Exercise session 6

Maximum a posteriori estimation

ELEC-E5440 Statistical Signal Processing

Robin Rajamäki

December 1, 2021

An example of Bayesian estimation

See Example 11.2 on p. 351 in Kay's book [1] for an intuitive understanding of the interplay between the prior, likelihood, and posterior. In the following, we solve a similar problem, namely Exercise 11.4 in [1, p. 370].

11.4 The data $x[n] = A + w[n]$ for $n = 0, 1, \ldots, N-1$ are observed. The unknown parameter A is assumed to have the prior PDF

$$p(A) = \begin{cases} \lambda \exp(-\lambda A) & A \geqslant 0 \\ 0 & A < 0 \end{cases}$$

where $\lambda > 0$, and $w[n]$ is WGN with variance σ^2 and is independent of A. Find the MAP estimator of A.

Solution: Since the noise is Gaussian, its PDF is

$$p(w[n]) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{w^2[n]}{2\sigma^2} \Big).$$

Consequently, the conditional probability of $x[n]$ given A is also a Gaussian process with mean A, i.e.,

$$p(x[n]|A) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x[n]-A)^2}{2\sigma^2} \right)$$

For simplicity, collect the observations into vector $\boldsymbol{x} = [x[0], x[1], \ldots, x[n-1]]^{\mathrm{T}}$. The likelihood of the i.i.d. observations is therefore the conditional PDF

$$p(\boldsymbol{x}|A) = \prod_{n=0}^{N-1} p(x[n]|A) = \frac{1}{\sqrt{2\pi\sigma^2}^N} \exp\left(-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n]-A)^2 \right).$$

The MAP estimator of A maximizes the posterior, that is,

$$\hat{A}_{\mathrm{MAP}} = \arg\max_A p(A|\boldsymbol{x}) = \arg\max_A p(\boldsymbol{x}|A)p(A) = \arg\max_A \big(\log p(\boldsymbol{x}|A) + \log p(A) \big).$$

For the given $p(A)$ and $p(\boldsymbol{x}|A)$, the MAP becomes (ignoring irrelevant constants)

$$\hat{A}_{\text{MAP}} = \arg\max_{A} \underbrace{\left(-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2 - \lambda A \right)}_{f(A)}.$$

Setting the derivative of $f(A)$ w.r.t. A equal to zero yields

$$\frac{df(A)}{dA} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n] - A) - \lambda = 0 \iff A = \bar{x} - \frac{\lambda\sigma^2}{N},$$

where $\bar{x} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$ denotes the sample mean. Indeed, the above argument maximizes $f(A)$, since

$$\frac{d^2 f(A)}{dA^2} = -\frac{N}{\sigma^2} - \lambda < 0.$$

However, we also need to take into account[1] that $A \geq 0$, since it may occur that $\bar{x} < \lambda\sigma^2/N$. Consequently, the MAP estimator is

$$\hat{A}_{\text{MAP}} = \max\left(0, \bar{x} - \frac{\lambda\sigma^2}{N}\right).$$

Fig. 1 illustrates the MAP estimate, together with the prior and posterior PDFs, as well as the likelihood function, for the case $N = 1$ and $x = 2.0558$. We see that the prior biases the MAP estimate towards smaller values of A compared to the mode of the likelihood function, i.e., the maximum likelihood estimate (MLE). The MLE equals the measurement x in this single observation case. In contrast, the minimum mean squared error (MMSE) estimator (not depicted) is, by definition, the mean of the posterior distribution.

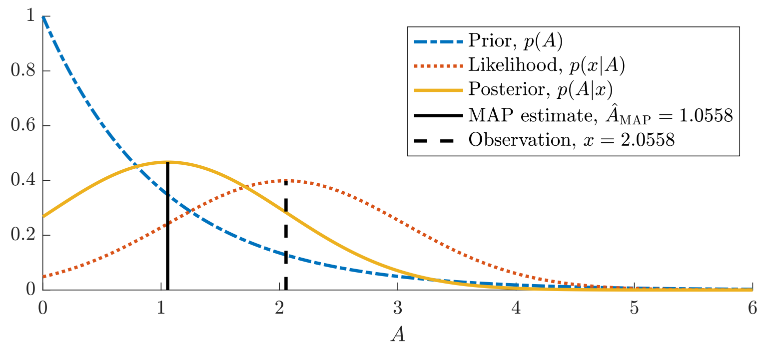

Figure 1: MAP estimate together with prior, likelihood, and posterior PDFs.

References

[1] S. M. Kay, *Fundamentals of statistical signal processing: Estimation theory.* Prentice Hall PTR, 1993.

[1] Since we previously solved an *unconstrained* optimization problem, for simplicity.