EMAT10100 Engineering Maths I
Lectures 4&5 Introduction to Probability: Random Variables

Rosalyn Moran & Nikolai Bode

Plan of action

Because you have all been doing your homework practice (!)

- **You are now supposed to be fluent with the material in section 13.3 and we will not be doing further re-enforcement of it. Use drop-in classes for help!**

(We mean it!)

You are now supposed to know about: *randomness*, *probability*, *random trials*, *(elementary) outcomes*, *sample space*, *events*, *axioms of probability*, esp. *general addition rule*; and also *conditional probability* and *independence*

- This week: discrete and continuous random variables (parts of) *James* section 13.4 in particular
- Next week: lots and lots of practical *distributions* (parts of) *James* section 13.4 and section 13.5

Random Variables (RVs)

Simple man's version: a RV describes what you get from a *random trial* in which the *elementary outcomes* are numbers.

Most basic example: throwing a six-sided die
sample space $S = \{1, 2, 3, 4, 5, 6\}$

When the *outcomes* are numbers — then we can talk about (e.g.)

- whether one outcome is bigger than another
- what the average outcome is
- and put the outcomes inside other useful Mathematical functions and procedures

* NB most generally - we could have vector-valued outcomes and even more exotic objects - beyond the scope here.

Random variables - more formally and generally

- Let S be the sample space for a given random trial.
- Define a function $\phi : S \to \mathbf{R}$.
- Then ϕ acting on S defines a *random variable* X.

NB there is no need for the original sample space to be numeric.

Example: Tossing a single coin. $S = \{H, T\}$.

We might define a random variable by (e.g.) $\phi(H) = 0$ and $\phi(T) = 1$.

(Note that this is not particularly useful in this case and many other choices are possible — so this is really for purposes of illustration only).

A More Interesting Example

Trial: draw a card at random from a deck of playing cards.

- The sample space is

$$\begin{aligned} S = \{&2\clubsuit,\ 3\clubsuit,\ 4\clubsuit,\ \ldots\ 10\clubsuit, J\clubsuit, Q\clubsuit, K\clubsuit, A\clubsuit,\\ &2\diamondsuit,\ 3\diamondsuit,\ 4\diamondsuit,\ \ldots\ 10\diamondsuit, J\diamondsuit, Q\diamondsuit, K\diamondsuit, A\diamondsuit,\\ &2\heartsuit,\ 3\heartsuit,\ 4\heartsuit,\ \ldots\ 10\heartsuit, J\heartsuit, Q\heartsuit, K\heartsuit, A\heartsuit,\\ &2\spadesuit,\ 3\spadesuit,\ 4\spadesuit,\ \ldots\ 10\spadesuit, J\spadesuit, Q\spadesuit, K\spadesuit, A\spadesuit\}\end{aligned}$$

  and is non-numeric!
- In lots of card games: a "high" card beats a lower card — so we might set up an RV X by defining ϕ with $\phi(2\clubsuit) = 2$, $\phi(3\clubsuit) = 3$, and so on and $\phi(J\clubsuit) = 11$, $\phi(Q\clubsuit) = 12$, $\phi(K\clubsuit) = 13$, $\phi(A\clubsuit) = 14$ (and likewise for other suits).
- The RV X gives a framework for asking questions such as: "what is the probability my card beats another card if . . ."

Random variables - two notes

- The numeric values $\phi(\omega)$ for each $\omega \in S$ are different to the probabilities $P(\omega)$ of those outcomes*. The interesting stuff happens when we combine $\phi(\omega)$ with $P(\omega)$.

- If you are ever confused about how to apply the axioms of probability; whether two RVs are independent etc. — write down the sample space carefully and apply the principles from previous lectures.

* We are being a little slack here by defining probabilities on *outcomes* — recall they should be defined on *events*. This distinction becomes important shortly.

Notation and Terminology — Probability function

- Random trial: throw one (fair) six-sided die
  Let the RV X denote the value of the result.

- The RV X is completely described by the *probability function*

$$P_X(x) := P(X = x)$$

- So here
  $P_X(1) = 1/6$, $P_X(2) = 1/6$, $P_X(3) = 1/6$, $P_X(4) = 1/6$,
  $P_X(5) = 1/6$, $P_X(6) = 1/6$.

* Note the convention: upper case for the RV and lower case for the values it can take.

Notation and terminology — the (cumulative) distribution

- Recall the probability function $P_X(x) := P(X = x)$.

- The *cumulative distribution function* or *CDF*, $F_X(x)$, is defined by:

$$F_X(x) := P(X \leq x).$$

Example: Random Trial: throw one (fair) six-sided die. Then

$$F_X(4) = P(X \leq 4) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = \frac{2}{3}$$

Exercise: In general — what is the probability that an RV X lies between a and b;

i.e., what is $P(3 < X \leq 5)$ in the example above?

Extended Exampercise: Throwing two fair dice

- Random trial: throw two fair six-sided dice
- Let X denote the RV for the sum of the two values showing.
- Compute the *probability function* $P_X(x)$.
- Plot the probability function and the cumulative distribution for the "two dice"

* This is a special case of a more general problem:
Let Y, Z be independent RVs.
Compute the probability function of the RV X given by $X = Y + Z$.

Worked solution

Discrete and Continuous RVs

- *Discrete RVs.*
  - Either: defined on a *finite* sample space (throwing die etc.)
  - Or: (possibly) on an *infinite* sample space, but somehow values are well-separated, i.e. *discrete*

    E.g.: Pick a positive integer at random in some way.
- *Continuous RVs*
  - Not discrete!!!

    E.g.: Pick any (real) number at random between 0 and 1.

Today: concepts taught through discrete RVs only.
Wednesday: how to extend ideas to cover continuous RVs.

