Approximate marginalization / Bayesian updates / inference
- Variational method
- Sampling methods

ORIE 4742 - Info Theory and Bayesian ML

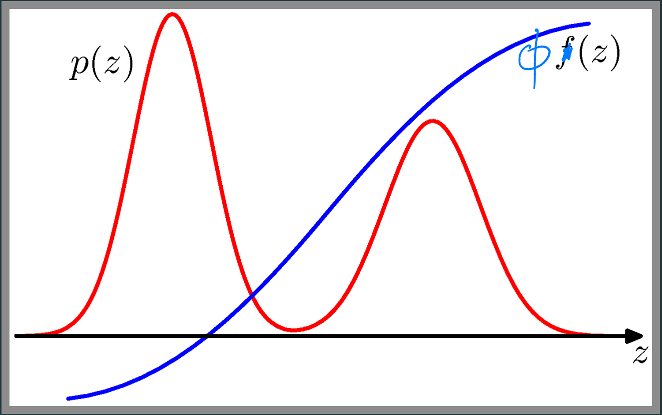
Monte Carlo Techniques

April 16, 2020

Sid Banerjee, ORIE, Cornell

Today's lecture - Bishop Ch 11
($\approx$ Mackay Ch 29)

why Monte Carlo?



- Task - compute $\mathbb{E}[\phi(X)]$, $X \sim F$
- Approximation - $X_1, X_2, \ldots, X_L \sim F$, $\mathbb{E}[\phi(X)] \approx \frac{1}{L} \sum_{i=1}^{L} \phi(X_i)$

monte carlo basics

- primitive: rand()

Any simulation / randomized algo $\equiv$ regular algorithm + fn to generate $\underline{U[0,1]}$

- pseudorandomness and seed

- $X_1, X_2, \ldots X_L = \boxed{g(\text{seed})}$ deterministic fn whose outputs 'look random'

- advantage. calculations are repeatable (ie, any calculation is approx the same)

- estimate variance

$X \sim F$ · $\underbrace{\mathbb{E}[\phi(X)]}_{\mu} = \mathbb{E}\Big[\underbrace{\frac{1}{L}\sum_{i=1}^{L}\phi(X_i)}_{Z_L}\Big] = \frac{1}{L}\sum_{i=1}^{L}\mathbb{E}[\phi(X)]$ if $X_i \sim F$

assume X_i are ⫫ · $\mathrm{Var}(Z_L) = \frac{1}{L}\mathbb{E}[(\mu - \phi(x_i))^2]$

(ie, $Z_L \in \mu \mp 2\sqrt{\mathrm{Var}(Z_L)}$ wp 95%)

- use CIs to indicate how good your estimate is

- issues: non-independence, 'rare-events'

- MCMC - $X_i \sim F$, dependent

- ϕ may be very large for low probability X

monte carlo sampling techniques

basic methods (iid samples)

last class
- inversion
- distribution-specific techniques (Box-Muller for Gaussians)
- rejection sampling
- importance sampling
- advanced techniques (adaptive rejection sampling, SIR)

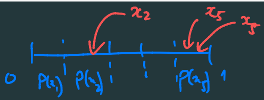
markov-chain monte carlo (MCMC)

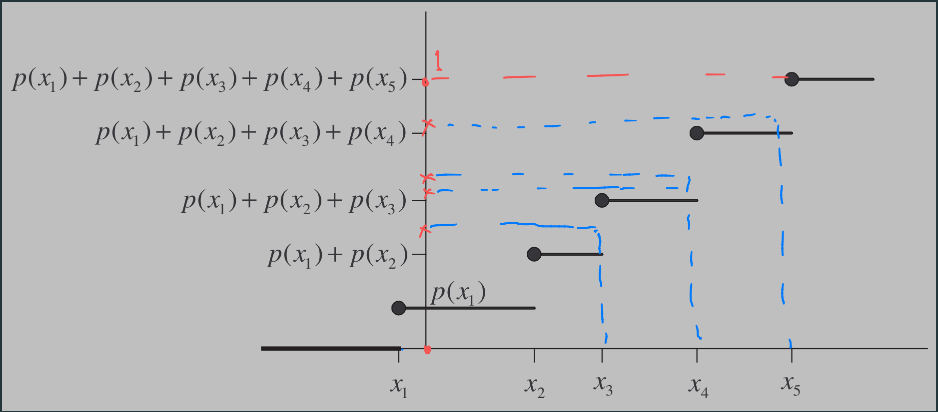
today
- Metropolis-Hastings
- Gibbs sampling
- advanced techniques (slice sampling, Hamiltonian MC)

basic monte carlo techniques

warmup: simulating discrete rv

X takes values $x_1 \leq x_2 \leq \ldots \leq x_5$, $\mathbb{P}[X = x_i] = p(x_i)$





the inversion method

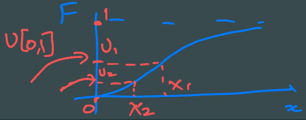
X continuous r.v. with pdf f and c.d.f. $F(\cdot)$

