Homework 1

Winter 2023

Jaiden Atterbury

Instructions

- This homework is due in Gradescope on Wednesday Jan 11 by midnight PST.
- Please answer the following questions in the order in which they are posed.
- Don't forget to knit the document frequently to make sure there are no compilation errors.
- When you are done, download the PDF file as instructed in section and submit it in Gradescope.

Exercises

1. (Gizmo) The internal temperature in a gizmo (device) is a random variable X with PDF (in appropriate units)
$$f(x) = 11 \cdot (1-x)^{10}, \quad 0 < x < 1$$

a. The gizmo has a cutoff feature, so that whenever the temperature exceeds the cutoff - call it k - it turns off. It is observed that the gizmo shuts off with probability 10^{-22}. What is k?

Let the random variable X denote the internal temperature in a gizmo. In order to find the value of k we must use the information given to us about the temperature exceeding the cutoff. In particular, the problem statement says that whenever the temperature exceeds the cutoff k it turns off, and it does so with probability 10^{-22}. This translates to the statement $P(X > k) = 10^{-22}$, which can be rewritten as $1 - P(X \leq k) = 10^{-22}$. This translates to $1 - F(k) = 10^{-22}$. Thus we must start by fnding the CDF of X.

Finding the CDF of X:

$$\begin{aligned}
F(k) &= \int\limits_0^k 11(1-x)^{10} dx \\
&= 11 \int\limits_0^k (1-x)^{10} dx \\
&= -11 \int\limits_1^{1-k} u^{10} du \quad \text{Let } u = 1 - x \\
&= 11 \int\limits_{1-k}^1 u^{10} du \\
&= \left[u^{11}\right]_{1-k}^1 \\
&= 1 - (1-k)^{11}
\end{aligned}$$

Solving for k:
Now going back to the information about the cutoff we computed, we know that $1 - F(k) = 10^{-22}$. Thus all we need to do is plug in our value of F(k) and solve for k.

$$
\begin{aligned}
1 - (1 - (1 - k)^{11}) &= 10^{-22} \\
1 - 1 + (1 - k)^{11} &= 10^{-22} \\
(1 - k)^{11} &= 10^{-22} \\
1 - k &= \sqrt[11]{10^{-22}} \\
k &= 1 - \sqrt[11]{10^{-22}}
\end{aligned}
$$

Therefore, $\boxed{k = 1 - \sqrt[11]{10^{-22}}}$.

b. Fill in the blank: the number k is the $\boxed{1 - 10^{-22}}$ quantile of the distribution of X.

