3 Basic Definitions of Probability Theory

3defprob.tex: Feb 10, 2003

Classical probability
Frequency probability
axiomatic probability

Historical developement: Classical → Frequency → Axiomatic

The Axiomatic definition encompasses the Classical and Frequency definitions of probability

Note that in a number of places in these notes I use *axiom* as a synomom for *assumption*.

Classical probability

Definition *Classical probability (MBG): If a random experiment (process with an uncertain outcome) can result in n mutually exclusive and equally likely outcomes, and if n_A of these outcomes has an attribute A, then the probability of A is the fraction (n_A/n).*

This notion of probabilty had its conception in the study of games of chance; in particular, fair games of chance.

E.g. if one tosses a coin there are two mutually exclusive outcomes: head or tail. Of these two outcomes, one is associated with the attribute heads; one is associated with the attribute tails. If the coin is fair each outcome is equally likely. In which case, $\Pr[head] = \frac{n_A}{n} = \frac{1}{2}$, where $n = 2$ and n_A is the number of possible outcomes associated with a head (1).

Consider some other examples:

1. The roll of a die: There are 6 equally likely outcomes. The probability of each is 1/6.
2. Draw a card from a deck: There are 52 equally likely outcomes.
3. The roll of two die: There are 36 equally likely outcomes ($6x6$): 6 possibilities for the first die, and 6 for the second. The probability of each outcome is 1/36.

4. Drawing (with replacement) four balls from an urn with an equal number of red, white, and blue balls: There are 81 possible outcomes ($3x3x3x3 = 3^4$). For example, {red, white, white, blue} is an outcome which is a different outcome from {white, white, red, blue}. The probability associated with each outcome is 1/81.
5. The toss of two coins: The four possible outcomes are (H,H), (H,T), (T,H) and (TT). The probability of each is 1/4.
6. The draw of two cards: There are 52^2 possible outcomes.

Terms to note in the definition of classical probability are *random*, *n*, *mutually exclusive*, and *equally likely*.

Axiom *A Basic assumption in the definition of classical probability is that n is a finite number; that is, there is only a finite number of possible outcomes. If there is an infinite number of possible outcomes, the probability of an outcome is not defined in the classical sense.*

Definition *mutually exclusive: The random experiment result in the occurance of only one of the n outcomes. E.g. if a coin is tossed, the result is a head or a tail, but not both. That is, the outcomes are defined so as to be mutually exclusive.*

Definition *equally likely: Each outcome of the random experiment has an equal chance of occuring.*

Definition *random experiment (Gujarati p*23*): A random experiment is a process leading to at least two possible outcomes with uncertainty as to which will occur.*

Definition *sample space: The collection of all possible outcomes of an experiment. If I had to guess, I would say it is called "sample space" because it is the collection (set) of all possible samples.*

As important thing to note is that classical probabilities can be deduced from knowledge of the sample space and the assumptions. Nothing has to be observed in terms of outcomes to deduce the probabilities.

Frequency Probability

What if n is not finite? In that case, the Classical definition is not applicable. What if the outcomes are not equally likely? Again, the Classical defintion of probability is not applicable. In such cases, how might we define the probablity of an outcome that has attribute A.

Definition *We might take a random sample from the population of interest and identify the proportion of the sample with attribute A. That is, calculate*

$$\text{Relative freq of } A \text{ in the sample} = \frac{\text{number of obser in the sample that possess attribute } A}{\text{number of obser in the sample}}$$

Then assume "Relative freq of A in the sample" is an estimate of $Pr[A]$

The foundation of this approach is that there is some $\Pr[A]$. We cannot deduce it, as in Classical probability, but we can estimate it.

For example, one tosses a coin, which might or might not be fair, 100 times and observes heads on 52 of the tosses. One's estimate of the probability of a head is $.52$. Frequency probability allows to estimate probabilities when Classical probability provides no insight.

Axiomatic Approach to Probability

Put simply, the axiomatic approach build up probability theory from a number of assumptions (axioms).

Axiom *there is some sample space,* Ω*: the collection of all possible outcomes of an experiment*

For example, if the experiment is tossing a coin $\Omega = \{H, T\}$. If the experiment is rolling a die $\Omega = \{1, 2, 3, 4, 5, 6\}$. If the experiment is tossing two coins $\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$. If the experiment is taking a sample of one from the U.S. population, $\Omega = \{$population of the U.S.$\}$. If the experiment is drawing a random interger, Ω is the set of all intergers (there are an infinite number of these).

Definition *Event A (MGB p*15*): a subset of the sample space. The set of all events associated with an experiment is defined as the "event space"*

Note that the above is not truly a defintion (it does not identify the necc and suff conditions). Being a subset of Ω ($A \subseteq \Omega$) is a necessary but not a sufficient condition for being an event. That is, for some sample spaces, there are subset that are not events; event space can be a strict subset of sample space. If the number of outcomes in Ω is finite, then all subsets of Ω are events.

