- Random variables

- Probability mass/density functions

- Cumulative mass/density functions

- Expectations, moments

A random variable is…

- a variable, and it is random…

A random variable is…

- a variable who takes on its value by chance. A random variable can take on a set of possible values, each with an associated probability.

- To fully characterise a random variable (r.v.) we need to know:
  - all its possible outcomes (domain/support)
  - the probability of hitting each particular outcome (probability mass/density function)

- Let X be the outcome from tossing a fair coin.
  - X is a random variable
  - two possible outcomes: $\{head, tail\}$
  - $\Pr(X = head) = 0.5, \Pr(X = tail) = 0.5$

- Let X be the outcome from rolling a fair die.
  - six possible outcomes: $\{1, 2, 3, 4, 5, 6\}$
  - $\Pr(X = 1) = 1/6, \Pr(X = 2) = 1/6, \ldots$

- Let X be tomorrow's temperature.
  - possible outcomes: from -15°C to 35°C
  - how can we quantify the probabilities then…?

Discrete and Continuous r.v.

- A quantity X is called a discrete r.v. if 1) it can only take a discrete collection of values, and 2) it is random.

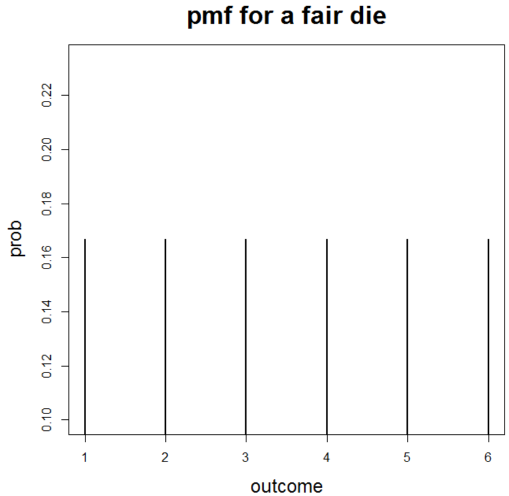
- A quantity X is called a continuous r.v. if 1) it can take a whole range of real-numbered values, and 2) it is random.

Probability mass function for discrete r.v.

- A probability mass function (or pmf) for a discrete r.v. X is a function that describes the relative probability that X takes each of its possible values.

- Denoted by $f_X(x)$ or $f(x)$.

- pmf is in form of vertical bars



Probability density function for continuous r.v.

- A probability density function (or pdf) for a continuous r.v. X is a function that describes the relative probability that X takes each value in the range of possible values.

- The range of possible values (with non-zero probabilities) is called the *support* of X.

Some common discrete r.v.

Bernoulli r.v.

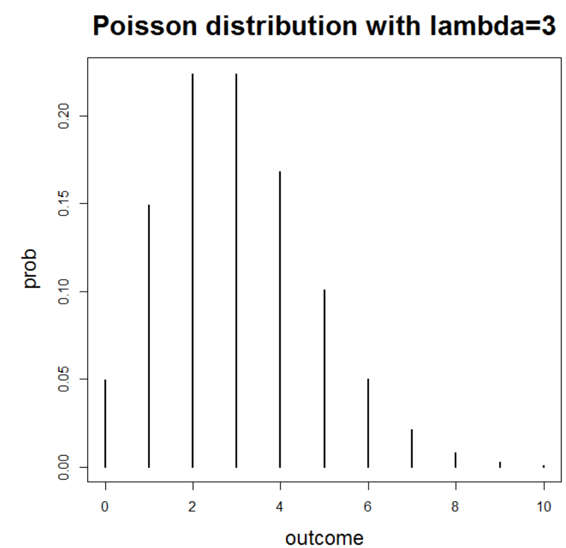
- Binary outcome: success (1) or failure (0).
- One parameter: p, probability of success.
- pmf: $\Pr(X = 1) = p, \Pr(X = 0) = 1 - p$
  - alternative expression: $f_X(x) = p^x(1-p)^{1-x}$
- $X \sim Bernoulli(p)$

Binomial r.v.

