MCMC
Numerical methods for Bayes
What are the chances that a solitaire laid out with 52 cards will come out successfully?
Goal: Find the laws that govern the functioning and the interactions among nature’s living organisms.

How? Through a model, $M(\theta)$

What do we know? $P(\theta)$

What do we have? Some data $(Y)$

What do we want to know? $P(\theta|Y) = \frac{P(Y|\theta) P(\theta)}{P(Y)}$
Rejection sampling

Model:
\[ y = b + a \times x \]

Priors:
\[ P(a) \sim U[a_{\text{min}}, a_{\text{max}}] \]
\[ P(b) \sim U[b_{\text{min}}, b_{\text{max}}] \]
Rejection sampling

Model:
\[ y = b + a \times x + \varepsilon \]
y \sim N(b + a \times x, \text{sd})

Priors:
\[ P(a) \sim U[a_{\text{min}}, a_{\text{max}}] \]
\[ P(b) \sim U[b_{\text{min}}, b_{\text{max}}] \]
\[ P(\text{sd}) \sim U[\text{sd}_{\text{min}}, \text{sd}_{\text{max}}] \]
Rejection sampling

Model:
\[ y \sim N(a \times x, 10) \]

Priors:
\[ P(a) \sim U[a_{\text{min}}, a_{\text{max}}] \]
\[ b = 0 \]
\[ \text{sd} = 10 \]
Rejection sampling

Model:
\[ y \sim N(a \times x, 10) \]

Priors:
\[ P(a) \sim U[a_{\text{min}}, a_{\text{max}}] \]
\[ b = 0 \]
\[ \text{sd} = 10 \]
Rejection sampling

Model:
\( y \sim \mathcal{N}(a \times x, 10) \)

Priors:
\[
P(a) \sim U[a_{\text{min}}, a_{\text{max}}] \\
b = 0 \\
sd = 10
\]

1. draw random values from priors
2. calculate likelihood
   i.e. the probability of observing data \( y \) given the model and drawn parameter values:
   \[
y \sim \mathcal{N}(a_1 \times x, 10)
   \]
Rejection sampling

Model:
\( y \sim N(a \times x, 10) \)

Priors:
\( P(a) \sim U[a_{\text{min}}, a_{\text{max}}] \)
\( b = 0 \)
\( sd = 10 \)

1. Draw random values from priors
2. Calculate likelihood
   i.e. the probability of observing data \( y \) given the model and drawn parameter values:

\[
y \sim N(a_{1} \times x, 10)
\]

\[
1 \quad e^{-\frac{(y-(b+a_{1}\times x))^{2}}{2sd^{2}}}
\]

\[
\frac{1}{sd\sqrt{2\pi}}
\]

in R: `sum(dnorm(y, mean = 0 + a_{1}\times x, sd = 10))`
Rejection sampling

1. draw random values from priors
2. calculate likelihood
3. accept value proportional to likelihood

Model:
\( y \sim N(ax, 10) \)

Priors:
\( P(a) \sim U[a_{\text{min}}, a_{\text{max}}] \)
\( b = 0 \)
\( \text{sd} = 10 \)

\[
P(\theta | Y) = \frac{P(Y|\theta) P(\theta)}{P(Y)}
\]
Rejection sampling

Model:
\[ y \sim N(ax, 10) \]

Priors:
\[ P(a) \sim U[a_{\text{min}}, a_{\text{max}}] \]
\[ b = 0 \]
\[ \text{sd} = 10 \]

1. draw random values from priors
2. calculate likelihood
3. accept value proportional to likelihood

\[ P(a\mid y) \propto P(y\mid a) P(a) \]

unnormalized posterior PDF
Rejection sampling

Model:
\( y \sim N(a \times x, 10) \)

Priors:
\( P(a) \sim U[a_{\text{min}}, a_{\text{max}}] \)
\( b = 0 \)
\( \text{sd} = 10 \)

1. draw random values from priors
2. calculate likelihood
3. accept value proportional to likelihood
   a. Sample many values from prior
   \( \text{Asamples} \leftarrow \text{runif}(1000, a_{\text{min}}, a_{\text{max}}) \)
Rejection sampling

Model:
\[ y \sim N(a \cdot x, 10) \]

1. Draw random values from priors
2. Calculate likelihood
3. Accept value proportional to likelihood

a. Sample many values from prior
b. Calculate unnormalized posterior values

```
for(i in 1:1000)
  LL        <- sum(dnorm(y, mean = Asamples[i]*x, sd =10))
  prior     <- dunif(Asamples[i], a_{min}, a_{max})
  Psamples  <- LL * prior
```
Rejection sampling

1. draw random values from priors
2. calculate likelihood
3. accept value proportional to likelihood

- Sample many values from prior
- Calculate unnormalized posterior values

This is proportional to what I want to sample from
Rejection sampling

1. draw random values from priors
2. calculate likelihood
3. accept value proportional to likelihood

- Sample many values from prior
- Calculate unnormalized posterior values
- Choose a distribution from which sampling is easy

\[ \mathcal{N}(5,5) : \text{proposal or envelope distribution} \]
Rejection sampling

1. Draw random values from priors
2. Calculate likelihood
3. Accept value proportional to likelihood

a. Sample many values from prior
b. Calculate unnormalized posterior values
c. Choose a distribution from which sampling is easy
d. Determine scaling constant: \( C = \max\left(\frac{g(a)}{f(a)}\right) \)
Rejection sampling

