Introduction to Probability Theory

Outline

- Basic concepts in probability theory
- Elementary properties of probability
- Conditional Probability
- Bayes' Theorem

Definition of Probability

- ***Random Experiment***: is an experiment whose result would not be predicted but the list of possible outcomes are known.
- ***Sample space***: possible outcomes of an experiment
  - S = {HH, HT, TH, TT}
- ***Event***: a subset of possible outcomes
  - A={HH}, B={HT, TH}
- ***Probability of an event*** : an number assigned to an event Pr(A)
  - Axiom 1: Pr(A) ≥ 0
  - Axiom 2: Pr(S) = 1
  - Axiom 3: For every sequence of disjoint events

    $$\Pr(\bigcup_i A_i) = \sum_i \Pr(A_i)$$

  - Example: Pr(A) = n(A)/N: frequentist statistics

The Axiomatic "Definition" of Probability

Suppose that for experimental model M, the sample space S of possible outcomes is defined as: $\{A_1, \ldots, A_n\} \in S$.

Let $Pr(A_i)$ = the probability of an event A_i in the sample space S.

A probability distribution on a sample space S is a specification of numbers $Pr(A_i)$ which satisfy A1, A2, A3.

A1. For any outcome A_i, $Pr(A_i) \geq 0$.

A2. $Pr(S) = 1$.

A3. For any infinite sequence of disjoint events $A_1, \ldots, A_n$:

$$Pr(\cup_{i=1 \text{ to } \infty} A_i) = \sum_{i=1 \text{ to } \infty} Pr(A_i)$$

Note: it turns out that each of these three axioms can be justified using the coherence criterion.

Some Theorems Based on the Definition of Probability and a Few Proofs

Theorem 1. Pr(∅) = 0

Proof:

By definition, A_j and A_k are disjoint if $A_j \cap A_k = \varnothing$.
Further, it is obvious that: $\varnothing \cap \varnothing = \varnothing$.
Thus, if $A_j = \varnothing$ and $A_k = \varnothing$, then A_j and A_k are disjoint.

Let $A_1 \ldots A_n$ define the set of events such that $A_j = \varnothing$.

By the above definitions, it follows that the events A_j are disjoint.

Since the A_j are disjoint, we can exploit A3 such that:

$$Pr(\varnothing) = Pr(\cup_i A_i) = \Sigma_i Pr(A_i) = \Sigma_i Pr(\varnothing) = n\,Pr(\varnothing)$$

In order that $Pr(\varnothing) = n\,Pr(\varnothing)$, $Pr(\varnothing)$ must equal 0.

Some Theorems cont.

Theorem 2. For any sequence of n disjoint events $A_1,\ldots,A_n$,

$$Pr(\cup_{i \text{ to } n} A_i) = \Sigma_{i \text{ to } n} Pr(A_i)$$

Proof:

Let $A_1,\ldots,A_n$ define the n disjoint events and let $A_k = \varnothing$ for events $k \in \{n+1,\ldots, \infty\}$.

By the definition of disjoint events, we have an infinite series of disjoint events.

By A3 and Theorem 1 which states that $Pr(\varnothing)=0$:

$$Pr(\cup_{i \text{ to } n} A_i) = Pr(\cup_{i \text{ to } \infty} A_i) = \Sigma_{i \text{ to } \infty} Pr(A_i).$$

$$= \Sigma_{i \text{ to } n} Pr(A_i) + \Sigma_{n+1 \text{ to } \infty} Pr(A_i)$$

$$= \Sigma_{i \text{ to } n} Pr(A_i) + 0$$

$$= \Sigma_{i \text{ to } n} Pr(A_i)$$

Some Theorems cont.

Theorem 3. For any event A, $Pr(A^C) = 1 - Pr(A)$

Theorem 4. For any event A, $0 \leq Pr(A) \leq 1$

Proof:

By contradiction in two parts:

Part 1. Suppose $Pr(A) < 0$. Then that would violate axiom A1, a contradiction.

Part 2. Suppose $Pr(A) > 1$. Then by Theorem 3, $Pr(A^C) < 0$, which also contradicts A1.

Thus, $0 \leq Pr(A) \leq 1$.

Some Theorems cont.

Theorem 5. For any two events A and B,

$$\mathbf{Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)}$$

Proof:

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

Since all three elements in the equation are disjoint, Theorem 2 implies:

$Pr(A \cup B) = Pr(A \cap B^C) + Pr(A \cap B) + Pr(A^C \cap B)$

$= Pr(A \cap B^C) + Pr(A \cap B) + Pr(A^C \cap B) + Pr(A \cap B) - Pr(A \cap B)$

Further, we know that $Pr(A) = Pr(A \cap B^C) + Pr(A \cap B)$ and that $Pr(B) = Pr(A \cap B) + Pr(A^C \cap B)$

Thus, $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$

Independent Events

Intuitively, we define independence as:

Two events A and B are independent if the occurrence or non-occurrence of one of the events has no influence on the occurrence or non-occurrence of the other event.

Mathematically, we write define independence as:

Two events A and B are independent if $\Pr(A \cap B) = \Pr(A)\Pr(B)$.

