概率图模型理论及应用

Theory and Applications of Probabilistic Graphical Models

(Lesson 9)

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

含有连续变量的图模型推理

树消除算法(Cluster-tree elimination)依然成立

算法本身并没有对联合分布中的变量是离散还是连续做限制

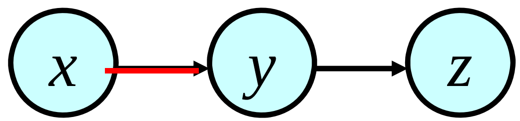
- 树消除算法的本质
  - 对连乘积求和/求积分的有效组织
  - 基础是，乘法对加法/积分的分配律 $\int f(x)g(y)dy = f(x)\int g(y)dy$
- 推理问题 $p(x_Q \mid X_E = x_E)$：只需关心隐变量 x_Q 中含有连续变量

Example

$$p(x) = N(x|0,1) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}$$

$$p(y|x) = N(y|x,1) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(y-x)^2}{2}\right\}$$

$$p(z|y) = N(z|y,1) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(z-y)^2}{2}\right\}$$



$$p(x,y,z) = p(x)\,p(y|x)\,p(z|y)$$

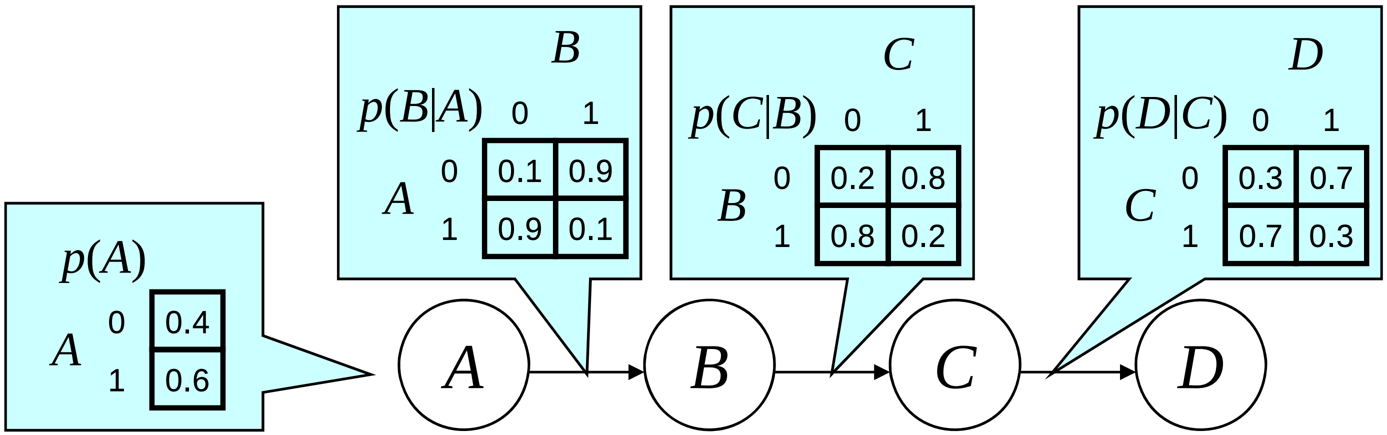
试求：$p(x \mid z = 1.5)$ $\propto p(x, z=1.5) = \int p(x,y,z=1.5)\,dy$

$$p(x,y,z=1.5) = p(x)\,p(y|x)\,p(z=1.5|y)$$

$$p(x|z=1.5) \propto \Downarrow_x \left[\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} \cdot \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(y-x)^2}{2}\right\} \cdot \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(1.5-y)^2}{2}\right\}\right]$$

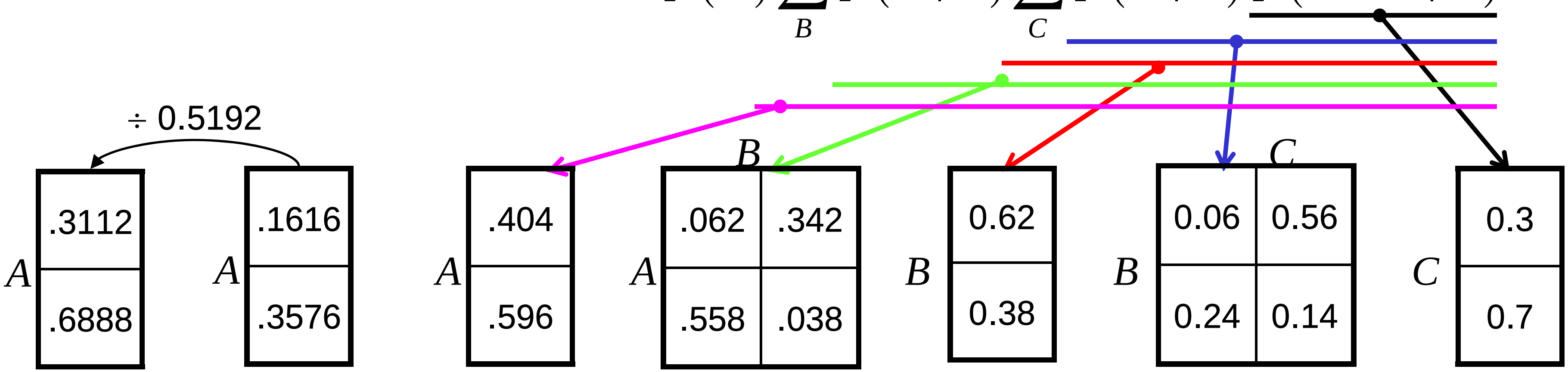
Operations on factors: product, marginalize

