STAT 352: Data Modeling and Inference

Homework 4. Due Friday, February 18th. I encourage you to type all of your solutions, though this is not necessary. However, you must scan (or photograph) any handwritten portions and upload the files to Canvas. For questions that require `R` code, you must turn in your `R` code on Canvas. Your code must in a .Rmd file.

Question 1 (Computation): *Preamble:* In Lab 4, you saw MLE functions for the Exponential and Gamma distributions and coded your own MLE functions for alternative parameterizations of the Gamma distribution as well as for the Weibull distributions. These three distributions generate only positive random variables and their density functions are given by

$$\begin{aligned} f_E(y;\beta) &= \beta \exp(-\beta y) \\ f_G(y;\alpha,\beta) &= \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} \exp(-\beta y) \\ f_W(y;k,\lambda) &= \frac{k}{\lambda^k} y^{k-1} \exp\left(-\left(\frac{y}{\lambda}\right)^k\right) \end{aligned}$$

The Exponential distribution has one parameter, a rate parameter β, which is positive. The Exponential distribution can only really be used in very specific settings because in only has one parameter.

The Gamma distribution has two parameters, a shape parameter α and a rate parameter β, both of which are positive. The Exponential distribution is a special case of the Gamma distribution with $\alpha = 1$. The Gamma distribution generalizes the Exponential distribution by changing the power of y multiplying $\exp(-\beta y)$ in the density. This is done to allow the density to have different kinds of origin behavior (for small y values). This also changes the tail behavior of the density (for large y values).

The Weibull distribution has two parameters, a shape parameter k and a scale parameter λ, both of which are positive. This parameterization of the Weibull distribution is to match that for the built-in suite of functions for the Weibull distribution in `R` . Similarly, the Exponential distribution is a special case of the Weibull distribution with $k = 1$ and $\beta = 1/\lambda$. The Weibull distribution generalizes the Exponential distribution by changing the power of y in the exponential function in the density. In fact, it is obtained by raising a Exponential random variable to the power $1/k$. This is done to allow the density to have different kinds of tail behavior (for large y values). This also changes the origin behavior of the density (for small y values).

An alternative distribution to the Exponential, Gamma, and Weibull distributions is the Log-Normal distribution. This is defined as the name suggests. If $X = \log(Y)$ for $Y > 0$ and X is given a Normal distribution, then Y is said to be log-normally distributed. The density is given by

$$f_{LN}(y;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{y} \exp\left(-\frac{1}{2\sigma^2}\left(\log(y) - \mu\right)^2\right).$$

`R` has built-in density, probability, quantile, and random generation function for the Log-Normal Distribution (`?dlnorm`) that match this parameterization.

We have already seen two distributions here that are defined via transformations, the Weibull and the Log-Normal. Two other common distributions for modeling positive data are the Inverse-Gamma distribution and Inverse-Weibull distribution (commonly called the Frechet distribution).

The Inverse-Gamma is obtained by taking $Y = 1/X$ where X is Gamma distributed and the Inverse-Weibull (Frechet) is obtained by taking $Y = 1/X$ where X is Weibull Distributed. Their densities are given by

$$f_{IG}(y;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{-\alpha-1} \exp\left(-\frac{\beta}{y}\right)$$
$$f_{IW}(y;k,\lambda) = \frac{k}{\lambda^k} y^{-k-1} \exp\left(-\left(\frac{1}{\lambda y}\right)^k\right)$$

In these distributions, β is a scale and λ is a rate (scales divide y and rates multiply y in the density). There are myriad other distributions that we could make by changing variables, for instance the Generalized-Gamma is obtained by taking a Gamma distributed random variable to a power. However, I think we have probably ventured far enough for now.

Useful functions: There are built-in functions for the Exponential, Gamma, Weibull, and Log-Normal distributions (`?dexp`, `?dgamma`, `?dweibull`, `?dlnorm`) that we can use when trying to fit these distributions to some data. Unfortunately, there are no built in functions for the Inverse-Gamma or Inverse-Weibull (Frechet) distribution. In order to not make this HW too difficult, I am going to provide the density functions for these distributions.

```
dinvgamma = function(x, shape, scale, log = FALSE) {
#scale is beta
#shape is alpha
y=1/x
out = dgamma(y,shape,rate=scale,log=TRUE)-2*log(x)
if(log){out}else{exp(out)}
}
dinvweibull = function(x, shape, rate = 1, log = FALSE) {
#shape is k
#rate is lambda
y=1/x
out = dweibull(y,shape,scale=rate,log=TRUE)-2*log(x)
if(log){out}else{exp(out)}
}
```

Notice that these are DUMB functions; they do not do any error catching (if you give bad input, they are not going to try to stop you).

