

Expectation Maximization calculation for the DTQ method

We consider a parameteric SDE model:

$$dX_t = f(X_t; \boldsymbol{\theta})dt + g(X_t; \boldsymbol{\theta})dW_t \quad (1)$$

In this model $f(X_t; \boldsymbol{\theta})$ is the drift function and $g(X_t; \boldsymbol{\theta})$ is the diffusion function. A concrete example of such an SDE is the Ornstein-Uhlenback SDE (linear in nature),

$$dX_t = \theta_1(\theta_2 - X_t)dt + \theta_3dW_t \quad (2)$$

We start with the parameter inference problem where we have data available as a time series, denoted by $\mathbf{x} = (x_0, x_1, \dots, x_N)$. Since the observed data might be have large inter-observation times, we consider intermediate points which we consider as *missing data points*, denoted by \mathbf{z} . On the interval $[t_i, t_{i+1}]$, we have 2 observed data points, $X_{t_i} = x_i$ and $X_{t_{i+1}} = x_{i+1}$. We consider F missing data points on this interval, denoted by $z_{i,F}$, the first subscript corresponding to the interval and the second subscript for the missing data point on the interval. Thus the missing data on an interval $[t_i, t_{i+1}]$, can be represented as $\mathbf{z}_i = (z_{i1}, z_{i2}, \dots, z_{iF})$. The complete data on this interval would thus become $(x_i, z_{i1}, z_{i2}, \dots, z_{iF}, x_{i+1})$, comprising of the observed data and the unknown missing data that we introduced.

1 EM algorithm

The Expectation-Maximization algorithm consists of 2 steps, computing the expectation of the log likelihood function and maximizing this value with respect to the parameters.

1. Start with an initial guess for the parameter, $\boldsymbol{\theta}^{(0)}$
2. For the expectation step,

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) = \mathbb{E}_{\mathbf{z}|\mathbf{x}, \boldsymbol{\theta}^{(k)}}[\log p(\mathbf{x}, \mathbf{z} | \boldsymbol{\theta})] \quad (3)$$

$$= \sum_{\mathbf{z}} \underbrace{\log p(\mathbf{x}, \mathbf{z} | \boldsymbol{\theta})}_{\text{Part I}} \cdot \underbrace{p(\mathbf{z} | \mathbf{x}, \boldsymbol{\theta}^{(k)})}_{\text{Part II}} \quad (4)$$

3. For the maximization step, we start with the current iterate and a dummy variable $\boldsymbol{\theta}$, so that the next iterate of the parameters would be the maximal value of $\boldsymbol{\theta}$

$$\boldsymbol{\theta}^{(k+1)} = \arg \max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) \quad (5)$$

We can either use a numerical optimizer for the optimization step or differentiate the $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ function with respect to $\boldsymbol{\theta}$ vector and equate it to zero to get the maximal value.

4. Iterate Step 2 and 3 until convergence.

1.1 Computation of the complete log likelihood

The first part of the expectation is the complete likelihood, $\log p(\mathbf{x}, \mathbf{z} \mid \boldsymbol{\theta})$, which can be expanded as,

$$\begin{aligned} \log p(\mathbf{x}, \mathbf{z} \mid \boldsymbol{\theta}) = & \log p(x_0 \mid \boldsymbol{\theta}) + \underbrace{\sum_{i=0}^{N-1} \log p(z_{i1} \mid x_i, \boldsymbol{\theta})}_{(1)} + \underbrace{\sum_{i=0}^{N-1} \sum_{j=1}^{F-1} \log p(z_{i,j+1} \mid z_{ij}, \boldsymbol{\theta})}_{(2)} \\ & + \underbrace{\sum_{i=0}^{N-1} \log p(x_{i+1} \mid z_{iF}, \boldsymbol{\theta})}_{(3)} \end{aligned} \quad (6)$$

