Lecture 4

Assigning Probabilities via Independence

Notice that neither the three axioms of probability function, $\mathbb{P}$, nor the rules derived from them in the work horse Theorem 1.2.1, tell us how to obtain

$$\mathbb{P}(A \cap B) \quad \text{after knowing } \mathbb{P}(A) \text{ and } \mathbb{P}(B).$$

In fact, unfortunately, there is no general "*intersection rule*", which shows how one can obtain $\mathbb{P}(A \cap B)$ by only knowing $\mathbb{P}(A)$ and $\mathbb{P}(B)$. Now we introduce a fundamental notion of probability theory, called independence, which does give an intersection rule for $\mathbb{P}$.

As an important bonus, we will see that this notion of independence, in fact, allows us to assign $\mathbb{P}(A)$ in certain situations, which we could not have figured out by the counting techniques of our earlier lectures. So, in essence, independence is a method of assigning probabilities which is different from the equilikely approach.

4.1 Independent Events

Definition - 4.1.1 - *If A and B are two events, then $\mathbb{P}(A \cap B)$ is called the joint probability of the two events. If we have a sequence of events, $A_1, A_2, ..., A_n$, then their joint probability is $\mathbb{P}(A_1 \cap A_2 \cap ... \cap A_n)$. Sometimes the joint probability is also written as $\mathbb{P}(A_1, A_2, ..., A_n)$, i.e., a comma stands for set intersection, "$\cap$".*

The probability, $\mathbb{P}(A \cap B)$, measures the likelihood that events A and B will occur/happen together. Independence tells us how we can find the probability of $A \cap B$ by knowing only the individual probabilities of A and B.

Definition - 4.1.2 - *Two events, A, B, are called independent*[1] *if and only if*

$$\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).}$$

[1] Also called stochastically independent or probabilistically independent.

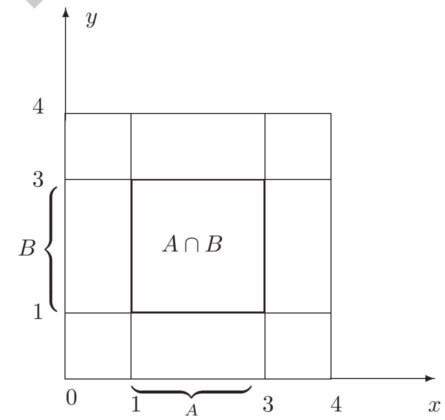
The ordinary probability, $\mathbb{P}(A)$, is also called the marginal probability. So, A, B are independent if the joint probability of A, B is equal to the product of the marginal probabilities of A and B. (Intuitively speaking, two events are independent if they do not influence or block the occurrence of each other.)

Example - 4.1.1 - Let two dice (a red and a green) be rolled so that all the 36 possible outcomes are equilikely. Let A be the event that the red die lands Four, and let B be the event that the green die lands Four. Now both A and B contain six outcomes each, while $A \cap B$ has only one outcome. So,

$$\mathbb{P}(A) = \frac{6}{36}, \quad \mathbb{P}(B) = \frac{6}{36}, \quad \mathbb{P}(A \cap B) = \frac{1}{36} = \mathbb{P}(A)\,\mathbb{P}(B).$$

Therefore, A and B are independent events. In this example the intuitive meaning of independence is clear since the red die has no interaction with the green die.

Example - 4.1.2 - Let a point be selected at random from a square, $[0, 4] \times [0, 4]$, be a square as shown in the following figure.



Let A be the event that the selected point lies in the rectangle $[1, 3] \times [0, 4]$. Let B be the event that the selected point lies in the rectangle $[0, 4] \times [1, 3]$. The event $A \cap B$ stands for the event that the point lies in $[1, 3] \times [1, 3]$, shown as the inside square in the above figure. Using the area method of computing probabilities, note that

$$\mathbb{P}(A) = \frac{8}{16}, \quad \mathbb{P}(B) = \frac{8}{16}, \quad \mathbb{P}(A \cap B) = \frac{4}{16} = \mathbb{P}(A)\,\mathbb{P}(B).$$

So, in this case A, B are independent events.

