

## Probability review

Sample space - the set of all possible outcomes

eg: roll a die, sample space:  $S =$

An event is a subset of the sample space

eg: A: roll a three  $A =$

B: roll an even  $B =$

C: roll something  $C =$

$A \cup B =$  roll a three or an even =

$A \cap B =$  roll a three and an even =

empty set / non-event

complement:  $\bar{A} =$  not A not rolling a three = {1, 2, 4, 5, 6}

If  $A \cap B = \emptyset$  then A and B are mutually exclusive or disjoint events.

For a discrete (i.e. finite or countable) sample space S a probability function is a function  $f: S \rightarrow \mathbb{R}$  such that

- for each outcome  $x \in S$ ,  $0 \leq f(x) \leq 1$

-  $\sum_{x \in S} f(x) = 1$ .

eg: probability function for rolling a fair die is  $f(x) = \frac{1}{6}$

Given a probability function  $f: S \rightarrow \mathbb{R}$ , the probability of an event  $A \subset S$  is

$$P(A) = \sum_{x \in A} f(x)$$

eg: probability of rolling an even =

Properties:

$$P(\emptyset) = 0$$

$$P(\bar{A}) = 1 - P(A)$$

$$P(\bar{A} \cap B) = P(B) - P(A \cap B)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

The conditional probability of an event A occurring given that B has occurred is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \begin{array}{l} \text{when } P(B) \neq 0 \\ \text{else } P(A|B) = 0 \end{array}$$

eg: roll a 2 given an even has been rolled

$\underbrace{\text{roll a 2}}$  given  $\underbrace{\text{an even has been rolled}}$

A B

$$P(A|B) =$$

We say A is independent of B if  $P(A|B) = P(A)$

In which case  $P(A) = P(A|B) = \frac{P(A \cap B)}{P(B)}$

$$\Rightarrow P(A \cap B) =$$

eg: toss a coin twice, A = second toss tails  
B = first toss tails

A is independent of B, so  $P(A|B) =$

and the probability of consecutive tails

$$P(A \cap B) =$$

## Discrete random variables

A random variable  $X$  is a function  $X: S \rightarrow \mathbb{R}$  from a sample space to the reals.

If we have a probability function then we denote

$$P(X(s) = x) = P(\quad)$$

$x \in \mathbb{R}$        $P(X = x)$  = probability of a value  $x \in \mathbb{R}$   
for the random variable  $X: S \rightarrow \mathbb{R}$

### Example

roll a blue die and a green die

sample space  $S = \{(1,1), (1,2), (2,1), \dots, (6,6)\}$

$6 \times 6 = 36$  possible outcomes

define a random variable

$$X: S \rightarrow \mathbb{R}$$

$$(a, b) \mapsto a + b \quad \text{range } X = \{2, 3, \dots, 12\}$$

$$P(X = 2) = \quad (\text{only one way to roll: } (1,1))$$

$$P(X = 3) =$$

$$P(X = 7) =$$

The above is an example of a discrete random variable because the range of  $X: S \rightarrow \mathbb{R}$  is a discrete subset of  $\mathbb{R}$ .

The probability mass function (pmf) for a discrete random variable  $X: S \rightarrow \mathbb{R}$  is the function  $p_X: \mathbb{R} \rightarrow \mathbb{R}$

$$p_X(x) = P(X = x) = P(\{s \in S \mid X(s) = x\})$$

From the axioms defining a probability function it follows that

$$0 \leq P_X(x) \leq 1 \quad \text{and} \quad \sum_{x \in \text{range}(X)} P_X(x) = 1$$

eg: pmf for the green and blue dice example

$$\begin{array}{c|cccccc|c} x & 2 & 3 & \dots & 12 & & & \text{otherwise} \\ P_X(x) & \frac{1}{36} & \frac{1}{18} & & \frac{1}{36} & & 0 & \end{array}$$

The cumulative distribution function (cdf) for a discrete r.v.  $X$  is

$$F_X(t) = P(X \leq t) = \sum_{x \leq t} P_X(x)$$

The expected value  $E[X]$  or  $\mu_X$  is

$$E[X] = \sum_{\substack{\text{range} \\ \text{of } X}} x P_X(x) \quad (\text{mean})$$

eg: a friend proposes a game where he rolls a die and gives you  $\$(4-s)$ , where  $s$  is the number rolled. What is the expected value of this game?

