Introduction to Probability

MATH 30530-01 SPRING 2012

Announcements

Course Information

Calendar: class activities, assignments, notes

Lecture Notes

Homework Solutions

Exam Information

Welcome to the website! Important course information will be posted on this web page and announced in class. You are responsible for all material that appears here, so you should check this page frequently for updates.

May 4 I have posted a quick summary of the course under exam information Also, I made a mistake in what I posted last night. Last fall's final is now posted. It has 10 problems.

May 3 I have posted a copy of the final from last fall's course on the website.

April 2 I have posted a clean copy of Midterm Exam II on the website. I have also posted instructions for rewriting incorrect answers to exam questions.

March 26 Wednesday's exam will be at our usual class time and in our usual classroom. Tuesday evening's review session will be 7--9PM in 127 Hayes-Healy. The exam will cover Chapter 4: §4.3--4.7, §4.8.1--§4.8.3, §4.9, and §4.10; and Chapter 5: §5.1--§5.4, §5.5 (omitting §5.5.1), and §5.7.

Midterm II from the fall term is posted under exam information.

February 27 Answers to the first midterm are posted on the course website. The grading scale for the exam follows. A: 90 and up; B: 80--89; C: 65--79. Fortunately, we need go no lower. The median was 89.5.

February 18 Monday's exam will be at our usual class time and in our usual classroom. Sunday evening's review session will be 7--9PM in 229 Hayes-Healy. The exam will cover Chapter 1; Chapter 2, omitting §2.6 and §2.7; Chapter 3, omitting §3.5; and Chapter 4, §4.1 and §4.2.

The first midterm from last fall is posted on the course website under Exam Information. You should know that we---students and faculty alike---thought the exam was too long. Also, it was given a little later in the term, so it covered more material than this exam will.

In addition to the exam from last semester, the textbook has complete solutions to the Self-Test questions at the end of each chapter.

February 16 The first midterm from last fall has been posted.

February 16 Solutions to Homework Assignment 03 have been posted.

February 8 Extra credit is available for the Math for Everyone talk tomorrow, February 9, at 5:00PM in Jordan 105. Details are posted on the calendar---today.

February 8 I have posted regular office hours on my home page.

January 18 The first midterm exam is scheduled for Monday, February 20. There was an error in the original version of the general information document.

Lecture 01: Random Phenomena
Wednesday, January 18, 2012

Probability is the branch of mathematics concerned with random phenomena. Probability spaces form the theoretical framework for studying experiments with unpredictable outcomes. Today we shall look at the following examples:

- the Monty Hall problem
- coin tossing
- dice
- radioactive decay

1 The Monty Hall Problem

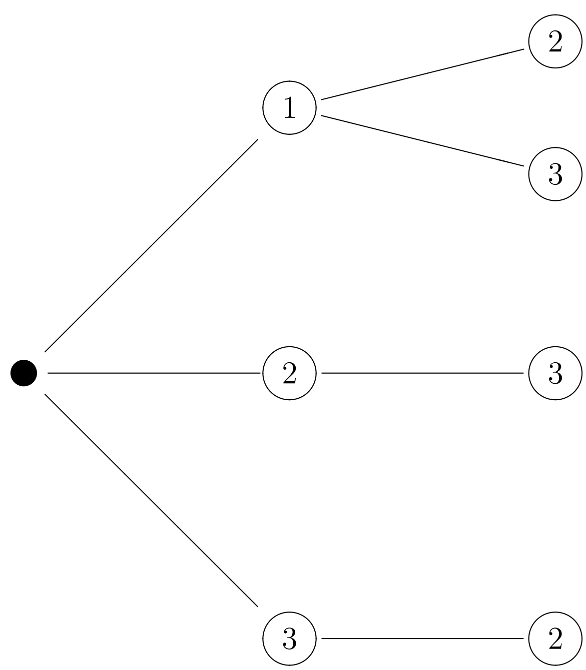
This problem attracted a fair amount of notoriety a few years ago. It comes in several forms. It was inspired by the game show *Let's Make a Deal,* which was hosted by Monty Hall. The problem was originally posed by Steve Selvin in a letter to *The American Statistician* in 1974. Marilyn Vos Savant stated it as follows in Parade Magazine in 1990:

> Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1 (but the door is not opened), and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?

Are you better off **switching** your choice or **sticking** with your original choice? (We are assuming that you would prefer to win the car.) To answer the question we must describe the possible outcomes of the experiment in a more detailed and precise way. It won't help much to say that there are two possibilities: you win a car or you win a goat. Describing the outcomes will also force us to make certain assumptions explicit.