Example - 4.1.3 - The sample space corresponding to the gender of three-children families is as follows.

$$S = \{BBB,\ BBG,\ BGB,\ GBB,\ BGG,\ GBG,\ GGB,\ GGG\}.$$

Assume S is a simple sample space. Let A be the event that the family has both boys and girls, and let B be the event that the family has at most one girl.

$$\begin{aligned} A &= \{BBG, BGB, GBB, BGG, GBG, GGB\}, \\ B &= \{BBB, BBG, BGB, GBB\}. \end{aligned}$$

Therefore, $\mathbb{P}(A) = \frac{6}{8}$, $\mathbb{P}(B) = \frac{4}{8}$ and $\mathbb{P}(A \cap B) = \frac{3}{8}$. Here we notice that A, B are independent events. Why A, B should be independent is not intuitively justifiable. But our definition of independence applies. So, probabilistic independence is a bit more general concept than the intuitive concept of independence.

Independence of more than two events comes in two varieties.

Definition - 4.1.3 - *Three events, A, B, C, are called pairwise independent if (i) $\mathbb{P}(A\cap B) = \mathbb{P}(A)\cdot\mathbb{P}(B)$, (ii) $\mathbb{P}(A\cap C) = \mathbb{P}(A)\cdot\mathbb{P}(C)$ and (iii) $\mathbb{P}(B\cap C) = \mathbb{P}(B)\cdot\mathbb{P}(C)$.*

Definition - 4.1.4 - *Three events, A, B, C, are called mutually independent if (i) they are pairwise independent, and (ii) $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C)$.*

Remark - 4.1.1 - The analogous definition of pairwise and mutual independence of more than three events should be clear. Indeed, if $A_1, A_2, \cdots$ are events, then this sequence is called mutually independent if

$$\mathbb{P}\left(\bigcap_{j\in J} A_j\right) = \prod_{j\in J} \mathbb{P}(A_j), \qquad \text{for every finite subset } J \text{ of } \{1, 2, 3, \cdots\}.$$

Of course, mutual independence implies pairwise independence. The converse is false as the following example shows. The word "independent" from now on will always mean mutually independent.

An infinite sequence of events is called independent if every finite collection of them is mutually independent. We will continue this topic in the next section.

Example - 4.1.4 - We roll a fair die twice. Let $A = \{$odd number occurs on the first roll$\}$, $B = \{$odd number occurs on the second roll$\}$, $C = \{$the sum of the two face values is odd$\}$. Their probabilities are (convince yourself):

$$\begin{aligned} \mathbb{P}(A) &= \frac{18}{36} = \frac{1}{2} = \mathbb{P}(B) = \mathbb{P}(C), \\ \mathbb{P}(A\cap B) &= \frac{9}{36} = \frac{1}{4} = \mathbb{P}(A)\cdot\mathbb{P}(B), \\ \mathbb{P}(A\cap C) &= \frac{9}{36} = \frac{1}{4} = \mathbb{P}(A)\cdot\mathbb{P}(C), \\ \mathbb{P}(B\cap C) &= \frac{9}{36} = \frac{1}{4} = \mathbb{P}(B)\cdot\mathbb{P}(C). \end{aligned}$$

Thus, A, B, C are pairwise independent. However, A, B, C are not mutually independent since, $\mathbb{P}(A \cap B \cap C) = 0 \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = \frac{1}{8}$.

4.2 Bernoulli Process & Binomial Model

An experiment is called a Bernoulli experiment if it can have only two possible outcomes. (Typically the outcomes are labeled Heads (H) and Tails (T), using a coin toss analogy.) When the same coin is tossed (i.e., a Bernoulli experiment is performed) several times, the resulting collection of Bernoulli experiments is called a Bernoulli process. If we count the number of heads in a fixed segment of a Bernoulli process the resulting experiment is called a binomial experiment.

One of the benefits of the independence concept is that we can find probabilities dealing with total number of heads in a binomial experiment.

Example - 4.2.1 - **(2 tosses — fair coins)** We toss a fair coin twice. Let $A = \{H$ occurs on the first toss$\}$ and let $B = \{H$ occurs on the second toss$\}$. Then A and B turn out to be independent. To see this, note that the sample space is $S = \{HH, HT, TH, TT\}$. It is quite reasonable to postulate that the singleton sets of S have equal probabilities — making S a simple sample space. Now

$$A = \{HH, HT\}, \qquad B = \{HH, TH\}, \qquad A \cap B = \{HH\}.$$

Therefore, $\mathbb{P}(A) = \frac{1}{2} = \mathbb{P}(B)$ and $\mathbb{P}(A \cap B) = \frac{1}{4}$. We see that A, B are independent events since $\mathbb{P}(A \cap B) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = \mathbb{P}(A) \cdot \mathbb{P}(B)$.

