Rosen's paper review

Daeyoung Lim*
Department of Statistics
Korea University

May 20, 2016

1 Model Specifications

$$\boldsymbol{y}_i\left(t_{ij}\right) = \boldsymbol{X}_{ij}\boldsymbol{\mu}\left(t_{ij}\right) + \boldsymbol{Z}_{ij}\boldsymbol{g}_i\left(t_{ij}\right) + \boldsymbol{\delta}_i\left(t_{ij}\right) \tag{1}$$

where

$$\boldsymbol{y}_i\left(t_{ij}\right): \; p \times 1 \qquad i = 1,\ldots,n, \; j = 1,\ldots,m_i \tag{2}$$
$$\boldsymbol{\mu}\left(t\right) = \left(\boldsymbol{\mu}_1'\left(t\right),\ldots,\boldsymbol{\mu}_p'\left(t\right)\right)' \tag{3}$$
$$\boldsymbol{\mu}_k'\left(t\right) = \left(\mu_{k1}\left(t\right),\ldots,\mu_{kr}\left(t\right)\right)': \qquad\qquad r \times 1 \tag{4}$$
$$\boldsymbol{g}_i\left(t\right) = \left(\boldsymbol{g}_{i1}'\left(t\right),\ldots,\boldsymbol{g}_{ip}'\left(t\right)\right)' \tag{5}$$
$$\boldsymbol{g}_{ik}\left(t\right) = \left(g_{ij1}\left(t\right),\ldots,g_{iks}\left(t\right)\right)': \qquad\qquad s \times 1 \tag{6}$$
$$\boldsymbol{x}_{ij}: \; r \times 1 \tag{7}$$
$$\boldsymbol{z}_{ij}: \; s \times 1 \tag{8}$$
$$\boldsymbol{X}_{ij} = \boldsymbol{I}_p \otimes \boldsymbol{x}_{ij}' \tag{9}$$
$$\boldsymbol{Z}_{ij} = \boldsymbol{I}_p \otimes \boldsymbol{z}_{ij}' \tag{10}$$
$$\boldsymbol{\delta}_i\left(t_{ij}\right) \sim \text{Ornstein-Uhlenbeck process} \tag{11}$$

2 Ornstein-Uhlenbeck process

An Ornstein-Uhlenbeck process is the second-order stationary process $\{X_t\}$, that satisfies the following differential equation:

$$dX\left(t\right) = -aX\left(t\right)\,dt + \sigma\,dB\left(t\right), \quad t \geq 0 \tag{12}$$

where $\{B\left(t\right)\}$ is standard Brownian motion, and a and $\sigma > 0$ are parameters and X_0 is a random variable that is independent of $\{B\left(t\right)\}$.

2.1 Multivariate Ornstien-Uhlenbeck process

The univariate version of the OU process naturally evolves to a multivariate form which reads as follows:

$$d\boldsymbol{X}\left(t\right) = -\boldsymbol{A}\boldsymbol{X}\left(t\right)\,dt + \boldsymbol{B}\,d\boldsymbol{W}\left(t\right), \tag{13}$$

where $\boldsymbol{A}, \boldsymbol{B}$ are constant matrices. The solution for this SDE is

$$\boldsymbol{X}\left(t\right) = \exp\left(-\boldsymbol{A}t\right)\boldsymbol{X}_0 + \int_0^t \exp\left\{-\boldsymbol{A}\left(t - t'\right)\right\}\boldsymbol{B}\,d\boldsymbol{W}\left(t\right). \tag{14}$$

The properties of such an OU process are as follows:

---

*Prof. Taeryon Choi

- If $\boldsymbol{A}$ has only eigenvalues with positive real part, a stationary solution exists of the form

$$\boldsymbol{X}_s(t) = \int_{-\infty}^t \exp\left\{-\boldsymbol{A}\left(t-t'\right)\right\} \boldsymbol{B}\, d\boldsymbol{W}(t). \tag{15}$$

The expected value $\mathrm{E}\left[\boldsymbol{X}_s(t)\right] = 0$ and the covariance matrix

$$\boldsymbol{\Sigma} = \mathrm{Cov}\left(\boldsymbol{X}_s(t), \boldsymbol{X}_s'(s)\right) = \int_{-\infty}^{\min(t,s)} \exp\left\{-\boldsymbol{A}\left(t-t'\right)\right\} \boldsymbol{B}\boldsymbol{B}' \exp\left\{-\boldsymbol{A}'\left(s-t'\right)\right\} dt'. \tag{16}$$

