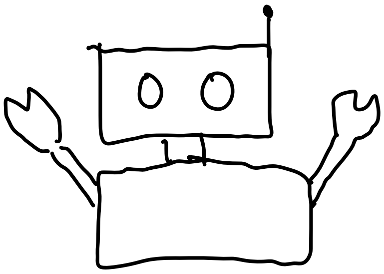
CS 331, Fall 2024
Lecture 21 (11/11)

Today:
- Reservoir sampling
- Rejection sampling
- MCMC

Reservoir sampling (Part VII, Section 4.1)

Goal in sampling: generate distributions from target π over universe Ω

Many apps!
- Parameter estimation
- Counting / integration
- Uncertainty quantification
- Generative models
- Privacy / fairness



ChatGPT / DALL-E / ...

Basic toy problem

$$\underbrace{w_1, w_2, w_3 \ldots \overset{\text{unknown end}}{w_n}}_{\text{arrive in a stream}}$$

Goal: return w_i for uniformly random $i \in [n]$

AKA: data structure supporting

- Stream(w): Add w to stream
- Report(): Output uniformly random streamed element

Trivial with $O(n)$ space: Store everything!

Reservoir: $O(1)$ space possible

(word RAM model, $w = O(1)$ space)

Implementation: store counter $i \in \mathbb{N}$ (init: 0)

reservoir element $\overline{w}$ (init: None)

Stream(w):

$(i+1)^{th}$ element

$$\overline{w} \leftarrow \begin{cases} \overline{w} & \text{w.p. } \frac{i}{i+1} \\ w & \text{w.p. } \frac{1}{i+1} \text{ (evict)} \end{cases}$$

$$i \leftarrow i+1$$

Report(): return $\overline{w}$

Why is $\overline{w} = w_j$ w.p. $\frac{1}{n}$, for all $j \in [n]$?

$$\Pr\left[\overline{w} = w_j\right] = \frac{1}{j} \cdot \frac{j}{j+1} \cdot \frac{j+1}{j+2} \cdot \ldots \cdot \frac{n-1}{n}$$

- evict on j
- no evict on $>j$

$$= \frac{1}{\cancel{j}} \cdot \frac{\cancel{j}}{\cancel{j+1}} \cdot \frac{\cancel{j+1}}{\cancel{j+2}} \cdot \ldots \cdot \frac{\cancel{n-1}}{n}$$

$$= \frac{1}{n}$$ as claimed!

Another interpretation: "fixing" distributions

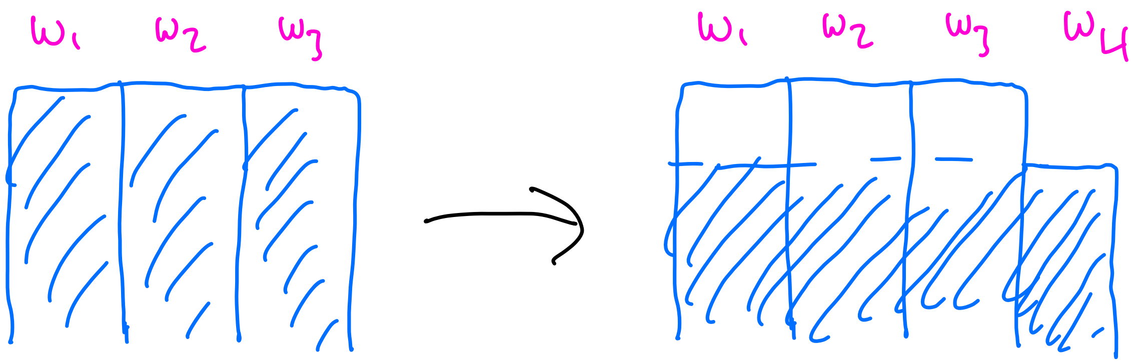
$\pi_k :=$ dist. of $\overline{w}$ after $w_1, w_2, \ldots, w_k$