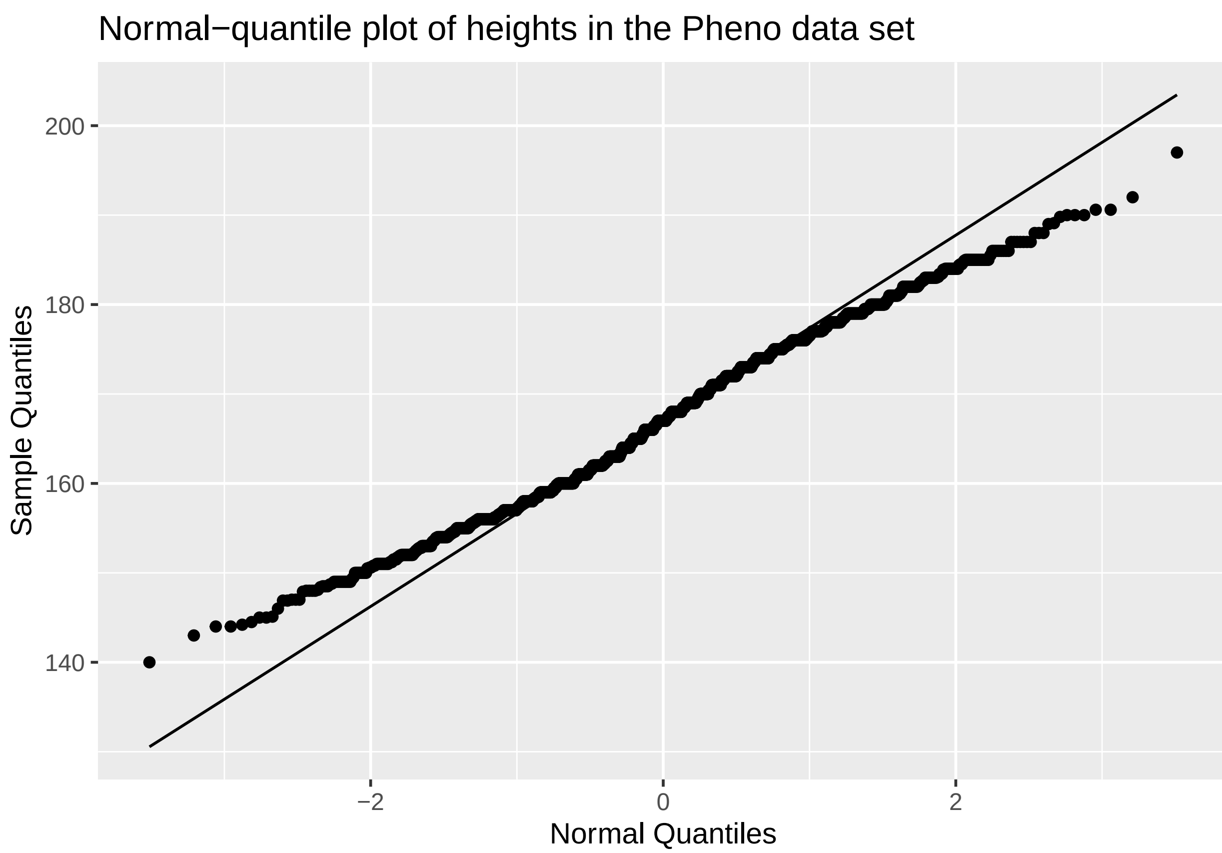
2. Pruim problem 3.46 on page 221 (Please review section 13.2 from notes for how to make QQplots. You can use `group = sex` in the mapping in order to create the conditional plots for part b.)

The Pheno data set contains phenotype information for 2457 subjects from the Finland-United States Investigation of NIDDM Genetics (FUSION) study of type 2 diabetes. Among the physical measurements recorded is the height (in cm) of the subjects.

a. Make a normal-quantile plot of the heights of all subjects in this data set. Do the heights appear to be normally distributed?

Normal-quantile plot of the heights:

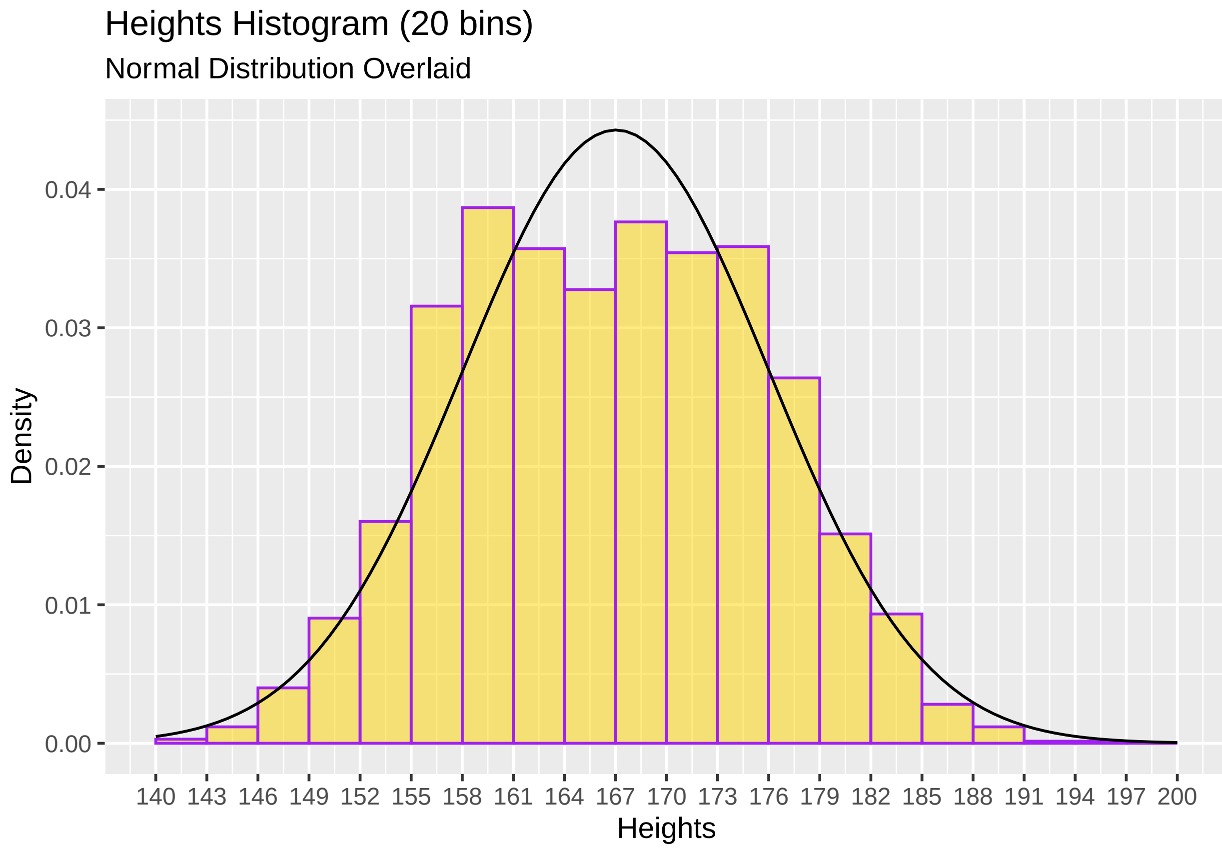
```
ggplot(data = Pheno, mapping = aes(sample = height)) +
  stat_qq(distribution = qnorm) +
  stat_qqline(distribution = qnorm) +
  labs(title = "Normal-quantile plot of heights in the Pheno data set",
       y = "Sample Quantiles",
       x = "Normal Quantiles",)
```



Histogram of heights with normal distribution overlaid:

```
height_data <- Pheno$height
sample_mean <- mean(height_data, na.rm=TRUE)
sample_sd <- sd(height_data, na.rm=TRUE)

ggplot() +
  geom_histogram(data = Pheno,
                 mapping = aes(x = height, y = ..density..),
                 binwidth = 3,
                 breaks = seq(140, 200, 3),
                 alpha = 0.5,
                 color = "purple",
                 fill = "gold") +
  labs(x = "Heights",
       y = "Density",
       title = "Heights Histogram (20 bins)",
       subtitle = "Normal Distribution Overlaid") +
  geom_function(fun = dnorm,
                args = list(mean = sample_mean, sd = sample_sd),
                xlim = c(140,200)) +
  scale_x_continuous(breaks = round(seq(140, 200, 3), 2))
```



