Probability Theory

$p(x)\begin{cases} \text{prob. mass func. (pmf), x-discrete}, & P(\zeta = x) = p(x) \\ \text{prob. density func. (pdf), x-continuous}, & P(a < \zeta < b) = \int_a^b p(x)\,dx \end{cases}$

$\boxed{\text{Th}}$ $F(x)$ - cumulative density function (cdf) $= P(\zeta < x)$

discrete: $\forall x \in \mathcal{A}\ (P(\zeta = x) \geq 0) \wedge \sum_{x \in \mathcal{A}} P(\zeta = X) = 1$

$\mathbb{E}[\zeta] = \sum_{x \in \mathcal{A}} x \cdot P(\zeta = x) \qquad \mathbb{E}[f(\zeta)] = \sum_{x \in \mathcal{A}} f(x) \cdot P(\zeta = x)$

continuous: $\forall x \in \mathcal{A}\ (p(x) \geq 0) \wedge \int_{\mathcal{A}} p(x)\,dx = 1$

$\mathbb{E}[\zeta] = \int_{\mathcal{A}} x \cdot p(x)\,dx \qquad \mathbb{E}[f(\zeta)] = \int_{\mathcal{A}} f(x)\,p(x)\,dx$

$Var(\zeta) = \mathbb{E}[\zeta^2] - (\mathbb{E}[\zeta])^2 = \sigma^2$

$\mathbb{E}[n\zeta] = n\mathbb{E}[\zeta] \quad \mathbb{E}[n+\zeta] = n + \mathbb{E}[\zeta] \quad Var(n\zeta) = n^2 Var(\zeta) \quad Var(n+\zeta) = Var(\zeta)$

$\boxed{\text{Th}}$ $P(A|B) = \dfrac{P(B|A) \cdot P(A)}{P(B)} = \dfrac{P(A \cap B)}{P(B)}$

Bayes

$\boxed{\text{Th}}$ $\xi_1, \xi_2, \ldots, \xi_n$ - independent random variables, $\mathbb{E}[\xi_i] = \mu$, $Var[\xi_i] = \sigma^2 < \infty$, $\overline{S_n} = \frac{\xi_1 + \xi_2 + \ldots + \xi_n}{n}$

Central Limit $\rightarrow \frac{\overline{S_n} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)$, $\overline{S_n} \sim N(\mu, \frac{\sigma^2}{n})$ for large enough n

Relative Error $= \dfrac{\sqrt{Var(\overline{S_n})}}{|\theta|} = \dfrac{\sigma}{\sqrt{N}\,|\theta|}$

Hoeffding's Inequality

$\boxed{\text{Def.}}$ $\overline{S_n}$ approximates μ, $\delta = P(|\overline{S_n} - \mu| > \varepsilon) \rightarrow \delta \leq 2e^{-2n\varepsilon^2}$

$P(|\overline{S_n} - \mu| > \varepsilon) \leq \delta \rightarrow P(-\varepsilon + \overline{S_n} < \mu < \varepsilon + \overline{S_n}) = 1 - \delta$

$2e^{-2n\varepsilon^2} \leq \delta \rightarrow n \geq \frac{1}{-2\varepsilon^2} \ln\left(\frac{\delta}{2}\right)$

1) $X_i \sim Ber(p)$, $\overline{X_n} = \frac{\sum_{i=1}^{n} X_i}{n}$ / $X \sim Ber(p)$, $\mathbb{E}[X] = p$, $Var[X] = p(1-p)$ /

$Var(\overline{X_n}) = Var\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{N\ Var(X_i)}{N^2} = \frac{p(1-p)}{N}$, $RE = \frac{\sqrt{Var(\overline{X_n})}}{p} = \frac{\sqrt{p(1-p)}}{\sqrt{n}\ p}$

