Stat 111 Spring 2011 Week 2
Continuous Probability Problems

1a. Let the random variable (r.v.) X be the time until the next event in a Poisson process with rate λ events per unit time.

a) Show that $X \sim \text{Exp}(\lambda)$. That is, the density function for X is

$$f_x(x) \;=\; \lambda e^{-\lambda x}, \quad x > 0.$$

b) Find $E(X)$ and $\text{Var}(X)$, the mean and variance of the r.v. X.

c) Find the density function for $Y = \sqrt{X}$.

d) Find the Moment Generating Function (MGF) for X:

$$M_x(t) \;\equiv\; E_x\left(e^{tx}\right) \;=\; \int_{-\infty}^{\infty} e^{tx} f_x(x)\, dx \;=\; \int_0^{\infty} e^{tx}\left(\lambda e^{-\lambda x}\right) dx.$$

e) Assuming an MGF $M_y(t)$ is differentiable in a neighborhood around 0, show that the kth derivative (with respect to t) of $M_y(t)$, evaluated at $t = 0$, yields the kth moment of Y:

$$M_y^{\{k\}}(0) = E(Y^k) \;=\; \int_{-\infty}^{\infty} y^k f_y(y)\, dx.$$

(Note my use of the *Law of the unconscious statistician* in both d and e.) Show this works for $k = 1$ and $k = 2$ with the Exponential distribution.

f) Show that the MGF of the sum of two independent random variables is the product of their two MGF's. Find the MGF for the waiting time until the kth event from a Poisson(λ) process.

1b. If $X \sim \text{Gamma}(\alpha, \lambda)$, with $\alpha > 0,\ \lambda > 0$, then the density function for X is

$$f_x(x) \;=\; \frac{\lambda^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1}\, e^{-\lambda x}, \quad x > 0.$$

a) Find the mean and variance of X. Hint: $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$, for $\alpha > 0$.

b) Find the distribution of $Y = cX$, for a positive constant c.

c) Find the Moment Generating Function for a Gamma(α, λ) random variable. What can you conclude (from this, the uniqueness of the MGF and from 1a part f) about the relationship between the Gamma and Exponential distributions?

d) Suppose $X \sim N(0, 1)$ and let $Y = X^2$. Show that $Y \sim \text{Gamma}(1/2, 1/2)$ (or Chi-square(1)).

2a. Suppose $U \sim \text{Uniform}(0, 1)$ and let $X = -\log(1 - U)/\lambda$.

a) Show that $X \sim \text{Exp}(\lambda)$ (and note this is also true of $X' = -\log(U)/\lambda$).

b) Show that this is the *inverse CDF method* of generating an Exponential random variable, and explain how and why this works in general.

c) If X and Y are independent N$(0, 1)$ random variables, find the distribution of the squared distance from the origin to the point (x, y).

d) Outline an algorithm for generating two independent N$(0, 1)$ r.v.'s, given two independent $U(0, 1)$ r.v.s.

2b. Let X be a random variable with MGF $M_x(t)$.

a) Find the MGF for $Y = a + bX$.

b) Find the MGF for $Z \sim N(0,1)$ and for $Y = \mu + \sigma Z \sim N(\mu, \sigma^2)$.

c) Argue, using the uniqueness of the MGF, that the sum of two independent Normal r.v.'s is Normal.

d) For $Y \sim N(\mu, \sigma^2)$, $T = \exp(Y)$ is said to have a *log-Normal* distribution. Use your result from b to find the mean and variance of T. What is the median?

3a. Suppose $U_1, U_2, \ldots, U_n$ are independent Uniform$(0,1)$ random variables. Let $U_{(n)} = \max(U_i)$ be the largest of the U_i's (the nth *order statistic*).

a) Derive the density function for $U_{(n)}$ and find it's mean.

b) If $X \sim$ Beta(a, b) then

$$f_x(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, \quad 0 < x < 1.$$

Show that the distribution of the kth order statistic $U_{(k)}$ follows a Beta distribution (hint: use a differential argument, and recall that $\Gamma(k+1) = k!$ if k is a positive integer).

c) Find the mean and variance of the Beta distribution.

3b. Suppose $X \sim$Beta(a, b). Then $Y = \left(\frac{b}{a}\right) \frac{X}{1-X}$ follows an $F(2a, 2b)$ distribution.

a) Find the density function for Y. Summarize the integrals we can do now that we know the Beta and F density functions.

b) Find the mean of Y.

c) Suppose $a = b$. What is the *median* of Y?

d) If $V_1 \sim$ Gamma(a, λ) independent of $V_2 \sim$ Gamma(b, λ), then $Y = \frac{V_1/a}{V_2/b} \sim F(2a, 2b)$. What does this imply about the relationship between the Beta random variable X and V_1 and V_2?

Problems to turn in:

1. Suppose $Y|\,\theta \sim$ Poisson(θ) and that $\theta \sim$ Gamma(α, λ). For example, θ might represent the mean number of goals scored in a match by a particular team, with the Gamma distribution describing the variability of different teams' scoring potentials. Then Y is the actual number of goals scored in a particular match for a randomly chosen team.

a) Find the marginal distribution of Y. That is, if you pick a team at random, what is the distribution of the number of goals scored in their next game? Compare this result to that for problem 6 on the week 1 problem set.

b) If you observe $Y = y$, what is the conditional distribution of θ?

2. In R, generate a vector of 1000 random uniform draws by typing **u = runif(1000)** and return. If you now type "u" and return, you will see the 1000 simulated values.

a) In R, type **hist(u); summary(u)**

You will get a histogram of the values in the vector u, as well as numerical summaries (typing ";" ends a line and begins a new command just like return). Export the histogram as a pdf file by typing **pdf("filename.pdf"); hist(u); dev.off()** and return.

You can now open and print the pdf file called filename.pdf (choose your own descriptive name) using a pdf browser such as acrobat reader, or import it into a word processing program such as Word. You can adjust the size of the output by including commands such as "height=6, width=10" to specify a 10 inch by 6 inch plot.

b) Now perform a transformation on x: type **x = sqrt(u)** and return.

Each element of the vector x is now the square root of the corresponding element in u. Print a histogram of x and contrast it to that of u. Give an intuitive explanation for the difference.

c) What proportion of the values in x are less than 0.5? Type **mean(x < 0.5)**

The comparison $x < 0.5$ is either true (1) or false (0) for each element, and the mean is the proportion of times the statement is true. Find the theoretical probability of $X < 0.5$ by rephrasing the problem as a statement about U, the $U(0,1)$ random variable.

d) Derive the exact CDF and density function for $X = \sqrt{U}$, for $U \sim U(0,1)$.

e) Suppose you generate $U_1, U_2 \overset{\text{i.i.d.}}{\sim} U(0,1)$ and let $Y = \max(U_1, U_2)$. Show that Y has the same distribution as $X = \sqrt{U_1}$. To demonstrate this, generate a 1000×2 matrix of uniforms, and take the largest element in each row:

```
U = matrix(runif(2000), ncol=2, nrow=1000); y = apply(U, 1, max)
```

The "apply" function applies a function ("max" in this case) to an array, preserving the designated dimension (if you entered "2" instead of "1", you would get a vector of length 2, each element being the maximum of 1000 uniforms). Compare the quantiles of your simulated x and y vectors by typing:

```
plot(sort(x), sort(y)); abline(0,1)
```

The command "abline(a,b)" draws a straight line with intercept a and slope b on to an existing graph. Print this graph and explain how it shows evidence that the distributions of x and y are similar.