Part 1

Method for simulating $x_1, \dots, x_n \overset{iid}{\sim} f(x)$

Rejection sampling inefficient for high-dimensional x

Monte Carlo integration (importance sampling), $\widehat{Eh(x)} = \frac{1}{n}\sum_{i=1}^{n} h(x_i)$

Intro to Bayesian inference.

Priors

- Conjugate
- Jeffreys
- Non-informative
- Improper
- Elicitation from expert knowledge

$$\underbrace{f(\theta|x)}_{\text{Posterior}} \propto \underbrace{f(x|\theta)}_{\text{Likelihood}} \underbrace{f(\theta)}_{\text{prior}}$$

Project 1: Rejection sampling from $f(\theta|x)$

Importance sampling estimate of $E(\theta|x)$

Part 2:

Sampling from high-dimensional distributions (Bayes posteriors)

Hierarchical Bayesian models

Ex.: Power-plant failures (George 1993)

y_i : Observed number of failures of plant $i = 1, 2, \dots, n$

t_i : Known operation time of plant i

Model: For $i = 1, 2, \dots, n$

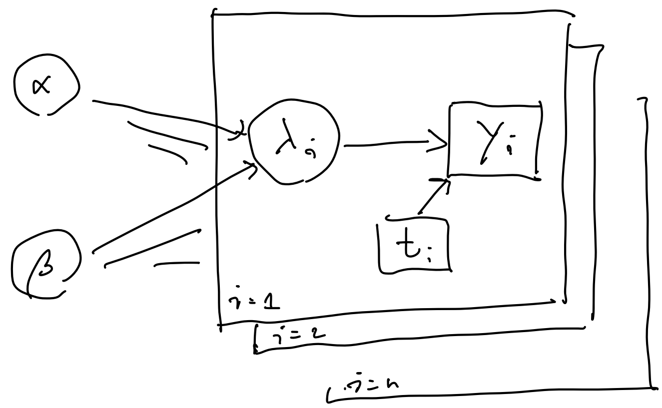
$$y_i | \lambda_i \overset{iid}{\sim} \text{Poisson}(\lambda_i t_i)$$

Prior (hierarchical):

$$\lambda_i | \alpha, \beta \overset{iid}{\sim} \text{Gamma}(\alpha, \beta)$$

$$\left.\begin{array}{l} \alpha \sim \text{Exp}(1.0) \\ \beta \sim \text{Gamma}(0.1, 1.0) \end{array}\right\} \text{Sometimes called a hyper-prior}$$

Directed acyclic graph (DAG):



Squares / circles denotes observed / unobserved quantities (data / parameters)

Joint pmf/pdf of all random variables given by general product rule:

$$f(\alpha, \beta, \lambda_1, \ldots, \lambda_n, y_1, \ldots, y_n)$$

$$= \underbrace{f(\alpha) f(\beta) \prod_{i=1}^{n} f(\lambda_i \mid \alpha, \beta)}_{\text{joint prior of } \alpha, \beta, \lambda_1, \ldots, \lambda_n} \underbrace{\prod_{i=1}^{n} f(y_i \mid \lambda_i)}_{\text{likelihood}}$$

Posterior proportional to joint:

$$f(\alpha, \beta, \lambda_1, \ldots, \lambda_n \mid y_1, \ldots, y_n)$$

$$\propto f(\alpha) f(\beta) \prod_{i=1}^{n} f(\lambda_i \mid \alpha, \beta) \prod_{i=1}^{n} f(y_i \mid \lambda_i)$$

$$\propto e^{-\alpha} \beta^{-0.9} e^{-0.1\beta} \prod_{i=1}^{n} \lambda_i^{\alpha-1} e^{-\beta \lambda_i} \prod_{i=1}^{n} \frac{(\lambda_i t_i)^{y_i} e^{-\lambda_i t_i}}{y_i!}$$

How do we sample from this?

Markov chain Monte Carlo

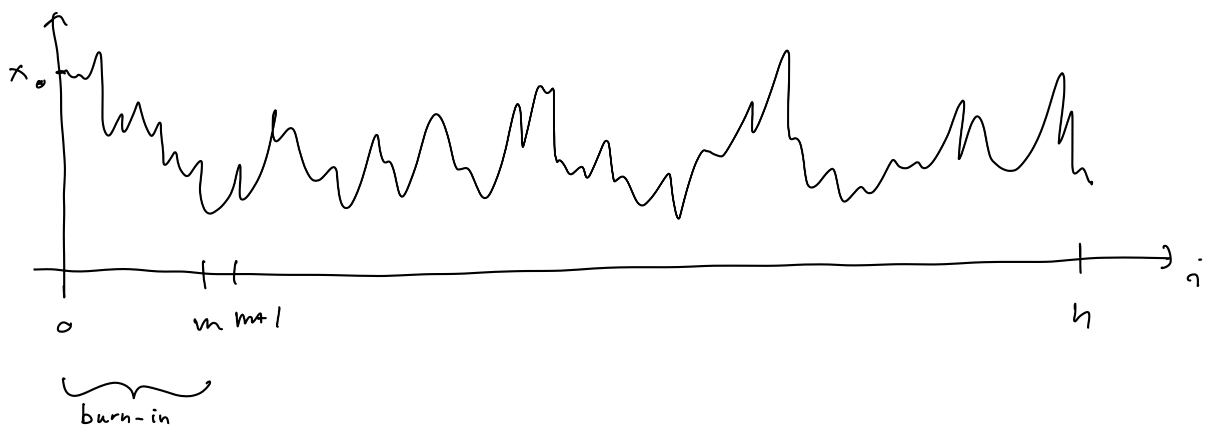
Aim: Generate samples from high-dimensional pmf / pdf $\pi(x)$

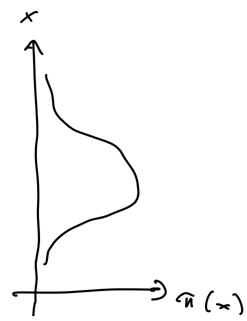
Idea: Construct and simulate Markov chain $\{X_i\}_{i=0}^{\infty}$ having $\pi(x)$ as its stationary distribution.

Estimate $Eh(X)$ (e.g. posterior mean / SDs) by Monte-Carlo integration

e.g.

$$\hat{\mu} = \frac{1}{n} \sum_{i=m+1}^{m+n} h(x_i)$$





Review: Discrete-time, discrete-state Markov chains

Markov chain: Stochastic process $\{X_i\}_{i=0}^{\infty}$ s.t.

$$P(X_{i+1}=x_{i+1} \mid X_i = x_i \cap X_{i-1}=x_{i-1} \cap \ldots)$$

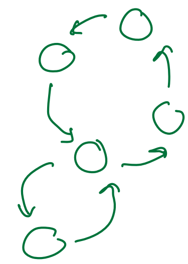
$$= P(X_{i+1}=x_{i+1} \mid X_i = x_i)$$

Stationary transition probs. if $P(X_{i+1}=y \mid X_i = x) = p(y|x)$
("going to y given current state x", "forward transition probs.")

