Supplementary Materials for
Fitting Birth-Death Processes to Panel Data with Applications to Bacterial DNA Fingerprinting

by Charles R. Doss, Marc A. Suchard, Ian Holmes, Midori Kato-Maeda, and Vladimir N. Minin

Appendix A. Here, we prove our main result. We repeat the theorem formulation first.

THEOREM 1. *Let $\{X_t\}$ be a linear BDI process with parameters $\lambda \geq 0$, $\mu \geq 0$, and $\nu \geq 0$. Over the interval $[0, t]$, let N_t^+ be the number of jumps up, N_t^- be the number of jumps down, and R_t be the total particle-time. Then $H_i(u, v, w, s, t) = E\left(u^{N_t^+} v^{N_t^-} e^{-wR_t} s^{X_t} | X_0 = i\right)$ satisfies the following partial differential equation:*

$$\text{(S-1)} \qquad \frac{\partial}{\partial t} H_i = \left[s^2 u\lambda - (\lambda + \mu + w)s + v\mu\right] \frac{\partial}{\partial s} H_i + \nu(us - 1)H_i,$$

subject to initial condition $H_i(u, v, w, s, 0) = s^i$. The Cauchy problem defined by equation (S-1) *and the initial condition has a unique solution. When $\lambda > 0$, the solution is*

$$H_i(u, v, w, s, t) = \left(\frac{\alpha_1 - \alpha_2 \frac{s-\alpha_1}{s-\alpha_2} e^{-\lambda(\alpha_2 - \alpha_1)ut}}{1 - \frac{s-\alpha_1}{s-\alpha_2} e^{-\lambda(\alpha_2-\alpha_1)ut}}\right)^i \left(\frac{\alpha_1 - \alpha_2}{s - \alpha_2 - (s - \alpha_1)\, e^{-\lambda(\alpha_2-\alpha_1)ut}}\right)^{\frac{\nu}{\lambda}} e^{-\nu(1-u\alpha_1)t},$$

where $\alpha_i = \frac{\lambda+\mu+w \mp \sqrt{(\lambda+\mu+w)^2 - 4\lambda\mu uv}}{2\lambda u}$ for $i = 1, 2$. When $\lambda = 0$, the solution is

$$H_i(u, v, w, s, t) = \left(se^{-(\mu+w)t} - \frac{v\mu\left(e^{-(\mu+w)t} - 1\right)}{\mu + w}\right)^i e^{\frac{\nu u[v\mu - (\mu+w)s]\left(e^{-(\mu+w)t}-1\right)}{(\mu+w)^2} + \nu\left(\frac{uv\mu}{\mu+w} - 1\right)t}.$$

PROOF. We consider a joint measure $V_{i,j}(n_1, n_2, x, t) = P(X_t = j, N_t^+ = n_1, N_t^- = n_2, R_t \leq x | X_0 = i)$. For ease of notation, we will let λ_{ij} be the instantaneous rate of transitioning from state i to state j for the BDI process and $\lambda_i = \sum_{j \neq i} \lambda_{ij}$. Also, we will let $a_i = i$ be the reward rate for R_t; that is, for staying in state i for time h, the process R_t increases by ih. Following Neuts (1995), we start with

$$\begin{aligned} V_{ij}(n_1, n_2, x, t) &= 1_{\{i=j\}} 1_{\{x \geq a_i t\}} 1_{\{n_1 = n_2 = 0\}} e^{-\lambda_i t} \\ &+ 1_{\{j \geq 1\}} 1_{\{n_1 \geq 1\}} \int_0^t V_{i,j-1}[n_1 - 1, n_2, x - (t-u)a_j, u] e^{-\lambda_j (t-u)} \lambda_{j-1,j}\, du \\ &+ 1_{\{n_2 \geq 1\}} \int_0^t V_{i,j+1}[n_1, n_2 - 1, x - (t-u)a_j, u] e^{-\lambda_j(t-u)} \lambda_{j+1,j}\, du, \end{aligned}$$

where $1_{\{\cdot\}}$ is the indicator function. Next, we derive differential equations for the Laplace-Stieltjes transform $V_{i,j}^*(n_1, n_2, w, t) = \int_0^\infty e^{-wx}\, dV_{i,j}(n_1, n_2, x, t)$:

$$\begin{aligned} \frac{\partial}{\partial t} V_{ij}^*(n_1, n_2, w, t) = &- jwV_{ij}^*(n_1, n_2, w, t) - [j(\lambda + \mu) + \nu]\, V_{ij}^*(n_1, n_2, w, t) \\ &+ 1_{\{n_1 \geq 1\}} 1_{\{j \geq 1\}} [\lambda(j-1) + \nu]\, V_{i,j-1}^*(n_1 - 1, n_2, w, t) \\ &+ 1_{\{n_2 \geq 1\}} \mu(j+1) V_{i,j+1}^*(n_1, n_2 - 1, w, t). \end{aligned}$$

We now write $H_i(u,v,w,s,t) = \sum_{j\geq \mathbf{0}} h_{i,j}(u,v,w,t)s^j$ where $h_{i,j}(u,v,w,t) := \sum_{n_1,n_2} V^*_{i,j}(n_1,n_2,w,t)u^{n_1}v^{n_2}$. The functions $h_{i,j}$ then satisfy

$$\frac{\partial}{\partial t}h_{ij}(u,v,w,t) = -\left[j(\lambda+\mu+w)+\nu\right]h_{i,j}(u,v,w,t) + \left[\lambda(j-1)+\nu\right]uh_{i,j-1}(u,v,w,t)1_{\{j\geq 1\}} + (j+1)\mu v h_{i,j+1}(u,v,w,t).$$

