# INFO 7390
# Advances in Data Sciences and Architecture
# Assignment 1 – Review Probability & Stats Solutions

Professor: Nik Bear Brown
Due: September 21, 2019

## Q1 (5 Points)

A certain disease has an incidence rate of 5%. If the false negative rate is 1% and the false positive rate is 3%, compute the probability that a person who tests positive actually has the disease.

S:

A certain disease has an incidence rate of 2%. If the false negative rate is 10% and the false positive rate is 1%, compute the probability that a person who tests positive actually has the disease.

Imagine 10,000 people who are tested. Of these 10,000, 200 will have the disease; 10% of them, or 20, will test negative and the remaining 180 will test positive. Of the 9800 who do not have the disease, 98 will test positive. So of the 278 total people who test positive, 180 will have the disease. Thus

$$P(\text{disease} \mid \text{positive}) = \frac{180}{278} \approx 0.647$$

so about 65% of the people who test positive will have the disease.

Using Bayes theorem directly would give the same result:

$$P(\text{disease} \mid \text{positive}) = \frac{(0.02)(0.90)}{(0.02)(0.90) + (0.98)(0.01)} = \frac{0.018}{0.0278} \approx 0.647$$

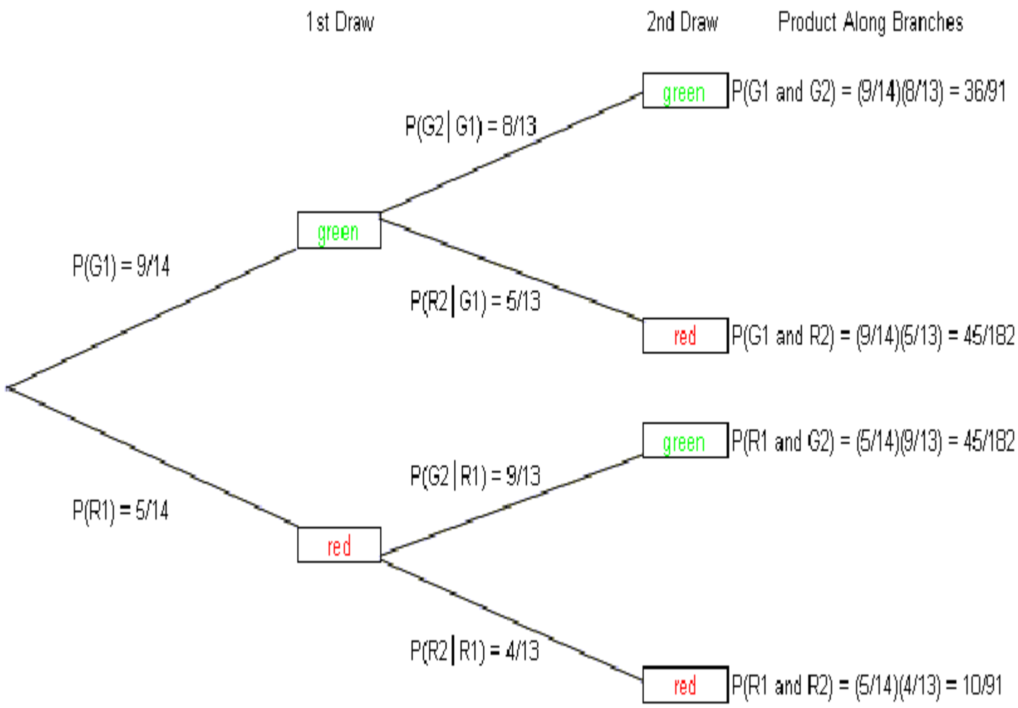
## Q2 (5 Points)

A box contains 5 red balls and 9 green balls. Two balls are drawn in succession **without** replacement. That is, the first ball is selected and its color is noted but it is not replaced, then a second ball is selected. What is the probability that:

**a.** the first ball is green and the second ball is green?
**b.** the first ball is green and the second ball is red?
**c.** the first ball is red and the second ball is green?
**d.** the first ball is red and the second ball is red?

S:

We will construct a tree diagram to help us answer these questions.




