Parameter Estimation

Saravanan Vijayakumaran
sarva@ee.iitb.ac.in

Department of Electrical Engineering
Indian Institute of Technology Bombay

March 15, 2013

Parameter Estimation

- Hypothesis testing was about making a choice between discrete states of nature
- Parameter or point estimation is about choosing from a continuum of possible states

Example

Consider the signal below

$$y(t) = A \sin(2\pi f_c t + \phi) + n(t)$$

where $n(t)$ is a noise signal.

- The amplitude A is a real number
- The frequency f_c is a positive real number in some known interval
- The phase ϕ can take any real value in the interval $[0, 2\pi)$
- We are interested in estimating A, f_c and ϕ

System Model for Parameter Estimation

- Consider a family of distributions

$$\mathbf{Y} \sim P_\theta, \quad \theta \in \Lambda$$

where the observation vector $\mathbf{Y} \in \Gamma \subseteq \mathbb{R}^n$ for $n \in \mathbb{N}$ and $\Lambda \subseteq \mathbb{R}^m$ is the parameter space

Example

$$Y = A + Z$$

where A is an unknown parameter and Z is a standard Gaussian RV. Here $\theta = A$.

- The goal of parameter estimation is to find θ given $\mathbf{Y}$
- An estimator is a function from the observation space to the parameter space

$$\hat{\theta} : \Gamma \to \Lambda$$

Which is the Optimal Estimator?

- Assume there is a cost function C

$$C : \Lambda \times \Lambda \to \mathbb{R}$$

  such that $C[a, \theta]$ is the cost of estimating the true value of θ as a

- Examples of cost functions

  Squared Error $C[a, \theta] = (a - \theta)^2$
  Absolute Error $C[a, \theta] = |a - \theta|$
  Threshold Error $C[a, \theta] = \begin{cases} 0 & \text{if } |a - \theta| \leq \Delta \\ 1 & \text{if } |a - \theta| > \Delta \end{cases}$

Which is the Optimal Estimator?

- Suppose that the parameter θ is the realization of a random variable Θ
- With an estimator $\hat{\theta}$ we associate a conditional cost or risk conditioned on θ

$$r_\theta(\hat{\theta}) = E_\theta \left\{ C \left[\hat{\theta}(\mathbf{Y}), \theta \right] \right\}$$

- The average risk or Bayes risk is given by

$$R(\hat{\theta}) = E \left\{ r_\Theta(\hat{\theta}) \right\}$$

- The optimal estimator is the one which minimizes the Bayes risk

Which is the Optimal Estimator?

- Given that

$$r_\theta(\hat{\theta}) = E_\theta\left\{C\left[\hat{\theta}(\mathbf{Y}),\theta\right]\right\} = E\left\{C\left[\hat{\theta}(\mathbf{Y}),\Theta\right]\middle|\Theta=\theta\right\}$$

the average risk or Bayes risk is given by

$$\begin{aligned} R(\hat{\theta}) &= E\left\{C\left[\hat{\theta}(\mathbf{Y}),\Theta\right]\right\} \\ &= E\left\{E\left\{C\left[\hat{\theta}(\mathbf{Y}),\Theta\right]\middle|\mathbf{Y}\right\}\right\} \end{aligned}$$

- The optimal estimate for θ can be found by minimizing for each $\mathbf{Y}=\mathbf{y}$ the posterior cost

$$E\left\{C\left[\hat{\theta}(\mathbf{y}),\Theta\right]\middle|\mathbf{Y}=\mathbf{y}\right\}$$

Minimum-Mean-Squared-Error (MMSE) Estimation

- $C[a, \theta] = (a - \theta)^2$
- The posterior cost is given by

$$
\begin{aligned}
E\left\{ (\hat{\theta}(\mathbf{y}) - \Theta)^2 \middle| \mathbf{Y} = \mathbf{y} \right\} &= \left[\hat{\theta}(\mathbf{y})\right]^2 \\
&\quad -2\hat{\theta}(\mathbf{y}) E\left\{ \Theta \middle| \mathbf{Y} = \mathbf{y} \right\} \\
&\quad +E\left\{ \Theta^2 \middle| \mathbf{Y} = \mathbf{y} \right\}
\end{aligned}
$$

