MCMC II

James M. Flegal

Agenda

- Markov chain Monte Carlo, again
- Gibbs sampling
- Output analysis for MCMC
- Convergence diagnostics
- Examples: Capture-recapture and toy example

Gibbs Sampling

1. Select starting values x_0 and set $t = 0$
2. Generate in turn (deterministic scan Gibbs sampler)
   - $x_{t+1}^{(1)} \sim f(x^{(1)}|x_t^{(-1)})$
   - $x_{t+1}^{(2)} \sim f(x^{(2)}|x_{t+1}^{(1)}, x_t^{(3)}, \ldots, x_t^{(p)})$
   - $x_{t+1}^{(3)} \sim f(x^{(3)}|x_{t+1}^{(1)}, x_{t+1}^{(2)}, x_t^{(4)}, \ldots, x_t^{(p)})$
   - $\ldots$
   - $x_{t+1}^{(p)} \sim f(x^{(p)}|x_{t+1}^{(-p)})$
3. Increment t and go to Step 2

Gibbs Sampling

- Common to have one or more components not available in closed form
- Then one can just use a MH sampler for those components known as a Metropolis within Gibbs or Hybrid Gibbs sampling
- Common to "block" groups of random variables

Example: Capture-recapture

- Data from a fur seal pup capture-recapture study for $i = 7$ census attempts
- Goal is to estimate the number of pups in a fur seal colony using a capture-recapture study

Count	Parameter	1	2	3	4	5	6	7
Captured	c_i	30	22	29	26	31	32	35
Newly Caught	m_i	30	8	17	7	9	8	5

Example: Capture-recapture

- Let N be the population size, I be the number of census attempts where c_i were captured ($I = 7$ in our case), and r be the total number captured ($r = \sum_{i=1}^{I} m_i = 84$)
- We consider a separate unknown capture probability for each census $(\alpha_1, \ldots, \alpha_I)$ where the animals are equally "catchable"
- Then

$$L(N, \alpha | c, r) \propto \frac{N!}{(N-r)!} \prod_{i=1}^{I} \alpha_i^{c_i} (1 - \alpha_i)^{N - c_i}$$

Example: Capture-recapture

- Assume N and α are apriori independent with

$$f(N) \propto 1 \text{ and } f(\alpha_i|\theta_1, \theta_2) \overset{i.i.d.}{\sim} \text{Beta}(\theta_1, \theta_2)$$

- We use $\theta_1 = \theta_2 = 1/2$, which is the Jeffrey's Prior
- The resulting posterior is proper when $I > 2$ and recommended when $I > 5$

Example: Capture-recapture

- Then it is easy to show the posterior is

$$f(N, \alpha | c, r) \propto \frac{N!}{(N-r)!} \prod_{i=1}^{I} \alpha_i^{c_i} (1-\alpha_i)^{N-c_i} \prod_{i=1}^{I} \alpha_i^{-1/2} (1-\alpha_i)^{-1/2}.$$

- Further, one can show

$$N - 84 | \alpha \sim \text{NB}\left(85, 1 - \prod_{i=1}^{I} (1-\alpha_i)\right) \text{ and}$$

$$\alpha_i | N \sim \text{Beta}\left(c_i + 1/2, N - c_i + 1/2\right) \text{ for all } i$$

Example: Capture-recapture

Then we can consider the chain

$$(N, \alpha) \to (N', \alpha) \to (N', \alpha')$$

or

$$(N, \alpha) \to (N, \alpha') \to (N', \alpha'),$$

where both involve a “block” update of α

Example: Capture-recapture

First, we can write the data into R

```
captured <- c(30, 22, 29, 26, 31, 32, 35)
new.captures <- c(30, 8, 17, 7, 9, 8, 5)
total.r <- sum(new.captures)
```

Example: Capture-recapture

The following R code implements the Gibbs sampler

```
gibbs.chain <- function(n, N.start = 94, alpha.start = rep(.5,7)) {
    output <- matrix(0, nrow=n, ncol=8)
    for(i in 1:n){
        neg.binom.prob <- 1 - prod(1-alpha.start)
        N.new <- rnbinom(1, 85, neg.binom.prob) + total.r
        beta1 <- captured + .5
        beta2 <- N.new - captured + .5
        alpha.new <- rbeta(7, beta1, beta2)
        output[i,] <- c(N.new, alpha.new)
        N.start <- N.new
        alpha.start <- alpha.new
    }
    return(output)
}
```

MCMC output analysis

How can we tell if the chain is mixing well?

