Non Bayesian Parameter Estimation

In many problems, it is unnatural to assign a prior distribution to the latent variable of interest

- We instead treat it as deterministic, but unknown
- Instead of having observation models conditioned on some variable x, we instead have it be parameterized by x
    - This is the notation $p(y|x)$ versus $p(y;x)$

The first thing to note is that if we look at our previous formulation of BLS:

$$\hat{x} = \arg\min_f \mathbb{E}\left[(x - f(y))^2\right]$$

- However since x is now fixed, we have this is minimized when we just choose $f(y) = x$
- We can't do this!
- We need to restrict our search to valid estimators that are just statistics of the data

Estimators themselves are functions $\hat{x}(y)$:

- Their bias and variance are functions of x because we take the expectation over all possible y

For the beginning, we will focus on just scalar values of x (but y can be vector valued)

Minimum Variance Unbiased (MVU) Estimators

MVU

A MVU estimator $\hat{x}_*$ is an estimator such that it has 0 bias and of all estimators that have 0 bias, it has the minimum variance $\lambda_{\hat{x}}(x)$ for every possible x

This could potentially not exist if:

- There are no unbiased estimators
- There are no estimators that uniformly have lower variance for all x

In general, there is no systematic procedure to find these or determine if they exist

Cramer-Rao Bound

Cramer-Rao Bound (Scalar Version)

Suppose $p_y(y;\cdot)$ for every y is positive and differentiable on $\mathcal{X} \in \mathbb{R}$ and satisfies the regularity condition for all x:

$$\mathbb{E}\left[\frac{\partial}{\partial x} \ln p_y(y;x)\right]$$

Then for any unbiased $\hat{x}(\cdot)$:

$$\lambda_{\hat{x}}(x) \geq \frac{1}{J_y(x)}$$

where $J_y(X)$ is the Fisher information in y about x:

$$J_y(x) = \mathbb{E}\left[S(y;x)^2\right]$$

where $S(y;x)$ is the score function for x based on y:

$$S(y;x) = \frac{\partial}{\partial x} \ln p_y(y;x)$$

With an extra regularity condition, we can express the Fisher information as:

Corollary

If $p_y(y;\cdot)$ also satisfies the regularity condition:

$$\mathbb{E}\left[\left(\frac{\partial}{\partial x}\ln p_y(y;x)\right)^2+\frac{\partial^2}{\partial x^2}\ln p_y(y;x)\right]=0$$

then the Fisher information can also be expressed as:

$$J_y(x)=-\mathbb{E}\left[\frac{\partial^2}{\partial x^2}\ln p_y(y;x)\right]$$