概率图模型理论及应用

Theory and Applications of Probabilistic Graphical Models
(Lesson 10 - sampling)

欧智坚

清华大学电子工程系

Addr: 罗姆楼 6-104

Tel: 62796193

Email: ozj@tsinghua.edu.cn

课前摘要

abs_lesson10_Sampling_章雅婷.

abs_lesson10_Sampling_贾浩歌.

abs_lesson10_sampling_王腾蛟.

课程章节

- 第一章 引言（1）
- 第二章 图模型的表示理论（2）
  - Semantics (DGM, UGM)
  - HMM, CRF
- 第三章 图模型的推理理论（6）
  - 精确推理：variable-elimination，cluster-tree，triangulate
  - 连续变量：Kalman
  - 采样近似：sampling
  - 变分近似：variational

树消除算法（Cluster-tree elimination）
理论上总是成立

- 第四章 图模型的学习理论（3）
  - 参数学习：maxlikelihoodEstimate，RFLearning，BayesEstimate
  - 结构学习：StructureLearning
- 第五章 一个综合例子（1）

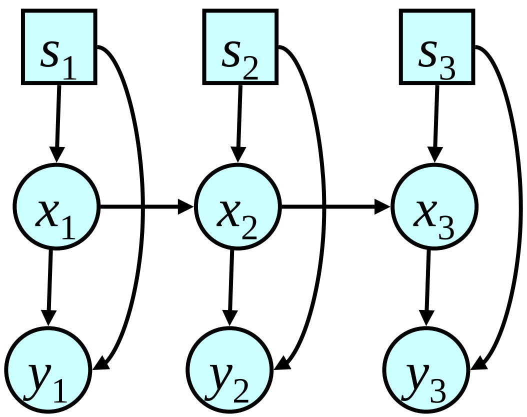
2 cases for approximate inference

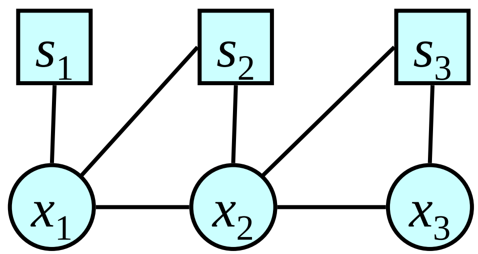
根源：消息计算实际上无法进行 $\lambda_{C_1 \to C_2} = \Downarrow_{sep(C_1,C_2)} \{\prod \cdots\}$

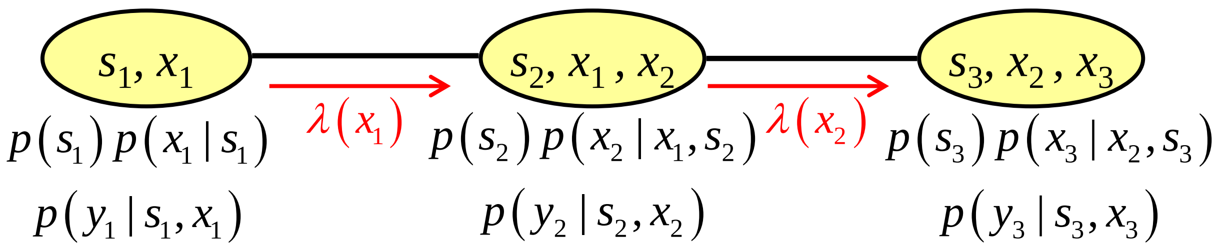
Case 1: containing continuous-var, nonlinear **though low induced-width**

Exact only when *either* LG, *or* CG with continuous-var elim first

例如：switching KF $p(x_t \mid y_{1:t}) = ?$







表示了M个高斯（假设 i 有M个可能取值）

$$\sum_i \phi(x, i \mid g_i, h_i, K_i) = \phi\left(x \mid \{g_i, h_i, K_i\}_{1 \le i \le M}\right)$$

消除变量 i ，得到M个高斯的混合

2 cases for approximate inference

根源：消息计算实际上无法进行

Case 2: High induced-width $w(G)$，G：连乘积的素图

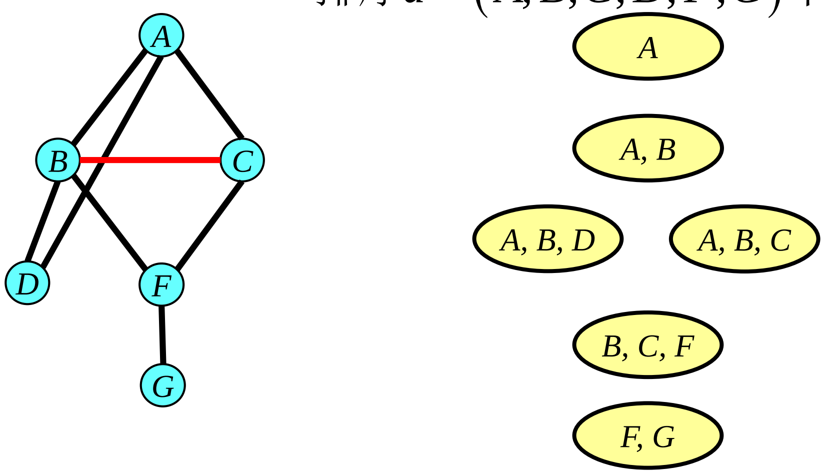
- 在**均是**离散变量情形下，推理计算的复杂度与连乘积的素图表示的诱导宽度成指数关系。

  最优树分解中最大的cluster

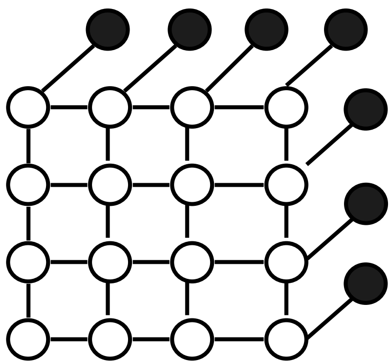
  - 图G的诱导宽度 $w(G)$ = 图G *在所有可能排序下* 的诱导宽度的最小者

    在最优排序下 的诱导宽度
  - 图G在排序d下的诱导宽度 $w(G_d)$
    = 在排序d下，消除结点过程中，结点的邻居数目的最大者
    = 在排序d下，消除式三角化过程中，结点的最大消除集的规模 - 1

排序$d = (A, B, C, D, F, G)$下消除集



Example: Image de-noising



- 带噪图像的概率模型
  - x_i：原始干净图像像素 $\in \{-1, 1\}$
  - y_i：带噪观测图像像素 $\in \{-1,1\}$
  - 相邻象素 x_i, x_j 取值的随机规律
  - y_i 与 x_i 取值的随机规律

$$p(x,y) \propto \exp\left\{\beta \sum_{i-j} x_i x_j + \gamma \sum_i x_i y_i \right\} \quad \beta > 0, \gamma > 0$$

$\phi(x_i, x_j) \overset{i-j}{=}$

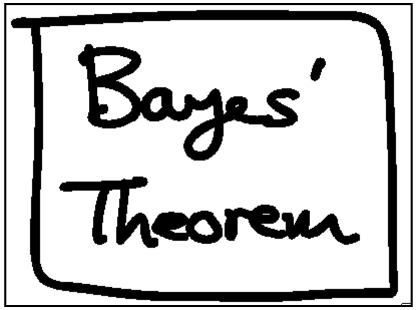
	-1	1
-1	e^{β}	$e^{-\beta}$
1	$e^{-\beta}$	e^{β}