Differences and similarities of discrete and continuous RVs

- The concept of the *probability distribution* $F_X(x) := P(X \leq x)$ works for both discrete and continuous RVs as it is defined using the probability of the *event* $\{X \leq x\}$
- The *probability function* $P_X(x) := P(X = x)$ only makes sense for discrete RV as it is defined on a specific *outcome*.
  - For continuous RVs — the probability of individual outcomes is usually zero, and $P_X(x)$ needs replacing with a *probability density function* (pdf) $f_X(x)$.
- All of the other concepts are similar except every time we write a sum Σ for discrete RVs — it gets replaced by an integral $\int$ for continuous RVs.

Examples of modelling with discrete and continuous RVs

- Random trial: select an individual at random from this room.
- Random variable X: take the 'day number' of their birthday.
- This is a discrete RV.
- Random variable Y: measure the height of the individual.
- This can take any value from a range — so is a continuous RV.

* - There are some philosophical / modelling issues here.
What if we measured height only to the nearest millimetre?

Measures of location and dispersion

We want to derive measures for

- The 'typical' value that an RV takes (a so-called *location* measure)
  - Maybe the typical height in this room is about 1.65 m?
  - Maybe the typical height in a room of 5 year olds is more like 1.2 m?
- The 'spread' of values that an RV takes (a so-called *dispersion* measure)
  - The spread of heights in a mixed room of adults and 5 year olds will be bigger than in rooms of adults only or 5 year olds only.

A large dispersion value indicates 'more uncertainty' and a bigger spread in the range of possible outcomes.

Location measure 1: Mean (average) μ of an RV

- Let X be an RV.
- Peform its trial n times and record the values $x_1, x_2, \ldots, x_n$.
- Then the average
$$\frac{1}{n}(x_1 + x_2 + \ldots + x_n)$$
will usually tend to 'settle down' (converge) to a fixed value μ as we take more and more trials (n bigger and bigger)

This value μ is called the mean (average) of X — sometimes written $E(X)$ (pronounced "expectation of X").
It turns out that for discrete RVs we may compute

$$\mu = E(X) := \sum_x x P_X(x)$$

where the sum is over all possible values x that X can take.

Exampercise

Compute the mean for

- a roll of one six-sided die
- a roll of two six-sided dice

Further location measures: *Median* m and *Mode*

Median $m \simeq$ the 'middle' value of an RV X.

Roughly speaking: the probability that X is less than m should be the same as the probability that X is bigger than m.

Formally we define the *median* as the value m such that

$$P(X \leq m) \geq 1/2 \quad \text{and} \quad P(X \geq m) \geq 1/2.$$

Exercise: What is the median for (i) a roll of one six-sided die? (ii) a roll of two six-sided dice?

The *mode* is the most common (i.e. most probable) value of an RV X — i.e., the value of x such that $P_X(x)$ is greatest.

Exercise: What is the mode for a roll of two six-sided dice?

Mean versus Median

- The *mean* is affected more than the *median* by rare extreme values of the RV (values that are too small or too large).

  E.g. consider the discrete RV defined by the probability function $P_X(1) = 0.33$, $P_X(2) = 0.33$, $P_X(3) = 0.33$, $P_X(1000000) = 0.01$.
  Compute the mean and median.

- Sometimes the *mean* doesn't even exist!
  But the *median* always exists (even if at times is hard to compute) e.g.,

  consider the discrete RV defined by the probability function

  $$P_X(x) = \frac{6}{\pi^2}\frac{1}{x^2},$$

  where x is any positive integer $1, 2, \ldots$.
  Compute the mean and median.

Dispersion measures: Variance and standard deviation

The *variance* σ^2 — sometimes called Var(X) — is a measure of how far an RV X typically deviates from its mean. We define

$$\sigma^2 := E((X-\mu)^2) = \sum_x \left[P_X(x)(x-\mu)^2\right],$$

so that σ^2 is the average of $(X-\mu)^2$.
(As before, the sum is over all values x that X can take.)

A little algebra (see last slides of this lecture) gives:

$$\sigma^2 = E(X^2) - \mu^2 = \sum_x \left[P_X(x)x^2\right] - \mu^2.$$

Sometimes we work with the *standard deviation* σ defined as the square root of the variance. It has the same dimensions (units) as X.

* Question: why don't we use $E(X-\mu)$ to analyse dispersion?

Exampercises

- Compute the variance and standard deviation for (i) the roll of one six-sided die; (ii) the roll of two six-sided dice.

Solutions

Functions of RVs

- Given an RV X, we can always define a new RV $Y = g(X)$, where $g : \mathbf{R} \mapsto \mathbf{R}$ is some given function.
- So each trial of X generates an x, and the value y of Y is given by $y = g(x)$.
- **Exercise:** Let X be the RV for the roll of a die. Let $Y = X^2$. What is the probability function $P_Y(y)$?

Linear Functions and Combinations of RVs

Note that linear functions imply special scaling properties for the mean. Specifically given two RVs X and Y and two constants a and b we have:

- $E(aX) = aE(X)$
- $E(X + b) = E(X) + b$
- $E(X + Y) = E(X) + E(Y)$
- $E(aX + bY) = aE(X) + bE(Y)$

Using these properties, we can show that

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

applying the properties above.

* Question: what about the variance of $X + Y$?

James Homework

- Read Sections 13.4.1 and 13.4.2 from *James* to recap basics.
- Attempt Exercises 13.4.5, pp. 998 (5th ed., pp. 1024) numbers 25, 26, 27.
- Read Sections 13.4.6 and 13.4.7 on mean and variance (but beware that James mixes continuous RV material into these sections).
- Skim Sections 13.4.3 and 13.4.4 for preparation on continuous RV material.