MAP Estimation is not a limit of Bayes Estimation

Robert Bassett *

Department of Mathematics
University of California Davis

In this article I will provide a counterexample to an often-cited result in estimation theory: that every maximum a posteriori (MAP) estimator is the limit of a sequence of Bayes Estimators.

Maximum a posteriori and Bayes estimators both seek to the find an estimate of an unobserved parameter x based on some observation z. The MAP estimator is defined as the mode of the posterior distribution $x \to f(x|z)$, so that

$$\hat{x}_{MAP}(z) = \text{argmax}_x f(x|z).$$

Bayes estimation seeks to minimize expected loss. Let $L : X \to \mathbb{R}$ be a loss function, and let $\hat{x} = \hat{x}(z)$ be an estimator of x based on some measurement z. An estimator $\hat{x}$ is said to be a Bayes estimator of x if it minimizes the Bayes risk functional

$$\mathbb{E}[L(x - \hat{x})].$$

Since $\hat{x}$ is measurable w.r.t. the observation variable z, an equivalent definition of a Bayes estimator is that it minimizes

$$\mathbb{E}[L(x - \hat{x})|z]$$

for each measurement z.

False Claim: Let $f : \mathbb{R}^n \to \mathbb{R}$ be a posterior density with a unique mode $\hat{x}_{MAP}$. Let $L_c : \mathbb{R}^n \to \mathbb{R}$ be a loss function parametrized by c, defined as

$$L_c(x) = \begin{cases} 0 & \text{if } ||x|| < \epsilon \\ 1 & \text{otherwise} \end{cases}$$

Let $\hat{x}_c$ be the Bayes estimate of x with respect to loss L_c, i.e.

$$\hat{x}_c = \text{argmin}_{\hat{x}} \mathbb{E}[L_c(x - \hat{x})].$$

*rbassett@math.ucdavis.edu

Then

$$\lim_{c\to 0} \hat{x}_c = \hat{x}_{MAP}.$$

This claim is widely circulated in the estimation community [3][4.1.2] [2][2.2.4] [1][1.5.2] where the proof goes as follows:

$$\begin{aligned}
\hat{x}_c &\in \text{argmin}_{\hat{x}}\ \mathbb{E}[L_c(x-\hat{x})] \\
&= \text{argmin}_{\hat{x}} \int_{s\in\mathbb{R}^n} L_c(s-\hat{x}) f(s) ds \\
&= \text{argmin}_{\hat{x}} \left(1 - \int_{||s-\hat{x}||<c} f(s) ds\right) \\
&= \text{argmax}_{\hat{x}} \int_{||s-\hat{x}||<c} f(s) ds
\end{aligned} \tag{1}$$

Therefore

$$\begin{aligned}
\lim_{c\to 0} \hat{x}_c &= \lim_{c\to 0} \text{argmax}_{\hat{x}} \int_{||s-\hat{x}||<c} f(x) ds \\
&= \text{argmax}_{\hat{x}} f(\hat{x})
\end{aligned}$$

It is this last equality that is unjustified.

The Counterexample

Consider a posterior density $f : \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \begin{cases} 1 - \sqrt{-x} & x \in [-1, 0] \\ 1 - \frac{1}{2^n} & x \in [n, n + \frac{1}{2^n}]\ \forall n \in \mathbb{N}_+ \end{cases}$$

First we will verify that this is indeed a density.

$$\begin{aligned}
\int_{-\infty}^{\infty} f(s) ds &= \int_{-1}^{0} 1 - \sqrt{-s} ds + \sum_{n=1}^{\infty} \frac{1}{2^n} \left(\frac{2^n - 1}{2^n}\right) \\
&= \frac{1}{3} + \sum_{n=1}^{\infty} \frac{1}{2^n} - \frac{1}{4^n} \\
&= \frac{1}{3} + 1 - \frac{1}{3} = 1
\end{aligned}$$

The mode of f exists, is unique, and is 0, where the maximum of 1 is attained. Therefore $\hat{x}_{MAP} = 0$. From (1), with loss L_c the Bayes error is minimized when the expression