Example of Independence

Are party id and vote choice independent in presidential elections?

Suppose Pr(Rep. ID) = .4, Pr(Rep. Vote) = .5, and Pr(Rep. ID $\cap$ Rep. Vote) = .35

To test for independence, we ask whether:

Pr Pr(Rep. ID) * Pr(Rep. Vote) = .35 ?

Substituting into the equations, we find that:

Pr Pr(Rep. ID) * Pr(Rep. Vote) = .4*.5 = .2 $\neq$ .35,

so the events are not independent.

Independence of Several Events

The events $A_1, \ldots, A_n$ are independent if:

$$Pr(A_1 \cap A_2 \cap \ldots \cap A_n) = Pr(A_1)Pr(A_2)\ldots Pr(A_n)$$

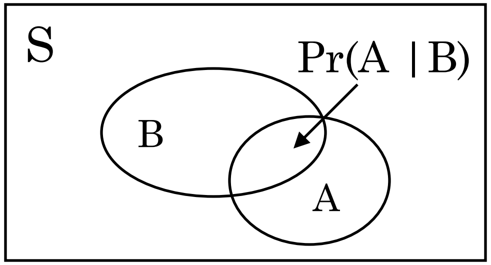
And, this identity must hold for any subset of events.

CONDITIONAL PROBABILITY

Conditional probabilities allow us to understand how the probability of an event A changes after it has been learned that some other event B has occurred.

The key concept for thinking about conditional probabilities is that the occurrence of B reshapes the sample space for subsequent events.

- That is, we begin with a sample space S
- A and B ∈ S
- The conditional probability of A given that B looks just at the subset of the sample space for B.



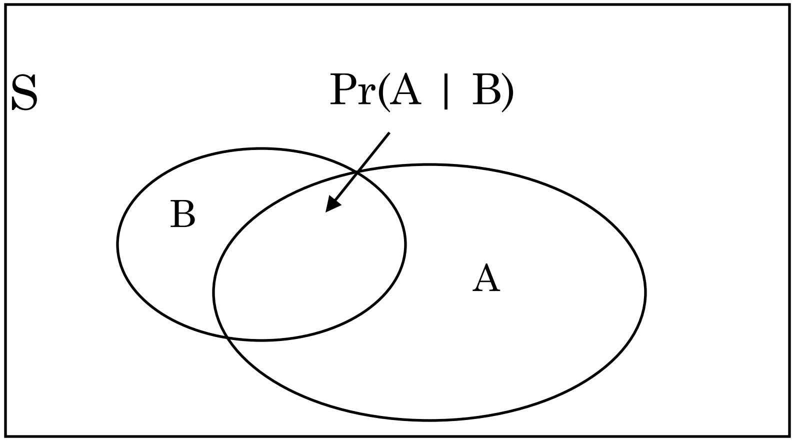
The conditional probability of A given B is denoted Pr(A | B).

- Importantly, according to Bayesian orthodoxy, all probability distributions are implicitly or explicitly conditioned on the model.

Conditional Probability Cont.

By definition: If A and B are two events such that Pr(B) > 0, then:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$



Example: What is the Pr(Republican Vote | Republican Identifier)?

$\Pr(\text{Rep. Vote} \cap \text{Rep. Id}) = .35$ and $\Pr(\text{Rep ID}) = .4$

Thus, Pr(Republican Vote | Republican Identifier) = .35 / .4 = .875

Useful Properties of Conditional Probabilities

Property 1. The Conditional Probability for Independent Events

If A and B are independent events, then:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{\Pr(A)\Pr(B)}{\Pr(B)} = \Pr(A)$$

Property 2. The Multiplication Rule for Conditional Probabilities

In an experiment involving two non-independent events A and B, the probability that both A and B occurs can be found in the following two ways:

$$\Pr(A \cap B) = \Pr(B)\Pr(A|B)$$

or

$$\Pr(A \cap B) = \Pr(A)\Pr(B|A)$$

Conditional Probability and Partitions of a Sample Space

The set of events $A_1,\ldots,A_k$ form a **partition of a sample space S** if: $\cup_{i=1 \text{ to } k} A_i = S$.

If the events $A_1,\ldots,A_k$ partition S and if B is any other event in S (note that it is impossible for $A_i \cap B = \varnothing$ for some i), then the events $A_1 \cap B, A_2 \cap B,\ldots,A_k \cap B$ will form a partition of B.

Thus, $B = (A_1 \cap B) \cup (A_2 \cap B) \cup \ldots \cup (A_k \cap B)$

$Pr(B) = \sum_{i=1 \text{ to } k} Pr(A_i \;\; B)$

Finally, if $Pr(A_i) > 0$ for all i, then:

$Pr(B) = \sum_{i=1 \text{ to } k} Pr(B \mid A_i) \, Pr(A_i)$

EXAMPLE OF CONDITIONAL PROBABILITY AND PARTITIONS OF A SAMPLE SPACE

$$\Pr(B) = \sum_{i=1 \text{ to } k} \Pr(B \mid A_i) \Pr(A_i)$$

Example. What is the Probability of a Republican Vote?