$\phi(x_i, y_i) =$

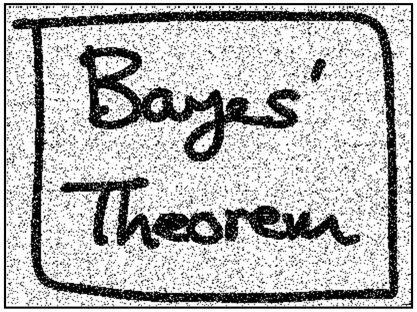
	-1	1
-1	e^{γ}	$e^{-\gamma}$
1	$e^{-\gamma}$	e^{γ}

- 给定带噪观测图像 y，恢复原始干净图像 x

$$\max_{x_i} p(x_i \mid y)$$



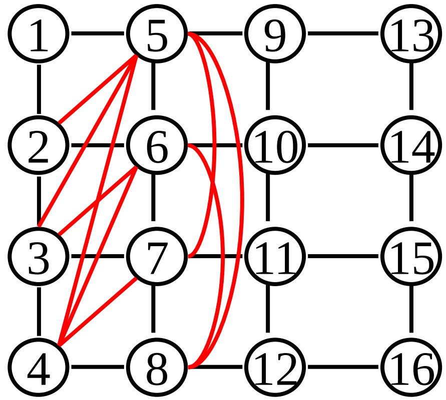
original clean



noisy observation

连乘积形式联合分布 $p(x,y) \propto \exp\left\{ \beta \sum_{i-j} x_i x_j + \gamma \sum_i x_i y_i \right\} \quad \beta > 0, \gamma > 0$

连乘积的素图表示 ***G***



induced-width $w(G)$ = ?

找出最优变量排序，保证尽可能小的消除集

1,2,5	5,6,7,8,9
2,3,5,6	6,7,8,9,10
3,4,5,6,7	7,8,9,10,11
4,5,6,7,8	8,9,10,11,12

Introduction - sampling

- 统计模拟方法—Monte Carlo 方法：非数值方法计算积分
  - 欲计算积分 $E_{p(x)}\left[\phi(x)\right]=\int \phi(x)\,p(x)dx$
  - 随机生成 R 个服从分布 $p(x)$ 的独立样本 $x^{(r)}$，对上述积分近似计算为

$$\int \phi(x)\,p\left(x \big/_{E}\right)dx \approx \frac{1}{R}\sum_{r=1}^{R}\phi\left(x^{(r)} \big/_{E}\right)$$

概率计算 可看成 是示性函数的期望

$$p\left(\overline{x} \big/_{E}\right)=\int 1\left(x=\overline{x}\right)p\left(x \big/_{E}\right)dx \approx \frac{1}{R}\sum_{r=1}^{R}1\left(x^{(r)}=\overline{x} \big/_{E}\right)$$

采样近似推理： $p(x \mid E)$无法直接计算，获取后验分布 $p(x \mid E)$ 的样本？

用概率分布 $p(x \mid E)$的样本去表示这个分布

Sampling approximation

MCMC: MH sampling, Gibbs sampling

图像去噪

Case 2: High induced-width $w(G)$

Importance sampling, Particle filtering

视频物体跟踪

Case 1: containing continuous-var, nonlinear

MCMC example: Metropolis–Hastings algorithm

We want to Draw samples from the target distribution $p(x)$?

Solution: Construct a Markov chain that has $p(x)$ as the stationary distribution.

1. Random initialize x_0
2. For $t = 1, \cdots$

   Generates x^* from proposal/transition kernel $q(x^*|x_{t-1})$,

   Accept $x_t = x^*$ with probability $min\left\{1, \frac{p(x^*)q(x_{t-1}|x^*)}{p(x_{t-1})q(x^*|x_{t-1})}\right\}$,

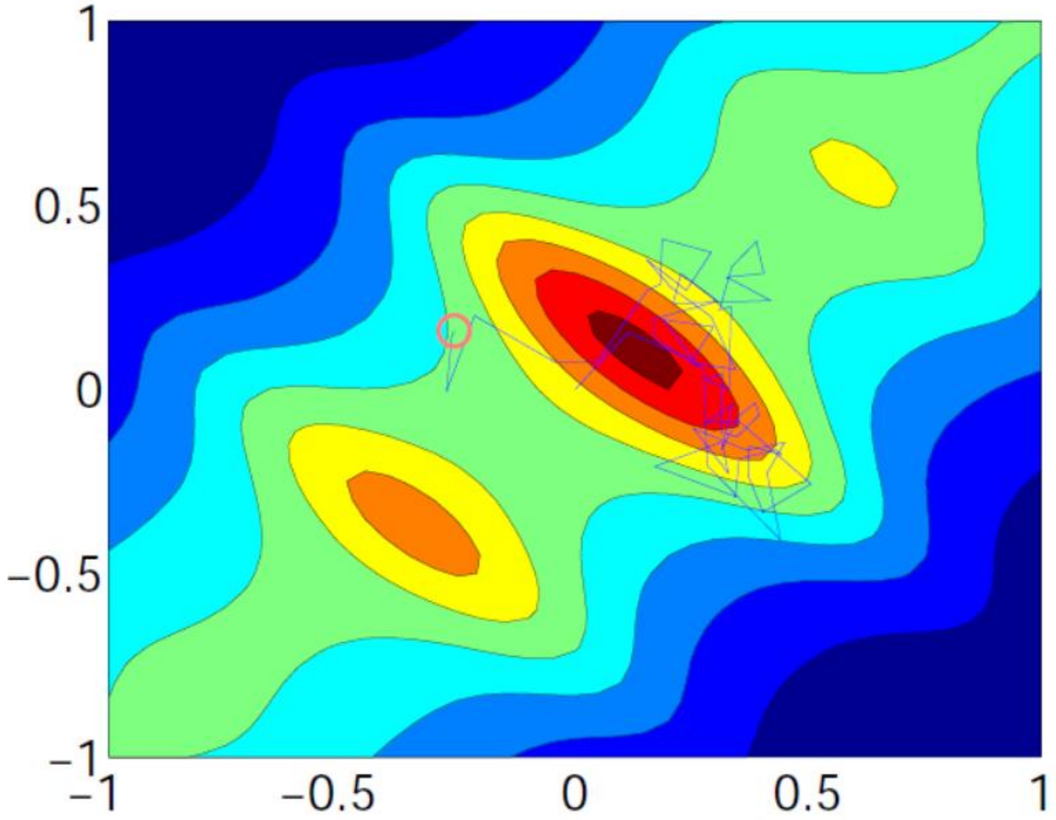
   otherwise set $x_t = x_{t-1}$

   Burn-in : first few samples are discarded.

- For an irreducible & ergodic Markov chain, there exist a stationary distribution π (i.e. satisfying equation $\pi = \pi P$).
- A sufficient condition: satisfy the detailed balance equation

