Practice Exam 2 Math 10 Spring 2006

(I wrote this **fast**, hence there may be tpyso (typos), if you are confused contact me ASAP - Prof. Leibon)

1. (15 points) The Masterfoods company says that before the introduction of purple, yellow candies made up 20% of their plain M&M's, red another 20%, and orange, blue and green each made of 10%. The rest were brown.

   (a) If you pick three M&M's in a row what is the probability that the third one is the first one that is red?

   Answer: Let

   $$E_r = \text{“the third one drawn is the first one that is red”}$$

   Independence is a reasonable assumption and assuming independence,

   $$P(E_r) = P(M \neq red, M_2 \neq red, M_3 = red) =$$

   $$P(M \neq red)P(M_2 \neq red)P(M_3 = red) = (.8)(.8)(.2) = 0.128$$

   (b) If you draw three M&M's, are the events that "an orange is drawn on the first draw" and "the third one drawn is the first one that is red" independent or disjoint or neither. Explain.

   Answer: Let

   $$E_o = \text{“an orange is drawn on the first draw”}$$

   E_o and E_r are not disjoint since both events can occur, and they are not independent since $P(E_o) = .1$ while $P(E_0|E_r) = P(M_1 = orange|(M_1 = red)^c) = .1/(.8) = 0.125 \neq .1$.

2. (10 points) Pricilla is flying form Manchester NH to San Diego CA with a connection in Dallas. The probability that her first flight leaves on time is 0.20. If her flight is on time, the probability that her luggage will make the connecting flight in Dallas is 0.95, but if the first flight is delayed, the probability that the luggage will make it is only 0.60.

   (a) Are the flights leaving on time and the luggage making the connection independent events? Explain.

   Answer: Let

   $$E_l = \text{“ flights leaves on time”}$$

   $$E_c = \text{“luggage makes the connection”}$$

   Notice $P(E_c|E_l) = .95$ while $P(E_c) = P(E_c|E_l)P(E_l)+P(E_c|!E_l)P(!E_l) = 0.67 \neq 0.95$. So the probability of E_c is **dependent** on E_l.

   (b) What is the probability that her luggage arrives in San Diego with her?

   Answer: As above , $P(E_c) = 0.67$.

3. (10 points) Which statements are true about the creation of confidence intervals:

   (a) For a given sample size, lower confidence means a lager margin of error.

   Answer: False. Even for a fixed $\hat{p}$, if I lower my confidence this will shrink z^* and hence the Margin of error.

   (b) For a specified confidence level, larger samples provide smaller margins of error.

   Answer: Logically this is False since size of the margin of error depends on $\hat{p}$ as well as N. For example, if we ran a trial with $N = 100$ and $k = 12$ and were using 95% confidence, then $ME = \pm 0.065$ while if we ran a trial with $N = 110$ and $k = 55$ and were using 95% confidence , then $ME = \pm 0.095$. So the actual marign of error is much larger though the number of trials have increased. However, this claim is True if we use the same (or roughly the same) $\hat{p}$. Namely the standard error $\sqrt{\frac{\hat{p}\hat{q}}{N}}$ gets smaller as N increases for fixed $\hat{p}$. I feel like the book would tacitly want you to make this assumption.

   (c) For a fixed margin of error, larger samples provide greater confidence.

   Answer: False in general, since the confidence will be determine by the number of standard deviations that our fixed ME equals, in other words $ME/SD(\hat{p}) = \frac{\sqrt{N}ME}{\sqrt{\hat{p}\hat{q}}}$. For example, $ME = 0.05$ corresponds to 1.54 standard deviations when $N = 100$ and $k = 12$, but only 1.05 standard deviations when $N = 110$ and $k = 55$. However for a fixed $\hat{p}$ this is True since then $ME/SD(\hat{p})$ increases as N increases. Once again, I feel like the book would tacitly want you to make the assumption that $\hat{p}$ is fixed.

   (d) For a given confidence level, halving the margin of error requires twice the sample size.

   Answer: This is False (both in general and definitely in spirit). For a fixed $\hat{p}$ and confidence it is $\frac{1}{\sqrt{N}}$ that determines the size of the ME; hence we would need to quadruple the sample size to half the error.

4. (15 points) A company with a fleet of 320 cars found that the emissions systems of 11 out of 30 they tested failed to meet pollution control guidelines.

   (a) Is this strong evidence that more than 25% of the fleet might be out of compliance?

