
MCMC convergencia

Modelos Bayesianos con aplicaciones ecológicas

Dr. Cole Monnahan

University of Concepción, Chile

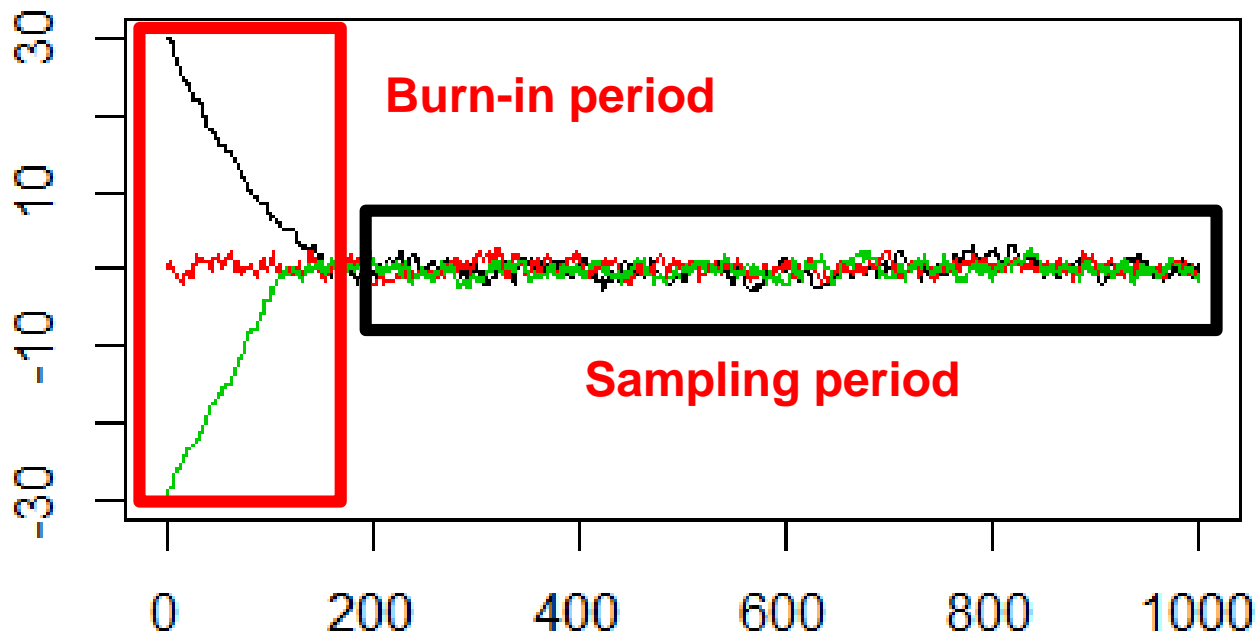
Enero, 2019

Convergencia de MCMC

- MCMC muestras converge como n va a infinito
- Pero solo tenemos una cantidad finita (e.g., 5000).
- Significa que hay que ser cuidadoso que la cadena genera muestras de la posterior
- Es decir, que la cadena (probablemente) ha convergido.

