Zero Variance MCMC

Antonietta Mira

joint with D. Bressanini e P. Tenconi

University of Insubria, Varese, Italy

Warwick, EPSRC Symposium, March 2009

MARKOV CHAIN MONTE CARLO

SETTING:

We are interested in evaluating

$$\boxed{\mu = E_\pi f(X)} \qquad X \in \mathcal{X}$$

We know π only up to a normalizing constant

POSSIBLE SOLUTION:

Construct an ergodic Markov chain

$$P(X, A) = \mathrm{Pr}(X_n \in A | X_{n-1} = X) \qquad A \subset \mathcal{X}$$

stationary with respect to π: $\pi P = \pi$

Simulate the Markov chain: $X_0, X_1 \ldots X_n \sim P$
MCMC estimator of μ:

$$\boxed{\widehat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} f(X_i)}$$

HOW GOOD IS THE MCMC ESTIMATE?

Under regularity conditions:

- π is also the unique limiting distribution

$$||P^n(X,\cdot) - \pi(\cdot)|| \to 0, \qquad n \to \infty$$

- LLN and CLT hold (bias of order $1/n$)

thus a measure of efficiency of the MCMC estimator is its ASYMPTOTIC VARIANCE

$$V(f,P) = \lim_{n\to\infty} n \, \mathrm{Var}_\pi[\widehat{\mu}_n]$$

$$= \sigma^2 + 2\sum_{k=1}^{\infty} \rho_k$$

where

$$\sigma^2(f) = \mathrm{Var}_\pi \, f(X)$$

$$\rho_k(f,P) = \mathrm{Cov}_\pi[f(X_0), f(X_k)]$$

We can reduce the asymptotic variance by:

- decreasing $\sigma^2(f)$
  $\longrightarrow$ substituting functions

- decreasing $\rho_k(f, P)$
  $\longrightarrow$ avoiding backtracking
  $\longrightarrow$ delaying rejection
  $\longrightarrow$ inducing negative correlation

Improve relative efficiency by decreasing $\sigma^2(f)$ via function substitution

Instead of estimateing $\mu = E_\pi(f)$ via $\widehat{\mu}_n(f)$ reduce variance by substituting f with $\tilde{f}$ s.t.:

$$\boxed{E_\pi(\tilde{f}\,) = E_\pi(f\,) = \mu}$$

$$\boxed{\sigma^2(\tilde{f}\,) << \sigma^2(f)}$$

Ideally: $\boxed{\tilde{f} = \mu} \implies \boxed{\sigma^2(\tilde{f}\,) = 0 \text{ !!!}}$

General recipe to construct $\tilde{f}$
(Assaraf & Caffarel, 1999, 2000):
use auxiliary operator $H(x,y)$ and function ϕ

H needs to be

- Hermitian (self adjoint)
- $\int H(x,y)\sqrt{\pi(y)}dy = 0$

ϕ needs to be integrable

Define

$$\tilde{f}(x) = f(x) + \frac{\int H(x,y)\phi(y)dy}{\sqrt{\pi(x)}} = f(x) + \Delta f(x)$$

By construction: $\boxed{E_\pi(f\,) = E_\pi(\tilde{f}\,) = \mu}$

$\tilde{f}(x) = f(x) + \Delta f(x)$

$\Delta f(x)$ = control variate

Could generalize:

$\tilde{f}(x) = f(x) + \theta_1 \Delta_1 f(x) + \theta_2 \Delta_2 f(x) + ...$

iid setting: optimal choice of θ_i is available

MCMC setting: hard to find non trivial control variates and to estimate optimal θ_i

The optimal choice for (H, ϕ) can be obtained by imposing

$$\boxed{\sigma(\tilde{f}\,) = 0}$$

or, equivalently

$$\boxed{\tilde{f} = \mu}$$

which leads to the fundamental equation:

$$\int H(x,y)\phi(y)dy = -\sqrt{\pi(x)}[f(x) - \mu_f]$$

hard to solve exactly but can find approximate solutions:

- select an operator H
- parametrize ϕ
- optimally choose the parameters by minimizing $\sigma(\tilde{f}\,)$ over an MCMC simulation
- run a new Markov chain and estimate μ by $\widehat{\mu}(\tilde{f}\,)$ instead of $\widehat{\mu}(f\,)$

Choice of H: Given a reversible kernel P

$$H(x,y) = \sqrt{\frac{\pi(x)}{\pi(y)}}[P(x,y) - \delta(x-y)]$$

and, letting $\tilde{\phi} = \frac{\phi}{\sqrt{\pi}}$ we get:

$$\tilde{f}(x) = f(x) - \int P(x,y)[\tilde{\phi}(x) - \tilde{\phi}(y)]dy$$

This choice it exploited by Dellaportas et al.

