Assignment #2 — Due Wednesday, February 4, 2008, by 5:00 P.M.

Turn in homework in lecture, discussion, or your TA's mailbox. Indicate the discussion section in which you expect to attend to pick up this assignment on the assignment.

311: Monday 1:20–2:10 **312:** Monday 12:05–12:55

For each problem, you may use a combination of paper/pencil calculations and or calculations with R. If you use R, include the commands you use for your solution. You do not need to include all of the output, but you can cut and paste if you wish.

The file `prob.R` has some functions for graphing and visualizing some probability distributions. You need to download this file and `source()` it to use the functions.

I will likely improve and add to these functions as the semester progresses. The graphing functions use a prefix of g, and are similar to the R functions for calculating probability densities/functions (prefix d), cumulative distribution functions (prefix p), quantile functions (prefix q), or random number generation (prefix r). The first argument is `endpoints` which is a vector with the range over which you wish to display the distribution. Additional arguments specify the parameter values. For example, the command `gexp(c(0,4000),1/1000)` will draw a graph of the Exponential(1/1000) density from 0 to 4000 and the command `ggamma(c(0,10),2,1)` graphs the Gamma$(2, 1)$ density from 0 to 10.

The functions require the `lattice` package. If not already installed, you can install this package with this command (assuming your computer is connected to the internet).

```
> install.packages("lattice",repos="http:cran.us.r-project.org")
```

1. A brand of lighbulbs is advertised to have a mean life of 1000 hours.
   (a) If the lifetime distribution is exponential, what is the value of the parameter λ? Find the shortest interval that contains 0.95 of the probability. *(Hint: this interval corresponds to where the density is highest.)*
   (b) If the lifetime T has a Gamma$(4, \lambda)$ distribution with density $f(t) = (\lambda^4/6)t^3 e^{-\lambda t}$ for $t > 0$ and mean $\mu = 4/\lambda$, find a value t for which $\mathsf{P}(T > t) = 0.9$. Write the answer in terms of the cumulative distribution function F (and/or its inverse, the quantile function F^{-1}) and find a numerical answer using R.
   (c) If the lifetime T of a new lightbulb has a Gamma$(4, \lambda)$ distribution, find a value t for which $\mathsf{P}(T > 500 + t \,|\, T > 500) = 0.9$. The value t represents a lower estimate of additional life with 90% probability given that the bulb has been already used for 500 hours and still works. Write the answer in terms of the cumulative distribution function F (and/or its inverse, the quantile function F^{-1}) and find a numerical answer using R.
   (d) Graph both the exponential and gamma densities from this problem. Verify by eye that the numerical answers you have found look reasonable.

2. Suppose that a random variable $X \sim \text{Beta}(\alpha, \beta)$, so

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 < x < 1$$

   and assume that $\alpha, \beta > 1$.

   (a) Find the mean (expected value) and mode (point that maximizes $f(x)$) in terms of α and β.

(b) Assume that $\alpha = 2$ and $\beta = 3$. Assess the accuracy of a predicted value x_0 for X by calculating the mean squared error,

$$\mathsf{E}((X - x_0)^2) = \int_0^1 (x - x_0)^2 f(x)\,\mathrm{d}x\ .$$

Compare the accuracy of the mean and the mode for prediction of X by computing the ratio of the mean square errors. Which estimate is better by the mean square error criterion? *Hint: Recall the definition of the Beta function.*

3. Do Exercise 5.3.1.

4. Do Exercise 5.3.3.

5. Do Exercise 5.3.8.

6. Do Exercise 5.4.5. Enter the data in R either directly,

```
> ex5.4.5 = c(3.9, 7.2, 6.9, 4.5, 5.8, 3.7, 4.4, 4.5, 5.6, 2.5,
+     4.8, 8.5, 4.3, 1.2, 2.3, 3.1, 3.4, 4.8, 1.8, 3.7)
```

or by putting the data into a file (say `ex5.4.5.dat`) and using `scan()`.

```
> ex5.4.5 = scan("ex5.4.5.dat")
```

Another useful trick is to use `file.choose()` and then use the operating system to read in the data.

```
> ex5.4.5 = scan(file.choose())
```

In the `lattice` package, there is a function `histogram()`. You can control the breaks with the arguments `endpoints`, `nint` (number of intervals), or `breaks`. Here are three ways to set the breaks to be at the values 0, 2, 4, 6, 8, 10.

```
> library(lattice)
> print( histogram(~ex5.4.5,endpoints=c(0,10),nint=5,type="density") )
> print( histogram(~ex5.4.5,breaks = c(0,2,4,6,8,10),type="density" ) )
> print( histogram(~ex5.4.5,breaks = seq(0,10,by=2),type="density") )
```

The file `stat310.R` has some functions of general purpose usefulness. I will add to this file during the semester. The function `easyLayout()` can be used with `lattice` to put multiple plots in a single figure. This function requires that you add the package `grid`. (There are very general ways to do this, but this function should be simpler to use.) Here is an example that puts two figures in an array of one row and two columns.

```
> figA = histogram(~ex5.4.5,breaks=c(1,4.5,5.5,6.5,10),type="density")
> figB = histogram(~ex5.4.5,breaks=c(1,3.5,4.5,6.5,10),type="density")
> easyLayout(list(figA,figB),1,2)
```

Work to do, but not turn in.

- Read sections 5.4–5 and 6.1–2.