STAT 431 — Applied Bayesian Analysis — Course Notes

Bayesian Computation: Deterministic Methods

Fall 2022

Overview

Notation:

$$
\begin{aligned}
\boldsymbol{\theta} &= (\theta_1, \ldots, \theta_p) = \text{parameters} \\
\boldsymbol{y} &= \text{observed data} \\
\pi(\boldsymbol{\theta}) &= \text{prior density} \\
p(\boldsymbol{\theta} \mid \boldsymbol{y}) &= \text{posterior density} \\
p(\theta_j \mid \boldsymbol{y}) &= \text{marginal posterior density for } \theta_j
\end{aligned}
$$

Notice: When $\boldsymbol{\theta}$ has a continuous posterior distribution, most Bayesian inference tasks involve integration —

- Computing the marginal data density $m(\boldsymbol{y})$:

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

  so

$$m(\boldsymbol{y}) = \int f(\boldsymbol{y} \mid \boldsymbol{\theta})\,\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta}$$

  (related to the normalizing constant)

- Computing a posterior expectation:

  For some function g, might want

$$\mathrm{E}\big(g(\boldsymbol{\theta}) \mid \boldsymbol{y}\big) \;=\; \int g(\boldsymbol{\theta})\, p(\boldsymbol{\theta} \mid \boldsymbol{y})\, d\boldsymbol{\theta}$$

  For example, a posterior mean of θ_j:

$$g(\boldsymbol{\theta}) \;=\; \theta_j$$

  or a posterior variance of θ_j:

$$g(\boldsymbol{\theta}) \;=\; \big(\theta_j - \mathrm{E}(\theta_j \mid \boldsymbol{y})\big)^2$$

- Computing a marginal posterior density:

$$p(\theta_j \mid \boldsymbol{y}) \;=\; \int_{\substack{\text{all } \theta_k \\ k \neq j}} p(\boldsymbol{\theta} \mid \boldsymbol{y}) \, d\boldsymbol{\theta}_{(-j)}$$

where $\boldsymbol{\theta}_{(-j)}$ is $\boldsymbol{\theta}$ with θ_j removed.

Note: This is generally a different integration problem (different integrand) for each possible value of θ_j.

- Computing a posterior probability:

  For $H_0 : \boldsymbol{\theta} \in \Theta_0$,

$$\mathrm{Prob}(H_0 \mid \boldsymbol{y}) \;\; = \;\; \int_{\Theta_0} p(\boldsymbol{\theta} \mid \boldsymbol{y}) \, d\boldsymbol{\theta}$$

  Alternatively, this can be computed as the posterior expectation of

$$g(\boldsymbol{\theta}) \;\; = \;\; I(\boldsymbol{\theta} \in \Theta_0)$$

  where I is an **indicator function**:

$$I(\boldsymbol{\theta} \in \Theta_0) \;\; = \;\; \begin{cases} 1, & \boldsymbol{\theta} \in \Theta_0 \\ 0, & \boldsymbol{\theta} \notin \Theta_0 \end{cases}$$

Other things that might involve integration (directly or indirectly):

- finding a posterior quantile (such as a median)
- obtaining (and working with) a posterior predictive distribution

Numerical Integration

Goal: Approximate

$$\int_{\Delta} f(\boldsymbol{x})\, d\boldsymbol{x}$$

Idea: Partition Δ into m regions $\Delta_1, \ldots, \Delta_m$, with representative points

$$\boldsymbol{x}_1 \in \Delta_1, \quad \ldots, \quad \boldsymbol{x}_m \in \Delta_m$$

and use

$$\sum_{j=1}^{m} f(\boldsymbol{x}_j) \cdot \mathrm{area}(\Delta_j) \qquad \text{where} \quad \mathrm{area}(\Delta_j) \;=\; \int_{\Delta_j} d\boldsymbol{x}$$

[Draw in 3-D ...]

For example, the midpoint rule (in one dimension):

$$\int_a^b f(x)\,dx \quad \approx \quad \sum_{j=1}^{m} f(x_j) \cdot \frac{b-a}{m}$$

$$\text{where} \quad x_j \;=\; a \,+\, (b-a)\,\frac{j-\frac{1}{2}}{m}$$

[Draw example ...]

