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 \( H_i(u, v, w, s, t) = E \left( (u^{N_t^+}v^{N_t^-}e^{-wR_t}s^X)|X_0 = i \right) \) satisfies the following partial differential equation:

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

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

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

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

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

**Proof.** We consider a joint measure \( V_{ij}(n_1, n_2, x, t) = P(X_t = j, N_t^+ = n_1, N_t^- = n_2, R_t \leq x|X_0 = i) \). For ease of notation, we will let \( \lambda_{ij} \) 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 \( \alpha_i = i \) 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

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

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

\[
\frac{\partial}{\partial t} V_{ij}^*(n_1, n_2, w, t) = -j w V_{ij}^*(n_1, n_2, w, t) - [j(\lambda + \mu) + \nu] V_{ij}^*(n_1, n_2, w, t)
+ 1_{\{n_1 \geq 1\}} 1_{\{j \geq 1\}} [\lambda(j-1) + \nu] V_{ij-1}^*(n_1 - 1, n_2, w, t)
+ 1_{\{n_2 \geq 1\}} [\mu(j+1)] V_{ij+1}^*(n_1, n_2 - 1, w, t).
\]
We now write \( H_i(u,v,w,s,t) = \sum_{j \geq 0} h_{i,j}(u,v,w,t)s^j \) where \( h_{i,j}(u,v,w,t) := \sum_{n_1,n_2} V^s_{ij}(n_1,n_2,w,t)u^{n_1}v^{n_2} \). The functions \( h_{i,j} \) then satisfy

\[
\frac{\partial}{\partial t} h_{i,j}(u,v,w,t) = - [J(\lambda + \mu + w) + \nu] h_{i,j}(u,v,w,t) + [\lambda(j-1) + \nu] u h_{i,j-1}(u,v,w,t) 1_{(j \geq 1)}
\]

\[
+ (j+1) \nu v h_{i,j+1}(u,v,w,t).
\]

Using this fact, we arrive at

\[
\frac{\partial}{\partial t} H_i = - s \sum_{j \geq 1} s^j \frac{j(\lambda + \mu + w)}{h_{i,j}} + \sum_{j \geq 1} s^j (-\nu) h_{i,j} - \nu h_{i,0} + s \sum_{j \geq 1} s^j u \nu h_{i,j-1}
\]

\[
+ \sum_{j \geq 1} s^j v (j+1) u h_{i,j+1} + \nu h_{i,1} + s^2 \sum_{j \geq 1} s^j u (j-1) \lambda h_{i,j-1}
\]

\[
= - (\lambda + \mu + w) s \frac{\partial}{\partial s} H_i - \nu H_i + sw H_i + \nu \mu \frac{\partial}{\partial s} H_i + s^2 u \lambda \frac{\partial}{\partial s} H_i,
\]

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

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

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

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

\[ \square \]

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 \( E[N^+_i | X_0 = i, X_t = j] \), we now instead compute one of

\[
\frac{\partial^2 H_i(u,1,0,s,t)}{\partial u \partial w} |_{u=1,w=0},
\]

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

\[
\frac{\partial^2 H_i(u,v,0,s,t)}{\partial u \partial v} |_{u=1,v=1},
\]

when we want to compute one of \(-E[N^+_i R_t | X_0 = i, X_t = j]\), \(-E[N^-_i R_t | X_0 = i, X_t = j]\), or
In this section, we provide details for calculating the observed information matrix. Louis (1982) shows that, in problems with incomplete observations, the observed information form of the BDI generating function (Lange, 2004, page 168). Given a probability generating function, the 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 known 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 \( I_Y(\gamma) \) can be calculated as

\[
\frac{-\partial^2}{\partial \gamma^2} l_o(Y; \gamma) = E_Y \left[ -\frac{\partial^2}{\partial \gamma^2} l_c(X; \gamma) - \frac{\partial}{\partial \gamma} l_c(X; \gamma) \frac{\partial}{\partial \gamma} l_c(X; \gamma)' \right] Y \\
+ \frac{\partial}{\partial \gamma} l_o(Y; \gamma) \frac{\partial}{\partial \gamma} l_o(Y; \gamma)',
\]

