University of BRISTOL

WEEK 2

Random variables – main definitions

Felipe Campelo

Adapted from Montgomery and Runger's
Applied Statistics and Probability for Engineers
lecture notes (Ch 3)

In this lecture…

In this lecture, we will briefly go over some main definitions related to *random* variables, including:

- Definition of discrete and continuous probability distributions, probability mass functions / density functions, cumulative distribution functions
- Mean and variance of random variables, and the expected value of functions of random variables.

Random variables

A ***random variable*** (RV) can be defined as a function that assigns a **real number** to each outcome in the **sample space** of a **random experiment**.

Commonly denoted by uppercase letters, such as X or Y. After an experiment is conducted, the **observed value** of the random variable is commonly denoted by the corresponding lowercase letter, such as $x = 70$ or $y = 0.123$.

Example: Let X denote the RV representing the total rainfall in Bristol tomorrow (in mm). The sample space $\Omega = \{x \in \mathbb{R}: x \geq 0\}$ represents all possible observable values. If tomorrow we have 12.3mm of rain, then we'd have an observation $x = 12.3$.

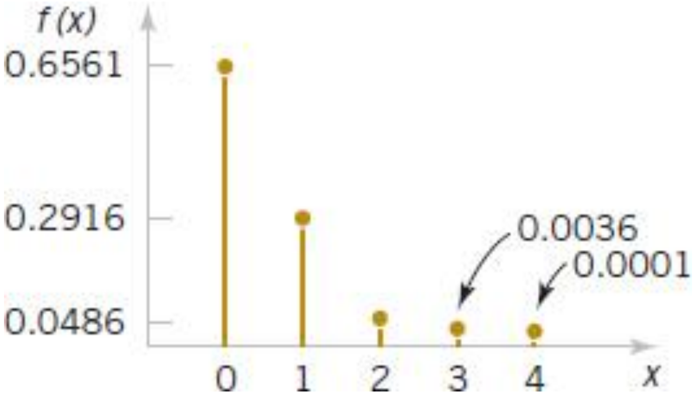
RVs can be **discrete** if they have a countable sample space; or **continuous** if its sample space is defined as an interval (finite or infinite) of real numbers.

Discrete random variables

A **random variable** is a function that assigns a real number to each outcome in the sample space of a random experiment.

The **probability distribution** of a random variable X is a description of the probabilities associated with the possible values of X.

A **discrete random variable** has a probability distribution that specifies the list of possible values of X along with the probability of each. Alternatively, it can also be expressed in terms of a function or formula.



$P(X=0) = 0.6561$

$P(X=1) = 0.2916$

$P(X=2) = 0.0486$

$P(X=3) = 0.0036$

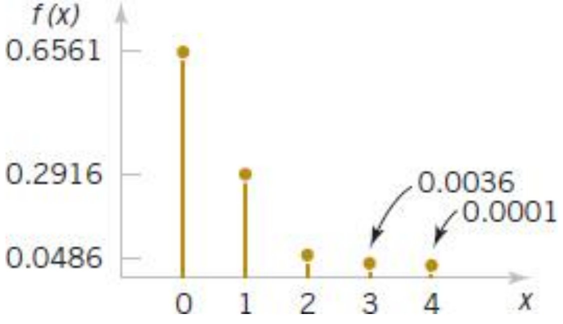
$P(X=4) = 0.0001$

1.0000

Probability mass function

For a discrete random variable X with possible values $\{x_1, x_2, \dots, x_K\}$, a **probability mass function** (pmf) $f(x)$ is a function such that:

