`EWF 2.0`: Exact sampling from the Wright–Fisher diffusion with time-varying demography

Jaromir Sant[1,2], Paul A. Jenkins[3,4], Jere Koskela[4,5], Dario Spanò[4]

ESOMAS Department, University of Turin[1] & Collegio Carlo Alberto[2], Italy

Department of Computer Science[3] & Department of Statistics[4]

University of Warwick, Coventry CV4 7AL, United Kingdom

School of Mathematics, Statistics and Physics[5],

Newcastle University, Newcastle upon Tyne NE1 7RU, United Kingdom

October 21, 2025

Abstract

Accurate modelling of allele frequency trajectories requires incorporation of both genetic mechanisms such as selection and mutation, as well as realistic population demography. Accounting for a non-constant demography within a Wright–Fisher diffusion framework induces a time-inhomogenous drift coefficient, a regime falling outside the scope of existing exact simulation routines. To address this gap, we introduce `EWF 2.0`, an exact simulation algorithm that accommodates time-varying demography within Wright–Fisher diffusions whilst retaining all the functionality of previous `EWF` versions. This extension is particularly valuable for bridge simulation, where existing methods cannot handle differing mutation and selection rates. For a specified demographic history, mutation parameters, selection function and sampling times, `EWF 2.0` generates exact draws from the law of the corresponding Wright–Fisher diffusion or diffusion bridge. We validate correctness using distributional tests (Kolmogorov–Smirnov, QQ plots), confirming agreement with theoretical expectations. In spite of its greater generality, `EWF 2.0` retains the same runtime as in previous versions, ensuring computational efficiency and scalability. All software is available at `https://github.com/JaroSant/EWF`

1 Background

The Wright–Fisher diffusion is a cornerstone of population genetics, valued for its flexibility in modelling the combined effects of mutation, selection, and drift on allele frequency trajectories. Its probabilistic properties are well understood (Ethier and Kurtz (1986, Chapter 10); Karlin and Taylor (1981, Chapter 15)), and it has become the theoretical gold standard for describing allele frequency dynamics. Many inference methods in population genomics build on this process to analyse temporal variation in genetic

data (Bollback *et al.*, 2008; Malaspinas *et al.*, 2012; Živković *et al.*, 2015; Schraiber *et al.*, 2016; Ferrer-Admetlla *et al.*, 2016; Sohail *et al.*, 2021).

Nonetheless, working with the Wright–Fisher diffusion remains difficult in practice. Although an explicit infinite-series representation of the transition density has been available for decades (Griffiths, 1979; Tavaré, 1984), it is computationally unwieldy, forcing most methods to rely on approximations. These challenges are further compounded when demography varies over time, as the drift term becomes time-inhomogeneous. Frameworks that accommodate non-constant demographies are now common, with spectral methods providing pointwise evaluations of the transition density under varying effective population sizes (Steinrücken *et al.*, 2015), and inferential approaches directly incorporating demographic change into the methodology (Schraiber *et al.*, 2016; He *et al.*, 2020). However, to date no software exists to permit *exact* simulation under a time-varying demography.

A major step forward was the introduction of an exact sampling technique for Wright–Fisher diffusions (Jenkins and Spanò, 2017), which was subsequently implemented as `EWF` in Sant *et al.* (2023). This enabled efficient simulation across a wide class of models, but only under the assumption of constant population size. That assumption is restrictive, since demographic variation is ubiquitous in natural populations and, if ignored, can be confounded with changes in mutation or selection pressures.

Here we extend this class of simulation schemes to accommodate piecewise constant demographies, presenting both the mathematical framework and its subsequent implementation in `EWF 2.0`. Such demographic changes translate into time-inhomogeneous mutation and selection parameters, thereby requiring non-trivial modifications to the existing exact simulation scheme, particularly in the case of diffusion bridges where one cannot simply chain together independent simulations, but needs to explicitly account for the temporal dependence of the underlying model parameters. Bridge sampling is especially relevant for genomic selection analyses, where pre-screening loci for large allele-frequency shifts introduces ascertainment bias; conditioning on endpoints via bridges provides a principled correction, preventing spurious inflation of mutation or selection estimates.

Although we focus solely on piecewise constant demographies, we point out that this choice allows us to retain a tractable exact simulation scheme whilst also offering a direct way of addressing non-constant demographic scenarios, given that the latter can be arbitrarily well-approximated through stepwise demographies. The new algorithm enables exact sampling of Wright–Fisher diffusion bridges when mutation and selection parameters differ across epochs, while preserving all features of the original package (including boundary absorption/reflection and arbitrary frequency-dependent selection). Despite the added generality, `EWF 2.0` retains the same computational complexity and runtime as previous versions, ensuring efficient simulation. The above-detailed capabilities broaden the scope of exact Wright–Fisher simulation and strengthen its utility for modern population-genetic inference.