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 \( \hat{\gamma} \) of \( \gamma \). Recalling log \( \lambda_p = z_p' \gamma \), we get for \( 1 \leq j \leq c_1 \),

\[
\frac{\partial}{\partial \gamma_{p,j}} l_c(X; \gamma) = \sum_{p=1}^{m} -(R_{t,p,n(p)} + t_{p,n(p)} \beta) e^{z_p' \gamma} z_{p,\lambda,j} + N_{t,p,n(p)} z_{p,\lambda,j},
\]

or in matrix form,

\[
\frac{\partial}{\partial \gamma_{\lambda}} l_c(X; \gamma) = Z_{\lambda} \left( -\text{diag}(R + \beta T) e^{Z_{\lambda} \gamma} + N^+ \right),
\]

respectively, for some \( i, j \in \mathbb{N} \). Similarly, we compute one of

\[
\frac{\partial^2 H_i(u, 1, 0, s, t)}{\partial u^2} \bigg|_{u=1}, \quad \frac{\partial^2 H_i(1, v, 0, s, t)}{\partial v^2} \bigg|_{v=1}, \quad \text{or} \quad \frac{\partial^2 H_i(1, 1, w, s, t)}{\partial w^2} \bigg|_{w=0},
\]

when we want to compute one of \( \mathbb{E} [N^+_t X_0 = i, X_t = j] \), \( \mathbb{E} [N^-_t X_0 = i, X_t = j] \), or \( \mathbb{E} [R^+_t X_0 = i, X_t = j] \), 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 \( P(s) = \sum_{j=0}^N p_j s^j \), we can extend \( P \) to the complex plane by defining \( P(e^{2\pi \sqrt{-1} t}) = \sum_{j=0}^\infty p_j e^{2\pi \sqrt{-1} j t} \), and then the \( j \)th coefficient \( p_j \) is equal to \( p_j = \int_0^1 P(e^{2\pi \sqrt{-1} t}) e^{-2\pi \sqrt{-1} j t} dt \). We can thus approximate \( p_j \) by

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

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 \( X_t \).
where for \( x = (x_1, \ldots, x_s) \in \mathbb{R}^s \), some natural number \( s \), we set \( \text{diag}(x) \) 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 \( e^{Z_\lambda \gamma_\lambda} \) to be the exponential function applied componentwise to the vector \( Z_\lambda \gamma_\lambda \). Next, for \( 1 \leq j \leq c_2 \), we see

\[(S-5) \quad \frac{\partial}{\partial \gamma_{\mu,j}} l_c(\mathbf{X}; \gamma) = \sum_{p=1}^{m} -R_{t_p, n(p)} e^{x_{p, \mu, \gamma_\mu} z_{p, \mu,j}} + N_{t_p, n(p)}^- z_{p, \mu,j}\]

and, in matrix form,

\[(S-6) \quad \frac{\partial}{\partial \gamma_{\mu}} l_c(\mathbf{X}; \gamma) = Z_\mu' (- \text{diag}(\mathbf{R}) e^{Z_\mu \gamma_\mu} + \mathbf{N}^-),\]

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

\[(S-7) \quad \frac{\partial}{\partial \gamma} l_c(\mathbf{X}; \gamma) \frac{\partial}{\partial \gamma} l_c(\mathbf{X}; \gamma)' = \begin{pmatrix} A & B \\ B' & C \end{pmatrix},\]

where we can calculate the matrix blocks by

\[(S-8) \quad A = Z_\lambda' (- \text{diag}(\mathbf{R} + \beta \mathbf{T}) \lambda + \mathbf{N}^+) (- \text{diag}(\mathbf{R} + \beta \mathbf{T}) \lambda + \mathbf{N}^+)' Z_\lambda \]

\[(S-8) \quad B = Z_\lambda' (- \text{diag}(\mathbf{R} + \beta \mathbf{T}) \lambda + \mathbf{N}^+) (- \text{diag}(\mathbf{R}) \mu + \mathbf{N}^-)' Z_\mu \]

\[(S-8) \quad C = Z_\mu' (- \text{diag}(\mathbf{R}) \mu + \mathbf{N}^-) (- \text{diag}(\mathbf{R}) \mu + \mathbf{N}^-)' Z_\mu.\]

