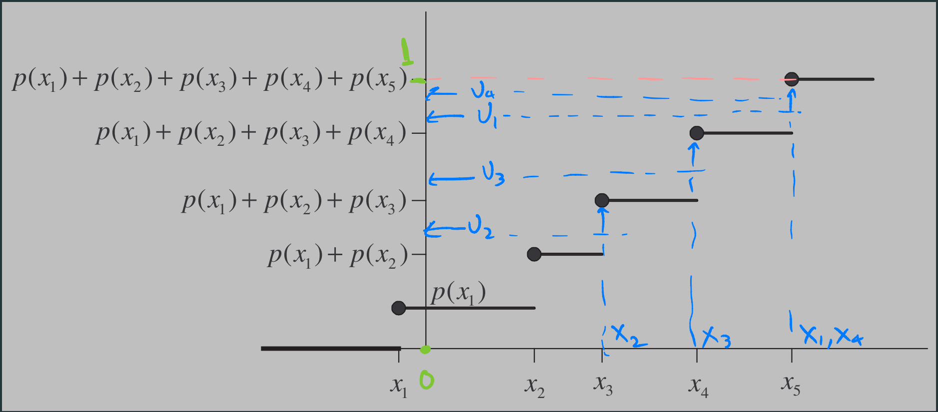
basic monte carlo techniques

warmup: simulating discrete rv

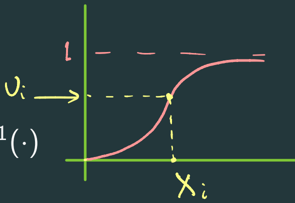
X takes values $x_1 \leq x_2 \leq \ldots \leq x_5$, $\mathbb{P}[X = x_i] = p(x_i)$



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 $F^{-1}(\cdot)$
- $F(x) = u \iff F^{-1}(u) = x$

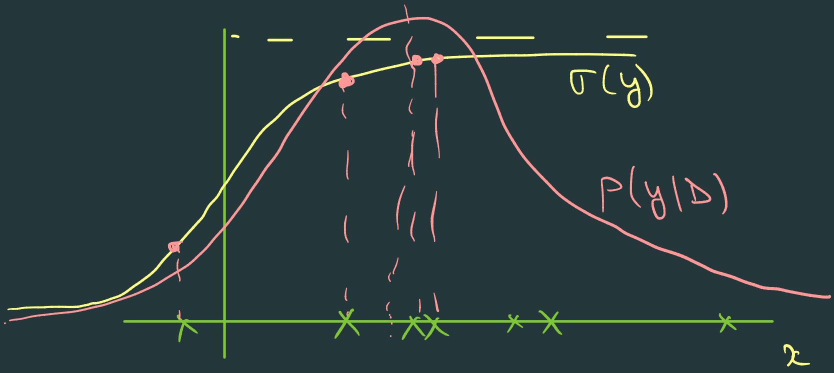


inversion method

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

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

Application - Compute integrals



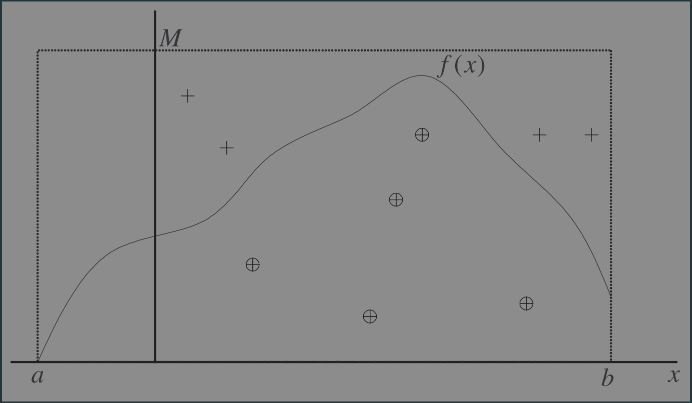
$$I = \mathbb{E}_{x \sim F}\left[\sigma(x)\right] \approx \frac{1}{N} \sum_{i=1}^{N} \sigma(x_i)$$

rejection sampling

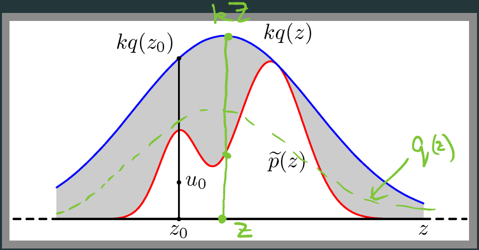
want samples of a rv $X \in [a, b]$, with pdf $f(x) \leq M$

