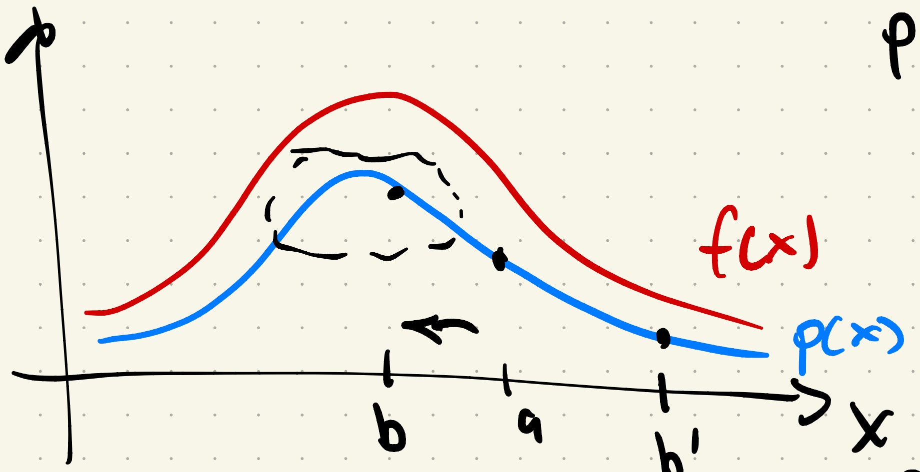
MCMC methods

- **Metropolis-Hastings Algorithm:** M-H

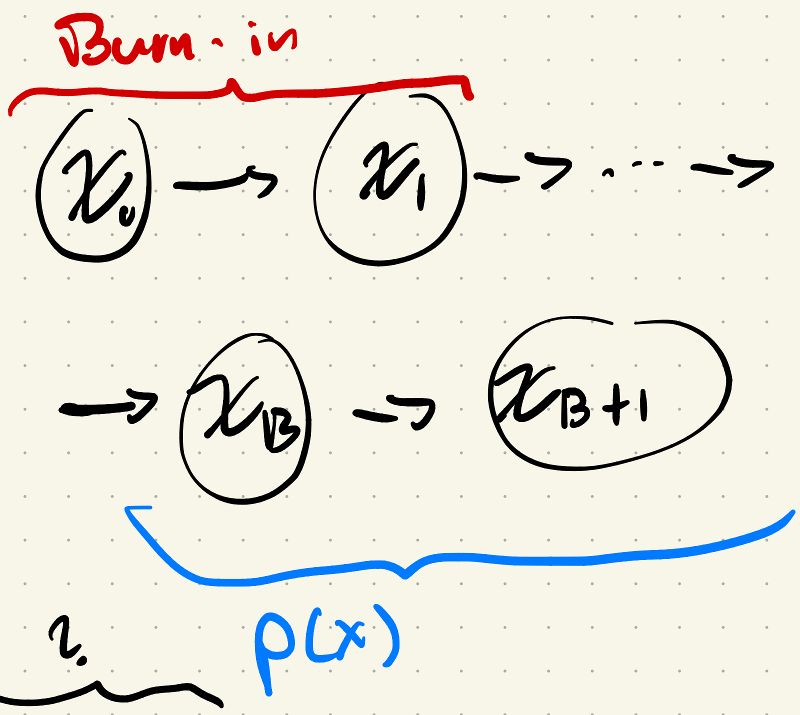
Pop. MCMC algo

Goal: Sample from $p(x)$

$$p(x) = \frac{f(x)}{NC}$$

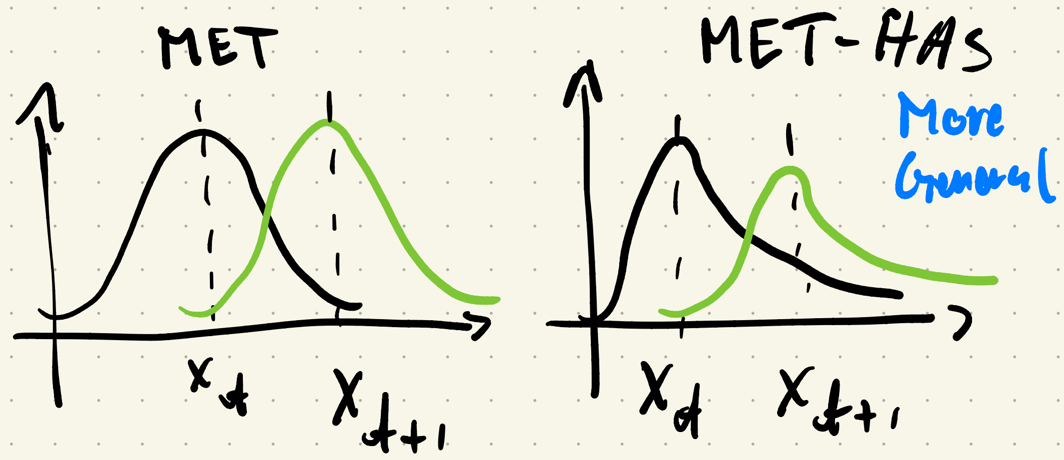


$A(a \to b) = 1$ $\quad A(b \to a) = \frac{p(b)}{p(a)}$



① Sample from $r(x_{t+1} | x_t)$

② Candidate next state $\overbrace{P_A(x_t \to x_{t+1})}^{?}$



~ Detailed Balance cond. $\forall a, b$

$$p(a) T(a \to b) = p(b) T(b \to a)$$

$$\frac{f(a)}{NC} g(b|a) A(a \to b) = \frac{f(b)}{NC} g(a|b) A(b \to a)$$

$$\frac{A(a \to b)}{A(b \to a)} = \underbrace{\frac{f(b)}{f(a)}}_{r_f} \underbrace{\frac{g(a|b)}{g(b|a)}}_{r_g}$$

MCMC

Global balance

