Probability review

Sample space – the set of all possible outcomes

eg: roll a die, sample space: $S = \{1, 2, 3, 4, 5, 6\}$

An event is a subset of the sample space

eg: A : roll a three $A = \{3\} \subset S = \{1, 2, 3, 4, 5, 6\}$

B : roll an even $B = \{2, 4, 6\} \subset S$

C : roll something $C = S \subset S$

$A \cup B$ = A or B   roll a three or an even $= \{2, 3, 4, 6\}$

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

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 $f : S \to \mathbb{R}$ such that

- for each outcome $x \in S$, $0 \le f(x) \le 1$
- $\sum_{x \in S} f(x) = 1$.

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

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

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

eg: probability of rolling an even $= \sum_{x \in \{2,4,6\}} f(x) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6}$

Properties:

$P(\phi) = 0$

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

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

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

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

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \qquad \text{when } P(B) \neq 0$$
$$\text{else } P(A \mid B) = 0$$

eg: $\underbrace{\text{roll a 2}}_{A}$ given $\underbrace{\text{an even}}_{B}$ has been rolled

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{1/2} = \frac{1}{3}$$

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

In which case $P(A) = P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$

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

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

A is independent of B, so $P(A \mid B) = P(A) = \frac{1}{2}$

and the probability of consecutive tails

$$P(A \cap B) = P(A)\,P(B) = \frac{1}{4}$$

Discrete random variables

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

If we have a probability function then we denote

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

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

<u>Example</u>

roll a blue die and a green die

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

$6 \times 6 = 36$ possible outcomes

define a random variable

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

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

$P(X = 2) = \dfrac{1}{36}$ (only one way to roll: $(1,1)$)

$P(X = 3) = \dfrac{2}{36} = \dfrac{1}{18}$ $\qquad (1,2), (2,1)$

$P(X = 7) = \dfrac{6}{36} = \dfrac{1}{6}$

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

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

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

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

x	2	3	...	12	otherwise
$p_X(x)$	$1/36$	$1/18$		$1/36$	0

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 μ_X is

$$E[X] = \sum_{\substack{\text{range} \\ \text{of } X}} x \, p_X(x) \qquad \text{(mean)}$$

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?

$X(s) = 4 - s$ ← r.v. representing how much \$ you get

pmf:

x	-2	-1	0	1	2	3
$p_X(x)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

$$E[X] = \sum x \, p_X(x) = -2\left(\tfrac{1}{6}\right) - 1\left(\tfrac{1}{6}\right) + 0 + \frac{1}{6} + \frac{2}{6} + \frac{3}{6}$$
$$= \frac{1}{2}$$

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, $c \in \mathbb{R}$:

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

The variance of a discrete r.v. X, $Var(X)$ or σ_X^2 is

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

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

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

Properties of variance

$$Var(X) \geq 0$$

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

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

Standard discrete r.v.s

A Bernoulli trial is any experiment with a binary sample space

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

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

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

Define a random variable $Y: S \to \mathbb{R}$ by $Y(\text{success}) = 1$

$Y(\text{failure}) = 0$

Then Y has pmf

$$p_Y(x) = \begin{cases} 1-p & x=0 \\ p & x=1 \\ 0 & \text{else} \end{cases}$$

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

The Bernoulli distribution has $E[X] = p$ and $Var[X] = pq$

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, $X \sim Bin(n, p)$ with pmf

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

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

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 ~186cm, so let's say:

$$P(185 < H < 187) = 1$$

Suppose I assign $P(H = 186) = 1/2$

and $P(H = 186.1) = 1/4$

Problem: there are infinitely many numbers between 186 and 186.1

eg: 186.05, 186.025, 186.001, 186.0001, ...

