Homework 3

Gabriel Lahman

2/4/2019

Question 1

Although the Weibull and gamma distributions do not nest each other, they are both nested by the generalized gamma distribution. Specifically, the Weibull is a generalized gamma with $\alpha = 1$ and the gamma is a generalized gamma with $k = 1$. Although we cannot directly compare the gamma and the Weibull, we can compare how well the restricted models compare to the generalized gamma. More specifically, we first compute the likelihood ratio statistic of the gamma versus the generalized gamma. Using the maximum likelihood estimators of α and θ from the fitted gamma model, the maximum estimator of the k for the generalized gamma with the set α and θ would be found, then the likelihood ratio statistic would be computed (Denoted as lr_{gamma}). Second, the likelihood ratio statistic for generalized gamma versus the Weibull would be found by using the maximum likelihood estimators of k and θ from the fitted Weibull model as set parameters in the generalized gamma, then finding the maximum likelihood estimator for α in the generalized gamma and computing the likelihood ratio statistic of the Weibull versus generalized gamma (Denoted as $lr_{Weibull}$). The model which better fit the data would be the model that best matched the generalized gamma using the fixed parameters, or more specifically, $min(lr_{gamma}, lr_{Weibull})$.

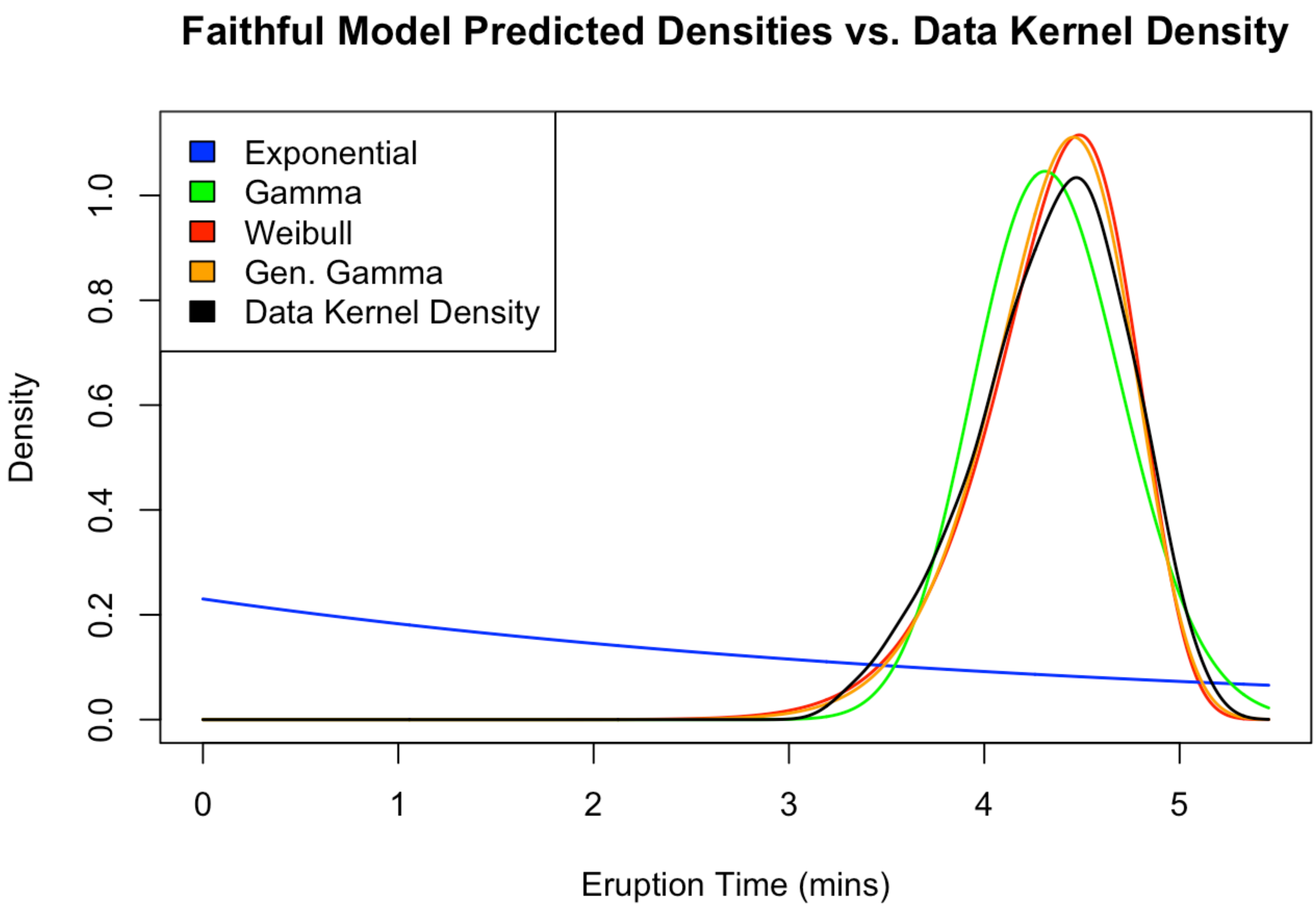
Question 2

```
data("faithful")
y = faithful$eruptions[faithful$waiting > 71]

MLE_exp = MLE_fun_exp(y)
MLE_gamma = MLE_fun_gamma(y)
MLE_weibull = MLE_fun_weibull(y)
MLE_gen_gamma = MLE_fun_gen_gamma(y)
```

```
## [1] "Log Likelihoods: Exp= -395.02 Gamma= -72.89 Weibull= -67.16 Gen. Gamma= -66.9
2"
```