2 Implementation

Denote by $X := (X_t)_{t\geq 0}$ the two-allele non-neutral Wright–Fisher diffusion on $[0,1]$ with time-varying mutation parameter $\boldsymbol{\theta}(t) = (\theta_1(t), \theta_2(t))$, selection coefficient $\sigma(t)$, and frequency-dependent selection function $\eta(x)$. Then X is a solution to the stochastic differential equation

$$\begin{aligned} dX_t = &\frac{1}{2}\left[\sigma(t)X_t(1-X_t)\eta(X_t) - \theta_2(t)X_t + \theta_1(t)(1-X_t)\right]dt \\ &+ \sqrt{X_t(1-X_t)}dW_t \end{aligned} \tag{1}$$

for $t \geq 0$ with $X_0 \in [0,1]$. We further assume that the selection function is of the form $\eta(x) = \sum_{i=0}^{n} a_i x^i$ for some finite n, which encompasses a wide class of selective regimes including genic ($\eta(x) = 1$) and diploid selection ($\eta(x) = h + x(1-2h)$ with h the dominance parameter).

The effective population size is assumed to be given by a known piecewise constant function $N_e(t)$. Changes in $N_e(t)$ induce corresponding changes in the diffusion parameters via $\theta_j(t) = 2N_e(t)\mu_j$ for $j = 1, 2$, and $\sigma(t) = 2N_e(t)\gamma$, where μ_j denote the corresponding pre-limiting parameters expressed per generation, and γ is the selection coefficient.

We refer to each interval over which $N_e(t)$ is constant as an *epoch*. For illustration, consider two epochs $(-\infty, s)$ and $[s, \infty)$ for some $s > 0$, with corresponding parameters $\boldsymbol{\theta}, \sigma$ and $\boldsymbol{\theta'}, \sigma'$ over the two intervals. Suppose the diffusion starts at $x \in (0,1)$ at time 0 and ends at $z \in (0,1)$ at time $t > s$. In the neutral case (i.e. $\sigma = 0 = \sigma'$) with strictly positive mutation parameters, the bridge diffusion transition density at change point s admits the mixture representation

$$\begin{aligned} p^{x,z;t}_{\boldsymbol{\theta},\boldsymbol{\theta'}}(y;s) = &\sum_{m=0}^{\infty}\sum_{k=0}^{\infty}\sum_{l=0}^{m}\sum_{j=0}^{k} q^{\theta}_m(s) q^{\theta'}_k(t-s)\mathrm{Bin}_{m,x}(l) \\ &\times \mathrm{Beta}_{\theta'_1+j,\theta'_2+k-j}(z)\mathrm{DM}_{\theta_1+l,\theta_2+m-l;k}(j) \\ &\times \mathrm{Beta}_{\theta_1+l+j,\theta_2+m-l+k-j}(y) \\ =: &\sum_{m=0}^{\infty}\sum_{k=0}^{\infty}\sum_{l=0}^{m}\sum_{j=0}^{k} p^{(x,z,s,t,\boldsymbol{\theta},\boldsymbol{\theta'})}_{m,k,l,j} \\ &\times \mathrm{Beta}_{\theta_1+l+j,\theta_2+m-l+k-j}(y) \end{aligned} \tag{2}$$

where $(q^{\theta}_g(u))_{g\in\mathbb{N}}$ for $u > 0$ defines a probability distribution on the integers, $p^{(x,z,s,t,\boldsymbol{\theta},\boldsymbol{\theta'})}_{m,k,l,j}$ defines a probability distribution on $\mathbb{N}^4$ (see Supplementary Information, Section 1), and $\theta := \theta_1 + \theta_2$.

By adapting Jenkins and Spanò (2017, Section 3), the mixture terms can be sampled using suitable alternating-series bounds. The key extension is that these bounds remain valid when mutation parameters differ across epochs:

Proposition 1. *For any $(m,k,l,j) \in \mathbb{N}^4$, the weights $p^{(x,z,s,t,\boldsymbol{\theta},\boldsymbol{\theta'})}_{m,k,l,j}$ admit 'eventually monotone' sequences of lower and upper bounds, ensuring* exact *simulation from the mixture distribution is possible.*

A more precise statement (together with accompanying proof) can be found in the Supplementary Information, Section 2, Proposition 4. Additionally, expressions for the bridge transition density for all possible combinations of starting and ending frequency x and z, including the possibility that these frequencies sit *at* the boundaries, $x, z \in [0, 1]$, can be found in equations (5)–(13) in Section 1 of the Supplementary Information. These boundary cases are especially relevant for *de novo* mutations, where alleles enter at (effectively) zero frequency and begin spreading through the population. In (2) the Beta densities are singular at 0 or 1 when the corresponding mutation parameters are smaller than 1. Thus bridges with boundary endpoints require a boundary-specific treatment rather than a naïve evaluation of the mixture.

