SI Text

MLE Estimates. The transition probability function, $P_{sf}(T) = P\{N(T) = f \mid N(0) = s\}$, that a pure birth process with birth intensities μ_0, μ_1 . . . presently in state *s* will be in state *f* after a fixed time interval *T* is given by (1)

$$P_{sf}(T) = \prod_{j=s}^{f-1} \mu_j \int\int_{\sum_{j=s}^{f} t_j = T} \exp\left(-\sum_{j=s}^{f} \mu_j t_j\right) dt_s \ldots dt_f \quad . \qquad \textbf{[7]}$$

The right-side integral evaluates to give

$$C_{sf}(\mu_s \ldots \mu_j, T) = \int\int_{\sum_{j=s}^{f} t_j = T} \exp\left(-\sum_{j=s}^{f} \mu_j t_j\right) dt_s \ldots dt_f = \sum_{k=s}^{f} \frac{\exp(-\mu_k T)}{\prod_{j \neq k} (\mu_j - \mu_k)} \quad . \qquad \textbf{[8]}$$

The intensity vector for the parametric model is substituted into Eq. **8** $\mu_j \rightarrow \pi_j$, and the log likelihood function for the basic model with *n* observed individuals is

$$\ln L = \sum_j a_j \ln(\pi_j) + \sum_{i=1}^{n} \ln\left(C_{s_i f_i}(\pi_{s_i} \ldots \pi_{f_i}, T)\right) \ , \qquad \textbf{[9]}$$

where a_j is the number of observed jumps from state $j = 0, 1. \ldots$

The partial deriaties of ln(*C*) are connected to the mean time the *i*th process spends in state *j*. This is seen from Eqs. **7** and **8**. Hence,

$$-\frac{\partial \ln\left(C_{s_i f_i}(\pi_{s_i} \ldots \pi_{f_i}, T)\right)}{\partial \pi_j} = \mathrm{E}\left[T_{ij} \mid N_i(0) = s_i, N_i(T) = f_i\right]; \quad s_i \le j \le f_i \quad . \qquad \textbf{[10]}$$

The maximum likelihood estimates of the model parameters $\pi(\theta) = \{\beta, \delta, \varepsilon\}$ are found from the simultaneous solutions to $\frac{\partial \ln \mathrm{L}}{\partial \pi_{\mathrm{j}}} = 0$:

$$
\begin{aligned}
a_0 &= \beta \sum_{i=1}^{n} E_{\pi(\beta,\delta,\varepsilon)}\left(T_{ij} \middle| N_i(0), N_i(T)\right) \\
\sum_{j\geq 1} a_j &= \varepsilon \sum_{j\geq 1} \sum_{i=1}^{n} E_{\pi(\beta,\delta,\varepsilon)}\left(T_{ij} \middle| N_i(0), N_i(T)\right) j^{\delta} \qquad \textbf{[11]} \\
\sum_{j\geq 1} a_j \ln(j) &= \sum_{j\geq 1} \sum_{i=1}^{n} E_{\pi(\beta,\delta,\varepsilon)}\left(T_{ij} \middle| N_i(0), N_i(T)\right) j^{\delta} \ln(j)
\end{aligned}
$$

The system of equations is solved by using the expectation maximization (EM) algorithm. Based on an initial guess of the parameter values, the expected interarrival times are calculated by using Eq. **10**. New parameter estimates are found by solving the update equations in Eq. **11**. The whole process is repeated until convergence. At each update, the δ parameter is found iteratively using a Newton–Raphson algorithm.

For the general model with birth intensities $\kappa\,\mu_0$, $\kappa\,\mu_1$ . . ., where κ is Gamma-distributed, the transition probability function is given by

$$
P_{sf}(T) = \prod_{j=s}^{f-1} \mu_j \int_0^{\infty} \left[\kappa^{f-s} \int\int_{\sum_{j=s}^{f} t_j = T} \exp\left(-\kappa \sum_{j=s}^{f} \mu_j t_j \right) dt_s \ldots dt_f \right] g_{\alpha}(\kappa) d\kappa \ . \qquad \textbf{[12]}
$$

The right-side integral evaluates to

$$D_{sf}(T) = \int_0^\infty \left[\kappa^{f-s} \int\int_{\sum_{j=s}^{f} t_j = T} \exp\left(-\kappa \sum_{j=s}^{f} \mu_j t_j \right) dt_s \ldots dt_f \right] g_\alpha(\kappa) d\kappa$$
$$= \sum_{k=s}^{f} \frac{(1+\mu_k T/\alpha)^{-\alpha}}{\left(\prod_{j \neq k} (\mu_j - \mu_k) \right)} \qquad \textbf{[13]}$$

For the random factor model the likelihood equation is obtained by replacing $C_{s_i f_i} \rightarrow D_{s_i f_i}$ in Eq. **9**. Here and in the following the parameter dependence of the functions C_{sf}, D_{sf} is omitted for simplicity. The expected interarrival times $\kappa_i T_{ij}$ are calculated from

$$-\frac{\partial \ln D_{s_i f_i}}{\partial \pi_j} = \mathrm{E}\left(\kappa_i T_{ij} \mid N_i(0) = s_i, N_i(T) = f_i \right); \quad s_i \leq j \leq f_i \qquad \textbf{[14]}$$

The MLE of $\pi(\Theta) = \{\alpha, \beta, \delta, \varepsilon\}$ is found by calculating the profile likelihood $\ln L_p(\alpha)$:

$$\ln L_p(\alpha) = \max_{(\beta,\delta,\varepsilon)} \ln L(\alpha, \beta, \delta, \varepsilon) \qquad \textbf{[15]}$$

For each value of α, the estimates for $(\beta, \delta, \varepsilon)$ are generated by using the MLE approach as described above. Convergence was obtained within 10 EM iterations and 30 iterations of the Newton–Rhapson algorithm. The confidence intervals for the parameters were found by parametric bootstrap and based on 1,000 simulations.

