More on Importance Sampling (Chapter 8)

Formulation in terms of expectations completely general, since
\[ \int h(x) \, dx = \int \frac{h(x)}{f(x)} f(x) \, dx = E_{X \sim f} \left( \frac{h(X)}{f(X)} \right) \]
where the density \( f(x) > 0 \) where \( h(x) \neq 0 \). If \( X^{(1)}, \ldots, X^{(n)} \sim f \), \( \int h(x) \, dx \) can be estimated by
\[ S^{-1} \sum_{i=1}^{n} h(X^{(i)})/f(X^{(i)}). \]

Choice of \( f \) is very important; locating modes of \( h \) and examining dispersion about modes often useful.

Importance Sampling in Bayesian Inference (Section 8.2)
likelihood = \( L(\beta) \) data = \( y \) prior = \( \pi(\beta) \), posterior = \( \pi(\beta|y) = c^{-1} L(\beta) \pi(\beta), \)
where \( c = \int L(\beta) \pi(\beta) \, d\beta \)

Want to compute \( q^* = \int q(\beta) \pi(\beta|y) \, d\beta \)

for eg \( q(\beta) = \beta, \) \( q(\beta) = \beta^2, \) \( q(\beta) = \exp(\beta), \)
\( q(\beta) = I(\beta \leq x) \) (posterior quantiles).

Often difficult to sample directly from \( \pi(\beta|y) \), so might consider importance sampling.

Don’t usually know \( c \). Given a sample \( \beta^{(1)}, \ldots, \beta^{(S)} \sim g \), then an estimator of \( q^* \) is
\[ \tilde{q} = \frac{\sum_{s=1}^{S} w_s q(\beta^{(s)})}{\sum_{s=1}^{S} w_s}, \]
where \( w_s = L(\beta^{(s)}) c(\beta^{(s)}) / g(\beta^{(s)}) \). \( g \) need not be normalized.

In general (ie even when \( g \) is not normalized), define
\( c = E_g(w_s) = \int L(\beta) \pi(\beta) \, g(\beta) \, d\beta = \int L(\beta) \pi(\beta) \, d\beta \int g(\beta) \, d\beta \)

Properties of \( \tilde{q} \):
\[ \tilde{q} - q^* = \frac{\sum_{s=1}^{S} w_s q(\beta^{(s)}) - q^*}{\sum_{s=1}^{S} w_s} \]

By WLLN, \( \sum_{s=1}^{S} w_s \tilde{q} = E_g(w_s) = c. \) Since
\[ E[w_s q(\beta^{(s)})] = \int q(\beta) L(\beta) \pi(\beta) \, g(\beta) \, d\beta \]

If \( \pi(\beta|y) \) is multimodal, consider a mixture
\( \sum_{k=1}^{K} \psi(\beta; \beta_k, V_k) \), where the \( \beta_k \) are the local modes, and each \( g \) is a multivariate \( T \) or normal.

Importance sampling can often be combined with a control variate:
Suppose \( E_g(\beta) = \mu \) is known. Then the components of \( \beta^{(s)} - \mu \) can be used as control variates. For example, \( E_g(\beta_k|y) \) can be estimated with
\[ \tilde{\beta}_k = \frac{\sum_{s=1}^{S} w_s \beta^{(s)} - \mu_k}{\sum_{s=1}^{S} w_s \beta^{(s)} - \mu_k} \]

An approximately optimal value of \( r \) is
\[ r = S \frac{\text{Cov}(w_s \beta^{(s)} - \beta_k^* \mu)}{\sum_{s=1}^{S} w_s (\beta^{(s)} - \mu_k)^2} \frac{\text{Var}(\beta_k^*)}{\text{Var}(\beta_k^*-\mu_k)} \]
where \( \beta_k^* \) is the true posterior mean. Then
\[ \text{Var}(\tilde{\beta}_k) = S \left( \frac{\sum_{s=1}^{S} w_s (\beta^{(s)} - \beta_k^*)^2}{\sum_{s=1}^{S} w_s (\beta^{(s)} - \mu_k)^2} \right) \frac{\text{Cov}(w_s \beta^{(s)} - \beta_k^* \mu)}{\sum_{s=1}^{S} w_s (\beta^{(s)} - \mu_k)^2} \frac{\text{Var}(\beta_k^*)}{\text{Var}(\beta_k^*-\mu_k)} \]
Estimate posterior means and quantiles of posterior of $\hat{\beta}$.

\[ \text{Posterior mean:} \]
\[ \text{wtsum} \leftarrow \sum \text{wt} \]
\[ \text{wt} \leftarrow \exp(\text{flog}(\text{bhat}) - \text{w2} + \text{w}/2) \]
\[ \text{w2} \leftarrow \sum \exp(\text{flog}(\text{bhat}) - \text{w2} + \text{w}/2) \]
\[ \text{wtsum} \leftarrow \sum \text{wtsum} \]

Standard errors of estimated posterior means:

\[ \sqrt{\text{diag}(\text{solve(inf)}/\text{nt})} \]

Approximate posterior with discrete distribution with mass
\[ w_n = \frac{L(\hat{\beta}) \pi(\hat{\beta})}{L(\hat{\beta}) \pi(\hat{\beta})} \]

where $\pi(\hat{\beta})$ is the prior density of $\beta$.

The posterior is
\[ \pi(\beta|y) \propto L(\beta)\pi(\beta) \]

Estimate posterior means and quantiles of $\hat{\beta}$, $\exp(\hat{\beta})$, and
\[ P(y = 1|x) = p(x|bhat) = \frac{\exp(x'\hat{\beta})}{1 + \exp(x'\hat{\beta})} \]

Below, $n = 200$ with $y_i = 1$ for 132 cases; 5 covariates plus a constant term for $p = 6$; $\sigma^2 = 1000$.

Slide 5

Use $g(\beta) = N(\hat{\beta}, I(\hat{\beta})^{-1})$. Compute $\hat{\beta}$ and $I(\hat{\beta})$:

\[ \text{bhat} \leftarrow \text{nlminb}(\text{rep}(0, \text{np}), \text{flog}) \]
\[ \text{bhat} \leftarrow \text{z$\text{theta}$} \]

Approximation:

\[ \text{wtsum} \leftarrow \sum \text{wtsum} \]

Approximate posterior with discrete distribution with mass
\[ w_n = \frac{L(\hat{\beta}) \pi(\hat{\beta})}{L(\hat{\beta}) \pi(\hat{\beta})} \]

where $\pi(\hat{\beta})$ is the prior density of $\beta$.