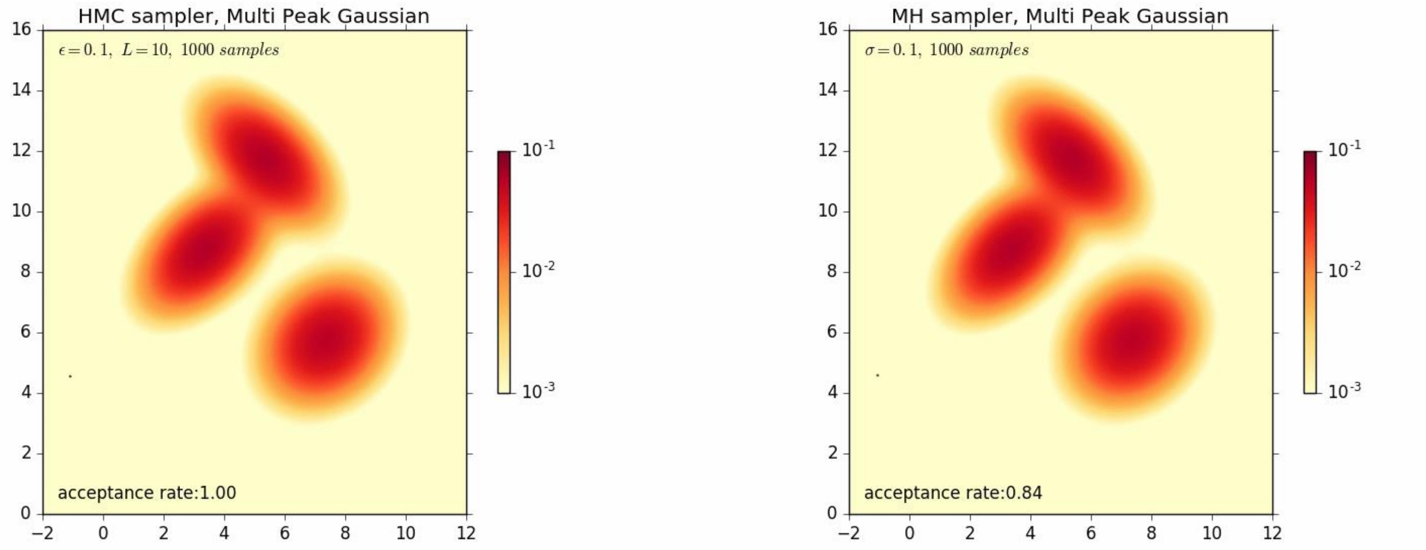
$$\pi_i P_{ij} = \pi_j P_{ji}$$

Metropolis–Hastings example



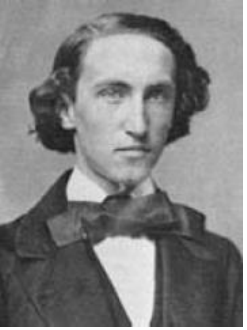
e.g. $q(x^*|x_{t-1}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x^*-x_{t-1})^2}{2\sigma^2}}$

Hamiltonian Monte Carlo (HMC) vs MH



GIF demo for sampling from a 2-dim 3-component GMM

from 张亦弛本科毕设 2017



J.W. Gibbs, 1839-1903

Gibbs sampling

- 背后原理：Markov Chain Monte Carlo (MCMC)
- 用于多维分布的采样
- 考虑对一个多维分布$p(z) = p(z_1, \cdots, z_M)$的采样：

1. $Initialize\ \left\{z_1^{(1)}, z_2^{(1)}, z_3^{(1)}, z_4^{(1)}\right\}$

2. A sweep generates a sample of $z^{(t)} = \left\{z_1^{(t)}, z_2^{(t)}, z_3^{(t)}, z_4^{(t)}\right\}$

每次固定其它分量的当前采样值，从一个分量的条件分布采样

– $Sample\ z_1^{(2)} \sim p\left(z_1 \mid z_2^{(1)}, z_3^{(1)}, z_4^{(1)}\right)$

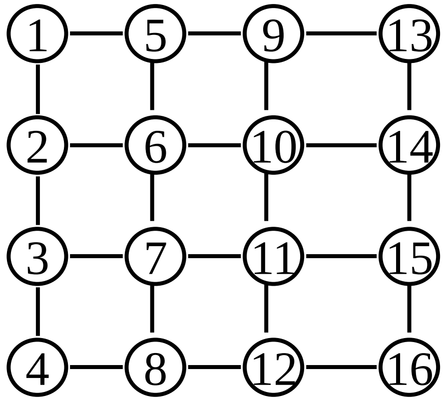
– $Sample\ z_2^{(2)} \sim p\left(z_2 \mid z_1^{(2)}, z_3^{(1)}, z_4^{(1)}\right)$

– $Sample\ z_3^{(2)} \sim p\left(z_3 \mid z_1^{(2)}, z_2^{(2)}, z_4^{(1)}\right)$

– $Sample\ z_4^{(2)} \sim p\left(z_4 \mid z_1^{(2)}, z_2^{(2)}, z_3^{(2)}\right)$

连乘积形式联合分布 $p(x,y) \propto \exp\left\{ \beta \sum_{i-j} x_i x_j + \gamma \sum_i x_i y_i \right\} \quad \beta > 0, \lambda > 0$

连乘积的素图表示



- 给定带噪观测图像 y，恢复原始干净图像 x $\quad \max_{x_i} p\left(x_i \mid y\right)$

1. $Initialize \ \left\{ x_1^{(1)}, \cdots, x_{16}^{(1)} \right\}$

sample from $p\left(x_1 \mid x_{\neg 1}, y\right)$

sample from $p\left(x_2 \mid x_{\neg 2}, y\right)$

$\vdots$

sample from $p\left(x_{16} \mid x_{\neg 16}, y\right)$

$\rightarrow \left\{ x_1^{(2)}, \cdots, x_{16}^{(2)} \right\} \rightarrow \left\{ x_1^{(3)}, \cdots, x_{16}^{(3)} \right\}$

Sampling approximation

Gibbs sampling

图像去噪

Case 2: High induced-width $w(G)$

Importance sampling, Particle filtering

视频物体跟踪

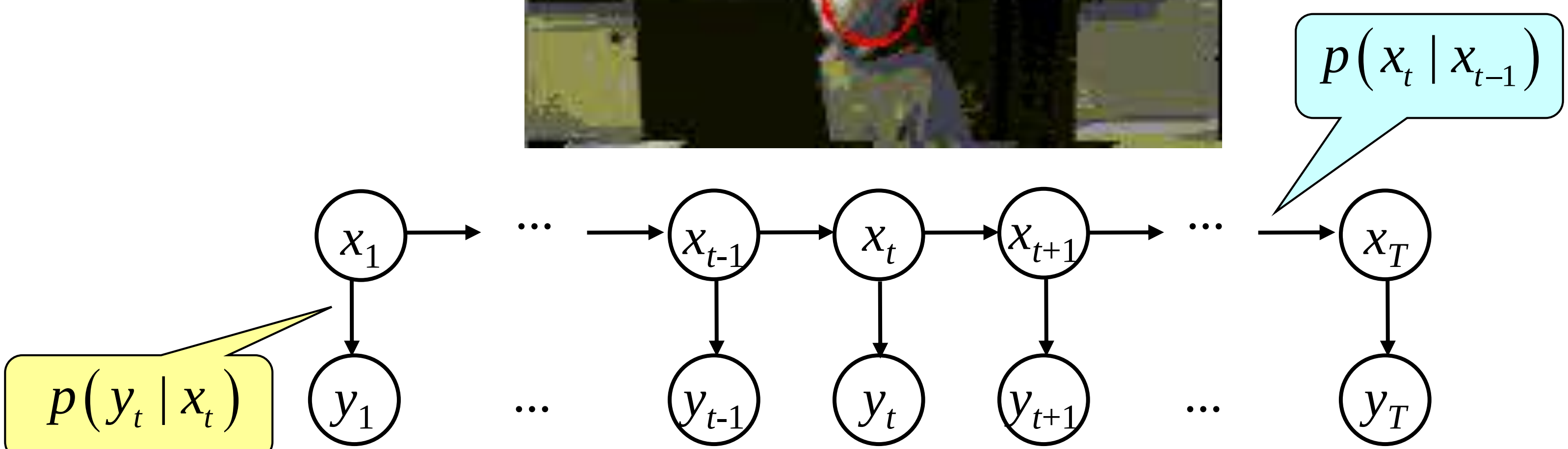
Case 1: containing continuous-var, nonlinear

视频目标跟踪

→ 轮廓跟踪　　求：$p(x_t \mid y_{1:t}) = ?$

- 观测量 $y_1, y_2, \ldots$ 为视频图像序列(帧序列)
- 状态量 $x_1, x_2, \ldots$ 为所关心物体的位置(轮廓线)

$$x_t = \begin{pmatrix} x_t^1 \\ x_t^2 \\ x_t^3 \\ \vdots \\ x_t^K \end{pmatrix}$$