- Need closed form expression for conditional expectation of $\tilde{\phi}$ or a rnd scan Gibbs sampler to estimate it
- They argue that $\tilde{\phi}$ should be close to the solution to Poisson equation
- f and Δf should be highly correlated
- They find the optimal θ

General setting: $X \in \Re^d$

$$H = -\frac{1}{2} \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2} + V(x)$$

where

$$V(x) = \frac{1}{2\sqrt{\pi(x)}} \sum_{i=1}^{d} \frac{\partial^2 \sqrt{\pi(x)}}{\partial x_i^2}$$

so that:

$$\tilde{f}(x) = f(x) + \frac{H\phi(x)}{\sqrt{\pi(x)}}$$

The fundamental equation in this setting becomes:

$H\phi(x) = -\sqrt{\pi(x)}[f(x) - \mu_f]$

Choice of ϕ

optimal choice: exact solution of the fundamental equation

sub-optimal choice: parametrize ϕ and choose the parameters to minimize $\sigma(\tilde{f}\ )$

If we parametrize ϕ in terms of a multiplicative constant c and then minimize $\sigma(\tilde{f}\ )$ with respect to c, the optimal choice of c is

$$c = \frac{[E_\pi(f(x)\Delta f(x))]^2}{E_\pi(\Delta f(x))^2}$$

and, for this value of the parameter we obtain

$$\sigma^2(\tilde{f}\ ) = \sigma^2(f\ ) - \frac{[E_\pi(f(x)\Delta f(x))]^2}{E_\pi(\Delta f(x))^2}$$

thus, regardless of the choice of ϕ, a variance reduction in the MCMC estimator is obtained by going from f to $\tilde{f}$

Useful R functions:

- construction of H:
  "fdHess" is used to get the Hessian
  (uses finite differences)

- construction of ϕ:
  "optim" is used to get the parameters
  (uses simulated annealing, quasi-Newton or conjugate gradients methods)

TOY EXAMPLES:
Gaussian and Student-T target distributions to gain insight on functional form of ϕ

Univariate case:

functions of interest:
$f_1(x) = x, \quad f_2(x) = x^2$

Bivariate case:

functions of interest:
$f_1(x) = x_1, \quad f_2(x) = x_1^2, \quad f_3(x) = x_1 x_2$

Length of simulations

first MC (to estimate ϕ parameters): T = 100
second MC (to estimate μ via $\tilde{f}$): n = 150

UNIVARIATE STD GAUSSIAN

$\pi(x) = \exp(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2})$

$f_1(x) = x$

Exact solution to the fundamental equation is available

$\phi_1(x) = (-2\sigma^2 x)\sqrt{\pi(x)}$

$\phi_1(x) = -a(x-c)\exp\{-b(x-c)^2\}$

	f_1	$\tilde{f}_1$		a	b	c
mean	0.030	0.0005	Exact sol.	2.00	0.25	0.00
var	1.022	0.001	Estimated sol.	2.00	0.25	0.01