Inference example



$$p(A \mid D=0) \propto p(A, D=0) = \sum_B \sum_C p(A)\, p(B \mid A)\, p(C \mid B)\, p(D=0 \mid C)$$

$$= p(A) \sum_B p(B \mid A) \sum_C p(C \mid B)\, p(D=0 \mid C)$$



A	
	.3112
	.6888

A	
	.1616
	.3576

A	
	.404
	.596

A \ B		
	.062	.342
	.558	.038

B	
	0.62
	0.38

B \ C		
	0.06	0.56
	0.24	0.14

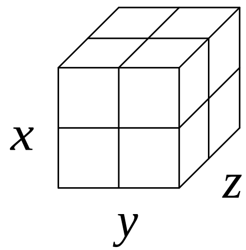
C	
	0.3
	0.7

Two basic operations: product, marginalization

Factor/Potential representation

- Pure discrete: tables (multi-dimensional arrays)

$$\phi(x, y, z)$$



- Multivariate Gaussian can be represented in two forms

Moment form
矩表示

$$\phi(x \mid \mu, \Sigma) = \frac{1}{(2\pi)^{N/2} |\Sigma|^{1/2}} \exp\left\{ -\frac{1}{2} (x-\mu)^T \Sigma^{-1} (x-\mu) \right\}$$

$$g = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \log|\Sigma| - \frac{1}{2} \mu^T \Sigma^{-1} \mu$$

$$K = \Sigma^{-1}$$

$$h = \Sigma^{-1} \mu$$

$$\mu = K^{-1} h$$

$$\Sigma = K^{-1}$$

Canonical form
典范表示

$$\phi(x \mid g, h, K) = \exp\left\{ g + x^T h - \frac{1}{2} x^T K x \right\}$$

典范函数（canonical potential）：
指数肩膀上 x 的零次型、一次型、二次型的线性组合，表征参数为(g, h, K)

① Operations on canonical potentials: easier

- Converting a linear-Gaussian CPD to a canonical potential

$$p(x \mid z) = N(x \mid Bz + \mu, \Sigma)$$

$$= \frac{1}{(2\pi)^{N/2} |\Sigma|^{1/2}} \exp\left\{ -\frac{1}{2} (x - Bz - \mu)^T \Sigma^{-1} (x - Bz - \mu) \right\}$$

↓ 重新整理，利用 $x - Bz = (I, -B)\begin{pmatrix} x \\ z \end{pmatrix}$

$$f(x \mid z) = \phi(x, z \mid g, h, K)$$

g: $-\frac{N}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}\mu^T \Sigma^{-1} \mu$

h: $\begin{pmatrix} I \\ -B^T \end{pmatrix} \Sigma^{-1} \mu$

K: $\begin{pmatrix} I \\ -B^T \end{pmatrix} \Sigma^{-1} (I, -B) = \begin{pmatrix} \Sigma^{-1} & -\Sigma^{-1}B \\ -B^T\Sigma^{-1} & B^T\Sigma^{-1}B \end{pmatrix}$

② Operations on canonical potentials: easier

- Entering evidence

如果变量 y 有观测值 $\bar{y}$，则将 $\bar{y}$ 代入到包含 y 的势函数，得到新的势函数。

$$\phi(y,z) = \exp\left\{ g + \left(y^T, z^T\right)\begin{pmatrix} h_y \\ h_z \end{pmatrix} - \frac{1}{2}\left(y^T, z^T\right)\begin{pmatrix} K_{yy} & K_{zy} \\ K_{yz} & K_{zz} \end{pmatrix}\begin{pmatrix} y \\ z \end{pmatrix} \right\}$$

↓ 代入观测值 $\bar{y}$

$$\phi(\bar{y},z) = \exp\left\{ \left(g + h_y^T \bar{y} - \frac{1}{2}\bar{y}^T K_{yy}\bar{y} \right) + z^T\left(h_z - K_{zy}\bar{y}\right) - \frac{1}{2} z^T K_{zz} z \right\}$$

③ Operations on canonical potentials: easier

- Product

$$\phi(x \mid g_1, h_1, K_1) \cdot \phi(x \mid g_2, h_2, K_2) = \phi(x \mid g_1 + g_2, h_1 + h_2, K_1 + K_2)$$

$$\phi(x, y \mid g_1, h_1, K_1) \cdot \phi(y, z \mid g_2, h_2, K_2) = ?$$

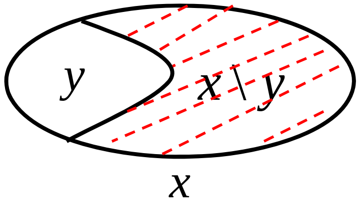
$$\phi\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix} \middle| g_1, \begin{pmatrix} h_{1,x} \\ h_{1,y} \\ 0 \end{pmatrix}, \begin{pmatrix} K_{1,xx} & K_{1,xy} & 0 \\ K_{1,yx} & K_{1,yy} & 0 \\ 0 & 0 & 0 \end{pmatrix}\right) \times \phi\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix} \middle| g_2, \begin{pmatrix} 0 \\ h_{2,y} \\ h_{2,z} \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & K_{2,yy} & K_{2,yz} \\ 0 & K_{2,zy} & K_{2,zz} \end{pmatrix}\right) = ?$$