In general, ”none of the outcomes” is an event, typically denoted ϕ, and ”one of the outcomes” is an event, denoted Ω. That is, Ω is both the sample space and an event, a certain event.

The event set is difficult to formally define.

Some examples of event sets:

1. the toss of a single coin: $\Omega = \{H,T\}$. The possible events are H, T, neither H nor T (ϕ), and either H or T (Ω).
2. the toss of two coins: $\Omega = \{(H,H), (H,T), (T,H), (T,T)\}$. How many events are there? (H,H), (H,T), (T,H), (T,T), ”one of the these outcomes”, ”none of these outcomes”, ”at least one head” ((H,H), (H,T) or (T,H)), ”at least one tail” ((H,T), (T,H) or (T,T)). Can you think of others?
3. drawing 4 cards from a deck: Events include all spades, sum of the 4 cards is ≤ 20 (assuming face cards have a value of zero), a sequence of integers, a hand with a $2,3,4$ and 5. There are many more events.
4. the example on page 16 of MGB: randomly choose a lightbulb to observe and record time it takes to burn out. $\Omega = \{x : x \geq 0\}$
   Define event $A(k,m)$ as $A \equiv \{x : k \leq x \leq m\}$

In econometrics we will be concerned with the probablility of events, and need only to define those events we care about. For example, we might be concerned with the probability that the winner is a female or that the sample mean is 5 given that the population mean is 0.

Typically, one uses capital letters to represent events. E.g. A, B, C; or A_1, A_2, A_3, etc. I will use $\mathcal{A}$ to repesent the set of all possible events.

Axiom $\Omega \in \mathcal{A}$. *That is, the event that one of the outcomes is occurs in an event. Note that* $Pr[\Omega] = 1$.

Axiom *if* $A \in \mathcal{A}$ *then* $\bar{A} \in \mathcal{A}$ *where* $\bar{A}$ *is the compliment of* A. *That is, if* A *is an event, then not* A *is an event*

Axiom *if* A_1 *and* $A_2 \in A$ *then* $A_1 \cup A_2 \in \mathcal{A}$. *That is two events together is an event.*

These three assumptions imply the following:

1. $\phi \in \mathcal{A}$ For example, this implication follows from the first two Axioms.
2. if $A_1, A_2, \ldots A_n \in \mathcal{A}$ then $\bigcup_{i=1}^{n} A_i$ and $\bigcap_{i=1}^{n} A_i \in \mathcal{A}$

Any collection of events that fulfills the above three assumptions/axioms is called a Boolean algebra

Definition *The axiomatic defintion of probability: With these definitions and assumptions in mind, a probability function* $Pr[.]$ *maps events in* $\mathcal{A}$ *onto the* $0,1$ *inteval. That is, there exists some function that identifies the probabilty associated with any event in* $\mathcal{A}$*. It fufills the following axioms:*

$$Pr[A] \geq 0 \; \forall A \in \mathcal{A}$$

$$Pr[\Omega] = 1$$

If $A_1, A_2, \ldots, A_n$ *is a sequence of mutually exclusive events* $(A_i \cap A_j = \phi \; \forall i,j \; i \neq j)$ *and if* $\cup_{i=1}^{n} A_i \in \mathcal{A}$ *then* $Pr[\cup_{i=1}^{n} A_i] = \sum_{i=1}^{n} Pr[A_i]$.

From this definition with its three axioms, it is possible to deduce a bunch of properties that one would expect a probability function to have (MGB 24). Including

- $\Pr[\phi] = 0$, and
- $\Pr[\bar{A}] = 1 - \Pr[A]$
- If A and $B \in \mathcal{A}$ and $A \subseteq B$ then $\Pr[A] \leq \Pr[B]$
- If $A_1, A_2, \ldots A_n \in \mathcal{A}$ then $\Pr[A_1 \cup A_2 \cup \ldots \cup A_n] \leq \Pr[A_1] + \Pr[A_2] + \ldots + \Pr[A_n]$. This is called Boole's inequality.
- If A and $B \in \mathcal{A}$ then $\Pr[A] = \Pr[A \cap B] + \Pr[A \cap \bar{B}]$. That is, $\Pr[A] = \Pr[AB] + \Pr[A\bar{B}]$
- If A and $B \in \mathcal{A}$ then $\Pr[A \cap \bar{B}] = \Pr[A] - \Pr[A \cap B]$. That is, $\Pr[A - B] = \Pr[A] - \Pr[AB]$

Put simply, the axiomatic definition builds up the notion of a probability function from a number of assumptions/axioms. The properties of that probablity function follow from the assumptions.