Before we give the description, let's make the following convention: When we say that someone chooses one possibility "at random" from a finite set of possibilities, we mean that each of the possibilities is equally likely to be chosen. For example, if you choose one of the three doors at random, each door has probability 1/3 of being chosen.

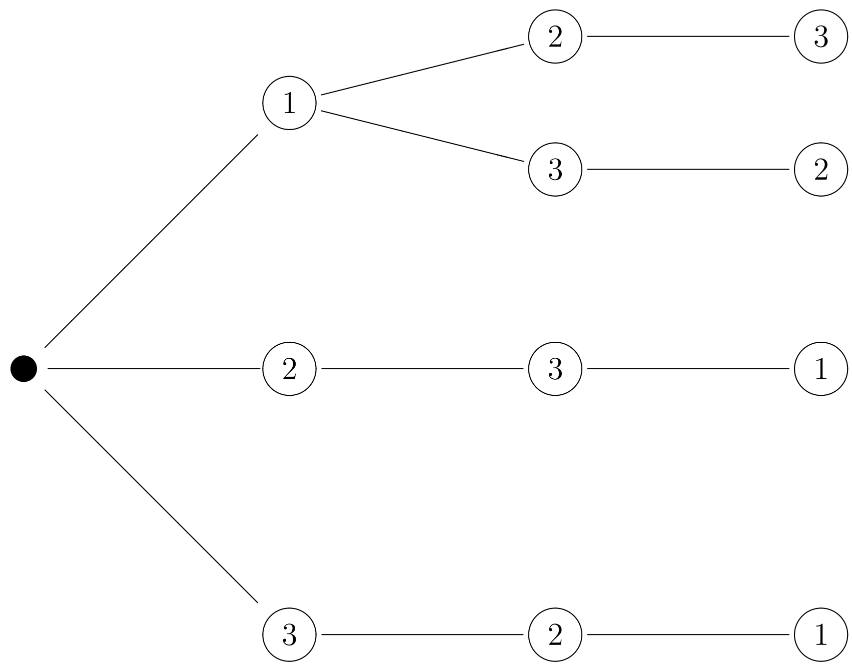
Let's assume that the doors are painted red, white, and blue. We also assume that the car and the goats are already in position. You choose one of the doors at random, indicating your choice of door by its color. Unknown to you but known to the host, the doors are numbered 1, 2, and 3: The car is behind door number 1, and the goats are behind doors 2 and 3. The host then opens a door that has a goat behind it. If possible, the host chooses this door at random. We describe the outcome of the experiment with the tree diagram below.



On each path through the tree from the root, the first node gives the number of the door you choose, and the second node gives the number of the door the host opens. At this point the rest of the game is determined by the strategy you follow: stick or switch. If you stick with your original choice, you win the car if that choice was door

1, and you win a goat otherwise.

Here is what will happen if you switch your choice (the third node shows your new choice):



So if you switch, you win a goat if your initial choice was door 1, and you win the car if you chose door 2 or door 3.

How likely is each of these outcomes? Let's describe a path through the first tree by an ordered pair (a, b), where a is the number of the door you chose and b is the number of the door the host opened. The set of possible outcomes is thus

$$\Omega = \{(1,2), (1,3), (2,3), (3,2)\}$$

Since you are equally likely to open each door, each of the outcomes $(2,3)$ and $(3,2)$ has probability $1/3$. Since the host also chooses at random, each of the outcomes $(1,2)$ and $(1,3)$ has probability $1/6$.

If your strategy is to stay with your original choice, you win the car if the outcome is in the subset $\{(1,2), (1,3)$ of Ω, thus you win the car with probability $1/3$.

If your strategy is to switch your choice, you win the car if the outcome is in the subset $\{(2,3),(3,2)\}$ of Ω, thus you win the car with probability $2/3$.

Questions: 1. We made the assumption that, when possible, the host chooses which door to open at random. Does this matter?

2. Can you think of other ways to set up the problem? For example, the doors need not be identified in two ways. They could simply be identified by their colors, and the car and goats could be placed randomly.

References. These also contain further references.

- `http://en.wikipedia.org/wiki/Monty_Hall_problem`
- Rosenhouse, Jason, *The Monty Hall Problem,* Oxford University Press 2009, ISBN 978-0-19-536789-8

There are also many websites with simulations of the problem. You can find them by googling "Monty Hall problem".

