

today: sampling Markov chains

How to sample?

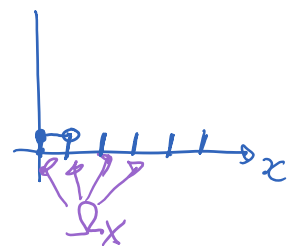
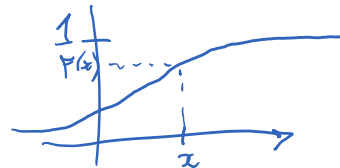
- 1)  $X \sim \text{Unif}([0,1]) \rightarrow$  pseudo-random generator "rand"
- 2)  $X \sim \text{Bernoulli}(p) \quad X = \mathbb{1}\{U \leq p\}$  where  $U \sim \text{Unif}([0,1])$
- 3) inverse transform sampling cum. dist. fct.

let  $F$  be target c.d.f. of distribution  $p$  for  $X \quad F(x) \triangleq P\{X \leq x\}$

(first, suppose  $F$  is invertible)

let  $X \triangleq F^{-1}(U)$  with  $U \sim \text{Unif}([0,1])$

claim that  $X$  has cdf  $F(x)$



$F$  is invertible and monotone

proof:  $P\{X \leq y\} = P\{F^{-1}(U) \leq y\} \stackrel{F \text{ is invertible and monotone}}{=} P\{U \leq F(y)\} = F(y)$

[ if  $F$  is not invertible, define

$X \triangleq \min\{x : F(x) \geq U\}$

(recall that  $F$  is cdf, from right)

example:

want  $X \sim \text{Exp}(\lambda)$

density  $p(x) = \lambda \exp(-\lambda x) \mathbb{1}_{\mathbb{R}^+}(x)$

$F(x) = 1 - \exp(-\lambda x)$

inverse  $F^{-1}(u) = -\frac{1}{\lambda} \log(1-u)$

multivariate distribution?

can generalize above trick using "chain rule"

$X_{1:p}$  (dim  $p$ )

from  $p(x_{1:p}) = \prod_{i=1}^p p(x_i | x_{1:i-1})$

use cdf for this conditional:

(conditional density sense when  $\downarrow$  etc)

$F_{X_i | X_{1:i-1}}(x_i | x_{1:i-1}) \triangleq P\{X_i \leq x_i | X_{1:i-1} = x_{1:i-1}\}$

could use  $U_1, \dots, U_p \stackrel{iid}{\sim} \text{Unif}([0,1])$

$$\begin{cases} X_1 = F_{X_1}^{-1}(U_1) \\ X_2 = F_{X_2 | X_1}^{-1}(U_2 | X_1) \end{cases}$$

inverse of  $F_{X_2 | X_1}(\cdot | X_1)$    
 use this argument   
 fixed

$$\left\{ \begin{array}{l} X_1 = F_{X_1}(U_1) \\ X_2 = F_{X_2|X_1}^{-1}(U_2 | X_1) \\ \vdots \\ X_p = F_{X_p|X_1, \dots, X_{p-1}}^{-1}(U_p | X_{1:p-1}) \end{array} \right.$$

*inverse of  $F_{X_1|X_1}(q | X_1)$*   
*inverse of this argument*  
*is a very complicated fct.*  
*(curse of dimensionality?)*

[aside: "copulas" → model for multivariate data with uniform marginals]

exception is multivariate Gaussian

$$N(\mu, \Sigma) \quad \Sigma = U \Lambda U^T$$

(where  $U U^T = I_p$   
 $\Lambda$  is diagonal)

generate  $V \sim N(0, I_p)$

( $V_i \stackrel{iid}{\sim} N(0, 1)$ )

$$X = U \Lambda^{1/2} V + \mu$$

$$E X = \mu$$

$$\text{cov}(X) = U \Lambda^{1/2} \underbrace{\text{cov}(V)}_{I_p} \Lambda^{1/2} U^T = \Sigma$$