Do the following: We are going to analyze a dataset of eruption times of the Old Faithful geyser. To load the dataset, use the code

```
data(faithful); y = faithful$eruptions[faithful$waiting<71]
```

a) Fit the MLE for the Gamma distribution to the data. Report the log-likelihood and a 98% confidence interval for the parameters using the asymptotic normal distribution of the MLE and an estimate of the Fisher Information obtained using the output from `optimHess`. Do you think you can distinguish α from 1?

b) Fit the MLE for the Inverse-Gamma distribution to the data. Report the log-likelihood and a 98% confidence interval for the parameters using the asymptotic normal distribution of the

MLE and an estimate of the Fisher Information obtained using the output from `optimHess`. Do you think you can distinguish α from 1?

c) Fit the MLE for the Weibull distribution to the data. Report the log-likelihood and a 98% confidence interval for the parameters using the asymptotic normal distribution of the MLE and an estimate of the Fisher Information obtained using the output from `optimHess`. Do you think you can distinguish k from 1?

d) Fit the MLE for the Inverse-Weibull distribution to the data. Report the log-likelihood and a 98% confidence interval for the parameters using the asymptotic normal distribution of the MLE and an estimate of the Fisher Information obtained using the output from `optimHess`. Do you think you can distinguish k from 1?

e) Fit the MLE for the Log-Normal distribution to the data. Report the log-likelihood and a 98% confidence interval for the parameters using the asymptotic normal distribution of the MLE and an estimate of the Fisher Information obtained using the output from `optimHess`. Does this model fit the data as well as (or better than) the previous four? Do you think this distribution fits the data sufficiently well to model the data?

Question 2 (Theory): *Preamble:* We are going to think about the CLT for MLEs in the context of the five different models for different kinds of data from HW 3. For each question, I will give you the density (mass) function for a single data point, its expectation, the log-likelihood for an observed data vector, the second derivative of the log-likelihood, and the MLE. You will notice that these are all examples where the sample mean is sufficient for the parameter, as evidenced by the part of the log-likelihood function that I have sectioned off as $n \times [\cdots]$. All of the second derivatives of the log-likelihoods are of the form $-n \times [\cdots]$. Your task is to compute an estimate of the Fisher information and then use its limiting value to state the CLT for the MLE.

Useful formulas and facts:

$$
\begin{aligned}
\exp(a+b) &= \exp(a)\exp(b) \\
\exp(\sum a_i) &= \prod \exp(a_i) \\
\log(ab) &= \log(a)+\log(b) \\
\log(\prod a_i) &= \sum \log(a_i) \\
\log(a^b) &= b\log(a) \\
\exp(a)^b &= \exp(ab) \\
\log(\exp(b)) &= b \\
\exp(\log(a)) &= a \\
k! &= 1\times 2\times 3\times\cdots\times k \\
\tbinom{n}{k} &= \tfrac{n!}{k!(n-k)!}
\end{aligned}
$$

Do the following:

a) A random variable Y follows a Bernoulli(p) distribution if Y can take values only 0 or 1 and $p = P(Y = 1)$. This is the most basic model for $0-1$ events like disease diagnosis or whether a machine will work properly when you turn it on. The mass function for a Bernoulli(p)

$$f(y;p) = p^y(1-p)^{1-y}.$$

The parameter p must be between 0 and 1. The expectation of Y is given by $\mathrm{E}[Y] = p$. Let $Y_1, \ldots, Y_n \overset{iid}{\sim}$ Bernoulli(p) with realizations $y_1, \ldots, y_n$. The log-likelihood for p is

$$\ell(p; y_1, \ldots, y_n) = n \times [\bar{y} \log(p) + (1 - \bar{y}) \log(1 - p)]$$

and the MLE is $\hat{p} = \bar{y}$. The second derivative of the log-likelihood with respect to p is

$$\ell''(p; y_1, \ldots, y_n) = -n \times \left[\frac{\bar{y}}{p^2} + \frac{1 - \bar{y}}{(1-p)^2} \right].$$

Compute an estimate of the Fisher Information for one observation $I_0(p)$ by computing $-\ell''(\hat{p}; y_1, \ldots, y_n)/n$. What is the limiting value of this estimate? What is the limiting distribution of $\sqrt{n}\,(\hat{p} - p)$?

b) A random variable K follows a Negative-Binomial(r, p) distribution if K counts the number successes are observed before the r-th failure in independent and sequential Bernoulli(p) trials. The random variable K can take any non-negative integer value and r is a fixed positive integer. This is often used to model the number of times a certain type of machine will function properly in sequential tasks before you might want to think about replacing or repairing it. The mass function for a Negative-Binomial(r, p) random variable is

$$f(k; p) = \binom{r + k - 1}{k} (1 - p)^r p^k.$$

The parameter p must be between 0 and 1. The expectation of K is $\mathrm{E}[K] = pr/(1 - p)$. Let $K_1, \ldots, K_n \overset{iid}{\sim}$ Negative-Binomial(r, p) with realizations $k_1, \ldots, k_n$. The log-likelihood for p is

$$\ell(p; k_1, \ldots, k_n) = \sum_{i=1}^{n} \log\left(\binom{r + k_i - 1}{k_i} \right) + n \times \left[\bar{k} \log(p) + r \log(1 - p) \right]$$

and the MLE is $\hat{p} = \bar{k}/(r + \bar{k})$. The second derivative of the log-likelihood with respect to p is

$$\ell''(p; k_1, \ldots, k_n) = -n \times \left[\frac{\bar{k}}{p^2} + \frac{r}{(1-p)^2} \right].$$

