Math 6262: Statistical Estimation
Homework 2

Due: **Tue 12/03**

1^o: *Estimating the support of* $\mathsf{Uniform}([0,\theta])$.

Consider the task of estimating the (right boundary of the) support of the uniform distribution $\mathbb{P}_\theta = \mathsf{Uniform}([0,\theta])$—i.e. one with p.d.f. $f(x|\theta) = \frac{1}{\theta}\mathbb{1}\{x \in [0,\theta]\}$ for some $\theta > 0$—from an i.i.d. sample $X_{1:n} := (X_1, X_2, ..., X_n)$.

(a) Compute the Fisher information $\mathbb{I}(\theta^*)$ as an explicit function of true parameter value θ^*. (Note that since $\theta \in \mathbb{R}$, $\mathbb{I}(\theta^*)$ is a 1-by-1 matrix, i.e. just a number.) Write down the resulting Cramér-Rao bound (CRB). Pinpoint the assumption whose violation invalidates the CRB.

(b) Our next goal is to show that CRB is indeed "violated" in this statistical model, by exhibiting an estimator with a smaller asymptotic variance.

- Find the MLE $\hat{\theta}_n$ as a closed-form function of $X_{1:n}$.
- Compute the bias of $\hat{\theta}_n$, and adjust $\hat{\theta}_n$ to get an unbiased estimator $\tilde{\theta}_n$.
- Computing the variance of $\tilde{\theta}_n$ and compare it with the CRB.

2^o: *Estimating the median in a location family.*

For any $\theta \in \mathbb{R}$, let $\mathbb{P}_\theta$ be the distribution of $\theta + Z$, where the distribution of Z is absolutely continuous, with (known) p.d.f. $f(\cdot)$ such that $f(0) > 0$, and such that 0 is the median:

$$\mathbb{P}\{Z < 0\} = \mathbb{P}\{Z > 0\} = \frac{1}{2}.$$

The k^{th} *order statistic* of $X_{1:n} := (X_1, ..., X_n)$, denoted with $X_{(k)}$, is defined as the k-th largest among $X_{1:n}$.[1] Assuming that n is odd, we define the *sample median* $\widehat{\text{Med}}_n$ of $X_{1:n}$ as the $(\frac{n+1}{2})^{\text{nd}}$ order statistic. Note that $\widehat{\text{Med}}_n$ is a reasonable candidate estimate for θ^*. (Could you explain why?)

(a) Show that $\widehat{\text{Med}}_n$ is unbiased when f is symmetric, i.e. when $f(-x) = f(x)$ for all $x \in \mathbb{R}$.

Hint. *Use the tower rule:* $\mathbb{E}[\widehat{\text{Med}}_n] = \mathbb{E}[\mathbb{E}[\widehat{\text{Med}}_n|Y]]$ *for any random variable* Y. *Try to find a suitable random variable* Y*—supported on* $\{1, ...n\}$*—for which* $\mathbb{E}[\widehat{\text{Med}}_n|Y] = 0$ *almost surely.*

(b) Show that $\widehat{\text{Med}}_n$ is the MLE when Z has the standard Laplace distribution: $f(u) = \frac{1}{2}e^{-|u|}$.
Hint: *find the derivative of* $\ell(u) := \log f(u)$*? Does it matter that* ℓ *is not differentiable at* 0*?*

[1] Ties broken arbitrarily; note, however, that with probability 1 all X_i's are different, and so ties do not happen.

(c′) Compute the variance of $\widehat{\mathrm{Med}}_n$ in this situation (i.e. Laplace distribution) in the cases $n = 1$ and $n = 3$. Compare with the Cramér-Rao bound.

Hint. *Use the following fact: if $X_{1:n}$ is an i.i.d. sample from a law with c.d.f. $F_X(x)$, then the c.d.f. of its k^{th} order statistic is*

$$F_{X_{(k)}}(x) = \sum_{j=0}^{k-1} \binom{n}{j} F_X(x)^{n-j}(1 - F_X(x))^j$$

Another useful fact is that $\widehat{\mathrm{Med}}_n$ is shift-equivariant*: if we shift the distribution by a constant $a \in \mathbb{R}$, then the distribution of $\widehat{\mathrm{Med}}_n$ will simply be shifted in the same way. This allows to assume w.l.o.g. that $\mu_* = 0$ in your calculations.*

(d) Find the *asymptotic variance* of $\widehat{\mathrm{Med}}_n$ in the general situation, i.e. only assuming that $f(0) > 0$. To this end, use the so-called "delta-method," described as follows:

If $\widehat{\theta}_n$ estimates $\theta \in \mathbb{R}$ in such a way that

$$\sqrt{n}(\widehat{\theta}_n - \theta) \underset{n\to\infty}{\rightsquigarrow} \mathcal{N}(0, \sigma^2),$$

and $g(\cdot)$ is differentiable at θ, then $\sqrt{n}[g(\widehat{\theta}_n) - g(\theta)] \underset{n\to\infty}{\rightsquigarrow} \mathcal{N}(0, \sigma_g^2)$ with $\sigma_g^2 = \sigma^2(g'(\theta))^2$.

What you can say about this in the light of the previous example (with the Laplacian density)?

(e) Give another example of a symmetric $f(u)$ such that $\widehat{\mathrm{Med}}_n$ does not attain the CRB in the corresponding location family.

3^o: *MLE for the ratio of two independent exponential distributions.*
Let $X \sim \mathsf{Exp}(\lambda)$ and $Y \sim \mathsf{Exp}(\mu)$ be independent, where $\mathsf{Exp}(\lambda)$ is the distribution with p.d.f.

$$\lambda e^{-\lambda x}, \quad x > 0.$$

Let $Z = X/Y$, and consider an i.i.d. sample $(Z_1, ..., Z_n)$ with each Z_i being distributed as Z.

(a) Without deriving the distribution of Z explicitly, argue that it depends only on $\theta := \mu/\lambda$, not on λ, μ separately.

Hint: for $\alpha > 0$, what is the distribution of αX?

(b) Show that the p.d.f. of Z is

$$f(z|\theta) = \frac{\theta}{(z+\theta)^2}, \quad z > 0.$$

Does this distribution have an expectation?

Hint: *you might want to start with the c.d.f.*

(c) Show that $\widehat{\theta}_n$, the MLE of θ from $Z_1, ..., Z_n$, satisfies the following equation:

$$\frac{1}{n}\sum_{i=1}^{n} F(Z_i|\widehat{\theta}_n) = \frac{1}{2} \tag{1}$$

where $F(\cdot|\theta)$ is the c.d.f. of Z. Using the properties of c.d.f., show that (1) has a *unique* solution.

(d) Comment on the above equation, explaining why the right-hand side has the factor $\frac{1}{2}$. To this end, show that for any continuous distribution $\mathbb{P}_\theta$ with c.d.f. $H(t;\theta)$, it holds that

$$\mathbb{E}_{T\sim\mathbb{P}_\theta}[H(T;\theta)] = \frac{1}{2}.$$

Then explain the MLE equation in this context.

Hint: *You may draw an analogy with the method of moments.*

(e) Show that for the distribution whose p.d.f. you found in (b), θ also happens to be the median.