(Cholesky decomposition)

$$\Sigma = L L^T$$

Box-Muller transformation to sample (2d) Gaussian

$$\begin{array}{l} R^2 \sim \text{Exp}(1) \\ \Theta \sim \text{Unif}([0, 2\pi]) \end{array} \Rightarrow \begin{array}{l} X \stackrel{\Delta}{=} R \cos \Theta \\ Y \stackrel{\Delta}{=} R \sin \Theta \end{array} \quad (X, Y) \sim N(0, I_2)$$

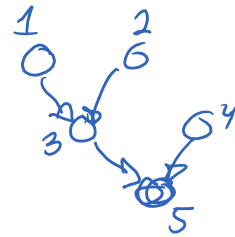
17h40

Sampling from a DGM is (relatively) easy: ancestral sampling:

$(X_1, \dots, X_p) \sim p \in \mathcal{L}(G)$  where  $G$  is a DAG

$$p(x_1, \dots, x_p) = \prod_{i=1}^p p(x_i | x_{\pi(i)})$$

suppose wlog  $1, \dots, p$  is a tp. sort of  $G$



ancestral sampling:

$$\left[ \begin{array}{l} \text{for } i=1, \dots, p \\ \quad \text{sample } X_i \sim p(X_i = \cdot | X_{\pi(i)}) \\ \text{end} \end{array} \right.$$

*these are already observed by top sort.*

can show by induction that  $(X_1, \dots, X_p)$  has dist.  $p$

⊗ important side note: when you sample from joint,

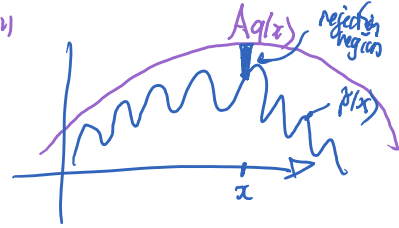
*you are also sampling from "marginals" by just ignoring the joint aspect*

i.e.  $(X, Y) \sim p(x, y)$

then looking at  $X$  itself,  $X \sim p(x)$

rejection sampling:

say  $p(x) = \frac{\tilde{p}(x)}{Z_p}$  ; let's say can find a  $q(x)$  "proposal" which is easy to sample from  
 s.t.  $Aq(x) \geq \tilde{p}(x) \forall x$



- rule:
- sample  $X \sim q(x)$
  - Accept with prob.  $\frac{\tilde{p}(x)}{Aq(x)} \in [0, 1]$
  - reject o.w.  $\rightarrow$  start again

let's show that accepted sample has correct prob.

(say  $X$  is discrete)

$$\underbrace{P\{X=x, X \text{ is accepted}\}}_{\text{sampling mechanism}} = \underbrace{P\{X \text{ is accepted} | X=x\}}_{\frac{\tilde{p}(x)}{Aq(x)}} \underbrace{P\{X=x\}}_{q(x)} = \frac{\tilde{p}(x)}{A}$$

$$P\{X \text{ is accepted}\} = \sum_x \frac{\tilde{p}(x)}{A} = \frac{Z_p}{A}$$

(marginal prob of acceptance  $\rightarrow$  want this to be high)

$$P\{X=x | X \text{ is accepted}\} = \frac{(\tilde{p}(x)/A)}{(Z_p/A)} = \frac{\tilde{p}(x)}{Z_p} = p(x)$$

application to conditioning in a DBM

say want to sample  $p(x | \bar{x}_E)$

here, could use  $\tilde{p}(x) = p(x_{E^c}, x_E) \delta(x_E, \bar{x}_E) \Rightarrow p(x) = p(x_{E^c} | \bar{x}_E) \delta(x_E, \bar{x}_E)$

$$Z_p = p(\bar{x}_E)$$

let  $q(x)$  be original joint in DBM (sample using ancestral sampling)

$$q(x) = p(x_{E^c}, x_E)$$

$$q(x) \geq \tilde{p}(x) \forall x \quad [\text{take } A=1]$$

acceptance prob.  $\frac{\tilde{p}(x)}{Aq(x)} = \delta(x_E, \bar{x}_E)$

do: • do ancestral sampling (rejection sampling)

alg. <sup>A(q|x)</sup> :

