Web-based Supplementary Materials for

Likelihood-Based Inference for Discretely Observed Birth-Shift-Death and Multi-Type Branching Processes

By Jason Xu, Peter Guttorp, Midori Kato-Maeda, Vladimir Minin

Appendix A

Here we derive and solve the Kolmogorov backward equations of the two-type branching process necessary for evaluating the generating functions whose coefficients yield transition probabilities. See [Bailey, 1990] for an exposition on this solution technique.

Our two-type branching process is represent by a vector $(X_1(t), X_2(t))$ that denotes the numbers of particles of two types at time t. Recall the quantities $a_1(k,l)$, the rates of producing k type 1 particles and l type 2 particles, starting with one type 1 particle, and $a_2(k,l)$, analogously defined but beginning with one type 2 particle. Then we may introduce respective pseudo-generating functions $u_i(s_1, s_2) = \sum_k \sum_l a_i(k,l) s_1^k s_2^l$ for $i = 1, 2$, and the probability generating functions can be expressed

$$\phi_{10}(t, s_1, s_2) = E\left[s_1^{X_1(t)} s_2^{X_2(t)} \mid X_1(0) = 1, X_2(0) = 0\right] = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} P_{(1,0),(k,l)}(t) s_1^k s_2^l$$
$$= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \left[\mathbf{1}_{k=1,l=0} + a_1(k,l)t + o(t)\right] s_1^k s_2^l = s_1 + u_1(s_1, s_2)t + o(t). \tag{A-1}$$

An analogous expression for $\phi_{01}(t, s_1, s_2)$ is obtained similarly. For short, we write $\phi_{10} := \phi_1, \phi_{01} := \phi_2$, and thus we have the following relations between ϕ and u

$$\frac{d\phi_1(t, s_1, s_2)}{dt}\Big|_{t=0} = u_1(s_1, s_2), \qquad \frac{d\phi_2(t, s_1, s_2)}{dt}\Big|_{t=0} = u_2(s_1, s_2).$$

By particle independence, $\phi_{i,j} = \phi_1^i \phi_2^j$, so it suffices to work with only ϕ_1, ϕ_2. We now derive the backward equations for ϕ_1 and ϕ_2. Chapman-Kolmogorov equations yield the symmetric relations

$$\phi_1(t + h, s_1, s_2) = \phi_1(t, \phi_1(h, s_1, s_2), \phi_2(h, s_1, s_2)) \tag{A-2}$$
$$= \phi_1(h, \phi_1(t, s_1, s_2), \phi_2(t, s_1, s_2)). \tag{A-3}$$

To derive the backward equations, we begin by expanding $\phi_1(t + h, s_1, s_2)$ around t and applying

(A-3):

$$\begin{aligned}\phi_1(t+h, s_1, s_2) &= \phi_1(t, s_1, s_2) + \frac{d\phi_1(t+h, s_1, s_2)}{dh}\Big|_{h=0} h + o(h) \\ &= \phi_1(t, s_1, s_2) + \frac{d\phi_1(h, \phi_1(t, s_1, s_2), \phi_2(t, s_1, s_2))}{dh}\Big|_{h=0} h + o(h) \\ &= \phi_1(t, s_1, s_2) + u_1(\phi_1(t, s_1, s_2), \phi_2(t, s_1, s_2))h + o(h).\end{aligned}$$

Since an analogous argument applies for ϕ_2, we arrive at the system

$$\begin{cases} \frac{d}{dt}\phi_1(t, s_1, s_2) = u_1(\phi_1(t, s_1, s_2), \phi_2(t, s_1, s_2)), \\ \frac{d}{dt}\phi_2(t, s_1, s_2) = u_2(\phi_1(t, s_1, s_2), \phi_2(t, s_1, s_2)), \end{cases}$$

subject to initial conditions $\phi_1(0, s_1, s_2) = s_1, \phi_2(0, s_1, s_2) = s_2$.

We now substitute the rates specific to our birth-shift-death model into this general form: recall the rates defining the two-type branching process formulation presented in Section 2.4 of the main paper are

$$\begin{aligned} &a_1(1,1) = \lambda, \quad &&a_1(0,1) = \nu, \quad &&a_1(0,0) = \mu \quad &&a_1(1,0) = -(\lambda + \nu + \mu), \\ &a_2(0,2) = \lambda, \quad &&a_2(0,1) = -(\lambda + \mu), \quad &&a_2(0,0) = \mu, && \end{aligned} \tag{A-4}$$

so that the pseudo-generating functions and backward equations are

$$\begin{cases} u_1(s_1, s_2) = \lambda s_1 s_2 + \nu s_2 + \mu - (\lambda + \nu + \mu)s_1, \qquad & \frac{d}{dt}\phi_1 = \lambda\phi_1\phi_2 + \nu\phi_2 + \mu - (\lambda + \nu + \mu)s_1, \\ u_2(s_1, s_2) = \lambda s_2^2 - (\lambda + \mu)s_2 + \mu, & \frac{d}{dt}\phi_2 = \lambda\phi_2^2 - (\lambda + \mu)\phi_2 + \mu. \end{cases} \tag{A-5}$$

