Bayesian Estimation*

Clayton Scott

We are interested in estimating θ given the observation x. Naturally then, any estimation strategy will be based on the posterior distribution $p(\theta \mid x)$. Furthermore, we need a criterion for assessing the quality of potential estimators.

1 Loss

The quality of an estimate θ is measured by a real-valued **loss function**: $L\left(\theta, \theta\right)$. For example, squared error or quadratic loss is simply $L\left(\theta, \theta\right) = \left(\theta - \theta\right)^T \left(\theta - \theta\right)$

2 Expected Loss

Posterior Expected Loss:

$$E\left[L\left(\theta, \theta\right) \mid x\right] = \int L\left(\theta, \theta\right) p(\theta \mid x)\, d\theta$$

Bayes Risk:

$$\begin{aligned} E\left[L\left(\theta, \theta\right)\right] &= \int \int L\left(\theta, \theta\right) p(\theta \mid x)\, p(x)\, d\theta dx \\ &= \int \int L\left(\theta, \theta\right) p(x \mid \theta)\, p(\theta)\, dx d\theta \\ &= E\left[E\left[L\left(\theta, \theta\right) \mid x\right]\right] \end{aligned} \quad (1)$$

The "best" or optimal estimator given the data x and under a specified loss is given by

$$\theta = \underset{\theta}{argmin} E\left[L\left(\theta, \theta\right) \mid x\right]$$

Example 1: Bayes MSE

$$\text{BMSE}\left(\theta\right) \equiv \int \int \left(\theta - \theta\right)^2 p(\theta \mid x)\, d\theta p(x)\, dx$$

*Version 1.2: Dec 14, 2003 11:43 pm -0600
†http://creativecommons.org/licenses/by/1.0

Since $p(x) \geq 0$ for every x, minimizing the inner integral $\int (\theta - E[\theta])^2 p(\theta \mid x)\, d\theta = E\left[L\left(\theta, \theta\right) \mid x\right]$ (where $E\left[L\left(\theta, \theta\right) \mid x\right]$ is the posterior expected loss) for each x, minimizes the overall BMSE.

$$\begin{aligned} \frac{\partial \int \left(\theta - \theta\right)^2 p(\theta \mid x) d\theta}{\partial \theta} &= \int \frac{\partial \left(\left(\theta - \theta\right)^2 p(\theta \mid x)\right)}{\partial \theta d\theta} \\ &= -2 \int \left(\theta - \theta\right) p(\theta \mid x)\, d\theta \end{aligned} \quad (2)$$

Equating this to zero produces

$$\theta = \int \theta p(\theta \mid x)\, d\theta \equiv E[\theta \mid x]$$

The conditional mean (also called **posterior mean**) of θ given x!

Example 2

$$\forall n, n \in \{1, \dots, N\} : (x_n = A + W_n)$$

$$W_n \sim \mathcal{N}\left(0, \sigma^2\right)$$

prior for unknown parameter A:

$$p(a) = U(-A_0, A_0)$$

$$p(x \mid A) = \frac{1}{\left(2\pi\sigma^2\right)^{\frac{N}{2}}} e^{\frac{-1}{2\sigma^2} \sum_{n=1}^{N} (x_n - A)^2}$$

$$p(A \mid x) = \begin{cases} \frac{\frac{1}{2A_0 (2\pi\sigma^2)^{\frac{N}{2}}} e^{\frac{-1}{2\sigma^2} \sum_{n=1}^{N} (x_n - A)^2}}{\int_{-A_0}^{A_0} \frac{1}{2A_0 (2\pi\sigma^2)^{\frac{N}{2}}} e^{\frac{-1}{2\sigma^2} \sum_{n=1}^{N} (x_n - A)^2} dA} & \text{if } |A| \leq A_0 \\ 0 \text{ if } |A| > A_0 \end{cases}$$

Minimum Bayes MSE Estimator:

$$\begin{aligned} A &= E[A \mid x] \\ &= \int_{-\infty}^{\infty} a p(A \mid x)\, dA \\ &= \frac{\int_{-A_0}^{A_0} A \frac{1}{2A_0 (2\pi\sigma^2)^{\frac{N}{2}}} e^{\frac{-1}{2\sigma^2} \sum_{n=1}^{N} (x_n - A)^2} dA}{\int_{-A_0}^{A_0} \frac{1}{2A_0 (2\pi\sigma^2)^{\frac{N}{2}}} e^{\frac{-1}{2\sigma^2} \sum_{n=1}^{N} (x_n - A)^2} dA} \end{aligned} \quad (3)$$