$$\int_{||s-\hat{x}||<c} f(s) ds \tag{2}$$

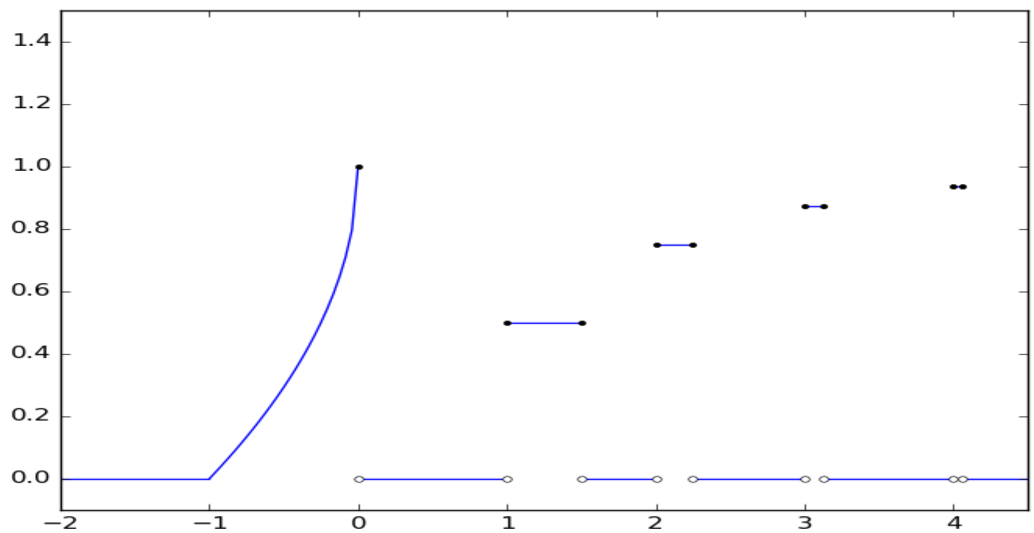

Figure 1: The Posterior Density f

is maximized. Hence

$$\hat{x}_c \in \text{argmax}_{\hat{x}} \int_{||s-\hat{x}||<c} f(s)ds.$$

Take (c^ν) to be the sequence $(\frac{1}{4^\nu})$, $\nu \in \mathbb{N}_+$. We will prove that $\hat{x}_{c^\nu} \nrightarrow \hat{x}_{MAP}$, which shows that the claim is false. In order to find the maximum in (2), we can consider the maximum on each continuous piece of f. For the piece defined on $[-1, 0]$, since the function is monotonically increasing we see that the maximum occurs at $\frac{-1}{2^{2\nu+1}}$, where (2) is

$$\int_{-\frac{1}{4^\nu}}^{0} 1 - \sqrt{-s}ds = \frac{1}{4^\nu} - \frac{2}{3}\left(\frac{1}{4^\nu}\right)^{3/2} = \frac{1}{4^\nu} - \frac{2}{3 \cdot 8^\nu}. \tag{3}$$

On the other hand, on the continuous piece of f defined on $[2^\nu, 2^\nu + \frac{1}{2^{2\nu}}]$, the maximizer of (2) occurs at $2\nu + \frac{1}{2^{2\nu+1}}$, the value of (2) equals

$$\frac{1}{2^{2\nu}}\left(1 - \frac{1}{2^{2\nu}}\right) = \frac{1}{4^\nu} - \frac{\left(\frac{1}{2^\nu}\right)}{8^\nu} \tag{4}$$

When $\nu > 1$, the Bayes estimate $\hat{x}_{c^\nu}$ is not in $[-1, 0]$ because (4) is strictly greater than (3). We have shown that there is a point outside of $[-1, 0]$ with Bayes error stricly less than the minimum Bayes error in this region.

With a little more effort, we can show that the Bayes estimate $\hat{x}_{c^\nu} = 2\nu + \frac{1}{2^{2\nu+1}}$. We consider the other continuous pieces of f and divide into two cases; the region $[n, n + \frac{1}{2^n}]$ for $n < 2\nu$, and $[n, n + \frac{1}{2^n}]$ for $n > 2\nu$. The Bayes estimate $\hat{x}_{c^\nu}$ obviously does not occur where f is zero, because integration is increasing with respect to increasing functions. Hence it suffices to consider these two cases.

Case 1: $n < 2\nu$. The maximum of (2) in this region occurs (non-uniquely) at the midpoint $n + \frac{1}{2^{n+1}}$ of the interval $[2n, 2n + \frac{1}{2^n}]$. The value of (2) is

$$\frac{1}{2^{2\nu}}\left(1 - \frac{1}{2^n}\right) < \frac{1}{2^{2\nu}}\left(1 - \frac{1}{2^{2\nu}}\right) = \frac{1}{4^\nu} - \frac{\left(\frac{1}{2^\nu}\right)}{8^\nu}$$

Where the inequality follows from $n < 2\nu$. By (4), the Bayes error at any point in this region is strictly greater than that at $2\nu + \frac{1}{2^{2\nu+1}}$.

Case 2: $n > 2\nu$. Again, $n + \frac{1}{2^{n+1}}$ is a non-unique maximizer of (2) in this region. The value of (2) is

$$\frac{1}{2^n}\left(1 - \frac{1}{2^n}\right) < \frac{1}{2^{2\nu}}\left(1 - \frac{1}{2^{2\nu}}\right).$$

The inequality comes from basic calculus: the function $x(1-x)$ is strictly increasing when $x < \frac{1}{2}$. Similar to case 1, this shows that the Bayes estimate $\hat{x}_{c^\nu} \notin [n, \frac{1}{2^{n+1}}]$.

We have therefore shown that $\hat{x}_{c^\nu} = 2\nu + \frac{1}{2^{\nu+1}}$, so that $\lim_{\nu\to\infty} \hat{x}_{c^\nu} \to \infty$. We conclude that $\lim_{c\to\infty} \hat{x}_c \neq \hat{x}_{MAP}$, and have established that the claim is false.

References

[1] Mário AT Figueiredo. “Lecture notes on bayesian estimation and classification”. In: *Instituto de Telecomunicacoes-Instituto Superior Tecnico* (2004).

[2] Wolfgang Hoegele et al. “Bayesian Estimation Applied to Stochastic Localization with Constraints due to Interfaces and Boundaries”. In: *Mathematical Problems in Engineering* 2013 (2013).

[3] Christian Robert. *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media, 2007.