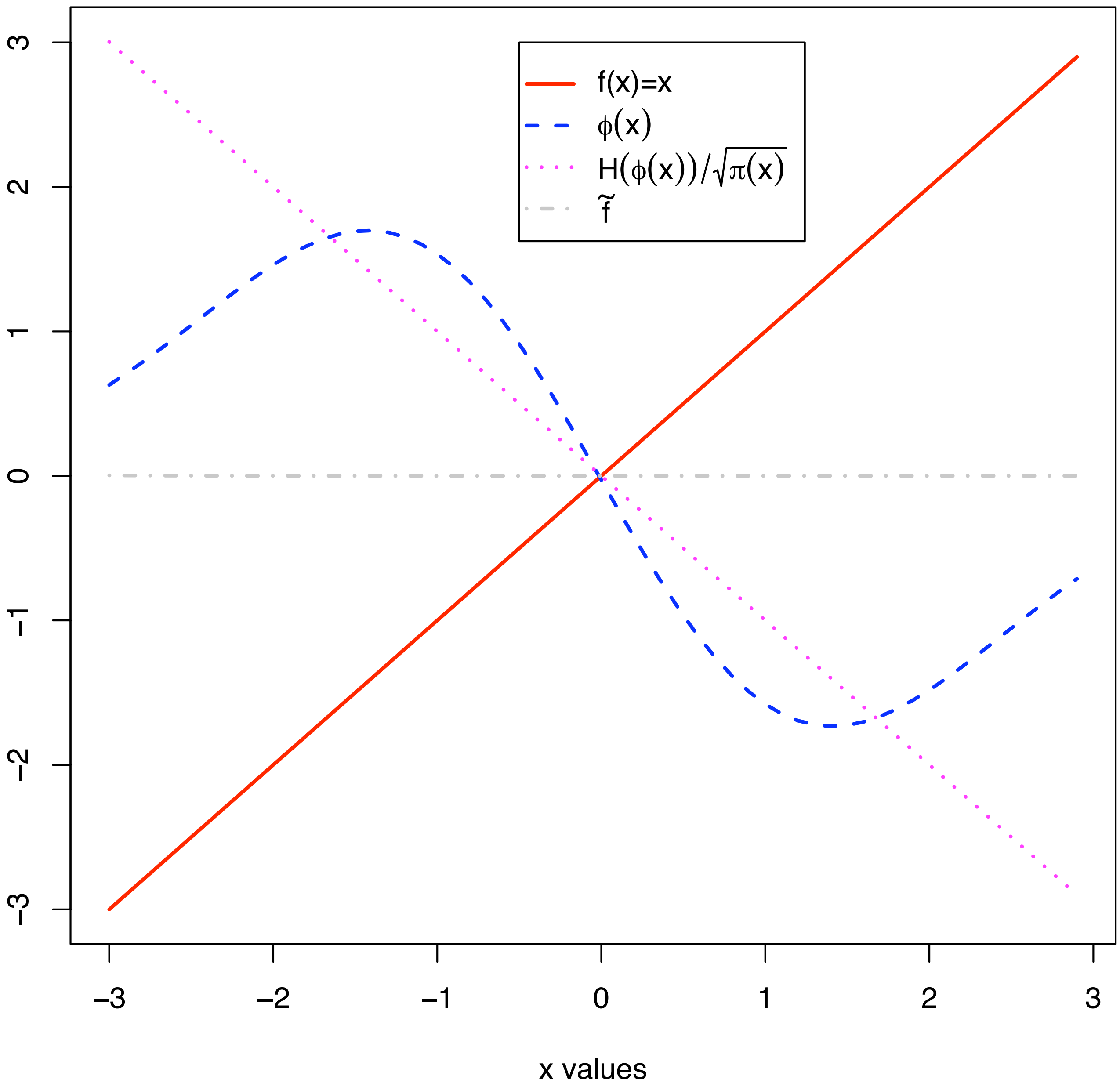
$f_2(x) = x^2$

$\phi_2(x) = (-\sigma^2 x^2 - 2\mu\sigma^2 x)\sqrt{\pi(x)}$

$\phi_2(x) = -a(x-c)^2 \exp\{-b(x-c)^2\}$

	f_2	$\tilde{f}_2$		a	b	c
mean	0.901	1.000	Exact sol.	1.00	0.25	0.00
var	1.387	0.044	Estimated sol.	0.985	0.247	-0.015

Target = N(0,1), $f(x) = x$



$N(\mu = 1, \sigma^2 = 2)$

$f_1(x) = x$

$f_2(x) = x^2$

exact $\tilde{f}_1$ and $\tilde{f}_2$, MC simulation n = 150

	f_1	$\tilde{f}_1$	f_2	$\tilde{f}_2$
$\widehat{\mu}_f$	0.912	1	2.824	3
$\widehat{\sigma}^2_f$	2.013	2.28e-22	9.377	3.53e-21

Univariate Student-T with $g = 5$

$f_1(x) = x$

$f_2(x) = x^2$

exact $\tilde{f}_1$ and $\tilde{f}_2$, MC simulation n = 150

	f_1	$\tilde{f}_1$	f_2	$\tilde{f}_2$
$\widehat{\mu}_f$	-0.271	1.65e-12	1.834	1.666
$\widehat{\sigma}^2_f$	1.778	5.19e-22	20.536	1.32e-23

Robustness of ϕ

For Student-T when $f(x) = x$,

$$\phi_1(x) = \underbrace{\left(\frac{2}{3}\frac{1}{1-g}x^3 + 2\frac{g}{1-g}x\right)}_{P(x)}\sqrt{\pi(x)}.$$

The same structure as in the normal case, but with a higher degree polynomial.

We verified robustness against misspecification of $P(x)$: despite we imposed a first order $P(x)$ we still obtained 93% variance reduction

From MC to MCMC

$$V(f,P) = \left(2\sum_{k=1}^{\infty}\frac{\rho_k}{\sigma^2}+1\right)\sigma^2 = \sigma^2 + 2\sum_{k=1}^{\infty}\rho_k$$

$\Downarrow$

$\tau =$ integrated autocor. time

$\Downarrow$

$\widehat{\tau} =$ Sokal's adaptive truncated correlogram estimate

TARGET: Student-T(5 df)

Same ϕ functions as for the Gaussian case

We used random walk MCMC sampler with different σ_{RW}

We report mean of $\widehat{\tau}$ (variances) over 10 MC simulations

$f_1(x) = x$

$\widehat{\tau}$	$\sigma_{RW} = 0.1$	$\sigma_{RW} = 0.2$	$\sigma_{RW} = 0.5$	$\sigma_{RW} = 1$
f_1	100.16 (33.2)	80.39 (34.1)	45.23 (23.1)	13.32 (7.2)
$\tilde{f}_1$	7.73 (1.8)	3.45 (1.9)	1.48 (0.1)	1.23 (0.2)

$f_2(x) = x^2$

$\widehat{\tau}$	$\sigma_{RW} = 0.1$	$\sigma_{RW} = 0.2$	$\sigma_{RW} = 0.5$	$\sigma_{RW} = 1$
f_2	79.14 (20.8)	63.66 (32.5)	23.84 (11.5)	14.18 (14.5)
$\tilde{f}_2$	1.86 (2.3)	8.17 (2.7)	1.30 (0.36)	2.58 (2.0)

MCMC for $N(\mu = 1, \sigma^2 = 2)$

$f_1(x) = x$

$f_2(x) = x^2$

exact $\tilde{f}_1$ and $\tilde{f}_2$, MCMC simulation n = 150

	f_1	$\tilde{f}_1$	f_2	$\tilde{f}_2$
$\widehat{\mu}_f$	0.080	1	3,193	3
$\widehat{\sigma}^2_f$	2.563	4.8e-20	13.209	1.31e-19