$$\pi(x) = \int q(x|z)\,\pi(z)\,dz$$

Local Balance

Let X_k q and $\lambda(x)$

$$\lambda(x)\,q(z|x) = \lambda(z)\,q(x|z) \qquad x \overset{q}{\text{———}} z$$

Flow

$$\int \lambda(x)\,q(z|x)\,dz = \int \lambda(z)\,q(x|z)\,dz$$

$$\Longleftrightarrow \quad \lambda(x) = \int \lambda(z)\,q(x|z)\,dz$$

DONE

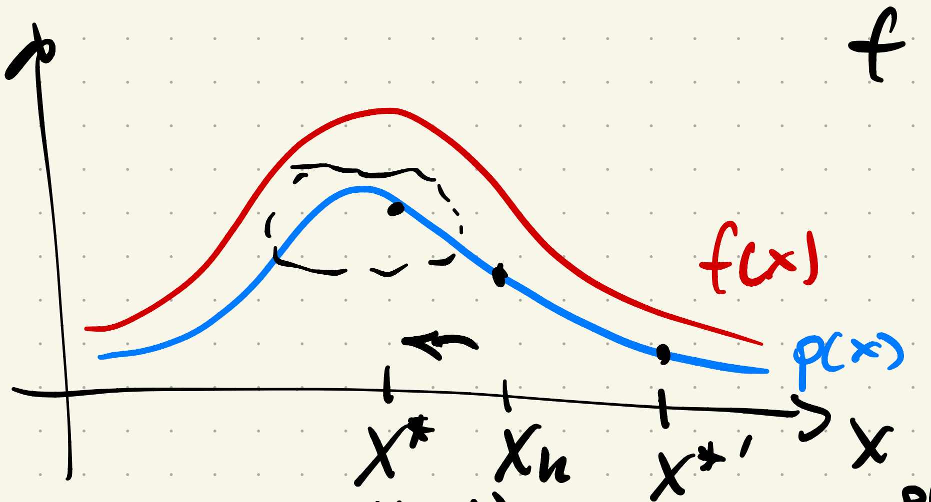
MCMC methods

- **Metropolis-Hastings Algorithm:** M-H

Pop. MCMC algo

Goal: Sample from $f(x)$

$$f(x) = \frac{Z(x)}{C}$$



$A(a\to b) = 1$ $\quad A(b\to a) = \frac{p(b)}{p(a)}$

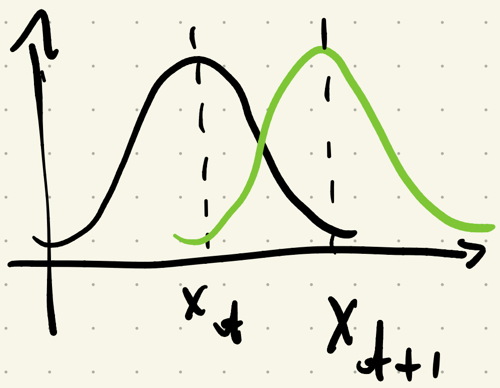
$$\underbrace{x_0 \to x_1}_{\text{Burn-in}} \to \cdots \to$$

$$\to \underbrace{x_B \to x_{B+1}}_{f(x)}$$

① Sample from $x^* \sim r(x_{t+1} | x_t)$.