We prove inductively π_k uniform over $\downarrow$ $\forall k$

Base case: $k=1$, $\overline{w} \leftarrow w_1$ w.p. 1

Induct:

$$\pi_{k+1}(w_j) = \begin{array}{ll} \overset{(\text{induct})}{\frac{1}{k}} \cdot \overset{(\text{no evict})}{\frac{k}{k+1}} & j \in [k] \\ \underset{(\text{evict})}{\frac{1}{k+1}} & j = k+1 \end{array}$$



$$\pi_3 \times \frac{3}{4} + \frac{1}{4} \delta_{w_4} = \pi_4$$

Rejection Sampling (Part VII, Section 4.2)

General strategy in sampling: you want $\sim \pi$

Let μ be some dist. such that

- We can sample $\sim \mu$
- We can relate μ to π

Crazy stuff... Box-Muller transform

$U_1, U_2 \sim \text{Unif}([0,1))$

then, $\sqrt{-2\log(U_1)}\cos(2\pi U_2) \sim \text{Nor}(0,1)$

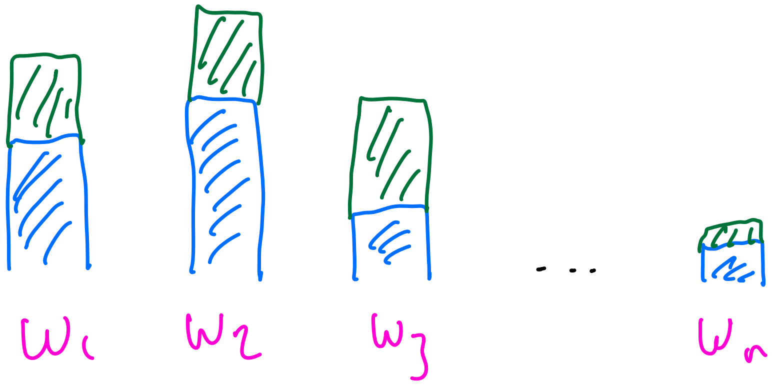
- requires explicit dist. formulas
- not very flexible

} we give simple framework w/ neither problem!

Idea: suppose $\forall\, \omega \in \Omega$,

$$\pi(\omega) \leq \mu(\omega)$$

then we could sample $\sim \mu$ and reject



Problem: $$1 = \sum_{\omega \in \Omega} \pi(\omega)$$
$$\leq \sum_{\omega \in \Omega} \mu(\omega) = 1$$

This idea only works if $\pi = \mu$...

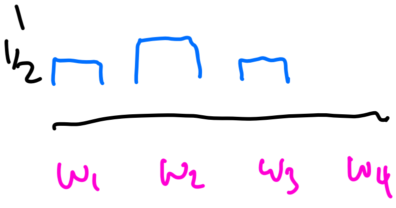
Fix: scaling / unnormalizing

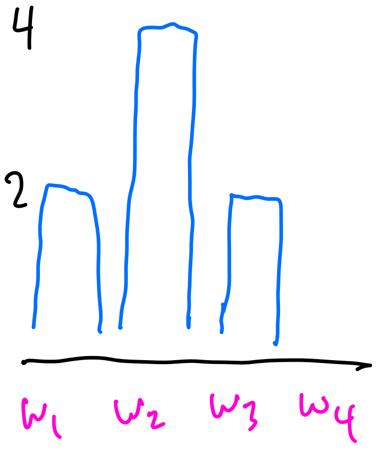
We write $\underset{\text{distribution}}{\pi} \propto \underset{\text{unnormalized distribution}}{P}$ for some $P : \Omega \to \mathbb{R}_{\geq 0}$

if: $$\pi(\omega) = \frac{\overset{\text{"weight"}}{P(\omega)}}{N_P} \quad \forall \omega \in \Omega$$

where $$N_P = \sum_{\omega \in \Omega} P(\omega)$$ "total weight" a.k.a. normalizing constant

Intuition: Unnormalized dists.





... all have same normalized dist.