1. $f(x_i) \geq 0$
2. $\sum_{i=1}^{K} f(x_i) = 1$
3. $f(x_i) = P(X = x_i)$



P(X =0) = 0.6561
P(X =1) = 0.2916
P(X =2) = 0.0486
P(X =3) = 0.0036
P(X =4) = 0.0001

1.0000

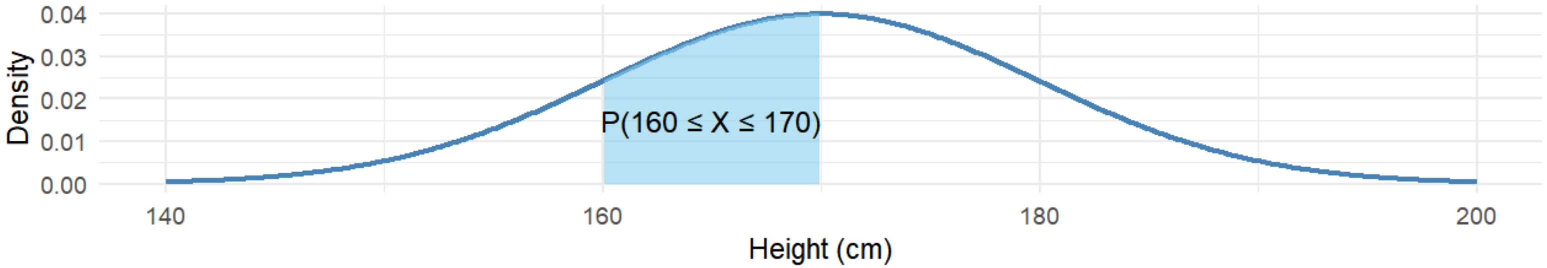
$$f(X) = \begin{cases} 0.6561\text{ , if } X = 0 \\ 0.2916\text{, if } X = 1 \\ 0.0486\text{, if } X = 2 \\ 0.0036\text{, if } X = 3 \\ 0.0001\text{, if } X = 4 \end{cases}$$

Continuous random variables

A **continuous random variable** has infinite possible values in its sample space, which is commonly defined as an interval (finite or infinite) of Real numbers.

A **probability density function** (pdf) $f(X)$ can be used to describe the probability distribution of a continuous random variable X. A pdf is a function such that:

1. $\mathrm{f(x)} \geq 0, \forall x$
2. $\int_{-\infty}^{\infty} \mathrm{f(x)d(x)} = 1$
3. $\mathrm{P}(\mathrm{a} \leq x \leq b) = \int_{\mathrm{a}}^{\mathrm{b}} \mathrm{f(x)d(x)}$, i.e., the area under the curve $f(x)$ between a and b.



Cumulative distribution function – discrete RV

The **cumulative distribution function** (cdf), $F(x)$ is the probability that a random variable X, with probability distribution $f(x)$, will be found at a value less than or equal to x:

$$F(x) = P(X \leq x) = \sum_{\forall x_i \leq x} f(x_i)$$

For a discrete random variable X, $F(x)$ satisfies the following **properties**:

1. $F(x) = P(X \leq x) = \sum_{\forall x_i \leq x} f(x_i)$
2. $0 \leq F(x) \leq 1$
3. If $x \leq y$, then $F(x) \leq F(y)$

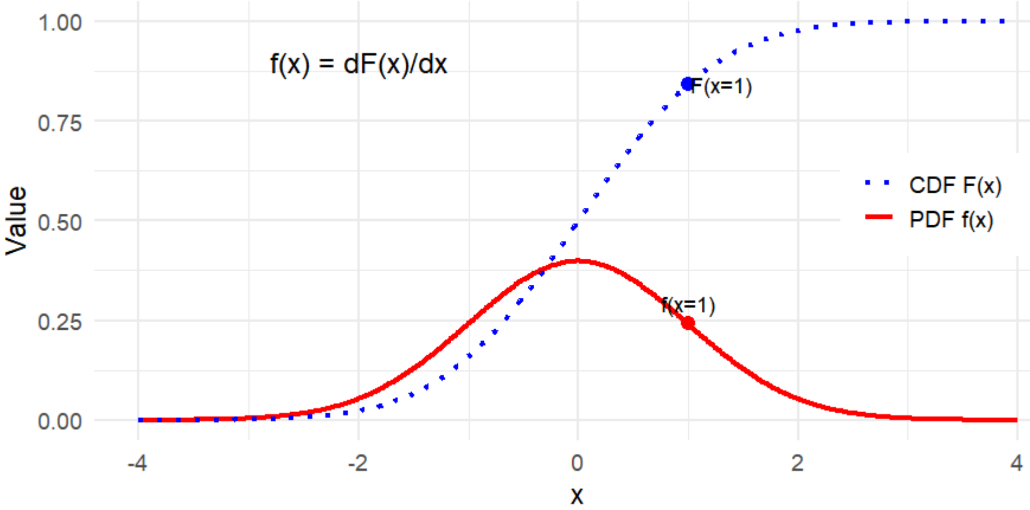
Cumulative distribution function – continuous RV