MCMC for univariate Student-T with $g = 5$

$f_1(x) = x$

$f_2(x) = x^2$

exact $\tilde{f}_1$ and $\tilde{f}_2$, MCMC simulation n = 150

	f_1	$\tilde{f}_1$	f_2	$\tilde{f}_2$
$\hat{\mu}_f$	0.095	2.08e-12	1.55	1.666
$\hat{\sigma}_f^2$	1.551	1.08e-22	4.077	6.51e-24

BIVARIATE CASE:

functions of interest:

$$f_1(x) = x_1$$

$$f_2(x) = x_1^2$$

$$f_3(x) = x_1 x_2$$

auxiliary functions:

$$\phi_1(x) = -a(x_1 - c)\,\mathrm{exp}\{-[d(x_1 - c)^2 + b(x_2 - f)^2]\}$$

$$\phi_2(x) = -a(x_1 - c)^2\,\mathrm{exp}\{-[d(x_1 - c)^2 + b(x_2 - f)^2]\}$$

$$\phi_3(x) = -a(x_1 \cdot x_2 - c \cdot f)\,\mathrm{exp}\{-[b(x_1 - c)^2 + d(x_2 - f)^2]\}$$

MCMC for bivariate Normal

$(\mu_1, \mu_2) = (2, 1)$
$(\sigma_1, \sigma_2) = (4, 1)$, $\rho = 0.6$

exact $\tilde{f}_1$, $\tilde{f}_2$, $\tilde{f}_3$, MCMC simulation n = 150

	f_1	$\tilde{f}_1$	f_2	$\tilde{f}_2$	f_3	$\tilde{f}_3$
$\widehat{\mu}_f$	1.683	2.549	5.366	8	2.136	3.2
$\widehat{\sigma}_f^2$	2	2.01e-16	33.937	1.19e-14	7.14	7.11e-17

Bivariate Student-T, $g = 7$

exact $\tilde{f}_1$, $\tilde{f}_2$, $\tilde{f}_3$, MCMC simulation n = 150

	f_1	$\tilde{f}_1$	f_2	$\tilde{f}_2$	f_3	$\tilde{f}_3$
$\widehat{\mu}_f$	-0.09	7.29e-10	1.049	1.4	-0.038	-4,31e-12
$\widehat{\sigma}_f^2$	1.04	1.02e-17	5.44	1.92e-17	1.25	1.95e-21

Other examples considered:

- Simple Bayesian models
- Credit risk models

Note: Rao-Blackwellization can be seen as a special case of this:
replace $f(x^i)$ by a conditional expectation
naturally reduces the variance

Poisson-Gamma model

$$l(y_i|\theta) \sim Po(\theta), \quad i = 1, \cdots, s = 30;$$
$$h(\theta) \sim Ga(\alpha = 4, \beta = 4).$$

We are interested in the first moment of the posterior distribution, in this case we have the exact solution: $\frac{\beta + \sum_{i=1}^{s} y_i}{\alpha + s} = 4.058824$.