Compute an estimate of the Fisher Information for one observation $I_0(p)$ by computing $-\ell''(\hat{p}; k_1, \ldots, k_n)/n$. What is the limiting value of this estimate? What is the limiting distribution of $\sqrt{n}\,(\hat{p} - p)$?

c) A random variable $X > 0$ follows an Exponential(λ) distribution if it is memoryless, which means that $P(X > t + s | X > t) = P(X > s)$. As a waiting time to an event, this means that if you have already waited for t minutes then the probably of waiting a further s minutes is the same as the probability of waiting s minutes if you are just starting to wait. It is often used as a simple first model for waiting times for things like internet queues. The density function for an Exponential(λ) random variable is

$$f(x; \lambda) = \lambda \exp(-\lambda x).$$

The parameter λ must be positive. The expectation of X is $\mathrm{E}[X] = 1/\lambda$. Let $X_1, \ldots, X_n \overset{iid}{\sim}$ Exponential(λ) with realizations $x_1, \ldots, x_n$. The log-likelihood for λ is

$$\ell(\lambda; x_1, \ldots, x_n) = n \times [\log(\lambda) - \lambda \bar{x}]$$

and the MLE is $\hat{\lambda} = 1/\bar{x}$. The second derivative of the log-likelihood with respect to λ is

$$\ell''(\lambda; x_1, \ldots, x_n) = -n \times \left[\frac{1}{\lambda^2}\right].$$

Compute an estimate of the Fisher Information for one observation $I_0(\lambda)$ by computing $-\ell''\left(\hat{\lambda}; x_1, \ldots, x_n\right)/n$. What is the limiting value of this estimate? What is the limiting distribution of $\sqrt{n}\left(\hat{\lambda} - \lambda\right)$?

d) A random variable $G > 0$ follows a Gamma(m, λ) distribution if G is the total amount of waiting time for m independent and sequential events to occur where the waiting time for each event is distributed as Exponential(λ). An example would be something like the total amount of time for m customers to get through a queue. The density function for an Gamma(m, λ) random variable is

$$f(g; \lambda) = \frac{\lambda^m}{(m-1)!} g^{m-1} \exp(-\lambda g).$$

The parameter λ must be positive and m is a fixed positive integer. The expectation of G is $\mathrm{E}[G] = m/\lambda$. Let $G_1, \ldots, G_n \overset{iid}{\sim}$ Gamma(m, λ) with realizations $g_1, \ldots, g_n$. The log-likelihood for λ is

$$\ell(\lambda; g_1, \ldots, g_n) = (m-1)\sum_{i=1}^{n} \log(g_i) + -n\log((m-1)!) + n \times [m\log(\lambda) - \lambda\bar{g}]$$

and the MLE is $\hat{\lambda} = m/\bar{g}$. The second derivative of the log-likelihood with respect to λ is

$$\ell''(\lambda; g_1, \ldots, g_n) = -n \times \left[\frac{m}{\lambda^2}\right].$$

Compute an estimate of the Fisher Information for one observation $I_0(\lambda)$ by computing $-\ell''\left(\hat{\lambda}; g_1, \ldots, g_n\right)/n$. What is the limiting value of this estimate? What is the limiting distribution of $\sqrt{n}\left(\hat{\lambda} - \lambda\right)$?

e) A random variable C follows a Poisson(λ) distribution if C counts the number of sequential events that are observed in a fixed window of time if the waiting times between events are independent and follow an Exponential(λ) distribution. This is often used as a basic model for the number of customers that get through a queue in an hour. The random variable C can take any non-negative integer value. The mass function for a Poisson(λ) random variable is

$$f(c; \lambda) = \frac{\lambda^c}{c!}\exp(-\lambda).$$

The parameter λ must be positive. The expectation of C is $\mathrm{E}[C] = \lambda$. Let $C_1, \ldots, C_n \overset{iid}{\sim}$ Poisson(λ) with realizations $c_1, \ldots, c_n$. The log-likelihood for λ is

$$\ell(\lambda; c_1, \ldots, c_n) = -\sum_{i=1}^{n} \log(c_i!) + n \times [\log(\lambda)\bar{c} - \lambda]$$

and the MLE is $\hat{\lambda} = \bar{c}$. The second derivative of the log-likelihood with respect to λ is

$$\ell''(\lambda; c_1, \ldots, c_n) = -n \times \left[\frac{\bar{c}}{\lambda^2}\right].$$

Compute an estimate of the Fisher Information for one observation $I_0(\lambda)$ by computing $-\ell''\left(\hat{\lambda}; c_1, \ldots, c_n\right)/n$. What is the limiting value of this estimate? What is the limiting distribution of $\sqrt{n}\left(\hat{\lambda} - \lambda\right)$?