Samples drawn at an arbitrary time τ from the law of a diffusion bridge spanning a demographic breakpoint are generated by (i) sampling the corresponding allele frequency at the time $\tau + s$ of the demographic breakpoint, and (ii) sampling a bridge over the interval $[0, \tau + s]$ conditional on the value of the bridge at time $\tau + s$ and the appropriate start or end points. To see this, suppose we are interested in generating allele frequencies at time τ from the law of a diffusion bridge going from x to z over a time interval $\tau + s + t$ where $\tau, s, t > 0$, and $\tau + s$ denotes a time at which we have a demographic changepoint. Then the joint density of bridge points y, w drawn at times τ and $\tau + s$ is given by

$$\begin{aligned} p^{x,z;\tau+s+t}_{\boldsymbol{\theta},\boldsymbol{\theta'}}(y, w; \tau, \tau + s) &= \frac{p_{\boldsymbol{\theta}}(x, y; \tau) p_{\boldsymbol{\theta}}(y, w; s) p_{\boldsymbol{\theta'}}(w, z; t)}{p_{\boldsymbol{\theta},\boldsymbol{\theta'}}(x, z; \tau + s + t)} \\ &= p^{x,w;\tau+s}_{\boldsymbol{\theta},\boldsymbol{\theta}}(y; \tau) \frac{p_{\boldsymbol{\theta}}(x, w; \tau + s) p_{\boldsymbol{\theta'}}(w, z; t)}{p_{\boldsymbol{\theta},\boldsymbol{\theta'}}(x, z; \tau + s + t)} \\ &= p^{x,w;\tau+s}_{\boldsymbol{\theta},\boldsymbol{\theta}}(y; \tau) p^{x,z;\tau+s+t}_{\boldsymbol{\theta},\boldsymbol{\theta}}(w; \tau + s). \end{aligned}$$

Thus diffusion bridge samples at time τ can be generated simply by integrating out the corresponding samples generated at time $\tau + s$.

In view of Proposition 1, draws from the neutral transition density $p^{x,z;t}_{\boldsymbol{\theta},\boldsymbol{\theta'}}(y; s)$ can be obtained through the following procedure:

1. Draw $(M, K, L, J) \sim \{p^{(x,z,s,t,\boldsymbol{\theta},\boldsymbol{\theta'})}_{m,k,l,j} : (m, k, l, j) \in \mathbb{N}^4\}$
2. Conditional on $(M, K, L, J) = (m, k, l, j)$, draw $Y \sim \text{Beta}(\theta_1 + l + j, \theta_2 + m - l + k - j)$

Step 2 is straightforward but Step 1 is non-trivial. However, Proposition 1 guarantees that the infinite mixture weights admit computable monotone bounds, enabling exact rejection sampling even when mutation parameters differ across epochs, extending the techniques employed in Jenkins and Spanò (2017).

2.1 Non-neutral simulation

The algorithm of Jenkins and Spanò (2017, Section 5) for simulating non-neutral Wright–Fisher bridges extends naturally to the setting with changing effective population size. The key idea is to exploit the Radon–Nikodym derivative between the non-neutral and neutral diffusion laws, as given by a Girsanov

transform, and to re-interpret this quantity as (proportional to) the probability of an event that can be simulated explicitly.

2.2 Small time approximations

For very short intervals s or $t-s$, the quantities $q_m^{\theta}(s)$ and $q_k^{\theta'}(t-s)$ in (2) become numerically unstable (Griffiths, 1984; Jenkins and Spanò, 2017). In these cases we approximate the transition by a Gaussian distribution, using a central limit theorem of Griffiths (1984, Theorem 4), whenever the interval length falls below a user-defined threshold (default 0.1). Further details on the approximations used as well as implementation can be found in Section 4 of the Supplementary Information.

2.3 Implementation and output validation

We implemented our method in C++, building upon the software `EWF`. The relevant sampling routines were suitably modified to allow for non-constant demographies whilst retaining all previous functionality. The package includes Python bindings and example scripts for integrating `EWF 2.0` into population-genetic workflows.

To gauge the implementation's accuracy and correctness, we generated 10,000 samples of the trajectory of a diffusion for a range of sampling time and parameter regimes, and for comparison we numerically computed the corresponding transition density by suitably truncating the infinite sums in (2). Across all tested setups, both QQ plots and Kolmogorov–Smirnov tests showed close agreement between simulated and theoretical distributions (see Supplementary Section 5), confirming the correctness of our implementation.

Figure 1 contrasts bridge realisations under a constant and bottleneck demography. Relative to the constant-size baseline, we observe that samples generated under a bottleneck demography are noisier than those generated under a constant population size, in line with what would be expected under successively alternating high and low effective population sizes. Accounting for and explicitly modelling these shifts is essential to avoid confounding demographic effects with other genetic forces, underscoring the practical value of `EWF 2.0`.

To further illustrate the practical use of `EWF 2.0` in simulating data under realistic demographies and plausible parameter regimes, we simulated 100 candidate allele-frequency trajectories for the horse coat coloration dataset (Ludwig *et al.*, 2009), using demography inferred in Der Sarkissian *et al.* (2015) (see Section 7 in the Supplementary Information for full details). We further assumed a generation time of five years as well as biologically plausible per generation mutation and selection parameters of $\mu_1 = \mu_2 = 6.2 \times 10^{-9}$ and $\gamma = 7 \times 10^{-4}$ respectively (Lee *et al.*, 2014; Ludwig *et al.*, 2009).