轮廓跟踪—状态转移 $p(x_t \mid x_{t-1})$

- 领域随机扰动

$$x_t = x_{t-1} + \omega_{t-1} \qquad \omega_{t-1} \sim N(0, Q)$$

- 匀速运动的一阶模型

$$\begin{pmatrix} x_t \\ \dot{x}_t \end{pmatrix} = \begin{pmatrix} 1 & \Delta \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ \dot{x}_{t-1} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \omega_{t-1,1} \\ \omega_{t-1,2} \end{pmatrix}$$

轮廓跟踪—观测模型

观测模型 $p(y_t \mid x_t)$ 表示：给定图像 y_t 下，出现特定轮廓线 x_t 的似然值

沿上次跟踪结果法线的边缘特征

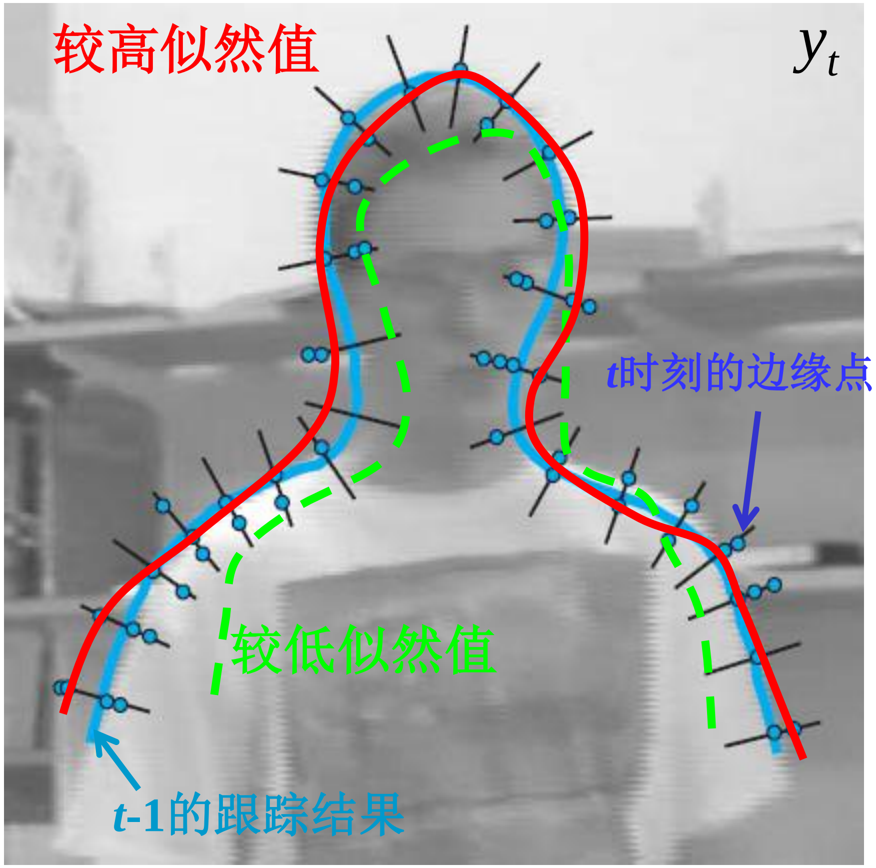
轮廓线
contour

$$\begin{pmatrix} y_t^1 \\ y_t^2 \\ y_t^3 \\ \vdots \\ y_t^K \end{pmatrix} \qquad \begin{pmatrix} x_t^1 \\ x_t^2 \\ x_t^3 \\ \vdots \\ x_t^K \end{pmatrix}$$

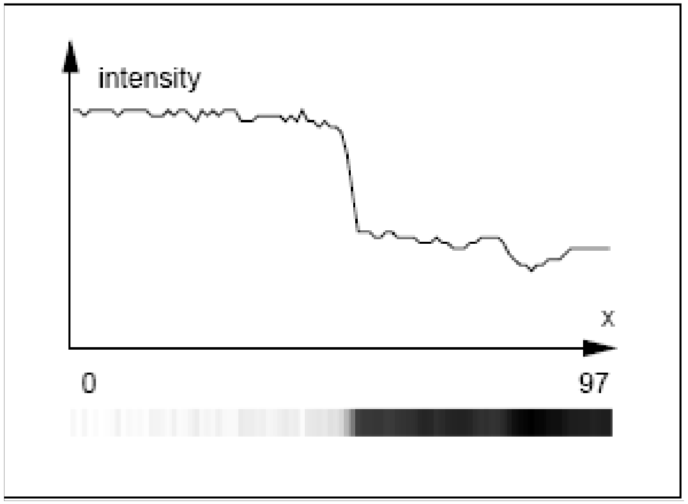
沿着上次轮廓线的第 k 条法线上的边缘点

轮廓线的K个支点—轮廓点

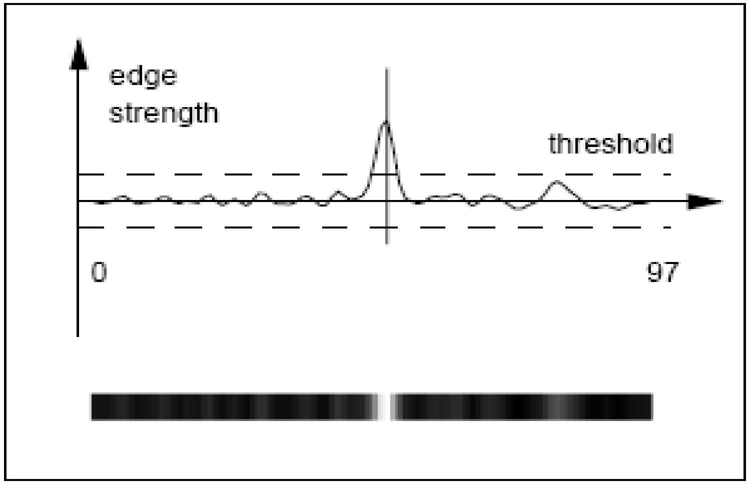
$$p\left(y_t \mid x_t\right) = \prod_k p\left(y_t^k \mid x_t^k\right)$$



