ESD.86

Models, Data, Inference for Socio-Technical Systems

Massachusetts Institute of Technology
Engineering Systems Division
Problem Set #1
Issued: Wednesday February 7, 2007. Due: Tuesday February 20, 2007.

1. Twenty-Six Doors. Redo the analysis of the '3-Door" problem (actually, we used 3 envelopes), using the full alphabet of 26 letters. That is, the set up is the same as in class, but there are 26 envelopes in front of the class. In a closed room before class, the TA and the professor placed a 20-dollar bill inside one envelope. Only they know which envelope contains the money. The student contestant picks, say, envelope "L" as her initial choice. Then the TA dramatically opens each envelope one at a time, starting with "A", showing to the class the emptiness of each. But when the TA gets to a particular envelope other that "L", say envelope "Q", he skips that one and works through all remaining envelopes R through Z, showing their emptiness. The contestant is then left with two alternatives: "L" and "Q", both unopened. All others are now known to be empty. The student is allowed to change her mind or to stick with "L". What should she do and why?

2. Weird Country. Suppose there is a country in which all citizens want to construct their families in a particular way. Each married couple agrees to a very unusual family planning strategy. The mother and father (to-be) decide to keep having children until the first girl is born. Then they stop. That is, there is a 'stopping rule': *Keep having children until the first girl is born, then STOP.* For simplicity we assume that all couples can have children and in theory, at least, the number of children in each family is limitless (until, of course, the first girl is born). For modeling purposes assume that the gender of each child is determined by the outcome of a Bernoulli trial, with 50-50 chance of being boy or girl. All such Bernoulli trials are mutually independent.

(a) Determine average family size of completed families.
(b) In steady state, determine the fraction of the population that is female.
(c) Suppose there is a government policy issued that dictates that no family shall have more than five children, and that all citizens strictly obey this edict. So the new stopping rule is: *Keep having babies until the first girl is born or until I have five sons, whichever comes first*. Determine the proportion of sons and daughters within families operating in this new restricted environment.

3. Baseball Championships. In American baseball, the World Series is a 'best of seven' sequence of games, the team that first wins four games being anointed as the "World Champions". Suppose in the next World Series the Boston Red Sox play the San Francisco Giants. Suppose that a Bernoulli trial decides the outcome of any particular game, with probability that the Red Sox will win any given game equal to p. All Bernoulli trials are mutually independent. So, for instance, the chance that the Red Sox would win the World Series in precisely four games is p^4. Suppose that the Red Sox are the superior team, meaning that $p > 0.5$.

(a) What is the probability that the inferior team, i.e., the San Francisco Giants, will win the World Series? Create a graph of this as a function of p. (MS Excel is good for this.)
(b) Some baseball playoff series are 'best of five games,' not 'best of seven,' as the World Series. Answer part (a) again, but this time for a best of five series. Do you have any comment about this? About how reliable are the results of a 'short series"?
(c) In 1903, 1919, 1920 and 1921 the World Series winner was determined through a best-of-nine playoff. Redo part (a) for a best-of-nine series. Are you seeing a trend here? (It might be nice to plot the results as a function of p, with one plot each for best of 5, best of 7, and best of 9.)
(d) For a Best-of-Seven World Series having evenly matched teams (i.e., $p = 0.5$), what percentage of World Series would you expect would require the full seven games to determine the champion? If the teams were not evenly matched, how would this figure change? In fact since 1905, 35 out of 94 World Series have gone to Game 7 - a rate of 37.2 percent. How does this compare with your number(s)? Are the results 'significantly different? Might the differences be explained by inadequacies in our simple Bernoulli model? See http://www.aip.org/isns/reports/2003/080.html .

4. Birthdays. Count the number of people (students and faculty) in our class during the next class period. What is the probability that at least two of us share the same birthday? Hint: see http://www.erin.utoronto.ca/~aweir/107/hw_assigns/bday.pdf .

5. Guru. Suppose over the course of 6 months you receive 10 unsolicited emails from a stranger, each predicting the outcome of some future event. Each event for which he forecasts the results has just two possible outcomes. For instance, one of his predictions could be the name of the team winning the Super Bowl. Another could be the name of the winner in a closely matched 2-person election. Each of these predictions occurs before the event in question, and each turns out to be correct! After you verify that the 10th prediction is correct, you get another email from him saying that he will give you the outcome of some future 11th event, but that you must pay him $100,000 for the information. If his prediction turns out to be true, you could make $1,000,000 with the information. What do you do? How do you think about this problem? Can you construct a simple model explaining why this person may be a total fraud?