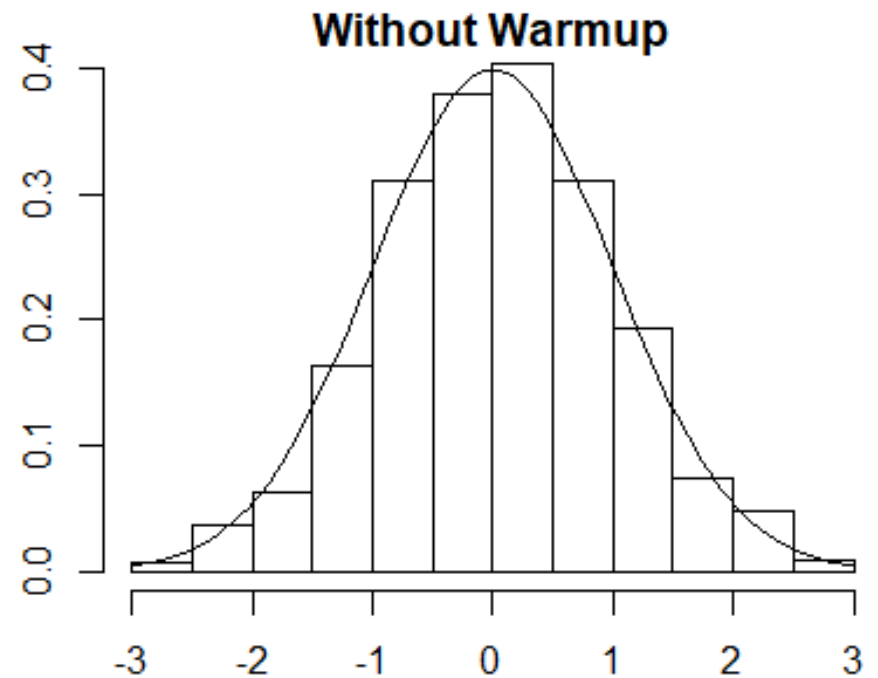
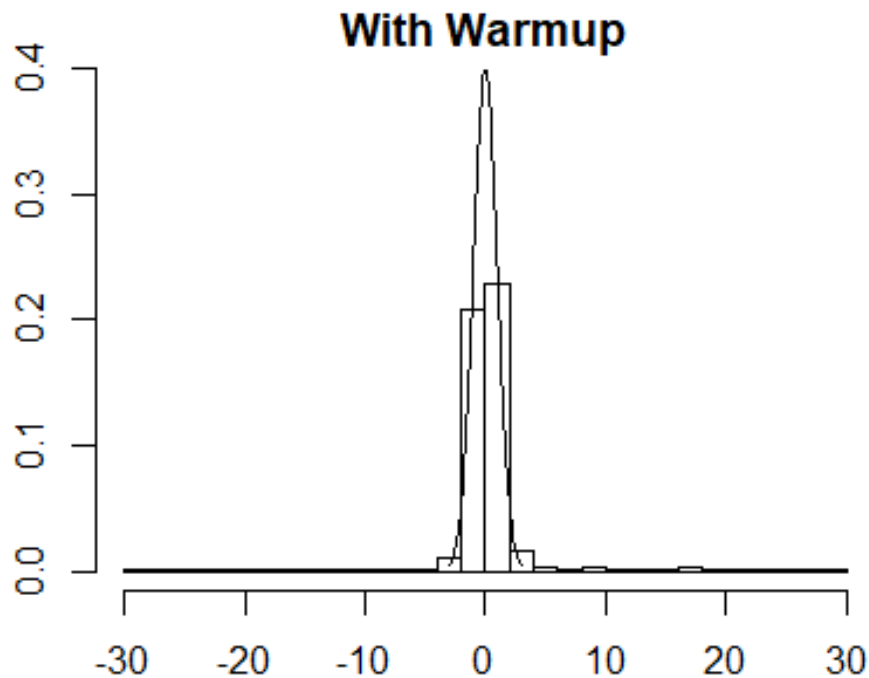
Periodo burn-in o warmup

- Qué pasa si usamos inicializaciones muy amplias?



Periodo burn-in o warmup

- Podemos usar las muestras del periodo 'burn-in'?



