Statistics 357 : Axioms of Probability

Probability

Ω : set of possible outcomes (sample space)

$\mathcal{F}$: set of possible events

P : a rule or function that assigns probability to each event

In finite outcomes games, and some infinite outcome games, one can specify Ω and then $\mathcal{F}$ = set of all possible subsets of Ω. This setting is very useful for studying simple gambling games and trying to sort out the classical paradoxes of probability.

In some games or experiments the set of possible outcomes is the set of real numbers $\mathcal{R}$, a set of ordered pairs of real numbers $\mathcal{R} \times \mathcal{R} = \mathcal{R}^2$. In this case the set of possible events is not the set of possible subsets of $\mathcal{R}$.

The axioms of probability are some basic rules or properties that P needs to satisfy. All other properties of probability functions can be deduced from these. These will be useful when we try to understand CDF's and how they can be used to derive certain formulae for pdf's and probability mass functions of random variables that are functions of simpler underlying random variables. This is useful for understanding how to price various insurance policies and for Monte Carlo simulation methods which are used extensively in insurance and finance amongst other areas.

Properties of events; $A, B, \ldots$ will denote events

1. Ω and $\emptyset$ (= empty set) are events

2. If A is an event then A^c (= complement of the set A) is an event

3. If A_1, A_2, A_3, ... are events then
$$E = \bigcup_{i=1}^{\infty} A_i$$
is an event.

Property 3 deals with countable unions of events. Since $\emptyset$ is an event, then taking $A_1 = A$, $A_2 = B$ and $A_i = \emptyset$ for all $i \geq 3$ we then have that
$$A \cup B = \bigcup_{i=1}^{\infty} A_i$$
is an event.

Property 3 is also useful later. For example when we will deal with a Poisson random variable X, we might consider the event that the outcome is an even integer. For this we can consider the *event* A_i = event that X takes on the value $2(i-1)$. Then the outcome that X is even (call this the set E) can be written as
$$E = \bigcup_{i=1}^{\infty} A_i \ .$$
The rules of probability then allow us to obtain a formula for the probability of E in terms of the *distribution* of X.

Probability Measures or Probability Functions

A probability function P is a rule (function) to assign to each event a number in $[0,1]$. Formally it is a mapping from the set of events to $[0,1]$

$$P : \mathcal{F} \mapsto [0,1]$$

In general it is an into mapping, not an onto mapping. The probability function has some properties for reasons of mathematical or logical consistency. These can be memorized on a case by case basis, but then what happens if some new situation comes up. There is a set of Axioms or basic defining properties of P for any probability model. From these Axioms all other properties can be derived, some easily and some with careful mathematical manipulations.

Three axioms of Probability.

1. $P(\Omega) = 1$

2. $P(A) \geq 0$ for any event A

3. For countable sequences of disjoint events $A_1, A_2, A_3, \ldots$

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Here and in the text we derive some of the basic familiar properties of probability.

Axiom 3 implies that $P(\emptyset) = 0$.

First we need to show : Suppose A and B are events and that $A \subseteq B$. Then $P(A) \leq P(B)$

Proof : $B = A \cup (A^c \cap B)$, and the RHS is the union of disjoint sets. Since $A^c \cap B$ is an event, then Axiom 2 applies to it. Therefore

$$\begin{aligned} P(B) &= P(A) + P(A^c \cap B) \\ &\geq P(A) + 0 \quad \text{using Axiom 2} \\ &= P(A) \end{aligned}$$

Let $A_i = \emptyset$ for all i. By definition the A_i are disjoint. By Axiom 3 we then have

$$\begin{aligned} P(\emptyset) &= P\left(\bigcup_{i=1}^{\infty} A_i\right) \\ &= \sum_{i=1}^{\infty} P(A_i) \\ &= \sum_{i=1}^{\infty} P(\emptyset) \end{aligned}$$

Suppose $\alpha = P(\emptyset)$. Either $\alpha = 0$ or $0 < \alpha \leq 1$. If $0 < \alpha \leq 1$ then

$$\alpha = \sum_{i=1}^{\infty} \alpha = \infty$$

which is impossible. Therefore $\alpha = 0$.

As before Axiom 3 has simplifies to a nice form for finite unions. Suppose A and B are disjoint events. We have seen above that their union is an event. Since $\emptyset$ is an event, then taking $A_1 = A$, $A_2 = B$ and $A_i = \emptyset$ for all $i \geq 3$ we then have that

$$A \cup B = \bigcup_{i=1}^{\infty} A_i$$

and all the A_i are disjoint.

