MCMC - INTRODUCTION

$$p(x) = \frac{1}{Z}\tilde{p}(x) \quad , Z = \int \tilde{p}(x)\,dx$$

$$\boxed{\text{• BUILD A DTMC } (X_t), t \geq 0 \text{ ON } \mathcal{X} \text{ S.T. } p(x) \text{ IS THE STATIONARY DENSITY OF } (X_t).}$$

$\mathcal{X}$ can be a continuous space.

$p(y|x)$, $x, y \in \mathcal{X}$, TRANSITION KERNEL

$$p(X_1) = \int p(X_1|X_0)\,p(X_0)\,dX_0 \quad \text{on}$$

• ERGODIC CHAIN $\exists t \geq 0 : p(X_t = y \mid X_0 = x) > 0 \quad \forall x, y$

$$\underbrace{\pi(y) = \int p(y|x)\,\pi(x)\,dx}$$ IS INVARIANT

IF (X_t) $t \geq 0$ IS ERGODIC + ...

THEN $\lim_{t \to \infty} p(X_t = x) = \pi(x)$

REVERSIBLE MARKOV CHAIN

$(X_t)_{t \geq 0}$ $\quad$ $p(y|x)$ transition kernel

$$\pi(y) = \int p(y|x)\pi(x)dx$$

A M.C. IS REVERSIBLE (SATISFIES THE BALANCE CONDITION)

$\exists\, \pi$ distribution s.t.

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

Then π is STATIONARY

$$\int p(y|x)\pi(x)dx = \int p(x|y)\pi(y)dx = \pi(y)\int p(x|y)dx = \pi(y) \quad ✓$$

MCMC

$P(x) = \frac{1}{Z} \tilde{P}(x)$

PROPOSAL KERNEL

FIX $q(y|x)$ TRANSITION KERNEL , easy to sample from
and makes M.C. ERGODIC.

$\Rightarrow X_t = x$

1) SAMPLE y FROM $q(y|x)$

2) SET $X_{t+1} = \begin{cases} y & \text{with probability } \alpha(y|x) = \min\left\{1, \dfrac{\tilde{P}(y) \cdot q(x|y)}{\tilde{P}(x) \cdot q(y|x)}\right\} \\ x & \end{cases}$

$\frac{P(y)}{P(x)}$

METROPOLIS-HASTINGS ACCEPTANCE CRITERION

transition kernel of MC

$\boxed{\underbrace{P(y|x)}_{} P(x)} = \alpha(y|x) q(y|x) P(x)$

$y \neq x$

$= \min\left\{1, \frac{P(y)}{\cancel{P(x)}} \frac{q(x|y)}{\cancel{q(y|x)}}\right\} \cdot q(y|x) \cancel{p(x)}$

if $q(x|y) = q(y|x)$, this becomes the METROPOLIS criterion

$= \min\{q(y|x)p(x), q(x|y)p(y)\}$

$= \min\left\{\frac{q(y|x)}{q(x|y)} \frac{p(x)}{p(y)}, 1\right\} q(x|y)p(y) = \alpha(x|y) q(x|y) P(y) =$

$= \boxed{P(x|y) P(y)}$

detailed balance condition holds.

GIBBS SAMPLING

$p(x) = p(x_1, \dots, x_n) \qquad x = (x_1, \dots, x_n)$

ASSUMPTION: we can <u>sample from 1D conditionals</u>

$p(x_i \mid x_{\setminus i}) \qquad x_{\setminus i} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$

$$\boxed{(X^{(t)})_{t\geq 0}}$$

1) Pick $k \in \{1, \dots, n\}$ — ROUND-ROBIN STRATEGY / UNIFORMLY AT RANDOM

2) Set $x_j^{(t+1)} = x_j^{(t)}$ for $j \neq k$

3) Sample $x_k^{(t+1)} \sim p(x_k \mid x_{\setminus k}^{(t)})$

ISSUES

- ERGODICITY
- IF VARIABLE are STRONGLY CORRELATED, CONVERGENCE is SLOW

$$q_k(y \mid x) = \begin{cases} p(y_k \mid x_{\setminus k}) & \text{, when } y_{\setminus k} = x_{\setminus k} \\ 0 & \text{, otherwise} \end{cases}$$

$\Rightarrow \alpha_k(y \mid x) = 1$ acceptance probability

$$\boxed{x_{\setminus k} = y_{\setminus k}}$$

$$\text{MH } \alpha_k: \frac{p(y)\, q_k(x \mid y)}{p(x)\, q_k(y \mid x)} = \frac{\cancel{p(y_k \mid y_{\setminus k})}\, \cancel{p(y_{\setminus k})}}{\cancel{p(x_k \mid x_{\setminus k})}\, \cancel{p(x_{\setminus k})}} \; \frac{\cancel{p(x_k \mid y_{\setminus k})}}{\cancel{p(y_k \mid x_{\setminus k})}} = 1$$

- SAMPLE BLOCKS $x_j, x_k \subseteq x$
- IF $p(x_i \mid x_{\setminus i})$ is NOT KNOWN, then
  - Rejection sampling
  - METROPOLIS_WITHIN_GIBBS

CONVERGENCE DIAGNOSTICS

OUTPUT $\psi : \mathcal{X} \to \mathbb{R}$, $\Psi(x)$. we assume that Ψ has values in $\mathbb{R}$, and transform it otherwise.

