Monte Carlo methods for hierachical models

Rasmus Waagepetersen
Department of Mathematics
Aalborg University

Purpose of workshop

- Discuss likelihood-based inference for non-linear and non-Gaussian mixed models.
- Review methods for computation of the likelihood function for such models (to provide knowledge of the background technology used in software).
- In particular consider Monte Carlo methods and how to obtain Monte Carlo samples using MCMC.
- Discuss Bayesian inference for non-linear and non-Gaussian mixed models and how it can be implemented in BUGS.

Outline

Lecture I
- Data examples
- Generalized linear mixed models
- Computation of likelihood function

Lecture II
- importance sampling
- maximization of likelihood (derivatives)
- Markov chain Monte Carlo

Lecture III
- Markov chain Monte Carlo using BUGS
- Bayesian inference using BUGS

Lecture IV
- Analysis of sensitivity to specification of priors
- Non-normal random effects
- ?

Data example: cherries

$$Y_{stBb} = \begin{cases} 1 \text{ if bud is fresh} \\ 0 \text{ if bud is dead} \end{cases}$$

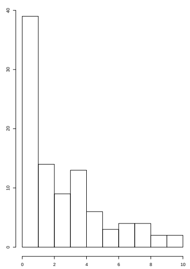
$s = 1, \ldots, 5$ STOCK (variety), $t = 1, \ldots, 4$ TREE, $B = 1, 2, 3$ BRANCH, $b = 1, 2, \ldots$ BUD.

Are some stocks more sensitive to night frost than others ?
Logistic regression but observations on same tree/branch may be correlated.....

Data example: behavioral data for piglets

Y_{zlwr} number of aggressions under shower within 48 minutes for treatment $z = 0, 1$ Z, treatment $l = 0, 1$ L, week $w = 1, \dots, 4$ WEEK, and replicate $r = 1, \dots, 6$ REPL.

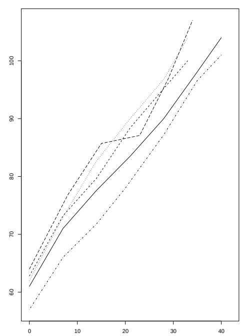
Histogram of observations:



Non-nested random effects REPL REPL*WEEK and REPL*Z*L (=PEN)

Data example: growth curves for pigs

Y_{pt} weight of pig p at time t.



Growth curve with random coeffient ?

Normal linear mixed model (cherries design):

$$U_{st} \sim N_k(0, \tau^2_{\mathrm{TR}}) \quad U_{stB} \sim N(0, \tau^2_{\mathrm{BR}}) \quad \epsilon_{stBb} \sim N(0, \sigma^2) \text{ independent}$$
$$Y_{stBb} = \beta_s + U_{st} + U_{stB} + \epsilon_{stBb}.$$

General formulation:

$$U \sim N\big(0, \Sigma(\tau)\big)$$
$$Y = X\beta + ZU + \epsilon \sim N\big(X\beta, Z\Sigma(\tau)Z^{\mathsf{T}} + \sigma^2 I\big)$$

multivariate normal.

Likelihood function ($\theta = (\beta, \sigma^2, \tau)$):

$$L(\theta) = f(y;\theta) = \int_{\mathbb{R}^k} f(y,u;\theta)\mathrm{d}u = \int_{\mathbb{R}^k} f(y|u;\beta,\sigma^2) f(u;\tau)\mathrm{d}u$$
$$= |\Sigma(\tau) + \sigma^2_\epsilon I|^{-1/2} \exp\big(-(y - X\beta)^{\mathsf{T}}(\Sigma(\tau) + \sigma^2_\epsilon I)^{-1}(y - X\beta)/2\big)$$

Compute maximum likelihood (ML) or restricted maximum likelihood (REML) estimates e.g. using `PROC MIXED`.

Non-normal example: logistic regression with random effects (simple exp. design)

$$U_j \sim N(0, \tau^2),\ j = 1, \dots, n$$
$$Y_{jl}|U_j = u_j \sim \mathrm{Bern}(p_j) = \mathrm{binomial}(1, p_j),\ l = 1, \dots, n_j$$
$$\log(p_j/(1-p_j)) = \eta_j = \beta + U_j$$
$$p_j = \exp(\eta_j)/(1 + \exp(\eta_j))$$

Conditional density:

$$f(y|u;\beta) = \prod_{j,l} p_j^{y_{jl}}(1-p_j)^{1-y_{jl}} = \prod_j \frac{\exp(\beta + u_j)^{y_{j+}}}{(1 + \exp(\beta + u_j))^{n_j}}$$

Likelihood function ($u = (u_1, \dots, u_n)$)

$$\int f(y|u;\beta) f(u;\tau^2)\mathrm{d}u = \prod_j \int_{\mathbb{R}} \frac{\exp(\beta + u_j)^{y_{j+}}}{(1 + \exp(\beta + u_j))^{n_j}} \frac{1}{\sqrt{2\pi\tau^2}} \exp(-u_j^2/(2\tau^2))\mathrm{d}u_j$$

No analytical solution.

