Supplementary material to "A three-term recurrence relation for accurate evaluation of transition probabilities of the simple birth-and-death process"

Alberto Pessia[1,*] and Jing Tang[1]
[1]Research Program in Systems Oncology (ONCOSYS), Faculty of Medicine, University of Helsinki, Helsinki, Finland

S1 Algorithm for evaluating ${}_2F_1(-a, -b; -(a+b-k); -z)$

Algorithm 1 Hypergeometric evaluation

Input: Integers $a \geq 0$, $b \geq 0$, $k \leq 1$. Real $z > -1$.
Output: ${}_2F_1(-a, -b; -(a+b-k); -z)$
Initialization:
  $m \leftarrow \min(a, b)$
  $M \leftarrow \max(a, b)$
if $z = 0$ or $m = 0$ **then**
  return 1
end if
$y \leftarrow 1 + \dfrac{Mz}{M + 1 - k}$
if m = 1 **then**
  return y
end if
// To avoid overflow define $R_b = y_b / y_{b-1}$, that is $y_b = R_b y_{b-1}$. Note that $R_1 = y_1/y_0 = y_1$.
$R \leftarrow y$
for $n = 2, \ldots, m$ **do**
  $R \leftarrow 1 + \dfrac{z}{M + n - k}\left(M - n + 1 + \dfrac{(n-1)(n-1-k)}{(M+n-k-1)R}\right)$
  $y \leftarrow Ry$
end for
return y

S2 Gradient and Hessian of the log-transition probability

Partial derivatives of the log-transition probability are simple but cumbersome. To simplify notation we will drop function arguments (unless required for clarity) and denote the first and second order partial derivatives of a function $f(x, y)$ with f_x, f_y, f_{xx}, f_{xy}, f_{yx}, and f_{yy}. We will also use the

*To whom correspondence should be addressed. Email: academic@albertopessia.com

substitutions

$$x = e^{(\lambda-\mu)t}, \quad c^{(h)} = \binom{i}{h}\binom{i+j-h-1}{i-1}, \qquad \theta^{(h)} = \mu^{i-h}\lambda^{j-h}\phi(t,\lambda,\mu)^{i+j-2h}\gamma(t,\lambda,\mu)^{h},$$

$$u = \frac{{}_2F_1\left[\begin{smallmatrix}-(i-1),\ -(j-1)\\ -(i+j)\end{smallmatrix};-z(t,\ \lambda,\ \mu)\right]}{{}_2F_1\left[\begin{smallmatrix}-i,\ -j\\ -(i+j-1)\end{smallmatrix};-z(t,\ \lambda,\ \mu)\right]}, \qquad v = \frac{{}_2F_1\left[\begin{smallmatrix}-(i-2),\ -(j-2)\\ -(i+j+1)\end{smallmatrix};-z(t,\ \lambda,\ \mu)\right]}{{}_2F_1\left[\begin{smallmatrix}-i,\ -j\\ -(i+j-1)\end{smallmatrix};-z(t,\ \lambda,\ \mu)\right]}$$

Partial derivatives of the log-transition probability, in their most general form, are simply

$$(\log p)_\lambda = \frac{\sum_h c^{(h)}\theta^{(h)}_\lambda}{\sum_k c^{(k)}\theta^{(k)}} \qquad (\log p)_\mu = \frac{\sum_h c^{(h)}\theta^{(h)}_\mu}{\sum_k c^{(k)}\theta^{(k)}}$$

$$(\log p)_{\lambda\lambda} = \frac{\sum_h c^{(h)}\theta^{(h)}_{\lambda\lambda}}{\sum_k c^{(k)}\theta^{(k)}} - (\log p)^2_\lambda \qquad (\log p)_{\lambda\mu} = \frac{\sum_h c^{(h)}\theta^{(h)}_{\lambda\mu}}{\sum_k c^{(k)}\theta^{(k)}} - (\log p)_\lambda(\log p)_\mu$$

$$(\log p)_{\mu\lambda} = \frac{\sum_h c^{(h)}\theta^{(h)}_{\mu\lambda}}{\sum_k c^{(k)}\theta^{(k)}} - (\log p)_\mu(\log p)_\lambda \qquad (\log p)_{\mu\mu} = \frac{\sum_h c^{(h)}\theta^{(h)}_{\mu\mu}}{\sum_k c^{(k)}\theta^{(k)}} - (\log p)^2_\mu$$