2 Coin Tossing

When we toss a coin, we (who aren't serious magicians) can't say in advance whether the coin will turn up heads or tails. This is what we mean when we say that the outcome is "unpredictable" or "random". Despite this unpredictability, there is some statistical regularity in the outcomes of tossing a coin. Suppose we keep tossing the coin and we keep track of the **relative frequency** of heads, that is, of

$$\frac{\text{number of heads}}{\text{number of tosses}}.$$

It appears that there is a number p such the relative frequency of heads settles in around p. (For most coins in our actual experience p is near to $1/2$, but we can certainly imagine that p could be any number such that $0 \leq p \leq 1$.) We take p to be the **probability** that the coin will turn up heads when it is tossed. Not surprisingly this is referred to as a **relative frequency** interpretation of probability. (As you might guess, there are other interpretations of probability.)

3 Dice

Next, we shall consider rolling a pair of dice. We shall assume that one die is green and one red. When we write down an outcome of the experiment, we use an ordered pair with first component the number showing on the green die and second component the number showing on the red die. Thus the outcome $(2,3)$ means that the green die turned up 2 and the red die turns up 3.

We can make a list of the possible outcomes of the experiment:

$$\begin{array}{llllll}
(1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6) \\
(2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6) \\
(3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6) \\
(4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6) \\
(5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6) \\
(6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6)
\end{array}$$

We call the set Ω of all possible outcomes the **sample space** associated with this experiment. Sometimes we refer to the individual outcomes as **sample points.** Here, Ω has 36 elements.

We shall assume that our dice are perfect and are identical except for their colors, so that it is reasonable to assume that all the outcomes have equal probability—namely, probability $1/36$.

We can ask the question, "What is the probability that the sum of the faces is 7?" We look at the corresponding set of outcomes:

$$E = \{(1,6),\ (2.,5),\ (3,4),\ (4,5),\ (5,2),\ (6,1)\}$$

We call E an **event**. An event is a subset of the sample space. In simple examples like this one, we can take any subset of the sample space to be an event. We take the probability $P(E)$ of the event E to be the sum of the probabilities of the sample points in E. Thus

$$P(E) = \frac{6}{36} = \frac{1}{6}.$$

As another example, we let F be the event that the green die turns up odd. Then F contains 18 sample points, and so $P(F) = 1/2$.

We could also ask for the probability that the green die turns up odd and the sum of the faces is 7. This would be the probability $P(E \cap F)$ of the intersection of E

and F. Since the event $E \cap F$ contains 3 sample points, $P(E \cap F) = 1/12$. In general, performing common set theoretic operations (union, intersection, complement, relative complement, etc.) on events produces further events.

4 Radioactive Decay

A radioactive atom may change into an atom of a different kind by emitting particles or radiation. For example, ${}^{238}_{92}\mathrm{U}$ decays to ${}^{234}_{90}\mathrm{Th}$ by emitting an α particle, and ${}^{14}_{6}\mathrm{C}$ decays to ${}^{14}_{7}\mathrm{N}$ by emitting a β particle and an antineutrino. According to physicists, we cannot say when an individual radioactive atom will decay. However, if initially we have a large number of atoms, say N_0, of atoms of a radioactive isotope, we can say something about the number $N(t)$ of atoms that have not decayed by time t, but nothing about which ones have not decayed. A relative frequency interpretation of probability leads us to take

$$\frac{N(t)}{N_0}$$

as the probability that an individual atom will not decay by time t.

The rate at which the isotope decays is approximately proportional to the number of atoms of the isotope that are present. (This approximation is very good.) Thus $N(t)$ satisfies the initial value problem

$$\frac{dN}{dt} = -\lambda N, \quad N(0) = N_0,$$

where λ is a positive constant. Thus,

$$N(t) = N_0 e^{-\lambda t},$$

and

$$\frac{N(t)}{N_0} = e^{-\lambda t}$$

is the probability that an atom has not decayed by time t. We choose an atom at random and let T denote the time at which it decays. Now T is a positive number, and so the set of possible outcomes is the set $(0, \infty)$ of positive real numbers. From our last equation, we see that

$$P\{a < T\} = e^{-\lambda a}.$$

Thus, if $0 < a < b$,

$$\begin{aligned} P\{a < T \leq b\} &= P(\{a < T\} \setminus \{b < T\}) \\ &= P\{a < T\} - P\{b < T\} \\ &= e^{-\lambda a} - e^{-\lambda b} \end{aligned}$$

(Draw a picture!)