Upon rearranging, the expression for ϕ_2 becomes a Ricatti equation

$$\phi_2' - \lambda\phi_2^2 + (\lambda + \mu)\phi_2 = \mu,$$

and the constant solutions $\phi_2 = 1, \mu/\lambda$ are both particular solutions. Using the simpler root $\phi_2 = 1$, we can reduce the above Ricatti equation to a linear ODE by making a substitution $z = \frac{1}{\phi_2 - 1}$, so that $\phi_2 = 1 + \frac{1}{z}$:

$$\begin{aligned}\phi_2' = -\frac{z'}{z^2} = \mu - (\lambda + \mu)(\frac{1}{z} + 1) + \lambda(1 + \frac{1}{z})^2 \qquad &= \mu - \frac{\lambda + \mu}{z} - (\lambda + \mu) + \lambda(\frac{1}{z^2} + \frac{2}{z} + 1) \\ = -\frac{\mu - \lambda}{z} + \frac{\lambda}{z^2}.\end{aligned}$$

Multiplying through by $-z^2$ and rearranging, we arrive at a linear equation that is easily solved via the integrating factor method:

$$z' + (\lambda - \mu)z = -\lambda \Rightarrow z = -\frac{\lambda}{\lambda - \mu} + Ce^{-(\lambda - \mu)t}.$$

Substituting ϕ_2 back into the expression, we obtain

$$\phi_2 = 1 + \frac{1}{\frac{\lambda}{\mu-\lambda} + Ce^{(\mu-\lambda)t}},$$

and plugging in the initial condition $\phi_2(0, s_1, s_2) = s_2$, we see $C = \frac{1}{s_2-1} + \frac{\lambda}{\lambda-\mu}$. Thus, we arrive at the closed form solution

$$\phi_2(t, s_1, s_2) = 1 + \left[\frac{\lambda}{\mu-\lambda} + (\frac{1}{s_2-1} + \frac{\lambda}{\lambda-\mu})e^{(\mu-\lambda)t}\right]^{-1} := g(t, s_1, s_2) \tag{A-6}$$

We can now plug this solution into the ODE for ϕ_1 to obtain

$$\frac{d}{dt}\phi_1 + (\lambda + \nu + \mu - \lambda g)\phi_1 = \nu g + \mu. \tag{A-7}$$

Closed form solution for ϕ_1

Equation (A-7) is linear with variable coefficients, and can again be solved by multiplying by an integrating factor. If we define the integrating factor $\psi := \exp\left[\int (\lambda + \nu + \mu - \lambda g)dt\right]$, then

$$\frac{d}{dt}(\phi_1\psi) = \psi(\nu g + \mu),$$

and after integration and rearranging,

$$\phi_1 = \psi^{-1}\left[\int \psi(\nu g + \mu)dt + C\right]. \tag{A-8}$$

After further simplification, we may write

$$\psi = e^{(\nu+\mu)t}(\lambda s_2 - \mu) + \lambda e^{(\lambda+\nu)t}(1 - s_2),$$

and the integrand becomes

$$\psi(\nu g + \mu) = (\nu + \mu)\psi + \frac{\nu\psi}{\frac{\lambda}{\mu-\lambda} + (\frac{1}{s_2-1} + \frac{\lambda}{\lambda-\mu})e^{(\mu-\lambda)t}}. \tag{A-9}$$

Integrating (A-9) and plugging into (A-8) with initial condition $\phi_1(0, s_1, s_2) = s_1$, we ultimately obtain a closed form expression

