Bayesian Statistics

Markov Chain Monte Carlo (MCMC) Methods

Brani Vidakovic

Professor

School of Industrial and Systems Engineering

Before We Begin…

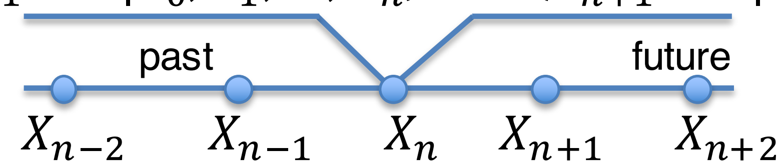
In this unit:

- Background + History
- Metropolis Algorithm
- Gibbs Sampler

Markov Chain Monte Carlo

$X_0, X_1, X_2, \dots, X_{n-1}, X_n, X_{n+1}, \dots$ forms a Markov Chain if

- $P(X_{n+1} \in A | X_0, X_1, \dots, X_n) = P(X_{n+1} \in A | X_n)$



- Given X_n, future (events defined on $X_{n+1}, X_{n+2}, \dots$) is independent of past (events defined on $\dots, X_{n-2}, X_{n-1}$)
- $P(X_{n+1} \in A | X_n) = Q(A | X_n)$ (transition kernel)
- $Q(A | X_n = x) = \int_A q(x, y) dy = \int_A q(y|x) dy$
- Π is invariant distribution if

$$\Pi(A) = \int Q(A|x) \Pi(dx)$$

- π is density for Π, it is stationary if

$$q(x|y)\pi(y) = q(y|x)\pi(x)$$

(detailed balance equation)

- If $Q^n(A|x) = P(X_n \in A|X_0 = x)$,

$$\lim_{n\to\infty} Q^n(A|x) = \Pi(A)$$

Π is equilibrium distribution

❖ Construct Markov Chain so that the equilibrium distribution corresponds to the posterior

→ Initial condition $X_0 = x$ is “forgotten” and when n is large, X_n is a random variable sampled from the posterior

GTx

Monte Carlo

- Term coined by Metropolis for approximation methodology based on sampling
- $\mu_\pi(g) = \int g(\theta)\pi(\theta)d\theta$?
- If we can sample $\theta_1, \theta_2, \ldots, \theta_n \overset{\text{iid}}{\sim} \pi$ then

$$\mu_\pi(g) \approx \frac{1}{n}\sum_{i=1}^{n} g(\theta_i) \qquad [\text{SLLN}]$$