We will now list all partial derivatives of basic functions to be used later in the Section. Partial derivatives of function $\log\phi(t,\lambda,\mu)$ are

$$(\log\phi)_\lambda = -\frac{x}{\lambda x-\mu}\left(1-\frac{(\lambda-\mu)t}{x-1}\right), \qquad (\log\phi)_\mu = \frac{1}{\lambda x-\mu}\left(1-\frac{(\lambda-\mu)tx}{x-1}\right)$$

$$(\log\phi)_{\lambda\lambda} = \frac{(x+\mu t)x}{(\lambda x-\mu)^2}\left(1-\frac{(\lambda-\mu)t}{x-1}\right) + \frac{tx}{(\lambda x-\mu)(x-1)}\left(1-\frac{(\lambda-\mu)tx}{x-1}\right)$$

$$(\log\phi)_{\mu\mu} = \frac{1+\lambda tx}{(\lambda x-\mu)^2}\left(1-\frac{(\lambda-\mu)tx}{x-1}\right) + \frac{tx}{(\lambda x-\mu)(x-1)}\left(1-\frac{(\lambda-\mu)t}{x-1}\right)$$

$$(\log\phi)_{\lambda\mu} = -\frac{(1+\mu t)x}{(\lambda x-\mu)^2}\left(1-\frac{(\lambda-\mu)t}{x-1}\right) - \frac{tx}{(\lambda x-\mu)(x-1)}\left(1-\frac{(\lambda-\mu)tx}{x-1}\right)$$

$$(\log\phi)_{\mu\lambda} = (\log\phi)_{\lambda\mu}$$

Partial derivatives of function $z(t,\lambda,\mu)$ are

$$z_\lambda = \frac{(\lambda-\mu)x}{\lambda\mu(x-1)^2}\left(\frac{\lambda+\mu}{\lambda} - \frac{(\lambda-\mu)t(x+1)}{x-1}\right)$$

$$z_\mu = -\frac{(\lambda-\mu)x}{\lambda\mu(x-1)^2}\left(\frac{\lambda+\mu}{\mu} - \frac{(\lambda-\mu)t(x+1)}{x-1}\right)$$

$$z_{\lambda\lambda} = \frac{x}{\lambda\mu(x-1)^2}\left(2\left(\frac{\mu}{\lambda}\right)^2 - \frac{(\lambda-\mu)t}{x-1}\left(2\left(\frac{\lambda+\mu}{\lambda}\right)(x+1) - \frac{(\lambda-\mu)t}{x-1}(x^2+4x+1)\right)\right)$$

$$z_{\mu\mu} = \frac{x}{\lambda\mu(x-1)^2}\left(2\left(\frac{\lambda}{\mu}\right)^2 - \frac{(\lambda-\mu)t}{x-1}\left(2\left(\frac{\lambda+\mu}{\mu}\right)(x+1) - \frac{(\lambda-\mu)t}{x-1}(x^2+4x+1)\right)\right)$$

$$z_{\lambda\mu} = -\frac{x}{\lambda\mu(x-1)^2}\left(\frac{\lambda^2+\mu^2}{\lambda\mu} - \frac{(\lambda-\mu)t}{x-1}\left(\frac{(\lambda+\mu)^2}{\lambda\mu}(x+1) - \frac{(\lambda-\mu)t}{x-1}(x^2+4x+1)\right)\right)$$