$$
\begin{aligned}
\phi_1(t,s_1,s_2) = \Big[& e^{(\nu+\mu)t}(\lambda s_2-\mu)+\lambda e^{(\lambda+\nu)t}(1-s_2)\Big]^{-1} \\
& \cdot \Bigg\{ \nu(\mu-\lambda)e^{\nu t}\Bigg[\frac{e^{\mu t}(\lambda s_2-\mu){}_2F_1\Big(1,\frac{\mu+\nu}{\mu-\lambda},\frac{\lambda-2\mu-\nu}{\lambda-\mu},\frac{e^{(\mu-\lambda)t}(\lambda s_2-\mu)}{\lambda(s_2-1)}\Big)}{\lambda(\mu+\nu)} \\
& + \frac{e^{\lambda t}(1-s_2){}_2F_1\Big(1,\frac{\lambda+\nu}{\mu-\lambda},\frac{\mu+\nu}{\mu-\lambda},\frac{e^{(\mu-\lambda)t}(\lambda s_2-\mu)}{\lambda(s_2-1)}\Big)}{\lambda+\nu}\Bigg] + (\lambda s_2-\mu)e^{(\mu+\nu)t} \\
& + \frac{\lambda(\nu+\mu)(1-s_2)e^{(\lambda+\nu)t}}{\lambda+\nu} + \mu + s_1(\lambda-\mu) - \lambda s_2 + \frac{\lambda(s_2-1)(\nu+\mu)}{\lambda+\nu} \\
& + \nu(\lambda-\mu)\Bigg[\frac{\lambda s_2-\mu}{\lambda(\mu+\nu)}{}_2F_1\Big(1,\frac{\mu+\nu}{\mu-\lambda},\frac{\lambda-2\mu-\nu}{\lambda-\mu},\frac{\lambda s_2-\mu}{\lambda(s_2-1)}\Big) \\
& + \frac{1-s_2}{\lambda+\nu}{}_2F_1\Big(1,\frac{\lambda+\nu}{\mu-\lambda},\frac{\mu+\nu}{\mu-\lambda},\frac{\lambda s_2-\mu}{\lambda(s_2-1)}\Big)\Bigg]\Bigg\},
\end{aligned}
\tag{A-10}
$$

where ${}_2F_1$ indicates the hypergeometric function. In practice, we solve for ϕ_1 numerically rather than using this closed form solution: evaluating (A-7) via Runge-Kutta methods proves more stable than evaluation of the hypergeometric functions arising in (A-10), and only requires numerically solving a single linear ordinary differential equation.

Appendix B

Here we derive the equations in our main theorem. The formulation is repeated below:

Theorem 1 *Let $\{X_t\}$ be a two-type branching defined by the rates in equation* (A-4)*. Denote particle time and the number of births, shifts, and deaths over the interval $[0,t)$ by R_t, b_t, f_p, and d_t respectively. Define the generating functions corresponding to births as*

$$
\begin{aligned}
H_1^+(r,s_1,s_2,t) &= E\left[r^{b_t}s_1^{X_1(t)}s_2^{X_2(t)} \mid \mathbf{X}(0)=(1,0)\right] \quad \textit{and} \\
H_2^+(r,s_1,s_2,t) &= E\left[r^{b_t}s_1^{X_1(t)}s_2^{X_2(t)} \mid \mathbf{X}(0)=(0,1)\right].
\end{aligned}
$$

Then

$$
H_2^+ = y_b + \left[\frac{-\lambda r}{2\lambda r y_b-\lambda-\mu} + \left(\frac{1}{s_2-y_b}+\frac{\lambda r}{2\lambda r y_b-\lambda-\mu}\right)e^{-(2y_b\lambda r-\lambda-\mu)t}\right]^{-1},
$$

where $y_b = (\lambda+\mu+\sqrt{\lambda^2+2\lambda\mu+\mu^2-4\lambda\mu r})/(2\lambda r)$, and H_1^+ satisfies the following differential equation:

$$
\frac{d}{dt}H_1^+(t,s_1,s_2,r) = \lambda r H_1^+H_2^+ + \nu H_2^+ + \mu - (\lambda+\mu+\nu)H_1^+, \tag{B-1}
$$

subject to initial condition $H_1(r,s_1,s_2,0)=s_1$.

The analogous generating functions for shifts, deaths, and particle time satisfy the following equations:

$$H_2^-(t,s_1,s_2,r) = y_d + \left[\frac{-\lambda}{2\lambda y_d - \lambda - \mu} + \left(\frac{1}{s_2 - y_d} + \frac{\lambda}{2\lambda y_d - \lambda - \mu}\right) e^{-(2y_d\lambda - \lambda - \mu)t}\right]^{-1},$$
$$H_2^{\rightarrow}(t,s_1,s_2,r) = 1 + \left[\frac{\lambda}{\mu - \lambda} + (\frac{1}{s_2 - 1} + \frac{\lambda}{\lambda - \mu})e^{(\mu-\lambda)t}\right]^{-1},$$
$$H_2^*(t,s_1,s_2,r) = y_* + \left[\frac{-\lambda}{2\lambda y_* - \lambda - \mu - r} + \left(\frac{1}{s_2 - y_*} + \frac{\lambda}{2\lambda y_* - \lambda - \mu - r}\right) e^{-(2y_*\lambda - \lambda - \mu - r)t}\right]^{-1},$$