Theorem: An irreducible (one class), aperiodic $(\gcd\{n \in \mathbb{N} : P(X_n = i \mid X_0 = i) > 0\} = 1)$
positive recurrent (finite expected time to return to all states)
has a unique limiting distribution

$$\pi(x) = \lim_{i \to \infty} P(X_i = x \mid X_0 = x_0)$$

Ex: Periodicity = 2



satisfying (from law of total probability)

$$\pi(y) = \sum_{x \in S} p(y|x)\,\pi(x) \quad \text{for all } y \in S$$

$$\sum_{x \in S} \pi(x) = 1$$

Problem statement

TMA4265 Stochastic modelling:

Given $p(y|x)$, find $\pi(x)$

MCMC: The opposite:

Given $\pi(x)$ find $p(y|x)$.

Many solutions! ($n = |S|$ eqs. and $n(n-1)$ unknowns)

The "backward transition probs." are

$$P(X_i = x \mid X_{i+1} = y)$$

$$= \frac{P(X_{i+1} = y \mid X_i = x)\, P(X_i = x)}{P(X_{i+1} = y)} \quad \text{(Bayes theorem)}$$

$$\to \frac{\pi(x)\, p(y|x)}{\pi(y)} \overset{\text{def.}}{=} p^*(x|y)$$

as $i \to \infty$.

Def.: If $p^*(x/y) = p(x/y)$ the Markov chain is time reversible.

Time reversibility thus implies the detailed balance eq.

$$\pi(x)\, p(y|x) = \pi(y)\, p(x|y) \quad \text{for all } x, y \in S$$

We will restrict solutions $p(y|x)$ to those of time reversible chains.

Metropolis-Hastings algorithm (discrete state space version)

Tentative forward transition probs.:

$$p(y|x) = \begin{cases} Q(y|x)\,\alpha(y|x) & \text{for } y \neq x \\ 1 - \sum\limits_{y \neq x} Q(y|x)\,\alpha(y|x) & \text{for } y = x \end{cases}$$

where $Q(y|x)$ is a proposal conditional pmf and

$0 < \alpha(y|x) \leq 1$ is the probability of accepting the transition

from x to y.

How can we chose $\alpha(y|x)$?

Detailed balance eq. implies

$$Q(y|x)\,\alpha(y|x)\,\pi(x) = Q(x|y)\,\alpha(x|y)\,\pi(y)$$

$$\frac{\alpha(y|x)}{\alpha(x|y)} = \frac{Q(x|y)\,\pi(y)}{Q(y|x)\,\pi(x)} \qquad (*)$$

Let

$$\alpha(y|x) = \min\left(1, \frac{Q(x|y)\,\pi(y)}{Q(y|x)\,\pi(x)}\right).$$

If $Q(x|y)\pi(y) < Q(y|x)\pi(x)$ then

$$\alpha(y|x) = \frac{Q(x|y)\,\pi(y)}{Q(y|x)\,\pi(x)} \quad \text{and} \quad \alpha(x|y) = 1$$

and so $(*)$ is satisfied.

If $Q(x|y)\pi(y) \geq Q(y|x)\pi(x)$ then

$$\alpha(y|x) = 1, \quad \alpha(x|y) = \frac{Q(y|x)\,\pi(x)}{Q(x|y)\,\pi(y)}$$

and $(*)$ is again satisfied.

Alg.: Hastings (1970)

Initialise x_0

for $i = 1, 2, \ldots, n$

Generate $y \sim Q(y | x_{i-1})$

$\alpha \leftarrow \min\left(1, \frac{Q(x_{i-1} | y)\pi(y)}{Q(y | x_{i-1})\pi(x_{i-1})}\right)$

Generate $u \sim unif(0, 1)$

if $(u < \alpha)$

$x_i \leftarrow y$

else

$x_i \leftarrow x_{i-1}$

end for

return $x_1, \ldots, x_n$

Metropolis algorithm (1953)

$Q(y|x) = Q(x|y)$ (symmetric so that $\alpha = \min\left(1, \frac{\pi(y)}{\pi(x)}\right)$

Independance sampler (Tierney, 1994)

$Q(y|x) = Q(y)$

Suitable if we can find a $Q(y) \approx \pi(y)$

Ex. (toy): $X \sim \text{Pois}(10)$, $\pi(x) = \frac{\lambda^x e^{-\lambda}}{x!} \propto \frac{10^x}{x!} I(x \geq 0)$

$$Q(y|x) = \begin{cases} \frac{1}{2} & \text{for } y = x \pm 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\alpha(y|x) = \min\left(1, \frac{Q(x|y)\,\pi(y)}{Q(y|x)\,\pi(x)}\right)$$

$$= \min\left(1, \frac{\frac{1}{2} \cdot 10^y \cdot x!\; I(y \geq 0)}{\frac{1}{2} \cdot 10^x \cdot y!\; I(x \geq 0)}\right)$$

$$= \min\left(1, \left(\frac{10}{\max(x,y)}\right)^{y-x} I(y \geq 0)\right)$$

Ok to propose y's outside support of $\pi(x)$

Metropolis - Hastings

Conditional proposal $Q(y|x)$ given current state x.
chosen by user.
Accept y as next state x with probability

$$\alpha(y|x) := \min\left(1, \frac{\pi(y)\,Q(x|y)}{\pi(x)\,Q(y|x)}\right)$$

otherwise keep current state x.

Toy ex.

$$Q(y|x) = \begin{cases} \frac{1}{2} & y = x \pm 2 \\ 0 & \text{otherwise} \end{cases}$$

would not work

Converges to $\pi(x)$ if chain is

1) Irreducible (one communicating class)
2) Aperiodic: Sufficient (but not necessary) conditions:

$p(x|x) > 0$ for some $x \in S$

$\alpha(y|x) < 1$ for some $x \in S$

3) Recurrent: Return to all states with prob. 1.
Positive recurrent: Expected return (hitting) time finite
Follows from 1) if $|S|$ is finite

Not true if $\pi(x)$ is improper (chain may behave as random walk)

Guaranteed if $\pi(x)$ is proper ?

Continuous state space

Proposal $Q(y|x)$ is now a probability density function.
The transition kernel becomes continuous and discrete
and can be represented by the generalized pdf

$$p(y|x) = Q(y|x)\,\alpha(y|x) + \delta(y-x)\left[1 - \int_{-\infty}^{\infty} Q(y|x)\,\alpha(y|x)\,dy\right]$$

where $\delta(\,)$ is the delta Dirac function whose integral $\int_A \delta(x)\,dx = 1$
for sets $A \ni 0$.

Rate of convergence, tuning of $Q(y|x)$

Random walk proposal e.g. $y \sim N(x_{i-1}, \sigma^2)$

$$\alpha = \min\left(1, \frac{\pi(y)}{\pi(x_{i-1})}\right)$$

σ large: large steps but high rejection rate

σ small: high rejection rate but small steps

$\boxed{\text{R demo}}$

Average α around 20% – 50% typically best

Independence sampler

$Q(y|x) = Q(y)$ leads to

Chain gets "stuck" if $\frac{Q(x)}{\pi(x)}$ is small

$$\alpha(y|x) = \min\left(1, \frac{\pi(y)\, Q(x)}{\pi(x)\, Q(y)}\right)$$

Optimal $Q(y) = \pi(y)$ for which $\alpha(y|x) = 1$.

Ex. Rao 1973, Recombination data

$$L(\theta|y) \propto \left(\frac{1}{2} + \frac{\theta}{4}\right)^{y_1} \left(\frac{1-\theta}{4}\right)^{y_2} \left(\frac{1-\theta}{4}\right)^{y_3} \left(\frac{\theta}{4}\right)^{y_4}$$

$\pi(\theta) = 1$ for $0 \le \theta \le 1$

Let $\hat{\theta}$ denote the MLE = Mod$(\theta|y)$ and

$$\widehat{SE}(\hat{\theta}) = H(\hat{\theta})^{-1/2}$$

where is the observed Fisher information $H(\theta) = \frac{d^2 \ln L}{d\theta^2}$ at $\theta = \hat{\theta}$.

