Applications - Lecture 6

Introduction to Bayesian Estimation

Ingénierie Mathématique & Informatique

Section	Notion	Used in
6.1	Prior Q, posterior $Q(\cdot \mid x_n)$, Bayes formula	App. 1, App. 2
6.1.3	Conjugate prior (Beta-Bernoulli)	App. 1
6.2.1	Posterior mean (PM) and MAP	App. 1, App. 2
6.2.3	Credible intervals	App. 1

1. **Update sequentially** a success probability: adaptive clinical trials with the Beta-Bernoulli conjugate pair.
2. **Regularise and quantify uncertainty**: a Gaussian conjugate prior turns MAP into ridge training of an interatomic potential, with credible bands on new predictions.

App. 1 - Sequential Bayesian Updating in Adaptive Clinical Trials

Tracking a response rate as data accumulate

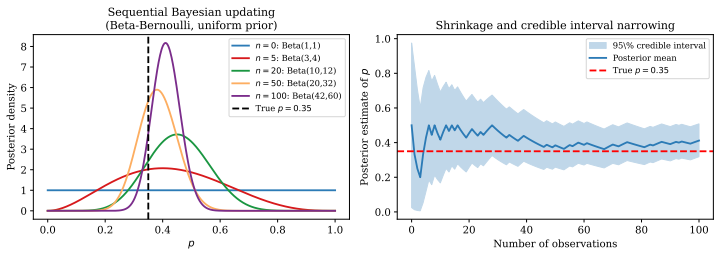
Problem. In an adaptive clinical trial, the credibility of each value of the response rate $p \in (0,1)$ of a new treatment must be updated after every patient, in order to decide whether to continue, stop for efficacy, or stop for futility.

Beta-Bernoulli conjugate pair. Put the prior $p \sim \beta(a,b)$, with density $q_{a,b}(u) = \frac{1}{B(a,b)} u^{a-1}(1-u)^{b-1}$ on $[0,1]$. Given $X_1, \dots, X_n \overset{\text{iid}}{\sim} \text{Bernoulli}(p)$ and $s = X_1 + \cdots + X_n$, the posterior is again Beta:

$$Q_{a,b}(\cdot \mid x_n) = \beta(s + a,\ n - s + b).$$

The family $\{\beta(a,b),\ a,b > 0\}$ is thus a conjugate prior for the Bernoulli model.

Posterior concentration and take-aways



Left: posterior density of p after $n \in \{0, 5, 20, 50, 100\}$ observations (true $p = 0.35$, prior $\beta(1, 1)$).

Right: posterior mean $\hat{p}_n^{\mathrm{PM}}$ and 95% credible interval as n grows.

- The PM $\hat{p}_n^{\mathrm{PM}} = (s + a)/(n + a + b)$ shrinks the MLE s/n towards the prior mean $a/(a + b)$; the shrinkage vanishes as $n \to \infty$.
- The credible interval $[q_{0.025}, q_{0.975}]$ has a direct posterior-probability reading, unlike a frequentist confidence interval.

App. 2 - Bayesian Training of an Interatomic Potential

A prior for a data-scarce potential

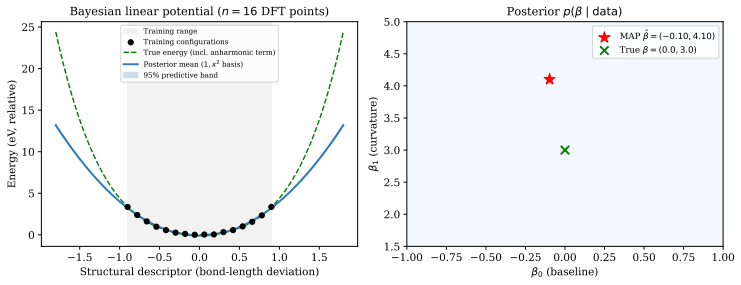
Problem. DFT calculations are expensive, so a linear interatomic potential is often trained on few configurations n relative to its basis size p; OLS then overfits, and predictions on new configurations carry no error bar.

Model. In the Gaussian linear model $y_n = x_n\beta + \varepsilon_n$ (y_i = DFT energy, x_i = descriptor vector), $\varepsilon_n \sim \mathcal{N}_n(0, \sigma^2 I_n)$ with σ^2 known, put the conjugate prior $\beta \sim \mathcal{N}_{p+1}(0, kI_{p+1})$. Taking the log-posterior,

$$\hat{\beta}^{\mathrm{MAP}} = \arg\min_{\beta}\Big\{\|y_n - x_n\beta\|^2 + h\|\beta\|^2\Big\}, \qquad h = \frac{\sigma^2}{k}.$$

So the MAP under this Gaussian prior is exactly the ridge-trained potential, with ridge parameter $h = \sigma^2/k$.

Predictive uncertainty and take-aways



Left: posterior mean energy and 95% predictive band from $n = 16$ toy DFT points, fitted with a basis missing a true anharmonic term. Right: joint posterior of (β_0, β_1); red star $= \hat{\beta}^{\mathrm{MAP}}$.

- The full posterior on β yields a predictive credible band on new configurations.
- Outside the training range the true curve escapes this band: a narrow posterior only reflects parameter uncertainty, not missing basis functions, so extrapolation still needs care.