Project I

Instructions: Complete the following exercises.

1. To better understand the second problem on the Midterm about the gamma density, run the available R code to see the effect of increasing α.

   (a) What happens to the gamma density as α increases? What kind of parameter is α?

   ```
   # plot the gamma p.d.f

   x = seq(0,10,0.1)

   alpha = 10
   lambda = 5

   curve(dgamma(x, shape = alpha, rate = lambda), 0, 20, add = TRUE)


   alpha = 20      # bigger alpha here
   lambda = 5      # bigger lambda here

   curve(dgamma(x, shape = alpha, rate = lambda), 0, 20, add = TRUE, lty = 1)

   for(i in 1:10){

   alpha = 20 + 10*i     # bigger alpha here
   lambda = 5            # bigger lambda here

   curve(dgamma(x, shape = alpha, rate = lambda), 0, 20, add = TRUE, lty = i)

   }
   ```

   (b) Change the R code to investigate the effect of changing λ. What happens to the gamma density as λ increases? What kind of parameter is λ?

2. To better understand the fifth problem on the Midterm about the logistic distribution, run the available R code.

   ```
   # plot the logistic p.d.f.

   x = seq(0,1,0.0001)
   #x

   y = log(x/(1-x))
   #y
   ```

```
X11()
par(mfrow=c(2,2))
plot(x,y,type='l',main="transformation")

F_y = exp(y)/(1+exp(y))
#F_y

plot(y,F_y,type='l',main="cdf")

f_y = exp(y)/(1+exp(y))^2
#f_y

plot(y,f_y,type='l',main="pdf")

mu = 1
sigma = .5
f_y = (1/sigma)*exp((y-mu)/sigma)/(1+exp((y-mu)/sigma))^2
f_y

plot(y,f_y,type='l',main="pdf, mu = 1, sigma=0.5")
```

(a) Print the pictures for each part of problem 5. Clearly label each picture for each part.

3. Read the two BVN Handouts from the class website.

   - BVNmatrixnotation.pdf
   - 2dRotationalTransformation.pdf

(a) What are the two transformations that take two independent normal random values to two correlated/dependent normal random values?

(b) What are the two rotational transformations that take two correlated normal random values to two independent normal random values?

(c) Run the code provided in the RBVNsim3.R handout. Verify that the transformations perform as described.

4. Investigate the use of the R packages, IDPmisc and gplots. In the IDPmisc package there is a command ipairs that is useful for plotting large amounts of data using color. In the gplots package there is a command hist2d that is also useful for large data and uses color.

   Run the code provided at the bottom of the RBVNsim3.R handout. Use the output to better visualize the correlation/dependence between the random values that have been simulated in the previous part.

Name:

Section: 2p.m. 6p.m.

CALIFORNIA STATE UNIVERSITY, EAST BAY

STATISTICS DEPARTMENT

Statistics 6401 Advanced Probability

Fall 2012

Instructions: This is a closed book and closed notes exam. You may use a pencil or a pen (blue or black), and write your answers on your own paper. Please write you name on the exam sheet and turn it in on top of your written solutions on your own paper. Please assemble your papers in the order of the problems on the test. The test has a total of 75 points. Use your time appropriately.

Please see the class website for PART I of the class project.

1. **(5 points)** If you were to choose a bivariate p.d.f. to model the joint distribution of weights and lengths of new born children, what would it be? Sketch the contour lines and give guesses of the parameters in your model.

2. **(15 points)** Let X be a gamma random variable with p.d.f.

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\lambda x) \quad (1)$$

for $x \geq 0$ and $f(x) = 0$ otherwise, where $\alpha > 0$ and $\lambda > 0$.

   (a) For $\alpha = 10$ and $\lambda = 5$ sketch the p.d.f. of X.

   (b) For $\alpha = 20$ and $\lambda = 5$ sketch the p.d.f. of X. (Hint: Sketch the p.d.f.'s on the same set of axes.)

   (c) Describe the change in the gamma p.d.f. as α is increased, holding λ fixed.

3. **(10 points)**

   (a) Find the joint density of $X_1, X_2, ..., X_n$, independent random variables sampled from the $Gamma(\alpha, \lambda)$.

   (b) Find the joint density of $X_1, X_2, ..., X_n$, independent random variables sampled from the $Normal(\mu, \sigma^2)$.

4. **(25 points)** A large corporation has a policy to randomly test its employees for the use of a certain illegal drug. Suppose that each employee can be classified uniquely into one of three groups:

- Those who never use the drug (G_1): 75% of all employees.
- Those who occasionally use the drug (G_2): 20% of all employees, and
- Those regularly use the drug (G_3) 5% of all employees.

Thus $G_1 \cup G_2 \cup G_3 = S$, the sample space of all employees, and $G_i \cap G_j = \phi$, for $i \neq j$. The probability that this testing procedure used will signal use of the drug (T) differs depending on an individual employee's use of the drug: 10% for those in group G_1 (false accusation), 60% for those in group G_2, and 95% for those in group G_3. Suppose employees are chosen at random for this test.

(a) From the information given above, evaluate the following five probabilities: $P(S)$, $P(G_1)$, $P(G_1^c)$, $P(T|G_1)$, $P(T^c|G_1)$. Here superscript c indicates complement and the vertical bar indicates a conditional probability.

(b) Evaluate the probabilities $P(G_1 \cap T)$ and $P(T)$. Say what rule of manipulating probabilities you are using in each case.

(c) Now we look at some "inverse" conditional probabilities that help in assessing the usefulness of the drug test results. Recall the following definition of the conditional probability $P(A|B)$ for two arbitrary events A and B with $P(B) > 0$:

$$P(A|B) = P(A \cap B)/P(B).$$

Use this equation to find $P(G_1|T)$ and $P(G_3|T)$.

(d) Consider the three probabilities $P(G_1 \cap T)$, $P(G_1|T)$, and $P(T|G_1)$. As you can see, all three events involve both of the events G_1 and T. However, each probability refers to a different sample space. Name the sample space in each case, either with a description or a letter.

(e) The head of an employees union wants to argue that random use of this drug test is unfair to employees. Which one or two of the probabilities above do you believe is most supportive of his argument, and why?

5. **(20 points)**

Suppose that $X \sim U(0, 1)$. Consider the transformation

$$Y = \log\left(\frac{X}{1-X}\right).$$

(a) Determine the cumulative distribution function of Y, $F_Y(y)$. This distribution of Y is known as the standard logistic distribution.

(b) Determine the probability density function of Y, $f_Y(y)$.

(c) Is the standard logistic distribution a symmetric distribution? Hint: consider the probability density function.

(d) Determine the probability density function of $Z = \mu + \sigma Y$. This distribution of Z is known as the logistic distribution with parameters μ and σ.