$\pi(\theta|y)$ looks Gaussian $\Rightarrow$ $y \sim N(\underset{\mu}{\hat{\theta}}, \underset{\sigma^2}{\widehat{SE}(\hat{\theta})^2})$ good choice.

$\boxed{\text{R demo}}$ 96% acceptance, lag 1 acf = 0.07!

Numerical issues

- Numerical over- and underflow in $\pi(x)$ (likelihood $\times$ prior)

Solution: Work with $\log \pi(x)$ and $\log Q(y|x)$ and compute

$$\alpha <- \min\left(1, \exp\left(\log \pi(y) - \log \pi(x) + \log Q(x|y) - \log Q(y|x)\right)\right)$$

- Proposal y is outside support of $\pi(x)$

Evaluate $\log \pi(y)$ to -Inf

- $\pi(x)$ finite mixture (log sum exp-trick)

$$\ln \pi(x) = \ln \sum_i \underbrace{p_i \pi_i(x)}_{\text{may under/overflow.}}$$

$$= \ln \sum_i a_i$$

$$= \ln \sum_i e^{\overbrace{\ln a_i}^{\text{ok}} - \ln a_{max} + \overbrace{\ln a_{max}}^{\text{ok}}}$$

$$= \ln\left(e^{\ln a_{max}} \sum_i e^{\ln a_i - \ln a_{max}}\right)$$

$$= \underbrace{\ln a_{max}}_{\text{ok}} \cdot \underbrace{\sum_{i=1}^{n} e^{\ln a_i - \ln a_{max}}}_{\text{some but not all terms may underflow so ok.}}$$

Iterative conditioning in MCMC.

Apply MH randomly(*) or cyclically on components x^j of $\underline{x}$, conditioning on remaining components

$$\underline{x}^{-j} = (x^1, x^2, \ldots, x^{j-1}, x^{j+1}, \ldots, x^p)$$

For (*) $Q(\underline{y} \mid \underline{x})$ is a mixture.

Componentwise proposals

$$Q_j(y^j \mid x^j, \underline{x}^{-j}), \quad j = 1, \ldots, p$$

with acceptance

$$\alpha(y^j \mid x^j, \underline{x}^{-j}) = \min\left(1, \frac{\pi(\underline{y})\, Q(\underline{x} \mid \underline{y})}{\pi(\underline{x})\, Q(\underline{y} \mid \underline{x})}\right)$$

We have $\pi(\underline{y}) = \pi(y^j \mid \underline{y}^{-j})\, \pi(\underline{y}^{-j})$

$$= \pi(y^j \mid \underline{x}^{-j})\, \pi(\underline{x}^{-j}) \quad \text{etc.}$$

so that the acceptane prob. simplifies (via cancelling terms) to

$$\min\left(1, \frac{\pi(y^j \mid \underline{x}^{-j})\, Q(x^j \mid y^j, \underline{x}^{-j})}{\pi(x^j \mid \underline{x}^{-j})\, Q(y^j \mid x_j^j, \underline{x}^{-j})}\right)$$

i.e. only factors containg y^j in $\pi(\underline{y})$ etc need to be evalnated.

Gibbs sampling

8.1

$$Q_j(y^j \mid x^j, \underline{x}^{-j}) = \pi(y^j \mid \underline{x}^{-j})$$

"Componentwise independence sampler with optimal choice of Q"

Acceptance prob $= 1$

8.1

Ex. Simple linear regression:

$$Y_i \mid a, b, \tau \sim N(a + b x_i, \tau^{-1})$$

$$a \sim N(0, \tfrac{1}{\tau_a})$$

$$b \sim N(0, \tfrac{1}{\tau_b})$$

$$\tau \sim \text{Gamma}(\alpha, \beta)$$

Prior.
Note, this is not conjugate, see below

Joint posterior

$$\pi(a, b, \tau \mid y) \propto \underbrace{\tau^{\frac{n}{2}} e^{-\frac{1}{2}\tau \sum_{i=1}^{n}(y_i - a - b x_i)^2}}_{f(y \mid a, b, \tau)} \underbrace{\tau_a^{\frac{1}{2}} e^{-\frac{1}{2}\tau_a a^2}}_{\pi(a)} \underbrace{\tau_b^{\frac{1}{2}} e^{-\frac{1}{2}\tau_b b^2}}_{\pi(b)} \underbrace{\tau^{\alpha-1} e^{-\beta\tau}}_{\pi(\tau)}$$

Componentwise conditionals:

$$\pi(\tau \mid a, b, y) \propto \tau^{\alpha + \frac{n}{2} - 1} e^{-\tau\left[\frac{1}{2}\sum(y_i - a - b x_i)^2 + \beta\right]}$$

i.e. $\tau \mid a, b, y \sim \text{Gamma}\left(\alpha + \frac{n}{2}, \beta + \frac{1}{2}\sum(y_i - a - b x_i)^2\right)$

$$(y_i - bx_i - a)^2$$

$$\pi(a \mid b, \tau, y) \propto e^{-\frac{1}{2}\left(\tau_a a^2 + \tau \sum_{i=1}^{n} (y_i - a - bx_i)^2\right)}$$

$$= e^{-\frac{1}{2}\left(\tau_a a^2 + \cancel{\overbrace{\tau \sum (y_i - bx_i)^2}^{\text{const}}} - 2\tau a \sum (y_i - bx_i) + \tau n a^2\right)}$$

$$= e^{-\frac{1}{2}(\tau_a + n\tau)\left(a - \frac{\tau}{\tau_a + n\tau} \sum (y_i - bx_i)\right)^2}$$

i.e. $a \mid b, \tau, y \sim N\left(\frac{\tau}{\tau_a + n\tau} \sum_{i=1}^{n} (y_i - bx_i), (\tau_a + n\tau)^{-1}\right)$

$$\pi(b \mid a, \tau, y) \propto e^{-\frac{1}{2}\left(\tau_b b^2 + \tau \sum (y_i - a - bx_i)^2\right)}$$

$$= e^{-\frac{1}{2}\left(\tau_b b^2 + \cancel{\tau \sum (y_i - a)^2} - 2\tau b \sum (y_i - a)x_i + \tau b^2 \sum x_i^2\right)}$$

$$\propto e^{-\frac{1}{2}\left(\tau_b + \tau \sum x_i^2\right)\left(b - \frac{\tau \sum x_i (y_i - a)}{\tau_b + \tau \sum x_i^2}\right)^2}$$

i.e. $b \mid a, \tau, y \sim N\left(\frac{\tau \sum x_i (y_i - a)}{\tau_b + \tau \sum x_i^2}, \left(\tau_b + \tau \sum x_i^2\right)^{-1}\right)$

$\boxed{\text{R demo}}$

As $\alpha, \beta \to 0$ $\quad \pi(\tau) = \frac{\beta^\alpha}{\Gamma(\alpha)} \tau^{\alpha - 1} e^{-\beta\tau} \xrightarrow{\text{q-vaguely}} \frac{1}{\tau}$ (scale prior)

As $\tau_a \to 0$ $\quad \pi(a) = \sqrt{\frac{\tau_a}{2\pi}} e^{-\frac{1}{2}\tau_a a^2} \xrightarrow{\text{q-vaguely}} 1$ (uniform prior)

sensu Bioche & Druilhet 2017.

