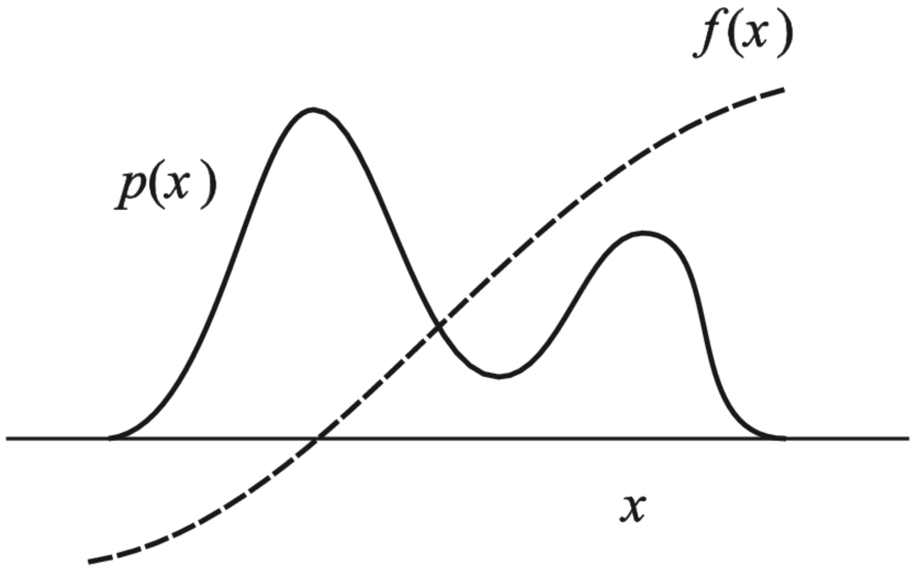
Sampling

Goal: Independent samples $x^{(1)}, x^{(2)}, \ldots$ from target probability distribution p

Estimating expectation $E[f(X)]$ where X from p

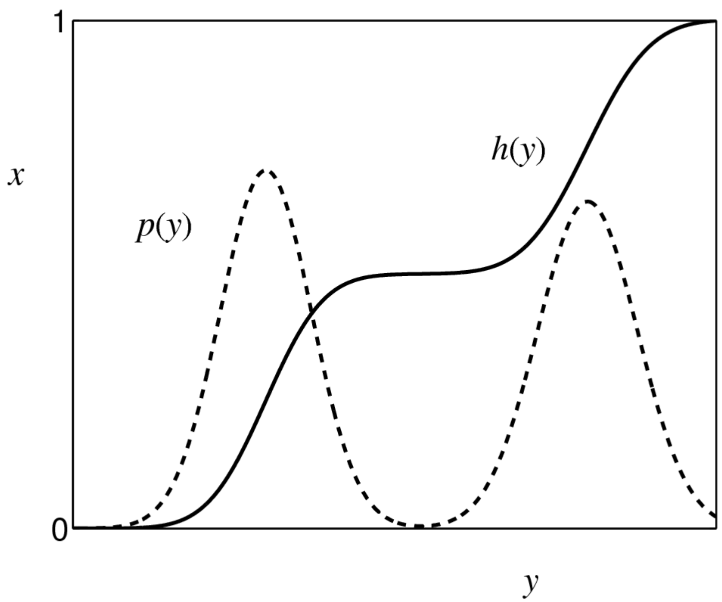
$\hat{f} = \frac{1}{M} \sum_{m=1}^{M} f(x^{(m)})$

$E[\hat{f}] = E[f(X)]$ even if samples not independent, but each from p



Sampling from basic distributions, transformation method

- Pseudorandom number generators: uniform samples from $\{0, \ldots, N-1\}$
- Divide by N to get "uniform" samples from $[0, 1)$
- Other simple distributions by transformation $y = h^{-1}(x)$
  x is from $U(0, 1)$, h is CDF of the target distribution p



Sampling from categorical (finite discrete) distribution

- Vector of probabilities $p(0), \dots, p(m-1)$
- CDF - prefix sums, transformational method by binary search in $O(\log m)$ per sample, $O(m)$ preprocessing
- Alias method by Kronmal and Peterson (1979) needs $O(1)$ per sample, $O(m)$ preprocessing
  (alternatively many samples at once by multinomial distrib.)
- Split probability space to m slices,
  each size $1/m$, with ≤ 2 different elements from $\{0, \dots m-1\}$
- Uniformly choose slice, then sample from items within the slice
- To create slices, split elements into two lists: with prob. $\leq 1/m$ and with prob. $> 1/m$. Always use one from the first list completely and a part of one from the second list.