Extension: 将定义在 (x, y) 上的典范函数视为 (x, y, z) 上的典范函数

④ Operations on canonical potentials: easier

- 假设 x 和 y 为变量集，且 $y \subseteq x$。$\phi(x)$为定义在x上的函数，则 $\phi(x)$ 在 y 上的边缘化为

$$\Downarrow_y \phi(x) = \sum_{x \setminus y} \phi(x)$$



- Marginalization of a canonical potential

$$\phi(y,z \mid g, h, K) \sim g, \begin{pmatrix} h_y \\ h_z \end{pmatrix}, \begin{pmatrix} K_{yy} & K_{zy} \\ K_{yz} & K_{zz} \end{pmatrix}$$

$$\Downarrow_y \phi(y,z \mid g, h, K) = \phi\left(y \mid \hat{g}, \hat{h}, \hat{K}\right)$$

$$g + \left\{ \frac{|y|}{2}\log(2\pi) - \frac{1}{2}\log\left|K_{zz}\right| + \frac{1}{2} h_z^T K_{zz}^{-1} h_z \right\} \quad \left(h_y - K_{yz} K_{zz}^{-1} h_z\right) \quad \left(K_{yy} - K_{yz} K_{zz}^{-1} K_{zy}\right)$$

典范函数—总结

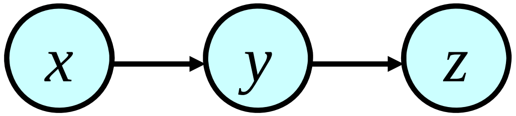
- 运算法则
  - 如果 x 的密度函数 $p(x)$ 正比于一个典范函数（表征参数为g, h, K），那么 $p(x)$ 就是一个高斯密度函数，并且 h, K 就是这个高斯密度函数的典范表示中的h, K。
  - Converting a linear-Gaussian CPD to a canonical potential
  - Entering evidence
  - Product
  - Marginalization
- 优点
  - 典范函数对于引入证据、相乘、边缘化操作是封闭。
  - 推理计算中，函数的操作 简化为 参数的操作。

Example - revisited

$$p(x) = N(x \mid 0,1) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}$$

$$p(y \mid x) = N(y \mid x,1) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(y-x)^2}{2}\right\}$$

$$p(z \mid y) = N(z \mid y,1) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(z-y)^2}{2}\right\}$$



试求：$p(x \mid z = 1.5) = ?$

$$p(x,y,z) = p(x)\,p(y \mid x)\,p(z \mid y)$$

$$p(x \mid z = 1.5) \propto p(x, z = 1.5)$$

$$\propto \Downarrow_x \; p(x)\,p(y \mid x)\,p(z = 1.5 \mid y)$$

$$\propto \Downarrow_x \; \exp\left\{-\frac{x^2}{2} - \frac{(y-x)^2}{2} - \frac{(1.5-y)^2}{2}\right\}$$

$$\propto \Downarrow_x \; \exp\left\{(x,y)\begin{pmatrix}0\\1.5\end{pmatrix} - \frac{1}{2}(x,y)\begin{pmatrix}2 & -1\\-1 & 2\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}\right\}$$

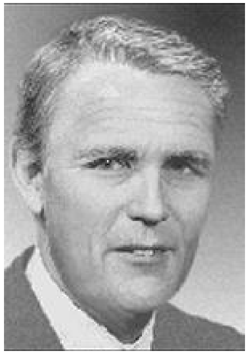
$$= N\left(x \mid \frac{1}{2}, \frac{2}{3}\right)$$

含有连续变量的图模型推理

Cluster-tree elimination算法依然成立

算法本身并没有对联合分布中的变量是离散还是连续做限制

Linear-Gaussian KalmanFilter	Conditional Gaussian (CG) Mixed discrete-Gaussian	Non-linear non-Gaussian



Kalman filter - introduction

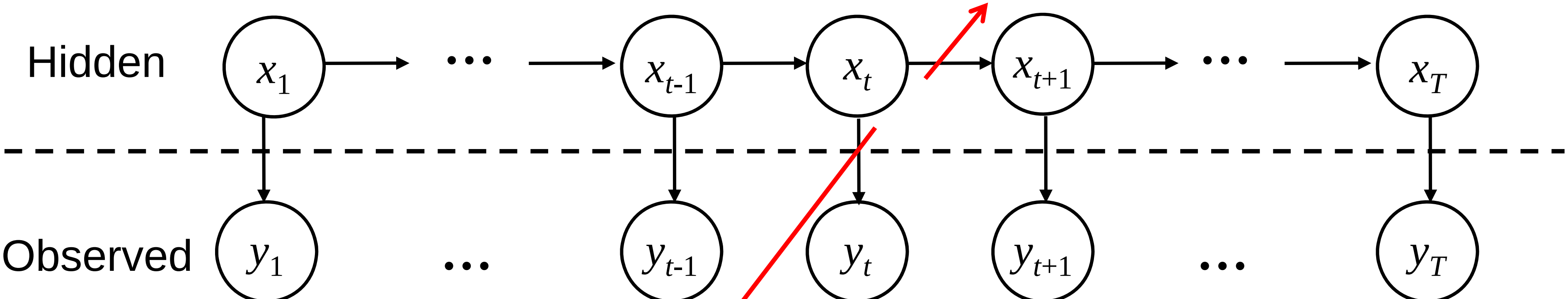
KalmanFilter is structurally identical to HMM