Note that if the number of outcomes are finite and equally likely then one has the Classical world of probability. Also note that the Frequency defintion assumes the existence of the probability function $\Pr[A]$. The axiomatic approach subsumes the Classical and Frequency approaches.

Probability of event A conditional on event B occuring

Let $\Pr[A|B] \equiv$ Probability of event A conditional on event B occuring, assuming $\Pr[B] > 0$. Otherwise $\Pr[A|B]$ is not defined.

Consider the following: $\Omega \equiv \{(x,y) : 0 \leq x \leq 100,\ 0 \leq y \leq 100\}$. Now define two sets: $A \equiv \{(x,y) : 20 \leq x \leq 50,\ 40 \leq y \leq 60\}$ and $B \equiv \{(x,y) : 0 \leq x \leq 40,\ 50 \leq y \leq 100\}$. Note that these two sets define events. Note that in this example the two sets partially intersect. Consider other examples where $A \subset B$ or $A = B$. Draw some pictures.

Definition *of conditional probability (*1*)*

$$Pr[A|B] = \frac{Pr[AB]}{Pr[B]}$$

if $Pr[B] > 0$*. Read* $Pr[AB] \equiv Pr[A \cap B] \equiv Pr[A \text{ and } B]$.

So (2)

$$\Pr[B|A] = \frac{\Pr[AB]}{\Pr[A]}$$

if $\Pr[A] > 0$. One can rearrange (1) and (2) to obtain (1a) and (2a)

$$\Pr[AB] = \Pr[A|B]\Pr[B]$$

and

$$\Pr[BA] = \Pr[B|A]\Pr[A]$$

Combining (1a) and (2a), one obtains (3)

$$\Pr[AB] = \Pr[A|B]\Pr[B] = \Pr[B|A]\Pr[A] = \Pr[BA]$$

What is the intuition behind this being the definition of conditional probability? In conditional probability, the sample space is effectively reduced to B; that is, what is the probability that A will happen given that ones lives in a world where B prevails. So, $\Pr[A|B]$ is the probability that both A and B occur, $\Pr[AB]$, as a proportion of the probability of B, $\Pr[B]$.

Note the following, which following from (2) and (3):

$$\Pr[B|A] = \frac{\Pr[AB]}{\Pr[A]} = \frac{\Pr[A|B]\Pr[B]}{\Pr[A]}$$

This says that if one knows, $\Pr[A|B]$, $\Pr[B]$, and $\Pr[A]$, one can determine $\Pr[B|A]$. What is this result called? Baye's theorem?

When are events A and B independent?

Start by considering a case where $\Pr[A] > 0$ and $\Pr[B] > 0$, but $A \cap B = \phi$, so $\Pr[AB] = 0$. If these assumptions hold, $\Pr[A|B] = \Pr[B|A] = 0$; that is, the two events are mutually exclusive. Draw a Venn diagram. In this case, are A and B independent? NO. One precludes the other.

When do we say A and B are independent?

Definition *A and B are independent iff any of the following are true:*

$$Pr[AB] \equiv Pr[A \cap B] = Pr[A]\,Pr[B]$$

$$Pr[A|B] = Pr[A] \text{ if } Pr[B] > 0$$

$$Pr[B|A] = Pr[B] \text{ if } Pr[A] > 0$$

Note that these are three equivalent statements: each implies the other two. Can you show this?

The following can be deduced:
$\Pr[A] > 0$, $\Pr[B] > 0$ and $A \cap B = \phi \Rightarrow A$ and B are not independent

$\Pr[A] > 0$, $\Pr[B] > 0$ and A and B are independent $\Rightarrow A \cap B \neq \phi$

Independence of n events is more complicated than independence of 2 events.

Some student examples of independence and dependence

Example 1

Suppose a survey classified the population as male or female, and as favoring or opposing the death penalty. Suppose, the proportions in each category were

	Death	not Death
Male	.459	.441
Female	.051	.049

. There are four events: male, female, favor death penalty, don't favor death penalty. In this case, the probability of an individual favoring the death penalty, conditional on being male is

$$\Pr(D|M) = \frac{\Pr(D,M)}{\Pr(M)} = \frac{.459}{.459 + .441} = .51$$

and $\Pr(D) = .459 + .051 = .51$, so the the probability of favoring the death penalty equals the probability of favoring the death penalty conditional on one being male, so the two events (favoring death and being male) are independent. However if the proportions were

	Death	not Death
Male	.27	..21
Female	.24	.28

$$\begin{aligned} \Pr(D|M) &= \frac{\Pr(D,M)}{\Pr(M)} \\ &= \frac{.27}{.27 + .21} = .5625 \end{aligned}$$

and $\Pr(D) = .27 + .24 = .51$, so the the probability of favoring the death penalty does not equal the probability of favoring the death penalty conditional on one being male, so the two events (favoring death and being male) are dependent. Males are more likely to prefer the death penalty.