

## ① Laplace Approximation:

Used for integrals of the form:

$$\int h(\theta) q(\theta) d\theta \quad \text{central assumption.}$$

where  $q(\theta)$  is a possibly unnormalized smooth concentrated around its mode  $\theta_0$ .

Idea: Taylor Expand  $h(\theta)$  and  $q(\theta)$  around its mode.  $\theta_0 = \arg \max_{\theta} \log(q(\theta))$ .

$$\begin{aligned} q(\theta) & (\text{2 Order Taylor Expansion}): \\ l(\theta) & (\text{1. Order Taylor Expansion}): \end{aligned} \quad \left\{ \begin{array}{l} \text{expand for} \\ \log(\cdot) \text{ and} \\ \text{then transform} \\ \text{back for} \\ \text{simplification.} \end{array} \right.$$

Put the two together and leverage the symmetry of  $\frac{1}{2}(\theta - \theta_0)^2 l'(\theta_0)(\theta - \theta_0)$  that cancels with  $\theta - \theta_0$  you get

$$\int h(\theta) q(\theta) d\theta = h(\theta_0) q(\theta_0) \det l'(\theta_0)^{-\frac{1}{2}} \cdot (2\pi)^{\frac{1}{2}} \quad \left\{ \begin{array}{l} \text{This means that you can} \\ \text{use the mean of} \\ \text{such normal} \end{array} \right.$$

Given a posterior:

$$E(h(\theta) | x) = \int \frac{h(\theta) \pi(\theta) f(x|\theta) d\theta}{\int \pi(\theta) f(x|\theta) d\theta} \quad \left\{ \begin{array}{l} \text{Take } q(\theta) \text{ either} \\ = f(x|\theta) \text{ or } f(x|\theta) \cdot \pi(\theta) \end{array} \right.$$

you can use the Laplace approx. [separately] in the num. and denominator.

With this you can then show for instance consistency of the BIC estimator leveraging the fact that it stems from a Bayesian estimation.

## ② Monte Carlo Methods

Simple idea:

You have a target dist. such as the posterior dist.  
i.e.  $\theta$  in Bayesian stat.

Assume now that  $X \sim \pi$   
 $\pi(\theta|x)$  in Bayesian.

You are interested in the first moment of the posterior.

$$\mu_n = E_{\pi} (h(\theta)) = \int h(\theta) \pi(\theta) d\theta$$

first moment  
 $d\pi(\theta)$  under  
 $\pi(\cdot)$

Where  $h(\theta)$  has to have a bounded second moment i.e.  $\int h(\theta)^2 \pi(\theta) d\theta$ .

Then it is possible to get the desired first moment by taking the sample average.

$$\hat{\mu}_n \approx \bar{h}_n = \frac{1}{N} \sum_{i=1}^N h(X_i)$$

by CLT this converges at speed  $\frac{1}{\sqrt{N}}$  no matter the dimension of  $\theta$ !!

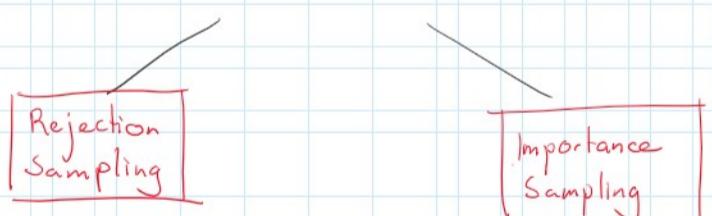
{ important however is that you sample from iid.  $X$ .

To guarantee this: pseudo random numbers & quantile transformation.

Problem: Quantile transformation is rarely feasible in Bayesian Stat.

b) Necessity to find other Methods.

Solution to difficulty of quantile transformation in Bayesian:



### ③.1) Rejection Sampling

The idea here is that instead of sampling from the target dist. you use a different dist. called the proposal dist.

In a second step you correct then such to get to the desired target.

So essentially: Step 1: Sample from a proposal

Step 2: Adjust the prop. to get to the target.

concretely this means:

Step 1: Sample independently from  $Y \sim \pi \sim U \sim \text{Uniform}(0,1)$

Step 2: Adjust according to the following rejection ratio:

$$a(x) = \frac{\pi(x)}{\pi(x) + (1-\pi(x))} \leq 1$$

i.e. accept if  $U \leq a(x)$  and set  $X = Y$ , otherwise continue with Step 1.