Using the log likelihoods for each distribution, it appears as the generalized gamma with the maximized parameters best fits the data, since its log likelihood is the highest. The Weibull also appears to be marginally less fitting than the generalized gamma.

```
y_density = density(y, from = 0)
exp_density = pred_dens_exp(y_density$x, MLE_exp)
gamma_density = pred_dens_gamma(y_density$x, MLE_gamma)
weibull_density = pred_dens_weibull(y_density$x, MLE_weibull)
gen_gamma_density = pred_dens_gen_gamma(y_density$x, MLE_gen_gamma)
plot(x = y_density$x,
     y = weibull_density,
     type = "l",
     col = "red",
     lwd = 1.5,
     xlab = "Eruption Time (mins)",
     ylab = "Density",
     main = "Faithful Model Predicted Densities vs. Data Kernel Density")
lines(x = y_density$x,
     y = exp_density,
     col = "blue",
     lwd = 1.5)
lines(x = y_density$x,
      y = gamma_density,
      col = "green",
      lwd = 1.5)
lines(x = y_density$x,
      y = gen_gamma_density,
      col = "orange",
      lwd = 1.5)
lines(x = y_density$x,
      y = y_density$y,
      lwd = 1.5)
legend("topleft",
       c("Exponential", "Gamma", "Weibull", "Gen. Gamma", "Data Kernel Density"),
       fill = c("blue", "green", "red", "orange", "black"))
```




```
lr_exp_gamma = -2*(MLE_exp$log_lik - MLE_gamma$log_lik)
lr_exp_weibull = -2*(MLE_exp$log_lik - MLE_weibull$log_lik)
lr_gamma_gen_gamma = -2*(MLE_gamma$log_lik - MLE_gen_gamma$log_lik)
lr_weibull_gen_gamma = -2*(MLE_weibull$log_lik - MLE_gen_gamma$log_lik)

p_exp_gamma = 1 - pchisq(lr_exp_gamma, 1)
p_exp_weibull = 1 - pchisq(lr_exp_weibull, 1)
p_gamma_gen_gamma = 1 - pchisq(lr_gamma_gen_gamma, 1)
p_weibull_gen_gamma = 1 - pchisq(lr_weibull_gen_gamma, 1)
```

```
## [1] "LR Statistic, P-Value"
```

```
## [1] "Exp/Gamma: 644.27 , 0"
```

```
## [1] "Exp/Weibull: 655.71 , 0"
```

```
## [1] "Gamma/Gen.Gamma: 11.93 , 0.001"
```

```
## [1] "Weibull/Gen.Gamma: 0.49 , 0.485"
```