轮廓跟踪—边缘特征

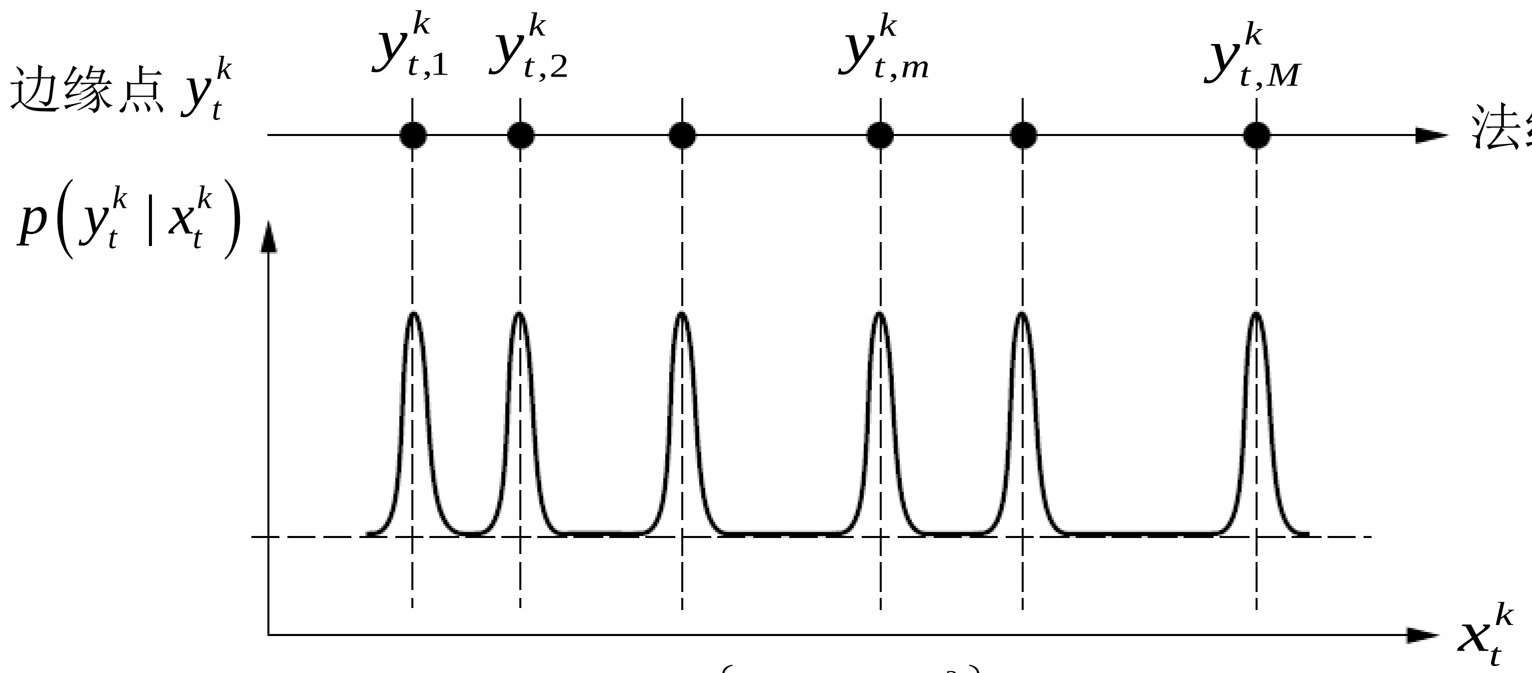


轮廓点的法线方向上灰度值



轮廓点的法线方向上灰度的梯度

轮廓跟踪—观测模型



$$p\left(y_t^k \mid x_t^k\right) = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{\left\|y_{t,m}^k - x_t^k\right\|^2}{2\sigma^2}\right\}$$

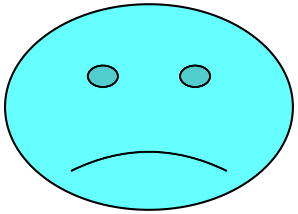
轮廓线与第k条法线的交点位置—轮廓点

$$p\left(y_t \mid x_t\right) = \prod_{k=1}^{K} p\left(y_t^k \mid x_t^k\right)$$

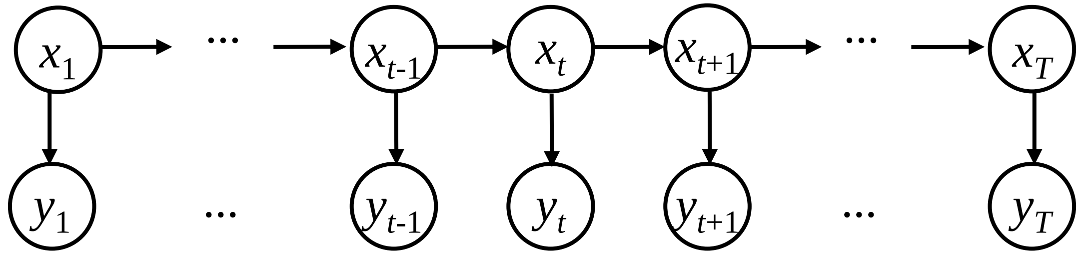
$p(y_t \mid x_t)$ 是 M^K 个典范函数的混合

多峰

非高斯



$$p(x_t \mid y_{1:t}) = ?$$



测量更新

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

时间更新

$$p(x_{t+1} \mid y_{1:t}) = \int p(x_{t+1} \mid x_t)\, p(x_t \mid y_{1:t})\, dx_t$$

特别地，Kalman Filter :

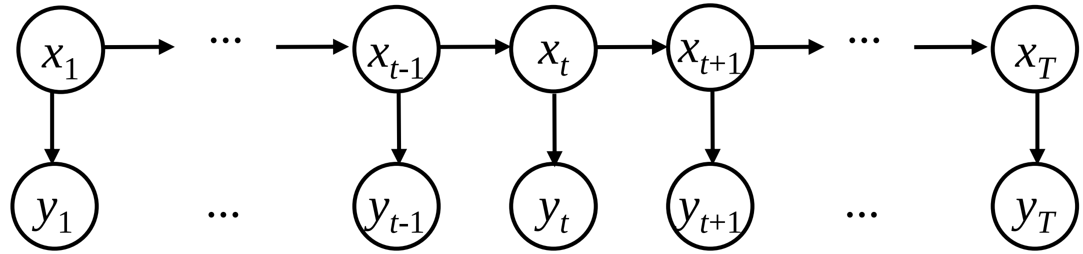
$p(x_t \mid x_{t-1})$ $p(y_t \mid x_t)$ 局部函数均是典范函数

$$x_t = Ax_{t-1} + G\omega_{t-1} \qquad \omega_t \sim N(0, Q)$$

$$y_t = Cx_t + v_t \qquad v_t \sim N(0, R)$$

函数的递归计算化归为典范参数的递归计算 ☺

$p(x_t \mid y_{1:t}) = ?$



测量更新

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

$$p(x_1 \mid y_1) = \frac{1}{p(y_1)} p(y_1 \mid x_1) p(x_1) \qquad p(x_2 \mid y_2) = \frac{p(y_{1,2})}{p(y_1)} p(y_2 \mid x_2) p(x_2 \mid y_1)$$

时间更新

$$p(x_{t+1} \mid y_{1:t}) = \int p(x_{t+1} \mid x_t) p(x_t \mid y_{1:t}) dx_t$$

$$p(x_2 \mid y_1) = \int p(x_2 \mid x_1) p(x_1 \mid y_1) dx_1$$

$$p(x_1) \longrightarrow p(x_1 \mid y_1) \longrightarrow p(x_2 \mid y_1) \longrightarrow p(x_2 \mid y_{1:2}) \dashrightarrow p(x_t \mid y_{1:t})$$

1 个 典范函数	M^K 个 典范函数 的混合	M^K 个 典范函数 的混合	$(M^K)^2$ 个 典范函数 的混合	$(M^K)^t$ 个 典范函数 的混合

Importance Sampling (unnormalized)

很难从 $p(x)$ － target density 直接采样！

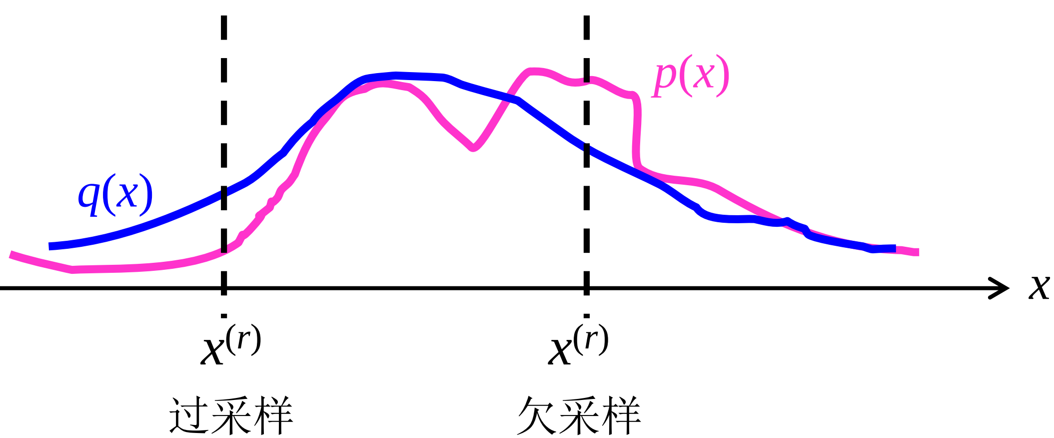
- 从另一个分布 $q(x)$ － proposal density 采样，然后加权
  - 随机生成 N 个服从分布 $q(x)$ 的独立样本 $x^{(r)}$

$$E_{p(x)}\left[\phi(x)\right] = \int \frac{p(x)}{q(x)} \phi(x) q(x) dx = E_{q(x)}\left[\frac{p(x)}{q(x)} \phi(x)\right]$$

$$\approx \frac{1}{R} \sum_{r=1}^{R} w\left(x^{(r)}\right) \phi\left(x^{(r)}\right)$$