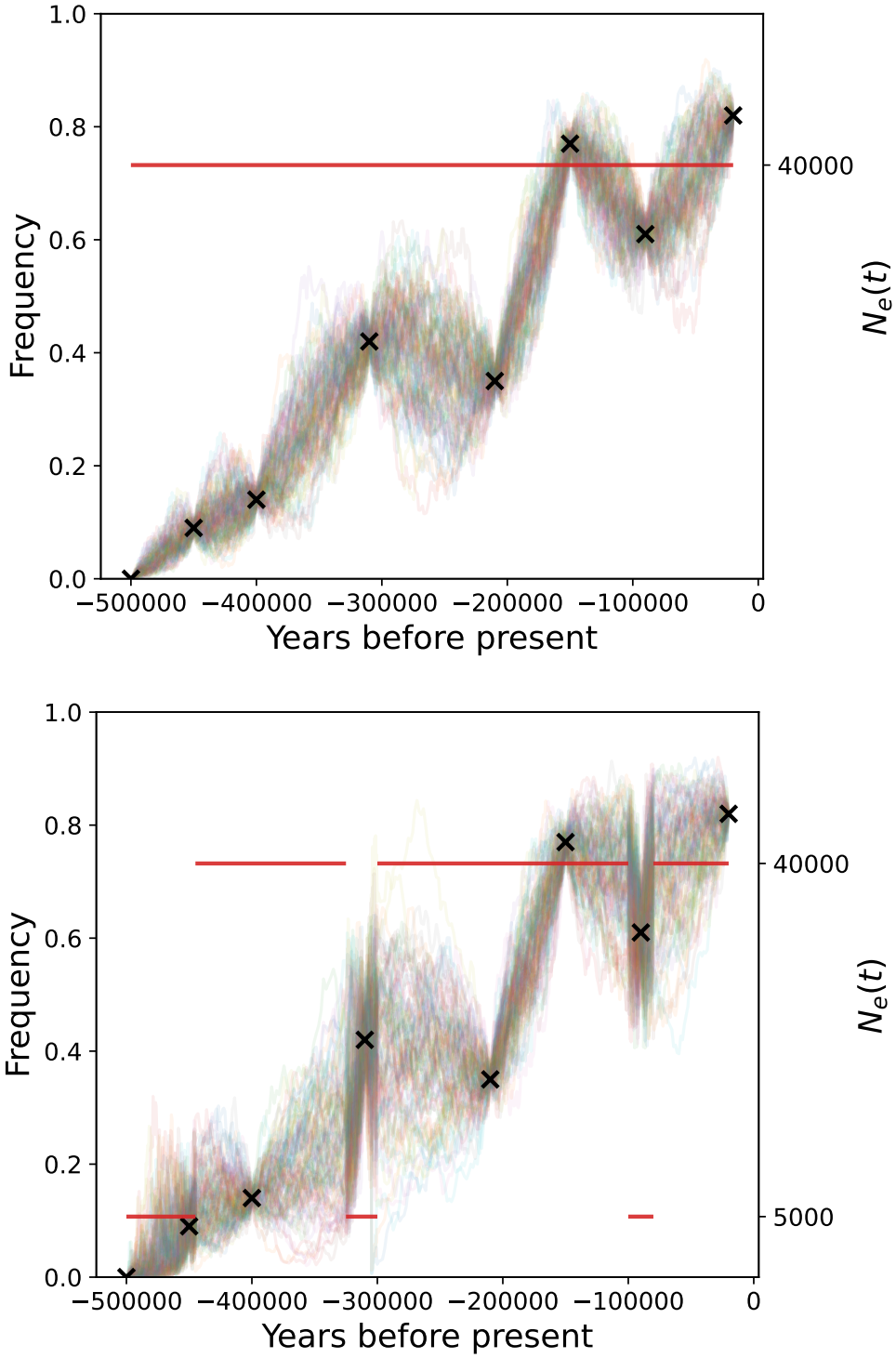

Figure 1: Bridge diffusion samples generated under a constant $N_e = 5000$ demography (top) and under a bottleneck demography (bottom). True observations are denoted by black crosses, with the corresponding demographies plotted in red.

