Probability & Data

Week 3 – Discrete Distributions

DR NICK FEWSTER-YOUNG

Discrete Distributions

- Key Discrete Distributions
  - Bernoulli
  - Geometric
  - Binomial
  - Poisson

- Calculating Probabilities using these distributions
- Simulations in R

Assignment 1

- Probability (basic questions)
- Data set – Discrete Data Types
- Use the data set and R to do exploratory analysis, and do probabilities based on these distributions and questions.
- Due in Week 6, Tuesday night!
- Submission is online.
- Plagiarism – all work is individual and should be by yourself.

Motivation

- You are working as a data scientist for a large retail company and must decide whether or not to open a new store in a particular area. You are uncertain about the potential sales of the proposed store, and you would like to obtain additional information before you make a decision on opening the store, so you collect and analyse data.
- In analsying the data, you have quantitative discrete data types and hence the plan of attack to analyse the data requires the knowledge of discrete distributions ☺

Sample Spaces

- A sample space is the collection of all possible elementary events.
- A random variable is a real-valued function that is defined on a sample space. Moreover a random variable (r.v) is a function that associates a single real number with each elementary event in a sample space.
- Examples:
  a. If we toss a coin then two events could happen, S:= {Heads, Tails}. In this case, we could assign "Heads" with the number 1 and "Tails" with the number 0.
  b. The same idea can be done with the events, "Rain tomorrow" and "No rain tomorrow".
  c. Tossing a dice, "tossing a one" -> 1, "tossing a two" -> 2, "tossing a three" -> 3, etc.
  d. Other examples of random variables with numerical values equal to the numerical values associated with the corresponding events are:
     i. The sales of a particular product,
     ii. The number of vehicles passing a certain point
     iii. The total number of points scored in an AFL game.

Discrete Random Variable

A **Discrete Random variable** is a variable which can only take on a countable number of values of an event.

For example, consider an AFL game between the Adelaide Crows and Port Adelaide, and let's select three fans randomly from the game. The simulation yields the following:

$S \coloneqq \{CPC, CCC, PPP, CPP, CCP, PCC, PPC, PCP\}$.

Let the random value, X := number of Crow fans selected and it is known that P(C) = 0.7 and P(P)=0.3.

Thus, P(X=0) = P({PPP}) = 0.3*0.3*0.3 = 0.027.

P(X=1) = P({CPP,PPC,PCP}) = 3*(0.3*0.3*0.7)= 0.189.

P(X=2) = 3*(0.7*0.7*0.3) = 0.441.

P(X=3) = 0.7*0.7*0.7 = 0.343.

Probability (Density)

We know that probability follows has a few rules! Suppose that S is the space (of all events), then if X is a discrete random variable there is a probability mass function,

$$P(X = x) = f(x) > 0 \;\; where \; x \in S$$

and

$$\sum_{x \in S} f(x) = 1 \qquad\qquad and \qquad\qquad P(X \in V) = \sum_{x \in \mathrm{V}} f(x)\,.$$

So what does $f(x)$ look like, it is a function, it can be presented:

- tabular form;
- graphical form; or
- formula.

Probability

So we have a understanding of how to calculate probability for a set of values for a random variable. A common question that naturally arises is what is the probability of the event having a number greater (less) than a particular value. Well…meet the cumulative distribution function:

The cumulative distribution function (CDF) of the random variable X has the following definition:

$$F_X(t) = P(\ X \leq t)$$

Consequently, it has a few properties which we can deduce from its definition.

- $F_X(t)$ is a positive nondecreasing function of t, for $-\infty<t<\infty$.
- If X is a discrete random variable whose minimum value is a and maximum value b then

$$F_X(a) = P(X \leq a) = P(\ X = a)$$

and $F_X(b) = P(X \leq b) = 1$.

Examples – AFL

Consider the AFL game between the Adelaide Crows and Port Adelaide again, and we selected three fans randomly from the game. The simulation yielded the following:

$$S \coloneqq \{CPC, CCC, PPP, CPP, CCP, PCC, PPC, PCP\}$$

where $P(C) = 0.7$ and $P(P) = 0.3$.