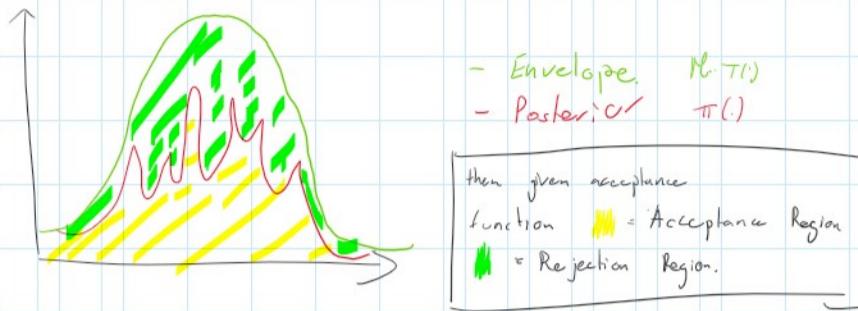
continue with Step 1:

Proof of equivalence:

$$\begin{aligned} P(X \in A) &= P(Y \in A \mid U \leq \pi(y)) \\ &\stackrel{\text{general measurable set } A}{=} \frac{P(Y \in A, U \leq \pi(y))}{P(U \leq \pi(y))} \quad \left\{ \begin{array}{l} \int_A \int_{U \leq \pi(x)} \pi(x) dU dx = \int_A \pi(x) \cdot \pi(x) dx = \frac{\pi(A)}{M} \\ \int_{(0,1)} \int_{U \leq \pi(x)} \pi(x) dU dx = \int_{(0,1)} \pi(x) \cdot \pi(x) dx = \frac{1}{M} \end{array} \right. \\ &\quad \text{thus } \pi \approx \frac{\pi(x)}{\pi(x) \cdot M} \end{aligned}$$

Intuition of Rejection Sampling:

- You are sampling from an envelope  $\boxed{\pi(x) \cdot M}$



! Important: You do not know the shape of the posterior ex-ante.

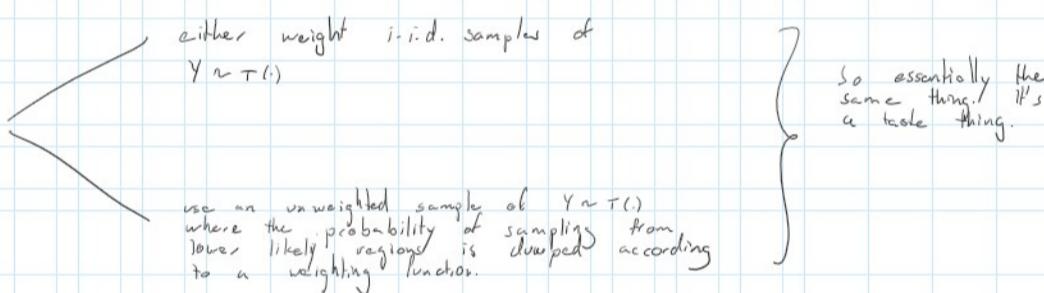
↳ Important is the choice of the  $M$  hyperparameter.

- Large  $M$ 
  - You ensure the necessary envelope property.
  - If  $M$  too big, you waste important computational resources. → too many rejections.

### 3.2 Importance Sampling

The idea is closer to the one of Rejection Sampling.

Instead of rejecting you:



#### 3.2.1 Weighted Importance Sampling

$$\text{Weight accordingly to: } w(x) = \frac{\pi(x)}{\pi(x)}, \text{ i.e.}$$

Step 1: Generate  $Y \sim T()$  and scale them according to  $w(x)$ .

Step 2: Take the first moment of weighted  $Y$ . CLT assures convergence.

Proof of equivalence:

$$E_T(h(Y) \cdot w(Y)) = \int_X h(x) \cdot \frac{\pi(x)}{\pi(x)} T(x) dx = \int_X h(x) \pi(x) dx \quad \square$$

The only requisite for applying the Method is  $h(x) \pi(x) > 0 \Rightarrow T(x) > 0$ .  
i.e. the support of the proposal must be at least as big as the one of the target.

