Caltech Department of Mathematics

Ma 3/103 KC Border
Introduction to Probability and Statistics Winter 2017

Assignment 2: Exercises on counting

Due Tuesday, January 17 by 4:00 p.m at 253 Sloan

Instructions: For each exercise please record how much time you spent on it.

When asked for a probability, give both a formula and an explanation for why you used that formula, and also give a numerical value when available. When asked for a numerical probability, evaluate the formula numerically.

This year I am continuing the following experiment. Each assignment will contain zero or more optional exercises. They are optional in the following sense: Grades will calculated without taking the optional exercises into account, but the maximum grade will be an A. If you want an A+, you will have to earn an A and also accumulate sufficiently many optional points. **No collaboration is allowed on optional exercises**.

When a problem says that an element of a set is selected **at random**, assume that each element is equally likely to be chosen.

Exercise 1 (The World Series) (50 pts) The "World Series" is a tournament between the champion of the USA's National League and American League to decide the U.S. Major League Baseball champion. At present, it is won by the first team to win four games out of a possible seven. Since baseball games do not end in ties, at most seven games are ever played.[1]

[1]This is a lie. In the dim past (before night baseball) there were three World Series that had a tie game, when the game was called (shortened) on account of darkness. There were also three Series that used a best-of-nine format.

It is often said that “baseball is a game of inches.” This means that small changes in the physical outcomes of a given play can lead to loss or victory. It also means that the outcome of a game between two teams is effectively random. Let us say that if the probability p that Team A beats Team B is strictly greater than $1/2$, then Team A is *better* than Team B. Note that it is possible (with probability $1 - p$) for the better team to lose a game. Frederick Mosteller [2] estimated (based on data from 44 Series from the first half of the 20$^{\text{th}}$ century) that the probability that the better team wins any given World Series game is 0.65 and that the outcomes of the games are stochastically independent. A few years ago I redid his calculation for all 108 Series through 2012 and came up with 0.59. (You will have a chance to figure this out later in the course with data through the 2016 Series.)

Let p be the probability that Team A wins any given game. Assume that it is the same for every game, and that the outcomes of the game are stochastically independent. The probability that Team B wins is thus $1 - p$.

We can describe the general rule to determine the winner in two ways. Either the winner is first team to win m games, or as the team that wins the most out of $2m - 1$ games. In practice, the series is over as soon as one team wins m games.

1. What is the probability that Team A wins the series in exactly m games? (Give the formula, and explain it.)

2. What is the probability that Team B wins the series in exactly m games?

3. What is the probability that the series is over in exactly m games?

4. What is the probability that Team A wins in exactly $m + 1$ games?

5. What is the probability that the series lasts exactly $2m - 1$ games?

6. What is the probability that Team A wins the series by being the first team to win m games?

7. Suppose the rule was that the teams had to play all $2m - 1$ games. What is the probability that Team A wins the series? What interesting algebraic fact does this prove?

8. What is the probability that the Team A wins a best-of-7 series ($m = 4$) if $p = 0.65$? (Remember to give both a formula and a numeric answer.)

□

Exercise 2 (30 pts) In 1693 Samuel Pepys[2] posed the following question to Isaac Newton.[3]

> In game A I throw 3 dice and win if I get at least 1 ace (an ace is the case where a face shows a single dot or pip). In game B I throw 6 dice and win if I get at least 2 aces. In game C I throw 9 dice and win if I get at least 3 aces. In which game are my chances best?

Newton was unable to convince Pepys of the answer.

1. What is the numerical probability of losing in game A?
2. What are the two mutually exclusive ways to lose game B? What is the probability of each? What is the numerical probability of losing game B?
3. What are the three mutually exclusive ways to lose game C? What is the probability of each? What is the numerical probability of losing game C?
4. Which game has the smallest probability of losing (and therefore the largest probability of winning)?

□

Exercise 3 (Sampling with and without replacement) (30 pts) There is a finite set of n_0 objects, of which the number $d_0 < n_0$ are defective.

1. A sample of size $n < \min\{d_0, n_0 - d_0\}$ is chosen at random *without* replacement. For $d = 0, \dots, n$ what is the probability that exactly d members of the sample are defective?
2. What if the sample is chosen at random *with* replacement?
3. Compare these probabilities for the case $n_0 = 100$, $d_0 = 20$, $n = 10$.

□

[2]Samuel Pepys (rhymes with "peeps") is most noted for the diary he he kept of life in London during a period (1660–1669) that included the Great Fire of London and the Great Plague of London. http://en.wikipedia.org/wiki/Samuel_Pepys

[3]Isaac Newton is least noted for being the Master of the Royal Mint, a position he held for about the last thirty years of his life. http://en.wikipedia.org/wiki/Isaac_Newton

Exercise 4 (30 pts) There are three urns holding various numbers of black, white, and yellow balls. Here is a table showing their composition:

Urn	B	W	Y
1	10	16	4
2	18	10	2
3	17	12	1

An urn is selected at random, and a ball is drawn out. Use Bayes' Law to answer the following questions.

1. For each $i = 1, 2, 3$, what is the posterior probability that urn i was selected, given that a yellow ball was drawn
2. For each $i = 1, 2, 3$, what is the posterior probability that urn i was selected, given that a white ball was drawn?
3. For each $i = 1, 2, 3$, what is the posterior probability that urn i was selected, given that the ball that was drawn is not black?

□

Exercise 5 (10 pts) How much time did you spend on the previous exercises? □

Exercise 6 (Optional Exercise) (50 pts) In a finite sequence of s independent tosses of a fair coin, what is the average number of runs of length r? A **run** is a *maximal* consecutive subsequence of the same symbol. (The symbols for a coin are typically denoted H (Heads) and T (Tails), but they could just as well be 0 and 1.) Let $N(r, s)$ denote the sum of the number of runs of length r over all 2^s sample of length s. The question reduces to finding the formula for $N(r, s)$ and dividing by 2^s to get the average number.

Just to make sure we are on the same page, here are the counts for $N(r, 3)$.

sequence	runs of length 1	runs of length 2	runs of length 3
$\overline{TTT}$	0	0	1
$\overline{TT}\,\overline{H}$	1	1	0
$\overline{T}\,\overline{H}\,\overline{T}$	3	0	0
$\overline{T}\,\overline{HH}$	1	1	0
$\overline{H}\,\overline{TT}$	1	1	0
$\overline{H}\,\overline{T}\,\overline{H}$	3	0	0
$\overline{HH}\,\overline{T}$	1	1	0
$\overline{HHH}$	0	0	1
Total	N(1,3) = 10	N(2,3) = 4	N(3,3) = 2

□

References

[1] W. Feller. 1968. *An introduction to probability theory and its applications*, 3d. ed., volume 1. New York: Wiley.

[2] F. Mosteller. 1952. The world series competition. *Journal of the American Statistical Association* 47(259):355–380. http://www.jstor.org/stable/2281309

[3] E. D. Schell. 1960. Samuel Pepys, Isaac Newton, and probability. *American Statistician* 14(4):27–30. http://www.jstor.org/stable/2681382