Sampling from graphical models

Easy in acyclic directed models if no variables observed

Sampling state paths in HMMs (stochastic traceback)

Goal: sample a state path given observations

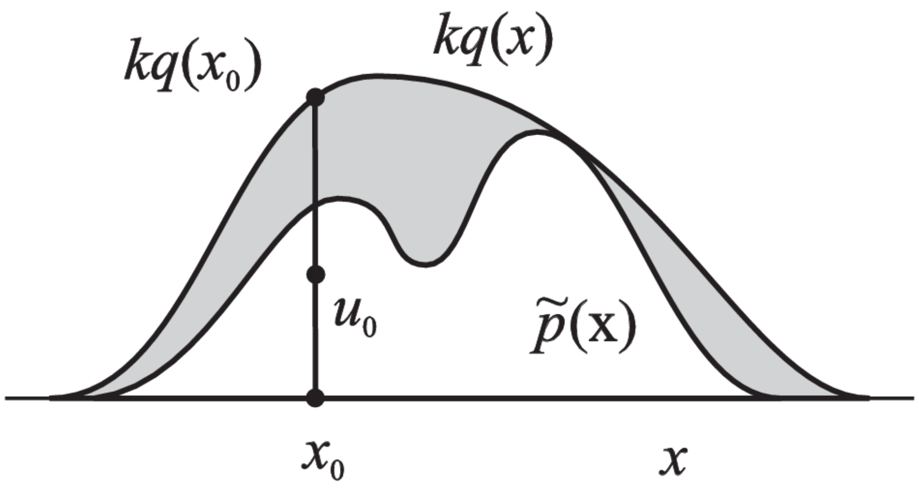
Alternative formulation:

- Given DAG with a single source s and single sink t
- Edges weighted by positive numbers
- Weight of a path product of edge weights
- Sample a path proportionally to path weights

Can be generalized to a junction tree?

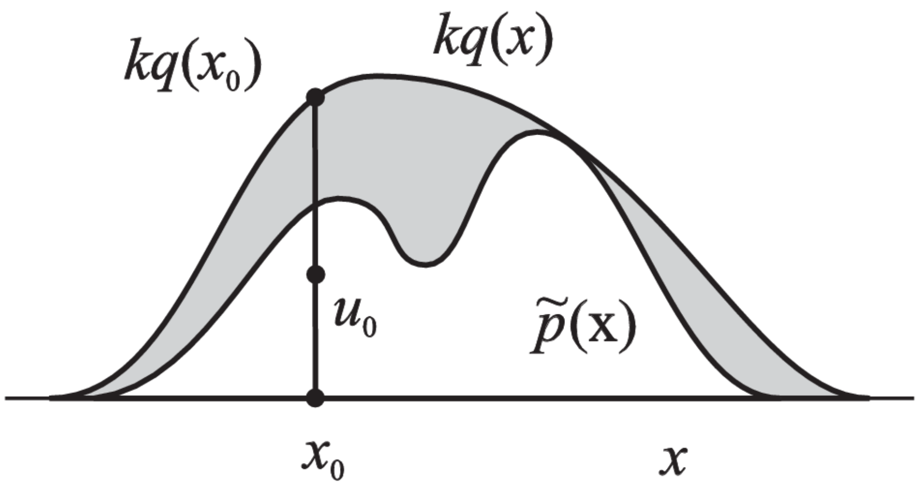
Rejection sampling

- Assume we cannot sample from p,
  but we can evaluate $p(x)$ up to a constant $p(x) = \frac{1}{Z}\widetilde{p}(x)$
- Proposal distribution $q(x)$ and constant k s.t. $kq(x) \geq \widetilde{p}(x)$
- Sample x_0 from q, then u_0 from $U(0, kq(x_0))$
- If $u_0 \geq \widetilde{p}(x_0)$, keep sample $x = x_0$, otherwise reject



