M5MS04 - Computational Statistics; Coursework 2-Autumn 2018

Question 1

This question requested an implementation of Newton's method that found quantiles for the Cauchy distribution. The probability density function (pdf) and cumulative distribution function (cdf) for the centred Cauchy distribution with unit scale parameter are, respectively:

$$F'(x) = \frac{1}{\pi}\frac{1}{1+x^2}, x \in \mathbb{R} \tag{1}$$

$$F(x) = \frac{1}{\pi}\arctan(x) + \frac{1}{2} \tag{2}$$

A plot of the Cauchy cdf is shown in Figure 1. The Cauchy distribution is convenient because its quantile function has a closed-form expression[1],

$$F^{-1}(p) = \tan\left(\pi(p - \frac{1}{2})\right) \tag{3}$$

This makes it straightforward to check whether a quantile proposed by the root-finding procedure is accurate.

The chosen functional iteration was:

$$x_n = f(x_{n-1}) = x_{n-1} - \frac{F(x_{n-1}) - \alpha}{F'(x_{n-1})}, \tag{4}$$

This is implemented on the following page in the function `CauchyQuantileNM`. `CauchyQuantileAnalytic` determines the exact solution for the quantile using the expression in Equation 3.

Several (α, x_0) pairs were tested as a first assessment of the method's performance.

```
> CauchyQuantileAnalytic <- function(alpha){
+   if (!(alpha < 1) || !(alpha > 0)){
+     stop('Alpha must be in (0, 1).')
+   }
+   return(tan(pi*(alpha - 0.5)))
+ }
> print(CauchyQuantileAnalytic(0.5))

[1] 0

> print(CauchyQuantileNM(alpha=0.5, x0=1)) # converges

[1] -9.237995e-17

> print(CauchyQuantileNM(alpha=0.5, x0=10)) # does not converge

[1] NaN

> print(CauchyQuantileAnalytic(1e-4))

[1] -3183.099

> print(CauchyQuantileNM(alpha=1e-4, x0=1)) # converges

[1] -3183.099

> print(CauchyQuantileNM(alpha=1e-4, x0=100)) # does not converge

[1] NaN

> print(CauchyQuantileNM(alpha=1e-4, x0=-100)) # converges
```

[1] This only corresponds to a quantile because the Cauchy cdf is strictly increasing.

```
[1] -3183.099
```

```
> print(CauchyQuantileNM(alpha=1 - 1e-4, x0=100)) # converges
```

```
[1] 3183.099
```

```
> print(CauchyQuantileNM(alpha=1 - 1e-4, x0=-100)) # does not converge
```

```
[1] NaN
```

I think that the domain of convergence for the method can be determined by finding the smallest absolute value of x_0 for which $|x_1| > |x_0|$. Unfortunately I did not have time to include this in the report.

```
> CauchyQuantileNM <- function(alpha, x0, tol=1e-5){
+   # x0 is initial value for Newton's method
+   # tol is an absolute convergence criterion in x
+
+   h <- function(x, alpha){ # Right-hand term in Newton's method
+         g      <- pcauchy(x) - alpha
+         g_dash <- dcauchy(x)
+         return(g/g_dash)
+   }
+
+   x     <- x0
+   del_x <- 1e32
+
+   while(abs(del_x) > tol){
+     del_x <- -h(x, alpha)
+     x     <- x + del_x
+
+     if(is.na(del_x)){ # Return NaN if the method diverges
+       # stop('Newton\'s method diverged. Try a different starting value.')
+       break
+     }
+   }
+   return(x)
+ }
```

Question 2

This question involved

- deriving an expression for the maximum likelihood estimate (MLE) of a parameter given a particular p.d.f.,
- verifying this expression by optimizing the log-likelihood directly,
- computing an estimate of the MLE's standard error under a Normal approximation,
- stating an appropriate confidence interval for α under this approximation,
- comparing the estimated standard error to the actual standard deviation of the MLE across simulated datasets.

The key outcome of this question was that the sample size was too small for the Gaussian approximation to be usable.

The pdf given in the question was

$$f(x) = \frac{\alpha}{2}\left(\frac{x}{2}\right)^{\alpha-1} : 0 < x < 2 \tag{5}$$

where $\alpha > 0$.