Generalized linear mixed models

$$U \sim N_p(0, \Sigma(\tau))$$

Linear predictor

$$\eta = X\beta + ZU$$

Link function

$$g(\mathbb{E}[Y|U = u]) = \eta$$

Conditional distribution parametrised by η and dispersion parameter ϕ

$$Y|U = u \sim f(y|u; \beta, \phi)$$

Examples:

- $Y|U$ Poisson with $g(x) = \log(x)$ and $\phi = 1$
- $Y|U$ Bernouilli with $g(x) = \log(x/(1-x))$ and $\phi = 1$

Likelihood intractable except when $Y|U$ is normal and $g(\eta) = \eta$.

Penalized quasilikelihood

$\theta = (\beta, \tau, \phi)$

PQL estimates $\hat{\theta}$ and $\hat{u}$ maximize joint density

$$f(y, u; \theta) = f(y|u; \beta, \phi) f(u; \tau).$$

PQL estimates less accurate than ML.

Asymptotic results require increasing number of observations for each random effect.

Implemented in SAS macro `glimmix`.

Nonlinear normal mixed models

Example (growth curve for pigs):

$$U_p \sim N(0, \tau^2) \text{ and } \epsilon_{pt} \sim N(0, \sigma^2) \text{ independent}$$
$$m(t; U_p, a, b, c) = \exp\left(a - b\exp(-t(c + U_p))\right)$$
$$Y_{pt} = m(t; U_p, a, b, c) + \epsilon_{pt}$$

Likelihood ($\theta = (a, b, c, \tau^2, \sigma^2)$):

$$L(\theta) = \prod_p \int_{\mathbb{R}} [\prod_t \frac{\exp(-(y_{pt} - m(t|u_i; a, b, c))^2/(2\sigma^2))}{\sqrt{2\pi\sigma^2}}] \frac{\exp(-u_p^2/(2\tau^2))}{\sqrt{2\pi\sigma^2}} \mathrm{d}u_p$$

The computational problem

$$U = (U_1, \ldots, U_n) \sim N_n(0, \Sigma(\tau))$$
$$\eta = X\beta + ZU$$
$$Y_i|U = u \sim f_i(\cdot|u; \beta, \phi) \quad \text{independent}$$

n-dimensional integral:

$$L(\theta) = \int_{\mathbb{R}^n} \Big[\prod_i f_i(y_i|u; \beta, \phi)\Big] f(u_1, \ldots, u_n; \tau) \mathrm{d}u_1 \cdots \mathrm{d}u_n$$

May factorize into lower-dimensional integrals depending on structure of $\Sigma(\tau)$ and Z.

Examples:

Growth curves: one dimensional

$$L(\theta) = \int_{\mathbb{R}^n} \prod_{p=1}^{n} \Big[f(u_p) \prod_{t=1}^{n_p} f(y_{pt}|u_p)\Big] \mathrm{d}u_1 \cdots \mathrm{d}u_n = \prod_{p=1}^{n} \int_{\mathbb{R}} \Big[\prod_{t=1}^{n_p} f(y_{pt}|u_p)\Big] f(u_p)\mathrm{d}u_p$$

Cherries: 4 dimensional

$$L(\theta) = \prod_{s,t} \int_{\mathbb{R}^4} \Big[\prod_{B,b} f(y_{stBb}|u_{stB}, u_{st})\Big]\Big[\prod_{B=1}^{3} f(u_{stB})\Big] f(u_{st})\mathrm{d}u_{st1}\mathrm{d}u_{st2}\mathrm{d}u_{st3}\mathrm{d}u_{st}$$

Behavioral: 9 dimensional

$$L(\theta) = \prod_r \int_{\mathbb{R}^9} \Big[\prod_{zlw} f(y_{zlwr}|u_{zlr}, u_{wr}, u_r)\Big]\Big[\prod_{zl} f(u_{zlr})\mathrm{d}u_{zlr}\Big]\Big[\prod_{w} f(u_{wr})\mathrm{d}u_{wr}\Big] f(u_r)\mathrm{d}u_r$$

One-dimensional case

Compute

$$L(\theta) = \int_{\mathbb{R}} f(y|u;\beta) f(u;\tau)\mathrm{d}u$$

Possibilities:

- Laplace approximation.
- Numerical integration/quadrature (e.g. Gaussian quadrature as in `PROC NLMIXED`) (one level of random effects, dimensions one or two).
- Simple Monte Carlo (small dimensions).
- Importance sampling.

