Worksheet 06 - Expected Value and Random Variables

Directions*: Please upload a PDF to Gradescope that includes both your written responses and corresponding R code inputs/outputs (if requested) for each problem.*

Special Directions Be sure to demonstrate the correct use of improper integral notation (and mathematical notation in general) in your work.

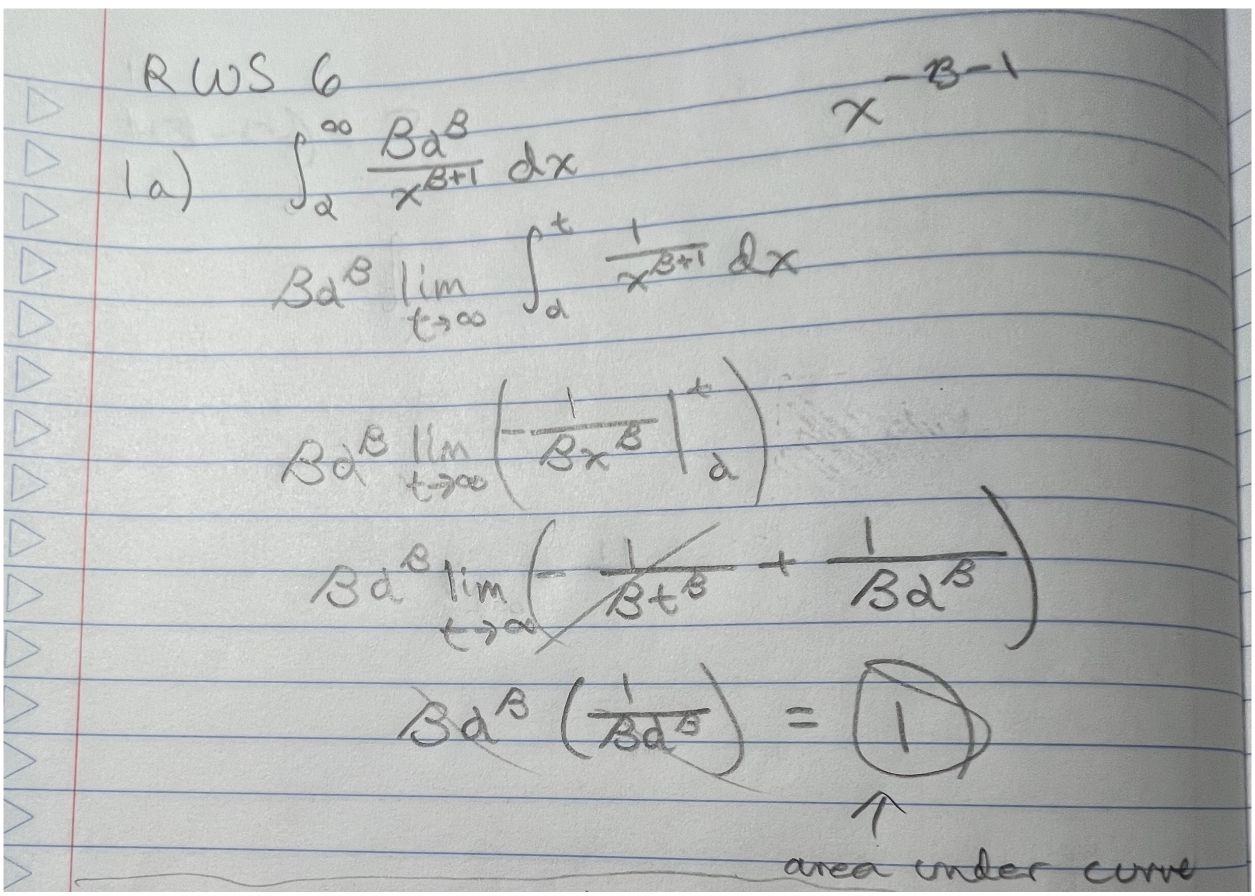
Probem 1. The **Pareto random variable** with parameters $\alpha > 0$ and $\beta > 0$ has the probability density function

$$f_X(x) = \frac{\beta\alpha^\beta}{x^{\beta+1}}, \text{ where } \alpha < x < \infty.$$

Problem 1 Part a) Verify that $f_X(x)$ is a density function.