The random variable X is the number of Crow fans. We can write down the probability density function over S as follows

X	0	1	2	3
$P(X = x) = f(x)$	0.027	0.189	0.441	0.343

If you add up the probabilities then you see that it sums to 1!

Example - AFL

The CDF for this set of data is as follows as well:

t	0	1	2	3
$F_X(t) = P(X \leq t)$	0.027	0.216	0.657	1

Example – Oil

- An oil company conducts a study that indicates that an exploratory oil well should have a 20% chance of striking oil. What is the probability that the first strike comes on the third well drilled?
  - What this means is, X is the number of times until we strike oil first.

X	1	2	3	4	....	x
$P(X = x) = f(x)$	0.2	$0.2 \times 0.8 = 0.16$	$0.8^2 \times 0.2 = 0.128$	$0.8^3 \times 0.2 = 0.1024$	....	$0.8^{x-1} \times 0.2$

  - It can be shown that as we keep "missing", the probability goes to zero.

X	1	2	3	4	....	t
$F_X(t) = P(X \leq t)$	0.2	0.36	0.488	0.5904	....	$\sum_{r=1}^{t-1} 0.8^r \times 0.2$

  - It can be shown that as t goes to infinity then $F_X(t)$ goes to 1.

Expectation

What are we expecting? That is what you should be asking yourself when it comes to expectation. So let's say you do an experiment, toss two fair, six-sided die as many times as you can with each time recording the sum. ***After a long time (lots of flips)***, what would the average sum of numbers (or "**mean**") of the tosses be?

2, 10, 6, 8, 5, 6, 7, 4, 12, 9, 2, 6, 5, 8, 10, …..

The probability density functions like this:

X	2	3	4	5	6	….	12
$P(X = x) = f(x)$	1/36	2/36	3/36	4/36	5/36	….	1/36

The highest probability event here is that you roll a total of 7.

Let's look at the formula for the **Expected value.**

Expectation

Definition. If *f*(*x*) is the probability mass function of the discrete random variable *X* over S, then the **expected value** is given by

$$E(X) = \sum_{x \in S} x f(x)$$

Example – Dice sums

If we apply the formula for expectation then we get the following expected value:

$$E(X) = \sum_{x \in S} x f(x) = \frac{1}{36} + 2 \times \frac{2}{36} + 3 \times \frac{3}{36} + 4 \times \frac{4}{36} + \cdots + 12 \times \frac{1}{36} = 7.$$

Example

Consider a slot machine (pokies) where we assume a player has a 20% chance of winning $5 and an 80% chance of losing $1. The probability mass function of the random variable *X*, the amount won or lost on a single play is:

X	$5	-$2
$P(X = x) = f(x)$	0.2	0.8

By using the formula, we can see that overall long term average is that someone will

$$E(X) = 5\times 0.2 - 2\times 0.8 = -0.6.$$

This means that by playing this game thousands of times, you will in the end of the day still lose 60 cents. That is how the pokies make their money off you.

Variance

Definition. Let $u(X) = (X - \mu)^2$, the expectation of $u(X)$:

$$E\big(u(X)\big) = E\big((X-\mu)^2\big) = \sum_{x \in S} (x-\mu)^2 f(x) = E(X^2) - E(X)^2$$

is called the **variance of *X***, and is denoted as ***Var*(*X*)** or $\boldsymbol{\sigma^2}$ ("sigma-squared"). The variance of *X* can also be called the **second moment of *X* about the mean $\boldsymbol{\mu}$**.

The positive square root of the variance is called the **standard deviation of *X***, and is denoted $\boldsymbol{\sigma}$ ("sigma").

Variance

Exercises: Calculate the variance of the previous examples.

