

Experiment of Assigning Probability by University Students

Patrick K. K. Chu

Faculty of Business Administration, University of Macau,

Avenida Padre Tomas Pereira,

Macao SAR, China.

E-mail: patrickc@umac.mo

Abstract

Students who have taken the introductory statistics course were surveyed regarding their knowledge in assigning probability using three different approaches: classical, frequency, and subjective approach. The findings show that most of the students are able to solve the classical probability problems; on the other hand, the students seem not to be confident in specifying the probabilities, which involve frequency or subjective approaches. This study discusses the results and presents the interpretations to the statistics instructors in business schools.

1. Introduction

Interpretation of probability is one of the most important concepts in understanding statistical inference. This survey will investigate how students interpret the probability and they are also required to give their explanations for their answers. By conducting this survey, we get some insights on the students' interpretations of probabilities.

2. Three views of probability

There are three different approaches of probability taught in most of the introductory statistics courses. The first one is **classical** or **priori classical probability** approach. Here the probability of success is based on prior knowledge of the process involved. The probability may be found by counting the number of successes in the sample and dividing the number of outcomes in the sample space. This approach is limited to the situation where the sample space is defined and each outcome is equally likely.

The second approach to probability is the **frequency** or **empirical** approach. Although the probability is still defined as the ratio of the number of factorable outcomes to the total number of outcomes, these outcomes are based on observed data, rather than upon prior knowledge of a process. The assumption of this approach is that the random process can be replicated many times under identical conditions.

The third approach to probability is the **subjective** approach. The probability of success in the previous two approaches is computed objectively, either from prior knowledge or actual data. The probability of favour in the subjective approach is regarded as a numerical measure of a person's opinion of the likelihood of an event. Since the interpretation of the probability is personal, different individuals will have different opinions and assign different probabilities to a given event.

3. Coverage of approach to probability in most popular statistics textbooks

Much of the probability chapters in the textbook used for business students are devoted to the classical viewpoint. On the contrary, few paragraphs have been contributed to the discussion of the empirical approach and subjective approach. Some textbooks which are written exclusively for MBA students even haven't any chapters that discuss the basic probability concepts.

4. What probability is needed in order to understand statistical inference?

Usually, most of the students do not understand why they have to learn probability. In fact, what probability knowledge is needed to communicate basic tenets of statistical inference?

The techniques that students have to learn in inferential statistics are sampling distribution, confidence interval estimation and hypothesis testing. The construction of confidence interval and the testing of hypothesis involve sampling distribution. To understand the concept behind the sampling distribution, students need to know that the sample statistic, such as sample mean and sample proportion, has a predictable probability distribution in repeated sampling. Probability statements about the sampling distribution are reinterpreted as inferential statements such as "95% confidence" and "5% level of significance or type one error".

The interpretation of 95% confidence interval means that there is a 95% chance that the interval may contain the population parameter. It is desirable that students may view a 95% confidence interval as one particular result from a method that is correct 95% of the time in the long run. Looking at the above teaching goals, it is clear that students should understand the frequentist interpretation of probability in order to interpret the statistical inference procedures.

Bayesian statistics is a new stream of statistics in recent years. The Bayesian framework is based on the subjective interpretation of probability. The Bayesian statistician has to specify the prior probability of a certain event subjectively. After the specification of prior probability, the data are observed and this prior probability, simply speaking, the belief will be revised. The students should understand that the subjective probability is conditional on the current knowledge of the individual.

The previous two examples have illustrated that aside from the classical approach, the other two approaches, frequency and subjective approaches, are also important in the introductory course.

5. Classroom survey

To learn about the students' understanding of probability, a sheet with six probability questions was distributed to 136 students from 6 classes of statistics II in May 2004. For each question, the students were asked to (1) make an intelligent guess at the probability, and (2) choose the reasons to explain how he or she obtained the probability.

Question 1: Suppose a card is drawn from a standard deck of 52 cards, what is the probability that a spade card may be drawn?

This question is an easy question for the students. Practically, most of the students know there are 13 spades in a standard deck of 52 cards. So the probability of interest is $13/52=0.25$. The bar chart in figure 1 here shows that 101 out of 130 students with valid answers gave a correct answer. The correct rate is about 74.3%.