- The Bayes estimate is given by

$$
\hat{\theta}_{MMSE}(\mathbf{y}) = E\left\{ \Theta \middle| \mathbf{Y} = \mathbf{y} \right\}
$$

Example 1: MMSE Estimation

- Suppose X and Y are jointly Gaussian random variables
- Let the joint pdf be given by

$$p_{XY}(x,y) = \frac{1}{2\pi|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{s}-\mu)^T \Sigma^{-1} (\mathbf{s}-\mu)\right)$$

where $\mathbf{s} = \begin{bmatrix} x \\ y \end{bmatrix}$, $\mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$ and $\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$

- Suppose Y is observed and we want to estimate X
- The MMSE estimate of X is

$$\hat{X}_{MMSE}(y) = E\left[X \middle| Y = y\right]$$

Example 1: MMSE Estimation

- The conditional distribution of X given $Y = y$ is a Gaussian RV with mean
$$\mu_{X|y} = \mu_x + \frac{\sigma_x}{\sigma_y}\rho(y - \mu_y)$$
and variance
$$\sigma^2_{X|y} = (1 - \rho^2)\sigma^2_x$$
- Thus the MMSE estimate of X given $Y = y$ is
$$\hat{X}_{MMSE}(y) = \mu_x + \frac{\sigma_x}{\sigma_y}\rho(y - \mu_y)$$

Example 2: MMSE Estimation

- Suppose A is a Gaussian RV with mean μ and known variance v^2
- Suppose we observe $Y_i, \; i = 1, 2, \ldots, M$ such that

$$Y_i = A + N_i$$

where N_i's are independent Gaussian RVs with mean 0 and known variance σ^2

- Suppose A is independent of the N_i's
- The MMSE estimate is given by

$$\hat{A}_{MMSE}(\mathbf{y}) = \frac{\frac{Mv^2}{\sigma^2}\hat{A}_1(\mathbf{y}) + \mu}{\frac{Mv^2}{\sigma^2} + 1}$$

where $\hat{A}_1(\mathbf{y}) = \frac{1}{M}\sum_{i=1}^{M} y_i$

Maximum A Posteriori (MAP) Estimation

- In some situations, the conditional mean may be difficult to compute
- An alternative is to use MAP estimation
- The MAP estimator is given by

$$\hat{\theta}_{MAP}(\mathbf{y}) = \underset{\theta}{\operatorname{argmax}}\, p\left(\theta \middle| \mathbf{Y} = \mathbf{y}\right)$$

  where p is the conditional density of Θ given $\mathbf{Y}$.

- It can be obtained as the optimal estimator for the threshold cost function

$$C[a, \theta] = \begin{cases} 0 & \text{if } |a - \theta| \leq \Delta \\ 1 & \text{if } |a - \theta| > \Delta \end{cases}$$

  for small $\Delta > 0$

Maximum A Posteriori (MAP) Estimation

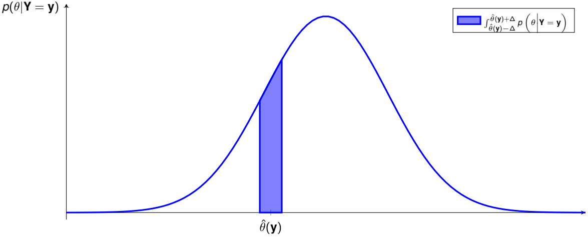
- For the threshold cost function, we have[1]

$$
\begin{aligned}
& E\left\{ C\left[\hat{\theta}(\mathbf{y}), \Theta\right] \middle| \mathbf{Y} = \mathbf{y} \right\} \\
= & \int_{-\infty}^{\infty} C[\hat{\theta}(\mathbf{y}), \theta] p\left(\theta \middle| \mathbf{Y} = \mathbf{y}\right) \, d\theta \\
= & \int_{-\infty}^{\hat{\theta}(\mathbf{y})-\Delta} p\left(\theta \middle| \mathbf{Y} = \mathbf{y}\right) \, d\theta + \int_{\hat{\theta}(\mathbf{y})+\Delta}^{\infty} p\left(\theta \middle| \mathbf{Y} = \mathbf{y}\right) \, d\theta \\
= & \int_{-\infty}^{\infty} p\left(\theta \middle| \mathbf{Y} = \mathbf{y}\right) \, d\theta - \int_{\hat{\theta}(\mathbf{y})-\Delta}^{\hat{\theta}(\mathbf{y})+\Delta} p\left(\theta \middle| \mathbf{Y} = \mathbf{y}\right) \, d\theta \\
= & 1 - \int_{\hat{\theta}(\mathbf{y})-\Delta}^{\hat{\theta}(\mathbf{y})+\Delta} p\left(\theta \middle| \mathbf{Y} = \mathbf{y}\right) \, d\theta
\end{aligned}
$$

