DSA 8010 - Random variables and discrete probability distributions

Random variables

Random variables

A random variable is a function that associates a real number to each outcome in a sample space.

Notation:

X (or another capital letter). A random variable.

$X = x$. The event containing all outcomes such that X is equal to x.

$\mathcal{X}$. Set of all possible values that X can take. Sometimes called the *support* of X.

Random variables

Consider the following probability experiment: roll one die and flip one coin.

Here are two random variables we could define:

1. $Y = \begin{cases} 1 & \text{if die is odd} \\ 0 & \text{if die is even} \end{cases}$
2. $X = \begin{cases} \text{number on die} & \text{if heads} \\ \text{2x number on die} & \text{if tails} \end{cases}$

Random variables

Experiment: roll one die and flip one coin.

Sample space:

$$\mathcal{S} = \{H1, H2, H3, H4, H5, H6, T1, T2, T3, T4, T5, T6\}$$

1. $$Y = \begin{cases} 1 & \text{if die is odd} \\ 0 & \text{if die is even} \end{cases}$$

   Write out $\mathcal{Y}$.

   Find $P(Y = 0)$.

Random variables

Experiment: roll one die and flip one coin.

Sample space:

$$\mathcal{S} = \{H1, H2, H3, H4, H5, H6, T1, T2, T3, T4, T5, T6\}$$

2. $X = \begin{cases} \text{number on die} & \text{if heads} \\ \text{2x number on die} & \text{if tails} \end{cases}$

   Write out $\mathcal{X}$.

   Find $P(X = 6)$.

Random variables

A random variable is a function that associates a real number to each outcome in a sample space.

Notation:

X (or another capital letter). A random variable.

$X = x$. The event containing all outcomes such that X is equal to x.

$\mathcal{X}$. Set of all possible values that X can take. Sometimes called the *support* of X.

Types of random variables

Discrete random variables take on at most a countable number of values.

Examples: X, Y from the previous example, new cases of disease, traffic fatalities in a period, number of supporters of a policy.

Continuous random variables can take on any value in a given range (uncountable number of values).

Examples: Water levels in a lake, percent deforestation, response time, miles per gallon, BMI, temperature.

Probability distribution functions

Probability distributions

The probability distribution function of a random variable is a function that determines the probability for any possible value.

Probability mass function (PMF). For discrete random variables. PMF is a formula, graph, or table that gives the probability that $X = x$ for every $x \in \mathcal{X}$.

$$P(X = x) = \begin{cases} \frac{e^{-2}2^x}{x!} & \text{if } x = 0, 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

Probability distributions

The probability distribution function of a random variable is a function that determines the probability for any possible value.

Probability mass function (PMF). For discrete random variables. PMF is a formula, graph, or table that gives the probability that $X = x$ for every $x \in \mathcal{X}$.

Probability density function (PDF). For continuous random variables. A PDF is a formula for a curve. The probability of the random variable falling in a given interval is represented by an area under the curve.

$$f(x) = \begin{cases} 3e^{-3x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Probability distributions

The probability distribution function of a random variable is a function that determines the probability for any possible value.

Probability mass function (PMF). For discrete random variables. PMF is a formula, graph, or table that gives the probability that $X = x$ for every $x \in \mathcal{X}$.

Probability density function (PDF). For continuous random variables. A PDF is a formula for a curve. The probability of the random variable falling in a given interval is represented by an area under the curve.

Cumulative distribution function (CDF). For a random variable X, its CDF is the function $F(x) = P(X \leq x)$.

Probability mass functions

A probability mass function must follow two rules:

1. $0 \leq P(X = x) \leq 1$ for all x.
2. $\sum_{x \in \mathcal{X}} P(X = x) = 1$. (The sum of all probabilities equals one.)

Probability mass function

Example 1. Make a table of the PMF of X from the coin and die experiment.

Find $P(X = 2)$ and $P(X \leq 6)$.

Probability mass function

Example 1. Make a table of the PMF of X from the coin and die experiment.

x	1	2	3	4	5	6	7	8	9	10	11	12
$P(X = x)$	$\frac{1}{12}$	$\frac{2}{12}$	$\frac{1}{12}$	$\frac{2}{12}$	$\frac{1}{12}$	$\frac{2}{12}$	$\frac{0}{12}$	$\frac{1}{12}$	$\frac{0}{12}$	$\frac{1}{12}$	$\frac{0}{12}$	$\frac{1}{12}$

Find $P(X = 2)$.

and $P(X \leq 6)$.

