

today: sampling  
MCMC

How to sample?

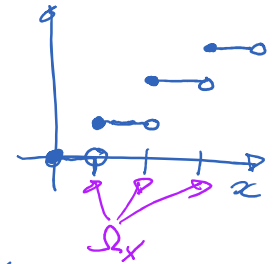
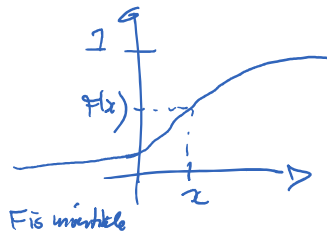
- 1)  $X \sim \text{Unif}([0,1]) \rightarrow$  pseudo-random generator "rand"
- 2)  $X \sim \text{Bernoulli}(p) \quad X = \mathbb{1}\{U \leq p\}$  where  $U \sim \text{Unif}([0,1])$
- 3) inverse transform sampling: cumulative dist. fct.

let  $F$  be target cdf of distribution  $p$  for  $X \quad F(x) \triangleq \mathbb{P}\{X \leq x\} \quad (x \in \mathbb{R})$

(first, suppose  $F$  is invertible)

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

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



proof:  $\mathbb{P}\{X \leq y\} = \mathbb{P}\{F^{-1}(U) \leq y\} = \mathbb{P}\{U \leq F(y)\} = F(y)$

[if  $F$  is not invertible, define  $X \triangleq \min\{x : F(x) \geq U\}$

(recall that  $F$  is dist. 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  $F^{-1}(u) = -\frac{1}{\lambda} \ln(1-u)$

Multivariate distribution?

can generalize above trick using "chain rule"

$X_{1:p}$  (dim  $p$ ) cdf  $F(x_{1:p}) \triangleq \mathbb{P}\{X_1 \leq x_1, \dots, X_p \leq x_p\}$

$F_{X_{1:p}}(x_{1:p}) = F_{X_1}(x_1) \underbrace{F_{X_2|X_1}(x_2|x_1)} \dots F_{X_p|X_{1:p-1}}(x_p|x_{1:p-1})$

$$\hookrightarrow F_{X_2|X_1}(x_2|x_1) \triangleq P\{X_2 \leq x_2 | X_1 = x_1\}$$

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

$$X_1 = F_{X_1}^{-1}(U_1)$$

$$X_2 = F_{X_2|X_1}^{-1}(U_2 | X_1)$$

$\vdots$

$$X_p = F_{X_p|X_{1:p-1}}^{-1}(U_p | X_{1:p-1})$$

is very complicated function

(curse of dimensionality?)

[aside: "copulas"  $\rightarrow$  model for multivariate dependence with uniform marginals]

exception is multivariate Gaussian:

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

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

(Cholesky decomposition)

$$L = U \Lambda^{1/2}$$

$$\Sigma = LL^T$$

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

( $V_p \stackrel{\text{iid.}}{\sim} N(0,1)$ )

$$X \triangleq \underbrace{U \Lambda^{1/2}}_L V + \mu$$

$$EX = \mu$$

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

Box-Muller transform to sample (2d) Gaussian

$$\begin{aligned} R^2 &\sim \text{Exp}(1) & \Rightarrow & X \triangleq R \cos \Theta \\ \Theta &\sim \text{Unif}(0, 2\pi) & & Y \triangleq R \sin \Theta \end{aligned}$$

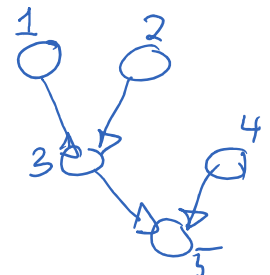
$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N(0, I)$$

sampling for DGM is 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 top. sort for  $G$



ancestral sampling:

for  $i=1, \dots, p$

those are already sampled

for  $i=1, \dots, p$   
 sample  $X_i \sim p(X_i = \cdot | X_{1:i})$   
 end

these are already sampled  
 by top sort.

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

aside: formal tool to show that  $(X_1, X_2)$  has right joint dist.

to show that two r.v.'s are equal in dist. i.e.  $U \stackrel{d}{=} V$

$\Leftrightarrow E_U[f(U)] = E_V[f(V)]$  for all functions in a  
 big enough class  
 (e.g. cts & bounded fct.)