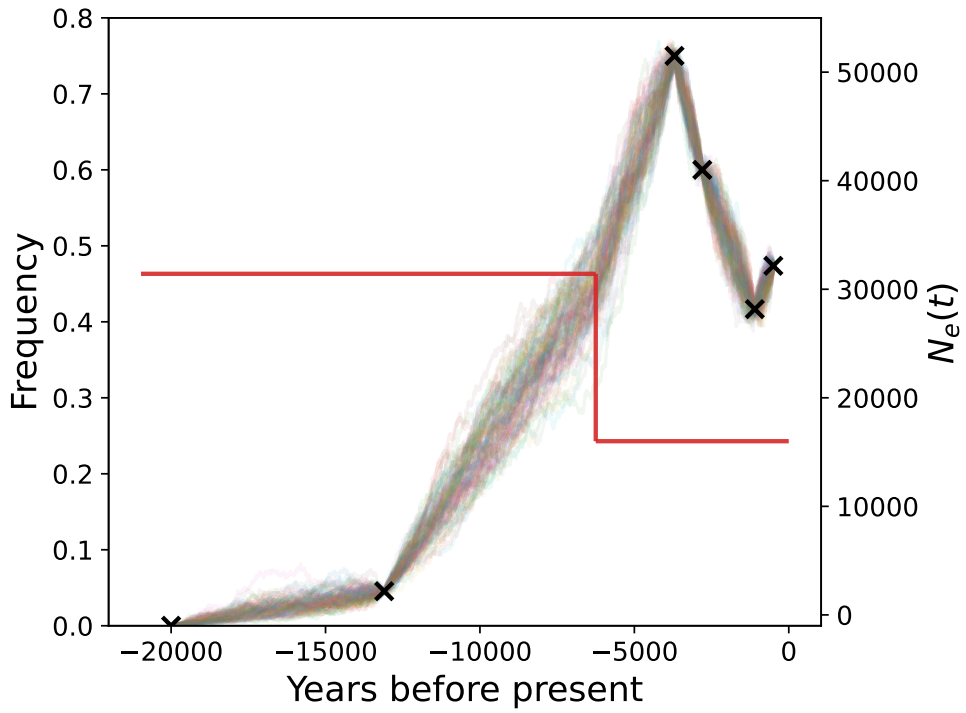

Figure 2: 100 candidate trajectories for the horse coat colouration data found in Ludwig *et al.* (2009) simulated using `EWF 2.0`, using the demography inferred in Der Sarkissian *et al.* (2015). The observed frequencies (black crosses) are assumed to be exact observations of the underlying diffusion. Note how the above simulated paths differ from those in Figure 1 in Sant *et al.* (2023), in particular how the kink in the paths observed here at 6250 years before present is absent there.

3 Conclusions

Demographic changes alter the balance of evolutionary forces shaping populations, and accurate models of allele frequency evolution must explicitly account for these dynamics. Here we have shown how exact simulation of the Wright–Fisher diffusion can be extended to settings with variable effective population size. Our software `EWF 2.0` generalizes the framework of Sant *et al.* (2023) by enabling exact draws from diffusions and diffusion bridges even when mutation and selection parameters differ across epochs. This capability is especially important for correcting ascertainment bias, as it allows conditioning on observed endpoints under realistic demographic histories.

Beyond its theoretical contribution, `EWF 2.0` provides a practical tool for population genetic simulation. It can be readily integrated into pipelines for inference from time-series data, benchmarking of approximate algorithms, and exploration of evolutionary scenarios under demographic change. With Python bindings and efficient C++ routines, it is accessible for use in both methodological and empirical studies, and is particularly timely given the increasing availability of temporal genomic datasets across diverse species (Ludwig *et al.*, 2009; Wutke *et al.*, 2016; Loog *et al.*, 2017; Fages *et al.*, 2019; Akbari *et al.*, 2024; Mallick *et al.*, 2024).

Funding

JS has been supported by MUR - Prin 2022 - Grant no. 2022CLTYP4, funded by the European Union – Next Generation EU.

Supplementary Information

1 Expressions for the bridge transition density

Consider a Wright–Fisher diffusion bridge started from $x \in (0,1)$ at time 0 and ending at $z \in (0,1)$ at time t, having mutation parameter $\boldsymbol{\theta}$ over the time interval $[0,s)$ and $\boldsymbol{\theta'}$ over $[s,s+t)$ for some $s \in (0,t)$. The density (with respect to Lebesgue measure) of a point y sampled at time s is given (Jenkins and Spanò, 2017, Proposition 2) by

$$\begin{aligned}
p^{x,z,t}_{\boldsymbol{\theta},\boldsymbol{\theta'}}(y;s) &:= \frac{p_{\boldsymbol{\theta}}(x,y;s)p_{\boldsymbol{\theta'}}(y,z;s)}{\int_0^1 p_{\boldsymbol{\theta}}(x,y;s)p_{\boldsymbol{\theta'}}(y,z;t)dy} \\
&= \sum_{m=0}^{\infty}\sum_{k=0}^{\infty}\sum_{l=0}^{m}\sum_{j=0}^{k} \frac{q^{\theta}_m(s)q^{\theta'}_k(t)\mathcal{B}_{m,x}(l)\mathcal{D}_{\theta'_1+j,\theta'_2+k-j}(z)\mathcal{DM}_{\theta_1+l,\theta_2+m-l;k}(j)\mathcal{D}_{\theta_1+l+j,\theta_2+m-l+k-j}(y)}{\int_0^1 p_{\boldsymbol{\theta}}(s,x,y)p_{\boldsymbol{\theta'}}(t,y,z)dy} \\
&=: \sum_{m=0}^{\infty}\sum_{k=0}^{\infty}\sum_{l=0}^{m}\sum_{j=0}^{k} p^{(x,z,s,t,\boldsymbol{\theta},\boldsymbol{\theta'})}_{m,k,l,j}\mathcal{D}_{\theta_1+l+j,\theta_2+m-l+k-j}(y),
\end{aligned} \tag{3}$$