**Using the tree diagram, we see that:**

**a.** the probability the first ball is green and the second ball is green = $P(G1 \text{ and } G2) = \frac{36}{91}$

**b.** the probability the first ball is green and the second ball is red = $P(G1 \text{ and } R2) = \frac{45}{182}$

**c.** the probability the first ball is red and the second ball is green = $P(R1 \text{ and } G2) = \frac{45}{182}$

**d.** the probability the first ball is red and the second ball is red = $P(R1 \text{ and } R2) = \frac{10}{91}$

**<u>Formula for Conditional Probability</u>**:
The probability that the second event *B* occurs given that the first event *A* has occurred can be found by:

$$P\left(B|A\right)=\frac{P(A\text{ and }B)}{P(A)},\ \text{ where } P\left(A\right)\neq 0$$

Note: This formula is obtained from the Multiplication Rule for two dependent events.
(Using algebra, we solve for $P(B|A)$ by dividing both sides of the equation by $P(A)$)

**<u>The key to solving conditional probability problems is to</u>**:

1. Define the events.
2. Express the given information and question in probability notation.
3. Apply the formula.

Q3 (5 Points)

How many people must there be before the probability that at least two people have a birthday on October 31 is greater than 1/2?

S:

Note: This is NOT the birthday problem or the probability that, in a set of n randomly chosen people, some pair of them will have the same birthday and a 50% probability with 23 people.

http://en.wikipedia.org/wiki/Birthday_problem

In that problem we are comparing every person to all of the others. We think of that problem in terms of number of possible pair-wise combinations.

In the question asked we are comparing n people with a fixed date, October 31, not making all possible pair-wise combinations.

- Probability that someone in the room has a birthday on October 31, denoted by P(B) is 1− probability that no one in the room has a birthday on October 31, $P(B) = 1 - (364/365)^n$.
- We wish $P(B) \geq 1/2$,

thus $1 - (364/365)^n \geq 1/2$. (*This is fine for full credit*)
Taking logs,
$\log(1/2) \geq \log(364/365)^n$
$-\log 2 \geq n \log(364/365)$
$\log 2 \leq n \log(364/365)$
$253 \leq n$

Q4 (5 Points)

What is the probability of getting exactly 2 heads after flipping three coins?

S:

If one is head and 0 tails:
0 0 0
0 0 1
0 1 0
0 1 1
1 0 0
1 0 1
1 1 0
1 1 1

There are three events in getting exactly 2 heads
{0 1 1, 1 0 1, 1 1 0}
So 3/8 is the probability of getting exactly 2 heads after flipping three coins.

We can also the formula for a discrete random variable based on a binomial distribution:

$$\binom{n}{k} p^k (1-p)^{n-k}$$

X = 1, with probability $\binom{3}{2}(\frac{1}{2})^1(\frac{1}{2})^2 = \frac{3}{8}$

Note: 3 choose 2 is 3. The 3/8 comes from 3* $1/2^3$

Q5 (5 Points)

Consider a six-sided die that gets a 1 with probability p = 1/6. How confident are you that you can get a 1 after rolling the die 3 times?

S:

Because these are independent events (the roll of one die doesn't affect another) we have p = 1/6 chance on each of the die throws. You want probability of at least a 1 in 3 rolls that means total probability (1 - none of the dice has 1). Hence the probability comes out to be $1-(5/6)^3=1-0.58=0.42$

However if we want exactly one success (a roll of 1) in three tries this is just the binomial expansion. If we run n trials, where the probability of success for each single trial is p, what is the probability of exactly k successes?

$$\frac{1-p}{1} \quad \frac{p}{2} \quad \frac{p}{3} \quad \frac{1-p}{4} \quad \frac{p}{5} \quad \ldots\ldots. \quad \frac{\ }{n}$$

k slots where prob. success is $p$, n-k slots where prob. failure is $1-p$

Thus, the probability of obtaining a specific configuration as denoted above is $p^k(1-p)^{n-k}$. From here, we must ask ourselves, how many configurations lead to exactly k successes. The answer to this question is simply, "the number of ways to choose k slots out of the n slots above. This is $\binom{n}{k}$. Thus, we must add $p^k(1-p)^{n-k}$ with itself exactly $\binom{n}{k}$ times.

This leads to the formula:

$$\binom{n}{k} p^k (1-p)^{n-k}$$

X = 1, with probability $\binom{3}{1}(\frac{1}{6})^1(\frac{5}{6})^2 = \frac{25}{72}$

Note: 3 choose 1 is 3

25/72 = 0.3472 see WolframAlpha - (3 choose 1)* 1/6*(5/6)^2 [http://po.st/ZztZy1](http://po.st/ZztZy1)

Q6 (5 Points)

Look up the Boolean satisfiability problem (SAT)
https://en.wikipedia.org/wiki/Boolean_satisfiability_problem

Consider a randomized version of SAT called Max-SAT which tries to satisfy as many clauses as possible with a random polynomial-time algorithm. More precisely, we define Max-SAT as follows: Given a set of k clauses C= {$C_1$, $C_2$ … $C_k$} and n literals X = {$X_1$, $X_2$ … $X_n$} find a truth assignment satisfying as many clauses as possible. Each clause must have at least on literal in it, and all of the literals within a single clause are distinct.

