14.30 Introduction to Statistical Methods in Economics
Spring 2009

Problem Set #5 - Solutions

14.30 - Intro. to Statistical Methods in Economics

Instructor: Konrad Menzel

Due: Tuesday, March 31, 2009

Question One

The convolution formula is a useful trick when we are interested in the sum or average of independent random variables. In the last problem set, we dealt with the random variable X, below.

$$f(x) = \begin{cases} e^{-x} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}.$$

Now, suppose that $X = X_1 = X_2 = \cdots = X_k$ are independent, identically distributed random variables.

1. Using the convolution formula, determine the PDF of $Y_2 = \frac{1}{2}(X_1 + X_2)$. *Hint: Define $Z_1 = X_1$ and $Z_2 = X_1 + X_2$ and then use the transformation method to get back to Y_2 from Z_2.*

    - Solution to 1: To use the convolution formula, we need the joint PDF of X_1 and X_2 and x_2 as a function of y_2 and x_1. The function is $x_2 = 2y_2 - x_1$. Also, since they are independent, we can just construct the joint PDF by multiplying the two marginals, $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$. This gives a joint PDF of $f(x_1, x_2) = e^{-(x_1+x_2)}$ for $x_1 > 0$ and $x_2 > 0$ and zero otherwise. The convolution formula adapted to this problem (taking into account the limits) is (where we take into account the change in variables formula for x_2 as a function of x_1 and y_2 using the Jacobian)

    $$\begin{aligned} f_{Y_2}(y_2) &= \int_0^{y_2} f(x_1, 2y_2 - x_1)\,|J|\,dx_1 \\ &= \int_0^{y_2} e^{-(x_1+2y_2-x_1)}\,|2|\,dx_1 \\ &= 2\int_0^{y_2} e^{-2y_2}dx_1 \\ &= 2\left(x_1 e^{-2y_2}\right)_{x_1=0}^{x_1=2y_2} \\ f_{Y_2}(y_2) &= 4y_2 e^{-2y_2}. \end{aligned}$$

2. Compute its expectation: $\mathbb{E}[Y_2]$.

- Solution to 2: To compute the expectation, we use the standard formula for continuous random variables:

$$\mathbb{E}[Y_2] = \int_{-\infty}^{\infty} y_2 f_{Y_2}(y_2) dy_2.$$

This integral must be computed using two iterations of integration by parts.

$$\begin{aligned}
\mathbb{E}[Y_2] &= \int_0^{\infty} y_2 \left(4y_2 e^{-2y_2}\right) dy_2 \\
&= \int_0^{\infty} (y_2)^2 \, 4e^{-2y_2} dy_2 \\
&= \left[(y_2)^2 \left(-2e^{-2y_2}\right)\right]_0^{\infty} - \int_0^{\infty} (2y_2)\left(-2e^{-2y_2}\right) dy_2 \\
&= \left[(y_2)^2 \left(-2e^{-2y_2}\right)\right]_0^{\infty} - \left[(2y_2)\left(-e^{-2y_2}\right)\right]_0^{\infty} + \int_0^{\infty} 2\left(-e^{-2y_2}\right) dy_2 \\
&= \left[(y_2)^2 \left(-2e^{-2y_2}\right)\right]_0^{\infty} - \left[(2y_2)\left(-e^{-2y_2}\right)\right]_0^{\infty} + \left[e^{-2y_2}\right]_0^{\infty} \\
&= (0-0) + (0-0) + (0+1) \\
\mathbb{E}[Y_2] &= 1
\end{aligned}$$

Alternatively, we could have used the properties of expectations (no convolution necessary) to obtain

$$\begin{aligned}
\mathbb{E}[Y_2] &= \mathbb{E}[\frac{1}{2}(X_1 + X_2)] \\
&= \frac{1}{2}\mathbb{E}[X_1 + X_2] \\
&= \frac{1}{2}\left(\mathbb{E}[X_1] + \mathbb{E}[X_2]\right) \\
&= \frac{1}{2}\left(2\mathbb{E}[X]\right) \\
&= \mathbb{E}[X] \\
&= \int_{-\infty}^{\infty} x f_X(x) dx \\
&= \int_0^{\infty} x e^{-x} dx \\
&= (xe^{-x})_0^{\infty} - \int_0^{\infty} -e^{-x} dx \\
&= 0 + (0+1) \\
\mathbb{E}[Y_2] &= 1
\end{aligned}$$

3. Using the convolution formula, determine the PDF for $Y_3 = \frac{1}{3}(X_1 + X_2 + X_3)$. *Hint: Use the hint from part 1 to define $Z_3 = X_1 + X_2 + X_3$ and perform a convolution with X_3 and Z_2 to transform the problem into Z_2 and Z_3.*