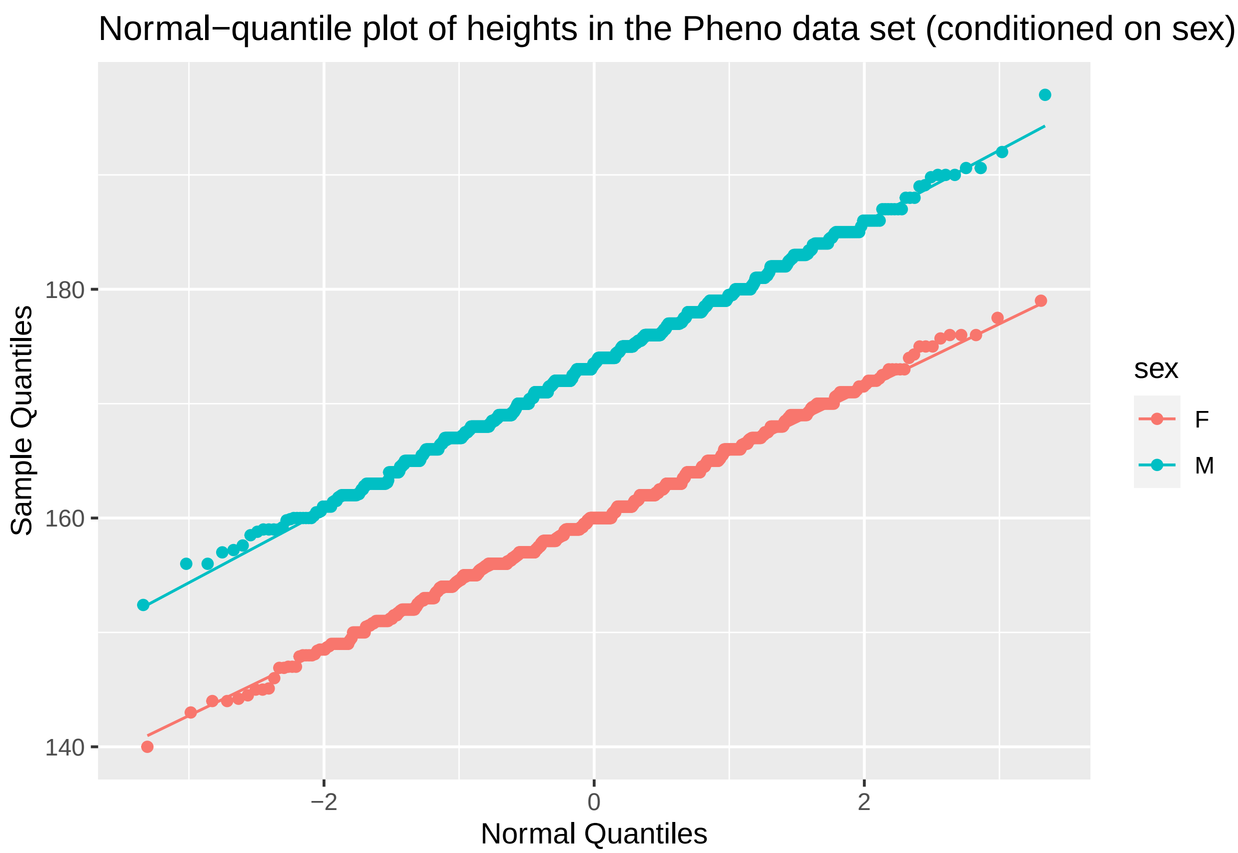
Assessing the fit:
Since we don't see a strong linear relationship between the percentiles of our sample data and the percentiles of the normal distribution, we don't have assurance that our data comes from a normal distribution. Furthermore, we don't have basis for using the normal distribution for modeling our data.

Furthermore, looking at the histogram of the heights with the normal distribution overlaid we see that there seems to be a bimodal aspect to the histogram. Further emphasizing that the heights do not seem to be normally distributed. This bimodal aspect could be caused by the fact that the male and female heghts could be coming from a normal distribution with a different mean (since males are taller on average).

b. Now make conditional normal-quantile plots, conditioned on sex. Do the heights of the men and women in this study appear to be approximately normally distributed? If not, how do the distributions differ from a normal distribution?

Normal-quantile plot of the heights (conditioned on sex):

```
ggplot(data = Pheno, mapping = aes(sample = height, group = sex, color = sex)) +
  stat_qq(distribution = qnorm) +
  stat_qqline(distribution = qnorm) +
  labs(title = "Normal-quantile plot of heights in the Pheno data set (conditioned on sex)",
       y = "Sample Quantiles",
       x = "Normal Quantiles",)
```



Assessing the fit:
Since we see a strong linear relationship between the percentiles of our sample data and the percentiles of the normal distribution for both sexes, we have some assurance that the heights of the men and women in this study appear to be approximately normally distributed.

