Introduction to Probability Theory for Graduate Economics

Brent Hickman

November 20, 2009

- **Additional Readings:** Most introductory texts on mathematical statistics will mirror significant portions of these notes fairly closely; one such text is *Introduction to Probability and Mathematical Statistics* 2nd *Ed.* by Lee J. Bain and Max Engelhardt. Exceptions include topics like stochastic dominance (see Section 2, Part I), affiliation (see Section 3), and stochastic processes (see Section 5). An overview of stochastic dominance and affiliation is given in the appendices of Krishna's book *Auction Theory*. For a good introduction to discrete-time stochastic processes, see Chapters 4 and 5 of *Introduction to Probability Models*, by Sheldon Ross.

1 Fundamentals of Probability

There are many real-world phenomena that are ruled by some component of chance. In order to better understand **stochastic** phenomena, we construct mathematical **probability models**. A natural way to view such models is in terms of a random **experiment**, where each instance or **trial** results in an observed result called an **outcome**.

Definition 1 *The set of all possible outcomes of an experiment is called the **sample space**, commonly denoted by S Being a set, S can have properties such as **finiteness**, **countability** or **uncountability**. If S is a countable set, then it is said to be **discrete**.*

Exercise 1 Identify the sample space, *S*, in the following experiments.

    a. Three 6-sided die are cast and the outcomes are recorded.

b. A researcher records how many patients enter an emergency room between November and March with respiratory problems.

c. Another researcher records the amount of time it takes for an ER patient to complete triage and see a doctor.

d. A new coin is drawn at random from a minting machine. It is flipped once and its mass in grams is measured. The coin flip outcome and the mass are both recorded. ■

Definition 2 *An **event** is a subset of the sample space, S, and it is said to have **occurred** if it contains the observed outcome of a particular trial.*

▶ *Being subsets, we can talk about taking **unions or intersections of events**.*

▶ *An **elementary event** is a subset of S for which its only non-empty subset is itself. In other words, an elementary event is a singleton.*

▶ *A **compound event** is the union of at least two elementary events.*

The whole sample space, S, is called the **sure event** and the empty set, $\varnothing$, is called the **null event**.

Definition 3 *Two events, A and B are said to be **mutually exclusive** if $A \cap B = \varnothing$ and a collection of events $\{A_1, A_2, \ldots, A_k\}$ is said to be mutually exclusive if its members are **pairwise mutually exclusive**, or $A_i \cap A_j = \varnothing$ when $i \neq j$.*

Exercise 2 For each of the four cases listed in Example 1, identify some elementary and compound events, and construct a collection of three mutually exclusive events. ■

We usually speak of events in the context of the probabilities with which they will occur. The concept of classical probability is based on the notion of **relative frequency**: in M trials of an experiment, if $m(A)$ is the number of trials in which event A occurred, then $f_A = \frac{m(A)}{M}$ is the relative frequency of A. The probability of event A is denoted and defined by $P(A) \equiv \lim_{M \to \infty} f_A$.[1]

Definition 4 *A function $P : \{A \| A \subseteq S\} \to \mathbb{R}_+$ is a **probability function** if it satisfies the following properties:*

1. $P(A) \geq 0 \ \forall \ A \subseteq S$

---

[1]I should mention that this is merely ONE concept of the notion of probability, but not the only one. For example, a Bayesian econometrician might view probability as a subjective measure of randomness which is updated as new information is encountered. However, for the purposes of this course, we will stick with the objective classical view of probability.

2. $P(S) = 1$

3. $P\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} P(A_i)$ *for every mutually exclusive collection of events* $\{A_1, A_2, \ldots, A_k\}$.

1.1 Properties of Probability

I. Let A be an event and let A' be it's compliment; then $P(A) \leq 1$ and $P(A) = 1 - P(A')$

II. For any two events A and B, if $A \subseteq B$ then $P(A) \leq P(B)$

III. For any two events A and B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

IV. For any three events A and B and C,

$$
\begin{aligned}
P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\
&\quad - P(A \cap B) - P(A \cap C) - P(B \cap C) \\
&\quad + P(A \cap B \cap C)
\end{aligned}
$$

V. For any collection of events $\{A_1, A_2, A_3, \ldots\}$, in order to determine the probability of their union, we begin by defining a collection of related sets in the following way: $B_1 = A_1$, and $B_i = A_i \cap \left(\bigcup_{j=1}^{i-1} A_j\right)'$ for $i \geq 2$. This gives us the following:

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

VI. **Boole's Inequality**: for any collection of events $\{A_1, A_2, A_3, \ldots\}$, we have

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

This is an immediate consequence of (*II*) and (*V*).