	HMM	Kalman filter
Different implementation	Discrete hidden variable	Continuous hidden variable
	transition matrix	Linear-Gaussian

状态方程: $x_{t+1} = Ax_t + G\omega_t \quad \omega_t \sim N(0, Q)$, and $\omega_{t_1} \perp \omega_{t_2}$ for $t_1 \neq t_2$

$$p(x_{t+1} \mid x_t) = N\left(Ax_t, GQG^T\right)$$

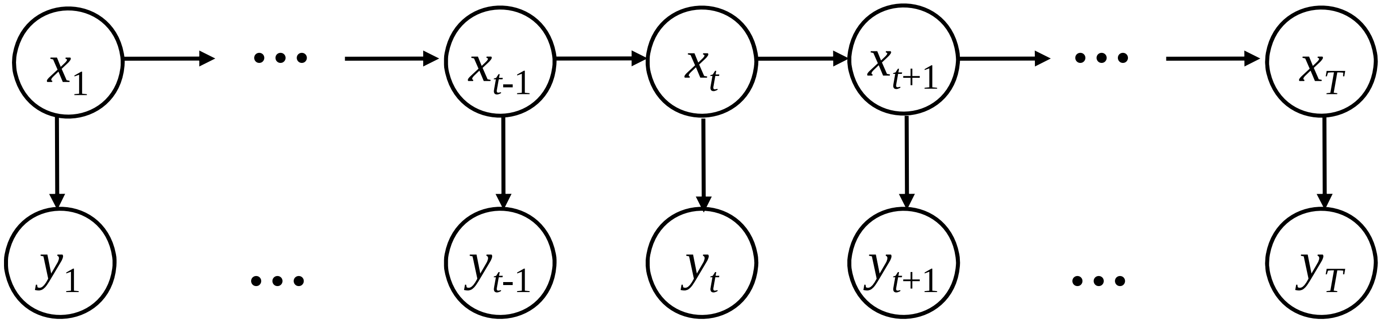
Initialized as $x_1 \sim N(0, \Sigma_1)$



$$p(y_t \mid x_t) = N(Cx_t, R)$$

观测方程: $y_t = Cx_t + v_t \quad v_t \sim N(0, R)$, and $v_{t_1} \perp v_{t_2}$ for $t_1 \neq t_2$

Example: constant-velocity model



state equation: $x_{t+1} = Ax_t + G\omega_t \quad \omega_t \sim N(0, Q)$, and $\omega_{t_1} \perp \omega_{t_2}$ for $t_1 \neq t_2$

$$\begin{pmatrix} \rho_{t+1} \\ \dot{\rho}_{t+1} \\ \theta_{t+1} \\ \dot{\theta}_{t+1} \end{pmatrix} = \begin{pmatrix} 1 & \Delta & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \rho_t \\ \dot{\rho}_t \\ \theta_t \\ \dot{\theta}_t \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \omega_{t,1} \\ \omega_{t,2} \end{pmatrix}$$

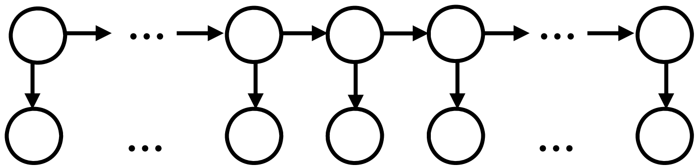
策动噪声

observation equation: $y_t = Cx_t + v_t \quad v_t \sim N(0, R)$, and $v_{t_1} \perp v_{t_2}$ for $t_1 \neq t_2$

$$\begin{pmatrix} \rho_t^o \\ \theta_t^o \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \rho_t \\ \dot{\rho}_t \\ \theta_t \\ \dot{\theta}_t \end{pmatrix} + \begin{pmatrix} v_{t,1} \\ v_{t,2} \end{pmatrix}$$

观测误差

Inference problem for KalmanFilter



- Filtering: $p(x_t \mid y_{1:t}) \propto p(x_t, y_{1:t}) \quad 1 \le t \le T$
  - Similar to calculate α variables in the HMM, $\alpha(q_t) = p(q_t, y_{1:t})$
  - HMM (L.E. Baum, *et al*, 1966), KalmanFilter (R.E. Kalman, 1960) developed separately

- 递归滤波的思想（Kalman, 1960）
  - 随着观测量的进入，不停更新对状态量的估计
  - 在时刻 t，对状态量 x_t 的估计：计算 $p(x_t \mid y_{1:t})$
  - 观测到新的数据 y_{t+1}，如何计算 $p(x_{t+1} \mid y_{1:t+1})$?