Sample means and sample variances

- In most scenarios, we collect data and build a sample of the experiment in question. When calculating the statistics, we call them sample mean and sample variance. For example,

Eg. A random sample of 10 students were reported sleeping 7, 6, 8, 4, 2, 7, 6, 7, 6, 5 hours, respectively. What is the sample mean and sample variance?

Coming soon ☺

Sample Means and variances

Definition. The **sample mean**, denoted $\bar{x}$ and read "x-bar," is the average of the n data points x_1, x_2, …, x_n:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

The sample mean summarizes the "location" or "center" of the data.

Definition. The **sample variance**, denoted s^2 and read "s-squared," summarizes the "spread" or "variation" of the data:

$$s^2 = \frac{1}{n-1}((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2) = 1/(n-1)\ \ \sum_{i=1}^{n}(x_i - \bar{x})^2$$

The **sample standard deviation**, denoted s is simply the positive square root of the sample variance. That is:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Sample means and sample variances

- In most scenarios, we collect data and build a sample of the experiment in question. When calculating the statistics, we call them sample mean and sample variance. For example,

Eg. A random sample of 10 students were reported sleeping 7, 6, 8, 4, 2, 7, 6, 7, 6, 5 hours, respectively. What is the sample mean and sample variance?

$$mean = E(X) = \frac{7+6+8+4+2+7+6+7+6+5}{10} = 5.8$$

$$Var(X) = \frac{\sum\left(X-E(X)\right)^2}{n-1} = 3.06$$

Key Discrete Distributions

- Bernoulli
- Geometric
- Binomial
- Poisson
- More ☺

Bernoulli Distribution

- The random variable X = x is called Bernoulli if:

It takes only two values: 1, 0 (success, failure).

- The parameter for the Bernoulli distribution is p, $0 < p < 1$ which is the probability of success.
- We write $X \sim$ Bernoulli(p).

Thus the probability function of p is

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

The Bernoulli distribution is used to code experiments with a dichotomous outcome.

Examples

1. A single coin toss gives either a head ($X=1$) or tail ($X=0$) ;
2. A randomly sampled component may be defective ($X=1$) or not defective ($X=0$) ;
3. You will pass this course ($X=1$) or not pass ($X=0$) ; .....

Evolving this distribution

- By itself, the Bernoulli Distribution isn't that exciting.
- But if we repeat our experiments over and over, up to a total of n times, then we can build up a distribution of Bernoulli trials.
- In statistics this is called a Binomial Distribution (we saw this earlier, yes???).
- Also, we can re visit the Geometric Distribution.

Geometric Distribution

Assume Bernoulli trials:

1) there are two possible outcomes;
2) the trials are independent; and
3) *p*, the probability of success, remains the same from trial to trial.

Let X denote the number of trials until the first success. Then, the probability mass function of X is:

$$f(x) = p(x) = P(X = x) = (1 - p)^{x-1}p$$

where $x = 1, 2, ...$ In this case, we say that X follows a **geometric distribution**.

Example – Oil (Revisited)

- An oil company conducts a study that indicates that an exploratory oil well should have a 20% chance of striking oil. What is the probability that the first strike comes on the third well drilled?
  - What this means is, X is the number of times until we strike oil first.

X	1	2	3	4	….	x
$P(X = x) = f(x)$	0.2	$0.2 \times 0.8 = 0.16$	$0.8^2 \times 0.2 = 0.128$	$0.8^3 \times 0.2 = 0.1024$	….	$0.8^{x-1} \times 0.2$

  - It can be shown that as we keep "missing", the probability goes to zero.

X	1	2	3	4	….	t
$F_X(t) = P(X \leq t)$	0.2	0.36	0.488	0.5904	….	$\sum_{r=1}^{t-1} 0.8^r \times 0.2$

  - It can be shown that as t goes to infinity then $F_X(t)$ goes to 1.