2) Y - RV, CDF $= F_Y$, RV $T = F_Y^{-1}(X)$, $X \sim U(0,1)$

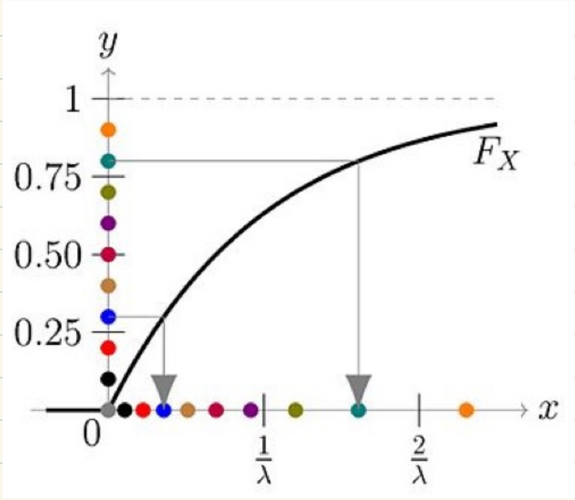
$F_T(a) = P(T \leq a) = P(F_Y^{-1}(X) \leq a) = P(X \leq F_Y(a)) = \int_0^{F_Y(a)} 1\ dx = F_Y(a)$

3) $f(x) = \frac{\sqrt{x(1-x)}}{Cx}$, $0 \leq x \leq 1$, $C \geq 1$

$\lim_{x \to 1} f(x) = 0$, $\lim_{x \to 0} f(x) = \infty$

Monte Carlo

Approximating an expectation by the sample mean of a function of simulated random variables.



Inverse Transform

Sample any probability distribution with an invertible cumulative distribution F.

sample $F^{-1}(X)$, $X \sim U(0,1)$

2) Fibonacci sequence:

$$F(n) = \begin{cases} 1 & \text{if } n=1,2 \\ F(n-1)+F(n-2) & \text{else} \end{cases}$$

F^*, $F^*(1) = X_1 \sim U(0,4)$, $F^*(2) = X_2 \sim U(4,9)$

$P(F^*(10) > 300) \sim 0.28$, $F^*(10) = 21F^*(1) + 34F^*(2) = 21X_1 + 34X_2$

3) $X \sim N(0,1)$ – continuous $\Rightarrow P(X=0.3) = \int_{0.3}^{0.3} f(x)\,dx = F(x)\Big|_{0.3}^{0.3} = F(0.3) - F(0.3) = 0$

4) a) Sample from circle, center $\vec{x}$, radius r, $f(x) = 2x$

$\vec{x}' = \vec{x} + rY\begin{bmatrix}\cos\theta \\ \sin\theta\end{bmatrix}$, $\theta \sim U(0, 2\pi)$, Y has density $f(x)$

Inverse Transform for sampling Y:

$F(x) = \int_0^x f(t)\,dt = \int_0^x 2t\,dt = t^2\Big|_0^x = x^2 \Rightarrow F^{-1}(x) = \sqrt{x}$

$\Rightarrow \vec{x}' = \vec{x} + r\sqrt{X}\begin{bmatrix}\cos\theta \\ \sin\theta\end{bmatrix}$, $\theta \sim U(0,2\pi)$, $X \sim U(0,1)$

b) 2 samplers, sampling A or B respectively, estimate $p(A|B)$

1) sample X_i using sampler for B

2) Y_i – RV, 1 if $X_i \in A$, otherwise 0

3) Repeat 1 & 2 N times, $\hat{p} = \bar{Y}_n = \frac{\sum_{i=1}^{N} Y_i}{N}$, $\hat{p} \sim p$ for large N

c) $Y_i \sim Ber(p) \Rightarrow$ Relative Error $= \frac{\sqrt{Var(\bar{Y}_n)}}{p} = \frac{\sqrt{p(1-p)}}{\sqrt{N}\,p}$

d) move A to the left s.t. $P(A \cap B) > 0$ decreases

$P(A \cap B) \downarrow \Rightarrow p \downarrow \Rightarrow$ Relative Error $\uparrow$

