

ESE 524: Exponential Family and Cramer-Rao Bound Example

January 29, 2019

Holding out for a Hero

- Some probability distributions do not look like they are "exponential", but looks can be deceiving.
- Let $x[n] \in \{1, 2, 3, \dots\}$, for $n = 0, \dots, N - 1$ be integer valued samples from the discrete distribution:

$$p(x[n]; \theta) = \frac{1}{1 + \theta} \left(\frac{\theta}{1 + \theta} \right)^{x[n]-1}$$

- **Goal:** Find a sufficient statistic for θ , and unbiased estimator for θ , and the cramer rao bound on estimators of θ .
- Source: Problem 4.18 in *Statistical Methods for Signal Processing* by Alfred Hero.

Step 1: Find a Sufficient Statistic

- The joint distribution is given as the product of the independent distributions:

$$p(\mathbf{x}; \theta) = \prod_{n=0}^{N-1} \frac{1}{1+\theta} \left(\frac{\theta}{1+\theta} \right)^{x[n]-1} = \left(\frac{1}{1+\theta} \right)^N \left(\frac{\theta}{1+\theta} \right)^{\sum_{n=0}^{N-1} (x[n]) - N}$$

- First, check that the distribution is in the exponential family, we need to find $h(x), \eta(\theta), T(x), B(\theta)$:

Step 1: Find a Sufficient Statistic

- The joint distribution is given as the product of the independent distributions:

$$p(\mathbf{x}; \theta) = \prod_{n=0}^{N-1} \frac{1}{1+\theta} \left(\frac{\theta}{1+\theta} \right)^{x[n]-1} = \left(\frac{1}{1+\theta} \right)^N \left(\frac{\theta}{1+\theta} \right)^{\sum_{n=0}^{N-1} (x[n]) - N}$$

- First, check that the distribution is in the exponential family, we need to find $h(x), \eta(\theta), T(x), B(\theta)$:

- ▶ $h(x) = 1, \forall x$
- ▶ $\eta(\theta) = \ln \left(\frac{\theta}{1+\theta} \right)$
- ▶ $T(x) = \sum_{n=0}^{N-1} (x[n] - 1)$
- ▶ $B(\theta) = \ln(\theta + 1)$

Step 1: Find a Sufficient Statistic

- The joint distribution is given as the product of the independent distributions:

$$p(\mathbf{x}; \theta) = \prod_{n=0}^{N-1} \frac{1}{1+\theta} \left(\frac{\theta}{1+\theta} \right)^{x[n]-1} = \left(\frac{1}{1+\theta} \right)^N \left(\frac{\theta}{1+\theta} \right)^{\sum_{n=0}^{N-1} (x[n]-1)}$$

- First, check that the distribution is in the exponential family, we need to find $h(x), \eta(\theta), T(x), B(\theta)$:
 - ▶ $h(x) = 1, \forall x$
 - ▶ $\eta(\theta) = \ln \left(\frac{\theta}{1+\theta} \right)$
 - ▶ $T(x) = \sum_{n=0}^{N-1} (x[n] - 1)$
 - ▶ $B(\theta) = \ln(\theta + 1)$
- By using the fact that $a = \exp(\ln(a))$, we have shown that this distribution is in the exponential family - this is a good trick for distributions involving exponents.
- Since this is an exponential family distribution, $T(\mathbf{x}) = \sum_{n=0}^{N-1} (x[n] - 1)$ is the natural sufficient statistic to use here.

Step 2: Find an Unbiased estimator

- Oftentimes, we can use the sufficient statistic from Slide 2 to easily build an estimator.
- Calculate the expected value of T (See board for details):

$$\mathbb{E}(T(x)) = \mathbb{E}\left(\sum_{n=0}^{N-1} (x[n] - 1)\right) = \sum_{n=0}^{N-1} \mathbb{E}(x[n] - 1) =$$

$$\sum_{n=0}^{N-1} \sum_{j=1}^{\infty} (j - 1)p(x[n] = j; \theta) = N \sum_{j=1}^{\infty} (j - 1)p(x[0] = j; \theta)$$

