Sampled representations:

crucial, useful fact

$x_i \sim g(x)$

then

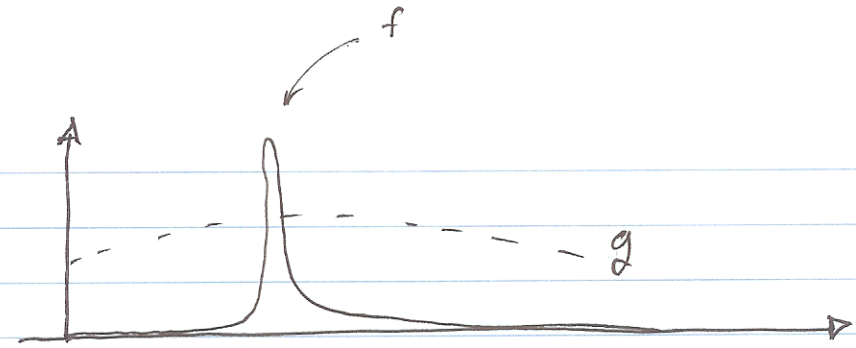
$$\frac{1}{N} \sum f(x_i) \longrightarrow \int f g \, dx$$

This gives a method to represent Prob densities by samples.

→ it's a <u>representation</u>, because we can <u>compute expectations</u>

~~Notice this~~

But we may not get good ests:



$x_i \sim g$ will most likely give a high variance estimate of

$$E_g[f] = \int f g \, dx$$

<u>Importance sampling:</u>

assume we know h, which is "similar" to fg (big when fg is big).

1) $x_i \sim h$

2) $$\frac{1}{N} \sum f(x_i) \cdot \frac{g(x_i)}{h(x_i)} \longrightarrow \int fg \, dx$$

will be a better estimate (lower variance)

This suggests a representation

$$\{(x_i, w_i)\} = \left\{\left(x_i, \frac{g(x_i)}{h(x_i)}\right)\right\}$$

Representing a posterior from a prior

we have

$$\underset{\text{likelihood}}{P(x/\theta)} \qquad \underset{\text{prior}}{P(\theta)}$$

- assume we have $\theta_i \sim P(\theta)$
- want to represent

$$P(\theta/x) = \frac{P(x/\theta)\,P(\theta)}{P(x)}$$

$$= \frac{P(x/\theta)\,P(\theta)}{\int P(x/\theta)\,P(\theta)\,d\theta}$$

<u>Now</u>

if $\quad w_i = p(x/\theta_i)$

then

$$\frac{1}{N}\sum_i w_i \longrightarrow \int p(x/\theta_i)\, p(\theta)\, d\theta$$

and

$$\frac{1}{N}\sum_i f(\theta_i)\, w_i \longrightarrow \int f(\theta)\, p(x/\theta)\, p(\theta)\, d\theta$$

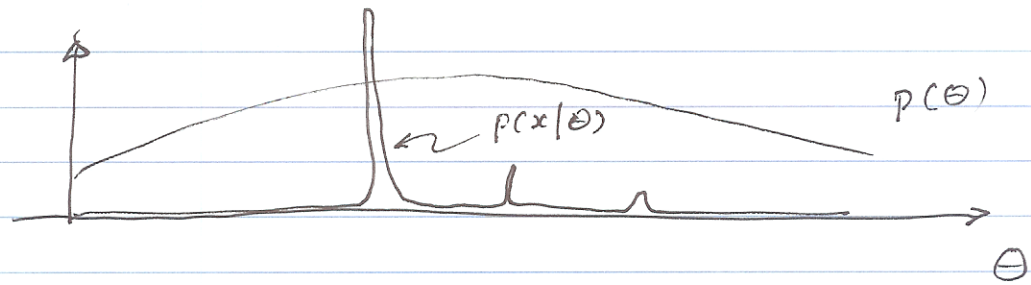
So

$$\frac{\sum_i f(\theta_i)\, w_i}{\sum_i w_i} \longrightarrow \int f(\theta)\; p(\theta/x)\, d\theta$$

<u>This gives one amazingly simple recipe for representing a posterior</u> !

