Probability and Statistics: From Classical to Bayesian

Week 2

8th April-15th April 2021

1 Total probability and Bayes' rule

Total probability rule and Bayes' rule: Let $A_1, ..., A_n$ be pairwise disjoint and mutually exhaustive events in a probability space. Assume P(A_i) > 0 for all i. This means that $A_i \cap A_j = \phi$; for any i $\neq$ j and $A_1 \cup A_2 \cup . \quad . \quad . \quad . \cup A_n = \Omega$. We also refer to such a collection of events as a partition of the sample space.
Let B be any other event.
(1) (Total probability rule).

$$P(B) = P(A_1)P(B|A_1) + \cdots + P(A_n)P(B|A_n)$$

(2) (Bayes' rule) Assume that P(B) > 0. Then, for each k = 1, 2, . . . n, we have

$$P(A_k|B) := \frac{P(A_k)P(B|A_k)}{P(A_1)P(B|A_1) + ... + P(A_n)P(B|A_n)}$$

Proof: The proof is merely by following the definition.
(1) The right hand side is equal to

$$P(A_1)\frac{P(B\cap A_1)}{P(A_1)}+...+P(A_n)\frac{P(B\cap A_n)}{P(A_n)} = P(B\cap A_1)+...+P(B\cap A_n),$$

which is equal to P(B) since A_i are pairwise disjoint and exhaustive.
(2) Try as an exercise.

2 Independence for general number of events

Events $A_1, \ldots, A_n$ in a common probability space are said to be independent if

$$P(A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_m}) = P(A_{i_1})P(A_{i_2})\ldots P(A_{i_m})$$

for every choice of m $\leq$ n and every choice of $1 \leq i_1 < i_2 < ... < i_m \leq n$. Check out example for clarity.

3 Random Variables and Probability Distribution

Any function $X : \Omega \to R$ is called a random variable.
If X is a discrete random variable then its range R_X is a countable set. So, we can list the elements in R_X. In other words, we can write

$$R_X = \{x_1, x_2, x_3, ...\}$$

Note that here $x_1, x_2, x_3, ...$ are possible values of the random variable X. While random variables are usually denoted by capital letters, to represent the numbers in the range we usually use lowercase letters such as x, x_1, y, z, etc. For a discrete random variable X, we are interested in knowing the probabilities of $X = x_k$. Note that here, the event $A = \{X = x_k\}$ is defined as the set of outcomes s in the sample space S for which the corresponding value of X is equal to x_k. In particular,

$$A = \{s \in S | X(s) = x_k\}$$

The probabilities of events $X = x_k$ are formally shown by the **Probability mass function (pmf)** of X.
Although the PMF is usually defined for values in the range, it is sometimes convenient to extend the PMF of X to all real numbers. If $x \notin R_X$, we can simply write $P_X(x) = P(X = x) = 0$. Thus, in general we can write :

$$P_X(x) = \begin{cases} P(X = x) & \text{if } x \text{ is in } R_X \\ 0 & \text{otherwise} \end{cases}$$

Properties of PMF:

- $0 \leq P_X(x) \leq 1 \quad \text{for all } x \in \mathbb{R}$;
- $\sum_{x \in R_X} P_X(x) = 1$;
- for any set A $\subset R_X, P(X \in A) = \sum_{x \in A} P_X(x)$.

Cumulative Distribution Function

Define $F_X : R \to [0, 1]$ by

$$F_X(t) = P\{\omega : X(\omega) \leq t\} \quad \text{for t} \in \mathbb{R}$$

In general, let X be a discrete random variable with range $R_X = \{x_1, x_2, x_3, ...\}$, such that $x_1 < x_2 < x_3 < ....$ Here, for simplicity, we assume that the range R_X is bounded from below, i.e., x_1 is the smallest value in R_X. If this is not the case then $F_X(x)$ approaches zero as $x \to -\infty$ rather than hitting zero. In particular, note that the CDF starts at 0; i.e., $F_X(-\infty) = 0$. Then, it jumps at each point in the range. In particular, the CDF stays flat between x_k and x_{k+1}.

Basic Properties of CDF : The following observations are easy to notice

- F is an increasing function on $\mathbb{R}$.
- $\lim_{t\to+\infty} F(t) = 1$ and $\lim_{t\to-\infty} F(t) = 0$
- F is right continuous, that is, $\lim_{h\to 0+} F(t+h) = F(t+) = F(t)$ for all t $\in \mathbb{R}$.

We take the these three properties of CDF proved in the previous section as the definition of a CDF or distribution function, in general.

Definition. A (cumulative) distribution function (or, CDF for short) is any function F : R $\to$ [0, 1] be a non-decreasing, right continuous function such that F(t) $\to$ 0 as t $\to -\infty$ and F(t) $\to$ 1 as t $\to +\infty$.

If (Ω, p) is a discrete probability space and X : $\Omega \to R$ is any random variable, then the function $F(t) = P\{\omega : X(\omega) \leq t\}$ is a CDF, as discussed in the previous section. However, there are distribution functions that do not arise in this manner.

As discussed for general definition of CDF, a CDF need not increase only in 'jumps', but it could be continuous summation (integral) of a continuous function too. i.e. :

$$F(x) = \int_{-\infty}^{x} f(x)\, dx$$

where $f(x)$ is continuous on $\mathbb{R}$ (could be discontinuous on a finite number of points(Why?)).
It can be seen that such a F(x) will satisfy all three properties required to be a CDF.
Such $f(x)$ (continuous, such that its integral follows properties of CDF) are known as Probability Distribution Function(pdf). You will see a lot more of this in Assignment made for this week.

Follow this link, if you are facing difficulty in understanding how this PDF arises: `https://www.youtube.com/watch?v=Fvi9A_tEmXQ`

4 Expectation

This is the part where things go interesting. Now that you are equipped with basics of Distributive Functions, lets see some of the most used terms in probability.
Let X be a random variable with distribution F_X. We shall assume that it has pmf (or, pdf) denoted by f_X.
Definition 1. The expected value of $g(X)$, defined as the quantity

$$E[g(X)] = \sum_t g(t)f(t) \text{ if f is a pmf, or}$$

$$E[g(X)] = \int_{-\infty}^{+\infty} g(t)f(t)\,dt \text{ if g is a pdf}$$

(provided the sum, or the integral converges absolutely), where g is a function from $R_x \to \mathbb{R}$.

Mean: If g is the identity map ($g(x) = x$ for all x), then E[$g(X)$] is written as E[X] and known as the mean, denoted commonly by the symbol μ.

Variance: If g is the map: $g(x) = (x - E[X])^2$,(E[X] is the mean). Then its expectation is known as variance, commonly denoted by σ^2 or $Var(X)$.

Exercise: Show that $Var(x) = E[X^2] - E[X]^2$