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

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

or, in matrix form,

\[
\frac{\partial^2}{\partial \gamma_{\lambda}^2} l_c(\mathbf{X}; \gamma) = - Z_\lambda' \text{diag}(\mathbf{R} + \beta \mathbf{T}) \text{diag}(e^{Z_\lambda \gamma_\lambda}) Z_\lambda.
\]

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

\[
\frac{\partial^2}{\partial \gamma_{\mu,j} \partial \gamma_{\mu,k}} l_c(\mathbf{X}; \gamma) = - \sum_{p=1}^{m} R_{t_p, n(p)} e^{x_{p, \mu, \gamma_\mu} z_{p, j} z_{p, k}} = - \sum_{p=1}^{m} R_{t_p, n(p)} \lambda_p z_{p, j} z_{p, k},
\]

or, in matrix form,

\[
\frac{\partial^2}{\partial \gamma_{\mu}^2} l_c(\mathbf{X}; \gamma) = - Z_\mu' \text{diag}(\mathbf{R}) \text{diag}(e^{Z_\mu \gamma_\mu}) Z_\mu.
\]

Since for all \( 1 \leq j \leq c_1 \) and \( 1 \leq k \leq c_2 \),

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

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 \( U, D, \) and \( P \) that are defined
in (12) of the main paper by \( U_k, D_k \) and \( P_k \), respectively. We see for \( 1 \leq p, q \leq c_1 \) that \( E_\gamma \left[ A_{p,q} | Y \right] \) is equal to

\[
\sum_{k=1}^{m} \sum_{l=1, l \neq k}^{m} \left( (P_k P_l + \sum_{p,q} P_k T_{l,p} \beta + P_l T_{k,q} \beta + T_{k,l} \beta^2) \lambda_{k,l} \right)
- P_k \lambda_k U_l - T_k \beta \lambda_k U_l - P_l \lambda_l U_k - T_k \beta \lambda_l U_k + U_k U_l \right) z_{k,\lambda,p} z_{l,\lambda,q}
+ \sum_{k=1}^{m} \left( \left( E_\gamma \left[ R_k \right] Y \right) - 2P_k T_{k,\beta} + T_{k,2} \beta^2 \right) \lambda_k^2 - 2E_\gamma \left[ R_k N_k^+ | Y \right] \lambda_k
- 2T_k \beta \lambda_k U_k + E_\gamma \left[ N_k^2 | Y \right) \right) z_{k,\lambda,p} z_{k,\lambda,q}.
\]

For \( 1 \leq p \leq c_1 \) and \( 1 \leq q \leq c_2 \) we see that \( E_\gamma \left[ B_{p,q} | Y \right] \) is equal to

\[
\sum_{k=1}^{m} \sum_{l=1, l \neq k}^{m} \left( P_k P_l \lambda_k \lambda_l + P_k T_{l,\beta} \lambda_k \lambda_l - U_k P_l \lambda_l - (P_k + T_k \beta) U_k \lambda_l + U_k D_l \right) z_{k,\lambda,p} z_{l,\mu,q}
+ \sum_{k=1}^{m} \left( - E_\gamma \left[ R_k \right] Y \right) \lambda_k \mu_k - \beta T_k \lambda_k \mu_k - E_\gamma \left[ R_k N_k^- | Y \right) \lambda_k
- \beta T_k \lambda_k E_\gamma \left[ N_k^- | Y \right) - E_\gamma \left[ N_k^+ R_k | Y \right] \mu_k + E_\gamma \left[ N_k^+ N_k^- | Y \right) \right) z_{k,\lambda,p} z_{k,\mu,q}.
\]