$$z_{\mu\lambda} = z_{\lambda\mu}$$

Partial derivatives of function $\log({}_2F_1(-i, -j; -(i+j-1); -z(t, \lambda, \mu)))$ are

$$
\begin{aligned}
(\log {}_2F_1)_\lambda &= \frac{iju}{i+j-1} z_\lambda, \qquad\qquad (\log {}_2F_1)_\mu = \frac{iju}{i+j-1} z_\mu \\
(\log {}_2F_1)_{\lambda\lambda} &= \frac{iju}{i+j-1} \left(z_{\lambda\lambda} + z_\lambda^2 \left(\frac{(i-1)(j-1)}{i+j-2} \frac{v}{u} - \frac{iju}{i+j-1} \right) \right) \\
(\log {}_2F_1)_{\mu\mu} &= \frac{iju}{i+j-1} \left(z_{\mu\mu} + z_\mu^2 \left(\frac{(i-1)(j-1)}{i+j-2} \frac{v}{u} - \frac{iju}{i+j-1} \right) \right) \\
(\log {}_2F_1)_{\lambda\mu} &= \frac{iju}{i+j-1} \left(z_{\lambda\mu} + z_\lambda z_\mu \left(\frac{(i-1)(j-1)}{i+j-2} \frac{v}{u} - \frac{iju}{i+j-1} \right) \right) \\
(\log {}_2F_1)_{\mu\lambda} &= (\log {}_2F_1)_{\lambda\mu}
\end{aligned}
$$

We can now study the shape of the partial derivatives of the log-transition probability in the various sub-domains. Considering that the binomial coefficient $\binom{a}{b}$ is equal to zero for all $b > a$, we will use the convention that $\binom{a}{b}/\binom{a}{b}$ is always equal to 1 for all a and b.

Parameters greater than zero

When t, λ, and μ are all greater than zero we can safely use representation (3.1). We need to distinguish the case $\mu \neq \lambda$ from the case $\mu = \lambda$.

Unequal rates

If $\mu \neq \lambda$ the partial derivatives are simply

$$(\log p)_\lambda = \frac{j}{\lambda} + (i+j)(\log \phi)_\lambda + (\log {}_2F_1)_\lambda \tag{1}$$

$$(\log p)_\mu = \frac{i}{\mu} + (i+j)(\log \phi)_\mu + (\log {}_2F_1)_\mu \tag{2}$$

$$(\log p)_{\lambda\lambda} = -\frac{j}{\lambda^2} + (i+j)(\log \phi)_{\lambda\lambda} + (\log {}_2F_1)_{\lambda\lambda} \tag{3}$$

$$(\log p)_{\mu\mu} = -\frac{i}{\mu^2} + (i+j)(\log \phi)_{\mu\mu} + (\log {}_2F_1)_{\mu\mu} \tag{4}$$

$$(\log p)_{\lambda\mu} = (i+j)(\log \phi)_{\lambda\mu} + (\log {}_2F_1)_{\lambda\mu} \tag{5}$$

$$(\log p)_{\mu\lambda} = (\log p)_{\lambda\mu} \tag{6}$$

Equal rates

Apply the limit $\mu \to \lambda$ directly to equations (1)-(6) to get

$$(\log p)_\lambda|_{\mu=\lambda} = \frac{j}{\lambda} + (i+j)\ (\log \phi)_\lambda|_{\mu=\lambda} + (\log {}_2F_1)_\lambda|_{\mu=\lambda} \tag{7}$$

$$(\log p)_\mu|_{\mu=\lambda} = \frac{i}{\lambda} + (i+j)\ (\log \phi)_\mu|_{\mu=\lambda} + (\log {}_2F_1)_\mu|_{\mu=\lambda} \tag{8}$$

$$(\log p)_{\lambda\lambda}|_{\mu=\lambda} = -\frac{j}{\lambda^2} + (i+j)\ (\log \phi)_{\lambda\lambda}|_{\mu=\lambda} + (\log {}_2F_1)_{\lambda\lambda}|_{\mu=\lambda} \tag{9}$$