With this improper prior credible intervals matches freq. conf. int.

Conjungate normal inverse gamma - prior on $\sigma^2 = \tau^{-1}$ and $\underline{\beta}$:

$$\tau \sim \Gamma(\alpha_0, \lambda_0), \quad \beta | \tau \sim N(\beta_0, \tau^{-1} Q_0^{-1}) \quad \Leftrightarrow \quad \underline{\beta} \sim N\text{-}G(\beta_0, Q_0^{-1}, \alpha_0, \lambda_0)$$

(Each component of $\underline{\beta}$ has a marginal loc-scale t-distribution, $\underline{\beta}$ is multivariate loc-scale student-t)

$$\pi(\tau, \beta | y) \propto \underbrace{\tau^{\frac{n}{2}} e^{-\frac{\tau}{2}(y - X\beta)^T (y - X\beta)}}_{\text{likelihood}} \underbrace{\tau^{\alpha_0 - 1} e^{-\lambda_0 \tau}}_{\pi(\tau)} \underbrace{e^{-\frac{\tau}{2}(\beta - \beta_0)^T Q_0 (\beta - \beta_0)}}_{\pi(\beta | \tau)}$$

$$\beta^T X^T X \beta - y^T X \beta - \beta^T X^T y + y^T y$$
$$+ \beta^T Q_0 \beta - \beta_0^T Q_0 \beta - \beta^T Q_0 \beta_0 + \beta_0^T Q_0 \beta_0$$

$$= (\beta^T - \beta_1)^T Q_1 (\beta - \beta_1) + C$$

$$= \beta^T Q_1 \beta - \beta_1^T Q_1 \beta - \beta^T Q_1 \beta_1 + \beta_1^T Q_1 \beta_1 + C$$

Equating coefficients,

$$Q_1 = X^T X + Q_0$$

$$Q_1 \beta_1 = X^T y + Q_0 \beta_0 \Rightarrow \beta_1 = (X^T X + Q_0)^{-1} (X^T y + Q_0 \beta_0)$$

$$C = y^T y + \beta_0^T Q_0 \beta_0 - \beta_1^T Q_1 \beta_1$$

Thus,

$$\pi(\tau, \beta | y) \propto \pi(\beta | \tau, y) \pi(\tau | y)$$

$$\propto \underbrace{|Q_1|^{\frac{1}{2}} \tau^{\frac{p}{2}} e^{-\frac{\tau}{2}(\beta - \beta_1)^T Q_1 (\beta - \beta_1)^T}}_{\propto \pi(\beta | \tau, y)} \underbrace{\tau^{\overbrace{\alpha_0 - 1 + \frac{n-p}{2}}^{\alpha_1}} e^{-\tau \overbrace{(\lambda_0 + \frac{1}{2} C)}^{\lambda_1}}}_{\propto \pi(\tau | y)}$$

Can easily simulate from this posterior (no need for MCMC).

Ex. cont. Power-plants:

$$\pi(\alpha, \beta, \lambda_1, \dots, \lambda_n | y) = e^{-\alpha} \beta^{-0.9} e^{-1.0\beta} \prod_{i=1}^{n} \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda_i^{\alpha-1} e^{-\beta\lambda_i} \prod_{i=1}^{n} (\lambda_i t_i)^{y_i} e^{-\lambda_i t_i}$$

Componentwise conditionals

$$\pi(\alpha | \beta, \underline{\lambda}, \underline{y}) \propto e^{-(1 - \sum \ln\lambda_i - n\ln\beta)\alpha} \Gamma(\alpha)^{-n}$$

For Metropolis random-walk proposals, e.g.

$$\alpha_{new} | \alpha, \beta, \underline{\lambda} \sim N(\alpha, \underbrace{\sigma_\alpha^2}_{\text{needs tuning}})$$

the acceptance probability is

$$\alpha(\alpha_{new} | \alpha) = \min\left(1, \exp\left(-(1 - \sum \ln\lambda_i - n\ln\beta)(\alpha_{new} - \alpha)\right) \Gamma(\alpha_{new})^{-n} \Gamma(\alpha)^{n}\right)$$

$$\pi(\beta | \alpha, \underline{\lambda}, y) \propto \beta^{-0.9 + n\alpha} e^{-(1.0 + \sum\lambda_i)\beta} \sim \text{Gamma}(0.1 + n\alpha, 1 + \sum\lambda_i)$$

$$\pi(\lambda_i | \alpha, \beta, \underline{\lambda}^{-i}, y) \propto \lambda_i^{\alpha-1} e^{-\beta\lambda_i} \lambda_i^{y_i} e^{-\lambda_i t_i}$$

$$= \lambda_i^{\alpha + y_i - 1} e^{-(\beta + t_i)\lambda_i} \sim \text{Gamma}(\alpha + y_i, \beta + t_i)$$

Sanity check: Compare posterior mean α and β to method of moment estimates (integrating out the λ_i's, each y_i has a neg. bin. distribution)

Ex.: Korsbetningen, Battle of Visby: Archeological excavation.

$x_1 = 256$ left femurs

$x_2 = 237$ right femurs

How many people N buried in total?

Model

$$x_1, x_2 \overset{iid}{\sim} \text{bin}(N, p)$$

Prior

$$N \sim \text{unif}(N_{min}, N_{max})$$

$$p \sim \text{beta}(a, b)$$

$$\pi(N, p | \underline{x}) \propto \underbrace{I(N_{min} \le N \le N_{max})}_{\pi(N)} \underbrace{p^{a-1}(1-p)^{b-1}}_{\pi(p)} \underbrace{\prod_{i=1}^{2} \binom{N}{x_i} p^{x_1+x_2} (1-p)^{2N - x_1 - x_2}}_{L(N, p; x)}$$

$$\propto I(N_{min} \le N \le N_{max})\, p^{a+x_1+x_2-1} (1-p)^{b+2N-x_1-x_2-1} \frac{N!^2}{(N-x_1)!\,(N-x_2)!}$$

Metropolis within Gibbs (hybrid sampler), single site updates

Gibbs step for p:

$$p' | N, \underline{x} \sim \text{beta}(a + x_1 + x_2,\ b + 2N - x_1 - x_2)$$

Metropolis for N:

$$N' | N, p, \underline{x} \sim \text{unif}(N-d, N+d) \quad \text{(random walk)}$$

$$\alpha(N'/N) = \min\left(1, \frac{\pi(N', p | \underline{x})}{\pi(N, p | \underline{x})}\right)$$

$$= \min\left(1, \frac{\frac{N'!^2}{(N'-x_1)!(N'-x_2)!}(1-p)^{2N'}}{\frac{N!^2}{(N-x_1)!(N-x_2)!}(1-p)^{2N}}\right) = \min\left(1, \left(\frac{N'!}{N!}\right)^2 \prod_{i=1}^{2} \frac{(N-x_i)!}{(N'-x_i)!} (1-p)^{2(N'-N)}\right)$$

Metropolis with single (N, p) block update

Joint proposal

$$Q(N', p' \mid N, p, \underline{x}) \propto \underbrace{I(N-d \leq N' \leq N+d)}_{\text{random walk prop. for } N'} \; \underbrace{\frac{p'^{(a+x_1+x_2-1)} (1-p')^{(b+2N'-x_1-x_2-1)}}{B(a+x_1+x_2,\, b+N'-x_1-x_2)}}_{\text{Full conditional of } p \text{ given } N}$$

This leads to

$$\frac{\pi(N', p' \mid \underline{x})\, Q(N, p \mid N', p', \underline{x})}{\pi(N, p \mid \underline{x})\, Q(N', p' \mid N, p, \underline{x})}$$

$$= \frac{I(N_{min} \leq N' \leq N_{max}) \; \dfrac{N'!^2}{(N'-x_1)!\,(N'-x_2)!} \; B(a+x_1+x_2,\, b+N'-x_1-x_2)}{\underbrace{\dfrac{N!^2}{(N-x_1)!\,(N-x_2)!}}_{\text{non-cancelling terms from } \pi(\quad)} \; \underbrace{B(a+x_1+x_2,\, b+N-x_1-x_2)}_{\text{non-cancelling terms from } Q(\quad)}}$$

$\boxed{\text{R demo}}$ block updates much faster convergence!