- Solution to 3: Apply the same method to the results from part 1, where we obtained $f_{Y_2}(y_2)$. We write the joint PDF of X_1, X_2, and X_3 as

$$\begin{aligned} f_{X_1,X_2,X_3}(x_1,x_2,x_3) &= f_{X_1}(x_1)f_{X_2}(x_2)f_{X_3}(x_3) \\ &= e^{-(x_1+x_2+x_3)} \end{aligned}$$

We'll take a slightly different approach for this problem, in order to help us solve part 5 later. Define $Z_2 = X_1+X_2$ and $Z_3 = X_1+X_2+X_3$. Now, we can ignore any Jacobian transformation and directly apply the convolution formula. We can first solve for Z_2's distribution by only using the convolution on the joint distribution of X_1 and X_2.

$$\begin{aligned} f_{Z_2}(z_2) &= \int_0^{z_2} f(x_1, z_2 - x_1)dx_1 \\ &= \int_0^{y_2} e^{-(x_1+z_2-x_1)}dx_1 \\ &= \int_0^{z_2} e^{-z_2}dx_1 \\ &= \left(x_1 e^{-z_2}\right)_{x_1=0}^{x_1=z_2} \\ f_{Z_2}(z_2) &= z_2 e^{-z_2}. \end{aligned}$$

Now, we use the convolution on the joint distribution of Z_2 and X_3 to obtain the distribution of Z_3:

$$\begin{aligned} f_{Z_3}(z_3) &= \int_{-\infty}^{\infty} f_{X_3,Z_2}(x_3, z_2)dz_2 \\ &= \int_0^{z_3} f_{X_3}(z_3 - z_2)f_{Z_2}(z_2)dz_2 \\ &= \int_0^{z_3} z_2 e^{-z_3}dz_2 \\ &= \int_0^{z_3} z_2 e^{-z_3}dz_2 \\ f_{Z_3}(z_3) &= \frac{1}{2}\left(z_3\right)^2 e^{-z_3} \end{aligned}$$

But, we were interested in Y_3, not Z_3. So, we perform a simple continuous inverse transformation of random variables using $Z_3 = 3Y_3$:

$$\begin{aligned} f_{Y_3}(y_3) &= f_{Z_3}(3y_3)|3| \\ &= \frac{3}{2}\left(3y_3\right)^2 e^{-3y_3} \\ f_{Y_3}(y_3) &= \frac{3}{2}\left(3y_3\right)^2 e^{-3y_3}. \end{aligned}$$

4. Compute its expectation: $\mathbb{E}[Y_3]$.

- Solution to 4: Very easily we can apply the properties of expectations to determine again that $\mathbb{E}[Y_3] = 1$. However, we can also determine this using the convolution formula. What we should have learned from part 2 was that all of the leading integration by parts terms will not matter as $y^k e^{-y} = 0$ for $y = \infty$ and for $y = 0$ for any $k \in \mathbb{N}$. So, we only need to worry about the last term in the integration by parts sequence, which ends up being e^{-3y_3} (verify it if you like).

$$\begin{aligned}
\mathbb{E}[Y_3] &= \int_0^\infty y_3 \left(\frac{3}{2}(3y_3)^2 e^{-3y_3}\right) dy_3 \\
&= \int_0^\infty y_3 \left(\frac{3}{2}(3y_3)^2 e^{-3y_3}\right) dy_3 \\
&= \left[e^{-3y_3}\right]_0^\infty \\
&= (0+1) \\
\mathbb{E}[Y_3] &= 1
\end{aligned}$$

5. Using the convolution formula, determine the PDF for $Y_k = \frac{1}{k}(X_1 + X_2 + \cdots + X_k) = \frac{1}{k}\sum_{i=1}^k X_k$. *Hint: Try to determine a pattern from part 1 and part 3 using the methods from their hints.*

   - Solution to 5: We use the convolution formula using the sequence determined from part 3. We define more partial sums $Z_1 = X_1$, $Z_2 = X_1 + X_2$, $Z_3 = X_1 + X_2 + X_3$, ... , $Z_k = \sum_{i=1}^k X_i$. This allows us to easily construct the inverse functions for X's as function of Z's as they are just the difference between adjacent Z's. This gives the following Jacobian:

$$J = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

which, conveniently has a determinant of 1. So, we can just ignore the Jacobian component. We now write the joint distribution:

$$\begin{aligned}
f(z_1, z_2 - z_1, \ldots, z_k - z_{k-1}) &= f(x_1(z_1, ..., z_k), \ldots, x_k(z_1, ..., z_k)) \\
&= e^{-z_k}.
\end{aligned}$$

