Analyzing Response Time Data

Statistical Modeling in Psychology

October 9-10 2024

Supported by NSF awards SES-1424481 and DMS-1407604.

The Ohio State University

Yesterday

- Random Variables
- Probability Density Functions
- Function Estimation
- Parameter Estimation
  - Method of Moments
  - Least Squares
  - Maximum Likelihood
- Model Comparison (discussion)

Outline

Outline

Bayesian Inference

Three components:

- The prior $\pi(\theta)$
- The likelihood $\mathcal{L}(\theta \mid Y)$
- The posterior $\pi(\theta \mid Y)$

Likelihood

Given an *iid* sample of data $T = \{T_1, T_2, \ldots, T_N\}$, the *likelihood* is

$$\mathcal{L}(\theta \mid T) = \prod_{i=1}^{n} f(T_i \mid \theta),$$

where θ is the vector of model parameters that we would like to estimate/infer.

Bayes' Theorem

Describe the uncertainty about parameter θ with a probability distribution with pdf $\pi(\theta)$ – the prior.
Then

$$\begin{aligned}\pi(\theta \mid T) &= \frac{\mathcal{L}(\theta \mid T)\pi(\theta)}{\int_{\theta\in\Omega_\theta} \mathcal{L}(\theta \mid T)\pi(\theta)d\theta} \\ &\propto \mathcal{L}(\theta \mid T)\pi(\theta)\end{aligned}$$

is the posterior distribution of θ given the observed data.

Posterior

The posterior describes our uncertainty about θ given the observed data. Given a posterior distribution, we can make probability statements about θ that would not otherwise be possible.

Bayes as Updating

$$\begin{aligned}\pi(\theta \mid Y) &= \frac{\mathcal{L}(Y \mid \theta)\pi(\theta)}{f_Y(y)} \\ &= \pi(\theta)\frac{\mathcal{L}(Y \mid \theta)}{f_Y(y)}\end{aligned}$$

Exercise: The base rate problem.

Example: The Beta-Binomial Model

A variable Y_i follows a binomial distribution with parameters N (the observed number of trials in a statistical experiment) and q, where q is the probability of “success” on any trial. Consider $Y = \{Y_1, Y_2, \ldots, Y_M\}$, so the likelihood is

$$\mathcal{L}(q \mid N, Y) = \prod_{i=1}^{M} \binom{N}{Y_i} q^{Y_i} (1-q)^{N-y_i}.$$

Beta Prior

Let

$$\pi(q) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} q^{a-1}(1-q)^{b-1}.$$

Then

$$\begin{aligned}
\pi(q \mid Y, N, a, b) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} q^{a-1}(1-q)^{b-1} \times \\
& \prod_{i=1}^{M} \binom{N}{Y_i} q^{Y_i}(1-q)^{N-y_i} \\
&\propto q^{\sum_i Y_i + a - 1}(1-q)^{NM - \sum_i Y_i + b - 1}.
\end{aligned}$$

Beta Prior

So

$$q \mid Y, N, a, b \sim \mathcal{B}\left(\sum_{i=1}^{M} Y_i + a, NM - \sum_{i=1}^{M} Y_i + b\right),$$

a beta distribution with parameters $\sum_{i=1}^{M} Y_i + a$ and $NM - \sum_{i=1}^{M} Y_i + b$.

THE OHIO STATE UNIVERSITY

Exercise: Giving an Exam

- Choose values for N, a and b, and simulate a value for q. (Don't peek!)
- Choose a value for M and simulate a data set $Y = \{Y_1, Y_2, \ldots, Y_M\}$ using the unknown value of q and the value of N that you selected.
- Compute the posterior distribution of q given the data Y, N, a and b.
- Estimate the value of q (which you still don't know because you didn't peek!) from the posterior.
- Compute a "credible set" for the parameter q.

Error of Estimate

Definition (Error)

The *error* of an estimate $\hat{\theta}$ is simply

$$\hat{\theta} - \theta.$$

Recall: $\hat{Y} - Y$ in linear regression.

THE OHIO STATE UNIVERSITY

Error of Estimate

Definition (Error)

The *error* of an estimate $\hat{\theta}$ is simply

$$\hat{\theta} - \theta.$$

Recall: $\hat{Y} - Y$ in linear regression.