- ~~iid~~ Markovian dependence in $\theta_1, \ldots, \theta_n$

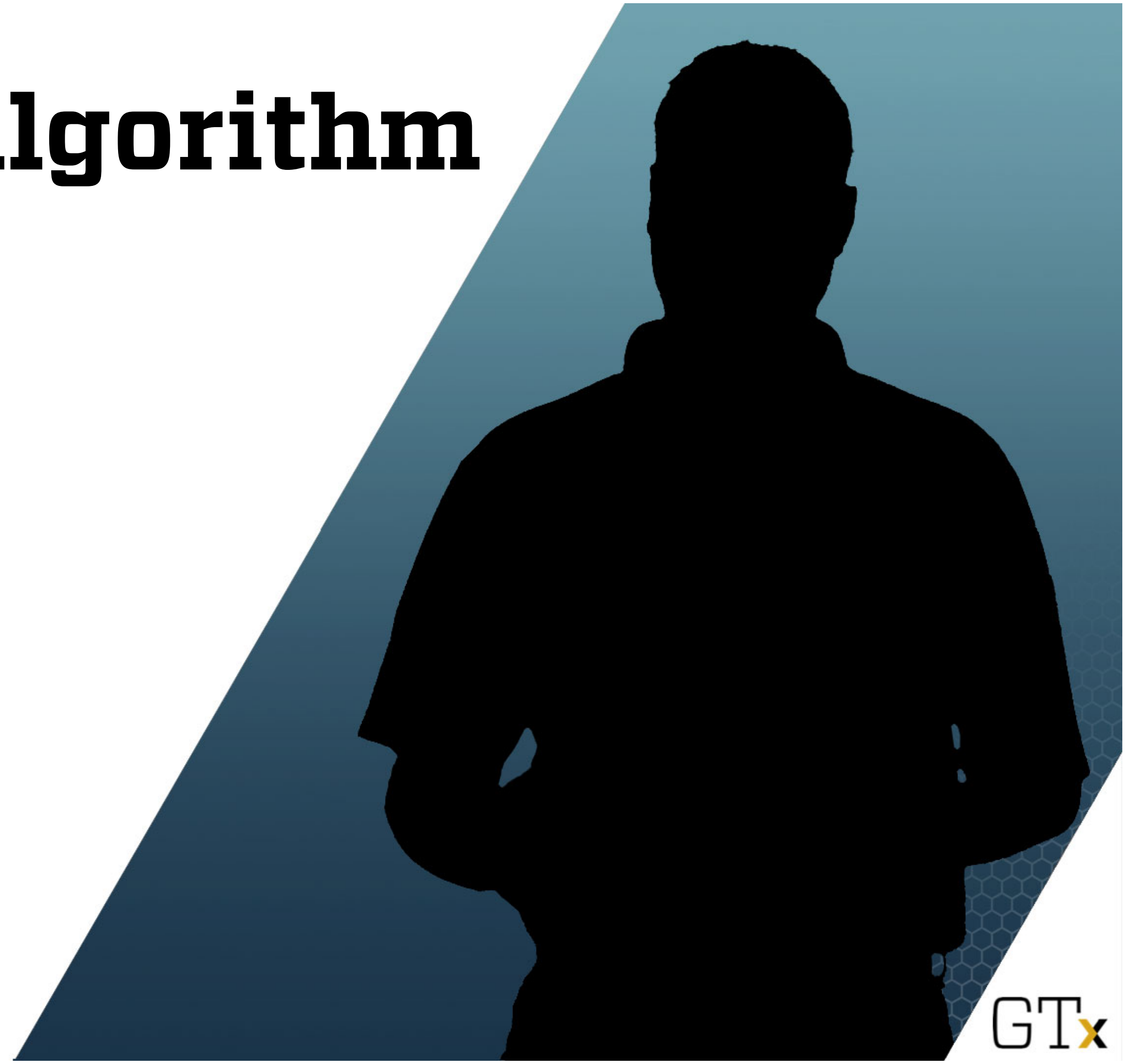
$$\mu_\pi(g) \approx \frac{1}{n}\sum_{i=1}^{n} g(\theta_i) \qquad [\text{ergodic} - \text{type theorems}]$$

- Metropolis algorithm – Gibbs Sampling

- Some History
  - Metropolis, Rosenbluth, Rosenbluth, Teller “Equation of state calculations by fast computing machines”, Journal of Chemical Physics, 21, 6, 1953
  - Hastings (1970) in Biometrika paper considers continuous case.
  - Besag (1974) Hammersley-Clifford theorem, start of Gibbs sampling
  - Geman & Geman (1984)
  - Tanner and Wong (1987)
  - Gelfand and Smith (1990)!
  - Casella and George (1992)

Metropolis Algorithm

- Theoretical background
- Examples (Octave)



Metropolis Algorithm

- Detailed balance equation

$$q(y|x)f(x) = q(x|y)f(y)$$

- π is target, q is kernel density of MC

Call q admissible if support $(\pi_x) \subset \bigcup_x$ support $q(\cdot\,|x)$

In general: $q(y|x)\pi(x) \neq q(x|y)\pi(y)$

(say $>$, wlog)

$$q(y|x)\rho(x,y)\pi(x) = q(x|y)\pi(y)\times 1$$

$$\Rightarrow \rho(x,y) = \frac{q(x|y)\pi(y)}{q(y|x)\pi(x)} \wedge 1$$

- Since $\rho(x,y)$ depends on $\frac{\pi(y)}{\pi(x)}$, when target is the posterior, normalizing marginal distributions cancel!

