Basics of Probability

August 27 and September 1, 2009

1 Introduction

A phenomena is called **random** if the exact outcome is uncertain. The mathematical study of randomness is called the **theory of probability**.

A probability model has two essential pieces of its description.

- S, the sample space, the set of possible outcomes.
  - An **event** is a collection of **outcomes**.

    $$A = \{s_1, s_2, \cdots, s_n\}$$

    and a subset of the sample space

    $$A \subset S.$$

- P, the probability assigns a number to each event.

Thus, a probability is a function. We are familiar with functions in which both the domain and range are subsets of the real numbers. The domain of a probability function is the collection of all possible outcomes. The range is still a number. We will see soon which numbers we will accept as possible probabilities of events.

The operations of **union**, **intersection** and **complement** allow us to define new events. Identities in set theory tell that certain operations result in the same event. For example, if we take events A, B, and C, then we have the following:

1. **Commutivity**. $A \cup B = B \cup A, \quad A \cap B = B \cap A$.
2. **Associativity**. $(A \cup B) \cup C = A \cup (B \cup C), \quad (A \cap B) \cap C = A \cap (B \cap C)$.
3. **Distributive laws**. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
4. **DeMorgan's Laws** $(A \cup B)^c = A^c \cap B^c, \quad (A \cap B)^c = A^c \cup B^c$.

A third element in a probability model is a σ-algebra $\mathcal{F}$. $\mathcal{F}$ is a collection of subsets of S satisfying the following conditions:

1. $\emptyset \in \mathcal{F}$.
2. If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.
3. If $\{A_j; j \geq 1\} \subset \mathcal{F}$, then $\cup_{j=1}^{\infty} A_j \in \mathcal{F}$.

2 Set Theory - Probability Theory Dictionary

Event Language	Set Language	Set Notation	Venn Diagram
sample space	universal set	S	
event	subset	$A, B, C, \cdots$	
outcome	element	s	
impossible event	empty set	$\emptyset$	
not A	A complement	A^c	
A or B	A union B	$A \cup B$	
A and B	A intersect B	$A \cap B$	
difference	A but not B	$A \backslash B$ $= A \cap B^c$	
symmetric difference	either A or B but not both	$A \Delta B$ $= (A \backslash B) \cup (B \backslash A)$	
A and B are mutually exclusive	A and B are disjoint	$A \cap B = \emptyset$	
if A then B	A is a subset of B	$A \subset B$	

Whenever S is finite or countable, then we take $\mathcal{F}$ to be all subsets of S. When S is uncountable, then, in general, we cannot make this choice for $\mathcal{F}$ and maintain other more desirable properties. The best known example is to take $S = [0, 1]$ and let the probability of an interval be equal to the length of the interval. Then we cannot define a probability P on all of the subsets of $[0, 1]$ so that $P([a, b]) = b - a$.

3 Examples of sample spaces and events

1. Toss a coin $S = \{H, T\}$ $\#(S) = 2$

   Toss heads $A = \{H\}$ $\#(A) = 1$

2. Toss a coin three times.

   $S =$ {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT} $\#(S) =$

   Toss at least two heads in a row.

   $A =$ {HHH, HHT, THH } $\#(A) =$

   Toss at least two heads.

   $B =$ {HHH, HHT, HTH, THH } $\#(B) =$

3. Toss a coin 100 times. $\#(S) =$

   $A =$ {67 heads} $\#(A) =$

4. Roll two dice. $\#(S) =$

   $A =$ {sum is 7} $\#(A) =$

   $B =$ {maximum value is 4} $\#(B) =$

   $C =$ {sum is not 7} $\#(C) =$

   $D =$ {sum is 7 or maximum value is 4} $\#(D) =$

5. Roll three dice. $\#(S) =$

   $A =$ {sum is 9} $\#(A) =$

   $B =$ {sum is 10} $\#(B) =$

   $C =$ {sum is 9 or 10} $\#(B) =$

6. Pick a card from a deck. $\#(S) =$

   $A =$ {pick a $\heartsuit$} $\#(A) =$

   $B =$ {level is 4} $\#(B) =$

   $C =$ {level is not 4} $\#(C) =$

7. Pick two cards form the deck.

   (a) Replacing the first before choosing the second. $\#(S) =$

   (b) Choosing the first, then the second without replacing. $\#(S) =$

   (c) Choosing two cards simultaneously. $\#(S) =$

   $A =$ {pick two aces} find $\#(A)$ in each of the three circumstances.

4 Equally Likely Outcomes

If S is a finite sample space, then if each outcome is equally likely, we define the probability of A as the fraction of outcomes that are in A.

$$P(A) = \frac{\#(A)}{\#(S)}.$$

Thus, computing $P(A)$ means counting the number of outcomes in the event A and the number of outcomes in the sample space Ω and dividing.

