Review of Basic Probability Theory

James H. Steiger

Department of Psychology and Human Development
Vanderbilt University

Review of Basic Probability Theory

Introduction

If you took Psychology 310, you had a thorough introduction to set theory and probability theory "from the ground up."

RDASA Chapter 3 provides a very condensed account of many of the key facts we developed.

I'll review these facts here, but lecture slides and recordings are also available of material from Psychology 310 should you need them.

MWL use the notation $p(A)$ to stand for the probability of an event A, while I use the notation $\Pr(A)$.

Set Theory

Some grounding in basic set theory is very useful to gaining a solid understanding of probability theory.

Many books attempt to discuss probability theory while never formally defining set theory, and instead relying on an "informal" version of set theory. This saves time, but also reduces precision, depending on the sophistication of the reader.

MWL use "informal" language and notation while I shall stick with the formal notation of set theory. In the long run, it will be useful to you to be able to read texts using either notational variation.

Set Theory

Definition of a Set

A set is any well-defined collection of "objects."

The term "object" is used loosely. It could refer to numbers, ideas, people, or fruits.

The *elements* of a set are the objects in the set.

Set Theory

Complement, Union, Intersection, and Difference

The *universal set* Ω is the set of all objects currently under consideration.

The *null (empty) set* $\emptyset$ is a set with no elements.

The *complement* of A, i.e., $\bar{A}$, is the set of all elements not in A (but in Ω).

Consider two sets A and B:

The *union* of A and B, i.e., $A \cup B$, is the set of all elements in A, B, or both.

The *intersection* of A and B, i.e., $A \cap B$, is the set of all elements in A and B. Informal discussions of probability will refer to $\Pr(A \text{ and } B)$ to signify $\Pr(A \cap B)$.

The *set difference*, $A - B$, is defined as the set of elements in A but not in B. That is, $A - B = A \cap \overline{B}$.

Set Theory

Set Partition and Mutually Exclusive Sets

Two sets A and B are said to be *mutually exclusive* if they have no elements in common, i.e., $A \cap B = \emptyset$

A set of events is *exhaustive* of another set if their union is equal to that set.

Set Theory

Set Partition and Mutually Exclusive Sets

A group of n sets $E_i, i = 1, \ldots n$ is said to *partition* the set A if the E_i are mutually exclusive and exhaustive with respect to A.

Note the following:

1. Any set is the union of its elements.
2. Any set is partitioned by its elements.

Probabilistic Experiments and Events

Probabilistic Experiments and their Sample Spaces

A probabilistic experiment E is a situation in which

1. Something can happen.
2. What can happen is well-defined.
3. The outcome is potentially uncertain.

Probabilistic Experiments and Events

Probabilistic Experiments and their Sample Spaces

Associated with every probabilistic experiment E_i is a *sample space* S_i.

The sample space is defined as the universal set of all possible outcomes of E_i.

Probabilistic Experiments and Events

Elementary and Compound Events

The *elementary events* in a sample space are the elements of S.

As such, they constitute the "finest" possible partition of S.

Example (Elementary Events)

You throw a fair die, and observe the number that comes up. The elementary events are 1,2,3,4,5,6.

Probabilistic Experiments and Events

Elementary and Compound Events

Compound events are the union of two or more elementary events.

Example (Compound Events)

In the die throw experiment, the event Even $= 2 \cup 4 \cup 6 = \{2, 4, 6\}$ is a compound event.

Axioms and Basic Theorems of Probability

Given a sample space S and any event A_i within S, we assign to each event a number called the *probability* of A. The probabilities in S must satisfy the following:

1. $\Pr(A) \geq 0$
2. $\Pr(S) = 1$
3. If two events A and B in S are mutually exclusive, then $\Pr(A \cup B) = \Pr(A) + \Pr(B)$.

Axioms and Basic Theorems of Probability

From the axioms, we can immediately derive the following theorems:

$\Pr(A) = 1 - \Pr(\overline{A})$

$\Pr(\emptyset) = 0$

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

Basic Rules for Computing Probability

The General Rule

For any event A composed of elementary events e_i, i.e., $A = \cup_{i=1}^{n} e_i$, the probability of A is the sum of the probabilities of the elementary events in A, i.e., $\Pr(A) = \sum_{i=1}^{n} \Pr(e_i)$

Basic Rules for Computing Probability

The General Rule

Since the probabilities of the elementary events must sum to 1, if the elementary events are equally likely, then every one of the n_S elementary event has a probability of $1/n_S$.

Consequently, the total probability of an event A can be computed as

$$\Pr(A) = \frac{n_A}{n_S} \tag{1}$$

where n_A is the number of elementary events in A, and n_S is the number of elementary events in the sample space.

Joint Events, Conditional Probability, and Independence

Joint and Marginal Events

In many situations, two or more physical processes are occurring simultaneously, and we can define the elementary events of the sample space to be the intersection of the outcomes on the two processes.

For example, I might throw a die and a coin simultaneously.

