Chapter 7: Estimation

Sections

Properties of MLE's

Theorem 7.6.2: MLE's are invariant

If $\hat{\theta}$ is the MLE of θ and $g(\theta)$ is a function of θ
then $g(\hat{\theta})$ is the MLE of $g(\theta)$

Example:

- Let $\hat{p}$ be the MLE of a probability parameter, e.g. the p in $\mathrm{Binomial}(n, p)$.
  Then the MLE of the odds, $\frac{p}{1-p}$ is $\frac{\hat{p}}{1-\hat{p}}$

In general this does not hold for Bayes estimators.

- E.g. for square error loss $E(g(\theta)|\mathbf{x}) \neq g(E(\theta|\mathbf{x}))$

Consistency:

- Under fairly typical conditions, e.g. MLEs are unique, the sequence of MLEs is a consistent sequence of estimators of θ, i.e. $\hat{\theta}_n(\mathbf{X}) \xrightarrow{P} \theta$

Computation

For MLE's

- In many practical situations the maximization we need is not available analytically or too cumbersome
- There exist many numerical optimization methods, Newton's Method (see definition 7.6.2) is one example.

For Bayesian estimators

- In many practical situations the posterior distribution is not available in closed form
  - This happens if we cannot evaluate the integral for the marginal distribution
  - In stead people either approximate the posterior distribution or take random samples from it, e.g. using Markov Chain Monte Carlo (MCMC) methods

Method of Moments (MOM)

- Let $X_1, \dots, X_n$ be i.i.d. from $f(x|\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is k dimensional.
- The *j^{th} sample moment* is defined as $m_j = \frac{1}{n}\sum_{i=1}^n X_i^j$

Method of moments (MOM) estimator: match the theoretical moments and the sample moments and solve for parameters:

$$m_1 = E(X_1|\boldsymbol{\theta}),\ m_2 = E(X_1^2|\boldsymbol{\theta}),\ \dots,\ m_k = E(X_1^k|\boldsymbol{\theta})$$

Example:

- Let $X_1, \dots, X_n$ be i.i.d. $\mathrm{Gamma}(\alpha, \beta)$. Then

$$E(X) = \frac{\alpha}{\beta} \quad \text{and} \quad E(X_2) = \frac{\alpha(\alpha+1)}{\beta^2}$$

Find the MOM estimator of α and β

MOM estimators are consistent.

Sufficient Statistics

- A statistic: $T = r(X_1, \dots, X_n)$

Def: Sufficient Statistics

Let $X_1, \dots, X_n$ be a random sample form $f(x|\theta)$ and let T be a statistic. If the conditional distribution of

$$X_1, \dots, X_n | T = t$$

does not depend on θ then T is called a *sufficient statistic*

- The idea: Just as good to have the observed sufficient statistic as it is to have the individual observations of $X_1, \dots, X_n$
- Can limit our search for a good estimator to sufficient statistics

Sufficient Statistics

Theorem 7.7.1: Factorization Criterion

Let $X_1, \dots, X_n$ be a random sample form $f(x|\theta)$ where $\theta \in \Omega$ is unknown. A statistic $T = r(X_1, \dots, X_n)$ is a sufficient statistic for θ if and only if for all $\mathbf{x} \in \mathbb{R}^n$ and all $\theta \in \Omega$, the joint pdf/pf $f_n(\mathbf{x}|\theta)$ can be factored as

$$f_n(\mathbf{x}|\theta) = u(\mathbf{x})v\left(r(\mathbf{x}), \theta\right)$$

where function u and v are nonnegative.

- The function u may depend on $\mathbf{x}$ but not on θ
- The function v depends on θ but depends on $\mathbf{x}$ only through the value of the statistic $r(\mathbf{x})$

Both MLEs and Bayesian estimators depend on data only through sufficient statistics.

END OF CHAPTER 7