$X(s)$  ← r.v. representing how much \$ you get

pmf:

$x$		
$P_X(x)$		

$$E[X] = \sum x P_X(x) =$$

=

Expected value of a function  $f$  of a discrete r.v.  $X$

$$E[f(X)] = \sum_{\substack{x \text{ in} \\ \text{range}(X)}} P_X(x) f(x)$$

Properties:  $X, Y$  discrete r.v.'s,  $c \in \mathbb{R}$ :

$$E(c) = c \quad E[cX] = cE[X] \quad E[X+Y] = E[X] + E[Y]$$

The variance of a discrete r.v.  $X$ ,  $\text{Var}(X)$  or  $\sigma_X^2$  is

$$\text{Var}(X) = E[(X - \mu_X)^2] = \sum_{\text{range}(X)} (x - \mu_X)^2 p_X(x)$$

The standard deviation is  $\sigma_X = \sqrt{\text{Var}(X)}$

Note:  $(x - \mu_X)^2$  is the squared distance between the outcome  $x$  and the mean (or expected value)  $\mu_X$ , so the variance is the expected value of squared distance from the mean.  
- a measure of spread.

Properties of variance

$$\text{Var}(X) \geq 0$$

$$\text{Var}(aX + b) = a^2 \text{Var}(X) \quad a, b \in \mathbb{R}$$

$$\text{Var}(X) = E[X^2] - (E[X])^2 \quad \leftarrow \text{very useful formula.}$$

Typical discrete r.v.s

A Bernoulli trial is any experiment with a binary sample space

$$S = \{\text{success, failure}\} \\ \text{happy, sad etc.}$$

Probabilities  $P(\text{success}) = p$

$$P(\text{failure}) = 1-p$$

Define a random variable  $Y: S \rightarrow \mathbb{R}$  by  $Y(\text{success}) =$   
 $Y(\text{failure}) =$

Then  $Y$  has pmf  $P_Y(x) = \begin{cases} & x=0 \\ & x=1 \\ & \text{else} \end{cases}$

If a random variable  $X$  has the above pmf for some  $p$   
we say it has the Bernoulli distribution  $X \sim$

The Bernoulli distribution has  $E[X] = p$  and  $\text{Var}[X] = p(1-p)$

If  $X$  is the random variable representing the number of successes in  $n$  independent Bernoulli trials with success probability  $p$  then  $X$  is a Binomial random variable,  $X \sim \text{Bin}(n, p)$

with pmf

$$p_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, 2, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

(combinations:  $\binom{n}{x} = {}^n C_x$ )

$X \sim \text{Bin}(n, p)$  then  $E[X] = np$

$$\text{Var}(X) = np(1-p)$$

## Continuous random variables

A continuous random variable is a random variable whose range is an infinite, uncountable subset of  $\mathbb{R}$ .

(eg:  $[0, 1]$  - there are infinitely many numbers here and you can't count them)

Example

What is the probability that my height is exactly 186cm?

Last time I measured it was  $\sim 186\text{cm}$ , so let's say:

$$P(185 < H < 187) =$$

Suppose I assign  $P(H = 186) =$

and  $P(H = 186.1) =$

Problem: there are infinitely many numbers between 186 and 186.1

Eg:

we can't assign all of these a non-zero probability

why not? total probability will exceed 1

(and it doesn't make sense to assign 186.1 a non-zero probability but not, say 186.025.)

Solution: work with probability densities instead.

A probability density function (pdf) is a function  $f: \mathbb{R} \rightarrow \mathbb{R}$  such that

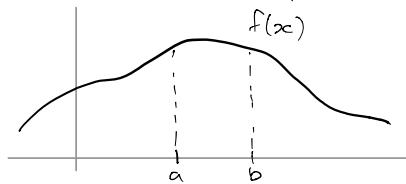
1.  $f(x) \geq 0$

2.  $\int_{-\infty}^{\infty} f(x) dx = 1$       improper integral

If  $X$  is a continuous random variable such that

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

we say  $X$  has probability density function  $f$ , denote  $f_X$   
i.e.: probabilities correspond to areas under  $f$ :



Note:  $P(X = a) =$

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

For a crv  $X$  with PDF  $f_X$ :

cumulative distribution function (cdf)

$F_X: \mathbb{R} \rightarrow [0, 1]$  defined by

$$F_X(t) = P(X \leq t) = \int_{-\infty}^t f_X(x) dx$$

expected value

$$E[X] = \mu_X =$$

compare  
- discrete versions.

variance

$$\sigma_X^2 = \text{Var}(X) = E[(X - \mu_X)^2] =$$

Properties of  $E[X]$  and  $\text{Var}[X]$  also hold for crv's.