Problem: In Bayesian Statistics we often know the target  $\pi(x)$  up to proportionality.

Solution: In this case you need to adjust importance sampling according to:

$$\frac{1}{N} \sum_{i=1}^N \frac{h(Y^i) \cdot v(Y^i)}{\sum_{j=1}^N v(Y^j)} \quad \text{, where now: } w(x) \propto \frac{\pi(x)}{v(x)}$$

So you basically normalize through this ratio term.

! Important this second version is baised but consistent.

↳ To see that consider:

$$\frac{1}{N} \sum_{x \in \mathcal{X}} h(Y^i) \cdot v(Y^i) \xrightarrow[\text{a.s.}]{\text{CLT}} \frac{\int_X h(x) w(x) \pi(x) dx}{\int_X w(x) \pi(x) dx}$$

$$\frac{\frac{1}{N} \sum_{i=1}^N h(Y^i) w(Y^i)}{\frac{1}{N} \sum_{i=1}^N w(Y^i)} \xrightarrow{\text{a.s.}} \frac{\int h(x) w(x) \tau(x) dx}{\int w(x) \tau(x) dx}$$

defining  $\pi^* = \frac{\pi(x)}{\tau(x)}$  it follows  $w(x) = \frac{\pi(x)}{\tau(x)}$

normalized.

$$\frac{\int h(x) \cdot \pi(x) dx}{\int \pi(x) dx} = \frac{\int h(x) \cdot \pi^* dx}{\int \pi^* dx} = c$$

### (3.2.2.) Unweighted Importance Sampling (i.e. Prob. Weighting Importance Sampling). Sampling Importance Resampling

Step 1: Given  $\{y_1, \dots, y_N\}$  draws from i.i.d.  $\tau(\cdot)$

$$P(Y^i = s) = \frac{w(Y^i)}{\sum_{r=1}^N w(Y^r)}$$

Probability of realization  $y^i$

Step 2: Set  $Z^+ = Y^i$  { i.e.  $y^i$  weighted (by exponential) with its probability of appearing in the proposed dist. }

$$\text{then } \frac{1}{N} Z^+ \rightarrow E_{\pi}(X)$$

Proof:

$$\frac{1}{N} \sum_{i=1}^N \sum_{s=1}^N P(Y^i = s) \cdot h(Y^i)$$

$$= \frac{\sum_{s=1}^N w(Y^s)}{\sum_{r=1}^N w(Y^r)} \cdot h(Y^s) \xrightarrow{\text{a.s.}} \frac{\int h(x) w(x) \tau(x) dx}{\int w(x) \tau(x) dx} = \mu_n \text{ (by previous proof).}$$

### Problems with Importance & Rejection Sampling.

Both require the proposed distribution to be close to the target distribution to reduce the computation power.

While in the case of rejection sampling the issue was already mentioned stating that the envelope should be as minimal as possible for importance sampling the reasoning is that if the proposal distribution is too different from the target; then:

$$\frac{w(Y^s)}{\sum_r w(Y^r)} = \frac{\pi(Y^s)}{\sum_r \pi(Y^r)}$$

is far from a uniform and we must sample multiple times to sample for the entire dist. space.  
(in case of the SIR) or simply because there is too much variance of the weights in importance sampling.

Solution for high-dimension where the problem of finding a similar proposal is especially tedious: MCMC.

### (4.) Markov Chain Monte Carlo.

The idea of MCMC is that instead from sampling from independent RV of the proposal distribution you can leverage:

- (i) Irreducible n state recurrent Markov Chains
  - or (ii) Reversible Markov Chains.
- For the both basic Markov Chain theory shows that a unique stationary / steady state solution exists.

#### Proof: Reversibility $\Rightarrow$ Stationary MC

Start from definition of stationary MC in cont. space:  $\pi(A) = \int_A \pi(x) \cdot P(x, A) dx \quad \forall A$

Then it is possible to see that given reversibility and a set  $B = \mathbb{R}^n$ :

$$\pi(A) = \int_A \pi(x) dx = \int_B \pi(x) \cdot P(x, A) dx \stackrel{\text{as prob}}{=} \int_B \pi(x) \cdot P(x, A) dx \stackrel{\text{given def. above.}}{=} \pi(A)$$

over entire set, the MC must move somehow there is the set so that prob=1.

