

Stochastic Version of EM Algorithm for Nonlinear Random Change-Point Models

Hongbin Zhang^{1,2}, Binod Manandhar²

¹Department of Epidemiology and Biostatistics, Graduate School of Public Health and Health Policy
City University of New York
55 West 125th Street, New York, United States
hongbin.zhang@sph.cuny.edu

²Institute of Implementation Science for Population Health
City University of New York
55 West 125th Street, New York, United States

Abstract - Random effect change-point models are commonly used to infer individual-specific time of event that induces trend change of longitudinal data. Linear models are often employed before and after the change point. However, in applications such as HIV studies, a mechanistic nonlinear model can be derived for the process based on the underlying data-generation mechanisms and such nonlinear model may provide better "predictions". In this article, we propose a random change-point model in which we model the longitudinal data by segmented nonlinear mixed effect models. Inference wise, we propose a maximum likelihood solution where we use the Stochastic Expectation-Maximization (StEM) algorithm coupled with independent multivariate rejection sampling through Gibbs's sampler. We evaluate the method with simulations to gain insights.

Keywords: Random change-point model, Nonlinear mixed effects model, Stochastic version of EM, Gibbs's sampler, Multivariate rejection sampling

1. Introduction

In human immunodeficiency virus (HIV) care, the time from diagnosis to antiretroviral therapy (ART) initiation is a critical determinant of success and is not uniform across jurisdictions and patients due to various barriers and other factors. While time-updated laboratory testing data on viral loads (RNA copies per millilitre of blood plasma) are reported in routine HIV surveillance activities, information on ART is not collected [1]. We consider a statistical modelling method, e.g., using change-point model to estimate and infer ART initiation time. Such change-point model has biological interpretation where the ART initiation induces trajectory change of the underlying process of HIV dynamics.

Random change-point models which allow for individual-specific changes for longitudinal outcomes have been widely used in medical research [2]. A general way to define a class of random change-point models is to assume a linear mixed effects model for the longitudinal data before and after the change point, i.e., segmented linear mixed effects model (see, e.g., [3], [4], [5]). In some applications such as the motivation example above, however, the longitudinal data may be missing or censored due to detection limits. In this case, the assumed linear model based on the observed longitudinal data may be inappropriate for the (unobserved) data. If, on the other hand, a mechanical or scientific model is available for the longitudinal data, such mechanical model can be used to better "predict" the unobserved data, leading to better estimate of the change points. Such a mechanistic model is often nonlinear. In this article, we consider random change-point models where we model the longitudinal data by nonlinear mixed effects (NLME) models.

The aim of this paper is to propose and evaluate a segmented-NLME model for longitudinal data. we consider a full likelihood-based inference via the implementation of the stochastic version of the EM algorithm, a.k.a. StEM algorithm, proposed by Diebolt and Celeus [6]. The StEM algorithm has been proved to be computationally efficient algorithm as only one realization of the missing data is required for each iteration [7]. The major challenge of accounting random change point in likelihood framework is regarding the inference (see, e.g., [8]). In the case of StEM, it is further complicated by the issue of convergence [9].

In Section 2, we describe the HIV dynamics, the nonlinear random change-point model in general form, and the estimation and inference procedure. We evaluate the method with a designated simulation study in Section 3, focusing on the assessment of a Geweke-based convergence criteria. We conclude the article with some discussions in Section 4.

2. Nonlinear Random Change-point Model and Estimation Procedure

2.1. HIV Dynamics, Notations and the Models

Typically, viral load shows a dramatic fluctuation after HIV infection before reaching to a set-point and will then increase with a steady rate until the development of AIDS if without treatment [10]. ART initiation, however, will induces substantial reductions in HIV RNA. There is extensive research for HIV dynamics after ART, essentially associated with the purpose of assessing antiretroviral drug's therapeutic effect. Based on biological and clinical arguments, Wu and Ding [11] proposed to approximate viral load data pattern by the virological model $V(t) = P_1 + P_2 e^{-\lambda t}$ where $V(t)$ is the total virus at time t and P_1 and P_2 are baseline values for each phase. Parameters λ is the viral decay rates and may be interpreted as the turnover rates of productively infected cells and long-lived and/or latently infected cells if the therapy is perfect.