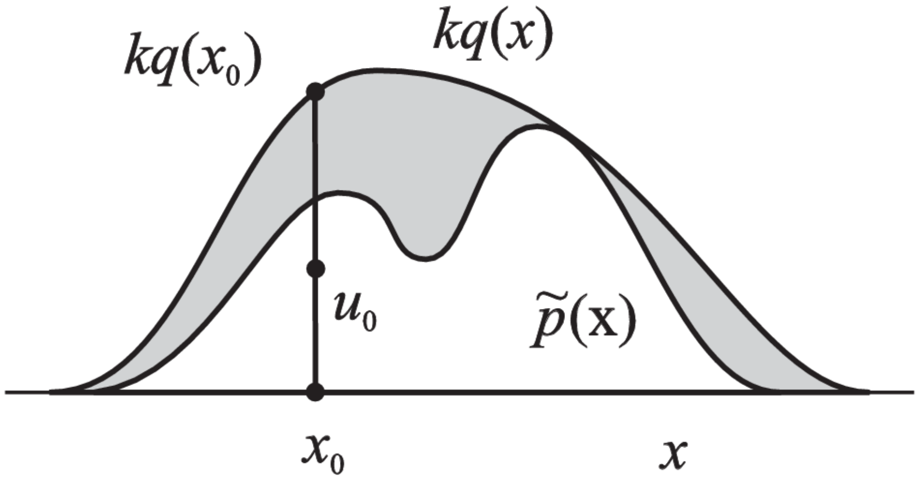
Rejection sampling

- $p(x) = \frac{1}{Z}\widetilde{p}(x)$, $kq(x) \geq \widetilde{p}(x)$
- Sample x_0 from q, then u_0 from $U(0, kq(x_0))$
- If $u_0 \leq \widetilde{p}(x_0)$, keep sample $x = x_0$, otherwise sample again
- $\Pr(X = x) \propto \Pr(x_0 = x) \Pr(u_0 \leq \widetilde{p}(x) \mid x_0 = x) = \frac{q(x)\widetilde{p}(x)}{kq(x)} = \frac{\widetilde{p}(x)}{k} \propto p(x)$



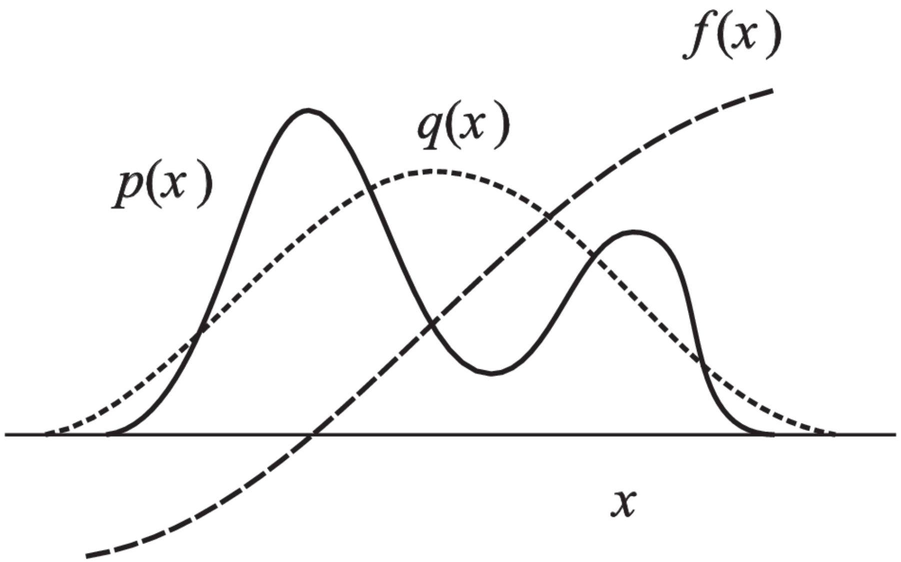
Rejection sampling (cont.)