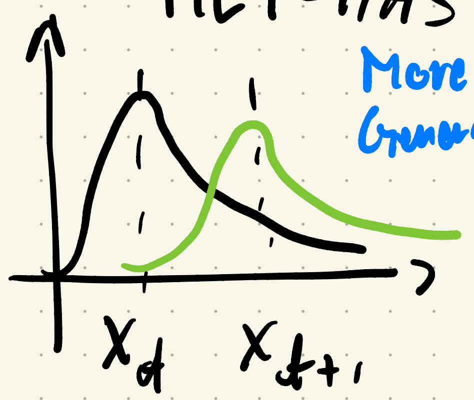
② Candidate next state $P_A(x_t \to x_{t+1})$

MET



MET-HAS

More General



$$X_{k+1} = \begin{cases} x^* & \text{if } \dfrac{f(x^*)\,r(x_k|x^*)}{f(x_k)\,r(x^*|x_k)} \\ x_k & \end{cases}$$

~ Local Balance cond. $\forall a, b$ Global Balance

$$\lambda(x)\, q(z|x) = \lambda(z)\, q(x|z)$$

$$\frac{f(a)}{NC} g(b|a) A(a\to b) = \frac{f(b)}{NC} g(a|b) A(b\to a)$$

$$\frac{A(a\to b)}{A(b\to a)} = \underbrace{\frac{f(b)}{f(a)}}_{r_f} \underbrace{\frac{g(a|b)}{g(b|a)}}_{r_g}$$

M-H algorithm

Init

$X_1 \sim \mathcal{X}$

for $k = 1 \to (N-1)$

$x^* \sim r(z \mid X_k)$

$$\alpha(X_k, X^*) \leftarrow 1 \wedge \frac{f(X^*)\, r(X_k \mid X^*)}{f(X_k)\, r(X^* \mid X_k)}$$

Draw $U \sim \mathcal{U}(0,1)$

if $U \leq \alpha$

$X_{k+1} \leftarrow X^*$

else

$X_{k+1} \leftarrow X_k$

set $\tau_N = \sum_{k=1}^{N} \frac{\phi(X_k)}{N}$

return τ_N

MCMC methods

$r_f r_g < 1$

$A(a \to b) = r_f r_g$

$A(b \to a) = 1$

$r_f r_g \geq 1$

$A(a \to b) = 1$

$A(b \to a) = \frac{1}{r_f r_g}$

So, $A(a \to b) = \text{Min}(1, r_f r_g)$

Intuition

$\boxed{\begin{array}{l}\text{Always accept } p(b) > p(a) \\ \text{Sometimes accept } p(a) > p(b)\end{array}}$

g : sym

$g(a|b) = g(b|a)$

$A(a \to b) = \text{Min}\left(1, \frac{f(a)}{f(b)}\right)$

proportional

$= \text{Min}\left(1, \frac{p(b)}{p(a)}\right)$

$p(b) > p(a) \Rightarrow A(a \to b) = 1$

$p(a) > p(b) \Rightarrow A(b \to a) = \frac{p(b)}{p(a)}$

MCMC methods

• Proof stationarity M–H

$f(x)$ – target

$\pi(x)$ – MC stationary distr.

π satisfies Detailed Balance Condition

Def: $\pi(x)P(x,y) = \pi(y)P(y,x), \ \forall x,y$

"In M–H $P(x,y)$ is prop prob $g(y|x)$ and accept prob $\alpha(x,y)$ and staying at x when y rejected.

We have

$$\pi(x)\overbrace{r(y|x)\alpha(x,y)}^{f(y|x)} = \pi(x)r(y|x)\min\left(1, \frac{\pi(y)r(x|y)}{\pi(x)g(y|x)}\right) =$$

$$= \min\big(\pi(x)g(y|x), \pi(y)g(x|y)\big)$$

Reverse transition

$$\pi(y)g(x|y)\alpha(y,x) = \pi(y)g(x|y)\min\left(1, \frac{\pi(x)g(y|x)}{\pi(y)g(x|y)}\right) =$$

$$= \min\big(\pi(y)g(x|y), \pi(x)g(y|x)\big)$$

$\forall x,y$ we have $\pi(x)P(x,y) = \pi(y)P(y,x)$

Detailed Balance Condition

fullfilled, thus π stationary distr of MC.