Just like in the discrete case, the **cumulative distribution function** (cdf) $F(x)$ is a function that denotes the probability that the variable assumes any values smaller than x,

$$F(x) = P(X \leq x) = \int_{-\infty}^{\infty} f(u)du\,, \text{for } x \in \mathbb{R}$$

The probability density function of a continuous RV can be determined from the cumulative distribution function by differentiating function by differentiating (if the derivative exists),

$$f(x) = \frac{dF(x)}{dx}$$

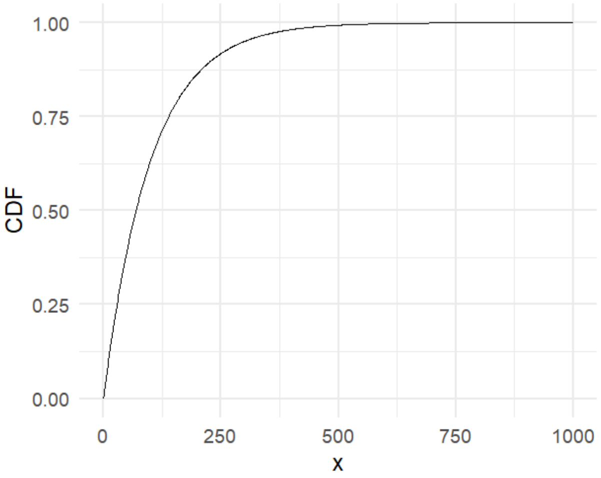


Example

The time until convergence of an optimisation algorithm (in milliseconds) is approximated by this cdf,

$$F(x) = \begin{cases} 0, & \text{for x} < 0 \\ 1 - e^{-0.01x} & \text{for } x \geq 0 \end{cases}$$

What is the pdf?



Example

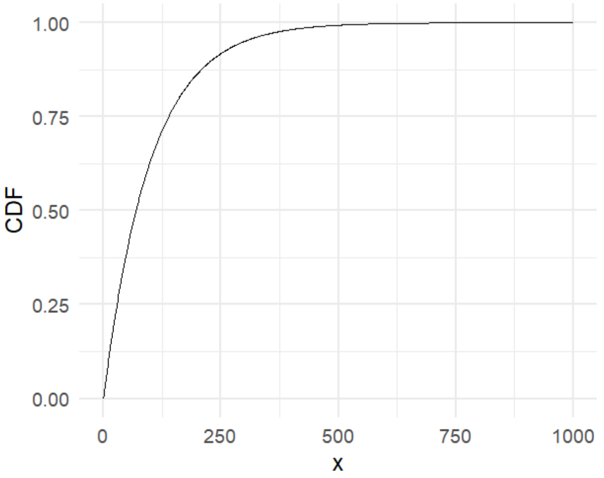
The time until convergence of an optimisation algorithm (in milliseconds) is approximated by this cdf,

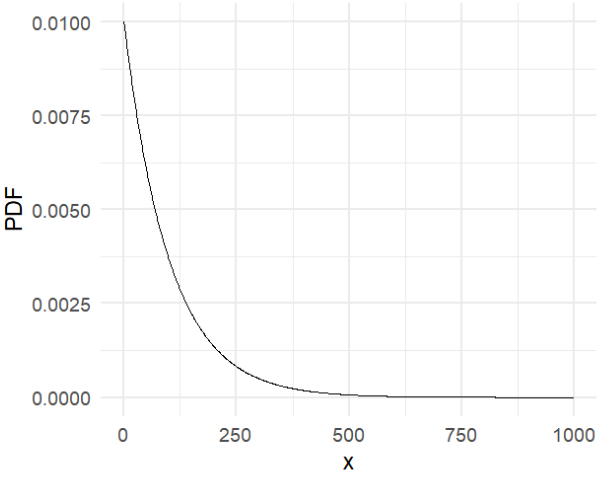
$$F(x) = \begin{cases} 0, & \text{for x} < 0 \\ 1 - e^{-0.01x} & \text{for } x \geq 0 \end{cases}$$

What is the pdf? Taking the derivative,

$$f(x) = \frac{dF(x)}{dx} = \begin{cases} 0, & \text{for x} < 0 \\ 0.01e^{-0.01x} & \text{for } x \geq 0 \end{cases}$$

What is the probability that the algorithm converges faster than 200ms?