Example - 4.2.2 - **(2 tosses — unfair coins)** Let us take the last example one notch higher. Suppose now the coin is unfair, so that the probability of seeing H on a toss is p and it is not necessarily equal to $\frac{1}{2}$. The issue is how to assign probabilities to the singleton sets of the sample space $S = \{HH, HT, TH, TT\}$. We will go backwards. Once again let us take the two events $A = \{H$ occurs on the first toss$\}$ and $B = \{H$ occurs on the second toss$\}$. We may argue that the two events should be independent regardless the coin is fair or unfair since the two tosses should not "influence each other". This postulate turns out to be enough to give us a probability space. Here is how. To find the probability space that makes the events A and B independent, let $\mathbb{P}(\{HH\}) = \alpha$, $\mathbb{P}(\{HT\}) = \beta$, $\mathbb{P}(\{TH\}) = \gamma$, $\mathbb{P}(\{TT\}) = \delta$. We must have $\alpha + \beta + \gamma + \delta = \mathbb{P}(S) = 1$. We need three more equations. The postulated independence of A and B gives that

$$\alpha = \mathbb{P}(\{HH\}) = \mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B) = (\alpha + \beta)(\alpha + \gamma),$$

since $\mathbb{P}(A) = \mathbb{P}(\{HH, HT\}) = \alpha + \beta$ and $\mathbb{P}(B) = \mathbb{P}(\{HH, TH\}) = \alpha + \gamma$. Using the fact that the probability of a head on a toss is, say p,

$$\begin{aligned} \alpha + \beta &= \mathbb{P}(A) = \mathbb{P}(H \text{ on the first toss}) = p, \\ \alpha + \gamma &= \mathbb{P}(B) = \mathbb{P}(H \text{ on the second toss}) = p. \end{aligned}$$

These four equations with four unknowns give us the unique solution (verify it!)

$$\mathbb{P}(\{HH\}) = p^2, \quad \mathbb{P}(\{HT\}) = \mathbb{P}(\{TH\}) = p(1-p), \quad \mathbb{P}(\{TT\}) = (1-p)^2.$$

The moral of the story is that the probability of any singleton event consists of the product of the probabilities of H's and T's that make up the singleton set. For instance, $\mathbb{P}(\{TT\}) = (1-p)(1-p)$ since the probability of a T on a single toss is $(1-p)$. The next example considers the same coin but tossed 3 times.

Example - 4.2.3 - **(3 tosses — unfair coins)** Suppose now the unfair coin is tossed three times. Again, the probability of seeing an H on a toss is p and it is not necessarily equal to $\frac{1}{2}$. The issue is how to assign probabilities to the singleton sets of the sample space

$$S = \{HHH,\ HHT,\ HTH,\ HTT,\ THH,\ THT,\ TTH,\ TTT\}.$$

By repeating an argument similar to the one used in the last example, the probabilities of singleton events are

$$\begin{aligned}&\mathbb{P}(\{HHH\}) = p^3, \quad \mathbb{P}(\{HHT\}) = \mathbb{P}(\{HTH\}) = p^2(1-p),\\ &\mathbb{P}(\{HTT\}) = p(1-p)^2, \quad \mathbb{P}(\{THH\}) = p^2(1-p),\\ &\mathbb{P}(\{THT\}) = \mathbb{P}(\{TTH\}) = p(1-p)^2, \quad \mathbb{P}(\{TTT\}) = (1-p)^3.\end{aligned}$$

Again, the moral of the story is that the probability of any singleton event consists of the product of the probabilities of H's and T's that make up the singleton event. For instance, $\mathbb{P}(\{HTT\}) = p(1-p)(1-p)$ since the probability of H on a single toss is p and a T on a single toss is $(1-p)$. This shows that the probabilities of various events of interest are as follows.