Note: Accuracy requires m large and f (somewhat) continuous and smooth.

R function `integrate()` uses an adaptive algorithm for one-dimensional integration.

Example: Proportion of people like us with pets

Recall: $y = 12$ out of $n = 70$

$$Y \mid \theta \;\sim\; \text{Binomial}(n, \theta)$$

$$\text{likelihood} \;\propto\; \theta^{12}(1-\theta)^{58}$$

Let's use the JP:

$$\theta \;\sim\; \text{Beta}(1/2,\ 1/2)$$

Using conjugacy, the posterior is

$$\theta \mid y \;\sim\; \text{Beta}(12.5,\ 58.5)$$

so

$$\text{E}(\theta \mid y) \;=\; \frac{12.5}{12.5 + 58.5} \;\approx\; 0.17606$$

Let's pretend we don't know the posterior, and we want to approximate

$$
\begin{aligned}
\mathrm{E}(\theta \mid y) &= \int \theta \cdot p(\theta \mid y)\, d\theta \\
&= \int \theta\, \frac{f(y \mid \theta)\, \pi(\theta)}{m(y)}\, d\theta \\
&= \frac{\displaystyle\int \theta\, f(y \mid \theta)\, \pi(\theta)\, d\theta}{\displaystyle\int f(y \mid \theta)\, \pi(\theta)\, d\theta}
\end{aligned}
$$

So we approximate two integrals ...

R Example 3.1:

Population Proportion: Numerical Integration

Remark:

Can avoid evaluating two different integrals by instead using normalized *weights* to approximate the posterior density.

See BSM, Section 3.1.2.

Beyond one dimension, need either to make nested calls to R function `integrate` or to use some other function.

For example, R package `cubature` has functions for (adaptive) numerical integration of a multivariate function over a hypercube.

Example: Are Bike Owners Less Likely to Ride the Bus?

Data (from a class survey):

- Among $n_1 = 19$ bike owners, $y_1 = 8$ ride the bus
- Among $n_2 = 51$ non-bike owners, $y_2 = 25$ ride the bus

$$
\begin{aligned}
\theta_1 &= \text{population proportion of owners who ride} \\
\theta_2 &= \text{population proportion of non-owners who ride} \\
\boldsymbol{y} &= (y_1, y_2)
\end{aligned}
$$

Want

$$\mathrm{Prob}(\theta_1 < \theta_2 \mid \boldsymbol{y})$$

Likelihood:

$$\begin{aligned} f(\boldsymbol{y} \mid \theta_1, \theta_2) &= f(y_1 \mid \theta_1)\, f(y_2 \mid \theta_2) \\ &\propto \theta_1^{y_1}(1-\theta_1)^{n_1-y_1}\, \theta_2^{y_2}(1-\theta_2)^{n_2-y_2} \end{aligned}$$

We'll use a product-JP:

$$\theta_1, \theta_2 \quad \sim \quad \text{indep. } \mathrm{Beta}(1/2, 1/2)$$

Need to compute

$$\begin{aligned}
\text{Prob}(\theta_1 < \theta_2 \mid \boldsymbol{y}) &= \int_0^1 \int_0^{\theta_2} p(\theta_1, \theta_2 \mid \boldsymbol{y})\, d\theta_1\, d\theta_2 \\
&= \frac{\displaystyle\int_0^1 \int_0^{\theta_2} f(\boldsymbol{y} \mid \theta_1, \theta_2)\, \pi(\theta_1, \theta_2)\, d\theta_1\, d\theta_2}{\displaystyle\int_0^1 \int_0^1 f(\boldsymbol{y} \mid \theta_1, \theta_2)\, \pi(\theta_1, \theta_2)\, d\theta_1\, d\theta_2}
\end{aligned}$$

[Draw integration region ...]

R Example 3.2:

Comparing Population Proportions: Numerical Integration

A different approach:

Approximate a continuous posterior with a multivariate generalization of the normal distribution.

Under some conditions, this is justified by a “Bayesian central limit theorem” when enough data are observed.

See BSM, Section 3.1.3.