$$(\log p)_{\mu\mu}|_{\mu=\lambda} = -\frac{i}{\lambda^2} + (i+j)\ (\log \phi)_{\mu\mu}|_{\mu=\lambda} + (\log {}_2F_1)_{\mu\mu}|_{\mu=\lambda} \tag{10}$$

$$(\log p)_{\lambda\mu}|_{\mu=\lambda} = (i+j)\ (\log \phi)_{\lambda\mu}|_{\mu=\lambda} + (\log {}_2F_1)_{\lambda\mu}|_{\mu=\lambda} \tag{11}$$

$$(\log p)_{\mu\lambda}|_{\mu=\lambda} = (\log p)_{\lambda\mu}|_{\mu=\lambda} \tag{12}$$

where

$$(\log \phi)_\lambda|_{\mu=\lambda} = (\log \phi)_\mu|_{\mu=\lambda} = -\frac{t}{2(1+\lambda t)}$$

$$(\log \phi)_{\lambda\lambda}|_{\mu=\lambda} = (\log \phi)_{\mu\mu}|_{\mu=\lambda} = \frac{(1-2\lambda t)t^2}{12(1+\lambda t)^2}$$

$$(\log \phi)_{\lambda\mu}|_{\mu=\lambda} = (\log \phi)_{\mu\lambda}|_{\mu=\lambda} = \frac{(5+2\lambda t)t^2}{12(1+\lambda t)^2}$$

$$(\log {}_2F_1)_\lambda|_{\mu=\lambda} = (\log {}_2F_1)_\mu|_{\mu=\lambda} = -\frac{iju}{(i+j-1)\lambda^3 t^2}$$

$$(\log {}_2F_1)_{\lambda\lambda}|_{\mu=\lambda} = \frac{ij}{(i+j-1)\lambda^4 t^2}\left(\frac{(12-\lambda^2t^2)u}{6} + \frac{(i-1)(j-1)v}{(i+j-2)\lambda^2 t^2} - \frac{iju^2}{(i+j-1)\lambda^2 t^2}\right)$$

$$(\log {}_2F_1)_{\mu\mu}|_{\mu=\lambda} = (\log {}_2F_1)_{\lambda\lambda}|_{\mu=\lambda}$$

$$(\log {}_2F_1)_{\lambda\mu}|_{\mu=\lambda} = \frac{ij}{(i+j-1)\lambda^4 t^2}\left(\frac{(6+\lambda^2t^2)u}{6} + \frac{(i-1)(j-1)v}{(i+j-2)\lambda^2 t^2} - \frac{iju^2}{(i+j-1)\lambda^2 t^2}\right)$$

$$(\log {}_2F_1)_{\mu\lambda}|_{\mu=\lambda} = (\log {}_2F_1)_{\lambda\mu}|_{\mu=\lambda}$$

Note that functions u and v must be evaluated at the point $z(t,\lambda,\lambda) = (\lambda t)^{-2} - 1$.

Parameters equal to zero

When any of t, λ, or μ equal zero it is easier to compute the partial derivatives starting from the standard representation (1.2) instead of (3.1). However, derivatives of $\theta^{(h)}$ are long and complicated, especially the second-order partial derivatives. Considering that intermediate results are not of interest we won't write them here. Instead, we will only provide the required final solutions.

Observation time is zero

When $t = 0$ the partial derivatives are always zero regardless of the values of i, j, λ, or μ. This is a consequence of the fact that the transition probability, equations (1.6) and (1.8), does not depend on the process rates.