Using this fact, we arrive at

$$\begin{aligned}\frac{\partial}{\partial t}H_i =& -s\sum_{j\geq 1}s^{j-1}j(\lambda+\mu+w)h_{ij} + \sum_{j\geq 1}s^j(-\nu)h_{i,j} + -\nu h_{i0} + s\sum_{j\geq 1}s^{j-1}u\nu h_{i,j-1} \\ &+ \sum_{j\geq 1}s^j v(j+1)\mu h_{i,j+1} + v\mu h_{i,1} + s^2\sum_{j\geq 1}s^{j-2}u(j-1)\lambda h_{i,j-1} \\ =& -(\lambda+\mu+w)s\frac{\partial}{\partial s}H_i - \nu H_i + su\nu H_i + v\mu\frac{\partial}{\partial s}H_i + s^2u\lambda\frac{\partial}{\partial s}H_i,\end{aligned}$$

which proves that H_i satisfies equation (S-1).

Using the method of characteristics, we solve the above PDE with initial condition $H_i(u,v,w,s,0) = s^i$. When $\lambda > 0$, the solution is

$$H_i(u,v,w,s,t) = \left(\frac{\alpha_1 - \alpha_2\frac{s-\alpha_1}{s-\alpha_2}e^{-\lambda(\alpha_2-\alpha_1)ut}}{1-\frac{s-\alpha_1}{s-\alpha_2}e^{-\lambda(\alpha_2-\alpha_1)ut}}\right)^i \left(\frac{\alpha_1-\alpha_2}{s-\alpha_2-(s-\alpha_1)e^{-\lambda(\alpha_2-\alpha_1)ut}}\right)^{\frac{\nu}{\lambda}} e^{-\nu(1-u\alpha_1)t},$$

where $\alpha_i = \frac{\lambda+\mu+w\mp\sqrt{(\lambda+\mu+w)^2-4\lambda\mu uv}}{2\lambda u}$, for $i = 1,2$. In the case of $\lambda = 0$ (death-immigration model), the solution is

$$H_i(u,v,w,s,t) = \left(se^{-(\mu+w)t} - \frac{v\mu\left(e^{-(\mu+w)t}-1\right)}{\mu+w}\right)^i e^{\frac{\nu u[v\mu-(\mu+w)s]\left(e^{-(\mu+w)t}-1\right)}{(\mu+w)^2}+\nu\left(\frac{uv\mu}{\mu+w}-1\right)t}.$$

□

In Section 3.1 of the main document, we show how to use this generating function to compute the conditional moments we need for the EM algorithm. We use the same strategy to compute the joint and cross-moments we need for computation of the information matrix in Appendix B. For instance, in the main document on page 7 where we compute $G_i^+(t,s) = \partial H_i(u,1,0,s,t)/\partial u|_{u=1}$ when we want to compute $\mathbf{E}\left[N_t^+|X_0=i,X_t=j\right]$, we now instead compute one of

$$\frac{\partial^2 H_i(u,1,w,s,t)}{\partial u\partial w}\Big|_{u=1,w=0},$$
$$\frac{\partial^2 H_i(1,v,w,s,t)}{\partial v\partial w}\Big|_{v=1,w=0}, \text{ or}$$
$$\frac{\partial^2 H_i(u,v,0,s,t)}{\partial u\partial v}\Big|_{u=1,v=1},$$

when we want to compute one of $-\mathbf{E}\left[N_t^+R_t|X_0=i,X_t=j\right]$, $-\mathbf{E}\left[N_t^-R_t|X_0=i,X_t=j\right]$, or

$\mathbf{E}\left[N_t^+ N_t^- | X_0 = i, X_t = j\right]$, respectively, for some $i, j \in \mathbb{N}$. Similarly, we compute one of

$$\frac{\partial^2 H_i(u,1,0,s,t)}{\partial u^2}\Big|_{u=1},$$
$$\frac{\partial^2 H_i(1,v,0,s,t)}{\partial v^2}\Big|_{v=1}, \text{ or}$$
$$\frac{\partial^2 H_i(1,1,w,s,t)}{\partial w^2}\Big|_{w=0},$$

when we want to compute one of $\mathbf{E}\left[N_t^{+2} | X_0 = i, X_t = j\right]$, $\mathbf{E}\left[N_t^{-2} | X_0 = i, X_t = j\right]$, or $\mathbf{E}\left[R_t^2 | X_0 = i, X_t = j\right]$, respectively. Then, we use the Riemann approximation to the Fourier transform as described on page 8 of the main document to access the power series coefficient corresponding to the specific j that we want. The Riemann approximation approach is also the method we use to compute transition probabilities (Henrici, 1979; Sehl et al., 2011). Given a probability generating function $P(s) = \sum_{j \geq 0} p_j s^j$, we can extend P to the complex plane by defining $P(e^{2\pi\sqrt{-1}t}) = \sum_{j\geq 0} p_j e^{2\pi\sqrt{-1}t}$, and then the jth coefficient p_j is equal to $p_j = \int_0^1 P(e^{2\pi\sqrt{-1}t})e^{-2\pi\sqrt{-1}jt}dt$. We can thus approximate p_j by

$$\frac{1}{K}\sum_{k=0}^{K-1} P(e^{2\pi\sqrt{-1}t})e^{-2\pi\sqrt{-1}jk/K},$$

for large K. Thus, to compute the transition probabilities, which we need to convert joint moments into conditional ones as in (7) in the main document, we simply apply the above formula where P is the generating function for the BDI process X_t. This generating function is given by $H_i(1,1,0,s,t)$ where H_i is from Theorem 1. This yields the know form of the BDI generating function (Lange, 2004, page 168).