$$\frac{d}{dt}H_1^-(t,s_1,s_2,r) = \lambda H_1^- H_2^- + \nu H_2^- + \mu r - (\lambda + \mu + \nu)H_1^-,$$
$$\frac{d}{dt}H_1^{\rightarrow}(t,s_1,s_2,r) = \lambda H_1^{\rightarrow} H_2^{\rightarrow} + \nu r H_2^{\rightarrow} + \mu - (\lambda + \mu + \nu)H_1^{\rightarrow},$$
$$\frac{d}{dt}H_1^*(t,s_1,s_2,r) = \lambda H_1^* H_2^* + \nu H_2^* + \mu - (\lambda + \mu + \nu + r)H_1^*,$$

where $y_d = (\lambda + \mu + \sqrt{\lambda^2 + 2\lambda\mu + \mu^2 - 4\lambda\mu r})/(2\lambda)$, $y_* = (\lambda + \mu + r + \sqrt{(\lambda + \mu + r)^2 - 4\lambda\mu})/(2\lambda)$ *and* $H_1^-(r,s_1,s_2,0) = H_1^{\rightarrow}(r,s_1,s_2,0) = H_1^*(r,s_1,s_2,0) = s_1$.

Proof Begin by expanding

$$H_{10}^+(t,r,s_1,s_2) = \sum_n \sum_k \sum_l Pr(b_t = n, x_t = (k,l) | x_0 = (1,0)) s_1^k s_2^l r^n.$$

Recall the jump rates of the process in equation (A-4): a_1 correspond to the process beginning with 1 type one particle, and a_2 are jump rates starting with 1 type two particle. We can express the probability terms in H_{10}^+ using the same type of first-order decomposition as in equation (A-1); for instance, in the event of a birth,

$$Pr(b_t = 1, x_t = (1,1) | x_0 = (1,0)) = a_1(1,1) + o(t) = \lambda + o(t)$$

and for other values of $n > 1$,

$$Pr(b_t = n, x_t = (1,1) | x_0 = (1,0)) = o(t).$$

In the case of a shift,

$$Pr(b_t = 0, x_t = (0,1) | x_0 = (1,0)) = a_1(0,1) + o(t) = \nu + o(t)$$

and for other values of $n \neq 0$,

$$Pr(b_t = n, x_t = (0,1) | x_0 = (1,0)) = o(t).$$

We see that the r^n term in the series H^+_{10} is either $r^1 = r$ if exactly one birth occurs, or $r^0 = 1$ as other powers correspond to more than one event and are absorbed into the $o(t)$ term. Thus,

$$H^+_{10}(t, r, s_1, s_2) = \sum_k \sum_l g_{10,kl}(r, t) s_1^k s_2^l = \sum_n \sum_k \sum_l Pr(b_t = n, x_t = (k, l)|x_0 = (1, 0)) s_1^k s_2^l r^n$$
$$= s_1 + \lambda s_1 s_2 r + \nu s_2 + \mu - (\lambda + \nu + \mu)s_1 + o(t) := s_1 + u_1^b(s_1, s_2)t + o(t)$$

with u_1^b denoting the pseudo-generating function, similarly to (A-1). With an analogous derivation for u_2^b, we arrive at the system

$$\begin{cases} u_1^b(s_1, s_2) = \lambda r s_1 s_2 + \nu s_2 + \mu - (\lambda + \nu + \mu)s_1 \\ u_2^b(s_1, s_2) = \lambda r s_2^2 - (\lambda + \mu)s_2 + \mu, \end{cases} \tag{B-2}$$

and since

$$\frac{dH^+_{10}(t, r, s_1, s_2)}{dt}\bigg|_{t=0} = u_1^b(s1, s2, r), \qquad \frac{dH^+_{01}(t, r, s_1, s_2)}{dt}\bigg|_{t=0} = u_2^b(s1, s2, r),$$

we obtain the backward equations system

$$\begin{cases} \frac{d}{dt} H^+_{10}(t, s_1, s_2, r) = u_1^b(H^+_{10}(t, s_1, s_2, r), H^+_{01}(t, s_1, s_2, r)), \\ \frac{d}{dt} H^+_{01}(t, s_1, s_2, r) = u_2^b(H^+_{10}(t, s_1, s_2, r), H^+_{01}(t, s_1, s_2, r)) \end{cases} \tag{B-3}$$

by the same Chapman-Kolmogorov argument used for transition probabilities, subject to initial conditions $H_{10}(t = 0, s_1, s_2, r) = s_1$ and $H_{01}(t = 0, s_1, s_2, r) = s_2$. The systems for deaths and shifts are derived analogously beginning with this first-order expansion technique, and are respectively given by

$$\begin{cases} u_1^d(s_1, s_2) = \lambda s_1 s_2 + \nu s_2 + r\mu - (\lambda + \nu + \mu)s_1 \\ u_2^d(s_1, s_2) = \lambda s_2^2 - (\lambda + \mu)s_2 + r\mu, \end{cases} \tag{B-4}$$

$$\begin{cases} u_1^{\rightarrow}(s_1, s_2) = \lambda s_1 s_2 + r\nu s_2 + \mu - (\lambda + \nu + \mu)s_1, \\ u_2^{\rightarrow}(s_1, s_2) = \lambda s_2^2 - (\lambda + \mu)s_2 + \mu. \end{cases} \tag{B-5}$$

