S1 Appendix: Rate quantiles and operators

1 Piecewise linear approximation

In this article we introduced a linear piecewise approximation of the i-CDF (inverse cumulative distribution function) to improve the computational performance of the quant parameterisation. Let $\hat{F}^{-1}(\mathcal{R}_i)$ be the piecewise approximation of the i-CDF $F^{-1}(\mathcal{R}_i)$. The approximation consists of $n$ pieces (where $n = 100$ is fixed). Due to the nonlinear nature of small and large quantiles in a log-normal distribution, the first and last pieces are not linear approximations but rather equal to the underlying distribution itself.

$$\hat{F}^{-1}(q) = \begin{cases} F^{-1}(q) & \text{if } q \leq \frac{1}{n} \text{ or } q \geq \frac{n-1}{n} \\ F^{-1}([v]) + \left( F^{-1}([v] + 1) - F^{-1}([v]) \right) \left( v - [v] \right) & \text{otherwise.} \end{cases}$$

where $v = q(n-1)$ indexes quantile $q$ into piece number $[v]$. Values from the underlying function $F^{-1}$ are cached, enabling rapid computation.

2 Tree operators for rate quantiles

Zhang and Drummond 2020 introduced several tree operators for the real parameterisation — including ConstantDistance, SimpleDistance, and SmallPulley [1]. In this appendix, these three operators are extended to the quant parameterisation. Following the notation presented in the main article, let $t_i$ be the time of node $i$, let $0 < q_i < 1$ be the rate quantile of node $i$, and let $r_i = \hat{F}^{-1}(q_i)$ be the real rate of node $i$ where $\hat{F}^{-1}$ is the linear approximation of the i-CDF.
**Constant Distance**

Let $X$ be a uniformly-at-random sampled internal node on tree $T$. Let $L$ and $R$ be the left and right child of $X$, respectively, and let $P$ be the parent of $X$. Under the quant parameterisation, the **ConstantDistance** operator works as follows:

**Step 1.** Propose a new height for $t_X$:

$$ t'_X \leftarrow t_X + s\Sigma $$

where $\Sigma$ is drawn from a proposal transition distribution (Uniform or Bactrian), and $s$ is a tunable step size. Ensure that $\max\{t_L, t_R\} < t'_X < t_P$, and if the constraint is broken then reject the proposal.

**Step 2.** Recalculate $q_X$ as:

$$ q'_X \leftarrow \hat{F}\left(r'_X\right) $$

$$ \leftarrow \hat{F}\left(\frac{t_P - t_X}{t_P - t'_X} r_X\right) $$

$$ \leftarrow \hat{F}\left(\frac{t_P - t_X}{t_P - t'_X} F^{-1}(q_X)\right). $$

This ensures that the genetic distance between $X$ and $P$ remains constant after the operation by enforcing the constraint:

$$ r_X(t_P - t_X) = r_X'(t_P - t'_X). $$

**Step 3.** Similarly, propose new rate quantiles for the two children $C \in \{L, R\}$:

$$ q'_C \leftarrow \hat{F}\left(r'_C\right) $$

$$ \leftarrow \hat{F}\left(\frac{t_X - t_C}{t'_X - t_C} r_C\right) $$

$$ \leftarrow \hat{F}\left(\frac{t_X - t_C}{t'_X - t_C} F^{-1}(q_C)\right). $$

Ensure that $0 < q'_i < 1$ for all proposed nodes $i \in \{X, L, R\}$, and if the constraint is broken then reject the proposal. This constraint can only be broken from numerical issues.

**Step 4.** Finally, in order to calculate the Metropolis-Hastings-Green ratio, return the determinant of the Jacobian matrix:
The derivatives $\hat{D}$ and $\hat{F}$ are computed using numerical approximations for the first and last pieces, or as the gradient of the linear approximation for internal pieces. As its final step, the operator returns $\ln |J|$.

\[
\begin{align*}
\ln |J| &= \ln \left( \frac{\partial X'}{\partial t} \times \frac{\partial q'}{\partial q} \times \frac{\partial q''}{\partial p} \times \frac{\partial \Sigma}{\partial \Sigma} \right) \\
&= \ln \left( \frac{\partial t'}{\partial t} \times \frac{\partial q'}{\partial q} \times \frac{\partial q''}{\partial p} \times \frac{\partial \Sigma}{\partial \Sigma} \right) \\
&= \ln \left( \frac{\partial t'}{\partial t} \times \frac{\partial q'}{\partial q} \times \frac{\partial q''}{\partial p} \times \frac{\partial \Sigma}{\partial \Sigma} \right)
\end{align*}
\]