VII. **Bonferroni's Inequality**: for any collection of events $\{A_1, A_2, \ldots, A_k\}$, we have

$$
P\left(\bigcap_{i=1}^{k} A_i\right) \geq 1 - \sum_{i=1}^{k} P(A_i').
$$

This is an immediate consequence of (*I*) and (*VI*).

Boole's and Bonferroni's inequalities give us potentially useful bounds on the probabilities of unions or intersections of events. For example, if an upper bound on the probability of the union

of events was sufficient and easier to calculate than the actual probability, Boole's inequality would serve. Note, however, that the bounds are also potentially uninformative, because the upper bound in (*VI*) could be greater than 1 and the lower bound in (*VII*) could be negative.

1.2 Conditional Probability, Total Probability and Bayes' Rule

One of the main objectives in probability theory is to use available information to update beliefs about how random events occur. For example, we might ask questions like, "how likely is it that a football team will win, given that it lost its previous game by 3 points?," or, "what is the probability that demand will be high tomorrow, given that it is low today?" To address question, we use the concept of conditional probability.

Definition 5 *For two events, A and B, the **conditional probability** of A given the occurrence of B is given by*

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

Note that from the previous definition it immediately follows that

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A).$$

From the second equality, we see that $P(A|B)$ can also be expressed as

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

This formula is widely known as **Bayes' Rule**.

When computing the probability of an event A, it is often easier to break up A into a collection of mutually exclusive events, and then sum their probabilities. To do this, one might think of partitioning the sample space S into a collection $\{B_1, \ldots, B_k\}$ of mutually exclusive and **exhaustive** events, meaning that the members of the collection do not overlap, and their union covers all of S. When this is true, we can express A in the following way:

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_k).$$

This leads us to the Law of Total Probability.

Definition 6 *The Law of Total Probability States that if $\{B_1, \ldots, B_k\}$ is a set of mutually exclusive and*

4

exhaustive events, then for any event A, we have

$$P(A) = \sum_{i=1}^{k} P(A|B_i)P(B_i).$$

With the law of total probability on hand, we can restate Bayes' rule in the following useful way:

Theorem 1 *If $\{B_1, \ldots, B_k\}$ is a set of mutually exclusive and exhaustive events, then for any event A, we have*

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{k} P(A|B_i)P(B_i)}.$$

Exercise 3 (*Exercise 32 from Capter 1 of "Introduction to Probability and Mathematical Statistics, 2nd Ed.,"* *Bain and Engelhardt, Duxbury, 1992*) A baseball team has 3 pitchers, *A*, *B*, and *C*, with winning percentages of 0.4, 0.6 and 0.8, respectively. These pitchers pitch with frequency 2, 3, and 5 out of every 10 games, respectively. In other words, for a randomly selected game, $P(A) = 0.2$, $P(B) = 0.3$, and $P(C) = 0.5$. Compute the following probabilities:

a. Pr[team wins game] = $P(W)$

b. Pr[*A* pitched game|team won] = $P(A|W)$ ∎

Exercise 4 (*Belief updating in strategic games*): A drunken sailor walks into a bar, itching to pick a fight. He sees a man sitting at the bar and he considers the choice of whether to fight him. The man could either be a whimp or a tough-guy, but whimps always dress like tough guys, so the sailor can't tell them apart. However, he thinks that whimps and tough-guys hang out in bars with equal probability. When whimps are antagonized, they fight with probability w and they win with probability .4. When tough-guys are antagonized, they fight with probability t and they win with probability .7. The sailor's ship will be in port for two days, so he will have the opportunity to return the next day and fight the man again, regardless of what happens on the first day. For the next three questions, assume that the sailor decides to pick a fight on the first day.

a. What is the probability that the sailor's opponent is a tough-guy, given that his opponent fought and won on the first day?

b. What is the sailor's probability of losing a fight on the second day, conditional on having fought and lost on the first day?

c. Assume that the sailor antagonized the man at the bar on the first day, and that the man fought and won. On the second day, the sailor's payoff is 0 if there is no fight, 10 if he fights and wins,

and -10 if he fights and looses. What choice gives a higher expected payoff on the second day, challenging the man at the bar to a re-match or backing down? ∎

