ECSE 6520: Estimation and Detection Theory

Sufficient Statistic

Class Notes - 4

January 26, 2014

Contents

1 Sufficient Statistics

Many detection/estimation/classification problems have the following common structure.A continuous time waveform $\{x(t)|t \in \mathbb{R}\}$ is measured at n time instants $t_1, ..., t_n$ producing the vector $x = [x_1, ..., x_n]^T$ where $x_i = x(t_i)$. The vector x is modelled as a realization of a random vector X with a joint distribution which is of known form but depends on a handful (p) of unknown parameters $\theta = [\theta_1, ..., \theta_p]^T$.

- $X = [X_1, ..., X_n]^T$, $X_i = X(t_i)$, is a vector of random measurements or observations taken over the course of the experiment
- $\mathcal{X}$ is sample or measurement space of realizations x of X.
- $\mathcal{B}$ is the event space induced by X, e.g., the Borel subsets of $\mathbb{R}^n$.
- $\theta \in \Theta$ is an unknown parameter vector of interest.
- Θ is parameter space for the experiment.
- P_θ is a probability measure on $\mathcal{B}$ for given θ. $\{P_\theta\}_{\theta\in\Theta}$ is called the statistical model for the experiment. Note that the probability model induces the joint cumulative distribution function of X.
- Our objective is to infer properties of θ knowing only the parametric form of the statistical model, i.e, the pdf $f_\theta(x)$ of X, given a realization x of X. In other words, we want to come up with a function, called inference function, which maps X to a subset of the parameter space, say for the purpose of designing an estimator, classifier or a detector.
- Is it possible to compress the measurement x into a low dimensional statistic without effecting the quality of the inference about θ?

- In other words, does there exist $T = \tau(X)$, where the dimension of T is $M < N$, such that T carries all the useful information on θ?

- If so, for the purpose of studying θ, we could discard the raw measurements x and retain only the compressed statistics t.

Definition 1 ***Sufficient Statistics* -** *Let $f_\theta(x)$ be the pdf of the random variable X. The statistic $T = \tau(X)$ is a sufficient statistics for θ if the conditional distribution of X given T is independent of θ. Equivalently, the functional form of $f_{X|T}(x|t)$ does not involve θ.*

1.1 The Fisher-Neyman Factorization Theorem

In general, it is difficult to verify the definition of sufficient statistic directly since it involves derivation of the conditional probability. The Fisher-Neyman Factorization theorem allows us to verify sufficient statistics more readily.

Theorem 1 ***The Fisher-Neyman Factorization Theorem* -** *Let $f_\theta(x)$ be the pdf or the pmf of the random variable X. The statistic $T = \tau(X)$ is a sufficient statistics for θ iff there exists functions $b_\theta(t)$ and $a(x)$ such that*

$$f_\theta(x) = b_\theta(\tau(x))a(x).$$

- The theorem gives us a formula for $f_{X|T}(x|t)$, namely

$$f_{X|T}(x|t) = \frac{a(x)}{\sum_{\{x'|\tau(x')=t\}} a(x')}.$$

- The theorem states that $\tau(X)$ is sufficient for θ if and only if the pdf or pmf of X may be written as a scale constant, dependent on x and t, but independent on θ, times the pdf or pmf of T.

- The theorem also shows that the scale constant is the conditional density of x, given t, and it shows how to compute the pdf or pmf for the sufficient statistics T.

- If an invertible function is applied to a sufficient statistics, the result is again a sufficient statistic.

1.2 Rao-Blackwell Theorem

The importance of the sufficient statistics is reflected in the following theorem.

Theorem 2 ***Rao-Blackwell Theorem* -** *Let $f_\theta(x)$ be the pdf or the pmf of the random variable X. Let $T = \tau(X)$ be a sufficient statistic for θ. Let $\hat{\theta}_1(x)$ be an estimator of θ and define the mean square error*

$$MSE(\hat{\theta}_1) = E[\|\hat{\theta}_1(x) - \theta\|^2].$$

Define

$$\hat{\theta}_2(x) = E[\hat{\theta}_1(x)|T = \tau(x)].$$

Then

$$MSE(\hat{\theta}_2) \leq MSE(\hat{\theta}_1)$$

with equality if and only if $\hat{\theta}_1 = \hat{\theta}_2$ with probability 1.

Remarks -

- Given any estimate $\hat{\theta}_1$ that is NOT a function of a sufficient statistic, there exist a better estimate with respect to MSE.
- We may restrict out search for estimators to functions of sufficient statistic.

2 Minimal Sufficient Statistics

- What is the maximum possible amount of reduction one can apply to the data sample without losing information concerning how the model depends on θ? The answer to this question lies in the notion of a *minimal sufficient statistic*.

- Minimal sufficient statistic cannot be reduced any further without loss in information. In other words, any other sufficient statistic can be reduced down to a minimal sufficient statistic without information loss. Since reduction of a statistic is accomplished by applying a functional transformation we have the following definition:

Definition 2 ***Minimal Sufficient Statistics -****: T_{min} is a minimal sufficient statistic if it can be obtained from any other sufficient statistic T by applying a functional transformation to T. Equivalently, if T is any sufficient statistic there exists a function q such that $T_{min} = q(T)$.*

Minimal sufficient statistics are not unique: if T_{min} is minimal sufficient $h(T_{min})$ is also minimal sufficient if h is any invertible function. Minimal sufficient statistics can be found in a variety of ways. One way is to find a *complete sufficient statistic.*

3 Complete Sufficient Statistic

Definition 3 ***Complete Sufficient Statistic -*** *A sufficient statistic T is complete if*

$$E_\theta[g(T)] = 0, \quad for \quad all \quad \theta \in \Theta$$

implies that the function g is identically zero, i.e., $g(t) = 0$ for all values of t.

Theorem 3 *Under general conditions, if T is a complete sufficient statistic, then T is minimal.*

4 The Exponential Family

In general, sufficient statistics, especially ones that are minimal and complete, can be difficult to find. For a special family of distributions, however, we can immediately identify a complete minimal sufficient statistic.

Definition 4 ***The Exponential Family*** **-** *The distribution of the random variable X is said to belong to the exponential family of distributions if its pmf/pdf can be expressed as*

$$f_\theta(x) = a(\theta)b(x)\exp\{c(\theta)^T\tau(x)\}$$

for some a, b, c, and τ, where the dimension p of θ is also the dimension of $c(\theta)$ and $\tau(x)$.

- Many common distributions belong to the exponential family, including Gaussian with unknown mean and/or variance, Poisson, exponential, gamma, binomial, and multinomial.

Theorem 4 *If the distribution of X belongs to the exponential family, then $T = \tau(X)$ is a sufficient statistic.*

Note the $\tau(x)$ in the above theorem is the $\tau(x)$ in the definition of the exponential family.

Theorem 5 *Under certain "reasonable" conditions, $T = \tau(x)$ is a complete and minimal sufficient statistic for θ.*