Example

The time until convergence of an optimisation algorithm (in milliseconds) is approximated by this cdf,

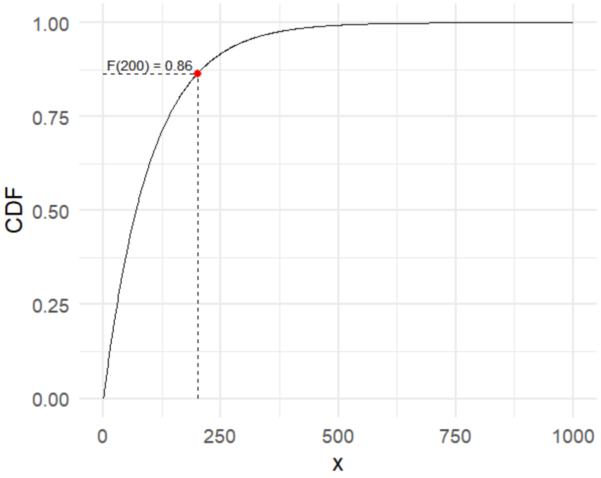
$$F(x) = \begin{cases} 0, & \text{for x} < 0 \\ 1 - e^{-0.01x} & \text{for } x \geq 0 \end{cases}$$

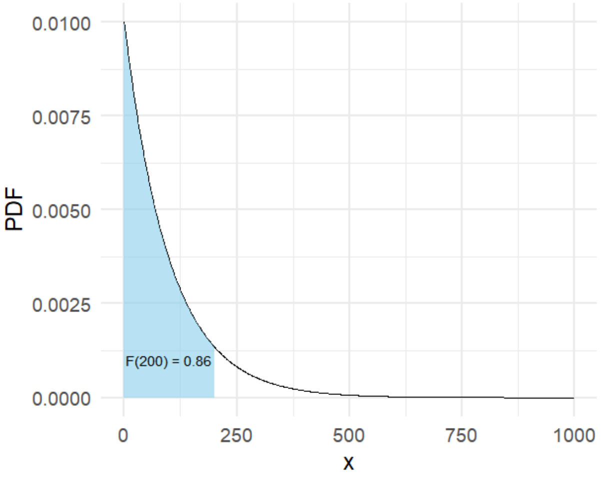
What is the pdf? Taking the derivative,

$$f(x) = \frac{dF(x)}{dx} = \begin{cases} 0, & \text{for x} < 0 \\ 0.01e^{-0.01x} & \text{for } x \geq 0 \end{cases}$$

What is the probability that the algorithm converges faster than 200ms?

$$P(x \leq 200) = F(200) = 1 - e^{-2} = 0.8647$$





Mean and variance of RVs

Mean:

$$\mu = E(X) = \begin{cases} \sum_x xf(x), & discrete\ RVs \\ \int_{-\infty}^{\infty} xf(x)dx, & continuous\ RVs \end{cases}$$

Variance:

$$\sigma^2 = V(X) = E(x-\mu)^2$$

$$= \begin{cases} \sum_x (x-\mu)^2 f(x) = \sum_x x^2 f(x) - \mu^2, & discrete\ RVs \\ \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2, & continuous\ RVs \end{cases}$$

Standard deviation: $\sigma = \sqrt{\sigma^2}$

Mean: *measure of central tendency (center or middle) of the probability distribution.*

Variance: *measure of the dispersion (variability) of the distribution.*

Standard deviation: *square root of the variance. Another way to measure dispersion. Has the convenient property of being defined in the same units as the mean.*

Expected value of a function of a RV

If X is a discrete random variable with pmf $f(x)$, and $h(X)$ is a function of X, then:

$$E[h(x)] = \sum_{x} h(x) f(x)$$

Alternatively, for a continuous random variable with pdf $f(x)$,

$$E[h(x)] = \int_{-\infty}^{\infty} h(x) f(x) dx$$

Notice that the Variance $V(X)$ can be seen as an example of this, for the case in which $h(x) = (X - \mu)^\wedge 2$.