where $\theta := \theta_1 + \theta_2$, $p_{\boldsymbol{\theta}}(\cdot,\cdot;\cdot)$ denotes the transition density for a Wright–Fisher diffusion with mutation parameter $\boldsymbol{\theta}$,

$$\mathcal{B}_{m,x}(l) = \binom{m}{l}x^l(1-x)^{m-l}, \qquad \mathcal{D}_{a,b}(x) = \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)},$$
$$\mathcal{DM}_{a,b;d}(c) = \binom{d}{c}\frac{B(a+c,b+d-c)}{B(a,b)}, \qquad B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)},$$

denote the usual binomial probability mass function, beta density, beta-binomial probability mass function and beta function respectively, and $\Gamma(x)$ denotes the gamma function. We observe further that

$$\int_0^1 p_{\boldsymbol{\theta}}(x,y;s)p_{\boldsymbol{\theta'}}(y,z;t)dy = \sum_{m,k=0}^{\infty} q^{\theta}_m(s)q^{\theta'}_k(t)\sum_{l=0}^{m}\mathcal{B}_{m,x}(l)\sum_{j=0}^{k}\mathcal{D}_{\theta'_1+j,\theta'_2+k-j}(z)\mathcal{DM}_{\theta_1+l,\theta_2+m-l;k}(j). \tag{4}$$

The cases corresponding to when either of x or z are at the boundary $\{0,1\}$ can be derived from (3) by taking the corresponding limit as x or z go to 0 or 1 in both the denominator and numerator. We illustrate the calculations for the case $x = 0$ and $z = 1$; all other cases follow using similar suitably modified arguments. Observe that if $x = 0$, the only non-zero term in the sum over l in (3) and (4)

corresponds to the case when $l = 0$, such that

$$
\begin{aligned}
p^{0,1;t}_{\boldsymbol{\theta},\boldsymbol{\theta'}}(y;s) &= \lim_{z\to 1} p^{0,z;t}_{\boldsymbol{\theta},\boldsymbol{\theta'}}(y;s) \\
&= \lim_{z\to 1} \sum_{m=0}^{\infty}\sum_{k=0}^{\infty}\sum_{j=0}^{k} \frac{q^{\theta}_m(s)q^{\theta'}_k(t)\mathcal{D}_{\theta'_1+j,\theta'_2+k-j}(z)\mathcal{DM}_{\theta_1,\theta_2+m;k}(j)}{\sum_{d,f=0}^{\infty} q^{\theta}_d(s)q^{\theta'}_f(t)\sum_{g=0}^{d}\mathcal{D}_{\theta'_1+g,\theta'_2+f-g}(z)\mathcal{DM}_{\theta_1,\theta_2+d;f}(j)} \\
&= \lim_{z\to 1} \sum_{m=0}^{\infty}\sum_{k=0}^{\infty} \frac{q^{\theta}_m(s)q^{\theta'}_k(t)\mathcal{DM}_{\theta_1,\theta_2+m;k}(k)\frac{(1-z)^{\theta'_2-1}}{B(\theta'_1+k,\theta'_2)}\left(1+o(1-z)\right)}{\sum_{d,f=0}^{\infty} q^{\theta}_d(s)q^{\theta'}_f(t)\mathcal{DM}_{\theta_1,\theta_2+d;f}(f)\frac{(1-z)^{\theta'_2-1}}{B(\theta'_1+f,\theta'_2)}\left(1+o(1-z)\right)} \\
&= \sum_{m=0}^{\infty}\sum_{k=0}^{\infty} \frac{q^{\theta}_m(s)q^{\theta'}_k(t)\frac{B(\theta_1+k,\theta_2+m)}{B(\theta_1,\theta_2+m)B(\theta'_1+k,\theta'_2)}}{\sum_{d,f=0}^{\infty} q^{\theta}_d(s)q^{\theta'}_f(t)\frac{B(\theta_1+f,\theta_2+d)}{B(\theta_1,\theta_2+d)B(\theta'_1+f,\theta'_2)}}
\end{aligned}
$$

All possible bridge transition densities can be found below. Note that in $p^{x,z;t}_{\boldsymbol{\theta},\boldsymbol{\theta'}}(y;s)$, the superscripts x, z denote the bridge endpoints; writing x, z means that $x, z \in (0,1)$, whereas boundary endpoints are written explicitly as 0 or 1. In the next section we show that we can obtain monotonically converging sequences of upper and lower bounds, as in Jenkins and Spanò (2017) for the case when $x, z \in (0,1)$, whereas the edge cases when either of $x, z \in \{0,1\}$ are tackled in Section 3, and follow by adapting similar arguments.