- want to generate samples of X.
- $F(\cdot)$ non-decreasing $\implies$ can define inverse $F^{-1}(\cdot)$
- $F(x) = u \iff F^{-1}(u) = x$

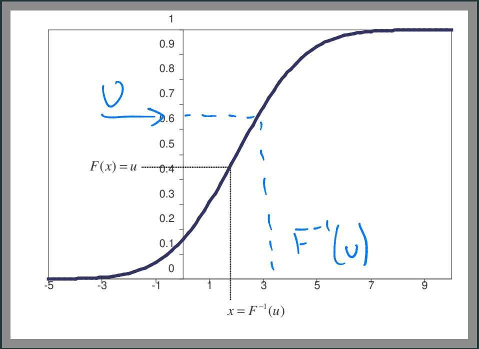
inversion method

given desired cdf F (continuous, increasing), generate sample $X_0 \sim F$ as:

1. generate $U \sim U[0, 1]$.
2. return $X_o = F^{-1}(U)$.



intuition/proof for inversion method



$X \sim F^{-1}(U)$

$$\mathbb{P}[X \le x] = \mathbb{P}[F^{-1}(U) \le x]$$
$$= \mathbb{P}[U \le F(x)]$$
$$= F(x)$$

$\Rightarrow$ CDF of X is F

- Problem - $F^{-1}(\cdot)$ may not be simple

inversion method example

generate samples of an exponential r.v. with parameter λ, with cdf

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

$\Rightarrow \quad F^{-1}: \quad y = 1 - e^{-\lambda x} \Rightarrow \underline{x = -\frac{1}{\lambda} \ln(1-y)}$

$\Rightarrow \quad X = -\frac{1}{\lambda} \ln(1 - U) \sim \mathrm{Exp}(\lambda)$

$\sim U[0,1]$

Box-Muller method for Gaussian r.v.

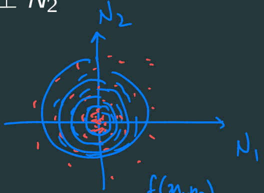
- generate $N_1 \sim \mathcal{N}(0,1)$, $N_2 \sim \mathcal{N}(0,1)$, $N_1 \perp\!\!\!\perp N_2$
- in polar coordinates $(N_1, N_2) = (R\cos\theta, R\sin\theta)$

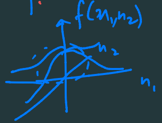
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-x^2/2} e^{-y^2/2}\, dx\, dy = 1$$

$$x = r\cos\theta, \quad y = r\sin\theta$$

$$\Rightarrow \int_0^{\infty}\int_0^{2\pi} \underbrace{\frac{1}{2\pi}}_{f(\theta)} \cdot \underbrace{r e^{-r^2/2}}_{f(r)}\, dr\, d\theta = 1$$

$$= \left(\int_0^{2\pi} \frac{d\theta}{2\pi}\right)\left(\int_0^{\infty} \underbrace{r e^{-r^2/2}} dr\right)$$





$$\Theta \sim \mathrm{Unif}[0, 2\pi]$$

$$\frac{R^2}{2} \sim \mathcal{E}xp(1)$$

the Box-Muller Method

$$(N_1, N_2) = (R\cos\theta, R\sin\theta)$$, $N_1 \perp\!\!\!\perp N_2$

$\theta \sim U[0, 2\pi]$, and independent of R.

$R = \sqrt{N_1^2 + N_2^2} = \sqrt{2X}$, where $X \sim Exp(1)$

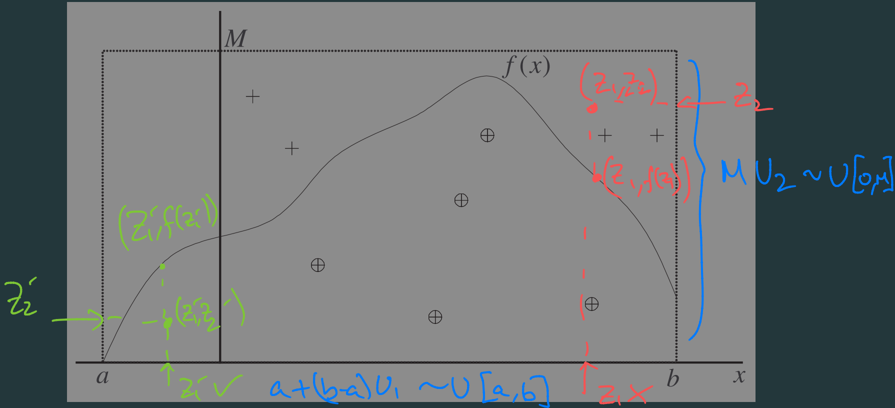
- Generate $\frac{R^2}{2} = -\ln(1-U_1)$, $\Theta = 2\pi U_2$
- Set $N_1 = R\cos\Theta$, $N_2 = R\sin\Theta$