Laplace approximation

Let $g(u) = \log(f(y|u)f(u))$ and choose $\hat{u}$ so $g'(\hat{u}) = 0$ ($\hat{u} = \arg\max g(u)$).

Taylor expansion around $\hat{u}$:

$$g(u) \approx \tilde{g}(u) = g(\hat{u}) + (u-\hat{u})g'(\hat{u}) + \frac{1}{2}(u-\hat{u})^2 g''(\hat{u}) = g(\hat{u}) - \frac{1}{2}(u-\hat{u})^2\big(-g''(\hat{u})\big)$$

I.e. $\exp(\tilde{g}(u))$ proportional to density for $N\big(\hat{u}, -1/g''(\hat{u})\big)$.

$$L(\theta) = \int_{\mathbb{R}} \exp(g(u))\mathrm{d}u \approx \int_{\mathbb{R}} \exp(\tilde{g}(u))\mathrm{d}u$$

$$= \exp(g(\hat{u}))\sqrt{\frac{-2\pi}{g''(\hat{u})}} \int_{\mathbb{R}} \frac{\exp\big(-\frac{1}{2}(u-\hat{u})^2\big(-g''(\hat{u})\big)}{\sqrt{2\pi(-1/g''(\hat{u}))}}\mathrm{d}u = \exp(g(\hat{u}))\sqrt{\frac{-2\pi}{g''(\hat{u})}}$$

Also possible for higher dimensions.

NB: $f(u|y) = f(y|u)f(u)/f(y) \propto \exp(g(u))$ so $U|Y = y \approx N\big(\hat{u}, -1/g''(\hat{u})\big)$.

Simple Monte Carlo:

$$L(\theta) = \int_{\mathbb{R}} f(y|u;\beta) f(u;\tau^2)\mathrm{d}u = \mathbb{E}_{\tau^2} f(y|U;\beta) \approx L_{SMC}(\theta) = \frac{1}{M}\sum_{m=1}^{M} f(y|U^m;\beta)$$

where $U^m \sim N(0,\tau^2)$ independent.

Monte Carlo variance:

$$\mathbb{V}\mathrm{ar}(L_{SMC}(\theta)) = \frac{1}{M}\mathbb{V}\mathrm{ar} f(y|U^1;\beta)$$

Often $\mathbb{V}\mathrm{ar} f(y|U^1;\beta)$ is large so large M is needed.

Estimate $\mathbb{V}\mathrm{ar} f(y|U^1;\beta)$ using empirical variance estimate based on $f(y|U^m;\beta)$, $m = 1,\ldots,M$:

$$\frac{1}{M-1}\sum_{m=1}^{M}\big(f(y|U^m;\beta) - L_{SMC}(\theta)\big)^2$$

Importance sampling

$h(\cdot)$ probability density on $\mathbb{R}$.

$$L(\theta) = \int_{\mathbb{R}} f(y|u;\beta)f(u;\tau^2)\mathrm{d}u = \int_{\mathbb{R}} \frac{f(y|u;\beta)f(u;\tau^2)}{h(u)}h(u)\mathrm{d}u = \mathbb{E}\frac{f(y|V;\beta)f(V;\tau^2)}{h(V)}$$

where $V \sim h(\cdot)$.

$$L(\theta) \approx L_{IS,h}(\theta) = \frac{1}{M}\sum_{m=1}^{M} \frac{f(y|V^m;\beta)f(V^m;\tau^2)}{h(V^m)} \text{ where } V^m \sim h(\cdot), p = 1,\ldots,M$$

Find h so $\mathbb{V}\mathrm{ar}\frac{f(y|V;\beta)f(V;\tau^2)}{h(V)}$ small.

$\mathbb{V}\mathrm{ar}L_{IS,h}(\theta) < \infty$ if $f(y|v;\theta)f(v;\theta)/h(v)$ bounded (i.e. use $h(\cdot)$ with heavy tails).

Possibility: Note

$$f(y|u,\beta)f(u;\tau^2) = \mathrm{const}f(u|y;\theta) \Leftrightarrow \frac{f(y|u,\beta)f(u;\tau^2)}{f(u|y;\theta)} = \mathrm{const}$$

(const=$1/L(\theta)$).