Looking at the resulting values for the likelihood ratio statistics and their corresponding p-values (based off of χ^2_1), the very small p-values for the exponential/gamma comparison and exponential/Weibull comparison show that the more restricted exponential model does not fit the data as well as the models with free α or k. Additionally, the gamma/generalized gamma comparison also yields a significant p-value, showing that the model with free k and α fits better than just free α. Looking at the graph, the non-significant result for the Weibull/generalized gamma comparison is pretty obvious. Their graphs are incredibly close, which echos the closeness of their log-likelihoods. This may allow us to make an assumption that keeping k unrestricted is allows for much more improvement in model fit than keeping α unrestricted. Visually, both the Weibull and generalized gamma appear to be very close to actual data. Additionally, their log-likelihoods are relatively good. As result of this, I would be comfortable using either the Weibull or generalized gamma to predict new data.

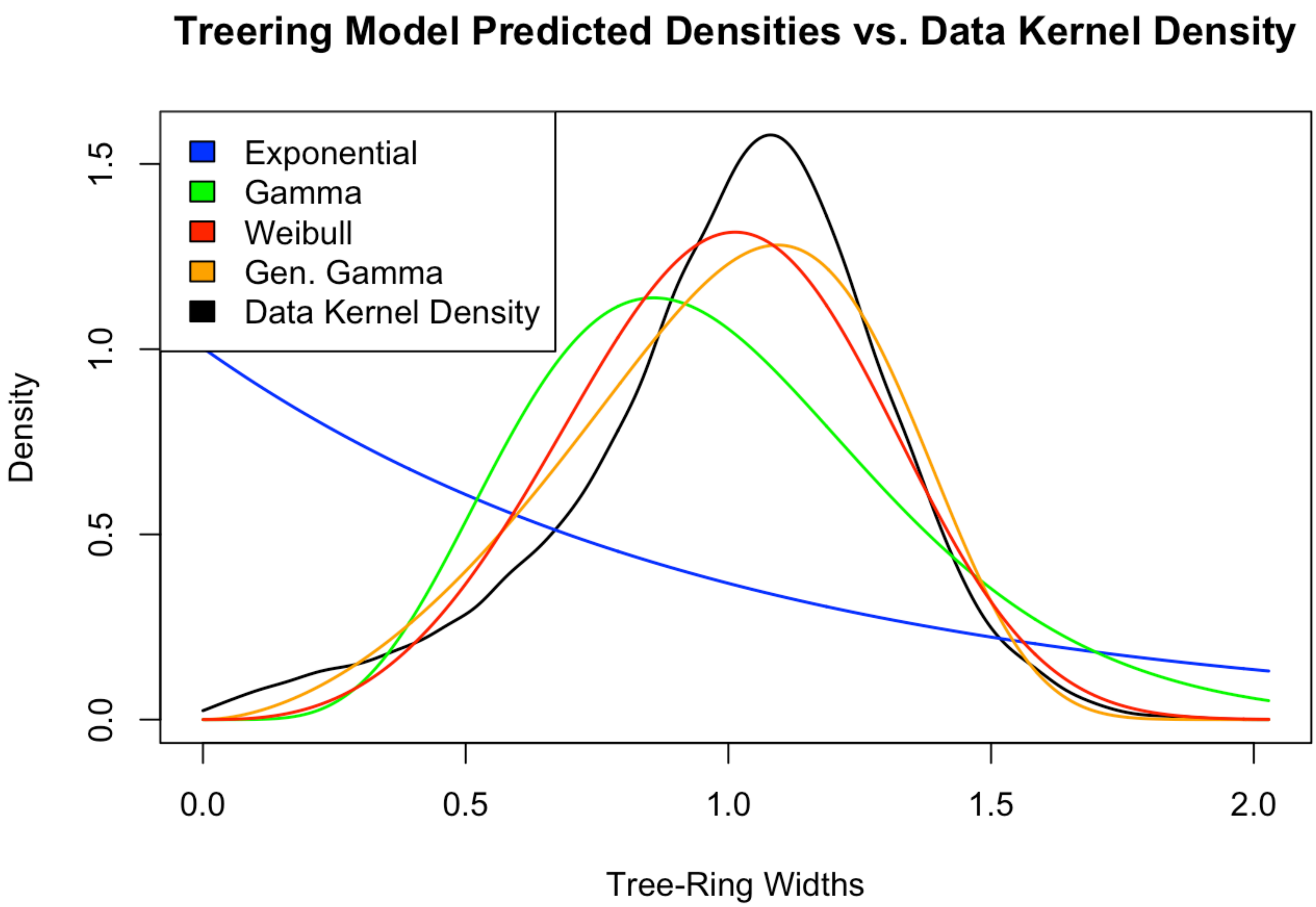
Question 3

```
data("treering")

y = treering[treering > 0]

MLE_exp = MLE_fun_exp(y)
MLE_gamma = MLE_fun_gamma(y)
MLE_weibull = MLE_fun_weibull(y)
MLE_gen_gamma = MLE_fun_gen_gamma(y)
```

```
## [1] "Log Likelihoods: Exp= -7954.72 Gamma= -3062.11 Weibull= -1855.52 Gen. Gamma=
-1629.43"
```

