson}(0.0288)$, where the rate parameter comes from $\frac{1.5}{52} = 0.0288$. Then we have:

$$P(X = 0) = e^{-0.0288} = 0.9716.$$

Problem 2.3

iv) What is the probability of going 18 months without an earthquake over 6.7?

Solution In 18 months, the number of such earthquakes follows $X \sim \text{Poisson}(2.25)$, where the rate parameter comes from $1.5 \times \frac{18}{12} = 2.25$. Then we have:

$$P(X = 0) = e^{-2.25} = 0.1054.$$

v) What is the expected number of earthquakes over 6.7 in the next 4 years?

Solution In the next 4 years, the number of such earthquakes follows $X \sim \text{Poisson}(6)$. Therefore, the expected number is 6.

Problem 2.4

Let the random variable $X \sim \text{Poisson}(\lambda)$. Show that

$$E(X) = \lambda \quad \text{and} \quad Var(X) = \lambda.$$

Solution Recall that the pdf of X is

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad \text{for integers } x = 0, 1, 2, \ldots$$

Therefore, by definition, we have

$$\begin{aligned}
E(X) = &\sum_{x=0}^{\infty} x f(x) = e^{-\lambda} \sum_{x=1}^{\infty} \frac{x\lambda^x}{x!} \\
=& e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!} = \lambda e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \\
=& \lambda e^{-\lambda} \cdot e^{\lambda} = \lambda.
\end{aligned}$$

Problem 2.4

Similarly, we have

$$
\begin{aligned}
E(X^2) =& e^{-\lambda}\sum_{x=1}^{\infty}\frac{x^2\lambda^x}{x!} = e^{-\lambda}\sum_{x=1}^{\infty}\frac{x(x-1)\lambda^x}{(x-1)!} + e^{-\lambda}\sum_{x=1}^{\infty}\frac{x\lambda^x}{(x-1)!} \\
=& e^{-\lambda}\sum_{x=2}^{\infty}\frac{\lambda^x}{(x-2)!} + \lambda e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} \\
=& \lambda^2 e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} + \lambda e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = \lambda^2 + \lambda.
\end{aligned}
$$

As a result, we get

$$
Var(X) = E(X^2) - (E(X))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.
$$

Problem 2.5

[Please answer the question and then use R to check your answer, and provide your R codes.]

The probability that a patient recovers from a rare blood disease is $p = 0.004$. If 1500 people are known to have contracted this disease.

i) What is the probability that (a) at least 10 survive? (b) from 3 to 8 survive? (c) exactly 5 survive?

Solution The binomial pmf is given by:

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

where $n = 1500$ and $p = 0.004$.

(a) $P(X \geq 10) = 1 - P(X \leq 9) = 1 - \sum_{k=0}^{9} P(X = k)$.

(b) $P(3 \leq X \leq 8) = P(X \leq 8) - P(X \leq 2)$.

(c) $P(X = 5)$.

Problem 2.5

R code

Set up parameters (a) (b) (c)

```
n <- 1500
p <- 0.004
```

Problem 2.5

R code

Set up parameters (a) (b) (c)

```
## Use pmf
prob_a <- 1
for (k in 0:9) {
  prob_a <- prob_a - dbinom(x = k, size = n, prob = p)
}
prob_a
```

[1] 0.08351016

```
## Use cdf
prob_a <- 1 - pbinom(q = 9, size = n, prob = p)
prob_a
```

[1] 0.08351016

Problem 2.5

R code

Set up parameters (a) **(b)** (c)

```r
## Use pmf
prob_b <- 0
for (k in 3:8) {
  prob_b <- prob_b + dbinom(x = k, size = n, prob = p)
}
prob_b
```

[1] 0.7860399

```r
## Use cdf
prob_b <- pbinom(q = 8, size = n, prob = p) -
  pbinom(q = 2, size = n, prob = p)
prob_b
```

[1] 0.7860399

Problem 2.5

R code

Set up parameters (a) (b) (c)

```r
## Use pmf
prob_c <- dbinom(x = 5, size = n, prob = p)
prob_c
```

[1] 0.1608377

```r
## Use cdf
prob_c <- pbinom(q = 5, size = n, prob = p) -
  pbinom(q = 4, size = n, prob = p)
prob_c
```

[1] 0.1608377

The relationship between Binomial & Poisson distributions

The binomial distribution tends towards the Poisson distribution as $n \to \infty$, $p \to 0$ and np stays constant.

Poisson Limit Theorem

(Poisson approximation to Binomial Distribution)

Let p_n be a sequence of real numbers in $[0, 1]$ such that the sequence np_n converges to a finite limit λ. Then:

$$\lim_{n\to\infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{\lambda^k e^{-\lambda}}{k!}$$

The Poisson distribution with $\lambda = np$ closely approximates the binomial distribution if n is large and p is small.

Problem 2.5

ii) Use Poisson distribution to approximate the probabilities above.

Solution The Poisson probability mass function is given by:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!},$$

where the rate parameter is $\lambda = np = 1500 \times 0.004 = 6$.

Then we can replace the binomial pmf in i) with the Poisson pmf to approximate the probabilities.

Problem 2.5

R code

Set up parameters | (a) | (b) | (c)

```
n <- 1500
p <- 0.004
lambda <- n * p
```

Problem 2.5

R code

Set up parameters | (a) | (b) | (c)

```
## Use pmf
prob_a <- 1
for (k in 0:9) {
  prob_a <- prob_a - dpois(x = k, lambda = lambda)
}
prob_a
```

[1] 0.08392402

```
## Use cdf
prob_a <- 1 - ppois(q = 9, lambda = lambda)
prob_a
```

[1] 0.08392402

Problem 2.5

R code

Set up parameters (a) (b) (c)

```
## Use pmf
prob_b <- 0
for (k in 3:8) {
  prob_b <- prob_b + dpois(x = k, lambda = lambda)
}
prob_b
```

[1] 0.7852687

```
## Use cdf
prob_b <- ppois(q = 8, lambda = lambda) -
  ppois(q = 2, lambda = lambda)
prob_b
```

[1] 0.7852687

Problem 2.5

R code

Set up parameters (a) (b) (c)

```r
## Use pmf
prob_c <- dpois(x = 5, lambda = lambda)
prob_c
```

[1] 0.1606231

```r
## Use cdf
prob_c <- ppois(q = 5, lambda = lambda) -
  ppois(q = 4, lambda = lambda)
prob_c
```

[1] 0.1606231

Problem 2.6

Assume that the number of flu a person gets in a year follows a Poisson distribution with parameter $\lambda = 5$. Now a medicine to prevent flu is developed and tested. Suppose that the medicine is effective for 75% of the people, by reducing the parameter of the Poisson distribution to 3; but it does not work for the other 25% of the people. If someone takes this medicine and gets flu twice in one year, what is the probability that the medicine is effective for this person?

Thank you!

Slides created via Yihui Xie's R package **xaringan**.

Theme customized via Garrick Aden-Buie's R package **xaringanthemer**.

Tabbed panels created via Garrick Aden-Buie's R package **xaringanExtra**.

The chakra comes from remark.js, **knitr**, and R Markdown.