Laplace: $U|Y = y \approx N\big(\hat{u}, -1/g''(\hat{u})\big)$

Use $h(\cdot)$ density for $N\big(\hat{u}, -1/g''(\hat{u})\big)$ or $t_\nu\big(\hat{u}, -1/g''(\hat{u})\big)$-distribution:

$$\frac{f(y|u,\beta)f(u;\tau^2)}{h(u)} \approx \mathrm{const}$$

so small variance.

Simulation straightforward.

Possibility: Consider fixed θ_0:

$$h(u) = f(u|y,\theta_0) = f(y|u;\theta_0)f(u;\theta_0)/L(\theta_0)$$

Then

$$L(\theta) = \int_{\mathbb{R}} \frac{f(y|u;\beta)f(u;\tau^2)}{h(u)}h(u)\mathrm{d}u = L(\theta_0)\int_{\mathbb{R}} \frac{f(y|u;\beta)f(u;\tau^2)}{f(y|u;\beta_0)f(u;\tau_0^2)}f(u|y,\theta_0)\mathrm{d}u$$

$$= L(\theta_0)\mathbb{E}\Big[\frac{f(y|U;\beta)f(U;\tau^2)}{f(y|U;\beta_0)f(U;\tau_0^2)}|Y = y\Big] \Leftrightarrow \frac{L(\theta)}{L(\theta_0)} = \mathbb{E}\Big[\frac{f(y|U;\beta)f(U;\tau^2)}{f(y|U;\beta_0)f(U;\tau_0^2)}|Y = y\Big]$$

So we can estimate ratio $L(\theta)/L(\theta_0)$ where $L(\theta_0)$ is unknown constant.

This suffices for finding MLE:

$$\arg\max_\theta L(\theta) = \arg\max_\theta \frac{L(\theta)}{L(\theta_0)}$$

Maximization of likelihood

Want to solve

$$u(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta}\log L(\theta) = 0$$

Newton-Raphson iteration:

$$\theta^{m+1} = \theta^m + u(\theta^m)i(\theta^m)^{-1}$$

Score and information ($f(y,u;\theta) = f(y|u;\beta)f(u;\tau^2)$):

$$u(\theta) = \mathbb{E}_\theta[\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(y,U;\theta)|Y = y]$$

$$i(\theta) = -\frac{\mathrm{d}}{\mathrm{d}\theta}u(\theta) = u(\theta)^2 - \mathbb{E}_\theta[\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(y,U;\theta) + (\frac{\mathrm{d}}{\mathrm{d}\theta}\log f(y,U;\theta))^2|Y = y]$$

Replace $u(\theta^m)$ and $i(\theta^m)$ by Monte Carlo estimates.

(NB coincides with derivatives of Monte Carlo approximation of likelihood if same importance sampling distribution used).

EM-algorithm:

Suppose θ^m is current parameter value. Obtain θ^{m+1} by maximizing

$$\mathbb{E}_{\theta^m}[\log f(y,U;\theta)|Y=y]$$

MCEM (Monte Carlo EM): use Monte Carlo approximation

$$\mathbb{E}_{\theta^m}[\log f(y,U;\theta)|Y=y] \approx \frac{1}{M}\sum_{m=1}^{M} \log f(y,U^m;\theta)$$

where $U^m \sim f(u|y;\theta^m) = f(y|u;\beta^m)f(u;\tau^{2,m})/L(\theta^m)$.

or importance sampling

$$\mathbb{E}_{\theta^m}[\log f(y,U;\theta)|Y=y] \approx \frac{1}{M}\sum_{m=1}^{M} \log f(y,U^m;\theta)\frac{f(y|U^m;\theta^m)f(U^m;\tau^{2,m})/L(\theta^m)}{h(U^m)}$$

where $U^m \sim h(\cdot)$.

Further use of Monte Carlo: Prediction of U

Consider pair of random variables (X,Y).

Minimum mean square error predictor $\hat{X}$ of X given observation of Y is conditional expectation $\mathbb{E}(X|Y)$.

I.e.

$$\mathbb{E}(X-\hat{X})^2$$

is minimal for $\hat{X} = \mathbb{E}(X|Y)$.

Compute Monte Carlo estimate of $\mathbb{E}(U|Y=y)$ using conditional simulations of $U|Y=y$ or importance sampling.

$$\mathbb{E}(U|Y=y) = \frac{1}{M}\sum_{m=1}^{M} U^m, \quad U^m \sim f(u|y)$$

Conditional simulation of $U|Y=y$

Note

$$f(y|u;\beta_0)f(u;\tau_0^2) \le Bf(u;\tau_0^2)$$

where $B = \sup_u f(y|u;\tau_0^2)$.