Hint: Recall that the area under the density function over its entire domain must be one. If you are rusty on improper integral notation, see an example in part (c) first.

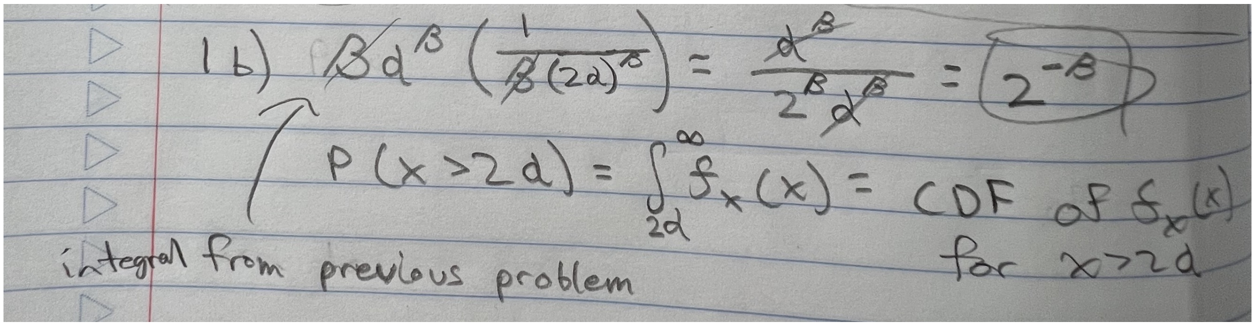
NOTE*: Upload an image of your work for part (a) by replacing "upload_image.jpg" with your appropriately titled .jpg file in the R chunk below.*



Problem 1 Part b) Determine $P(X > 2\alpha)$.

Hint: Recall that $P(X > 2\alpha)$ would be the area under the density function over the portion of its domain that is greater than 2α. If you are successful, you will see that $P(X > 2\alpha)$ is simply a power of 2.

NOTE: Upload an image of your work for part (b) by replacing "upload_image.jpg" with your appropriately titled .jpg file in the R chunk below.



Problem 1 Part c) Determine the variance of X, given $E[X] = \frac{\alpha\beta}{\beta-1}$. What restriction do we require on β for convergence in the computation of the variance?

Hint: The calculation of $E[X]$, the mean of X, is given below. You do not need to repeat it in your work. Notice that we require $\beta > 1$ for convergence.

Hint: There is a different restriction on β for the variance. Additionally, you will need to determine $E[X^2]$ before you can determine $\text{Var}(X)$. You will not be required to simplify your final answer for $\text{Var}(X)$, but if you do then you will see that $\text{Var}(X) = \dfrac{\alpha^2\beta}{(\beta-1)^2(\beta-2)}$

Calculation of E[X]:

$$\begin{aligned}
E[X] &= \int_\alpha^\infty x \cdot \frac{\beta\alpha^\beta}{x^{\beta+1}} dx \\
&= \beta\alpha^\beta \lim_{t\to\infty} \left[\int_\alpha^t x^{-\beta} dx\right] \\
&= \beta\alpha^\beta \lim_{t\to\infty} \left[\frac{x^{1-\beta}}{1-\beta}\right]_\alpha^t \\
&= \frac{\alpha^\beta\beta}{1-\beta} \lim_{t\to\infty} \left[t^{1-\beta} - \alpha^{1-\beta}\right] \text{ so we need } 1-\beta < 0 \Rightarrow \beta > 1 \text{ for convergence.} \\
&= \frac{\alpha^\beta\beta}{1-\beta} \left[0 - \alpha^{1-\beta}\right] \\
&= \frac{-\alpha^\beta\beta}{1-\beta} \cdot \alpha^{1-\beta} \\
&= \frac{-\alpha\beta}{1-\beta} \\
&= \frac{\alpha\beta}{\beta-1}
\end{aligned}$$

NOTE: Upload an image of your work by replacing "upload_image.jpg" with your appropriately titled .jpg file in the R chunk below.