We point out that here we are only interested in cases when both $\boldsymbol{\theta}$ and $\boldsymbol{\theta'}$ have strictly positive entries, since in practice changes in the diffusion-rescaled mutation parameters are driven by corresponding changes in the effective population size rather than in the pre-limiting mutation rate itself. Thus whilst it is theoretically possible to have $\boldsymbol{\theta} = (\theta_1, \theta_2)$ with $\theta_1, \theta_2 > 0$ and $\boldsymbol{\theta'} = \mathbf{0}$, we shall not consider such scenarios as this would correspond to the case where $N_e(t) = 0$ for some time t (i.e. a population collapse, in which case a diffusion approximation is ill-suited). Nonetheless similar calculations to those presented

here and in Sant *et al.* (2023) still hold, and thus exact simulation is still possible.

$$p^{0,0;t}_{\boldsymbol{\theta},\boldsymbol{\theta'}}(y;s) = \sum_{m,k=0}^{\infty} \frac{q^{\theta}_m(s) q^{\theta'}_k(t) \frac{B(\theta_1,\theta_2+m+k)}{B(\theta_1,\theta_2+m)B(\theta'_1,\theta'_2+k)}}{\sum_{d,f=0}^{\infty} q^{\theta}_d(s) q^{\theta'}_f(t) \frac{B(\theta_1,\theta_2+d+f)}{B(\theta_1,\theta_2+d)B(\theta'_1,\theta'_2+f)}} \mathcal{D}_{\theta_1,\theta_2+m+k}(y) \tag{5}$$

$$p^{0,z;t}_{\boldsymbol{\theta},\boldsymbol{\theta'}}(y;s) = \sum_{m,k=0}^{\infty} \sum_{j=0}^{k} \frac{q^{\theta}_m(s) q^{\theta'}_k(t) \mathcal{DM}_{\theta_1,\theta_2+m;k}(j) \mathcal{D}_{\theta'_1+j,\theta'_2+k-j}(z)}{\sum_{d,f=0}^{\infty} \sum_{g=0}^{f} q^{\theta}_d(s) q^{\theta'}_f(t) \mathcal{DM}_{\theta_1,\theta_2+d;f}(g) \mathcal{D}_{\theta'_1+g,\theta'_2+f-g}(z)} \mathcal{D}_{\theta_1+j,\theta_2+m+k-j}(y) \tag{6}$$

$$p^{0,1;t}_{\boldsymbol{\theta},\boldsymbol{\theta'}}(y;s) = \sum_{m,k=0}^{\infty} \frac{q^{\theta}_m(s) q^{\theta'}_k(t) \frac{B(\theta_1+k,\theta_2+m)}{B(\theta_1,\theta_2+m)B(\theta'_1+k,\theta'_2)}}{\sum_{d,f=0}^{\infty} q^{\theta}_d(s) q^{\theta'}_f(t) \frac{B(\theta_1+f,\theta_2+d)}{B(\theta_1,\theta_2+d)B(\theta'_1+f,\theta'_2)}} \mathcal{D}_{\theta_1+k,\theta_2+m}(y) \tag{7}$$

$$p^{1,0;t}_{\boldsymbol{\theta},\boldsymbol{\theta'}}(y;s) = \sum_{m,k=0}^{\infty} \frac{q^{\theta}_m(s) q^{\theta'}_k(t) \frac{B(\theta_1+m,\theta_2+k)}{B(\theta_1+m,\theta_2)B(\theta'_1,\theta'_2+k)}}{\sum_{d,f=0}^{\infty} q^{\theta}_d(s) q^{\theta'}_f(t) \frac{B(\theta_1+d,\theta_2+f)}{B(\theta_1+d,\theta_2)B(\theta'_1,\theta'_2+f)}} \mathcal{D}_{\theta_1+m,\theta_2+k}(y) \tag{8}$$

$$p^{1,z;t}_{\boldsymbol{\theta},\boldsymbol{\theta'}}(y;s) = \sum_{m,k=0}^{\infty} \sum_{j=0}^{k} \frac{q^{\theta}_m(s) q^{\theta}_k(t) \mathcal{DM}_{\theta_1+m,\theta_2;k}(j) \mathcal{D}_{\theta'_1+j,\theta'_2+k-j}(z)}{\sum_{d,f=0}^{\infty} \sum_{g=0}^{f} q^{\theta}_d(s) q^{\theta'}_f(t) \mathcal{DM}_{\theta_1+d,\theta_2;f}(g) \mathcal{D}_{\theta'_1+g,\theta'_2+f-g}(z)} \mathcal{D}_{\theta_1+m+j,\theta_2+k-j}(y) \tag{9}$$