- The stationary covariance matrix satisfies the following equation:

$$\boldsymbol{A}\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\boldsymbol{A}' = \boldsymbol{B}\boldsymbol{B}'. \tag{17}$$

The solution to this equation is given by

$$\boldsymbol{\Sigma} = \frac{|\boldsymbol{A}|\,\boldsymbol{B}\boldsymbol{B}' + [\boldsymbol{A} - \mathrm{Tr}\,(\boldsymbol{A})\,\boldsymbol{I}]\,\boldsymbol{B}\boldsymbol{B}'\,[\boldsymbol{A} - \mathrm{Tr}\,(\boldsymbol{A})\,\boldsymbol{I}]}{2\,(\mathrm{Tr}\,(\boldsymbol{A}))\,|\boldsymbol{A}|} \tag{18}$$

- *(Time Correlation Matrix in the Stationary State)* The following relations hold.

$$\mathrm{Cov}\left(\boldsymbol{X}_s(t), \boldsymbol{X}_s'(s)\right) = \exp\left\{-\boldsymbol{A}\left(t-s\right)\right\} \boldsymbol{\Sigma}, \qquad\qquad t > s \tag{19}$$
$$= \boldsymbol{\Sigma}\exp\left\{-\boldsymbol{A}'\left(s-t\right)\right\}, \qquad\qquad t < s \tag{20}$$

- If all its transition densities depend only on the time differences, then the Markov process is *homogeneous*. OU process is homogeneous. Therefore, the transition probabiliy of an OU process is given by

$$\mathrm{P}\left(\boldsymbol{X}_s(t)\,|\,\boldsymbol{X}_s(s), t-s\right) = |2\pi\boldsymbol{\Omega}|^{-1/2} \exp\left\{-\frac{1}{2}\boldsymbol{\gamma}'\boldsymbol{\Omega}^{-1}\boldsymbol{\gamma}\right\} \tag{21}$$

where

$$\boldsymbol{\gamma} = \boldsymbol{X}_s(t) - \exp\left(-\boldsymbol{A}\left(t-s\right)\right)\boldsymbol{X}_s(s) \tag{22}$$
$$\boldsymbol{\Omega} = \boldsymbol{\Sigma} - \exp\left\{-\boldsymbol{A}\left(t-s\right)\right\}\boldsymbol{\Sigma}\exp\left\{-\boldsymbol{A}'\left(t-s\right)\right\}. \tag{23}$$

3 Cubic Spline Interpolation

When the data come in the form of pairs (x_i, y_i), it is often our interest to find out what underlying *function* they fall onto. There are many ways of achieving such a goal but one of the most popular methods is called *the cubic spline*. It is a nonparametric way of intepolation guaranteeing the continuous twice-differentiability. Unlike linear interpolation, the cubic spline is differentiable even at the end point of each interval. Let's pick an arbitrary interval between x_j and x_{j+1}. Following the notation of the book *Numerical Recipes in C*, $y_i = y(x_i)$, $i = 1, \ldots, N$. It is proven that the function can be uniquely constructed as

$$y = Ay_j + By_{j+1} + Cy_j'' + Dy_{j+1}'' \tag{24}$$

where A, B, C, D are

$$A \equiv \frac{x_{j+1} - x}{x_{j+1} - x_j} \qquad\qquad\qquad B \equiv \frac{x - x_j}{x_{j+1} - x_j} \tag{25}$$

$$C \equiv \frac{1}{6}\left(A^3 - A\right)(x_{j+1} - x_j)^2 \qquad\qquad D \equiv \frac{1}{6}\left(B^3 - B\right)(x_{j+1} - x_j)^2. \tag{26}$$