3. (CDF method) Suppose X is a gamma random variable with shape parameter $k\,(>0)$ and rate parameter $\lambda\,(>0)$. In other words, the PDF of X is given by:

$$f(x) = \frac{\lambda^k}{\Gamma(k)}\, x^{k-1}\, e^{-\lambda\, x}, \qquad x > 0,$$

where $\Gamma(k)$ is the gamma function. Review 13.3 from the notes if you need a reminder.

Define

$$Y = \frac{1}{X}.$$

Show, using the CDF method, that Y has PDF

$$f(y) = \frac{\lambda^k}{\Gamma(k)}\, \frac{1}{y^{k+1}} e^{-\frac{\lambda}{y}} \qquad y > 0$$

Goal: Find PDF of Y using the CDF method

Step 1: Find analogous ways to represent the CDF of Y:

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P(\frac{1}{X} \leq y) \\
&= P(\frac{1}{y} \geq X) \\
&= 1 - P(X \leq \frac{1}{y}) \\
&= 1 - F_X(x) \quad \text{where } x = \frac{1}{y}
\end{aligned}$$

Step 2: Differentiate the CDF to obtain the PDF:

$$\begin{aligned}
f(y) &= \frac{d}{dy} F_Y(y) \\
&= \frac{d}{dy}(1 - F_X(x)) \\
&= \frac{d}{dy} 1 - \frac{d}{dy} F_X(x) \\
&= 0 - \frac{d}{dx} F_X(x) \cdot \frac{d}{dy} x \\
&= -f(x) \cdot \frac{d}{dy} \frac{1}{y} \\
&= \frac{1}{y^2} f(x) \quad \text{where } x = \frac{1}{y}
\end{aligned}$$

Step 3: Plug into the PDF of X and clean up:

$$\begin{aligned}
f(y) &= \frac{1}{y^2} \frac{\lambda^k}{\Gamma(k)} (\frac{1}{y})^{k-1} e^{-\lambda(\frac{1}{y})} \\
&= \frac{\lambda^k}{\Gamma(k)} \frac{1}{y^2 y^{k-1}} e^{-\lambda(\frac{1}{y})} \\
&= \frac{\lambda^k}{\Gamma(k)} \frac{1}{y^{k+1}} e^{-\frac{\lambda}{y}}
\end{aligned}$$

Thus, we have shown that that Y has PDF $f(y) = \frac{\lambda^k}{\Gamma(k)} \frac{1}{y^{k+1}} e^{-\frac{\lambda}{y}}, y > 0$.

4. (Beta dist) The beta distribution is a probability distribution that is often used in applications where the random variable is a proportion. The beta PDF depends on two *shape* parameters - call them α (> 0) and β (> 0) - and is given by:

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

a. Show that the uniform distribution on (0,1) is a special case of the beta when $\alpha = \beta = 1$.

In order to show that the uniform distribution on (0,1) is a special case of the beta when $\alpha = \beta = 1$. We must first find the PDF of the uniform distribution on (0,1) and make sure it is identical to the beta when $\alpha = \beta = 1$.

PDF of Uniform on (0,1):
Let $X \sim Unif(0,1)$. Then using the fact that the PDF of a uniform random variable is $f(x) = \frac{1}{b-a}$, it follows that the PDF of the uniform distribution on (0,1) is $f(x) = 1 \quad 0 < x < 1$.

PDF of Beta when $\alpha = \beta = 1$:
Let $X \sim Beta(\alpha = 1, \beta = 1)$. Then using the fact that the PDF of a beta random variable is $f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\, x^{\alpha-1}\,(1-x)^{\beta-1}$, it follows that the PDF of the beta distribution with $\alpha = \beta = 1$ is $f(x) = \frac{\Gamma(1+1)}{\Gamma(1)\,\Gamma(1)}\, x^{1-1}\,(1-x)^{1-1}$. Using the three facts that $\Gamma(1) = 1$, $\Gamma(k+1) = k\Gamma(k)$, and $x^0 = 1$, we can simplify the PDF to $f(x) = \frac{1 \cdot \Gamma(1)}{1 \cdot 1} \cdot 1 \cdot 1$. Thus, $f(x) = 1 \quad 0 < x < 1$.