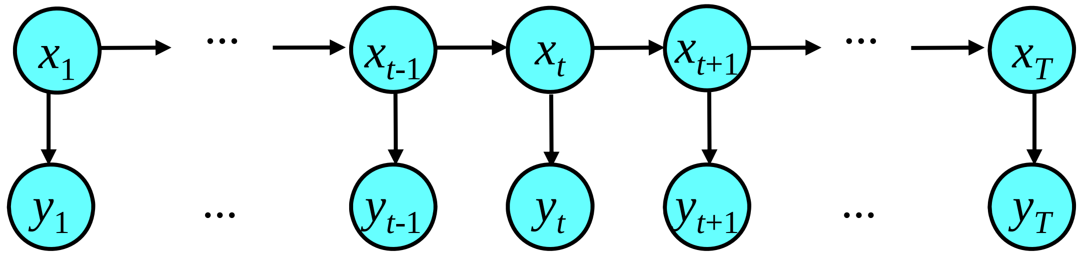
  - 时间更新: $p(x_t \mid y_{1:t}) \rightarrow p(x_{t+1} \mid y_{1:t})$

    t 时刻的估计　　t 时刻的预测

  - 测量更新: $p(x_{t+1} \mid y_{1:t}) \rightarrow p(x_{t+1} \mid y_{1:t+1})$

    t 时刻的预测　　t+1 时刻的估计

$$p(x_T \mid y_{1:T}) = ?$$

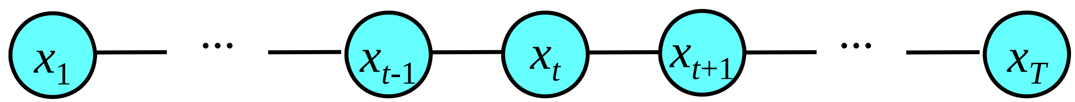


连乘积形式联合分布

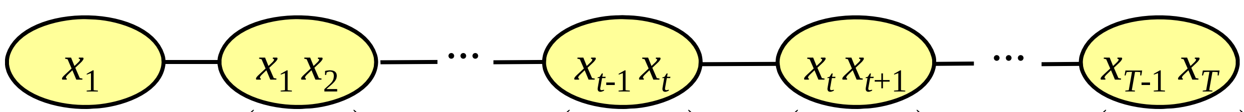
$\Downarrow x_T$

$$p(x_{1:T}, y_{1:T}) = \begin{matrix} p(x_1) \\ \times p(y_1 \mid x_1) \end{matrix} \quad \begin{matrix} p(x_2 \mid x_1) \\ \times p(y_2 \mid x_2) \end{matrix} \quad \cdots \quad \begin{matrix} p(x_t \mid x_{t-1}) \\ \times p(y_t \mid x_t) \end{matrix} \quad \cdots \quad \begin{matrix} p(y_T \mid x_{T-1}) \\ \times p(y_T \mid x_T) \end{matrix}$$

连乘积的素图表示

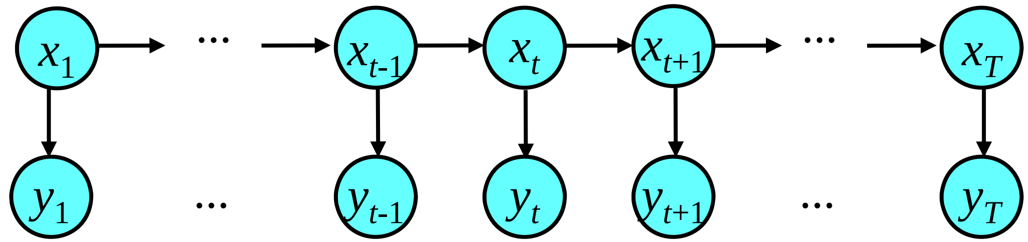


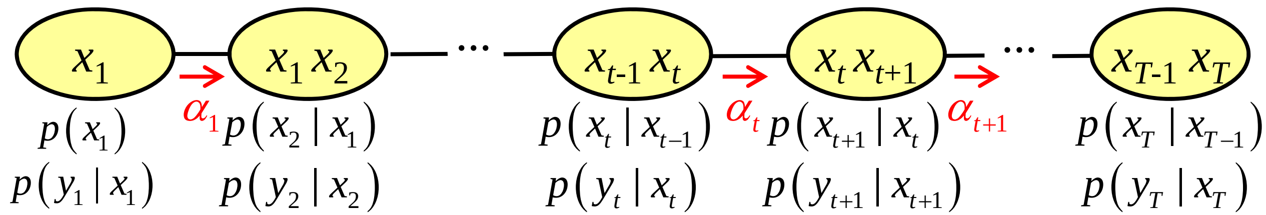
编译得到连接树：以素图的最大簇作为cluster，建一棵连接树

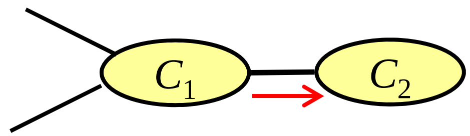


$$\begin{matrix} p(x_1) \\ p(y_1 \mid x_1) \end{matrix} \quad \begin{matrix} p(x_2 \mid x_1) \\ p(y_2 \mid x_2) \end{matrix} \quad \begin{matrix} p(x_t \mid x_{t-1}) \\ p(y_t \mid x_t) \end{matrix} \quad \begin{matrix} p(x_{t+1} \mid x_t) \\ p(y_{t+1} \mid x_{t+1}) \end{matrix} \quad \begin{matrix} p(x_T \mid x_{T-1}) \\ p(y_T \mid x_T) \end{matrix}$$