Probability mass function

Example 2: The following table gives the PMF of a random variable W.

w	20	21	22	23	24
$P(W = w)$	0.43	0.25	0.17	0.11	0.04

Find $P(W = 21)$.

Find $P(W > 22)$.

Probability mass function

Example 3: Assume that the random variable V has the following PMF:

$$P(V = v) = \frac{5^v e^{-5}}{v!}, \quad v = 0, 1, 2, \ldots$$

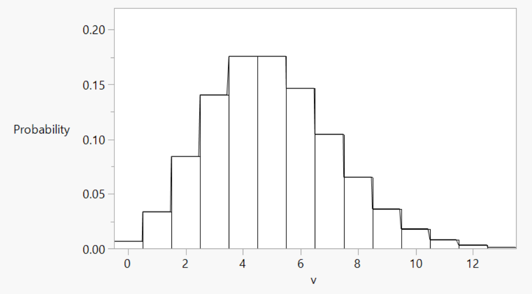
Find $P(V = 3)$.

Probability mass function

Example 3: Assume that the random variable V has the following PMF:

$$P(V = v) = \frac{5^v e^{-5}}{v!}, \quad v = 0, 1, 2, \ldots$$

Graph of the PMF:



Expectation and variance

Expected value of discrete random variable

The mean or "expected value" of a discrete random variable X is defined as

$$E(X) = \sum_{x \in \mathcal{X}} xP(X = x).$$

- Sometimes μ will be used to denote $E(X)$.
- $E(X)$ can be thought of as the average value of the random variable.

Expected value of discrete random variance

The variance of a discrete random variable X is defined as

$$\text{Var}(X) = \sum_{x \in \mathcal{X}} (x - E(X))^2 P(X = x).$$

- Sometimes σ^2 will be used to denote $\text{Var}(X)$.
- The square root of $\text{Var}(X)$ is the standard deviation of X.

Example: calculating mean and variance.

The PMF of W is given below.

w	20	21	22	23	24
$P(W = w)$	0.43	0.25	0.17	0.11	0.04

Find the mean of W.

Example: calculating mean and variance.

The PMF of W is given below.

w	20	21	22	23	24
$P(W = w)$	0.43	0.25	0.17	0.11	0.04

Find the standard deviation of W.

Popular discrete distributions

Bernoulli distribution

A Bernoulli trial is an experiment with a binary outcome. One of the two possible outcomes is (perhaps arbitrarily) labeled a "success" and the other is a "failure."

Examples: flip a fair coin; ask a yes/no question.

Bernoulli distribution

A Bernoulli trial is an experiment with a binary outcome. One of the two possible outcomes is (perhaps arbitrarily) labeled a "success" and the other is a "failure."

The Bernoulli distribution is used to model data from Bernoulli trials.

Random variable.

$$X = \begin{cases} 1 & \text{if success} \\ 0 & \text{if failure} \end{cases}$$

Parameter. $\pi =$ probability of success.

$$E(X) = \pi \text{ (expected number of successes)}$$
$$\mathrm{Var}(X) = \pi(1-\pi)$$

PMF. $P(X = x) = \pi^x (1-\pi)^{1-x}, \quad x \in \{0, 1\}.$

Bernoulli distribution: mean and variance

Show that $E(X) = \pi$ and $\mathrm{Var}(X) = \pi(1 - \pi)$ for a Bernoulli random variable.

Binomial experiment

Binomial experiments.

A **binomial experiment** has the following features.

1. n independent trials
2. each trial has two possible outcomes (success, failure)
3. probability of success, π, is constant for all trials

Example: flipping a coin

I flip a fair coin three times. This is a binomial experiment. What are n and π?

What is the sample space?

What is the probability of the outcome HTT?

Example: flipping a coin

I flip a fair coin three times. This is a binomial experiment. What are n and π?

What is the sample space?

All outcomes in the sample space are equally likely. What is the probability of one getting exactly 1 tails?

Binomial distribution

Random variable. The number of successes in n trials.

Parameters. n = number of trials and π = probability of success.

$$E(X) = n\pi \text{ (expected number of successes)}$$
$$\text{var}(X) = n\pi(1 - \pi)$$

PMF. $P(X = x) = \binom{n}{x}\pi^x(1 - \pi)^{n-x}, \quad x = 0, 1, \ldots, n.$

Binomial distribution

Random variable. The number of successes in n trials.