1c) $\beta > 2$ is required for convergence in calculating the variance

$$Var(x) = E(x^2) - E(x)^2$$

$$E(x^2) = \int_\alpha^\infty x^2 \frac{\beta\alpha^\beta}{x^{\beta+1}}\,dx$$

$$\beta\alpha^\beta \lim_{t\to\infty} \int_\alpha^t x^2 x^{-\beta-1}\,dx$$

$$x^{-\beta+2}$$

$$\beta\alpha^\beta \lim_{t\to\infty} \int_\alpha^t x^{-\beta+1}\,dx$$

$$\beta\alpha^\beta \lim_{t\to\infty} \left(\frac{1}{(2-\beta)x^{\beta-2}} \Bigg|_\alpha^t \right)$$

$$-\beta\alpha^\beta \frac{1}{(2-\beta)\alpha^{\beta-2}}$$

$$= \frac{\beta\alpha^\beta}{(\beta-2)\alpha^\beta\alpha^{-2}}$$

$$Var(x) = \frac{\beta\alpha^2}{(\beta-2)} - \left(\frac{\alpha\beta}{\beta-1}\right)^2$$

$$\frac{\beta\alpha^2}{(\beta-2)} - \frac{\alpha^2\beta^2}{(\beta-1)^2} = \frac{\alpha^2\beta}{(\beta-1)^2(\beta-2)}$$

Problem 1 Part d) Find the probability transform for the Pareto distribution with $\alpha = 1$ and $\beta = 4$.

Hint: Recall that the probability transform is the inverse of the distribution function of Pareto random variables. You will first need to find the distribution function from the given density function where $\alpha = 1$ and $\beta = 4$. You should find that the probability transform is $x = \frac{1}{(1-u)^{1/4}}$.

NOTE: Upload an image of your work by replacing "upload_image.jpg" with your appropriately titled .jpg file in the R chunk below.

1 a) $f_x(x) = \frac{\beta \alpha^\beta}{x^{\beta+1}}$ $\alpha = 1$ $\beta = 4$

$$CDF = \int_1^u \frac{\beta \alpha^\beta}{x^{\beta+1}} \, dx$$

$$\int_1^u \frac{4}{x^5} \, dx$$

$$\left. \frac{4}{-4x^4} \right|_1^u$$

$$-\frac{1}{u^4} + 1$$

$$1 - \frac{1}{u^4} = x$$

$$1 - x = \frac{1}{u^4}$$

$$u^4 = \frac{1}{1-x}$$

$$x = \frac{1}{(1-u)^{1/4}}$$

Problem 1 Part e) Using R and the probability transform in part (d), simulate 1000 Pareto random variables with $\alpha = 1$ and $\beta = 4$. Find the sample mean and variance of this simulation. Next, compute the population mean and variance, based on earlier work in part (c), where $\alpha = 1$ and $\beta = 4$. Finally, comment on the comparison of the population mean and variance with the sample mean and variance.

Hint: Recall, when describing how close simulated values are to the true population value, be sure to comment using specific values and consider how many standard deviations the simulated mean is from the population mean.

```
set.seed(2022)  # for repeatability

# NOTE: simulate 1000 uniformly distributed random variables on [0, 1]
u = runif(1000, min = 0, max = 1)

# NOTE: use the probability transform to find the quantiles NOTE: these
# quantiles are the simulated Pareto random variables
pareto_sim = 1/(1 - u)^(1/4)

# NOTE: compute the mean and variance of the simulated values NOTE: we ask R to
# print the output of these lines by wrapping them in parenthesis


(mean(pareto_sim))
```

```
## [1] 1.335709
```

```
(var(pareto_sim))
```

```
## [1] 0.2737066
```

```
# mean_pop = E(X)
(mean_pop = 4/3)
```

```
## [1] 1.333333
```

```
# var_pop is given previous part, just plug in values
(var_pop = 4/(9 * 2))
```

```
## [1] 0.2222222
```

```
# sd_pop
(sd_pop = var_pop^0.5)
```

```
## [1] 0.4714045
```

Space for response: Both the sample mean and variance are very close to the population mean and variance which makes sense since we are using a population of 1000 and bigger population improves the resolution and accuracy of the data. The sample mean is almost identical to the population mean and falls well within one standard deviation of the population mean which is 0.47. The variances are also close to each other and only differ from each other by 0.05.