Effective sample size, variance of $\hat{\mu}_{MCMC}$

MC estimator of $EY = Eh(X)$ is $\hat{\mu}_{MC} = \frac{1}{n}\sum_{i=1}^{n} h(x_i) = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y}$

For iid samples

$$Var(\hat{\mu}_{MC}) = \frac{1}{n} Var(y_i)$$

where

$$\widehat{Var(y_i)} = s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

MCMC: Estimator $\hat{\mu}_{MCMC} = \frac{1}{n}\sum h(x_i) = \bar{y}$ has variance

$$Var(\hat{\mu}_{MCMC}) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} y_i\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} Cov(y_i, y_j)$$

$$= \frac{Var(y_i)}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\rho(|i-j|)$$

$$= \frac{Var(y_i)}{n^2}\left(n\rho_0 + 2(n-1)\rho_1 + \ldots + 2\rho_{n-1}\right)$$

$$\approx \frac{Var(y_i)}{n}\left(1 + 2\sum_{k=1}^{\infty}\rho_i\right) \qquad (*)$$

Plugging in

$$\hat{\rho}(k) = \frac{\sum_{i=1}^{n-k}(x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

estimator of ρ we obtain "time-series SE" of $\hat{\mu}_{MCMC}$.
(but R-package coda instead based on Heidelberger et. al. (1981))

Effective sample size ESS defined by equating $\frac{Var(\bar{y}_\cdot)}{ESS}$ to (*) giving

$$ESS = \frac{n}{\underbrace{1 + 2\sum_{k=1}^{\infty} \rho(i)}_{\text{"auto-correlation time } \tau\text{"}}}$$

(Kass et. al. 1998)

Other diagnostic tools in R-package coda.

Ex.: Gibbs sampling, random intercept linear mixed model (LMM)

For $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n_i$, $\sum_{i=1}^{m} n_i = n$,

$$y_{ij} = \mu + \gamma_i + \varepsilon_{ij} \qquad (1)$$

where

$$\varepsilon_{ij} \mid \tau \overset{iid}{\sim} N(0, \tau^{-1}) \qquad (2)$$

and

$$\gamma_i \mid \nu \overset{iid}{\sim} N(0, \nu^{-1}) \qquad (3)$$

As in TMA4315

Prior

$\pi(\mu) \propto 1$ (improper)

$\left.\begin{array}{l}\tau \sim \text{Gamma}(\alpha, \beta) \\ \nu \sim \text{Gamma}(\alpha, \beta)\end{array}\right\}$ $(\alpha = 0.01, \beta = 0.01)$ (weakly informative)

To eliminate the deterministic relation (1) and ε_{ij}, (1) and (2) can be replaced by the statement

$$y_{ij} \mid \mu, \gamma_i, \tau \overset{iid}{\sim} N(\mu + \gamma_i, \tau^{-1})$$

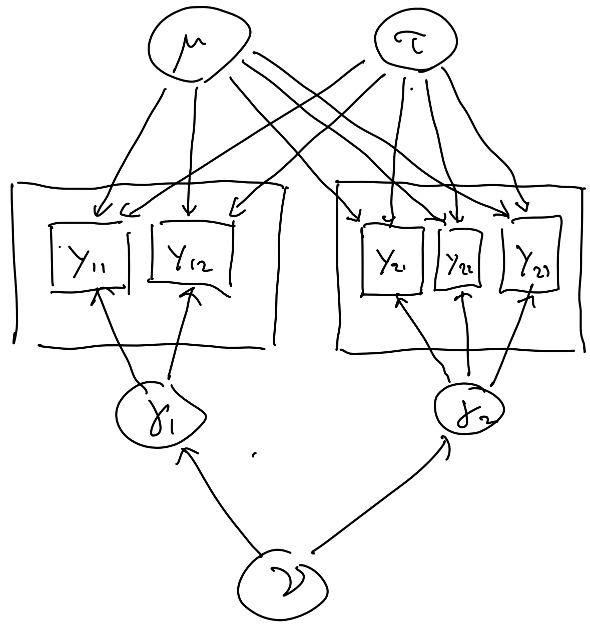
Model DAG (directed acyclic graph):

$m = 2$

$n_1 = 2$

$n_2 = 3$

$n = 5$



In general, letting $\underline{x} = (\underline{\theta}, \underline{y}) = (x^1, x^2, \dots, x^k)$ where $\underline{\theta}$ are parameters (circular nodes) and $\underline{y}$ data (square nodes), the posterior

$$\pi(\underline{\theta} \mid \underline{y}) \propto \pi(\underline{\theta}, \underline{y}) = \pi(\underline{x}) = \prod_i p(x^i \mid \text{parents}(x^i))$$

Componentwise conditionals simplify to

$$\pi(x^i \mid \underline{x}^{-i}) = \text{terms in } \pi(\underline{x}) \text{ containing } x^i$$

$$= p(x^i \mid \text{parents}(x^i)) \prod_{j \in \text{children}(i)} p(x^j \mid \text{parents}(j))$$

Thus, for the above LMM,

$$\pi(\nu \mid \ ) \propto \underbrace{\nu^{\alpha-1} e^{-\beta\nu}}_{\pi(\nu \mid \text{parents}(\nu)) \atop \overset{\shortparallel}{\emptyset}} \prod_{i=1}^{m} \underbrace{\nu^{\frac{1}{2}} e^{-\frac{1}{2}\nu \gamma_i^2}}_{\pi(\gamma_i \mid \text{parents}(\gamma_i))}$$

$$= \nu^{\alpha + \frac{m}{2} - 1} e^{-(\beta + \frac{1}{2}\sum \gamma_i^2)\nu} \sim \text{Gamma}\left(\alpha + \frac{m}{2}, \beta + \frac{1}{2}\sum \gamma_i^2\right)$$