Notes

1. No closed-form estimator
2. As $A_0 \to \infty$, $A \to \frac{1}{N} \sum_{n=1}^{N} x_n$
3. For smaller A_0, truncated integral produces an A that is a function of x, σ^2, and A_0
4. As N increases, $\frac{\sigma^2}{N}$ decreases and posterior $p(A \mid x)$ becomes tightly clustered about $\frac{1}{N} \sum x_n$. This implies $A \to \frac{1}{N} \sum_n x_n$ as $n \to \infty$ (the data "swamps out" the prior)

3 Other Common Loss Functions

3.1 Absolute Error Loss

(Laplace, 1773)

$$L\left(\theta,\theta\right) = |\theta- \theta\,|$$

$$\begin{array}{rcl} E\left[L\left(\theta,\theta\right)\mid x\right] & = & \int_{-\infty}^{\infty} |\theta- \theta\,|p\left(\theta\mid x\right)d\theta \\ & = & \int_{-\infty}^{\theta\left(\theta-\theta\right)p(\theta\mid x)d\theta+\int_{\theta}^{\infty}\left(\theta-\theta\right)p(\theta\mid x)d\theta} \end{array} \quad (4)$$

Using integration-by-parts it can be shown that

$$\int_{-\infty}^{\theta\left(\theta-\theta\right)p(\theta\mid x)d\theta=\int_{-\infty}^{\theta P(\theta<y\mid x)dy}}$$

$$\int_{\theta}^{\infty}\left(\theta-\theta\right)p(\theta\mid x)d\theta=\int_{\theta}^{\infty}P(\theta>y\mid x)dy$$

where $P\left(\theta < y\mid x\right)$ and $P\left(\theta > y\mid x\right)$ are a cumulative distributions. So,

$$E\left[L\left(\theta,\theta\right)\mid x\right] = \int_{-\infty}^{\theta P(\theta<y\mid x)dy+\int_{\theta}^{\infty}P(\theta>y\mid x)dy}$$

Take the derivative with respect to θ implies $P\left(\theta <\theta\mid x\right) = P\left(\theta >\theta\mid x\right)$ which implies that the optimal θ under absolute error loss is **posterior median**.

3.2 '0-1' Loss

$$L\left(\theta,\theta\right) = \begin{cases} 0 & \text{if} \quad \theta = \theta \\ 1 & \text{if} \quad \theta \neq \theta \end{cases} = I_{\{\hat{\theta}\neq\theta\}}$$

$$E\left[L\left(\theta,\theta\right)\mid x\right] = E\left[I_{\{\hat{\theta}\neq\theta\}}\mid x\right] = Pr\left[\theta \neq \theta\ \mid x\right]$$

which is the probability that $\theta \neq \theta$ given x. To minimize '0-1' loss we must choose θ to be the value of θ with the highest posterior probability, which implies $\theta \neq \theta$ with the smallest probability.

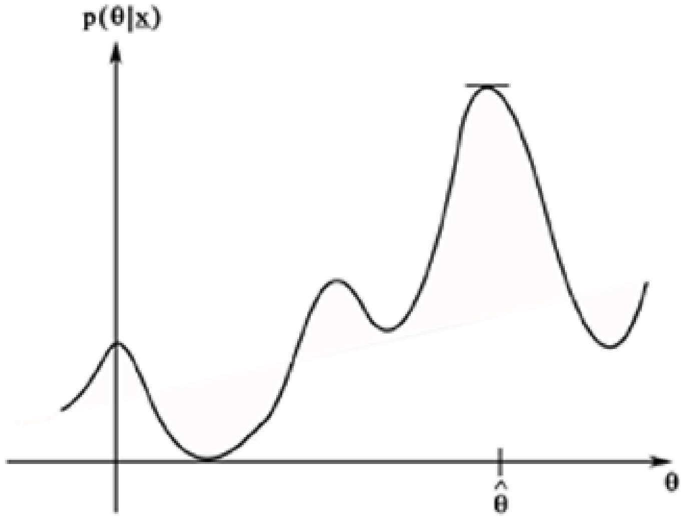

Figure 1

The optimal estimator θ under '0-1' loss is the **maximum *a posteriori*** (MAP) estimator–the value of θ where $p\left(\theta \mid x\right)$ is maximized.