Rejection sampling:

1. Generate $V \sim f(\cdot;\tau_0^2)$ and $W \sim \text{Unif}(]0,1[)$
2. Return V if $W \le f(y|V;\tau_0^2)/B$; otherwise go to 1.

Typically small probability for accept (cf. simple Monte Carlo Method).

Extremely small accept probability in high dimensions (curse of dimensionality).

Instead we can use MCMC (later).

Curse of dimensionality for a toy example

Suppose

$$(U_1,\dots,U_n) \sim \text{Unif}([0,1]^n)$$

and

$$Y_i|U_i = u_i \sim \text{Unif}([\theta_i - \epsilon/2, \theta_i + \epsilon/2])$$

Then

$$(U_1,\dots,U_n)|Y_1=y_1,\dots,Y_n=y_n \sim \text{Unif}(\times_{i=1}^n [y_i-\epsilon/2, y_i+\epsilon/2] \cap [0,1]^n)$$

I.e. $(U_1,\dots,U_n)|Y_1=y_1,\dots,Y_n=y_n$ lives on set of volume $< \epsilon^n$.

Very small probability $< \epsilon^n$ that a simulation from $\text{Unif}([0,1]^n)$ falls in $\times_{i=1}^n [y_i-\epsilon/2, y_i+\epsilon/2] \cap [0,1]^n$.

Conditional simulation for high-dimensional U: Markov chain Monte Carlo

Consider

$$U = (U_1, \ldots, U_n) \sim N_n(0, \Sigma)$$

with

$$\mathbb{C}\mathrm{ov}(U_i, U_j) \neq 0 \text{ for all } i, j$$

Then n-dimensional conditional density for $U|Y = y$:

$$f(u_1, \ldots, u_n|y) \propto [\prod_i f(y_i|u_1, \ldots, u_n)] f(u_1, \ldots, u_n)$$

Can not be factorized into lower dimensional densities.

Example: spatial statistics (Christensen and Waagepetersen 2001)

Observations Y_i are weed counts at spatial locations (x_i, y_i) $i = 1, \ldots, 270$

$Y_i|U_i$ is Poisson where U_i random effect associated with (x_i, y_i) (soil properties).

$$\mathbb{C}\mathrm{ov}(U_i, U_j) = \tau^2 \exp(-d_{ij}/\alpha)$$

where d_{ij} is distance between (x_i, y_i) and (x_j, y_j)

Example: quantitative genetics (Sorensen and Waagepetersen 2003)

Y_{ij} size of jth litter of ith pig.

U_i, $\tilde{U}_i$ random genetic effects influencing size and variability of Y_{ij}:

$$Y_{ij}|U_i = u_i, \tilde{U}_i = \tilde{u}_i \sim N\left(\mu + u_i, \exp(\tilde{\mu}_i + \tilde{u}_i)\right)$$

$$(U_1, \ldots, U_n, \tilde{U}_1, \ldots, \tilde{U}_n) \sim N(0, G \otimes A)$$

A: additive genetic relationship matrix (depending on pedigree).

$$G = \begin{bmatrix} \sigma_a^2 & \rho\sigma_a\sigma_{\tilde{a}} \\ \rho\sigma_a\sigma_{\tilde{a}} & \sigma_{\tilde{a}}^2 \end{bmatrix}$$

ρ: coefficient of genetic correlation between U_i and $\tilde{U}_i$.

NB: high dimension $n > 6000$.

MCMC

Suppose $U = (U_1, \ldots, U_n) \sim \pi(\cdot)$ where $\pi(\cdot)$ is a complicated probability distribution.

Markov chain Monte Carlo:

Generate ergodic Markov chain

$$U^1, U^2, U^3, \ldots \qquad (U^m = (U_1^m, \ldots, U_n^m)$$

so that

$$\text{distribution of } U^m \to \pi(\cdot)$$

I.e. for large m, $U^m \sim \pi(\cdot)$ and

$$\mathbb{E}_\pi k(U) \approx \frac{1}{M} \sum_{m=1}^{M} k(U^m)$$

Joint updating Metropolis-Hastings algorithm:

Basic ingredient: proposal density

$$q(v|u), v \in \mathbb{R}^n$$

defined for all $u \in \mathbb{R}^p$ and easy to sample.