For \( 1 \leq p, q \leq c_2 \) we see that \( E_\gamma \left[ C_{p,q} | Y \right] \) is equal to

\[
\sum_{k=1}^{m} \sum_{l=1, l \neq k}^{m} \left( P_k P_l \mu_k \mu_l - D_k P_l \mu_l - D_l P_k \mu_k + D_k D_l \right) z_{k,\mu,p} z_{l,\mu,q}
+ \sum_{k=1}^{m} \left( \mu_k^2 E_\gamma \left[ R_k^2 | Y \right) - 2 \mu_k E_\gamma \left[ N_k^- R_k | Y \right] + E_\gamma \left[ N_k^- N_k^- | Y \right) \right) z_{k,\mu,p} z_{k,\mu,q}.
\]

For the Hessian term, since for \( 1 \leq j \leq c_1 \) and \( 1 \leq k \leq c_2 \), \( \frac{\partial^2}{\partial \gamma_j \gamma_k} l_c (X; \gamma) = 0 \), we see that

\[
E_\gamma \left[ - \frac{\partial^2}{\partial \gamma_j \gamma_k} l_c (X; \gamma) \right] \text{ equals}
\]

\[
(S-9) \quad \left( \begin{array}{cc}
Z_\lambda & 0 \\
0 & Z_\mu
\end{array} \right) \text{diag}(P + \beta^T) \text{diag}(\lambda) Z_\lambda
- \left( \begin{array}{cc}
Z_\mu & 0 \\
0 & Z_\mu
\end{array} \right) \text{diag}(\mu) Z_\mu
\].

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, \( H_i(u, v, w, s, t) = \mathbb{E} \left( u^{N_t^+} v^{N_t^-} e^{-w R_t s X_t} | X_0 = i \right) \), for the death-immigration model is

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

Suppose we are interested in computing \( \mathbb{E} \left( N_t^+ 1_{\{X_t = j\}} | X_0 = i \right) \). 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:

\[
\mathbb{E} \left( N_t^+ 1_{\{X_t = j\}} | X_0 = i \right) = \left. \frac{\partial}{\partial u} \frac{\partial^j}{\partial s^j} H_i(u, 1, 0, s, t) \right|_{u = 1, s = 0}
\]

and

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

where

\[
A = 1 - e^{-\mu t}, \quad B = e^{-\mu t}, \\
C(u) = -\frac{\nu u (e^{\mu t} - 1)}{\mu}, \\
D(u) = \frac{\nu u (e^{\mu t} - 1)}{\mu} + \nu (u - 1)t.
\]

Therefore,

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

where

\[
C'(1) = -\frac{\nu (e^{\mu t} - 1)}{\mu}, \quad D'(1) = \frac{\nu (e^{\mu t} - 1)}{\mu} + \nu t.
\]
Now, the derivatives with respect to $s$ can be recovered by expanding $\frac{\partial}{\partial u} H_i(u, 1, 0, s, t)\bigg|_{u=1}$ into a power series:

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

$$\times \left[ \sum_{m=0}^{i} \binom{i}{m} A^{i-m} B^m s^m \right] \left[ \sum_{k=0}^{\infty} \frac{C_k}{k!} s^k \right]$$

$$= e^{D(1)} \left\{ \sum_{n=0}^{\infty} e^{D(1)} \left[ \sum_{k=0}^{n} \frac{C_k}{k!} \right] \right\} 1$$

$$= e^{D(1)} \sum_{n=0}^{\infty} e^{D(1)} \left[ \sum_{k=0}^{n} \frac{C_k}{k!} \right]$$