rejection sampling

want samples of a rv $X \in [a, b]$, with pdf $f(x) \leq M$

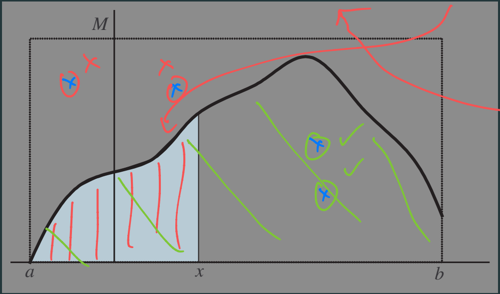
rejection sampling

1. Generate $U_1, U_2 \sim U[0,1]$, and set $Z_1 = a + (b-a)U_1$, $Z_2 = MU_2$
2. if $Z_2 \leq f(Z_1)$, return $X_o = Z_1$; else, reject and repeat



rejection sampling: proof of correctness

observe: $\mathbb{P}[Z_1 \leq x,\ Z_2 \leq f(Z_1)] = \mathbb{P}[(Z_1, Z_2) \in \text{shaded region in figure}]$



$\mathbb{P}[Z_1 \leq x \mid Z_1 \text{ accepted}]$

$= \dfrac{\mathbb{P}[Z_1 \leq x \text{ AND } Z_1 \text{ accepted}]}{\mathbb{P}[Z_1 \text{ accepted}]}$

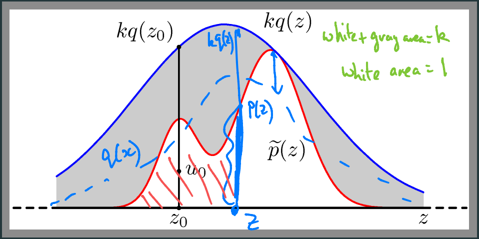
$= \dfrac{\text{shaded area in red}}{\text{shaded area in green}}$

$= F(x)$

$\mathbb{P}[Z_1 \text{ accepted}] \propto$ green shaded area

$\mathbb{P}[Z_1 \text{ accepted}, Z_1 \leq x] \propto$ red shaded area

generalized rejection sampling

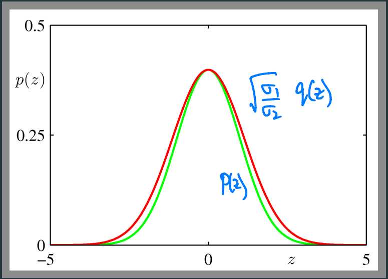


- Suppose we have a Sampler for $Z \sim Q$
- Want - $X \sim P$
- Suppose $\max_x \frac{p(x)}{q(x)} \leq k \ (\geq 1)$

$\Rightarrow \quad p(x) \leq kq(x) \ \forall x \in \mathbb{R}$

- Algo - generate $Z_1 \sim Q$.
  - accept Z_1 with probability $\frac{p(Z_1)}{kq(Z_1)}$, else reject

  (equiv, generate $Z_2 = kq(Z_1)\, U_2$, accept if $Z_2 < p(Z_1)$)



Eg - $Z \sim N(0,\sigma_1)$, want $X \sim N(0,\sigma_2)$

- $q(z) = \frac{1}{\sqrt{2\pi\sigma_1}} e^{-z^2/2\sigma_1}$, $p(z) = \frac{1}{\sqrt{2\pi\sigma_2}} e^{-z^2/2\sigma_2}$

- $$\frac{p(z)}{q(z)} \leq \sqrt{\frac{\sigma_1}{\sigma_2}}$$

$\Rightarrow Z \sim N(0,\sigma_1)$, accept w.p. $\frac{e^{-z^2/2\sigma_2}}{e^{-z^2/2\sigma_1}} \cdot \sqrt{\frac{\sigma_1}{\sigma_2}}$

- $\boxed{\mathbb{P}[\text{Acceptance}] = \frac{1}{k}} = \sqrt{\frac{\sigma_2}{\sigma_1}} \Rightarrow \mathbb{E}[\#\text{ of samples for 1 acceptance}]$
$= k = \sqrt{\sigma_1/\sigma_2}$

- Q: What if $Z \sim N(0,\sigma_1 I_d)$, $X \sim N(0,\sigma_2 I_d)$

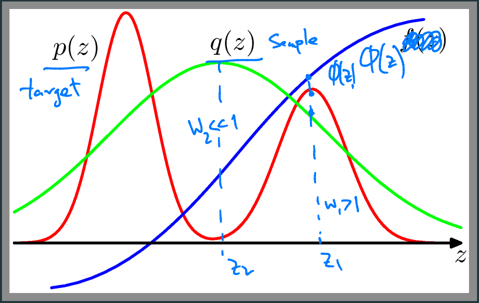
$\Rightarrow \mathbb{P}[\text{Acceptance}] = 1/k^d = \left(\frac{\sigma_2}{\sigma_1}\right)^{d/2} \ll 1 \Rightarrow$ Need many samples

importance sampling

- given function $\phi(\cdot)$, want $\mathbb{E}[\phi(X)]$ where $X \sim P$
- can generate samples $Z \sim Q$