- Trace plots or time-series plots
- Autocorrelation plots
- Plot of estimate versus Markov chain sample size
- Effective sample size (ESS)

$$\mathrm{ESS}(n) = \frac{n}{1 + 2\sum_{k=1}^{\infty} \rho_k(g)},$$

  where $\rho_k(g)$ is the autocorrelation of lag k for g
- Alternative, ESS can be written as

$$\mathrm{ESS}(n) = \frac{n}{\sigma^2/\mathrm{Var}g}$$

  where σ^2 is the asymptotic variance from a Markov chain CLT

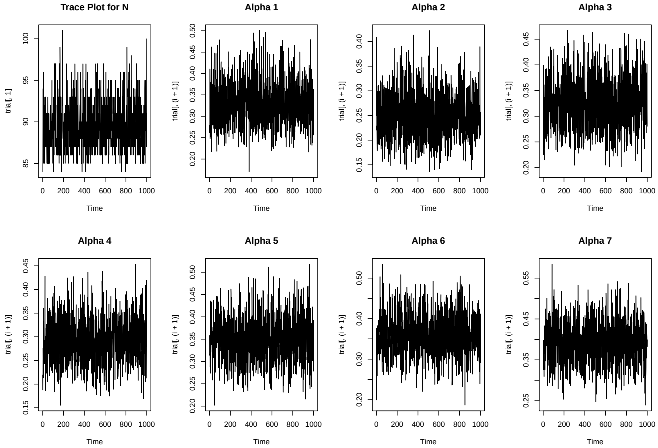
Example: Capture-recapture

Then we consider some preliminary simulations to ensure the chain is mixing well

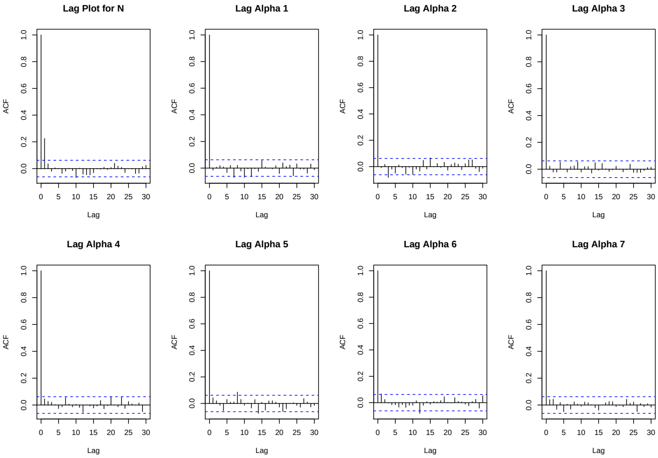
```
trial <- gibbs.chain(1000)
plot.ts(trial[,1], main = "Trace Plot for N")
for(i in 1:7){
    plot.ts(trial[,(i+1)], main = paste("Alpha", i))
    }

acf(trial[,1], main = "Lag Plot for N")
for(i in 1:7){
    acf(trial[,(i+1)], main = paste("Lag Alpha", i))
    }
```