Therefore,

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

$$+ D'(1) \binom{i}{j-k} A^{i-j+k} B^{j-k} 1_{\{j-k \leq i\}}$$

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/TkIfIndelModelPathSummaries), 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. $E(N_t^+1_{\{X_t=j\}} | X_0 = 0)$, $E(N_t^-1_{\{X_t=j\}} | X_0 = 0)$, and $E(R_t1_{\{X_t=j\}} | X_0 = 0)$ when $\nu = \lambda$,

2. $E(N_t^+1_{\{X_t=j\}} | X_0 = 1)$, $E(N_t^-1_{\{X_t=j\}} | X_0 = 1)$, and $E(R_t1_{\{X_t=j\}} | X_0 = 1)$ when $\nu = 0$.

Notice that in the the sequence alignment BDI model, $X_t$ is a BDI 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 $E(N_t^+1_{\{X_t=j\}} | X_0 = 0)$ ( $\nu = \lambda$) and $E(N_t^+1_{\{X_t=j\}} | X_0 = 1)$ ( $\nu = 0$). The other expectations can be derived analogously.

1. **Objective:** $E(N_t^+1_{\{X_t=j\}} | X_0 = 0)$ ($\nu = \lambda$):

   First,

   $$E(N_t^+1_{\{X_t=j\}} | X_0 = 0) = \frac{1}{j!} \frac{\partial}{\partial r} \frac{\partial}{\partial s^j} H_0^+(r, s, t) \bigg|_{s=0, r=1},$$
where
\[ H_0^+(r, s, t) = H_0(r, 1, 0, s, t) = \frac{(\alpha_1 - \alpha_2) e^{-\lambda(1-ra_1)t}}{s - \alpha_2 - (s - \alpha_1) e^{-\lambda(\alpha_2-\alpha_1)rt}} \text{ and } \alpha_{1,2} = \frac{\lambda + \mu + \sqrt{(\lambda + \mu)^2 - 4\lambda\mu r}}{2\lambda r}. \]

We find the formula for this partial derivative by explicit differentiation:
\[
\frac{\partial^j}{\partial s^j} H_0^+(r, s, t) = \frac{(-1)^j j! (\alpha_1 - \alpha_2) e^{-\lambda(1-ra_1)t} \left(1 - e^{-\lambda(\alpha_2-\alpha_1)rt}\right)^j}{(s - \alpha_2 - (s - \alpha_1) e^{-\lambda(\alpha_2-\alpha_1)rt})^{j+1}} A(r)
\]
\[
\left. \frac{\partial^j}{\partial s^j} H_0^+(r, s, t) \right|_{s=0} = \frac{(-1)^j j! (\alpha_1 - \alpha_2) e^{-\lambda(1-ra_1)t} \left(1 - e^{-\lambda(\alpha_2-\alpha_1)rt}\right)^j}{(\alpha_1 e^{-\lambda(\alpha_2-\alpha_1)rt} - \alpha_2)^{j+1}} B(r)
\]
\[
\frac{\partial}{\partial r} \frac{\partial^j}{\partial s^j} H_0^+(r, s, t) \bigg|_{s=0, r=1} = \frac{A'(1)B(1) - A(1)B'(1)}{B^2(1)},
\]
where
\[
A(1) = \left(1 - e^{(\lambda-\mu)t}\right)^j \left(1 - \frac{\mu}{\lambda}\right),
\]
\[
B(1) = \left(e^{(\lambda-\mu)t} - \frac{\mu}{\lambda}\right)^{j+1},
\]
\[
A'(1) = \left(1 - e^{(\lambda-\mu)t}\right)^{j-1} \left[j2\mu e^{(\lambda-\mu)t} + \left(1 - e^{(\lambda-\mu)t}\right) \left(\frac{\lambda^2 + \mu^2}{\lambda(\mu - \lambda)} - \mu t\right)\right],
\]
\[
B'(1) = (j + 1) \left(e^{(\lambda-\mu)t} - \frac{\mu}{\lambda}\right)^j \left(\frac{\lambda(1 + 2\mu t)}{\mu(\mu - \lambda)} e^{(\lambda-\mu)t} + \frac{\mu^2}{\lambda(\mu - \lambda)}\right).
\]