Motivated by the problem of estimating ART initiation time after HIV diagnosis [1], we may consider the following nonlinear random effects change-point model for the reported viral load for individual i at time t_{ij} after diagnosis in a typical HIV surveillance data

$$\log_{10}(V(t_{ij})) = a_{1i}(t_{ij} - \tau_i)^- + \log_{10}(b_{1i} + b_{2i}e^{-b_{3i}(t_{ij}-\tau_i)^+}) + \epsilon_{ij},$$

where the \log_{10} transformation is used to stabilize the variance and makes the data more normally distributed. The error term is represented by ϵ_{ij} , τ_i is the change point, functions x^- and x^+ correspond to $\min(x, 0)$ and $\max(x, 0)$, respectively. The quantity a_{1i} is the subject-specific regression coefficient representing the slope of viral load before change point, and b_{1i} , b_{2i} , and b_{3i} are subject-specific mixed effects for the viral load trajectory after the change point. We may define

$$a_{1i} = \alpha_1 + \alpha_{1i}, \quad b_{1i} = \beta_1 + \beta_{1i}, \quad b_{2i} = \beta_2 + \beta_{2i}, \quad b_{3i} = \beta_3 + \beta_{3i},$$

where $\alpha_1, \beta_1, \beta_2$, and β_3 are the population parameters (fixed effects), $\alpha_{1i}, \beta_{1i}, \beta_{2i}$, and β_{3i} are random effects which are usually assumed to follow normal distribution with zero mean. The random change points can be conventionally assumed to follow a normal distribution. A log-normal distribution can also be assumed to ensure positive value of the change point be estimated.

In what follows, we present the models and the methods in general forms, illustrating that our methods may be applicable in other applications. Let y_{ij} be the measurement (can be left-censored) $j = 1, 2, \dots, n_i$ for subject $i = 1, 2, \dots, n$ and let t_{ij} be the time when y_{ij} is measured. We consider a general segmented-NLME model

$$y_{ij} = g((t_{ij} - r(\tau_i))-, a_i) + h((t_{ij} - r(\tau_i))^+, b_i) + \epsilon_{ij}, \quad (2.1)$$

with

$$\tau_i \sim N(\boldsymbol{\tau}, \boldsymbol{\sigma}_\tau^2), \quad \mathbf{a}_i \sim N(\boldsymbol{\alpha}, \mathbf{A}), \quad \mathbf{b}_i \sim N(\boldsymbol{\beta}, \mathbf{B}), \quad \boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \boldsymbol{\sigma}_\epsilon^2),$$

where $g()$ and $h()$ are nonlinear functions, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are vectors of population parameters, $\boldsymbol{\tau}$ is the population mean and $\boldsymbol{\sigma}_\tau^2$ is the variance for the random change point τ_i , \mathbf{A} and \mathbf{B} are the variance-covariance matrix for random effects \mathbf{a}_i and \mathbf{b}_i , respectively, and $\boldsymbol{\sigma}_\epsilon^2$ is the within-individual variance. The $r()$ function is introduced to allow possible parameterization around the change point. For the segmented-NLME model, it is reasonable to assume that $\mathbf{a}_i, \mathbf{b}_i$ are independent and both are independent of τ_i which is usually introduced by external force.

We consider a likelihood inference procedure for the model (2.1) based on the observed data $y_i = (y_{i1}, \dots, y_{in_i})$. Let $\theta = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\sigma}_\tau^2, \mathbf{A}, \mathbf{B})$ be the collection of all unknown parameters and $f()$ be a generic density function, and let $f(X|Y)$ denote a conditional density of X given Y . The observed data likelihood can be written as

$$L(\theta) = \prod_{i=1}^n \left\{ \iiint \left[\prod_{j=1}^{n_i} f(y_{ij} | \tau_i, a_i, b_i; \theta) \right] * f(\tau_i) f(a_i) f(b_i) d\tau_i da_i db_i \right\}. \quad (2.2)$$

Directly maximizing the likelihood (2.2) is challenging due to the nonlinear models involved and the nested integrals. By treating the unobserved random effects, τ_i, a_i, b_i as "missing data", we have "complete data" $\{(y_i, \tau_i, a_i, b_i), i = 1, \dots, n\}$ and complete-data log-likelihood function for individual i can be expressed as

$$l_c(\theta) = \log f(\tau_i; \theta) + \log f(a_i; A) + \log f(b_i; B) + \log f(y_i | \tau_i, a_i, b_i; \theta). \quad (2.3)$$