Since the joint distribution of all x's is just a function of z_k, we would say that z_k is a sufficient statistic for the distribution of the sum of k independent x's. Now, all we need to do is integrate out each of z_1 through z_{k-1} in order to get the marginal of z_k. From part 3, we learned that the convolution will yield:

$$f_{Z_2}(z_2) = z_2 e^{-z_2}$$

and

$$f_{Z_3}(z_3) = \frac{1}{2}(z_3)^2 e^{-z_3}$$

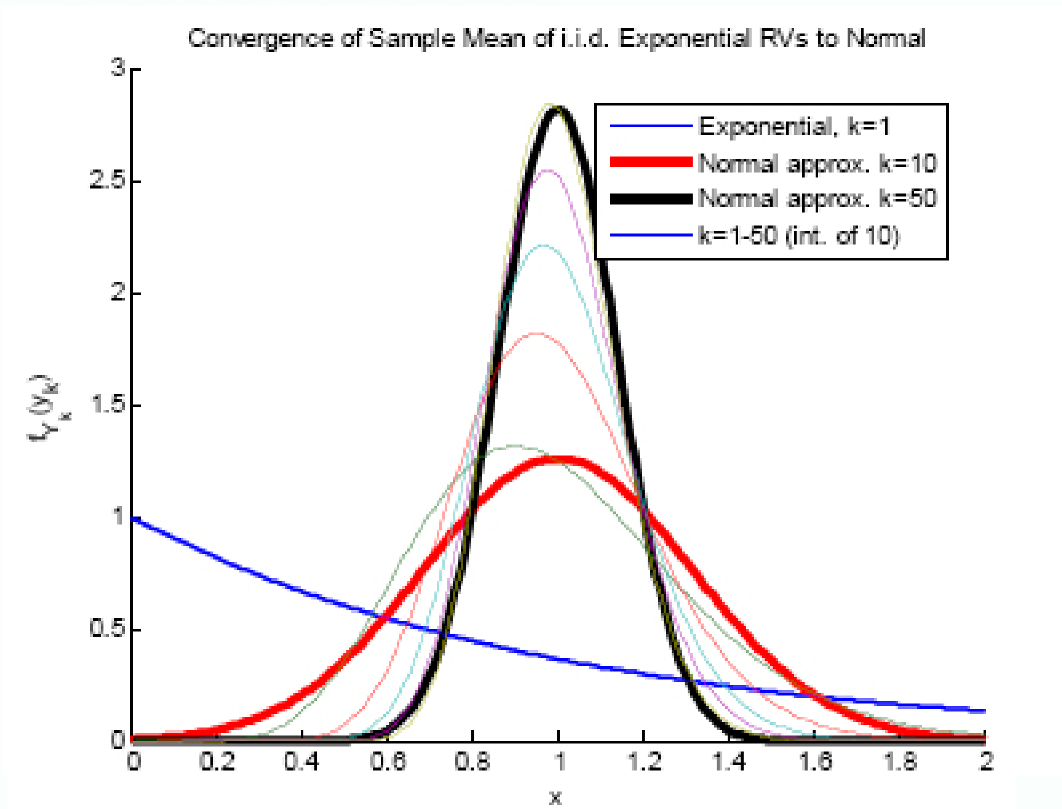
which, if repeated, will yield the general formula for Z_k:

$$f_{Z_k}(z_k) \quad = \quad \left(\prod_{i=1}^{k-1} \frac{1}{i}\right) (z_k)^{k-1} \, e^{-z_k}$$

which, upon applying the inverse transform method for $Z_k = kY_k$:

$$f_{Y_k}(y_k) \quad = \quad \left(\prod_{i=1}^{k-1} \frac{1}{i}\right) (y_k)^{k-1} \, k^k e^{-ky_k}$$

which is what we were looking for. Interestingly enough, this converges to a Normal or Gaussian distribution as we send k to infinity. This is just a special case of a more general theorem: the Central Limit Theorem. Below we see a plot of the convergence of the mean of k random draws from the exponential distribution to the Normal distribution:



6. Compute its expectation: $\mathbb{E}[Y_k]$.

  - Solution to 1: See part 4. The expectation is 1. :)

7. What does this tell us about the mean of a sample of size k? Is this property specific to the exponential distribution? Explain.

  - Solution to 7: The mean of an i.i.d. (independent and identically distributed) sample of size k is the same as the expectation of a single draw. What this means

is that the average of a sample of size k is an unbiased estimate of the average or mean of a distribution.
This property is not specific to the exponential distribution. You may have not noticed this through all of the integration, but if you computed the expectation using the properties of the expecation operator, you would have recognized that you didn't use any properties of the exponential distribution to discover this.