To derive the system governing the particle time generating function, recall the quantity $q^*_{ij,kl}(x; t) := Pr(R_t \le x, X(t) = (k, l)|X(0) = (i, j))$, and consider its Laplace-Stieltjes transform

$$V_{ij,kl}(r; t) = \int_0^\infty e^{-rx} \mathrm{d}q^*_{ij,kl}(x; t). \tag{B-6}$$

The Laplace-Stieltjes transform of such a probability distribution corresponding to a *reward function*, where a_{ij} is the reward accrued per unit time spent in state (i,j), satisfies the forward equation

$$\frac{d}{dt}V_{ij,kl}(r;t) = -a_{ij}rV_{ij,kl}(r;t) + \sum_{m=1}^{K}\sum_{n=1}^{K} Q_{ij,mn}V_{ij,kl}(r;t), \tag{B-7}$$

where $\mathbf{Q}$ is the infinitesimal generator of the Markov chain, with finite or countable number of rows and columns and entries $Q_{ij,kl}$ the instantaneous rates of transitioning from state (i,j) to (k,l), and $Q_{ij,ij} = -\sum_{m,n\neq i,j} Q_{ij,mn}$. Following Neuts [Neuts, 1995], we derive the following integral equation:

$$q^*_{ij,kl}(x,t) = \mathbf{1}_{\{ij=kl\}}\mathbf{1}_{\{x\geq a_{ij}t\}}e^{Q_{ij,ij}t} + \sum_{m,n\neq i,j}\int_0^t e^{Q_{ij,ij}u}Q_{ij,mn}q^*_{mn.kl}(x-a_{ij}u,t-u)\mathrm{d}u.$$

Taking the Laplace transform of both sides and denoting $\widetilde{V}_{ij,kl}(r;t) = \int_0^\infty e^{-rx}q^*_{ij,kl}(x;t)\mathrm{d}x$, we obtain

$$\widetilde{V}_{ij,kl}(r,t) = \mathbf{1}_{\{ij=kl\}}r^{-1}\exp\left[(Q_{ij,ij}-a_{ij})t\right] + \sum_{m,n\neq i,j}\int_0^t e^{Q_{ij,ij}u}Q_{ij,mn}\mathrm{d}u\int_{a_{ij}u}^\infty e^{-rx}q^*_{mn,kl}(x-a_{ij}u;t-u)\mathrm{d}x.$$

Making a change of variables $y = x - a_{ij}u$ in the rightmost integral and multiplying both sides by $\exp\left[-(Q_{ij,ij}-a_{ij})t\right]$ yields

$$\exp\left[-(Q_{ij,ij}-a_{ij})t\right]\widetilde{V}_{ij,kl}(r,t) = \frac{1}{r} + \sum_{m,n\neq i,j}\int_0^t \exp\left[-(Q_{ij,ij}-a_{ij})(t-u)\right]Q_{ij,mn}\widetilde{V}_{mn,kl}(r;t-u).$$

Next, make another substitution $v = t-u$ and simplify after differentiating the above equation with respect to t: we arrive at

$$\frac{\partial}{\partial t}\widetilde{V}_{ij,kl}(r;t) = -a_{ij}r\widetilde{V}_{ij,kl}(r;t) + \sum_{m=1}^{K}\sum_{n=1}^{K} Q_{ij,mn}\widetilde{V}_{mn,kl}(r;t).$$

Equation (B-7) then follows from $V_{ij,kl}(r;t) = s\widetilde{V}_{ij,kl}(r;t)$, with $V_{ij,kl}(t)(r;0) = \mathbf{1}_{\{ij=kl\}}$.

The matrix $\mathbf{V}(r;t) := \{V_{ij,kl}(r;t)\}$ can therefore be written as a matrix exponential

$$\mathbf{V}(r;t) := \exp[\mathbf{Q} - \mathrm{diag}(\mathbf{a})r)t] := \exp(\widetilde{\mathbf{Q}}t), \tag{B-8}$$

where $\mathrm{diag}(\mathbf{a})$ is the diagonal matrix with diagonal entries a_{ij}. In our case, $a_{ij} = 1$, since we are interested in particle time and the “reward” that accumulates per unit of time is that quantity of time itself. Strictly speaking we don’t need infinite dimensional matrix algebra here, but we use it to simplify our notation.