The expression can be simplified under the assumption that F is sufficiently large so that we can make an assumption that one-step transition densities in (1), (2) and (3) follow Gaussian distribution. Thus all the terms, $p(z_{i1} \mid x_i, \boldsymbol{\theta})$, $p(z_{i,j+1} \mid z_{ij}, \boldsymbol{\theta})$ and $p(x_{i+1} \mid z_{iF}, \boldsymbol{\theta})$ can be expressed with a Gaussian function

$$G(x, y, \boldsymbol{\theta}) = p(x \mid y, \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi g^2(y, \boldsymbol{\theta})h}} \exp \left(-\frac{1}{2g^2(y, \boldsymbol{\theta})h} (x - y - f(y, \boldsymbol{\theta})h)^2 \right)$$

1.2 Computation of the density of the missing data points

Looking back at the expectation equation (3), the expected value if computed by summing over all \mathbf{z} values which is a nested integral. Since the log likelihood can be expanded in 4 terms, so the density $p(\mathbf{z} \mid \mathbf{x}, \boldsymbol{\theta}^{(k)})$ gets multiplied by each of these terms. Upon summing over all the values of \mathbf{z} , there will be 3 steps of terms remaining, corresponding to the respective terms in the equation (6), as described below,

1. Corresponding to term (1), we have the term $p(z_{i1} \mid \mathbf{x}, \boldsymbol{\theta}^{(k)})$. Using Bayes theorem we get,

$$\begin{aligned} p(z_{i1}, \mathbf{x} \mid \boldsymbol{\theta}^{(k)}) &= p(z_{i1} \mid \mathbf{x}, \boldsymbol{\theta}^{(k)}) \cdot p(\mathbf{x} \mid \boldsymbol{\theta}^{(k)}) \\ \implies p(z_{i1} \mid \mathbf{x}, \boldsymbol{\theta}^{(k)}) &= \frac{p(z_{i1}, \mathbf{x} \mid \boldsymbol{\theta}^{(k)})}{p(\mathbf{x} \mid \boldsymbol{\theta}^{(k)})} = \frac{p(z_{i1}, \mathbf{x} \mid \boldsymbol{\theta}^{(k)})}{p(x_0 \mid \boldsymbol{\theta}^{(k)}) \prod_{j=0}^{N-1} p(x_{j+1} \mid x_j, \boldsymbol{\theta}^{(k)})} \end{aligned} \quad (7)$$

The numerator for the expression can be expanded using Markov property as,

$$\begin{aligned} p(z_{i1}, \mathbf{x} \mid \boldsymbol{\theta}^{(k)}) &= p(z_{i1}, x_0, x_1, \dots, x_N \mid \boldsymbol{\theta}^{(k)}) \\ &= p(x_0 \mid \boldsymbol{\theta}^{(k)}) \prod_{j=i+1}^{N-1} p(x_{j+1} \mid x_j, \boldsymbol{\theta}^{(k)}) p(x_{i+1} \mid z_{i1}, \boldsymbol{\theta}^{(k)}) p(z_{i1} \mid x_i, \boldsymbol{\theta}^{(k)}) \prod_{j=0}^{i-1} p(x_{j+1} \mid x_j, \boldsymbol{\theta}^{(k)}) \end{aligned}$$

Substituting the expression for $p(z_{i1}, \mathbf{x} \mid \boldsymbol{\theta}^{(k)})$ in equation (7) and expanding the denominator using Markov property in a similar way gives,

$$p(z_{i1}, \mathbf{x} \mid \boldsymbol{\theta}^{(k)}) = \frac{p(x_{i+1} \mid z_{i1}, \boldsymbol{\theta}^{(k)}) p(z_{i1} \mid x_i, \boldsymbol{\theta}^{(k)})}{p(x_{i+1} \mid x_i, \boldsymbol{\theta}^{(k)})} \quad (8)$$