Pr(Rep. Vote) = Pr(Rep. Vote | Rep. ID) Pr(Rep. ID)
+ Pr(Rep. Vote | Ind. ID) Pr(Ind. ID)
+ Pr(Rep. Vote | Dem. ID) Pr(Dem. ID)

Note: the definition for Pr(B) defined above provides the denominator for Bayes' Theorem.

Bayes' Theorem (Rule, Law)

Bayes' Theorem: Let events $A_1,\ldots,A_k$ form a partition of the space S such that $\Pr(A_j) > 0$ for all j and let B be any event such that $\Pr(B) > 0$. Then for i = 1,..,k:

$$\Pr(A_i \mid B) = \frac{\Pr(A_i)\Pr(B \mid A_i)}{\sum_k \Pr(A_k)\Pr(B|A_k)}$$

Proof:

$$\Pr(A_i | B) = \frac{\Pr(A_i \cap B)}{\Pr(B)} = \frac{\Pr(A_i)\Pr(B | A_i)}{\sum_k \Pr(A_k)\Pr(B | A_k)}$$

Bayes' Theorem is just a simple rule for computing the conditional probability of events A_i given B from the conditional probability of B given each event A_i and the unconditional probability of each A_i

Interpretation of Bayes' Theorem

$\Pr(A_i)$ = Prior distribution for the A_i. It summarizes your beliefs about the probability of event A_i before A_i or B are observed.

$\Pr(B \mid A_i)$ = The conditional probability of B given A_i. It summarizes the *likelihood* of event B given A_i.

$$\Pr(A_i \mid B) = \frac{\Pr(A_i)\Pr(B \mid A_i)}{\sum_k \Pr(A_k)\Pr(B \mid A_k)}$$

$\Pr(A_i \mid B)$ = The posterior distribution of A_i given B. It represents the probability of event A_i after B has been observed.

$\sum_k \Pr(A_k)\Pr(B \mid A_k)$ = The normalizing constant. This is equal to the sum of the quantities in the numerator for all events A_k. Thus, $P(A_i \mid B)$ represents the likelihood of event A_i relative to all other elements of the partition of the sample space.

Example of Bayes' Theorem

What is the probability in a survey that someone is black given that they respond that they are black when asked?

- Suppose that 10% of the population is black, so Pr(B) = .10.
- Suppose that 95% of blacks respond Yes, when asked if they are black, so $\Pr(Y_1 \mid B) = .95$.
- Suppose that 5% of non-blacks respond Yes, when asked if they are black, so $\Pr(Y_1 \mid B^C) = .05$

$$\Pr(B \mid Y_1) = \frac{\Pr(B)\Pr(Y_1 \mid B)}{\Pr(B)\Pr(Y_1 \mid B) + \Pr(B^C)\Pr(Y_1 \mid B^C)}$$

$$\Pr(B \mid Y_1 = 1) = \frac{(0.1)(.95)}{(.1)(.95) + (.9)(.05)} = \frac{.095}{.14} = .68$$

We reach the surprising conclusion that even if 95% of black and non-black respondents correctly classify themselves according to race, the probability that someone is black given that they say they are black is less than .7.

EXAMPLE CONT.

Continuing the last example, suppose that the interviewer also makes an estimate of the respondent's race. Let's say the interviewer correctly classifies 90 percent of respondents, and her classification is independent of the self-classification.

Thus, $\Pr(Y_2 \mid B) = 0.9$ and $\Pr(Y_2 \mid B^C) = 0.1$.

One way to incorporate information is to recalculate our estimates from scratch.

$$\Pr(B \mid Y_1, Y_2) = \frac{\Pr(B)\Pr(Y_1 \mid B)\Pr(Y_2 \mid B)}{\Pr(B)\Pr(Y_1 \mid B)\Pr(Y_2 \mid B) + \Pr(B^C)\Pr(Y_1 \mid B^C)\Pr(B^C)\Pr(Y_2 \mid B^C)}$$

$$\Pr(B \mid Y_1 = 1, Y_2 = 1) = \frac{(.10)(.95)(.90)}{(.10)(.95)(.90) + (.9)(.05)(.10)} = \frac{.0855}{.09} = .95$$

Alternatively, we can just update our last set of results:

$$\Pr(B \mid Y_1, Y_2) = \frac{\Pr(B \mid Y_1 = 1)\Pr(Y_2 \mid B, Y_1 = 1)}{\Pr(B \mid Y_1 = 1)\Pr(Y_2 \mid B, Y_1 = 1) + \Pr(B^C \mid Y_1 = 1)\Pr(Y_2 \mid B^C, Y_1 = 1)}$$

$$\Pr(B \mid Y_1 = 1, Y_2 = 1) = \frac{(.68)(.90)}{(.68)(.90) + (.32)(.10)} = \frac{.612}{.644} = .95$$

THANK YOU

Submitted by:
Rakesh Kumar,
(Deptt. Of Mathematics),
P.G.G.C.G, Sec-11,
Chandigarh.