From using only the log-likelihood, the generalized gamma appears to best fit the data by a large margin, as it’s value is much higher relative to any other distribution.

```
y_density = density(y, from=0)
exp_density = pred_dens_exp(y_density$x, MLE_exp)
gamma_density = pred_dens_gamma(y_density$x, MLE_gamma)
weibull_density = pred_dens_weibull(y_density$x, MLE_weibull)
gen_gamma_density = pred_dens_gen_gamma(y_density$x, MLE_gen_gamma)

plot(x = y_density$x,
     y = y_density$y,
     type = "l",
     lwd = 1.5,
     xlab = "Tree-Ring Widths",
     ylab = "Density",
     main = "Treering Model Predicted Densities vs. Data Kernel Density")
lines(x = y_density$x,
     y = exp_density,
     col = "blue",
     lwd = 1.5)
lines(x = y_density$x,
      y = gamma_density,
      col = "green",
      lwd = 1.5)
lines(x = y_density$x,
      y = gen_gamma_density,
      col = "orange",
      lwd = 1.5)
lines(x = y_density$x,
      y = weibull_density,
      col="red",
      lwd = 1.5)
legend("topleft",
       c("Exponential", "Gamma", "Weibull", "Gen. Gamma", "Data Kernel Density"),
       fill = c("blue", "green", "red", "orange", "black"))
```




```
lr_exp_gamma = -2*(MLE_exp$log_lik - MLE_gamma$log_lik)
lr_exp_weibull = -2*(MLE_exp$log_lik - MLE_weibull$log_lik)
lr_gamma_gen_gamma = -2*(MLE_gamma$log_lik - MLE_gen_gamma$log_lik)
lr_weibull_gen_gamma = -2*(MLE_weibull$log_lik - MLE_gen_gamma$log_lik)

p_exp_gamma = 1 - pchisq(lr_exp_gamma, 1)
p_exp_weibull = 1 - pchisq(lr_exp_weibull, 1)
p_gamma_gen_gamma = 1 - pchisq(lr_gamma_gen_gamma, 1)
p_weibull_gen_gamma = 1 - pchisq(lr_weibull_gen_gamma, 1)
```

```
## [1] "LR Statistic, P-Value"
```

```
## [1] "Exp/Gamma: 9785.2 , 0"
```

```
## [1] "Exp/Weibull: 12198.4 , 0"
```

```
## [1] "Gamma/Gen.Gamma: 2865.37 , 0"
```

```
## [1] "Weibull/Gen.Gamma: 452.18 , 0"
```

Looking at the likelihood ratio statistics and their corresponding p-values, it is very obvious that as you move from most-restricted model (exponential) to least restricted (generalized gamma), the addition of free parameters greatly improves model fit. However, it is important to realize that these improvements are relative. Even the generalized gamma, the “best fitting” model, has a very bad log-likelihood and visually does not compare very well to the actual data. Because of this, I would not be comfortable using any of the models to predict new data.