Definition (Mean Squared Error of Estimate)

The *mean squared error of estimate* is the average sum of squared error:

$$\begin{aligned} MS(\hat{\theta}) &= E(\hat{\theta} - \theta)^2 \\ &= \int_{\hat{\theta} \in \Theta} (\hat{\theta} - \theta)^2 f(\hat{\theta} \mid \theta) d\hat{\theta}. \end{aligned}$$

The Ohio State University

Error of Estimate

Definition (Mean Squared Error of Estimate)

The *mean squared error of estimate* is the average sum of squared error:

$$\begin{aligned} MS(\hat{\theta}) &= E(\hat{\theta} - \theta)^2 \\ &= \int_{\hat{\theta} \in \Theta} (\hat{\theta} - \theta)^2 f(\hat{\theta} \mid \theta) d\hat{\theta}. \end{aligned}$$

After some algebra,

$$MS(\hat{\theta}) = \text{bias}(\hat{\theta})^2 + Var(\hat{\theta}).$$

Note: Sometimes a biased estimator will have smaller MSE than an unbiased estimator.

Outline

Bayesian Estimators

From the Posterior

- Posterior mean
- Posterior mode (MAP)
- Credible sets
- etc.

Note: Bayesian estimators often have smaller MSE than frequentist estimators.

Beta-Binomial Model

Goal: Estimate success probability q

Let $Y \sim \mathcal{B}(N, q)$ and $\pi(q) \sim \mathrm{B}(1, 1)$. Then

$$\pi(q \mid Y) \sim \mathrm{B}(1 + Y, 1 + N - Y).$$

Frequentist Estimate

$$\hat{q} = \frac{Y}{N}$$

So $E(\hat{q}) = q$ and $Var(\hat{q}) = \dfrac{q(1-q)}{N} = MS(\hat{q})$.

Posterior Mean

B$(1+Y, 1+N-Y)$

Let $\tilde{q} = \frac{1+Y}{(1+Y)+(1+N-Y)} = \frac{1+Y}{N+2}$, the posterior mean.

Posterior Mean

This estimator is biased

Write

$$\tilde{q} = \frac{1+Y}{N+2} = \frac{Y}{N+2} + \frac{1}{N+2}.$$

Then

$$\begin{aligned} E(\tilde{q}) &= E\left(\frac{Y}{N+2} + \frac{1}{N+2}\right) \\ &= \frac{E(Y)}{N+2} + \frac{1}{N+2} \\ &= \frac{Nq}{N+2} + \frac{1}{N+2} \neq q. \end{aligned}$$

THE OHIO STATE UNIVERSITY

Posterior Mean

This estimator is biased:

$$E(\tilde{q}) = \frac{Nq}{N+2} + \frac{1}{N+2} \neq q.$$

Posterior Mean

This estimator is biased:

$$E(\tilde{q}) = \frac{Nq}{N+2} + \frac{1}{N+2} \neq q.$$

Variance

Write

$$\tilde{q} = \frac{1+Y}{N+2} = \frac{Y}{N+2} + \frac{1}{N+2}.$$

Then

$$\begin{aligned} Var(\tilde{q}) &= Var\left(\frac{Y}{N+2} + \frac{1}{N+2}\right) \\ &= \frac{Var(Y)}{(N+2)^2} = \frac{Nq(1-q)}{(N+2)^2}. \end{aligned}$$

Posterior Mean

This estimator is biased:

$$E(\tilde{q}) = \frac{Nq}{N+2} + \frac{1}{N+2} \neq q.$$

Variance

$$Var(\tilde{q}) = \frac{Nq(1-q)}{(N+2)^2}.$$

THE OHIO STATE UNIVERSITY

Posterior Mean

Mean Squared Error

$$
\begin{aligned}
MS(\tilde{q}) &= Var(\tilde{q}) + \text{bias}(\tilde{q})^2 \\
&= \frac{Nq(1-q)}{(N+2)^2} + (\frac{Nq}{N+2} + \frac{1}{N+2} - q)^2 \\
&= \left(\frac{1-2q}{N+2}\right)^2 + \frac{Nq(1-q)}{(N+2)^2}.
\end{aligned}
$$

Mean Squared Error

Comparison

$$\frac{q(1-q)}{N} \text{ vs. } \left(\frac{1-2q}{N+2}\right)^2 + \frac{Nq(1-q)}{(N+2)^2}$$