You can therefore sample from such Markov Chains and once the chain converged to the stable stationary equilibrium average from such sample and get your  $\mu_{\pi}$  due to the CLT.

To summarize in an algorithmic way it consists of:

Step 1: Generate  $X^+ = G(X^{t+1}, U^+)$

where  $G$  is the transition rule for the markov chain being irreducible and state recurrent.

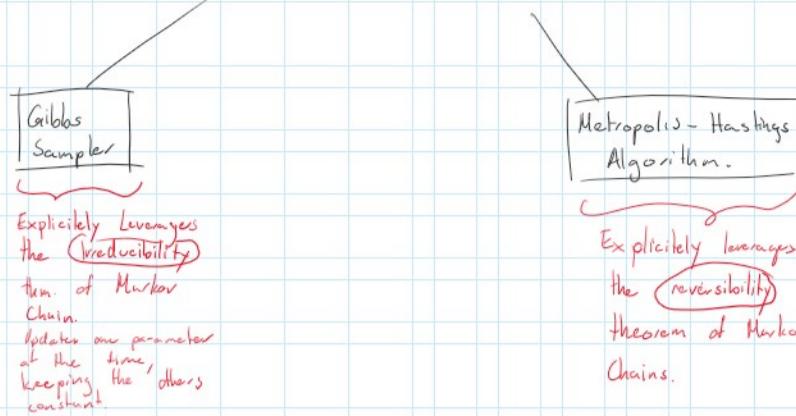
Step 2: Leverage the CLT by computing

$$\frac{1}{N-r} \sum_{i=r+1}^N h(X^i)$$

burn-in period necessary for the Markov chain to converge to its steady state.

There are now essentially two algorithms to generate the map  $G(X^{t+1}, U^+)$  that result in a transition kernel with the desired distribution as the

There are now essentially two algorithms to generate the map  $G(X^{t+1}, U^t)$  that results in a stationary distribution, these are:



### 4.1 Gibbs - Sampler

The idea of the sampler consists in drawing dependent samples depending on a visiting schedule  $t = 1, 2, \dots$ ; this can be either deterministic or random

You would sample then from:

$$X_{i,t}^+ \sim \pi_i(x_{i,t} | X_{-i,t}^{t+1}) \quad \text{all but } i \text{ entry}$$

and then update  $X_{-i,t}^{t+1} = \{X_{-i,t}^{t+1}, X_{i,t}^+\}$

such that then:

$$X_{i,t+1}^{t+1} \sim \pi_{i,t+1}(x_{i,t+1} | X_{-i,t+1}^t)$$

etc...

Given that this algorithm visit each component infinitely often you have a Markov Chain with the desired properties converging to the stationarity equilibrium.

$\pi_{i,t}$  are called full conditionals

and given  $\pi_{i,t} \propto \pi(x)$ ; we can leverage  $\pi(x)$  to estimate  $\pi_{i,t}$

Notice that to properly leverage the Gibbs sampler, you need to leverage the full conditional, i.e.

$$\pi(x_{i,t} | X_{-i,t}) = \frac{\pi(x_{i,t}, \dots, x_{n,t})}{\pi(x_{i+1}, \dots, x_{i-1}, x_{i+1,t}, \dots, x_{n,t})}$$

This has to be computed analytically!

### 4.2 Metropolis-Hastings Algorithm

The Metropolis-Hastings Algorithm explicitly leverages the reversibility theorem of Markov chains. It tries to explicitly build a Markov Chain such that the chain is reversible under the constraint that  $\int p(x,y) dy = 1$ ; i.e. the transition kernel is properly defined.

So that the logic is as follows:

Step 1: Choose an arbitrary transition density  $q(x,y) \approx q(y,x)$ .

Step 2: Assure  $\pi(y)p(y,x) = \pi(x)p(x,y)$  given the above

This can be achieved by choosing:

$$p(x,y) = q(x,y) \min(1, \frac{\pi(y) \cdot q(y,x)}{\pi(x) \cdot q(x,y)})$$

$$p(y,x) = q(y,x) \cdot \min(1, \frac{\pi(x) \cdot q(x,y)}{\pi(y) \cdot q(y,x)})$$

You have then  $\int p(x,y) dy \leq \int q(x,y) dy = 1$  so that you can put the missing mass on the diagonal  $p(x,x) = 1 - \int p(x,y) dy$ .