Binomial Distribution

- The random variable *X* is called Binomial if:
  - an experiment consists of n independent replications of a Bernoulli experiment
- We write $X \sim B(n, p)$, where *n* and *p* are called **parameters** of the distribution.
- The Binomial Distribution has probability function

$$f(x) = p(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

Example: in a manufacturing process of computer chips 45% of chips are defective (and must be discarded). If 1000 chips are chosen at random, the number of defective chips, X, write down the distribution.

Example

Example: In a manufacturing process of computer chips 45% of chips are defective (and must be discarded). If 1000 chips are chosen at random, the number of defective chips, X, write down the distribution.

Solution: The distribution here has parameters, n=1000, p=0.45 and thus is the B(1000,0.45) distribution.

Analyse the B(n,p)

- Consider n independent Bernoulli trials and the event {X = x} (the occurrence of x successes).

One way this occurs is to have x successes followed by n − x failures. The probability of this outcome is

$$p \times p \times p \ldots p(1-p) \times (1-p) \times (1-p) \ldots (1-p) = p^x(1-p)^{n-x}$$

Note: the x successes and n − x failures could occur in many ways:

- There are such arrangements possible;

$$\binom{n}{k}$$

Therefore the probability of exactly x successes is

$$p^x(1-p)^{n-x}$$

Thus, the Binomial distribution forms in the following way

$$f(x) = p(x) = \binom{n}{k} p^x(1-p)^{n-x}$$

Example

Example: 1% of tiles are broken during packing and shipping. Find the probability that a randomly chosen carton of 25 tiles contains more than one broken tile.

Solution: Let X be the number of broken tiles and we can write $X \sim B(25, 0.01)$. Then

$$P(X > 1) = 1 - P(X \leq 1) = 1 - P(X = 0) - P(X = 1)$$

From the binomial distribution

$$P(X = 0) = 0.7778$$
$$P(X = 1) = 0.1964$$

Hence $P(X > 1) = 1 - 0.7778 - 0.1964 = 0.0258$.

Binomial Distribution

- Consider a finite population, in which each individual either has a given attribute (success) or does not (failure).
- Example: in a batch of 200,000 tile each tile may or may not be defective.
- If n objects are selected randomly with replacement then *X* has the binomial distribution.
- If n objects are selected randomly without replacement then X does not have the binomial distribution.
- Why? The assumption of independence is violated. We would actually have a hypergeometric distribution in this case.
- However, for large N, X is approximately binomial and distributed according to p, ie the proportion of successes.

Poisson Distribution

- The random variable X is called Poisson event if:
  - the number of occurrences of some event in a fixed interval of space or time occur independently according to a constant rate.
- We write $X \sim Poisson\ (\lambda)$ where $\lambda > 0$ is the rate.
- The probability function is given by

$$f(x) = p(x) = P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}.$$

Examples

- The number of accidents at a certain plant in a given year;
- The number of phone calls arriving at a switchboard in a given hour;
- The number of defects in a length of rope (space)
- The number of plating faults on a metal sheet (space)
- The number of particles emitted from a radioactive specimen in a given time.

Example

Example: Accidents occur at a given intersection at a rate of 7.2 accidents per year. Find the probability that exactly two accidents will occur in a given four month period.

Solution: Let X be the number of accidents in the 4 month period. We can see that

$\lambda = \frac{7.2}{3} = 2.4$

Thus $X \sim Poisson(2.4)$. The probability of exactly two accidents:

$$P(X = 2) = \frac{e^{-2.4} 2.4^2}{2!} = 0.2613.$$

Example

Example: A trucking company operates a large fleet of trucks and last year had 103 breakdowns. Find the probability there are at least 2 breakdowns on any given day.

Solution: Let X be the number of breakdowns on a given day. Convert the number of breakdowns in one year to a daily breakdown rate

$$\lambda = \frac{103}{365} = 0.282$$

assuming a non-leap year. Thus $X \sim Poisson(0.282)$.