$w(x)$

$x^{(r)}$ 连同$w(x^{(r)})$ 视为 概率分布 $p(x)$ 的粒子表示



Importance Sampling (normalized)

- $p(x)$ is only known up to a normalization constant α
  - $p(x) = p'(x) / \alpha$
  - 方便得到与其成正比的 $p'(x)$

$$E_{p(x)}\left[\phi(x)\right] = \int \frac{p(x)}{q(x)}\phi(x)q(x)dx = \frac{1}{\alpha}\int \frac{p'(x)}{q(x)}\phi(x)q(x)dx \approx \frac{1}{\alpha}\frac{1}{R}\sum_r w'\left(x^{(r)}\right)\phi\left(x^{(r)}\right)$$

$w'(x)$

$$\alpha = \int p'(x)dx = \int \frac{p'(x)}{q(x)}q(x)dx \approx \frac{1}{R}\sum_r \frac{p'\left(x^{(r)}\right)}{q\left(x^{(r)}\right)} = \frac{1}{R}\sum_r w'\left(x^{(r)}\right)$$

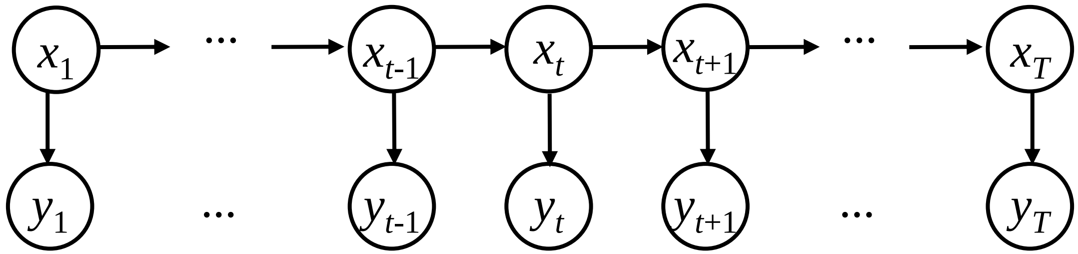
=p'\q

$$E_{p(x)}\left[\phi(x)\right] \approx \frac{\sum_r w'\left(x^{(r)}\right)\phi\left(x^{(r)}\right)}{\sum_r w'\left(x^{(r)}\right)} = \sum_r w^{(r)}\phi\left(x^{(r)}\right) \qquad w^{(r)} = w\left(x^{(r)}\right) = \frac{w'\left(x^{(r)}\right)}{\sum_r w'\left(x^{(r)}\right)}$$

$x^{(r)}$ 连同 $w^{(r)}$ 视为概率分布 $p(x)$ 的粒子表示

只要能求出 $\dfrac{\text{目标分布} p\left(x^{(r)}\right)}{\text{举荐分布} q\left(x^{(r)}\right)}$ 的相对值！

$$p(x_t \mid y_{1:t}) = ?$$



测量更新

$$p\left(x_t \mid y_{1:t}\right) = \frac{p\left(y_{1:t-1}\right)}{p\left(y_{1:t}\right)} p\left(y_t \mid x_t\right) p\left(x_t \mid y_{1:t-1}\right) \qquad \frac{p\left(x_t^{(r)} \mid y_{1:t}\right)}{p\left(x_t^{(r)} \mid y_{1:t-1}\right)} \propto p\left(y_t \mid x_t^{(r)}\right) \propto w_t^{(r)}$$

$$p\left(x_1 \mid y_1\right) = \frac{1}{p\left(y_1\right)} p\left(y_1 \mid x_1\right) p\left(x_1\right) \qquad \frac{p\left(x_1^{(r)} \mid y_1\right)}{p\left(x_1^{(r)}\right)} \propto p\left(y_1 \mid x_1^{(r)}\right) \propto w_1^{(r)}$$

时间更新

$$p\left(x_{t+1} \mid y_{1:t}\right) = \int p\left(x_{t+1} \mid x_t\right) p\left(x_t \mid y_{1:t}\right) dx_t = \sum_r w_t^{(r)} p\left(x_{t+1} \mid x_t^{(r)}\right)$$

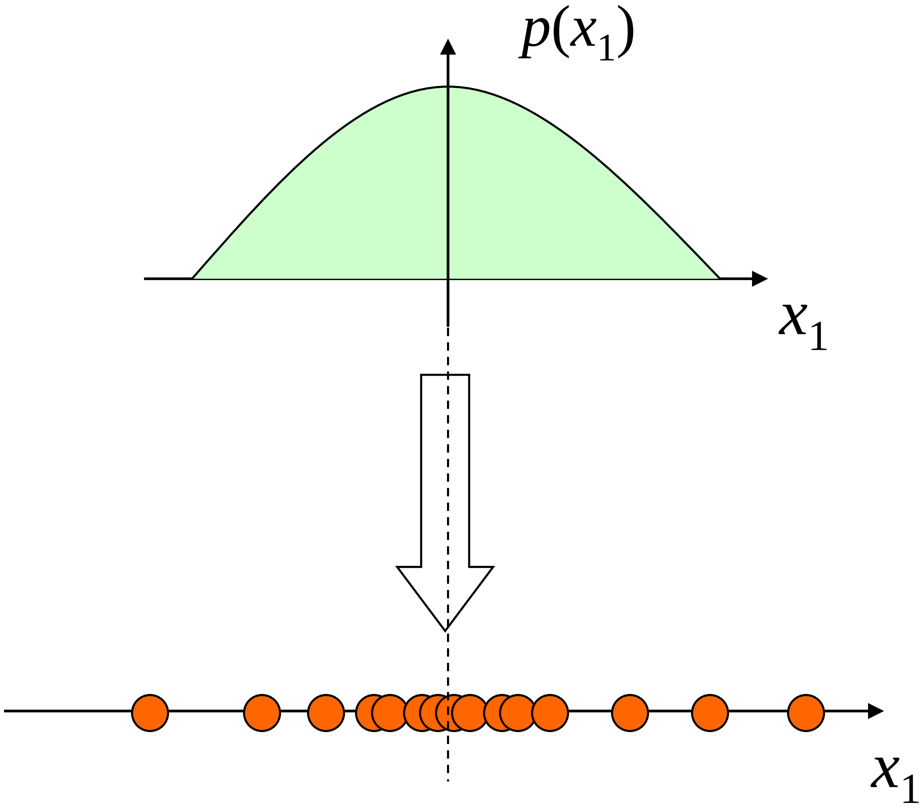
$$p\left(x_2 \mid y_1\right) = \int p\left(x_2 \mid x_1\right) p\left(x_1 \mid y_1\right) dx_1 = \sum_r w_1^{(r)} p\left(x_2 \mid x_1^{(r)}\right)$$