Example

Example: Suppose that a scratch lottery ticket is sold by £2 and has 3 possible outcomes:

- No win ($X = 0$), with $f(0) = 0.9$.
- Small prize ($X = 1$), with $f(1) = 0.09$.
- Large prize ($X = 2$), with $f(2) = 0.01$.

Assume a payoff function $g(x)$ with $g(0) = £0, g(1) = £10, g(2) = £50$. What is the expected payoff?

Example

Example: Suppose that a scratch lottery ticket is sold by £2 and has 3 possible outcomes:

- No win ($X = 0$), with $f(0) = 0.9$.
- Small prize ($X = 1$), with $f(1) = 0.09$.
- Large prize ($X = 2$), with $f(2) = 0.01$.

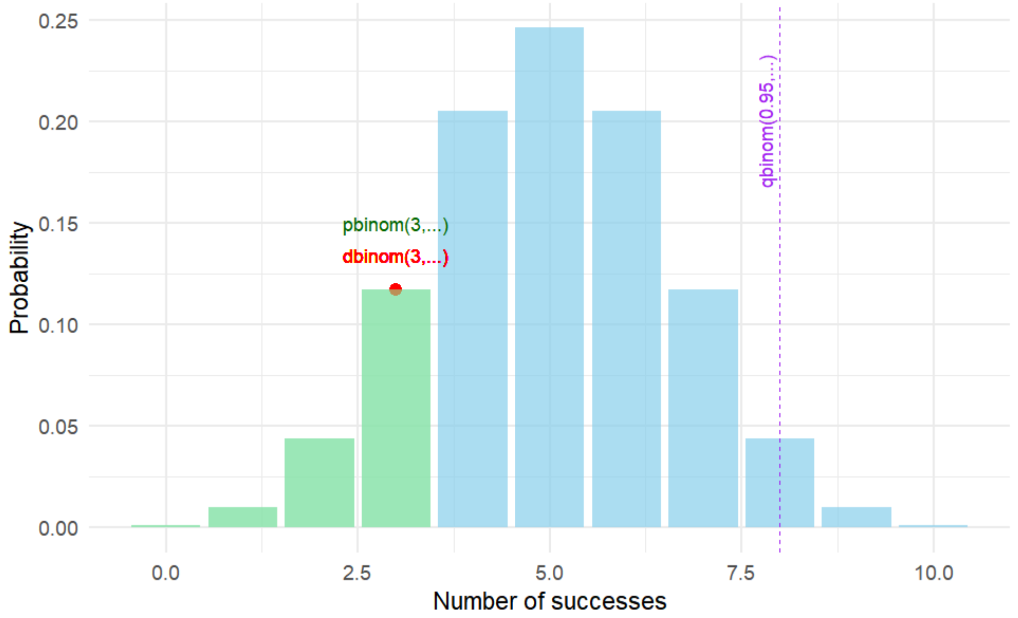
Assume a payoff function $g(x)$ with $g(0) = £0, g(1) = £10, g(2) = £50$. What is the expected payoff?

$$E[g(x)] = \sum_x g(x)f(x) = 0 + 0.9 + 0.5 = £1.40$$

Since the ticket costs $£2$, the lottery operator is expected to make an average profit of $£2 \ - £1.40 = 60\text{p}$ / ticket.

R functions for random variables

R makes it really simple to simulate different aspects of random variables. Most families of RVs have a set of functions starting with the letters d/p/q/r.



For example, for the Binomial distribution:

Function	Meaning	Example	Output
dbinom(x, N, prob)	**pmf, $P(X = x)$**	dbinom(3, 10, 0.5)	Probability of exactly 3 heads in 10 tosses
pbinom(q, N, prob)	**cdf, $P(X \leq x)$**	pbinom(3, 10, 0.5)	Probability of at most 3 heads in 10 tosses
qbinom(p, N, prob)	**Quantile fn.**	qbinom(0.95 10, 0.5)	Smallest x such that $P(X \leq x) \geq 0.95$
rbinom(n, N, prob)	**Sampling fn.**	rbinom(5, 10, 0.5)	5 simulated tosses.

Further reading

DC Montgomery, GC Runger, *Applied statistics and probability for engineers, 5th ed.*
[Chapters 2.8, 3.1 - 3.4, 4.1 - 4.4]

- *Read the sections and try to solve the worked examples by yourself.*