   Answer: Well if we viewed the samples as independent trials then we see that a standard deviation assuming $p_0 = .25$ corresponds to $\sqrt{\frac{(.25)(.75)}{30}} \approx 0.08$, and so our data is only $(11/30 - .25)/0.08 \approx 1.45$ standard deviation above the mean. Fortunately, we can safely view our trials as independent since $30 < (0.10)320 = 32$ from our rule of thumb regarding independence. So we want to compute $P(\hat{p} > 11/30|p_0 = 1/4)$. Unfortunately, $1/4(30) = 7.5 < 10$ so the CLT has not necessarily kicked in according to our rule of thumb concern the use of the normal approximation. This would force us to compute

$$P(\hat{p} \geq 11/30|p_0 = 1/4) = 1 - \sum_{k=0}^{k=10} \binom{30}{k} (1/4)^k (3/4)^{30-k} = 0.106.$$

(On an exam I would not have you do this!) This is not very impressive so we must regard our data as consistent with the possibility that the true success rate p satisfies $p = p_0$. Hence this is not "strong evidence".

(b) Test an appropriate hypothesis and state your conclusion.

Answer: This is simply a rephrasing of our solution to part (a). Namely we ran a test where H_0: $p = 1/4$ and H_1: $p > 1/4$. The test's results are conclusions are described above. Once again, we could not safely use the CLT - forcing us to use the binomial, as above.

5. (10 points) It is widely believed that regular mammogram screening may detect breast cancer early, resulting in fewer deaths from that disease. One study that investigated this issue over a period of 18 years was published during the 1970s. Among 30,565 women who had never had mammograms, 196 died of breast cancer, while only 153 of 30,131 who had undergone screening died of breast cancer.

   (a) Do these results suggest that mammograms may be an effective screening tool to reduce breast cancer deaths?

   Answer: We can state a null hypothesis H_0 that $p_N - p_M = 0$ where p_N is the rate of death with no mammograms and p_M is the rate of death with mammograms. The alternate hypothesis H_1 is that $P_N - p_M > 0$. In testing this we need to imagine that under our null that our samples are a random sample of less than 10% of the population of women. Under this assumption we can see how many standard errors our difference of $\frac{196}{30565} - \frac{153}{30131} = \hat{p}_N - \hat{p}_M \approx 0.0013$ is from 0. The pooled percent is $p_p \approx 0.0057$ so we have the standard error is $\sqrt{1/N_N + 1/N_M}\sqrt{p_p q_p} = 0.0006$. So We find our difference is 2.17 standard deviations above 0, and hence I feel comfortable rejecting H_0.

   This assumption of randomness is a very serious one here since it hard to imagine that the experiment was not performed in a self selecting manner and hence in this case their may be many confounding factors (like age, race, previous health conditions, diet or etc). We cannot assess this from what we are told, and are forced to assume that their are no such confounding factors in order to interpret that it was the mammogram screening that led to the reduction in the mortality rate. I would want to look at the method of choosing the two population and the demographics of these samples. If I was satisfied that the randomness criteria was well approximated, then I would conclude that the empirical observed 20 percent reduction in mortality due to breast cancer was likely at least in part due to the mammogram screening.

(b) If your conclusion is incorrect, then what type of error have you committed?

Answer: A type 1 error - in other words the the null is true and we rejected it.

6. (10 points) Recall an American roulette wheel has 18/38 red slots, and imagine that a Tasmanian roulette wheel has 10/38 red slots. You wish to determine whether certain roulette data was produced by a Tasmanian or American Roulette wheel by examining whether the number of times that the wheel landed on a red slot.

(a) Recall that assuming a symmetric risks forces

$$p^* = p_0 + \frac{\sqrt{p_0 q_0}(p_1 - p_0)}{(\sqrt{p_0 q_0} + \sqrt{p_1 q_1})}$$

(where p_0 is the smaller of the two successes rates). Suppose you wish to run an experiment with symmetric risks and that the maximal risk you are willing to take is a 5% chance of an error. What is the experiment's minimal cost? In other words, how large will your sample size need to be?

Answer: Well we have that $p_0 = 10/38$ and $p_1 = 18/18$ and we can plug in to find $p^* \approx 0.362$. We need to find N so that $p_1 - p^* = 1.68(\sqrt{p_1 p_2 N})$, or $N \approx \left(\frac{1.645\sqrt{p_1 q_1}}{p_1 - p^*}\right)^2 \approx 53.9$. Hecne $N = 54$ is the smallest suitable N.