Similarly,

$$\pi(\tau \mid \ ) \propto \underbrace{\nu^{\alpha-1} e^{-\beta\nu}}_{\pi(\nu)} \underbrace{\prod_{i=1}^{m} \prod_{j=1}^{n_i} \tau^{\frac{1}{2}} e^{-\frac{\tau}{2}(y_{ij} - \mu - \gamma_i)^2}}_{\pi(y_{ij} \mid \text{parents}(y_{ij}))}$$

$$= \nu^{\alpha + \frac{n}{2} - 1} e^{-\left(\beta + \frac{1}{2}\sum_i \sum_j (y_{ij} - \mu - \gamma_i)^2\right)}$$

$$\sim \text{Gamma}\left(\alpha + \frac{n}{2}, \beta + \frac{1}{2}\sum_i \sum_j (y_{ij} - \mu - \gamma_i)^2\right),$$

$$\pi(\mu \mid \;) \propto \underbrace{1}_{\pi(\mu)} \prod_i \prod_j \underbrace{\tau^{\frac{1}{2}} e^{-\frac{\tau}{2}(\mu - y_{ij} + \gamma_i)^2}}_{\pi(y_{ij} \mid \text{parents}(y_{ij}))}$$

$$\sum_i \sum_j (\mu - y_{ij} + \gamma_i)^2 = \sum_i \sum_j \left[\mu^2 - 2(y_{ij} - \gamma_i)\mu + (y_{ij} - \gamma_i)^2\right]$$

$$= n\mu^2 - 2\mu \underbrace{\sum_i \sum_j (y_{ij} - \gamma_i)} + \sum_i \sum_j (y_{ij} - \gamma_i)^2$$

$$= n\left(\mu - \frac{1}{n}\sum_i \sum_j (y_{ij} - \gamma_i)\right)^2 + c'$$

Thus $\mu \mid \underline{y}, \underline{\gamma}, \tau \sim N\left(\frac{1}{n}\sum_i \sum_j (y_{ij} - \gamma_i), (n\tau)^{-1}\right)$

Finally

$$\pi(\gamma_i \mid \;) \propto \underbrace{\nu^{\frac{1}{2}} e^{-\frac{\nu}{2}\gamma_i^2}}_{\pi(\gamma_i \mid \text{parents}(\gamma_i))} \prod_{j=1}^{n_i} \underbrace{\tau^{\frac{1}{2}} e^{-\frac{1}{2}(\gamma_i - y_{ij} + \mu)^2}}_{\pi(y_{ij} \mid \text{parents}(y_{ij}))}$$

$$\propto e^{-\frac{1}{2}\left[\nu\gamma_i^2 + \tau \sum_{j=1}^{n_i}\left(\gamma_i^2 - 2\gamma_i(y_{ij} - \mu) + (y_{ij} - \mu)^2\right)\right]}$$

$$[\;\;] = (\nu + n_i\tau)\gamma_i^2 - 2\tau\gamma_i \sum_{j=1}^{n_i} y_{ij} - \mu + c$$

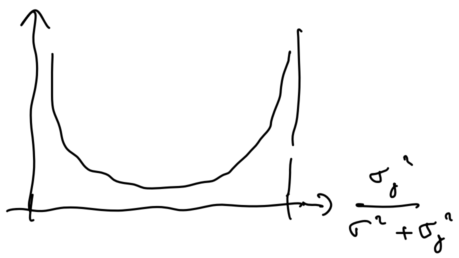
$$= (\nu + n_i\tau)\left(\gamma_i - \frac{\tau \sum (y_{ij} - \mu)}{\nu + n_i\tau}\right)^2$$

i.e. $\gamma_i \mid \ldots \sim N\left(\frac{\tau \sum (y_{ij} - \mu)}{\nu + n_i\tau}, (\nu + n_i\tau)^{-1}\right)$

The prior $\tau \sim \text{Gamma}(\alpha, \beta)$ indep. of $\nu \sim \text{Gamma}(\alpha, \beta)$ translates to the proportion of variance explained by the random intercept,

$$\frac{\sigma_\gamma^2}{\sigma^2 + \sigma_\gamma^2} \quad \frac{\frac{1}{\nu}}{\frac{1}{\nu} + \frac{1}{\tau}} = \frac{\tau}{\nu + \tau} \sim \text{Beta}(\alpha, \alpha)$$

with large mass near 0 and 1 for small α

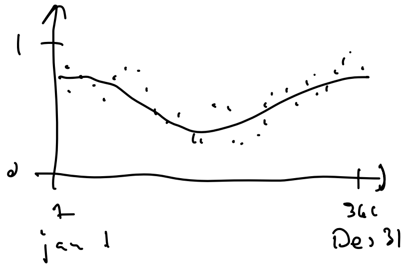


Since the likelihood tends to a positive limit as random intercept variance $\sigma_\gamma^2 \to 0$ the large mass near zero will be present also in the posterior. Sending $\alpha \to 0$, the posterior is improper as noted by Hobert & Casella 1996.

Fuglstad et. al. 2020 instead propose independent priors on $\sigma_\gamma^2 / (\sigma^2 + \sigma_\gamma^2)$ and $\sigma_{tot}^2 = \sigma^2 + \sigma_\gamma^2$ making σ_γ^2 and σ^2 a priori dependent.

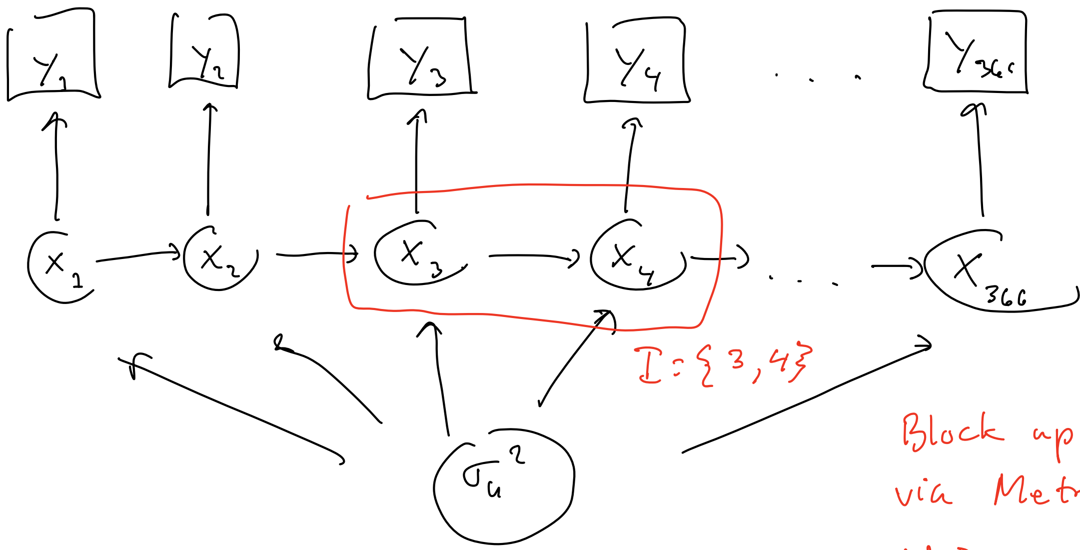
Project 2

DAG



39 years of data
1951 – 1989

Estimate seasonal trend



$$Y_t \mid x_t \overset{iid}{\sim} \text{bin}(n_t, \pi(x_t))$$

$$x_t = x_{t-1} + u_t, \quad u_t \overset{iid}{\sim} N(0, \sigma_u^2)$$

$$\sigma_u^2 \sim \text{invGamma}(\alpha, \beta)$$

Block update of $\underline{x}_{(a,b)}$ via Metropolis-Hastings:

Using proposal

$$p(\underline{x}'_I \mid \underline{x}_{-I}, \sigma_u^2)$$

i.e. target density unconditional on $\underline{y}_I$ leads to

$$\alpha(\underline{x}'_I \mid \underline{x}_{-I}, \underline{y}, \sigma_u^2)$$

$$= \min\left(1, \frac{p(\underline{y}_I \mid \underline{x}'_I)}{p(\underline{y}_I \mid \underline{x}_I)}\right)$$