importance sampling

1. generate $Z_1, Z_2, \ldots, Z_L \sim Q$
2. compute $\mathbb{E}[\phi(X)] = \frac{1}{L}\sum_{i=1}^{L} \underbrace{w_i\phi(Z_i)}_{w(z_i)}$, where $\underline{w_i = p(Z_i)/q(Z_i)}$ importance sampling weigts



$$\mathop{\mathbb{E}}_{P}[\phi(X)] = \int_{\mathbb{R}^d} \phi(x) p(x)\, dx$$

$$= \int_{\mathbb{R}^d} \underbrace{\phi(x) \cdot \left(\frac{p(x)}{q(x)}\right) q(x)}\, dx$$

$$= \mathop{\mathbb{E}}_{Q}\left[\underbrace{W(Z)}\phi(Z)\right]$$

$$\approx \frac{1}{L}\sum_{i=1}^{L} \left(\frac{p(z_i)}{q(z_i)}\right)\phi(z_i)$$

importance sampling: unknown normalization

- Suppose $p(x) = \frac{1}{Z_p}\tilde{p}(x)$, $q(x) = \frac{1}{Z_Q}\tilde{q}(x)$, $Y_i \sim Q$

- $$\mathbb{E}[\phi(X)] = \int \phi(x)\frac{\tilde{p}(x)}{Z_p}\,dx$$

$$= \left(\frac{Z_Q}{Z_p}\right)\int \phi(x)\underbrace{\left(\frac{\tilde{p}(x)}{\tilde{q}(x)}\right)}_{w(x)=\tilde{p}(x)/\tilde{q}(x)}\cdot\frac{\tilde{q}(x)}{Z_Q}\,dx$$

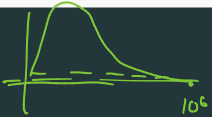
$$= (Z_Q/Z_p)\,\mathbb{E}_Q[\phi(Y)\,w(Y)]$$

- $$1 = \int p(x)\,dx = \frac{Z_Q}{Z_p}\int\frac{\tilde{p}(x)}{\tilde{q}(x)}q(x)\,dx = \frac{Z_Q}{Z_p}\mathbb{E}[w(Y)]$$

$$\Rightarrow \mathbb{E}[\phi(X)] = \frac{\mathbb{E}_Q[\phi(Y)\,w(Y)]}{\mathbb{E}[w(Y)]} \approx \sum_{i=1}^{L}\frac{w_i}{\sum_j w_j}\cdot\phi(Y_i)$$

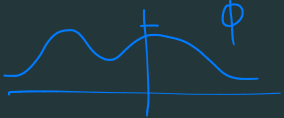
importance sampling: comments

$$\mathbb{E}[\phi(x)] \approx \frac{1}{L}\sum \underbrace{\frac{p(z_i)}{q(z_i)}}_{\gg 1} \underbrace{\phi(z_i)}_{\phi(x) \leq 1}$$



$$\frac{1}{L}\mathrm{Var}\big(\underbrace{\phi(z_i)\,w(z_i)}\big)$$

depends on max $\phi(z_i)w(z_i)$



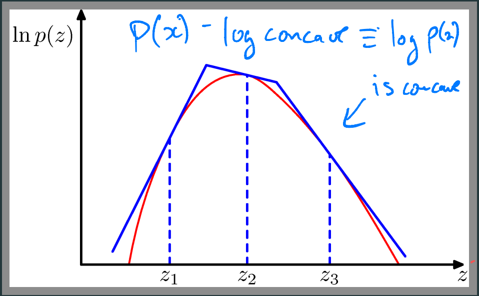
- Problem – If $p(z)/q(z)$ is large for some z, then high variance

- Practice – choose q s.t. $\boxed{\dfrac{p(z)}{q(z)}} \cdot \boxed{\phi(z)}$ is not too large

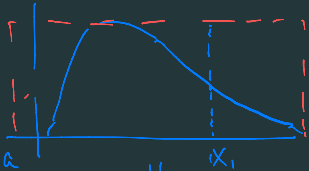
want this to be small

$1/10^6$

(advanced) adaptive rejection sampling

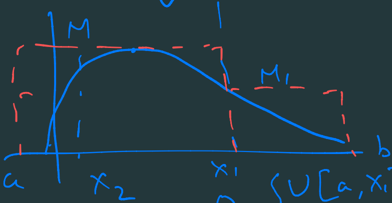


Idea - Keep improving the 'proposal dist' Q



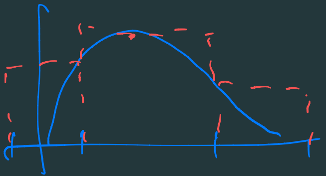
$Q \sim U[a,b]$

$\Downarrow$



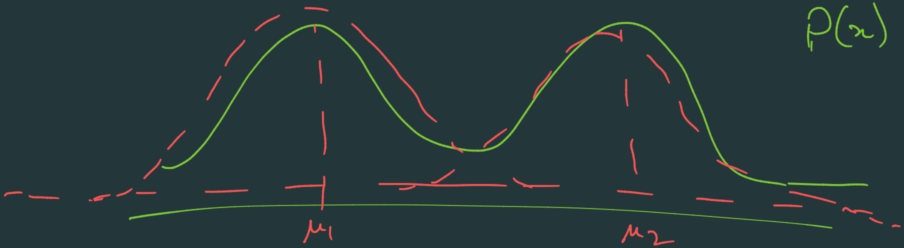
$Q \sim \begin{cases} U[a, x_1]. \\ U[x_1, b] \end{cases}$