The observations are assumed to be independent, therefore the likelihood of the data under this probability density function is:

$$P(x_1, ..., x_n|\alpha) = \prod_{i=1}^{n} \frac{\alpha}{2}\left(\frac{x_i}{2}\right)^{\alpha-1} \tag{6}$$

such that the corresponding log-likelihood is

$$\mathcal{L}(X, \alpha) = n \log \frac{\alpha}{2} + (\alpha - 1) \sum_{i=1}^{n} \log\left(\frac{x_i}{2}\right) \tag{7}$$

Differentiating with respect to α, we obtain

$$\frac{\partial \mathcal{L}}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n} \log\left(\frac{x_i}{2}\right) \tag{8}$$

Denote the MLE as $\hat{\alpha}$, equate the derivative of the log-likelihood to zero, and solve for $\hat{\alpha}$:

$$\hat{\alpha} = -\sum_{i=1}^{n} \frac{n}{\log\left(\frac{x_i}{2}\right)} \tag{9}$$

Computing the MLE's value for this dataset by both Equation 9 and by directly optimizing our expression for the log-likelihood (Equation 7), we find the values agree with one another.

```
>   x <- c(1.30, 1.33, 1.75, 0.19, 1.13, 1.17, 0.68, 1.32, 1.89, 1.52)
>   n <- length(x)
>   getMLEalpha <- function(x){-n/sum(log(x/2))}
>   MLE_alpha_cf <- getMLEalpha(x)
>   print(MLE_alpha_cf)
```

```
[1] 1.59781
```

```
>   negative_log_likelihood <- function(alpha){
+     return(-sum(log((alpha/2)*(x/2)^(alpha -1) )))
+   }
>   MLE_alpha_numeric <- optim(par=20, fn=negative_log_likelihood)
>   MLE_alpha_numeric <- MLE_alpha_numeric$par
>   print(MLE_alpha_numeric)
```

```
[1] 1.597656
```

To construct an asymptotic confidence interval for α, we exploit the fact that the distribution of the MLE will converge on a Normal distribution as the sample size becomes larger[2]. This is a corollary of the Central Limit Theorem and may be written

$$\hat{\alpha}_n - \alpha \xrightarrow{d} \mathcal{N}\Big(0, \frac{1}{n} I(\alpha)^{-1}\Big) \quad \text{as } n \to \infty \tag{10}$$

To use this approximation, an estimate of the Fisher information I is needed. We use a sample mean as described on page 15 of the notes, which is a consistent estimator for I:

$$\hat{I}(\hat{\alpha}) = -\frac{1}{n}\frac{\partial^2}{\partial \alpha^2} \sum_{i=1}^{n} \log f_\alpha(x_i)\Big|_{\alpha=\hat{\alpha}} \tag{11}$$

$$= -\frac{1}{n}\frac{\partial^2}{\partial \alpha^2} \mathcal{L}(X, \hat{\alpha}) \tag{12}$$

$$= -\frac{1}{n}\frac{\partial}{\partial \alpha}\Big(\frac{n}{\hat{\alpha}} + \sum_{i=1}^{n} \log\big(\frac{x_i}{2}\big)\Big) \tag{13}$$

$$= \frac{1}{\hat{\alpha}^2} \tag{14}$$

We use Equation 9 to derive an expression for the confidence interval. From this approximation it follows that

$$P\Big(-z_{\beta/2} < \frac{\sqrt{n}(\hat{\alpha}_n - \alpha)}{\sqrt{I^{-1}}} < z_{\beta/2}\Big) = \beta \tag{15}$$