$$\phi(x_1) \qquad p \qquad \phi\left(x_1^{(r)}\right)$$

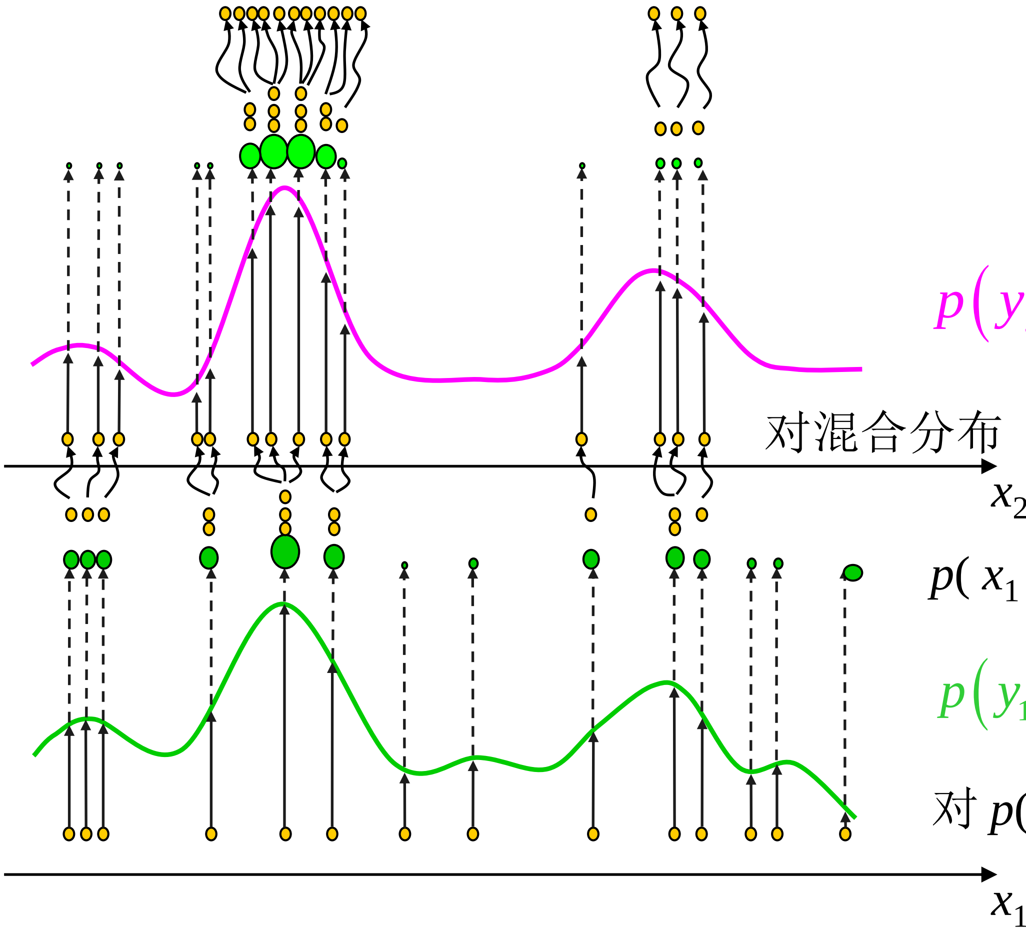
$$p\left(x_1\right) \rightarrow p\left(x_1 \mid y_1\right) \rightarrow p\left(x_2 \mid y_1\right) \rightarrow p\left(x_2 \mid y_{1:2}\right) \dashrightarrow p\left(x_t \mid y_{1:t}\right)$$

样本表示	粒子表示	样本表示	粒子表示	粒子表示
$x_1^{(r)}$	$x_1^{(r)}, w_1^{(r)}$	$x_2^{(r)}$	$x_2^{(r)}, w_2^{(r)}$	$x_t^{(r)}, w_t^{(r)}$

初始状态分布



Particle filter—直观认识



$p\left(y_2 \mid x_2^{(r)}\right) \propto w_2^{(r)}$

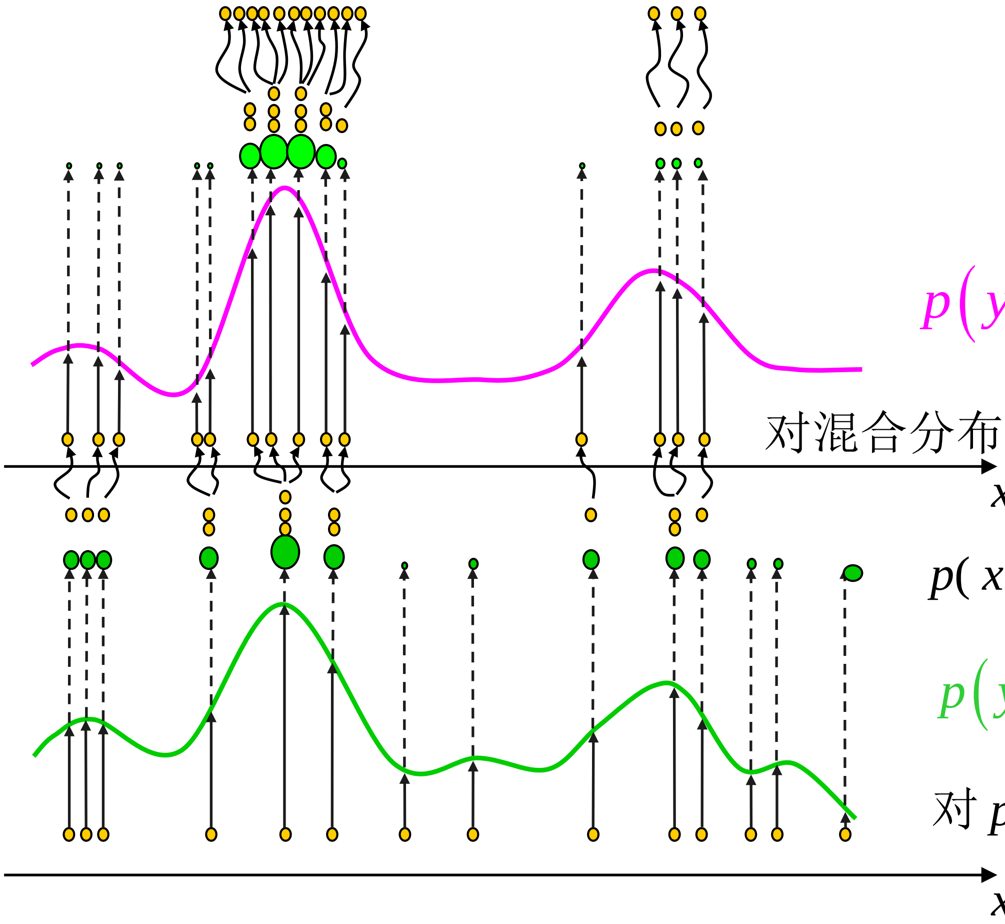
对混合分布 $p\left(x_2 \mid y_1\right) = \sum_r w_1^{(r)} p\left(x_2 \mid x_1^{(r)}\right)$ 的采样

$p(x_1 \mid y_1)$的粒子表示 $x_1^{(r)}, w_1^{(r)}$

$p\left(y_1 \mid x_1^{(r)}\right) \propto w_1^{(r)}$

对 $p(x_1)$的采样 $x_1^{(r)}$

Particle filter—直观认识



$p\left(y_{t+1} \mid x_{t+1}^{(r)}\right) \propto w_{t+1}^{(r)}$

对混合分布 $p\left(x_{t+1} \mid y_{1:t}\right) = \sum_r w_t^{(r)} p\left(x_{t+1} \mid x_t^{(r)}\right)$ 的采样