The posterior is
\[ \pi(\beta|y) \propto L(\beta)\pi(\beta) \]

Estimate posterior means and quantiles of $\hat{\beta}$, $\exp(\hat{\beta})$, and
\[ P(y = 1|x) = p(x|bhat) = \frac{\exp(x'\hat{\beta})}{1 + \exp(x'\hat{\beta})} \]

Below, $n = 200$ with $y_i = 1$ for 132 cases; 5 covariates plus a constant term for $p = 6$; $\sigma^2 = 1000$.

Slide 6

Use $g(\beta) = N(\hat{\beta}, I(\hat{\beta})^{-1})$. Compute $\hat{\beta}$ and $I(\hat{\beta})$:

\[ \text{bhat} \leftarrow \text{nlminb}(\text{rep}(0, \text{np}), \text{flog}) \]
\[ \text{bhat} \leftarrow \text{z$\text{theta}$} \]

Approximation:

\[ \text{wtsum} \leftarrow \sum \text{wtsum} \]

Approximate posterior with discrete distribution with mass
\[ w_n = \frac{L(\hat{\beta}) \pi(\hat{\beta})}{L(\hat{\beta}) \pi(\hat{\beta})} \]

where $\pi(\hat{\beta})$ is the prior density of $\beta$.

The posterior is
\[ \pi(\beta|y) \propto L(\beta)\pi(\beta) \]

Estimate posterior means and quantiles of $\hat{\beta}$, $\exp(\hat{\beta})$, and
\[ P(y = 1|x) = p(x|bhat) = \frac{\exp(x'\hat{\beta})}{1 + \exp(x'\hat{\beta})} \]

Below, $n = 200$ with $y_i = 1$ for 132 cases; 5 covariates plus a constant term for $p = 6$; $\sigma^2 = 1000$.

Slide 7

Use $g(\beta) = N(\hat{\beta}, I(\hat{\beta})^{-1})$. Compute $\hat{\beta}$ and $I(\hat{\beta})$:

\[ \text{bhat} \leftarrow \text{nlminb}(\text{rep}(0, \text{np}), \text{flog}) \]
\[ \text{bhat} \leftarrow \text{z$\text{theta}$} \]

Approximation:

\[ \text{wtsum} \leftarrow \sum \text{wtsum} \]

Approximate posterior with discrete distribution with mass
\[ w_n = \frac{L(\hat{\beta}) \pi(\hat{\beta})}{L(\hat{\beta}) \pi(\hat{\beta})} \]

where $\pi(\hat{\beta})$ is the prior density of $\beta$.

The posterior is
\[ \pi(\beta|y) \propto L(\beta)\pi(\beta) \]

Estimate posterior means and quantiles of $\hat{\beta}$, $\exp(\hat{\beta})$, and
\[ P(y = 1|x) = p(x|bhat) = \frac{\exp(x'\hat{\beta})}{1 + \exp(x'\hat{\beta})} \]

Below, $n = 200$ with $y_i = 1$ for 132 cases; 5 covariates plus a constant term for $p = 6$; $\sigma^2 = 1000$.

Slide 8
\begin{align*}
+ & \text{out} <- \text{NULL} \\
+ & \text{for} (i \text{ in } \text{pct}) \\
+ & \indent <- \max((1:1:length(p))[wtp<i])+0:1 \\
+ & \text{out} <- \text{rbind(out,c(p[o][ind],wtp[ind]))} \\
+ & \} \\
+ & \text{out} \\
+ \end{align*}

```r
> qntl(H[1,],wt)

\begin{tabular}{rrrr}
[1,] & 0.4961963 & 0.4964762 & 0.02490576 & 0.02500103 \\
[2,] & 0.8590689 & 0.8590836 & 0.49999317 & 0.50009727 \\
[3,] & 1.2611304 & 1.2612538 & 0.97479210 & 0.97502347 \\
\end{tabular}

\text{Slide 9}
```

\begin{align*}
\text{For } q(\bar{\alpha}) = \exp(\bar{\alpha}^2), \text{ all that is needed is to calculate the } q(\bar{\alpha}(s)) \text{ and proceed as before:} \\
+ & \text{qq} <- \exp(H[2,]) \\
+ & \text{sum(qq*wt)/wtsum} \\
+ & \text{[1]} & 0.4616637 \\
+ \end{align*}

```r
> qntl(qq,wt)

\begin{tabular}{rrrr}
[1,] & 0.2996718 & 0.2998061 & 0.02452151 & 0.02515693 \\
[2,] & 0.4524621 & 0.4524712 & 0.49997168 & 0.50006287 \\
[3,] & 0.6670736 & 0.6671004 & 0.97498370 & 0.97505701 \\
\end{tabular}
```

\text{Slide 10}

\begin{align*}
\text{The success probability } p(x_0) x_0 & = (1; 5; 1; 0; -5; -1) \\
+ & \text{p} <- \exp(x_0 \times H) \\
+ & \text{pm} <- \frac{\text{sum(wt*p)/wtsum}}{\text{pm}} \\
+ & \text{[1]} & 0.799744 \\
+ & \text{sqrt(sum((wt*(p-pm)/wtsum)^2))} \\
+ & \text{[1]} & 0.0007656173 \\
+ \end{align*}

```r
> H2 <- H-bhat \\
+ \text{for (i in 1:np) } \\
+ \rho <- \text{cor(H2[i,],bb2[,i])} \\
+ \text{cc} <- \text{nt*var(H2[i,],bb2[,i])}/\text{var(H2[i,])} \\
+ \text{bpmc} <- \text{sum(bb[,i]/wtsum-cc*H2[i,]/nt)} \\
+ \text{print(c(rho=rho,correct=cc,mean=bpmc,} \\
+ \text{se=se.mean[i]*sqrt(1-rho^2)))} \\
+ \} \\
+ \text{rho correct mean se} \\
+ \text{0.7122092} & 1.079125 & 0.8670506 & 0.001969069 \\
+ \text{0.6623438} & 1.057443 & -0.7931535 & 0.00236496 \\
+ \text{0.6862681} & 1.094498 & 1.195384 & 0.002572485 \\
+ \text{0.6930217} & 1.058342 & -0.4124472 & 0.001982007 \\
+ \text{0.5696292} & 1.107328 & 0.566097 & 0.003288518 \\
+ \text{0.8613971} & 1.033094 & 0.04610898 & 0.001159218 \\
\text{With } \frac{1}{\sqrt{\Sigma}} = .707, \text{ the standard error is reduced by } \\
\text{1/\sqrt{\Sigma}, which is equivalent to doubling } S.
```

