A PRIMER ON BAYESIAN STATISTICS

BY

T. S. MEANS

A PRIMER ON BAYESIAN STATISTICS

BY

T. S. MEANS

DEPARTMENT OF ECONOMICS
SAN JOSE STATE UNIVERSITY
SAN JOSE, CA 95192-0114

This material is not to be quoted without permission from the author. Lawrence L. Lapin and Roger Nils Folsom provided helpful comments.

Chapter One

Probabilities: Objective or Subjective?

1.1 Introduction

The basic definitions of probability are presented in this chapter. As the title suggests, there is a question of whether probabilities are defined objectively or subjectively. What you will learn is that the so-called objective definitions contain subjective elements. That is, the distinction between objective and subjective is a matter of form rather than substance.

Before we proceed to the probability definitions it is useful to state a few definitions:

D.1 Experiment: An experiment is any procedure that leads to a well-defined and observable event.
D.2 Elementary Event: The outcome from the experiment.
D.3 Composite Event: A collection of one or more elementary events.
D.4 Probability (Sample) Space: The probability space is a collection of all the possible outcomes from the experiment.

In statistics an experiment must have well defined observable outcomes so that everyone observing the experiment will agree on the outcome. For each experiment there is only one outcome. Thus the elementary events that make up the probability space are mutually exclusive--two events cannot occur simultaneously. However, in most cases we are interested in a group or class of events, which we will call a composite event.

To illustrate the above definitions consider a simple experiment. Toss a six-sided die and record the number on top if it lands flat. For this experiment there are six elementary events: 1, 2, 3, 4, 5, and 6. The composite event, "an even number", contains three elementary events: 2, 4, and 6.

The definition of an experiment may exclude certain outcomes. If the die lands so that it is leaning against an object we will not record the outcome. The reason for this is fairly obvious. If the die lands so that it is tilted, how do we agree on which number to record? Another reason for excluding an outcome is that certain outcomes are not interesting. If a coin is tossed before the start of a football game to determine who will receive the football first, a third outcome--the coin lands on its edge--will not aid in deciding who kicks off.

It is easy to see the importance of a well-defined experiment. While we may rule out certain outcomes, it is not to say that they are impossible and/or cannot occur.

Rather, the exclusion of certain outcomes is to simplify and form a more complete experiment.

The above definitions do not provide a way of assigning a probability to a specific event. From the four definitions it seems reasonable to expect the probability of an event to conform to the following axioms.

Probability Axioms:

A.1 For event A, $0 \leq P(A) \leq 1$.

A.2 For event S, the probability space, which consists of all possible events, P(S) = 1.

A.3 If A and B are mutually exclusive events, then P(A or B) = P(A) + P(B).

The first axiom tells us that probabilities must be positive and bounded. For the second axiom P(S) represents the probability of a certain event, which is equal to one. The third axiom shows how to combine two events. Mutually exclusive events means that during an experiment either event A can occur or event B can occur, but both cannot occur. Since the events are mutually exclusive P(A or B) is the sum of the individual probabilities for each event. The third axiom can easily be extended to cover all mutually exclusive events in S.

The above axioms can apply to a number of situations or experiments, which may or may not make sense in terms of assigning probabilities. For example, if I slice a pie into six equal pieces should the number 1/6 be assigned as the probability of one slice? As we will see in the next section those who use the objective definitions to assign probabilities do not use the term probability to describe all types of events that follow the above definitions and axioms. Furthermore, as pointed out by those who define probabilities subjectively, the objective definitions admit a subjective interpretation.

1.2 Objective Definitions of Probability:

In this section we present the objective definitions used to assign a probability to an event. The first definition is referred to as the Classical definition. The second definition, which has two parts, is the Relative Frequency definition.

D.1 Classical Definition**: If an experiment has N equally likely outcomes and if K of these outcomes are favorable for event A, then**

P(A) = (# of outcomes favorable to A)/(# of equally likely outcomes)
= K/N

D.2 Relative Frequency Definition: If an experiment is repeated N times and S is the number of successes for event A,

$$P(A) = S/N$$

D.3 If we hypothetically let N (the number of trials in D.2) grow and approach infinity then the limit of the Relative Frequency definition approaches P(A).

$$P(A) = \lim_{N \to \infty} (S/N)$$

The first definition is applied to experiments where the outcomes are equally likely. For example, a fair coin is flipped once. The probability that heads will appear is 1/2.

The second definition is applied to experiments that can be repeated. Some textbooks describe repeatable experiments as those that are "standardized". For example flipping a coin may be thought of as a repeatable experiment. Suppose a coin is flipped ten times and four heads is observed. The second definition would have us assign 0.4 as the probability of flipping a head. Since we believe that heads and tails are equally likely events, we don't want the probability of flipping a head to depend on the number of trials. The third definition takes care of this problem by viewing the limit of the observed frequency, so that as $N \to \infty$, S/N equals P(A).