$\Leftarrow$

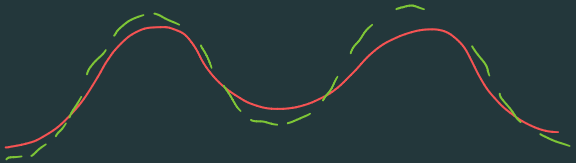


- $R = \dfrac{\text{Area under red}}{\text{Area under blue}}$ ↘

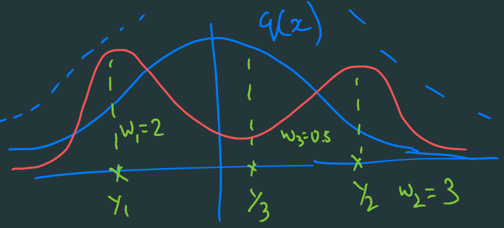
Eg



$$q = \frac{0.9}{2} N(\mu_1, 1) + \frac{0.9}{2} N(\mu_2, 1) + 0.1\, U[-10^6, 10^6]$$



(advanced) sampling-importance-resampling (SIR/bootstrap)



- SIR heuristic
  - Generate $Y_1, Y_2, \ldots, Y_L \sim Q$
  - Compute $W_i = \dfrac{p(Y_i)}{q(Y_i)}$
  - Resample $\underbrace{\tilde{Y}_1, \tilde{Y}_2, \ldots, \tilde{Y}_L}$ from dist $\tilde{P} \sim \{ Y_i$ wp $W_i / \Sigma W_i \}$

- Eg - $\tilde{P} = \begin{cases} Y_1 & \text{wp } 2/5.5 \\ Y_2 & \text{wp } 3/5.5 \\ Y_3 & \text{wp } 0.5/5.5 \end{cases}$

  - Claim - As $L \to \infty$, $\tilde{P} \to P$

disadvantage

$\tilde{P} \approx P$

(not exact samples)

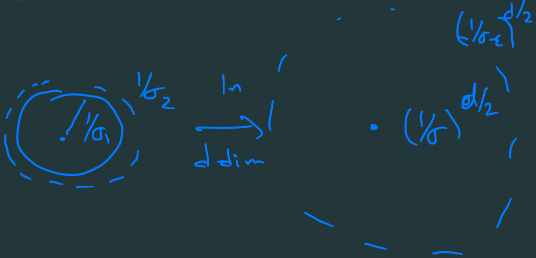
advantage

(no sample rejected)

Aside

$N(0, \sigma I_d) \approx$ Unif dist on sphere of radius $1/\sigma^{d/2}$ in $\mathbb{R}^d$

$Z \sim N(0, \sigma_1 I_d)$, $X \sim N(0, (\sigma-\varepsilon) I_d)$



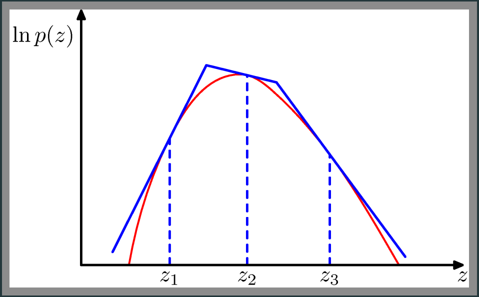
the Box-Muller Method

$$(N_1, N_2) = (R\cos\theta, R\sin\theta)$$

$\theta \sim U[0, 2\pi]$, and independent of R.

$R = \sqrt{N_1^2 + N_2^2} = \sqrt{2X}$, where $X \sim Exp(1)$

(advanced) adaptive rejection sampling



markov chain monte carlo

MCMC: the basic idea

- Method 1 - Generate $X_1, X_2, \ldots X_L$ iid from P
  (Inversion sampling, Importance sampling)
- Method 2 - Generate $X_1, \ldots, X_L$ iid, accept/reject samples