Note the similarity of equation (B-8) to the matrix exponential corresponding to transition probabilities $\mathbf{P}(t) = \exp(\mathbf{Q}t)$: thus, the system of backward equations for $V_{ij,kl}$ are almost identical to those for transition probabilities $p_{ij,kl}$. The generators $\widetilde{\mathbf{Q}} \neq \mathbf{Q}$ differ only in diagonal entries: instantaneous rates of no event occurring are augmented by an extra r term $\widetilde{Q}_{ij,ij} = -\sum_{m,n\neq i,j} Q_{ij,mn} - r$. The system of backward equations is thus given by

$$\begin{cases} u_1^*(s_1, s_2) = \lambda s_1 s_2 + \nu s_2 + \mu - (\lambda + \nu + \mu + r)s_1, \\ u_2^*(s_1, s_2) = \lambda s_2^2 + \mu - (\lambda + \mu + r)s_2, \end{cases} \tag{B-9}$$

and as we have seen in the derivation for expected births in equation (B-3), this implies that the generating function

$$H_{10}^*(r, s_1, s_2, t) = \sum_k \sum_l \int_0^\infty e^{-rx} \mathrm{d}q_{ij,kl}^*(x; t) = \sum_k \sum_l V_{10,kl}(r, t) s_1^k s_2^l$$

also satisfies the same system.

Reducing the systems

Each of the four systems for births, shifts, deaths, and particle time can be reduced to a single ODE by first solving the second equation analytically. We demonstrate this in the case of the birth equations (B-3), and abbreviate $H_{10}^+ := H_1, H_{01}^+ := H_2$. Plugging (B-2) into (B-3),

$$\begin{cases} \frac{d}{dt} H_1(t, s_1, s_2, r) = \lambda r H_1 H_2 + \nu H_2 + \mu - (\lambda + \nu + \mu)H_1, \\ \frac{d}{dt} H_2(t, s_1, s_2, r) = \lambda r H_2^2 - (\lambda + \mu)H_2 + \mu. \end{cases}$$

The second equation is a Ricatti equation. To solve it, we first identify a constant solution

$$y_b = \frac{\lambda + \mu + \sqrt{\lambda^2 + 2\lambda\mu + \mu^2 - 4\lambda\mu r}}{2\lambda r}$$

obtained by setting

$$\frac{d}{dt} H_2 = 0 = \lambda r H_2^2 - (\lambda + \mu)H_2 + \mu.$$

Next, perform a change of variables $z = \frac{1}{H_2 - y_b}$ so that $H_2 = y_1 + \frac{1}{z}$, and thus

$$\frac{dz}{dt} + (2y_b\lambda r - \lambda - \mu)z = -\lambda r$$

Using the multiplier method with multiplier $\exp\{(2y_b\lambda r - \lambda - \mu)t\}$, we obtain

$$z = e^{-(2y_b\lambda r - \lambda - \mu)t} \left[\int -\lambda r e^{(2\lambda r y_b - \lambda - \mu)t} dt + C \right] = \frac{-\lambda r}{2\lambda r y_b - \lambda - \mu} + Ce^{-(2y_b\lambda r - \lambda - \mu)t}.$$

Thus,

$$H_2 = y_b + \frac{1}{z} = y_b + \left[\frac{-\lambda r}{2\lambda r y_b - \lambda - \mu} + Ce^{-(2y_b\lambda r - \lambda - \mu)t}\right]^{-1}$$

and from $H_2(0, r, s_1, s_2) = s_2$, we see $C = \frac{1}{s_2 - y_b} + \frac{\lambda r}{2\lambda r y_b - \lambda - \mu}$. Finally, we arrive at the full solution to the second ODE

$$H_2 := g^b(t, s_1, s_2, r) = y_b + \left[\frac{-\lambda r}{2\lambda r y_b - \lambda - \mu} + \left(\frac{1}{s_2 - y_b} + \frac{\lambda r}{2\lambda r y_b - \lambda - \mu}\right) e^{-(2y_b\lambda r - \lambda - \mu)t}\right]^{-1}.$$

Plugging this solution into the equation for H_1, we have a single ODE that is numerically solvable:

$$\frac{d}{dt} H_1^+(t, s_1, s_2, r) = \lambda r H_1^+ g^b + \nu g^b + \mu - (\lambda + \mu + \nu) H_1^+.$$

An analogous solution beginning with Equations (B-4), (B-5), and (B-9) instead of (B-3) and solving the second Ricatti equation is used to simplify the other equation systems, yielding the results presented in Theorem 1.

Appendix C

Here we include additional figures that support, but are not crucial to, illustrating our simulation results.