Advanced Sampling Methods

Acceptance Rejection

$$\varphi(x) = \frac{f(x)}{C} = \frac{f(x)}{\int_{-\infty}^{\infty} f(x)\,dx} \longrightarrow \text{cannot always compute}$$

C - normalizing constant, $\int \varphi(x)\,dx = 1$

1) pick $g(x)$ - spans same range as $f(x)$
2) pick M s.t. $\forall x,\ Mg(x) \geq f(x)$
3) sampling: s from $g(x)$, $u \sim U(0,1)$, accept if $u \leq \frac{f(s)}{Mg(s)}$

Efficiency: C/M

Simplex

n-dimensional simplex $A = \{\vec{x} = (x_1, \dots, x_{n+1})^T \mid \sum_{i=1}^{n+1} x_i = 1 \wedge \forall i,\ 0 \leq x_i \leq 1\}$

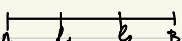
1) n samples of $\log(X_i)$, $X_i \sim U(0,1)$
2) normalize by dividing by sum
3) return normalized

Importance Sampling

$$\mathbb{E}_p(f(X)) = \int_R f(x)p(x)\,dx = \int_R q(x)\left(f(x)\frac{p(x)}{q(x)}\right)dx = \int_R H(x)q(x)\,dx = \mathbb{E}_q[H(x)] \approx \frac{1}{n}\sum_{i=1}^{n} H(x_i^q)$$

1) pick $q(x)$
2) generate $X_1^q, \dots, X_n^q$, calculate $y = \frac{1}{n}\sum_{i=1}^{n} H(x_i^q)$
3) construct confidence interval on $y \in \left[-z_{\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}} + y,\ z_{\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}} + y\right]$

Importance Splitting



$P(B|A) = P(B|l_2)P(l_2|l_1)P(l_1|A)$

1) pre-determined # samples from A
2) sample reaches l_i, split it into k to start from l_i, if $\varphi(X) < \tau$, discard
3) $P(l_{i+1}|l_i) = \frac{\#\text{ samples reach } l_{i+1} \text{ from } l_i}{\#\text{ samples simulated from } l_i}$
4) $P(B|A) = P(B|l_n) \cdot \prod_{i=1}^{n} P(l_{i+1}|l_i) \cdot P(l_1|A)$

Strong Markov Property must hold.

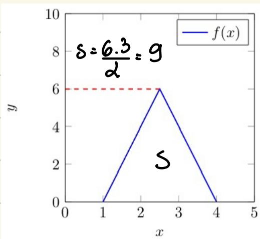
1) $X \sim N(\mu, \sigma^2)$, Importance Sampling for $\int_{-\infty}^{\infty} x f(x)\,dx$

$q(x) = \frac{x f(x)}{\mu} \Rightarrow \int_{-\infty}^{\infty} x f(x) \frac{q(x)}{q(x)}\,dx = \int_{-\infty}^{\infty} q(x)\left[\frac{x f(x)}{q(x)}\right] dx = \int_{-\infty}^{\infty} q(x)\mu\,dx = \mathbb{E}_q[f'(X)] = \sum_{i=1}^{N} f'(X_i^q)$

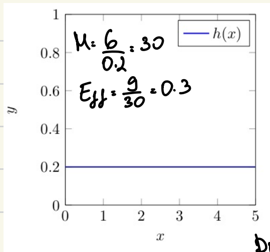
$f'(x) = \mu \quad Var(f'(X_i^q)) = \mathbb{E}[f(X_i^q)^2] - (\mathbb{E}[f(X_i^q)])^2 = \mu^2 - \mu^2 = 0$

2) Importance Splitting

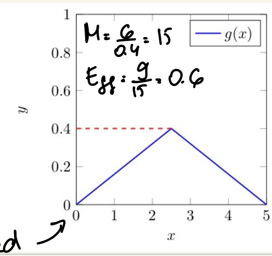
3) $\varphi(x) = \frac{f(x)}{NC}$ on $[0,5]$



a)



prefered



b) $x_1 = 3, \quad x_2 = 2, \quad x_3 = 2, \quad x_4 = 2.5, \quad x_5 = 1$

random draws from $U(0,1)$: $\underline{0.8,\ 0.2,\ 0.95,\ 1,\ 0.25}$, 0.4, 0.7, 0.3