As $J$ is triangular, its determinant $|J|$ is equal to the product of diagonal elements:

\[
\begin{align*}
\ln |J| &= \ln\left(\frac{\partial t'}{\partial t} \times \frac{\partial q'}{\partial q} \times \frac{\partial q''}{\partial p} \times \frac{\partial \Sigma}{\partial \Sigma}\right) \\
&= \ln 1 + \ln \left(\frac{\partial t'}{\partial t} \times \frac{\partial q'}{\partial q} \times \frac{\partial q''}{\partial p} \times \frac{\partial \Sigma}{\partial \Sigma}\right) \\
&= \ln \left(\frac{\partial t'}{\partial t} \times \frac{\partial q'}{\partial q} \times \frac{\partial q''}{\partial p} \times \frac{\partial \Sigma}{\partial \Sigma}\right)
\end{align*}
\]

The derivatives $\hat{D}$ and $\hat{F}$ are computed using numerical approximations for the first and last pieces, or as the gradient of the linear approximation for internal pieces. As its final step, the operator returns $\ln |J|$.

**Simple Distance**

While `ConstantDistance` proposes internal node heights, `SimpleDistance` operates on the root. Let $X'$ be the root node and let $L$ and $R$ be its two children.

**Step 1.** Propose a new height for $t_X$:

\[
t_X' \leftarrow t_X + s \Sigma.
\]

Ensure that $\max\{t_L, t_R\} < t_X'$, and if the constraint is broken then reject the proposal.
**Step 2.** Propose new rate quantiles for the two children \( \mathcal{C} \in \{L, R\} \):

\[
q_C' \leftarrow \hat{F}(r_C') \\
\leftarrow \hat{F}(\frac{t_X - t_C}{t_X' - t_C} \times r_C) \\
\leftarrow \hat{F}(\frac{t_X - t_C}{t_X' - t_C} \times \hat{F}^{-1}(q_C)).
\]

These proposals ensure that the genetic distance between \( X \) and its children \( \mathcal{C} \) remain constant after the operation by enforcing the constraint:

\[
r_C(t_X - t_C) = r_C'(t_X' - t_C). \quad (10)
\]

Ensure that \( 0 < q_C' < 1 \), and if the constraint is broken then reject the proposal.

**Step 3.** Finally, in order to calculate the Metropolis-Hastings-Green ratio, return the determinant of the Jacobian matrix:

\[
J = \begin{bmatrix}
    \frac{\partial t_{X'}}{\partial t_X} & \frac{\partial t_{X'}}{\partial q_L} & \frac{\partial t_{X'}}{\partial q_R} \\
    \frac{\partial q_L'}{\partial t_X} & \frac{\partial q_L'}{\partial q_L} & \frac{\partial q_L'}{\partial q_R} \\
    \frac{\partial q_R'}{\partial t_X} & \frac{\partial q_R'}{\partial q_L} & \frac{\partial q_R'}{\partial q_R}
\end{bmatrix} = \begin{bmatrix}
    0 & 0 & 0 \\
    \frac{\partial t_{X'}}{\partial t_X} & \frac{\partial q_L'}{\partial q_L} & \frac{\partial q_L'}{\partial q_R} \\
    \frac{\partial q_R'}{\partial t_X} & 0 & \frac{\partial q_R'}{\partial q_R}
\end{bmatrix}. \quad (11)
\]

As \( J \) is triangular, its determinant \( |J| \) is equal to the product of diagonal elements:

\[
\ln |J| = \ln \left\{ \frac{\partial t_{X'}}{\partial t_X} \times \frac{\partial q_L'}{\partial q_L} \times \frac{\partial q_R'}{\partial q_R} \right\} = \ln \frac{\partial t_{X'}}{\partial t_X} + \ln \frac{\partial q_L'}{\partial q_L} + \ln \frac{\partial q_R'}{\partial q_R} = \ln 1 \\
+ \ln \frac{\partial t_{X'}}{\partial t_X} \frac{\partial q_L'}{\partial q_L} \frac{\partial t_{X'}}{\partial q_R} \frac{\partial q_R'}{\partial q_R} = \ln D \hat{F} \left( \frac{t_X - t_L}{t_X' - t_L} \times \hat{F}^{-1}(q_L) \right) + \ln \frac{\partial t_{X'}}{\partial t_X} \frac{\partial q_L'}{\partial q_L} \frac{\partial t_{X'}}{\partial q_R} \frac{\partial q_R'}{\partial q_R} = \ln D \hat{F} \left( \frac{t_X - t_L}{t_X' - t_L} \times \hat{F}^{-1}(q_L) \right) + \ln D \hat{F} \left( \frac{t_X - t_L}{t_X' - t_L} \times \hat{F}^{-1}(q_R) \right) = \ln \left\{ \frac{\partial t_{X'}}{\partial t_X} \times \hat{F}^{-1}(q_L) \right\} + \ln \frac{\partial t_{X'}}{\partial t_X} \frac{\partial q_R'}{\partial q_R} = \ln \left\{ \frac{\partial t_{X'}}{\partial t_X} \times \hat{F}^{-1}(q_R) \right\} + \ln \frac{\partial t_{X'}}{\partial t_X} \frac{\partial q_R'}{\partial q_R}.
\]