Death rate is zero

When μ approaches zero also the partial derivatives of $\theta^{(h)}$, in general, approach zero. Only exceptions are $\theta^{(i)}_{\lambda}$, $\theta^{(i)}_{\mu}$, $\theta^{(i-1)}_{\mu}$, $\theta^{(i)}_{\lambda\lambda}$, $\theta^{(i)}_{\mu\mu}$,$\theta^{(i-1)}_{\mu\mu}$, $\theta^{(i-2)}_{\mu\mu}$, $\theta^{(i)}_{\lambda\mu}$, $\theta^{(i-1)}_{\lambda\mu}$, $\theta^{(i)}_{\mu\lambda}$, and $\theta^{(i-1)}_{\mu\lambda}$. Partial derivatives become

$$
\begin{aligned}
(\log p)_{\lambda}|_{\mu=0} &= -\left(\frac{ix-j}{x-1}\right)t && (13)\\
(\log p)_{\mu}|_{\mu=0} &= \left(\frac{ix-j}{x-1}\right)t - \frac{i(i-1)x + j(j+1)x^{-1} - 2ij}{(i-j-1)\lambda} && (14)\\
(\log p)_{\lambda\lambda}|_{\mu=0} &= \frac{(i-j)t^2x}{(x-1)^2} && (15)\\
(\log p)_{\mu\mu}|_{\mu=0} &= \frac{(i-j)t^2x}{(x-1)^2} + \frac{i(i-1)j(j+1)(x-1)^4}{(i-j-1)^2(i-j-2)\lambda^2x^2} + \\
&\quad - \frac{i(i-1)(x-2\lambda t)x + j(j+1)(x^{-1}+2\lambda t)x^{-1} - 2ij}{(i-j-1)\lambda^2} && (16)\\
(\log p)_{\lambda\mu}|_{\mu=0} &= -\frac{(i-j)t^2x}{(x-1)^2} + \frac{i(i-1)(1-\lambda t)x + j(j+1)(1+\lambda t)x^{-1} - 2ij}{(i-j-1)^2\lambda^2} && (17)\\
(\log p)_{\mu\lambda}|_{\mu=0} &= (\log p)_{\lambda\mu}|_{\mu=0} && (18)
\end{aligned}
$$

where $x = e^{\lambda t}$. Note that equation (14) has a discontinuity at the value $j = i - 1$ where

$$\lim_{j\to(i-1)^-} (\log p)_{\mu}|_{\mu=0} = -\infty, \qquad \lim_{j\to(i-1)^+} (\log p)_{\mu}|_{\mu=0} = \infty$$

Equation (16) has a discontinuity at the value $j = i - 2$ where

$$\lim_{j\to(i-2)^-} (\log p)_{\mu\mu}|_{\mu=0} = \infty, \qquad \lim_{j\to(i-2)^+} (\log p)_{\mu\mu}|_{\mu=0} = -\infty$$

Equations (17) and (18) have a discontinuity at the value $j = i - 1$ where

$$\lim_{j\to(i-1)^-} (\log p)_{\lambda\mu}|_{\mu=0} = -\infty, \qquad \lim_{j\to(i-1)^+} (\log p)_{\lambda\mu}|_{\mu=0} = \infty$$

Birth rate is zero

When λ approaches zero also the partial derivatives of $\theta^{(h)}$, in general, approach zero. Only exceptions are $\theta_{\lambda}^{(j)}$, $\theta_{\lambda}^{(j-1)}$, $\theta_{\mu}^{(j)}$, $\theta_{\lambda\lambda}^{(j)}$, $\theta_{\lambda\lambda}^{(j-1)}$, $\theta_{\lambda\lambda}^{(j-2)}$, $\theta_{\mu\mu}^{(j)}$, $\theta_{\lambda\mu}^{(j)}$, $\theta_{\lambda\mu}^{(j-1)}$, $\theta_{\mu\lambda}^{(j)}$, and $\theta_{\mu\lambda}^{(j-1)}$. Partial derivatives become

$$(\log p)_{\lambda}|_{\lambda=0} = \left(\frac{jx-i}{x-1}\right)t - \frac{j(j-1)x + i(i+1)x^{-1} - 2ij}{(j-i-1)\mu} \tag{19}$$