<u>But</u>:

it's often not very good.



In particle filtering, the effect that results is sometimes called "sample impoverishment"

- Many w_i are v. small, few are big

$\Rightarrow$ very few θ_i contribute to the estimate of the integral in any significant way.

$\rightarrow$ <u>test</u>:

look at variance of w_i

How do we get samples?

- Uniform, Normal, Some (few) others:

  there are standard algs

- Finite mixtures of above:
  - draw a sample from mixture weights
  - then from component

Rejection sampling:

- wish to draw a sample from $p(x)$
- know how to draw from $q(x)$
- $p(x) \leq c\, q(x)$

<u>Algorithm</u>:

- repeat until accepted
- draw $x_i \sim q$
- draw $y \sim U[0,1]$
- if $y \leq \dfrac{p(x_i)}{c\, q(x_i)}$ accept

else go again.

1) Why does this work?

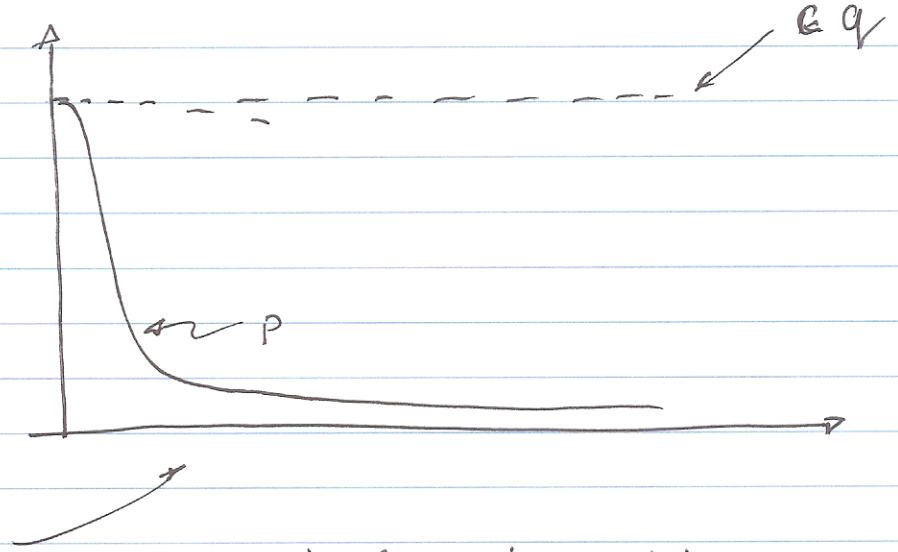
$P\{\text{sample at } x_i\} = P\{\text{generate sample at } x_i \text{ and accept}\}$

$$= q(x_i) \cdot \frac{p(x_i)}{c\, q(x_i)}$$

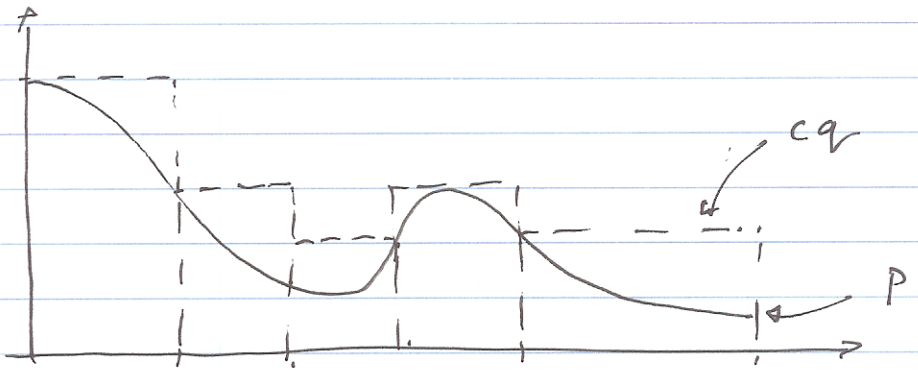
So $P\{\text{sample at } x_i\} \propto p(x_i)$

2) How efficient is it?