Appendix B. In this section, we provide details for calculating the observed information matrix. Louis (1982) shows that, in problems with incomplete observations, the observed information $\hat{I}_Y(\boldsymbol{\gamma})$ can be calculated as

$$-\frac{\partial^2}{\partial\boldsymbol{\gamma}^2} l_o(\mathbf{Y};\boldsymbol{\gamma}) = \mathrm{E}_{\boldsymbol{\gamma}}\left[-\frac{\partial^2}{\partial\boldsymbol{\gamma}^2} l_c(\mathbf{X};\boldsymbol{\gamma}) - \frac{\partial}{\partial\boldsymbol{\gamma}} l_c(\mathbf{X};\boldsymbol{\gamma})\frac{\partial}{\partial\boldsymbol{\gamma}} l_c(\mathbf{X};\boldsymbol{\gamma})' \Big| \mathbf{Y}\right] + \frac{\partial}{\partial\boldsymbol{\gamma}} l_o(\mathbf{Y};\boldsymbol{\gamma})\frac{\partial}{\partial\boldsymbol{\gamma}} l_o(\mathbf{Y};\boldsymbol{\gamma})', \tag{S-2}$$

where l_o and l_c are the observed-data and complete-data likelihoods, as defined in (2) and (3) of the main paper, and where the last term is 0 when we plug in the MLE $\hat{\boldsymbol{\gamma}}$ of $\boldsymbol{\gamma}$. Recalling $\log \lambda_p = \mathbf{z}'_{p,\lambda}\gamma_\lambda$, we get for $1 \leq j \leq c_1$,

$$\frac{\partial}{\partial \boldsymbol{\gamma}_{\lambda,j}} l_c(\mathbf{X};\boldsymbol{\gamma}) = \sum_{p=1}^{m} -(R_{t_{p,n(p)}} + t_{p,n(p)}\beta)e^{\mathbf{z}'_{p,\lambda}\gamma_\lambda} z_{p,\lambda,j} + N^+_{t_{p,n(p)}} z_{p,\lambda,j}, \tag{S-3}$$

or in matrix form,

$$\frac{\partial}{\partial \gamma_\lambda} l_c(\mathbf{X};\boldsymbol{\gamma}) = \mathbf{Z}'_\lambda \left(-\operatorname{diag}(\mathbf{R} + \beta\mathbf{T})e^{\mathbf{Z}_\lambda\gamma_\lambda} + \mathbf{N}^+\right), \tag{S-4}$$

where for $\mathbf{x} = (x_1, \ldots, x_s) \in \mathbb{R}^s$, some natural number s, we set $\text{diag}(\mathbf{x})$ to be the matrix with 0's on the non-diagonal elements and x_i as the (i,i)th element, and we define the vectors $\mathbf{R} := (R_1, \ldots, R_m) := (R_{1,t_{1,n(1)}}, \ldots, R_{m,t_{m,n(m)}})'$, $\mathbf{T} := (T_1, \ldots, T_m) := (t_{1,n_1}, \ldots, t_{m,n(m)})'$, and $\mathbf{N}^+ := (N_1^+, \ldots, N_m^+) := (N^+_{1,t_{1,n(1)}}, \ldots, N^+_{m,t_{m,n(m)}})'$ all in $\mathbb{R}^m$. Here that we take $e^{\mathbf{Z}_\lambda \gamma_\lambda}$ to be the exponential function applied componentwise to the vector $\mathbf{Z}_\lambda \gamma_\lambda$. Next, for $1 \le j \le c_2$, we see

$$\text{(S-5)} \qquad \frac{\partial}{\partial \boldsymbol{\gamma}_{\mu,j}} l_c(\mathbf{X}; \boldsymbol{\gamma}) = \sum_{p=1}^{m} -R_{t_{p,n_p}} e^{\mathbf{z}'_{p,\mu}\gamma_\mu} z_{p,\mu,j} + N^-_{t_{p,n(p)}} z_{p,\mu,j}$$

and, in matrix form,

$$\text{(S-6)} \qquad \frac{\partial}{\partial \gamma_\mu} l_c(\mathbf{X}; \boldsymbol{\gamma}) = \mathbf{Z}'_\mu \left(-\operatorname{diag}(\mathbf{R}) e^{\mathbf{Z}_\mu \gamma_\mu} + \mathbf{N}^- \right),$$

where $\mathbf{N}^- := (N_1^-, \ldots, N_m^-) := (N^-_{1,t_{1,n(1)}}, \ldots, N^-_{m,t_{m,n(m)}})' \in \mathbb{R}^m$.