Question Two

(Bain/Engelhardt, p. 228)

Suppose that $X_1, X_2, \ldots, X_k$ are independent random variables and let $Y_i = u_i(X_i)$ for $i = 1, 2, \ldots, k$. Show that $Y_1, Y_2, \ldots, Y_k$ are independent. Consider only the case where X_i is continuous and $y_i = u_i(x_i)$ is one-to-one. *Hint:* If $x_i = w_i(y_i)$ is the inverse transformation, then the Jacobian has the form

$$J = \prod_{i=1}^{k} \frac{d}{dy_i} w_i(y_i).$$

For extra credit, prove the *Hint* about the Jacobian.

- Solution: We can write $X_i = w_i(Y_1, Y_2, ..., Y_k)$ where $w_i(\cdot)$ is the inverse function of $u_i(\cdot)$, i.e. $w_i(\cdot) = u_i^{-1}(\cdot)$. Since u_i is only a function of X_i, the inverse transform will only be a function of Y_i, or $X_i = w_i(Y_i)$. With this intuition, we can show the independence by showing that pdfs can be factored:

$$f_{X_1,...,X_k}(x_1, ..., x_k) = f_{X_1}(x_1) \cdots f_{X_k}(x_k).$$

We now apply the transformation method:

$$f_{Y_1,...,Y_k}(y_1, ..., y_k) = f_{X_1,...,X_k}(x_1, ..., x_k)\,|J|$$

where

$$\begin{aligned} J &= \begin{vmatrix} \frac{d}{dy_1} w_1(y_1) & \cdots & \frac{d}{dy_k} w_1(y_1) \\ \vdots & \ddots & \vdots \\ \frac{d}{dy_1} w_k(y_k) & \cdots & \frac{d}{dy_k} w_k(y_k) \end{vmatrix} \\ J &= \begin{vmatrix} \frac{d}{dy_1} w_1(y_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{d}{dy_k} w_k(y_k) \end{vmatrix} \end{aligned}$$

since each w_i is only a function of y_i and not y_{-i}. So, all of the off diagonal derivatives are zero. Thus, the Jacobian can be easily computed since we have a diagonal matrix: $J = \prod_{i=1}^{k} \frac{d}{dy_i} w_i(y_i)$.

With this fact, we can now substitute in all of the pieces:

$$\begin{aligned} f_{Y_1,...,Y_k}(y_1,...,y_k) &= f_{X_1}(w_1(x_1))\cdots f_{X_k}(w_k(y_k))\left|\prod_{i=1}^{k}\frac{d}{dy_i}w_i(y_i)\right| \\ &= f_{X_1}(w_1(x_1))\cdots f_{X_k}(w_k(y_k))\prod_{i=1}^{k}\left|\frac{d}{dy_i}w_i(y_i)\right| \\ &= \left(f_{X_1}(w_1(y_1))\left|\frac{d}{dy_1}w_1(y_1)\right|\right)\cdots\left(f_{X_k}(w_k(y_k))\left|\frac{d}{dy_k}w_k(y_k)\right|\right) \\ &= f_{Y_1}(y_1)\cdots f_{Y_k}(y_k) \end{aligned}$$

where we use the independence of the X's and the *Hint* about the Jacobian in the first line, the absolute value of a product is equal to the product of the absolute values in the second line, commutative and associative properties of multiplication in the third line, and the definition of the transformation of a single random variable in the last line. So, $Y_1, ..., Y_k$ are independent and we are done.

Question Three

Moved to a later problem set.

Question Four

Order statistics are very useful tools for analyzing the properties of samples.

1. Write down the general formula of the pdf and cdf for the k^{th} order statistic of a sample of size n of a random variable X with CDF $F_X(x)$.

   - Solution to 1: The general formula for the k^{th} order statistic, Y_k, with pdf $f(x)$ ($f(x) > 0$ on $a < x < b$ and zero otherwise) and cdf $F(x)$ is:

     $$g_k(y_k) = \frac{n!}{(k-1)!(n-k)!}\left[F(y_k)\right]^{k-1}\left[1 - F(y_k)\right]^{n-k} f(y_k)$$

     if $a < y_k < b$ and zero otherwise. The marginal cdf can be written as

     $$G_k(y_k) = \sum_{j=k}^{n}\binom{n}{j}\left[F(y_k)\right]^{j}\left[1 - F(y_k)\right]^{n-j}.$$

Question Five

(Bain/Engelhardt p. 229)

Let X_1 and X_2 be a random sample of size $n = 2$ from a continuous distribution with pdf of the form $f(x) = 2x$ if $0 < x < 1$ and zero otherwise.