STEP 1. Start with arbitrary x_0 from the support of target π

STEP 2. At stage n, generate proposal y from $q(y|x_n)$

STEP 3. $x_{n+1} = y$, with prob. $\rho(x_n, y)$

$x_{n+1} = x_n$, with prob. $1 - \rho(x_n, y)$

(Generate $U \sim \mathrm{U}(0,1)$ and accept proposal y if $U \leq \rho(x_n, y)$)

STEP 4. Increase n and go to STEP 2.

$$\rho(x,y) = \frac{q(x|y)\pi(y)}{q(y|x)\pi(x)} \wedge 1$$

How to select q?

Any admissible choice would do! But…

GTx

- If $q(x|y) = q(y|x)$, i.e. if the kernel density is symmetric, then

$$\rho(x, y) = \frac{\pi(y)}{\pi(x)} \wedge 1$$

- If $q(x|y) = q(y|x) = q(|x - y|)$, the algorithm is called Metropolis random walk (original proposal by Metropolis)

- $q(y|x) \equiv q(y)$ [Free of x]

  Algorithm is called **Independence Metropolis**

Example 1. Metropolis algorithm for

$X|\theta \sim \mathrm{N}(\theta, 1)$ and $\theta \sim \mathrm{Ca}(0,1)$.

$\pi(\theta|x) \propto \frac{e^{-\frac{(x-\theta)^2}{2}}}{1+\theta^2}$; θ' is the proposal, θ is the current status

$q(\theta'|\theta)$ density of $\mathrm{N}(x, \tau^2)$: $q \propto e^{\frac{1}{2\tau^2}(\theta'-x)^2}$

Take $\tau^2 = 1$.

$$\gamma = \frac{\pi(\theta')q(\theta|\theta')}{\pi(\theta\ )q(\theta'|\theta\ )} = \frac{\frac{e^{-\frac{(x-\theta')^2}{2}}}{1+\theta'^2} e^{-\frac{(\theta-x)^2}{2}}}{\frac{e^{-\frac{(x-\theta)^2}{2}}}{1+\theta^2} e^{-\frac{(\theta'-x)^2}{2}}} = \frac{1+\theta^2}{1+(\theta')^2}.$$

$$\rho = 1 \wedge \frac{1+\theta_n^2}{1+(\theta')^2}; \qquad \theta_{n+1} = \begin{cases} \theta' & \text{w.p. } \rho \\ \theta_n & \text{w.p. } 1-\rho \end{cases}$$

$X = 2$

$\theta_0 = 1 \Rightarrow$ *norcaumet.m*

Result: $\delta(2) = 1.2825$

for a fixed random number-generator seed.

GTx

Example 2.

Weibull Distribution

$$T_1, T_2, \ldots, T_n \sim \mathrm{Wei}(\alpha, \eta)$$

$$f(t|\alpha, \eta) = \alpha\eta t^{\alpha-1} e^{-\eta t^{\alpha}}$$

(for $\alpha = 1$, Weibull≡Exponential)

$$\pi(\alpha, \eta) \propto e^{-\alpha} \eta^{\beta-1} e^{-\xi\eta}$$

Proposal, product of two exponentials

$$q(\alpha', \eta'|\alpha, \eta) = \frac{1}{\alpha\eta} \exp\left\{-\frac{\alpha'}{\alpha} - \frac{\eta'}{\eta}\right\}$$

α, η old, current | accept proposal

α', η' proposed | with probability ρ

GTx

$$\rho = 1 \wedge \frac{\left[\prod_{i=1}^{n} \alpha' \eta' t_i^{\alpha'-1} e^{-\eta' t_i^{\alpha'}}\right] e^{-\alpha'} (\eta')^{\beta-1} e^{-\xi \eta'} \frac{1}{\alpha' \eta'} e^{-\frac{\alpha}{\alpha'}-\frac{\eta}{\eta'}}}{\left[\prod_{i=1}^{n} \alpha\, \eta\, t_i^{\alpha\, -1} e^{-\eta\, t_i^{\alpha}}\right] e^{-\alpha}\, \eta^{\beta-1} e^{-\xi \eta} \frac{1}{\alpha\, \eta} e^{-\frac{\alpha'}{\alpha}-\frac{\eta'}{\eta}}}$$