The first and second derivatives of y with respect to x are given by

$$\frac{dy}{dx} = \frac{y_{j+1} - y_j}{x_{j+1} - x_j} - \frac{3A^2 - 1}{6}(x_{j+1} - x_j)\,y_j'' + \frac{3B^2 - 1}{6}(x_{j+1} - x_j)\,y_{j+1}'' \tag{27}$$

$$\frac{d^2y}{dx^2} = Ay_j'' + By_{j+1}''. \tag{28}$$

3.1 Technical details regarding cubic splines

Accepting that the above equations are given, the only problem left for us is that we actually do not know y_i'' yet assumed them to be known when modeling. It is also given for $j = 2, \ldots, N-1$

$$\frac{x_j - x_{j-1}}{6} y_{j-1}'' + \frac{x_{j+1} - x_{j-1}}{3} y_j'' + \frac{x_{j+1} - x_j}{6} y_{j+1}'' = \frac{y_{j+1} - y_j}{x_{j+1} - x_j} - \frac{y_j - y_{j-1}}{x_j - x_{j-1}}. \tag{29}$$

In order to uniquely determine the curve, 2 more conditions need to be made. The most common ways of doing this are

- set one or both of y_1'' and y_N'' equal to zero, giving the so-called *natural cubic spline*, which has zero second derivative on one or both of its boundaries, or

- set either of y_1'' and y_N'' to values calculated from equation (27) so as to make the first derivative of the interpolating function have a specified value on either or both boundaries.

Part of cubic splines' popularity is due to the linearity of the equations as well as the tridiagonality. In particular, the tridiagonality makes the computation so much easier, enabling it to be solved in $O(N)$ operations instead of $O(N^3)$ for the general linear problem. Refer to *tridiagonal matrix algorithm* or equivalently *Thomas algorithm* for further details.

3.2 Cubic spline in the paper

Cubic splines can be rewritten as a specific case of B-splines. In the paper, the author constructs a general setting where the model is

$$y_i = f(x_i) + \epsilon_i, \tag{30}$$

where $\mathrm{E}(\epsilon_i) = 0$ and f is an unknown smooth function. Upon choosing the knots $\kappa_1, \ldots, \kappa_K$, the cubic spline is reexpressed as a linear combination of the basis functions $1, x, |x - \kappa_1|^3, \ldots, |x - \kappa_K|^3$:

$$f(x) = \beta_0 + \beta_1 x + \sum_{k=1}^{K} u_k |x - \kappa_k|^3. \tag{31}$$

It is, at all times, not a good idea to overload a notation for multiple different things but the author does not seem to care. Hence, again uses $\boldsymbol{\Omega}_K$ whose (k, ℓ)th element is $|\kappa_k - \kappa_\ell|^3$. The author of the paper draws another representation of the fitting problem in the form of a mixed model by recasting the function in a matrix form and writing down the frequentist loss function. The loss function is

$$L(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\theta}) = \sum_{i=1}^{n} (y_i - f(x_i))^2 + \frac{1}{\lambda} \boldsymbol{\theta}' \boldsymbol{D} \boldsymbol{\theta} \tag{32}$$

where $\boldsymbol{\theta} = (\beta_0, \beta_1, u_1, \ldots, u_K)'$ and the matrix $\boldsymbol{D}$ is

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{0}_{2 \times 2} & \boldsymbol{0}_{2 \times K} \\ \boldsymbol{0}_{K \times 2} & \boldsymbol{\Omega}_K \end{bmatrix}. \tag{33}$$

Now let's define $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{1} & \boldsymbol{x} \end{bmatrix}$ and $\boldsymbol{Z}_K = \begin{bmatrix} |x_i - \kappa_1|^3 & \cdots & |x_i - \kappa_K|^3 \end{bmatrix}_{1 \leq i \leq n}$. Now we can reformulate the equation (32):

$$\frac{1}{\sigma_\epsilon^2} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}_K \boldsymbol{u}\|^2 + \frac{1}{\lambda \sigma_\epsilon^2} \boldsymbol{u}' \boldsymbol{\Omega}_K \boldsymbol{u} \tag{34}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ and $\boldsymbol{u} = (u_1, \ldots, u_K)'$. We have just converted the cubic spline fitting problem into an equivalent linear mixed model fitting problem with a common setting of

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}_K \boldsymbol{u} + \boldsymbol{\epsilon}, \qquad \mathrm{Cov}(\boldsymbol{u}) = \sigma_u^2 \boldsymbol{\Omega}_K^{-1}. \tag{35}$$

