

Probability Tutorial Using Dice

Bill McNeill

billmcn@u.washington.edu

January 21, 2005

The theory of probability takes a few basic intuitive ideas about the relative likelihood of events and expresses them in a mathematically rigorous manner using set theory. This tutorial explains the basic concepts using the concrete example of rolling a pair of dice.

1 Basic Outcomes

Say that a *trial* is when we roll two dice and read the numbers that come up. The *basic outcome* of the trial is an ordered pair of numbers.

Here are some basic outcomes.

- Die #1 is 3. Die #2 is 6.
- Die #1 is 4. Die #2 is 4.
- Die #1 is 6. Die #2 is 1.

As a shorthand, write the basic outcomes in the form (Die #1, Die #2). For example

- Die #1 is 3. Die #2 is 6 is (3,6).
- Die #1 is 4. Die #2 is 4 is (4,4).
- Die #1 is 6. Die #2 is 1 is (6,1).

2 Sample Space

There are only a finite number of possible basic outcomes. For example if we're using six-sided dice, (4,4) is possible but (7, 23) is not. Call the set of all possible outcomes the *sample*

space of this experiment. Each trial is an *element* of the set. By convention we refer to the sample space by the Greek letter Ω .

$$\Omega = \left\{ \begin{array}{cccccc} (1, 1) & (1, 2) & (1, 3) & (1, 4) & (1, 5) & (1, 6) \\ (2, 1) & (2, 2) & (2, 3) & (2, 4) & (2, 5) & (2, 6) \\ (3, 1) & (3, 2) & (3, 3) & (3, 4) & (3, 5) & (3, 6) \\ (4, 1) & (4, 2) & (4, 3) & (4, 4) & (4, 5) & (4, 6) \\ (5, 1) & (5, 2) & (5, 3) & (5, 4) & (5, 5) & (5, 6) \\ (6, 1) & (6, 2) & (6, 3) & (6, 4) & (6, 5) & (6, 6) \end{array} \right\} \quad (1)$$

In this sample space there are $6 \times 6 = 36$ outcomes.

3 Event

An *event* is any subset of Ω . Here are some events.

- Die #1 is 3 and die #2 is 6 is event $\{(3, 6)\}$
- Both dice are the same is event $\{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$
- The two dice sum to 5 is event $\{(1, 4), (2, 3), (3, 2), (4, 1)\}$

Note that all individual outcomes are events, but events can also be combinations of outcomes. The set of all outcomes Ω and the empty set \emptyset are also possible events.

By convention, events are referred to with capital Roman letters from the beginning of the alphabet. For example $A = \{(2, 4), (5, 6)\}$.

4 Probability Distribution

Define a mapping between the set of all events and the set of real numbers. If every event has one and only one corresponding number, we call this mapping a *function*. Write a function as $f(\text{event}) = \text{number}$. For example

- $f(\{(3, 6)\}) = 42$
- $f(\{(2, 4), (5, 6)\}) = 3.75$
- $f(\Omega) = 99$

We can define any arbitrary function we want, but some classes of functions are more interesting than others. One class of functions that is interesting is that for which the following *axioms of probability* hold.

$$P(\Omega) = 1 \quad (2)$$

$$P(A) \geq 0 \quad (3)$$

$$P(A \cup B) = P(A) + P(B) \text{ for all } A \cap B = \emptyset \quad (4)$$

A function that obeys all three of these conditions for all $A \subseteq \Omega$ and $B \subseteq \Omega$ is called a *probability function* or (equivalently) a *probability distribution*. By convention we write the name of a probability distribution with a P .

Statements (5) and (6) are easily proved corollaries to the axioms of probability. They will be true for any P that satisfies (2)-(4).

$$P(A) \leq 1 \tag{5}$$

$$P(\emptyset) = 0 \tag{6}$$

There are many ways to define valid probability functions. Usually when working with discrete sample spaces like we have here we do the following.

$$P(A) = \frac{|A|}{|\Omega|} \tag{7}$$

The vertical lines around A denote the number of elements in A , known as the *cardinality* of A . For example

- $|\{(1, 6), (3, 2)\}| = 2$
- $|\{(1, 1), (2, 2), (3, 3)\}| = 3$
- $|\Omega| = 36$
- $|\emptyset| = 0$

It is easy to prove that (7) satisfies the conditions in (2)-(6). Here are some examples of each condition.

- $P(\{(1, 6), (3, 2)\}) = \frac{|\{(1,6),(3,2)\}|}{|\Omega|} = \frac{2}{36} = \frac{1}{18}$ and $0 \leq \frac{1}{18} \leq 1$
- $P(\{(1, 6), (3, 2)\}) = \frac{1}{18} = \frac{1}{36} + \frac{1}{36} = P(\{(1, 6)\}) + P(\{(3, 2)\})$
- $P(\Omega) = \frac{|\Omega|}{|\Omega|} = \frac{36}{36} = 1$
- $P(\emptyset) = \frac{|\emptyset|}{|\Omega|} = \frac{0}{36} = 0$

The mathematical definition of probability is useful because it resembles our intuitive notion of how likely or unlikely an event should be. For example, we know from our experience with dice in the real world that we are more likely to make a roll that sums to seven than to roll two sixes. Using the mathematical scheme we just defined, the probability of rolling a seven is

$$P(\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}) = \frac{1}{6} \tag{8}$$

While the probability of the event of rolling two sixes is

$$P(\{(6, 6)\}) = \frac{1}{36} \tag{9}$$

That the probability of rolling two sixes is a smaller number squares with our intuition that this is a less likely event.

