Technical Supplement for 'Exploring patterns of demand in bike sharing systems via replicated point process models'

Daniel Gervini and Manoj Khanal
Department of Mathematical Sciences
University of Wisconsin–Milwaukee

June 29, 2018

1 Estimation

1.1 The model

In Gervini and Khanal (2018) it is assumed that for each bike station j we have n daily replications of a temporal Poisson process, $X_1, \ldots, X_n$, with corresponding intensity functions λ_i that follow the model

$$\log \lambda_i(t) = \mu(t) + \sum_{k=1}^{p} u_{ik}\phi_k(t), \quad i \in \{1, \ldots, n\}, \tag{1}$$

with $t \in [a, b]$; specifically, $a = 0$ and $b = 24$. For simplicity of notation we omit the station subindex j here, since estimation is done independently for each bike station, but keep in mind that μ, the ϕ_ks and the u_{ik}s will be different for each station j. The principal components $\{\phi_k\}$ are constrained to be orthonormal on $[a, b]$. For generative purposes we can assume the u_{ik}s are uncorrelated (across k) zero-mean random variables, but for estimation purposes we will treat them as fixed effects. For fixed effects, the zero-mean and uncorrelated conditions would translate into the constraints

$$\frac{1}{n}\sum_{i=1}^{n} u_{ik} = 0, \quad k \in \{1, \ldots, p\}, \tag{2}$$

$$\frac{1}{n}\sum_{i=1}^{n} u_{ik}u_{ik'} = 0, \quad k, k' \in \{1, \ldots, p\},\ k' \neq k. \tag{3}$$

Furthermore, the mean and principal components are modeled as spline functions: $\mu(t) = \mathbf{c}_0^T \boldsymbol{\beta}(t)$ and $\phi_k(t) = \mathbf{c}_k^T \boldsymbol{\beta}(t)$, where $\boldsymbol{\beta}(t) = (\beta_1(t), \ldots, \beta_q(t))^T$ is a spline basis; we use B-splines, but other choices are possible. From (1) we can express

$$\log \lambda_i(t) = (\mathbf{c}_0 + \mathbf{C}\mathbf{u}_i)^T \boldsymbol{\beta}(t) \tag{4}$$

where $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_p]$ and $\mathbf{u}_i = (u_{i1}, \ldots, u_{ip})^T$. The orthonormality of the ϕ_ks can be expressed as

$$\mathbf{C}^T \mathbf{J}_0 \mathbf{C} = \mathbf{I}, \tag{5}$$

with $\mathbf{J}_0 = \int_a^b \boldsymbol{\beta}(t)\boldsymbol{\beta}^T(t)dt$. Moreover, it is natural in this application to require periodicity of the intensity functions: $\lambda_i(a) = \lambda_i(b)$. A sufficient condition for this is that $\mu(a) = \mu(b)$ and $\phi_k(a) = \phi_k(b)$ for all k. Then, in terms of the basis coefficients, we have the additional constraints

$$\mathbf{a}^T \mathbf{c}_k = \mathbf{0}, \quad k \in \{0, \ldots, p\}, \tag{6}$$

with $\mathbf{a} = \boldsymbol{\beta}(a) - \boldsymbol{\beta}(b)$.

1.2 The estimator

The parameters $\mathbf{c}_k$s and the fixed effects $\mathbf{u}_i$s are estimated by penalized maximum likelihood, where the penalization terms control the smoothness of μ and the ϕ_ks. Let $x_i = \{t_{i1}, \ldots, t_{im_i}\}$ denote a realization of the process; then the density function is

$$f_i(m_i, t_{i1}, \ldots, t_{im_i}) = \exp\left\{-\int_a^b \lambda_{\mathbf{u}_i}(t)dt\right\} \frac{1}{m_i!} \prod_{l=1}^{m_i} \lambda_{\mathbf{u}_i}(t_{il})$$

with $\lambda_{\mathbf{u}_i}$ denoting a λ that follows model (4). Then the log-likelihood function, ignoring the constant factor $1/m_i!$, is

$$\begin{aligned} \ell &= -\sum_{i=1}^{n} \int_a^b \lambda_{\mathbf{u}_i}(t)dt + \sum_{i=1}^{n} \sum_{l=1}^{m_i} \log \lambda_{\mathbf{u}_i}(t_{il}) \\ &= -\sum_{i=1}^{n} \int_a^b \exp\{(\mathbf{c}_0 + \mathbf{C}\mathbf{u}_i)^T \boldsymbol{\beta}(t)\}dt + \sum_{i=1}^{n} (\mathbf{c}_0 + \mathbf{C}\mathbf{u}_i)^T \sum_{l=1}^{m_i} \boldsymbol{\beta}(t_{il}), \end{aligned} \tag{7}$$