Periodo burn-in o warmup

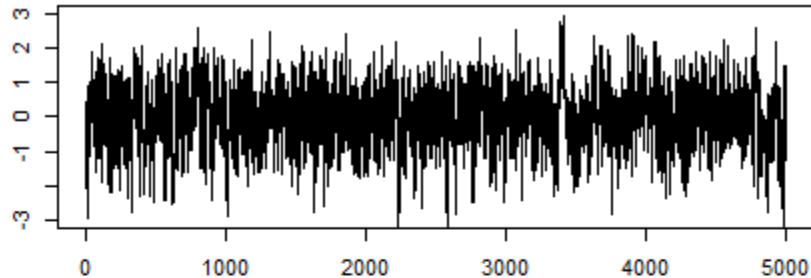
- MCMC convergencia es cuando las cadenas son indistinguible
- Tenemos que quitar las muestras antes de convergencia, porque no son de la distribución a posteriori.
- Cómo conocimos si las cadenas no han convergido?
- Chequeamos con estadísticas como “Rhat” y otras (más tarde)

La calibración (tuning)

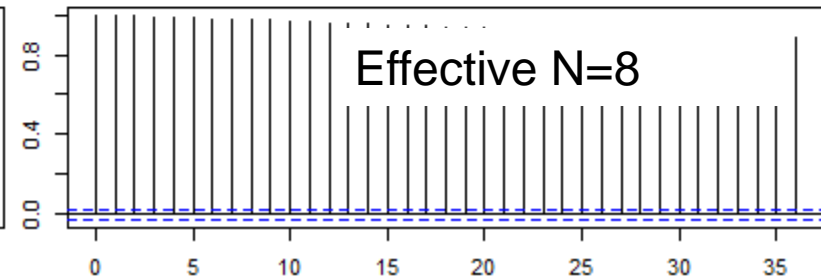
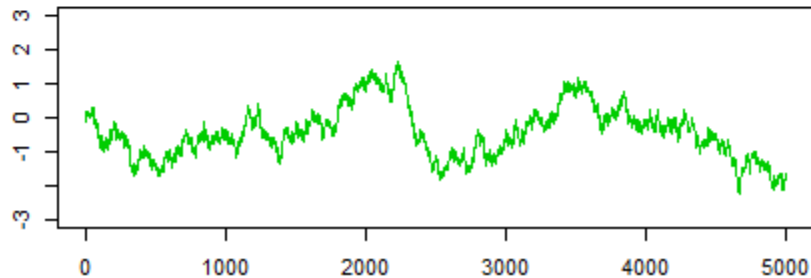
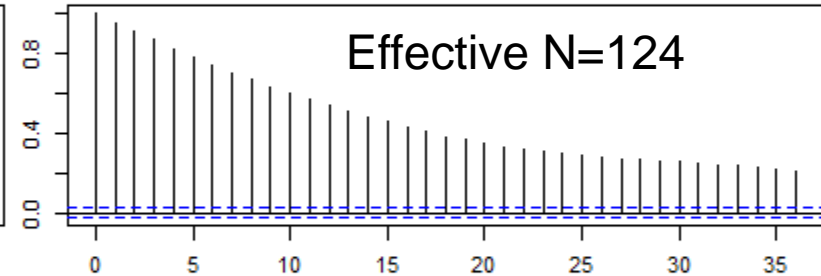
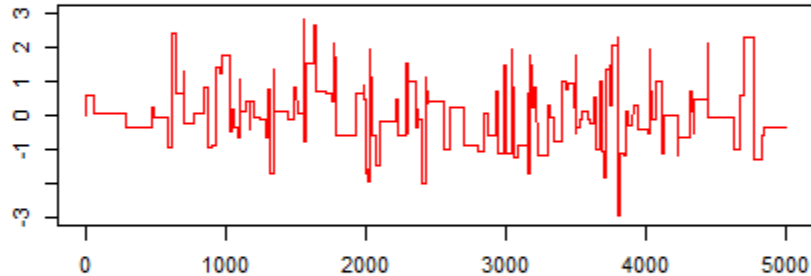
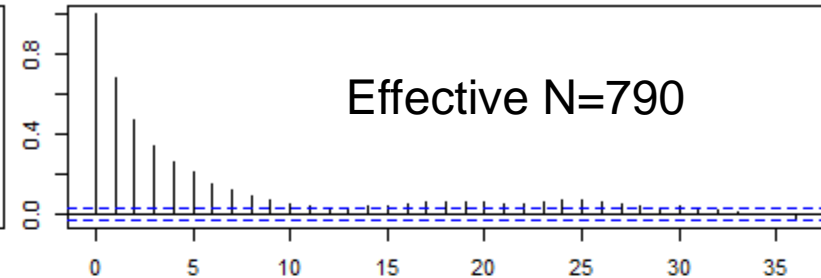
- Tuvimos el variable “U” en el algoritmo:
`new <- x[i-1]+runif(1, -U, U)`
- Cuál es el efecto con diferente valores?
- [Mostrar en R]
- Durante el periodo burn-in, el algoritmo se afina
- Por eso debemos usar un periodo suficientemente largo para afinar