Assume: $\pi \propto P$ (target) $\qquad \dfrac{P(\omega)}{Q(\omega)} \le 1 \quad \forall \omega$

$\mu \propto Q$ (reference)

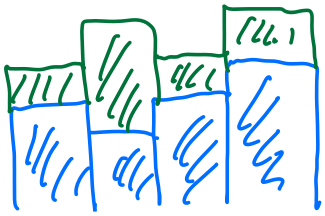
```
RejectionSample(P, Q, μ):
    While True:
        $\omega \sim \mu$
        $Z \leftarrow 1$ w.p. $\frac{P(\omega)}{Q(\omega)}$ else $0$
        If $Z == 1$: return $\omega$
```

Claim 1: $\mathbb{E}$ terminates in $\dfrac{\sum_{\omega \in \Omega} Q(\omega)}{\sum_{\omega \in \Omega} P(\omega)}$ loops

Claim 2: Samples exactly from P

Want: $P \le Q$

$$\Sigma Q = O(\Sigma P)$$



"total areas close"

Proof: $\Pr(Z=1) = \sum_{\omega\in\Omega} \mu(\omega)\cdot\frac{P(\omega)}{Q(\omega)}$

$$\boxed{\mu = \frac{Q}{N_Q}}$$

$$= \left(\sum_{\omega\in\Omega} P(\omega)\right)\cdot\frac{1}{N_Q}$$

$$= \frac{N_P}{N_Q}$$ chance / loop

We keep looping until first $Z=1$,

Like coin flipping: first heads $\mathbb{E}$ in $\frac{N_Q}{N_P}$

Done w/ Claim 1.

Conditional on $Z=1$, weight $\propto Q(\omega)\frac{P(\omega)}{Q(\omega)}$

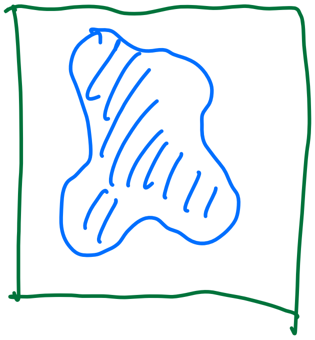
$$= P(\omega)$$

Done w/ Claim 2. $\blacksquare$

Examples

π = Uniform on complicated shape$_1$

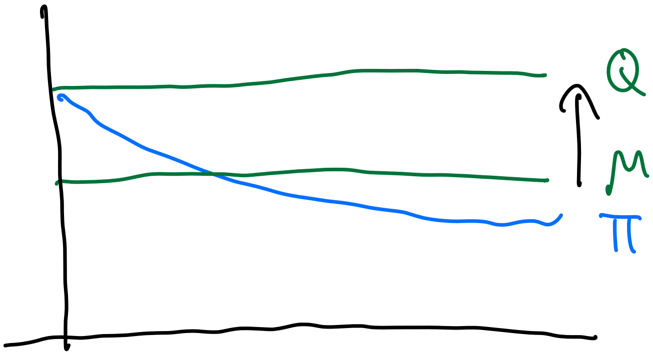
μ = uniform on simple shape$_2$ $\supseteq$



$P(\omega) = 1 \quad \forall\, \omega \in \text{shape}_1$

$Q(\omega) = 1 \quad \forall\, \omega \in \text{shape}_2$

Acceptance prob. = ratio of areas



π is "flat"

over domain Ω

$= [n]$

Idea: if $\frac{1}{2n} \leq \pi(i) \leq \frac{2}{n}$

Let μ = uniform on $[n]$, $Q = 2 \times \mu$

Then, $\frac{\Sigma Q}{\Sigma \pi} \leq \frac{2}{1} = 2$, accepts quickly!

Metropolis-Hastings (Part VII, Section 4.3)