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



generalized rejection sampling



- Given a 'sampler' $Z \sim Q$
- Want samples $X \sim P$

- find k s.t. $kq(x) \geq p(x) \; \forall x$ (ie, $k \geq \max \frac{p(x)}{q(x)}$)
- Generate $Z \sim Q$
- Accept (ie set $X = Z$) w.p $\underbrace{\frac{p(Z)}{k q(Z)}}_{<1 \text{ by defn of } k}$, else repeat

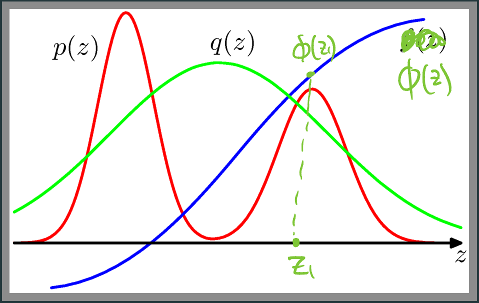
Problems
1) Need to know k 2) Need $p(x)$ exactly
3) Inefficient in high dimensions

importance sampling (for estimating integrals)

- 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 $\mathbb{E}_p[\phi(X)] = \frac{1}{L}\sum_{i=1}^{L} w_i \phi(Z_i)$, where $w_i = p(Z_i)/q(Z_i)$



$$\mathbb{E}_{X\sim P}[\phi(x)] = \int \phi(x)\, p(x)\, dx$$

$$= \int \phi(x) \left(\frac{p(x)}{q(x)}\right) q(x)\, dx$$

$$= \mathbb{E}_{Z\sim Q}\Big[\phi(Z)\, \underset{\uparrow\ p(z)/q(z)}{W(Z)}\Big]$$

$$= \frac{1}{L}\sum_{i=1}^{L} w_i \phi(z_i)$$

importance sampling: unknown normalization

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

  want $X \sim P$

- $$\mathbb{E}[\phi(X)] = \int \phi(x) \frac{\tilde{p}(x)}{Z_P} dx = \int \phi(x) \left(\frac{\tilde{p}(x)}{\tilde{q}(x)}\right)\left(\frac{\tilde{q}(x)}{Z_Q}\right)\left(\frac{Z_Q}{Z_P}\right) dx$$

  $$= \underbrace{\left(\frac{Z_Q}{Z_P}\right)}_{\text{need an estimate}} \mathbb{E}_{Y\sim Q}[\phi(Y) w(Y)], \quad w(Y) = \frac{\tilde{p}(Y)}{\tilde{q}(Y)}$$

- Suppose $\phi(x) = 1 \Rightarrow \mathbb{E}[\phi(x)] = 1 = \left(\frac{Z_Q}{Z_P}\right) \cdot \mathbb{E}_{Y\sim Q}[1 . w(Y)]$

$$\Rightarrow \text{For any } \phi: \mathbb{E}_{X\sim P}[\phi(X)] = \frac{\mathbb{E}_{Y\sim Q}[\phi(Y) w(Y)]}{\mathbb{E}_{Y\sim Q}[w(Y)]} \approx \frac{\frac{1}{L}\sum_{i=1}^{L} \phi(Y_i) w(Y_i)}{\frac{1}{L}\sum_{i=1}^{L} w(Y_i)}$$

importance sampling: comments

- $\mathbb{E}[\phi(x)] \approx \sum_{i=1}^{L} \frac{\tilde{w}(y_i)}{\sum_{j=1}^{L} \tilde{w}(y_j)} \cdot \phi(y_i)$, where $\tilde{w}(y_i) = \frac{\tilde{p}(y_i)}{\tilde{q}(y_i)}$

- Quality of MCMC approx depends on Variance of estimator

- $\mathrm{Var}\left(\frac{1}{L} \sum w(y_i)\phi(y_i)\right) = \mathrm{Var}\left(\phi(y)\, w(y)\right) \leftarrow$ is smaller if $\phi(y)\, \underline{w(y)}$ has smaller range

Eg- Ber(p) vs 100 Ber$\left(\frac{p}{100}\right)$

$\Rightarrow$ Good proposal dist^n q is suchthat

$\frac{p(z)}{q(z)}$ is small

- Ideally choose $q \approx p$ (definitely $q(z) \neq 0$ for any z s.t. $p(z) > 0$)

markov chain monte carlo

MCMC: the basic idea

- Method 1 – Generate $X_1, X_2 \dots X_n$ iid from P (inversion, importance sampling)
- Method 2 – Generate $Y_1, Y_2, \dots, Y_n$ iid from Q, accept/reject samples to get $Y_1, \cancel{Y_2}, Y_3 \dots \cancel{Y_n} \longrightarrow X_1, X_2, \dots X_K, K<n$
- MCMC – Generate X_1, generate $\tilde{X}_2 = f(X_1)$,
  - Accept/Reject $\tilde{X}_2$ (ie, $X_2 = X_1$ or $X_2 = \tilde{X}_2$) with prob $A(X_1, \tilde{X}_2)$