- do ancestral sampling
- accept if  $X_E = z_E$

(rejection sampling for DGM conditioning)

$$P\{\text{accept}\} = \frac{Z_p}{Z_q} = p(z_E)$$

### Importance sampling:

in context of computing  $\mathbb{E}_p[f(x)] = \mu$   $X \sim p$

$\leadsto$  can "weight" sample  $X^{(i)}$

$$\mathbb{E}_p[f(x)] = \int_x f(x) p(x) = \int_x f(x) \frac{p(x) \cdot q(x)}{q(x)} \quad \text{for some dist. } q \text{ s.t. } \text{supp}(q) \supseteq \text{supp}(p)$$

$$= \mathbb{E}_q \left[ f(y) \frac{p(y)}{q(y)} \right] \quad \text{where } Y \sim q$$

$$\approx \frac{1}{n} \sum_{i=1}^n g(y_i) \quad \text{where } y^{(i)} \stackrel{\text{iid}}{\sim} q$$

and  $g(y) \triangleq f(y) w(y)$

where  $w(y) \triangleq \frac{p(y)}{q(y)}$  "weights"

$$\hat{\mu}_{\text{IS}} = \frac{1}{n} \sum_{i=1}^n f(x_i) w_i \quad y_i \stackrel{\text{iid}}{\sim} q$$

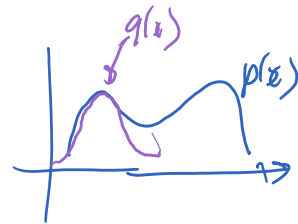
*"importance weights"*  $w_i \triangleq \frac{p(y_i)}{q(y_i)}$

indeed  $\text{var}[\hat{\mu}]$  can be too sometimes!

$$\mathbb{E}[\hat{\mu}] = \mu$$

$$\text{Var}[\hat{\mu}] = \frac{1}{n} \left[ \mathbb{E}_p \left[ f(x) \frac{p(x)}{q(x)} \right]^2 - \mu^2 \right]$$

issues arise when  $q$  is small and  $p$  is big



intuitively you want  $q(x) \propto f(x)p(x)$

extension to un-normalized dist. :

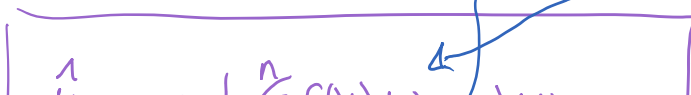
$$p(x) = \frac{\tilde{p}(x)}{Z_p} \quad q(x) = \frac{\tilde{q}(x)}{Z_q}$$

$$\mu = \mathbb{E}_q \left[ f(y) \frac{p(y)}{q(y)} \right]$$

$$= \mathbb{E}_q \left[ f(y) \frac{\tilde{p}(y)}{\tilde{q}(y)} \right] \frac{Z_p}{Z_q}$$

estimate  $\frac{Z_p}{Z_q}$  with  $\hat{\frac{Z_p}{Z_q}} \triangleq \frac{1}{n} \sum_{i=1}^n \frac{\tilde{p}(y_i)}{\tilde{q}(y_i)}$

$$= \frac{1}{n} \sum_{i=1}^n w_i$$



$$\hat{\mu}_{\text{UIS}} = \frac{\sum_{i=1}^n f(x_i) w_i}{\sum_{i=1}^n w_i}$$

$x_i \sim q$   
 $w_i \propto \frac{p(y_i)}{q(y_i)}$

note:  $\hat{\mu}_{\text{UIS}}$  is (slightly) biased, but asymptotically unbiased

- this estimator has often lower variance than  $\hat{\mu}_{\text{IS}}$ , even when  $Z_p = Z_q = 1$

(normalized "stabilized" estimator) new weights  $\tilde{w}_i = \frac{w_i}{\sum_{j=1}^n w_j} \in [0, 1]$

see 2017 notes for

- variance reduction (link with SAGA)
- Rao-Blackwellization

Good reference on sampling:

Monte Carlo Statistical Methods, Robert & Casella, 2004