Rearranging the inequality, it's possible to arrive at an asymptotic $1-\beta$ confidence interval for α:

$$\Big(\hat{\alpha} - \frac{\hat{\alpha}}{\sqrt{n}} z_{\beta/2}, \hat{\alpha} + \frac{\hat{\alpha}}{\sqrt{n}} z_{\beta/2}\Big) \tag{16}$$

where $z_{\beta/2}$ denotes the standard Normal quantile associated with level $1 - \frac{\beta}{2}$. Ignoring the uncertainty associated with our estimate of the Fisher information, this confidence interval will contain α with probability $1-\beta$. Setting $\hat{\alpha} = 1.598...$, $n = 10$, and $\beta = 0.05 : z_{\beta/2} = 1.96...$, we find that the symmetric 95% confidence interval is

$$(0.607, 2.588) \tag{17}$$

How useful this confidence is depends upon how accurately $\mathcal{N}(0, \frac{1}{n}I(\alpha)^{-1})$ approximates the distribution of $\hat{\alpha} - \alpha$. To investigate this, we evaluate our estimate of the variance of $\hat{\alpha}$. We generate 1000 data sets with $\alpha = 2$, compute their MLEs in closed form, compute the sample variance across these MLEs, then compare this variance to the value given by

$$\mathrm{Var}(\hat{\alpha}) = \frac{1}{n}\hat{I}(\hat{\alpha})^{-1} = \frac{\hat{\alpha}^2}{n} \tag{18}$$

As can be seen in Figure 2, the distribution of $\hat{\alpha}_n - \alpha$ for $n = 10$ is long-tailed and asymmetric. The assymmetry is a consequence of the MLE having high variance and the $\alpha > 0$ boundary. It is therefore clear that for these hyperparameter values of n and α, the Gaussian approximation is inaccurate. This suggests that the closed-form estimate of the MLE's standard error given by Equation 17 will inaccurately approximate the variance obtained by simulation. This is borne out by our results:

[2]We need to interpret this approximation carefully since $\alpha > 0$ but the approximation will assign non-zero probabilities to α values less than or equal to 0.

```
>   print(alpha_variance) # True variance of the parameter estimator
[1] 0.6706709
>   print(mean(mle_alpha^2/n)) # Approximation assuming normality
[1] 0.5565809
```

The limit theorem states that as n increases the distribution of $\hat{\alpha}$ becomes Gaussian, implying that our estimate of the standard error should improve as it does so. Repeating the exact same simulation, but changing the sample size to $n = 1000$, we find that this is the case. Figure 3 illustrates that $\hat{\alpha}$'s distribution becomes a shapely Gaussian, while the values printed below highlight that the variance approximation becomes much more accurate.

```
>   print(alpha_variance) # True variance of the parameter estimator
[1] 0.004008547
>   print(mean(mle_alpha^2/n)) # Approximation assuming normality
[1] 0.004012156
```