$$X_1 \longrightarrow \cancel{\tilde{X}_2}$$
$$X_1 \longrightarrow X_2 = X_1 \longrightarrow \tilde{X}_3 \longrightarrow X_3 = \tilde{X}_3 \; \dots$$

markov chains: basic definition

- Seq of rv $X_1, X_2, \ldots, X_n$ is a MC iff they have following Bayes Net

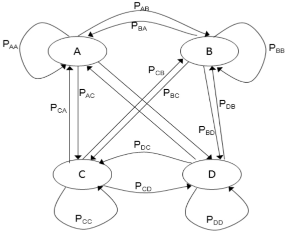
$$X_1 \to X_2 \to X_3 \to \cdots \to X_n$$

(ie. $\mathbb{P}[X_i = y \mid X_1 = x_1, X_2 = x_2 \ldots X_{i-1} = x_{i-1}] = \mathbb{P}[X_i = y \mid X_{i-1} = x_{i-1}]$
$\forall i \geq 1, \ \forall (x_1, x_2, \ldots, x_{i-1}, y)$)

- Exercise - Use d-separation to check $\forall s < t$

$$P(X_i, X_{i+t} \mid X_{i+s}) = P(X_i \mid X_{i+s}) \, P(X_{i+t} \mid X_{i+s})$$

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



$$P = \begin{bmatrix} p_{AA} & p_{AB} & p_{AC} & p_{AD} \\ p_{BA} & p_{BB} & p_{BC} & p_{BD} \\ p_{CA} & p_{CB} & p_{CC} & p_{CD} \\ p_{DA} & p_{DB} & p_{DC} & p_{DD} \end{bmatrix}$$

transition matrix

- State-space = $\{A, B, C, D\}$
- $P(X_{t+1} = A \mid X_t = B) = P_{BA}$
- This is called a state-transition diagram (NOT a Bayes Net)

Facts

1) $\sum_{x \in \{A,B,C,D\}} P_{Ax} = 1 \Leftrightarrow P$ is stochastic (ie, row sums are all 1)

2) Directed, can have cycles and self-loops

3) Let π_t = Distribution of $X_t = \mathbb{P}[X_t = x \mid X_0]$

Then $\pi_{t+1}^T = \pi_t^T P = [\pi_t[A] \ \dots] \begin{bmatrix} P_{AB} & P_{AC} & \dots \\ & \ddots & \end{bmatrix}$

markov chains: steady-state

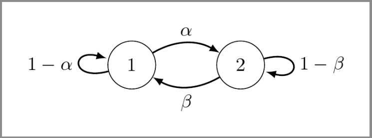
Given $X_0, X_1, X_2, \ldots$ from a time-invariant MC, then

as $t \to \infty$, we have $\pi_t \to$ fixed distr π s.t $\boxed{\pi^T = \pi^T P}$

(ie, for all states x, and initial x_0, $\lim_{t\to\infty} \pi_t[x] = \pi[x]$)

- under 'mild' conditions (finite, 'strongly-connected' (irreducible), has self loops (aperiodic))

example:



$$\pi_0^T = \overset{\pi[1]\ \pi[2]}{(1,\ 0)} \quad \text{(ie, start in 1)}$$

$$\pi_{t+1}^T = \left(\pi_t[1]\ \ \pi_t[2]\right)\begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

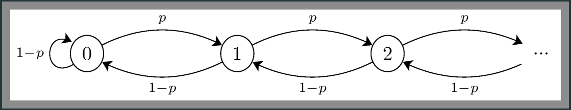
$$\to \quad \pi = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right)$$

Want: π = target dist Q

markov chains: example

2 questions
1) How do we design P to get desired π
2) How do we compute $\mathbb{E}[\Phi(X)]$, where $X \sim \pi$

the 1-d random walk



$(p < \frac{1}{2})$

- Steady-state $\pi \equiv \underbrace{\sum_j \pi(i) P_{ij}}_{\pi(i)} = \sum_j \pi(j) P_{ji} \quad \forall i \quad (*)$

<u>Claim</u> - $\pi(i) = \frac{1}{Z}\left(\frac{p}{1-p}\right)^i \quad \forall i$

<u>'Proof'</u> - For any i, LHS of $(*)$ $= \frac{1}{Z}\left(\frac{p}{1-p}\right)^i \left(p + (1-p)\right)$

RHS of $(*)$ $= \frac{1}{Z}\left(\left(\frac{p}{1-p}\right)^{i-1} p + \left(\frac{p}{1-p}\right)^{i+1}(1-p)\right) = \frac{1}{Z}\left(\frac{p}{1-p}\right)^i$