Probability Fundamentals: Part 1

math primer: Combinatorics, Sets, Bayes, Distributions

2019DMB02 - PGDM

NULL SPACE

Contents

0.1 Combinatorics

Starting with the fundamental principle of counting, we can assume that **experiment 1** results in any of m possible outcomes and **experiment 2** results in any of n possible outcomes. Then, if these two experiments are performed in succession, we would observe that there are a total of mn outcomes possible. Note that the below matrix lists out all the possible pairs of outcomes from experiment 1 and 2. The item (i, j) corresponds to the pair in which i was obtained in experiment 1 and j was obtained in experiment 2.

$$\begin{bmatrix} (1,1) & (1,2) & \cdots & (1,n) \\ \vdots & \vdots & \vdots & \vdots \\ (m,1) & (m,2) & \cdots & (m,n) \end{bmatrix} \tag{1}$$

As a general rule, remember that if there are a total of r experiments to be performed and the first has n_1 possibilities as outcomes, the second experiment has n_2 possibilities and the r^{th} experiment has n_r possibilities, then we will have a total possibilities from all the experiments together as:

$$n_1 \times n_2 \cdots \times n_r \tag{2}$$

0.1.1 Permutations

Ordered arrangements of elements are called **permutations**. For example if we have letters a, b, c then the permutations of these letters (elements) is given as:

$$abc, acb, bac, bca, cab, cba \tag{3}$$

Each such arrangement is called a **permutation**. Note that as a general rule, for n objects there are $n!$ permutations:

$$n(n-1)(n-2)\cdots 3.2.1 \tag{4}$$

Things become a bit more involved when we are permuting elements in which there are some objects that are alike. For example if we want to find different arrangements of the word **PEPPER** then obviously we will have a total of 6! permutation possible since there are six letters in the word. But, what if we simply interchange the alike elements in the word ? For example, if we simply interchange the two middle P's, then it wouldn't really change our permutation. For this reason we calculate the total number of permutations of PEPPER by adjusting for the permutations among the alike elements as well. So we final number of permutations would become:

$$\frac{6!}{3!2!} \tag{5}$$

Note that 3! refers to the number of permutations among the P's (which are three in number) and 2! refers to the number of permutations among the E's. As a general rule we can say:

$$\frac{n!}{n_1! n_2! \cdots n_r!} \tag{6}$$

Where there are n_1 alike elements of type 1, n_2 alike elements of type 2 and so on.

0.1.2 Combinations

If we want to form groups of r objects from a total of n objects where essentially our permutations are now **order irrelevant** then we call them **combinations**. The number of such combinations are given by:

$$\binom{n}{r} = \frac{n!}{(n-r)!r!} \tag{7}$$

0.1.3 Binomial Theorem

The Binomial theorem is a general rule that applies to polynomial expansion of a sum of two variables. It is given by:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} \tag{8}$$

As a simple example, consider finding the expansion of $(x+y)^3$. We can use the binomial theorem to expand this expression:

$$(x+y)^3 = \binom{3}{0} x^0 y^3 + \binom{3}{1} x^1 y^2 + \binom{3}{2} x^2 y + \binom{3}{3} x^3 y \tag{9}$$

$$= y^3 + 3xy^2 + 3x^2 y + x^3 \tag{10}$$

0.2 Basic Sets

The Union of many events given by $E_1, E_2, \cdots, E_n$ can be expressed as:

$$\bigcup_{n=1}^{\inf} E_n \tag{11}$$

This set of all unions consists of all outcomes in atleast one of the E_i events. In a similar manner, the event consisting of outcomes in all of the E_i events is given by a continuous intersection of these sets:

$$\bigcap_{n=1}^{\inf} E_n \tag{12}$$

Note the all important **De-Morgans** laws given by the following expressions. Also note that the superscript c refers to **complement** of a set (which is nothing but a set of elements not in the set).

$$(A \cup B)^c = A^c \cap B^c \rightarrow \left(\bigcup_{n=1}^{n} E_i \right)^c = \bigcap_{n=1}^{n} E_i^c \tag{13}$$

$$(A \cap B)^c = A^c \cup B^c \rightarrow \left(\bigcap_{n=1}^{n} E_i \right)^c = \bigcup_{n=1}^{n} E_i^c \tag{14}$$

