TP5: Fisher Information

Francesco Saverio Pezzicoli[*], Guillaume Charpiat

March, 2023

Credits: Vincenzo Schimmenti

Recall that for a distribution $p(x|\theta)$ with parameters $\theta = \{\theta_1, \ldots, \theta_n\}$ the Fisher information matrix reads:

$$I_{i,j}(\theta) = -\mathbb{E}\left[\frac{\partial^2 \log p(x|\theta)}{\partial\theta_i \partial\theta_j}\right] \tag{1}$$

where the expected value is taken w.r.t. to $p(x|\theta)$, keeping θ fixed. The Cramer-Rao bound for an unbiased estimator $\hat{\theta}$ of a parameter θ is:

$$\mathrm{Var}(\hat{\theta}) \geqslant \frac{1}{NI(\theta)} \tag{2}$$

where N is the number of i.i.d. samples in the estimator. Recall that un biased estimator $\hat{\theta}$ is such that:

$$\mathbb{E}[\hat{\theta} - \theta] = 0 \tag{3}$$

where θ is the true parameter.

Example: Exponential distribution $p(x) = \lambda e^{-\lambda x}$. The log probablity is:

$$-\log p(x|\lambda) = -\log\lambda + \lambda x$$

By taking two time the derivatives w.r.t. λ we get:

$$-\partial_\lambda^2 \log p(x|\lambda) = \frac{1}{\lambda^2}$$

Since the second derivative does not depend on x:

$$\mathbb{E}\left[-\partial_\lambda^2 \log p(x|\lambda)\right] = \frac{1}{\lambda^2}$$

Hence the Fisher information reads:

$$I(\lambda) = \frac{1}{\lambda^2}$$

[*]francesco.pezzicoli@universite-paris-saclay.fr

If we want rephrase the Fisher information using the average parameter of the exponential distribution $\mu = 1/\lambda$ we need to use the change of parameter property of the Fisher information, namely:

$$I(\eta) = I(\theta(\eta)) \left(\frac{d\theta}{d\eta}\right)^2$$

By choosing $\theta = \lambda$ and $\eta = \mu$ we have:

$$\frac{d\lambda(\mu)}{d\mu} = -1/\mu^2$$

and

$$I(\mu) = I(\lambda(\mu)) \left(\frac{d\lambda(\mu)}{d\mu}\right)^2 = 1/\mu^2$$

When we estimate μ from data using Maximum Likelihood approach we know that the unbiased estimator for μ is:

$$\hat{\mu} = \frac{1}{N}\sum_{k=1}^{N} x_k$$

If the true parameter is μ, the Variance of the estimator is:

$$\mathrm{Var}(\hat{\mu}) = \frac{\mu^2}{N}$$

By substituting our results in the Cramer-Rao bound:

$$\frac{\mu^2}{N} \geqslant \frac{\mu^2}{N}$$

In this case the inequality is saturated.

Problem 1. *Compute the Fisher information matrix for a Gaussian distribution with parameters μ and σ^2:*

$$p(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \tag{4}$$

The matrix is given by three elements:

$$I_{\mu,\mu} = -\mathbb{E}\left[\frac{\partial^2 \log p(x|\mu, \sigma^2)}{\partial \mu^2}\right] \tag{5}$$

$$I_{\sigma^2,\sigma^2} = -\mathbb{E}\left[\frac{\partial^2 \log p(x|\mu, \sigma^2)}{\partial (\sigma^2)^2}\right] \tag{6}$$

$$I_{\mu,\sigma^2} = -\mathbb{E}\left[\frac{\partial^2 \log p(x|\mu, \sigma^2)}{\partial \mu \partial (\sigma^2)}\right] \tag{7}$$

We treat σ^2 as the parameter so we take derivatives w.r.t. σ^2 and not σ (since two parameters σ and $-\sigma$ would characterize the same model).

- *Repeat the discussion about the Cramer-Rao bound for the parameter μ. Recall that an unbiased estimator for μ is:¡/li¿*

$$\hat{\mu} = \frac{1}{N} \sum_{k=1}^{N} x_k$$

- *(NOT MANDATORY) Do the same for the parameter σ^2 knowing that the unbiased estimator is:*

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{k=1}^{N} (x_k - \hat{\mu})^2$$

For solving the second point, it is necessary to compute the variance of $\hat{\sigma}^2$ which might not be straightforward.

Problem 2. *Repeat the same analysis for a Bernoulli random variable:*

$$P(X = 0|\theta) = 1 - \theta \tag{8}$$
$$P(X = 1|\theta) = \theta \tag{9}$$

Remember that the unbiased estimator for θ is:

$$\hat{\theta} = \frac{1}{N} \sum_{k=1}^{N} x_k$$

where $x_k = 0, 1$. The variance of X w.r.t. the Bernoulli distribution is:

$$\begin{aligned} \mathrm{Var}(X) = \sum_{x=0,1} x^2 p(X = x|\theta) - \left(\sum_{x=0,1} x p(X = x|\theta) \right)^2 = \\ = \theta(1 - \theta) \end{aligned}$$

Comments