Sample Problems from Previous AM 165 Final Exams

December 4, 2001

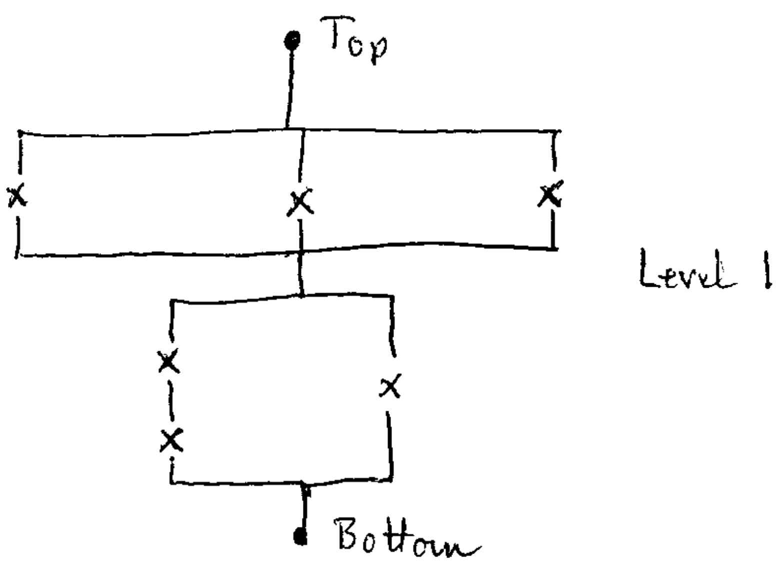
These problems were given on two previous AM165 final exams.

Sample solutions will not be provided for these problems, but you may ask any of the course staff for help and explanations.

A1. Let S be a randomly chosen 10-element subset of $\{1, 2, 3, \ldots, 20\}$. Let Y be the number of elements of S which are divisible by 3.

1. What is the distribution of Y?
2. What is $E(Y)$?

A2. Consider the pictured diagram where each x represents a random variable which either blocks or doesn't block the path — both with probability 1/2. What is the probability of at least one unblocked path existing from the top to bottom?
Hint: Try conditioning on the number of unblocked paths to the first level.



A3. At a theatrical production the program notes instruct us that, after the opening curtain, we will wait Y hours before Godot appears on stage. The notes also advise us that Y has an exponential distribution with mean 1/2. The play lasts 1 hour and Godot never appears. What is the probability of this?
Someone says that if the play had continued 15 more minutes she is *sure* Godot would have appeared. What is the probability that her opinion would be verified?

A4. A density function of the form

$$f(y) = \frac{\theta}{(1+y)^{\theta+1}},$$

where $0 < y < \infty$ and $\theta > 0$, is often used to model the distribution of incomes over a certain level and the distribution of sizes of insurance claims. Suppose we have two independent observations Y_1 and Y_2 from such a distribution.
What is the maximum likelihood estimate of θ?

A5. A public health survey is to be done in a large city to determine the proportion, p, of children, ages 7 to 16, having adequate measles immunization. It is desired that the final estimate, $\hat{p}$, be within 0.03 of p with probability 0.98, no matter what the actual value of p is.
What is the minimum sample size required?

A6. In the article "Clinical assessment of antiplaque agents," in *Compendium of Continuing Education in Dentistry*, 1984, the author reported on a study of the effect of an oral antiplaque rinse on reducing plaque buildup on teeth. Fourteen subjects were randomly assigned to a Treatment Group and a Control Group (7 subjects each). The 7 subjects in the Treatment Group used a rinse that contained an antiplaque agent. The 7 subjects in the Control Group received a similar rinse that contained no antiplaque agent. A plaque buildup score y was measured for each subject after 14 days. Let μ_T denote the true (unknown) mean score for subjects receiving the antiplaque agent and let μ_C denote the true (unknown) mean score for subjects receiving no antiplaque agent.
Assume that the scores are normally distributed and that the (unknown) variance of the scores is the same for all subjects.
The reported results are as follows:

	Treatment Group	Control Group
Sample size	7	7
Sample mean	.78	1.26
Sample standard deviation	.32	.32

a) Find a 95% confidence interval estimate for μ_T.
b) Find a 99% confidence interval estimate for $\mu_C - \mu_T$.
c) What does the result of part (b) show about the effectiveness of the treatment?

A7. A random sample of 100 children from a common age group participated in a special language skills program designed to build vocabulary. Measurements at the end of the program showed that children's vocabularies had a mean of $\overline{X} = 1115$ words and a sample standard deviation of $S = 270$ words. It is known from earlier studies that the mean vocabulary in this age group ordinarily is 1000 words.
Could a result like this possibly have occurred (probability greater than 0.01) if the language-skills program were totally ineffective? (Null hypothesis *no effect* and probability of Type I error $\alpha = 0.01$.)