$$E(c) = c \quad E[cX] = E[X] \quad E[X+Y] = E[X] + E[Y]$$

$$\text{Var}(X) \geq 0$$

$$\text{Var}(aX + b) = a^2 \text{Var}(X) \quad a, b \in \mathbb{R}$$

$$\text{Var}(X) = E[X^2] - (E[X])^2 \quad \leftarrow \text{very useful formula.}$$

## Some useful pdf's

The uniform distribution

Suppose  $X$  is a crv with range  $X = [a, b] \subset \mathbb{R}$

define  $f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$

$f$  is a pdf called the uniform distribution

check:  $f(x) \geq 0$ ?

$$\int_{-\infty}^{\infty} f(x) dx =$$

Example: spinning a wheel



model probability of stopping with orientation within a given range.

$$[0, 2\pi] \quad f(x) = \begin{cases} \frac{1}{2\pi} & 0 \leq x \leq 2\pi \\ 0 & \text{otherwise} \end{cases}$$

for  $a \in [0, 2\pi]$ ,  $\epsilon > 0$

$$P(a - \epsilon < X < a + \epsilon) = \int_{a-\epsilon}^{a+\epsilon} \frac{1}{2\pi} dx =$$

$$=$$

=

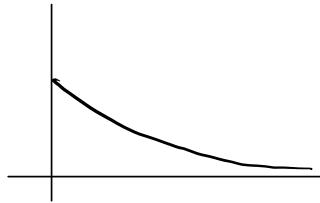
i.e. the probability of the wheel stopping in a given interval depends only on the size of the interval - not where it is hence "uniform" distribution.

## The exponential distribution

Let  $T$  be a continuous random variable which takes positive values (range  $T = [0, \infty)$ )

define, for any  $\lambda \geq 0$ ,

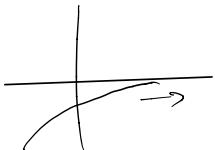
$$f_T(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$



then  $f_T$  is a pdf called the exponential distribution.

check:  $f_T(x) \geq 0$ ?

$$\int_{-\infty}^{\infty} f_T(x) dx = \int_0^{\infty} \lambda e^{-\lambda x} dx = \lim_{t \rightarrow \infty} \int_0^t \lambda e^{-\lambda x} dx$$



FOR THE EXPONENTIAL DISTRIBUTION:  $T \sim \exp(\lambda)$

cdf:  $F_T(t) = 1 - e^{-\lambda t}$

expected value  $M_T = \frac{1}{\lambda}$

variance  $\text{Var}(T) = \frac{1}{\lambda^2}$

$$G_T = \frac{1}{\lambda}$$

The exponential distribution is used to model things like wait times.

$M_T = \frac{1}{\lambda}$  so as  $\lambda$  increases the expected value for wait time decreases  $\rightarrow$  think of  $\lambda$  as the "rate" of occurrence

## The normal distribution

Let  $\mu, \sigma$  be some parameters and define

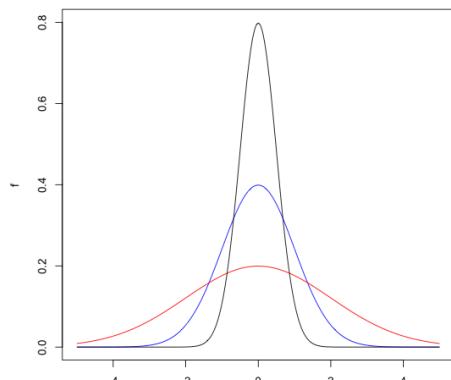
$$f_X(x) = \frac{1}{\sqrt{\sigma^2 2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

then  $f_X$  is a pdf called the normal (or Gaussian) distribution  
(won't prove this)

If  $X$  has the normal distribution we write  $X \sim N(\mu, \sigma^2)$   
and then, as the notation suggests:

$$E(X) =$$

$$\text{Var}(X) =$$



$$\begin{aligned} \mu &= 0 & \sigma &= 0.5 \\ \sigma &= 1 \\ \sigma &= 2 \end{aligned}$$

Used to model things like height in a random sample of people.

The standard normal distribution is  $N(0, 1)$  ( $\mu = 0$ ,  $\sigma = 1$ )

If  $X$  is a random variable with  $X \sim N(\mu, \sigma_x^2)$  then

$Z =$  is a rv. with the standard normal dist.

i.e.  $Z \sim N(0, 1)$ . Applying this transformation  $Z = \frac{X - \mu_x}{\sigma_x}$   
is called standardising the r.v.  $X$ .