A Posterior Mode is the Bayes Estimator Under Zero-One Loss

Suppose that we have a scalar parameter, θ, and an estimator, $\hat{\theta}$. Also, suppose that we have a "zero-one" loss function (sometimes called a "step loss" function):

$$L[\theta,\hat{\theta}] = 0 \quad ; \quad |\theta - \hat{\theta}| < \varepsilon$$
$$= 1 \quad ; \quad |\theta - \hat{\theta}| \geq \varepsilon$$

where ε is a small positive number.

Let our posterior p.d.f. be $p(\theta \mid y)$, where y is a vector of observations for our random data. We want to prove that the $\hat{\theta}$ that minimizes posterior expected loss will correspond to a mode of $p(\theta \mid y)$.

To prove this result we need to recall how to differentiate an integral when the range of integration depends on the variable with respect to which we are differentiating.

Lemma: ("Leibniz's Rule")

Let $$\phi(\alpha) = \int_{u_1}^{u_2} f(x,\alpha)dx \quad ; \quad a \leq \alpha \leq b$$

where u_1 and u_2 may depend on α. Then

$$\frac{d\phi}{d\alpha} = \int_{u_1}^{u_2} \frac{\partial f}{\partial \alpha} dx + f(u_2,\alpha)\frac{du_2}{d\alpha} - f(u_1,\alpha)\frac{du_1}{d\alpha}$$

for $a \leq \alpha \leq b$. ■

Now, if the support of the posterior pd.f. is [A , B], then the posterior expected loss is:

$$EL = \int_A^B L[\theta,\hat{\theta}]p(\theta \mid y)d\theta = \int_{\hat{\theta}+\varepsilon}^{B} p(\theta \mid y)d\theta + \int_A^{\hat{\theta}-\varepsilon} p(\theta \mid y)d\theta \ .$$

To minimize EL, we need to solve the following equation for $\hat{\theta}$:

$$\frac{d}{d\hat{\theta}} \int_{\hat{\theta}+\varepsilon}^{B} p(\theta \mid y)d\theta + \frac{d}{d\hat{\theta}} \int_A^{\hat{\theta}-\varepsilon} p(\theta \mid y)d\theta = 0 \ . \qquad (1)$$

Applying Leibniz's rule, the first term on the LHS of (1) is

$$\int_{\hat{\theta}+\varepsilon}^{B} \frac{d}{d\hat{\theta}} p(\theta \mid y) d\theta + p(B \mid y)\frac{dB}{d\hat{\theta}} - p(\hat{\theta}+\varepsilon \mid y)\frac{d(\hat{\theta}+\varepsilon)}{d\hat{\theta}},$$

which is just $-p(\hat{\theta}+\varepsilon \mid y)$.

Applying Leibniz's rule, the second term on the LHS of (1) is

$$\int_{A}^{\hat{\theta}-\varepsilon} \frac{d}{d\hat{\theta}} p(\theta \mid y) d\theta + p(\hat{\theta}-\varepsilon \mid y)\frac{d(\hat{\theta}-\varepsilon)}{d\hat{\theta}} - p(A \mid y)\frac{dA}{d\hat{\theta}} ,$$

which is just $p(\hat{\theta}-\varepsilon \mid y)$.

So, the condition in (1) becomes:

$$p(\hat{\theta}-\varepsilon \mid y) = p(\hat{\theta}+\varepsilon \mid y) . \qquad (2)$$

The value of $\hat{\theta}$ that satisfies (2) can be found graphically, as follows (Leonard and Hsu, 1999, pp.158-159):

1. Plot the posterior pd.f., and suppose that it is uni-modal.
2. Choose a small positive value for ε.
3. Draw a horizontal line, parallel to the θ axis. Raise or lower the line until the distance between the two points where the line intersects $p(\theta \mid y)$ is 2ε.
4. Drop a vertical line down from the mid-point of this line of length 2ε, so that it crosses the horizontal axis at $\theta = \hat{\theta}$.
5. If ε is made arbitrarily small, then $\hat{\theta}$ will locate the mode of the posterior pd.f.

Note that if the posterior is both uni-modal and symmetric, then this method will locate the mode for *any* (positive) choice of ε. In this case, of course the mode, median and mean (if it exists) of the posterior p.d.f. will all coincide.