Since $f(x) = 1 \quad 0 < x < 1$ for both distributions we can say the uniform distribution on (0,1) is a special case of the beta when $\alpha = \beta = 1$.

b. The beta distribution offers a very flexible array of shapes for modeling data. Run the following code interactively from your Console to see how it changes with different parameter values. Feel free to try other values for α and β. Then describe the impact of the parameters α and β on the shape. Specifically when is it symmetric? When is it skewed to the right? Skewed to the left? Bowl shaped? (Don't just state the α, β values, but rather provide a meaningful summary of the behavior)

When is the beta distribution symmetric?:
As it appears from the above beta distributions with different α and β values, the pdf of the beta distribution is symmetric whenever $\alpha = \beta$. In particular, the distribution is symmetric about 1/2 in all cases.

When is the beta distribution increasing?:
As it appears from the above beta distributions with different α and β values, the pdf of the beta distribution is strictly increasing when one of two cases occur: 1.) $\alpha = 1$ and $\beta < 1$ and 2.) $\alpha > 1$ and $\beta \leq 1$.

When is the beta distribution decreasing?:
As it appears from the above beta distributions with different α and β values, the pdf of the beta distribution is strictly decreasing when one of two cases occur: 1.) $\beta \geq 1$ and $\alpha < 1$ and 2.) $\beta > 1$ and $\alpha = 1$.

When is the beta distribution unimodal?:
As it appears from the above beta distributions with different α and β values, the pdf of the beta distribution is unimodal whenever $\alpha > 1$ and $\beta > 1$.

When is the beta right skewed?:
As it appears from the above beta distributions with different α and β values, the pdf of the beta distribution is right skewed whenever $\alpha > 1$ and $\beta > 1$ (the pdf is unimodal) and $\alpha < \beta$.

When is the beta left skewed:
As it appears from the above beta distributions with different α and β values, the pdf of the beta distribution is right skewed whenever $\alpha > 1$ and $\beta > 1$ (the pdf is unimodal) and $\alpha > \beta$.

When is the beta distribution bowl shaped?:
As it appears from the above beta distributions with different α and β values, the beta distribution is bowl shaped whenever $\alpha < 0$ and $\beta < 0$.

c. Since the beta distribution is a valid PDF, it must integrate to 1. In other words, it must be the case that

$$\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

Use this fact to show that the rth moment of the Beta is given by

$$E\left[X^r\right] = \frac{\Gamma(\alpha+r)}{\Gamma(\alpha)} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+r)}.$$

Finding the rth moment of the Beta:

$$\begin{aligned}
E[X^r] &= \int\limits_{-\infty}^{\infty} x^r f(x)dx \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int\limits_{0}^{\infty} x^r x^{\alpha-1}(1-x)^{\beta-1}dx \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int\limits_{0}^{\infty} x^{r+\alpha-1}(1-x)^{\beta-1}dx
\end{aligned}$$

Notice that $\int\limits_{0}^{\infty} x^{r+\alpha-1}(1-x)^{\beta-1}dx$ has the same from as the above fact. Therefore we can say that $\int\limits_{0}^{\infty} x^{r+\alpha-1}(1-x)^{\beta-1}dx$ can be rewritten as $\frac{\Gamma(\alpha+r)\Gamma(\beta)}{\Gamma(\alpha+\beta+r)}$. Thus when we plug this into our rth moment calculation we get:

$$\begin{aligned}
E[X^r] &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \times \frac{\Gamma(\alpha+r)\Gamma(\beta)}{\Gamma(\alpha+\beta+r)} \\
&= \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+r)\Gamma(\beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+r)} \\
&= \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+r)}{\Gamma(\alpha)\Gamma(\alpha+\beta+r)} \\
&= \boxed{\frac{\Gamma(\alpha+r)}{\Gamma(\alpha)} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+r)}}
\end{aligned}$$

d. Use your result from part c. to show that

$$\begin{aligned}
\mu = E\left[X\right] &= \frac{\alpha}{\alpha+\beta}, \\
\sigma^2 &= Var\left[X\right] \\
&= \frac{\mu\left(1-\mu\right)}{\alpha+\beta+1}
\end{aligned}$$

Finding E[X]:
Since E[X] is the 1st moment, we can use the result from part c and rewrite E[X] as:

$$\begin{aligned}
E[X] &= \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+1)} \quad \text{(HW1 Problem 4c)} \\
&= \frac{\alpha\Gamma(\alpha)}{\Gamma(\alpha)} \times \frac{\Gamma(\alpha+\beta)}{(\alpha+\beta)\Gamma(\alpha+\beta)} \quad \text{(Lemma 13.2.2)} \\
&= \boxed{\frac{\alpha}{\alpha+\beta}}
\end{aligned}$$

Finding Var[X]:
To find Var[X] we must use the fact that $Var[X] = E[X^2] - \mu^2$ and then use the fact from part c to find $E[X^2]$.

Finding $E[X^2]$:

$$\begin{aligned}
E[X^2] &= \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+2)} \quad \text{(HW1 Problem 4c)} \\
&= \frac{\Gamma(\alpha+1+1)}{\Gamma(\alpha)} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+1+1)} \\
&= \frac{(\alpha+1)\Gamma(\alpha+1)}{\Gamma(\alpha)} \times \frac{\Gamma(\alpha+\beta)}{(\alpha+\beta+1)\Gamma(\alpha+\beta+1)} \quad \text{(Lemma 13.2.2)} \\
&= \frac{\alpha(\alpha+1)\Gamma(\alpha)}{\Gamma(\alpha)} \times \frac{\Gamma(\alpha+\beta)}{(\alpha+\beta)(\alpha+\beta+1)\Gamma(\alpha+\beta)} \quad \text{(Lemma 13.2.2)} \\
&= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} \\
&= \frac{\mu(\alpha+1)}{(\alpha+\beta+1)}
\end{aligned}$$

Putting it all together:

$$\begin{aligned}
Var[X] &= E[X^2] - \mu^2 \\
&= \frac{\mu(\alpha+1)}{(\alpha+\beta+1)} - \mu^2 \\
&= \mu(\frac{(\alpha+1)}{(\alpha+\beta+1)} - \mu) \\
&= \mu(\frac{(\alpha+1) - (\alpha+\beta+1)\mu}{(\alpha+\beta+1)}) \\
&= \mu(\frac{(\alpha+1) - (\alpha+\beta+1)\frac{\alpha}{\alpha+\beta}}{(\alpha+\beta+1)}) \\
&= \mu(\frac{(\alpha+1) - \frac{\alpha^2+\alpha\beta+\alpha}{\alpha+\beta}}{(\alpha+\beta+1)}) \\
&= \mu(\frac{\frac{\alpha^2+\alpha+\beta+\beta\alpha-\alpha^2-\alpha\beta-\alpha}{\alpha+\beta}}{(\alpha+\beta+1)}) \\
&= \mu(\frac{\frac{\beta}{\alpha+\beta}}{(\alpha+\beta+1)}) \\
&= \mu(\frac{1-\mu}{(\alpha+\beta+1)}) \quad \text{since } 1-\mu = 1 - \frac{\alpha}{\alpha+\beta} = \frac{\alpha+\beta}{\alpha+\beta} - \frac{\alpha}{\alpha+\beta} = \frac{\beta}{\alpha+\beta} \\
&= \boxed{\frac{\mu\,(1-\mu)}{\alpha+\beta+1}}
\end{aligned}$$