STA 580 – Spring 2011 – Dr. Charnigo

Lecture 2

Probability

Introduction. Recall that in Lecture 1 we discussed how to summarize data numerically and graphically. These summaries can help us to make inferences about the population from which a sample is drawn. Lectures 4 through 14 will develop several inferential techniques. However, for those techniques to be comprehensible (i.e., for them to be more than "black box" equations), I need to introduce a probabilistic framework. That framework will be the focus of Lectures 2 and 3. Some students find Lectures 2 and 3 to be challenging, so do not be worried if you have to ask questions and spend some time absorbing the material.

Lecture 2 focuses on basic definitions and computational formulas for probability and conditional probability. But what is probability? There are three ways in which we may characterize probability. I describe them presently and refer to them as the subjective characterization, the frequentist characterization, and the mathematical characterization.

Subjective characterization. The probability of an event is a number between 0 and 1 (sometimes expressed as a percentage between 0% and 100%) indicating to what extent the event is anticipated by a well-trained or well-informed individual. An event with probability 1 (or 100%) is deemed certain to occur, while an event with probability 0 (or 0%) is deemed certain not to occur. A classic example is weather forecasting. The meterologist with the NBC affiliate may say that there is a 20% chance of rain, while the meterologists with the ABC and CBS affiliates may quote figures of 30% and 40%. Another example is setting odds for sporting contests.

Frequentist characterization. Suppose that we could conduct the same experiment over and over again, each time recording whether some event of interest took place. The proportion of occasions on which the event took place would converge to some number between 0 and 1, which would then be defined as the probability of the event. A classic example is flipping a coin repeatedly. The probability of the coin coming up heads is defined as 1/2 if, in the long run, the proportion of flips yielding heads approaches 1/2. This frequentist characterization is implicit in the inferential techniques that we will develop this semester.

Mathematical characterization. Events of interest to us can be conceptualized as mathematical objects called sets. With this conceptualization, the probability of an event is the size of the corresponding set. No set is permitted to have size smaller than 0 or size larger than 1. This mathematical characterization facilitates visualization of events and, as we will see, is indispensable for calculations.

Set-theoretic concepts and notation. When viewed as sets, events are often denoted by capital letters such as A and B. The probability of event A is written $P(A)$ or $Pr(A)$.

• We define $A \cup B$ (read "A union B", Cf. Definition 3.4) to be the event that at least one of A and B occurs; that is, "union" corresponds to the English word "or" (in the inclusive sense).

• We define $A \cap B$ (read "A intersection B", Cf. Definition 3.5) to be the event that both A and B occur; that is, "intersection" corresponds to the English word "and".

• We define $\bar{A}$ (read "A complement", Cf. Definition 3.6) to be the event that A does not occur; that is, "complement" is negation and correspoinds to the English word "not". Some people prefer to write A^c instead of $\bar{A}$.

Example (set-theoretic concepts and notation). Refer to "Genetics" on page 69. Let A denote the event that the first sibling inherits the disease, and let B denote the event that the second sibling inherits the disease. Then
• $A \cup B$ means that at least one of the siblings inherits the disease;
• $A \cap B$ means that both of the siblings inherit the disease;
• $\bar{A}$ means that the first sibling does not inherit the disease;
• $\bar{A} \cap B$ means that the first sibling does not inherit the disease but the second sibling does;
• $\bar{A} \cap \bar{B}$ means that neither sibling inherits the disease; and,
• $\overline{A \cup B}$ also means that neither sibling inherits the disease (because it is the negation of at least one sibling inheriting the disease).

Independent events. We say that events A and B are independent if the occurrence of A does not make B any more or less likely to occur. Such a relationship is governed by the multiplication (Definition 3.7)

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

Examples (independent events). Here are two examples, one to illustrate independence and one to illustrate a lack thereof.

• Example #1. Suppose that I flip a fair coin and then roll a fair die. Let A be the event that the coin comes up heads and B be the event that the die comes up six, so that $P(A) = 1/2$ and $P(B) = 1/6$. What is $P(A \cap B)$? If we repeat this experiment (say) 12000 times, then on about 6000 occasions the coin will come up heads ($12000 \times 1/2 = 6000$).

Intuitively, we know that the coin coming up heads does not alter the chance that the die will come up six. Hence, among the 6000 or so times the coin comes up heads, we will get a six about 1000 times ($6000 \times 1/6 = 1000$). In summary, we anticipate getting a heads and a six about 1000 times out of 12000, which is 1/12. The multiplication $1/12 = 1/2 \times 1/6$ corresponds to Definition 3.7.

Before continuing to the second example, I want to emphasize what makes the multiplication work: the coin coming up heads does not affect the chance that the die will come up six.

• Example #2. Suppose that in a certain community there are 10000 two-parent families. Moreover, suppose that both parents are hypertensive in 3500 of these families, the mother only is hypertensive in 1800 families, the father only is hypertensive in 1700 families, and neither parent is hypertensive in 3000 families. Now I randomly select a two-parent family, with each family having the same probability of selection (namely, 1 in 10000). Let A be the event that the mother is hypertensive, and let B be the event that the father is hypertensive. We have $P(A \cap B) = 3500/10000 = 0.35$, $P(A) = \{3500 + 1800\}/10000 = 5300/10000 = 0.53$, and $P(B) = \{3500 + 1700\}/10000 = 5200/10000 = 0.52$. Yet,

$$P(A) \times P(B) = 0.53 \times 0.52 = 0.2756 \neq 0.35.$$

The multiplication fails! Why is that?

