Bayesian Networks with Continious Distributions

Sven Laur

March 17, 2009

1 Theoretical Background

Although it is common to consider Bayesian Networks consisting of nodes with discrete variables, there are no theoretical reasons why a Bayesian network cannot contain continuous variables. The main limiting reason is technical. If a distribution is continuous, then marginalisation becomes a challenging problem. The resulting solution is analytical only for few types of distributions. As an illustrating example, consider a Bayesian network depicted in Figure 1.

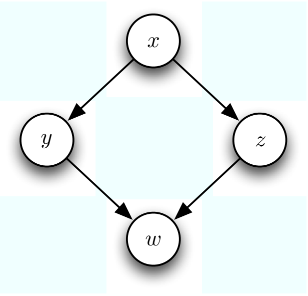

Figure 1: Diamond shape Bayesian network

When random variables x, y, z and w are continuous, we have to integrate out x, y and z in order to compute the marginal density of w:

$$\begin{aligned} p(w) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(w|y,z)p(y|x)p(z|x)dxdydz \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(w|y,z)\cdot\underbrace{\left(\int_{-\infty}^{\infty} p(y|x)p(z|x)dx\right)}_{\phi_1(y,z)}\cdot dydz \enspace . \end{aligned}$$

The solution is analytically tractable if we can find all integrals symbolically. Moreover, if we want to compute the marginal distributions with a computer program and not by hand, all potential functions $\phi_i(\ldots)$ must be in the form

$$\phi_i(\ldots) = f_{\boldsymbol{\theta}}(\ldots)$$

where $\mathcal{F} = \{f_{\boldsymbol{\theta}}\}$ is a function family indexed by a parameter vector $\boldsymbol{\theta}$. In this case, all standard evidence propagation algorithms work—instead of tables for functions $\phi_i(\ldots)$ you have to pass parameter vectors $\boldsymbol{\theta}_i$ through the network.

The necessary symbolic computations can be implemented as algorithm if the function family $\mathcal{F}$ is closed under the multiplication and marginalisation and the effects can be directly expressed in terms of parameter vectors:

$$\begin{aligned} f_{\alpha(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)}(\boldsymbol{x}) &= f_{\boldsymbol{\theta}_1}(\boldsymbol{x}) f_{\boldsymbol{\theta}_2}(\boldsymbol{x}) \\ f_{\beta(\boldsymbol{\theta},\mathcal{I})}(\boldsymbol{x}_{\mathcal{I}}) &= \int\limits_{-\infty}^{\infty} \cdots \int\limits_{-\infty}^{\infty} f_{\boldsymbol{\theta}}(\boldsymbol{x}) d\boldsymbol{x}_{\mathcal{J}} \end{aligned}$$

where $\alpha(\cdot,\cdot)$ and $\beta(\cdot,\cdot)$ are computable functions, $\boldsymbol{x}_{\mathcal{I}} = (x_{i_1}, \ldots, x_{i_k})$ for any index set $\mathcal{I} \subseteq \{1, \ldots, n\}$ and its complement $\mathcal{J} = \{1, \ldots, n\} \setminus \mathcal{I}$.

Concrete example. Although many families of exponential distributions are known to be closed under the multiplication, the multivariate normal distribution is also closed under the marginalisation. Recall that the density function of a multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the following

$$p(x_1, \ldots, x_n) = \frac{1}{\sqrt{2\pi \det \boldsymbol{\Sigma}}} \cdot \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right)$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$ is the vector of expected values of $\boldsymbol{x}$ and $n \times n$ square matrix $\boldsymbol{\Sigma} = (\sigma_{ij})$ consists of pairwise correlations between variables x_i and x_j.