  $X_1, X_2, \cancel{X_3}, \ldots, X_L \longrightarrow Y_1, Y_2, Y_4 \ldots Y_L$
- Method 3 - Generate X_1, $X_2 = f(X_1)$, $X_3 = f(X_2)$, ...
  - Accept/Reject X_t w.p. $A(X_{t-1}, X_t)$
    (ie, set $Y_t = X_t$ or set $Y_t = X_{t-1}$)

$$\underset{(Y_0)}{X_0} \longrightarrow \underset{(Y_1)}{X_1 = f(X_0)} \xrightarrow{\times} X_2 = f(X_1) \qquad \underset{(Y_2)}{X_1} \longrightarrow \underset{(Y_3)}{X_3 = f(X_1)} \ldots$$

markov chains: basic definition

Seq of r.v. $X_1, X_2, \ldots, X_n$ is a MC

iff $$P(X_i \mid X_1, X_2, \ldots, X_{i-1}) = P(X_i \mid X_{i-1})$$

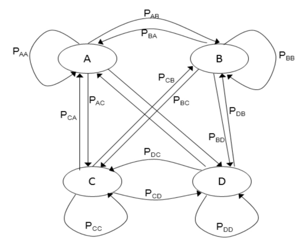
BayesNet for MC

$$X_1 \to X_2 \to X_3 \to \cdots \to X_n$$

- Check - $$P(X_i, X_{i+2} \mid X_{i+1}) = P(X_i \mid X_{i+1}) \cdot P(X_{i+2} \mid X_{i+1})$$

(d-separation)

markov chains: state-space and transition matrix (time-invariant)



$$\begin{array}{cc} & X_{t+1} \\ X_t & \mathbf{P}=\begin{bmatrix} p_{AA} & p_{AB} & p_{AC} & p_{AD} \\ p_{BA} & p_{BB} & p_{BC} & p_{BD} \\ p_{CA} & p_{CB} & p_{CC} & p_{CD} \\ p_{DA} & p_{DB} & p_{DC} & p_{DD} \end{bmatrix} \end{array}$$

transition prob matrix

<u>Note</u> - This is not a BayesNet!

- State-space = $\{A, B, C, D\}$
- $P(X_{t+1}=A \mid X_t=B) = \underline{\underline{P_{BA}}}$

does not depend on X_{t-1}

- $\sum P_{Ax} = 1 \;\forall A$ ≡ Each row-sum of $P = 1$, $P_{ij} \geq 0 \;\forall ij$ (stochastic matrix)

x : outgoing edges

- Directed (not necessarily acyclic) graph

row vector

- Let $\pi_t \equiv$ Distribution of $X_t = P[X_t = x \mid X_0]$

Then $\pi_{t+1} = \pi_t \cdot P = [\;\pi_t\;]\begin{bmatrix} \cdot & P & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix}$

markov chains: steady-state

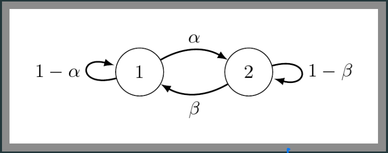
Q - Given $X_1, X_2, X_3 \dots$ from a time invariant MC, what can we say about $\lim_{t\to\infty} X_t$ (or X_t for large t)

1) X_t can get 'absorbed' (not interesting for us ...)

2) π_t can 'converge' to a fixed distrib π 'steady-state dist'

(recall $\pi_t = \pi_{t-1} P \Rightarrow \pi = \pi P$)

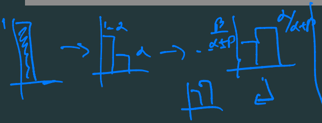
example:



$\pi_0 = (1, 0)$ (i.e., start at '1')

$$\pi_1 = (1 \;\; 0)\begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} = (1-\alpha \;\; \alpha)$$

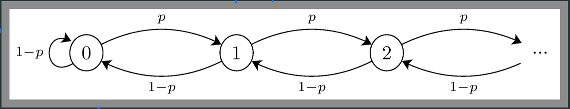
$\vdots$



$$\lim_{t\to\infty} \pi_t = \pi = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right)$$

markov chains: example (infinite Markov Chain)

the 1-d random walk on $\{0, 1, 2, ---\}$



$P[X_t = i+1 \mid X_{t-1} = i] = p$

$P[X_t = i-1 \mid X_{t-1} = i] = 1-p$

all else is 0

$(P[X_t = 0 \mid X_{t-1} = 0] = 1-p)$

$$P = \begin{pmatrix} 1-p & p & 0 & \cdots & \\ 1-p & & p & \cdots & 0 \\ 0 & & 1-p & p & \cdots \\ & & & \ddots & \end{pmatrix}$$

- Steady-state $\pi(i)$

$$\sum_j \pi(i) P_{ij} = \sum_j \pi(j) P_{ji} \quad \forall i$$

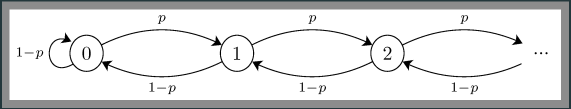
- Claim - $\pi(i) = \frac{1}{Z}\left(\frac{p}{1-p}\right)^i$ (assuming $p < 1/2$ $\Rightarrow Z = \sum (p/_{1-p})^i < \infty$)

'Proof': $\forall i > 0$, $\sum_j \pi(i) P_{ij} = \frac{1}{Z}\left(\frac{p}{1-p}\right)^i \cdot (p + 1-p)$

$$\sum_j \pi(j) P_{ji} = \frac{1}{Z}\left(\left(\frac{p}{1-p}\right)^{i-1} \cdot p + \left(\frac{p}{1-p}\right)^{i+1}(1-p)\right)$$

markov chains: reversibility

(idea 'borrowed' from physics)