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 $f : \mathbb{R} \to \mathbb{R}$ such that

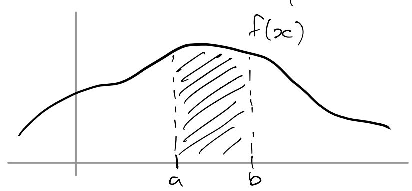
1. $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$ improper integral

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: $P(X = a) = 0$

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

FOR A CRV X WITH PDF f_X:

cumulative distribution function (cdf)

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

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

expected value

$$E[X] = \mu_X = \int_{-\infty}^{\infty} x\, f_X(x)\,dx$$

variance

$$\sigma_X^2 = Var(X) = E\left[(X-\mu_X)^2\right] = \int_{-\infty}^{\infty} (x-\mu_X)^2 f_X(x)\,dx$$

compare
- discrete versions:

$$\sum_{x \in \text{range } X} x\, p_X(x)$$

$$\sum (x-\mu_X)^2\, p_X(x)$$

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

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

$Var(X) \geq 0$

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

$Var(X) = E[X^2] - (E[X])^2$ ← very useful formula.

Some useful pdf's

The uniform distribution

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

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

f is a pdf called the uniform distribution

check: $f(x) \ge 0$? $b > a$ ✓

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_a^b \frac{1}{b-a}\,dx = \left[\frac{x}{b-a}\right]_a^b = \frac{b}{b-a} - \frac{a}{b-a} = \frac{b-a}{b-a} = 1$$

Example: spinning a wheel

model probability of stopping with orientation within a given range.

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

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

$$P(a-\varepsilon < X < a+\varepsilon) = \int_{a-\varepsilon}^{a+\varepsilon} \frac{1}{2\pi}\,dx = \left[\frac{x}{2\pi}\right]_{a-\varepsilon}^{a+\varepsilon}$$

$$= \frac{a+\varepsilon}{2\pi} - \frac{(a-\varepsilon)}{2\pi}$$

$$= \frac{\varepsilon}{\pi}$$

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

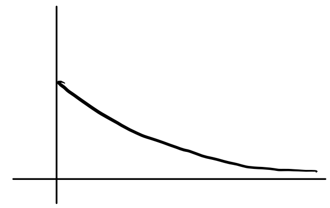
hence "uniform" distribution.

The exponential distribution

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

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

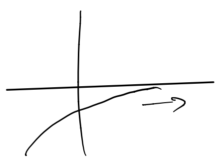
$$f_T(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$



then f_T is a pdf called the exponential distribution.

check : $f_T(x) \geq 0$? ✓ $\lambda \geq 0$, $e^a > 0$

$$\int_{-\infty}^{\infty} f_T(x) = \int_0^{\infty} \lambda e^{-\lambda x} dx = \lim_{t \to \infty} \int_0^t \lambda e^{-\lambda x} dx = \lim_{t \to \infty} \left[-e^{-\lambda x} \right]_0^t$$



$$= \lim_{t \to \infty} \left(-e^{-\lambda t} + 1 \right) = 0 + 1 = 1$$

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

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

expected value $\mu_T = 1/\lambda$

variance $Var(T) = 1/\lambda^2$

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

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

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

The normal distribution

Let μ, σ be some parameters and define

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

then f_X is a pdf called the normal (or Gaussian) distribution

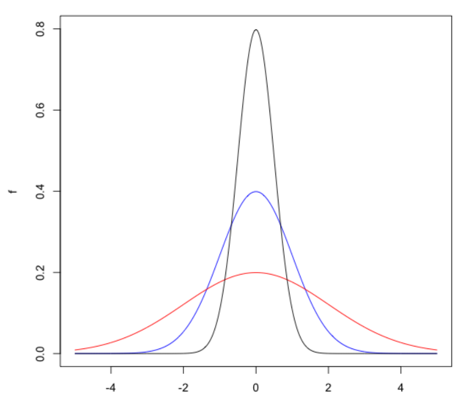
(won't prove this)

If X has the normal distribution we write $X \sim N(\mu, \sigma^2)$

and then, as the notation suggests:

$$E(X) = \mu$$

$$Var(X) = \sigma^2$$



$\mu = 0$ $\sigma = 0.5$

$\sigma = 1$

$\sigma = 2$

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

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

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

$Z = \dfrac{X - \mu_X}{\sigma_X}$ is a r.v. with the standard normal dist.

i.e. $Z \sim N(0,1)$. Applying this transformation $Z = \dfrac{X - \mu_X}{\sigma_X}$

is called standardising the r.v. X.