Stat 201A, Fall 2012
HOMEWORK 8 (due Thursday 10/4)

1. Complete the argument that we started in class, to show that $E(Y|X)$ is the "least squares" estimate of Y based on X.

2. A box contains a large number of widgets. One-third of them have lifetimes that are i.i.d. with mean μ and variance σ^2. The remaining two-thirds of the widgets have lifetimes that are i.i.d. with mean ν and variance τ^2. Find the mean and variance of the lifetime of one widget picked at random from all the widgets in the box. If this is not possible with the information given, explain why not.

3. A box contains b black balls and w white balls. I add to it d balls picked at random without replacement from a bag that has b_0 black balls and w_0 white balls. Then I pick n balls at random without replacement from the $d+b+w$ balls in the box. Find the expectation and variance of the number of black balls among the n balls.

4. Let N be a non-negative integer valued random variable, and let $X, X_1, X_2, \ldots$ be i.i.d. and independent of N. Define the "random sum" S by

$$\begin{aligned} S &= 0 \text{ if } N = 0 \\ &= X_1 + X_2 + \cdots + X_n \text{ if } N = n > 0 \end{aligned}$$

a) Let ψ be our usual notation for moment generating functions. By conditioning on N, show that

$$\psi_S(t) = \psi_N(\log \psi_X(t))$$

assuming that all the quantities above are well defined. [The identity $w = e^{\log(w)}$ might be handy.]

b) Let N have the Poisson (λ) distribution. Find ψ_N.

c) Recall the formula for the m.g.f. of an indicator, and use the results of the previous parts to find the distribution of the number of heads in a Poisson (λ) number of tosses of a coin that lands heads with parameter p.

5. I am going to toss a coin, but I don't know its probability of landing heads. I put a beta (r, s) prior on that probability, and then I toss the coin. It takes k tosses for me to see the first head.

a) Given this information, what is the posterior density of the coin's probability of landing heads? Are the beta densities a family of conjugate priors in this setting?

b) For fixed r and s, in what way does the shape of the posterior density depend on k? Does the relation make sense intuitively?

6. Normal approximations are often used, without continuity correction, for discrete data. This is not bad if the distribution of the data is indeed roughly bell shaped with a large SD. In this spirit, assume that MSAT and VSAT scores of a large population of students have a bivariate normal distribution with correlation 0.5. Suppose the MSAT scores have an average of 500 and an SD of 90, and the VSAT scores have an average of 480 and an SD of 100. In **a** and **b** below, give your answer in terms of the normal c.d.f. Φ, and then use R to get numerical answers.

a) About what percent of the students scored higher on the MSAT than on the VSAT?

b) Of the students who scored 550 on the MSAT, about what percent scored higher on the

MSAT than on the VSAT?

7. Let X have the exponential distribution with rate λ, and let Y have exponential distribution with rate μ, independent of X.

a) Find $P(X > Y)$ by conditioning on Y. Does your answer make sense when $\lambda = \mu$?

b) Fix $c > 0$, and use the result of **a** to find $P(X > cY)$ without a new calculation. [Yes, you need the distribution of cY, but you know what it is from class.]

c) Stare at your answer to **b** and write down the c.d.f. of X/Y.

d) Use **c** to find the median of the distribution of X/Y. In what way is it related to the medians of the distributions of X and Y? In what way is it related to the means of the distributions of X and Y?

e) Use **c** again, and the "tail integral" formula you derived for expectations in an earlier homework, to notice something interesting about $E(X/Y)$. No, I'm not going to tell you what it is.

8. Suppose you believe that a sample consists of i.i.d. observations from the gamma (r, λ) distribution for some unknown parameters $r > 0$ and $\lambda > 0$. The goal is to use the data to estimate r and λ.

a) Recall the mean and variance of the gamma (r, λ) distribution. Let $X_1, X_2, \ldots X_n$ be the sample, and recall the notation for the kth sample moment:

$$\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k$$

Find a function $\hat{r}$ of $\hat{\mu}_1$ and $\hat{\mu}_2$ that you think will be a good estimate of r if the sample is large; justify your choice. Also find a good estimate $\hat{\lambda}$. This way of estimation is called "the method of moments."

b) It should be clear that it's going to be hard to figure out the distributions of $\hat{r}$ and $\hat{\lambda}$ analytically. So let's start a simulation study. First, pick r uniformly from $(2, 5)$, and pick λ uniformly from $(1, 2)$. Round these numbers so that each has only two decimal places. Assign these as your own personal r and λ.

c) Pretend you are Tyche, and generate 100 i.i.d. gamma observations with your r and λ as parameters. Draw the histogram of this sample, and plot the gamma density over the histogram. Compute $\hat{r}$ and $\hat{\lambda}$ from the sample, and compare the values to the real r and λ. Comment (this will be qualitative and subjective) on whether your histogram looks roughly like the density, and whether your estimates are close to the true values.

d) To understand the variability in your estimates, generate 999 more samples of size 100 each, and compute $\hat{r}$ and $\hat{\lambda}$ for each one. You should end up with 1000 values of $\hat{r}$ and 1000 of $\hat{\lambda}$. Draw the histograms of the $\hat{r}$'s and the $\hat{\lambda}$'s, and comment on their shapes. Find their means and say whether the answers are as you expected. Also find the SD of the $\hat{r}$'s and the SD of the $\hat{\lambda}$'s. These numbers quantify the variability in your estimates.

9 (continuing Problem 8). Keep the same r and λ that you had in the previous problem, and repeat **c** and **d**, but with the sample size changing to 1000. That is, get 1000 samples of size 1000 each. Do the same analysis as in the previous problem, and also comment on any difference that arises due to the increase in sample size.

10 (continuing Problem 9). Sadly, you're not Tyche. So when you are analysing data that you didn't generate, you can't expect to know what the real r and λ are, which means that you can't create lots of samples from the true distribution to see how your estimates behave. So then what? Well then you can use the "bootstrap" method, as follows.

Revert to your status as human being and suppose you don't know the real r and λ. Also suppose that the first sample (size 1000) that you generated in Problem 9 is the entire set of data that you have. Get your $\hat{r}$ and $\hat{\lambda}$ from this sample. Then *pretend that these are the real values of r and λ,* and generate 1000 samples of size 1000 each, each i.i.d. gamma with parameters $\hat{r}$ and $\hat{\lambda}$ estimated from your original sample. Compute $\hat{r}$ and $\hat{\lambda}$ from each of these samples.

Draw the histogram of the new $\hat{r}$'s. For the purposes of estimation, you could say that you're about 95% confident that the true r is in the range given by the "central 95%" interval of the distribution you generated; that is, the interval that's left after you've chopped off tails of 2.5% on both sides.

To see if your estimated interval is reasonable, compare the histogram of $\hat{r}$'s in this problem to that in Problem 9. How different they are will depend on how your original sample turned out, but it is unlikely that they are very far off.

Now do a similar analysis for the $\hat{\lambda}$'s.