

SAMPLING AND MCMC

$$E[f] = \int_{\mathcal{X}} f(x) p(x) dx$$

$p(y, x)$ joint

$$p(y) = \int p(y, x) dx = \int \underbrace{p(y|x)}_{f(x)} p(x) dx = E[p(y|x)]$$

Sample $\underbrace{x_1 \dots x_N}_{\text{INDEPENDENT}}$ from $p(x)$ and compute

$$\boxed{\hat{E}_N[f]} = \frac{1}{N} \sum_{i=1}^N f(x_i) \xrightarrow{N \rightarrow \infty} E[f]$$

is a random variate as it depends on x_1, \dots, x_N .

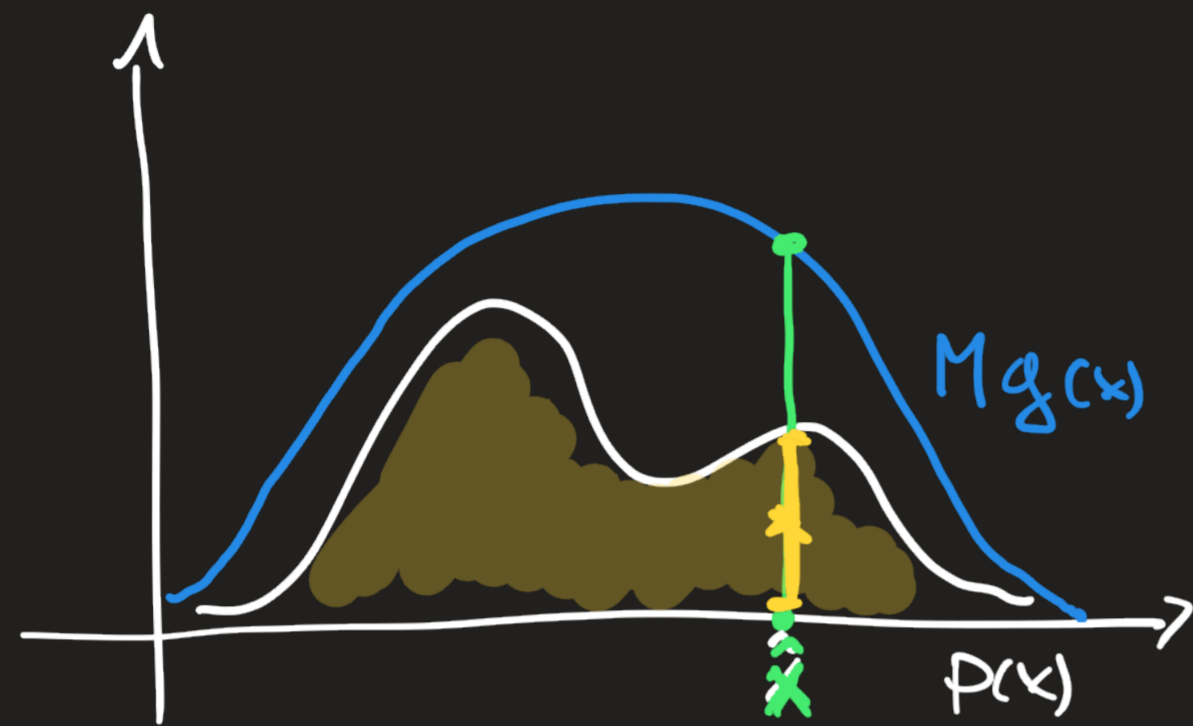
$$p(x) = \frac{1}{Z} \underbrace{\tilde{p}(x)}_{\text{computable}}$$

non-computable

$$\left(p(x|y) = \frac{p(y|x)p(x)}{\rightarrow \boxed{p(y)}} \right)$$

TWO STRATEGIES: - sample from $q \approx p$ (similar) and correct on
- MCMC. \Leftarrow MARKOV CHAIN MONTE CARLO

REJECTION SAMPLING



use $q(x)$ PROPOSAL DISTRIBUTION

$$\exists M > 0 : Mq(x) \geq p(x), \forall x \in \mathcal{X}$$

$$\left(\Rightarrow \frac{p(x)}{Mq(x)} \leq 1 \right)$$

1) SAMPLING \hat{x} FROM $q(x)$

2) ACCEPT \hat{x} WITH PROBABILITY $\alpha \leq \frac{p(\hat{x})}{Mq(\hat{x})}$

IF WE REJECT, WE REPEAT UNTIL ACCEPTANCE.

EXPECTED # of SAMPLES FROM q PER p -SAMPLE IS M

INEFFICIENT IN HIGH DIMENSIONS.

