Gatsby Bridging Program

Probability: Discrete Distributions

Cameron Stewart

Gatsby Computational Neuroscience Unit

Table of contents

Random Variables, Probability Mass Functions, and Cumulative Distribution Functions

What is a Random Variable?

Random variables are functions which map the possible outcomes of an experiment to numerical values. How we define these functions is up to us.

Example 1

Consider a bag containing 3 red balls (R) and 5 green balls (G). In our experiment, we are going to draw 2 balls from the bag, with replacement. The sample space is $\Omega = \{\text{RR}, \text{RG}, \text{GR}, \text{GG}\}$.

We are interested in determining the probabilities of drawing various numbers of red balls. To to this, we could start by defining a random variable $X : \Omega \to \{0, 1, 2\}$ such that

$$X(\omega) = \begin{cases} 0 & \text{if } \omega = \text{GG} \\ 1 & \text{if } \omega \in \{\text{RG, GR}\} \\ 2 & \text{if } \omega = \text{RR} \end{cases}.$$

What is a Random Variable?

Random variables are often utilised without any explicit reference to the sample space. Instead of writing $X(\omega)$, we will typically just write X. Instead of writing $\mathbb{P}(E)$ for some event E, we typically write $\mathbb{P}(X \in A)$ for some set of numerical values A.

Example 1 Cont.
What is the probability of drawing one red ball? We could express this as $\mathbb{P}(\{\text{RG}, \text{GR}\})$, but it is common and accepted notation to instead write $\mathbb{P}(X \in \{1\})$, or $\mathbb{P}(0 < X < 2)$, or most preferably in this case $\mathbb{P}(X = 1)$.

$$\begin{aligned}\mathbb{P}(X = 1) &= \frac{3}{8}\frac{5}{8} + \frac{5}{8}\frac{3}{8} \\ &= \frac{15}{32}.\end{aligned}$$

Discrete and Continuous Random Variables

Discrete random variables can take a countable number of values. For example:

- $X \in \{0, 1\}$
- $X \in \{0, 1, 2, \dots\}$
- X representing the number of buses arriving within an hour.

Continuous random variables can take values in continuous ranges. For example:

- $X \in [0, 1]$
- $X \in \mathbb{R}$
- X representing the waiting time until the next bus.

Probability Mass Functions

Discrete distributions can be defined by their probability mass function (PMF). The PMF of random variable X is often denoted by f_X or p_X, and is defined as

$$f_X(x) = \mathbb{P}(X = x) \ .$$

f_X is simply a function name. It is fine to use a different name, as long as it is clear how the function is defined. Occasionally, the same name is used for PMFs if it is clear from context how these are defined, but I'd advise against this practice for the sake of clarity. E.g. it is clearer to write $p_X(x)$ and $p_Y(y)$ than $p(x)$ and $p(y)$ if these correspond to 2 different PMFs.

Probability Mass Functions

Probabilities can't be negative and must sum to 1 over the set of all possible values $\mathcal{X}$, so we have the following constraints:

$$f_X(x) \geq 0 \text{ for all } x$$

and

$$\sum_{x \in \mathcal{X}} f_X(x) = 1\,.$$

$\mathcal{X}$ is referred to as the support of X. $f_X(x) = 0$ for $x \notin \mathcal{X}$.

Example 1 Cont.
Let's continue with our red ball example. What is the PMF of X? First, we recognise that $\mathcal{X} = \{0, 1, 2\}$. You can verify for yourself that

$$f_X(x) = \begin{cases} \frac{25}{64} & \text{if } x = 0 \\ \frac{15}{32} & \text{if } x = 1 \\ \frac{9}{64} & \text{if } x = 2 \\ 0 & \text{otherwise} \end{cases},$$

and that this function satisfies the conditions for a PMF on the previous slide.

Cumulative Distribution Functions

A probability distribution can also be defined in terms of its cumulative distribution function (CDF). The CDF of a random variable X is a monotonically increasing function defined as

$$F_X\left(x\right) = \mathbb{P}\left(X \leq x\right) \, .$$

For discrete random variables, we can relate this definition to the PMF as follows:

$$F_X\left(x\right) = \sum_{y \leq x} f_X\left(y\right) \, .$$

Cumulative Distribution Functions

As probabilities must sum to 1 over the support, we have the following emergent properties for CDFs:

$$\lim_{x \to -\infty} F_X\left(x\right) = 0$$
$$\lim_{x \to \infty} F_X\left(x\right) = 1$$

We can also see that the following is true:

$$\mathbb{P}\left(a < X \leq b\right) = F_X\left(b\right) - F_X\left(a\right) \, .$$

Cumulative Distribution Functions

Example 1 Cont.
Back to the red ball example. What is the CDF of X? Previously, we found that

$$f_X(x) = \begin{cases} \frac{25}{64} & \text{if } x = 0 \\ \frac{15}{32} & \text{if } x = 1 \\ \frac{9}{64} & \text{if } x = 2 \\ 0 & \text{otherwise} \end{cases}.$$

Hence, the CDF is

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{25}{64} & \text{if } 0 \leq x < 1 \\ \frac{55}{64} & \text{if } 1 \leq x < 2 \\ 1 & \text{if } x \geq 2 \end{cases}.$$