2.2. The Estimation Procedure

The EM algorithm introduced by Dempster et al. [12] is a classical approach to estimate parameters of models with non-observed or incomplete data. Let us briefly cover the EM principle. Denote z be the vector of non-observed data and the complete data of the model is (y, z) . The EM algorithm maximizes the $Q(\theta | \theta') = E(L_c(y, z; \theta) | y; \theta')$ function in two steps, where the $L_c(y, z; \theta)$ is the log-likelihood of the complete data. At the k^{th} iteration, the E-step is the evaluation of $\theta^{(k)}(\theta) = Q(\theta | \theta^{(k-1)})$, where the M-step updates $\theta^{(k-1)}$ by maximizing $\theta^{(k)}(\theta)$.

For the cases in which the E step has no analytic form, Wei and Tanner [13] introduce Monte Carlo EM (MCEM) which calculates the conditional expectations at the E step via large number of simulations with each iteration and hence is quite computationally intensive. Diebolt and Celeus [6] introduce stochastic versions of the EM algorithm, namely the stochastic EM (StEM) which replaces the E-step by a single imputation of the complete data, and then averages the last batch of M

estimates in the resulting Markov Chain iterative sequence to obtain a point estimate of the parameter. The imputed data $z^{(k)}$ at the k^{th} iteration are a random draw from the conditional distribution of the missing data given the observed data and the estimated parameter values at the $(k-1)^{th}$ iteration, $f(z^{(k)} | y, \theta^{(k-1)})$. As $z^{(k)}$ only depends on $z^{(k-1)}$, $\{z^{(k)}\}_{k \geq 1}$ is a Markov chain. Assuming that $z^{(k)}$ take values in a compact space and the kernel of the Markov chain is positive continuous with respect to a Lebesgue measure, the Markov chain is ergodic and that ensures the existence of a unique stationary distribution (see, e.g., [14], [15]). We now detail the StEM algorithm for the segmented-NLME previously presented. At each iteration $k+1$:

Imputation: Draw missing data (τ_i, a_i, b_i) from the conditional distribution $[\tau_i, a_i, b_i | y_i; \theta^{(k)}]$. Specifically, we use the Gibbs sampler to generate samples from $[\tau_i, a_i, b_i | y_i; \theta^{(k)}]$ by iteratively sampling from the full conditionals $[\tau_i | y_i, a_i, b_i; \theta^{(k)}]$, $[a_i | y_i, \tau_i, b_i; \theta^{(k)}]$, and $[b_i | y_i, \tau_i, a_i; \theta^{(k)}]$ as follows:

$$\begin{aligned} f(\tau_i | y_i, a_i, b_i; \theta^{(k)}) &\propto f(\tau_i; \theta^{(k)}) f(y_i | \tau_i, a_i, b_i; \theta^{(k)}), \\ f(a_i | y_i, \tau_i, b_i; \theta^{(k)}) &\propto f(a_i; \theta^{(k)}) f(y_i | \tau_i, a_i, b_i; \theta^{(k)}), \\ f(b_i | y_i, \tau_i, a_i; \theta^{(k)}) &\propto f(b_i; \theta^{(k)}) f(y_i | \tau_i, a_i, b_i; \theta^{(k)}). \end{aligned}$$

Monte Carlo samples from each of the full conditionals can be generated using multivariate rejection sampling methods [16].

Maximization: After data augmentation, the maximization step involves maximizing the complete likelihood from (2.3). For the "complete data" $\{(y_i, \tau_i, a_i, b_i), i = 1, \dots, n\}$, the complete log-likelihood no longer involves integrals, which substantially simplifies the maximization. Also, due to the mutual independence among τ_i, a_i, b_i the maximization can be done by part. Solving the score equations yields the following equations:

$$\begin{aligned} \tau &= \frac{1}{n} \sum_{i=1}^n r(\tau_i), \quad \sigma_\tau^2 = \frac{1}{n} \sum_{i=1}^n (r(\tau_i) - \tau)^2, \\ \alpha &= \frac{1}{n} \sum_{i=1}^n a_i, \quad A = \frac{1}{n} \sum_{i=1}^n (a_i - \alpha)(a_i - \alpha)^T, \end{aligned}$$

$$\beta = \frac{1}{n} \sum_{i=1}^n b_i, \quad B = \frac{1}{n} \sum_{i=1}^n (b_i - \beta)(b_i - \beta)^T,$$

$$\sigma_\epsilon^2 = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{1}{n_i} \sum_{j=1}^{n_i} \left[y_{ij} - g\left((t_{ij} - r(\tau_i))^- , a_i\right) - h\left((t_{ij} - r(\tau_i))^+ , b_i\right) \right]^2 \right\},$$