Parameters. n = number of trials and π = probability of success.

$$E(X) = n\pi \text{ (expected number of successes)}$$
$$\text{var}(X) = n\pi(1-\pi)$$

PMF. $P(X = x) = \binom{n}{x}\pi^x(1-\pi)^{n-x}, \quad x = 0, 1, \ldots, n.$

Is this a binomial random variable?

- Count the number of cars passing my billboard in an hour.
- Count the number of rainy days in the month of September.
- In a sample of 100 students, count the number who plan to graduate in May.

Example: Mendel's peas

Mendel showed that if green and yellow inbred lines of peas are crossed that the ratio of yellow to green peas is 3 to 1. Randomly choose 10 such offspring and check if they are yellow or green.

Is this a binomial experiment? What are n and π?

What is the mean number of yellow peas?

Example: Mendel's peas

Find $P(X = 8)$.

Example: Mendel's peas

Find the probability of at least 8 yellow offspring.

Poisson distribution

Random variable. A discrete value greater than or equal to 0. Often used to model counts.

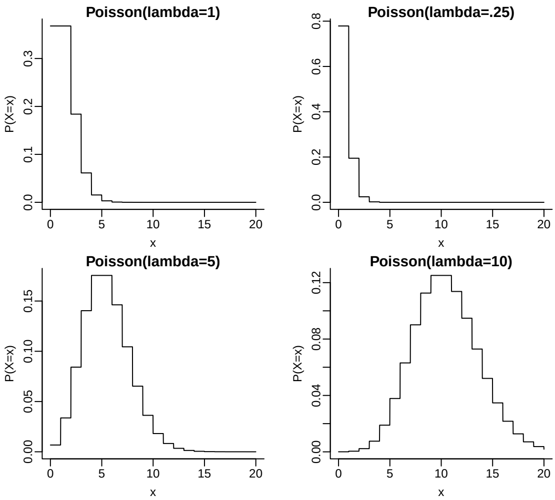
Parameters. $\lambda =$ rate parameter. $\lambda > 0$.

$E(X) = \lambda$ (expected count)
$\text{var}(X) = \lambda$

PMF. $P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots.$

Poisson distribution

Some Poisson pmfs:



Example: customer counts

Suppose the number of customers in a gift shop between 6-7pm has a Poisson distribution with rate $\lambda = 3.75$. Find the probability of 5 customers visiting the gift shop during that time.

Example: customer counts

Suppose the number of customers in a gift shop between 6-7pm has a Poisson distribution with rate $\lambda = 3.75$. Find the probability of more than 1 customer visiting the gift shop during that time.

Discrete probability distributions in R

Family	functions	parameters
Binomial	`pbinom`, `dbinom`, `qbinom`, `rbinom`	size=n, prob = probability of success.
Poisson	`ppois`, `dpois`, `qpois`, `rpois`	lambda
Normal	`pnorm`, `dnorm`, `qnorm`, `rnorm`	mean, sd=standard deviation
t	`pt`, `dt`, `qt`, `rt`	df
Chi-square	`pchisq`, `dchisq`, `qchisq`, `rchisq`	df
f	`pf`, `df`, `qf`, `rf`	df1, df2

Discrete probability distributions in R

- The functions beginning with 'p' are used to find interval probabilities. Use the `lower.tail` option to indicate whether you want $\leq$ or $>$ probabilities.

  Example (Poisson with $\lambda = 1.4$):

  - To find $P(X \leq 1)$, use `ppois(1,lambda=1.4, lower.tail=TRUE)`.
  - To find $P(X > 1)$, use `ppois(1, lambda=1.4, lower.tail=FALSE)`.

- The functions beginning with 'd' are used to find point probabilities.
- The functions beginning with 'q' are used to find percentiles.
- The functions beginning with 'r' generate random numbers from the distribution.

Example: customer counts

Suppose the number of customers in a gift shop between 6-7pm has a Poisson distribution with rate $\lambda = 3.75$. Find the probability of 5 customers visiting the gift shop during that time.

Example: customer counts

Suppose the number of customers in a gift shop between 6-7pm has a Poisson distribution with rate $\lambda = 3.75$. Find the probability of more than 1 customer visiting the gift shop during that time.

Example: Mendel's peas

Recall that the number of yellow peas has a binomial distribution with $n = 10$ and $\pi = 0.75$. Find the probability of at least 8 yellow offspring.

Example: binomial

Let Y be a binomial random variable with $\pi = 0.65$ and $n = 21$. Find the probability that $17 < Y \leq 20$.

Example: binomial

Let Y be a binomial random variable with $\pi = 0.65$ and $n = 21$. Find the probability that $17 \leq Y < 20$.