It is then possible to see that reversibility is guaranteed by:

$$\text{Case I: } \pi(y) \cdot q(y,x) < \pi(x) \cdot q(x,y)$$

$$\Rightarrow p(x,y) = \frac{\pi(y)}{\pi(x)} \cdot q(y,x) \quad p(y,x) = q(y,x)$$

and  $\pi(x) \cdot p(x,y) = \pi(y) \cdot p(y,x)$  or

$$\text{Case II: } \pi(y) \cdot q(y,x) > \pi(x) \cdot q(x,y)$$

$$\Rightarrow p(x,y) = q(x,y) \quad p(y,x) = \frac{\pi(x)}{\pi(y)} \cdot q(x,y)$$

and  $\pi(x) \cdot p(x,y) = \pi(y) \cdot p(y,x)$  given  $\Rightarrow$   
 $= \pi(y) \cdot p(y,x)$

The algorithmic way of implementing the above consists of a rejection like step where:

Step 1: Simulate  $x^0$

Step 2: Generate  $y \sim q(x^{t+1}, x) \wedge U^t \sim U(0,1)$  independently

Step 3: Set

$$x^+ = \begin{cases} y^+ & \text{if } U^t \leq \min(1, \frac{\pi(y) \cdot q(y, x^{t+1})}{\pi(x) \cdot q(x^{t+1}, y)}) \\ x^{t+1} & \text{else} \end{cases}$$

Notice that in contrast to standard rejection sampling if you

$$x^+ = \begin{cases} y^+ & \text{if } U \leq \min\left(1, \frac{\pi(y) \cdot q(x^{t-1}, y^+)}{\pi(x) \cdot q(y^+, x^{t-1})}\right) \\ x^{t-1} & \text{else} \end{cases}$$

3 keep the current value in accordance with  $P(x, y) = 1 - S_p(y|x)$  above.

Notice that in contrast to standard rejection sampling if you reject you stay with  $x^{t-1}$  and don't start from the scratch.

Practical Advice: Important for the algorithm above is the choice of  $q(x^{t-1}, \cdot)$ . A common choice is for a symmetric PDF such that the acceptance probability simplifies to

$$\min\left(1, \frac{\pi(y)}{q(x, y)}\right) = \min\left(1, \frac{\pi(y)}{\pi(x)}\right).$$

i.e. always accept if the proposed  $y^+$  is more likely in the target distribution we want sample from than  $x^{t-1}$ . Otherwise just accept with prob as to likelihood ratio  $\frac{\pi(y^+)}{\pi(x^{t-1})}$ .

A common choice for such as a symmetric PDF is the normal; i.e.  $y^+ \sim N(x^{t-1}, \Sigma)$ .

Notice finally that it is possible to combine the two methods above in Applications.

Practical Advice: In theory no matter the choice of  $\Sigma$  for the RW-Metropolis-Hastings Algo you are assured that as  $N \rightarrow \infty$   $x^{t-1} \sim \pi \rightarrow x^+ \sim \pi$ .

In practice however as  $N$  is finite the choice of  $\Sigma$  matters.

Numerical experiment have proven that sensible choices for such hyper-parameters are:

$$(i). \quad \Sigma = \frac{2,38^2}{P} \text{Cov}_{\pi}(x)$$

(ii) Average  $P(Y)$  being accepted = 0,239; i.e.

$$P(Y \text{ accepted} | x^{t-1}) = \int q(x^{t-1}, y^+) \cdot a(x^{t-1}, y^+) dy$$

$$P(Y \text{ accepted}) = \int_{\pi(x)} \int q(x^{t-1}, y^+) \cdot a(x^{t-1}, y^+) dy dx = 0,239$$

As you cannot solve (i) & (ii) directly as an analytic formula for  $\pi$  is not known you must find alternative solutions.