Given initial state U^1 generate $U^2, U^3, \ldots$ as follows:

1. Conditional on $U^m = u^m$ generate proposal $V^{m+1} \sim q(\cdot|u^m)$.
2. With probability
$$\min\{1, \frac{\pi(V^{m+1})q(u|V^{m+1})}{\pi(u^m)q(V^{m+1}|u^m)}\}$$
accept $U^{m+1} = V^{m+1}$; otherwise $U^{m+1} = u^m$.

Under mild conditions of irreducibility and aperiodicity this produces an ergodic Markov chain with stationary distribution given by $\pi(\cdot)$.

Ex: random walk Metropolis:

$$V^{m+1} \sim N(u^m, \sigma^2_{\text{prop}})$$

where σ^2_{prop} is the proposal variance.

Then

$$q(v|u) = q(u|v)$$

so Metropolis-Hastings ratio reduces to Metropolis ratio:

$$\min\{1, \frac{\pi(V^{m+1})q(u|V^{m+1})}{\pi(u^m)q(V^{m+1}|u^m)}\} = \min\{1, \frac{\pi(V^{m+1})}{\pi(u^m)}\}$$

Simple example (Exercise 1):

$$\pi(u|y) = \prod_{i=1}^{10} f(y_i|\exp(u+\beta))f(u;\tau^2)/L(\theta)$$

where $f(y_i|\lambda)$ density for Poisson distribution of intensity λ.

Random walk Metropolis ratio (normalising constant $L(\theta) = f(y;\theta)$ cancels out):

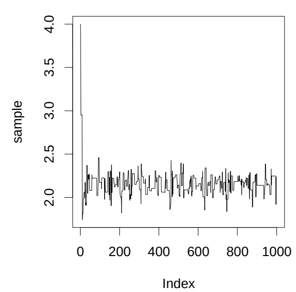
$$\frac{\prod_{i=1}^{10} f(y_i|\exp(V^{m+1}+\beta))f(V^{m+1};\tau^2)/L(\theta)}{\prod_{i=1}^{10} f(y_i|\exp(u^m+\beta)f(u^m;\tau^2)/L(\theta)} = \frac{\prod_{i=1}^{10} f(y_i|\exp(V^{m+1}+\beta))f(V^{m+1};\tau^2)}{\prod_{i=1}^{10} f(y_i|\exp(u^m+\beta)f(u^m;\tau^2)}$$

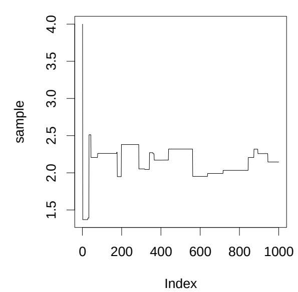
NB: need only know π up to constant of proportionality.

Convergence of Markov chains for simple example

Plots of $U^1, U^2, U^3, \ldots$:

$\sigma^2_{\text{prop}} = 0.2$ (accept rate 24 %) $\quad \sigma^2_{\text{prop}} = 40$ (accept rate 1 %)

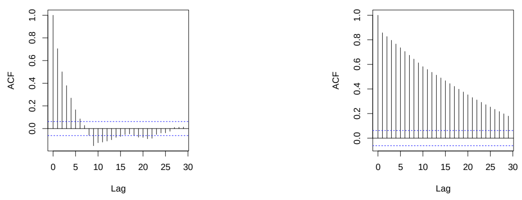




Autocorrelation/mixing:

Plot of autocorrelation $\rho(k) = \mathbb{C}\text{orr}(U^m, U^{m+k})$:

$\sigma^2_{\text{prop}} = 0.2$ (quick mixing) $\quad \sigma^2_{\text{prop}} = 40$ (slow mixing)



Note:

$$\mathbb{V}\text{ar}\frac{1}{M}\sum_{m=1}^{M} U^m = \frac{\mathbb{V}\text{ar}U}{M^2}\sum_{m=1}^{M}\sum_{n=1}^{M}\rho(|m-n|)$$

so small autocorrelation advantageous.

Single-site Metropolis-Hastings

Update one component in each iteration.

Update of ith component:

1. Conditional on $U^m = u^m$ generate $V_i^{m+1} \sim q_i(\cdot|u^m)$ and let

$$V^{m+1} = (u_1^m, \dots, u_{i-1}^m, V_i^{m+1}, u_{i+1}^m, \dots, u_n^m)$$

2. With probability

$$\min\{1, \frac{\pi(V^{m+1})q_i(u_i^m|V^{m+1})}{\pi(u^m)q_i(V_i^{m+1}|u^m)}\}$$

accept $U^{m+1} = V^{m+1}$; otherwise $U^{m+1} = u^m$.