$$\begin{aligned}\mathbb{P}(\{0 \text{ heads in 3 tosses}\}) &= \mathbb{P}(\{TTT\}) = (1-p)^3,\\ \mathbb{P}(\{1 \text{ head in 3 tosses}\}) &= \mathbb{P}(\{TTH,\ THT,\ HTT\}) = 3p(1-p)^2,\\ \mathbb{P}(\{2 \text{ heads in 3 tosses}\}) &= \mathbb{P}(\{HHT,\ HTH,\ THH\}) = 3p^2(1-p),\\ \mathbb{P}(\{3 \text{ heads in 3 tosses}\}) &= \mathbb{P}(\{HHH\}) = p^3.\end{aligned}$$

We may write this compactly as follows,

$$\mathbb{P}(\{k \text{ heads in 3 tosses}\}) = \binom{3}{k} p^k (1-p)^{3-k}, \qquad k = 0, 1, 2, 3.$$

Continuing this reasoning a bit further, the probability of k heads, in an n-tosses-long binomial experiment, is as follows, called the binomial model.

$$\boxed{\mathbb{P}(k\ H\text{'s in } n \text{ tosses}) \;=\; \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, 2, \cdots, n.}$$

4.3 Geometric & Negative Binomial Models

Recall that "a Bernoulli process" stands for a sequence of trials for which

- the trials are independent, (i.e., trials do not influence each other)
- each trial can have only two possible outcomes, (called Heads and Tails, denoted by H and T), and
- the probability of an H remains the same from trial to trial, and we will usually denote this probability by p.

For a Bernoulli process, several questions of interest can be raised — resulting into several different types of random experiments. We have already seen one example, a binomial experiment. In the binomial experiment we restricted our attention over the first n Bernoulli trials, and asked "what is the probability of obtaining k number of heads over these n trials"? Answer:

$$\boxed{\mathbb{P}(k \text{ heads in } n \text{ tosses}) = \binom{n}{k} p^k (1-p)^{n-k}; \qquad k = 0, 1, 2, \cdots, n.}$$

We derived this result in the last section. Binomial experiment is just one of several experiments one can encounter over a Bernoulli process. Now we present a few more experiments.

Example - 4.3.1 - **(Geometric experiment)** Over the whole Bernoulli process what is the probability that k tails will be observed to get the first head? Answer:

$$p(1-p)^k, \qquad k = 0, 1, 2, \cdots.$$

The justification is quite easy. The sample space is

$$S = \{H,\ TH,\ TTH,\ TTTH,\ TTTTH,\ TTTTTH,\ \cdots\}.$$

The event, "k tails before the first head", consists of only one element $\underbrace{TTT\cdots T}_{k \text{ of them}}H$.

Since the trials are independent, its probability is

$$\underbrace{(1-p)(1-p)\cdots(1-p)}_{k \text{ of them}}\,p = p(1-p)^k.$$

Example - 4.3.2 - **(Geometric & negative binomial series)** An easy way to verify the following expression is by multiplying both sides with $(1-r)$,

$$\frac{1-r^{n+1}}{1-r} \;=\; 1 + r + r^2 + r^3 + \cdots + r^{n-1} + r^n.$$

This equality holds for all values of r. However, if we restrict $r \in (-1, 1)$, and let n get large, we get the expression for a geometric series:

$$\boxed{\frac{1}{1-r} \;=\; 1 + r + r^2 + r^3 + \cdots \;=\; \sum_{k=0}^{\infty} r^k, \qquad -1 < r < 1.}$$

Taking a derivative with respect to r gives another series:

$$\frac{1}{(1-r)^2} \;=\; 1 + 2r + 3r^2 + 4r^3 + \cdots \;=\; \sum_{k=0}^{\infty} \binom{2+k-1}{2-1} r^k, \qquad -1 < r < 1.$$

More derivatives give more varieties. For instance, when we take another derivative of this (and divide both sides by 2) we get

$$\frac{1}{(1-r)^3} \;=\; 1 + 3r + 6r^2 + 10r^3 + \cdots \;=\; \sum_{k=0}^{\infty} \binom{3+k-1}{3-1} r^k, \qquad -1 < r < 1.$$

After taking $n-1$ derivatives of the geometric series, and diving by $(n-1)!$, the general expression (called a negative binomial series) is as follows.

$$\boxed{(1-r)^{-n} = \sum_{k=0}^{\infty} \frac{n(n+1)\cdots(n+k-1)}{k!}\, r^k = \sum_{k=0}^{\infty} \binom{n+k-1}{n-1} r^k.}$$

Example - 4.3.3 - **(Negative binomial experiment)** A coin is tossed until the 10-th head is observed. Let p be the probability of observing a head on a toss.