- Needs good q and k
- Rejection probability tends to increase exponentially with the number of dimensions



Importance sampling

- Sample from proposal q, then estimate $E_p[f(X)]$
- $\hat{f} = \frac{1}{M}\sum_{m=1}^{M} f(x^{(m)})\frac{p(x^{(m)})}{q(x^{(m)})}$
- $E_q\left[f(X)\frac{p(X)}{q(X)}\right] = \sum_x f(x)\frac{p(x)}{q(x)}q(x) = \sum_x f(x)p(x) = E_p[f(X)]$



Importance sampling (cont.)

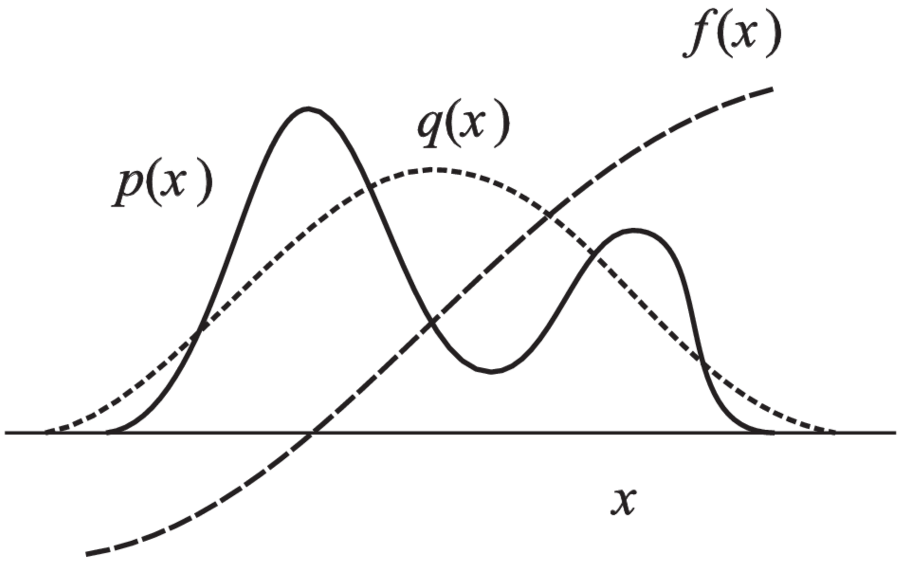
- If p and q can be evaluate only up to constant: $p(x) = \widetilde{p}(x)/Z_p$, $q(x) = \widetilde{q}(x)/Z_q$
- Let $r_m = \widetilde{p}(x^{(m)})/\widetilde{q}(x^{(m)})$
- $\hat{f} = \dfrac{\sum_{m=1}^{M} r_m f(x^{(m)})}{\sum_m r_m}$
- $E_q(r_m) = \sum_x \dfrac{\widetilde{p}(x)}{\widetilde{q}(x)} q(x) = \sum_x \dfrac{Z_p \cdot p(x)}{Z_q \cdot q(x)} q(x) = \dfrac{Z_p}{Z_q}$
- $E_q\left[r_m(X) f(X)\right] = \sum_x \dfrac{Z_p \cdot p(x)}{Z_q \cdot q(x)} f(x) q(x) = \dfrac{Z_p}{Z_q} E_p[f(X)]$

Importance sampling (cont.)

- If p and q differ too much, small weights
- If $q(x)$ very low where $f(x)p(x)$ high,

  no sample might come from this region,

  then estimate might be very bad, but we do not see it



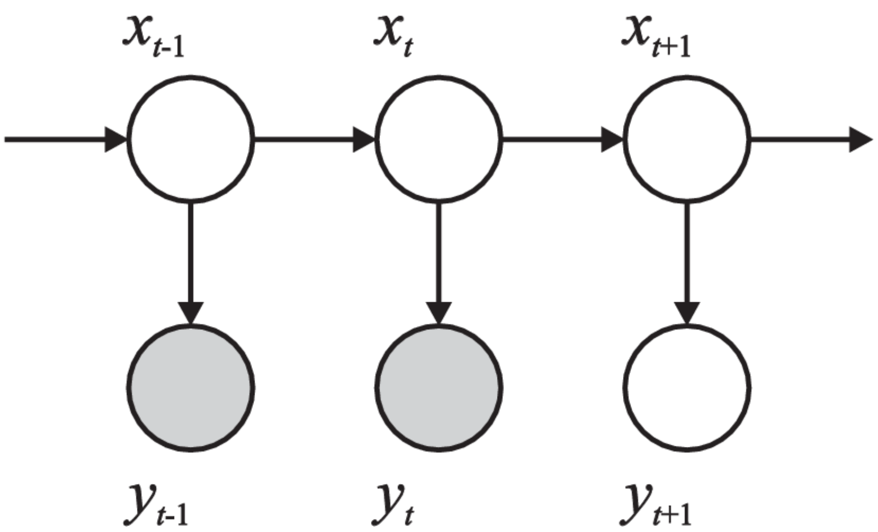
Particle filter (sequential Monte Carlo)