MCMC

Software

BUGS
WinBUGS
JAGS

$\Big\}$ Gibbs sampling, often slow

Stan (Hamiltonian MCMC)

Laplace approximation

$$\int_{\mathbb{R}^p} e^{-f(\underline{x})} d\underline{x} \approx \int_{\mathbb{R}^p} e^{-f(\hat{\underline{x}}) - \frac{1}{2}(x-\hat{x})^T H(\hat{x})(x-\hat{x})} d\underline{x} \quad \text{where } \hat{\underline{x}} = \arg\max f(\underline{x})$$

$$= e^{-f(\hat{\underline{x}})} (2\pi)^{\frac{p}{2}} \left| H(\hat{x}) \right|^{-\frac{1}{2}}$$

Ex.: $n! = \Gamma(n-1) = \int_0^\infty x^n e^{-x} dx$

$$= \int_0^\infty e^{-x + n\ln x} dx$$

$f(x) = x - n\ln x$

$f'(x) = 1 - \frac{n}{x}$ $\qquad \hat{x} = n$

$f''(x) = \frac{n}{x^2}$

$f''(\hat{x}) = \frac{1}{n}$

$f(\hat{x}) = n(1 - \ln n)$

$$\overset{n \text{ large}}{=} \int_{-\infty}^{\infty} e^{-f(\hat{x}) - \frac{1}{2} f''(\hat{x})(x-\hat{x})} dx$$

$$= e^{-f(\hat{x})} (2\pi)^{\frac{1}{2}} f''(\hat{x})^{-\frac{1}{2}}$$

$$= e^{-n(1-\ln n)} (2\pi n)^{\frac{1}{2}}$$

$$\simeq (2\pi n)^{\frac{1}{2}} \left(\frac{n}{e}\right)^n \quad \text{(Stirling's approximation)}$$

Latent variable models, e.g. LMMs, GLMMs, state space models

Model of latent vector $\underline{x}$:

$\pi(\underline{x}|\underline{\theta})$, often high dimensional and Gaussian

Model for observations $\underline{y}$ conditional on $\underline{x}$ and $\underline{\theta}$

$\pi(\underline{y}|\underline{x},\underline{\theta})$

Ex.: Random intercept logistic regression (GLMM)

For $i=1,\dots,m$, $j=1,\dots,n_i$

$$y_{ij}|\gamma_i \sim \text{bin}(n_{ij}, \text{logit}^{-1}(\eta_{ij}))$$

where

$$\eta_{ij} = \beta_0 + \gamma_i$$

and

$$\gamma_i \sim N(0,\sigma^2)$$

Ex.: Project 2A

To do inference (Bayesian or maximum likelihood) we need the marginal likelihood

$$\pi(\underline{y}|\underline{\theta}) = \int \pi(\underline{y},\underline{x}|\underline{\theta})\,d\underline{x}$$

$$= \int \pi(\underline{x}|\theta)\pi(\underline{y}|\underline{x},\theta)\,d\underline{x} \quad \propto \int e^{-\frac{1}{2}\underline{x}^T Q \underline{x}^T + \sum \ln \pi(y_i|x_i,\theta)}\,d\underline{x}$$

$$= \int e^{-f(\underline{x},\underline{y},\theta)}\,d\underline{x}$$

$$\approx \int e^{-f(\hat{x}(\theta),\underline{y},\theta) - \frac{1}{2}(x-\hat{x}(\theta))^T H(\hat{x}(\theta))(x-\hat{x}(\theta))}\,d\underline{x}$$

$$= \pi(y,\hat{x}(\theta))(2\pi)^{\frac{p}{2}}\left(H(\hat{x}(\theta))\right)^{-\frac{1}{2}} = \tilde{\pi}(\tilde{y}|\theta)$$

Exact if $\pi(\underline{y},\underline{x}|\theta)$ Gaussian in $\underline{x}$.

Sparsity

If $\pi(\underline{x} \mid \theta)$ is Gaussian with precision matrix Q and Q is sparse (e.g. if $\underline{x}$ is a Gaussian Markov random field), and $y_1, y_2, \ldots, y_n \mid \underline{x}$ are conditionall independent, then $H(\underline{x}(\underline{\theta}))$ is also sparse which simplifies/speeds up numerical optimization to find $\hat{x}(\underline{\theta})$ and computation of $|H(\hat{x}(\underline{\theta}))|$

Required in INLA. Computation of $\tilde{\pi}(\underline{y} \mid \underline{\theta})$ "hand-coded". for specific model classes.

TMB/RTMB

R packages TMB/RTMB computes an R-funtion computing $\hat{x}(\underline{\theta})$ (via numerical optimization) and $\tilde{\pi}(\underline{y} \mid \underline{\theta})$

via AD from a C++/R-function computing $f(\underline{x}, \underline{y}, \underline{\theta})$ supplied by the user.
Automatic sparsity detection.

MLEs computed by optimizing $\tilde{\pi}(\underline{y} \mid \underline{\theta})$ numerically (outer optimization)

(Stan/TMBstan: Hamiltonian MCMC sampling from $\pi(\underline{\theta} \mid \underline{y}) \propto \tilde{\pi}(\underline{y} \mid \underline{\theta})\, \pi(\underline{\theta})$)

Automatic differentiation (AD)

Toy example: Suppose $f(x_1, x_2, x_3) = x_1 x_2 \sin(x_3) + \exp(x_1 x_2)/x_3$

For $x_1 = 1$

$x_2 = 2$

$x_3 = \frac{\pi}{2}$

find $f(x_1, \ldots, x_3)$ and $\frac{\partial f}{\partial \underline{x}}$

$x_1 \qquad \frac{\partial x_1}{\partial \underline{x}} = (1, 0, 0)^T \qquad \frac{\partial^2 x_1}{\partial \underline{x} \partial \underline{x}^T} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

$x_2 \qquad \frac{\partial x_2}{\partial \underline{x}} = (0, 1, 0)^T$

$x_3 \qquad \frac{\partial x_3}{\partial \underline{x}} = (0, 0, 1)^T$

$w_1 = x_1 x_2 \qquad \frac{\partial w_1}{\partial \underline{x}} = \frac{\partial x_1}{\partial \underline{x}} \cdot x_2 + \frac{\partial x_2}{\partial \underline{x}} \cdot x_1 = (1, 0, 0)^T x_2 + (0, 1, 0)^T x_1$

$= (x_2, x_1, 0)^T$

$w_2 = \sin(x_3) = \cos(x_3) \frac{\partial x_3}{\partial \underline{x}} = \cos(x_3)(0, 0, 1)^T$

$w_3 = w_1 w_2 = \frac{\partial w_1}{\partial \underline{x}} \cdot w_2 + \frac{\partial w_2}{\partial \underline{x}} \cdot w_1$

$\vdots$

Software implementation work with new class of dual number $(w, \frac{\partial w}{\partial \underline{x}})$ and define new methods (basic functions and operators) for this class $\boxed{\text{R demo}}$

INLA in more detail

Aim: Compute posterior marginals of each θ_i and x_i.