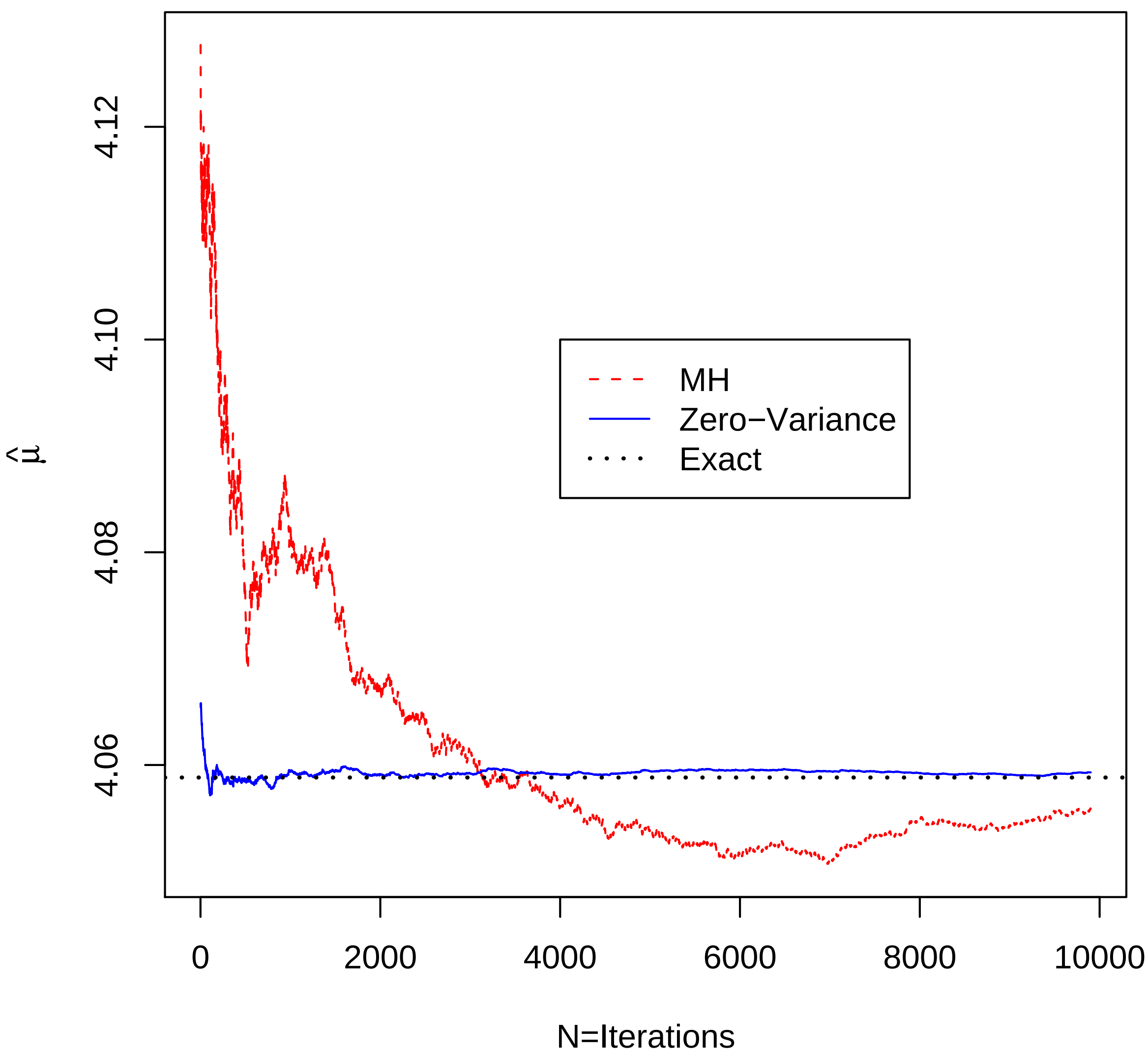
1. run a first MCMC simulation of length 1000 (burn-in of 100);

2. minimize the variance of $\tilde{f}$, obtained using ϕ_1 (case univ. normal)

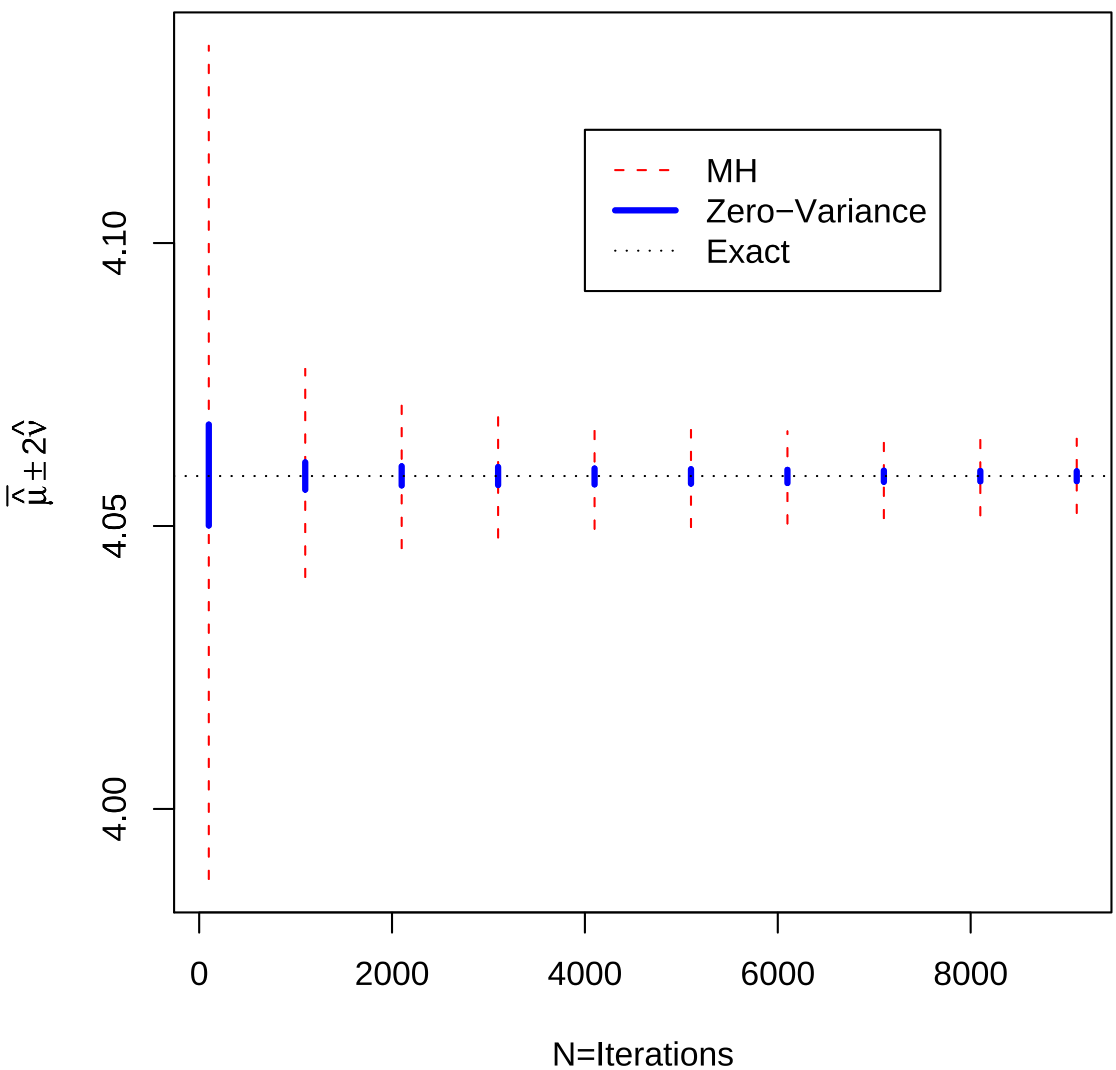
3. run 100 parallel MCMC chains, each of length 10000 (burn-in of 150 steps);