Example: Capture-recapture



Example: Capture-recapture



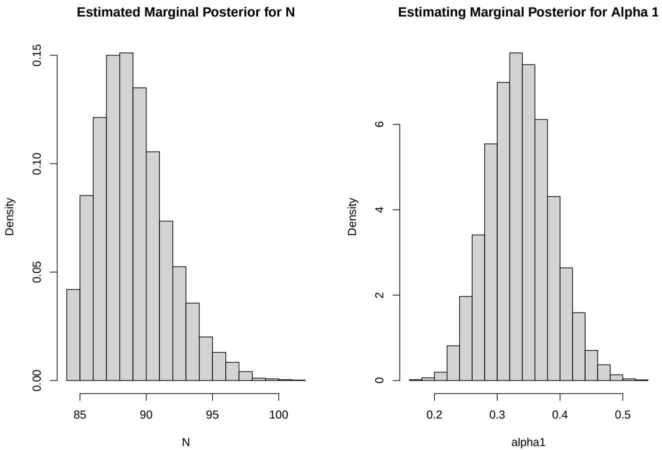
Example: Capture-recapture

Now for a more complete simulation to estimate posterior means and a 90% Bayesian credible region

```
sim <- gibbs.chain(10000)
N <- sim[,1]
alpha1 <- sim[,2]
```

Example: Capture-recapture

```r
par(mfrow=c(1,2))
hist(N, freq=F, main="Estimated Marginal Posterior for N")
hist(alpha1, freq=F, main ="Estimating Marginal Posterior for Alpha 1")
```




```r
par(mfrow=c(1,1))
```

Example: Capture-recapture

```r
library(mcmcse)
```

```
## Warning: package 'mcmcse' was built under R version 4.0.2

## mcmcse: Monte Carlo Standard Errors for MCMC
## Version 1.4-1 created on 2020-01-29.
## copyright (c) 2012, James M. Flegal, University of California, Riverside
##                     John Hughes, University of Colorado, Denver
##                     Dootika Vats, University of Warwick
##                     Ning Dai, University of Minnesota
##  For citation information, type citation("mcmcse").
##  Type help("mcmcse-package") to get started.
```

```r
ess(N)
```

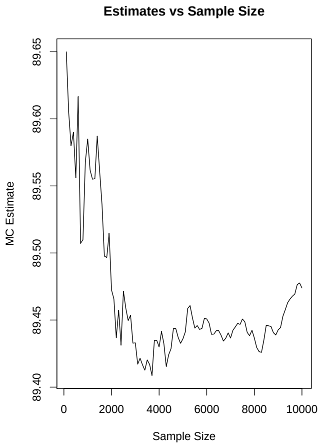
```
## [1] 6049.838
```

```r
ess(alpha1)
```

```
## [1] 9058.551
```

Example: Capture-recapture

```
par(mfrow=c(1,2))
estvssamp(N)
estvssamp(alpha1)
```




```
par(mfrow=c(1,1))
```

Example: Capture-recapture

```r
mcse(N)
```

```
## $est
## [1] 89.4739
##
## $se
## [1] 0.03496931
```

```r
mcse.q(N, .05)
```

```
## $est
## [1] 86
##
## $se
## [1] 0.03786093
```

```r
mcse.q(N, .95)
```

```
## $est
## [1] 94
##
## $se
## [1] 0.05379948
```

Example: Capture-recapture

```r
mcse(alpha1)
```

```
## $est
## [1] 0.3369423
##
## $se
## [1] 0.0005334134
```

```r
mcse.q(alpha1, .05)
```

```
## $est
## [1] 0.2550433
##
## $se
## [1] 0.00104159
```

```r
mcse.q(alpha1, .95)
```

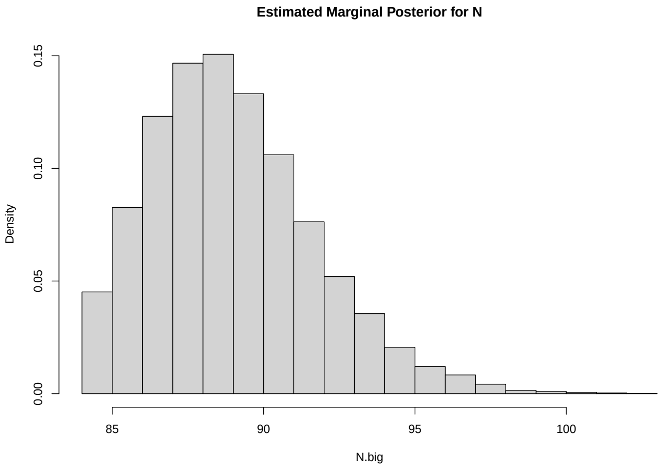
```
## $est
## [1] 0.4241185
##
## $se
## [1] 0.001152417
```