Laplace of marginal likelihood:

$$\tilde{\pi}(\underline{y} \mid \underline{\theta}) \approx \pi(y, \hat{x} \mid \theta)\,(2\pi)^{\frac{p}{2}}\,\left|H(\hat{x}(\theta))\right|^{-\frac{1}{2}}$$

Approximate joint posterior of hyperparameters can then be written

$$\tilde{\pi}(\underline{\theta} \mid \underline{y}) \propto \tilde{\pi}(\underline{y} \mid \underline{\theta})\,\pi(\underline{\theta}) \qquad \text{(Bayes)}$$

$$= \pi(y, \hat{\underline{x}} \mid \theta)\,\pi(\underline{\theta})\,(2\pi)^{\frac{p}{2}}\,\left|H(\hat{x})\right|^{-\frac{1}{2}}$$

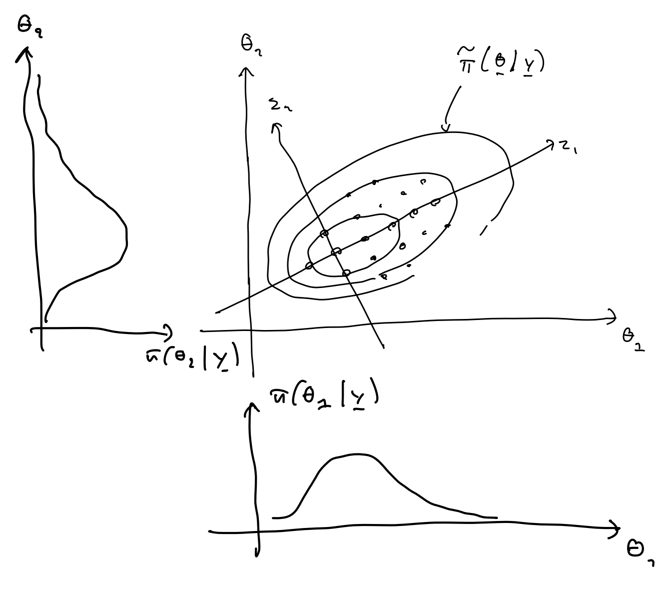
$$= \frac{\pi(\underline{y}, \hat{\underline{x}} \mid \underline{\theta})}{\tilde{\pi}_G(\hat{\underline{x}} \mid y, \underline{\theta})} \qquad \text{(Rue et al. 2009, eq. 3)}$$

where $\tilde{\pi}_G(y, \theta)$ is the Gaussian approx. of $\pi(\underline{x} \mid \underline{y}, \underline{\theta})$

Hyperparameter marginals given by

$$\pi(\theta_i \mid \underline{y}) = \int \pi(\underline{\theta} \mid \underline{y})\, d\underline{\theta}_{-i}$$

are computed by numerical integration (strategic placement of evaluation grid points $\underline{\theta}_1, \underline{\theta}_2, \ldots$ with weights $\Delta_1, \Delta_2, \ldots$ followed by numerical integration of interpolant)



Latent variable marginal posteriors:

In general

$$\pi(x_i \mid \underline{y}) = \int \pi(x_i \mid \underline{y}, \underline{\theta})\, \pi(\theta \mid \underline{y})\, d\underline{\theta} \approx \sum_k \underbrace{\pi(x_i \mid \underline{y}, \underline{\theta}_k)}_{*} \Delta_k$$

Gaussian approximation of $(*)$:

$$\pi(x_i \mid \underline{y}, \underline{\theta}_k) \approx \tilde{\pi}_G(x_i \mid \underline{y}, \underline{\theta}_k)$$

Inner Laplace approximation of $(*)$ w.r.t. $\underline{x}_{-i}$

$$\pi(x_i \mid \underline{y}, \underline{\theta}) \propto \left. \frac{\pi(\underline{x}, \underline{y}, \underline{\theta})}{\tilde{\pi}_{GG}(\underline{x}_{-i} \mid x_i, \underline{y}, \underline{\theta})} \right|_{\underline{x}_{-i} = \hat{\underline{x}}_{-i}} \qquad (*)$$

$$\text{where } \hat{\underline{x}}_{-i} = \underset{x_{-i}}{\operatorname{argmax}}\ \pi(\underline{x}, \underline{y}, \underline{\theta}) \approx E_{\tilde{\pi}_G}(\underline{x}_{-i} \mid x_i)$$

Evaluated at $x_i^{(s)}$, $s = 1, 2, \ldots$ chosen by Gauss-Hermite quadrature rule (why?) representing final density by

$$\tilde{\pi}_{LA}(x_i \mid \underline{\theta}, \underline{y}) \propto N\left(x_i; \mu_i(\underline{\theta}), \sigma_i^2(\underline{\theta})\right) \exp\left(\text{cubic spline}(x_i)\right)$$

Simplified Laplace

See section 3.2.3 in Rue et. al. (2009)

INLA, glmmTMB and RTMB example

Hype park purse snatching data:

y_i: Number of purses snatch on day i.

Model: $y_i | x_i \sim \text{Pois}(\lambda_i)$

where

$$\ln \lambda_i = \eta_i = \mu + x_i \;, \quad \lambda_i = e^{\mu + x_i}$$

and

$$x_i \sim AR(1)$$

i.e. the stationary solution of

$$x_i = \rho x_{i-1} + \varepsilon_i$$

where $\varepsilon_i \overset{iid}{\sim} N(0, \tau^{-1})$. Stationary variance satisfy,

$$K^{-1} = Var(x_i) = Var(\rho x_{i-1} + \varepsilon_i) = \rho^2 Var(x_{i-1}) + \tau^{-1} = \rho^2 K^{-1} + \tau^{-1}$$

thus

$$K^{-1} = \frac{\tau^{-1}}{1-\rho^2}$$

$$K = \tau(1-\rho^2)$$

Joint density of $\underline{x}$ is

$$\pi(\underline{x} | \underline{\theta}) = \underbrace{\pi(x_1 | \theta)}_{\sim N(0, K^{-1})} \prod_{i=2}^{n} \underbrace{\pi(x_i | x_{i-1}, x_{i-2}, \ldots)}_{\sim N(\rho x_{i-1}, \tau^{-1})}$$

(multivariate normal)

Joint likelihood / density of $\underline{x}, \underline{y}$:

$$\pi(\underline{y}, \underline{x} | \underline{\theta}) = \pi(x_1 | \theta) \prod_{i=1}^{2} \pi(x_i | x_{i-1}, \theta) \prod_{i=1}^{n} \pi(y_i | x_i, \underline{\theta})$$

Specifying priors in INLA via hyper argument to f( )

Specified on internal parameters, e.g. for f(model="ar1")

on

	Default	
$\theta_1 = \ln \kappa$	loggamma	1, 5e-5
$\theta_2 = \ln \frac{1+\rho}{1-\rho}$	normal	0, 0.15
~~$\theta_3 = \mu$~~	normal	fixed = TRUE

For detail, see ?inla.models, inla.doc("ar1")

Fixed effect priors via ?control.fixed

Sanity check:

Marginal variance of each y_i predicted by model is

$$\mathrm{Var}(y_i) = E\,\mathrm{Var}(y_i|x_i) + \mathrm{Var}\,E(y_i|x_i)$$
$$= E\left(e^{\mu + x_i}\right) + \mathrm{Var}\left(e^{\mu + x_i}\right)$$
$$= e^{\mu + \kappa^{-1}/2} + e^{2\mu + \kappa^{-1}}\left(e^{\kappa^{-1}} - 1\right)$$

Does it match $\widehat{\mathrm{Var}(y_i)} = S_y^2$?