THE UNIVERSITY OF SYDNEY
SCHOOL OF MATHEMATICS AND STATISTICS

Computer Exercise Week 6

STAT3023: Statistical Inference — Semester 2, 2021

Lecturers: Neville Weber and Michael Stewart

Suppose $\mathbf{X} = (X_1, \ldots, X_n)$ is a vector of iid RVs with common PDF $f_\theta(\cdot)$ where

$$f_\theta(x) = \frac{1}{\theta} g\left(\frac{x}{\theta}\right)$$

for a known PDF $g(\cdot)$ which possesses a continuous derivative. The family $\mathcal{F} = \{f_\theta(\cdot) \colon \theta > 0\}$ is thus a *scale family* and θ is a *scale parameter*, like the standard deviation in the normal family. The Cramér-Rao Lower Bound for variance of an unbiased estimator of θ in such a family based on n iid observations is given by

$$\frac{1}{nI_\theta} \quad \text{where} \quad I_\theta = \frac{J(g)-1}{\theta^2} \quad \text{and} \quad J(g) = \int \frac{[xg'(x)]^2}{g(x)}\, dx \tag{1}$$

(this is verified in the Advanced Workshop).

We shall study what happens when

$$g(x) = \frac{1}{\pi(1+x^2)}$$

is the Cauchy density (same as Student's-t with 1 degree of freedom, is also the density of the ratio of two independent $N(0,1)$ random variables); note that the quartiles of $g(\cdot)$ are ± 1, and also that neither the mean nor the variance exist! We shall consider two estimators of θ based on $\mathbf{X}$:

- $\hat{\theta}_{\text{IQR}}(\mathbf{X}) = \frac{\text{IQR}(\mathbf{X})}{2}$;
- $\hat{\theta}_{\text{MLE}}(\mathbf{X})$, the maximum likelihood estimator (obtained *numerically* using R);

1. Generate a sample of size `n=100` from the Cauchy distribution with `scale=0.75` (see the help for `rcauchy()`). Print its `summary()` and draw a (horizontal) boxplot (try setting `fig.height=3` in the Rmarkdown chunk options). You might also want to use `set.seed()` to make sure you get the same generated sample each time you "knit".

2. Compute the value of the estimator $\hat{\theta}_{\text{IQR}}$ for your generated sample using the `IQR()` function. Save it as `th.hat.IQR`.

3. We shall find the maximum likelihood estimate numerically. First, write a function that computes the Cauchy scale log-likehood:

```
cauchy.scale.logL=function(th,x) {
    sum(dcauchy(x,scale=th,log=T))
}
```

We verify in the Advanced Workshop that $\min_i |X_i| \leq \hat{\theta}_{\text{MLE}} \leq \max_i |X_i|$. Create a vector `log.th` of length 100 equally spaced values between the `log()` of these two extreme possibilities, then write a loop to create a vector `logL`, so that `logL[i]` contains the log-likelihood for $\theta =$ `exp(log.th[i])`. Finally, plot `logL` versus `log.th` and add a vertical blue line indicating the value of `log(th.hat.IQR)`. **Why are we plotting against $\log(\theta)$ instead of θ?**

4. The R function `optimise()` can find the maximum (or minimum) value of a function as well as the corresponding maximiser (or minimiser). Execute the command

```r
optimise(cauchy.scale.logL,lower=min(abs(x)),upper=max(abs(x)),x=x,maximum=T)
```

```
## $maximum
## [1] 0.6617755
##
## $objective
## [1] -215.1418
```

Write a function which takes as argument a vector `x` and then outputs the mle of the Cauchy scale parameter. Call this function `cauchy.scale.mle`. Use it to obtain the value of $\hat{\theta}_{\text{MLE}}$; call it `th.hat.MLE`. Recreate the plot above and add to it a vertical red line indicating the value of `log(th.hat.MLE)`.

5. We now perform a loop to see how the two estimators perform over a range of θ values. Following the same basic method as in week 5, at each value of `th=(1:39)/40`, generate `N=10000` (you may reduce this to `N=1000` if it is taking too long) Cauchy samples of size `n=100` with that scale parameter, saving the estimates in `th.hat.IQR` and `th.hat.MLE` respectively and save the average squared errors in the corresponding elements of `MSE.IQR` and `MSE.MLE` respectively. Finally, plot `MSE.IQR` against `th` as a blue curve, and also add a plot of `MSE.MLE` against `th` as a red curve.

6. We are going to see how these curves compare with theory.

(a) We verify in the Advanced Workshop that for a sample of size n from a continuous density $f(\cdot)$, symmetric about zero with quartiles at $\pm q$, the sample IQR is $AN\left(2q, \frac{1}{4nf(q)^2}\right)$. Use this to deduce an (approximate) MSE for $\hat{\theta}_{\text{IQR}}$ and add this as a dotted line to the plot.

(b) According to our theory for "nice" models, the maximum likelihood estimator is AMVU, which in this case means

$$\hat{\theta}_{\text{MLE}}(\mathbf{X}) \sim AN\left(\theta, \frac{1}{nI_\theta}\right),$$

where I_θ is given at (1) above. We can also *numerically* determine the value of the integral $J(g)$ in this case using the R function `integrate()`:

(i) Write an appropriate function of the form

```r
integrand=function(x) { ... }
```

defining the integrand of the integral $J(g)$.

(ii) Then, use

```r
J=integrate(integrand,lower=-Inf,upper=Inf)$value
```

to numerically approximate the integral.

(iii) Use the value of `J` computed above and the function `lines(...,lty=2)` to add a plot of the asymptotic variance against `th` to the MSE plots.

(c) Add an appropriate legend, axis labels and heading to the graph (**hint:** try `xlab=expression(theta)`).

7. Comment on the plot.