Applied Bayesian Analysis for the Social Sciences

Day 2: Posteriors, MCMC and the Gibbs Sampler

Ryan Bakker
University of Georgia

Sampling from and Summarizing Posterior Distributions

- Recall that MLE finds the parameter value(s) that maximize the likelihood function for the observed data

- We use this information to produce point estimates and their standard errors.

- Then do standard hypothesis test, assuming normality, to assess statistical significance.

Posteriors

- In contrast, Bayesian models derive posterior distributions for parameters
  - Not a single point estimate.
  - Not reliant on CLT for inference.
  - Quantities of interest are integrals of posteriors, like means, medians, variances and various quantiles.

Why are posteriors better?

- No need to rely on asymptotics and/or assumptions of normality. The posterior dist of parameters can be directly assessed
- Allows computation of additional tests and summaries not available through classical methods
- Distributions can be easily transformed into quantities of interest not directly estimated as part of the model.

Is this hard to do?

- Usually not!
- We’ve seen the ease with which this can be done when analytical solutions are available.
  - Results from using conjugate priors
  - Easy to sample from univariate posteriors with random number generators
  - Easy to program even if ‘canned’ routines not available

What about when it is hard-ish?

- If analytical solutions are not available we can:
  - Still sample using R, for example
    - We may need to program (type in) the posterior.
- use approximation methods
  - Quadrature
  - Taylor series expansions
  - Mixture of normals for multi-modal dists.
- Still rely on CLT-which undermines utility of Bayesian modeling in many situations.
  - Small samples, for example

Sampling methods

- We can, instead, simulate a sample from the distribution(s) of interest and use discrete formulae to approximate integrals of interest:

Mean: $$\int x f(x) dx \approx \frac{1}{n} \sum_{i=1}^{n} x_i$$

Variance: $$\int (x - \mu)^2 f(x) dx \approx \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$$

- Quantiles can be computed by noting the value of x for which some % of the sample falls below/above.

Summarizing Posteriors

- So…Bayesian inference usually involves three steps:
  - Specifying a model and obtaining a posterior distribution for parameters
  - Generating samples from this posterior
  - Using discrete formulae to summarize our knowledge of these parameters
- We've seen that the analytical mean and the simulated means are VERY similar in the simple examples/homework from last week.

Basic sampling methods

- Inversion sampling. To sample from *f(x):*
  - Draw a uniform random number *u* between 0 and 1. This represents the area under f(x) to that point.
  - Then $z = F^{-1}(u)$ is a draw from *f(x).*
- We are looking for a *z* such that:

$$u = \int_L^z f(x)dx$$

Inversion sampling cont…

- Let's say we have a linear density function:

$$f(x) = \frac{1}{40}(2x+3) \qquad \text{(with } 0 < x < 5)$$

- First draw u ~ U(0,1) and then compute *z* that satisfies:

$$u = \int_0^z \frac{1}{40}(2x+3)dx$$

Inversion…

- Evaluting this integral at 0 and z yields:

$$40u = z^2 + 3z$$

$$40u + 9/4 = (z + 3/2)^2$$

- Taking the square root of both sides

$$z = \frac{-3 \pm \sqrt{160u + 9}}{2}$$

Inversion cont.

- Inversion is efficient and relatively easy, but…
  - Some inverse functions are impossible to derive.
    - E.g. the normal integral
  - Does not work with mulitvariate distributions as the inverse is not unique beyond one dimension.
    - More unknowns than equations problem.

Rejection Sampling

- Sample a value z from a distribution *g(x)* which is easy to sample from and for which all values of *m* * *g(x)* are greater than *f(x)* at all points.
- Compute the ratio: $R = \frac{f(z)}{m * g(z)}$
- Sample *u ~ U(0,1).* If *R* > *u*, then accept *z* as a draw from *f(x)*. Otherwise return to step 1.

Rejection continued

- We call *m*g(x)* an ‘envelope function’ as it envelopes *f(x).*

- Then we compute the ratio of densities of *f(x) to m * g(x)* for a given value of x.