1. Find the marginal pdfs of the smallest and largest order statistics, Y_1 and Y_2.

   - Solution to 1: The marginal pdfs of the smallest and largest order statistics from a sample of size $n = 2$ can be derived by using the marginal pdf formula, but we first need the CDF of X:

   $$F(x) = \int_0^x 2x' dx' = x^2$$

   Now, we use the marginal pdf formula:

   $$\begin{aligned} g_1(y_1) &= \frac{2!}{(1-1)!(2-1)!} \left[F(y_1)\right]^{1-1} \left[1 - F(y_1)\right]^{2-1} f(y_1) \\ &= 2\left[1 - (y_1)^2\right] 2y_1 \\ g_1(y_1) &= 4y_1 \left[1 - (y_1)^2\right] \end{aligned}$$

   and

   $$\begin{aligned} g_2(y_2) &= \frac{2!}{(2-1)!(2-2)!} \left[F(y_2)\right]^{2-1} \left[1 - F(y_2)\right]^{2-2} f(y_2) \\ &= 2\left[(y_2)^2\right] 2y_2 \\ g_2(y_2) &= 4(y_2)^3 \end{aligned}$$

2. Compute their expectations, $\mathbb{E}[Y_1]$ and $\mathbb{E}[Y_2]$.

   - Solution to 2: We apply the formula for the expectation of a random variable:

   $$\begin{aligned} \mathbb{E}[Y_1] &= \int_0^1 y_1 g_1(y_1) dy_1 \\ &= \int_0^1 y_1 4y_1 \left[1 - (y_1)^2\right] dy_1 \\ &= \int_0^1 \left[4(y_1)^2 - 4(y_1)^4\right] dy_1 \\ &= \left[\frac{4}{3}(y_1)^3 - \frac{4}{5}(y_1)^5\right]_0^1 \\ &= 4\left(\frac{1}{3} - \frac{1}{5}\right) \\ \mathbb{E}[Y_1] &= \frac{8}{15}. \end{aligned}$$

   So, the expectation of the first order statistic is $\frac{8}{15}$. For the second (maximal) order statistic, we get

$$\begin{aligned}
\mathbb{E}[Y_2] &= \int_0^1 y_2 g_2(y_2) dy_2 \\
&= \int_0^1 y_2 4(y_2)^3 dy_2 \\
&= \int_0^1 \left[4(y_2)^4\right] dy_2 \\
&= \left[\frac{4}{5}(y_2)^5\right]_0^1 \\
&= 4\left(\frac{1}{5}\right) \\
\mathbb{E}[Y_2] &= \frac{4}{5}.
\end{aligned}$$

Fortunately, we find the expectation of the maximal order statistic to be larger than the minimal (or first) order statistic.

3. Find the joint pdf of Y_1 and Y_2.

   - Solution to 3: The joint pdf of Y_1 and Y_2 can be obtained from the pdf $f(x)$ (the joint pdf can be written as $g(y_1, ..., y_n) = n! f(y_1) \cdots f(y_n)$ for n independently sampled observations from the distribution defined by $f(x)$) taking into account the permutations of y_1 and y_2:

$$\begin{aligned}
f_{Y_1,Y_2}(y_1, y_2) &= 2! f(y_1) f(y_2) \\
&= 2! \,(2y_1)\,(2y_2) \\
f_{Y_1,Y_2}(y_1, y_2) &= 8y_1 y_2
\end{aligned}$$

4. Find the pdf of the sample range $R = Y_2 - Y_1$.

   - Solution to 4: The pdf of the sample range is just a transformation of the joint PDF, similar to all of the convolutions that we've been doing. Consider the inverse transforms: $Y_1 = Y_1$ and $Y_2 = R + Y_1$. Then we can use the transform methods:

$$f_{Y_1,R}(y_1, r) = f_{Y_1,Y_2}(y_1, r + y_1)\,|J|$$

where $J = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1$. So, applying the formula for the joint pdf from part 3, we get:

$$f_{Y_1,R}(y_1, r) = 8y_1(r + y_1)\,|1|$$

which we now use to integrate out y_1 to get the marginal of R. We proceed, paying careful attention to the change in bounds (draw a picture to convince yourself–first with Y_1 and Y_2 and then with R and Y_1):

$$\begin{aligned} f_R(r) &= \int_0^{1-r} 8y_1(r+y_1)dy_1 \\ &= \left[8(\frac{1}{2}y_1^2 r + \frac{1}{3}y_1^3)\right]_0^{1-r} \\ &= 8\left[\frac{1}{2}(1-r)^2 r + \frac{1}{3}(1-r)^3\right] \\ f_R(r) &= \frac{4}{3}(1-r)^2(2+r) \end{aligned}$$