Repeat for $i = 1, \dots, n$

Examples

Random walk Metropolis: $\quad V_i^{m+1} \sim N(u_i^m, \sigma^2_{\text{prop}})$ and

$$\min\{1, \frac{\pi(V^{m+1})q_i(u_i^m|V^{m+1})}{\pi(u^m)q_i(V_i^{m+1}|u^m)}\} = \min\{1, \frac{\pi(V^{m+1})}{\pi(u^m)}\}$$

Gibbs sampler: $\quad V_i^{m+1} \sim U_i|U_j = u_j^m,\ j \neq i.$

Then $q(V_i^{m+1}|u^m) = \pi_i(V_i^{m+1}|u_{-i}^m)$ and

$$\frac{\pi(V^{m+1})q_i(u_i^m|V^{m+1})}{\pi(u^m)q_i(V_i^{m+1}|u^m)} = \frac{\pi_i(V_i^{m+1}|u_{-i}^m)\pi(u_{-i}^m)\pi_i(u_i^m|u_{-i}^m)}{\pi_i(u_i^m|u_{-i}^m)\pi(u_{-i}^m)\pi_i(V_i^{m+1}|u_{-i}^m)} = 1$$

so all proposals are accepted.

No need to choose a proposal variance.

Implementation of MCMC using BUGS (Bayesian analysis using Gibbs sampling).

Model specification in BUGS: hierarchical/directed acyclic graph (DAG).

Example:

$$\tau^2 = 1 \quad \sigma^2 = 0.05$$
$$U|\tau^2, \sigma^2 \sim N(0, \tau^2)$$
$$Y_1, Y_2|U = u, \tau^2, \sigma^2 \sim N(u, \sigma^2), \text{ conditionally independent}$$

DAG:

BUGS code

```
model {
    tau2rec <- 1
    sigma2rec <- 1/0.05
    u ~ dnorm(0.0,tau2rec )
    y1 ~dnorm(u,sigma2rec)
    y2 ~ dnorm(u,sigma2rec)
}
```

Bayesian inference for hierarchical model

Introduce prior $p(\theta)$ for unknown parameters.

Posterior distribution (knowledge obtained by observing data y):

$$p(\theta|y) = \frac{f(y|\theta)p(\theta)}{f(y)} \propto L(\theta)p(\theta)$$

Likelihood $L(\theta)$ given data y is unknown so consider augmented posterior (demarginalize)

$$p(\theta,u|y) \propto f(y|u,\theta)f(u|\theta)p(\theta)$$

Compute posterior expectations, variances etc. using samples from $p(\theta,u|y)$.

Note: if (θ^m, U^m) sample from $p(\theta,u|y)$ then θ^m sample from $p(\theta|y)$.

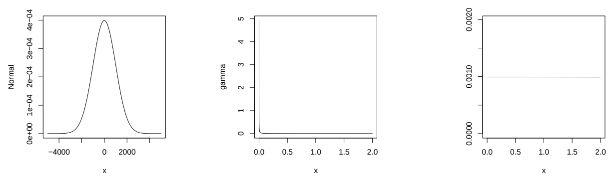
Hence, computation of $L(\theta)$ avoided.

Example: Bayesian analysis for cherries

Priors:

$$1/\tau^2_{\mathrm{TR}}, 1/\tau^2_{\mathrm{BR}} \sim \Gamma(0.001, 0.001) \quad \beta_s \sim N(0, 10^6)$$

$N(0,10^6)$ $\Gamma(0.001,0.001)$ Prior for $\log(\tau^2_{\mathrm{BR}})$



Mean 1 and variance 1000 for $\Gamma(0.001, 0.001)$.

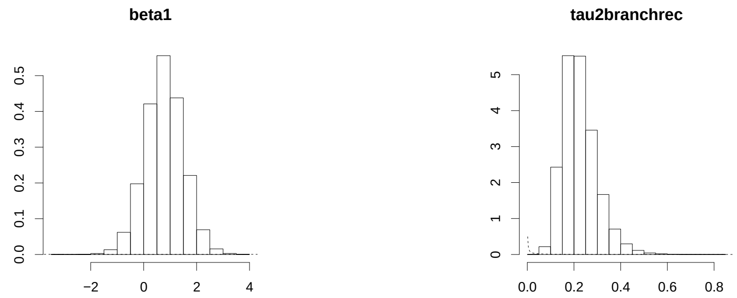
$U_{st} \sim N_k(0,\tau^2_{\mathrm{TR}})$ $U_{stB} \sim N(0,\tau^2_{\mathrm{BR}})$ independent

$Y_{stBb}|U_{st}, U_{stB}, \tau^2_{\mathrm{TR}}, \tau^2_{\mathrm{BR}}, \beta_s \sim \mathrm{Bern}\big(\exp(\beta_s + U_{st} + U_{stB})/(1+\exp(\beta_s+U_{st}+U_{stB})\big)$

Some posterior results for cherries

Priors and posteriors for β_1 and $1/\tau^2_{\mathrm{BR}}$:

β_1 $1/\tau^2_{\mathrm{BR}}$



Posterior mean 0.77 and variance 0.52 Posterior mean 0.23 and variance 0.005

Transformation

Suppose X has density f_X and $Y = g(X)$.