- (i) (Negative binomial experiment) What is the probability that k tails will be observed to get the 10-th head?
- (ii) What is the probability that an even number of tosses will be needed to observe the 10-th head?

(i) Write the sample space, S, as

$$\left\{ \underbrace{HH\cdots H}_{10\,H's},\ \underbrace{--\cdots-}_{9\,H's\text{ and }1T} H,\ \underbrace{--\cdots-}_{9\,H's\text{ and }2T's} H,\ \cdots \right\}$$

Observing k tails to get 10 heads involves outcomes of the type $\underbrace{--\cdots-}_{9\,H's\text{ and }k\ T's} H$.

Each one of such outcomes has the same probability, namely

$$p^9\,(1-p)^k\,p.$$

There are exactly $\binom{9+k}{9} = \binom{10+k-1}{10-1}$ such outcomes. Hence, the probability of observing k tails to get 10 heads is

$$\binom{10+k-1}{10-1} p^{10}\,(1-p)^k, \qquad k = 0, 1, 2, \cdots.$$

(ii) The probability of the event is

$$\binom{10-1}{9} p^{10} + \binom{10+2-1}{9} p^{10}(1-p)^2 + \binom{10+4-1}{9} p^{10}(1-p)^4 + \cdots$$

To simplify this expression, we use the negative binomial series for $x = 1-p$ and $x = -(1-p)$, with $n = 10$,

$$\begin{aligned} p^{-10} &= (1-(1-p))^{-10} = \sum_{k=0}^{\infty} \binom{10+k-1}{9} (1-p)^k, \\ (2-p)^{-10} &= (1-(-(1-p)))^{-10} = \sum_{k=0}^{\infty} \binom{10+k-1}{9} (-(1-p))^k. \end{aligned}$$

Adding these, dividing by 2, and multiplying by p^{10}, the probability is

$$p^{10} \sum_{k=0}^{\infty} \binom{10+2k-1}{9} (1-p)^{2k} = \frac{1+\{p/(2-p)\}^{10}}{2}.$$

Example - 4.3.4 - **(Shifted geometric and negative binomial experiments)** Over the entire (infinite) Bernoulli process, what is the probability that k trials will be observed to get the first head? Answer:

$$p(1-p)^{k-1}, \qquad k = 1, 2, 3, \cdots.$$

The justification is quite similar to the one used in the geometric experiment. The difference being that now we count the trial that results into the first head as well.

Similarly, over the entire (infinite) Bernoulli process, what is the probability that k trials will be observed to get the first 10 heads? Answer: (justify it)

$$\binom{k-1}{10-1} p^{10}(1-p)^{k-10}, \qquad k = 10, 11, 12, \cdots.$$

By the way, the names geometric and shifted geometric are some times interchanged in various books. The same goes with the negative binomial and shifted negative binomial experiments, one of which is sometimes also called a Pascal experiment.

4.4 Exercises

Exercise - 4.4.1 - *If we select a card at random from the standard deck of 52 cards, let A be the event that a face-card is drawn (i.e., a Jack, or Queen or King is drawn). Let B be the event that the drawn card is a club. Show that A and B are independent events.*

Exercise - 4.4.2 - *If $\mathbb{P}(\{H\}) = p$ and we keep tossing the coin until we see the first head, then find the probability of $\{13\ T$'s before the first $H\}$.*

Exercise - 4.4.3 - *If two events with positive probabilities are disjoint, can they be independent? Justify your answer.*

Exercise - 4.4.4 - *Show that if A, B are independent events then so are A, B^c.*

Exercise - 4.4.5 - *A fair coin is tossed and then a coin, whose probability of heads is 0.6, is tossed. Write the sample space as a product set.*

Exercise - 4.4.6 - *For Exercise 4.4.5 write the probability measure when every subset of the sample space is an event.*

Exercise - 4.4.7 - *For Exercise 4.4.5 find the probability that at least one head will occur.*

Exercise - 4.4.8 - *For Exercise 4.4.5 either write or borrow the following `Matlab` code to simulate the experiment one million times and obtain a relative frequency estimate of the probability of seeing at least one head.*