Stat 321 - Day 2
Equal Likeliness

Activity 2: Random Babies

Suppose that on one night at a certain hospital, four mothers (named Johnson, Miller, Smith, and Williams) give birth to baby boys. Each mother gives her child a first name alliterative to his last: Jerry Johnson, Marvin Miller, Sam Smith, and Willy Williams. As a very sick joke, the hospital staff decides to return babies to their mothers completely at random. We will first use *simulation* to investigate what will happen in the long run.

Simulation Analysis:

(a) Take four index cards and one sheet of scratch paper. Write a baby's first name on each index card, and divide the sheet of paper into four areas with a mother's last name written in each area. Shuffle the four index cards well, and then deal them out randomly with one index card going to each area of the sheet. Finally, turn over the cards to reveal which babies were randomly assigned to which mothers. Record how many mothers got the right baby.

# of matches:

(b) Do the random "dealing" of babies a total of five times, recording in each case the number of matches:

Repetition #	1	2	3	4	5
# of matches					

(c) Combine your results on the number of matches with the rest of the class, obtaining a tally of how often each result occurred. Record the counts and proportions in the table below:

# of matches	0	1	2	3	4	Total
Count						
Proportion						1.00

(d) In what proportion of these simulated cases did *at least one* mother get the correct baby?

The *probability* of a random event is the long-run proportion (relative frequency) of times the event would occur if the random process were repeated over and over under identical conditions. One can *approximate* a probability by *simulating* the process a large number of times. Simulation leads to an *empirical* estimate of the probability.

(Exact) Enumeration Analysis:

In situations where the outcomes of a random process are *equally likely* to occur (e.g. tossing a fair coin), exact probabilities can be calculated by listing all of the possible outcomes and determining the proportion of these outcomes which correspond to the event of interest. The listing of all possible outcomes is called the *sample space*.

The sample space for the "random babies" consists of all possible ways to distribute the four babies to the four mothers. Let *xyzw* mean that the baby *x* went to the first mother, baby *y* to the second mother, baby *z* to the third mother, and baby *w* to the fourth mother. For example, 1243 would mean that the first two mothers got the right baby, but the third and fourth mothers had their babies switched.

(e) Below is the beginning of a list of the sample size for the "random babies" process. Fill in the remaining possibilities, using this same notation. [Try to be systematic about how you list these outcomes so that you don't miss any. One sensible approach is to list in a second row the outcomes for which mother 1 gets baby 2 and then in the third row the cases where mother 1 gets baby 3 and so on.]

Sample Space:

1234	1243	1324	1342	1423	1432
2134					

(f) How many possible outcomes are there in this sample space? That is, in how many different ways can the four babies be returned to their mothers?

You could have determined the number of possible outcomes without having to list them first. For the first mother to receive a baby, she could receive any one of the four. Then there are three babies to choose from in giving a baby to the second mother. The third mother receives one of the two remaining babies and then the last baby goes to the fourth mother. Since the number of possibilities at one stage of this process does not depend on the outcome (which baby) of earlier stages, the total number of possibilities is the product $4 \cdot 3 \cdot 2 \cdot 1 = 24$. This is also known as $4!$, read "4 factorial."

(g) For each of the above outcomes in your sample space, indicate how many mothers get the correct baby.

(h) In how many outcomes is the number of "matches" equal to exactly:

4: 3: 2: 1: 0:

(i) Calculate the (exact) probabilities by dividing your answers to (h) by your answer to (f). Comment on how closely the exact probabilities correspond to the empirical estimates from the simulation above.

4: 3: 2: 1: 0:

The "number of matches" is an example of a *random variable*, which is a function assigning a numerical output to each outcome in a sample space. Here, each of the 24 outcomes has a corresponding numerical value for "number of matches." This is a *discrete* random variable in that it that can assume only a finite, or countably infinite, number of values. The *probability distribution* of a discrete random variable is given by its set of possible values and their associated probabilities.

(j) Are the possible values of the "number of matches" random variable equally likely? Explain.

(k) For your class simulation results, calculate the average (mean) number of matches per repetition of the process.

The long-run average value achieved by a numerical random process is called the *expected value* of the random variable. To calculate this expected value from the (exact) probability distribution, multiply each outcome of the random variable by its probability, and then add these up over all of the possible outcomes.

(l) Calculate the *expected* number of matches from the (exact) probability distribution, and compare that to the average number of matches from the simulated data.

(m) What is the probability that the number of matches equals this expected value exactly? Based on this probability, do you literally expect to find this number of matches in one realization of the process? Explain.

<u>Computer Simulation:</u>
Now that you have determined the (exact) probability distribution of this random variable, it might be instructive to simulate its behavior in order to examine how closely the results of a given sample match (or do not match) the long-run probabilities. You can use a Java applet to simulate the random assignment of babies to mothers much more quickly. Go to the listing of Java applets and click on "random babies."

(n) Click on "randomize" five times to repeat what you did in class with the index cards. Report the number of matches obtained with these five randomizations.

(o) Turn off the "animation" button and ask for 995 more trials (for a total of 1000). Report the results in the table below:

# of matches:	0	1	2	3	4	mean
Count:						
Proportion:						
Theoretical prob:	.375	.333333	.25	0	.041667	1.00

(p) Do the simulation results come close to the theoretical probabilities?

(q) Click on the 0 bar in the histogram to produce a time plot of the proportion versus the number of repetitions. Then click on the "theoretical values" button to superimpose the theoretical probability on this plot. Then ask for another 1000 trials and click on "randomize" again. Comment on what the resulting graph reveals about interpreting probability as long-term relative frequency (proportion).

(r) Click on "histogram" and then "average" and then "theoretical values" to see a graph of the average number of matches versus the number of repetitions. Comment on what this graph reveals about the interpretation of expected value.