Swarthmore Honors Exam 2014: Probability

John W. Emerson, Yale University

NAME: ____

Instructions:

This is a closed-book three-hour exam having 7 questions. You may not refer to notes or textbooks. You may use a calculator that does not do algebra or calculus. Normal, t, and χ^2 tables should be supplied with this exam. Please note:

- A few of the questions have a considerable amount of description and background which is intended to help you. Read this material carefully and completely before working on the problem.
- Most questions have multiple parts. Number the questions and parts clearly in your work, and start each of the questions on a new page.
- Questions that explicitly ask for (or imply the need for) discussion are a chance to demonstrate your understanding of the material. As a guideline, a short paragraph is likely appropriate and preferable to either a single sentence or a longer essay.

1. A market research company employs a large number of typists to enter data into a computer. The time taken for new typists to learn the computer system is well-approximated by a normal distribution with a mean of 90 minutes and a standard deviation of 20 minutes.

a. Calculate the proportion of new typists that take more than two hours (120 minutes) to learn the computer system.

b. Calculate the time below which which 25% of new typists take to learn the computer system.

c. Two typists start learning the computer system at the same time. What is the probability they both learn the system before 70 minutes have passed? Explain any assumptions required for your calculation.

2. Nobel Laureate Linus Pauling (1901-1994) conducted a randomized experiment to study whether taking vitamin C supplements helps prevent the common cold. The results were reported in the *Proceedings of the National Academy of Sciences.* He randomly assigned 279 French skiers to two groups, group C (that took vitamin C supplements) and group S (that took a sugar pill placebo). Here are the results:

```
> pauling

        Caught a cold Did not catch a cold
Group C            17                  122
Group S            31                  109
```

The ultimate question: is there evidence that Vitamin C helps reduce the incidence rate of colds in this population? Let p_c denote the (unknown) population incidence rate of colds for people taking vitamin C supplements, and let p_s denote the (unknown) population incidence rate of colds for people taking the sugar placebo.

a. State the null (H_0) and alternative (H_A) hypotheses for an appropriate test.

b. What test statistic will you use and what is its (approximate) sampling distribution, assuming your null hypothesis H_0 is true? Explain your assumptions, and draw a rough picture of the sampling distribution (but clearly label the picture). Please carefully define any notation you introduce.

c. Find or approximate the p-value of the test and state and justify your conclusions.

3. The total lifetime in days of a certain very delicate mechanical component of a machine is known to be approximately $N(\mu = 100, \sigma^2 = 100)$. After 95 days, your component is still working. How much longer do you expect the component to work?

4. For fixed positive constants w and h let T denote the triangle with vertices at the points $(0, h)$, $(-w, 0)$, and $(w, 0)$. Suppose points (X_i, Y_i) for $i = 1, \ldots, N$ are generated by a Poisson process on T with intensity λ (and thus N has a Poisson distribution and the points are distributed uniformly with respect to area on the triangle). Find the expected value of $\max_i Y_i$. You should interpret the maximum as zero if $N = 0$.

5. Suppose $X_1 \sim U(0, 1)$ and $X_2 \sim U(0, 2)$; and Y_1 and Y_2 are $U(0, a)$ for some constant $a > 0$. Everything is independent. Find a such that

$$\mathbb{E} \max(X_1, X_2) = \mathbb{E} \max(Y_1, Y_2).$$

6. I have a coin that lands heads with probability p_1; you have a coin that lands heads with probability p_2. I toss my coin until I get a head. Each time I toss my coin and get a tail, you toss your coin. Let Y denote the number of heads you get.

a. Find the expected value of Y.

b. Find the distribution of Y.

7. Consider an independent, identically distributed set of random variables X_1, ..., X_n that are known to be uniform on the interval $[0, 1]$. Let $\overline{X}$ denote the sample mean, and $\widetilde{X}$ denote the sample median. You are familiar with the Central Limit Theorem for the sample mean $\overline{X}$. However, there is also a Central Limit Theorem for the sample median $\widetilde{X}$.

Under certain conditions (satisfied here for the median of the X_i), stated casually:

$$\widetilde{X} \sim N\left(0.5, \sigma^2_{\widetilde{X}} = \frac{1}{4n}\right),$$

approximately, for large enough n. Or in general, if $f(m) > 0$, $F(m) = 1/2$, and F is differentiable at m then

$$\sqrt{n}(\widehat{m} - m) \to_d N\left(0, \frac{1}{[2f(m)]^2}\right)$$

where m is the population median, $\widehat{m}$ is the sample median from a sample of size n, and f and F are the population density and cumulative distribution functions, respectively. The conditions essentially ensure the uniqueness of the median.

Now suppose that Y_1, ..., Y_n are independent from the uniform distribution on the set $[-2, -1] \cup [1, 2]$. In this case the conditions above do not apply to the median of the Y_i. You should attempt parts (a) through (e) on this exam. You may choose to answer parts (f) through (j); if you choose not to do so, please come to the oral exam prepared to discuss them.

a. Show that the variance of X_1 is $1/12$.

b. What is $\mathbb{E}(X_1^2)$?

c. What is $\mathbb{E}\left((1 + X_1)^2\right)$?

d. Find the variance of Y_1.

e. Suppose $n = 100$. What is $P(\overline{X} < 0.45)$, approximately?

f. Suppose $n = 100$. What is $P(\widetilde{X} < 0.45)$, approximately?

g. Suppose $n = 100$. Can you find $P(\overline{Y} < -0.05)$, approximately? If so, do it.

h. Suppose $n = 100$. Can you find $P(\widetilde{Y} < -0.05)$, approximately? If so, do it.

i. Suppose $n = 100$. Can you find $P(\overline{Y} < -1.05)$, approximately? If so, do it.

j. Suppose $n = 100$. Can you find $P(\widetilde{Y} < -1.05)$, approximately? If so, do it.