Applied Probabilistic Programming & Bayesian Machine Learning

# §2-1: Model Checking

MIT IAP 2017 | Max W. Shen

In the first lecture, you became acquainted with Stan and built a variety of models. We also discussed a toolbox of prior distributions that are useful in practice. We largely discussed topics 1 and 2 in the **three steps of Bayesian data analysis**:

1. Write a model
2. Obtain posteriors by running inference
3. Check the model, and repeat from step 1 if needed

In this lecture, we focus on checking the validity of our models, topic 3.

## 1 All Models are Wrong, Only Some are Useful

Bayesian statistics is commonly criticized for relying too much on subjective prior knowledge for modeling. This is a key attribute, a great strength, and indeed a great weakness. Bayesian modeling is a powerful approach for machine learning given small-to-medium sized datasets and handles missing data with ease, due in part to the ability to incorporate substantive prior knowledge.

However, to build models that are worthy of our trust, it is crucial to check the assumptions embedded in the model - both the explicit assumptions and the implicit assumptions.

Assumptions, or prior knowledge, enter Bayesian modeling from two sources: (1) Choices of prior distributions over parameters, and (2) the choice of the model itself. The second source is typically far more flexible, powerful and important than the first source.

Similarly, uncertainty in posterior estimates that are output by Bayesian modeling comes from noise in the data due to finite random sampling. But we as interpreters of the model should have further uncertainty arising from uncertainty in the choice of model.

### 1.1 Sensitivity Analysis

Sensitivity analysis is the process of testing our model's robustness to assumptions. A warning sign of an untrustworthy model is if our posterior estimates are highly sensitive to the specific choice of prior distributions, our specific model, or even to data (particularly in the case of outlier data - addressing outliers is discussed later in this lecture).

Sensitivity analysis is the idea of rerunning your model with a variety of different priors to test your model's sensitivity to any particular prior. The information in your posterior is a combination of information from your data and information from your prior, and ideally we would like nearly all of the important information to come from your data and not your prior.

Often, there are multiple sensible models for any particular real-world process or data. It makes sense to try them all. In fact, the recommended approach of writing models - by starting with simple models and slowly adding complexity - should already give you an idea of a particular model's sensitivity to certain modeling assumptions.

When possible, it's preferred to perform "continuous model expansion" (discussed more in detail below) than comparing discrete models to each other.

## 1.2 Posterior Predictive Checking

Bayesian models produce posterior distributions on its parameters after seeing data. It is easy, and useful, to simulate data (equivalently, **posterior predictions**) using the inferred posterior parameters of your model, and comparing them with your data.

You can do this by drawing values of $\hat{\theta}$ from your posterior distribution $p(\theta|y)$, then simulating data from your model's likelihood by using $p(y_{sim}|\hat{\theta})$.

If your model is hierarchical, it's possible to start your posterior predictions by sampling at different layers. For example, consider a hierarchical model describing SAT performance from students belonging to different schools, such that we have 10 different schools each with 50 students. Once we have trained the model, we can ask the model to simulate a new student from a particular school, or simulate a new school with new students.

A warning sign of a **misspecified model** is if your posterior predictions are "dissimilar" from your data.

There are many different ways to measure "similarity" between your posterior predictions and your data. Some common quantities to compare include the mean, median, maximum, minimum, and standard deviations. These comparisons can be plotted graphically to quickly compare different quantities of interest, as well as enabling a more qualitative comparison. Some quantities are less sensitive to differences in models, like the median, while other quantities like the min or max are highly sensitive.

One benefit of Bayesian statistics is the ability to get confidence intervals on any quantity of interest. These confidence intervals can be used to detect model misspecification. For example, if the value for your quantity of interest from your real data lies outside a 95% confidence interval, then your model may have an issue.

**Exercise 1**

Model checking is one of the most important aspects of Bayesian modeling, so we really ought to give an exercise here, but unfortunately we haven't prepared an exercise that'll hold your hand and walk you through the concepts. We emphasize that it's crucial to keep in mind when designing models for real data.

You can get some practice on your own by playing around with some of the toy datasets - can you generate posterior predictions from a model? Is your model sensitive to specific prior distributions?

### 1.2.1 Multiple Comparisons

There are nearly infinite statistical quantities you could compare between your posterior predictions and your data. The more quantities you compare, the more likely you are to find a quantity where the value obtained from your data lies outside of your posterior predictive's confidence interval.

However, in Bayesian modeling, we are not attempting to *reject* some null hypothesis that our specified model is the true model - indeed, many Bayesians are of the philosophy that all models are wrong.

Instead, in practice, we suggest graphical checks of sanity for your posterior predictions, and asking how well your model performs in representing a few particularly important quantities of interest and parameters.

# 2 Continuous Model Expansion

Continuous model expansion is the idea of allowing your data to decide between models $M_1$ and $M_2$ for you. To do this, augment your model such that $M_1$ and $M_2$ are special cases of your new model $M_{new}$.

As an example, consider this situation: You have 22 studies, each with 100 observations, on the effect of having children on happiness, and you wish to perform a meta-analysis. Your quantity of interest is the mean effect of children on parental happiness. You decide to use a normal distribution with separate means $\mu_i$ and identical standard deviation $\sigma$ for each study $i$. At this point, two models make sense to you:

- (Model A) All studies are exactly the same, so you can pool all the data together into a single dataset of size $22 \times 100$
- (Model B) The mean effects $\mu_i$ for each study are identically, independently distributed according to a normal distribution

To write a model that includes both models as special cases, it makes sense to start with the more complex model and try to fit the simpler model into it.

**Exercise 2**

Write a single model that includes both models written above as special cases.

We provide our answer on the next page.

## 2.1 Answer

We define $y_{i,j}$ such that $i$ indexes $1, ..., 22$ for the 22 studies, and $j$ indexes $1, ..., 100$ for the 100 observations per study.

$$\begin{aligned} y_{i,j} &\sim Normal(\mu_i, \sigma) \\ \mu_i &\sim Normal(\alpha, \tau) \end{aligned} \tag{1}$$

Model 1 describes model B. Here, we notice that if $\tau = 0$, then model 1 becomes equivalent to model A.

**Exercise 2**

Recalling the relationships between the Cauchy distribution, the $t$-distribution, and the Gaussian distribution, write a model that combines the two following models. If you introduce any new parameters, place a weakly-informative prior on them.

$$\begin{aligned} y_{i,j} &\sim Normal(\mu_i, \sigma) \\ \mu_i &\sim Normal(\alpha, \tau) \end{aligned} \tag{2}$$

$$\begin{aligned} y_{i,j} &\sim Normal(\mu_i, \sigma) \\ \mu_i &\sim Cauchy(\alpha, \tau) \end{aligned} \tag{3}$$

## 3 Bibliography and Additional Resources

Chapters 6 and 7 of Bayesian Data Analysis 3 discuss model checking and information criterion.

The Stan reference is available here: http://mc-stan.org/documentation/