Note an important point that **events are nothing but sets of outcomes** and hence we can denote events as sets and perform set manipulation on them. For example, we can denote the concept of **mutually exclusive events** using set notation as follows:

$$E_1 \cap E_2 = E_1 E_2 = \phi \tag{15}$$

The above equation basically means that if the intersection of two sets is the **disjoint set** then they are effectively, mutually exclusive sets or events. We can also compute the probability of the union of many mutually exclusive events as follows:

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) \rightarrow P\left(\bigcup_{n=1}^{\text{inf}}\right) = \sum_{n=1}^{\text{inf}} P(E_i) \tag{16}$$

Moving on, we can state the basic **expansion of a union of sets** that are not mutually exclusive as:

$$E \cup F = E + F - EF \tag{17}$$

Expanding the **union of three sets**:

$$E \cup F \cup G = E + F + G - EF - EG - FG + EFG \tag{18}$$

Applying the probability operator and we simply get:

$$P(E \cup F \cup G) = P(E) + P(F) + P(G) - P(EF) - P(EG) - P(FG) + P(EFG) \tag{19}$$

Notice hard enough and you'll see that a pattern emerges in terms of signs in the above summation. The combined union is the sum of all (positive sign) sets taken one at a time, all (negative sign) sets taken two at a time and all (positive sign) sets taken three at a time. We can generalize this to the union of n sets as follows in terms of probability:

$$P(E_1 \cup E_2 \cdots E_n) = \sum_{i=1}^{n} P(E_i) - \sum_{i_1 < i_2} P(E_1 E_2) + \cdots + (-1)^{n+1} P(E_1 E_2 \cdots E_n) \tag{20}$$

0.3 Conditional probability

This means finding the probability of an event E occurring given the fact that event F has occurred. Since F has already occurred we can say that this is now our new sample space, instead of the entire sample space. So essentially, we want to find the probability that E and F both occur simultaneously given that F has already occurred. It is given by:

$$P(E|F) = \frac{P(EF)}{P(F)} \tag{21}$$

Note that the above expression can alternatively be written as:

$$P(EF) = P(E|F)P(F) \tag{22}$$

Now note an important point. Suppose that there are two sets or events called E and F. Now we know that when only these two sets exist in our world, then the set E can be defined as - **the union of the intersection of E with F and the intersection of E with the complement of F.**

$$E = EF \cup EF^c \tag{23}$$

Applying the probability operator to the above sets we get:

$$P(E) = P(EF) + P(EF^c) = P(E|F)P(F) + P(E|F^c)P(F^c) \tag{24}$$

Therefore from the above expression we can say that the total probability of event E is the weighted average of the conditional probability of E that F has occurred and the conditional probability of E such that F^c has occurred or F has not occurred.

0.3.1 Bayes

We will introduce the concept of **Bayes theorem** with the help of a common exmaple. Suppose that D is the event that a person has a disease and E is the event that upon testing for the disease, the test comes out positive (Note that there can be a false positive test also - if a person does not have the disease then the test comes positive). Now if we want to find the probability that - **the person has the disease given that the result if positive**.

$$P(D|E) = \frac{P(DE)}{P(E)} \tag{25}$$

$$= \frac{P(E|D)P(D)}{P(E|D)P(D) + P(E|D^c)P(D^c)} \tag{26}$$

0.3.2 Odds

As a quick note, odds are defined as the ratio of probability of occurrence of an event to the probability of the non-occurrence of the event. It is given as:

$$\frac{P(A)}{P(A^c)} = \frac{P(A)}{1 - P(A)} \tag{27}$$

0.4 Distributions

Starting with the **Bernoulli random variable**, we define this random variable as the outcome of a single trial when the outcomes are of only two types - success and failure, encoded as 1 and 0 respectively.

$$p(0) = P(X = 0) = 1 - p \tag{28}$$

$$p(1) = P(X = 1) = p \tag{29}$$

Extending the same concept a little further, if supposing we have n independent trials, each of which is associated with a probability of success of p and probability of failure of $(1-p)$ and if we define random variable X as the **Number of successes in n trials** then what we have is essentially a **binomial random variable**.

