BML: exercise sheet

Stars indicate the difficulty level, from 1 to 3. One star means that everyone should be able to do it without too much effort.

1 Lecture 1

1.1 Conjugate priors 101: Gaussians ($\star$)

Let $y|\mu \sim \mathcal{N}(\mu, I_N)$ and $\mu \sim \mathcal{N}(0, aI_N)$, for some $a > 0$. Show that

$$\mu|x \sim \mathcal{N}(by, bI_N), \text{ where } b = a/(a+1). \tag{1}$$

1.2 A conjugate prior on probability vectors ($\star$)

Let

$$\Delta_d = \{\theta \in [0,1]^d \text{ such that } \sum_{k=1}^{d} \theta_d = 1\}.$$

Let further $\alpha \in (\mathbb{R}_+)^d$. The Dirichlet pdf is defined by

$$\mathrm{Dir}(\theta|\alpha) = \frac{1}{B(\alpha)} \prod_{k=1}^{d} \theta_k^{\alpha_k - 1} 1_{\theta \in \Delta_d},$$

where $B(\alpha) = \prod_{k=1}^{d} \Gamma(\alpha_k)/\Gamma(\sum_{k=1}^{d} \alpha_k)$ is the so-called beta function. To what likelihood is the Dirichlet conjugate? What is the posterior?

1.3 Estimating the mean of a Gaussian ($\star\star$)

Let $\mu = (\mu_1, \dots, \mu_N) \in \mathbb{R}^N$, and consider N i.i.d. real variables $y_i|\mu \sim \mathcal{N}(\mu_i, 1)$. We wish to infer μ.

1. What is the maximum likelihood estimator $\hat{\mu}_{\mathrm{MLE}}$?
2. Henceforth, we judge estimators by the square loss. The frequentist risk of an estimator $\hat{\mu}$ is
$$R(\hat{\mu}) = \mathbb{E}_{y|\mu} \|\mu - \hat{\mu}\|^2.$$
show that $R(\hat{\mu}_{\mathrm{MLE}}) = N$.

3. Suppose we have prior belief that μ lies near 0, and we choose to represent it by $\mu \sim \mathcal{N}(0, aI_N)$, $a > 0$. What is the Bayes estimator $\hat{\mu}_{\text{Bayes}}$? What is its (frequentist) risk $R(\hat{\mu}_{\text{Bayes}})$? What is its Bayes risk $\mathbb{E}_\mu R(\hat{\mu}_{\text{Bayes}})$?

4. Since we actually have no idea what a should be, we propose to estimate it from data using empirical Bayes. Show that the marginal of y is

$$\int p(y, \mu)\mathrm{d}\mu = \mathcal{N}(y|0, (a+1)I_N).$$

In particular, what is the law of $S = \|y\|^2$? Deduce from it that $(N-2)/S$ is an unbiased estimator of $a+1$, and consider the empirical Bayes estimator

$$\hat{\mu}_{\text{EB}} = \left(1 - \frac{N-2}{S}\right) y.$$

What is its Bayes risk?

5. (harder, but elementary; do this only if you have solved all the preceding exercises; see Efron, 2012, Section 1.2 for a solution) Show that for $N \geqslant 3$, for every $\mu \in \mathbb{R}^N$,

$$R(\hat{\mu}_{\text{EB}}) < R(\hat{\mu}_{\text{MLE}}). \quad (2)$$

Frequentists say that $\hat{\mu}_{\text{EB}}$ dominates μ_{MLE}, in the sense that whatever the value of μ, the risk of $\hat{\mu}_{\text{EB}}$ is the smallest of the two. This happens even when μ is far from zero, in which case one might have thought that our $\mathcal{N}(0, aI_N)$ prior would have been a poor choice. Finally, if you are a strict Waldian, you should thus prefer $\hat{\mu}_{\text{EB}}$ to $\hat{\mu}_{\text{MLE}}$. Many people still use $\hat{\mu}_{\text{MLE}}$; see Efron, 2012, Section 1.3 for a tentative answer.

Equation 2 is called the James-Stein effect, and is a standard example of why following Bayesian guidelines can end up giving good frequentist estimators. Shrinkage like in $\hat{\mu}_{\text{EB}}$ are now commonplace in large-dimensional, penalized regression. For more on frequentist guarantees for Bayesian estimators and shrinkage, see (Parmigiani and Inoue, 2009, Sections 7, 8, 9).

1.4 For more exercises on Bayesian derivations

- Exercises 5.1 to 5.4 of (Murphy, 2012).
- Go through Sections 4.4 to 4.6 of (Murphy, 2012) with pen and paper. Linear Gaussian models appear all the time.
- Exercises 2.6, 2.9, 2.10, 2.13, 2.14, and 2.15 of (Marin and Robert, 2007). Solutions are here.

References

[1] B. Efron. *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*. Vol. 1. Cambridge University Press, 2012.

[2] J.-M. Marin and C.P. Robert. *Bayesian Core: A Practical Approach to Computational Bayesian Statistics*. New York: Springer-Verlag, 2007.

[3] K. Murphy. *Machine learning: a probabilistic perspective*. MIT Press, 2012.

[4] G. Parmigiani and L. Inoue. *Decision theory: principles and approaches*. Vol. 812. John Wiley & Sons, 2009.