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

As before,
\[
E\left( N_t^+1_{\{X_t=j\}} | X_0 = 1 \right) = \frac{1}{j!} \frac{\partial}{\partial r} \frac{\partial^j}{\partial s^j} H_1^+(r, s, t) \bigg|_{s=0, r=1},
\]
where
\[
H_1^+(r, s, t) = H_1(r, 1, 0, s, t) = \frac{\alpha_1(s - \alpha_2) - \alpha_2(s - \alpha_1) e^{-\lambda(\alpha_2-\alpha_1)rt}}{s - \alpha_2 - (s - \alpha_1) e^{-\lambda(\alpha_2-\alpha_1)rt}} = \alpha_2 + \frac{\alpha_1 - \alpha_2}{1 - \left(\frac{s-\alpha_1}{s-\alpha_2}\right) e^{-\lambda(\alpha_2-\alpha_1)rt}}
\]

For \( j = 0 \), we just need to plug in \( s = 0 \):
\[
H_1^+(r, 0, t) = \alpha_2 + \frac{C(r)\alpha_2\alpha_1 - \alpha_2^2}{\alpha_2 - \alpha_1 e^{-\lambda(\alpha_2-\alpha_1)rt} D(r)} = \alpha_2 + \frac{C(r)}{D(r)}.
\]
Then
\[ \frac{d}{dr} \mathbf{H}_1(r, 0, t) \bigg|_{r=1} = -\frac{\mu^2}{\lambda(\mu - \lambda)} + \frac{C'(1)D(1) - C(1)D'(1)}{D^2(1)}, \]
where
\begin{align*}
C(1) &= \frac{\mu}{\lambda} \left(1 - \frac{\mu}{\lambda}\right), \\
D(1) &= \frac{\mu}{\lambda} - e^{(\lambda-\mu)t}, \\
C'(1) &= \frac{\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}(1 + 2\mu t)\right].
\end{align*}

For \( j > 0 \)
\[ \frac{\partial^j}{\partial s^j} \mathbf{H}_1(r, s, t) = (-1)^{j+1} j! \left[ \begin{array}{c}
\frac{E(r)}{\alpha_1 e^{\lambda(\alpha_2-\alpha_1)rt} - \alpha_2} \\
\frac{G(r)}{\alpha_1 e^{\lambda(\alpha_2-\alpha_1)rt} - \alpha_2}
\end{array} \right]. \]

Then
\[ \frac{\partial}{\partial r} \frac{\partial^j}{\partial s^j} \mathbf{H}_1(r, s, t) \bigg|_{s=0, r=1} = (-1)^{j+1} j! \left[ \begin{array}{c}
\frac{E'(1)F(1) - E(1)F'(1)}{F^2(1)} \\
\frac{G'(1)I(1) - G(1)I'(1)}{I^2(1)}
\end{array} \right], \]
where
\begin{align*}
(S-10) \quad E(1) &= \frac{\mu}{\lambda} \left(1 - \frac{\mu}{\lambda}\right) \left(1 - e^{(\lambda-\mu)t}\right)^j, \\
(S-11) \quad F(1) &= \left(e^{(\lambda-\mu)t} - \frac{\mu}{\lambda}\right)^{j+1}, \\
(S-12) \quad G(1) &= \left(1 - \frac{\mu}{\lambda}\right) \left(1 - e^{(\lambda-\mu)t}\right)^{j-1}, \\
(S-13) \quad I(1) &= \left(e^{(\lambda-\mu)t} - \frac{\mu}{\lambda}\right)^j, \\
(S-14) \quad 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}\right], \\
(S-15) \quad 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], \\
(S-16) \quad 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], \\
(S-17) \quad 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].
\end{align*}
References.