Given observations $y_{(t)} = (y_1, \dots y_t)$,

sample from $p(x_t \mid y_{(t)})$ to evaluate $E_{p(x_t \mid y_{(t)})}[f(x_t)]$

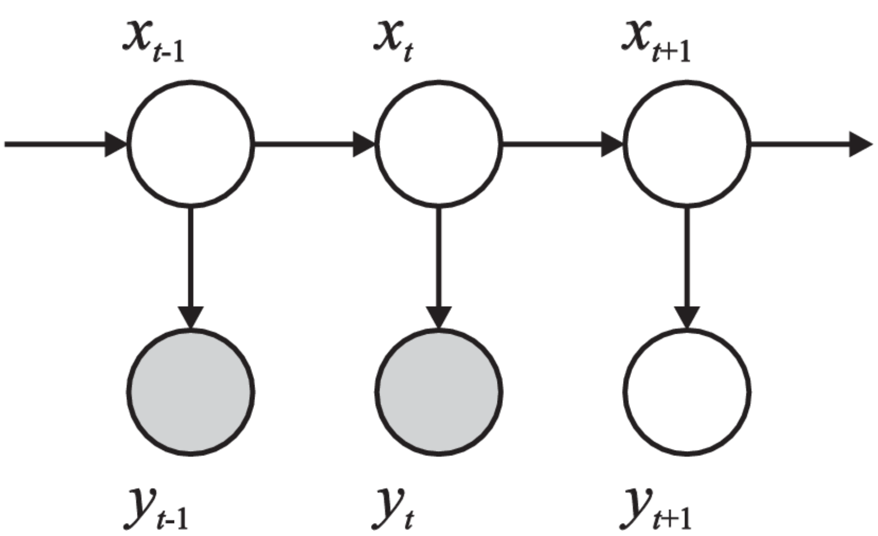
Can be done exactly for HMMs, but what if $p(y_t \mid x_t)$ complicated,

and x_t continuous or from a huge space?



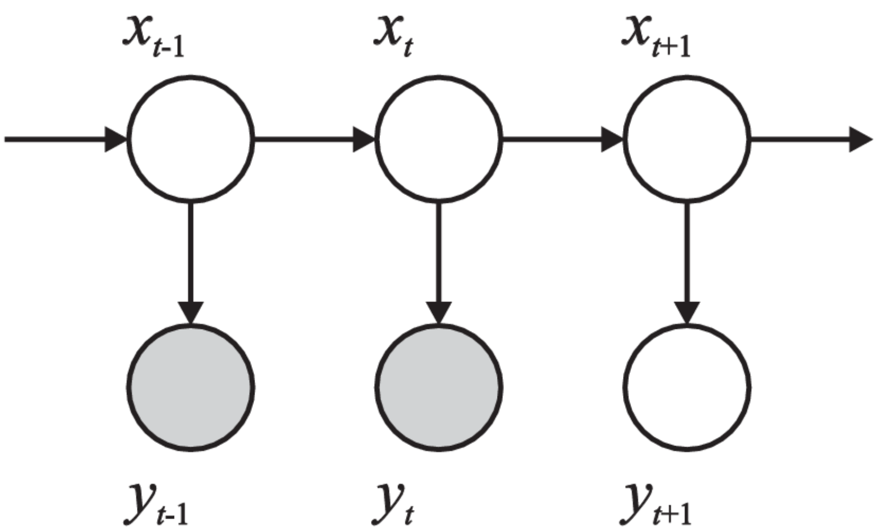
Particle filter (cont.)