1) $x_1 = 3$, $u = 0.8$

$\frac{f(x_1)}{Mg(x_1)} = \frac{4}{4.5} \approx 0.88 > 0.8 \Rightarrow$ accept $\qquad \frac{f(x_1)}{Mh(x_1)} = \frac{4}{6} \approx 0.66 < 0.8 \Rightarrow$ reject

2) $x_2 = 2$, $u = 0.2$

$\frac{f(x_2)}{Mg(x_2)} = \frac{4}{4.5} \approx 0.88 > 0.2 \Rightarrow$ accept $\qquad \frac{f(x_2)}{Mh(x_2)} = \frac{4}{6} \approx 0.66 > 0.2 \Rightarrow$ accept

3) $x_3 = 2$, $u = 0.95$

$\frac{f(x_3)}{Mg(x_3)} = \frac{4}{4.5} \approx 0.88 < 0.95 \Rightarrow$ reject $\qquad \frac{f(x_3)}{Mh(x_3)} = \frac{4}{6} \approx 0.66 < 0.95 \Rightarrow$ reject

4) $x_4 = 2.5$, $u = 1$

$\frac{f(x_4)}{Mg(x_4)} = \frac{6}{6} = 1 = 1 \Rightarrow$ accept $\qquad \frac{f(x_4)}{Mh(x_4)} = \frac{6}{6} = 1 = 1 \Rightarrow$ accept

5) $x_5 = 1$, $u = 0.25$

$\frac{f(x_5)}{Mg(x_5)} = \frac{0}{2} = 0 < 0.25 \Rightarrow$ reject $\qquad \frac{f(x_5)}{Mh(x_5)} = \frac{0}{6} = 0 < 0.25 \Rightarrow$ reject

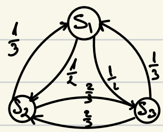
Markov Chain Monte Carlo

Markov Chains

$p_{i,j} = P(X_{t+1} = j \mid X_t = i)$

$S = \{s_1, \dots, s_n\}$ $n \in \mathbb{N}$ states

$X = (X_1, \dots, X_N)$ $X_i \in S$



Def. Markov Property: $P(X_{t+1} = x_{t+1} \mid X_t = x_t, X_{t-1} = x_{t-1}, \dots, X_0 = x_0) = P(X_{t+1} = x_{t+1} \mid X_t = x_t)$

$$M = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} & 0 \end{bmatrix}$$

$\pi_t = [P(X_t = s_1) \quad P(X_t = s_2) \quad P(X_t = s_3)]$ $\quad \pi_1 = [1 \quad 0 \quad 0]$

$P(X_{t+1} = s_1) = P(X_{t+1} = s_1 \mid X_t = s_1) P(X_t = s_1) + \dots + P(X_{t+1} = s_1 \mid X_t = s_3) P(X_t = s_3)$

$\pi_2 = \pi_1 M = [0 \quad \frac{1}{2} \quad \frac{1}{2}]$ $\quad \pi_3 = \pi_2 M = \pi_1 M^2 = [\frac{1}{3} \quad \frac{1}{3} \quad \frac{1}{3}]$

$\pi_k = \pi_1 M^{k-1}$, $k \to \infty \Rightarrow \pi M = \pi$, $Mv = \lambda v$ for $\lambda = 1$, $\sum_{i=1}^{n} \pi[i] = 1 \wedge \forall i \, (\pi[i] \geq 0)$

M-irreducible (strongly connected) $\Big\}$ $\to \exists! \, v \mid \pi$
M-aperiodic (no pattern in states)

