5.4.1 Proof of the Cramer-Rao-Inequality

We use the following definitions from the previous section. Let $p(y_N|\theta)$ be the probability density function of obtaining the measurements y_N given the parameters θ, and let $L(\theta, y_N) := -\log p(y_N|\theta)$ be the negative log-likelihood. In addition, assume $p(y_N|\theta_0)$ for the true θ_0 is known and that $\hat{\theta}(y_N)$ is an unbiased estimator. We define the Fischer information matrix as

$$M = \mathbb{E}\{\nabla_\theta^2 L(\theta_0, y_N)\} \tag{5.30}$$

The covariance of our estimator $\hat{\theta}(y_N)$ is given by

$$C = \mathrm{cov}(\hat{\theta}) = \mathbb{E}\{(\hat{\theta}(y_N) - \theta_0)(\hat{\theta}(y_N) - \theta_0)^\top\} \tag{5.31}$$

The Cramer-Rao inequality states that

$$C \succeq M^{-1} \tag{5.32}$$

and this is what we will prove. The main idea of the proof is to define a random vector $B \in \mathbb{R}^{2d}$ as follows:

$$B(\theta_0, y_N) = \begin{bmatrix} \hat{\theta}(y_N) - \theta_0 \\ \nabla_\theta L(\theta_0, y_N) \end{bmatrix} \tag{5.33}$$

Because for any matrix or vector B holds that $BB^\top \succeq 0$, and because this positive semi-definiteness property is preserved by the expectation operator, we know that

$$0 \quad \preceq \quad \mathbb{E}\{B(\theta_0, y_N)B(\theta_0, y_N)^\top\} \quad = \quad \begin{bmatrix} C & \underbrace{\mathbb{E}\{(\hat{\theta}(y_N) - \theta_0)\nabla_\theta L(\theta_0, y_N)^\top\}}_{:=Z} \\ Z^\top & \underbrace{\mathbb{E}\{\nabla_\theta L(\theta_0, y_N)\nabla_\theta L(\theta_0, y_N)^\top\}}_{:=\tilde{M}} \end{bmatrix} =: A \tag{5.34}$$

The matrix $\tilde{M}$ is positive semi-definite by construction. But to be able to use the Schur complement lemma below, we will from now on assume that $\tilde{M}$ is strictly positive definite, which is the case for any well-posed estimation problem.[1] The Schur complement lemma states that, if $\tilde{M}$ is positive definite, then the matrix A is positive semi-definite if and only if the Schur complement of $\tilde{M}$ in A, which is given by $C - Z\tilde{M}^{-1}Z^\top$, is also positive semi-definite. The Schur complement lemma can be expressed in compact mathematical notation as follows:

$$\text{If} \quad \tilde{M} \succ 0 \quad \text{then} \quad \left(\begin{bmatrix} C & Z \\ Z^\top & \tilde{M} \end{bmatrix} \succeq 0 \quad \Longleftrightarrow \quad C - Z\tilde{M}^{-1}Z^\top \succeq 0 \right) \tag{5.35}$$

Thus, from $A \succeq 0$ follows that $C \succeq Z\tilde{M}^{-1}Z^\top$. This would be equivalent to the desired Cramer-Rao inequality (5.32) if we could show that $Z\tilde{M}^{-1}Z^\top = M^{-1}$. We will do this in two steps, first computing $\tilde{M}$ and second computing Z. Both steps use the following basic fact from calculus:

$$\frac{\mathrm{d}\log f(x)}{\mathrm{d}x} = \frac{1}{f(x)} \cdot \frac{\mathrm{d}f(x)}{\mathrm{d}x} \quad \text{which allows us to express any derivative as} \quad \frac{\mathrm{d}f(x)}{\mathrm{d}x} = f(x) \cdot \frac{\mathrm{d}\log f(x)}{\mathrm{d}x} \tag{5.36}$$