Message passing for $p(x_T \mid y_{1:T})$







$$\lambda_{C_1 \to C_2} = \Downarrow_{\text{隔离子}} \{\text{位于发送端的所有cluster的函数集中函数连乘积}\}$$

$$\alpha_t(x_t) = \Downarrow_{x_t} p(x_{1:t}, y_{1:t}) = p(x_t, y_{1:t})$$

$$\alpha_{t+1}(x_{t+1}) = \Downarrow_{x_{t+1}} \alpha_t(x_t)\, p(x_{t+1} \mid x_t)\, p(y_{t+1} \mid x_{t+1})$$ HMM's α recursion

时间更新

$$p(x_t \mid y_{1:t}) \to p(x_{t+1} \mid y_{1:t})$$

测量更新

$$p(x_{t+1} \mid y_{1:t}) \to p(x_{t+1} \mid y_{1:t+1})$$

时间更新： $\Downarrow_{x_{t+1}} \quad p(x_t \mid y_{1:t})\, p(x_{t+1} \mid x_t) = p(x_{t+1} \mid y_{1:t})$

$$p(x_t \mid y_{1:t}) \to p(x_{t+1} \mid y_{1:t})$$

$$\phi(x_t \mid h_{t|t}, K_{t|t}) \to \phi(x_{t+1} \mid h_{t+1|t}, K_{t+1|t})$$

$$p(x_{t+1} \mid x_t) = N(Ax_t, GQG^T)$$

$$\begin{pmatrix} x_{t+1} \\ x_t \end{pmatrix}: \quad \begin{pmatrix} 0 \\ h_{t|t} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ h_{t|t} \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0 \\ 0 & K_{t|t} \end{pmatrix} + \begin{pmatrix} H^{-1} & -H^{-1}A \\ -A^T H^{-1} & A^T H^{-1} A \end{pmatrix} = \begin{pmatrix} H^{-1} & -H^{-1}A \\ -A^T H^{-1} & K_{t|t} + A^T H^{-1} A \end{pmatrix}$$

$$H \triangleq GQG^T \qquad h_{t+1|t} = H^{-1}A\left(K_{t|t} + A^T H^{-1} A\right)^{-1} h_{t|t}$$

$$K_{t+1|t} = H^{-1} - H^{-1}A\left(K_{t|t} + A^T H^{-1} A\right)^{-1} A^T H^{-1}$$

$$\text{测量更新：} p(x_{t+1} \mid y_{1:t})\, p(y_{t+1} \mid x_{t+1}) \frac{p(y_{1:t})}{p(y_{1:t+1})} = p(x_{t+1} \mid y_{1:t+1})$$

$$p(x_{t+1} \mid y_{1:t}) \to p(x_{t+1} \mid y_{1:t+1})$$

$$\phi(x_{t+1} \mid h_{t+1|t}, K_{t+1|t}) \to \phi(x_{t+1} \mid h_{t+1|t+1}, K_{t+1|t+1})$$

$p(y_{t+1} \mid x_{t+1}) = N(Cx_{t+1}, R)$ 视为 $\begin{pmatrix} y_{t+1} \\ x_{t+1} \end{pmatrix}$ 的典范函数： $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ $\begin{pmatrix} R^{-1} & -R^{-1}C \\ -C^T R^{-1} & C^T R^{-1} C \end{pmatrix}$

引入证据 y_{t+1} 后，视为 x_{t+1} 的典范函数： $C^T R^{-1} y_{t+1}$ $C^T R^{-1} C$

$$h_{t+1|t} + C^T R^{-1} y_{t+1} = h_{t+1|t+1}$$

$$K_{t+1|t} + C^T R^{-1} C = K_{t+1|t+1}$$

Kalman filter: Linear-Gaussian

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

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

Assume that at time t, we have available $p(x_t \mid y_{1:t}) \sim N\left(\mu_{t|t}, \Sigma_{t|t}\right)$

Calculate $p(x_{t+1} \mid y_{1:t+1}) \sim N\left(\mu_{t+1|t+1}, \Sigma_{t+1|t+1}\right)$ recursively :

$$\mu_{t+1|t} = A\mu_{t|t} \qquad \mu_{t+1|t+1} = \mu_{t+1|t} + \Sigma_{t+1|t}C^T\left(C\Sigma_{t+1|t}C^T + R\right)^{-1}\left(y_{t+1} - C\mu_{t+1|t}\right)$$

$$\Sigma_{t+1|t} = A\Sigma_{t|t}A^T + GQG^T \qquad \Sigma_{t+1|t+1} = \Sigma_{t+1|t} - \Sigma_{t+1|t}C^T\left(C\Sigma_{t+1|t}C^T + R\right)^{-1}C\Sigma_{t+1|t}$$

Time update

Measurement update

含有连续变量的图模型推理

Cluster-tree elimination算法依然成立

算法本身并没有对联合分布中的变量是离散还是连续做限制

Linear-Gaussian KalmanFilter	Conditional Gaussian (CG) Mixed discrete-Gaussian	Non-linear non-Gaussian