$x_1 \_\_\_ x_N \dots$

$\psi_j = \Psi(x_j)$ (NOTATION)

$$\boxed{\bar{\psi} = \frac{1}{N} \sum_j \psi_j}$$ estimate of $\mathbb{E}_p[\Psi] = \int \Psi(x)\, p(x)\, dx$

- WE NEED MC TO BE STATIONARY
  (KEEP SAMPLES ONLY WHEN STATIONARY)

- SAMPLE $\frac{m}{2} \geq 1$ TRAJECTORIES FROM OVER-DISPERSED INITIAL POINTS
- SAMPLE FOR $4n$ STEPS
- THROW AWAY FIRST HALF $\Rightarrow$ we have $2n$ points left (BURN-IN / WARM UP PHASE)
- WE SPLIT THE REMAINED TRAJ IN TWO:

$\boxed{m \text{ SEQUENCES OF LENGTH } n \text{ EACH}}$

x_{ij} — sample — sequence

$1 \leq j \leq m$

$1 \leq i \leq w$

$\psi(x_{ij}) = \psi_{ij}$

$$\begin{cases} \bar{\psi}_{\cdot j} = \frac{1}{w}\sum_{i=1}^{w} \psi_{ij} \\ \bar{\psi} = \frac{1}{m}\sum_{j=1}^{m} \bar{\psi}_{\cdot j} \end{cases}$$

$\mathrm{VAR}(\psi)$; $\mathrm{VAR}(\bar{\psi})$

[ψ is a r.v. $\psi(x)$]

$$W = \frac{1}{m}\sum_{j=1}^{m} s_j^2, \quad s_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\psi_{ij} - \bar{\psi}_{\cdot j}\right)^2$$

WITHIN VARIANCE

$$W \leq \mathrm{VAR}(\psi)$$

$$B = \frac{n}{m-1}\sum_{j=1}^{m}\left(\bar{\psi}_{\cdot j} - \bar{\psi}\right)^2$$ – BETWEEN VARIANCE

$$\mathrm{VAR}(\psi) \leq \mathrm{VAR}^{+}(\psi) = \frac{n-1}{n}W + \frac{1}{n}B$$

$\hat{R} = \sqrt{\mathrm{VAR}^{+}(\psi)/W}$: $\hat{R} > 1$, $\hat{R} \xrightarrow[n\to\infty]{} 1$, when $\hat{R} \leq 1.1$ then converged

EFFECTIVE SAMPLE SIZE

$$\mathrm{VAR}[\bar{\Psi}]$$

if ~~m·n samples are independent~~, then $\mathrm{VAR}(\bar{\Psi}) = \dfrac{\mathrm{VAR}(\Psi)}{n \cdot m}$

$$n \cdot m \, \mathrm{VAR}(\bar{\Psi}) \simeq \left(1 + 2\sum_{k=1}^{\infty} \rho_k\right) \mathrm{VAR}(\Psi)$$

ρ_k IS THE AUTOCORRELATION OF LAG K:

$$\rho_k = \mathrm{CORR}\left[\Psi(x_i), \Psi(x_{i+k})\right].$$

$$n_{eff} = \frac{n \cdot m}{\left(1 + 2\sum_{k=1}^{\infty} \rho_k\right)} \qquad \mathrm{VAR}(\bar{\Psi}) = \frac{\mathrm{VAR}(\Psi)}{n_{eff}}$$

$$E\left[(\psi_i - \psi_{i-k})^2\right] = 2(1-\rho_k)\,\mathrm{VAR}(\psi)$$

$$V_k = \frac{1}{m(n-k)} \sum_{j=1}^{m} \sum_{i=k+1}^{n} (\psi_{i,j} - \psi_{i-k,j})^2$$

VARIOGRAM AT LAG K

$$\boxed{\hat{\rho}_k = 1 - \frac{V_k}{2\,\mathrm{VAR}^+(\psi)}}$$

FOR LARGE K we have few samples => very noisy estimates

$$T = \min\{K \mid K \text{ is odd},\ \hat{\rho}_{K+1} + \hat{\rho}_{K+2} < 0\}$$

$$\sum_{K=1}^{\infty} \rho_K \approx \boxed{\sum_{K=1}^{T} \hat{\rho}_K}$$

$n_{eff} \simeq 100$