We next need to compute the squared gradient. We can write it as

$$\text{(S-7)} \qquad \frac{\partial}{\partial \boldsymbol{\gamma}} l_c(\mathbf{X}; \boldsymbol{\gamma}) \frac{\partial}{\partial \boldsymbol{\gamma}} l_c(\mathbf{X}; \boldsymbol{\gamma})' = \begin{pmatrix} A & B \\ B' & C \end{pmatrix},$$

where we can calculate the matrix blocks by

$$\text{(S-8)} \qquad \begin{aligned} A &= \mathbf{Z}'_\lambda(-\operatorname{diag}(\mathbf{R} + \beta\mathbf{T})\boldsymbol{\lambda} + \mathbf{N}^+)(-\operatorname{diag}(\mathbf{R} + \beta\mathbf{T})\boldsymbol{\lambda} + \mathbf{N}^+)'\mathbf{Z}_\lambda \\ B &= \mathbf{Z}'_\lambda(-\operatorname{diag}(\mathbf{R} + \beta\mathbf{T})\boldsymbol{\lambda} + \mathbf{N}^+)(-\operatorname{diag}(\mathbf{R})\boldsymbol{\mu} + \mathbf{N}^-)'\mathbf{Z}_\mu \\ C &= \mathbf{Z}'_\mu(-\operatorname{diag}(\mathbf{R})\boldsymbol{\mu} + \mathbf{N}^-)(-\operatorname{diag}(\mathbf{R})\boldsymbol{\mu} + \mathbf{N}^-)'\mathbf{Z}_\mu. \end{aligned}$$

To get the Hessian of l_c, we differentiate again, starting with (S-3), to see for $1 \le j, k \le c_1$,

$$\frac{\partial^2}{\partial \boldsymbol{\gamma}_{\lambda,j} \partial \boldsymbol{\gamma}_{\lambda,k}} l_c(\mathbf{X}; \boldsymbol{\gamma}) = -\sum_{p=1}^{m} (R_{t_{p,n(p)}} + \beta t_{p,n(p)}) e^{\mathbf{z}'_{p,\lambda}\gamma_\lambda} z_{p,\lambda,j} z_{p,\lambda,k} = -\sum_{p=1}^{m} (R_{t_{p,n(p)}} + \beta t_{p,n(p)}) \lambda_p z_{p,\lambda,j} z_{p,\lambda,k},$$

or, in matrix form,

$$\frac{\partial^2}{\partial \boldsymbol{\gamma}_\lambda^2} l_c(\mathbf{X}; \boldsymbol{\gamma}) = -\mathbf{Z}'_\lambda \operatorname{diag}(\mathbf{R} + \beta\mathbf{T}) \operatorname{diag}(e^{\mathbf{Z}_\lambda \gamma_\lambda}) \mathbf{Z}_\lambda.$$

Next, differentiating (S-5) we get for $1 \le j, k \le c_2$,

$$\frac{\partial^2}{\partial \boldsymbol{\gamma}_{\mu,j} \partial \boldsymbol{\gamma}_{\mu,k}} l_c(\mathbf{X}; \boldsymbol{\gamma}) = -\sum_{p=1}^{m} R_{t_{p,n(p)}} e^{\mathbf{z}'_{p,\mu}\gamma_\mu} z_{p,j} z_{p,k} = -\sum_{p=1}^{m} R_{t_{p,n(p)}} \mu_p z_{p,j} z_{p,k},$$

or, in matrix form,

$$\frac{\partial^2}{\partial \boldsymbol{\gamma}_\mu^2} l_c(\mathbf{X}; \boldsymbol{\gamma}) = -\mathbf{Z}'_\mu \operatorname{diag}(\mathbf{R}) \operatorname{diag}(e^{\mathbf{Z}_\mu \gamma_\mu}) \mathbf{Z}_\mu.$$

Since for all $1 \le j \le c_1$ **and** $1 \le k \le c_2$**,**

$$\frac{\partial^2}{\partial \gamma_{\lambda,j} \gamma_{\mu,k}} l_c(\mathbf{X}; \boldsymbol{\gamma}) = 0,$$

we have now calculated the squared gradient and the Hessian of l_c. We need only to take expectations. For the squared gradient, second- and cross-moments appear, whereas for the Hessian only first moments appear. For the squared gradient, we now take expectations element-by-element. We denote the kth element of the sufficient-statistic expectation vectors $\boldsymbol{U}, \boldsymbol{D}$, and $\boldsymbol{P}$ that are defined in (12) of the main paper by U_k, D_k and P_k, respectively. We see for $1 \le p, q \le c_1$ that $\mathrm{E}_{\tilde{\gamma}}\left[A_{p,q}\middle|\mathbf{Y}\right]$ is equal to

$$\sum_{k=1}^{m}\sum_{l=1,l\neq k}^{m}\big((P_kP_l + P_kT_l\beta + P_lT_k\beta + T_lT_k\beta^2)\lambda_k\lambda_l - P_k\lambda_kU_l - T_k\beta\lambda_kU_l - P_l\lambda_lU_k - T_l\beta\lambda_lU_k + U_kU_l\big)z_{k,\lambda,p}z_{l,\lambda,q}$$
$$+\sum_{k=1}^{m}\Big((\mathrm{E}_{\tilde{\gamma}}\left[R_k^2\middle|\mathbf{Y}\right] - 2P_kT_k\beta + T_k^2\beta^2)\lambda_k^2 - 2\mathrm{E}_{\tilde{\gamma}}\left[R_kN_k^+\middle|\mathbf{Y}\right]\lambda_k - 2T_k\beta\lambda_kU_k + \mathrm{E}_{\tilde{\gamma}}\left[N_k^{+2}\middle|\mathbf{Y}\right]\Big)z_{k,\lambda,p}z_{k,\lambda,q}.$$