- The Bayes estimate is obtained by maximizing the integral in the last equality

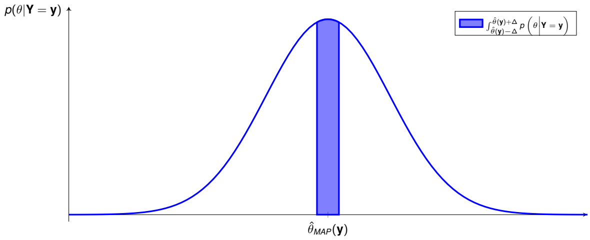
[1] Assume a scalar parameter θ for illustration

Maximum A Posteriori (MAP) Estimation



- The shaded area is the integral $\int_{\hat{\theta}(\mathbf{y})-\Delta}^{\hat{\theta}(\mathbf{y})+\Delta} p\left(\theta \middle| \mathbf{Y}=\mathbf{y}\right)\, d\theta$
- To maximize this integral, the location of $\hat{\theta}(\mathbf{y})$ should be chosen to be the value of θ which maximizes $p(\theta|\mathbf{Y}=\mathbf{y})$

Maximum A Posteriori (MAP) Estimation



- This argument is not airtight as $p(\theta|\mathbf{Y} = \mathbf{y})$ may not be symmetric at the maximum
- But the MAP estimator is widely used as it is easier to compute than the MMSE estimator

Maximum Likelihood (ML) Estimation

- The ML estimator is given by

$$\hat{\theta}_{ML}(\mathbf{y}) = \underset{\theta}{\operatorname{argmax}}\, p\left(\mathbf{Y} = \mathbf{y} \middle| \theta\right)$$

  where p is the conditional density of $\mathbf{Y}$ given Θ.
- It is the same as the MAP estimator when the prior probability distribution of Θ is uniform
- It is also used when the prior distribution is not known

Example 1: ML Estimation

- Suppose we observe $Y_i, \ i = 1, 2, \ldots, M$ such that

$$Y_i \sim \mathcal{N}(\mu, \sigma^2)$$

  where Y_i's are independent, μ is unknown and σ^2 is known

- The ML estimate is given by

$$\hat{\mu}_{ML}(\mathbf{y}) = \frac{1}{M} \sum_{i=1}^{M} y_i$$

Example 2: ML Estimation

- Suppose we observe $Y_i,\ i = 1, 2, \ldots, M$ such that

$$Y_i \sim \mathcal{N}(\mu, \sigma^2)$$

  where Y_i's are independent, both μ and σ^2 are unknown

- The ML estimates are given by

$$\begin{aligned}
\hat{\mu}_{ML}(\mathbf{y}) &= \frac{1}{M}\sum_{i=1}^{M} y_i \\
\hat{\sigma}^2_{ML}(\mathbf{y}) &= \frac{1}{M}\sum_{i=1}^{M} (y_i - \hat{\mu}_{ML}(\mathbf{y}))^2
\end{aligned}$$

Example 3: ML Estimation

- Suppose we observe $Y_i,\ i = 1, 2, \ldots, M$ such that

$$Y_i \sim \text{Bernoulli}(p)$$

  where Y_i's are independent and p is unknown

- The ML estimate of p is given by

$$\hat{p}_{ML}(\mathbf{y}) = \frac{1}{M} \sum_{i=1}^{M} y_i$$

Example 4: ML Estimation

- Suppose we observe $Y_i, \ i = 1, 2, \ldots, M$ such that

$$Y_i \sim \text{Uniform}[0, \theta]$$

  where Y_i's are independent and θ is unknown

- The ML estimate of θ is given by

$$\hat{\theta}_{ML}(\mathbf{y}) = \max\left(y_1, y_2, \ldots, y_{M-1}, y_M\right)$$

Reference

- Chapter 4, *An Introduction to Signal Detection and Estimation*, H. V. Poor, Second Edition, Springer Verlag, 1994.

Questions?