Numerical Approximation of Interarrival Times. Direct calculation of the interarrival times using Eqs. **10** and **14** may cause numerical problems as the expressions are evaluated as a sum of positive and negative numbers that may be quite large.

Alternatively, the functions C_{sf},D_{sf} can be approximated by Monte Carlo integration. Consider a random vector $(z_1 \ldots z_{v+1})$ uniformly distributed on $[0;1]^{v+1[}$, where $v = N(T) - N(0)$. Then, using the inverse transform method, exponential vectors we can use $t_i = -\ln(z_i) / \sum_{i=1}^{(v+1)} -\ln(z_i)$ to approximate C_{sf}, $\partial C_{sf} / \partial \pi_j$.

Repeating the procedure m times we have

$$\hat{C}_{sf} \approx \frac{1}{m}\sum_{k=1}^{m}\exp\left(-\sum_{i=1}^{v+1}\pi_{s-1+i}t_{ik}\right); \quad \frac{\partial \hat{C}_{sf}}{\partial \pi_j} \approx \frac{1}{m}\sum_{k=1}^{m}t_{jk}\exp\left(-\sum_{i=1}^{v+1}\pi_{s-1+i}t_{ik}\right) \qquad \textbf{[16]}$$

The interarrival times are estimated by replacing $-\partial \ln C_{sf} / \partial \pi_j$ with $-\left(\partial \hat{C}_{sf} / \hat{C}_{sf}\partial \pi_j\right)$ in Eq. **10**.

Likewise, the same type of vectors can approximate D_{rs} and $\partial D_{rs} / \partial \pi_j$:

$$\hat{D}_{rs} \approx \frac{1}{m}\sum_{k=1}^{m}\left(\frac{\Gamma(\alpha+\nu)\alpha^{\alpha}}{\Gamma(\alpha)\left(\alpha+\sum_{i=1}^{v+1}\pi_{r-1+i}t_{ik}\right)^{\alpha+\nu}}\right)$$

$$\frac{\partial \hat{D}_{rs}}{\partial \pi_j} \approx \frac{1}{m}\sum_{k=1}^{m}t_{jk}\left(\frac{\Gamma(\alpha+\nu+1)\alpha^{\alpha}}{\Gamma(\alpha)\left(\alpha+\sum_{i=1}^{v+1}\pi_{r-1+i}t_{ik}\right)^{\alpha+\nu+1}}\right) \qquad \textbf{[17]}$$

where Γ is the gamma function. As before, the interarrival times are approximated by ratios of $\hat{D}_{rs}$ and $\partial D_{rs} / \partial \pi_j$, *cf.* (Eq. **14**). The procedure is repeated for each individual with initial partner numbers $s_i \geq 7$–10, using a vector sample size of m = 1,000.

Goodness-of-Fit. The goodness-of-fit for the parametric models is assessed by deviance (likelihood ratio) statistics. Let n_s denote the number of persons with initially s partners,

$N_i(0) = s$, and n_{sf} is the number of persons that starts in state s and ends in state f, [$N_i(T) = f | N_i(0) = s$]. The deviance for the parametric models versus a saturated model (i.e., an exact fit of the observed processes) is

$$D = -2 \sum_{\forall n_{sf}>0} n_{sf} \left(\ln\left(P_{sf}\left(N\left(T\right) = f \mid N(0) = s \right)\right) - \ln\left(\frac{n_{sf}}{n_s} \right) \right) \qquad \textbf{[18]}$$

Here, the probabilities $P_{sf}[N(T) = f|N(0) = s]$ are obtained from Eq. **7** or **12** and calculated by using the best-fitting models $\pi(\hat{\theta})$. A bootstrap simulation technique is used to generate a sample of deviances **D*** that is distributed according to $\pi(\mathbf{N}; \hat{\beta}, \hat{\delta}, \hat{\varepsilon})$ or $\pi(\mathbf{N}; \hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\varepsilon})$. First, data sets of initial states $\mathbf{N}_k^*\left(0\right)$ $k = 1,...m$ of $\dim\left(\mathbf{N}\right)$ are generated from the original data, then for each k the final states $\mathbf{N}_k^*\left(T\right)$ are simulated using the estimated transition probabilities in the models. The observed deviance D is tested against the distribution **D***; the sample size was chosen m = 10,000.

1. Taylor HM, Karlin S (1998) in *An Introduction To Stochastic Modeling* (Academic, San Diego, CA), pp 333–417.