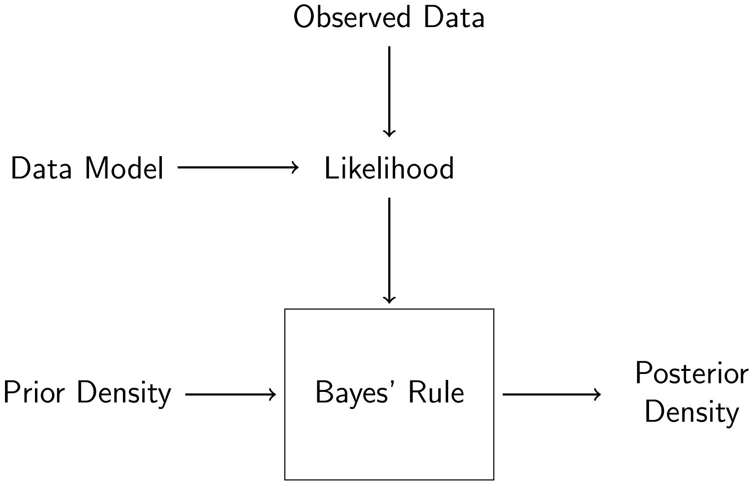
STAT 431 — Applied Bayesian Analysis — Course Notes

Bayesian Analysis Overview

Fall 2022

Main Idea of Bayesian Analysis



Data Notation

After we observe the data, we regard them as fixed:

$$\boldsymbol{y} \qquad \text{(or just } y \text{ if univariate)}$$

Before we observe the data, we regard them as random:

$$\boldsymbol{Y} \qquad \text{(or just } Y \text{ if univariate)}$$

For example, the data could represent a sample of n values:

$$\text{after observing:} \quad \boldsymbol{y} = (y_1, \ldots, y_n)$$

$$\text{before observing:} \quad \boldsymbol{Y} = (Y_1, \ldots, Y_n)$$

Parametric Models

Before observing the data, we may know only something about their distribution, based on some generating random process.

A parametric **model** for random data $\boldsymbol{Y}$ is a collection of possible distributions for $\boldsymbol{Y}$, indexed by a **parameter** $\boldsymbol{\theta}$ (or just θ, if univariate).

We don't know the "actual" value of $\boldsymbol{\theta}$.

In the frequentist (classical) perspective, $\boldsymbol{\theta}$ is *fixed*.

Bayesians can regard $\boldsymbol{\theta}$ as *random*.

Model Notation

Bayesians regard the model as the set of *conditional* distributions for $\boldsymbol{Y}$ given $\boldsymbol{\theta}$ and may write

$$\boldsymbol{Y} \mid \boldsymbol{\theta} \;\sim\; \mathcal{M}(\boldsymbol{\theta})$$

where $\mathcal{M}$ specifies the data model.

We assume each such conditional distribution has a density (PMF or PDF), denoted

$$f(\boldsymbol{y} \mid \boldsymbol{\theta}) \qquad\qquad \text{(or } f(\boldsymbol{Y} \mid \boldsymbol{\theta}) \text{ in BSM)}$$

For example, if $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ is known to be a normal (Gaussian) sample with unknown mean μ and unknown variance σ^2, then

$$\boldsymbol{\theta} \;=\; (\mu, \sigma^2)$$

and we write

$$Y_i \mid \mu, \sigma^2 \quad \sim \quad iid \;\, \mathrm{Normal}(\mu, \sigma^2)$$

so

$$f(\boldsymbol{y} \mid \mu, \sigma^2) \quad = \quad \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \mu)^2}{2\sigma^2}\right)$$

Likelihood

After observing the data, we can "plug them in" to the density

$$f(\boldsymbol{y} \mid \boldsymbol{\theta})$$

and call the resulting function of $\boldsymbol{\theta}$ the **likelihood**.

As we will see, a likelihood need be specified only up to proportionality in $\boldsymbol{\theta}$.

Likelihood can be used in non-Bayesian contexts, but we will use it as an ingredient in Bayes' rule.

Prior Notation

If $\boldsymbol{\theta}$ has a marginal distribution (as a Bayesian can assume), it is called the **prior distribution**.

If $\boldsymbol{\theta}$ has a marginal density (usually a PDF), it is the **prior density**, denoted in BSM as

$$\pi(\boldsymbol{\theta})$$

The prior is intended to represent our uncertainty about the value of $\boldsymbol{\theta}$ before seeing the data.

Posterior Notation

The conditional distribution of $\boldsymbol{\theta}$ given $\boldsymbol{Y} = \boldsymbol{y}$ is the **posterior distribution**.

It represents our uncertainty about the value of $\boldsymbol{\theta}$ after seeing the data.

The posterior distribution can have a **posterior density**, denoted

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}) \qquad\qquad \text{(or } p(\boldsymbol{\theta} \mid \boldsymbol{Y}) \text{ in BSM)}$$

The posterior is uniquely determined by the likelihood and prior, according to Bayes' rule ...

Bayes' Rule

Assuming there is a prior density,

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}) \;=\; \frac{f(\boldsymbol{y} \mid \boldsymbol{\theta})\,\pi(\boldsymbol{\theta})}{m(\boldsymbol{y})}$$

where the **marginal density** of $\boldsymbol{Y}$ is

$$m(\boldsymbol{y}) \;=\; \begin{cases} \displaystyle\sum_{\text{all } \boldsymbol{\theta}} f(\boldsymbol{y} \mid \boldsymbol{\theta})\,\pi(\boldsymbol{\theta}), & \boldsymbol{\theta} \text{ discrete} \\[2ex] \displaystyle\int f(\boldsymbol{y} \mid \boldsymbol{\theta})\,\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta}, & \boldsymbol{\theta} \text{ continuous} \end{cases}$$

Proportionality

Bayes' rule is written more simply as

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}) \quad \propto \quad f(\boldsymbol{y} \mid \boldsymbol{\theta})\, \pi(\boldsymbol{\theta})$$

where the proportionality is in $\boldsymbol{\theta}$, not $\boldsymbol{y}$.

That is, the proportionality constant (normalizing constant) may (and usually does) depend on $\boldsymbol{y}$.

(Q: Why is it OK to know the posterior density only up to proportionality?)

Note: In this form of Bayes' rule, we may drop any factors in the likelihood or the prior density that don't depend on $\boldsymbol{\theta}$.

Remark: Bayes' rule for *probabilities* can be interpreted as a special case in which both the data and parameter are discrete.

See the example in BSM Sec. 1.2.1.

Bayesian Inference Process

1. Define the data model(s).
2. Obtain the likelihood function.
3. Specify the prior density.
4. Compute the posterior density.

Then use the posterior to make inference.

Considerations in choosing a data model are the same as you may have seen in other statistical courses.

Considerations in choosing a prior relate to the kind of prior information you think you have (specific? vague? none?) and computational ease. We will revisit this in more detail later (BSM Chapter 2).