

PA 818
Professor Geoffrey Wallace
TA: Marc Ratkovic
Fall 2008

Homework #2

Question 1. Newborn Screening

In Wisconsin, all Newborn receive a screening that indicates possible presence of certain genetic disorders. One of the diseases that the newborn screening is intended to detect is Cystic Fibrosis.

Cystic Fibrosis is a genetic disorder. People with CF make mucus that is thicker than normal. Because of the thicker mucus people with CF often develop chronic lung infections and have trouble digesting certain types of food. *In order to get CF you must have two CF mutations (One mutation from each parent since people with CF are typically sterile). If both of your parents are CF carriers (have one mutation) then you have a $\frac{1}{4}$ chance of having CF, a $\frac{1}{2}$ chance of being a CF carrier, and a $\frac{1}{4}$ chance of not having CF and not being a carrier. To be a CF carrier you only need to have one parent that is a CF carrier. If one of your parents is a CF carrier you will have a 50% chance of being a CF carrier. For whites, about 1 in 20 people are CF carriers. A significantly smaller fraction of blacks are CF carriers (1 in 65). To screen for Cystic Fibrosis a sample of blood is taken and tested for its level of Immunoreactive Trypsinogen (IRT). IRT is a chemical which is found in elevated levels in all infants with CF. All infants with CF have high levels of IRT, but IRT levels are independent of being a CF carrier. The vast majority of infants with high IRT levels do not have CF. *Because only a small fraction of infants with high IRT levels have CF, infants with IRT levels in the top 4% of all infants screened are provided with genetic testing. Only 1.5% of these infants will have CF.**

The next step in the CF screening for infants with high levels of IRT is to search for presence of genetic mutations commonly associated with CF. While there are more than 1000 possible mutations associated with CF, this test only looks for the presence 30 of the most common mutations. *Ninety percent of CF mutations fall into the group of 30 mutations that are tested for.* If two mutations are found then the infant has CF with probability one. If one mutation is found then it is known that the infant is at least a carrier for CF (i.e. at least one parent is a carrier) and may have CF (if they have another mutation that was not tested for). *Infants that are in to top 4% of IRT levels, have **exactly one** of the 30 CF mutations that are tested for, and **at least one CF mutation** are said to “**screen positive**” for CF and are referred for further testing.*

- Find the incidence of CF in the population infant with white parents. (**Hint:** Both parents must be carriers).
- Find the probability that an infant with white parents “screens positive” for CF given that he has CF?
- Find the probability that an infant with white parents “screens positive” given that he does not have CF?
- Find the probability that an infant with white parents has CF given that he has screened positive.

Question 2. Fundamentals of Expectations and Variance

Consider the rolling of a 6-sided die. The probability of rolling any given side of the die is uniformly distributed (i.e. the probabilities of rolling any given side of the die are equal). Define the random variable X :

$$X = \begin{cases} 3 & \text{if a 1 or 2 is rolled} \\ -2 & \text{if a 3, 4, or 5 is rolled} \\ 2 & \text{if a 6 is rolled} \end{cases}$$

Find the following:

- a) $E(X)$, $E(X^2)$, $E\left(\frac{1}{X}\right)$
- b) $\text{Var}(X)$:
 - i. Find the variance using the equation $\text{Var}(X) = E\{X - E(X)\}^2$
 - ii. Find the variance using the equation $\text{Var}(X) = E(X^2) - \{E(X)\}^2$
 - iii. Is it possible to have a negative variance? How do you know? Explain both mathematically and intuitively.
 - iv. In words, explain why adding a constant to a variance does not change the variance.

Given the above results, verify the rules of expectation and variance given in class.

- c) $E(4X) = 4E(X)$
- d) $E(4X+6) = 4E(X)+6$
- e) $\text{Var}(4X) = 16\text{Var}(X)$
- f) $\text{Var}(4X+6) = 16\text{Var}(X)$

Question 3. Roulette

A roulette wheel consists of 38 slots, each containing a color and a number. Eighteen of the slots in a roulette wheel are red, 18 are black and 2 are green. In addition to each slot being colored red, black, or green they are numbered 1 through 38. The likelihood of the ball landing in any of the wheels 38 slots is $1/38$. In roulette gamblers can make several types of bets. They can bet either red or black. These bets pay \$1 for each \$1 wagered. Gamblers can also bet on any specific number 1 through 38. These bets pay \$35 for each \$1 wagered.