metro2.m $\beta = 2,\ \xi = 2$ hyperparameters

$T = [0.2\ \ 0.1\ \ 0.25];\ \ n = 3$

Initial values: $\alpha = 2,\ \eta = 2$

Result: $\hat{\alpha} \cong 0.9$, $\hat{\eta} \cong 1.85$

Gibbs Sampler

Gibbs Sampler

- Special case of Metropolis algorithm
- Component-wise update with “proposals” being full conditional distributions of components
- It can be shown that $\rho = 1$, i.e. Gibbs “proposal” is accepted at each step.

Let $f\left(\underset{\sim}{X} \mid \underset{\sim}{\theta}\right)\pi\left(\underset{\sim}{\theta}\right)$ be the numerator of the posterior. Suppose that we can find all full conditionals for components of $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_p)$:

$$\pi\left(\theta_1 | \theta_2, \theta_3, \ldots, \theta_n, \underset{\sim}{X}\right)$$
$$\pi\left(\theta_2 | \theta_1, \theta_3, \ldots, \theta_n, \underset{\sim}{X}\right)$$
$$\vdots$$
$$\pi\left(\theta_n | \theta_1, \theta_2, \ldots, \theta_{n-1}, \underset{\sim}{X}\right)$$

GTx

The Gibbs sampler proceeds as follows:

Start $\underset{\sim}{\theta}^{0} = \left(\theta_1^0, \theta_2^0, \dots, \theta_p^0\right)$ [initials]

Sample θ_1^{n+1} from $\pi\left(\theta_1 | \theta_2^n, \theta_3^n, \dots, \theta_p^n, \underset{\sim}{X}\right)$

Sample θ_2^{n+1} from $\pi\left(\theta_2 | \theta_1^{n+1}, \theta_3^n, \dots, \theta_p^n, \underset{\sim}{X}\right)$

Sample θ_3^{n+1} from $\pi\left(\theta_3 | \theta_1^{n+1}, \theta_2^{n+1}, \theta_4^n, \dots, \theta_p^n, \underset{\sim}{X}\right)$

$\vdots$

Sample θ_p^{n+1} from $\pi\left(\theta_p | \theta_1^{n+1}, \dots, \theta_{p-1}^{n+1}, \underset{\sim}{X}\right)$

Increase n

- How to find full conditionals?
- Form a kernel of joint distribution of all parameters and data
- To find the full conditional for component θ_i, select only the parts of the kernel that contain θ_i, all other θ^{is} and data are considered constant
- Normalize the selected part as a distribution. Often, you can recognize what distribution it is from the form of kernel
- Important: You should be able to sample from all conditionals

For example:

$X_1, \dots, X_n \sim \mathrm{N}\left(\theta, \frac{1}{\tau}\right)$; τ precision, $\frac{1}{\tau} = \sigma^2$

$\mu \sim \mathrm{N}(0,1) \qquad \tau \sim \mathrm{Ga}(2,1)$

$$\boxed{\text{joint} \propto (2\pi)^{-\frac{n+1}{2}} \tau^{\frac{n}{2}} e^{-\frac{\tau}{2}\sum_{i=1}^{n}(x_i-\mu)^2}\, e^{-\frac{1}{2}\mu^2}\, \tau e^{-\tau}}$$

$$\pi\left(\mu | \tau, \underset{\sim}{X}\right) \propto e^{-\frac{\tau}{2}\sum_{i=1}^{n}(x_i-\mu)^2}\, e^{-\frac{1}{2}\mu^2} \propto e^{-\frac{1}{2}(1+n\tau)\left(\mu - \frac{\tau\sum x_i}{1+n\tau}\right)^2}$$

