Caltech Department of Mathematics

Ma 3/103 — KC Border
Introduction to Probability and Statistics — Winter 2020

Assignment 1: Exercises on counting
Corrected Version

Due Tuesday, January 14 by 4:00 p.m.
in the dropbox in the lobby of Linde Hall
(the building formerly known as Sloan).

Instructions: 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. Feel free to use scientific computing software (Mathematica, R, Matlab, NumPy) or a calculator for these calculations, but do explain what you are doing and why. A printout without an explanation is not an adequate answer.

As in previous years, 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 (50 pts)

In Seven-Card Stud poker, each player receives seven cards, and then chooses five of them to form a poker hand.

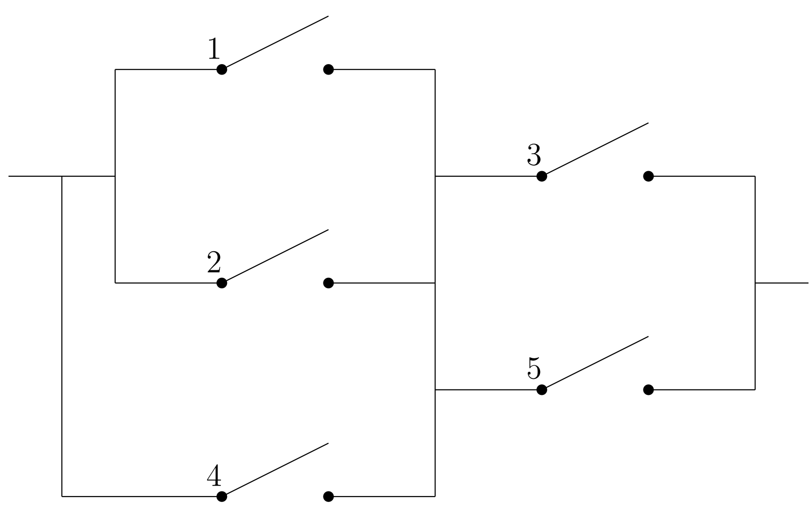
- (25 pts) What is the probability that a random seven-card stud poker hand produces a five-card hand with four-of-a kind?

In bridge, a hand has thirteen cards.

- (20 pts) What is the probability that a random bridge hand contains (exactly) three aces and two kings?
- (5 pts) What is the probability that two players each have at least three aces?

□

Exercise 2 (Inclusion-Exclusion) (20 pts) This is a variant of Pitman, Exercise 7, p. 71. The i^{th} switch in the following circuit is open with probability p_i (and closed with probability $q_i = 1 - p_i$) for each i. Assuming the switches function independently, find a formula for the probability that a current can flow from left to right through the circuit.



□

Exercise 3 (The World Series) (75 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. Under current rules, 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]

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 [1] estimated (based on data from 44 Series from the first half of the 20th century) that the probability that the better team wins any given World Series game is 0.65. 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 for all 111 best-of-7-game series through the 2019 Series.)

Mosteller's model assumes that $p > 1/2$ is the probability that Team A wins any given game. He also assumes that it is the same for every game, and that the outcomes of the game are stochastically independent. (Later we will test this aspect of the model.) The probability that Team B wins is thus $1-p$.

We can describe the general rule to determine the winner in two ways. Let m be the minimum number of games need to clinch the Series (currently $m=4$). 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.

For the first six questions, give the formula, and explain why it is true.

1. (25 pts) What is the probability that Team A wins the series in exactly n games, as a function of m, p, and n? Note that n ranges from m to $2m-1$. Equivalently, what is the probability that Team A wins the series while losing k games, where k ranges from 0 to $m-1$. ($n = m + k$)

2. (10 pts) What is the probability that Team B wins the series in exactly n games?

3. (20 pts) What is the probability that the series is over in exactly n games?

4. (5 pts) What is the probability that the series lasts exactly $2m-1$ games (the maximal length)?

[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 four Series that used a best-of-nine format.

5. (5 pts) What is the probability that Team A (the better team) wins the series by being the first team to win m games? So what is the probability that the poorer team wins the series?

6. (5 pts) 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?

7. (5 pts) What is the probability that Team A wins a best-of-7 series ($m = 4$) if $p = 0.6$? (Give both a formula and a numeric answer.)

□

Exercise 4 (30 pts) According to Schell [2], 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. 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?

(An ace is the case where a face shows a single dot or pip.) Newton was unable to convince Pepys of the correct 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)?

□

[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 5 (Sampling with and without replacement) (30 pts) A lake contains N fish, of which the number $W < N$ are walleyed pike (walleyes).

1. A catch of size $n < \min\{W, N - W\}$ is made at random *without* replacement. For $w = 0, \dots, n$ give a formula for the probability that exactly w members of the catch are walleyes.

2. What changes if the fish are returned to the lake immediately after being caught (sampling with replacement)?

3. Compare these probabilities for the case $N = 800$, $W = 40$, $n = 8$.

□

Exercise 6 (15 pts) How much time did you spend on the previous exercises? □

Exercise 7 (**Optional Exercise**) (50 pts) In a finite sample 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 computations 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] F. Mosteller. 1952. The world series competition. *Journal of the American Statistical Association* 47(259):355–380. http://www.jstor.org/stable/2281309

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