The Basic Example Solved with Variational Bayesian Inference

Umut Orguner

March 3, 2011

We consider the following scalar linear system.

$$x_{k+1} = \theta x_k + v_k \tag{1}$$
$$y_k = 0.5 x_k + e_k \tag{2}$$

where

- $x_k \in \mathbb{R}$ is the state variable with initial distribution $x_0 \sim \mathcal{N}(x_0; \bar{x}_0, \Sigma_0)$.
- $y_k \in \mathbb{R}$ is the measurement.
- θ is an unknown parameter with prior distribution $\theta \sim \mathcal{N}(\theta; 0, \sigma_\theta^2)$
- The white process noise $v_k \in \mathbb{R}$ and white measurement noise $e_k \in \mathbb{R}$ are distributed as

$$\begin{bmatrix} v_k \\ e_k \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} v_k \\ e_k \end{bmatrix}; \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_v^2 & 0 \\ 0 & \sigma_e^2 \end{bmatrix} \right) \tag{3}$$

The aim is to find a possibly approximate estimate of the posterior distribution $p(\theta, x_{0:N}|y_{0:N})$.

There is no exact formula for this posterior distribution. in the variational Bayesian inference we make the approximation

$$p(\theta, x_{0:N}|y_{0:N}) \approx q(\theta, x_{0:N}) \triangleq q_\theta(\theta) q_x(x_{0:N}) \tag{4}$$

It is known that, in the iterative scheme of minimizing $\mathrm{KL}(q(\theta, x_{0:N})||p(\theta, x_{0:N}|y_{0:N}))$, i.e., variational Bayes, one gets the following identities.

$$\log q_\theta(\theta) = E_{q_x}\left[\log p(y_{0:N}, x_{0:N}, \theta)\right] + \text{const.} \tag{5}$$
$$\log q_x(x_{0:N}) = E_{q_\theta}\left[\log p(y_{0:N}, x_{0:N}, \theta)\right] + \text{const.} \tag{6}$$

We can write the joint density $p(y_{0:N}, x_{0:N}, \theta)$ as follows.

$$p(y_{0:N}, x_{0:N}, \theta) = p(y_{0:N}|x_{0:N}) p(x_{1:N}|x_{0:N-1}, \theta) p(x_0) p(\theta) \tag{7}$$
$$= \prod_{i=0}^{N} p(y_i|x_i) \prod_{i=1}^{N} p(x_i|x_{i-1}, \theta) p(x_0) p(\theta) \tag{8}$$
$$= \prod_{i=0}^{N} \mathcal{N}(y_i; 0.5 x_i, \sigma_e^2) \prod_{i=1}^{N} \mathcal{N}(x_i; \theta x_{i-1}, \sigma_v^2) \mathcal{N}(x_0; \bar{x}_0, \sigma_0^2) \mathcal{N}(\theta; 0, \sigma_\theta^2) \tag{9}$$

Taking the logarithm of both sides, we get.

$$\log p(y_{0:N}, x_{0:N}, \theta) = \sum_{i=0}^{N} -\frac{0.5}{\sigma_e^2}(y_i - 0.5x_i)^2 + \sum_{i=1}^{N} -\frac{0.5}{\sigma_v^2}(x_i - \theta x_{i-1})^2 - \frac{0.5}{\sigma_0^2}(x_0 - \bar{x}_0)^2 - \frac{0.5}{\sigma_\theta^2}\theta^2 + \text{const.} \tag{10}$$

where we included the terms that do not involve any variables into the constant term.

- **Calculation of** $p_\theta(\theta)$:

$$\begin{aligned}
\log q_\theta(\theta) =& E_{q_x}\left[\log p(y_{0:N}, x_{0:N}, \theta)\right] + \text{const.} & (11)\\
=& E_{q_x}\left[\log p(x_{1:N}|x_{0:N-1}, \theta) + \log p(\theta)\right] + \text{const.} & (12)\\
=& E_{q_x}\left[\sum_{i=1}^{N} -\frac{0.5}{\sigma_v^2}(x_i - \theta x_{i-1})^2 + \log p(\theta)\right] + \text{const.} & (13)\\
=& E_{q_x}\left[\sum_{i=1}^{N} -\frac{0.5}{\sigma_v^2}(x_i - \theta x_{i-1})^2\right] + \log p(\theta) + \text{const.} & (14)\\
=& \sum_{i=1}^{N} -\frac{0.5}{\sigma_v^2} E_{q_x}(x_i - \theta x_{i-1})^2 + \log p(\theta) + \text{const.} & (15)\\
=& \sum_{i=1}^{N} \left[-\frac{0.5}{\sigma_v^2}(-2\theta\overline{x_i x_{i-1}} + \theta^2\overline{x_{i-1}^2})\right] + \log p(\theta) + \text{const.} & (16)\\
=& \sum_{i=1}^{N} \left[-\frac{0.5}{\sigma_v^2/\overline{x_{i-1}^2}}\left(\theta - \frac{\overline{x_i x_{i-1}}}{\overline{x_{i-1}^2}}\right)^2\right] + \log p(\theta) + \text{const.} & (17)\\
=& \sum_{i=1}^{N} \log \mathcal{N}\left(\theta; \frac{\overline{x_i x_{i-1}}}{\overline{x_{i-1}^2}}, \sigma_v^2/\overline{x_{i-1}^2}\right) + \log \mathcal{N}(\theta; 0, \sigma_\theta^2) + \text{const.} & (18)
\end{aligned}$$

where only the terms in (10) depending on only θ are kept and the rest are included into the constant. The terms $\overline{x_i x_{i-1}}$ and $\overline{x_{i-1}^2}$ are defined as

$$\overline{x_i x_{i-1}} \triangleq E_{q_x}(x_i x_{i-1}) \quad \text{and} \quad \overline{x_{i-1}^2} \triangleq E_{q_x}(x_{i-1}^2) \tag{19}$$