Question 2: What is the probability that getting an even number if an unbiased die is tossed?

This question is a slight variation of the past question. The correct probability is 0.5, which is found by the fact that there are three interested outcomes out of six outcomes in the sample space. The bar chart in figure 2 shows there are 88 out of 131 students with valid answer gave a correct answer. The correct rate is about 64.7%.

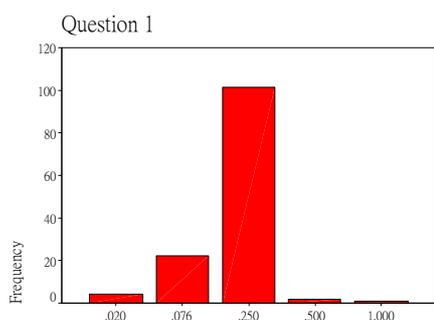


Figure 1

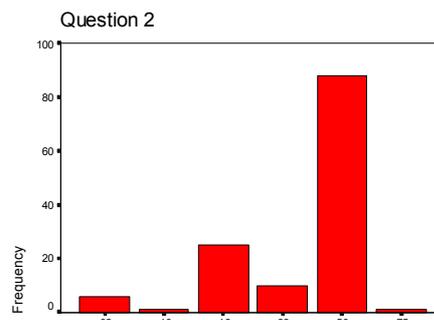


Figure 2

The probabilities of the following two questions may be found by using the frequency viewpoint. In each question, there is a relevant dataset or population that can be used to compute the probability.

Question 3: A previous experiment showed that there were 95 heads might be drawn out of 100 tosses. What is the probability that a head may be got in the next toss?

This probability actually can be found in two perspectives. Students certainly know that there are two possible outcomes when tossing a coin, either tail or head, and most of the students have a strong belief that the coin is fair and so the probability of having a head in the toss is 0.5. The distribution of the responses is portrayed in figure 3. However, the problem here is that the coin landed heads 95 times in 100 tosses before. The students may also use this information and they will estimate the probability of having head to

be $19/20=0.95$. Out of 136 students, 97 (71.3%) use the classical fair coin viewpoint, and only 31 (22.8%) use the frequency viewpoint.

It is interesting that six students gave the answer 0.75, which is between 0.5 and 0.95. I conjecture that these six students believed initially that the chance of landing a head should be 0.5. With the information of landing 19 heads in the previous 20 tosses, their estimation will then be a compromise between their prior belief and the information given.

Question 4(a): If you come across a student at campus, what is the probability that you come across a girl?

Question 4(b): Please state the reason of assigning probability in part (a).

Obviously, the probability in this question may be assigned by visualizing the proportion of girls at University. There are slightly more girls in university and the probability may be 0.6 if following such observations. However, this logic was only used by 24 students out of 135. The percentage rate is 17.6%. Among these 24 students, most of them state that their frequency probability estimate, 0.6, is based on their perceptions of the proportions of girls on campus. A very common response of this question is to say that as there are two only possible outcomes, boy or girl, the probability of coming across a girl on campus is half. There are 73.5% of the students give this answer. Among the students who gave this answer, 91% of them have this belief. Only 2% of them state that they give 0.5 is based on their subjective opinions.

The probabilities in the following two questions are designed to use the subjective view. The probability reflects the student's belief in the likelihood of the event.

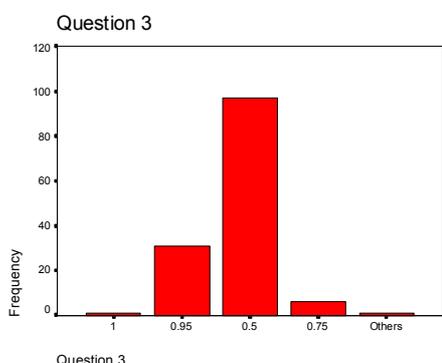


Figure 3

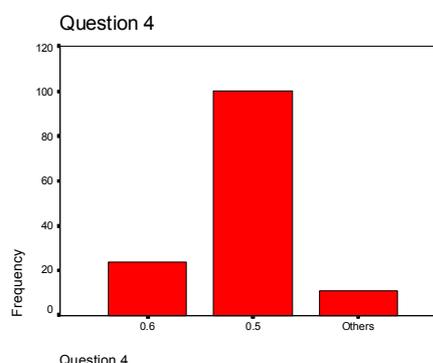


Figure 4

Question 5(a): What is the probability that Porto, a Portuguese soccer team, may win against Monaco in the final of UEFA Champions League this year (2004)?