Exercise 5 (*Exercise 1.2.1 from "Contemporary Bayesian Econometrics and Statistics," by John Geweke, Wiley, 2005*) Let D denote the event that a particular disease is present in an individual. A test for the disease is administered by a nurse and the result comes up as either $+$ or $-$, meaning, respectively, that the patient either has the disease or not. Like all tests, this one is not 100% accurate. The *sensitivity* of the test, denoted by q, is the probability of a "positive" result conditional on the disease being present. For this test, $q = P(+|D) = .98$. The *specificity* of the test, denoted by p, is the probability of a "negative" result conditional on the disease being absent; it is $p = P(-|D') = .90$. The *incidence* of the disease is the probability that the disease is present in a randomly selected individual; it is $\pi = .005$. Compute the probabilities of the following events:

a. The disease is present given a "negative" outcome

b. The disease is present given a "positive" outcome ∎

Exercise 6 (*Exercise 6 from the August 1, 2005 qualifying exam*) Suppose that a decision maker maximizes subjective expected utility and revises beliefs according to Bayes' rule. Assume that her utility function is $u(x) = x, \quad x \in \mathbb{R}$; *i.e.*, outcomes are monetary prizes. She has the opportunity to gamble on the toss of a coin: She can not participate, or she can say either "H" or "T" and she will win $30 if she is correct and she loses $50 otherwise. Without additional information, she thinks that it is equally likely that the coin is two-headed, fair or two-tailed. How much will she be willing to pay for the chance to observe one toss before playing? ∎

Of course, conditioning on available information assumes that the information will have the effect of altering beliefs about outcomes of an experiment. However, this is not always the case. In particular, it is not so for independent events.

Definition 7 *Two events A and B are called **independent events** if*

$$P(A \cap B) = P(A)P(B),$$

*otherwise, A and B are called **dependent events**. We say that collections of events, $\{A_1, A_2, \ldots, A_k\}$, are **independent** or **mutually independent** if for every subset of the collection, the probability of the intersection of the subset is the product of the probabilities of the events in the subset.*

Going back to the formula for conditional probability above, it is easy to see that if A and B are independent, then $P(A|B) = P(A)$ and likewise, $P(B|A) = P(B)$. In other words, knowledge of the

occurrence of one event tells us nothing about the likelihood of the other one.

Theorem 2 *Two events A and B are independent if and only if the following pairs of events are also independent:*

1. *A and B′*

2. *A′ and B*

3. *A′ and B′*

Exercise 7 Prove the above theorem. ∎

1.3 Counting Techniques

In many settings in probability theory, it will be necessary to systematically count objects. For example, maybe an event A can occur in N different ways, with the probability of any given instance being p. In that case, the probability of A would be Np. Thus, the ability to count the number of ways in which events occur is often crucial to computing their probabilities.

Theorem 3 *If there are N_i possible outcomes for the ith trial of an experiment, $i = 1, 2, \ldots, k$, then there are $\prod_{i=1}^{k} N_i$ possible outcomes in the sample space.*

Exercise 8 Consider a test with 5 true-false questions and 3 multiple choice questions with four choices each. How many total ways are there in which the test could be answered? ∎

In many cases in probability theory, one would like to count the number of ways in which a set of objects could be arranged or grouped. In other cases, one might be interested in arrangements or groupings of subsets of the objects. If specific **arrangements** are of interest, then the **ordering** will be important, but if mere **groupings** of objects are considered then ordering will not matter. This distinction is important, as it will have implications for the way in which the arrangements or groupings are counted. For example, if we randomly select two numbers from the set $\{0, 1, \ldots, 9\}$, then the selections $(3, 7)$ and $(7, 3)$ would be considered different arrangements, but they would be counted as representing the same grouping. Thus, there will be more arrangements than groupings.

Also, when counting arrangements of objects, the **distinguishability** of the objects will also be important. If some of the objects are indistinguishable, meaning there is no way to tell them apart, then this will also affect the way in which orderings are counted. For example, if we order two

identical blue balls before a red one, then exchanging the first two will produce the same ordering. Thus, there will be fewer arrangements possible if some objects are indistinguishable.

In technical terms, an ordered arrangement of a set of objects is called a **permutation** and a grouping of objects without respect to ordering is called a **combination**.