\text{Slide 11}

\begin{align*}
\text{H2} & <- H-bhat \\
+ \text{for (i in 1:np) } \\
+ \rho & <- \text{cor(H2[i,],bb2[,i])} \\
+ \text{cc} & <- \text{nt*var(H2[i,],bb2[,i])}/\text{var(H2[i,])} \\
+ \text{bpmc} & <- \text{sum(bb[,i]/wtsum-cc*H2[i,]/nt)} \\
+ \text{print(c(rho=rho,correct=cc,mean=bpmc,} \\
+ \text{se=se.mean[i]*sqrt(1-rho^2)))} \\
+ \} \\
+ \text{rho correct mean se} \\
+ \text{0.7122092} & 1.079125 & 0.8670506 & 0.001969069 \\
+ \text{0.6623438} & 1.057443 & -0.7931535 & 0.00236496 \\
+ \text{0.6862681} & 1.094498 & 1.195384 & 0.002572485 \\
+ \text{0.6930217} & 1.058342 & -0.4124472 & 0.001982007 \\
+ \text{0.5696292} & 1.107328 & 0.566097 & 0.003288518 \\
+ \text{0.8613971} & 1.033094 & 0.04610898 & 0.001159218 \\
\text{With } \frac{1}{\sqrt{\Sigma}} = .707, \text{ the standard error is reduced by } \\
\text{1/\sqrt{\Sigma}, which is equivalent to doubling } S.
```

\text{Slide 12}
Markov Chain Monte Carlo (MCMC) (Chapter 9)
Sample from a Markov chain (MC) to generate observations from its stationary distribution.

Markov Chains (Section 9.1)
Discrete time stochastic process
Finite number of states; \( S = \{x_1, \ldots, x_n\} \) is the state space, \( x_s \in \mathbb{R}^p \)
All distributions are discrete on computers (although can closely approximate continuous distributions).
\( X^{(n)} \) = state at time \( n \), \( X^{(0)} \) = initial state. Transition probabilities satisfy the Markov property:
\[
P(X^{(n+1)} = y|X^{(n)} = x) = P_i(j, y)
\]
(Time homogeneous) Let
\[
P^{(n)}(x, y) = P(X^{(n+m)} = y|X^{(n)} = x)
\]
Irreducible if \( P^{(m)}(x, y) > 0 \) for some \( m \) for each \( i, j \).
\( x_i \) is periodic if \( P^{(m)}(x_i, x_i) = 0 \) for \( m \) not divisible by \( d > 1 \). Otherwise aperiodic. For irreducible chains with finite state space, either all states are periodic or all are aperiodic.

Let \( p_{ij} = p(x_i, x_j), P = (p_{ij}) \), and \( P^{(n)} = (p^{(n)}(x_i, x_j)) \).
\[
p^{(2)}(x_i, x_j) = \sum_{k=1}^{s} p(x_i, x_k)p(x_k, x_j), \quad \Rightarrow
p^{(2)} = P^2, \quad \text{and in general } P^{(n)} = P^n.
\]
Let \( \pi^{(0)}(x_i) = P(X^{(0)} = x_i) \) = initial distribution,
\( \pi^{(n)}(x_i) = P(X^{(n)} = x_i) \) = marginal distribution of \( X^{(n)} \), and \( \pi^{(n)} = (\pi^{(n)}(x_1), \ldots, \pi^{(n)}(x_s))' \).
\[
\pi^{(1)}(x_j) = \sum_{i=1}^{S} \pi^{(0)}(x_i)p(x_i, x_j), \quad \Rightarrow
\pi^{(1)} = P\pi^{(0)}, \quad \Rightarrow \pi^{(n)} = (P^n)\pi^{(0)}.
\]
Irreducible, aperiodic chains, have a unique stationary distribution \( \pi = \pi(x_j) \) satisfying
\[
\pi = P\pi,
\]
where \( \pi = (\pi_1, \ldots, \pi_s)' \). If \( \pi^{(0)} = \pi \), then \( \pi^{(n)} = \pi \) for all \( n \). \( \pi \) is in the null space of \( I - P \).

Example. \( S = 4 \) states, with transitions
\[
gP(X_{n+1} = 2|X_0 = 3) =
(1/4)(1/2) + (1/4) \cdot 0 + (1/4) + (1/2)(1/4) = 1/4.
\]
Stationary distribution:
\[
> p \\
\{1, 2, 3, 4\} \\
0.25 \quad 0.50 \quad 0.25 \quad 0.00 \\
0.50 \quad 0.00 \quad 0.00 \quad 0.50 \\
0.25 \quad 0.25 \quad 0.25 \quad 0.50 \\
0.25 \quad 0.25 \quad 0.25 \quad 0.25
> \text{sta.dist <- } \text{svd(diag(scol(P)) - t(P))}$\&$'[, \text{scol(P)}]
> \text{print(sta.dist <- sta.dist/sum(sta.dist))}
[1] 0.3157896 0.2631579 0.1473684 0.2736842
> t(P) %*% sta.dist
[1,] 0.615 0.615
[2,] 0.615 0.615
[3,] 0.615 0.615
[4,] 0.615 0.615

Key property: for any \( \pi^{(0)} \),
\[
\lim_{n \to \infty} \pi^{(n)} = \pi.
\]
(Note: \( P^n \to \text{matrix with all rows = } \pi \).)
To sample from pmf \( f \) on \( S \):
- Construct an MC with \( f \) as its stationary distribution.
- Run a the chain from an arbitrary starting value until convergence.
- Subsequent \( X(t) \) \( \sim \) \( f \).

* How to construct the MC?
* How long to run the chain?
* Values not independent

For irreducible, aperiodic chains, \( \pi \) is the stationary distribution if the reversibility condition

\[
\pi_i p(x_i, x_j) = \pi_j p(x_j, x_i) \quad \text{for all } i, j
\]

holds (sufficient, but not necessary). This follows because

\[
\sum_{i=1}^g \pi_i p(x_i, x_j) = \sum_{i=1}^g \pi_i p(x_j, x_i) = \pi_j.
\]

The Metropolis-Hastings (M-H) Algorithm (Section 9.2)

Goal: Sample from pmf \( f(x) \). Let \( q(x, y) \) be an MC transition kernel. Set

\[
\alpha(x, y) = \min \left\{ \frac{f(y)q(y, x)}{f(x)q(x, y)}, 1 \right\}.
\]

Slide 17

### Random Walk Chain

For \( g \) defined on \( S \), set \( q(x, y) = g(y - x) \). Eg.,

\( g = N(0, H) \), or \( g = \mathcal{T}_r(0, H) \).

