Bayesian modelling

Final review

Léo Belzile

Last compiled Monday Apr 7, 2025

Fundamentals

- Bayesian inference uses likelihood based inference.
- It complements the likelihood $p(\boldsymbol{y} \mid \boldsymbol{\theta})$ with a prior $p(\boldsymbol{\theta})$.
- Provided that $p(\boldsymbol{\theta}, \boldsymbol{y})$ is integrable, we get

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}) \overset{\boldsymbol{\theta}}{\propto} p(\boldsymbol{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}).$$

Marginal likelihood

The normalizing constant

$$p(\boldsymbol{y}) = \int_{\boldsymbol{\Theta}} p(\boldsymbol{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}$$

to make the posterior a valid density is termed **marginal likelihood**.

Marginal likelihood

Moments of the posterior depend on $p(\boldsymbol{y})$.

It is hard to compute because $\boldsymbol{\Theta} \subseteq \mathbb{R}^p$, and the integral is often high-dimensional.

- Monte Carlo integration (does not typically work because prior need not align with likelihood)
- Numerical integration performance degrades with p, numerical overflow.

Bayes factors

The Bayes factor is the ratio of marginal likelihoods, as

$$p(\boldsymbol{y} \mid \mathcal{M}_i) = \int p(y \mid \boldsymbol{\theta}^{(i)}, \mathcal{M}_i) p(\boldsymbol{\theta}^{(i)} \mid \mathcal{M}_i) \mathrm{d}\boldsymbol{\theta}^{(i)}.$$

Values of $\mathsf{BF}_{ij} > 1$ correspond to model $\mathcal{M}_i$ being more likely than $\mathcal{M}_j$.

- Strong dependence on the prior $p(\boldsymbol{\theta}^{(i)} \mid \mathcal{M}_i)$.
- Must use proper priors.

Predictive distributions

Define the posterior predictive,

$$p(y_{\text{new}} \mid \boldsymbol{y}) = \int_{\boldsymbol{\Theta}} p(y_{\text{new}} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \boldsymbol{y}) \mathrm{d}\boldsymbol{\theta}$$

Bayesian inference

If we have samples from $p(\boldsymbol{\theta} \mid \boldsymbol{y})$ or an approximation of the joint/marginals, then we can

- use the **posterior** distribution to answer any question that is a function of $\boldsymbol{\theta}$ alone.
- use the **posterior predictive** $p(y_{\text{new}} \mid \boldsymbol{y})$ for prediction or forecasting, and checks of model adequacy.

Point estimators and credible regions

Interpretation is different from frequentist, but methods are similar:

- point estimators (MAP, posterior mean and median, etc.) derive from consideration of **loss functions** that return a summary of the posterior.
- credible interval or regions (interval for which the true parameter lies with a certain probability).

Stochastic approximations

Stochastic approximations rely on sampling methods (rejection sampling, MCMC)

- returns (correlated) posterior samples.
- Metropolis–Hastings acceptance ratio bypasses marginal likelihood calculation.
- Marginalization is straightforward.

Markov chains

- Need to assess convergence to the stationary distribution (traceplots)
- Autocorrelation reduces precision of Monte Carlo estimates (**effective sample size**)

Markov chain Monte Carlo algorithms

We covered in class the following (in decreasing order of efficiency).

- random walk Metropolis
- Metropolis-adjusted Langevin algorithm (MALA)
- Hamiltonian Monte Carlo

Better sampling performance, but the latter two require gradient and are more expensive to compute.

Model selection

- Bernstein-von Mises ensures convergence in total variation of the posterior under weak conditions.
- Distinguish between
  - $\mathcal{M}$-closed: true parameter is part of set considered or
  - $\mathcal{M}$-open: only misspecified models are considered.
- The model that gets selected minimizes the Kullback–Leibler divergence with the truth.
- In discrete parameter settings, we recover the truth with probability 1.

References