R code

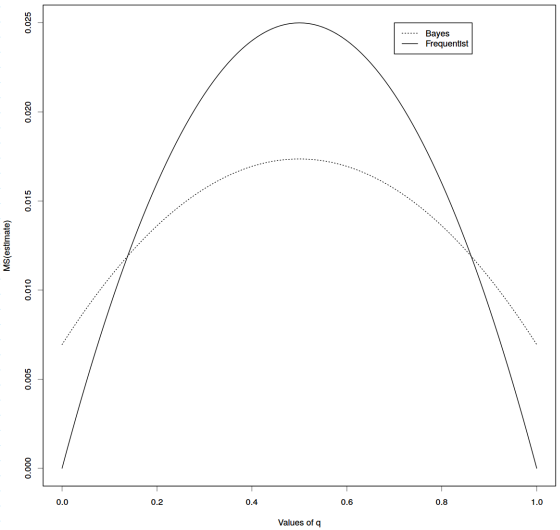
```
# R code for the beta-binomial model
# Contrast the MS of the frequentist MVUE estimator
# with the MS of the posterior mean estimator.

# Define a function for the MS of the MVUE estimator (Y/N):
mvue.ms <- function(q,N) q*(1-q)/N

# Define a function for the MS of the posterior mean
# ((Y+1)/(N+2)):
bayes.ms <- function(q,N) ((1-2*q)/(N+2))^2
                          + N*q*(1-q)/(N+2)^2

# Plot one on top of the other:
q <- seq(0,1,by=.01)
plot(q,mvue.ms(q,10),type='l',
     xlab='Values of q',ylab='MS(estimate)')
lines(q,bayes.ms(q,10),lty=3)
legend(.7,.025,c('Bayes','Frequentist'),lty=c(3,1))
```

Mean Squared Error Comparison



Objective and Uninformative Priors

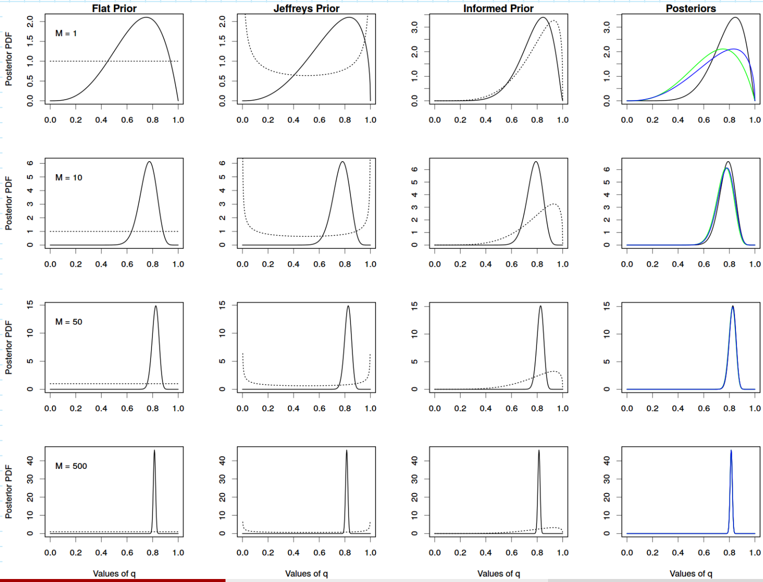
A Criticism of Bayesian Methods

The posterior depends on the choice of the prior, so our conclusions could be biased.

- An *uninformative* or *diffuse prior* tries to spread prior mass equally over the support of the parameter.
- An *objective prior* is chosen *a priori* on the basis of some criterion. E.g., a *reference prior* is chosen to maximize the “distance” between the prior and the posterior.
- As the amount of data grows, under general conditions, the posterior distribution of θ will tend toward a normal distribution with mean equal to the maximum likelihood estimate of θ - a Bayesian “Central Limit Theorem.”

The Ohio State University

Effects of Prior and Sample Size



Bayesian Central Limit Theorems

Definition (Bayesian Central Limit Theorem)

Consider a likelihood

$$\mathcal{L}(\theta \mid y) = \prod_{i=1}^{M} f_Y(y_i \mid \theta).$$

Under general conditions, as the sample size M grows, the posterior $\pi(\theta \mid y)$ will approach a normal distribution with mean equal to the maximum likelihood estimate of θ.

Outline

What's a Markov chain?

A random walk:

$$X_t \sim \mathcal{N}(X_{t-1}, \sigma^2) \equiv X_t = X_{t-1} + \epsilon_i,$$

where $\epsilon \sim \mathcal{N}(0, \sigma^2)$.