For $1 \le p \le c_1$ and $1 \le q \le c_2$ we see that $\mathrm{E}_{\tilde{\gamma}}\left[B_{p,q}\middle|\mathbf{Y}\right]$ is equal to

$$\sum_{k=1}^{m}\sum_{l=1,l\neq k}^{m}\big(P_kP_l\lambda_k\lambda_l + P_lT_k\beta\lambda_k\lambda_l - U_kP_l\lambda_l - (P_k + T_k\beta)D_l\lambda_k + U_kD_l\big)z_{k,\lambda,p}z_{l,\mu,q}$$
$$+\sum_{k=1}^{m}\Big(-\mathrm{E}_{\tilde{\gamma}}\left[R_k^2\middle|\mathbf{Y}\right]\lambda_k\mu_k - \beta T_k\lambda_k\mu_k - \mathrm{E}_{\tilde{\gamma}}\left[R_kN_k^-\middle|\mathbf{Y}\right]\lambda_k - \beta T_k\lambda_k\mathrm{E}_{\tilde{\gamma}}\left[N_k^-\middle|\mathbf{Y}\right] - \mathrm{E}_{\tilde{\gamma}}\left[N_k^+R_k\middle|\mathbf{Y}\right]\mu_k + \mathrm{E}_{\tilde{\gamma}}\left[N_k^+N_k^-\middle|\mathbf{Y}\right]\Big)z_{k,\lambda,p}z_{k,\mu,q}.$$

For $1 \le p, q \le c_2$ we see that $\mathrm{E}_{\tilde{\gamma}}\left[C_{p,q}\middle|\mathbf{Y}\right]$ is equal to

$$\sum_{k=1}^{m}\sum_{l=1,l\neq k}^{m}\big(P_kP_l\mu_k\mu_l - D_kP_l\mu_l - D_lP_k\mu_k + D_kD_l\big)z_{k,\mu,p}z_{l,\mu,q}$$
$$+\sum_{k=1}^{m}\big(\mu_k^2\mathrm{E}_{\tilde{\gamma}}\left[R_k^2\middle|\mathbf{Y}\right] - 2\mu_k\mathrm{E}_{\tilde{\gamma}}\left[N_k^-R_k\middle|\mathbf{Y}\right] + \mathrm{E}_{\tilde{\gamma}}\left[N_k^{-2}\middle|\mathbf{Y}\right]\big)z_{k,\mu,p}z_{k,\mu,q}.$$

For the Hessian term, since for $1 \le j \le c_1$ and $1 \le k \le c_2$, $\frac{\partial^2}{\partial\gamma_{\lambda,j}\gamma_{\mu,k}}l_c(\mathbf{X};\boldsymbol{\gamma}) = 0$, we see that $\mathrm{E}_{\boldsymbol{\gamma}}\left[-\frac{\partial^2}{\partial\boldsymbol{\gamma}^2}l_c(\mathbf{X};\boldsymbol{\gamma})\right]$ equals

$$\begin{pmatrix} \mathbf{Z}_\lambda' \,\mathrm{diag}(\mathbf{P} + \beta\mathbf{T})\,\mathrm{diag}(\boldsymbol{\lambda})\mathbf{Z}_\lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_\mu' \,\mathrm{diag}(\mathbf{P})\,\mathrm{diag}(\boldsymbol{\mu})\mathbf{Z}_\mu \end{pmatrix}. \tag{S-9}$$

The generating function presented in Theorem 1 can be used to compute the conditional means of all the needed cross-products and square terms in the gradient and Hessian, as shown in Appendix A. Thus we have now computed all the terms in (S-2).

Appendix C. In this section, we show that for two important special cases of the BDI model the E-step of the EM algorithm does not require any numeric approximations. **Notice that the first of these two models, the death-immigration model, is not a BDRI model — the focus of the main document. Nonetheless, we think it is important to illustrate that our theoretical developments apply to the death-immigration process, because this model plays an important role in applications of BDI processes (Crespi, Cumberland and Blower, 2005).**

Death-Immigration Model. We have shown that the generating function, $H_i(u,v,w,s,t) = \mathrm{E}\left(u^{N_t^+} v^{N_t^-} e^{-wR_t} s^{X_t} | X_0 = i\right)$, for the death-immigration model is

$$H_i(u,v,w,s,t) = \left(se^{-(\mu+w)t} - \frac{v\mu\left(e^{-(\mu+w)t}-1\right)}{\mu+w}\right)^i e^{\frac{\nu u[v\mu-(\mu+w)s]\left(e^{-(\mu+w)t}-1\right)}{(\mu+w)^2}+\nu\left(\frac{uv\mu}{\mu+w}-1\right)t}.$$

