ESE 524: Detection and Estimation Theory

Recitation 2

Washington University in St. Louis

Topics for Today

- Exponential family of distributions
- Cramér-Rao bound
- Mean value parameterization

Useful Formulas

- Multi-parameter exponential family of distributions:

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = h(\boldsymbol{x}) \exp\left[\sum_{i=1}^{k} \eta_i(\boldsymbol{\theta}) T_i(\boldsymbol{x}) - B(\boldsymbol{\theta})\right]$$

- One-parameter exponential family of distributions:

$$p(\boldsymbol{x}|\theta) = h(\boldsymbol{x}) \exp\left[\eta T(\boldsymbol{x}) - A(\eta)\right]$$

$$\mathrm{E}\left(T(\boldsymbol{x})\right) = \frac{dA(\eta)}{d\eta}, \quad \mathrm{var}\left(T(\boldsymbol{x})\right) = \frac{d^2 A(\eta)}{d\eta^2}$$

Useful Formulas (Cont.)

- Fisher information number:

$$\mathcal{I}(\theta) = \mathrm{E}_X \left[\left(\frac{\partial}{\partial \theta} \log\, p(\boldsymbol{x}, \theta) \right)^2 \right] = -\mathrm{E} \left(\frac{\partial^2}{\partial \theta^2} \log\, p(\boldsymbol{x}, \theta) \right)$$

- Cramér-Rao bound for a single parameter:

$$\frac{1}{\mathcal{I}(\theta)}$$

- Multi-parameter CRB:

$$\mathrm{cov}\ \ (\boldsymbol{T}(\boldsymbol{x})) \geq \frac{\partial \psi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \mathcal{I}(\boldsymbol{\theta})^{-1} \frac{\partial \psi(\boldsymbol{\theta})^T}{\partial \boldsymbol{\theta}}$$

- Where $\psi(\boldsymbol{\theta}) = \mathrm{E}\left(\boldsymbol{T}(\boldsymbol{x})\right)$

Holding out for a Hero

- Some probability distributions do not look like they are "exponential", but looks can be deceiving.
- Let $x[n] \in \{1, 2, 3, \ldots\}$, for $n = 0, ..., N-1$ 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 a sufficient statistic for θ, and unbiased estimator for θ, and the Cramér-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:
  $p(\boldsymbol{x};\theta) = \prod_{n=0}^{N-1} \frac{1}{1+\theta}\left(\frac{\theta}{1+\theta}\right)^{x[n]-1} = \left(\frac{1}{1+\theta}\right)^{N}\left(\frac{\theta}{1+\theta}\right)^{\sum_{n=0}^{N-1}(x[n])-N}$
- First, check that the distribution is in the exponential family, we need to find $h(\boldsymbol{x}), \eta(\theta), T(\boldsymbol{x}), B(\theta)$:
  - $h(\boldsymbol{x}) = 1,\ \forall x$
  - $\eta(\theta) = \ln\left(\frac{\theta}{1+\theta}\right)$
  - $T(\boldsymbol{x}) = \sum_{n=0}^{N-1}(x[n]-1)$
  - $B(\theta) = \ln(\theta+1)$
- 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, $T(\boldsymbol{x}) = \sum_{n=0}^{N-1}(x[n]-1)$ is the natural sufficient statistic to use here.

Step 2: Find an Unbiased estimator

- Oftentimes, we can use the sufficient statistic from slide no. 2 to easily build an estimator.
- Calculate the expected value of T (see board for details):

$$\mathrm{E}(T(\boldsymbol{x})) = \mathrm{E}(\sum_{n=0}^{N-1}(x[n]-1)) = \sum_{n=0}^{N-1}\mathrm{E}(x[n]-1) =$$

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

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

Step 3: Find the Cramér-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:
  $\ln(p(x[n];\theta)) = \ln(\frac{1}{1+\theta}) + (x[n]-1)\ln(\frac{\theta}{1+\theta}) =$
  $-\ln(1+\theta) + (x[n]-1)(\ln(\theta) - \ln(1+\theta) =$
  $-x[n]\ln(1+\theta) + (x[n]-1)\ln(\theta)$
- 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: $I_1(\theta) = \mathrm{E}[(\frac{\partial}{\partial\theta}(\ln(p(x[n];\theta))))^2] =$
  $\sum_{j=1}^{\infty}\left(\frac{j-1-\theta}{(1+\theta)\theta}\right)^2 \frac{1}{1+\theta}\left(\frac{\theta}{1+\theta}\right)^{j-1} = \frac{1}{\theta(\theta+1)}$