\* 2 r.v. example:  $U_1 \sim p(x_1)$   
 $U_2 | U_1 \sim p(x_2 | X_1 = U_1)$

now, want to show that

want to show this

$$E[f(X_1, X_2)] \stackrel{d}{=} E[f(U_1, U_2)]$$

tower property of expect.

by sampling mechanism

$$= E_{U_1} [ E_{U_2 | U_1} [f(U_1, U_2) | U_1] ]$$

$$= \int_{u_1} p(u_1) du_1 \left[ \int f(u_1, u_2) p(u_2 | u_1) du_2 \right]$$

$$= \int_{u_1} \int_{u_2} f(u_1, u_2) \underbrace{[p(u_1)p(u_2 | u_1)]}_{p(u_1, u_2)} du_1 du_2$$

\* important sidebar: when sample from joint, you are also sampling from "marginal"  
 by just ignoring the joint aspect

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

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

15h36

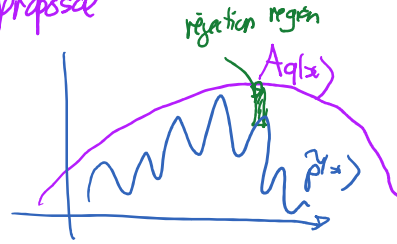
rejection sampling:

say  $p(x) = \frac{\tilde{p}(x)}{Z_p}$

let's say we can find a  $q(x)$  which is easy to sample from

"proposal"

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 samples have correct dist.

(say  $X$  is discrete)

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

sampling mechanism

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

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

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

Application on conditioning in DGM:

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

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

$$Z_p = p(\bar{x}_E) \quad p(x) = p(x_{E^c} | \bar{x}_E) \delta(x_E, \bar{x}_E)$$

let  $q(x)$  be original joint in DGM (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.  $\frac{\tilde{f}(x)}{Aq(x)} = \delta(x_E, \bar{x}_E)$

alg.  $\left\{ \begin{array}{l} \cdot \text{do ancestral sampling} \\ \cdot \text{accept if } x_E = \bar{x}_E \end{array} \right.$  (rejection sampling for DM)

$$P\{\text{accept}\} = \frac{Z_p}{A} = p(\bar{x}_E)$$

### Importance sampling:

in context of computing  $E_p[f(x)] = \mu$   $X \sim p$

$\rightsquigarrow$  (can "weight" sample  $X^i$ )

$$E_p[f(x)] = \int f(x)p(x)dx = \int f(x) \frac{p(x)}{q(x)} q(x) dx \quad \text{for some dist. } q \text{ s.t. } \text{supp}(q) \supseteq \text{supp}(p)$$

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

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

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

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

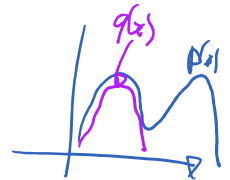
$$\hat{\mu}_{IS} = \frac{1}{n} \sum_{i=1}^n f(X_i) w_i \quad X_i \sim q$$

"importance weights"

$$w_i \triangleq \frac{p(Y_i)}{q(Y_i)}$$

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

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



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

indeed,  $\text{Var}_q[\dots]$  can sometimes be  $\infty$

intuitively, you want  $q(x)$  or  $f(x)p(x)$

extension to un-normalized dist.:

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

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

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

$$\hat{\mu}_{ISU} = \frac{\frac{1}{n} \sum_{i=1}^n f(y_i) w_i}{\frac{1}{n} \sum_{i=1}^n w_i} \quad y_i \sim q$$

$w_i \triangleq \frac{\tilde{p}(y_i)}{\tilde{q}(y_i)}$

- note:
- $\hat{\mu}_{ISU}$  is (slightly biased), but asymptotically unbiased (as  $n \rightarrow \infty$ )
  - this estimator often has lower variance than  $\hat{\mu}_{IS}$  even when  $Z_p = Z_q = 1$  (normalization "stabilizes" 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 SGA)
- Rao-Blackwellization