In that case, we have 12 elementary events, each one of which is an intersection. They are $H \cap 1$, $H \cap 2 \ldots$, $T \cap 1, \ldots T \cap 6$.

The *marginal* events are the outcomes on the individual processes, i.e., $H, T, 1, 2, 3, 4, 5, 6$.

Joint Events, Conditional Probability, and Independence

Joint and Marginal Events

Here is a tabular representation.

Note that the probability of a marginal event in a row or column is the sum of the probabilities of the joint events in that row or column. (Why?)

	Die					
Coin	1	2	3	4	5	6
H	$H \cap 1$	$H \cap 2$	$H \cap 3$	$H \cap 4$	$H \cap 5$	$H \cap 6$
T	$T \cap 1$	$T \cap 2$	$T \cap 3$	$T \cap 4$	$T \cap 5$	$T \cap 6$

Joint Events, Conditional Probability, and Independence

Conditional Probability

The *conditional probability* of A given B, written as $\Pr(A|B)$, is the probability of A within the reduced sample space defined by B.

To evaluate conditional probability, we simply move inside the class of events defined by B, and calculate what proportion of the events in B are examples of A.

For example, suppose the probabilities of our joint events were as given in the table on the next slide.

Joint Events, Conditional Probability, and Independence

Conditional Probability

		Die					
		1/6	1/6	1/6	1/6	1/6	1/6
	Coin	**1**	**2**	**3**	**4**	**5**	**6**
1/2	**H**	1/6	0	1/6	0	1/6	0
1/2	**T**	0	1/6	0	1/6	0	1/6

- What is $\Pr(1|H)$? $\Pr(1|T)$?
- What is $\Pr(H|1)$? $\Pr(H|2)$?
- Notice that the conditional probabilities within any conditionalizing event are simply the joint probabilities inside that event, re-standardized so that they add up to 1.
- They are re-standardized to add up to 1 by dividing by what they currently add up to, i.e., $\Pr(B)$.
- This leads to the formula for conditional probability often given as its definition:

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

Joint Events, Conditional Probability, and Independence

Conditional Probability

Conditional probability reflects how the probabilities of the outcomes on one process change after we are informed about the status of the outcome of the other process.

If knowledge of one process changes the probability structure for the other, we say that the processes are *dependent*.

If knowledge of one process does not change the probability structure for the other, then we say that the processes are *independent*.

This leads to two formally equivalent definitions of independence. Two events A and B are independent if

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

or

$$\Pr(A \cap B) = \Pr(A)\Pr(B) \tag{4}$$

Sequences and the Multiplicative Rules

Sequences as Intersections

Sequences of events are, of course, events themselves.

Consider when we shuffle a poker deck and draw two cards off the top.

The sequence of cards involves Card_1 followed by Card_2.

If we now consider the event "Ace on the first card followed by Ace on the second card," we quickly realize that this is the intersection of two events, i.e. $A_1 \cap A_2$. How can we compute the probability of this event?

Sequences and the Multiplicative Rules

The Multiplicative Rules

Recall that

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

This implies that

$$\Pr(A \cap B) = \Pr(A) \times \Pr(B|A) \tag{5}$$

If A and B are independent, the rule simplifies to

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

.

Sequences and the Multiplicative Rules

The Multiplicative Rules

These rules are sometimes referred to as the *multiplicative rules of probability*. They generalize to sequences of any length. For example, suppose you draw 3 cards from a poker deck without replacement. Since the outcome on each card affects the probability structure of the succeeding draws, the events are not independent, and the probability of drawing 3 aces can be written using the general multiplicative rule as:

$$\Pr(A_1 \cap A_2 \cap A_3) = \Pr(A_1) \times \Pr(A_2|A_1) \times \Pr(A_3|A_1 \cap A_2) \tag{6}$$

Probability Distributions

There are two fundamental types of random variables, *discrete* and *continuous*

With discrete random variables, the number of events is countable.

With continous random variables, the number of events is not countable and infinite, and continuous over at least some range on the number line.

Probability Distributions

Discrete Probability Distributions

When the number of events is countable and finite, the probability of each elementary event can be calculated, and the sum of these probabilities is 1.

In this case, each event has a probability, and the probability density function, or pdf, is usually denoted $p(x)$. This is the probability that the random variable X takes on the value x, i.e., $p(x) = \Pr(X = x)$.

The *cumulative probability function*, $F(x)$, is the cumulative probability up to and including x, i.e., $F(x) = \Pr(X \leq x)$.

Probability Distributions

Continuous Probability Distributions

When a random variable X is continuous, it takes on all values over some continuous range, and so the number of outcomes is uncountably infinite.

For continuous random variables, the probability of a particular outcome x cannot be defined, even if x can occur. The probability is infinitesimally small.

Rather than defining the probability of x, we instead define an alternative concept, *probability density*.

Probability density $f(x)$ is not probability. Rather, it is the instantaneous rate at which the cumulative probability F is increasing at x. That is, it is the slope (derivative) of $F(x)$ at x.

In a similar fashion, $F(x)$ is the area under the probability density curve. It is the integral of $f(x)$.