Suppose that I have selected a two-parent family in which the mother is hypertensive. There are $3500 + 1800 = 5300$ such families. The father is also hypertensive in 3500 out of these 5300 families, but 3500 out of 5300 is 0.6604. Thus, once I know that the mother is hypertensive, the chance that the father is hypertensive has escalated from 0.52 to 0.6604. So, the mother being hypertensive and the father being hypertensive are not independent events.

If you think about that from a public health perspective, there is really no surprise: people who live in the same household often have similar diets and exercise habits, so many fathers in families with hypertensive mothers are also hypertensive (and vice versa).

Mutually exclusive events. If A and B cannot both occur, then (Definition 3.2) we say that A and B are mutually exclusive. In this case, we have

$$P(A \cap B) = 0.$$

A common mistake is to confuse "mutually exclusive" with "independent". The two concepts are, in a sense, opposites. If A and B are mutually exclusive, then the occurrence of A not only affects the chance that B occurs but in fact reduces that chance to 0. Thus, knowing that A has occurred enables you to say with certainty that B has not occurred.

Calculating Probabilities

Two simple rules. The following two rules (Cf. Figure 3.5, Equation 3.3; Figure 3.3, Definition 3.6) are often sufficient to calculate probabilities of interest.

1. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ [If A and B are mutually exclusive, then this first rule simplifies to $P(A \cup B) = P(A) + P(B)$.]

2. $P(\bar{A}) = 1 - P(A)$

The subtraction of $P(A \cap B)$ in the first rule prevents double counting if A and B are not mutually exclusive. For instance, if A is the event that I obtain an even number on a single roll of a fair die and B is the event that I obtain a multiple of three, then $A \cup B$ is the event that I obtain a 2, 3, 4, or 6 and $A \cap B$ is the event that I obtain a 6. Hence, $P(A \cup B) = 4/6 = 3/6 + 2/6 - 1/6 = P(A) + P(B) - P(A \cap B)$. If I did not subtract out the 1/6, I would be double counting the possibility of getting a 6.

The second rule is intuitive. For instance, if there is a 1/3 chance of obtaining a multiple of three, then there must be a $2/3 = 1 - 1/3$ chance of not obtaining a multiple of three.

Examples (two simple rules). Consider problems 3.31, 3.32, and 3.33 on page 69. Let A denote the event that the first sibling inherits the disease, and let B denote the event that the second sibling inherits the disease.

• Provided that the siblings are not twins, we may reasonably treat A and B as independent events. Then

$$P(A \cap B) = P(A) \times P(B) = 1/2 \times 1/2 = 1/4.$$

• By the first rule, we have

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 1/2 + 1/2 - 1/4 = 3/4.$$

Applying the first rule again, we have

$$P(A \cup B) = P(\text{exactly 1 sibling affected}) + P(\text{both siblings affected}).$$

[To see this equality, take $C := \{\text{exactly 1 sibling affected}\}$ and $D := \{\text{both siblings affected}\}$. Since C and D are mutually exclusive, we have $P(A \cup B) = P(C \cup D) = P(C) + P(D)$.] Hence, we obtain

$$3/4 = P(\text{exactly 1 sibling affected}) + 1/4,$$

so that the requested probability is 1/2.
• By the second rule,

$$P(\bar{A} \cap \bar{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B) = 1 - 3/4 = 1/4.$$

Conditional Probability

Introduction. In many situations of interest to us, two events A and B will not be independent. That is, A having occurred makes B either more or less likely to occur. Here are two such situations.

• Situation #1. Let B be the event that an individual has a certain disease for which a definitive diagnosis is expensive, while A is the event that the individual is identified as "positive" by an inexpensive but non-definitive screening test. Decisions about medical treatment and care are based on the idea that B is more likely to have occurred if A has occurred than if A has not occurred.

• Situation #2. Let B be the event that an individual develops a certain disease by some specified future date, while A is the event that the individual has a particular behavioral risk factor in the present. The idea that B is more likely to occur if A has occurred motivates public health intervention programs.

Definition of conditional probability. The conditional probability of B given A (Definition 3.9) is denoted $P(B|A)$ and, assuming that $P(A) > 0$, is defined by

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

Note that multiplying both sides by $P(A)$ yields

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

which shows how the multiplication in Definition 3.7 is modified when A and B are not independent: we simply replace $P(B)$ by $P(B|A)$. Indeed, in our earlier example on hypertension in two-parent families, we can recover $0.35 = P(A \cap B)$ from the product $P(A) \times P(B|A) = 0.53 \times 0.6604$. Intuitively, the conditional probability $P(B|A)$ is an "updated" version of $P(B)$ given the information that A has occurred. Note that two events A and B are independent if (Equation 3.5) $P(B|A) = P(B)$.