Figure C-1 displays the transition probabilities $p_{(10,0),(ij)}$ for 25 randomly sampled (i, j) pairs with $0 \leq i, j \leq 32$, calculated by our generating function approach alongside their Monte Carlo estimates and confidence intervals. Monte Carlo estimates are based on 5000 realizations beginning with an initial count of 10 with $dt = 1.0, \lambda = .5, \mu = .45$ and ν ranging from 0.3 to 2.0.

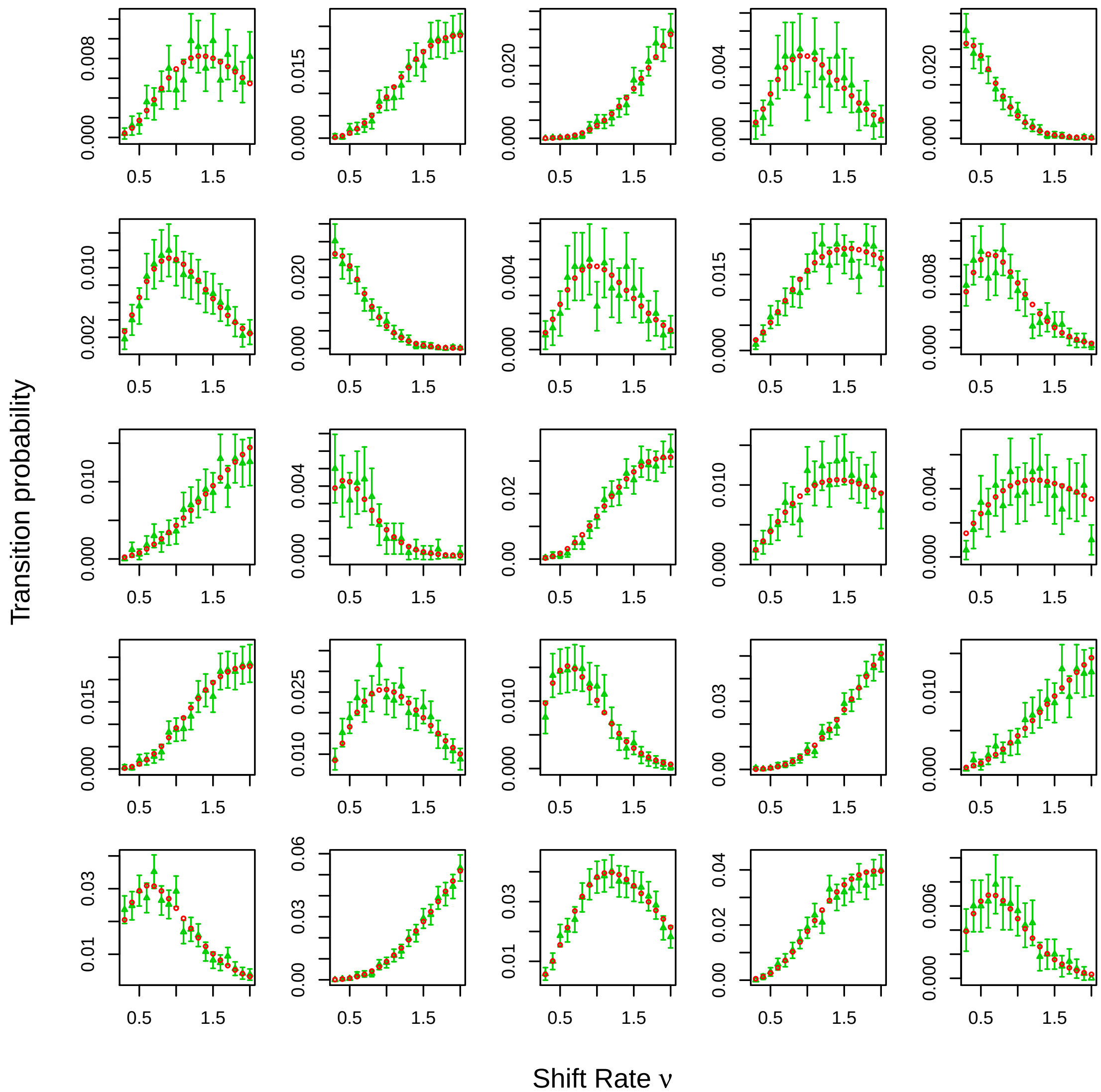

Figure C-1: Transition probabilities remain accurate when increasing all rates of process, presented over a wide range of ν values. Green points and intervals correspond to Monte Carlo estimates of transition probabilities and corresponding 95% confidence intervals. The red points denote probabilities computed with our generating function method.

