Problem Set 4 (Lectures 7-8)

Problem 1 (Cramer-Rao Lower Bound for a scalar parameter, 50 points). Let $\widehat{\theta}(x)$ be the estimator of a real-valued parameter θ in the statistical model with pd.f. $f(x,\theta)$ (we will assume also that x is a scalar). Unfortunately, there is no theorem that says that Maximum Likelihood estimators will always be unbiased or that the ML estimators will achieve the lowest possible variance among unbiased estimators (if, by chance, they happen to be unbiased).

I would like you to show that in parametric models (meaning, statistical models where θ is finite-dimensional) we can provide a lower bound on the variance of *any* estimator. The lower bound we present depends on the bias. The bound will be important, as in large samples, there are theorems that guarantee that ML estimators (which are "asymptotically" unbiased) will eventually approach the lower bound on the variance.

Here is what I would like you to show:

Proposition 1 (Cramér-Rao Bound)**.** *Suppose that the estimator $\widehat{\theta}$ and the statistical model satisfy:*

$$\int_{\mathbb{R}} \Big[\widehat{\theta}(x)\frac{\partial}{\partial\theta}f(x,\theta)\Big]dx = \frac{\partial}{\partial\theta}\int_{\mathbb{R}}\Big[\widehat{\theta}(x)f(x,\theta)\Big]dx \tag{0.1}$$

and

$$\int_{\mathbb{R}} \Big[\frac{\partial}{\partial\theta}f(x,\theta)\Big]dx = \frac{\partial}{\partial\theta}\int_{\mathbb{R}}\Big[f(x,\theta)\Big]dx = 0, \tag{0.2}$$

(both of which require that we can change the order in which we take integrals and derivatives). If these conditions are satistfied:

$$Var_{P_\theta}\Big[\widehat{\theta}(x)\Big] \geq \Big[\frac{\partial}{\partial\theta}\mathbb{E}_{P_\theta}[\widehat{\theta}(x)]\Big]^2 \Big/ Var_{P_\theta}\Big[S_\theta(x)\Big],$$

where

$$S_\theta(x) \equiv \frac{\partial}{d\theta}\ln f(x,\theta)$$

is the <u>*score*</u> *of the statistical model $\{f(x,\theta)\}_{\theta\in\Theta}$ and $Var_{P_\theta}\Big[S_\theta(x)\Big]$ is called the Fisher information of the statistical model at θ.*

I will help you a bit with the proof. Just fill in the blanks (if you can give a different proof of this result, go for it!)

Proof. (**5 points for each box**). The covariance between any estimator $\widehat{\theta}$ and the score (which is a random variable) is

$$\begin{aligned}
\mathbb{E}_\theta\left[\widehat{\theta}(x) S_\theta(x)\right] &= \int_{\mathbb{R}} \widehat{\theta}(x) \frac{\partial}{d\theta} \ln f(x,\theta) f(x,\theta) dx \\
&= \int_{\mathbb{R}} \widehat{\theta}(x) \frac{\partial}{d\theta} f(x,\theta) dx \\
&= \boxed{\quad \int_{\mathbb{R}} \qquad} \\
&\quad \text{(where we have used Equation 0.1)} \\
&= \frac{\partial}{d\theta} \boxed{\qquad\qquad}.
\end{aligned}$$

where $\mathbb{E}_\theta[\widehat{\theta}(x)]$ is the bias of the estimator $\widehat{\theta}(x)$ at θ. Assumption 0.2 implies

$$\mathbb{E}_\theta[S_\theta(x)] = \boxed{\qquad\qquad},$$

which implies

$$\mathbb{E}_\theta\left[\widehat{\theta}(x) S_\theta(x)\right] = \mathbb{E}_\theta\left[\left(\widehat{\theta}(x) - \mathbb{E}_\theta[\widehat{\theta}(x)]\right) S_\theta(x)\right]$$

Hence, by the Cauchy-Scharwz inequality:[1]

$$\begin{aligned}
\mathbb{E}_\theta\left[\widehat{\theta}(x) S_\theta(x)\right]^2 &\leq \boxed{\qquad\qquad}\boxed{\qquad\qquad} \\
&= \mathrm{Var}_\theta\left[\widehat{\theta}(x)\right] \boxed{\qquad\qquad}
\end{aligned}$$

Therefore,

$$\mathrm{Var}_\theta\left[\widehat{\theta}(x)\right] \geq \frac{\partial}{d\theta} \boxed{\qquad\qquad}.$$

□

Corollary (**15 Points**) Let $\widehat{\theta}$ be any *unbiased* estimator for the mean parameter θ in the model for the data $(x_1, \ldots, x_n)$, where $x_i \sim \mathcal{N}(\theta, \sigma^2)$, i.i.d. and σ^2 is known. Show that the sample mean estimator $\widehat{\theta}(x_1, \ldots, x_n) = (1/n) \sum_{i=1}^n x_i$ maximizes the likelihood and it achieves the smallest mean

[1]For any two random variables X and Y:

$$\mathbb{E}_{\mathbb{P}}[XY] \leq \mathbb{E}_{\mathbb{P}}[X^2]^{1/2} \mathbb{E}_{\mathbb{P}}[Y^2]^{1/2}$$

See pg. 24 Durrett (2010).

squared error relative to all unbiased estimators (HINT: Use the Cramer-Rao Lower bound) .

Problem 2 (Maximum Likelihood in the Linear Regression model, 50 points) We have shown that the OLS estimator $\widehat{\beta}_{\text{OLS}}$ maximizes the likelihood for the Normal Linear Regression model, where $Y|X \sim \mathcal{N}_n(X\beta, \sigma^2 \mathbb{I}_n)$ and σ^2 is known. In this exercise I would like you to treat σ^2 as an unknown parameter, and derive the ML estimators of β and σ^2. To help you a bit to simplify the algebra, you can assume that the X's are non-stochastic.

1. (10 points) Treating σ^2 as known, derive the score $S(x, \beta)$ and the Fisher Information matrix $\mathcal{I}(\beta)$ in the Normal Linear Regression model. Show that, at a given parameter β,

$$\widehat{\beta}_{\text{ML}} = \mathcal{I}(\beta)^{-1} S(x, \beta) + \beta \sim \mathcal{N}_k(\beta, \mathcal{I}(\beta)^{-1}).$$

   I am making this connection because one of the large sample approximations that you will show for ML estimators in the next part of the course is:

$$\widehat{\theta}_{\text{ML}} = \mathcal{I}_n(\theta)^{-1} S_n(x_1, \dots, x_n, \theta) + \beta \approx \mathcal{N}_k(\theta, \mathcal{I}_n(\theta)^{-1}).$$

   This means that ML estimators will eventually achieve the Cramer-Rao Lower bound.

2. (15 points) Treating σ^2 as unknown, I would like you to derive the ML estimator for both parameters β and σ^2. I would suggest you to start by deriving the score for this model (you already have one part) and solve the F.O.C. You will note that the ML estimator for β is still the same as before. OPTIONAL: Derive the Fisher information matrix for this model. What is the distribution of $\widehat{\sigma}^2_{\text{ML}}$?

References

DURRETT, R. (2010): *Probability: Theory and Examples*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 4th ed.