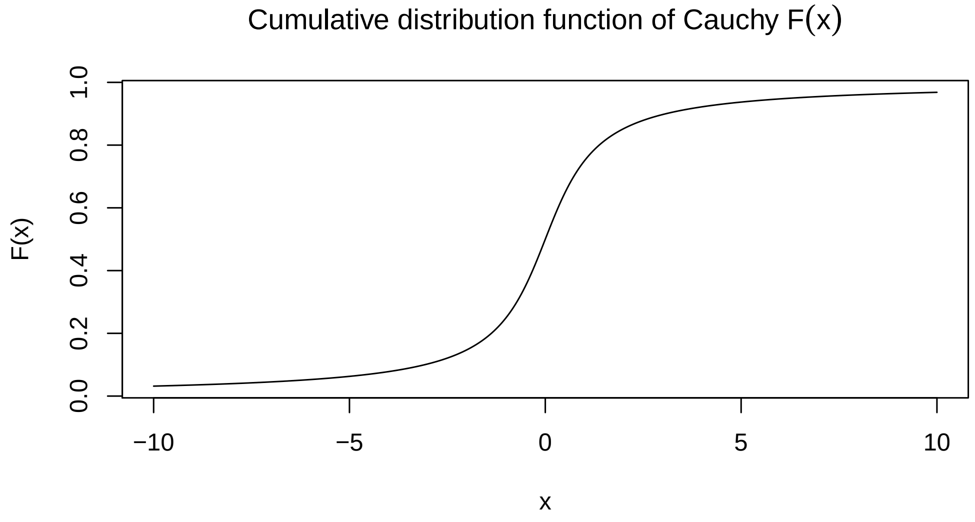

Figure 1: The cumulative distribution function of the Cauchy distribution.

```
>   N     <- 50000
>   n     <- 10
>   alpha <- 2
>   mle_alpha <- c()
>   getsample <- function(n, alpha) 2*runif(n)^(1/alpha) # Generates datasets
>   for (i in 1:N){
+     mle_alpha <- c( mle_alpha, getMLEalpha(getsample(n, alpha)) )
+   }
>   alpha_variance <- sum((mle_alpha - alpha)^2)/N
>   h1 <- hist(mle_alpha, breaks=100, border='white', col='darkblue',
+              main=expression(paste('Frequency distribution of ', hat(alpha),
+                                    ' for n=10')),
+              xlab=expression(paste('Maximum likelihood estimate of ', alpha,
+                                    ', ', hat(alpha))))
```

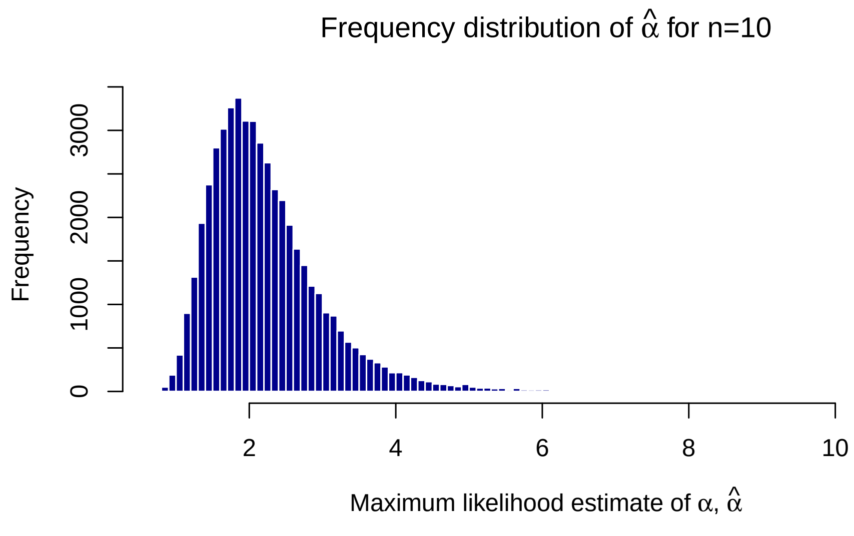

Figure 2: Distribution of the MLE $\hat{\alpha_n}$ for $n = 10$, $\alpha = 2$ (50 000 samples).

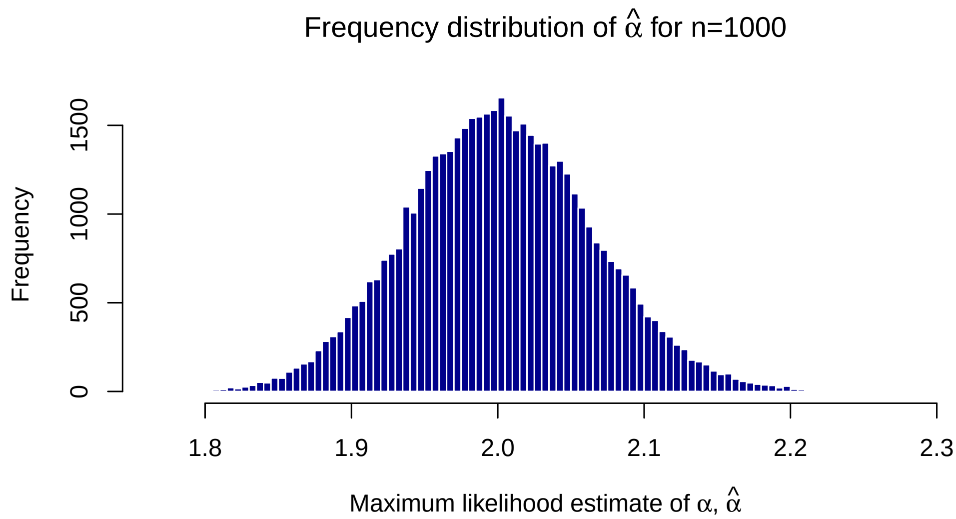

Figure 3: Distribution of the MLE $\hat{\alpha_n}$ for $n = 1000$, $\alpha = 2$ (50 000 samples).