The probability of at least two breakdowns:

$$P(X \geq 2) = 1 - P(X < 2) = 1 - P(\,X = 0) - P(X = 1)$$

$$= 1 - \exp\text{^}(-0.282) \times \frac{0.282}{0!} - e\text{^}(-0.282) \times 0.282/1! = 0.03.$$

Expectations for the key distributions

If X has the Bernoulli distribution with success probability p then

$$E(X) = \sum xp(x) = 0 \times (1-p) + 1 \times p = p$$

If X has the Geometric distribution with probability p then

$$E(X) = \sum xp(x) = \sum_{x=0} xp(1-p)^{x-1} = \frac{1}{p}.$$

	Mean	Variance	Std. Dev
Binomial *B(n,p)*	np	$np(1-p)$	$\sqrt{np(1-p)}$
Poisson $Po(\lambda)$	λ	λ	$\sqrt{\lambda}$

Example (St Petersburg)

Example: A casino is offering you to play the following game. You flip an unbiased coin until it lands on heads and the casino will pay you 2^k dollars where k is the number of flips. How much are you willing to pay in order to play?

Solution: By computing the expected return value, we see that the total number of flips X is a geometric random variable with $p_X(k) = \frac{1}{2^k}$. The gain is 2^k which means that

$$E(X) = \sum 2^k \times \frac{1}{2^k} = \infty.$$

You only get to play once, therefore the money you are willing to pay is bounded. This is known as the St Petersburg paradox.

Example

Example: Accidents occur at a given intersection at a rate of 7.2 accidents per year. What is the expected number of accidents that will occur in any given four month period.

Solution: Let X be the number of accidents in the 4 month period. We model X as a Poisson variable with rate parameter

$$\lambda = 7.2/3 = 2.4$$

Q: Why is this?

A: The accident rate is an annual rate, however we want the number of accidents in a 4 month period.

We needed to convert from year to months. Thus the expected number of accidents in any 4 month period:

$$E(X) = \lambda = 2.4.$$

Approximation

The Poisson distribution can be used to approximated the Binomial distribution when the sample size, n is very LARGE and p is very small. This is very useful.

Let's assume $X \sim B(n, p)$ (and n is large and p is small). The binomial probability function is well approximated by the Poisson probability function with $\lambda = np$.

The Poisson approximation to *B(n,p)* exists when

$$n \to \infty, p \to 0;$$
$$np \to \lambda$$

Example

Example: After bricks are manufactured, they are sorted into different grades. For a certain manufacturer, 1.5% of first grade bricks contain defects. Find the probability that a pallet of 500 bricks will contain exactly 5 defective bricks.

Solution: Let X be the number of defective bricks. Then X has the binomial distribution with n = 500 and p = 0.015.

The exact probability is $P(X = 5) = \binom{500}{5}(0.015)^5(1 - 0.015)^{495}$ = 0.10923.

Using the Poisson approximation, we take λ = np = 7.5 and

$$P(X = 5\,) = \frac{e^{-7.5}7.5^5}{5!} = 0.10937.$$

Poisson Process

A sequence of discrete events in continuous time can often be modelled as a Poisson process.

The defining characteristics are:

1) Events occur continuously and independently .
2) The time intervals between successive events are exponentially distributed (more later).
3) The probability of exactly one occurrence in a sufficiently short interval [t; t + h) is approximately h for small h .
4) The probability of more than one occurrence in [t; t + h) is essentially zero.

$X \sim Po(T)$ is the total number of occurrences in a time/space interval [0;T).

The Poisson process is a continuous-time process – its discrete-time counterpart is the Bernoulli process.

Processes

Let's consider a contrived Bernoulli example.

We own a new store that has a special entry policy. The door to the store will open once every minute exactly on the minute (allows for discretisation of arrivals) and will allow at most one customer to enter, that is, either a customer enters, or a customer does not enter the store each time the door opens (every minute).

- A customer arrives with probability p.
- Each customer entering the store is an independent arrival (e.g. no-one is shopping with friends).
- There is only two options; enters (denoted as 1) or not (denoted as 0).
- We decide to count the number of customers that enter the store over the time period [0; 31).