If \( g \) is symmetric, then

\[
\alpha(x, y) = \min \left\{ \frac{f(y)g(x - y)}{f(x)g(x - y)}, 1 \right\} = \min\{f(y)/f(x), 1\}.
\]

Independence Chain

\( q(x, y) = g(y) \) \((g = N(\bar{x}, H) \) density\).

\[
\alpha(x, y) = \min \left\{ \frac{f(y)/g(y)}{f(x)/g(x)}, 1 \right\} = \min\left\{ \frac{f(y)}{f(x)}, 1 \right\}.
\]

(ratio of importance sampling weights).

Rejection Sampling Chain

For rejection sampling with dominating function \( g \), may not be sure \( g \geq f \) everywhere.

Use rejection sampling to generate

\( X^{(n+1)} \sim h(x) \propto \min\{f(x), g(x)\} \Rightarrow q(x, y) = h(y) \), and

\[
\alpha(x, y) = \min \left\{ \frac{f(y)h(x)}{f(x)h(y)}, 1 \right\}
\]

If \( g \geq f \) everywhere, then \( \alpha(x, y) = 1 \). Points with \( g(x) < f(x) \) will have \( \alpha(x, y) < 1 \) for some \( y \)'s, \( \Rightarrow x \)

sometimes retained, increasing the frequency of non-dominated points.

Is importance sampling, independence chain sampling or rejection chain sampling better?

**Block-at-a-Time Algorithms** (Section 9.3)

Don't need to update all components of \( X^{(n)} \) simultaneously.

Suppose \( X = (U, Y) \sim f(x) \), and Suppose \( p_{12}(\cdot | r) \) and \( p_{21}(\cdot | w) \) are MC transition kernels with stationary distributions \( f_{12}(\cdot|v) \) and \( f_{21}(\cdot|w) \).

Given \( X^{(n)} = (U^{(n)}, V^{(n)}) \),

1. generate \( U^{(n+1)} \sim p_{12}(U^{(n)}, \cdot|V^{(n)}) \).
2. generate \( V^{(n+1)} \sim p_{21}(V^{(n)+1}, \cdot|U^{(n)+1}) \).

\( X^{(n+1)} = (U^{(n)+1}, V^{(n)+1}) \) has transition kernel

\[
p(x, y) = p(u, v)p(u, w)p_{12}(u, v)p_{21}(v, z|w),
\]

which has stationary distribution \( f(u, v) \). That is,

\[
\sum_{u} \sum_{v} f(u, v)p(u, v)p(u, w)p_{12}(u, v)p_{21}(v, z|w)
\]

Slide 19

1. Given \( X^{(n)} \), sample a trial value \( X^{(n+1)} \sim q(X^{(n)} \cdot) \).
2. For \( U^{(n)} \sim U(0, 1) \),

\[
X^{(n+1)} = \begin{cases} X^{(n+1)} & U^{(n)} \leq \alpha(X^{(n)}, X^{(n+1)}) \\ X^{(n)} & U^{(n)} > \alpha(X^{(n)}, X^{(n+1)}) \end{cases}
\]

Transition kernel is

\[
p(x, y) = \begin{cases} q(x, y) & y \neq x \\ 1 - \sum_{y \neq x} q(x, y) & y = x. \end{cases}
\]

\( f, p \) satisfy the reversibility condition; eg, suppose \( x \neq y \) and \( f(x)q(x, y) > f(y)q(y, x) \). Then

\[
\alpha(x, y) = \frac{f(y)q(y, x)}{f(x)q(x, y)} \quad \text{for all } x, y, z.
\]

\( f \) need not be normalized.

Can be slow if \( q(x, y) \) is not close to \( f(x) \).

Let \( \bar{x} \) be the mode of \( f(x) \), and let

\[
H = -(\partial^2 \log f(x)/\partial x^2)^{-1}.
\]

Slide 18
Transformations to reduce dependence among approximations.

Only need to find transition kernels for conditional distributions (often much easier to find good approximations).

Can use M-H for each block.

Block-at-a-time updates often induce greater autocorrelation.

Transformations to reduce dependence among components can help.

Gibbs Sampling (Section 9.3.1)

\[
p_{12}(u, w|v) = f_{2|1}(v|w), \quad p_{21}(v, z|w) = f_{1|2}(v|z)
\]

(sample directly from conditional distributions). Eg hierarchical normal random effects models, and incomplete data problems with normal distributions.

Implementation Issues (Section 9.4)

'burn-in': # steps until chain 'converges'.

Slide 21

\[
\begin{align*}
\sigma^2 & = \frac{1}{N - B} \sum_{j=1}^{N} \sum_{k=1}^{B} (2 \rho_{k,j} - \rho_{k}) \\
\text{Var}(\hat{b}) & = \frac{\sigma^2}{N - B} \sum_{j=1}^{N} \sum_{k=1}^{B} (N - B - j) \rho_{k}
\end{align*}
\]

Usual empirical moments are consistent for \( \sigma^2 \) and the \( \rho_k \).

Simpler approach: suppose \( N - B = Jm \), and let

\[
\hat{b}_j = \frac{\sum_{i=1}^{B+Jm} b(X^{(i)})/m}{Jm}, \quad j = 1, \ldots, J.
\]

Note \( \hat{b} = \sum_{i=1}^{B+Jm} b_j/\sqrt{N} \). If \( m \gg K \), then \( \rho = \text{cor}(\hat{b}, \hat{b}_j) \approx 0 \), and \( \text{Var}(\hat{b}) \approx \text{Var}(b_j)/J \).

\[
\text{cor}(\hat{b}_j, \hat{b}_j) = \left\{ \begin{array}{ll}
1 & |j - j'| = 1 \\
0 & |j - j'| > 1
\end{array} \right.
\]

\[
\text{Var}(\hat{b}) = \frac{\text{Var}(b_j)}{J} + \frac{2}{J} \sum_{j=1}^{J} \text{Cov}(\hat{b}_j, \hat{b}_{j+1})
\]

Slide 23

After burn-in, keep (a) 1 value, (b) every \( l \)th value, (c) all values.

Parallel chains vs 1 long run.

Convergence diagnostics:

- Any empirical test can fail
- Theoretical bounds of limited applicability
- Plot \( X^{(n)} \) and \( \sum_{i} X^{(i)} - N \) vs \( n \)
- Gelman and Rubin (1992): compare within and between chain variability of parallel chains