Muggeo et al. [8] discussed the inference issue for random change-point models and proposed a bootstrap-based approach, which is computationally intensive. In this paper, we propose to compute the Fisher information matrix by linearizing the $g() + h()$ function around the conditional expectation of the individual Gaussian parameters, $\{(\tau_i, a_i, b_i), i = 1, \dots, n\}$, extending the approach by Retout et al. [17].

For the convergence of a StEM algorithm, in literature, the commonly used approach is through visually examination of the trace plots (see, e.g., [18], [19], [20]). Recently, Zhang et al. [9] proposed a Geweke Statistics based method. Stationarity is claimed to achieve when the Geweke statistics is smaller than a designated threshold.

3. A Simulation Study

In this simulation study, the objectives are two folds: one is to evaluate the properties of the proposed method and the other is to assess the feasibility of convergence criteria. We first emulate the viral load reporting time, which is often irregular, in a typical jurisdiction by a progressive state-transition model with first order Markov assumption where the length of lagging for a reporting time depends on the previous reporting time. Specifically, the viral load reporting time $T_{|u}$ is assumed to follow an exponential distribution with parameter $\xi > 0$. Given the previous reporting time $u > 0$, the next reporting time, conditional on u , is $T_{|u} = \exp\left(\log\left(-\frac{1}{\xi} \log(X) + u\right)\right)$, for $X \sim \text{Uniform}(0,1)$.

We then generate viral loads based on the following model

$$y_{ij} = e^{a_i}(t_{ij} - e^{\tau_i})^- + \log_{10}\left(e^{b_{1i}-b_{2i}(t_{ij}-e^{\tau_i})^+} + e^{b_{3i}-b_{4i}(t_{ij}-e^{\tau_i})^+}\right) + \epsilon_{ij}.$$

The $g() = e^{a_i}(t_{ij} - e^{\tau_i})^-$ is a function for the viral load before change point where we parameterize a nonlinear form to ensure a positive slope and positive change point. The $h() = \log_{10}\left(e^{b_{1i}-b_{2i}(t_{ij}-e^{\tau_i})^+} + e^{b_{3i}-b_{4i}(t_{ij}-e^{\tau_i})^+}\right)$ function is a simplified analytical solution of a differential system describing HIV viral load decrease during antiretroviral treatment proposed by Perelson et al. [21]. The model has six individual random effects: $a_i, \tau_i, b_{1i}, b_{2i}, b_{3i}, b_{4i}$ which are assumed to be additive Gaussian with population mean (fixed) parameters as $\alpha, \tau, \beta_1, \beta_2, \beta_3, \beta_4$ and diagonal covariance matrix. The additive Gaussian error is assumed with a constant variance σ_ϵ^2 . For the fixed effects, the values are those proposed by Mei et al. [10] and Wu and Ding [11]: $(\alpha, \tau, \beta_1, \beta_2, \beta_3, \beta_4) = (-1.97, -1.23, 12, 8, 0.5, 0.05)$. The inter-subject variability is identical for the six parameters and set to be 0.3 ($=\sigma_\alpha^2 = \sigma_\tau^2 = \sigma_{\beta_1}^2 = \sigma_{\beta_2}^2 = \sigma_{\beta_3}^2 = \sigma_{\beta_4}^2$), corresponding to a variation coefficient of 55%, which is a realistic inter-subject variability in the context of HIV dynamics. We chose a variance $\sigma_\epsilon^2 = 0.065$, which corresponds to a constant variation coefficient of 15% for the viral load.