Suppose we are interested in computing $\mathrm{E}\left(N_t^+ 1_{\{X_t=j\}} | X_0 = i\right)$. First, we fix $v = 1$ and $w = 0$. Next, we differentiate the generating function once with respect to u and j times with respect to s, plugging in 1 and 0 respectively:

$$\mathrm{E}\left(N_t^+ 1_{\{X_t=j\}} | X_0 = i\right) = \frac{\partial}{\partial u}\frac{\partial^j}{\partial s^j} H_i(u,1,0,s,t)\Big|_{u=1,s=0}$$

and

$$H_i(u,1,0,s,t) = \left[1 + e^{-\mu t}(s-1)\right]^i e^{-\frac{\nu u(s-1)\left(e^{-\mu t}-1\right)}{\mu}+\nu(u-1)t} = (A+Bs)^i e^{C(u)s+D(u)},$$

where

$$\begin{aligned}
A &= 1 - e^{-\mu t}, \\
B &= e^{-\mu t}, \\
C(u) &= -\frac{\nu u\left(e^{\mu t}-1\right)}{\mu}, \\
D(u) &= \frac{\nu u\left(e^{\mu t}-1\right)}{\mu} + \nu(u-1)t.
\end{aligned}$$

Therefore,

$$\frac{\partial}{\partial u} H_i(u,1,0,s,t)\Big|_{u=1} = (A+Bs)^i e^{C(1)s+D(1)}[C'(1)s + D'(1)],$$

where

$$\begin{aligned}
C'(1) &= -\frac{\nu\left(e^{\mu t}-1\right)}{\mu}, \\
D'(1) &= \frac{\nu\left(e^{\mu t}-1\right)}{\mu} + \nu t.
\end{aligned}$$

Now, the derivatives with respect to s can be recovered by expanding $\frac{\partial}{\partial u} H_i(u,1,0,s,t)\Big|_{u=1}$ into a power series:

$$\frac{\partial}{\partial u} H_i(u,1,0,s,t)\Big|_{u=1} = (A+Bs)^i \, e^{C(1)s} e^{D(1)} [C'(1)s + D'(1)] = e^{D(1)} [C'(1)s + D'(1)]$$
$$\times \left[\sum_{m=0}^{i} \binom{i}{m} A^{i-m} B^m s^m \right] \left[\sum_{k=0}^{\infty} \frac{C^k}{k!} s^k \right]$$
$$= e^{D(1)} \left\{ \sum_{m=0}^{i+1} \left[C'(1) \binom{i}{m-1} A^{i-m+1} B^{m-1} 1_{\{m \geq 1\}} + D'(1) \binom{i}{m} A^{i-m} B^m 1_{\{m \leq i\}} \right] s^m \right\} \left[\sum_{k=0}^{\infty} \frac{C^k}{k!} s^k \right]$$
$$= \sum_{n=0}^{\infty} e^{D(1)} \left\{ \sum_{k=\max\{0,n-i-1\}}^{n} \frac{C^k}{k!} \left[C'(1) \binom{i}{n-k-1} A^{i-n+k+1} B^{n-k-1} 1_{\{n-k \geq 1\}} \right.\right.$$
$$\left.\left. + D'(1) \binom{i}{n-k} A^{i-n+k} B^{n-k} 1_{\{n-k \leq i\}} \right] \right\} s^n.$$

Therefore,

$$\mathrm{E}\left(N_t^+ 1_{\{X_t=j\}} | X_0 = i\right) = e^{D(1)} \sum_{k=\max\{0,j-i-1\}}^{j} \frac{C^k}{k!} \left[C'(1) \binom{i}{j-k-1} A^{i-j+k+1} B^{j-k-1} 1_{\{j-k \geq 1\}} \right.$$
$$\left. + D'(1) \binom{i}{j-k} A^{i-j+k} B^{j-k} 1_{\{j-k \leq i\}} \right].$$

One can derive expectations of N_t^- and R_t in a similar fashion.

Sequence Alignment BDI Model. Here we demonstrate that our generating function approach results in analytic formulae for the E-step in the evolutionary EM algorithm, developed by Holmes (2005). This is in contrast to the original Holmes (2005)'s implementation, which requires numerically solving a system of nonlinear ordinary differential equations. Holmes (2005)'s algorithm is based on a TKF91 model of sequence alignment evolution (Thorne, Kishino and Felsenstein, 1991). Instead of diving into the intricacies of this model, we refer the reader to Ian Holmes' web page (`http://biowiki.org/TkfIndelModelPathSummaries`), where he poses an open problem of deriving the E-step of Holmes (2005)'s algorithm in closed form and explicitly formulates this problem in terms of the BDI process. To derive the E-step of Holmes (2005)'s algorithm in closed form, using our BDI notation, one needs to find analytic expressions of the following expectations:

1. $\mathrm{E}\left(N_t^+ 1_{\{X_t=j\}} | X_0 = 0\right)$, $\mathrm{E}\left(N_t^- 1_{\{X_t=j\}} | X_0 = 0\right)$, and $\mathrm{E}\left(R_t 1_{\{X_t=j\}} | X_0 = 0\right)$ when $\nu = \lambda$,
2. $\mathrm{E}\left(N_t^+ 1_{\{X_t=j\}} | X_0 = 1\right)$, $\mathrm{E}\left(N_t^- 1_{\{X_t=j\}} | X_0 = 1\right)$, and $\mathrm{E}\left(R_t 1_{\{X_t=j\}} | X_0 = 1\right)$ when $\nu = 0$,

Notice that in the the sequence alignment BDI model, X_t is a BDRI process (with $\beta = 1$) when $X_0 = 0$ and X_t is a linear birth-death process (with no immigration) when $X_0 = 1$. We derive the analytic formulae for $\mathrm{E}\left(N_t^+ 1_{\{X_t=j\}} | X_0 = 0\right)$ ($\nu = \lambda$) and $\mathrm{E}\left(N_t^+ 1_{\{X_t=j\}} | X_0 = 1\right)$ ($\nu = 0$). The other expectations can be derived analogously.