A. What is the expected number of satisfied clauses if each clause has just one literal and we randomly assign the truth value by flipping a fair coin?
B. If each clause has just one literal can we always find a solution that will satisfy all k clauses?

S:

A. What is the expected number of satisfied clauses if each clause has just one literal and we randomly assign the truth value by flipping a fair coin?

Consider a clause C with just one literal probability that it is satisfied is 1/2. If you want to say that the literals assigned to clauses are independent, then the expectation of the sum of the clauses is the same the expected number of satisfied clauses which is k/2 clauses. This is just linearity of expectation slides.

B. If each clause has just one literal can we *always* find a solution that will satisfy all k clauses?
No. Counter example $C_1$ = x and $C_2$ is not x {x} ^ {~x}

Q7 (5 Points)

Compute the probability of randomly drawing 4 cards from a deck and getting exactly three Queens.

S:

$$P(\text{two Aces}) = \frac{({}_4C_2)({}_{48}C_3)}{{}_{52}C_5} = \frac{103776}{2598960} \approx 0.0399$$

Q8 (5 Points)

A group of people decide to do a Xmas gift exchange. In it a bag contains each person's name on a token. Each person selects a token at random. That is the name of the person to give a gift to secretly.

This only works if you get someone else's name, as you would not want to give a gift to yourself.

What is the probability the draw is successful? That is, what is the chance no one selects his or her own name?

S:

See ~37% or 1/e details in Secret Santa SURPRISING Probability https://youtu.be/7iNwyqeEH6Y

Q9 (5 Points)

$P(A) = 0.22$, $P(B) = 0.33$ , $P(A \cup B) = 0.55$

What is $P(A|B)$?

S:

Ans $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.22 + 0.33 - 0.55 = 0$

$P(A|B) = P(A \cap B)/P(B) = 0$

Q10 (5 Points)

Assume P(A) and P(B) are independent.

Which of the following are true?

a) $P(A \text{ and } B) = P(A)P(B)$
b) $P(A|B) = P(A)$
c) $P(B|A) = P(B)$

S:

All of the above.

For the following questions create a homework problem with solutions based on the worked problems in the jbstatistics videos. ONLY when a video has no example calculations can you create a conceptual question.

The solutions for Q11 through Q20 depend on your problem.

Q11 (5 Points)

Create a novel homework problem with solutions for a Z Test for One Mean. See the example in Z Tests for One Mean: An Example https://youtu.be/Xi33dGcZCA0

Q12 (5 Points)

Create a novel homework problem with solutions for creating a Confidence Interval. See the example Inference for One Variance: An Example of a Confidence Interval and a Hy... https://youtu.be/tsLGbpu_

Q13 (5 Points)

Create a novel homework problem with solutions for Deriving a Confidence Interval for a Variance (Assuming a Normal Distrubution). See the example Deriving a Confidence Interval for a Variance (Assuming a Normally Distr... https://youtu.be/q-cHZyOs5DQ

Q14 (5 Points)

Create a novel homework problem with solutions for A One-Way ANOVA Example. See the example A One-Way ANOVA Example https://youtu.be/WUoVftXvjiQ

Q15 (5 Points)

Create a novel homework problem with solutions for Finding the P-value in One-Way ANOVA. See the example Finding the P-value in One-Way ANOVA https://youtu.be/XdZ7BRqznSA

Q16 (5 Points)

Create a novel homework problem with solutions for example Inference for Two Means: Introduction. See the example Inference for Two Means: Introduction https://youtu.be/86ss6qOTfts

Q17 (5 Points)

Create a novel homework problem with solutions for Calculating Power and the Probability of a Type II Error. See the example Calculating Power and the Probability of a Type II Error (A One-Tailed E... https://youtu.be/BJZpx7Mdde4

Q18 (5 Points)

How does a t Distribution differ from a normal Distribution? See the example Introduction to the t Distribution (non-technical) https://youtu.be/Uv6nGIgZMVw

Q19 (5 Points)

Create a novel homework problem with solutions for a Conditional Probability worked problem. See the example Conditional Probability Example Problems https://youtu.be/ES9HFNDu4Bs

Q20 (5 Points)

Create a novel homework problem with solutions for calculating the Power of the Test. See the example Type I Errors, Type II Errors, and the Power of the Test https://youtu.be/7mE-K_w1v90