La afinación (tuning)

Trace



Autocorrelation



Thinning

- Muestras efectivas (n_{eff} /ESS o **effective samples**) son menos que N , y se disminuye como autocorrelación se aumenta.
- Si una cadena tiene mucha autocorrelación, tenemos que generar muchas muestras para tener suficiente muestras efectivas.
- En algunos casos, quitamos algunas muestras para no tener demasiado
- Pero perdimos información siempre cuando hacemos “thinning”

Diagnostic checks w/ CODA

- Si usas muestras sin convergencia, vas a obtener la respuesta incorrecta!
- Es **tu responsabilidad** asegurar que no hay evidencia que las cadenas no han convergido.
- En este curso usaremos: Rhat y ESS
- Se recomienda usar inicializaciones muy amplias

Geweke (1992)

- Geweke (1992) statistic – t test equivalent to assess the means of first (T_A) and last (T_B) parts of the Markov chain

$$\delta_A = \frac{1}{T_A} \sum_{t=1}^{T_A} h(x^{(t)}), \quad \delta_B = \frac{1}{T_B} \sum_{t=T-T_B+1}^T h(x^{(t)})$$

$$\sqrt{T}(\delta_A - \delta_B) / \sqrt{\frac{\sigma_A^2}{\tau_A} + \frac{\sigma_B^2}{\tau_B}} \quad , \tau_i = T_i / T, \quad \tau_i + \tau_i < 1$$

- Useful for a single chain or multiple chains
- Test statistic is standard Z score

Gelman-Rubin

Gelman and Rubin (1992) statistic (multiple chains): comparison of between-chain and within-chain variance, probably most commonly used diagnostic. Let $\xi_m^{(t)}$ be a sample from chain m in iteration t

$$B_T = \frac{1}{M-1} \sum_{m=1}^M (\bar{\xi}_m - \bar{\xi})^2, \text{ where } \bar{\xi}_m = \frac{1}{T} \sum_{t=1}^T \xi_m^{(t)}, \quad \bar{\xi} = \frac{1}{M} \sum_{m=1}^M \bar{\xi}_m$$

$$W_T = \frac{1}{M-1} \sum_{m=1}^M s_m^2 = \frac{1}{M-1} \sum_{m=1}^M \frac{1}{T-1} \sum_{t=1}^T (\xi_m^{(t)} - \bar{\xi}_m)^2$$

$$R_T^2 = \frac{\hat{\sigma}_T^2 + \frac{B_T}{M}}{W_T} \frac{\nu_T + 1}{\nu_T + 3} \quad \text{where} \quad \hat{\sigma}_T^2 = \frac{T-1}{T} W_T + B_T$$

and $\nu_T = 2(\hat{\sigma}_T^2 + \frac{B_T}{M})^2 / W_T$ is the estimated df

The “shrink factor” R - converges to 1 under stationarity, so large values indicate lack of convergence, quantiles calculated via t distribution

Effective sample size

- An MCMC chain has T iterations, but not T independent samples from the chain, thus we want a metric for helping determine the number of independent samples.
- Compute effective sample size as

$$\tau_T = T / \kappa(h)$$

$$\kappa(h) = 1 + 2 \sum_{t=1}^{\infty} \text{corr} \left(h(x^{(0)}), h(x^{(t)}) \right)$$

