# *Business Analytics: Data Analysis and Decision Making, 5e*

# Chapter 4: Probability and Probability Distributions

# Answers to Conceptual Questions

*Note to Instructors: Student answers will vary. The responses here are intended to provide general guidance in terms of concepts that could be discussed.*

1. If you add the individual probabilities of A and B, you double-count their intersection, the event where A and B both occur. So you should subtract the probability of their intersection from the sum of their probabilities to get the probability that at least one of them occurs.
2. When events are mutually exclusive, this is an extreme form of non-independence. If you are told that A occurs, the probability that B occurs is 0, but if you are told that A doesn’t occur, the probability that B occurs becomes positive.
3. The numberof people who show up is discrete because it is a count. However, there might be such a large number of possibilities that its distribution is *approximated* by a continuous distribution. In contrast, the time between flight arrivals is a continuous measurement, so it would probably be modeled as a continuous random variable. However, as an approximation, it might be discretized, say, into multiples of 5 minutes.
4. This is partly objective and partly subjective. The objective part is that similar epidemics have been observed in the past, so some historical data are available. The subjective part is that each epidemic is somewhat unique. In any case, there would almost surely be some subjectivity in the officials’ assessment.
5. This is true only if the *N* possible outcomes are equally likely. For example, you might be able to count that there are three weather possibilities for tomorrow: sunny, partly cloudy, and rainy. But there is probably no reason to believe that they are equally likely. These counting rules for probability usually arise in games of chance: dealing cards, throwing dice, and so on.
6. The mean and standard deviation are very important measures of a probability distribution, but the entire shape of the distribution, especially the tails, can be relevant in decision making. This is especially true when you are risk averse and large losses (or gains) are possible. The last section of Chapter 6 discusses this in more detail.
7. As of this writing (May 2013), Tiger Woods has stalled at 14 majors wins, but he has already won four tournaments this year (the only multiple winner on the 2013 tour so far). The big question is whether he will reach Jack Nicklaus’s record of 18 lifetime major wins. Even an avid golf fan such as me has difficulty assessing the probability of this event, but at this point, my guess is about 30%. Talk about subjective!
8. Much of today’s statistical analysis is based on statistical theory developed many years ago. The early researchers decided to use standard deviation, with its squared deviations from the mean, as their measure of variability, and almost everything since has been based on this assumption. It’s hard to say whether this decision was inevitable, or how far statistical theory could have progressed if a weighted sum of absolute deviations from the mean had been used instead. Your statistics textbooks might look considerably different if this alternative measure had been adopted.
9. The best way to answer this is that “uncertainty” is always relative to your current knowledge. So to you, the observer, the result of the coin that has already been flipped is still as uncertain as before it was flipped, and the rules of probability still apply in exactly the same way.
10. A distribution on *p* makes sense. For example, it might be based on interviews with many “experts” with different opinions and different sets of knowledge about California earthquakes. Each of them might give you their estimate of *p*, and then you could aggregate these into a distribution on *p*.
11. You and your friend could argue all day about this, even using mathematical arguments, but you wouldn’t get anywhere without data. The only way to resolve the independence issue is to gather data on many families with at least two children. You would divide them into two groups, one where the first child is a boy and the other where the first child is a girl. Then you would check whether the proportion in the first group that has a boy for the second child is the same as the similar proportion in the second group. If these two proportions are nearly the same, you can conclude independence.