Example: Capture-recapture

```
current <- sim[10000,] # start from here is you need more simulations
sim <- rbind(sim, gibbs.chain(10000, N.start = current[1], alpha.start = current[2:8]))
N.big <- sim[,1]
```

Example: Capture-recapture

```r
hist(N.big, freq=F, main="Estimated Marginal Posterior for N")
```



Example: Capture-recapture

```r
ess(N)
```

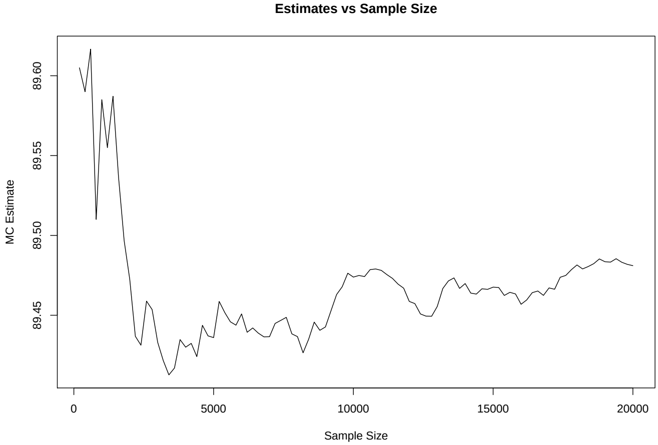
```
## [1] 6049.838
```

```r
ess(N.big)
```

```
## [1] 12326.6
```

Example: Capture-recapture

```r
estvssamp(N.big)
```



Example: Capture-recapture

```r
mcse(N)
```

```
## $est
## [1] 89.4739
##
## $se
## [1] 0.03496931
```

```r
mcse(N.big)
```

```
## $est
## [1] 89.48105
##
## $se
## [1] 0.02475476
```

Example: Capture-recapture

```r
mcse.q(N, .05)
```

```
## $est
## [1] 86
##
## $se
## [1] 0.03786093
```

```r
mcse.q(N.big, .05)
```

```
## $est
## [1] 86
##
## $se
## [1] 0.02952178
```

Example: Capture-recapture

```r
mcse.q(N, .95)
```

```
## $est
## [1] 94
##
## $se
## [1] 0.05379948
```

```r
mcse.q(N.big, .95)
```

```
## $est
## [1] 94
##
## $se
## [1] 0.04061
```

Convergence diagnostics

- A more popular method of MCMC output analysis
- There are many; Gelman and Rubin diagnostic, Geweke's diagnostic, Heidel diagnostic, Raferty diagnostic, and other in `coda` package
- Useful to detect problem with a sampler, but offer ***no*** guarantee you have converged
- Think of these like hypothesis testing

Gelman and Rubin diagnostic

Gelman and Rubin Diagnostic is also used a stopping criteria

- Most popular method for stopping the simulation, one of many convergence diagnostics
- Simulates m independent parallel Markov chains
- Considers a ratio of two different estimates of $\text{Var}_\pi g$, not σ_g^2 from the CLT
- Argue the simulation should continue until the diagnostic ($\widehat{R}_{0.975}$) is close to 1

Toy example

- Let $Y_1, \ldots, Y_m$ be i.i.d. $\mathrm{N}(\mu, \lambda)$ and let the prior for (μ, λ) be proportional to $1/\sqrt{\lambda}$
- The posterior density is characterized by

$$\pi(\mu, \lambda | y) \propto \lambda^{-\frac{m+1}{2}} \exp\left\{-\frac{1}{2\lambda}\sum_{j=1}^{m}(y_j - \mu)^2\right\}$$