Use of coda package in R

coda implements many diagnostics:

1. Cumulative distribution plot – `cumuplot()`
2. Cramer-von Mises – `heidel.diag()`
3. Geweke – `geweke.diag()`
4. Gelman-Rubin – `gelman.diag()`, and `gelman.plot()`
5. Effective sample size – `effectiveSize()`

Una introducción del software JAGS

Modelos Bayesianos con aplicaciones ecológicas

Dr. Cole Monnahan

University of Concepción, Chile

Enero, 2018

Software implementing MCMC

JAGS – Just Another Gibbs Sampler

- Sampler
 - Gibbs for conjugate priors
 - Adaptive-Rejection Metropolis for non-conjugate
 - Conditional updating (one parameter at a time)
 - Runs multiple chains at a time
 - Diagnostics via coda R package
-

JAGS syntax- a regression example

■ Model

$$Y_i \sim N(\mu_i, \tau)$$

$$\mu_i = \alpha + \beta x_i$$

■ Priors

$$\alpha \sim N(0, 1 \times 10^{-6})$$

$$\beta \sim N(0, 1 \times 10^{-6})$$

$$\tau \sim \Gamma(0.001, 0.001)$$

JAGS/BUGS code

```
#regression model
```

```
for(i in 1 : N) {  
  Y[i]~dnorm(mu[i], tau)  
  mu[i]<- alpha+beta*x[i]  
}
```

```
#PRIORS
```

```
alpha~ dnorm(0,1.0E-6)  
beta~dnorm(0,1.0E-6)  
tau~dgamma(0.001,0.001)  
sigma<-1/sqrt(tau)
```


JAG syntax- a GLM example

- Model

$$Y_i \sim \text{Poisson}(\lambda)$$

- Priors

$$\lambda \sim \text{Gamma}(1,1)$$

- [Demo in R]

JAGS code (save to insect1.jags)

```
model {  
  ## prior  
  lambda~dgamma(1,1)  
  ## likelihood  
  for(i in 1:N){  
    Y[i]~dpois(lambda)  
  }  
}
```

BUGS syntax- a GLM example

R code

```
library(R2jags)
dat <- dget('datos/insect_data2.txt')
plot(dat$x, dat$Y, xlab = "Female Egg Compliment",
      ylab = "Eggs laid on host", pch = 15)
inits <- function() list(lambda=runif(1,1,15))
fit1 <- jags(data=dat, inits=inits, parameters.to.save='lambda',
             model.file='modelos/insect1.jags', n.chains=3, n.iter=2000,
             n.burnin=500, n.thin=1)
traceplot(fit1)
effectiveSize(fit1)
gelman.diag(as.mcmc(fit1))
post <- data.frame(fit1$BUGSoutput$sims.matrix)
hist(post$lambda)
```

Homework

1. Copy the code from the previous slides and rerun the model. Copy the JAGS code into a separate file (insect1.jags).
2. Fit the normal conjugate example from lecture using JAGS and compare to exercise from R.

```
y <- .5 # observed  
sigma <- 1 # assume known  
mu0 <- -2 # prior mean  
tau0 <- .5 # prior SD
```

[Note: JAGS uses $\text{dnorm}(\mu, \tau)$ parameterization where $\tau^2 = 1/\sigma^2$.

References

- Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2014. Bayesian data analysis. Taylor & Francis.
- Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. CODA: Convergence Diagnostics and Out Analysis for MCMC. R News **6:7-11**.
- Link, W. A. and M. J. Eaton. 2012. On thinning of chains in MCMC. Methods in Ecology and Evolution **3:112-115**.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (with discussion). In Bernardo, J., Berger, J., Dawid, A., and Smith, A., editors, Bayesian Statistics 4, pages 169–193. Oxford University Press, Oxford.
- Gelman, A. and Rubin, D. (1992). Inference from iterative simulation using multiple sequences (with discussion). Statistical Science, 7:457–511.