- Many others (Cowles and Carlin, 1996)

Finite run could miss important subregions (eg if multi-modal). Could search for local modes, and start chains within each.

Precision of Estimates (Section 9.4.1)

\( X^{(1)} \), \( X^{(2)} \), \ldots, \( X^{(N)} \) has burn-in \( B \). Can estimate \( \text{E}(\hat{b}) = \int b(x)f(x)dx \) with

\[
\hat{b} = \frac{1}{N} \sum_{i=B+1}^{N} b(X^{(i)})/(N - B),
\]

(\( \text{E}(\hat{b}) = \text{E}(b) \)). Let \( \sigma^2 = \text{Var}(b(X^{(i)})) \), \( \rho_k = \text{cor}(b(X^{(i)}), b(X^{(i+k)})) \) (do not depend on \( i \) for

Slide 22

\[
\begin{align*}
\text{varest} & \leftarrow \text{function}(x, m=100) \{
\# \text{ estimate standard error of mean of correlated } \\
\# \text{ sequence } x \text{ using variation in means of blocks } \\
\# \text{ of size } n \text{. Assumes adjacent blocks are } \\
\# \text{ correlated, but blocks at greater lags are not} \\
\text{ng} & \leftarrow \text{floor(length}(x)/m) \\
\text{if}(\text{ng}==\text{length}(x)) & \text{ } x \leftarrow x[\{1:(\text{length}(x)-\text{ng}+1))] \\
\text{ind} & \leftarrow \text{rep}(1:ng, \text{rep}(m, ng)) \\
\text{gmx} & \leftarrow \text{tapply}(x, \text{ind}, \text{mean}) \\
\text{mx} & \leftarrow \text{mean}(\text{gmx}) \\
\text{rho} & \leftarrow \text{cor}(\text{gmx}[\text{-1}], \text{gmx}[\text{-length}(\text{gmx})]) \\
\text{c} & \leftarrow \text{mean}(\text{mx}, \text{sv}==\text{sqrt}(\text{var}(\text{gmx})*(1+2*\text{rho}/\text{ng}), \text{rho}==\text{rho}) \\
\}
\end{align*}
\]

One-way Random Effects Example (Section 9.5)

\[
\begin{align*}
\rho_{ij} & = \theta_i + \epsilon_{ij}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, J \\
\theta_i & \sim \text{iid } N(\mu, \sigma^2), \epsilon_{ij} & \sim \text{iid } N(0, \sigma^2). \text{ Priors:} \\
\mu & \sim N(\mu_0, \sigma^2_0), \nu_0 = 1/\sigma^2_0 \sim \Gamma(\alpha_1, b_1), \\
\nu_y = 1/\sigma^2_y & \sim \Gamma(a_2, b_2), \text{ where } \Gamma(\alpha, \beta) \text{ has density}
\end{align*}
\]
\( x \sim \mathcal{N}(\mu, \nu_0, \nu_0) \).

Conditional on \( (\mu, \nu_0, \nu_0) \),

\[
(Y_1, \ldots, Y_J)' \sim Y_i \sim \text{iid } \mathcal{N}(\mu J, \nu_0^{-1} I + \nu_0^{-1} J)' \]

Posterior means and variances could be calculated using G-H quad., but will use Gibbs sampling.

Think of \( \theta_i \) as parameters with a hierarchical prior.

Conditional on \( \theta \equiv (\theta_1, \ldots, \theta_J)' \) and \( \nu, \nu_0 \),

\[
Y_i \sim \text{iid } \mathcal{N}(\theta_i, \nu_0^{-1}) \Rightarrow L(\theta, \nu, \nu_0) \propto \nu_0^{J/2} \exp \left( - \sum_{i,j} (Y_{ij} - \theta_i)^2 / 2 \nu_0 \right),
\]

and the posterior

\[
\pi(\theta, \mu, \nu_0, \nu|Y) \propto L(\theta, \nu, \nu_0) g_0(\theta|\mu, \nu_0) g_1(\nu_0|\nu),
\]

where \( g_0, g_1, g_2 \), and \( g_3 \) are the prior densities of \( |\theta|, \mu, \nu_0, \) and \( \nu \).

Need to sample from the conditional distributions

\[
\begin{align*}
[\theta_i|\nu, \nu_0, \mu] & \sim N \left( \frac{J \overline{Y}_i / \nu_0 + \mu \nu_0}{J \nu_0 + \nu_0}, \frac{1}{J \nu_0 + \nu_0} \right), i = 1, \ldots, n, \\
[\mu|\theta, \nu_0] & \sim N \left( \frac{\sum_i \theta_i / \nu_0}{J \nu_0 + \nu_0}, \frac{1}{J \nu_0 + \nu_0} \right).
\end{align*}
\]

Slide 25

### Slide 26

Derivation of \( [\nu|\mu, \theta] \):

\[
L(\theta, \nu, \nu_0|Y) \propto \nu_0^{J/2} \exp \left( - \sum_{i,j} (Y_{ij} - \theta_i)^2 / 2 \nu_0 \right) \\
\propto \nu_0^{J/2} \nu_0^{-1} \exp \left( - \sum_{i,j} (Y_{ij} - \theta_i)^2 / 2 \right). \\
\]

\( m0 = \mu_0 = 0, nu0 = 1/\sigma_n^2 = 10^{-5} \), and \( a1, b1, a2, b2 \) (\( a1, b1, a2, b2 \)) all \( = 0.001 \). Generate data (data saved to file gibbs.dat).

- \( K \leftarrow 6 \); \( J \leftarrow 5 \); \( nu \leftarrow 5 \); \( vt \leftarrow 4 \); \( ve \leftarrow 16 \); \( .Random.seed \)
  
  [1] 1 53 41 59 44 3 59 6 53 42 5 3

- \( \theta \leftarrow \text{rnorm}(K, \text{mu}, \text{sqrt}(vt)) \)

- \( Y \leftarrow \text{rnorm}(K, J, \text{sqrt}(ve)) + \text{rep}(\theta, \text{rep}(J, K)) \)

- \( \text{group} \leftarrow \text{rep}(1:K, \text{rep}(J, K)) \)

- \( \text{ybar} \leftarrow \text{tapply}(Y, \text{group}, \text{mean}) \)

- \( \text{write}(\text{rbind}(Y, \text{group}, \text{"gibbs.dat"}, 2)) \)

- \( \text{ybar} \leftarrow 1 2 3 4 5 \)

- \( 5.097592 \ 7.457386 \ -0.2985133 \ 8.141821 \ -1.162981 \)

- \( \gamma(\alpha/2 + a_2, b_2 + \sum_i (\theta_i - \mu)^2 / 2) \)

- \( \gamma(\alpha J/2 + a_1, b_1 + \sum_i (Y_{ij} - \theta_i)^2 / 2) \)