5. Compute the expectation of the sample range, $\mathbb{E}[R]$.

- Solution to 5: Compute it:

$$\begin{aligned} \mathbb{E}[R] &= \int_0^1 r\frac{4}{3}(1-r)^2(2+r)\,dr \\ &= \frac{4}{3}\int_0^1 \left(1-2r+r^2\right)\left(2r+r^2\right)dr \\ &= \frac{4}{3}\int_0^1 \left(2r-4r^2+2r^3+r^2-2r^3+r^4\right)dr \\ &= \frac{4}{3}\int_0^1 \left(2r-3r^2+r^4\right)dr \\ &= \frac{4}{3}\left(r^2-r^3+\frac{1}{5}r^5\right)_0^1 \\ &= \frac{4}{3}\left(\frac{1}{5}\right) \\ \mathbb{E}[R] &= \frac{4}{15}. \end{aligned}$$

It is worth noting that the expectation for the sample range is equal to the difference between the expected first and last order statistics (minimal and maximal order statistics).

Question Six

(Bain/Engelhardt p. 229)

Consider a random sample of size n from a distribution with pdf $f(x) = \frac{1}{x^2}$ if $1 \leq x < \infty$; zero otherwise.

1. Give the joint pdf of the order statistics.

Solution to 1: The joint pdf can be represented as a relabeling of the original joint pdf of X_1 through X_n, the random variables representing the sample.

In particular, we can write:

$$
\begin{aligned}
g(y_1, ..., y_n) &= n! f(y_1) \cdots f(y_n) \\
&= n! \frac{1}{y_1^2} \cdots \frac{1}{y_n^2}
\end{aligned}
$$

2. Give the pdf of the smallest order statistic, Y_1.

   - Solution to 2: The pdf of the smallest order statistic is the PDF when we integrate out all of the other order statistics. We already have a formula for this, so we'll use it, rather than deriving it all over again. We use $k = 1$ and n:

$$
\begin{aligned}
g_k(y_k) &= \frac{n!}{(k-1)!(n-k)!} [F(y_k)]^{k-1} [1 - F(y_k)]^{n-k} f(y_k) \\
g_1(y_1) &= n [1 - F(y_1)]^{n-1} f(y_1) \\
&= n \left[1 - \int_1^{y_1} f(\tilde{y}_1) d\tilde{y}_1 \right]^{n-1} f(y_1) \\
&= n \left[1 - \left(-\frac{1}{y_1} + 1 \right) \right]^{n-1} \frac{1}{y_1^2} \\
&= n \left[\frac{1}{y_1} \right]^{n-1} \frac{1}{y_1^2} \\
g_1(y_1) &= \frac{n}{y_1^{n+1}}
\end{aligned}
$$

3. Compute its expectation, $\mathbb{E}[Y_1]$, or explain why it does not exist.

   - Solution to 3: Again, we have another integration to practice the formula for the expectation of a random variable using our answer from part 2:

$$
\begin{aligned}
\mathbb{E}[Y_1] &= \int_1^{\infty} y_1 \frac{n}{y_1^{n+1}} dy_1 \\
&= n \int_1^{\infty} \frac{1}{y_1^n} dy_1 \\
&= \frac{n}{n-1} \left[-\frac{1}{y_1^{n-1}} \right]_1^{\infty} \\
&= \frac{n}{n-1} [0 + 1] \\
\mathbb{E}[Y_1] &= 1 + \frac{1}{n-1}.
\end{aligned}
$$

4. Give the pdf of the largest order statistic, Y_n.

- Solution for 4: The pdf for the largest order statistic is computed from the same formula as for the smallest order statistic:

$$
\begin{aligned}
g_k(y_k) &= \frac{n!}{(k-1)!(n-k)!}\left[F(y_k)\right]^{k-1}\left[1-F(y_k)\right]^{n-k} f(y_k) \\
g_n(y_n) &= n\left[F(y_n)\right]^{n-1} f(y_n) \\
&= n\left[\int_1^{y_n} f(\tilde{y}_n)d\tilde{y}_n\right]^{n-1} f(y_n) \\
&= n\left[\left(-\frac{1}{y_n}+1\right)\right]^{n-1}\frac{1}{y_n^2} \\
g_n(y_n) &= n\left[1-\frac{1}{y_n}\right]^{n-1}\frac{1}{y_n^2}.
\end{aligned}
$$

5. Compute the expectation, $\mathbb{E}[X]$, of a single draw X from $f(x)$. Does the integral diverge? What does that say about the existence of $\mathbb{E}[Y_n]$? Explain.