(b) Imagine, that before seeing the data, you believe that the wheel is equally likely to be an American or a Tasmanian wheel. You are told that this roulette wheel came up red 12 times in 40 spins. After seeing this data, with what probability should you now believe that the wheel is Tasmanian?

Answer: Let $P(p = p_0) = .5$ and $P(p = p_1) = .5$ So

$$P(p = p_0|Data) = \frac{P(Data|p = p_0)P(p = p_0)}{P(Data|p = p_0)P(p = p_0) + P(Data|p = p_1)P(p = p_0 1},$$

and $P(Data|p = p_0) = \binom{N}{k} p_0^k q_0^{N-k}$ and $P(Data|p = p_1) = \binom{N}{k} p_1^k q_1^{N-k}$ with $N = 40$ and $k = 12$. So

$$P(p = p_0|Data) = \frac{p_0^k q_0^{N-k}}{p_0^k q_0^{N-k} + p_1^k q_1^{N-k}} = 0.914$$

7. (10 points) In an *efficient market* the fair price for a contract is the expected value and it lies between the price at which the contract can be bought and the price at which the contract can be sold. For a market to be efficient there can be no "free lunch", in other words, there can be no contracts available from which you could make money with no risk. Shares of the following contracts are offered in efficient market.

   Contract: "The party holding this contract receives one dollar from the party who sold this contract if event E_i occurs." The E_i include

$$
\begin{aligned}
E_1 &= \{\text{Hillary Clinton is the 2008 Democratic Presidential Nominee.}\} \\
E_2 &= \{\text{Mark Warner is the 2008 Democratic Presidential Nominee}\} \\
E_3 &= \{\text{John Edwards is the 2008 Democratic Presidential Nominee }\} \\
E_4 &= \{\text{Al Gore is the 2008 Democratic Presidential Nominee }\} \\
E_5 &= \{\text{US Attaks Iran}\} \\
E_6 &= \{\text{ US Attacks Iran and Hillary Clinton is the 2008 Democratic Presidential Nominee.}\}
\end{aligned}
$$

   (a) How are the following quoted prices at which you can buy or sell the below contract in our market related to the probability of the event that it is based upon? (Imagine that these prices include any transaction fees.)

Offering	Sell	Buy
E_1	0.45	0.48

   Answer: Let X_i be 1 if E_i occurs and or 0 otherwise. We find $E(X) = P(X = 1) = P(E_i) =$ Fair Price.

   (b) Is it possible that the market would offer the following opportunity for you to sell or buy the following contract? Why or why not?

Offering	Sell	Buy
E_1	0.45	0.42

   Answer: These prices are not offered since they would produce a free lunch: Simultaneously selling and buying X_i results in 0.03 per share profit with no risk.

   (c) Is it possible that the market would offer the following opportunities for you to sell or buy the following contracts? Why or why not?

Offering	Sell	Buy
E_1	0.45	0.48
E_2	0.33	0.37
E_3	0.15	0.18
E_4	0.13	0.15

   Answer: These prices are not offered since they would produce a free lunch: Simultaneously selling all four X_i results in results in 1.06 but when the nominee is determines you must give up at most 1 dollar.

(d) Is it possible that the market would offer the following opportunities for you to sell or buy the following contracts? Why or why not?

Offering	Sell	Buy
E_1	0.45	0.48
E_5	0.33	0.36
E_6	0.10	0.13

Answer: Yes they could be offered. From this we see $P(E_1 \& E_2) < 0.13$ and $P(E_1)P(E_2) \geq (0.45)(0.33) = 0.1485$, so $P(E_1 \& E_2) \neq P(E_1)P(E_2)$ and hence the events are not viewed by the market as independent. But that does not lead to risk free money, and in fact looking at the events this is quite a reasonable conclusion.

Help Sheet:

Table on page A-98 if needed.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

$$p^* = p_0 + \frac{\sqrt{p_0 q_0}(p_1 - p_0)}{(\sqrt{p_0 q_0} + \sqrt{p_1 q_1})}$$

Geometric: $P(X = x) = q^{x-1}p$, $E(X) = 1/p$, $Var(X) = q/p^2$.

Poisson: $P(X = x) = \frac{\lambda^x}{x!}e^{-\lambda}$, $E(X) = \lambda$, $Var(X) = \lambda$.

Binomial: $P(X = x) = \binom{n}{x}p^x q^{n-x} = \frac{n!}{(n-x)!x!}p^x q^{n-x}$.