The Classical and Frequency definitions easily apply to coin tossing experiments. However, what about experiments where we are not sure that all the outcomes are equally likely? What about experiments that are not repeatable? As the following discussion will demonstrate, there are a number of experiments with events that we believe should be assigned probabilities but do not exactly fit into any of the objective definitions. The discussion below will also point out how the objective definitions contain subjective elements.

In order to demonstrate the problems inherent in applying the objective definitions let us look at some examples.

Examples

E.1 A fair coin is flipped once. What is the probability that a head will show up?

E.2 The Miami Heat and the Dallas Mavericks are playing in the finals for the next NBA championship. What is the probability that the Heat will win the series?

E.3 What is the probability that the San Francisco 49ers will beat the Miami Dolphins in Super Bowl XIX?

E.4 What is the probability that Pablo Sandoval will get a base hit the next at bat?

E.5 What is the probability that there will be an earthquake tomorrow of magnitude 4.0 or greater?

Suppose we initially attempt to apply the Classical definition to the five examples. For E.1 we have two outcomes, heads and tails, so the probability of a head appearing is 1/2. A similar argument may be applied to the other examples. Either the Heat or Mavericks or the Knicks will win. Either the 49ers or the Dolphins will win. Pablo Sandoval will get a base hit in the next at bat or he will not get a hit. Finally, there will either be an earthquake tomorrow or there will not be one. Hence, we can assign a probability of 1/2 using the classical definition for all five examples.

At this point you should have serious objections to the above analysis. First, while flipping a fair coin may have two equally likely outcomes (heads or tail) it is unlikely that the outcomes for the other four examples are equally likely. In fact, for the fourth example, there are more than two outcomes. The event base hit consists of hitting a single, double, triple, or home run. Likewise, not getting a hit consists of the events striking out, walking, flying out, grounding out, etc. Second, Super Bowl XIX was already played and the 49ers won. The probability that the 49ers win should be one (alternatively, if the 49ers had lost the probability should be zero).

The first objection seems reasonable. Just because there are two teams playing each other in basketball does not necessarily mean that both teams are equally likely to win. In fact people from Miami (or Dallas) will tell you that the Heat (or Mavericks) are more likely to win. However, if it seems reasonable that the outcomes from two teams playing basketball are not equally likely, why does it seem reasonable that the outcomes from tossing a coin are equally likely? In tossing a coin we assumed that a fair coin was flipped. What does the term "fair" mean? To most of us, the term means that both outcomes, head and tail, are equally likely. In this context equally likely means equally probable.

If we apply the Classical definition to experiments where we assume the outcomes are equally likely then the logic is somewhat circular. In other words before we define the probability of an event we have a strong notion of which events are equally probable. Do we really know which events are equally probable before we have defined the probability of an event? It appears that we can apply the Classical definition to an event only if we presuppose a definition of what is meant by equally likely. That is, we define equally likely events before we have defined the probability of an event.

One might argue from past experience that the probability of tossing a head is 1/2. Anyone who has tossed coins before has noticed that the proportion of heads to total trials is very close to 1/2. In other words, an appeal is made to the Relative Frequency definition. When it is pointed out that in most coin tossing experiments, (especially when there is an odd number of trials) the proportion of heads is rarely equal to exactly 1/2, an appeal is usually made to D.3. The relative frequency of heads to trials, (S/N), approaches 1/2 as the number of trials becomes very large.

It is impossible to verify the limit of S/N since either no one can perform the experiment an infinite number of times or demonstrate that the limit exists. While it may be true that the proportion of heads approaches 1/2, it may also be true that the proportion is actually approaching 0.499999999.

There are two problems in employing the Relative Frequency definition to define a probability. The first problem concerns the notion of what constitutes a repeatable experiment. The second problem is how to apply the frequency definition to an experiment that will be performed only once.

Let's discuss the first problem. In a technical sense if we repeat the experiment exactly the same way each time we should get the same outcome. Referring back to the coin tossing experiment, suppose we toss the coin at different heights? Do we want to count all of these trials as part of the same experiment? Does the height of the toss have an affect on the outcome?

In tossing a coin it may be difficult to describe what is different between each toss of a coin. For the other examples it is easier to describe what is different between each trial. For example the Heat and the Mavericks have already played each for the NBA championship. In order to apply the relative frequency definition we need to know the number of trials. Do we include the past championship game, regular season games, exhibition games, games from previous years? Counting only the previous championship game makes very little sense since the teams probably had different players. Counting the games during the regular season may not be helpful if one believes that teams play harder or that the referees call fewer fouls during the play-offs .