$$\boxed{\mu | \tau, \underset{\sim}{X} \sim \mathrm{N}\left(\frac{\tau \sum x_i}{1+n\tau}, \frac{1}{1+n\tau}\right)}$$

$$\pi\left(\tau | \mu, \underset{\sim}{X}\right) \propto \tau^{\frac{n}{2}} e^{-\frac{\tau}{2}\sum_{i=1}^{n}(x_i-\mu)^2}\, \tau e^{-\tau} \propto \tau^{\frac{n}{2}+1} e^{-\tau\left[1+\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right]}$$

$$\boxed{\tau | \mu, \underset{\sim}{X} \sim Ga\left(\frac{n}{2}+2, 1+\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right)}$$

Example 1.

$X|\theta \sim \mathrm{N}(\theta, 1)$

$\theta \sim \mathrm{Ca}(0,1)$

Find $\delta(2)$ by Gibbs sampling

$$\pi(\theta) \propto \frac{1}{\tau^2 + (\theta-\mu)^2} \propto \int_0^{+\infty} \sqrt{\frac{\lambda}{2\pi\tau^2}} \exp\left\{-\frac{\lambda}{2\tau^2}(\theta-\mu)^2\right\} \lambda^{\frac{1}{2}-1} e^{-\frac{\lambda}{2}} d\lambda$$

$$\Rightarrow \theta \sim \mathrm{Ca}(\mu, \tau) \Leftrightarrow \theta|\lambda \sim \mathrm{N}\left(\mu, \frac{\tau^2}{\lambda}\right), \qquad \lambda \sim \mathrm{Ga}\left(\frac{1}{2}, \frac{1}{2}\right)$$

If $\mathrm{X}|\theta \sim \mathrm{N}(\theta, \sigma^2)$

$$\begin{cases} \theta|\lambda, x \sim \mathrm{N}\left(\frac{\tau^2}{\tau^2+\lambda\sigma^2} x + \frac{\lambda\sigma^2}{\tau^2+\lambda\sigma^2}\mu, \frac{\tau^2\sigma^2}{\tau^2+\lambda\sigma^2}\right) \\ \lambda|\theta, x \sim \mathrm{Exp}\left(\frac{\tau^2+(\theta-\mu)^2}{2\tau^2}\right) \end{cases}$$

norcaugibbs.m

Example 2.

Pumps. (George, Makov, and Smith, 1993)

	Pump	1	2	3	4	5	6	7	8	9	10
t_i	Time	94.32	15.72	62.88	125.76	5.24	31.44	1.048	1.048	2.096	10.48
x_i	# of failures	5	1	5	14	3	19	1	1	4	22

$$x_i|\theta_i \sim \mathrm{Poi}(\theta_i t_i)$$
$$f(x_i|\theta_i) \propto \theta_i^{x_i} e^{-\theta_i t_i}$$

$$\theta_i \sim \mathrm{Ga}(\alpha, \beta)$$
$$\alpha \equiv 1$$
$$\pi(\theta_i) \propto \beta e^{-\beta\theta_i}$$

$$\beta \sim \mathrm{Ga}(c, d)$$
$$\pi(\beta) \propto \beta^{c-1} e^{-d\beta}$$

joint distr $\propto \left(\prod_{i=1}^{n} \theta_i^{x_i} e^{-\theta_i t_i} e^{-\beta\theta_i}\right)\beta^n \beta^{c-1} e^{-d\beta}$

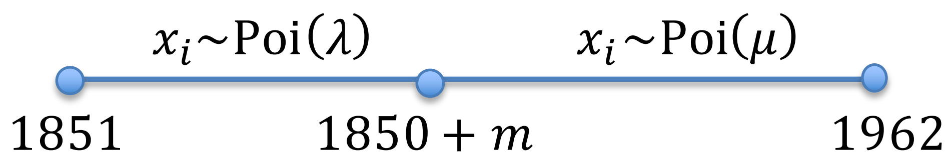
$$\begin{cases} \pi\left(\theta_i|\theta_{\neq i}, \beta, \underset{\sim}{x}\right) \propto \theta_i^{x_i} e^{-\theta_i t_i} e^{-\beta\theta_i} \equiv \mathrm{Ga}(x_i + 1, \beta + t_i) \\ \pi\left(\beta|\underset{\sim}{\theta}, \underset{\sim}{x}\right) \propto e^{-\beta\sum\theta_i} \beta^{n+c-1} e^{-d\beta} \equiv \mathrm{Ga}(n + c, \sum\theta_i + d) \end{cases}$$

Implemented in: *pumpsmc.m*

GTx

Example 3.