and the penalized log-likelihood is

$$\begin{aligned} P\ell &= \frac{\ell}{n} - \xi_1 \int_a^b \{\mu''(t)\}^2 dt - \xi_2 \sum_{k=1}^{p} \int_a^b \{\phi_k''(t)\}^2 dt \\ &= \frac{\ell}{n} - \xi_1 \mathbf{c}_0^T \boldsymbol{\Omega} \mathbf{c}_0 - \xi_2 \sum_{k=1}^{p} \mathbf{c}_k^T \boldsymbol{\Omega} \mathbf{c}_k, \end{aligned} \tag{8}$$

where $\boldsymbol{\Omega} = \int_a^b \boldsymbol{\beta}''(t) \boldsymbol{\beta}''^T(t)dt$. Note that (7) depends on the t_{il}s only through

$$\mathbf{b}_i := \sum_{l=1}^{m_i} \boldsymbol{\beta}(t_{il}).$$

The estimators of the $\mathbf{c}_k$s and the $\mathbf{u}_i$s are the maximizers of (8) subject to the constraints (2), (3), (5) and (6).

1.3 The algorithm

The algorithm computes the mean and components sequentially, starting with a mean-only model ($\mathbf{c}_0$) and adding the p components one at the time, updating only the last $\mathbf{c}_k$ and the corresponding u_{ik}s at each step. The estimation is done with a Newton–Raphson algorithm, where the updates of the component coefficients $\mathbf{c}_k$ are alternated with the updates of the component scores u_{ik}s. Moreover, the Newton–Raphson directions are projected to keep the constraints (5) and (6), and the u_{ik}s are re-centered and rotated to satisfy (2) and (3). More details are given next.

Updating $\mathbf{c}_0$: From (8) we have

$$\nabla_{\mathbf{c}_0} P\ell = \frac{1}{n}\sum_{i=1}^{n} \nabla_{\mathbf{c}_0} \log f_i(m_i, t_{i1}, \ldots, t_{im_i}) - \xi_1 2\mathbf{\Omega}\mathbf{c}_0.$$

Since

$$\nabla_{\mathbf{c}_0} \log \lambda_{\mathbf{u}_i}(t) = \boldsymbol{\beta}(t),$$

then

$$\nabla_{\mathbf{c}_0} P\ell = -\frac{1}{n}\sum_{i=1}^{n} \int_a^b \boldsymbol{\beta}(t)\lambda_{\mathbf{u}_i}(t)dt + \frac{1}{n}\sum_{i=1}^{n} \mathbf{b}_i - \xi_1 2\mathbf{\Omega}\mathbf{c}_0$$

and

$$\mathsf{H}_{\mathbf{c}_0} P\ell = -\frac{1}{n}\sum_{i=1}^{n} \int_a^b \boldsymbol{\beta}(t)\boldsymbol{\beta}(t)^T\lambda_{\mathbf{u}_i}(t)dt - \xi_1 2\mathbf{\Omega}.$$

Since we are aggregating the components sequentially, at this stage we have a mean-only model $\lambda_{\mathbf{u}_i}(t) = \exp\{\mathbf{c}_0^T\boldsymbol{\beta}(t)\}$, so there is in fact no $\mathbf{u}_i$ yet. Since $\mathbf{c}_0$ is constrained by (6) to satisfy $\mathbf{a}^T\mathbf{c}_0 = \mathbf{0}$, we use projected Newton–Raphson updates:

$$\hat{\mathbf{c}}_0^{\text{new}} = \hat{\mathbf{c}}_0^{\text{old}} - \mathbf{\Gamma}\{\mathbf{\Gamma}^T(\mathsf{H}_{\mathbf{c}_0} P\ell^{\text{old}})\mathbf{\Gamma}\}^{-1}\mathbf{\Gamma}^T(\nabla_{\mathbf{c}_0} P\ell^{\text{old}}),$$

where $\mathbf{\Gamma}$ is a $q \times (q-1)$ matrix whose columns form an orthonormal basis of $\mathbf{a}^{\perp}$.