Now, we can write $q_\theta(\theta)$ based on (18) as

$$\begin{aligned}
q_\theta(\theta) \propto& \prod_{i=1}^{N} \mathcal{N}\left(\theta; \frac{\overline{x_i x_{i-1}}}{\overline{x_{i-1}^2}}, \sigma_v^2/\overline{x_{i-1}^2}\right) \mathcal{N}(\theta; 0, \sigma_\theta^2) & (20)\\
=& \mathcal{N}\left(\theta; \overline{\theta}, \text{Var}(\theta)\right) & (21)
\end{aligned}$$

where

$$\text{Var}(\theta) = \left(\sum_{i=1}^{N} \frac{\overline{x_{i-1}^2}}{\sigma_v^2} + \frac{1}{\sigma_\theta^2}\right)^{-1} \tag{22}$$

$$\overline{\theta} \triangleq \text{Var}(\theta) \sum_{i=1}^{N} \frac{\overline{x_i x_{i-1}}}{\sigma_v^2} \tag{23}$$

- **Calculation of $p_x(x_{0:N})$:**

$$
\begin{aligned}
\log q_x(x_{0:N}) =& E_{q_\theta}\left[\log p(y_{0:N}, x_{0:N}, \theta)\right] + \text{const.} && (24)\\
=& E_{q_\theta}\left[\log p(y_{0:N}|x_{0:N}) + \log p(x_{1:N}|x_{0:N-1}, \theta) + \log p(x_0)\right] + \text{const.} && (25)\\
=& \log p(y_{0:N}|x_{0:N}) + E_{q_\theta}\left[\log p(x_{1:N}|x_{0:N-1}, \theta)\right] + \log p(x_0) + \text{const.} && (26)\\
=& \log p(y_{0:N}|x_{0:N}) + E_{q_\theta}\left[\sum_{i=0}^{N} -\frac{0.5}{\sigma_v^2}(x_i - \theta x_{i-1})^2\right] + \log p(x_0) + \text{const.} && (27)\\
=& \log p(y_{0:N}|x_{0:N}) + \sum_{i=1}^{N}\left[-\frac{0.5}{\sigma_v^2}(x_i - \overline{\theta} x_{i-1})^2 - \frac{0.5}{\sigma_v^2}\operatorname{Var}(\theta) x_{i-1}^2\right] \\
&+ \log p(x_0) + \text{const.} && (28)\\
=& \log p(y_{0:N}|x_{0:N}) + \sum_{i=1}^{N}\left[\log \mathcal{N}(x_i; \overline{\theta} x_{i-1}, \sigma_v^2) + \log \mathcal{N}(x_{i-1}; 0, \sigma_v^2/\operatorname{Var}(\theta))\right] \\
&+ \log p(x_0) + \text{const.} && (29)\\
=& \log p(y_{0:N}|x_{0:N}) + \log p(x_{1:N}|x_{0:N-1}, \overline{\theta}) + \sum_{i=1}^{N} \log \mathcal{N}(x_{i-1}; 0, \sigma_v^2/\operatorname{Var}(\theta)) \\
&+ \log p(x_0) + \text{const.} && (30)
\end{aligned}
$$

where only the terms in (10) depending on only $x_{0:N}$ are kept and the rest are included into the constant. The terms $\overline{\theta}$ and $\operatorname{Var}\theta$ are

$$
\begin{aligned}
\overline{\theta} =& E_{q_\theta}(\theta) && (31)\\
\operatorname{Var}(\theta) =& E_{q_\theta}((\theta - \overline{\theta})^2) && (32)
\end{aligned}
$$

which are equivalent to those defined in the calculation of $q_\theta(\theta)$. Now, we can write $q_x(x_{0:N})$ as

$$q_x(x_{0:N}) \propto p(y_{0:N}|x_{0:N}) p(x_{1:N}|x_{0:N-1}, \overline{\theta}) \prod_{i=0}^{N-1} \mathcal{N}(x_i; 0, \sigma_v^2/\operatorname{Var}(\theta)) p(x_0) \tag{33}$$

in which all of the terms are Gaussian. The densities $\prod_{i=1}^{N} \mathcal{N}(x_{i-1}; 0, \sigma_v^2/\operatorname{Var}(\theta))$ must be interpreted as a pseudo likelihood which provides extra measurements defined as

$$\tilde{y}_k = 0 = x_k + \tilde{e}_k \tag{34}$$

for $k = 0, \ldots, N-1$ where $\tilde{e}_k \sim \mathcal{N}(\tilde{e}_k; 0, \sigma_v^2/\operatorname{Var}(\theta))$. The final density estimate is written as

$$q_x(x_{0:N}) = p(x_{0:N}|y_{0:N}, \tilde{y}_{0:N-1}, \overline{\theta}) \tag{35}$$

which is a Gaussian density that can be calculated by running a Kalman smoother on the data $y_{0:N}$ and $\tilde{y}_{0:N-1}$ using the parameter estimate $\overline{\theta}$ in the state model.