Foundational ACTL2131 Concepts

An overview of basic probability and statistical theorems

Tadhg Xu-Glassop

UNSW

2024

Outline

Outline

Definition

Consider two events A and B, where B is a partition of of the sample space Ω^1.

Denote each possible outcome of B as $B_i, i \in \{1, 2, \ldots, n\}$ then, the Law of Total Probability (LOTP) states, for a discrete case,

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A|B_i) \cdot P(B_i).$$

---

[1]In essence, LOTP is usable if the scenario can be written as a tree diagram where all possible branches are present and probabilities sum to 1.

Visualising LOTP with a Tree Diagram

The Law looks far more complicated than it actually is. Consider the following tree diagram for an event B followed by an event A, whose probability of occurring is dependent on the outcome of B,
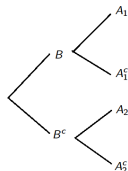
If we seek the probability $P(A)$, which is the **total** probability of A occurring, then it is simply the sum of each branch, that is,

$$P(A) = P(A_1|B) \cdot P(B) + P(A_2|B^c) \cdot P(B^c).$$

False Positive Example

Q. Consider a test for a virus which has a 5% chance to return a false positive (the test says you have the disease, when you actually don't). The test will return the correct result if the person has the disease. If it is estimated that 20% of the population has the virus, what is the probability of a single test returning a positive result?

False Positive Example (soln)

Ans. (Using a tree diagram may be useful) Denote the event of someone having the virus as V, and the event of a positive test P. We seek $P(P)$.

By LOTP, we have,

$$\begin{aligned}
P(P) &= P(P|V)P(V) + P(P|V^c)P(V^c) \\
&= 1 \cdot 0.2 + 0.05 \cdot 0.8 \\
&= 0.24.
\end{aligned}$$

Outline

Definition

Similar to LOTP, Consider two events A and B, where B is a partition of of the sample space Ω. We can denote each different outcome of B as $B_i, i \in \{1, 2, \ldots, n\}$ and a specific event of interest $B_j, j \subset i$.

Then, Bayes' Theorem states[2],

$$P(B_j|A) = \frac{P(A|B_j) \cdot P(B_j)}{\sum_i P(A|B_i) \cdot P(B_i)}$$

_____

[2]Like LOTP, similar conditions need to be met for Bayes' to be applicable, but generally we can use it if we can represent our situation as a tree diagram.

Derivation

This law looks complicated, but the derivation should hopefully clarify what is actually going on - it is no more than a combination of conditional probabality and LOTP.

From our scenario, we are intersted in the probability of a specific event $B_j|A$. We begin with the definition of conditional probability,

$$\begin{aligned} P(B_j|A) &= \frac{P(B_j \cap A)}{P(A)} \\ &= \frac{P(A|B_j) \cdot P(B_j)}{P(A)} \\ &= \frac{P(A|B_j) \cdot P(B_j)}{\sum_i P(A|B_i) \cdot P(B_i)}. \end{aligned}$$

False Positive Example (again!)

Q. Let us return to the same scenario. As a reminder, a test has a 5% chance to return a false positive, and it is estimated that 20% of the population has the virus. Someone takes a test and the result is positive. What is the probability they actually have the disease?

False Positive Example (again!) Solution

Ans. Using the same denotations as before, we seek $P(V|P)$. By Bayes' Theorem, we have,

$$
\begin{aligned}
P(V|P) &= \frac{P(A|B_j) \cdot P(B_j)}{\sum_i P(A|B_i) \cdot P(B_i)} \\
&= \frac{1 \cdot 0.2}{1 \cdot 0.2 + 0.05 \cdot 0.8} \\
&= 0.834.
\end{aligned}
$$

Outline

What is a Random Variable?

- A Random Variable (R.V.) is a quantity, typically denoted by X, whose value x is the outcome of an experiment/event.
  - It can be thought of that X is the variable itself, and x represents the possible values it can take.
  - $P(X = x)$ is the probability X takes the value x.
- For example, denote X as the number rolled on a standard die. Then, x represents the possible values of X. With this in mind, we read $P(X = 3)$ as the probability that we roll a 3.