For each scenario (see below), we generate 100 datasets. For each dataset, to obtain the initial values of the model parameters, we start with a modified version of the *logIplus* algorithm in [1], denoted as *logIplus**. The *logIplus** algorithm detects ART occurrence by empirically examining pairs of reported viral loads for an individual to see if viral load drops more than 1 log₁₀ based unit within certain observing window. We apply this algorithm and use the first reporting time of the detected pair as the initial change point. Initial τ and σ_τ^2 are the mean and variance across the sample. To obtain initial values for other parameters, e.g., those in $g()$ and $h()$ functions, we fit nonlinear mixed effects models using the popular iterative linearization method of [22], separately, on the viral loads segmented by the initial change points.

Once started the Markov chain with the initial values, the stationarity is determined by using a batch procedure based on the Geweke statistic [23]. Let batch size be M . We use a moving window for the Markov chain and compute the Geweke statistics at each increment of $\$w\$$ iterations. More precisely, we have the following batch procedure:

1. **Initialization.** Set $B=0$. Run the StEM algorithm to obtain the initial series of the estimates $\{\theta^{(w*B+1)}, \dots, \theta^{(w*B+M)}\}$.
2. **Check stationarity.** For each entry p in θ , we compute the Geweke statistic z_p based on the Markov chain $\{\theta_p^{(w*B+1)}, \dots, \theta_p^{(w*B+M)}\}$. based on the mean difference between first 10% and last 50% part of chain. We regard stationary being reached when all $|z_p|$ s are sufficiently smaller, i.e.

$$\sum_{p=1}^P z_p^2 < \varepsilon P$$

where P is the total number of parameter and we set $\varepsilon=1.5$ as in [9].

3. **Update.** If stationarity is not reached, execute w additional burn-in runs of the chain, increase the number B by 1 and then repeat step 2.

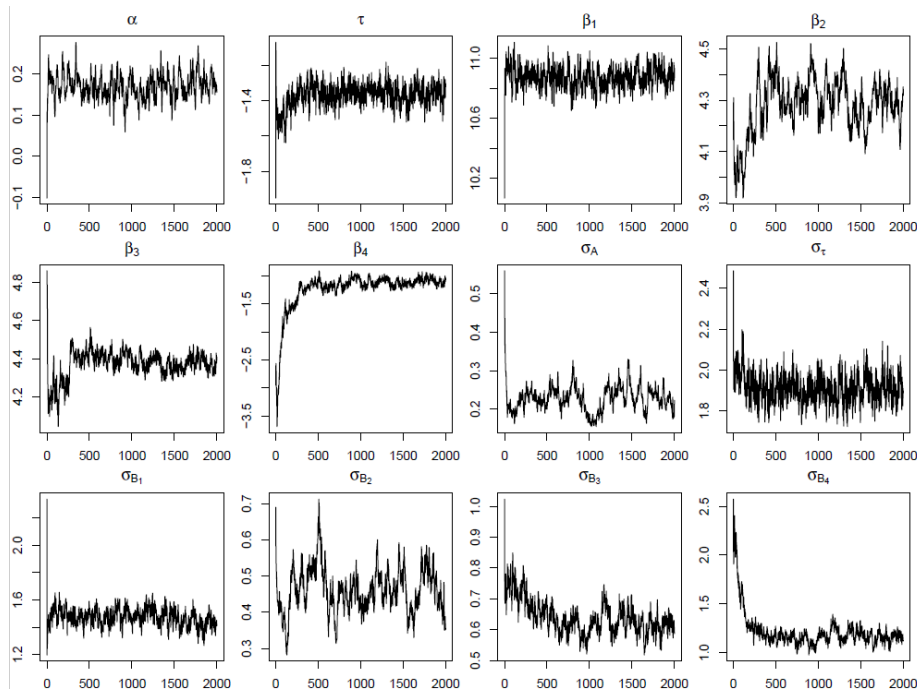


Fig. 1: Trace plots of all parameters in the segmented-NLME model for one typical simulation data. Results from 2000 iterations are displayed although for this sample, the convergence is arrived at the 1840th iteration.

