Supplement to the Article: 
Modeling Criminal Careers as Departures from a Unimodal Population Age-Crime Curve: The Case of Marijuana Use

May 4, 2012

Summary

This manuscript includes supplementary material associated with the article: “Modeling Criminal Careers as Departures from a Unimodal Population Age-Crime Curve: The Case of Marijuana Use”. In Section 1, we report a Monte Carlo study assessing coverage and estimation performance of the proposed method. In Section 2 we discuss unimodal smoothing in more details. Section 3 summarizes full conditional distributions used to develop the MCMC algorithm proposed in the main manuscript.
Figure 1: Simulation Study. Panel (a): Sample simulated counts. Panel (b): Sample time transformation functions. Panel (c) Simulated shape function (black line), posterior distribution (shaded area) and posterior mean shape function (red line). Panel (d): Aligned simulated counts.

1 Simulation Study

To assess our model we simulate 150 longitudinal count trajectories (see Figure 1, panel (a)), over 15 equally spaced time point between 0 and 1. Each simulated trajectory is of the form \( y_{ij} \sim \text{Poisson}\{\lambda_i(t_j)\} \) with intensity \( \lambda_i(t_j) = a_i(X_i)B\{\mu_i(t_j; X_i)\} \). For each trajectory we introduce a random pattern of missingness erasing between 3 and 10 observations. The
Figure 2: Monte Carlo Study. Panel (a): 90% posterior intervals associated with the cv parameter $b_0$ and amplitude regression parameters $b_{a}$. Panel (b): 90% posterior intervals associated with phase regression parameters $b_{\phi}$. In red we indicate the simulation truth.

covariate matrix $X$ defines a simple randomly balanced comparison experiment with two binary predictors. Population amplitude effects are coded, so that $E(a_i | b_a) = \exp\{X_i^\prime b_a\}$, with $b_a = (0, 1.5, -0.5)$ and population phase shifts are coded as $b_{\phi} = c(0, -0.05, 0.1)$.

The common shape function $B(t)$ is reported in Figure 1 (panel (c)) and was generated as a linear combination of (order 4) B-Spline basis functions, with 3 equally spaced interior knots spanning the extended time interval $[-0.2, 1.2]$. Finally, the time transformation functions $\mu_i(t_j)$ (Figure 1, panel (b)) were generated from a linear combination of 4 B–spline basis (order 4) over the sampling interval $[0, 1]$.

We fitted our Hierarchical Poisson Registration model to 200 synthetic data sets. We tested robustness to model mispecification overparametrizing both shape $B(\cdot)$ (7 equally spaced interior knots) and warping functions $\mu_i(\cdot)$ (3 interior knots). We placed relatively diffuse $\text{Gamma}(0.001, \text{rate} = 1000)$ prior on the precision parameter for the shape coefficients $\beta$ and $\text{Gamma}(0.1, 1)$ prior for the precision of the time transformation coefficients $\Phi$.

\footnote{See the main manuscript for a detailed description of the probability model.} For each data set, our inference is based on 5,000 samples (thinned by a factor of 2) from
the posterior distribution obtained after discarding the initial 10,000 MCMC iterations for burn-in. 2

Results from our Monte Carlo study are reported in Fig. 2. Here we represent 90% posterior intervals for the amplitude ($b_0, b_a$) and phase ($b_\phi$) regression parameters associated with each of 200 synthetic data sets. For all parameters observed coverage approaches reasonably well the nominal level of 90% with the lowest coverage of 82% associated with phase regression. In this case, shrinkage towards the prior mean of 0 phase effect is more pronounced due to sparseness. Coverage close to the nominal level (92%) was also observed for the functional estimate of the common shape function.

We illustrate the performance of the proposed model in more detail for a sample data set from our Monte Carlo study. Figure 1 (panel c) reports the posterior distribution (shaded area), simulation truth (black line) and posterior mean (red line) associated with the common shape function. Panel (d) in the same figure reports the aligned simulated counts using posterior mean estimate of the time transformation functions. We zoom into the estimation performance at the individual trajectory level in Fig. 3. Here we relate the posterior predictive distribution (shaded area) to the simulated counts (red dots), as well as the true (black lines) and posterior mean fit (red line). We observed good coverage of the simulated counts as well as containment of the true mean trajectory. In Fig. 4 we repeat the exercise for a sample of individual time transformation functions. As for the population parameters, for some transformation functions we report more pronounced shrinkage towards the prior mean (identity transformation) due to sparseness in the simulated counts.

---

2We base our inference on a relatively small number of MC samples in order to feasibly carry out a Monte Carlo study. In the analysis of DYS marijuana counts, we opt for a more conservative MCMC simulation size.
Figure 3: Posterior Predictive Performance: For each plot we represent in blue the posterior predictive distribution of counts over time. Red dots indicate the simulated data, the black line indicates the true trajectory and the red line indicates the estimated posterior mean trajectory.
Figure 4: Posterior Time Transformation: For each plot we represent in blue the posterior distribution of the time transformation functions. Black line indicates the true simulated time transformation, while red lines indicate the estimated posterior mean time transformation.
2 Unimodal Smoothing

Let $S(t, \gamma_K, r)$ be a set of B-spline bases of order $r$, evaluated at $t \in [0, T]$, with $K$ distinct interior knots indexed by $\gamma_K = (\gamma_1, \gamma_2, \ldots, \gamma_{K+2r})' = (0, \ldots, 0, \gamma_{r+1}, \ldots, \gamma_{r+K}, T, \ldots, T)'$. We represent a smooth function of time as $S(t, \beta) = S(t, \gamma_K, r)' \beta$, where $\beta = (\beta_1, \beta_2, \ldots, \beta_{K+r})'$ is a $(K+r)$ dimensional set of spline coefficients.