2. Corresponding to the F internal steps represented by term (2), we have the terms $p(z_{i,j+1}, z_{ij} \mid \mathbf{x}, \boldsymbol{\theta}^{(k)})$. We again use Bayes theorem in a similar way as before to get,

$$p(z_{i,j+1}, z_{ij} \mid \mathbf{x}, \boldsymbol{\theta}^{(k)}) = \frac{p(z_{i,j+1}, z_{ij}, \mathbf{x} \mid \boldsymbol{\theta}^{(k)})}{p(x_0 \mid \boldsymbol{\theta}^{(k)}) \prod_{j=0}^{N-1} p(x_{j+1} \mid x_j, \boldsymbol{\theta}^{(k)})} \quad (9)$$

The numerator can be expanded using Markov property as follows,

$$\begin{aligned} p(z_{i,j+1}, z_{ij}, \mathbf{x} \mid \boldsymbol{\theta}^{(k)}) &= p(x_0 \mid \boldsymbol{\theta}^{(k)}) \prod_{j=0}^{i-1} p(x_{j+1} \mid x_j, \boldsymbol{\theta}^{(k)}) \cdot p(z_{ij} \mid x_i, \boldsymbol{\theta}^{(k)}) \cdot p(z_{i,j+1} \mid z_{ij}, \boldsymbol{\theta}^{(k)}) \\ &\quad \cdot p(x_{i+1} \mid z_{i,j+1}, \boldsymbol{\theta}^{(k)}) \prod_{j=1}^{N-1} p(x_{j+1} \mid x_j, \boldsymbol{\theta}^{(k)}) \end{aligned} \quad (10)$$

$$\implies p(z_{i,j+1}, z_{ij} \mid \mathbf{x}, \boldsymbol{\theta}^{(k)}) = \frac{p(z_{ij} \mid x_i, \boldsymbol{\theta}^{(k)}) p(z_{i,j+1} \mid z_{ij}, \boldsymbol{\theta}^{(k)}) p(x_{i+1} \mid z_{i,j+1}, \boldsymbol{\theta}^{(k)})}{p(x_{i+1} \mid x_i, \boldsymbol{\theta}^{(k)})} \quad (11)$$

3. The last term (3) has the corresponding term $p(z_{iF} \mid \mathbf{x}, \boldsymbol{\theta}^{(k)})$, similar to the first term,

$$p(z_{iF} \mid \mathbf{x}, \boldsymbol{\theta}^{(k)}) = \frac{p(x_{i+1} \mid z_{iF}, \boldsymbol{\theta}^{(k)}) p(z_{iF} \mid x_i, \boldsymbol{\theta}^{(k)})}{p(x_{i+1} \mid x_i, \boldsymbol{\theta}^{(k)})} \quad (12)$$

1.3 Expectation Step

Combining the terms from Section 1.1 and Section 1.2, we can form a complete expression for the expectation. Going back to Section 1.1, we recall that the transition densities can be assumed to be Gaussian for sufficiently large enough F . Thus, the expectation expression can be rewritten as,

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)}) &= \sum_{\mathbf{z}} \left\{ \log p(x_0 \mid \boldsymbol{\theta}) + \sum_{i=0}^{N-1} \log G(z_{i1}, x_i, \boldsymbol{\theta}) \cdot p(z_{i1} \mid \mathbf{x}, \boldsymbol{\theta}^{(k)}) \right. \\ &\quad + \sum_{i=0}^{N-1} \sum_{j=1}^{F-1} \log G(z_{i,j+1}, z_{ij}, \boldsymbol{\theta}) \cdot p(z_{i,j+1}, z_{ij} \mid \mathbf{x}, \boldsymbol{\theta}^{(k)}) \\ &\quad \left. + \sum_{i=0}^{N-1} \log G(x_{i+1}, z_{iF}, \boldsymbol{\theta}) \cdot p(z_{iF} \mid \mathbf{x}, \boldsymbol{\theta}^{(k)}) \right\} \end{aligned} \quad (13)$$

1.4 Maximization Step

For the maximization step, there are 2 ways to maximize $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$, either through numerical optimizers or through equating the derivative with respect to the parameters to zero and using a root-finding solver if required.

