In lecture 16, we looked at precipitation amounts in Madison County (at Morrisville station). We found that the Weibull distribution had a good fit to the monthly precipitation amounts.

We found that the MLEs for the Weibull distribution were

$$\hat{a} = 2.1871$$
$$\hat{\sigma} = 3.9683$$

and

$$-\mathcal{L}(\{\hat{a}, \hat{\sigma}\}|\mathbf{x}) = 2166.496$$

is the realized negative log-likelihood. Note this means that the log-likelihood is

$$\mathcal{L}(\{\hat{a}, \hat{\sigma}\}|\mathbf{x}) = -2166.496,$$

and the usual likelihood is

$$L(\{\hat{a}, \hat{\sigma}\}|\mathbf{x}) = e^{[\mathcal{L}(\{\hat{a}, \hat{\sigma}\}|\mathbf{x})]} \approx= e^{-2166.496},$$

which `R` cannot differentiate from 0.

1. Someone asked "why Weibull?" in class. That is, why wouldn't we use another right-skewed distribution like the Gamma (see Lecture 15), or the Log-Normal (see Lecture 17).

   (a) Compute the MLEs for these data using a Gamma distribution.

   (b) Compute the MLEs for these data using the Log-Normal distribution.

   (c) Compute the likelihood ratio to compare the Weibull and the Gamma distribution. Which has a better fit according to the likelhiood ratio?

   $$Q = \frac{L(\{\hat{a}, \hat{\sigma}\}|\mathbf{x})}{L(\{\hat{\alpha}, \hat{\beta}\}|\mathbf{x})} = e^{\left[\mathcal{L}(\{\hat{a}, \hat{\sigma}\}|\mathbf{x}) - \mathcal{L}(\{\hat{\alpha}, \hat{\beta}\}|\mathbf{x})\right]}$$

   (d) Compute the likelihood ratio to compare the Weibull and the Log-Normal distribution. Which has a better fit according to the likelihood ratio?

   $$Q = \frac{L(\{\hat{a}, \hat{\sigma}\}|\mathbf{x})}{L(\{\hat{\mu}, \hat{\sigma}\}|\mathbf{x})} = e^{[\mathcal{L}(\{\hat{a}, \hat{\sigma}\}|\mathbf{x}) - \mathcal{L}(\{\hat{\mu}, \hat{\sigma}\}|\mathbf{x})]}$$

   (e) Compute the likelihood ratio to compare the Gamma and the Log-Normal distribution. Which has a better fit according to the likelhiood ratio?

   $$Q = \frac{L(\{\hat{\alpha}, \hat{\beta}\}|\mathbf{x})}{L(\{\hat{\mu}, \hat{\sigma}\}|\mathbf{x})} = e^{\left[\mathcal{L}(\{\hat{\alpha}, \hat{\beta}\}|\mathbf{x}) - \mathcal{L}(\{\hat{\mu}, \hat{\sigma}\}|\mathbf{x})\right]}$$

2. Optional Coding Challenge. Choose the "best" distribution and refit the model by season.

   (a) Fit the Distribution for Winter (December-February).

   (b) Fit the Distribution for Spring (March-May).

   (c) Fit the Distribution for Summer (June-August).

   (d) Fit the Distribution for Fall (September-November).

   (e) Plot the four distributions in one plot using `cyan3` for Winter, `chartreuse3` for Spring, `red3` for Summer, and `chocolate3` for Fall. Note any similarities/differences you observe across the seasons.

```r
library(tidyverse)
library(e1071)
library(nleqslv)

################################################################################
# Precipitation in Madison County
################################################################################
dat.precip <- read_csv(file = "agacis.csv")

####################################
# Clean Data
####################################
dat.precip.long <- dat.precip |>
  dplyr::select(-Annual) |>                   # Remove annual column
  pivot_longer(cols = c(Jan, Feb, Mar, Apr,   # pivot the column data into one col
                        May, Jun, Jul, Aug,
                        Sep, Oct, Nov, Dec),
               values_to = "Precipitation",   # store the values in Precipitation
               names_to = "Month") |>         # store the months in Month
  mutate(Precipitation = case_when(Precipitation == "M" ~ NA_character_,
                                   TRUE                 ~ Precipitation))|>
  mutate(Precipitation = as.numeric(Precipitation))

x = dat.precip.long$Precipitation

# 1.(a) MLE for Gamma Distribution

gamma = function(par, data, neg = FALSE){
  alpha = exp(par[1])
  beta = exp(par[2])
  loglik = sum(dgamma(data, shape = alpha, scale = beta, log = TRUE), na.rm = TRUE)
  return(ifelse(neg, -loglik, loglik))
}
gamma.mle = optim(par = c(1,1), fn = gamma, data = x, neg = TRUE)
alpha.hat = exp(gamma.mle$par[1])
beta.hat = exp(gamma.mle$par[2])
loglik.gamma = -gamma.mle$value

# 1.(b) MLE for Log-Normal Distribution

lognormal = function(par, data, neg = FALSE){
  mu = par[1]
  sigma = exp(par[2])
  loglik = sum(dlnorm(data, meanlog = mu, sdlog = sigma, log = TRUE), na.rm = TRUE)
  return(ifelse(neg, -loglik, loglik))
}
lognorm.mle = optim(par = c(1,1), fn = lognormal, data = x, neg = TRUE)
mu.hat = lognorm.mle$par[1]
sigma.hat = exp(lognorm.mle$par[2])
loglik.lognorm = -lognorm.mle$value

# 1.(c)

loglik.weibull = -2166.496
(Qwg = exp(loglik.weibull - loglik.gamma))

## [1] 2.161379e-07

# The likelihood ratio is approximately 2.16e-07, which is much less than 1.
# This means the Gamma distribution provides a better fit than the Weibull
# distribution.

# 1.(d)

(Qwln = exp(loglik.weibull - loglik.lognorm))

## [1] 2.370639e+16

# The liklihood ratio is approximately 2.37e+16, which is much larger than
# 1 so the Weibull distributions fits the data better than the Log-Normal
# distribution.

# 1.(e)

(Qgl = exp(loglik.gamma - loglik.lognorm))

## [1] 1.096818e+23

# The likelihood ratio is approximately 1.10e+23 which is much larger than
# 1 indicating that the Gamma distribution provides a much better fit than
# the Log-Normal distribution.
```