First, we show that $\tilde{M} = M$, starting from the trivial statement that $1 = \mathbb{E}\{1\} = \int p(y_N|\theta_0)\mathrm{d}y_N$. Differentiating left and right side of this statement with respect to θ_0 leads to the following:

$$\begin{aligned} 0 &= \frac{\mathrm{d}}{\mathrm{d}\theta_0} \int p(y_N|\theta_0)\, \mathrm{d}y_N = \int \frac{\mathrm{d}p(y_N|\theta_0)}{\mathrm{d}\theta_0} \mathrm{d}y_N \\ &= \int p(y_N|\theta_0) \frac{\mathrm{d}\log p(y_N|\theta_0)}{\mathrm{d}\theta_0} \mathrm{d}y_N = \mathbb{E}\{-\nabla_\theta L(\theta_0, y_N)^\top\} \end{aligned} \tag{5.37}$$

[1] If $\tilde{M}$ would be singular, one could imagine adding a small epsilon to all diagonal entries of it to ensure positive definiteness, then apply the same matrix operations, and finally take the limit as epsilon goes to zero. The result would be a Cramer-Rao lower bound that tends to infinity in the singular directions, reflecting the ill-posedness of the estimation problem.

We differentiate the negative transpose of the statement, $0 = \mathbb{E}\{\nabla_\theta L(\theta_0, y_N)\}$, a second time with respect to θ_0 and obtain the following

$$\begin{aligned} 0 &= \frac{\mathrm{d}}{\mathrm{d}\theta_0} \int p(y_N|\theta_0) \cdot \nabla_\theta L(\theta_0, y_N))\ \mathrm{d}y_N \\ &= \int p(y_N|\theta_0) \nabla_\theta^2 L(\theta_0, y_N) \mathrm{d}y_N + \int p(y_N|\theta_0) \nabla_\theta L(\theta_0, y_N) (-\nabla_\theta L(\theta_0, y_N))^\top \mathrm{d}y_N \\ &= \underbrace{\mathbb{E}\{\nabla_\theta^2 L(\theta_0, y_N)\}}_{=M} - \underbrace{\mathbb{E}\{\nabla_\theta L(\theta_0, y_N) \nabla_\theta L(\theta_0, y_N)^\top\}}_{=\tilde{M}} = M - \tilde{M} \end{aligned} \tag{5.38}$$

Thus, we have shown that $\tilde{M} = M$, i.e., that there are two equivalent ways to define the Fischer information matrix. Second, we will compute Z, in order to close the remaining gap in the proof. We start by noting that

$$\begin{aligned} Z &= \mathbb{E}\{(\hat{\theta}(y_N) - \theta_0) \nabla_\theta L(\theta_0, y_N)^\top\} \\ &= \mathbb{E}\{\hat{\theta}(y_N) \nabla_\theta L(\theta_0, y_N)^\top\} - \theta_0 \underbrace{\mathbb{E}\{\nabla_\theta L(\theta_0, y_N)^\top\}}_{=0} = \mathbb{E}\{\hat{\theta}(y_N) \nabla_\theta L(\theta_0, y_N)^\top\} \end{aligned} \tag{5.39}$$

Now, we will finally make use of the crucial assumption that the estimator is unbiased, which can be stated as

$$\theta_0 = \mathbb{E}\{\hat{\theta}(y_N)\} = \int \hat{\theta}(y_N) \cdot p(y_N|\theta_0)\ \mathrm{d}y_N \tag{5.40}$$

Differentiating both sides of this statement with respect to θ_0 yields

$$\begin{aligned} \mathbb{I} &= \int \hat{\theta}(y_N) \cdot p(y_N|\theta_0) \cdot (-\nabla_\theta L(\theta_0, y_N))^\top \mathrm{d}y_N \\ &= \mathbb{E}\{\hat{\theta}(y_N)(-\nabla_\theta L(\theta_0, y_N))^\top\} = -Z \end{aligned} \tag{5.41}$$

Thus, we have shown that $Z = -\mathbb{I}$, and with $\tilde{M} = M$ we can finally evaluate $Z\tilde{M}^{-1}Z^\top = (-\mathbb{I})\, M^{-1}(-\mathbb{I}) = M^{-1}$, proving the Cramer-Rao inequality (5.32). $\square$