**Theorem.** Let $\beta_k = \nu^* - (\nu_k - \nu^*)^2$, for $k = 1, \ldots, K + r$. For any scalar $\nu^* \geq 0$, if $0 \leq \nu_1 \leq \nu_2 \leq \cdots \leq \nu_{K+r} \leq 2\nu^*$, then the function $S(t, \beta) = S(t, \gamma_K, r)' \beta$ is positive and unimodal.

**Proof.** It is easy to show that $\beta_k \geq 0$ iff $0 \leq \nu_k \leq 2\nu^*$, $\forall k = 1, \ldots, K + r$, which implies positivity of $S(t, \beta)$. The first derivative with respect to $t$ of $S(t, \beta)$ can be expressed as $\partial S(t, \beta)/\partial t = (r-1)S(t, \tilde{\gamma}_K, (r-1)')\tilde{\beta}$, where $\tilde{\gamma}_K = (\gamma_2, \ldots, \gamma_{K+2r-1})'$ and $\tilde{\beta} = (\beta_1, \ldots, \beta_{K+r-1})t$, with $\tilde{\beta}_j = (\beta_{j+1} - \beta_j)/(\tilde{\gamma}_{j+r-1} - \tilde{\gamma}_j)$, for $j = 1, \ldots, K + r - 1$. It is easy to verify that $\tilde{\beta}_j \geq 0$ for all $j$ s.t. $(\nu_j - \nu^*)^2 \geq (\nu_{j+1} - \nu^*)^2$. Conversely, $\tilde{\beta}_j \leq 0$ for all $j$ s.t. $(\nu_j - \nu^*)^2 \leq (\nu_{j+1} - \nu^*)^2$. Because $\nu_{j+1} \geq \nu_j$, for all $j$, this implies that $\partial S(t, \beta)/\partial t$ exhibits only one sign change from positive to negative in $[0, T]$, providing unimodality of $S(t, \beta)$.

3 Web Appendix: Full Conditional Distributions

**Shape smoothing parameters $\sigma_{\beta}^2$:** For ease of notation we define the precision of the $j^{th}$ fixed effect $h_\beta = 1/\sigma_{\beta}^2$. We have that $P(h_\beta | a_\beta, b_\beta) \propto h_\beta^{-(a_\beta-1)}e^{\{b_\beta h_\beta\}}$. Given the fixed effect coefficients $\nu$, the conditional posterior density of $h_\beta$ can be written as $P(h_\beta | Y, \theta_{\setminus h_\beta}) \propto P(\nu | h_\beta)P(h_\beta | a_\beta, b_\beta)$, where $P(\nu | h_\beta) \propto h_\beta^{K/2}e^{\{-h_\beta/2\nu'\Omega\nu\}}$ and $\Omega$ is the banded concentration structure arising from a second order random walk. It is trivial
to show that: $P(h_\beta \mid Y, \theta_{\cdot h_\beta}) \propto h_\beta^{(a_\beta + 1)} \exp\{-b_\beta h_\beta\}$, corresponding to the density function of a Gamma random variable with shape $a_\beta = a_\beta + K/2$, and rate $b_\beta = b_\beta + \nu' \Omega \nu / 2$.

**Time transformation smoothing parameter $\sigma^2_\phi$:** For ease of notation we define the time transformation precision $h_\phi = 1/\sigma^2_\phi$. We have $P(h_\phi \mid a_\phi, b_\phi) \propto h_\phi^{a_\phi - 1} \exp\{-b_\phi h_\phi\}$. Given the matrix of time transformation coefficients $\phi$, the conditional posterior density of $h_\phi$ can be written as $P(h_\phi \mid Y, \theta_{\cdot h_\phi}) \propto P(\phi \mid \gamma, h_\phi) P(h_\phi \mid a_\phi, b_\phi)$, where $P(\phi \mid \gamma, h_\phi) \propto \prod_{i=1}^{N} h_\phi^{Q/2} \exp\{-h_\phi/2(\phi_i - \gamma_i)'\Xi(\phi_i - \gamma_i)\}$ and $\Xi$ is a banded concentration structure arising from a first order random walk, given $(\phi_i - \gamma_i) = 0$ for all $i = 1, ..., N$. It easily follows that $P(h_\phi \mid Y, \theta_{\cdot h_\phi}) \propto h_\phi^{a_\phi^* - 1} \exp\{-b_\phi^* h_\phi\}$, a Gamma random variable with shape $a_\phi^* = a_\phi + N \times Q/2$, and rate $b_\phi^* = b_\phi + 1/2 \times \sum_{i=1}^{N} (\phi_i - \gamma_i)'\Xi(\phi_i - \gamma_i)$.

**Phase Regression parameters $b_\phi$:** Using standard conjugate analysis it is easy to show that $P(b_\phi \mid \phi) = d N_p(m^*, V^*)$, where $V^* = (\sum_{i=1}^{N} \sum_{q=1}^{Q} X_i(X_i + \Omega)^{-1})$ and $m^* = V^*(\sum_i \sum_q X_i \tilde{\phi}_{iq})$. Where $\tilde{\phi}_{iq}$ is the $q$-th element of $(\phi_i - \Upsilon)(\Xi)^{1/2}$. and $\Xi$ is defined as before.

**Amplitude-Phase Covariance $\Sigma_b$:** using standard conjugate analysis it is easy to show that the conditional posterior distribution $P(\Sigma_b | b_a, b_\phi, v_0, \Phi_0) = d IW(p + v_0, (S_b + \Phi_0))$, where $S_b = \sum_{k=1}^{p}(b_{ak}, b_{\phi k})(b_{ak}, b_{\phi k})'$. 

8