Introduction and learning objectives

In Lecture 10, we approximated a posterior by choosing a family of distributions $\mathcal{Q}$ and maximising the evidence lower bound (ELBO) over its members. Here we practise the recipe, examine a limitation of a factorised approximation, and fit a posterior using samples.

By the end of the tutorial, you should be able to:

- Construct an ELBO and fit a Gaussian using analytic expectations.
- Explain how an independence assumption changes posterior uncertainty.
- Fit a non-Gaussian family using reparameterised samples and interpret the result.

The recipe. Write the joint; choose a variational family $\mathcal{Q} = \{q_\phi : \phi \in \Phi\}$; construct the ELBO for $q_\phi \in \mathcal{Q}$; evaluate its expectations; optimise ϕ; assess the fitted approximation. The model, observations and family are fixed during optimisation. All logarithms are natural.

Problem 1. Inferring an event rate. Suppose we count the number of events in a fixed period, such as calls arriving at a help desk. We want to infer the underlying event rate from this count and describe our uncertainty about it. Write the positive rate as $\lambda = e^z$ and use the Poisson model

$$p(z) = \mathcal{N}(z; 0, 1), \qquad p(x \mid z) = \frac{\exp(xz - e^z)}{x!}, \quad x \in \{0, 1, \ldots\}.$$

We observe $x = 4$ and choose the Gaussian family

$$\mathcal{Q} = \{q_{\mu,v} : \mu \in \mathbb{R},\ v > 0\}, \qquad q_{\mu,v}(z) = \mathcal{N}(z; \mu, v).$$

Here v is the variance; write q for a generic member. Keep x symbolic in the derivation. You may use

$$\mathbb{E}_q[e^z] = e^{\mu + v/2}, \qquad D_{\mathrm{KL}}[q \| \mathcal{N}(0,1)] = \tfrac{1}{2}(\mu^2 + v - 1 - \log v).$$

(a) Write $p(z \mid x)$ up to a normalising constant. Which term prevents this posterior from being Gaussian?

(b) Derive an explicit expression for

$$\mathcal{L}(\mu, v; x) = \mathbb{E}_q[\log p(x \mid z)] - D_{\mathrm{KL}}[q(z) \| p(z)].$$

Explain why maximising this expression does not require evaluating $p(x)$.

(c) Differentiate with respect to μ and v, and write the two equations an interior optimum must satisfy. You do not need to solve them. How could we parameterise v to keep it positive during optimisation?

Problem 2. What does independence cost? A sensor measures the sum of two quantities, but cannot measure either one separately. Its reading also contains noise. We want to infer both quantities and see how assuming independence in our approximation affects our uncertainty. Our model is

$$z_1, z_2 \overset{\text{ind}}{\sim} \mathcal{N}(0,1), \qquad x \mid z_1, z_2 \sim \mathcal{N}(z_1 + z_2, \tau^2), \qquad \tau > 0.$$

The exact posterior means are both $m = x/(2+\tau^2)$; take this result as given. Fix the means at these values and choose the family $\mathcal{Q} = \{q_v : v > 0\}$, where

$$q_v(z_1, z_2) = \mathcal{N}(z_1; m, v)\mathcal{N}(z_2; m, v).$$

We fit only the shared variance v.

(a) Use independence under q_v to show that

$$\mathbb{E}_{q_v}[(x - z_1 - z_2)^2] = (x - 2m)^2 + 2v.$$

Using the Gaussian KL from Problem 1 for each factor, show that the ELBO has the form

$$\mathcal{L}(v) = C - \frac{v}{\tau^2} - v + \log v,$$

where C does not depend on v. Differentiate once to obtain

$$v^* = \frac{\tau^2}{1+\tau^2}.$$

(b) Define the sum $S = z_1 + z_2$ and difference $D = z_1 - z_2$. Under the prior, S and D are independent and both have distribution $\mathcal{N}(0, 2)$. The measurement is $x = S + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \tau^2)$ is independent noise.

Explain why observing x reduces uncertainty about S but leaves the distribution of D unchanged. You may use the scalar Gaussian result

$$\mathrm{Var}_p(S \mid x) = \frac{2\tau^2}{2+\tau^2}, \qquad \mathrm{Var}_p(D \mid x) = 2.$$

(c) Under the fitted q_{v^*}, what are the variances of S and D? Show, by comparing denominators, that

$$\mathrm{Var}_p(S \mid x) < \mathrm{Var}_{q_{v^*}}(S) = \mathrm{Var}_{q_{v^*}}(D) < \mathrm{Var}_p(D \mid x).$$

Which uncertainty is overstated and which is understated? As $\tau \to 0$, what happens to uncertainty about D under the exact posterior and under q_{v^*}?

Discussion. What could a Gaussian with a nonzero covariance between z_1 and z_2 represent that this factorised Gaussian cannot?

Problem 3. Fitting a Laplace family with samples. A robot succeeds in eight out of ten attempts at the same task. We want to infer its success probability and describe our uncertainty. This time, we will fit a Laplace distribution using samples. Let z be the log odds of success. Given z, the trials are independent, with $x = 8$ and $n = 10$:

$$p(z) = \mathcal{N}(z; 0, 1), \qquad p(x \mid z) = \binom{n}{x} \sigma(z)^x [1 - \sigma(z)]^{n-x}, \qquad \sigma(z) = \frac{1}{1 + e^{-z}}.$$

Choose the *Laplace* family $\mathcal{Q} = \{q_{\mu,a} : \mu, a \in \mathbb{R}\}$, where

$$q_{\mu,a}(z) = \frac{1}{2e^a} \exp\left(-\frac{|z - \mu|}{e^a}\right).$$

Here $b = e^a$ is the scale; the variance is $2b^2$. This family is not Gaussian. You may use

$$D_{\mathrm{KL}}[q_{\mu,a} \| p] = \frac{\mu^2}{2} + e^{2a} - a + c_0, \qquad c_0 = \tfrac{1}{2}\log(2\pi) - 1 - \log 2.$$

The notebook provides the log likelihood

$$\ell(z) = \log \binom{n}{x} + xz - n \log(1 + e^z)$$

and the optimiser. We will estimate $\mathbb{E}_q[\ell(z)]$ by sampling from q.

(a) Let $\epsilon^{(k)} \overset{\text{iid}}{\sim} \text{Laplace}(0, 1)$ and $z^{(k)} = \mu + e^a \epsilon^{(k)}$, where $\text{Laplace}(0, 1)$ has location zero and scale one. Write a K-sample estimate of the ELBO, using a sample average for the expected log likelihood and the supplied KL. Explain why this estimator is unbiased for fixed μ, a.

(b) Complete the three expressions in the notebook's `sampled_elbo` function: the reparameterised samples, the average log likelihood, and the prior KL. Run the supplied check and optimiser once with $K = 10$.

The notebook supplies the gradients, a numerical reference posterior, and two plots: the optimisation trace and the fitted density. Compare the peak and tails of the fitted Laplace density with the reference. Which steps of the VI recipe changed when we chose a different family?

(c) Discuss the following using your fit:

- Must every update increase the true ELBO?
- Does increasing K change which posterior shapes $\mathcal{Q}$ can represent?
- Could a sampled ELBO estimate exceed $\log p(x)$ even though the ELBO is a lower bound?

What is being sampled? The samples come from the current approximation q, not from the exact posterior. The numerical reference is provided only to help us check the fit.