In many ways this example is more complicated than the previous two. The set of outcomes is no longer finite, but an interval of real numbers. We are naturally led to assign probabilities to intervals of outcomes, not to individual outcomes. Indeed the only plausible value of $P\{T = t_0\}$ is zero. It also turns out that we cannot assign probabilities to arbitrary subsets of $(0, \infty)$.

Lecture 02: Probability Spaces
Friday, January 20, 2012

1 DeMorgan's Laws

There is a discussion of the algebra of sets in §2.1. Unless it is causing some sort of trouble I won't go over most of it in class, but I will say a little bit about De Morgan's Laws. We will be working with a universal set Ω—usually the set of outcomes of some experiment–and events E and F—they are subsets of Ω. We take complements with respect to Ω:

$$E^c = \{x \in \Omega : x \notin E\}.$$

De Morgan's Laws are these two identities:

$$(E \cup F)^c = E^c \cap F^c$$
$$(E \cap F)^c = E^c \cup F^c$$

To prove that two events are equal, it suffices to show that each is a subset of the other. That is, we shall first show that if x is an element of the set on the left, then x is also an element of the set on the right. Then we shall show the converse.

For the first law we reason as follows.

$$
\begin{aligned}
x \in (E \cup F)^c &\implies x \notin E \cup F \\
&\implies x \notin E \text{ and } x \notin F \\
&\implies x \in E^c \text{ and } x \in F^c \\
&\implies x \in E^c \cap F^c
\end{aligned}
$$

We have shown that $(E \cup F)^c \subset E^c \cap F^c$. To prove the reverse inclusion, we observe that each of the steps above is reversible.

We can demonstrate the second of De Morgan's Laws similarly. This time we indicate immediately that each step in the reasoning is reversible.

$$
\begin{aligned}
x \in (E \cap F)^c &\iff x \notin E \cap F \\
&\iff x \notin E \text{ or } x \notin F \\
&\iff x \in E^c \text{ or } x \in F^c \\
&\iff x \in E^c \cup F^c
\end{aligned}
$$

We could also prove one of De Morgan's Laws as above, and then we could deduce the other law from the one just proved. For example, we can deduce the first law from the second. Assuming the second law, we have

$$(E^c \cap F^c)^c = (E^c)^c \cup (F^c)^c,$$

so

$$(E^c \cap F^c)^c = E \cup F.$$

The first law now follows by taking complements:

$$E^c \cap F^c = (E \cup F)^c.$$

The text proves DeMorgan's laws for finite collections of events. They actually hold for any collection of events. The arguments are essentially those given above.

2 Probability Spaces

A **probability space** consists of a nonempty set Ω, a class $\mathcal{A}$ of subsets of Ω, and an assignment P of a real number to each set in $\mathcal{A}$. The set Ω is called the **sample space,** the sets in $\mathcal{A}$ are called **events,** and the assignment P is called a **probability measure,** or simply a **probability.** When we model an experiment, the sample space is the set of possible outcomes, the events describe phenomena that can be observed, and we can think of the probability of an event as the chance the event will occur. We require that the following axioms for events be satisfied.

Axiom E1. There is at least one event, i.e., $\mathcal{A}$ is not empty.

Axiom E2. If E is an event, then its complement E^c is also an event.

Axiom E3. If E and F are events, then their union $E \cup F$ is also an event.

Remarks: 1. Also, the intersection EF is an event. (Probabilists use this notation; $EF = E \cap F$.)

It is sufficient to show that the complement of EF is an event. From DeMorgan's Law

$$(EF)^c = E^c \cup F^c$$

By Axiom E2, E^c and F^c are events, so their union is an event by Axiom E3. Thus we are done.

2. The union and intersection of any finite collection of events are events.

This follows by induction on the number of events in the collection.

3. The sets $\emptyset$ and Ω are events.

For, by Axiom E1, there exists an event E. and, by Axiom E2, E^c is also an event. Thus $\emptyset = EE^c$ and $\Omega = E \cup E^c$ are events, by Axiom E3 and Remark 1.

Axiom E4. If E_k, $k = 1, 2, \cdots$, is a sequence of events, then

$$\bigcup_{k=1}^{\infty} E_k$$

is also an event.

Remark. It follows from DeMorgan's Law that

$$\bigcap_{k=1}^{\infty} E_k$$

is also an event.

Just as there are axioms for events, there are axioms for the probability measure P.