Theorem 4 *The number of permutations of n distinguishable objects is n!.*

Theorem 5 *The number of permutations of n distinguishable objects taken r at a time is $_nP_r = \frac{n!}{(n-r)!}$.*

Theorem 6 *The number of combinations of n distinguishable objects chosen r at a time is $\binom{n}{r} = \frac{n!}{r!(n-r)!}$. The quantity $\binom{n}{r}$ is commonly referred to as "n-choose-r".*

Theorem 7 *The number of distinct permutations of n objects, of which r are of one kind and the rest are of another kind is $\binom{n}{r} = \frac{n!}{r!(n-r)!}$.*

Theorem 8 *Consider n objects divided into k groups, where the ith group contains r_i identical objects which are distinguishable from objects not in group i. Then the number of distinct permutations of the n objects is $\frac{n!}{r_1!r_2!\cdots r_k!}$.*

Suppose now that we have n distinct objects which we wish to separate into k bins, with r_i objects being assigned to the ith bin. This process is referred to as **partitioning**. Suppose the number of bins was two and we wanted to place r objects into the first bin and the rest into the second. Obviously, the number of ways to do this is $\binom{n}{r}$. More generally, the number of possible partitions is given in the following theorem:

Theorem 9 *The number of ways to partition n objects into k bins, with r_i being assigned to the ith bin and where $\sum_{i=1}^{k} r_i = n$ is $\frac{n!}{r_1!r_2!\cdots r_k!}$.*

Exercise 9 There are 5 indivisible units of a good to be allocated among 9 consumers. Assume that no one gets more than one unit.

   a. How many ways are there for consumer 1 to receive one?

   b. How many ways are there for both consumers 1 and 2 to receive a unit of the good?

   c. How many ways are there for the combined allocation to consumers 1 and 2 to be at least one unit?

   d. Assuming that a unit of the good is allocated to any given consumer with equal probability, what are the probabilities of the events in the previous three questions? ■

Exercise 10 *Order Statistics*: Consider a set of outcomes $\{X_i\}_{i=1}^n$ from n independent trials of the same experiment. In technical terms, we say that the X_is are **identically and independently distributed**, or **iid**. Suppose that they are real numbers and that we order them in increasing fashion in terms of their values. The kth value in the ordered sample, known as the kth **order statistic**, will be denoted by V_k, so that $V_1 \leq V_2 \leq \ldots \leq V_n$. Finally, let $F(v)$ represent the probability that any one observed outcome from the experiment is less than or equal to some benchmark, v. Derive the following probabilities using your knowledge of probability theory and counting techniques:

    a. $\Pr[V_n \leq v] = G_n(v)$

    b. $\Pr[V_{n-2} \leq v] = G_{n-2}(v)$

    c. $\Pr[V_{n-k} \leq v] = G_{n-k}(v)$, for arbitrary k.

HINT: For each event $(V_{n-k} \leq v)$, try breaking it up into a set of mutually exclusive events of the form $A_j = (V_{n-j} \leq v) \cap (V_{n-j+1} > v) \ \ j \leq k$. Then, for any given event A_j, try first to count the number of ways in which it could occur, and figure out what the probability of any given occurrence of A_j is. ∎

Exercise 11 *Endogenous State Transition in Dynamic Oligopoly* (Erickson & Pakes, *Review of Economic Studies*, 1995): Consider an industry in which there are N firms competing in each of infinitely many periods indexed by t. In each period, firms base their production decisions on an individual state and an aggregate state. Individual firm states are denoted by $\omega \in \{1, 2, \ldots, K\}$ and the aggregate state of the industry is $\mathbf{s} = (s_1, s_2, \ldots, s_K)$, where s_j is the number of firms with individual state $\omega = j$ in the current period. In each period, firms chose an input $x(\omega, s)$ which affects their transition to tomorrow's state through the stochastic function $\pi(i|j, x(\omega_t, \mathbf{s}^t)) = \Pr[\omega_{t+1} = i | \omega_t = j, x(\omega_t, \mathbf{s}^t)]$. Let $\mathbf{y}_j = (y_{1j}, y_{2j}, \ldots, y_{Kj})$ denote a profile of tomorrow's firms who started out in state j today, where y_{ij} is the number of firms who find themselves in state $\omega = i$ tomorrow. Finally, let $\mathbf{Y} = (\mathbf{y}_0, \ldots, \mathbf{y}_K)$ denote combined profile of all transitions from today's states to tomorrow's. Derive the following probabilities using your knowledge of probability theory and counting techniques:

    a. $m(\mathbf{y}_j | \mathbf{s}) = \Pr[\mathbf{y}_j | \mathbf{s}]$ (*i.e.*, the probability that s_j firms will transition to profile $\mathbf{y}_j$ tomorrow). *HINT:* use partitioning and independence of individual firm state transitions.

    b. $\mathcal{Q}(\mathbf{s}^{t+1} | \mathbf{s}^t) = \Pr[\mathbf{s}^{t+1} | \mathbf{s}^t]$ (*i.e.*, the probability that the industry will transition from aggregate state $\mathbf{s}^t$ today to $\mathbf{s}^{t+1}$ tomorrow). ∎