- depends on relationship between p, q.



many rejects in this case



here we have multiple intervals and draw from cq as a mixture

- Drawing a sample from a <u>discrete</u> distribution

- draw $x \sim U[0, 1]$ do point locatio

<u>Or</u>:

build a binary tree,
do rejection sampling

<u>Or</u>:

build a binary tree,
use a ~~binary~~ biased sample.

Gibbs sampling:

- assume we want to draw a sample from a multivar distrib.

$$p(u, v)$$

- assume $p(u|v)$ is easy
  $p(v|u)$ is easy.

- <u>Procedure</u>: start with (u_0, v_0)

  - draw $u_i \sim p(u|v_{i-1})$
    $v_i \sim p(v|u_i)$

  - do this many times

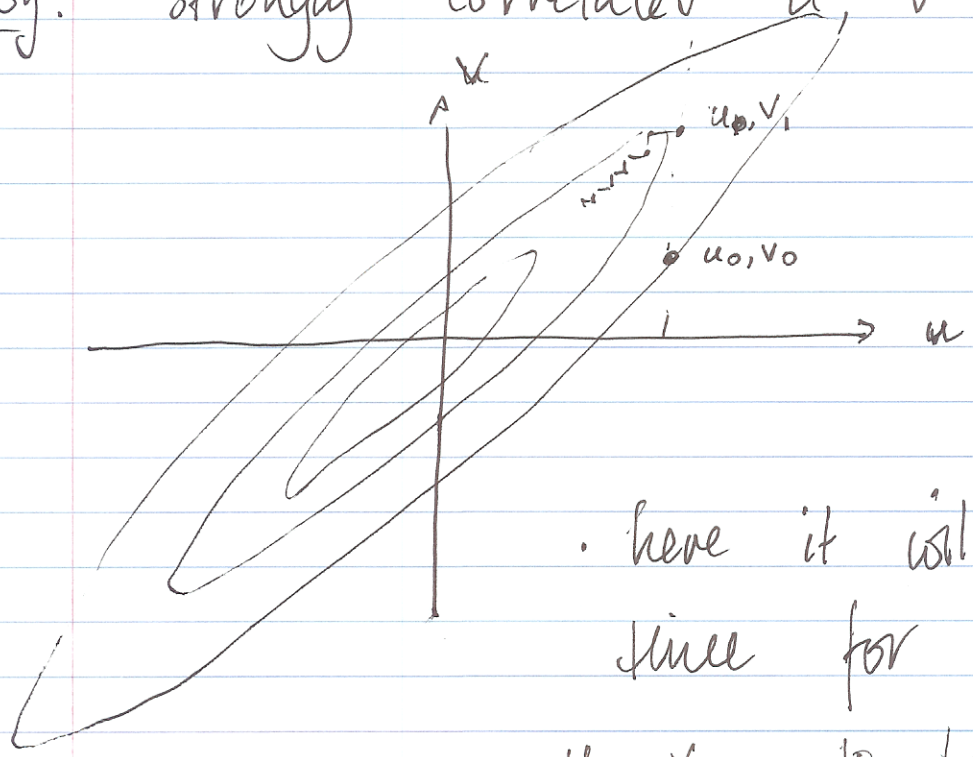
$$(u_n, v_n) \sim p(u, v)$$

(if we iterate often enough).

Eg $$p(u,v) \propto \exp\left[-\frac{1}{2}\left(\frac{u^2}{\sigma_1^2} + \frac{v^2}{\sigma_2^2}\right)\right]$$

- each stage is a draw from a univariate gaussian
- not much real advantage, because we could draw $(u,v) \sim N(0, Id)$ then scale.