which is proper as long as $m \geq 3$

Toy example

- A Gibbs sampler requires the full conditionals

$$\mu|\lambda, y \sim \mathrm{N}(\bar{y}, \lambda/m)$$

and

$$\lambda|\mu, y \sim \mathrm{IG}\left(\frac{m-1}{2}, \frac{s^2 + m(\bar{y}-\mu)^2}{2}\right) ,$$

where $\bar{y}$ is the sample mean and $s^2 = \sum(y_i - \bar{y})^2$

Toy example

- Consider the Gibbs sampler that updates λ then μ

$$(\lambda', \mu') \to (\lambda, \mu') \to (\lambda, \mu)$$

- This sampler is geometrically ergodic
- Suppose $m = 11$, $\bar{y} = 1$, and $s^2 = 14$
- Then $E(\mu|y) = 1$ and $E(\lambda|y) = 2$
- Consider estimating $E(\mu|y)$ and $E(\lambda|y)$ with $\bar{\mu}_n$ and $\bar{\lambda}_n$

Toy example

Stopped the simulation when

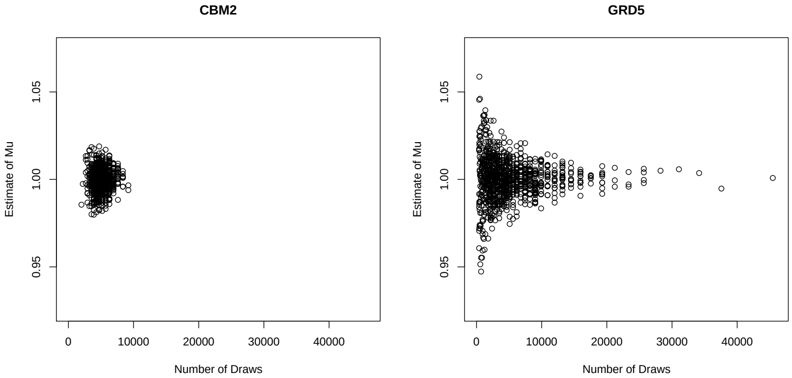
$$BM: \quad t_{.975,(a-1)} \frac{\hat{\sigma}_{BM}}{\sqrt{n}} + I(n < 400) < 0.04$$

$$GRD: \qquad \hat{R}_{0.975} + I(n < 400) < 1.005$$

- 1000 independent replications + Starting from $\bar{y}$ for BM + Starting from draws from π for GRD - Used 4 chains for GRD

Toy example

Plots of $\bar{\mu}_n$ vs. n for both stopping methods

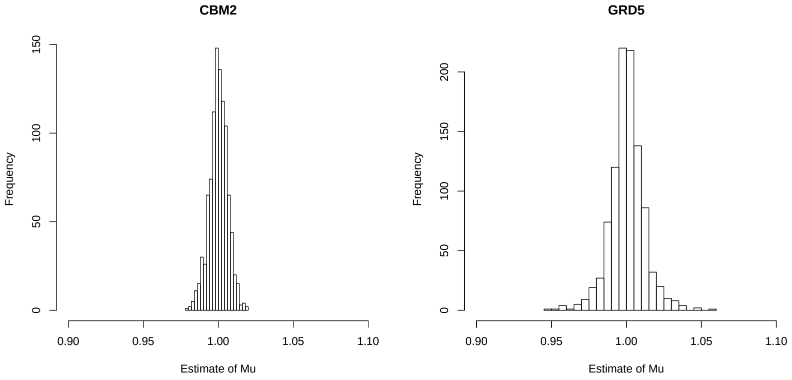


Toy example

MSE	BM	GRD
MSE for $E(\mu\|y)$	3.73e-05 (1.8e-06)	0.000134 (9.2e-06)
MSE for $E(\lambda\|y)$	0.000393 (1.8e-05)	0.00165 (0.00012)

Toy example

Histograms of $\bar{\mu}_n$ for both stopping methods.



Summary

- Bayesian inference usually requires a MCMC simulation
- Metropolis-Hastings algorithm and Gibbs samplers
- Basic idea is similar to OMC, but sampling from a Markov chain yields **dependent** draws
- MCMC output analysis is often ignored or poorly understood