Although the probability of a single value x cannot be defined with continuous random variables, the probability of an interval can be. To compute the probability of an interval, we simply take the difference of the two cumulative probabilities. That is,

$$\Pr(a \leq X \leq b) = F(b) - F(a)$$

Probability Distributions

Calculating Probability with R

R provides a set of functions for performing probability calculations for many of the most important distributions.

The key function types are

1. *d*, for calculating probability (discrete r.v.) or probability density (continuous r.v.)
2. *p*, for calculating cumulative probability.
3. *q*, for calculating inverse cumulative probabilities, or quantiles.

To perform these calculations, you need to know the code for the distribution you are working with. Some common codes are:

1. *norm*. The normal distribution.
2. *t*. Student's *t* distribution.
3. *f*. The *F* distribution.
4. *chisq*. The chi-square distribution.

You combine the function type and the distribution name to obtain the name of the function you need to perform a particular calculation.

Probability Distributions

Calculating Probability with R

What value of the standard normal distribution is at the 95th percentile?

```
> qnorm(0.95)

[1] 1.645
```

What is the probability that a *t*-variable with 19 degrees of freedom is less than or equal to 1.00?

```
> pt(1, 19)

[1] 0.8351
```

What is the probability that an observation from a normal distribution with mean 500 and standard deviation 100 will be between 600 and 700?

```
> pnorm(700, 500, 100) - pnorm(600, 500, 100)

[1] 0.1359
```

Random Variables and Their Characteristics

Random Variables

In many introductory courses, the informal notion of a random variable X is that it is a process that generates numerical outcomes according to some rule.

This definition suffices for practical purposes in such courses, and we can get by with it here.

A more formal definition is that, given a probabilistic experiment E with sample space of outcomes S, a random variable S is a function assigning a unique number to each outcome in S.

Because each outcome is connected to a unique number, each unique number inherits all the probabilities attached to its outcomes.

Here is a simple example from the discrete case of a die throw.

Random Variables and Their Characteristics

Random Variables

Suppose you throw a fair die, and code the outcomes as in the table below

Outcome (in S)	1	2	3	4	5	6
Value of X	1	0	1	0	1	0

The random variable X would then have the probability distribution shown in the following table

x	$P_X(x)$
1	1/2
0	1/2

Random Variables and Their Characteristics

Expected Value of a Random Variable

The *expected value*, or *mean* of a random variable X, denoted $\mathrm{E}(X)$ (or, alternatively, μ_X), is the long run average of the values taken on by the random variable.

Technically, this quantity is defined differently depending on whether a random variable is discrete or continuous.

For some random variables, $\mathrm{E}(|X|) = \infty$, and we say that the expected value *does not exist*.

Random Variables and Their Characteristics

Variance and Covariance of Random Variables

The variance of a random variable X is its average squared deviation score.

We can convert a random variable into deviation scores by subtracting its mean, i.e.,

$$X - \mathrm{E}(X)$$

So the squared deviation score is

$$(X - \mathrm{E}(X))^2$$

and the average squared deviation score is

$$\mathrm{Var}(X) = \sigma_X^2 = \mathrm{E}\left[(X - \mathrm{E}(X))^2\right]$$

An alternate, very useful formula is

$$\mathrm{Var}(X) = \sigma_X^2 = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$$

Random Variables and Their Characteristics

Variance and Covariance of Random Variables

In a similar vein, the covariance of two random variables X and Y is their average cross-product of deviations, i.e.

$$\text{Cov}(X, Y) = \text{E}\left[(X - \text{E}(X))(Y - \text{E}(Y))\right] \tag{7}$$

which may also be calculated as

$$\text{Cov}(X, Y) = \text{E}(XY) - \text{E}(X)\,\text{E}(Y) \tag{8}$$

Random Variables and Their Characteristics

The Algebra of Expected Values, Variances, Covariances and Linear Combinations

Properties of expected value and variance of a random variable are developed in detail in the Psychology 310 notes, and in the handout chapter on Covariance Algebra. For our purposes, the key properties are

1. Under linear transforms, the expected value (mean) of a random variables behaves the same as the mean of a list of numbers, i.e., for constants a, b and random variable X,

$$\mathrm{E}(aX + b) = a\,\mathrm{E}(X) + b \tag{9}$$

2. If X and Y are random variables and are uncorrelated,

$$\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y) \tag{10}$$

3. The mean, variance, and covariance of linear combinations follow the identical rules we developed earlier. For example

$$\begin{aligned}
\mathrm{E}(aX + bY) &= a\,\mathrm{E}(X) + b\,\mathrm{E}(Y) && (11)\\
\mathrm{Var}(X + Y) &= \mathrm{Var}(X) + \mathrm{Var}(Y) \\
&\quad + 2\,\mathrm{Cov}(X, Y) && (12)\\
\mathrm{Cov}(X + Y, X - Y) &= \mathrm{Var}(X) - \mathrm{Var}(Y) && (13)
\end{aligned}$$