Metropolis-Hastings

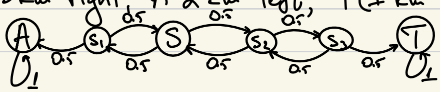
sample from probability distribution $P(x)$, proposal distribution $Q(q' \mid q)$

arbitrarily pick q_0, $q = q_0$

repeat $\begin{cases} \text{generate } q' \text{ from } Q(q' \mid q) \\ A(q' \mid q) = \min(1, \; P(q') Q(q \mid q') / P(q) Q(q' \mid q)) \\ u \sim U(0, 1), \quad u < A(q' \mid q), \; q = q' \end{cases}$

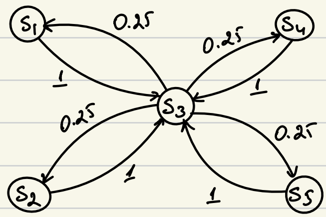
Metropolis Algorithm – $Q(q \mid q') = Q(q' \mid q)$, $A(q' \mid q) = \min(1, \; P(q') / P(q))$

1) T - 3km right, A - 2 km left, P(1 km left) = P(1 km right) = 0.5

a) 

b) absorbing states => not irreducible

2)
$$P = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0.25 & 0.25 & 0 & 0.25 & 0.25 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

a) 

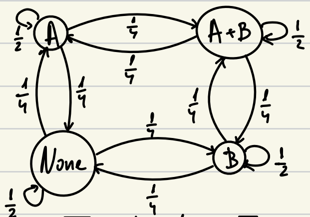
b) $X_0 = s_3 \Rightarrow \pi_0 = [0 \;\; 0 \;\; 1 \;\; 0 \;\; 0]$

i) $\pi_1 = \pi_0 P = [0.25 \;\; 0.25 \;\; 0 \;\; 0.25 \;\; 0.25]$

ii) $\pi_2 = \pi_1 P = \pi_0 P^2 = [0 \;\; 0 \;\; 1 \;\; 0 \;\; 0]$

c) $P(X_{10} = s_3) = 1$ because periodic

d) no stationary distribution because periodic

4) 

a)
$$M = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 \\ \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4} \\ \frac{1}{4} & 0 & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

b) M-irreducible and aperiodic $\Rightarrow \exists!\, v$ s.t. $vM = v$

uniform distribution $\Rightarrow v = [\frac{1}{4} \;\; \frac{1}{4} \;\; \frac{1}{4} \;\; \frac{1}{4}]$

$vM = [\frac{1}{4} \;\; \frac{1}{4} \;\; \frac{1}{4} \;\; \frac{1}{4}] = v$

c) f - fitness function, stationary distribution $\sim f$

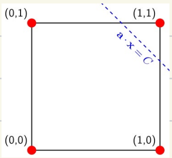
Metropolis Hastings with $P = f$.

Optimization & Planning

Knapsack Problem Model

$W = \{w_1, \dots, w_n\}$ weights, C capacity, $s_i = 1$ if item i in knapsack, otherwise $s_i = 0$

$\vec{x} = (s_1, s_2, \dots, s_n)^T$ solution, $\vec{a} = (w_1, w_2, \dots, w_n)^T \Rightarrow \vec{a} \cdot \vec{x} = \sum_{i=1}^{n} w_i \cdot s_i$, $\vec{a} \cdot \vec{x} \leq C \Rightarrow \vec{x}$-feasible



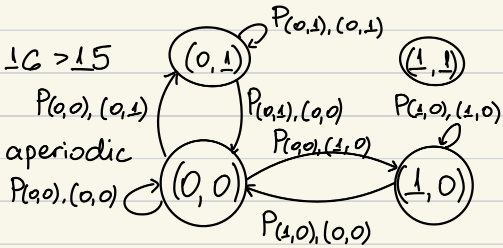
$\vec{a} = (5, 11)^T$

solution space $\Omega = \{\vec{x} \in \{0,1\}^n \mid \vec{a} \cdot \vec{x} \leq C\}$