An especially important in this sense is the one of iteratively adjusting the hyperparameters at each step (i.e. adaptive MCMC). Thus:

$$(i) \quad \Sigma^+ = \frac{2,38^2}{P} \frac{1}{t-1} \sum_{s=0}^{t-1} (x^s - \bar{x}^{t-1})(x^s - \bar{x}^{t-1})^T \quad (\text{i.e. sample covariance}).$$

(ii) Optimizer  $\Sigma^+$  scale i.e.  $\sigma^2 I_P$  given optimal acceptance level as above; i.e.

$$\sigma^{2+} = \begin{cases} r + \sigma^{2-1} & \text{if } \frac{1}{t-1} \sum_{s=0}^{t-1} \min\left(1, \frac{\pi(y^{s+1})}{\pi(x^s)}\right) > 0,239 \\ \frac{1}{r} \sigma^{2-1} & \text{if } \frac{1}{t-1} \sum_{s=0}^{t-1} \min\left(1, \frac{\pi(y^{s+1})}{\pi(x^s)}\right) < 0,239 \end{cases}$$

where  $r < l$

sample average of acceptance level.

#### 4.3. Hamiltonian MCMC.

The issue with Gibbs Sampler or the Metropolis-Hastings Algorithm is that the first just makes small movements and the second have often just a low acceptance probability for big movements.

A solution in this sense was found in applying a statistical physical analogy such as the one of Hamiltonian Mechanics.

Idea of Hamiltonian Mechanics:

You have a particle in different states with different probability, where the probability of being in a higher energy state is inversely proportional to the amount of energy in the system:  $P(E_i) \propto e^{-E_i/kT}$ .

Modeling the Energy level in the system as the sum of kinetic and potential energy it is then possible to rewrite

$$E(\theta, M) = \underbrace{U(\theta)}_{\substack{\text{position} \\ \text{of the particle} \\ \text{in the various} \\ \text{energy states}}} + \underbrace{K(M)}_{\substack{\text{potential} \\ \text{energy} \\ \text{kinetic} \\ \text{energy}}}$$

Modeling the two energy components as:

$$U(\theta) \propto -\log(P(X|\theta) \cdot P(\theta)) \quad [\text{unnormalized posterior}]$$

$$K(M) = \frac{M^2}{2\text{mass}}$$

so that the probability of each energy state becomes:

$$P(\theta, M) \propto e^{-\log(P(X|\theta) \cdot P(\theta)) - \frac{M^2}{2\text{mass}}} := H(\theta, M)$$

$$\propto P(X|\theta) \cdot P(\theta) \cdot e^{-\frac{M^2}{2\text{mass}}} \quad (1) \quad \text{Marginalizing}$$

$$\text{oc } P(X|O) \cdot P(O) \cdot e^{-\frac{m^2}{2\text{mass}}} \quad (1)$$

Normalizing constant of  $N(0, 1)$

$$\int P(O, M|x) d\mu = \frac{1}{Z} P(X|O) \cdot P(O) \int e^{-\frac{m^2}{2\text{mass}}} dm = 1$$

Marginalizing w.r.t. the momentum to get the prob. of each energy state

$$P(O|x) = \frac{1}{Z} P(X|O) \cdot P(O)$$

It follows therefore that generating samples  $(O, M)$  for the energy system described above you get samples from your desired posterior distribution. If  $(O, M) \sim \pi$  then  $O \sim \pi$  as from (1) you can see that  $M$  is a simple scaling factor; i.e. the one that allows to make big movements in the MC sampling and that are still accepted with high probability.

In order to do that you therefore have to:

- Generate  $M \sim N(0, 1)$

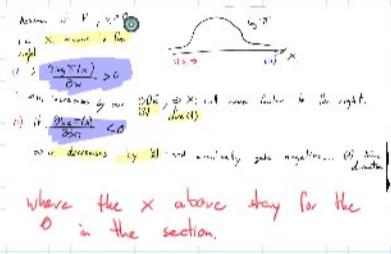
- Solve for the path of the particle  $v(O) \sim k(M)$  in time.

i.e. solve according to the Hamiltonian Mechanics: the ODE:

$$(i) \frac{dO}{dt} = \frac{\partial H(O, M)}{\partial M} = \frac{M}{\text{mass}}$$

$$(ii) \frac{dM}{dt} = \frac{\partial H(O, M)}{\partial O} = \frac{\partial \log(P(O) \cdot P(X|O))}{\partial M}$$

Behind this ODE lies the exact intuition of why such sampling Method works.



the solution of such an ODE will yield an invariant and volume preserving map  $G(O, M)$ .

The map can be obtained by solving the ODE above by a discretization method such as the leap frog method.