As its final step, the operator returns \( \ln |J| \).
Small Pulley

Just like the previous operator, SmallPulley operates on the root. Let $X$ be the root node and let $L$ and $R$ be its two children. However, unlike SimpleDistance, this operator alters the two genetic distances $d_L = r_L(t_X - t_L) = \hat{F}^{-1}(q_L)(t_X - t_L)$ and $d_R = r_R(t_X - t_R) = \hat{F}^{-1}(q_R)(t_X - t_R)$, while conserving their sum $d_L + d_R$.

**Step 1.** Propose new genetic distances for $d_L$ and $d_R$:

$$d_L' \leftarrow d_L + s\Sigma$$

$$d_R' \leftarrow d_R - s\Sigma$$

Ensure that $0 < d_L' < d_L + d_R$, and if the constraint is broken then reject the proposal.

**Step 2.** Propose new rate quantiles for the two children $L$ and $R$:

$$q_L' \leftarrow \hat{F}(\frac{d_L'}{t_X - t_L})$$

$$\leftarrow \hat{F}(\frac{\hat{F}^{-1}(q_L)(t_X - t_L) + \Sigma}{t_X - t_L})$$

$$q_R' \leftarrow \hat{F}(\frac{d_R'}{t_X - t_R})$$

$$\leftarrow \hat{F}(\frac{\hat{F}^{-1}(q_R)(t_X - t_R) - \Sigma}{t_X - t_R})$$

**Step 3.** Return the determinant of the Jacobian matrix:

$$J = \begin{bmatrix}
\frac{\partial q_L'}{\partial q_L'} & \frac{\partial q_L'}{\partial q_R'} \\
\frac{\partial q_R'}{\partial q_L'} & \frac{\partial q_R'}{\partial q_R'}
\end{bmatrix}$$

$$= \begin{bmatrix}
\frac{\partial q_L'}{\partial q_L'} & 0 \\
0 & \frac{\partial q_R'}{\partial q_R'}
\end{bmatrix}.$$  

As $J$ is triangular/diagonal, its determinant $|J|$ is equal to the product of diagonal elements:
\[
\ln |J| = \ln \left\{ \frac{\partial q'_L}{\partial q_L} \times \frac{\partial q'_R}{\partial q_R} \right\}
= \ln \frac{\partial q'_L}{\partial q_L} + \ln \frac{\partial q'_R}{\partial q_R}
= \ln \hat{D} \left( \frac{\hat{F}^{-1}(q_L)(t_X - t_L) + \Sigma}{t_X - t_L} \right) + \ln \frac{\partial q'_L}{\partial q_L} \frac{\hat{F}^{-1}(q_L)(t_X - t_L) + \Sigma}{t_X - t_L}
+ \ln \hat{D} \left( \frac{\hat{F}^{-1}(q_R)(t_X - t_R) - \Sigma}{t_X - t_R} \right) + \ln \frac{\partial q'_R}{\partial q_R} \frac{\hat{F}^{-1}(q_R)(t_X - t_R) - \Sigma}{t_X - t_R}
= \ln \hat{D} \left( \frac{\hat{F}^{-1}(q_L)(t_X - t_L) + \Sigma}{t_X - t_L} \right) + \ln \hat{D} \hat{F}^{-1}(q_L)
+ \ln \hat{D} \left( \frac{\hat{F}^{-1}(q_R)(t_X - t_R) - \Sigma}{t_X - t_R} \right) + \ln \hat{D} \hat{F}^{-1}(q_R). \tag{18}
\]

Thus, as its final step, the operator returns \( \ln |J| \).
3 CisScale operator

CisScale was originally introduced by Zhang and Drummond 2020 for the real parameterisation (therein named ucldstdevScaleOperator). Under the quant configuration, the CisScale operator works as follows.

**Step 1.** Propose a new value for the relaxed clock standard deviation $\sigma$

$$\sigma' \leftarrow \sigma \times e^{s\Sigma}. \quad (19)$$

**Step 2.** Recalculate all branch substitution rate quantiles $q$ such that their rates $r$ remain constant

let $r = \hat{F}^{-1}(q|\sigma)$

let $r' = r$ \quad (20)

let $q' \leftarrow \hat{F}(r'|\sigma'') = \hat{F}(\hat{F}^{-1}(q|\sigma)|\sigma'). \quad (21)$

**Step 3.** Return the log Hastings-Green ratio of this proposal. If $\Sigma$ was drawn from a symmetric proposal kernel (such as the Bactrian distribution), this is equal to:

$$|J| = \log(e^{s\Sigma}) + \log \left( \frac{\delta}{\delta q} \hat{F}(\hat{F}^{-1}(q|\sigma)|\sigma') \right) \quad (23)$$

$$= s\Sigma + \log D\hat{F}(\hat{F}^{-1}(q|\sigma)|\sigma') + \log D\hat{F}^{-1}(q|\sigma), \quad (24)$$

where derivatives $D\hat{F}$ and $D\hat{F}^{-1}$ can be approximated using either the piecewise linear model or standard numerical libraries.