Updating the $\mathbf{c}_k$s: Since the components are aggregated sequentially, we only update the last component, i.e. $\mathbf{c}_p$, at each stage. So we have

$$\nabla_{\mathbf{c}_p} P\ell = \frac{1}{n} \sum_{i=1}^{n} \nabla_{\mathbf{c}_p} \log f_i(m_i, t_{i1}, \ldots, t_{im_i}) - \xi_2 2 \boldsymbol{\Omega} \mathbf{c}_p.$$

Since

$$\nabla_{\mathbf{c}_p} \log \lambda_{\mathbf{u}_i}(t) = u_{ip} \boldsymbol{\beta}(t),$$

it follows that

$$\nabla_{\mathbf{c}_p} P\ell = -\frac{1}{n} \sum_{i=1}^{n} u_{ip} \int_a^b \boldsymbol{\beta}(t) \lambda_{\mathbf{u}_i}(t) dt + \frac{1}{n} \sum_{i=1}^{n} u_{ip} \mathbf{b}_i - \xi_2 2 \boldsymbol{\Omega} \mathbf{c}_p$$

and

$$\mathsf{H}_{\mathbf{c}_p} P\ell = -\frac{1}{n} \sum_{i=1}^{n} u_{ip}^2 \int_a^b \boldsymbol{\beta}(t) \boldsymbol{\beta}(t)^T \lambda_{\mathbf{u}_i}(t) dt - \xi_2 2 \boldsymbol{\Omega}.$$

Again, to update $\mathbf{c}_p$ we use projected Newton steps due to the constraints (5) and (6). Given that $\hat{\mathbf{c}}_1, \ldots, \hat{\mathbf{c}}_{p-1}$ have already been computed, let $\boldsymbol{\Gamma}$ be a $q \times (q-p)$ orthonormal basis of $\{\mathbf{a}, \mathbf{J}_0 \hat{\mathbf{c}}_1, \ldots, \mathbf{J}_0 \hat{\mathbf{c}}_{p-1}\}^{\perp}$; then

$$\begin{aligned} \tilde{\mathbf{c}}_p^{\text{new}} &= \hat{\mathbf{c}}_p^{\text{old}} - \boldsymbol{\Gamma} \{ \boldsymbol{\Gamma}^T (\mathsf{H}_{\mathbf{c}_p} P\ell^{\text{old}}) \boldsymbol{\Gamma} \}^{-1} \boldsymbol{\Gamma}^T (\nabla_{\mathbf{c}_p} P\ell^{\text{old}}), \\ \hat{\mathbf{c}}_p^{\text{new}} &= \tilde{\mathbf{c}}_p^{\text{new}} / \{ (\tilde{\mathbf{c}}_p^{\text{new}})^T \mathbf{J}_0 \tilde{\mathbf{c}}_p^{\text{new}} \}^{1/2}. \end{aligned} \tag{9}$$

Updating the u_{ik}s: Again, since the components are aggregated sequentially, we only update the component scores of the last component, i.e. the u_{ip}s, at each stage. Then

$$\frac{\partial P\ell}{\partial u_{ip}} = \frac{1}{n} \cdot \frac{\partial}{\partial u_{ip}} \log f_i(m_i, t_{i1}, \ldots, t_{im_i}).$$

Since

$$\frac{\partial}{\partial u_{ip}} \log \lambda_{\mathbf{u}_i}(t) = \phi_p(t),$$

we have

$$\frac{\partial P\ell}{\partial u_{ip}} = -\frac{1}{n} \int_a^b \phi_p(t) \lambda_{\mathbf{u}_i}(t) dt + \frac{1}{n} \mathbf{c}_p^T \mathbf{b}_i$$

and

$$\frac{\partial^2 P\ell}{(\partial u_{ip})^2} = -\frac{1}{n} \int_a^b \phi_p^2(t) \lambda_{\mathbf{u}_i}(t) dt.$$

We update each u_{ip} with a Newton step,

$$\tilde{u}_{ip}^{\text{new}} = \hat{u}_{ip}^{\text{old}} - \frac{\partial P\ell^{\text{old}}}{\partial u_{ip}} \Big/ \frac{\partial^2 P\ell^{\text{old}}}{(\partial u_{ip})^2}.$$