The simulation factors include n , the sample size; M , the batch size; and w , the burn-in incremental. We simulated with $n=300, 500$ and 1000 , $M=50, 100, 200$ and 300 , $w=10, 20, 30$. Figure 1 shows a typical trace plot. Table 1 presents the selected simulation results ($n=500$ or 1000 , $M=300$, $w=10$) for the fixed effects. Besides the averaged standard error (SE), sample based standard deviation (SD), we also calculate the mean squared error (MSE) and the percentage of biases (Bias%) by comparing the estimate of the parameter with the true value. In all our simulations, the StEM based method outperforms the *log1plus** approach which is understandable as the latter is essentially an empirical method. Also, we see the SEs and SDs are agreeable, indicating the linearization-based inference method works well for the finite sample. There are no difference in term of burn-in size but batch size does influence the convergence where smaller batch size like $B=50, 100$ and 200 do not seem to capture the distribution. Sample size wise, $n = 300$ does not seem to capture the heterogeneity of the model while there is slightly difference when $n=500$ and $n=1000$. We recommend the latter if computing power is not the issue.

Table 1: Simulation results for B=300, m=10 and n=1000/500.

	θ	α	β_1	β_2	β_3	β_4	τ
Method	True	-1.97	12.00	8.00	0.50	0.05	-1.23
n=1000							
log1plus*	Est	-2.36	10.22	5.81	1.56	-1.77	-0.98
	SE	0.16	2.45	1.49	0.15	0.12	0.18
	SD	0.14	2.46	1.17	0.21	0.19	0.18
	MSE	0.26	10.35	4.07	0.68	0.75	0.42
	Bias%	58.57	-30.69	-37.30	18.06	39.54	-66.95
StEM	Est	-2.11	11.56	7.77	0.56	0.12	-1.15
	SE	0.13	2.44	1.33	0.17	0.14	0.17
	SD	0.13	2.48	1.45	0.12	0.33	0.14
	MSE	0.11	1.12	1.41	0.12	0.14	0.15
	Bias%	19.57	-2.59	12.85	-1.87	11.56	-6.89
n=500							
log1plus*	Est	-2.45	11.08	6.98	1.47	-1.79	-0.99
	SE	0.21	2.60	2.15	0.27	0.32	0.21
	SD	0.24	2.76	2.14	0.45	0.39	0.26
	MSE	0.30	11.88	2.95	0.92	0.74	0.44
	Bias%	77.45	-31.97	-23.88	10.12	27.73	-65.44
StEM	Est	-2.22	11.57	8.06	0.64	0.15	-1.10
	SE	0.19	2.69	2.10	0.21	0.17	0.21
	SD	0.22	2.98	2.22	0.25	0.23	0.29
	MSE	0.24	1.37	1.44	0.15	0.16	0.28
	Bias%	22.23	-3.11	10.98	-1.88	9.17	-7.87
$\text{MSE} = (\hat{\theta}^{(s)} - \theta^{true})^2 + SE(\hat{\theta}^{(s)}), \text{ Bias\%} = \frac{1}{S} \sum_{s=1}^S 100 \times (\hat{\theta}^{(s)} - \theta^{true}) / \theta^{true}$							

4. Conclusion and Discussion

Nonlinear models have wide range of applications in HIV studies, cancer research, and pharmacokinetic-pharmacodynamic modelling [24]. In this paper, we extend the random change-point model to a more general class to allow for nonlinear mixed effects models for each segment. In addition, we establish a StEM algorithm-based solution and evaluate a convergence criterion of the Markov chain through simulations. The proposed method has conceptual simplicity, attributing to the EM algorithm.

StEM substantially improves the computation efficiency over the classic MCEM method for the problem when there is no existing analytic form for the E-step. Using high quality sample obtained from independent rejection sampling, we demonstrate that convergence is usually arrive within 2000 iterations for the size our problem. Due to the high dimension of the missing- data structure, a Gibbs sampler must be embedded within each StEM iteration. Even though computer power has been increasing at a tremendous rate, it is still wise to keep the number of simulations manageable. Hence, we shall continue in research in search of even more computationally efficient methods. One feasible solution is to use crude approximations, e.g., Metropolis sampling, at the burn-in period and gradually increase the accuracy of the approximation to the proper distribution.

Acknowledgements

The work is supported by NIH grant R21AI147933. It is also partially supported by the City University of New York High-Performance Computing Center, College of Staten Island, funded in part by the City and State of New York, City University of New York Research Foundation and National Science Foundation grants CNS-0958379, CNS-0855217, and ACI-112611.