1. **Objective**: $\mathrm{E}\left(N_t^+ 1_{\{X_t=j\}} | X_0 = 0\right)$ $(\nu = \lambda)$:
   First,
   $$\mathrm{E}\left(N_t^+ 1_{\{X_t=j\}} | X_0 = 0\right) = \frac{1}{j!} \frac{\partial}{\partial r} \frac{\partial^j}{\partial s^j} \mathbf{H}_0^+(r,s,t)\Big|_{s=0,r=1},$$

where

$$\mathbf{H}_0^+(r,s,t) = H_0(r,1,0,s,t) = \frac{(\alpha_1-\alpha_2)\,e^{-\lambda(1-r\alpha_1)t}}{s-\alpha_2-(s-\alpha_1)\,e^{-\lambda(\alpha_2-\alpha_1)rt}} \text{ and } \alpha_{1,2} = \frac{\lambda+\mu\mp\sqrt{(\lambda+\mu)^2-4\lambda\mu r}}{2\lambda r}.$$

We find the formula for this partial derivative by explicit differentiation:

$$\frac{\partial^j}{\partial s^j}\mathbf{H}_0^+(r,s,t) = \frac{(-1)^j j!\,(\alpha_1-\alpha_2)\,e^{-\lambda(1-r\alpha_1)t}\left(1-e^{-\lambda(\alpha_2-\alpha_1)rt}\right)^j}{\left(s-\alpha_2-(s-\alpha_1)\,e^{-\lambda(\alpha_2-\alpha_1)rt}\right)^{j+1}},$$

$$\frac{\partial^j}{\partial s^j}\mathbf{H}_0^+(r,s,t)\Big|_{s=0} = \frac{\overbrace{(-1)^j j!\,(\alpha_1-\alpha_2)\,e^{-\lambda(1-r\alpha_1)t}\left(1-e^{-\lambda(\alpha_2-\alpha_1)rt}\right)^j}^{A(r)}}{\underbrace{\left(\alpha_1 e^{-\lambda(\alpha_2-\alpha_1)rt}-\alpha_2\right)^{j+1}}_{B(r)}}$$

$$\frac{\partial}{\partial r}\frac{\partial^j}{\partial s^j}\mathbf{H}_0^+(r,s,t)\Big|_{s=0,r=1} = \frac{A'(1)B(1)-A(1)B'(1)}{B^2(1)},$$

where

$$\begin{aligned}
A(1) &= \left(1-e^{(\lambda-\mu)t}\right)^j\left(1-\frac{\mu}{\lambda}\right),\\
B(1) &= \left(e^{(\lambda-\mu)t}-\frac{\mu}{\lambda}\right)^{j+1},\\
A'(1) &= \left(1-e^{(\lambda-\mu)t}\right)^{j-1}\left[j2\mu t e^{(\lambda-\mu)t}+\left(1-e^{(\lambda-\mu)t}\right)\left(\frac{\lambda^2+\mu^2}{\lambda(\mu-\lambda)}-\mu t\right)\right],\\
B'(1) &= (j+1)\left(e^{(\lambda-\mu)t}-\frac{\mu}{\lambda}\right)^j\left(\frac{\lambda(1+2\mu t)}{\mu-\lambda}e^{(\lambda-\mu)t}+\frac{\mu^2}{\lambda(\mu-\lambda)}\right).
\end{aligned}$$

2. **Objective**: $\mathrm{E}\left(N_t^+ 1_{\{X_t=j\}}|X_0=1\right)$ $(\nu=0)$:
As before,

$$\mathrm{E}\left(N_t^+ 1_{\{X_t=j\}}|X_0=1\right) = \frac{1}{j!}\frac{\partial}{\partial r}\frac{\partial^j}{\partial s^j}\mathbf{H}_1^+(r,s,t)\Big|_{s=0,r=1},$$

where

$$\mathbf{H}_1^+(r,s,t) = H_1(r,1,0,s,t) = \frac{\alpha_1(s-\alpha_2)-\alpha_2\,(s-\alpha_1)\,e^{-\lambda(\alpha_2-\alpha_1)rt}}{s-\alpha_2-(s-\alpha_1)\,e^{-\lambda(\alpha_2-\alpha_1)rt}} = \alpha_2+\frac{\alpha_1-\alpha_2}{1-\left(\frac{s-\alpha_1}{s-\alpha_2}\right)e^{-\lambda(\alpha_2-\alpha_1)rt}}$$

For $j=0$, we just need to plug in $s=0$:

$$\mathbf{H}_1^+(r,0,t) = \alpha_2+\frac{\overbrace{\alpha_2\alpha_1-\alpha_2^2}^{C(r)}}{\underbrace{\alpha_2-\alpha_1 e^{-\lambda(\alpha_2-\alpha_1)rt}}_{D(r)}} = \alpha_2+\frac{C(r)}{D(r)}.$$