Problem 2. In this problem, we will use `R` to calculate probabilities and quantiles for random variables.

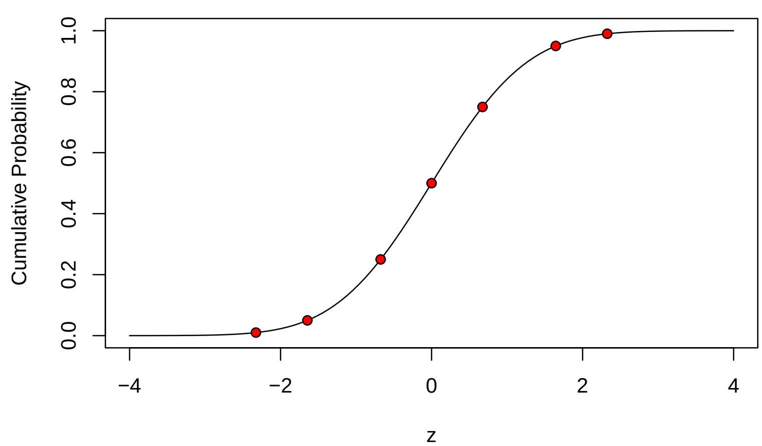
Problem 2 Part a) For Z, the standard normal random variable, determine the values for z, rounded to include three decimal places, such that $P\{Z \leq z\} = 0.01, 0.05, 0.25, 0.50, 0.75, 0.95,$ and 0.99, respectively. Indicate these z-values on a plot of the distribution function for Z.

Hint: Recall that the y-axis of the standard normal random variable cumulative distribution function $F_Z(z)$ denotes $P(Z \leq z)$. In order to compute the corresponding values for z, you will need to compute the quantile. Remember that the standard normal random variable has `mean=0` and `sd=1`.}

```r
# NOTE: Create a vector of probabilities
p = c(0.01, 0.05, 0.25, 0.5, 0.75, 0.95, 0.99)

# NOTE: Next, compute the quantiles

q = qnorm(p)

# NOTE: Create a data frame to display the probability P(Z <= z) with NOTE: the
# corresponding value z rounded to three decimal places
data.frame(`Cum Prob` = p, `z-value` = round(q, 3))
```

```
##   Cum.Prob z.value
## 1     0.01  -2.326
## 2     0.05  -1.645
## 3     0.25  -0.674
## 4     0.50   0.000
## 5     0.75   0.674
## 6     0.95   1.645
## 7     0.99   2.326
```

```r
# NOTE: The lines below will plot the cumulative distribution function for Z
# NOTE: Remember that the curve command must be defined in terms of x (not z)
curve(pnorm(x), xlim = c(-4, 4), ylim = c(0, 1), xlab = "z", ylab = "Cumulative Probability")

# NOTE: Finally, indicate these values on the plot.  NOTE: If you struggle with
# this, feel free to annotate on the plot instead.
points(q, p, pch = 21, bg = "red")
```



Problem 2 Part b) For X, a χ^2_4 random variable, determine the values for x, rounded to include three decimal places, such that $P\{X > x\} = 0.10, 0.05,$ and 0.01, respectively. Indicate these x-values values on a plot of the density function for X.