   - Solution to 5:

$$
\begin{aligned}
\mathbb{E}[X] &= \int_1^\infty x\frac{1}{x^2}dx \\
&= \int_1^\infty \frac{1}{x}dx \\
&= [\log x]_1^\infty \\
&= \infty - 0 \\
\mathbb{E}[X] &= \infty,
\end{aligned}
$$

   or, in other words, the integral diverges. This suggests that at least one order statistic cannot have an expectation, even though the smallest order statistic did since the expectation of X could be constructed from the order statistics as well by computing the expectation of each and then average over all n order statistics. By a bounding argument, we can say that since the smallest order statistic has an expectation, and the total expectation does not exist for X, then the largest order statistic must not have a finite expectation. While it would be possible for multiple order statistics to have nonexistent expectation, I do not determine which order statistic is the dealbreaker. Feel free to tinker with the formulas to figure it out for yourself.

6. Derive the pdf of the sample range, $R = Y_n - Y_1$, for $n = 2$. *Hint: Use partial fractions, searching on Yahoo for "QuickMath" will help you get the partial fractions via computer:*

http://www.quickmath.com/webMathematica3/quickmath/page.jsp?s1=algebra&s2=partialfractions&s3=basic

   - Solution to 6: We again define two random variables. Call them $S = Y_1$ and $R = Y_n - Y_1$ and the inverse transforms of $S = Y_1$ and $Y_2 = R + S$. The joint pdf from part 1 can now be used:

$$\begin{aligned} g(y_1, ..., y_n) &= n! \frac{1}{y_1^2} \cdots \frac{1}{y_n^2} \\ g(y_1, y_2) &= \frac{2}{y_1^2 y_2^2} \end{aligned}$$

We use the transformation formula, noting that the Jacobian has a determinant of 1: $J = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1$.

$$\begin{aligned} f_{R,S}(r,s) &= g(s, r+s)\,|1| \\ &= \frac{2}{s^2(r+s)^2} \\ f_R(r) &= \int_1^\infty \frac{2}{s^2(r+s)^2} ds \end{aligned}$$

which we can solve very painfully via partial fractions, unless you use a computer.*

This is the partial fraction decomposition:

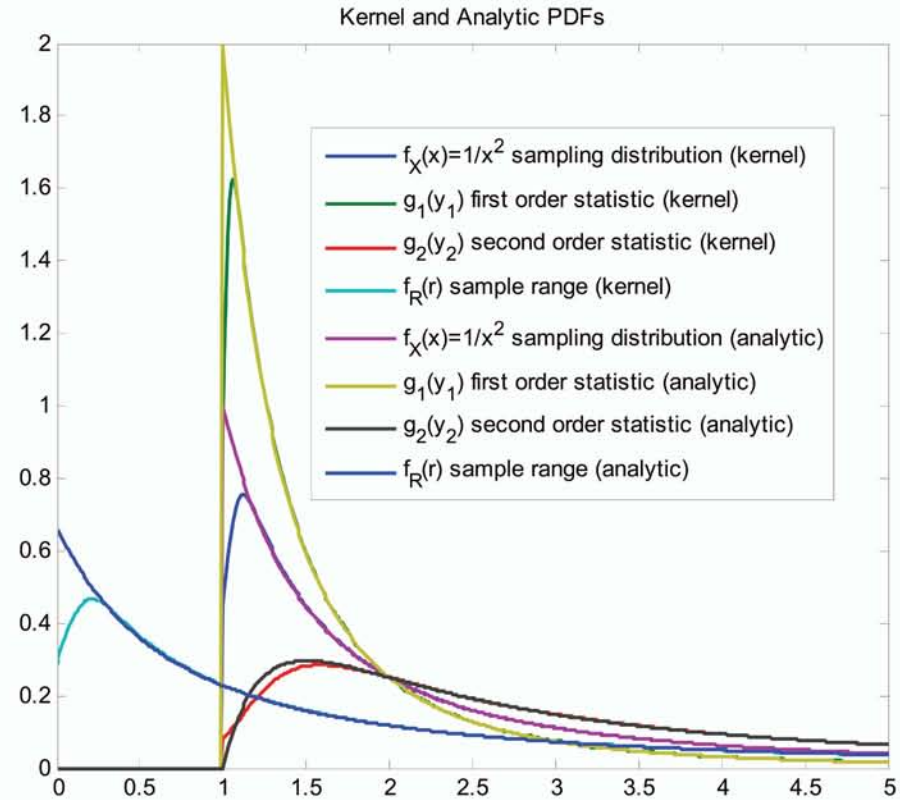
$$\frac{2}{s^2(r+s)^2} = \frac{4}{r^3(r+s)} + \frac{2}{r^2(r+s)^2} - \frac{4}{r^3 s} + \frac{2}{r^2 s^2}$$