Then

$$\frac{d}{dr}\mathbf{H}_1^+(r,0,t)\Big|_{r=1} = -\frac{\mu^2}{\lambda(\mu-\lambda)} + \frac{C'(1)D(1)-C(1)D'(1)}{D^2(1)},$$

where

$$\begin{aligned}
C(1) &= \frac{\mu}{\lambda}\left(1-\frac{\mu}{\lambda}\right),\\
D(1) &= \frac{\mu}{\lambda} - e^{(\lambda-\mu)t},\\
C'(1) &= \frac{\mu}{\mu-\lambda}\left(1+2\frac{\mu^2}{\lambda^2}-\frac{\mu}{\lambda}\right),\\
D'(1) &= -\left[\frac{\mu^2}{\lambda(\mu-\lambda)} + \frac{\lambda}{\mu-\lambda}e^{(\lambda-\mu)t}\left(1+2\mu t\right)\right].
\end{aligned}$$

For $j > 0$

$$\frac{\partial^j}{\partial s^j}\mathbf{H}_1^+(r,s,t) = (-1)^{j+1}j!\left[\frac{\overbrace{\alpha_2(\alpha_1-\alpha_2)\left(1-e^{\lambda(\alpha_2-\alpha_1)rt}\right)^j}^{E(r)}}{\underbrace{\left(\alpha_1 e^{\lambda(\alpha_2-\alpha_1)rt}-\alpha_2\right)^{j+1}}_{F(r)}} + \frac{\overbrace{(\alpha_1-\alpha_2)\left(1-e^{\lambda(\alpha_2-\alpha_1)rt}\right)^{j-1}}^{G(r)}}{\underbrace{\left(\alpha_1 e^{\lambda(\alpha_2-\alpha_1)rt}-\alpha_2\right)^{j}}_{I(r)}}\right].$$

Then

$$\frac{\partial}{\partial r}\frac{\partial^j}{\partial s^j}\mathbf{H}_1^+(r,s,t)\Big|_{s=0,r=1} = (-1)^{j+1}j!\left[\frac{E'(1)F(1)-E(1)F'(1)}{F^2(1)} + \frac{G'(1)I(1)-G(1)I'(1)}{I^2(1)}\right],$$

where

$$E(1) = \frac{\mu}{\lambda}\left(1-\frac{\mu}{\lambda}\right)\left(1-e^{(\lambda-\mu)t}\right)^j, \tag{S-10}$$

$$F(1) = \left(e^{(\lambda-\mu)t}-\frac{\mu}{\lambda}\right)^{j+1}, \tag{S-11}$$

$$G(1) = \left(1-\frac{\mu}{\lambda}\right)\left(1-e^{(\lambda-\mu)t}\right)^{j-1}, \tag{S-12}$$

$$I(1) = \left(e^{(\lambda-\mu)t}-\frac{\mu}{\lambda}\right)^{j}, \tag{S-13}$$

$$E'(1) = \left(1-e^{(\lambda-\mu)t}\right)^{j-1}\left[\frac{\mu}{\mu-\lambda}\left(1+2\frac{\mu^2}{\lambda^2}-\frac{\mu}{\lambda}\right)\left(1-e^{(\lambda-\mu)t}\right) + j2\frac{\mu^2}{\lambda}e^{(\lambda-\mu)t}t\right], \tag{S-14}$$

$$F'(1) = (j+1)\left(e^{(\lambda-\mu)t}-\frac{\mu}{\lambda}\right)^{j}\left[\frac{\lambda}{\mu-\lambda}e^{(\lambda-\mu)t}(1+2\mu t) + \frac{\mu^2}{\lambda(\mu-\lambda)}\right], \tag{S-15}$$

$$G'(1) = \left(1-e^{(\lambda-\mu)t}\right)^{j-2}\left[\frac{\lambda^2+\mu^2}{\lambda(\mu-\lambda)}\left(1-e^{(\lambda-\mu)t}\right) + (j-1)e^{(\lambda-\mu)t}2\mu t\right], \tag{S-16}$$

$$I'(1) = j\left(e^{(\lambda-\mu)t}-\frac{\mu}{\lambda}\right)^{j-1}\left[\frac{\lambda}{\mu-\lambda}e^{(\lambda-\mu)t}(1+2\mu t) + \frac{\mu^2}{\lambda(\mu-\lambda)}\right]. \tag{S-17}$$

References.

Crespi, C. M., Cumberland, W. G. and Blower, S. (2005). A queueing model for chronic recurrent conditions under panel observation. *Biometrics* **61** 194–199.

Henrici, P. (1979). Fast Fourier transform methods in computational complex analysis. *SIAM Review* **21** 481–527.

Holmes, I. (2005). Using evolutionary expectation maximization to estimate indel rates. *Bioinformatics* **21** 2294-2300.

Lange, K. (2004). *Applied Probability*, first ed. Springer-Verlag, New York.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B* **44** 226–233.

Neuts, M. F. (1995). *Algorithmic Probability: A Collection of Problems*. *Stochastic Modeling Series*. Chapman & Hall, London.

Sehl, M., Zhou, H., Sinsheimer, J. S. and Lange, K. L. (2011). Extinction models for cancer stem cell therapy. *Mathematical Biosciences* **234** 132–146.

Thorne, J. L., Kishino, H. and Felsenstein, J. (1991). An evolutionary model for maximum likelihood alignment of DNA sequences. *Journal of Molecular Evolution* **33** 114–124.