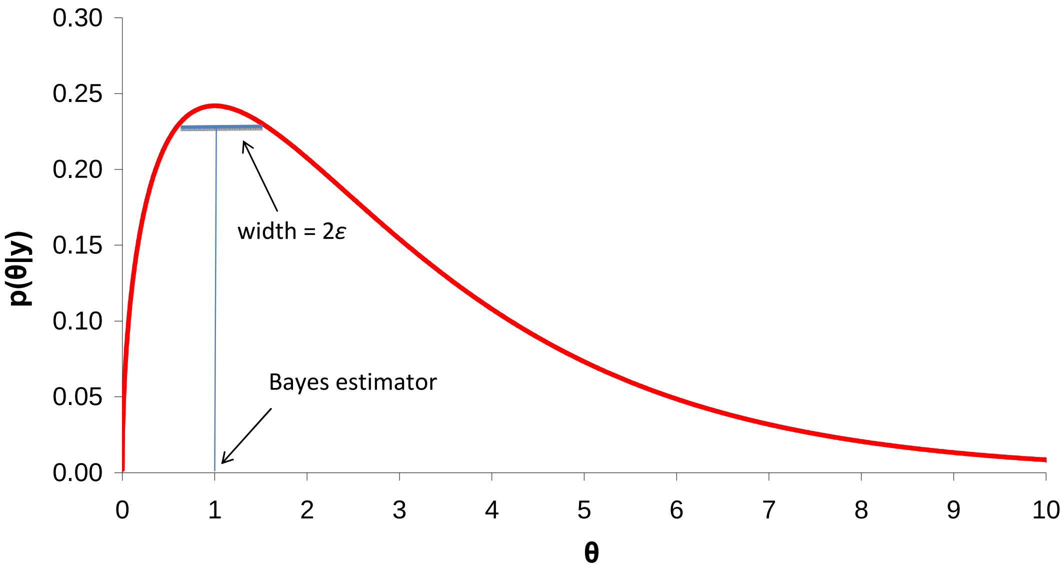
If the posterior density is multi-modal, the method above will locate local turning points in the density. Which turning points are found will depend on the choice of ε. In the multi-modal case, the value of the posterior expected loss will have to be computed for each "solution", and this will determine the choice of $\hat{\theta}$.

Figure 1 illustrates this procedure for the (asymmetric) uni-model case; and Figure 2 illustrates it for the case of a bi-modal posterior density. The latter is a mixture of two normals, with means of 1 and 4, and unit variances. The weights are 0.38 and 0.62 respectively. In Figure 2, the two modes occur at $\hat{\theta}_1 = 1$ and $\hat{\theta}_3 = 4$. There is a local minimum at $\hat{\theta}_2 = 2.2$. It can be shown (Leonard and Hsu, 1999, p.158) that in this particular case the value of the posterior expected loss is least when $\hat{\theta}_3 = 4$ is chosen as the estimator of θ.

Reference

Leonard, T. and J. S. J. Hsu (1999). *Bayesian Methods*. Cambridge University Press, Cambridge.

Figure 1: Uni-Modal Posterior p.d.f.



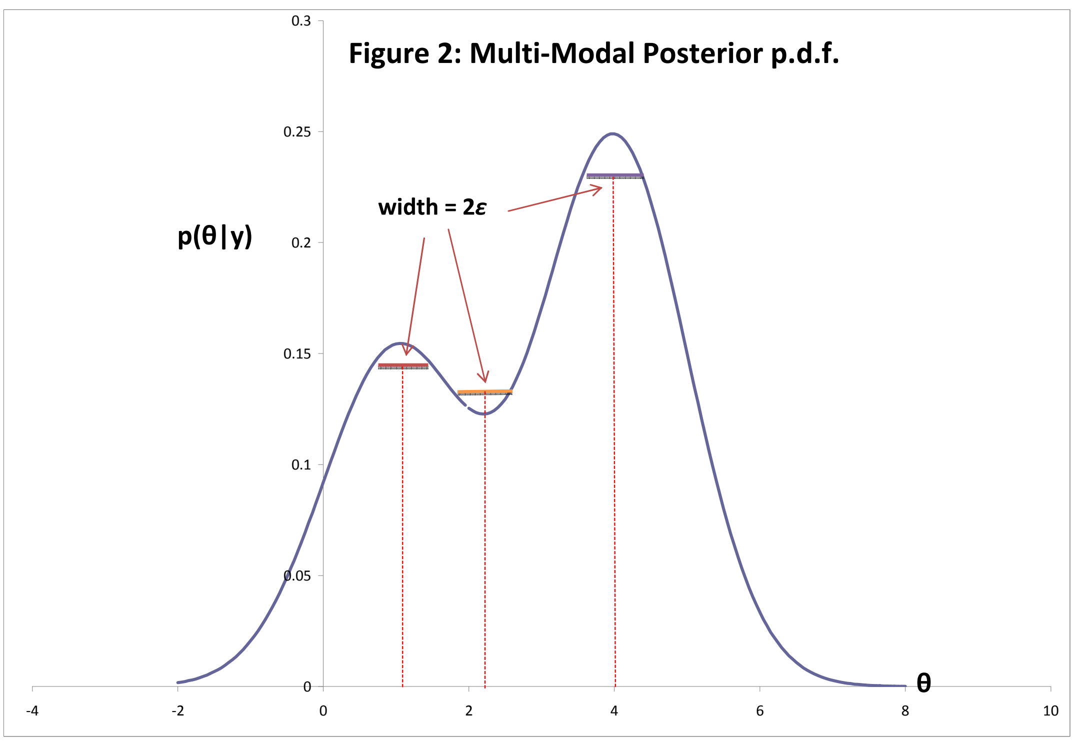


Figure 2: Multi-Modal Posterior p.d.f.

David E. Giles. Department of Economics, University of Victoria