Eg: strongly correlated u, v



- here it will take a long time for u_n, v_n to have high prob.

Markov chain

sequence of random vars,

$X_i \quad \cdots$ such that

$$P(X_{i+1} = x \mid X_i = \cdots \quad X_0 =)$$
$$= p(X_{i+1} = x \mid X_i = \cdot)$$

And this transition is stationary.

Discrete case:

Write $p(X_{i+1} = x_u \mid X_i = x_v) = P_{uv}$

some matrix

Now, start with $P(X_0 = x_w) = \pi_w$

$$P(X_1 = x_s) = \sum_{x_w} P(X_1 = x_s \mid X_0 = x_w)\, P(X_0 = x_w)$$

$$= \sum_{w} P_{uw} \cdot \pi_w = P\pi$$

Similarly:

$$P(X_i = x_s)$$

$$= \sum_{w_{i-1}, \ldots} P_{s w_{i-1}} P_{w_{i-1} w_{i-2}} \cdots P_{w_1 w_0} \pi_{w_0}$$

$$= P^k \pi$$

- Now P has at least 1 unit eigenvalue

  (because P^T does, $P^T 1 = 1$)

- Assume only 1; all others < 1

- then $P^k v \to \pi_s$ such that $P \pi_s = \pi_s$

$\boxed{\text{Stationary distribution}}$

This yields following procedure to sample from discrete dist π_S

- Build MC with π_S as stationary dist
- "run the chain".

eg to sample from

$1/2$

$1/2$

<u>chain</u>:

$$\begin{bmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{bmatrix}$$

- Something seems like a problem here $\longrightarrow$
- No guarantee that $\underline{x_i \perp\!\!\!\perp x_{i-1}}$!

Q: How do we build such a chain?

- Design requirement

$$\sum_{x_w} P(X_{i+1} = x_u \mid X_i = x_w)\, \pi(X_i = x_w) = \cancel{\pi_u}$$

$$= \pi(X_{i+1} = x_u)$$

or

$$\sum P_{uw}\, \pi_w = \pi_u$$

This is a weak condition, rather hard to achieve by choice of P

- <u>Detailed balance</u>

(stronger condition)

$$P_{uw}\, \pi_w = P_{wu}\, \pi_u$$

For each u, w

<u>Notice this implies above.</u>

Now we can build an MC in a <u>continuous domain</u> with stationary dist $\pi(x)$

1) obtain Proposal process $P(x \to y)$
(this is $P(X_{i+1} = y \mid X_i = x)$)

- $P(x \to y)$ should ~~have~~ satisfy some support conditions.
- Good enough if

$$\pi(y) > 0 \Rightarrow P(x \to y) > 0$$

$$P(x \to y) > 0 \Rightarrow P(y \to x) > 0$$

2) ~~Det~~ We achieve detailed balance by a form of rejection sampling

Algorithm

given x_i

- draw $x_{i+1}^{p} \sim P(x_i \longrightarrow y)$
- compute

$$\alpha = \min\left(1, \frac{P(x_{i+1}^{p} \to x_i)\,\pi(x_{i+1}^{p})}{P(x_i \to x_{i+1}^{p})\,\pi(x_i)}\right)$$

$$x_{i+1} = \begin{cases} x_{i+1}^{p} & \text{with prob } \alpha \\ x_i & \text{" " } 1-\alpha. \end{cases}$$

<u>Thm</u>: for large enough N,

$x_N \sim \pi$

whatever x_0

Amazingly powerful algorithm

+ • can draw samples from unnormalized dists

+ • easy; general

? + • can <u>sometimes</u> prove that chains are "fast mixing" (i.e. forget start point quickly)

? ~~m~~ • Samples are correlated

− • Correlation may be hard to est.

− • Can be hard to tell if sampler is "sticky"

− • "Sticky" samplers look better.