Then we re-center these updates to satisfy (2),

$$\check{u}_{ip}^{\text{new}} = \tilde{u}_{ip}^{\text{new}} - \frac{1}{n}\sum_{i=1}^{n} \tilde{u}_{ip}^{\text{new}},$$

and rotate them to satisfy (3),

$$\begin{bmatrix} \hat{u}_{1p}^{\text{new}} \\ \vdots \\ \hat{u}_{np}^{\text{new}} \end{bmatrix} = \mathbf{\Gamma}\mathbf{\Gamma}^T \begin{bmatrix} \check{u}_{1p}^{\text{new}} \\ \vdots \\ \check{u}_{np}^{\text{new}} \end{bmatrix}$$

where $\mathbf{\Gamma}$ is a $n \times \{n - (p-1)\}$ orthonormal basis of the orthogonal space to the columns of the current component score matrix,

$$\begin{bmatrix} \hat{u}_{11} & \cdots & \hat{u}_{1,p-1} \\ \vdots & \ddots & \vdots \\ \hat{u}_{n1} & \cdots & \hat{u}_{n,p-1} \end{bmatrix}.$$

1.4 Initialization

To initialize the mean we take a constant vector $\mathbf{c}_0 = c\mathbf{1}$ as initial estimator of $\mathbf{c}_0$. Since B-splines satisfy $\boldsymbol{\beta}(t)^T\mathbf{1} = 1$, this is basically like assuming that $\mu(t) \equiv c$ and then (for a mean-only model) that $\lambda_0(t) \equiv e^c$. Then $\int_a^b \lambda_0(t)dt = e^c(b-a)$. Since, for a mean-only model, the m_is are independent identically distributed $\mathscr{P}(\int_a^b \lambda_0)$, we can estimate $e^c(b-a)$ by $\overline{m}$ and then take $\hat{c} = \log(\overline{m}/(b-a))$.

Similarly, we initialize the first component as $\mathbf{c}_1 = \alpha\mathbf{1}$, which implies $\phi_1(t) \equiv \alpha$ and then $\|\phi_1\|^2 = \int_a^b \alpha^2 dt = \alpha^2(b-a) = 1$, so $\alpha = 1/(b-a)^{1/2}$. This is subsequently projected to satisfy (6).

For the other components we simply take $\hat{\mathbf{c}}_k = \tilde{\mathbf{c}}_k/(\tilde{\mathbf{c}}_k^T\mathbf{J}_0\tilde{\mathbf{c}}_k)^{1/2}$ with $\tilde{\mathbf{c}}_k$ the orthogonal projection of $\mathbf{1}$ on the space $\{\mathbf{a}, \mathbf{J}_0\mathbf{c}_1, \ldots, \mathbf{J}_0\mathbf{c}_{k-1}\}^{\perp}$. That is, if $\mathbf{\Gamma}$ is a $q \times \{q-(k-1)\}$ orthonormal basis of $\{\mathbf{J}_0\mathbf{c}_1, \ldots, \mathbf{J}_0\mathbf{c}_{k-1}\}^{\perp}$, then $\tilde{\mathbf{c}}_k = \mathbf{\Gamma}\mathbf{\Gamma}^T\mathbf{1}$.

1.5 Re-scaling of component scores

As explained in the paper, when the true u_{ik}s are random effects, the sample variances of the fixed-effect estimators $\hat{u}_{ik}$s obtained above tend to overestimate the true variances of the u_{ik}s. This problem is ameliorated by re-scaling the $\hat{u}_{ik}$s by a common factor τ, which is obtained by maximum likelihood of the counts m_is. Specifically, since $m_i \sim \mathscr{P}(\int_a^b \lambda_i(t)dt)$, the log-likelihood function is (up to constants)

$$\tilde{\ell}(\tau) = -\sum_{i=1}^{n} I_i(\tau) + \sum_{i=1}^{n} m_i \log I_i(\tau),$$

where $I_i(\tau) = \int_a^b \hat{\lambda}_i^{(\tau)}(t)dt$ and $\hat{\lambda}_i^{(\tau)}(t)$ follows model (1) with $\hat{u}_{ik}$ replaced by $\tau\hat{u}_{ik}$ (and μ and the ϕ_ks by their respective estimators). The $\hat{\tau}$ that maximizes $\tilde{\ell}(\tau)$ is obtained by a simple line search on the interval $[0, 2]$, since $\tau > 0$ and, generally, $\tau < 1$ (since the problem of the $\hat{u}_{ik}$s is inflation of variances).

Acknowledgement

This research was partly supported by US National Science Foundation grant DMS 1505780.

References

Gervini, D., and Khanal, M. (2018). Exploring patterns of demand in bike sharing systems via replicated point process models. *ArXiv* 1802.04755.