Question 5(b): Please state the reason of assigning probability in part (a).

If the student had information or if the students were interested in soccer, it would be reasonable to give a brief analysis of the teams' strengths and weaknesses before stating the probability that Porto may win against Monaco in UEFA Champions League. Although this match is a popular and annual important soccer match, it is clear from the responses that students generally knew little about the relative strengths of the two teams, and so it was probably difficult for most of them to give a subjective opinion regarding the outcome of this match.

Of the 135 students with valid answers, 82 (60.3%) gave the equally likely response. Mostly 85.4% of the students, who gave the answer 0.5, said that they have no idea or have no knowledge in soccer. They said that the winner was either going to be Porto or Monaco so the probability is 0.5. These students didn't say that this estimate was based on little knowledge about the teams – they just said that the probability was half since one of the teams was going to win. Of all 136 students, 15 (11% only) students gave the probability based on the subjective opinion. Such selection of criteria is obvious if the students said the Porto's probability of winning UEFA Champions League is 0.4. This opinion may be that Monaco has been playing well in recent games; especially Monaco has won Chelsea to get into the final. Only 20 students (14.7%) assign the probability based on their past results.

The bar chart of the probability estimates is shown in figure 5. As expected, there is a large spike at the value 0.5. Most of the students' subjective probability estimates were larger than half.

Question 6(a): May you guess the proportion chance that you may graduate within four years of study?

Question 6(b): Please state the reason of assigning probability in part (a).

Students should have some subjective opinions on the answer to this question. They should think the likelihood of graduating within four years. They may use a different frequency approach, they may also estimate the probability based on the graduate rates of the undergraduate students.

Of the 133 students who explained their probability, only 15 (11%) gave a frequentist response. Almost half of the students, 68 students (50%) give the probability according to how many credits they have taken already. This choice is especially when they assign the probability be 0 or 1.

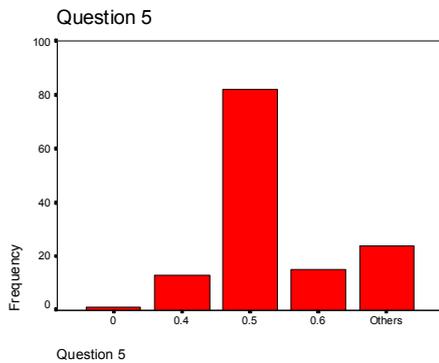


Figure 5

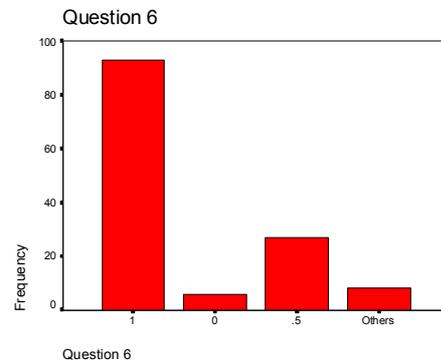


Figure 6

6. Summary of major findings

1. The classical probability questions are much easier for the students to answer than the frequentist and subjective questions. The students seem more comfortable with stylized probability questions.
2. There is a strong inclination for the students to assume the outcomes in the sample space are equally likely even when it is inappropriate to make this assumption.
3. Some students try to compute the probability, even a silly computation, to come up with a probability.
4. The students are reluctant to use their subjective opinion to state a probability.
5. There is a tendency for the students to specify subjective probability of 0 and 1 when they have some knowledge about the particular situation.

7. Implications for teaching

1. The instructors of the introductory statistics courses should discuss the situations where the classical approach cannot be applied.
2. The instructors should spend more time on the frequency approach to probability.
3. The instructors should discuss more on subjective assignment of probability as such practices are important when the students are studying some advance statistics courses, such as Bayesian Statistics.

Bibliography

Berenson, M. L., D. M. Levine, and T. C. Krehbiel, *Basic Business Statistics*, Prentice Hall.