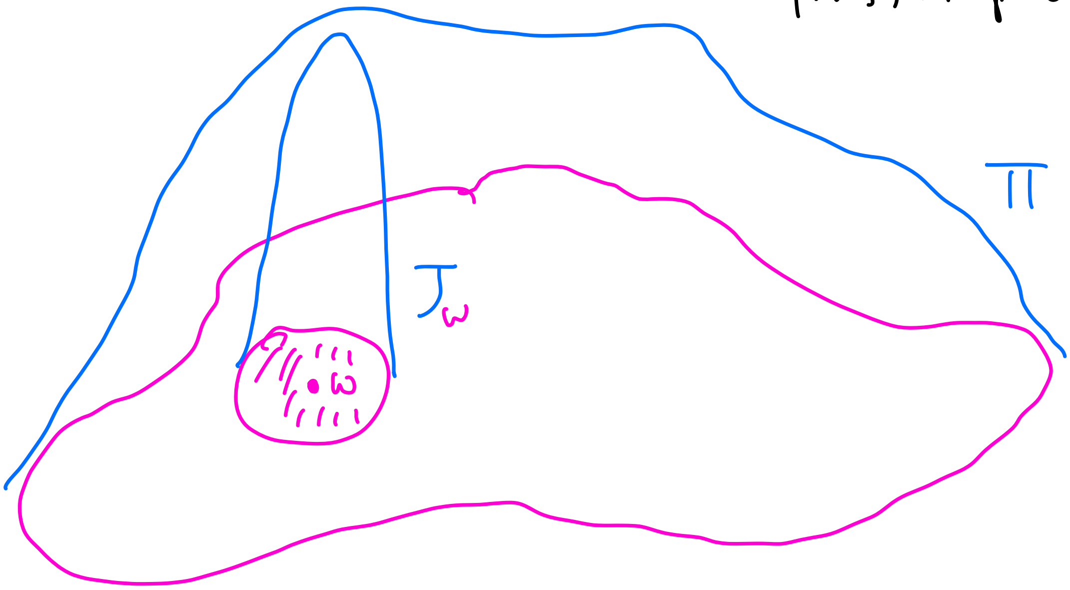
Predominant sampling framework today.

Top 10 algo of the 20th century.

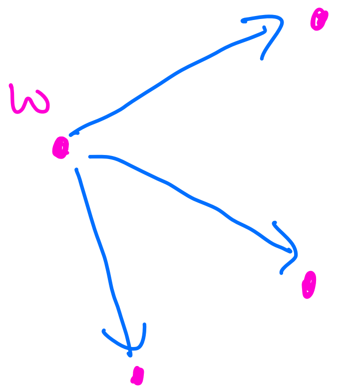
Idea: reduce sampling from large, global distribution $\pi : \Omega \to [0,1]$

to local distributions $T_w : \Omega \to [0,1]$

"transition probability"



Intuition: "random walk on graph"



One step: move $\sim T_w$

Many steps: hopefully $\sim \pi$?

Idea: T_w should look like π "zoomed in" @ w

Not the easiest to design / work with...

Fix: rejection sampling!

$w' \sim P_w$ (simple "proposal" dist.
e.g. uniform in small neighborhood)

Accept w.p. $\min\left(1, \frac{P_{w'}(w)\,\pi(w')}{P_w(w')\,\pi(w)}\right)$

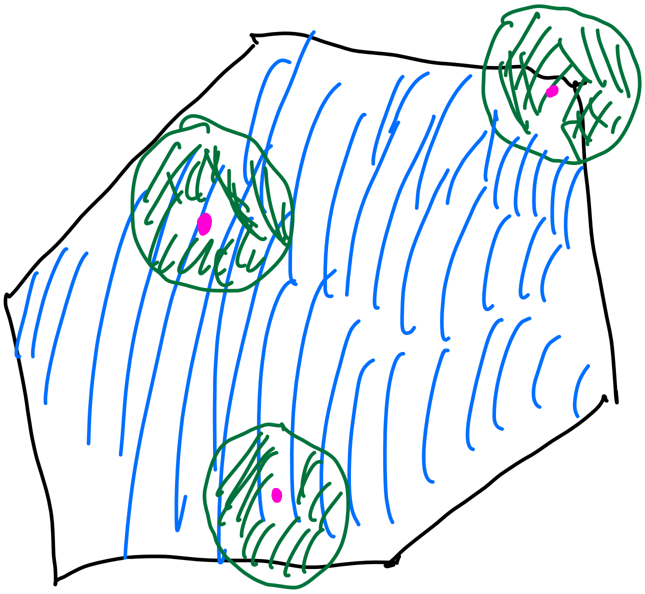
Main Claim (Metropolis-Hastings):

Correctness: π is the correct stationary, for *any* proposals $\{P_\omega\}_{\omega \in \Omega}$ (see notes)

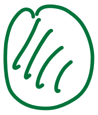
Runtime: Pick carefully to

- not reject too often
- "mix" quickly to π

e.g.



P_ω = unif. in ball @ ω

rejection = $\dfrac{\text{"surface area"}}{\text{"volume"}}$, mixing =  radius