A final step Metropolis step will transform the time-invariant map in a reversible Map such that  $O$  will satisfy the MC theorem.

Metropolis-step Acceptance probability:

$$a((O, M), (O^*, M^*)) = \min \left( 1, \frac{H(O^*, M^*)}{H(O, M)} \right)$$

Now generated  $(O, M)$  proposal

this does not keep the map exactly invariant and hence we require the Metropolis-Algo at a later step

Algorithmically all of the above summarizes to:

#### Algorithm (Hamiltonian Monte Carlo algorithm)

Simulate  $(X_0, U_0)$   
For  $t = 1, 2, \dots$

- 1a. Simulate  $U' \sim N(0, \text{diag}(m_i))$
- 1b. Use the leap frog method (or any other method that results into an invertible  $G$  that is volume preserving) to generate a proposal  $(X^*, U^*) = G(x_{t-1}, U')$  this can be considered a Gibbs step.
2. Simulate  $V \sim \text{uniform}(0, 1)$ . If  $V \leq a((x_{t-1}, U'), (X^*, U^*))$  set  $X_t = X^*$ , otherwise  $X_t = x_{t-1}$  this is the classical Metropolis step.

## 5. Sequential Monte Carlo

Instead of sampling from one single target  $\pi$ , one samples from a sequence of related targets. So this is especially advisable in the case of multimodal distribution as it has better chances to sample from regions around all modes.

The idea is to start from a simple distribution such as  $\pi_0$  (being a simple dist - for instance a uniform one).

Then it is possible to leverage a transition kernel to propagate a sample  $X^{0t} \sim \pi_0$  into a  $Y^{1t} \sim \int \pi_0(x) p(x^{0t}, y) dy$  with the new density  $p(Y^{1t}) \sim \int \pi_0(x) p(x^{0t}, y) dy$

Given that we are using a proposal distribution  $\pi_0$  to get to a target distribution  $\pi_1$  we need to apply an weighting function such as the importance sampling one as  $Y^{1t}$  will not have the target distribution  $\pi_1(Y^{1t})$

i.e. we could take the weighting function

$$w^{1t} \propto \frac{\pi_1(Y^{1t})}{\sum_{k=1}^K \pi_k(x^{0t}) p_k(x^{0t}, Y^{1t})} \quad \text{where this is the posterior; given the task } k_j \text{ in this case: } x^{0t}.$$

Problem: As usually it is not possible to integrate over the densities  $\pi_k$  the denominator is not analytically defined.

Problem: As usually it is not possible to integrate over the densities  $\pi_k$  the denominator is not analytically defined.

Solution: Instead of looking at the individual  $y^{k,t}$  look at the  $(x^{k-1,t}, y^{k,t})$ , pairs. These have  $\pi_{k-1}(x) p(x,y)$  as density. It is then possible to use:

$$\text{target: } q_{k-1}(y|x) \cdot \pi_k(y) \quad \{ \text{as this is simply rewriting } \int q_{k-1}(y|x) \pi_k(y) dx = \pi_k(y) \int q_{k-1}(y|x) dx = \pi_k(y) \}$$

proposed:  $\pi_{k-1}(x) p_k(x,y)$

Then you can use importance sampling weights:

$$w^{k,t} \propto \frac{q_k(y|x) \pi_k(y)}{\pi_{k-1}(x) p_k(x,y)} \quad \wedge \quad w^{k,t} = \frac{w^k(x^{k-1,t}, y^{k,t})}{\sum_{s=1}^N w^k(x^{k-1,s}, y^{k,s})}$$

and get

$$x^{k,t} = y^{k,t} \quad P(J^t = s) = w^{k,s}$$

The choice of the transition kernel is left to the reader. Some options involve:

for  $p_k(x,y)$ :

- independent moves  $p_k(y)$
- random walk moves:  $p_k(x,y)$  symmetric around the mean.
- MCMC moves: set  $p_k(x,y)$  as the density of MCMC kernel with  $\pi_k$  as the stationary dist.

for  $q_k(x,y)$ : (backward transition kernel)

optimal choice not available analytically.

$$\rightarrow \text{Solution: } q_k(y|x) = \frac{\pi(x)}{\pi(y)} \cdot p(x,y) \quad [\text{strong analogy to rejection sampling}].$$