5 Conditional Probability

If we have some additional information about an event this might influence our idea of its probability. We handle this case by introducing the concept of *conditional probability*.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (10)$$

Equation (10) is an expression for the conditional probability of A given B . It gives the likelihood of event A assuming that we already know event B has occurred.

For example, we already calculated the probability of rolling a 5 above. Now say we want to calculate the probability of rolling a 5 given that one or both of the dice rolled is a 2. We would calculate this conditional probability like so

$$A = \{(1, 4), (2, 3), (3, 2), (4, 1)\} \quad (11)$$

$$B = \left\{ \begin{array}{cccccc} (2, 1) & (2, 2) & (2, 3) & (2, 4) & (2, 5) & (2, 6) \\ (1, 2) & (3, 2) & (4, 2) & (5, 2) & (6, 2) & \end{array} \right\} \quad (12)$$

$$A \cap B = \{(2, 3), (3, 2)\} \quad (13)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (14)$$

$$= \frac{P(\{(2, 3), (3, 2)\})}{P(B)} \quad (15)$$

$$= \frac{\frac{1}{18}}{\frac{11}{36}} \quad (16)$$

$$= \frac{2}{11} \quad (17)$$

Equation (11) is the event of rolling a 5 and (12) is the event of one or both of the dice being a 2. The rest of the equations go through the details of calculating the conditional probability.

In this example $P(A|B) > P(A)$, but this will not always be true. In general conditioning may make an event more likely, less likely, or leave its likelihood unchanged depending on the events involved.

6 Random Variable

A random variable is a function that maps the set of events to \mathbb{R}^n . By convention random variables are written as upper case Roman letters from the end of the alphabet like X .

For example, define the random variable X to be the sum of the two dice. For every

element in the sample space, we can specify the value of X .

$$\Omega = \left\{ \begin{array}{cccccc} (1, 1) = 2 & (1, 2) = 3 & (1, 3) = 4 & (1, 4) = 5 & (1, 5) = 6 & (1, 6) = 7 \\ (2, 1) = 3 & (2, 2) = 4 & (2, 3) = 5 & (2, 4) = 6 & (2, 5) = 7 & (2, 6) = 8 \\ (3, 1) = 4 & (3, 2) = 5 & (3, 3) = 6 & (3, 4) = 7 & (3, 5) = 8 & (3, 6) = 9 \\ (4, 1) = 5 & (4, 2) = 6 & (4, 3) = 7 & (4, 4) = 8 & (4, 5) = 9 & (4, 6) = 10 \\ (5, 1) = 6 & (5, 2) = 7 & (5, 3) = 8 & (5, 4) = 9 & (5, 5) = 10 & (5, 6) = 11 \\ (6, 1) = 7 & (6, 2) = 8 & (6, 3) = 9 & (6, 4) = 10 & (6, 5) = 11 & (6, 6) = 12 \end{array} \right\} \quad (18)$$

If we know the probabilities of a set of events, we can calculate the probabilities that a random variable defined on those set of events takes on certain values. For example

- $P(X = 2) = P(\{(1, 1)\}) = \frac{1}{36}$
- $P(X = 5) = P(\{(1, 4), (2, 3), (3, 2), (4, 1)\}) = \frac{1}{9}$
- $P(X = 7) = P(\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}) = \frac{1}{6}$
- $P(X = 12) = P(\{6, 6\}) = \frac{1}{36}$

The expression for $P(X = 5)$ should be familiar, since we calculated it above as “the probability of the event that the two dice sum to five”.

Much of the theory of probability is concerned with defining functions of random variables and calculating the likelihood with which they take on their values.

7 Expectation Value

Once we know the probability distribution of a random variable we can use it to predict the average outcome of functions of that variable. This is done using *expectation values*.

The expectation value of a function $f(X)$ of a random variable X is defined to be

$$E[f(X)] = \sum_{x \in R} f(x)P(X = x) \quad (19)$$

where R is the range of all possible values of $f(X)$.

The most basic expectation value function is just $f(X) = X$. $E[X]$ is called the *mean* of the random variable X . The X defined in the previous section has the following mean value

$$E[X] = \sum_{x=2}^{12} xP(X = x) \quad (20)$$

$$= 2P(X = 2) + 3P(X = 3) + 4P(X = 4) \cdots + 12P(X = 12) \quad (21)$$

$$= 2\frac{1}{36} + 3\frac{1}{18} + 4\frac{1}{12} \cdots + 12\frac{1}{36} \quad (22)$$

$$= 7 \quad (23)$$

You can think of expectation values as taking a weighted average of the values of $f(X)$ where more likely values get a higher weight than less likely values.

References

- [1] Harold J. Larson. *Introduction to Probability Theory and Statistical Inference*. John Wiley and Sons, New York, 1982.
- [2] Christopher D. Manning and Hinrich Schütze. *Foundations of Statistical Natural Language Processing*. Massachusetts Institute of Technology, Cambridge, Massachusetts, 1999.