- \( \gamma(\alpha J/2 + a_1, b_1 + \sum_i (Y_{ij} - \theta_i)^2 / 2) \)

The following function runs the Gibbs sampler. Generated an initial sequence of 500 values.

- \( \text{gibbs.re} \leftarrow \text{function}(\text{nt}) \{ \)
  + \( \text{print}(\text{.Random.seed}) \)
  + \# prior distribution parameters
    + \( m0 \leftarrow 0; \nu0 \leftarrow 1 \times 10^{-5}; a1 \leftarrow .001; b_1 \leftarrow .001; b_2 \leftarrow .001 \)
  + \( \text{for} (i \in 2:nt+1) \{ \)
    + \( j \leftarrow i-1 \)
    + \( vt \leftarrow 1/(nu.\text{es}[j]*J+nu.\text{th}s[j]) \)
    + \( \text{th}[j] \leftarrow \text{rnorm}(K, J*ybar[nu.\text{es}[j]+nu.\text{th}s[j]*nu.\text{es}[j]], vt, \text{sqrt}(vt)) \)
    + \( \text{vt} \leftarrow 1/(K*nu.\text{th}s[j]*nu0) \)
    + \( \text{mu}[j] \leftarrow \text{rnorm}(1, nu.\text{th}s[j]*K*mean(nu.\text{es}[j]*nu.\text{th}s[j])) \)
  + \} \)
- \( \text{gibbs.re} \leftarrow \text{function}(\text{nt}) \{ \)
  + \( \text{print}(\text{.Random.seed}) \)
  + \# initial values for Gibbs Sampling
    + \( \text{mu} \leftarrow \text{rep}(0,nt+1) \)
    + \( \text{nu.es} \leftarrow \text{na.ths} \leftarrow \text{rep}(1,nt+1) \)
    + \( \text{th} \leftarrow \text{matrix}(0, \text{nu.es}[j]*K, \text{ncol}=nt+1) \)
    + \( \text{for} (i \in 2:nt+1) \{ \)
      + \( j \leftarrow i-1 \)
      + \( vt \leftarrow 1/(nu.\text{es}[j]*J+nu.\text{th}s[j]) \)
      + \( \text{th}[j] \leftarrow \text{rnorm}(K, J*ybar[nu.\text{es}[j]+nu.\text{th}s[j]*nu.\text{es}[j]], vt, \text{sqrt}(vt)) \)
      + \( \text{vt} \leftarrow 1/(K*nu.\text{th}s[j]*nu0) \)
      + \( \text{mu}[j] \leftarrow \text{rnorm}(1, nu.\text{th}s[j]*K*mean(nu.\text{es}[j]*nu.\text{th}s[j])) \)
  + \} \)
- \( \text{gibbs.re} \leftarrow \text{function}(\text{nt}) \{ \)
  + \( \text{print}(\text{.Random.seed}) \)
  + \# initial values for Gibbs Sampling
    + \( \text{mu} \leftarrow \text{rep}(0,nt+1) \)
    + \( \text{nu.es} \leftarrow \text{na.ths} \leftarrow \text{rep}(1,nt+1) \)
    + \( \text{th} \leftarrow \text{matrix}(0, \text{nu.es}[j]*K, \text{ncol}=nt+1) \)
    + \( \text{for} (i \in 2:nt+1) \{ \)
      + \( j \leftarrow i-1 \)
      + \( vt \leftarrow 1/(nu.\text{es}[j]*J+nu.\text{th}s[j]) \)
      + \( \text{th}[j] \leftarrow \text{rnorm}(K, J*ybar[nu.\text{es}[j]+nu.\text{th}s[j]*nu.\text{es}[j]], vt, \text{sqrt}(vt)) \)
      + \( \text{vt} \leftarrow 1/(K*nu.\text{th}s[j]*nu0) \)
      + \( \text{mu}[j] \leftarrow \text{rnorm}(1, nu.\text{th}s[j]*K*mean(nu.\text{es}[j]*nu.\text{th}s[j])) \)
\[ \text{nu.ths}[i] \leftarrow \frac{\text{rgamma}(1, K/2 + a2)}{\left( \sum((\text{ths}[,i] - \text{mus}[i])^2)/2 + b2 \right)} \]

\[ \text{nu.es}[i] \leftarrow \frac{\text{rgamma}(1, J*K/2 + a1)}{b1 + \sum((Y - \text{ths}[\text{group},i])^2)/2) \]

\} + 
+ \text{list}(\text{mus}=\text{mus}, \text{nu.es}=\text{nu.es}, \text{nu.ths}=\text{nu.ths}, \text{ths}=\text{ths}) \]

\[ \text{nt} \leftarrow 500 \]

\[ \text{run1} \leftarrow \text{gibbs.re(}\text{nt}) \]

\[ \begin{array}{c}
\text{[1]} 9 24 29 51 30 2 6 5 6 30 56 3 \\
\text{postscript(“fig1.ps”, pointsize=20)} \\
\text{plot(1:nt, 1/run1$\text{nu.ths}[-1], type=“l”, xlab=}
+ “\text{Iteration”, main=“Variance of Theta Dist”} \\
\text{dev.off()} \end{array} \]

Use a burn-in of 100, and generate a total of 5000 cycles.

\[ \text{run2} \leftarrow \text{gibbs.re}(5000) \]

\[ \begin{array}{c}
[1] 1 63 42 14 52 3 63 11 22 47 14 3 \\
\text{ths} \leftarrow \text{run2$ths[, -(1:101)]} \\
\text{mus} \leftarrow \text{run2$mus[, -(1:101)]} \\
\text{nu.ths} \leftarrow \text{run2$nu.ths[, -(1:101)]} \\
\text{nu.es} \leftarrow \text{run2$nu.es[, -(1:101)]} \\
\text{# How much correlation in the sequence?} \\
\text{acor} \leftarrow \text{function(x, k) cor(x[-(1:k)],} \\
+ x[-((\text{length(x)-k+1):length(x))}]) \\
\text{for (k in c(1, 10, 20, 30, 40, 50, 100, 200,} \\
+ 300)) \text{print(acor(1/nu.ths, k))} \\
[1] 0.2582468 \\
[1] 0.001480924 \\
[1] 0.001870187 \\
[1] -0.006411884 \\
[1] -0.009027028 \\
[1] -0.01824866 \\
[1] -0.001897716 \\
[1] -0.003997428 \\
[1] 0.008303639 \\
\text{for (k in c(1, 10, 20, 30, 40, 50, 100, 200,} \\
+ 300)) \text{print(acor(1/nu.es, k))} \\
[1] 0.3178706 \end{array} \]
Repeat runs with different priors:

\[ a_2 \leftarrow 2 \text{ # prior for var of random effects} \]
\[ \# \text{ has mode at 1, var of 2000000).} \]
\[ \text{run3} \leftarrow \text{gibbs.re(5000)} \]
\[ \text{for (k in c(1,10,20,30,40,50)) print(acor(1/nu.ths,k))} \]