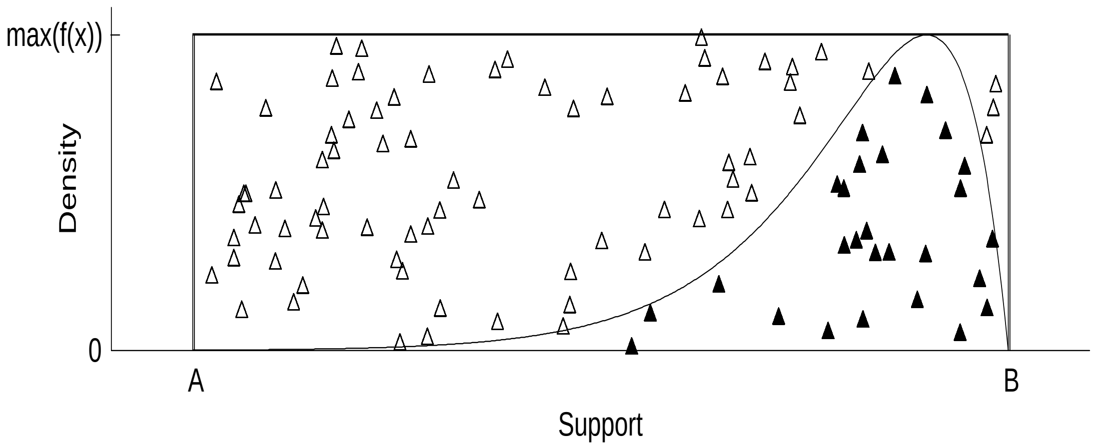
- Finally, we compare this ratio to a random uniform draw. This ratio is then the probability we accept a draw at a given value *x* as coming from *f(x).*

Rejection cont.

- Can be used for most distributions
  - Even if inversion cannot
  - Including multivariate distributions
- Some limitations
  - Finding an enveloping distribution may be hard (or impossible).
  - Can be inefficient—may take many draws from g(x) to get 'enough' draws from f(x).

Continuous Form with Bounded Support (cont.)

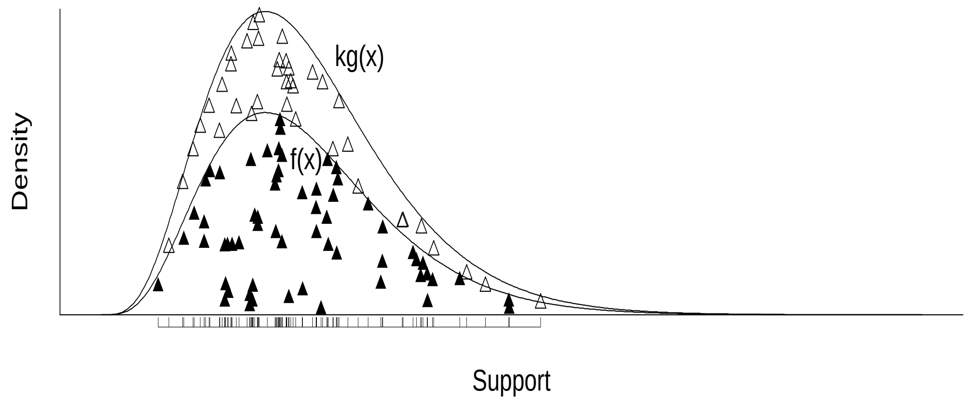
- 100 points sampled uniformly from the two-dimensional rectangle over: $[(A, B), (0, \max(f(x)))] = [(0, 10), (0, 0.4)]$. In total 27 values fell into the area we wish to integrate, so we obtain the size of the interval from: $(27/100)(10 \times 0.4) = 1.04$.



Continuous Form with Unbounded Support

- Now address the integral of some function $f(x)$ in which the analytical solution is difficult or impossible, and the form of $f(x)$ has unbounded tails.

- Specify a "majorizing function," $g(x)$, which for every value of x in the support of $f(x)$ has the property that $g(x) \geq f(x)$.



MCMC--briefly

- MCMC simulation methods can use either inversion or rejection sampling to sample from posterior densities.
  - Inversion not as likely given the impossibility of inverting multivariate distributions.
  - Gibbs Sampling usually relies on rejection (adaptive rejection, actually) sampling

However,

- Gibbs Sampling can be problematic if
  - We cannot derive the conditional distributions for the parameters of interest
  - If the form of the conditional distributions is unknown.
  - Inversion sampling is impossible
  - Cannot find appropriate envelope function

Metropolis-Hastings

- Algorithm that generates samples from the full joint density.
- Works on multivariate distributions
- Doesn’t require an envelope.

Metropolis Hastings in words

- Establish starting values for parameters.
- Draw a ‘candidate’ value from a proposal density
  - Similar to rejection, but doesn’t need to envelope
  - Use a distribution that is easy to sample from—like normal, uniform.
  - Assess probability this is from the target distribution—like rejection.