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.

```r
clean.data <- read.csv("cleandata.csv")
llgamma <- function(data, par, neg=F){
  alpha <- par[1]
  beta <- par[2]

  loglik <- sum(log(dgamma(x=data, shape=alpha, rate=beta)), na.rm=T)

  return(ifelse(neg, -loglik, loglik))
}

(mles <- optim(par = c(1,1),
           fn = llgamma,
           data=clean.data$Precipitation,
           neg=T))

## $par
## [1] 4.174581 1.189099
##
## $value
## [1] 2151.149
##
## $counts
## function gradient
##       97       NA
##
## $convergence
## [1] 0
##
## $message
## NULL

gamma.ll <- -mles$value[1]
alpha.hat.mle <- mles$par[1]
beta.hat.mle <- mles$par[2]
```

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

```
lllnorm <- function(par, data, neg = F) {
  x <- data
  mu <- par[1]
  sigma <- par[2]
loglik <- sum(log(dlnorm(x = x, meanlog = mu, sdlog = sigma)), na.rm=T)
ifelse(neg,-loglik, loglik)
}

(mle.lnorm <- optim(fn = lllnorm, par = c(1,1), data = clean.data$Precipitation, neg = T))

## $par
## [1] 1.1312609 0.5333417
##
## $value
## [1] 2204.201
##
## $counts
## function gradient
##       57       NA
##
## $convergence
## [1] 0
##
## $message
## NULL

norm.ll <- -mle.lnorm$value[1]
```

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

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

Answer: Because the liklihood ratio is extremely close to zero, we can say that the gamma distribution offers stronger support in representing the data than the weibull distribution does.

```
weibull.ll <- -2166.496
(wg.ratio <- exp(weibull.ll-gamma.ll))

## [1] 2.161318e-07
```

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

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

Answer: In this case, the ratio is significantly above one, indicating that the weibull distribution offers much stronger support for the data than the normal distribution does.

```
(wl.ratio <- exp(weibull.ll-norm.ll))

## [1] 2.370668e+16
```

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

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

Answer: This ratio is once again far above one, indicating that the Gamma distribution offers stronger support for the data then the log normal distribution does.

```
(gl.ratio <- exp(gamma.ll-norm.ll))

## [1] 1.096862e+23
```

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.