- Sample many values from prior
- Calculate unnormalized posterior values
- Choose a distribution from which sampling is easy
- Determine scaling constant: \( C = \max(g(a)/f(a)) \)
- Scale envelope distribution

You only need to determine “\( C \times f(a) \)” once!
Rejection sampling

1. draw random value from \( f(a) \)
2. calculate \( g(a) \) and \( f(a) \)
3. calculate \( u \)

\[
P(a_1 \mid y) \propto P(y \mid a_1) \cdot P(a_1)
\]

\[
g(a_1) = \frac{\text{sum}(\text{dnorm}(y, \text{mean} = 0 + a_1 \cdot x, \text{sd} = 10)) \ast \text{dunif}(a_1, a_{\text{min}}, a_{\text{max}})}{\text{diff}(\text{dnorm}(a_1, \text{mean} = 5, \text{sd} = 5))}
\]

\[
f(a_1) = \text{dnorm}(a_1, \text{mean} = 5, \text{sd} = 5)
\]

\[
u = \frac{g(a_1)}{C \ast f(a_1)}
\]
Rejection sampling

Draw random value: \( a_1 = 4 \)

Calculate \( f(a) \) and \( g(a) \):

\[ g(a_1), f(a_1) \]

Calculate \( u \):

\[ u = \frac{g(a_1)}{C \cdot f(a_1)} \]
Rejection sampling

Draw random value: $a_1 = 4$

Calculate $f(a)$ and $g(a)$: $g(a_1), f(a_1)$

Calculate $u$: $u = \frac{g(a_1)}{C* f(a_1)}$

Accept the proposed $a_1$ with probability $u$ based on a Bernoulli trial:

\[
\text{if } \text{runif}(1,0,1) < u \text{ accept}
\]
\[
\text{else reject}
\]
Rejection sampling

Unnormalized Posterior

Posterior PDF
Rejection sampling

Unnormalized Posterior

Posterior PDF

0 + 5*x
Rejection sampling

Importance sampling weights

\[ w = 0.00548 \]
\[ w = 1.59 \times 10^{-8} \]
\[ w = 9.65 \times 10^{-6} \]
\[ w = 0.371 \]
\[ w = 0.103 \]
\[ w = 1.01 \times 10^{-8} \]
\[ w = 0.111 \]
\[ w = 1.92 \times 10^{-9} \]
\[ w = 0.0126 \]
\[ w = 1.1 \times 10^{-51} \]
Rejection sampling

Pros:
- Parallelizable
- Easy to implement

Cons:
- Suffers from curse of dimensionality
• Perturb parameters
• Accept if new params are supported more (*also accept sometimes even if not)
• Otherwise keep old parameters
Markov Chain Monte Carlo

1) Start from some initial parameter value
2) Evaluate the unnormalized posterior
3) Propose a new parameter value
4) Evaluate the new unnormalized posterior
5) Decide whether or not to accept the new value
6) Repeat 3-5
Markov Chain Monte Carlo

1) Start from some initial parameter value
2) Evaluate the unnormalized posterior
3) Propose a new parameter value
4) Evaluate the new unnormalized posterior
5) Decide whether or not to accept the new value
6) Repeat 3-5
• Advantages
  – Multi-dimensional
  – Can be applied to
    • Whole joint PDF
    • Each dimension iteratively
    • Groups of parameters
  – Simple
  – Robust

• Disadvantages
  – Sequential samples not independent
  – Computationally intensive
  – Discard “Burn – in” period before convergence
  – Assessing convergence
How to assess convergence?

- Visual inspection
How to assess convergence?

- Visual inspection
- Multiple chains
How to assess convergence?

- Visual inspection
- Multiple chains
- Convergence stats
Convergence Statistics

- Brooks Gelman Rubin
  - Within vs among chain variance
  - Should converge to 1

\[ \hat{R} = \frac{B}{W} \]
Convergence Statistics

Trace Plot

GBR Diagnostic

MCMC Sample

MCMC Sample

mu

log(Shrink)
How to assess convergence?

- Visual inspection
- Multiple chains
- Convergence stats
- Quantiles
Quantiles
How to assess convergence?

- Visual inspection
- Multiple chains
- Convergence stats
- Quantiles
- Autocorrelation
Autocorrelation

- lag-\(k\) autocorrelation is the correlation between every sample and the sample \(k\) steps before
- AR should go down as \(k\) increases, i.e. samples can be considered as independent
How to assess convergence?

- Visual inspection
- Multiple chains
- Convergence stats
- Quantiles
- Autocorrelation
- Summary statistics
# Summary Statistics

## Analytical:

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>43.09901</td>
<td>9.95037</td>
</tr>
</tbody>
</table>

## MCMC:

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Naive SE</th>
<th>Time-series</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>43.05504</td>
<td>9.28108</td>
<td>0.05648</td>
<td>0.74503</td>
</tr>
</tbody>
</table>

## Quantiles:

<table>
<thead>
<tr>
<th></th>
<th>2.5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>24.98</td>
<td>36.46</td>
<td>43.39</td>
<td>49.99</td>
<td>60.01</td>
</tr>
</tbody>
</table>
How to assess convergence?

- Visual inspection
- Multiple chains
- Convergence stats
- Quantiles
- Autocorrelation
- Summary statistics
- Effective sample size
- Acceptance rate