The fourth example also points out the difficulty in defining the relevant set of trials. Pablo Sandoval's batting average is a relative frequency measure of base hits to "official" times at bat. Suppose his batting average is .300. Does this mean that the probability of a base hit in his next at bat is .300? Would information would you want concerning the next at bat? For the sake of discussion suppose you learn that Pablo is batting in the bottom of the ninth inning with the winning run at second base with no outs, facing a right-handed pitcher with the cleanup hitter on deck. A baseball expert might expect Sandoval to move the runner to third by grounding out to the right side of the infield and let the cleanup hitter win the game with a sacrifice fly. The expert would believe that the probability that Pablo Sandoval will get a basehit in his next at-bat is very close to zero.

As the above discussion demonstrates, the relative frequency definition provides very little help in assigning a probability if it is difficult to define the relevant set of repeatable trials. Defining the relevant set of trials requires one to define the relevant set of information on which to base the probability. Some of this information may be easy to identify but some information may be impossible to observe or individuals may disagree on its relevance. Clearly the determination of which trials are relevant is personal since it will depend on subjective factors.

The second problem is how to apply the relative frequency definition in assigning probabilities to individual events. For example, suppose a fair coin is tossed only once. The frequency definition would tell you that the ratio of heads to total trials is 1/2. However, the outcome of the next toss is either a head or a tail. You can't flip one-half of a head. Similarly in the baseball example, knowing that Pablo Sandoval gets a base hit 3 out of 10 times does not provide much help in determining the probability of getting a hit in the next at bat. The frequency definition avoids answering the question of how to apply a ratio to an individual event.

Suppose you listen to the following conversation between a student and a Frequentist (a person that believes in the frequency definition).

Student: What is the probability that Pablo Sandoval will get a basehit in his next at bat?

Frequentist: In his last 100 at bats he had 30 hits. If we repeat(?) the experiment 100 more times he will get 30 hits.

Student: You didn't answer my question. What is the probability of a basehit in the next at bat?

Frequentist: In the next 100 at bats he will get 30 hits.

Student: Your not answering my question. In the next at bat the outcome will either be a hit or not a hit. The relative frequency for one trial is either zero or one, and not 0.300.

Frequentist: Your statement is correct. All I can tell about the next at bat is that for the next 100 at bats, there will be 30 hits.

Student: It seems to me that there must be a way to define a probability for an individual trial. Is there a way that frequency statements can be applied to an individual trial?

Frequentist: Yes, there are three ways to apply frequency statements and define probabilities for individual events. However, each position will leave you more confused about using the frequency definition to define probabilities.

The three positions[1] are:

F.1 Probabilities are defined for individual events that have not occurred.
F.2 Probabilities are not defined for individual events.
F.3 Probabilities are defined for individual events if there are no recognizable subsets within the class of events.

The first position, F.1 would argue that probabilities are defined prior to the experiment. For example, (referring back to E.3) prior to the start of Super Bowl XIX it is legitimate to assign a probability that the 49ers will win. After the experiment is performed the probability is either one (the 49ers won) or zero (the 49ers lost). In other words, the probability of an event, given the experiment was already performed is either one (the event took place) or zero (the event did not take place).

---

[1] *Specification Searches*; Ad Hoc Inference with Nonexperimental Data, Edward E. Leamer, 1978, John Wiley & Sons, Inc.

The problem with the first position is how to determine when the probability ceases to be applied to the event. For example, suppose I take a die out of my pocket and toss it in the air and cover it with my hand when it lands. Since the experiment has already been performed, the probability of tossing a six is either a zero or one. When did the probability change from 1/6 to zero or one? Perhaps when I pulled the die out of my pocket? When I held the die a certain way in my hand? When I tossed it in the air? When I covered it with my hand?

Most statisticians subscribe to position F.1. However, they fail to apply the position consistently. Textbooks on statistics will discuss experiments on the probability of an event given some sample information. For example, what is the probability I will find gold on a piece of land given the test results of a geological survey. In a trivial sense nature has already determined the outcome of this experiment. There is either gold on the land or there is not. The probability of finding gold is either zero or one. The fact that you did not observe the outcome does not mean that the experiment has not taken place. I do not know who won the Oscar for best actor in 1961, but if I believed in position F.1, I would not assign a probability for a given actor to the outcome of this event.

The second position, F.2 is similar to the position taken in our hypothetical conversation by the Frequentist. Essentially this position reduces the number of experiments where the frequency definition can be applied. This "conservative" position rules out the notion of assigning a probability to describe the degree of likelihood for an event that will not be repeated. For example an individual adopting this position would not assign a probability as to which team will win the next Super Bowl.

