ESE 524: Mean Value Parameterization

February 7, 2019

From Last Tuesday

- Let $x[n] \in \{1, 2, 3, ...\}$, for $n = 0, ..., N-1$ be integer valued samples from the discrete distribution:

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

- We rewrote this distribution as a member of the exponential family:

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

- We found that the statistic $T(x) = \frac{1}{N}\sum_{n=0}^{N-1}(x[n]-1)$ is a sufficient statistic, unbiased estimator, and attains the CRB of

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

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 statment, Kay provides the condition that $T(x)$ is an efficient estimator of θ when:

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

Continuing Last Week's Example

- Using the exponential family representation:

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

- Taking the derivative yields:

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

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

Finding Efficient Estimators for Exponential Distibutions

- To build an efficient estimator based off a sufficient statistic $T(x)$ for exponential family variables:
  1. Change Variables to the Canonical parameterization $\theta \to \eta$.
  2. Find $\psi = \mathbb{E}(T(x)) = \frac{\partial B(\theta(\eta))}{\partial \eta}$ and variance $\mathrm{var}(T) = \frac{\partial^2 B(\theta(\eta))}{\partial \eta^2}$.
  3. Change variables again from $\eta \to \psi$. - $T(x)$ is now your MVU estimator.
- To check that $T(x)$ 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 $\mathbf{x} = [x[0], ..., x[N-1]]$ 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:

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

- Then the probability distribution becomes:

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

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

Mean Value Parameterization Cont.

- To find the Mean Value Parameterization, make a second change of variables.
- We know that $\mathbb{E}[T(\mathbf{x})] = \frac{\partial}{\partial\eta}B(\theta(\eta))$ (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:

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

- Notice that this is the same as the original probability distribution, the pdf was already the MVP.

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.

- Using -

$$\mathbb{E}[\frac{\partial^2}{\partial\theta^2}\ln(p(\mathbf{x};\theta))]$$

  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.