1. Toss a coin.
$$P\{\text{heads}\} = \frac{\#(A)}{\#(S)} = \frac{1}{2}.$$
2. Toss a coin three times.
$$P\{\text{toss at least two heads in a row}\} = \frac{\#(A)}{\#(S)} = \underline{\qquad}$$
3. Roll two dice.
$$P\{\text{sum is 7}\} = \frac{\#(A)}{\#(S)} = \underline{\qquad}$$

Because we always have $0 \leq \#(A) \leq \#(S)$,, we always have

$$0 \leq P(A) \leq 1 \tag{1}$$

and

$$P(S) = 1 \tag{2}$$

So, no we know that the range of the function we call the probability is a subset of the interval [0,1].

Toss a coin 4 times. $\#(S) = 16$
$A = \{\text{exactly 3 heads}\}$ $\#(A) = 4$
$= \{\text{HHHT, HHTH, HTHH, THHH}\}$

$$P(A) = \frac{4}{16} = \frac{1}{4}$$

$B = \{\text{exactly 4 heads}\}$ $\#(B) = 1$
$= \{\text{HHHH}\}$

$$P(B) = \frac{1}{16}$$

Now let's define the set $C = \{\text{at least three heads}\}$. If you are asked the supply the probability of C, your intuition is likely to give you an immediate answer.

$$P(C) = \underline{\qquad}.$$

Let's have a look at this intuition. The events A and B have no outcomes in common, they are mutually exclusive events, and thus,

$$\#(A \cup B) = \#(A) + \#(B).$$

If we take this **addition principle** and divide by $\#(S)$, then we obtain the following identity

If $A \cap B = \emptyset$, then

$$P(A \cup B) = P(A) + P(B). \tag{3}$$

Using this property, we see that

$$P\{\text{at least 3 heads}\} = P\{\text{exactly 3 heads}\} + P\{\text{exactly 4 heads}\} = \frac{4}{16} + \frac{1}{16} = \frac{5}{16}.$$

5 The Axioms of Probability

1. For any event A,
$$0 \leq P(A) \leq 1. \tag{1}$$
2. For the sample space S,
$$P(S) = 1. \tag{2}$$
3. If the events A and B are mutually exclusive ($A \cap B = \emptyset$), then
$$P(A \cup B) = P(A) + P(B). \tag{3}$$

We are saying that any function P that accepts events as its domain and returns numbers as its range and satisfies (1), (2), and (3) can be called a probability.

For example, if we toss a *biased* coin. We may want to say that

$$P\{\text{heads}\} = p$$

where p is not necessarily equal to 1/2. By necessity,

$$P\{\text{tails}\} = 1 - p.$$

If we iterate the procedure in Axiom 3, we can also state that if the events, $A_1, A_2, \cdots, A_n$, are mutually exclusive, then

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n). \tag{3$'$}$$

For the random experiment, flip a coin repeated until heads appears, we can write $A_j = \{$the first head appears on the j-th toss$\}$. We would like to say that

$$P\{\text{heads appears eventually}\} = P(A_1) + P(A_2) + \cdots + P(A_n) + \cdots.$$

This would call for an extension of Axiom 3 to an infinite number of mutually exclusive events. This is the general version of Axiom 3 we use when we want to use calculus in the theory of probability:

For $\{A_j; j \geq 1\}$, are mutually exclusive, then

$$P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j) \tag{3$''$}$$

6 Consequences of the Axioms

1. **The Complement Rule**. Because A and A^c are mutually exclusive

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

or

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

Toss a coin 4 times.

$$P\{\text{fewer than 3 heads}\} = 1 - P\{\text{at least 3 heads}\} = 1 - \frac{5}{16} = \frac{11}{16}.$$

We can extend this. If $A \subset B$, then the $P(B \backslash A) = P(B) - P(A)$.

2. **The Inclusion-Exclusion Rule**. For any two events A and B,

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

($P(A) + P(B)$ counts the outcomes in $A \cap B$ twice, so remove $P(A \cap B)$.)

Exercise 1. *Show that the inclusion-exclusion rule follows from the axioms. Hint:* $A \cup B = (A \cap B^c) \cup B$ *and* $A = (A \cap B) \cup (A \cap B^c)$.

Deal two cards.

$A = \{\text{ace on the second card}\}$, $\qquad B = \{\text{ace on the first card}\}$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$P\{\text{at least one ace}\} = \frac{1}{13} + \frac{1}{13} - \quad ?$$

To complete this computation, we will need to compute $P(A \cap B) = P\{\text{both cards are aces}\}$.

3. **The Bonferroni Inequality.** For any two events A and B,

$$P(A \cup B) \leq P(A) + P(B).$$

By induction we have the extended Bonferroni inequality:

Theorem 2. *For any events* $A_1, \ldots, A_n$

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