Bayes Model Selection and Assessment

Søren Lund Pedersen

In this article, I would like to discuss the Bayes factor, cross-validation, information criteria, posterior predictive checks, and stochastic search variable selection. The reason for exploring these topics is to understand how to select, assess, and improve Bayesian models to ensure they perform well on real-world data.

1 Bayes Factor

The Bayes factor (BF) is a metric for comparing two competing Bayesian models based on their predictive performance given the data. It can be thought of as a Bayesian alternative to the likelihood ratio test. Formally, the Bayes factor comparing model M_1 to model M_2 is defined as:

$$BF_{1,2} = \frac{P(D \mid M_1)}{P(D \mid M_2)},$$

where $P(D \mid M)$ is the marginal likelihood of the data D under model M.

- If $BF_{1,2} > 1$, the data provide stronger support for model M_1 over M_2.
- If $BF_{1,2} < 1$, the data provide stronger support for model M_2.

Bayes factors can be sensitive to the choice of priors, as the marginal likelihood integrates over the prior distribution. Hence, careful consideration of prior selection is crucial when using Bayes factors for model comparison.

2 Cross-validation

Cross-validation in a Bayesian context is conceptually similar to frequentist cross-validation. The key difference lies in the use of the posterior predictive distribution to evaluate model performance on unseen data. The process involves:

- Splitting the dataset into k folds.
- For each fold, training the model on $k - 1$ folds and testing on the remaining fold.
- Computing performance metrics (e.g., predictive accuracy, log-likelihood) using the posterior predictive distribution for the test fold.
- Averaging the performance metrics across all k folds.

While the mechanics of cross-validation remain similar, Bayesian cross-validation requires integrating over the posterior distribution, which can be computationally intensive. Nonetheless, it provides robust estimates of out-of-sample predictive performance.

3 Information Criteria

Information criteria offer ways to compare models by balancing goodness of fit and model complexity.

3.1 Deviance Information Criterion (DIC)

The Deviance Information Criterion is a Bayesian generalization of AIC. It is defined as:

$$DIC = \text{mean deviance} + p_D,$$

where

$$\text{mean deviance} = -2\mathbb{E}_{\boldsymbol{\theta}|\mathbf{y}}[\log P(\mathbf{y} \mid \boldsymbol{\theta})],$$

and

$$p_D = \mathbb{E}_{\boldsymbol{\theta}|\mathbf{y}}[-2 \log P(\mathbf{y} \mid \boldsymbol{\theta})] + 2 \log P(\mathbf{y} \mid \hat{\boldsymbol{\theta}}).$$

Here, p_D represents the effective number of parameters (model complexity). Lower DIC values suggest a better trade-off between model fit and complexity.

3.2 Watanabe-Akaike Information Criterion (WAIC)

WAIC is another Bayesian information criterion that fully averages over the posterior distribution, rather than relying on point estimates. It is defined as:

$$WAIC = -2 \sum_{i=1}^{n} \ln \left(\frac{1}{S} \sum_{s=1}^{S} p(y_i \mid \boldsymbol{\theta}^{(s)}) \right) + 2 \sum_{i=1}^{n} \text{Var}_s[\ln p(y_i \mid \boldsymbol{\theta}^{(s)})],$$

where $\boldsymbol{\theta}^{(s)}$ are samples from the posterior distribution and S is the number of samples. The first term measures model fit, while the second term penalizes model complexity. WAIC is asymptotically equivalent to Bayesian cross-validation and is more robust than DIC in many cases.

4 Posterior Predictive Checks

Posterior predictive checks (PPC) are diagnostic tools used to assess model fit by comparing observed data to data simulated from the model's posterior predictive distribution. The steps involved are:

1. Draw samples $\boldsymbol{\theta}^{(s)}$ from the posterior $p(\boldsymbol{\theta} \mid \mathbf{y})$.
2. For each $\boldsymbol{\theta}^{(s)}$, generate replicated data $\tilde{y}^{(s)}$ from the likelihood $p(y \mid \boldsymbol{\theta}^{(s)})$.
3. Compare summary statistics, plots, or other features of the replicated data $\tilde{y}^{(s)}$ to the observed data y.

If the model is appropriate, the replicated data should resemble the observed data. Discrepancies suggest model misfit, such as unaccounted skewness, overdispersion, or outliers.

5 Stochastic Search Variable Selection (SSVS) Using Spike-and-Slab Prior

Variable selection is crucial when dealing with high-dimensional data. **Stochastic Search Variable Selection** (SSVS) is a Bayesian technique for selecting relevant predictors by using a **spike-and-slab prior** on the regression coefficients.

The spike-and-slab prior for a coefficient β_j is a mixture of two distributions:

$$p(\beta_j \mid \gamma_j) = \begin{cases} \text{Spike: } \mathcal{N}(0, \tau_0^2) & \text{if } \gamma_j = 0, \\ \text{Slab: } \mathcal{N}(0, \tau_1^2) & \text{if } \gamma_j = 1, \end{cases}$$

with $\tau_0^2 \ll \tau_1^2$, meaning the spike component is concentrated near zero, and the slab component is more diffuse. Here, γ_j is a binary indicator that determines whether predictor j is included ($\gamma_j = 1$) or excluded ($\gamma_j = 0$) from the model.

SSVS proceeds by:

- Assigning spike-and-slab priors to each coefficient β_j.
- Using MCMC or other sampling methods to draw from the joint posterior of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$.
- Identifying predictors with high posterior probability of inclusion (i.e., high probability that $\gamma_j = 1$).

This approach effectively performs variable selection by shrinking irrelevant coefficients towards zero (the spike), while allowing relevant coefficients to vary freely (the slab). It integrates uncertainty in variable selection directly into the model, providing a principled way to handle model uncertainty and avoid overfitting.