- Sum of n independent and identically distributed (i.i.d.) Bernoulli r.v.
- Takes values on $\{0, 1, 2, \dots, n\}$
- Two parameters:
  - n: Number of independent Bernoulli trials
  - p: Probability of success (inherited from Bernoulli r.v.)
- $f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$
- $X \sim bin(n, p)$

Poisson r.v.

- Number of events occurring in a fixed interval of time.
- Possible outcomes: $\{0, 1, 2, 3, \dots\}$, all non-negative integers
- Rate parameter: $\lambda > 0$
- $f_X(x) = \frac{\lambda^x e^{-\lambda}}{x!}$
- $X \sim Poisson(\lambda)$



Some common continuous r.v.

Uniform r.v.

- Two parameters: a and b (the lower and upper bound)

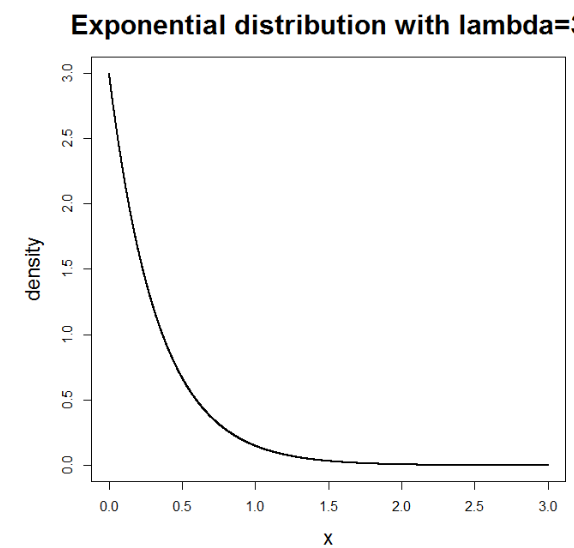
- $f_X(x) = \frac{1}{b-a}$
- $X \sim uniform(a, b)$

Exponential r.v.

- Time between events (remember Poisson?)
- Support: $[0, \infty)$
- λ: the rate parameter, $\lambda > 0$

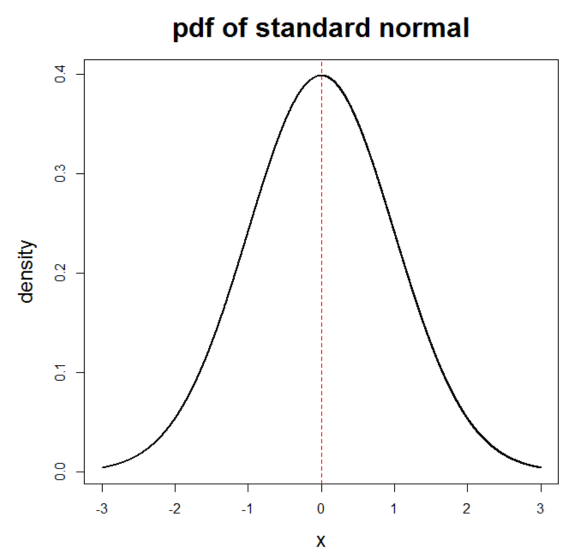
- $f_X(x) = \lambda\, exp(-\lambda x)$

- $X \sim exp(\lambda)$



Normal r.v.

- The most famous one (we'll see why)
- Takes values over the real number line
- Two parameters: μ, σ^2
- $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$
- $X \sim N(\mu, \sigma^2)$



Some notations

- We understand that the pdf/pmf can be written as $f_X(x)$ or $f(x)$. The former expression specifies the r.v. of interest through the subscript X.
  - "the pdf of the r.v. X is f_X"
- This can avoid confusion when handling more than one r.v., say, $f_X(x)$ and $f_Y(y)$, or $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$
- The lowercase (e.g. x or y) inside the round bracket indicates the value at which the pdf/pmf is evaluated.