Discrete Distributions

Consider a discrete case of X. We define the Probability Mass Function (p.m.f.) of X $p_X(x)$ as,

$$p_X(x) = P(X = x).$$

Further, we define the Cumulative Distribution Function (c.d.f.) $F_X(x)$ as,

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

We have the relationship between the two,

$$F_X(x) = \sum_{k=1}^{x} p_X(x_k),$$

as the cdf is simply the probability of all events less than and equal to x, so for a discrete case we can simply sum all lower cases.

Continuous Distributions

Now, consider a continuous case of X. We define the c.d.f. of X to be,

$$F_X(x) = P(X \leq x),$$

and the pdf to be,

$$f_X(x) = \frac{\partial}{\partial x} F_X(x).$$

Note that this implies the crucial relationship,

$$F_X(x) = \int_{-\infty}^{x} f_X(x_i) dx_i.$$

This is intuitive, as recall that an integral is an infinitesimal sum, so this relationship is just an continuous extension of our discrete case.

R.V.'s of Known Distributions

- Oftentimes, random variables follow a specific, known distribution.
    - For example, Binomial, Poisson, Normal and Exponential distributions are common ones you will encounter frequently.
- For known distributions, the PDF/PMF, CDF, key moments and more can be found in your Orange Book. Using these results from your Orange Book can reduce many tedious calculations and proofs.

Properties of R.V. Functions

Not every function can be a pdf / pmf / cdf! There are key
properties that each of these must follow:

- $\begin{cases} f_X(x) \geq 0 & \text{if } X \text{ is continuous,} \\ p_X(x) \geq 0 & \text{if } X \text{ is discrete.} \end{cases}$

- $p_X(x)$ is right continuous if X is discrete.

- $\begin{cases} \int_{-\infty}^{\infty} f_X(x) dx = 1 & \text{if } X \text{ is continuous,} \\ \sum_x p_X(x) = 1 & \text{if } X \text{ is discrete.} \end{cases}$

- $F(\infty) = 1$

- $F(-\infty) = 0$

R.V. Moments

When working with R.V.'s, we use **moments** to get a better understanding of the nature of our R.V..

We have non-central moments, where the nth non-central moment is defined as,

$$E\left[X^n\right],$$

and central moments, where the nth central moment is defined as

$$E\left[(X - E[X])^n\right].$$

R.V. Moments (cont.)

- Most commonly, we will be working with the first non-central moment $E[X] = \mu$, which is the mean (expected value) of a R.V..

- Another key moment is the second central moment $E[X^2] - E[X]^2 = \text{Var}[X] = \sigma^2$, which is the variance of a R.V..

  - The square root of variance gives the standard deviation, that is,
  $$\sigma = \sqrt{\text{Var}[X]}.$$

- Dividing $E\left[(X - \mu)^n\right]$ by σ^n gives the nth standardised moment.

  - The 3rd standardised moment is skewness, and the 4th is kurtosis.

Properties of Mean and Variance

Oftentimes, we will evoke the following properties for independent random variables X, Y and constants $a, b \in \mathbb{R}$,

- $E[X + Y] = E[X] + E[Y]$
- $E[XY] = E[X] \cdot E[Y]$
- $E[aX + b] = aE[X] + b$
- $\text{Var}[aX + b] = a^2\text{Var}[X]$

Finding Moments using the pmf / pdf

We can find the nth non-central moment using the pmf of a discrete R.V.,

$$E[X^n] = \sum_{\text{all } x} x^n p_X(x),$$

or by using the pdf of a continuous R.V.,

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

In fact, we can find the expected value of some function $g(X)$ with,

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx \quad \text{or} \quad \sum_{\text{all } x} g(x) p_X(x).$$

Moment of an R.V. Example

Q. Consider the continuous random variable X, which has the pdf,

$$f_X(x) = e^{-x}, x \geq 0.$$

Find the mean and variance of X.

Moment of an R.V. Example (Soln.)