Axiom P1. $0 \le P(E) \le 1$ for every event E.

Axiom P2. $P(\Omega) = 1$.

Events E and F are said to be **mutually exclusive** if they are disjoint, i.e., if $EF = \emptyset$.

Axiom P3. If $E_k,\ k = 1, 2, \cdots$, is a sequence of pairwise mutually exclusive events, then

$$P\left(\bigcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} P(E_k).$$

Remarks. 1.

$$P(\emptyset) = 0$$

To see this we take $E_1 = \Omega$ and $E_k = \emptyset$ for all $k \geq 2$. By Axiom P3 we have

$$\begin{aligned} P(\Omega) &= P\left(\bigcup_{k=1}^{\infty} E_k\right) \\ &= \sum_{k=1}^{\infty} P(E_k) \\ &= P(\Omega) + \sum_{k=2}^{\infty} P(\emptyset) \end{aligned}$$

Thus

$$0 = \sum_{k=2}^{\infty} P(\emptyset),$$

which is impossible if $P(\emptyset) > 0$. Thus $P(\emptyset) = 0$.

2. If E and F are mutually exclusive events, then

$$P(E \cup F) = P(E) + P(F).$$

This follows from Axiom P3 if we take $E_1 = E$, $E_2 = F$, and $E_k = \emptyset$ for all $k \geq 3$. For then,

$$\begin{aligned}
P(E \cup F) &= P(\bigcup_{k=1}^{\infty} E_k) \\
&= \sum_{k=1}^{\infty} P(E_k) \\
&= P(E) + P(F) + \sum_{k=3}^{\infty} P(\emptyset) \\
&= P(E) + P(F)
\end{aligned}$$

3. A similar argument shows that if $E_1, E_2, \ldots, E_n$ is a finite collection of pairwise mutually exclusive events, then

$$P(E_1 \cup E_2 \cup \cdots \cup E_n) = P(E_1) + P(E_2) + \cdots + P(E_n).$$

3 Some Simple Propositions

Proposition 4.1 *For any event E,*

$$P(E^c) = 1 - P(E)$$

Proof. We have that E^c is also an event, E and E^c are mutually exclusive, and $\Omega = E \cup E^c$. Thus

$$P(\Omega) = P(E) + P(E^c),$$

and so

$$P(E^c) = P(\Omega) - P(E),$$

and finally,

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

Proposition 4.2 *If E and F are events with $E \subset F$, then*

$$P(E) \leq P(F).$$

Proof. We have $F = E \cup (F \setminus E)$, and the events E and $F \setminus E$ are mutually exclusive. Thus,

$$P(F) = P(E) + P(F \setminus E).$$

Since $P(F \setminus E) \geq 0$, it follows that $P(F) \geq P(E)$.

Proposition 4.3 *If E and F are events, then*

$$P(E \cup F) = P(E) + P(F) - P(EF).$$

Proof. Now $E = (EF) \cup (EF^c)$, and the events EF and EF^c are mutually exclusive. Thus

$$P(E) = P(EF) + P(EF^c).$$

Similarly

$$P(F) = P(FE) + P(FE^c).$$

Adding these last two equations gives

$$P(E) + P(F) = 2P(EF) + P(EF^c) + P(FE^c).$$

Subtracting $P(EF)$ from each side gives

$$P(E) + P(F) - P(EF) = P(EF) + P(EF^c) + P(FE^c).$$

The three events on the right hand side of this equation, namely EF, EF^c, and FE^c, are pairwise mutually exclusive, and their union is EF. Thus

$$P(E) + P(F) - P(EF) = P(E \cup F).$$

Let's try to extend Proposition 4.3 to the case of three events. So suppose that we have events E, F, and G. Then

$$\begin{aligned}
P(E \cup F \cup G) &= P(E \cup (F \cup G)) \\
&= P(E) + P(F \cup G) - P(E(F \cup G)) \\
&= P(E) + P(F) + P(G) - P(FG) - P(EF \cup EG) \\
&= P(E) + P(F) + P(G) - P(FG) \\
&\quad -[P(EF) + P(EG) - P(EFG)] \\
&= P(E) + P(F) + P(G) \\
&\quad -P(FG) - P(EF) - P(EG) + P(EFG)
\end{aligned}$$

Thus we have the formula

$$\begin{aligned}
P(E \cup F \cup G) &= P(E) + P(F) + P(G) \\
&\quad -P(FG) - P(EF) - P(EG) + P(EFG).
\end{aligned}$$