- Assume we have M samples $x_t^{(m)}$ from $p(x_t \mid y_{(t-1)})$
- $w_t^{(m)} = \dfrac{p(y_t \mid x_t^{(m)})}{\sum_{m'} p(y_t \mid x_t^{(m')})}$
- $\hat{f} = \sum_m w_t^{(m)} f(x_t^{(m)})$

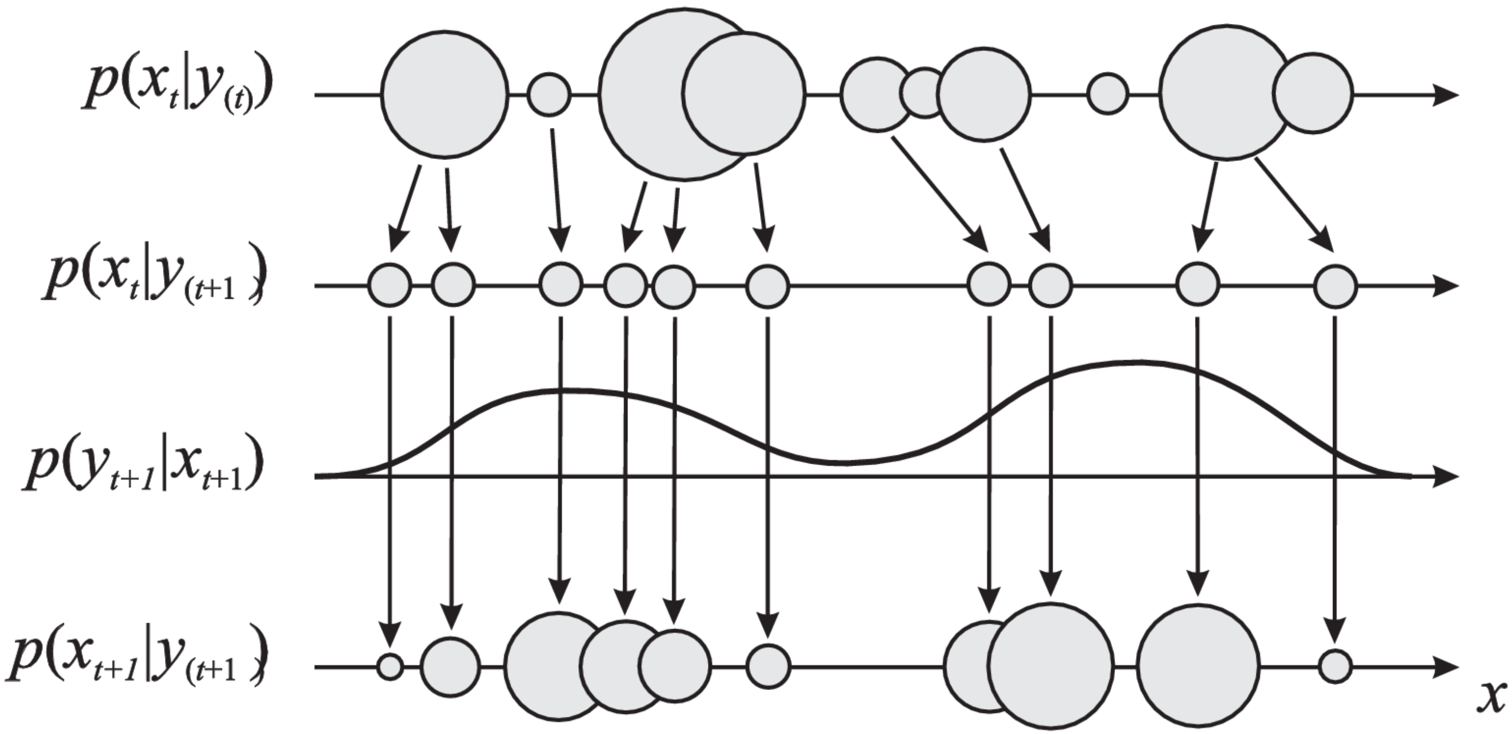


Particle filter (update)