IMPORTANCE SAMPLING

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

$$\begin{aligned} E_p[f] &= \int f(x) \frac{1}{Z} \tilde{p}(x) dx \\ &= \frac{\int f(x) \tilde{p}(x) dx}{\int \tilde{p}(x) dx} \end{aligned}$$

$\frac{g(x)}{g(x)}$
 $\frac{g(x)}{g(x)}$

CONSIDER A PROPOSAL DISTRIBUTION g , which we can sample from

$$\begin{aligned} E_p[f] &= \int f(x) p(x) dx = \int \boxed{f(x) \frac{p(x)}{g(x)}} g(x) dx = E_g \left[f \frac{p}{g} \right] \\ &= \frac{\int f(x) \frac{\tilde{p}(x)}{g(x)} g(x) dx}{\int \frac{\tilde{p}(x)}{g(x)} g(x) dx} \end{aligned}$$

We sample x_1, \dots, x_N from g

$$x_1, \dots, x_N \sim q(x)$$

$$\frac{\frac{1}{N} \sum_{i=1}^N f(x_i) \omega(x_i)}{\frac{1}{N} \sum_{i=1}^N \omega(x_i)}$$

$$\text{IMPORTANCE WEIGHTS } \omega(x_i) := \frac{\tilde{p}(x_i)}{q(x_i)}$$

IF $\frac{\tilde{p}}{q}$ is approx constant, Then estimates can be very good.

IF WEIGHTS vary a lot \Rightarrow large variance, poor estimate

$$\frac{1}{N} \sum \omega(x_i) \approx Z$$

MARKOV CHAINS

$$(X_t)_{t \geq 0}, t \in \mathbb{N} \quad X_0, X_1, X_2, \dots \quad X_i \in \mathcal{X}$$



$$\bullet P(X_n | X_{n-1}, \dots, X_0) = P(X_n | X_{n-1})$$

MARKOV or MEMORYLESS PROPERTY

$$\bullet P(X_n | X_{n-1}) = P(X_2 | X_0), \quad \forall n \geq 1$$

TIME HOMOGENEITY

ERGODIC $\forall x, y \in \mathcal{X}, \exists t \geq 0: P(X_t = y | X_0 = x) > 0$



$P(y|x)$ is TRANSITION KERNEL

STATIONARY DISTRIBUTION

$$\Pi(y) = \int P(y|x) \Pi(x) dx$$

Π is invariant for the MC dynamics

$$P_n(x) \equiv P(X_n = x | X_0) \xrightarrow{n \rightarrow \infty} \Pi(x), \quad \Pi \text{ is unique}$$

REVERSIBLE MARKOV CHAINS

$(X_t)_{t \geq 0}$ transition kernel

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

A M.C. IS REVERSIBLE (SATISFIES THE BALANCE CONDITION)
[π distribution s.t.

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

Then π IS STATIONARY

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

MCMC

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

PROPOSAL KERNEL

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

$$\rightarrow x_t = x$$

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

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

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

Transition Kernel of MC

$$P(y|x) P(x) = \alpha(y|x) q(y|x) P(x)$$

METROPOLIS-HASTINGS ACCEPTANCE CRITERION

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

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

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

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

$$= P(x|y) P(y)$$

detailed balance condition for MC.

GIBBS SAMPLING

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

ASSUMPTION: we can sample from 1D conditionals

$$p(x_i | x_{-i}) \quad x_{-i} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$$

$$(x^{(t)})_{t \geq 0}$$

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 $x_k^{(t+1)} \sim p(x_k | x_{-k}^{(t)})$

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

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

$$x_{-k} = y_{-k}$$

$$\text{Met } \alpha_k: \frac{p(y) q_k(x|y)}{p(x) q_k(y|x)} = \frac{\cancel{p(y_k | y_{-k})} p(y_{-k})}{\cancel{p(x_k | x_{-k})} p(x_{-k})} \frac{\cancel{p(x_k | y_{-k})}}{\cancel{p(y_k | x_{-k})}} = 1$$

• SAMPLE BLOCKS $x_j \rightarrow x_k \subseteq x$

• IF $p(x_k | x_{-k})$ IS NOT KNOWN, THEN
 - REJECTO-SAMPLING
 - METROPOLIS WITHIN-GIBBS

ISSUES

- ERGODICITY
- IF VARIABLES ARE STRONGLY CORRELATED, CONVERGENCE IS SLOW

CONVERGENCE DIAGNOSTICS

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

$x_1 \dots x_N \dots$

$$\psi_j = \psi(x_j) \text{ (NOTATION)}$$

$$\bar{\psi} = \frac{1}{N} \sum_j \psi_j \text{ estimate of } \mathbb{E}_{\pi}[\psi] = \int \psi(x) p(x) dx$$

- 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 $4M$ STEPS
- THROW AWAY FIRST HALF \Rightarrow we have $2M$ points left (BURN-IN WARM UP PHASE)
- WE SPLIT THE REMAINED TRAJ IN TWO:

M SEQUENCES OF LENGTH M EACH

x_{ij} $1 \leq j \leq m$ $1 \leq i \leq n$ $\psi(x_{ij}) = \psi_{ij}$

↑ sample ↑ sequence

$$\left\{ \begin{array}{l} \bar{\psi}_{\cdot j} = \frac{1}{n} \sum_{i=1}^n \psi_{ij} \\ \bar{\psi} = \frac{1}{m} \sum_{j=1}^m \bar{\psi}_{\cdot j} \end{array} \right. \leftarrow$$

$\rightarrow \text{VAR}(\psi) ; \text{VAR}(\bar{\psi})$
 $[\psi \text{ is a r.v. } \psi(x)]$

$\omega = \frac{1}{m} \sum_{j=1}^m s_j^2$, $s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (\psi_{ij} - \bar{\psi}_{\cdot j})^2$
 WITHIN VARIANCE

$\omega \leq \text{VAR}(\psi)$

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

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

$\hat{R} = \sqrt{\text{VAR}^+(\psi) / \omega}$; $\hat{R} \geq 1$, $\hat{R} \xrightarrow{n \rightarrow \infty} 1$, when $\hat{R} \leq 1.4$ then CONVERGES

EFFECTIVE SAMPLE SIZE

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

if ~~$m \cdot n$ samples are independent~~, then $\text{VAR}(\bar{\Psi}) = \frac{\text{VAR}(\Psi)}{n \cdot m}$ σ

$$n \cdot m \text{ VAR}(\bar{\Psi}) \approx \left(1 + 2 \sum_{k=1}^{\infty} \rho_k \right) \text{VAR}(\Psi) \quad \sigma$$

ρ_k IS THE AUTOCORRELATION OF u AT k :

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

$$n_{\text{eff}} = \frac{n \cdot m}{\left(1 + 2 \sum_{k=1}^{\infty} \rho_k \right)}$$

$$\text{VAR}(\bar{\Psi}) = \frac{\text{VAR}(\Psi)}{n_{\text{eff}}}$$

$$E[(\Psi_i - \Psi_{i-k})^2] = 2(1 - \rho_k) \text{VAR}(\Psi)$$

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

VARIOGRAM AT LAG k

$$\hat{\rho}_k = 1 - \frac{V_k}{2 \text{VAR}^+(\Psi)}$$

FOR LARGE k we have few samples \Rightarrow very noisy estimates

$\bar{k} = \min\{k \mid k \text{ is odd, } \hat{\rho}_{k+1} - \hat{\rho}_{k+2} \approx 0\}$

$$\sum_{k=1}^{\infty} \rho_k \approx \sum_{k=1}^{\bar{k}} \hat{\rho}_k$$

$$n_{\text{eff}} \approx 100$$

HAMILTONIAN MONTE CARLO

$$p(x) = \frac{1}{Z_x} \tilde{p}(x) = \frac{1}{Z_x} \exp(H_x(x))$$

• INTRODUCE MOMENTUM VARIABLES y $|y| = |x|$

$$p(y) = \frac{1}{Z_y} \exp(H_y(y)), \quad H_y(y) = -\frac{1}{2} y^T y \Rightarrow p(y) \text{ STANDARD GAUSSIAN}$$

$$p(x, y) = p(x) p(y) = \frac{1}{Z_x Z_y} \exp(H_x(x) + H_y(y)) = \frac{1}{Z} \exp(\underbrace{H(x, y)}_{\text{HAMILTONIAN}})$$

SAMPLE from $p(x, y)$ and forget y .

—

WE ARE IN POINT x_i

1) SAMPLE $y \sim p(y)$

2) WE CHOOSE A RANDOM DIRECTION IN TIME $(1, -1)$

3) WE MOVE ACCORDING TO H.d FROM (x_i, y) TO A CANDIDATE (x', y') doing L steps

4) M-H ACCEPTANCE: accept if $H(x', y') > H(x, y)$, otherwise accept with prob $\exp(H(x', y') - H(x, y))$

$$H(x', y') = H(x, y)$$

$$(x', y') = (x + \Delta x, y + \Delta y)$$

F.O. TAYLOR EXPANSION

$$H(x + \Delta x, y + \Delta y) \approx H(x, y) + \underbrace{\nabla_x H_x(x)^T \Delta x + \nabla_y H_y(y)^T \Delta y}_{=0}$$

$$\Delta x = \varepsilon \nabla_y H_y(y) = -\varepsilon y$$

$$\Delta y = -\varepsilon \nabla_x H_x(x)$$

($\varepsilon = +\varepsilon_0$ or $-\varepsilon_0$ with prob $\frac{1}{2}$)

We do L steps of this dynamics to get the final (x', y')

$$M.H. d := \min \left\{ 1, \frac{P(x', y')}{P(x, y)} \right\}$$

$$\hookrightarrow \exp(H(x', y') - H(x, y))$$

There are better integration schemes:
LEMP FROG integration