Therefore

$$\begin{aligned} P(A \cup B) &= P\left(\bigcup_{i=1}^{\infty} A_i\right) \\ &= \sum_{i=1}^{\infty} P(A_i) \\ &= P(A_1) + P(A_2) + \sum_{i=3}^{\infty} P(\emptyset) \\ &= P(A) + P(B) + 0 \\ &= P(A) + P(B) \ . \end{aligned}$$

1. If A is an event then A^c is an event.

   Proof : By Axiom 1 and 3

   $$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$$

   Thus $P(A^c) = 1 - P(A)$

2. Suppose A and B are events, not necessarily disjoint. Then

   $$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

   Proof :

   Notice that

   $$A \cup B = (A \cap B^c) \cup B$$

   Therefore

   $$P(A \cup B) = P(A \cap B^c) + P(B) \tag{1}$$

Next notice that

$$A = (A \cap B) \cup (A \cap B^c)$$

Therefore

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

and thus

$$P(A^c \cap B) = P(A) - P(A \cap B)\ .$$

Now substitute this expression into (1).

3. Suppose events $B_1, B_2, \ldots, B_n$ partition Ω. This means that the B_i are disjoint and their union is Ω. Then for any event A

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_n)\ .$$

Notice that the RHS is a union of disjoint events. Axiom 3 then gives

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

The analogous result holds for a countable partition $B_i, i = 1, 2, 3, \ldots$ of Ω.

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

4. Suppose events $B_1, B_2, \ldots, B_n$ partition Ω. This means that the B_i are disjoint and their union is Ω. Then for any event A

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_n)$$

Axiom 3 then gives

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

The analogous result holds for a countable partition $B_i, i = 1, 2, 3, \ldots$ of Ω.

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

The student should read Section 1.4. Proposition B and C are needed later. The notion of permutations are needed to study order statistics in Chapter 4.

Why are we interested in this formal idea of a probability space and probability measure?

- the definition dictates how and why we can use the notion of distribution function (cdf, pdf, pmf) and how joint distributions behave.
- the rules or definitions force the same answer no matter which sample space we decide to use, or perhaps not even work with one explicitly (eg the notion of distribution function), and that probabilities always apply to events or subsets of the sample space.
- sometimes an explicit probability space makes a certain problem easier - eg prisoner's paradox (which also uses conditional probability). In this course we will usually not need to be explicit in giving a sample space.

Section 1.5 Conditional Probability

The prototype simple example of this is choosing balls without replacement from an urn or box. See urn-eg1.pdf

For events A, B, with $P(B) > 0$ define the *conditional probability of event A occurring given that event B occurs* as

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{2}$$

When $P(B) = 0$ the conditional probability $P(A|B)$ is undefined (cannot divide by 0). This definition means that

$$P(A \cap B) = P(A|B)P(B) \tag{3}$$

which is a useful product rule. It is useful when the two terms on the RHS are know or given in some appropriate form.

A simple rewording of the urn example applies to insurance. Consider drivers between 20 and 25 years of age. Consider events

$$\begin{aligned} M &= \text{male driver} \\ F &= \text{female driver} \\ A &= \text{accident in next 12 months} \end{aligned}$$

Of course $M^c = F$. An insurance company will have the values of

$$P(A|M) \text{ and } P(A|F)$$

and from this can calculate $P(A)$ if they also have the probability of a driver being male.

$$\begin{aligned} P(A) &= P((A \cap M) \cup (A \cap F)) \\ &= P(A \cap M) + P(A \cap F) \\ &= \frac{P(A \cap M)}{P(M)} \times P(M) + \frac{P(A \cap F)}{P(F)} \times P(F) \\ &= P(A|M)P(M) + P(A|F)P(F) \end{aligned}$$

Thus

$$P(A) = P(A|M)P(M) + P(A|F)P(F)\ .$$

There are some interesting and useful consequences of conditional probability.

Suppose P is a probability measure on the sample space and set of events Ω and $\mathcal{F}$. Let B be a fixed event with $P(B)$

Notice that we can consider a new *Sample Space* as B and the events in $\mathcal{F}$ which are subsets of B, that is the collection of sets

$$\mathcal{B} = \{D : D = A \cap B, A \in \mathcal{F}\} \ .$$

Now consider

$$Q(D) = P(D|B) = P(A \cap B|B)$$

for some A such that $D = A \cap B$. (There is such an A by the definition of $\mathcal{B}$.) The function Q is then a probability measure (probability function) on the sample space B and set of events $\mathcal{B}$.