ECE531 Screencast 5.1: Introduction to Bayesian Estimation

D. Richard Brown III

Worcester Polytechnic Institute
Parameter Estimation Approaches

Two fundamentally different approaches to parameter estimation:

1. **Non-random (Classical)**: The parameter of interest $\theta$ is considered to be a deterministic but unknown constant. It does not possess any known prior distribution.

2. **Bayesian**: The parameter of interest $\theta$ is a realization of a random variable $\Theta$ with a known prior density $\pi(\theta)$.

Remarks:

- The performance of classical parameter estimators is usually a function of $\theta$.
- The Bayesian estimator gives the best possible estimate “on the average”, where the risk/cost is averaged over the joint pdf $p_{Y,\Theta}(y, \theta)$. Performance is not a function of $\theta$.
- If you have prior knowledge, you should use it. Prior knowledge will lead to a more accurate estimator.
Cost Assignments and Conditional Risk

Cost assignment: \( C_\theta(\hat{\theta}) : \Lambda \times \Lambda \mapsto \mathbb{R} \) is the cost of the parameter estimate \( \hat{\theta} \in \Lambda \) given the true parameter \( \theta \in \Lambda \). Let \( \epsilon := \hat{\theta}(y) - \theta \). Many cost assignments can be written as \( C_\theta(\hat{\theta}) = C(\epsilon) \). Recall:

- Squared error: \( C_\theta(\hat{\theta}) = \epsilon^2 \).
- Absolute error: \( C_\theta(\hat{\theta}) = |\epsilon| \).
- Uniform error ("hit or miss"):

\[
C_\theta(\hat{\theta}) = \begin{cases} 
0 & |\epsilon| \leq \frac{\Delta}{2} \\
1 & \text{otherwise}
\end{cases}
\]

Conditional risk of estimator \( \hat{\theta}(y) \) when the true parameter is \( \theta \):

\[
R_\theta(\hat{\theta}) := \mathbb{E} \left[ C_\theta(\hat{\theta}(Y)) \mid \Theta = \theta \right] = \int_\mathcal{Y} C_\theta(\hat{\theta}(y)) p_{Y \mid \Theta}(y \mid \Theta = \theta) \, dy
\]
The Bayesian Philosophy

We assume that the unknown parameter(s) are random with a known prior distribution $\Theta \sim \pi(\theta)$. The average/Bayes risk of estimator $\hat{\theta}(y)$ is then

$$
r(\hat{\theta}) = E[R_{\Theta}(\hat{\theta})] = \int_{\Lambda} R_{\theta}(\hat{\theta})\pi(\theta) d\theta$$

$$= \int_{\Lambda} \int_{Y} C_{\theta}(\hat{\theta}(y))p_{\theta}(y)\pi(\theta) dy d\theta$$

$$= \int_{Y} \int_{\Lambda} C_{\theta}(\hat{\theta}(y))p_{\theta}(y)\pi(\theta) d\theta dy$$

where $p_{\theta}(y)$ is shorthand notation for the conditional distribution

$p_{Y|\Theta}(y|\Theta = \theta)$.

The goal here is to find an estimator $\hat{\theta}(y)$ that minimizes the Bayes risk.
Let’s use Bayes’ rule to rewrite our conditional density

\[ p_\theta(y) := p_{Y \mid \Theta}(y \mid \Theta = \theta) = \frac{p_{Y, \Theta}(y, \theta)}{p_\Theta(\theta)} = \frac{p_{\Theta \mid Y}(\theta \mid Y = y)p_Y(y)}{p_\Theta(\theta)} = \frac{\pi_y(\theta)p(y)}{\pi(\theta)} \]

Hence, the Bayes risk can be written as

\[ r(\hat{\theta}) = \int_Y \int_\Lambda C_\theta(\hat{\theta}(y))p_\theta(y)p(\theta)\pi(\theta) \, d\theta \, dy \]

\[ = \int_Y \underbrace{\int_\Lambda C_\theta(\hat{\theta}(y))\pi_y(\theta) \, d\theta}_{\text{posterior cost of estimator } \hat{\theta}(y) \text{ when } Y=y} \, p(y) \, dy \]

For purposes of estimation, we can think of \( y \) as fixed. The Bayes estimate of the true parameter \( \theta \) can be found by specifying an estimator (a function of \( y \)) that minimizes this posterior cost for each \( y \in Y \).
Minimizing the Bayes Risk

We want to minimize

\[ r(\hat{\theta}) = \int_{\mathcal{Y}} \int_{\Lambda} C_{\theta}(\hat{\theta}(y)) \pi_y(\theta) d\theta \quad p(y) dy. \]

posterior cost of estimator \( \hat{\theta}(y) \) when \( Y = y \)

To do this, we can fix \( y \) and solve the minimization problem

\[ \hat{\theta}_{\text{opt}}(y) = \arg \min_{g(\cdot)} \int_{\Lambda} C_{\theta}(g(y)) \pi_y(\theta) d\theta \]

\[ = \arg \min_{g(\cdot)} \mathbb{E}[C_{\Theta}(g(y)) \mid Y = y] \]

for each \( y \in \mathcal{Y} \). The solution, of course, depends on our choice of \( C_{\theta}(\cdot) \).