Processes

Discrete vs Continuous time acts how we compute probabilities.

<u>Bernoulli process:</u>

A Bernoulli process is a discrete-time stochastic process consisting of a sequence of independent identically distributed Bernoulli trials. The number of successes has the Binomial distribution.

<u>Poisson process:</u>

A Poisson process is a continuous-time stochastic process consisting of a sequence of independent identically distributed Poisson arrivals. The number of successes has the Poisson distribution.

Bounding probabilities

We introduce two inequalities that allow us to characterise the behaviour of a random variables to some extent just from knowing its mean and variance.

Theorem *(Markov's Inequality)* Let X be a nonnegative random variable. For any positive constant $a > 0$,

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

Theorem *(Chebyshev's Inequality)* For any positive constant $a > 0$, and X *is a random variable with bounded variance then*

$$P(|X - E(X)| \geq a) \leq \frac{Var(X)}{a^2}$$

Example (Take off times)

Example: You know that after you board a plane, it never takes off early from the scheduled departure time. You hear that the mean take off time is 15 minutes after the scheduled time, however you know from your experience that quite a times it is actually 20 minutes. Apply *Markov's inequality* to obtain an upper bound on the probability that the plane leaves later than 20 minutes.

Solution: Let X be the random variable that a plane takes off late. Then

$$P(X \geq 20) \leq \frac{E(X)}{20} = \frac{15}{20} = \frac{3}{4}.$$

At most three quarters of the flights take off late of more than 20 minutes.

Example

Example: You are not satisfied with your bound on the planes taking off later than 20 minutes. You do some research and find out that the standard deviation of a plane taking off late is 3 minutes. Apply *Chebyshev's inequality* to obtain an upper bound on the probability that the plane leaves later than 20 minutes.

Solution: Let X be the random variable that a plane takes off late. Therefore,

$$P(X \geq 20) = P(|X - E(X)| \geq 20 - 15) = P(|X - E(X)| \geq 5)$$

$$\leq \frac{Var(X)}{25} = \frac{9}{25}.$$

Actually, at least 36% of the flights take off later than 20 minutes.

Conditional Expectation

Conditional expectations is a useful tool for manipulating random variables. There is one common confusion though and that it is a random variable and not an expectation – I know right!

Consider a function g of two random variables X, Y. The expectation of g conditioned on the event X=x for any fixed value x can be computed using the conditional pdf of Y given X.

$$E(g(X,Y)|X=x) = \sum_{y \in S} g(x,y) p_{Y|X}(y|x)$$

Theorem on Expectation

Theorem 4.4.2 (Iterated expectation). For any random variables X and Y and any function $g: R^2 \to R$,

$$E\big(g(X,Y)\big) \coloneqq E\Big(E\big(g(X,Y)\big)\Big|\, X\Big)$$

Iterated expectation allows to obtain the expectation of quantities that depend on several quantities very easily if we have access to the marginal and conditional distributions. This theorem has a lot of applications to examples which we will now illustrate.

Exercise (Conditional Expectation)

Example 4.4.3 (Desert) Nick and Lauren model the time which cars break down as an Poisson random variable T with a parameter that depends on the state of the motor M and the state of the road R. These three quantities are represented by random variables in the same probability space. In addition, they assume that the states of the road and the car are independent, and they are modelled by an uniform distribution and that the parameter of the Poisson random variable that represents the time in hours until there is a breakdown is equal to M +R. Let us compute the expected time (rate) at which a car breaks down in the desert, i.e. the mean of T. By iterated expectation

$$E(T) = E(E(T\,|M, R) = E(M + R)$$

because T is Poisson when conditioned on M and R. Thus,

$$E(T) = E\,(M + R) = (M + R)$$

Since M, R are uniform distributions.

R – Distributions and Next Week ☺

- Let's pull up R and plot some distributions and calculate probabilities from today's workshop.

- Open R :)

- That's it folks for this week.

- Next week, we will look at bivariate distributions! That is distributions with two or more variables.