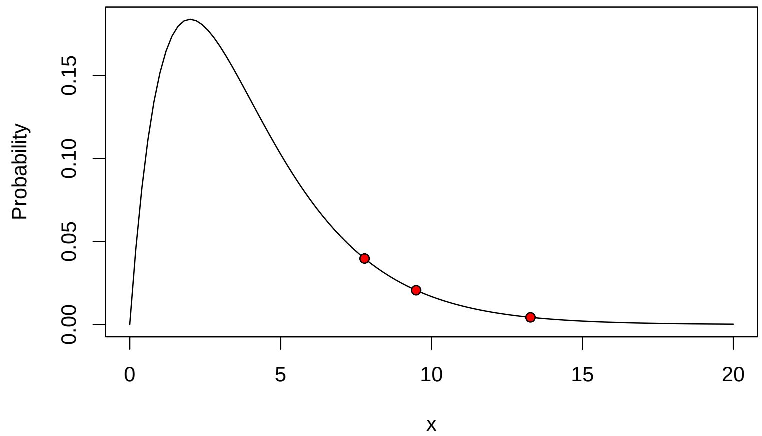
Hint: Recall that the y-axis of a χ^2_4 random variable cumulative distribution function denotes $P(X \leq x) = 1 - P(X > x)$. In order to compute the corresponding values for x, you will need to compute the quantile. Remember that a χ^2_4 random variable has `df=4`.}

NOTE: To do this task, enter your code in the R chunk below by filling in the blanks denoted with `FILL IN` *and uncommenting all non-NOTE lines. You may choose to upload an annotated density function, as in previous assignments, if you struggle to indicatee the values using the R-chunck.*

```
# NOTE: Create a vector of probabilities
p = c(0.1, 0.05, 0.01)
l = c(1 - 0.1, 1 - 0.05, 1 - 0.01)

# NOTE: Next, compute the quantiles. Keep in mind P(X \leq x)=1-P(X > x).
q = qchisq(l, 4)

# NOTE: Create a data frame to display the probability P(X <= x) with NOTE: the
# corresponding value x rounded to three decimal places
data.frame(`Surv Prob` = p, `x-value` = round(q, 3))
```

```
##   Surv.Prob x.value
## 1      0.10   7.779
## 2      0.05   9.488
## 3      0.01  13.277
```

```
# NOTE: The command dchisq(q, df) gives the density of chi-squared with df

# NOTE: The lines below will plot the density function for X NOTE: Remember
# that the curve command must be defined in terms of x
curve(dchisq(x, 4), from = 0, to = 20, xlab = "x", ylab = "Probability")

# NOTE: Finally, indicate these values on the plot.  NOTE: If you struggle with
# this, feel free to annotate on the plot instead.

# NOTE: The first step will be to define a vector of y-values on the density
# function NOTE: that correspond to the x-values. The fill ins in the points
# and segments NOTE: will be d's and q's.
d = dchisq(q, 4)
points(q, d, pch = 21, bg = "red")
```


Probability
0.00
0.05
0.10
0.15
0
5
10
15
20
x

Problem 2 Part c) Simulate 1000 independent Beta random variables with $\alpha = 2$ and $\beta = 4$. Determine the mean and variance of this sample. Are the sample mean and sample variance consistent with the distributional mean, $E[X]$, and distributional variance, $\text{Var}(X)$, defined below?

$$E[X] = \frac{\alpha}{\alpha + \beta}, \quad \text{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

```
set.seed(2022)
alpha = 2
beta = 4
N = 1000

# NOTE: Simulate the 1000 independent Beta random variables
beta_sim = rbeta(1000, 2, 4)

# NOTE: The lines below will compute the mean and variance of the simulation
sim_mean = mean(beta_sim)
sim_var = var(beta_sim)

# NOTE: The lines below will compute the distributional mean and variance
EX = (2/(2 + 4))
VarX = (2 * 4)/((6^2) * 7)

# NOTE: The line below will display the means and variances of the sample and
# distribution
data.frame(dist_mean = EX, sample_mean = sim_mean, dist_var = VarX, sample_var = sim_var)
```

```
##   dist_mean sample_mean   dist_var sample_var
## 1 0.3333333   0.3369563 0.03174603 0.03356632
```

Both the sample mean and variance are consistent with the distributional mean and variance as shown by the data frame above.