4 Narrow exchange rates

The NarrowExchangeRate operator is also compatible with rate quantiles. This operator behaves the same as presented in the main article however the Hastings-Green ratio requires further augmentation due to changes in dimension throughout the proposal.

Step 1. Apply NarrowExchange to the current tree topology as described in the main article. This will return a Hastings ratio $H$ due to the asymmetry of this proposal.

Step 2. Compute the relevant branch rates $r_i$ for $r \in \{A, B, C, D\}$ of the current state from their respective quantile parameters.

$$r_i = \hat{F}^{-1}(q_i). \quad (25)$$

Step 3. Propose new rates and node heights and compute the Hastings-Green ratio of the real-space component of the proposal (e.g. Algorithms 1-2 of the main article).

$$(r_A', r_B', r_C', r_D', t_D', |J_r|) \leftarrow \text{PROPOSAL}(r_A, r_B, r_C, r_D, t_D). \quad (26)$$

Step 4. Transform the rates back into quantiles.

$$q_i' = \hat{F}(r_i'). \quad (27)$$

Step 5. Compute the log Hastings-Green ratio of the interconversion between rates and quantiles.

$$\log |J_q| = \log \hat{F}(q) + \log \hat{F}^{-1}(r'). \quad (28)$$

Step 6. Return the total log Hastings-Green ratio of this proposal: $\log H + \log |J_r| + \log |J_q|$. 

## 5 Summary of proposal kernels

Operators whose proposal kernels are affected by the decision to use a Bactrian kernel, as opposed to a uniform kernel, are specified below.

| Operator(s)     | Proposal | Parameter $x$ |
|-----------------|----------|---------------|
| 1 RandomWalk    | $x' \leftarrow x + s\Sigma$ | $\mathcal{R}, \sigma$ |
| 2 Scale         | $x' \leftarrow x \times e^{s\Sigma}$ | $\mathcal{R}, \sigma$ |
| 3 Interval      | $y \leftarrow \frac{1-x}{x} \times e^{s\Sigma}$ | $\mathcal{R}$ |
|                 | $x' \leftarrow \frac{y}{y+1}$                  |               |
| 4 ConstantDistance | $x' \leftarrow x + s\Sigma$       | $t$       |
|     SimpleDistance |                                   |               |
| 5 SmallPulley   | $x' \leftarrow x + s\Sigma$ | $\mathcal{R}$ |
| 6 CisScale      | $x' \leftarrow x \times e^{s\Sigma}$ | $\sigma$ |

Table 1: Proposal kernels $q(x'|x)$ of clock model operators. In each operator, $\Sigma$ is drawn from either a Bactrian($m$) or uniform distribution. The scale size $s$ is tunable. **ConstantDistance** and **SimpleDistance** propose tree heights $t$. The **Interval** operator applies to rate quantiles and respects its domain i.e. $0 < x, x' < 1$. 


6 Supplementary NER Algorithm

A second NER algorithm is presented below. This operator was rejected by the screening protocol on simulated data.

Algorithm 1 The NER\{D_{BC}, D_{CE}\} operator.

1: procedure PROPOSAL\((t_A, t_B, t_C, t_D, t_E, r_A, r_B, r_C, r_D)\)
2: 3: \(s\Sigma \leftarrow \text{getRandomWalkSize()}\) \> Random walk size is 0 unless this is NERw
4: \(t'_D \leftarrow t_D + s\Sigma\) \> Propose new node height for \(D\)
5: 6: \(r'_A \leftarrow r_A\) \> Propose new rates
7: \(r'_B \leftarrow \frac{r_B(t_D-t_B)+r_D(t_E-t_D)+r_D(t_E-t'_D)}{t'_D-t_B}\)
8: \(r'_C \leftarrow \frac{r_C(t_E-t_C)-r_D(t_E-t'_D)}{t'_D-t_C}\)
9: \(r'_D \leftarrow r_D\)
10: 11: \(|J| \leftarrow \frac{(t_D-t_B)(t_E-t_C)}{(t'_D-t_B)(t'_D-t_C)}\) \> Calculate Jacobian determinant
12: return \((r'_A, r'_B, r'_C, r'_D, t'_D, |J|)\)

References

[1] Zhang R, Drummond A. Improving the performance of Bayesian phylogenetic inference under relaxed clock models. BMC Evolutionary Biology. 2020;20:1–28.