Furthermore, by defining $\boldsymbol{b} = \boldsymbol{\Omega}_K^{1/2} \boldsymbol{u}$ and $\boldsymbol{Z} = \boldsymbol{Z}_K \boldsymbol{\Omega}_K^{-1/2}$, the equation (35) is the same as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{\epsilon}, \qquad \mathrm{Cov}\begin{pmatrix} \boldsymbol{b} \\ \boldsymbol{\epsilon} \end{pmatrix} = \begin{pmatrix} \sigma_b^2 \boldsymbol{I}_K & \boldsymbol{0} \\ \boldsymbol{0} & \sigma_\epsilon^2 \boldsymbol{I}_n \end{pmatrix}. \tag{36}$$

3

4 Priors and sampling

Recall the original setting:

$$\boldsymbol{y}_i\left(t_{ij}\right) = \boldsymbol{X}_{ij}\boldsymbol{\mu}\left(t_{ij}\right) + \boldsymbol{Z}_{ij}\boldsymbol{g}_i\left(t_{ij}\right) + \boldsymbol{\delta}_i\left(t_{ij}\right). \tag{37}$$

Using the above cubic spline approach, we have two coefficient functions to estimate: $\boldsymbol{\mu}\left(t_{ij}\right)$ and $\boldsymbol{g}_i\left(t_{ij}\right)$. Thus, we will, as we have above, transform them into matrix multiplications:

$$\mu_{k\ell}\left(t_{ij}\right) = \boldsymbol{\phi}_{ij}'\boldsymbol{\beta}_{k\ell} + \boldsymbol{\psi}_{ij}'\boldsymbol{v}_{k\ell} \tag{38}$$

$$g_{ikm}\left(t_{ij}\right) = \boldsymbol{\phi}_{ij}'\boldsymbol{w}_{ikm} + \boldsymbol{\psi}_{ij}'\boldsymbol{u}_{ikm} \tag{39}$$

where $k = 1, \ldots, p$, $\ell = 1, \ldots, r$, $m = 1, \ldots, s$, $i = 1, \ldots, n$ and $\kappa_1, \ldots, \kappa_K$ are knots which are sample quantiles of t_{ij} $(j = 1, \ldots, m_i)$. Let $\boldsymbol{\Lambda}_K$ take the place of $\boldsymbol{\Omega}_K$ in the previous section. Therefore,

$$\boldsymbol{\Lambda}_K = \left[|\kappa_k - \kappa_\ell|^3\right]_{1 \le k, k' \le K}. \tag{40}$$

And the other parameters are given by

$$\boldsymbol{\phi}_{ij}' = \begin{bmatrix} 1 & t_{ij} \end{bmatrix} \tag{41}$$

$$\boldsymbol{\xi}_{ij}' = \left[|t_{ij} - \kappa_1|^3 \quad \cdots \quad |t_{ij} - \kappa_K|^3\right] \tag{42}$$

$$\boldsymbol{\psi}_{ij}' = \boldsymbol{\xi}_{ij}'\boldsymbol{\Lambda}_K^{-1/2}. \tag{43}$$

Then the priors are

$$\boldsymbol{\beta}_{k\ell} \sim \mathcal{N}\left(\boldsymbol{0}, \sigma_{\boldsymbol{\beta}_{k\ell}}^2\boldsymbol{I}_2\right) \tag{44}$$

$$\boldsymbol{v}_{k\ell} \sim \mathcal{N}\left(\boldsymbol{0}, \sigma_{\boldsymbol{v}_{k\ell}}^2\boldsymbol{I}_K\right) \tag{45}$$

$$\boldsymbol{w}_{ikm} \sim \mathcal{N}\left(\boldsymbol{0}, \mathrm{diag}\left(\sigma_{\boldsymbol{w}_{km0}}^2, \sigma_{\boldsymbol{w}_{km1}}^2\right)\right) \tag{46}$$

$$\boldsymbol{u}_{ikm} \sim \mathcal{N}\left(\boldsymbol{0}, \sigma_{\boldsymbol{u}_{km}}^2\boldsymbol{I}_K\right) \tag{47}$$

.

4.1 OU process priors

It is very unfamiliar for one to have OU process applied to a statistical model since we have thus far only considered Gaussian error terms. The paper states that the parameters that define an OU process are $\boldsymbol{A}$ and $\boldsymbol{BB}'$.