Computing Conditional Probabilities

Four more rules. Some useful rules involving conditional probabilities (Definition 3.9; Equation 3.6; Equation 3.9) are as follows.

1. $P(A \cap B) = P(A|B) \times P(B)$
2. $P(A) = P(A \cap B) + P(A \cap \bar{B}) = P(A|B) \times P(B) + P(A|\bar{B}) \times P(\bar{B})$
3. $$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B) \times P(B)}{P(A|B) \times P(B) + P(A|\bar{B}) \times P(\bar{B})}$$
4. $P(\bar{A}|B) = 1 - P(A|B)$

The first rule is an algebraic rearrangement of Definition 3.9 with the roles of A and B reversed. The second rule ("Law of Total Probability") applies the first rule to the mutually exclusive events $A \cap B$ and $A \cap \bar{B}$. The third rule ("Bayes' Theorem") combines the first and second rules. The fourth rule generalizes the rule $P(\bar{A}) = 1 - P(A)$ for ordinary probabilities.

Let me emphasize that the roles of A and B can be reversed in all four rules. Thus, for instance, the fourth rule entitles us to assert that $P(\bar{B}|A) = 1 - P(B|A)$.

Moreover, we may substitute $\bar{B}$ for B in any rule, as long as we simultaneously substitute B for $\bar{B}$ (since $\bar{\bar{B}} = B$). Thus, for instance, the first rule entitles us to assert that $P(A \cap \bar{B}) = P(A|\bar{B}) \times P(\bar{B})$.

Likewise, we may substitute $\bar{A}$ for A in any rule, as long as we simultaneously substitute A for $\bar{A}$.

Example (four more rules). The preceding rules[1] give us a powerful arsenal to solve problems involving conditional probability. Here is one such problem.

Suppose that in a certain population 10% of individuals have dementia, 70% of individuals who have dementia are identified as "positive" on a mental status test, and 90% of individuals who do not have dementia are identified as "negative". We would like to know, what percentage of individuals who test positive really do have dementia?

For most students, the hard part of a problem like this is to convert the verbal descriptions into symbols that can be manipulated using the four rules. A good first step is to define events A and B. In this example, we can let B denote the event that an individual has dementia, and we can let

[1] In fact, the second and third rules can be strengthened to accommodate more complicated situations. Let $B_1, B_2, \ldots, B_k$ be events with positive probability such that

$$P(B_i \cap B_j) = 0 \quad \text{if } i \neq j \quad \text{and} \quad \sum_{j=1}^{k} P(B_j) = 1.$$

We say that such events are mutually exclusive and collectively exhaustive. For example, we may take $k = 7$ and let B_1 through B_7 represent the seven diagnoses listed in Table 3.12 on page 74. If $B_1, B_2, \ldots, B_k$ are mutually exclusive and collectively exhaustive, then the Law of Total Probability becomes (Equation 3.7) $P(A) = \sum_{j=1}^{k} P(A|B_j) \times P(B_j)$ and Bayes' Theorem becomes (Equation 3.10) $P(B_1|A) = [P(A|B_1) \times P(B_1)]/[\sum_{j=1}^{k} P(A|B_j) \times P(B_j)]$. Note that we recover the second and third rules as originally stated in the special case that $k = 2$, $B_1 = B$, and $B_2 = \bar{B}$.

A denote that the event that an individual tests positive.

Then $P(B) = 0.10$, $P(A|B) = 0.70$, and $P(\bar{A}|\bar{B}) = 0.90$. That $P(B) = 0.10$ is not difficult to see. However, to understand why $P(A|B) = 0.70$, we must consider what is given and what is uncertain. When we say that 70% of individuals who have dementia test positive, what is given is dementia and what is uncertain is whether an individual tests positive (because 70% of those afflicted test positive and 30% do not). A conditional probability has the form $P(\text{what is uncertain}|\text{what is given})$, so we see that 0.70 is $P(A|B)$. Similar reasoning leads us to see that 0.90 is $P(\bar{A}|\bar{B})$.

Now, what we are looking for is $P(B|A)$. Which rule will help us? That will be the third rule. As a general strategy, try the third rule if you want to go from $P(A|B)$ to $P(B|A)$, try the second rule if you want to go from $P(A|B)$ to $P(A)$, and try the first rule if you want to go from $P(A|B)$ to $P(A \cap B)$.

Performing the arithmetic yields

$$P(B|A) = \frac{0.70 \times 0.10}{0.70 \times 0.10 + (1 - 0.90) \times (1 - 0.10)} = 0.07/0.16 = 0.4375.$$

We conclude that, among individuals who test positive, only 43.75% really have dementia.

Remarks. If we had started by defining A as the event that an individual has dementia and B the event that an individual tests positive, we would have arrived at the same conclusion. Remember that we can reverse the roles of A and B in the four rules.

Some students try to solve probability problems by asserting that "$P(A|B) = 1 - P(A|\bar{B})$", which sort of looks like the fourth rule. However, this is not a valid assertion because $P(A|B)$ and $1 - P(A|\bar{B})$ need not be equal.