Since both these methods require the gradient of $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ so we specify the gradients below. The derivative of $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ with respect to the $\boldsymbol{\theta}$ parameters would then be,

$$\begin{aligned} 0 = \frac{\partial Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})}{\partial \theta_\ell} &= \frac{p'(x_0 | \boldsymbol{\theta})}{p(x_0 | \boldsymbol{\theta})} + \sum_{i=0}^{N-1} \frac{H_\ell(z_{i1}, x_i, \boldsymbol{\theta})}{G(z_{i1}, x_i, \boldsymbol{\theta})} \cdot p(z_{i1} | x_i, \boldsymbol{\theta}^{(k)}) \\ &\quad + \sum_{i=0}^{N-1} \sum_{j=1}^{F-1} \frac{H_\ell(z_{i,j+1}, z_{ij}, \boldsymbol{\theta})}{G(z_{i,j+1}, z_{ij}, \boldsymbol{\theta})} \cdot p(z_{i,j+1}, z_{ij} | \mathbf{x}, \boldsymbol{\theta}^{(k)}) \\ &\quad + \sum_{i=0}^{N-1} \frac{H_\ell(x_{i+1}, z_{iF}, \boldsymbol{\theta})}{G(x_{i+1}, z_{iF}, \boldsymbol{\theta})} \cdot p(z_{iF} | \mathbf{x}, \boldsymbol{\theta}^{(k)}) \end{aligned}$$

where,

$$H_\ell(x, y, \boldsymbol{\theta}) = \frac{\partial G(x, y, \boldsymbol{\theta})}{\partial \theta_\ell} = \frac{\partial G}{\partial f} \cdot \frac{\partial f}{\partial \theta_\ell} + \frac{\partial G}{\partial g} \cdot \frac{\partial g}{\partial \theta_\ell}$$

1. With respect to θ_1, θ_2

$$\frac{H_1}{G}(x, y, \boldsymbol{\theta}) = \frac{\partial f}{\partial \theta_1} \left[\frac{(x - y - f(y)h)}{g^2(y)} \right], \frac{H_2}{G}(x, y, \boldsymbol{\theta}) = \frac{\partial f}{\partial \theta_2} \left[\frac{(x - y - f(y)h)}{g^2(y)} \right]$$

2. With respect to θ_3

$$\frac{H_3}{G}(x, y, \boldsymbol{\theta}) = \frac{\partial g}{\partial \theta_3} \left[\frac{(x - y - f(y)h)^2}{hg^3(y)} - \frac{1}{g(y)} \right]$$

1.5 Summary of terms that need to be computed

1. $p(x_{i+1} | z_{i1}, \boldsymbol{\theta}^{(k)}) \cdot p(z_{i1} | x_i, \boldsymbol{\theta}^{(k)})$
2. $p(z_{ij} | x_i, \boldsymbol{\theta}^{(k)}) \cdot p(z_{i,j+1} | z_{ij}, \boldsymbol{\theta}^{(k)}) \cdot p(x_{i+1} | z_{i,j+1}, \boldsymbol{\theta}^{(k)})$
3. $p(x_{i+1} | z_{iF}, \boldsymbol{\theta}^{(k)}) \cdot p(z_{iF} | x_i, \boldsymbol{\theta}^{(k)})$
4. $p(x_{i+1} | x_i, \boldsymbol{\theta}^{(k)})$ - the complete DTQ computation
5. $\partial G / \partial f, \partial G / \partial g$ - computed in Dtheta function
6. $\partial f / \partial \theta_1, \partial f / \partial \theta_2, \partial g / \partial \theta_3$ - computed in Dtheta function