- Coal Mining Disasters in UK (Carlin, Gelfand, and Smith, 1992)
- Change point problem



<u>Model:</u>

$$x_i | \lambda \sim \mathrm{Poi}(\lambda), \quad i = 1,2, \dots, m$$

$$x_i | \mu \sim \mathrm{Poi}(\mu), \quad i = m+1, \dots, n$$

$$m \sim \mathrm{DU}(n): \quad P(m = k) = \frac{1}{n}, \;\; k = 1, \dots, n \; (n = 112)$$

$$\lambda \sim \mathrm{Ga}(\alpha\, , \beta)$$

$$\mu \sim \mathrm{Ga}(\gamma, \delta)$$

posterior $\propto L\left(\lambda, \mu, m | \underset{\sim}{X}\right) \pi(\lambda)\pi(\mu)\pi(m)$

$$= \prod_{i=1}^{m} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \prod_{i=m+1}^{n} \frac{\mu^{x_i}}{x_i!} e^{-\mu} \frac{\beta^{\alpha}\lambda^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta\lambda} \frac{\delta^{\gamma}\mu^{\gamma-1}}{\Gamma(\gamma)} e^{-\delta\mu} \frac{1}{n}$$

$$\propto e^{-m\lambda} \lambda^{\sum_{i=1}^{m} x_i} e^{-(n-m)\mu} \mu^{\sum_{i=m+1}^{n} x_i} \lambda^{\alpha-1} e^{-\beta\lambda} \mu^{\gamma-1} e^{-\delta\mu}$$

$$= \lambda^{\alpha+\sum_{i=1}^{m} x_i - 1} e^{-(m+\beta)\lambda} \mu^{\gamma+\sum_{i=m+1}^{n} x_i - 1} e^{-(\delta+n-m)\mu}$$

$\Longrightarrow$

$$\lambda | \mu, m, \underset{\sim}{X} \sim \mathrm{Ga}\left(\alpha + \sum_{i=1}^{m} x_i, \beta + m\right)$$

$$\mu | \underset{\sim}{\lambda}, m, \underset{\sim}{X} \sim \mathrm{Ga}\left(\gamma + \sum_{i=m+1}^{n} x_i, \delta + (n-m)\right)$$

Full conditional for m?

$$\pi\left(m|\lambda,\mu,\underset{\sim}{X}\right) \propto \prod_{i=1}^{m}\frac{\lambda^{x_i}}{x_i!}e^{-\lambda}\prod_{i=m+1}^{n}\frac{\mu^{x_i}}{x_i!}e^{-\mu} =$$

$$= \left[\prod_{i=1}^{n}\frac{\mu^{x_i}}{x_i!}e^{-\mu}\right]e^{m(\mu-\lambda)}\left(\frac{\lambda}{\mu}\right)^{\sum_{i=1}^{m}x_i} = f(x|\mu)g(x|m)$$

$$\pi(m) \propto e^{m(\mu-\lambda)}\left(\frac{\lambda}{\mu}\right)^{\sum_{i=1}^{m}x_i}$$

$$P(m=k) = \frac{\pi(k)}{\sum_{i=1}^{n}\pi(i)}$$

disastersmc.m

Set hyperparameters:

$\alpha = 4,\ \beta = 1 \qquad \frac{\alpha}{\beta} = 4 \approx \lambda$

$\gamma = 1/2,\ \delta = 1 \qquad \frac{\gamma}{\delta} = \frac{1}{2} \approx \mu$

GTx

Summary