Step 2: Find an Unbiased estimator

- Oftentimes, we can use the sufficient statistic from Slide 2 to easily build an estimator.
- Calculate the expected value of T (See board for details):

$$\mathbb{E}(T(x)) = \mathbb{E}\left(\sum_{n=0}^{N-1} (x[n] - 1)\right) = \sum_{n=0}^{N-1} \mathbb{E}(x[n] - 1) =$$

$$\sum_{n=0}^{N-1} \sum_{j=1}^{\infty} (j - 1)p(x[n] = j; \theta) = N \sum_{j=1}^{\infty} (j - 1)p(x[0] = j; \theta)$$

- The sufficient statistic is a "natural" estimator for $N\theta$, so $\hat{\theta}(\mathbf{x}) = \frac{T(\mathbf{x})}{N}$ will be an unbiased estimator of θ !

Step 3: Find the Cramer-Rao Bound

- Since the samples are i.i.d, the fisher information number is $NI_1(\theta)$, where $I_1(\theta)$ is the fisher information number provided by a single sample.

Step 3: Find the Cramer-Rao Bound

- Since the samples are i.i.d, the fisher information number is $NI_1(\theta)$, where $I_1(\theta)$ is the fisher information number provided by a single sample.
- First compute the score function:

$$\begin{aligned}\ln(p(x[n]; \theta)) &= \ln\left(\frac{1}{1+\theta}\right) + (x[n] - 1) \ln\left(\frac{\theta}{1+\theta}\right) = \\ &= -\ln(1+\theta) + (x[n] - 1)(\ln(\theta) - \ln(1+\theta)) = -x[n] \ln(1+\theta) + (x[n] - 1) \ln(\theta)\end{aligned}$$

Step 3: Find the Cramer-Rao Bound

- Since the samples are i.i.d, the fisher information number is $NI_1(\theta)$, where $I_1(\theta)$ is the fisher information number provided by a single sample.
- First compute the score function:

$$\ln(p(x[n]; \theta)) = \ln\left(\frac{1}{1+\theta}\right) + (x[n] - 1) \ln\left(\frac{\theta}{1+\theta}\right) =$$

$$-\ln(1+\theta) + (x[n] - 1)(\ln(\theta) - \ln(1+\theta)) = -x[n] \ln(1+\theta) + (x[n] - 1) \ln(\theta)$$

- Then take the derivative:

$$\frac{\partial}{\partial \theta} (-x[n] \ln(1+\theta) + (x[n] - 1) \ln(\theta)) = \frac{x[n] - 1 - \theta}{(1+\theta)\theta}$$

Step 3: Find the Cramer-Rao Bound

- Since the samples are i.i.d, the fisher information number is $NI_1(\theta)$, where $I_1(\theta)$ is the fisher information number provided by a single sample.
- First compute the score function:

$$\ln(p(x[n]; \theta)) = \ln\left(\frac{1}{1+\theta}\right) + (x[n] - 1) \ln\left(\frac{\theta}{1+\theta}\right) =$$

$$-\ln(1+\theta) + (x[n] - 1)(\ln(\theta) - \ln(1+\theta)) = -x[n] \ln(1+\theta) + (x[n] - 1) \ln(\theta)$$

- Then take the derivative:

$$\frac{\partial}{\partial \theta} (-x[n] \ln(1+\theta) + (x[n] - 1) \ln(\theta)) = \frac{x[n] - 1 - \theta}{(1+\theta)\theta}$$

- Then take the expectation:

$$I_1(\theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial \theta} (\ln(p(x[n]; \theta)))\right)^2\right] = \sum_{j=1}^{\infty} \left(\frac{j-1-\theta}{(1+\theta)\theta}\right)^2 \frac{1}{1+\theta} \left(\frac{\theta}{1+\theta}\right)^{j-1} =$$

$$\frac{1}{\theta(\theta+1)}$$

Step 3: Find the Cramer-Rao Bound - Cont.

- Now that we know $I_1(\theta)$, $I(\theta) = \frac{N}{\theta(1+\theta)}$.
- Then the Cramer-Rao Bound is:

$$C(\theta) = \frac{\theta(1+\theta)}{N}$$

- Note that as the number of samples N increases, the Cramer-Rao Bound gets lower and theoretically estimators will perform better.
- Does $\hat{\theta} = \frac{T(\mathbf{x})}{N}$ reach the CRB?