- Some texts may even state the associated parameter(s) θ while quoting a pdf/pmf. E.g. $f(x;\theta)$ or $f(x|\theta)$

Properties of pmf/pdf

- Always above the horizontal axis (probabilities are non-negative)

- [Discrete r.v.] Sum of pmf (those vertical bars) = 1

- [Continuous r.v.] Area under the pdf = 1

- Other common distributions include:
  - negative binomial, geometric, hypergeometric, gamma, beta, Student’s t, chi-square, F, …

- Many of them are related

- It is better to understand their physical meanings than to recite their formulae…

Example

$X \sim Poisson(\lambda)$ and its pmf is $f_X(x) = \frac{\lambda^x e^{-\lambda}}{x!}$.

Show that $\sum_{x=0}^{\infty} f_X(x) = 1$

Example

$X \sim exp(\lambda)$ and its pdf is $f_X(x) = \lambda e^{-\lambda x}$. Show that $\int_{-\infty}^{\infty} f_X(x) dx = 1$

Cumulative mass/density function

- The probability of the r.v. X having the value less than or equal to x.
- $F(x) = \Pr(X \leq x)$, hence the name cumulative
- $F(-\infty) = 0\ and\ F(\infty) = 1$
- Always non-decreasing

- For discrete case,

$$F(x) = \sum_{x_i \leq x} f_X(x_i)$$

- For continuous r.v., $F(x)$ is the area under the pdf curve, from negative infinity to x.

$$F(x) = \Pr(X \leq x) = \int_{-\infty}^{x} f_X(t)dt$$

- Calculus!

Expectation

Expectation

- If you repeat the same experiment for many many times (e.g. keep tossing a coin, keep drawing r.v. from a distribution), then the expectation is the "average" from these repeated experiments.

- Note that the "average" here is the hypothetical average of infinitely many trials. Try not to confuse with the "sample average" that we calculate from data.
  – Population mean vs Sample mean

Expected value

- $E(X) = \sum_{all\ outcomes} xf(x)$

- $E(X) = \int_{-\infty}^{+\infty} xf(x)dx$

- "Average" value weighted according to the probability distribution. Often called the expected value of the r.v. X.

- $E(X)$ is the population mean or true mean of the r.v. X. It is a measure of central tendency.

Variance

- $Var(X) = E\left[\left(X - E(X)\right)^2\right]$
  $= E(X^2) - [E(X)]^2$
- The variance is the expected squared distance of the r.v. X from its population mean

- $E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x) dx$
  - if X is an r.v., then X^2 is also an r.v.
  - expected value of r.v. X^2

- Variance is a measure of dispersion

Example

- $X \sim Bernoulli(p)$, two possible outcomes: $\{0, 1\}$

$$E(X) = \sum xf(x)$$

$$= 0 * (1 - p) + 1 * p$$
$$= p$$

$$E(X^2) = \sum x^2 f(x)$$

$$= 0^2 * (1 - p) + 1^2 * p$$
$$= p$$

$$Var(X) = E(X^2) - [E(X)]^2$$
$$= p - p^2$$
$$= p(1 - p)$$

More on expectation

- $E(X^n) = \int_{-\infty}^{+\infty} x^n f_X(x) dx$

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

  $for\ any\ real\ function\ g$

- $E(X + Y) = E(X) + E(Y)$ for any r.v. X and Y (linearity)

- BTW, functions / transformations of r.v. are also r.v.

Statistical moments

$E(X)$: central tendency, mean

$E(X^2)$: dispersion, variance

$E(X^3)$: skewness

$E(X^4)$: kurtosis

- The n^{th} moment of a r.v. X is $E(X^n)$

Moment generating function

- Moment generating function (mgf) $M_X(t)$ can also be used to characterise a r.v.
  - t is a dummy variable. Subscript X is the r.v.
- The mgf "generates" statistical moments through its derivatives at $t = 0$:
- $n^{th}\ moment\ of\ X = E(X^n) = \frac{d^n M_X(t)}{dt^n}|_{t=0}$