$C = 15$

$$\vec{a} \cdot \vec{x} = \begin{pmatrix} 5 \\ 11 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 5 + 11 = 16 > 15$$

ergodic, aperiodic



Nearest Neighbor Random Walk

$X_t = (x_0, \dots, x_{n-1})$ at step t, $u \sim U(0,1)$

$$X_{t+1} = \begin{cases} X_t & \text{if } u < \frac{1}{2} \\ X_t' & \text{otherwise} \end{cases}$$

X_t': pick an item at random; if item in X_t, remove it; if not, add it

Metropolis-Hastings

$$P(\vec{x}) = \begin{cases} \frac{1}{Z} \exp(\beta \cdot \vec{a} \cdot \vec{x}) & \text{if } \vec{a} \cdot \vec{x} \leq C \\ 0 & \text{else} \end{cases}$$

β - "temperature", Z-normalizing const. $\begin{cases} \beta \uparrow = \text{exploitation} \\ \beta \downarrow = \text{exploration} \end{cases}$

$\hookrightarrow$ target distribution = fitness function (better solutions are more likely)

$$A(q' \mid q) = \min\left(1, \frac{P(q')}{P(q)}\right) = \min\left(1, \frac{\exp(\beta \cdot \vec{a} \cdot q')}{\exp(\beta \cdot \vec{a} \cdot q)}\right) = \min(1, \exp(\beta \cdot \vec{a} \cdot (q' - q)))$$

Simulated Annealing

$\beta(t) = c_0 \log(t)$, $c_0 > 0 \Rightarrow$ increasing temperature/exploration with time

1) $\mathcal{L} \leq \vec{a}\cdot\vec{x} \leq C$, minimize $\vec{a}\cdot\vec{x}$

a) $\vec{a} = (2,3,1,7)^T$, $C = 11$, $\mathcal{L} = 8$, $\vec{x} = (1,1,1,0)^T$

$\vec{a}\cdot\vec{x} = 6 < 8 \Rightarrow$ not feasible

b) $P(x) = \begin{cases} \frac{1}{Z}\exp\left(\frac{\beta}{\vec{a}\cdot\vec{x}}\right) & \text{if } \mathcal{L} \leq \vec{a}\cdot\vec{x} \leq C \\ 0 & \text{else} \end{cases}$

2) Vertex-Cover $S \subset V$, $G = (V, E)$

a) $\vec{x} = (x_1, x_2, \dots, x_n)^T = (1, 0, \dots, 1)^T$ if $x_i = 1$ i^{th} vertex in S

$\Omega = \{\vec{x} \mid \forall (u,v) \in E \ \ x_u + x_v \geq 1\}$

b) transition = remove/add vertex

c) $P(\vec{x}) = \begin{cases} \frac{1}{Z}\exp(-\vec{x}) & \text{if } \forall \{u,v\} \in E \ \ x_u + x_v \geq 1 \\ 0 & \text{else} \end{cases}$

d) exploitation, $\beta(t) = c_1 e^{c_2 t}$ $\begin{cases} A: c_1 = 0.16 & c_2 = 2 \\ T: c_1 = 0.04 & c_2 = 4 \end{cases}$

$\beta_A(t) = \beta_T(t) \Rightarrow 0.16e^{2t} = 0.04e^{4t} \Rightarrow 4 = e^{2t} \Rightarrow t = \ln 2$, $\beta'_A(t) = 0.32t\,e^{2t} < 0.16t\,e^{4t} = \beta'_T(t)$

3) a) Metropolis-Hastings, $\beta = 0$: Every transition picked is chosen.
Nearest-Neighbor Random Walk: $\frac{1}{2}$ probability of self loop.
} not equivalent

b) information about solution $\Rightarrow$ Nearest-Neighbor Random Walk in region > uniform sampling

c) Bias in random walks might make them less efficient than uniform sampling.

4) Inhomogeneous Markov Chain, $P(X_{t+1} = x_{t+1} \mid X_t = x_t)$ depends on t

Simulated Annealing, $P(X_t = b \mid X_{t-1} = a) \neq P(X_{t'} = b \mid X_{t'-1} = a)$, $t \neq t_1$