- Steady-state $\pi(i)$ (Global balance)

$$\sum_j \pi(i) P_{ij} = \sum_j \pi(j) P_{ji} \quad \forall i \qquad (**)$$

- (Easier) Local balance - $\forall i,j$

$$\pi(i) P_{ij} = \pi(j) P_{ji} \qquad (*)$$

- 'Magic Thm' - If local balance holds $\forall i,j \Rightarrow \pi$ is steady-state dist

Pf - Check $(*) \Rightarrow (**)$

$$\pi(i) = \frac{1}{2}\left(\frac{p}{1-p}\right)^i$$

- Check local balance $(*)$

$\forall i,j$, - if $i \neq j-1$ or $j+1$, then $P_{ij} = P_{ji} = 0 \Rightarrow$ True

- if $i = j+1$

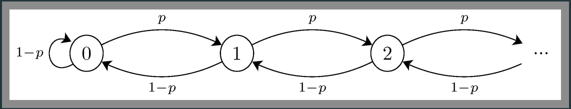
$$\frac{1}{2}\left(\frac{p}{1-p}\right)^i \cdot (1-p)$$

$\pi(i)$ P_{ij}

$$\frac{1}{2}\left(\frac{p}{1-p}\right)^{i-1} \cdot p$$

$\pi(i-1) = \pi(j)$ P_{ji}

Markov chains: the ergodic theorem



ergodic $\Rightarrow$ averages over time = averages over 'space'

- Let π be steady-state dist of MC with trans mat P (ie, $\pi = \pi P$)
- Then for any fn $\phi(x)$, $x \in S$ (state-space of MC),

$$\underset{X \sim \pi}{\mathbb{E}}\left[\phi(X)\right] = \lim_{T \to \infty} \frac{1}{T} \mathbb{E}\left[\sum_{t=0}^{T-1} \phi(X_t)\right]$$

space-average time average

where $X_0 \to X_1 \to X_2 \to \dots \to X_t$ are from the MC

Markov-chain monte carlo

Recall - Basic MCMC recipe

- Start at X_0
- At time t, compute $Y_t = f(X_{t-1})$
- Set $X_{t+1} = \begin{cases} Y_t & \text{w.p. } A(X_{t-1}, Y_t) \\ X_{t-1} & \text{otherwise} \end{cases}$

Some MC with trans mat R

- Compute $\underbrace{\frac{1}{L}\sum_{t=1}^{L}\phi(X_t)}$

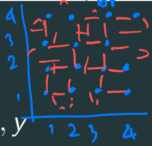
(want this to be $\mathbb{E}_{X\sim p}[\phi(X)]$)

- This requires the MC R to have steady-state dist as p (i.e., want $p = pR$)

the Metropolis algorithm

$Q \equiv$ move to random nbr

Eg -



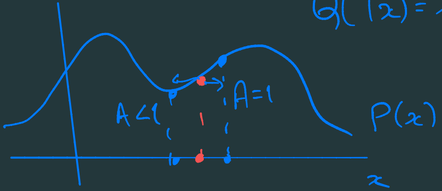
Symmetric

- target distribution $P(x) = \widetilde{P}(x)/Z$
- proposal distribution(s) $Q(x|y)$, with $Q(x|y) = Q(y|x)\,\forall\, x, y$

Metropolis sampling

1. choose initial Z_0
2. to obtain sample t, generate $Y_t \sim Q(\cdot|Z_{t-1})$
3. accept $Z_t = Y_{Y_t}$ with probability $A(Y_t, Z_{t-1}) = \max\left\{1, \frac{\widetilde{P}(Y_t)}{\widetilde{P}(Z_{t-1})}\right\}$
   else reject and set $Z_t = Z_{t-1}$

$Q(\ |x) = x \mp 1$ wp $\frac{1}{2}$

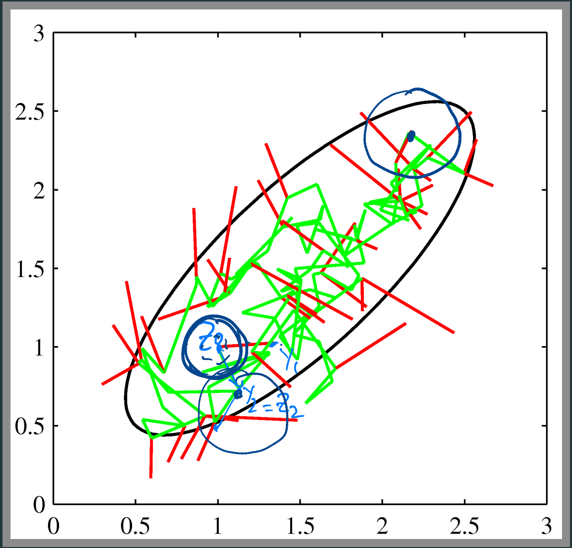


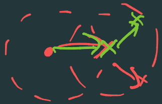
i.e., if $\widetilde{P}(Y_t) \geq \widetilde{P}(Z_{t-1})$, then move, else move with prob $\dfrac{\widetilde{P}(Y_t)}{\widetilde{P}(Z_{t-1})}$