$$p(i) = \binom{n}{i} p^i (1-p)^{n-i} \tag{30}$$

Some general properties:

$$E[X] = np \tag{31}$$

$$VAR[X] = npq = np(1-p) \tag{32}$$

$$P(X \leq i) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \tag{33}$$

0.4.1 Poisson

A random variable X taking on values $1, 2, \cdots$, is called a **Poisson random variable** with parameter λ if:

$$p(i) = P(X = i) = \frac{e^{-\lambda}\lambda^i}{i!} \tag{34}$$

Note that this is the approximation of a binomial variable when n is very large and p is small. Some general properties:

$$E[X] = \lambda \tag{35}$$

$$VAR[X] = \lambda \tag{36}$$

The derivation for **optional reading** can be presented below in a step wise manner:

- First, we assume a binomial random variable with parameters n and p such that n is very large and p quite small. Using the binomial distribution probability mass function, we can get the probability of i successes in n trials as:

$$P(X = i) = \frac{n!}{(n-i)!i!} p^i (1-p)^{n-i} \tag{37}$$

- Now we basically let $\lambda = np$ or $p = \lambda/n$ and with this we can rewrite the previous formula in terms of λ as follows :

$$P(X = i) = \frac{n!}{(n-i)!i!} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i} \tag{38}$$

$$= \frac{n(n-1)\cdots(n-i+1)}{n^i} \frac{\lambda^i}{i!} \frac{(1-\lambda/n)^n}{(1-\lambda/n)^i} \tag{39}$$

- Now we note the following approximations :

$$\left(1-\frac{\lambda}{n}\right)^n \approx e^{-\lambda} \tag{40}$$

$$\frac{n(n-1)\cdots(n-i+1)}{n^i} \approx 1 \tag{41}$$

$$\left(1-\frac{\lambda}{n}\right)^i \approx 1 \tag{42}$$

- And finally we end up with :

$$P(X=i) = \frac{e^{-\lambda}\lambda^i}{i!} \tag{43}$$

0.4.2 Geometric

Suppose now that there are many independent trials, each having a probability of success as p, such that these trials are performed until a success occurs. Our random variable X primarily defines the number of trials required until the first success is encountered.

$$P(X=n) = (1-p)^{n-1}p \tag{44}$$

Some key points:

$$E[X] = \frac{1}{p} \tag{45}$$

$$VAR[X] = \frac{1-p}{p^2} = \frac{q}{p^2} \tag{46}$$

0.4.3 Negative Binomial

Now suppose that we perform many independent trials, with each trial having the same probability of success as p and we perform trials until we accumulate r successes. Here let the primary random variable X denote the number of trials required to accumulate r successes.

$$P(X=n) = \binom{n-1}{r-1} p^r (1-p)^{n-r} \tag{47}$$

The main logic is that for us to stop conducting the trials, the r^th success has to happen at the n^th trial and therefore we count the combinations of the $r-1$ successes that must have occurred in the last $n-1$ trials. Some key points :

$$E[X] = \frac{r}{p} \tag{48}$$

$$VAR[X] = \frac{r(1-p)}{p^2} = \frac{rq}{p^2} \tag{49}$$

0.5 Cumulative Frequency distributions

The cumulative distribution function $F(X)$ for a random variable X is given by:

$$F(X) = P(X \leq x) \tag{50}$$

Note that for a distribution function $F(X)$, $F(b)$ denotes the probability that the random variable takes on values less than or equal to b. some properties about CDF functions are:

- F is non decreasing which essentially means that for $a < b$ we have $F(a) < F(b)$.
- The following can be defined as a limiting case:

$$\lim_{b \to \inf} F(b) = 1 \tag{51}$$

- The following can be defined as a limiting case:

$$\lim_{b \to -\inf} F(b) = 0 \tag{52}$$

- The CDF function is right continuous.

0.6 General points about discrete variables

Here is a quick list of some general pointers regarding expectations and variances regarding discrete random variables.

$$E[X] = \sum xp(x) \tag{53}$$

$$E[g(X)] = \sum g(x)p(x) \tag{54}$$

$$VAR[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2 \tag{55}$$

$$E[aX + b] = aE[X] + b \tag{56}$$

$$VAR[aX + b] = a^2 VAR[X] \tag{57}$$

References

[1] Sheldon Ross - A first course in probability