Figure C-2 shows that restricted moment calculations performed during the E-step are indeed accurate: the following figure corresponds to simulations with 3 times the rates in the Rosenberg-Tanaka paper: $(\lambda, \nu, \mu) = 3 \cdot (.0188, .0026, .0147)$, with 10 initial particles and varying time lengths.

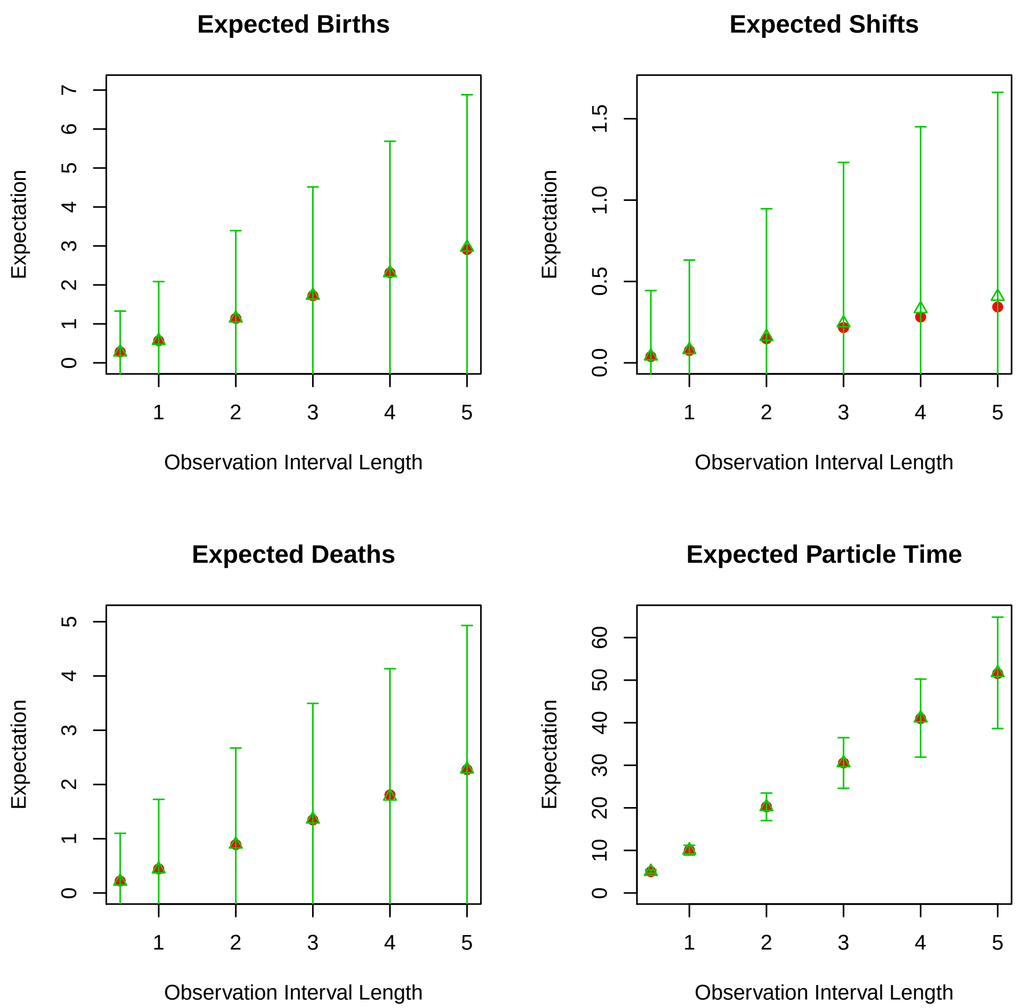

Figure C-2: Restricted moments calculated by our method (red) compared to approximation over 5000 Monte Carlo simulations and corresponding 95% confidence intervals (green).

Appendix D

The following table contains summary statistics for the San Francisco IS*6110* dataset.

Value	IS*6110* Data
Total Number of Observations	446
Total Number of Intervals	252
Average Interval Length	.35
Intervals with Deaths	13
Intervals with Shifts and Deaths	1
Intervals with Births	13
Intervals with Shifts and Births	1
Intervals with Only Shifts	3
Intervals with No Observed Change	221
Number of Individuals	196
Number of Individuals with EU lineage	109
Number of Individuals with EA lineage	54
Number of Individuals with IND lineage	25
Number of Individuals with HIV+	68
Number of Individuals with Drug Resistance	44

Table C-1: Summary statistics for *M. tuberculosis* IS*6110* dataset.

References

NTJ Bailey. The Elements of Stochastic Processes with Applications to the Natural Sciences, volume 25. John Wiley & Sons, 1990.

MF Neuts. Algorithmic Probability: A Collection of Problems, volume 3. CRC Press, 1995.