4. compute, on each chain, $\widehat{\mu}_f$, $\widehat{\mu}_{\tilde{f}}$ and the between chain variances, $\widehat{\nu}_f^2$ and $\widehat{\nu}_{\tilde{f}}^2$.

Poisson-Gamma model
single chain



Poisson-Gamma model
parallel chains



Simple credit risk model

We analyze a sample of 124 firms that gave rise to problematic credit and a sample of 200 healthy firms

Bayesian logistic regression model

$$\pi\left(\underline{\beta}|y,x\right) \propto \prod_{i=1}^{s} \theta_i^{y_i}\left(1-\theta_i\right)^{1-y_i} p\left(\underline{\beta}\right),$$

$$\ell\left(y_i|\theta_i\right) \sim Be(\theta_i)\ , \qquad \theta_i = \frac{\exp\left(\underline{x}_i^T\underline{\beta}\right)}{1+\exp\left(\underline{x}_i^T\underline{\beta}\right)}\ , \qquad i = 1,\cdots,s$$

where $\underline{\mathrm{x}}_i$ is a vector of four balance sheet indicators + intercept

We use a non informative improper prior on $\underline{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$

We run an initial M-H of length 300 (after a burn in of 700) and over this initial sample we estimate the optimal parameters of the ϕ function for $f_j(\underline{\beta}) = \beta_j$, $j = 1, \cdots, 5$

$$\phi^j\left(\underline{\beta}\right) = \left(\gamma_1^j\beta_1 + \gamma_2^j\beta_2 + \gamma_3^j\beta_3 + \gamma_4^j\beta_4 + \gamma_5^j\beta_5\right)\sqrt{\pi\left(\underline{\beta}|y,x\right)}$$

Estimated parameters

j	$\widehat{\mu}_{f_j}$	$\widehat{\mu}_{\tilde{f}_j}$	$\widehat{\sigma}^2_{f_j}$	$\widehat{\sigma}^2_{\tilde{f}_j}$	% var-red
1	-1.4761	-1.4339	0.0507	0.0015	97.04
2	-1.0337	-1.0138	0.0664	0.0018	97.28
3	-0.2858	-0.2830	0.0825	0.0043	94.78
4	-0.9687	-0.9746	0.0630	0.0007	98.88
5	0.8279	0.7756	0.0317	0.0012	96.21

With 50 000 iterations only, the zero-variance estimator is close to the 500 000 standard MCMC estimator

So one should run a 100 times longer Markov chain to achieve the same precision

Estimated ϕ parameters

j	$\hat{\gamma}_1^j$	$\hat{\gamma}_2^j$	$\hat{\gamma}_3^j$	$\hat{\gamma}_4^j$	$\hat{\gamma}_5^j$
1	-0.0946	-0.0133	-0.0575	-0.0464	0.0121
2	-0.0151	-0.1582	0.0593	0.0161	0.0551
3	-0.0563	0.0605	-0.1927	0.0147	-0.0355
4	-0.0461	0.0193	0.0141	-0.1011	0.0035
5	0.0106	0.0597	-0.0345	0.0001	-0.0625

This matrix is close to $-2\hat{\Sigma}$ where $\hat{\Sigma}$ is the var-cov matrix of MCMC sampled $\underline{\beta}$ so we argue we can skip the optimization phase of ϕ

Computational issues

When $\phi(\underline{x}) = P(\underline{x})\sqrt{\pi(\underline{x})}$
and $P(x)$ is a polynomial, then

$$(H\phi)(x) = -\frac{1}{2}\sum_{i=1}^{d}\left[\sqrt{\pi(x)}\frac{\partial^2}{\partial x_i^2}P(x) + 2\left(\frac{\partial}{\partial x_i}P(x)\right)\left(\frac{\partial}{\partial x_i}\sqrt{\pi(x)}\right)\right]$$

If $P(x)$ is a first order polynomial, then:

$$\tilde{f}(x) = f(x) - \frac{1}{2}\sum_{i=1}^{d}\left[a_i\left(\frac{\partial}{\partial x_i}\ln\pi(x)\right)\right]$$

the optimal ϕ parameters are close to $-2\hat{\Sigma}$

We can write a **fast computing** version of $\tilde{f}$

$$\tilde{f}_k(\underline{x}) = f_k(\underline{x}) - 2\hat{\Sigma}\times\nabla\ln(\pi(\underline{x}))$$