Q.E.D

MCMC methods

(A) Explain Gibbs Sampler, Hybrid Sampler.

(B) Hybrid Sampler valid MCMC algorithm.

(C) Why is good mixing important.

2.

(D) What can be said about mixing for MCMC samplers in course

---

(A)

- Gibbs Sampler: Multivariate prob distr $F(x_1, \ldots, x_n)$. Sequential sampling each variable from cond. distr. holding others fixed. For when cond. are known.

- Hybrid (Hamiltonian) sampler: Auxiliary momentum variables and sim. hamiltonian dyn to propose new states. Explores state space more efficiently. Larger steps and high accept rate maintained. Useful for continous prob. distr. with complex corr.

(B)

— Valid: Detailed balance and ergodicity.
Balance: always converge, ergodicity: can reach any state.

(C) Good mixing is exploring entire state space without getting stuck for too long.
Important as it becomes a better repr. of the entire distribution and more acc. estimater

MCMC methods

D

Skip lol

MCMC methods

Skip lol

Bootstrap

- Make inference of an estimand $\tau(\mathbb{P}_0)$ where $y \sim \mathbb{P}_0, f_0$
  $\hat{\tau} = t(y)$, statistic $t(\cdot)$
- Uncertainty of estimator : $\Delta(Y^*) = t(Y^*) - \tau$
- Bootstrap is data self sufficient with $\hat{\mathbb{P}}_0 \approx \mathbb{P}_0$

Non-Parametric

- $\hat{\mathbb{P}}_0$ is Empirical Distribution (ED). Y_i given weight $1/n$
- Does not assume any $Y \cancel{\sim} \mathbb{P}_D$ and instead directly resamples with replacement from population to create bootstrap samples. Purely data based approx

1. Generate Bootstrap sample from ED of size n
   $Y^* \sim \hat{\mathbb{P}}_0$ by drawing with replacement from y creating
   $Y^* = (y_{I_1}, \dots, y_{I_n})$ where $size(Y^*) = size(y)$.
2. Repeat above N times
3. Make analysis of interest on $\hat{\mathbb{P}}_0$ with $Y_1^*, \dots, Y_N^*$.
   $\tau(\mathbb{P}_0) \approx \tau(\hat{\mathbb{P}}_0)$.

Inference from histogram of $\Delta(Y_n^*) \approx t(Y^*) - \hat{\tau}, \forall n$

Flexible. Has no assumption of underlying D.

Dependent highly on original population sample, i.e $y \in Y$.

Bootstrap

- **Parametric**: assumes parametric approach, i.e data is drawn from a specific distribution

$Y \sim D(\theta)$ with density $f(\theta; x)$ and $F(\theta; x)$

1. Estimate $t(y) = \hat{\theta}$ from original dataset y
2. Generate Bootstrap samples $Y_1^*, \dots, Y_N^*$ by drawing with replacement from y.

   Ex: with $F^{-1}(\hat{\theta}; u)$ $u \sim \mathbb{U}$
3. Calculate new observations of $\hat{\theta}_n^*$, $n = 1, \dots, N$
4. Form $\Delta\hat{\theta}_n^* = \hat{\theta}_n^* - \hat{\theta}$ and find percentiles of $(1-\alpha)$ and form CI around $\hat{\theta}$. Potentially comp. effective.

Provides parameter estimates of model!

Useful when D assumption valid and improves!

May be unclear what D to assume.

Still dependent on sample observation. i.e $t(Y)$

Bootstrap

Semi-Parametric

A combination of both. Partly assumption on parametric model.

Ex: regression

$Y = \bar{\beta}\bar{X} + \varepsilon$, assume $\varepsilon \in N(0, D(\varepsilon))$ but no assumption on $\bar{\beta}$.

and check residuals

create boot resid w. resampl

$\varepsilon_b^* = (\varepsilon_1 \ldots \varepsilon_n)_b^*$ (no assumpt)

$Y_b^* \sim \hat{\beta}X + m$

$\hat{k}(y) =$ estimate params . . . .

$\hat{m}(Y) = (Y_i)_b^*$

$$I = \left(\hat{h} - t_{\alpha/2}(n-2)s_b, \ \hat{h} + t_{\alpha/2}(n-2)s_b \right)$$