The third position essentially amounts to a personal definition of probability. Recognizable subsets are classes of events where the frequency differs from the relative frequency of all the events. In the base hit example the fact that the batter is a switch hitter may lead you to formulate different probabilities of a base hit depending on whether the batter is batting left-handed or right-handed. Even though the overall frequency of hits is .300, the breakdown for left and right-handed hitting might be .275 and .325 respectively. With more information about the next at bat the Frequentist will identify the relevant frequency that applies to defining the probability of a basehit in the next at bat.

At his point, the Frequentist might argue that if one has all the available information concerning the next trial one can identify the class of events and the relevant frequency to use to define the probability. Unfortunately not everyone will agree on the set of information that is important. For example, in the base hit example does it matter if there are runners on base? What if the game is played at night? Suppose the next pitch is a curveball? What if the hitter is in a "slump"? Is the hitter using his "lucky" bat?

Clearly the set of information to use in defining the probability of a base hit will depend on objective information (i.e. that which we all agree on and can measure; runners on base, night game, E.R.A. of the pitcher, etc.) and also on subjective information (what we identify personally; lucky bat, slumping batter, batter has a will to win, etc.). It seems unlikely that everyone will always agree on the set of information that is relevant in determining the probability of a base hit. Perhaps that is why people would prefer to talk about baseball rather than about coin tossing.

We are now ready to proceed with the subjective definition of probability. Hopefully the above discussion has pointed out the objective definitions contain subjective elements. Though all definitions concerning probabilities are subjective, the subjective definition is not without its own problems.

1.3 Subjective Definition

The subjective definition of probability states that probabilities represent personal degrees of belief about the likelihood of a given event. It seems natural to apply the subjective definition to experiments that are not repeatable. However, what about experiments such as coin tossing? Shouldn't we expect everyone to hold to the same degree of belief? If everyone is observing the same coin toss, why should there be any difference of opinion about the probability of tossing a head?

Clearly a problem with using a personal definition is that people may hold what we think are "unreasonable" opinions about the likelihood of the occurrence of a given event. A second but related problem is whether it is possible to measure degrees of belief.

Concerning the first problem, suppose we have a friend that believes the probability of tossing a head from a coin toss is 0.75. Since talk is cheap we challenge his belief by offering him a gamble. The gamble consists of tossing a coin. If the outcome is a head our friend wins one dollar and if the outcome is a tail we win two dollars. Since our friend believes a head is more likely than a tail he accepts the bet. Unconvinced by his willingness to play for small stakes we increase the stakes to $1000 for a head and $2000 for a tail. At this point our friend refuses to accept the gamble. Did our friend refuse the gamble because of the increase in the stakes or because he really does not believe the probability of a head is 0.75? If our friend values winning in terms of happiness rather than in dollars it is difficult to answer this question. Winning $1000 may provide less increase in happiness than the decrease from losing $2000. Though we think our friend is being unreasonable about his opinion concerning the probability of a head, it maybe difficult to set up an experiment that reveals his "true" opinion.

The second problem is whether one can actually measure degrees of belief. Few people are experts at understanding and predicting earthquakes. While most of us believe the probability of an earthquake striking in our local area is very small it is doubtful that we could measure that belief and state a specific probability. In fact most of us would feel more comfortable if we could state a range of values rather than a specific point.

In some cases we even know that degrees of belief are measured with error but act as if there are not. For example, a scientist could analyze a die and conclude that all six outcomes are not equally likely. You also know that no die has six perfectly cut sides. However, the cost of measuring the exact degree of belief concerning each outcome of the die is very expensive. It is more convenient to believe that all six outcomes are equally likely even though you know that this belief is unlikely to be true.

A final comment concerning measurement is the difficulty in knowing which information to measure. For example, in baseball pitchers and hitters go on "hot" streaks and "cold" streaks. Baseball fans seem to think they know when a player is hot or cold. Similar to predicting the stock market, it's difficult to know when these streaks are going to start and/or stop. Players and fans use phrases like, "seeing the ball clearly," "timing is right," "curveball or slider is breaking hard," "pitchers got his fastball working," etc., to describe why the player is hot or cold. Unfortunately, these phrases fail to point out which variables are important in trying to measure the likelihood of an event; pointing out that it is very difficult to formulate a model and decide which factors are important in determining probabilities.

As an avid basketball player, my field goal percentage is about 50%. There are days when I sense that I can't miss a shot and there are other days when I'm sure that I can't make a shot. When I'm on a hot streak it doesn't seem to matter how hard I concentrate, the ball just goes through the hoop. Even though I shoot 50% overall, I'm not sure which variables are important in determining whether I'll be hot or cold the next time I play in a game.

The material in this chapter has attempted to convince you that all probabilities are based on subjective elements. In most cases we will use the Classical and Frequency definitions for convenience. The same probability laws will be used regardless of whether you use the objective or subjective definition. The intent of this material is to present the Classical methods but show that the alternative approaches (e.g. Bayesian) can add more insight into Classical procedures.