For the marginalisation, one has to drop all coordinates from $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ that are integrated out, see The Matrix Cookbook [PP08, p. 34] or any statistical handbook for further details. When two distributions are multiplied, then

$$p_1(\boldsymbol{x}) p_2(\boldsymbol{x}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_1)^{\mathrm{T}} \boldsymbol{\Sigma}_1^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_1) - \frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_2)^{\mathrm{T}} \boldsymbol{\Sigma}_2^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_2)\right)$$

and new parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ can be found by converting

$$(\boldsymbol{x} - \boldsymbol{\mu}_1)^{\mathrm{T}} \boldsymbol{\Sigma}_1^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_1) + (\boldsymbol{x} - \boldsymbol{\mu}_2)^{\mathrm{T}} \boldsymbol{\Sigma}_2^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_2)$$

into a canonical form

$$(\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \ .$$

2 Theoretical Task

First, derive formulae for multiplication and marginalisation of normal distributions. Second, verify that when the conditional distributions are specified through a normal distributions

$$p(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{\sqrt{2\pi \det \boldsymbol{\Sigma}}} \cdot \exp\left(-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})\right) \quad ,$$

then all computations can be done with formulae you derived as a solution for the first task. The condition $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}, \boldsymbol{\Sigma})$ has a natural interpretation that $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} + \boldsymbol{\varepsilon}$ where ε is possibly coloured Gaussian noise. Such models naturally occur in physics, economics and chemistry.

Thirdly, implement the evidence propagation algorithm for normal distributions. You can use any open or close source frameworks for building junction trees and propagating evidence in them. All results needed from matrix algebra can be found in The Matrix Cookbook [PP08].

3 Practical Task

Kalman filter is one of the most famous example of continuous dynamic Bayesian network, see Figure 2. The corresponding conditional probabilities are

$$\begin{aligned}\boldsymbol{u_k} &\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{P})\\ \boldsymbol{x}_k &\sim \mathcal{N}(\boldsymbol{F}\boldsymbol{x}_{k-1} + \boldsymbol{B}\boldsymbol{u}_k, \boldsymbol{Q})\\ \boldsymbol{z}_k &\sim \mathcal{N}(\boldsymbol{H}\boldsymbol{x}_k, \boldsymbol{R})\end{aligned}$$

where $\boldsymbol{u}_k$ is a control signal, $\boldsymbol{x}_k$ is the internal state of the system, $\boldsymbol{z}$ is the observable output, $\boldsymbol{B}$ is control-input model, $\boldsymbol{F}$ is a state transformation matrix, $\boldsymbol{H}$ is observation model and all other matrices describe noise in the system.

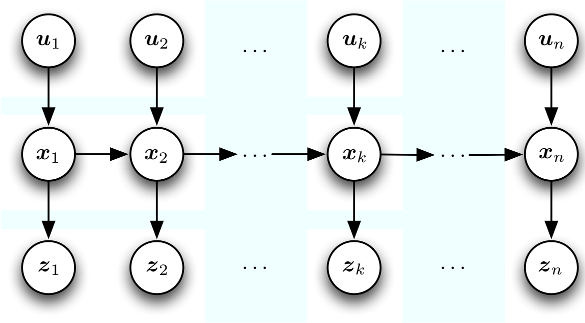

Figure 2: Kalman filter as a Bayesian network

In the standard setting, the control and output signals $\boldsymbol{u}_k$ and $\boldsymbol{z}_k$ are known and one tries to estimate internal state $\boldsymbol{x}_k$. Choose your favourite model of a linear system such as inverted pendulum, electronic circuit, and simulate the data. After that use evidence propagation algorithms in order to estimate the internal state and report how precise results you get.

Sometimes not all control and output signals are not available. Test the evidence propagation algorithm in that setting and characterise how does the precision of predictions depend on the amount of missing values. In particular, consider a case, where you have record inputs and outputs after ℓ time steps.

Finally, note that non-linear systems can always be linearised if we consider evolution of the system in small enough time steps. As a result, we can view such systems as Kalman filters, which are measured after ℓ time steps. Use the results obtained above and possibly new simulations to discuss whether such a technique leads to useful predictions.

References

[PP08] Kaare Brandt Petersen and Michael Syskind Pedersen. The matrix cookbook. Public draft `http://matrixcookbook.com`, February 16 2008.