Proposals and Fitness

The Metropolis-Hastings step:

$$\theta^* \sim \mathcal{N}(\theta_{t-1}, \sigma^2)$$

Compute

$$\frac{\mathcal{L}(\theta^* \mid Y)\pi(\theta^*)}{\mathcal{L}(\theta_{i-1} \mid Y)\pi(\theta_{i-1})}$$

If

$$\alpha \sim U(0,1) < \min\left\{1, \frac{\mathcal{L}(\theta^* \mid Y)\pi(\theta^*)}{\mathcal{L}(\theta_{i-1} \mid Y)\pi(\theta_{i-1})}\right\},$$

set $\theta_i = \theta^*$, otherwise $\theta_i = \theta_{i-1}$.

THE OHIO STATE UNIVERSITY

MCMC Approximation for the ex-Gaussian

Exercise: Estimate the posterior parameters of the ex-Gaussian distribution fit to your own data, the Wagenmakers data, or the mystery data. Think about the following:

- What role do the priors play? How are they implemented in the code? What happens when you change them?
- Consider how the proposal values θ^* are generated. How important is the variance of the proposals?
- Has the process resulted in a stable "target distribution" (posterior)? How do you know?
- Using the samples drawn from the posteriors, generate point estimates of the ex-Gaussian parameters.

Fitting Challenging Models

The brms package (Bürkner, 2017) fits a wide range of multilevel models, including RT models such as the exGaussian, drift diffusion, etc..

This package is a “wrapper” for the rstan package, which in turn is a wrapper for stan probabilistic programming language [https://mc-stan.org/].

- It has a pretty steep learning curve.
- It can take a very long time to generate parameter chains.
- There is some code in your notebook to look at.
- Some web resources to get you started:
    - http://singmann.org/wiener-model-analysis-with-brms-part-i/ (Part 1 of 4)
    - https://www.martinmodrak.cz/2021/04/01/using-brms-to-model-reaction-times-contaminated-with-errors/
    - https://solomonkurz.netlify.app/blog/2021-06-05-don-t-forget-your-inits/

Stan

Stan uses Hamiltonian Monte Carlo to sample from a posterior. A modification of the HMC, the No-U-Turn sampler (NUTS) allows the sampler to efficiently travel through highly correlated parameter spaces.

Open up the Stan section in today's notebook.

Exercise

1. Change the exGaussian priors to be informative and rerun the sampler.
2. Change the priors to something ridiculous, then examine the posterior estimates of q as a function of sample size. When does the prior choice no longer influence the posterior?
3. Use the Stan code to fit the exGaussian (or some other model) to your own data. Evaluate the fit of the model by way of
   - Determining convergence (Rhat, plot the chains, etc.)
   - Generate the posterior predictive distribution. Does it look like the data?
   - What else might you do?

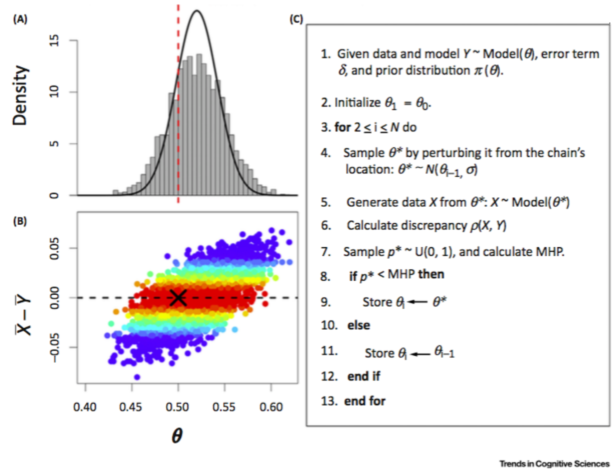
Outline

Computational Models

Some models have no explicit likelihood and predictions are generated by simulation.

Each evaluation of an objective function requires a new simulation.

Approximate Bayesian Computation



THE OHIO STATE UNIVERSITY

Exercise

Using the code in your notebook, estimate the posterior distributions of the parameters of the ex-Gaussian distribution using either the Wagenmakers data or by simulation.
Things to notice:

- What is the discrepancy function? Is it “sufficient”?
- How does the sampler get the proposal values in roughly the right place?
- What happens if the parameter chains get “stuck”?
- How did I choose the threshold distance?
- What happens when that distance gets smaller?

Outline

The Ohio State University

Final Discussion

- What did you learn?
- What was your most/least favorite part?
- What would you have liked to see that you didn't?
- What was surprising to you?