Then

$$f_Y(y) = f_X(g^{-1}(y))/g'(g^{-1}(y))$$

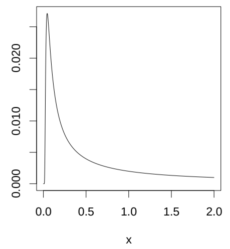
If $1/\tau^2 \sim \Gamma(\alpha,\beta)$ then density of $\omega = -\log(\tau^2)$ is

$$\frac{\beta^\alpha}{\Gamma(\alpha)} \exp(-\omega\alpha - \beta\exp(-\omega)) \approx \text{ const if } \alpha \text{ and } \beta \text{ small}$$

If $1/\tau^2 \sim \Gamma(\alpha,\beta)$ then density of $\tau = (1/\tau^2)^{-1/2}$ is

$$\frac{2\beta^\alpha}{\Gamma(\alpha)}\tau^{-2\alpha-1}\exp(-\beta\tau^{-2})$$

Plot of density for $\tau = (1/\tau^2)^{-1/2}$ when $1/\tau^2 \sim \Gamma(0.001, 0.001)$



Sensitivity to prior specification

Consider two priors $p_1(\theta)$ and $p_2(\theta)$ which yields posteriors

$$p_1(\theta|y) = f(y|\theta)p_1(\theta)/f_1(y)$$

and

$$p_2(\theta|y) = f(y|\theta)p_2(\theta)/f_2(y)$$

Suppose we have sample θ^m from $p_1(\theta|y)$ and we want to compute

$$\mathbb{E}_2[h(\theta)|y] = \int h(\theta)p_2(\theta|y)\mathrm{d}\theta$$

Reuse sample using importance sampling:

$$\mathbb{E}_2[h(\theta)|y] = \int h(\theta)p_2(\theta|y)\mathrm{d}\theta = \int h(\theta)\frac{p_2(\theta|y)}{p_1(\theta|y)}p_1(\theta|y)\mathrm{d}\theta$$
$$= \int h(\theta)\frac{f(y|\theta)p_2(\theta)f_1(y)}{f(y|\theta)p_1(\theta)f_2(y)}p_1(\theta|y)\mathrm{d}\theta = \frac{f_1(y)}{f_2(y)}\int h(\theta)\frac{p_2(\theta)}{p_1(\theta)}p_1(\theta|y)\mathrm{d}\theta$$
$$\approx \left(\frac{f_2(y)}{f_1(y)}\right)^{-1}\frac{1}{M}\sum_{m=1}^{M} h(\theta^m)\frac{p_2(\theta^m)}{p_1(\theta^m)}$$

Moreover:

$$\frac{f_2(y)}{f_1(y)} = \int \frac{f(y|\theta)p_2(\theta)}{f(y|\theta)p_1(\theta)}\frac{f_1(y|\theta)p_1(\theta)}{f_1(\theta)}\mathrm{d}\theta = \int \frac{p_2(\theta)}{p_2(\theta)}p_1(\theta|y)\mathrm{d}\theta \approx \frac{1}{M}\sum_{m=1}^{m}\frac{p_2(\theta^m)}{p_1(\theta^m)}$$

References

Booth, J. G. and Hobert, J. P. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm, *J. R. Statist. Soc.* B **61**: 265–285.

Christensen, O. F. and Waagepetersen, R. (2002). Bayesian prediction of spatial count data using generalized linear mixed models, *Biometrics* **58**: 280–286.

Evans, M. and Swartz, T. (1995). Methods for approximating integrals in statistics with special emphasis on bayesian integration problems, *Statistical Science* **10**: 254–272.

Geyer, C. J. and Thompson, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data, *Journal of the Royal Society of Statistics Series B* **54**: 657–699.

McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models, *J. Am. Statist. Ass.* **92**: 162–170.

Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*, Springer-Verlag, New York.

Sorensen, D. and Waagepetersen, R. (2003). Normal linear models with genetically structured variance heterogeneity: a case study, *Genetical Research* . to appear.