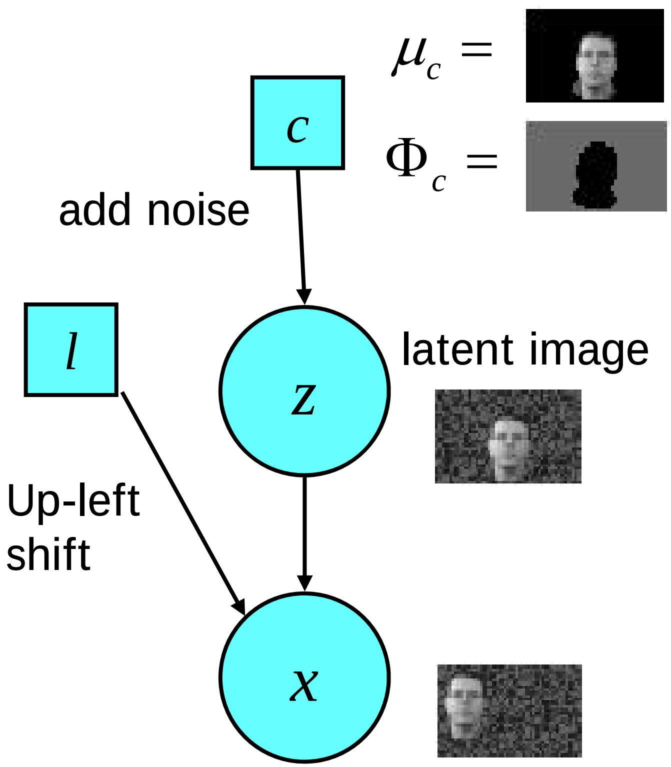
典范函数的运算法则

Linear-Gaussian CPD
Entering evidence
Product
Marginalization

使函数操作变得容易

Transformed mixture of Gaussians (TMG)

$$p(x,l,z,c) = p(c)\,p(l)\,p(z\mid c)\,p(x\mid l,z)$$
$$= \pi_c p_l \underbrace{N(z\mid \mu_c,\Phi_c)}_{\phi(c,z)} \underbrace{N(x\mid \Gamma_l z,\Psi)}_{\phi(l,x,z)}$$



试求：$p(c, l\mid x) = ?$

Conditional-Gaussian

Conditional-Gaussian 定义：$p(y\,|\,u,i)=N\left(y\,|\,B_i u+\mu_i,\Sigma_i\right),i=1,\cdots,M$

典范表示：$(\,g_i,\,h_i,\,K_i\,)\;i=1,\cdots,M$

条件典范函数定义：$\phi\left(x,\;i\,|\,g_i,h_i,K_i\right)=\exp\left\{g_i+x^T h_i-\frac{1}{2}x^T K_i x\right\}$

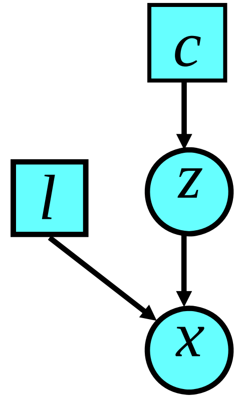
对离散变量 i 的每个可能取值，有一个连续变量 x 的典范函数

除了消除离散变量有所不同外，

条件典范函数的操作均与前面一样（对离散变量的每个可能取值分别进行）。

$$\phi\left(x,i\,|\,g_{1,i},h_{1,i},K_{1,i}\right)\cdot\phi\left(x,k\,|\,g_{2,k},h_{2,k},K_{2,k}\right)=\phi\left(x,i,k\,|\,g_{1,i}+g_{2,k},h_{1,i}+h_{2,k},K_{1,i}+K_{2,k}\right)$$

TMG Example



$p(c, l \mid x) = ?$

$$p(x,l,z,c) = \pi_c p_l N(z \mid \mu_c, \Phi_c) N(x \mid \Gamma_l z, \Psi)$$

$$p(c,l \mid x) \propto p(c,l,x)$$

$$= \int_z \pi_c p_l N(z \mid \mu_c, \Phi_c) N(x \mid \Gamma_l z, \Psi)$$

$$= \pi_c p_l \int_z \underbrace{\phi(c,z)\phi(l,x,z)}_{\phi(c,l,x,z)}$$

$$\propto \pi_c p_l N(x \mid \Gamma_l \mu_c, \ \Gamma_l \Phi_c \Gamma_l^T + \Psi)$$

条件典范函数的操作均与前面一样（对离散量的每个可能取值分别进行）

$$N(x \mid Bz + \mu, \Sigma) = \phi(x, z \mid g, h, K)$$

($B \to \Gamma_l$, $\mu \to 0$, $\Sigma \to \Psi$)

$$\begin{pmatrix} I \\ -B^T \end{pmatrix} \Sigma^{-1} \mu \qquad \begin{pmatrix} I \\ -B^T \end{pmatrix} \Sigma^{-1} (I, -B)$$

Conditional-Gaussian

条件典范函数定义：$\phi(x,\ i \mid g_i, h_i, K_i) = \exp\left\{ g_i + x^T h_i - \frac{1}{2} x^T K_i x \right\}$