- No optimization needed
- Only first derivative of target necessary

Extended credit risk model: estimate the default probability of companies that apply to banks for loan

DIFFICULTIES

- default events are rare events
- analysists may have strong prior opinions
- observations are exchangeable within sectors
- different sectors might present similar behaviors relative to risk

THE DATA

7520 companies
1.6 % of which defaulted
7 macro-sectors (identified by experts)
4 performance indicators (derived by experts from balance sheet)

	Dimension	% Default
Sector 1	63	0%
Sector 2	638	1.41%
Sector 3	1343	1.49%
Sector 4	1164	1.63%
Sector 5	1526	1.51%
Sector 6	315	9.52%
Sector 7	2471	0.93%

We used four explanatory variables

- **Variable 1** measures the overall economic performance of the firm

- **Variable 2** is related to the ability of the firm to pick-up external funds

- **Variable 3** is related to the ability of the firm to generate cash flow to finance its short term activities

- **Variable 4** measures the inefficiency in administrating commercial activities

THE MODEL

Bayesian hierarchical logistic regression model

Notation:

- n_j: number of companies belonging to sector $\boxed{j, \;\; j = 1, \cdots, 7}$

- $y(i_j)$: binary response of company i $\boxed{i = 1, \cdots, n_j}$ in sector j. $\boxed{y = 1 \Leftrightarrow \text{default}}$

- $\underline{x}(i_j)$: 4×1 vector of covariates (performance indicators) for company i in sector j

- $\underline{\alpha}$: 7×1 vector of intercepts
  one for each sector

- $\underline{\beta}$: 4×1 vector of slopes
  one for each performance indicator

PARAMETERS of INTEREST: $\underline{\alpha}$ and $\underline{\beta}$

PRIORS:

$$\alpha_j|\mu_\alpha,\sigma_\alpha \sim N_1(\mu_\alpha,\sigma_\alpha^2) \qquad \forall j$$

$$\mu_\alpha \sim N_1(0,64)$$

$$\sigma_\alpha^2 \sim G(25/9,5/9)$$

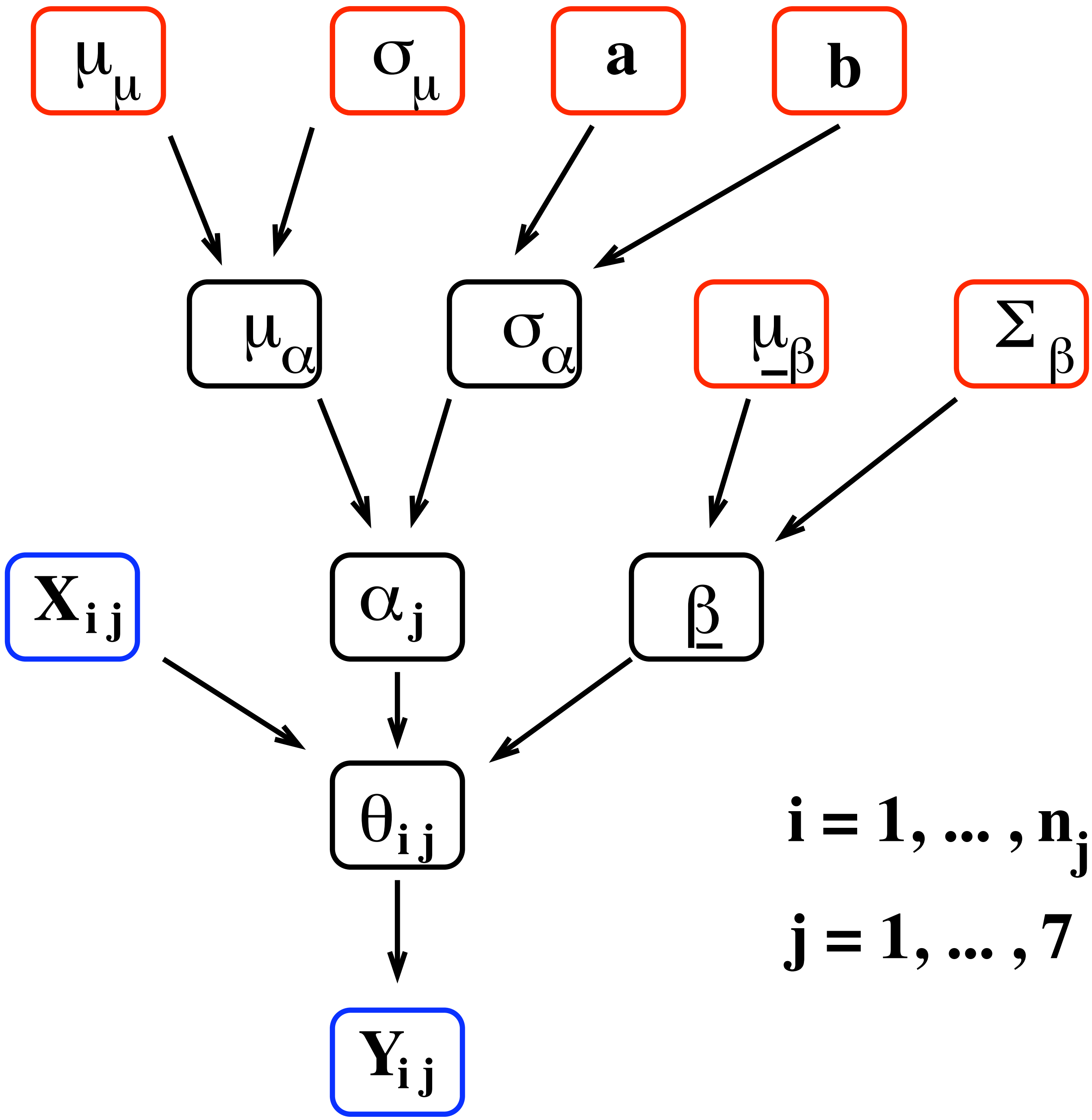
$$\underline{\beta} \sim N_4(\underline{0},64\times I_4)$$