- Assume we have M samples $x_t^{(m)}$ from $p(x_t \mid y_{(t-1)})$
- Want: samples from $p(x_{t+1} \mid y_{(t)})$
- $p(x_{t+1} \mid y_{(t)}) \approx \sum_m w_t^{(m)} p(x_{t+1} \mid x_t^{(m)})$
  mixture distribution, easy to sample from



Particle filter (update summary)



Particle filter (proofs)

- Assume we have M samples $x_t^{(m)}$ from $p(x_t \mid y_{(t-1)})$
- $w_t^{(m)} = \dfrac{p(y_t \mid x_t^{(m)})}{\sum_{m'} p(y_t \mid x_t^{(m')})}$
- $\sum_{m'} p(y_t \mid x_t^{(m')}) \approx M p(y_t \mid y_{(t-1)})$ because
  $E_{p(x_t|y_{(t-1)})}[p(y_t \mid x_t)] = \sum_{x_t} p(y_t \mid x_t) p(x_t \mid y_{(t-1)}) =$
  $\sum_{x_t} p(y_t \mid x_t, y_{(t-1)}) p(x_t \mid y_{(t-1)}) = \sum_{x_t} p(y_t, x_t \mid y_{(t-1)}) =$
  $\sum_{x_t} p(x_t \mid y_t, y_{(t-1)}) p(y_t \mid y_{(t-1)}) = p(y_t \mid y_{(t-1)})$

Particle filter (proofs cont.)

- Assume we have M samples $x_t^{(m)}$ from $p(x_t \mid y_{(t-1)})$
- $w_t^{(m)} = \dfrac{p(y_t \mid x_t^{(m)})}{\sum_{m'} p(y_t \mid x_t^{(m')})}$
- $\sum_{m'} p(y_t \mid x_t^{(m')}) \approx Mp(y_t \mid y_{(t-1)})$
- $\hat{f} = \sum_m w_t^{(m)} f(x_t^{(m)})$
- $E_{p(x_t|y_{(t-1)})}[w_t(x_t)f(x_t)] = \sum_{x_t} w_t(x_t)f(x_t)p(x_t \mid y_{(t-1)}) \approx \sum_{x_t} \frac{p(y_t|x_t)p(x_t|y_{(t-1)})}{Mp(y_t|y_{(t-1)})} f(x_t) = (1/M) \sum_{x_t} p(x_t \mid y_{(t)})f(x) = (1/M)E_{p(x_t|y_{(t)})}[f(x_t)]$

  because
  $\frac{p(y_t|x_t)p(x_t|y_{(t-1)})}{p(y_t|y_{(t-1)})} = \frac{p(y_t|x_t,y_{(t-1)})p(x_t|y_{(t-1)})}{p(y_t|y_{(t-1)})} = \frac{p(y_t,x_t|y_{(t-1)})}{p(y_t|y_{(t-1)})} = p(x_t \mid y_t, y_{(t-1)}) = p(x_t \mid y_{(t)})$

Particle filter (update proof)

- Assume we have M samples $x_t^{(m)}$ from $p(x_t \mid y_{(t-1)})$
- $w_t^{(m)} = \frac{p(y_t \mid x_t^{(m)})}{\sum_{m'} p(y_t \mid x_t^{(m')})}$
- $\sum_{m'} p(y_t \mid x_t^{(m')}) \approx M p(y_t \mid y_{(t-1)})$
- Proved $\frac{p(y_t|x_t)p(x_t|y_{(t-1)})}{p(y_t|y_{(t-1)})} = p(x_t \mid y_{(t)})$
- $p(x_{t+1} \mid y_{(t)}) = \sum_{x_t} p(x_{t+1}, x_t \mid y_{(t)})$
  $= \sum_{x_t} p(x_{t+1} \mid x_t, y_{(t)}) p(x_t \mid y_{(t)})$
  $= \sum_{x_t} p(x_{t+1} \mid x_t) \frac{p(y_t|x_t)p(x_t|y_{(t-1)})}{p(y_t|y_{(t-1)})}$
  $\approx \frac{1}{M} \sum_m p(x_{t+1} \mid x_t^{(m)}) \frac{p(y_t|x_t^{(m)})}{p(y_t|y_{(t-1)})} \approx \sum_m w_t^{(m)} p(x_{t+1} \mid x_t^{(m)})$