$$(\log p)_{\mu}|_{\lambda=0} = -\left(\frac{jx-i}{x-1}\right)t \tag{20}$$

$$(\log p)_{\lambda\lambda}|_{\lambda=0} = \frac{(j-i)t^2x}{(x-1)^2} + \frac{j(j-1)i(i+1)(x-1)^4}{(j-i-1)^2(j-i-2)\mu^2x^2} + \\ - \frac{j(j-1)(x-2\mu t)x + i(i+1)(x^{-1}+2\mu t)x^{-1} - 2ij}{(j-i-1)\mu^2} \tag{21}$$

$$(\log p)_{\mu\mu}|_{\lambda=0} = \frac{(j-i)t^2x}{(x-1)^2} \tag{22}$$

$$(\log p)_{\lambda\mu}|_{\lambda=0} = -\frac{(j-i)t^2x}{(x-1)^2} + \frac{j(j-1)(1-\mu t)x + i(i+1)(1+\mu t)x^{-1} - 2ij}{(j-i-1)^2\mu^2} \tag{23}$$

$$(\log p)_{\mu\lambda}|_{\lambda=0} = (\log p)_{\lambda\mu}|_{\mu=0} \tag{24}$$

where $x = e^{\mu t}$. Note that equation (19) has a discontinuity at the value $j = i + 1$ where

$$\lim_{j\to(i+1)^-} (\log p)_{\lambda}|_{\lambda=0} = \infty, \qquad \lim_{j\to(i+1)^+} (\log p)_{\lambda}|_{\lambda=0} = -\infty$$

Equation (21) has a discontinuity at the value $j = i + 2$ where

$$\lim_{j\to(i+2)^-} (\log p)_{\mu\mu}|_{\lambda=0} = -\infty, \qquad \lim_{j\to(i+2)^+} (\log p)_{\mu\mu}|_{\lambda=0} = \infty$$

Equations (23) and (24) have a discontinuity at the value $j = i + 1$ where

$$\lim_{j\to(i+1)^-} (\log p)_{\lambda\mu}|_{\lambda=0} = \infty, \qquad \lim_{j\to(i+1)^+} (\log p)_{\lambda\mu}|_{\lambda=0} = -\infty$$

Both rates are zero

The gradient of the log-transition probability at the origin is only defined when $j = i$. To prove it, we will compute the limit $(\lambda,\mu) \to (0,0)$ from different directions and observe whether they all converge to the same value or not. If $j \neq i$

$$\lim_{\lambda\to 0} (\log p)_{\lambda}|_{\mu=0} = \operatorname{sgn}(j-i)\infty, \quad \lim_{\mu\to 0} (\log p)_{\lambda}|_{\lambda=0} = -\frac{(i+j)t}{2}, \quad \lim_{\lambda\to 0} (\log p)_{\lambda}|_{\mu=\lambda} = 0$$

and

$$\lim_{\lambda\to 0} (\log p)_{\mu}|_{\mu=0} = -\frac{(i+j)t}{2}, \quad \lim_{\mu\to 0} (\log p)_{\mu}|_{\lambda=0} = \operatorname{sgn}(i-j)\infty, \quad \lim_{\lambda\to 0} (\log p)_{\mu}|_{\mu=\lambda} = 0$$

The same phenomenon can be observed with the second-order partial derivatives. When $j = i$ the first-order partial derivatives converge to $-it$. Second-order partial derivatives $(\log p)_{\lambda\lambda}$ and

$(\log p)_{\mu\mu}$ converge to 0 while $(\log p)_{\lambda\mu}$ and $(\log p)_{\mu\lambda}$ converge to $i^2 t^2$. If we interpret the transition probability as the likelihood of a single time point observation, these results are intuitive. Indeed, if $j \neq i$ the rates cannot be both equal to zero. If $j = i$, instead, the hypothesis $\lambda = \mu = 0$ is plausible because it is compatible with the observation.