- A) If the gambler bets red, the gambler wins if the ball rolls into a slot with a red number. If it rolls into a black or green slot, the gambler loses. If the amount bet in the i 'th game is \$5, and the random variable W_i represents the amount of money won or lost on the gamblers bet, find $P(W_i = \$5)$, $P(W_i = -\$5)$, $E(W_i)$ and $\text{Var}(W_i)$ assuming that the gambler bets red in the i 'th game.
- B) Find the gambler's expected value and variance of winnings after 100 bets the gambler always bets red.
- C) If the gambler bets her "lucky" number 13, she wins if the ball rolls into the 13 slot. If it rolls into any other slot she loses. If the amount bet in the i 'th game is \$5, and the random variable $W_i =$ represents the amount of money won or lost on the gamblers bet, find $P(W_i = \$175)$, $P(W_i = -\$5)$, $E(W_i)$ and $\text{Var}(W_i)$ assuming that the gambler bets 13 in the i 'th game.

- D) One of your close friends is planning a trip to Las Vegas and will be playing roulette. He knows you are taking a statistics course and ask for your advice. He is only working with a limited budget of \$500 and wants to know whether it would be best to bet specific numbers or to always bet red. His interest are in making sure the odds are as much in his favor as possible (his expected winnings per bet are highest) and that he can play for as long as possible without going broke. What advice do you give him? Justify your answer (It may be useful to draw a picture).
- E) Your close friend returns from Las Vegas. He ultimately decided to bet red on every wheel turn (he did not necessarily head your advice). After 2099 bets and countless hours of gambling fun he managed to lose his \$500. Shortly after he returns from his trip a roulette wheel scandal breaks. Apparently, the largest manufacture of roulette wheels (Roulette Wheel Inc.) has been making defective wheels. In the defective wheels the probability of the ball landing in any one of the 18 red slots is only $1/50$. Your friend “feels” strongly that the roulette wheel that he was playing was defective.
- i. Describe how you might use Bayes’ rule to calculate the probability that he was playing on a defective wheel. Be as specific as possible without actually doing the calculations. Write actual formulas you would use where appropriate.
 - ii. Do you believe he was playing on a defective wheel? Justify your answer.

Question 5. Binomial versus Hypergeometric

Historically, (over the past 10 national Presidential elections), 70% of votes from a city have been cast for the Republican candidate for President. You need to make an assumption about the proportion of the cities population that that votes Republican and decide to base your assumption on historical evidence (i.e. you assume the proportion that vote Republican is 0.7). Explain what factors might make it unrealistic.

- A) Explain what factors make it realistic.
- B) Assume the city has 1000 voters. If you take a sample of 10 eligible voters, find the probability that exactly 7 votes will be cast for the Republican candidate?
 - a. Use a hypergeometric probability function to solve this.
 - b. Use a binomial probability function to solve this and explain why or why not a binomial is a realistic density for modeling the data generating process.

Question 6. The Supreme Court I

The Supreme Court in the United States regularly relies on statistical evidence. Consider the following quote “Case comment: statistical tests for the analysis of data on peremptory challenges: clarifying the standard of proof needed to establish a *prima facie* case of discrimination in *Johnson v. California*” in [Law, Probability, and Risk](#) (2005) (Click for hyperlink):

“In two recent cases¹ the U.S. Supreme Court clarified the evidentiary burdens the defendant and prosecutor have when alleging that the peremptory challenges the other side made during the jury selection process were based on impermissible group bias. This comment will discuss how formal statistical testing could have assisted the courts in their examination of the peremptory challenges made by the prosecution in *Johnson v. California*² that eliminated all three Blacks eligible for jury service after the challenges for cause had been completed...

Recall that the courts noted that all 3 Blacks were included in the 12 people removed from the jury pool of 43 by the prosecutor. A natural question is: what is the probability that a race-neutral process in which 12 eligible jurors are removed from a pool of 3 Blacks and 40 others would include all 3 Blacks?”

(Hint: to make the calculations easily, use the HYPGEOMDIST(.) and BINOMDIST(.) formulas in Excel)

- a) Answer the question posed by the author, Joseph Gastwirth. (Hint: See question 3 of the section handout for an example.)
- b) What is the probability of including one or fewer Blacks in the jury pool? (Hint: Use $F(x)$, not $f(X)$ in Excel.)
- c) Assume instead that the twelve jurors followed a binomial distribution, drawn from a population that is 80% white, 10% black, and 10% Hispanic. What is the probability of selecting three or fewer jurors who are black?