POSTERIOR:

$$\begin{aligned}\pi(\underline{\alpha},\underline{\beta},\mu_\alpha,\sigma_\alpha|y,x) \;\propto\; & \prod_j\prod_i \theta_{ij}^{y(i_j)}(1-\theta_{ij})^{1-y(i_j)} \\ & \prod_j p(\alpha_j|\mu_\alpha,\sigma_\alpha)\; p(\mu_\alpha)p(\sigma_\alpha)\; p(\underline{\beta})\end{aligned}$$

where

$$\theta_{ij} = \frac{\exp[\alpha_j + \underline{x}'(i_j)\underline{\beta}]}{1+\exp[\alpha_j + \underline{x}'(i_j)\underline{\beta}]}$$


μ_μ
σ_μ
$\mathbf{a}$
$\mathbf{b}$
μ_α
σ_α
$\underline{\mu}_\beta$
Σ_β
$\mathbf{X}_{ij}$
α_j
$\underline{\beta}$
θ_{ij}
$\mathbf{Y}_{ij}$
$\mathbf{i = 1, \ldots, n_j}$
$\mathbf{j = 1, \ldots, 7}$

We focus on the functionals

$f_k\left(\underline{\eta}\right) = \eta_k$ where $\underline{\eta} = (\underline{\alpha}, \underline{\beta}, \mu_\alpha, \sigma_\alpha)$

ϕ as in the univariate normal case

1. A Markov chain of lenght 50 000 is run, (burn-in of 10 000) to sample $\pi\left(\underline{\eta}|y,x\right)$;

2. The target var-cov matrix of $\underline{\eta}$, Σ_π, is estimated along the simulated chain. This estimate, $\hat{\Sigma}$, is used to parametrize the ϕ functions to compute $\tilde{f}$ with the "fast version" of our algorithm, i.e.

$$\tilde{f}_k\left(\underline{\eta}\right) = f_k\left(\underline{\eta}\right) - 2\hat{\Sigma} \times \nabla \ln\left(\pi\left(\underline{\eta}|y,x\right)\right);$$

3. We evaluate $\tilde{f}_k\left(\underline{\eta}\right)$ on a second MCMC sample of length 3 000.

η_k	$\widehat{\mu}_{f_k}$	$\widehat{\mu}_{\tilde{f}_k}$	$\widehat{\sigma}^2_{f_k}$	$\widehat{\sigma}^2_{\tilde{f}_k}$	%var.red.
$\eta_1 = \alpha_1$	-6.5122	-6.4548	1.8261	0.7731	57.67
$\eta_2 = \alpha_2$	-5.3699	-6.5122	0.1546	0.0166	89.24
$\eta_3 = \alpha_3$	-5.1055	-5.1296	0.0884	0.0113	87.21
$\eta_4 = \alpha_4$	-4.8881	-4.9179	0.0876	0.0086	90.16
$\eta_5 = \alpha_5$	-5.2247	-5.2446	0.0869	0.0112	87.14
$\eta_6 = \alpha_6$	-3.9072	-3.9560	0.1057	0.0170	83.91
$\eta_7 = \alpha_7$	-6.3274	-6.3539	0.1097	0.0131	88.06
$\eta_8 = \beta_1$	-0.0942	-0.0901	0.0032	0.0005	83.83
$\eta_9 = \beta_2$	-1.2452	-1.2649	0.0999	0.0078	92.23
$\eta_{10} = \beta_3$	-1.4105	-1.4295	0.0415	0.0049	88.26
$\eta_{11} = \beta_4$	0.0870	0.0868	0.0027	0.0002	92.73
$\eta_{12} = \mu_\alpha$	-5.2806	-5.3548	0.3840	0.1114	70.98
$\eta_{13} = \sigma_\alpha$	1.3738	1.4248	0.1883	0.1601	15.00

General form of the ϕ solution of the fundamental equation in termf od π and f:
linear differential equation, not homogeneous with variable coefficients
find the associated Green function
intuition on the structure of ϕ