Performance of $\hat{\theta}$

- When $N = 1$, the variance of $\hat{\theta}$ is:

$$\mathbb{E}[(x[n] - 1)^2] \sum_{j=1}^{\infty} (j - 1 - \theta)^2 * \frac{1}{1 + \theta} \left(\frac{\theta}{1 + \theta} \right)^{j-1} = \theta(1 + \theta)$$

- So our estimator hit the CRB!
- IF $p(x; \theta)$ belongs to the exponential family, and IF $\mathbb{E}[T(x)] = \theta$, THEN AND ONLY THEN $\mathbb{E}[T(x)]$ will be the MVU Estimator for θ .
- This leads to the Mean Value Parameterization for exponential family distributions (source - 04_sufficiency.pdf on Canvas)

Mean Value Parameterization

- To find an MVU estimator for exponential family distributions, we usually have to change variables.
- Let $\mathbf{x} = [x[0], \dots, x[N-1]]$ be samples from an exponential family distribution with functions $h(x)$, $\eta(\theta)$, $T(\mathbf{x})$, $B(\theta)$.
- Then the **Canonical Form** is the result of a change of variables from θ to η .
- To do this, solve $\eta(\theta)$ for θ , with our probability distribution:

$$\eta(\theta) = \ln\left(\frac{1}{1+\theta}\right), \quad \theta(\eta) = \frac{\exp(\eta)}{1 - \exp(\eta)}$$

- Then the probability distribution becomes:

$$p(x[n]; \eta) = h(x) \exp(\eta T(\mathbf{x}) - B(\theta(\eta))) = \exp\left(\eta \left(\sum_{n=0}^{N-1} (x[n] - 1)\right) - \ln\left(\frac{\exp(\eta)}{1 - \exp(\eta)} + 1\right)\right) =$$

$$\exp\left(\eta \left(\sum_{n=0}^{N-1} (x[n] - 1)\right) - (\ln(\exp(\eta) - 1))\right)$$

Mean Value Parameterization Cont.

- To find the **Mean Value Parameterization**, make a second change of variables.
- We know that $\mathbb{E}[T(\mathbf{x})] = \frac{\partial}{\partial \eta} B(\theta(\eta))$ (here $A(\eta)$ from the notes is $B(\theta(\eta))$).
- Let $\psi = \frac{\partial}{\partial \eta} B(\theta(\eta))$ and rewrite the probability distribution in terms of ψ .
- In our examples $\psi = \frac{\partial}{\partial \eta} -\ln(\exp(\eta) - 1) = \frac{-\exp(\eta)}{\exp(\eta)-1} = \frac{\exp(\eta)}{1-\exp(\eta)}$, so $\eta(\psi) = \ln\left(\frac{\psi}{1+\psi}\right)$.
- Then the probability distribution becomes:

$$p(\mathbf{x}; \psi) = h(\mathbf{x}) \exp(\eta(\psi)T(\mathbf{x}) - B(\theta(\eta(\psi)))) = \exp\left(\ln\left(\frac{\psi}{1+\psi}\right) T(\mathbf{x}) - \ln(\psi+1)\right)$$

- Notice that this is the same as the original probability distribution

Comments of Mean Value Parameterization

- This problem was carefully chosen so that the original distribution was already in the MVP, the changes of variables will rarely reproduce the original distribution.
- $\mathbb{E}[T(\mathbf{x})]$ will be the MVU Estimator for ψ , **NOT θ** . This is **VERY IMPORTANT**. The transformation from θ to ψ is not necessarily 1-1 so you can't make conclusions about θ based off of this.
- This is our first "complete" problem in the class, where we used sufficient statistics to find an estimator, and then figured out the performance of that estimator.
- Using $-\mathbb{E}[\frac{\partial}{\partial \theta} \ln(p(\mathbf{x}; \theta))]$ to find the Fisher Information is usually easier than using the original definition.
- Fisher Information is ubiquitous in statistics, because it represents a metric of how much information the data carries. E.g. if you have two sets of samples, and one has higher fisher information, that set of samples is better for estimating θ .