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

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

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

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

Performance of $\hat{\theta}$

- When $N = 1$, the estimator is given by $\hat{\theta} = x[0] - 1$.
- the variance of $\hat{\theta}$ is: $\mathrm{E}[(\hat{\theta} - \theta)^2] = \mathrm{E}[((x[0] - 1) - \theta)^2] = \sum_{j=1}^{\infty} (j - 1 - \theta)^2 * \frac{1}{1+\theta} \left(\frac{\theta}{1+\theta}\right)^{j-1} = \theta(1 + \theta)$
- So our estimator hit the CRB!
- if $p(x; \theta)$ belongs to the exponential family, and if $\mathrm{E}[T(\boldsymbol{x})] = \theta$, then and only then $T(\boldsymbol{x})$ will be the MVU Estimator for θ.
- This leads to the mean value parameterization for exponential family distributions (source - 04_sufficiency.pdf on Canvas)

Efficient Estimators

- Recall from L2 that Efficient Estimators are estimators whose variance is the CRB.
- Theorem 2 from L2 states that if you can construct and efficient estimator of $\psi(\theta)$ for some function $\psi(\cdot)$, then we must be dealing with an exponential family distribution, and vice versa.
- In the proof of this statement, Kay provides the condition that $T(\boldsymbol{x})$ is an efficient estimator of $\boldsymbol{\theta}$ when:

$$\frac{\partial \ln(p(\boldsymbol{x};\theta))}{\partial \theta} = \mathcal{I}(\theta)(T(\boldsymbol{x}) - \theta)$$

Continuing the Previous Example

- Using the exponential family representation:

$$\ln(p(\boldsymbol{x};\theta)) = -N\ln(1+\theta) + \ln(\frac{\theta}{1+\theta})\sum_{n=0}^{N-1}(x[n]-1)$$

- Taking the derivative yields:

$$\frac{\sum_{n=0}^{N-1}(x[n]-1) - N\theta}{\theta(1+\theta)} = \frac{NT(\boldsymbol{x}) - N\theta}{\theta(1+\theta)} = \mathcal{I}(\theta)(T(\boldsymbol{x})-\theta)$$

- We satisfy the theorem! This confirms the claim made last week that $T(\boldsymbol{x})$ is the MVU for θ.
- But what if $\mathrm{E}(T(\boldsymbol{x}))$ is not θ?

Finding Efficient Estimators for Exponential Distibutions

- To build an efficient estimator based off a sufficient statistic $T(\boldsymbol{x})$ for exponential family variables:
    1. change variables to the canonical parameterization $\theta \to \eta$.
    2. Find $\psi = \mathrm{E}(T(\boldsymbol{x})) = \frac{\partial B(\theta(\eta))}{\partial \eta}$ and variance $\mathrm{var}(T) = \frac{\partial^2 B(\theta(\eta))}{\partial \eta^2}$.
    3. Change variables again from $\eta \to \psi$. $T(\boldsymbol{x})$ is now your MVU estimator.
- To check that $T(\boldsymbol{x})$ is efficient, again use the factorization condition (see board for proof).

Mean Value Parameterization

- To find an MVU estimator for exponential family distributions, we usually have to change variables.
- Let $\boldsymbol{x} = [x[0], ..., x[N-1]]$ be samples from an exponential family distribution with functions $h(\boldsymbol{x}), \eta(\theta), T(\boldsymbol{x}), B(\theta)$.
- 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:

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

- Then the probability distribution becomes:
$p(x[n];\eta) = h(\boldsymbol{x}))\exp(\eta T(\boldsymbol{x}) - B(\theta(\eta)) =$
$\exp(\eta(\sum_{n=0}^{N-1}(x[n]-1)) - \ln(\frac{\exp(\eta)}{1-\exp(\eta)}+1)) =$
$\exp(\eta(\sum_{n=0}^{N-1}(x[n]-1)) - (\ln(\exp(\eta)-1)))$

Mean Value Parameterization Cont.

- To find the mean value parameterization, make a second change of variables.
- We know that $\mathrm{E}[T(\boldsymbol{x})] = \frac{\partial}{\partial \eta} B(\theta(\eta))$ (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:$p(\boldsymbol{x};\psi) = h(\boldsymbol{x})\exp(\eta(\psi)T(\boldsymbol{x}) - B(\theta(\eta(\psi)))) = \exp(\ln\left(\frac{\psi}{1+\psi}\right) T(\boldsymbol{x}) - \ln(\psi + 1)))$
- 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.
- $T(\boldsymbol{x})$ will be the MVU Estimator for ψ, not θ. This is very important. Transforming from ψ to θ will change the variance of $T(\boldsymbol{x})$, so it might not be MVU anymore.
- 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 -$\mathrm{E}[\frac{\partial^2}{\partial\theta^2}\ln(p(\boldsymbol{x};\theta))]$ 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.