除了消除离散变量有所不同外，

条件典范函数的操作均与前面一样（对离散变量的每个可能取值分别进行）。

表示了M个高斯（假设 i 有M个可能取值）

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

消除变量 i，得到M个高斯的混合

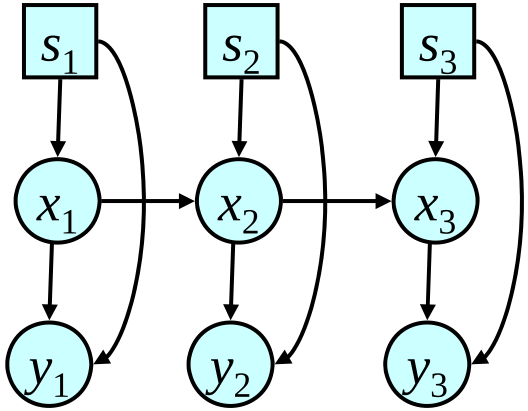
条件典范函数消除离散变量后得到的结果函数，不再具有典范函数的表示

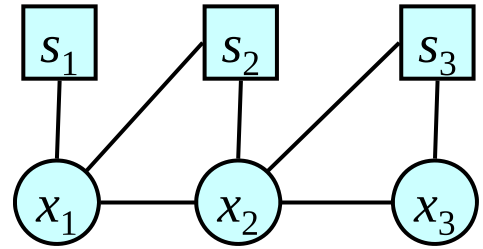
（i并没有真正被消除）

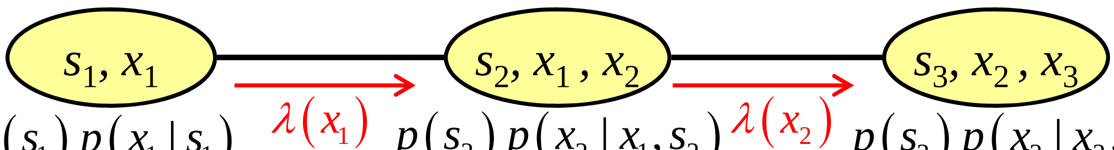
Problem with strict marginalization

Switching KF

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







$p(s_1)\,p(x_1 \mid s_1)\,p(y_1 \mid s_1, x_1)$

$p(s_2)\,p(x_2 \mid x_1, s_2)\,p(y_2 \mid s_2, x_2)$

$p(s_3)\,p(x_3 \mid x_2, s_3)\,p(y_3 \mid s_3, x_3)$

$$\lambda(x_1) = \sum_i p(s_1 = i)\,p(x_1 \mid s_1 = i)\,p(y_1 \mid s_1 = i, x_1)$$

$$\lambda(x_2) = \sum_j \int_{x_1} \lambda(x_1)\,p(s_2 = j)\,p(x_2 \mid x_1, s_2 = j)\,p(y_2 \mid s_2 = j, x_2)$$

$$= \sum_j \sum_i \int_{x_1} p(s_1 = i)\,p(x_1 \mid s_1 = i)\,p(y_1 \mid s_1 = i, x_1)\,p(s_2 = j)\,p(x_2 \mid x_1, s_2 = j)\,p(y_2 \mid s_2 = j, x_2)$$

消息函数 $\lambda(x_t)$ 是 M^t 个高斯的混合

Weak marginalization

- 条件典范函数消除离散变量后得到的新函数，不再具有典范函数的表示

表示了M个高斯（假设j有M个可能取值）

$$\sum_j \phi\left(x, j \mid p_j, \mu_j, \Sigma_j\right) = \phi\left(x \mid \left\{p_j, \mu_j, \Sigma_j\right\}_{1\le j\le M}\right)$$

消除变量j，得到M个高斯的混合

一种阻止这种指数增长的办法是：将 M 个高斯混合压成（collapse）1 个高斯

最小化 $\phi\left(x \mid \left\{p_j, \mu_j, \Sigma_j\right\}_{1\le j\le M}\right)$ 与 $\phi\left(x \mid \tilde{p}, \tilde{\mu}, \tilde{\Sigma}\right)$ 之间的 KL 距离

Weak marginalization

Moment matching

Assumed density filtering

$$\tilde{p} = \sum_j p_j$$

$$\tilde{p}_j = p_j / \tilde{p}$$

$$\tilde{\mu} = \sum_j \mu_j \tilde{p}_j$$

$$\tilde{\Sigma} = \sum_j \Sigma_j \tilde{p}_j + \sum_j \left(\mu_j - \tilde{\mu}\right)\left(\mu_j - \tilde{\mu}\right)^T \tilde{p}_j$$

含有连续变量的图模型推理

Cluster-tree elimination算法依然成立

算法本身并没有对联合分布中的变量是离散还是连续做限制

Linear-Gaussian KalmanFilter	Conditional Gaussian (CG) Mixed discrete-Gaussian	Non-linear non-Gaussian
典范函数的运算法则 Linear-Gaussian CPD Entering evidence Product Marginalization 使函数操作变得容易	Strict marginalization is feasible, when 先消除连续变量， 再消除离散变量 Weak marginalization Moment matching Assumed density filtering	Exact inference is usually not feasible ! Two classes of approximation: Stochastic Deterministic