Metropolis for 2-d Gaussian

$Q(x|y)$

$= N(0, \sigma I)$





Metropolis algorithm: proof of correctness

- Want to show P is a steady state dist^n

- $\forall\ x, y$, $R_{xy} = \mathbb{P}[x \to y] = Q(y|x) \cdot A(x,y)$

$$R_{xx} = \sum_y Q(y|x)\left(1 - A(x,y)\right)$$

- Now we check local balance, i.e., $\forall\ x, y$

$$P(x) R_{xy} = P(x) \cdot Q(y|x) \cdot \max\left(1, \frac{\tilde{P}(y)}{\tilde{P}(x)}\right)$$

// Symmetric

$$P(y) R_{yx} = P(y)\ Q(x|y) \cdot \max\left(1, \frac{\tilde{P}(x)}{\tilde{P}(y)}\right)$$

Also $\forall\ x, y$, if $P(x) \geq P(y) \Rightarrow A(x,y) = 1$, $A(y,x) = P(x)/P(y)$

Metropolis-Hastings

- target distribution $P(x) = \widetilde{P}(x)/Z$
- proposal distribution(s) $Q(x|y)$

Metropolis-Hastings sampling

1. choose initial Z_0
2. to obtain sample t, generate $Y_t \sim Q(\cdot|Z_{t-1})$
3. accept $Z_t = Y_y$ with prob $A(Y_t, Z_{t-1}) = \max\left\{1, \frac{\widetilde{P}(Y_t)Q(Z_{t-1}|Z_t)}{\widetilde{P}(Z_{t-1})Q(Z_t|Z_{t-1})}\right\}$
   else reject and set $Z_t = Z_{t-1}$

Gibbs sampling

- target distribution $P(x(1), x(2), \ldots, x(n))$

Gibbs sampling

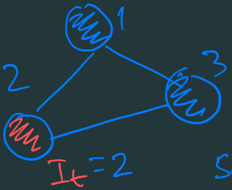
1. choose initial $X_0 = (X_0(1), X_0(2), \ldots, X_0(n))$
2. to obtain sample t:

   pick I_t uniformly at random (alternate- pick in round robin order)

   ie, $I_1 = 1, I_2 = 2, \ldots, I_n = n, I_{n+1} = 1, I_{n+2} = 2, \ldots$

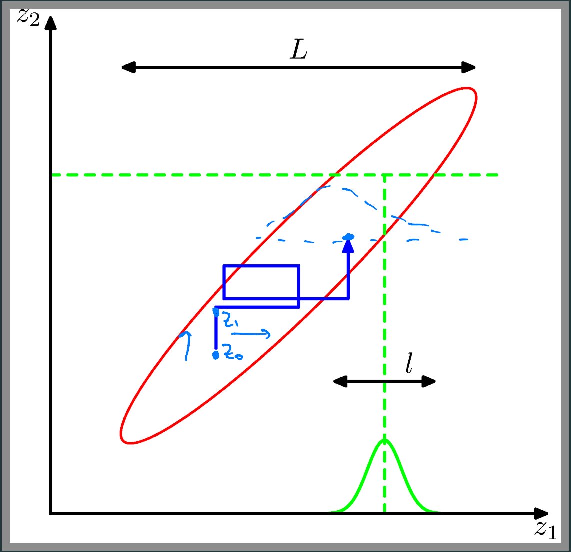
   set $X_t(i) = X_{t-1}(i)$ for $i \neq I_t$

   set $X_t(I_t) \sim P(\cdot | \underbrace{X_{t-1} \setminus X_{t-1}(I_t)}_{\text{ie, fix all } X(i) \text{ except } I_t})$



Sample $X_t(2) \sim P(x_2 | x_1, x_3)$

Gibbs sampling for 2-d Gaussian



$z_0 = [z_0(1), z_0(2)]$

$I_1 = 2$

$I_2 = 1$

$I_3 = 2$

$I_4 = 1$

$\vdots$

Gibbs sampling: proof of correctness

- $$R_{(x_1,x_2)\to(x_1',x_2)} = \underbrace{\mathbb{P}[I_t=1]}_{1/2} \cdot \mathbb{P}[x_1'|x_2]$$

$$R_{(x_1',x_2)\to(x_1,x_2)} = \frac{1}{2} \cdot P(x_1|x_2)$$

Claim. $\pi(x_1,x_2) = P(x_1,x_2)$ is a steady state dist

i.e. $$\underbrace{P(x_1,x_2)}_{P(x_2)P(x_1|x_2)} \cdot \frac{1}{2} \cdot P[x_1'|x_2] = \underbrace{P(x_1',x_2)}_{P(x_2)P(x_1'|x_2)} \cdot \frac{1}{2} \cdot P(x_1|x_2)$$