A8. Suppose X_1, X_2, X_3 are iid random variables from a distribution with mean value μ and variance σ^2. A rather weird estimator of μ is

$$\hat{\mu} = \frac{1}{2}X_1 + \frac{1}{3}X_2 + \frac{1}{4}X_3$$

.

a) Is $\hat{\mu}$ unbiased? Why?
b) What is the mean-square-error of this estimator?

B1. Two typists, Able and Breaker, each do half of a law firm's total typing. The number of typos that Able makes on a page is Poisson with parameter $\lambda = \lambda_1 = 1$ and the number of typos that Breaker makes on a page is Poisson with parameter $\lambda = \lambda_2 = 2$.

(a) A page is taken at random from a morning's output of the two typists and is found to contain 4 typos. What is the probability of finding 4 typos on a random page and what's the probability, after the fact, that Able typed this page?

(b) Let X denote the number of typos on a page taken at random from the combined output of Able and Breaker. Find the pdf of X.

(c) Find E(X).

B2. In a Gallup poll of 1,500 Americans in 1975 (from Gallup Opinion Index, October 1975), 45% answered "yes" to the question "Is there any area right around here—that is, within a mile—where you would be afraid to walk alone at night?" In a similar 1972 poll of 1,500 Americans, only 42% answered "yes."

(a) Test, at the 0.05 significance level, the hypothesis that there is no change vs. the hypothesis that more people are afraid to walk alone at night in 1975 than in 1972.

(b) Give a 95% confidence interval for the difference in proportions of people afraid to walk alone at night in 1975 and in 1972.

(c) If we wanted the confidence interval in (b) to be half as long as it turned out to be, how much larger should each of the sample sizes have been?

B3. In a test of a treatment for eradicating a weed that interferes with wheat production, it is assumed that the proportion Y of treated fields free from the weed will have a density function

$$f(y) = \theta y^{\theta-1}, \qquad 0 < y < 1,$$

where θ is an unknown positive parameter. Fields in five counties are sampled, and the resulting values of Y_1, Y_2, Y_3, Y_4 and Y_5 are $y_1 = 0.45$, $y_2 = 0.68$, $y_3 = 0.87$, $y_4 = 0.36$ and $y_5 = 0.54$.

(a) Find the mle of θ, expressed in terms of y_1, y_2, y_3, y_4 and y_5 and evaluate it for the observed data.

(b) Find the method of moments estimator of θ, expressed in terms of y_1, y_2, y_3, y_4 and y_5 and evaluate it for the observed data.

B4. An English lady claims to be able to detect whether tea was made by adding the milk first and the tea second, or vice versa. She says she is correct 80% of the time. As a test, over a 20-day period at tea time we give her 20 cups to classify. Let X denote the number of times she correctly determines how the tea was made.
We will test her claim that she is right 80% of the time against the possibility that she can't discern the difference at all (null hypothesis), in which case she would be right only 50% of the time.

(a) We will accept her claim (and reject the null hypothesis) if $X \geq c$, where c is an appropriately chosen critical level. How should c be chosen if we want the probability of incorrectly accepting her claim to be less than or equal to 0.05?

(b) For the value of c found in part (a), what is the probability that we incorrectly deny her claim, if in fact it is true?

B5. Suppose the true average growth μ of one type of plant during a one-year period is identical to that of a second type, but the variance of growth for the first type is σ^2, while for the second type the variance is $4\sigma^2$. Let $X_1, \ldots, X_m$ be m independent growth observations on the first type (so $\mathrm{E(X_i)} = \mu$ and $\mathrm{Var(X_i)} = \sigma^2$), and let $Y_1, \ldots, Y_n$ be n independent growth observations on the second type (so $\mathrm{E(Y_i)} = \mu$ and $\mathrm{Var(Y_i)} = 4\sigma^2$). We want to determine the best way to combine the separate sample means $\bar{X}$ and $\bar{Y}$ in order to estimate μ.

(a) Let δ be any constant value in $[0, 1]$. Find the bias of

$$\hat{\mu} = \delta\bar{X} + (1 - \delta)\bar{Y}.$$

(b) For fixed m and n, compute $\mathrm{Var}(\hat{\mu})$ and then find the value of δ that minimizes $\mathrm{Var}(\hat{\mu})$.

B6. A certain drug ("A") was administered to 5 individuals selected at random. After a fixed time period, the concentration of the drug was measured in appropriate units. Suppose the concentrations for the five individuals were found to be

22 28 24 30 26

Suppose that a second drug ("B") was administered to four different individuals, also selected at random. Concentrations were similarly measured with the following results

27 30 31 28

Assume that all the measurements have a normal distribution with a common unknown variance.

(a) Give a 90% confidence interval for the difference in mean blood concentrations of drugs A and B, following administration to randomly chosen individuals.

(b) Is there good evidence that the mean blood concentration of drug B following administration to a randomly chosen individual is higher than that of drug A? Test at the 0.1 level of significance.