\[ [1] \quad 0.6487311 \]
\[ [1] \quad -0.01415425 \]
\[ [1] \quad 0.004182847 \]
\[ [1] \quad -0.0003998655 \]
\[ [1] \quad -0.01843221 \]
\[ [1] \quad -0.006256902 \]
\[ \text{for (k in c(1,10,20,30,40,50,100, + \quad 200,300)) print(acor(mus,k))} \]
\[ [1] \quad 0.9991632 \]
\[ [1] \quad 0.9912071 \]
\[ [1] \quad 0.9827219 \]
\[ [1] \quad 0.9748673 \]
\[ [1] \quad 0.9662091 \]
\[ [1] \quad 0.9564646 \]
\[ [1] \quad 0.9043435 \]
\[ [1] \quad 0.8012627 \]
\[ [1] \quad 0.6749711 \]
\[ \quad \text{# the following se is not valid,} \]
\[ \quad \text{# since the correlation in blocks at} \]
\[ \quad \text{# lag more than 1 is much larger than 0.} \]
\[ \quad \text{varest(mus)} \]
\[ \quad \text{mean se rho} \]
\[ \quad 5.085794 \quad 0.1066138 \quad 0.9319637 \]
\[ \quad \text{varest(1/nu.ths)} \]
\[ \quad \text{mean se rho} \]
\[ \quad 0.0009894144 \quad 7.207767e-05 \quad 0.1943027 \]
\[ \quad \text{postscript(’plot2.ps’,pointsize=20)} \]
\[ \quad \text{plot(1:length(mus),mus,type=’l’,xlab= ’Iteration’,main=’Mean of Theta Dist’)} \]
\[ \quad \text{dev.off()} \]
> b2 <- 1
> run4 <- gibbs.re(5000)
  [1] 45 8 31 63 55 3 7 2 40 52 64 0
> ths <- run4%ths[-(1:101)]
> mus <- run4%mus[-(1:101)]
> nu.ths <- run4$nu.ths[-(1:101)]
> nu.es <- run4$nu.es[-(1:101)]
> mus <- run4$mus[-(1:101)]
> ths <- run4$ths[-(1:101)]
> run4 <- gibbs.re(5000)
> b2 <- 1

\[ \begin{align*}
\text{mean} & \quad \text{se} \quad \text{rho} \\
0.8282945 & \quad -0.7588482 \quad 1.1381240 \quad -0.3916353
\end{align*} \]

then \( \beta^{(i)} = \beta^{(j)} \), otherwise \( \beta^{(i)} = \beta^{(j-1)} \) (and \( w_j = w_{j-1} \)).

\[ 2.5\% \quad 50.0\% \quad 97.5\% \]

0.5320517 -0.0423460 0.1115138 0.1180735

\[ 0.7687718 \quad 4.392122 \quad 7.927748 \]

Bayesian Analysis of Logistic Regression (Section 9.6)

\[ y_i = \text{binary responses}, \quad x_i = p\text{-dimensional covariate,} \]

\[ L(\beta) = \prod_{i=1}^{n} \exp[y_i x_i^T \beta] \]

Prior is \( \pi(\beta) \propto \exp[-\beta^T \beta/(2\sigma^2)] \).

Data: \( n = 200 \), with \( y_i = 1 \) for 132 cases, 5 covariates \( (p = 6) \), \( \sigma^2 = 1000 \).

\[ \hat{\beta} = \text{posterior mode}, \quad I(\hat{\beta}) = -(\partial^2 / \partial \theta \partial \theta^T) \log[L(\theta) \pi(\theta)] \]

\[ w_{\beta} = w(\beta^{(i)}) = \frac{L(\beta^{(i)}) \pi(\beta^{(i)}) / [L(\hat{\beta}) \pi(\hat{\beta})]}{\exp[-(\beta^{(i)} - \beta)^T I(\beta^{(i)}) (\beta^{(i)} - \beta)/2]} \]

Independence Chain Sampling (Section 9.6.1)

\[ q(\beta^{(i)}, \beta) = g(\beta), \quad \text{with} \quad g = \text{the } N(\hat{\beta}, I^{-1}(\hat{\beta})) \text{ density.} \]

\[ a_\beta(\beta^{(i)}, \beta) = \frac{\pi(\beta^{(i)})g(\beta)}{\pi(\beta)g(\beta^{(i)})} = \frac{w(\beta)}{w_{\beta}}. \]

Generate \( \beta^{(i)} \) from \( g \) as before (omitted). For \( j = 2, \ldots, N \), sample \( U_j \sim U(0,1) \). If \( U_j < w_j/w_{j-1} \).

\[ \text{2.5\%} \quad 50.0\% \quad 97.5\% \]

0.7687718 4.392122 7.927748
Rejection Chain Sampling (Section 9.6.2)

Proposed dominating function: \(N(\bar{\beta}; \sigma^2 I - \beta^2 I)\) density, rescaled to \(= \frac{4}{\bar{\beta}}\) at \(\bar{\beta} = 10\,000\) generated for the initial rejection sampling.

3,846 are rejected. Sampling density is pointwise min of posterior and the normal density.

> ## Metropolis-Hastings rejection chain sampling
> nt <- 10000
> Z <- matrix(rnorm(nt*np),nrow=np)
> # normal approx, with variance inflated by 1.2
> H <- solve.tr(sqrt(1/1.2)*B,Z)+bhat #B as before
> # calculate ratios for rejection sampling
> w <- rep(1,np) %*% (Z*Z)
> w2 <- apply(H,2,flog) # flog as before
> # densities not normalized
> f <- exp(flog(bhat)-w2) # 'dominating' function
> g <- exp(-w/2)
> ratio <- f/g
> summary(ratio)

<table>
<thead>
<tr>
<th></th>
<th>Min.</th>
<th>1st Qu.</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>0.0001585</td>
<td>0.4579</td>
<td>0.6527</td>
<td>0.6166</td>
<td>0.7943</td>
<td>4.479</td>
</tr>
<tr>
<td>Median</td>
<td>0.010092943</td>
<td>0.009675866</td>
<td>0.055761542</td>
<td>0.008355903</td>
<td>0.0020893868</td>
<td>0.004500746</td>
</tr>
</tbody>
</table>

Of the 6,154 accepted by rejection sampling, all but 31 were accepted by M-H, so nearly independent.
Example: Institutional Effects in EST 1582 (not in notes)

Comparison of CAV vs. CAV-HEM induction for treatment of small cell lung cancer.

- 570 cases from 26 institutions
- /institution, range = 5 to 56, median = 18.5.
- Endpoint: survival (10/570 censored)