Soln. The mean is the first non-central moment of X, so we find it by using,

$$
\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} x \cdot f_X(x)dx \\
&= \int_{0}^{\infty} x \cdot e^{-x}dx \\
&= -xe^{-x}\big|_{0}^{\infty} + \int_{0}^{\infty} e^{-x}dx \\
&= 1
\end{aligned}
$$

So the mean of X is $\mu = 1$.

Moment of an R.V. Example (Soln.) (cont.)

To variance is the second central moment, which requires us to
first find the second non-central moment,

$$
\begin{aligned}
E\left[X^2\right] &= \int_{-\infty}^{\infty} x^2 f_X(x) dx \\
&= \int_0^{\infty} x^2 e^{-x} dx \\
&= -2xe^{-x}\big|_0^{\infty} + 2\int_0^{\infty} xe^{-x} dx \\
&= 2E[X] \\
&= 2
\end{aligned}
$$

So, to get the variance,

$$
\sigma^2 = E[X^2] - E[X]^2 = 2 - 1^2 = 1.
$$

Outline

Moment Generating Functions

Define the Moment Generating Function (MGF) of X to be,

$$M_X(t) = E\left[e^{X \cdot t}\right].$$

Using this function, we can generate non-central moments, where the nth moment is given by,

$$E\left[X^n\right] = M_X^{(n)}(t)\Big|_{t=0}.$$

A key feature of the MGF is that there is always a unique MGF for every PDF. This fact is integral later in the course!

Moment Generating Functions (cont.)

Why does that work?

First, consider the first derivative of our MGF,

$$
\begin{aligned}
M_X^{(1)}(t) &= \frac{\partial}{\partial t} E\left[e^{X \cdot t}\right] \\
&= E\left[\frac{\partial}{\partial t} e^{X \cdot t}\right] \\
&= E\left[X e^{Xt}\right].
\end{aligned}
$$

Now, if we substitute $t = 0$, we get,

$$
M_X^{(1)}(t)\Big|_{t=0} = E\left[X e^{X \cdot 0}\right] = E[X].
$$

Moment Generating Functions (cont.)

Now, as we have that,

$$\frac{\partial^n}{\partial t^n} e^{Xt} = X^n e^{Xt},$$

it follows that we can extend the case for $n = 1$ to any n to generate the nth moment of X, that is,

$$E[X^n] = M_X^{(n)}(t)\Big|_{t=0}.$$

As an exercise, prove the above by induction!

Application of MGF's

Oftentimes, we will find the MGF of a random variable by showing it shares an MGF with a known distribution (see Orange Book for these).

Otherwise, we derive the MGF of a random variable by using the defintion $M_X(t) = E\left[e^{Xt}\right]$.

Finding Moments with MGF

Q. Consider the continuous random variable X, which has the pdf,

$$f_X(x) = \frac{1}{2}e^{-\frac{1}{2}x}, x \geq 0.$$

By first finding the MGF of X, find $E[X^3]$. (Do not use Orange Book for this exercise).

Finding Moments with MGF (soln.)

Ans. First, we find the MGF of X,

$$
\begin{aligned}
M_X(t) &= E\left[e^{Xt}\right] \\
&= \int_0^\infty e^{xt}\frac{1}{2}e^{-\frac{1}{2}x}dx \\
&= \frac{1}{2}\int_0^\infty e^{-x\left(\frac{1}{2}-t\right)}dx \\
&= \frac{1}{2}\left(\frac{1}{2}-t\right)^{-1}\left[-e^{-x}\right]_0^\infty \\
&= \frac{1}{2}\left(\frac{1}{2}-t\right)^{-1}.
\end{aligned}
$$

Finding Moments with MGF (soln.) (cont.)

Now, to find our third non-central moment, we derive our MGF three times,

$$
\begin{aligned}
E\left[X^3\right] &= M_X^{(3)}(0) \\
&= 3\left(\frac{1}{2} - t\right)^{-4}\Bigg|_{t=0} \\
&= 48.
\end{aligned}
$$

Thus, we have found that $E\left[X^3\right] = 48$.