$p(x_t \mid y_{1:t})$的粒子表示 $x_t^{(r)}, w_t^{(r)}$

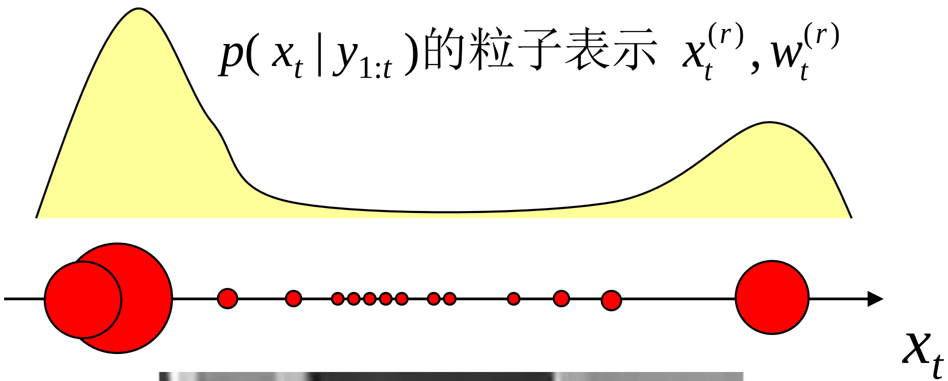
$p\left(y_t \mid x_t^{(r)}\right) \propto w_t^{(r)}$

对 $p(x_t \mid y_{1:t-1})$的采样 $x_t^{(r)}$

初始状态分布 $p(x_1)$

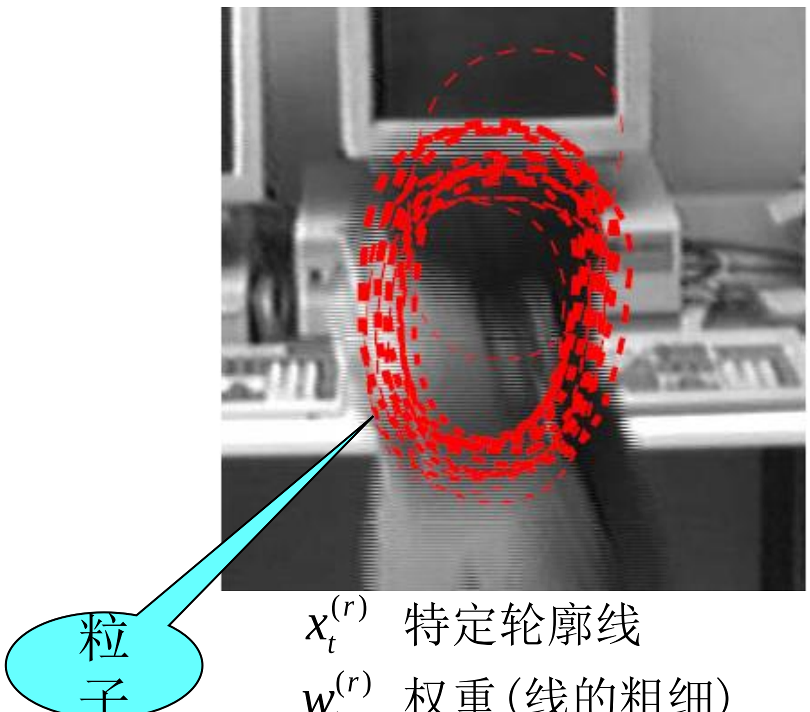
联系应用—粒子物理意义

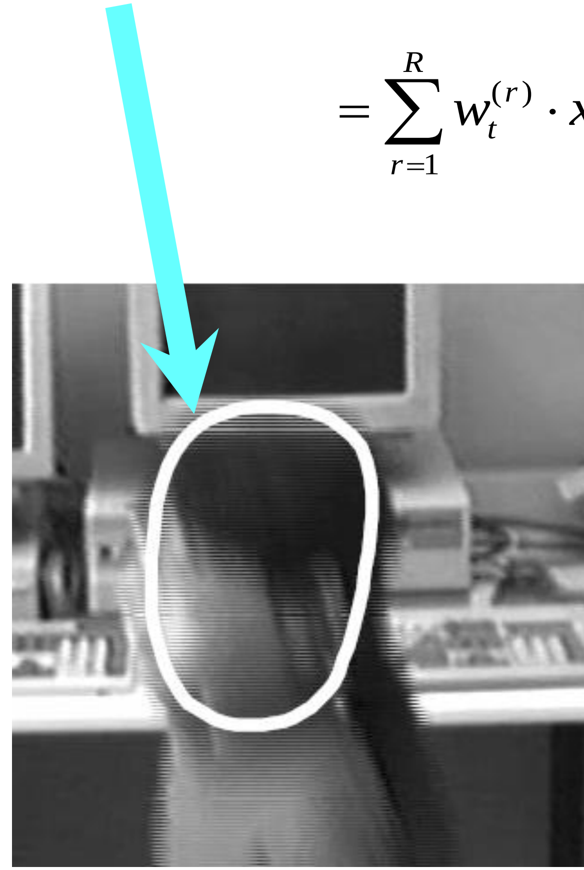
$p(x_t \mid y_{1:t})$的粒子表示 $x_t^{(r)}, w_t^{(r)}$



跟踪结果：条件均值 $E(x_t \mid y_{1:t})$

$$= \sum_{r=1}^{R} w_t^{(r)} \cdot x_t^{(r)}$$



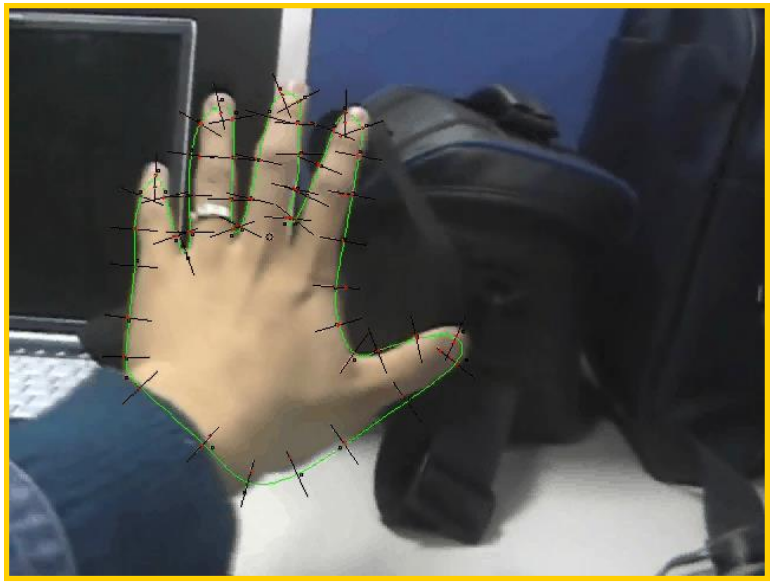


$x_t^{(r)}$ 特定轮廓线

$w_t^{(r)}$ 权重(线的粗细)

Particle filter—演示

- 基于概率图模型的运动物体跟踪系统
  - 2005年SRT项目校优秀奖
  - 无24 李亚斯等



课程章节

- 第一章 引言（1）
- 第二章 图模型的表示理论（2）
  - Semantics (DGM, UGM)
  - HMM, CRF
- 第三章 图模型的推理理论（6）
  - 精确推理：variable-elimination，cluster-tree，triangulate
  - 连续变量：Kalman
  - 采样近似：sampling
  - 变分近似：variational
- 第四章 图模型的学习理论（3）
  - 参数学习：maxlikelihoodEstimate，RFLearning，BayesEstimate
  - 结构学习：StructureLearning
- 第五章 一个综合例子（1）