- Significant Covariates: initial performance status, liver metastases, bone metastases, prior weight loss, & treatment.

Institutions: i = 1, . . . , N
Subjects / Institution: j = 1, . . . , ni.

Underlying hazard: piecewise constant

\[ \lambda_0(t) = \exp\left(\sum_{i=1}^N \alpha_i I_i(t)\right). \]

\[ z_{ij} = 2I(CAV-HEM) - 1, \] and \( x_{ij} \) denotes the other

Slide 45

\[ \left[\Sigma^{-1}\right] \sim W(N + d, [V^{-1} + \sum_{i=1}^N \theta_i \theta_i^T]^{-1}) \]

\( \left[\theta_i\right] = (\theta_0, \theta_1)_i \). Let \( \delta_{i,jk} = I(\text{subject } j \text{ is in interval } k), u_{ijk} = \text{total follow-up time for subject } j/i \text{ in interval } k. \)

\[ l(\eta) = \sum_{i,j,k} \delta_{i,jk} \psi_{ijk} - u_{ijk} \exp(\psi_{ijk}). \]

where \( \eta = (\alpha, \beta, \gamma, \theta) \) and

\[ \psi_{ijk} = \alpha_k + x_{ij} \beta + z_{ij}(\beta + \theta_0) + \theta_0. \]

Log conditional posterior density of \( \beta(\gamma, \alpha, \theta) \) is

\[ \log(\pi_\beta) \propto l(\eta) + \log(\text{prior}(\beta)) \]

\[ \propto \sum_{i,j,k} (\delta_{i,jk} \beta - \phi_{ijk} \exp(z_{ij} \beta)) - \beta^2/(2\sigma_\beta^2), \]

where

\[ \phi_{ijk} = \sum_{ij,k} \exp(\psi_{ijk} - z_{ij} \beta). \]

- Sample \( \beta \) by using rejection chain sampling: Given the current values of \( \gamma, \alpha, \theta \), find the mode \( \hat{\beta} \) of \( \pi_\beta(\beta) \) and compute \( h = (-\beta^2 \log(\pi_\beta(\beta))/\partial \beta^2)^{-1} \).

- ‘Dominating’ function is the \( N(\hat{\beta}, 2.25h) \) density, rescaled to be 5% larger than \( \exp(\pi_\beta) \) at \( \hat{\beta} \).

Slide 47

- A similar algorithm used for each component of \( \gamma \) (each updated separately)
- For \( \theta_0, \theta_1, \) used a similar rejection chain algorithm, with \( \theta_0, \theta_1 \) updated jointly.
- For \( \alpha_k \), set

\[ T_k = \sum_{i,j,k} \exp(z_{ij} \beta + z_{ij}(\beta + \theta_0) + \theta_0) \]

For \( k = 2, \ldots, m-1, \) for any \( m_k, \) density

\[ \propto \exp(-\frac{1}{2}(\alpha_k - \alpha_{k-1})^2 - \frac{1}{4}(\alpha_k - \alpha_{k-1})^2 + l(\eta)) \]

\[ \propto \exp(-\frac{1}{2}(2(\mu_k - m_k)^2 + 2(\alpha_k - m_k)(2m_k - m_{k-1} - \alpha_k + 1)) + \frac{1}{2}\alpha_k - T_k e^{\mu_k}) \]

\[ \propto \exp(\alpha_k \delta_{i,k} + 2(\mu_k - m_k) - T_k e^{\mu_k}) \times \exp(-\frac{1}{2}(\alpha_k - m_k)^2), \]

where \( \mu_k = (\alpha_k + \alpha_{k-1})/2 \). Choose \( m_k \) to match the modes of the normal and log gamma pieces, and sample from the piece that is more concentrated. The dominating function is the rescaled sampling density.

\( m = 30, \) with roughly equal numbers of failures in each interval

Slide 48
Parameters of Priors: \( \sigma_2^2 = 10,000, \sigma_3^2 = \infty \), 
\( a = 3, b = .05 \) (mean=60, var=1200; \( \nu \) should be large, to make var of jumps in log hazard small),
\( d = 2 \) and
\[
V = \begin{pmatrix}
125 & 0 \\
0 & 500
\end{pmatrix}
\]
Note \( dV = E(\Sigma^{-1}) \). Gave \( P(\text{Var}(\theta_0) > .364) = 12.1\% \) and \( P(\text{Var}(\theta_0) < .00334) = 12.4\% \).
Sampled in the order \( \gamma, \beta, \theta, \Sigma, \alpha, \nu \). Generated one stream of 21,000 cycles, discarded the first 1,000, retained every 4th cycle from the remaining 20,000. Calculations done in Fortran.

<table>
<thead>
<tr>
<th>( \beta )</th>
<th>Mean</th>
<th>Sim-SE</th>
<th>1st Auto</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta_{11} )</td>
<td>-137</td>
<td>.001</td>
<td>.15</td>
</tr>
<tr>
<td>( \Sigma_{22} )</td>
<td>.019</td>
<td>.001</td>
<td>.73</td>
</tr>
</tbody>
</table>

Some functions of interest:
(a) \( \exp(\theta_0 - \theta_{11}) \): ratio of hazard for CAV patients in the \( i \)th institution / overall hazard for CAV.
(b) \( \exp(\theta_0 + \theta_{11}) \): ratio of hazard for CAV-HEM patients in the \( i \)th institution / overall hazard for CAV-HEM.
(c) \( \exp(2\theta_{11}) \): ratio of treatment hazard ratio in the \( i \)th institution to the overall average.

First figure gives posterior percentiles of (a), (b) and (c). Second figure gives \( P(Q > R) \) and \( P(Q < 1/R) \) for \( Q = (a) \) (panels (a) and (b)) and \( Q = (b) \) ((c) and (d)).
Predictive Distribution:

\[ S_{ij}(t) = \int \exp \left\{ - \int_0^t \lambda(u|x_{ij}, z_{ij}, \eta) \, du \right\} \pi(\eta|y) \, dy. \]

Estimated by

\[ \sum_{s = B+1}^S \exp\left( - \int_0^t \lambda(u|x_{ij}, z_{ij}, \eta^{(s)}) \, du \right)/ (S - B). \]