STAT 578: Advanced Bayesian Modeling

Week 1 – Lesson 3

Bayesian Fundamentals

Fall 2019

Concepts, Notation, Terminology

Notation

$$
\begin{aligned}
y &= \text{random data (observed)} \\
\theta &= \text{parameter (unobserved)}
\end{aligned}
$$

In Bayesian analysis, θ is random. Its distribution expresses our uncertainty about its value.

Both y and θ can be vectors, e.g.,

$$y = (y_1, \dots, y_n)$$

All distributions we will use are defined through densities.

Reminder: We denote

$$\begin{aligned} p(\cdot) &= \text{any joint or marginal density} \\ p(\cdot \mid \cdot) &= \text{any conditional density} \end{aligned}$$

with the symbols of the arguments specifying which random variables are involved.

Note: Discrete densities and continuous densities will use the same notation – distinguish according to the type of random variable involved.

Distribution Concepts

A **Bayesian model** is specified by the prior and sampling distributions:

$$\begin{aligned} p(\theta) &= \text{the } \textbf{prior} \text{ density} \\ p(y \mid \theta) &= \text{the } \textbf{sampling} \text{ density} \end{aligned}$$

Together, these define any probability statement about y and θ, whether joint, marginal, or conditional.

Regard the prior as a summary of the uncertainty about θ *before* observing y.

Bayesian inference about θ is based on

$$p(\theta \mid y) \quad = \quad \text{the } \textbf{posterior} \text{ density}$$

which defines the posterior distribution.

Regard the posterior as a summary of the remaining uncertainty about θ, *after* observing y.

Classical non-Bayesian inference, in contrast, does not treat θ as random, so cannot summarize its uncertainty with a distribution.

Bayes' Rule

Bayes' rule specifies how to derive the posterior from the Bayesian model:

$$p(\theta \mid y) \;=\; \frac{p(\theta)\, p(y \mid \theta)}{p(y)}$$

The **normalizing factor**

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

ensures that $p(\theta \mid y)$ is a density in θ.

Since the normalizing factor is just a constant (calculated from the observed y), Bayes' rule can be written in an **unnormalized** form:

$$p(\theta \mid y) \quad \propto \quad p(\theta)\, p(y \mid \theta)$$

where the proportionality is in θ (not y).

Since the normalizing factor $p(y)$ can be difficult to compute, we will prefer methods that do not require it.

Factor $p(y \mid \theta)$, when regarded as a function of θ for fixed y, is the **likelihood**.

It is generally not a density in θ.

Since the unnormalized form of Bayes' rule requires only proportionality in θ, any factors in the likelihood that don't involve θ can be omitted (even if they involve y).

Therefore, we will often write

$$p(y \mid \theta) \quad \propto \quad \text{some expression in } \theta \text{ (and } y\text{)}$$

where the expression is guaranteed to be proportional only in θ (not y).