which we can now integrate. The solution we obtain for the pdf is

$$\begin{aligned} f_R(r) &= \int_1^\infty \left[\frac{4}{r^3(r+s)} + \frac{2}{r^2(r+s)^2} - \frac{4}{r^3 s} + \frac{2}{r^2 s^2}\right] ds \\ &= \left[\frac{4}{r^3}\log(\frac{r+s}{s}) - \frac{2}{r^2(r+s)} - \frac{2}{r^2 s}\right]_1^\infty \\ &= \left[\frac{4}{r^3}\log(1+\frac{r}{s}) - \frac{2}{r^2(r+s)} - \frac{2}{r^2 s}\right]_1^\infty \\ &= \left[0 - 0 - 0 - \frac{4}{r^3}\log(1+\frac{r}{s}) + \frac{2}{r^2(r+1)} + \frac{2}{r^2}\right] \\ f_R(r) &= \frac{2}{r^2}\left[1 + \frac{1}{1+r} - \frac{2}{r} log(1+r)\right] \end{aligned}$$

where $r > 0$. The PDF of the sample range and the first and second order statistics in a sample of size $n = 2$ are plotted below, and compared to the kernel estimates from 1 million samples of size $n = 2$ to benchmark the theory.

*A web site I found to do this is: http://www.quickmath.com/webMathematica3/quickmath/page.jsp?s1=algebra&s2=partialfractions&s3=basic



7. Compute its expectation, $\mathbb{E}[R]$, or explain why it does not exist.

   - Solution to 7: The expectation of the range cannot exist as the expectation of the maximal order statistic does not exist, and $\mathbb{E}[R] = \mathbb{E}[Y_2] - \mathbb{E}[Y_1]$. Even though $\mathbb{E}[Y_1]$ exists, we need both to exist to compute it.

8. Give the pdf of the sample median, Y_r, assuming that n is odd so that $r = (n+1)/2 \in \mathbb{N}$. Express the pdf as a function of just r and y_r (eliminate all n's and k's).

   - Solution to 8: The sample median is the order statistic with $r = k = (n+1)/2$. We can compute this in the same way as before, using the formula:

$$\begin{aligned} g_r(y_r) &= \frac{n!}{(r-1)!(n-r)!}\left[F(y_r)\right]^{r-1}\left[1-F(y_r)\right]^{n-r} f(y_r) \\ &= \frac{(2r-1)!}{(r-1)!(r-1)!}\left[1-\frac{1}{y_r}\right]^{r-1}\left[\frac{1}{y_r}\right]^{r-1}\frac{1}{y_r^2} \\ g_r(y_r) &= \frac{(2r-1)!}{\left[(r-1)!\right]^2}\frac{(y_r-1)^{r-1}}{y_r^{2r}} \end{aligned}$$

9. Compute its expectation, $\mathbb{E}[Y_r]$, or explain why it does not exist.

- Solution to 9: We will attempt to compute the expectation:

$$\begin{aligned}
\mathbb{E}[Y_r] &= \int_1^\infty y_r g_r(y_r) dy_r \\
&= \int_1^\infty y_r \frac{(2r-1)!}{[(r-1)!]^2} \frac{(y_r-1)^{r-1}}{y_r^{2r}} dy_r \\
&= \frac{(2r-1)!}{[(r-1)!]^2} \int_1^\infty \frac{(y_r-1)^{r-1}}{y_r^{2r-1}} dy_r \\
&= \frac{(2r-1)!}{[(r-1)!]^2} \left[\left(-\frac{1}{r} \frac{(y_r-1)^r}{y_r^{2r-1}} \right)_1^\infty - \int_1^\infty -\frac{2r-1}{r} \frac{(y_r-1)^r}{y_r^{2r}} dy_r \right] \\
&= \left[\frac{(2r-1)!}{[(r-1)!]^2} \left(-\frac{1}{r} \frac{(y_r-1)^r}{y_r^{2r}} \right)_1^\infty + \frac{2r-1}{r} \int_1^\infty (y_r-1) \left(\frac{(2r-1)!}{[(r-1)!]^2} \frac{(y_r-1)^{r-1}}{y_r^{2r}} \right) dy_r \right] \\
&= \left[0 + \frac{2r-1}{r} \mathbb{E}\left[Y_r - 1\right] \right] \\
\mathbb{E}[Y_r] &= \frac{2r-1}{r} \left[\mathbb{E}\left[Y_r\right] - 1 \right]
\end{aligned}$$

which we now solve for $\mathbb{E}[Y_r]$:

$$\mathbb{E}[Y_r] = \frac{2r-1}{r-1}.$$

A few numerical checks will yield that this is, in fact, the correct formula: for $r = 2$, we get a median of 3 and for $r = 3$, we get a median of 2.5. For $r = 5$, the median is 2.25.