References

- [1] SB Braunstein and MR Robertson and J Myers and B Abraham and D Nash. Increase in CD4+ T-Cell Count at the Time of HIV Diagnosis and Antiretroviral Treatment Initiation Among Persons With HIV in New York City. *Journal of Infection Disease*, 2016, vol. 214, pp. 1682-1686.
- [2] A Dominicus and S Ripatti and NL Pedersen and J Palmgren. A random change point model for assessing variability in repeated measures of cognitive function. *Statistics in Medicine*. 2008. 27(27): 5786-5798.
- [3] D Rudoy and SG Yuen and RD Howe and PJ Wolfe. Bayesian change-point analysis for atomic force microscopy and soft material indentation. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 2010. 59:573-593.
- [4] A Moss and E Juarez-Colunga and F Nathoo and B Wagner and S Sagel. A comparison of change point models with application to longitudinal lung function measurements in children with cystic fibrosis. *Statistics in Medicine*. 2016. 35:2058-2073.
- [5] O.D. Buhule and H. Choo-Wosoba and P.S. Albert. Modeling repeated labor curves in consecutive pregnancies: individualized prediction of labor progression from previous pregnancy data. *Statistics in Medicine*. 2020. 39(8):1068-1083.
- [6] J. Diebolt and G. Celeus. Asymptotic properties of a stochastic EM algorithm for estimating mixture proportions. *Stochastic Models*. 1996. 9:599-613.
- [7] E.H. IP. On single versus multiple imputation for a class of stochastic algorithm estimating maximum likelihood. *Computational Statistics*. 2002. 17:517-524.
- [8] MR Muggeo and CA David and JG Robert and D Sona. Segmented mixed models with random changepoints: a maximum likelihood approach with application to treatment for depression study. *Statistical Modelling*. 2014. 14(4):293-313.
- [9] S. Zhang and Y. Chen and Y. Liu. An improved stochastic EM algorithm for large-scale full-information item factor analysis. *British Journal of Mathematical and Statistical Psychology*. 2020. 73(1):44-71.
- [10] Y Mei and L Wang and SE Holte. A comparison of methods for determining HIV viral set point. *Statistics in Medicine*. 2008. 27:121-139.
- [11] H Wu and A Ding. Population HIV-1 dynamics in vivo: applicable models and inferential tools for virological data from AIDS clinical trials. *Biometrics*. 1999. 55:410-418.
- [12] A.P. Dempster and N.M. Laird and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B*. 1977. 39:1-38.
- [13] G.C. Wei and M.A. Tanner. A Monte-Carlo implementation of the EM algorithm and the poor man's data augmentation algorithm. *Journal of the American Statistical Association*. 1990. 85:699-704.
- [14] E.H. IP. A Stochastic EM Estimator in the Presence of Missing Data: Theory and Application. 1994. Stanford University. Phd Dissertation.
- [15] S.F. Nielsen. The stochastic EM algorithm: estimation and asymptotic results. *Bernoulli*. 2000. 6(3):457-489.
- [16] J Geweke. *Handbook of Computational Economics*. 1996. North-Holland. Amsterdam.
- [17] S. Retout and E. Comets and A. Samson and F. Mentre. Design in Nonlinear Mixed Effects Models: Optimization Using the Fedorov-Wynn Algorithm and Power of the Wald Test for Binary Covariates. *Statistics in Medicine*. 2007. 26(28):5162-5179.
- [18] F. Yang. A Stochastic EM Algorithm for Quantile and Censored Quantile Regression Models. *Compt Econ*. 2018. 52:555-582.
- [19] R. Wang and A. Bing and C. Wang and Y. Hu and R.J. Bosch and V. DeGruttola. A flexible nonlinear mixed effects model for HIV viral load rebound after treatment interruption. *Statistics in Medicine*. 2020. 39:2051-2066.
- [20] R. Huang and R. Xu and P. S. Dulai. Sensitivity analysis of treatment effect to unmeasured confounding in observational studies with survival and competing risk outcomes. *Statistics in Medicine*. 2020. 39:3397-3411.
- [21] AS Perelson and AU Neumann and M Markowitz. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science*. 1996. 271:1582-1586.
- [22] M Lindstrom and D Bates. Nonlinear mixed effects models for repeated measures data. *Biometrics*. 1990. 46:673-687.
- [23] A. Gelman and D.B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*. 1992. 7:457-472.
- [24] JK Lindsey. *Nonlinear Models in Medical Statistics*. 2001. Oxford. Oxford University Press.