$$p^{1,1;t}_{\boldsymbol{\theta},\boldsymbol{\theta'}}(y;s) = \sum_{m,k=0}^{\infty} \frac{q^{\theta}_m(s) q^{\theta'}_k(t) \frac{B(\theta_1+m+k,\theta_2)}{B(\theta_1+m,\theta_2)B(\theta'_1+k,\theta'_2)}}{\sum_{d,f=0}^{\infty} q^{\theta}_d(s) q^{\theta'}_f(t) \frac{B(\theta_1+d+f,\theta_2)}{B(\theta_1+d,\theta_2)B(\theta'_1+f,\theta'_2)}} \mathcal{D}_{\theta_1+m+k,\theta_2}(y) \tag{10}$$

$$p^{x,0;t}_{\boldsymbol{\theta},\boldsymbol{\theta'}}(y;s) = \sum_{m,k=0}^{\infty} \sum_{l=0}^{m} \frac{q^{\theta}_m(s) q^{\theta}_k(t) \mathcal{B}_{m,x}(l) \frac{B(\theta_1+l,\theta_2+m-l+k)}{B(\theta_1+l,\theta_2+m-l)B(\theta'_1,\theta'_2+k)}}{\sum_{d,f=0}^{\infty} \sum_{e=0}^{d} q^{\theta}_d(s) q^{\theta'}_f(t) \mathcal{B}_{d,x}(e) \frac{B(\theta_1+e,\theta_2+d-e+f)}{B(\theta_1+e,\theta_2+d-e)B(\theta'_1,\theta'_2+f)}} \mathcal{D}_{\theta_1+l,\theta_2+m-l+k}(y) \tag{11}$$

$$p^{x,z;t}_{\boldsymbol{\theta},\boldsymbol{\theta'}}(y;s) = \sum_{m,k=0}^{\infty} \sum_{l,j=0}^{m,k} \frac{q^{\theta}_m(s) q^{\theta}_k(t) \mathcal{B}_{m,x}(l) \mathcal{DM}_{\theta_1+l,\theta_2+m-l;k}(j) \mathcal{D}_{\theta'_1+j,\theta'_2+k-j}(z)}{p_{\boldsymbol{\theta},\boldsymbol{\theta'}}(x,z,s,t)} \mathcal{D}_{\theta_1+l+j,\theta_2+m-l+k-j}(y) \tag{12}$$

$$p^{x,1;t}_{\boldsymbol{\theta},\boldsymbol{\theta'}}(y;s) = \sum_{m,k=0}^{\infty} \sum_{l=0}^{m} \frac{q^{\theta}_m(s) q^{\theta}_k(t) \mathcal{B}_{m,x}(l) \frac{B(\theta_1+l+k,\theta_2+m-l)}{B(\theta_1+l,\theta_2+m-l)B(\theta'_1+k,\theta'_2)}}{\sum_{d,f=0}^{\infty} \sum_{e=0}^{d} q^{\theta}_d(s) q^{\theta'}_f(t) \mathcal{B}_{d,x}(e) \frac{B(\theta_1+e+f,\theta_2+d-e)}{B(\theta_1+e,\theta_2+d-e)B(\theta'_1+f,\theta'_2)}} \mathcal{D}_{\theta_1+l+k,\theta_2+m-l}(y) \tag{13}$$

2 Simulating exact draws from the density (12)

We can adopt the same approach as in Proposition 4 in Jenkins and Spanò (2017) to obtain monotonically converging sequences of upper and lower bounds for the numerator in (12), however the denominator differs significantly from the one present in Jenkins and Spanò (2017). Here we show how such bounds can

be obtained in the case when $x, z \in (0,1)$ via Proposition 2 below, whilst in Section 3 we extend these arguments to the remaining cases.

As in (Jenkins and Spanò, 2017, Section 3), these monotonically converging sequences of upper and lower bounds then allow us to generate draws $(m,k,l,j) \in \mathbb{N}^4$ (or a suitable subspace of $\mathbb{N}^4$), from which we can obtain samples from the corresponding bridge diffusion by considering an appropriate draw from a Beta distribution.

In view of the infinite series expansion (Griffiths, 1979; Tavaré, 1984)

$$q_m^{\theta}(s) = \sum_{h=m}^{\infty} \frac{(\theta+2h-1)\,(\theta+m)_{(h-1)}}{m!(h-m)!} e^{-\frac{h(h+\theta-1)s}{2}},$$

observe that we can re-write

$$\begin{aligned}
&\sum_{m=0}^{\infty}\sum_{k=0}^{\infty}\sum_{l=0}^{m}\sum_{j=0}^{k} q_m^{\theta}(s) q_k^{\theta'}(t) \mathcal{B}_{m,x}(l) \mathcal{D}_{\theta'_1+j,\theta'_2+k-j}(z) \mathcal{DM}_{\theta_1+l,\theta_2+m-l;k}(j) \\
&\quad = \sum_{m,k=0}^{\infty}\sum_{h=m}^{\infty} (-1)^{h-m} a_{hm}^{\theta} e^{-\frac{h(h+\theta-1)s}{2}} \sum_{g=k}^{\infty} (-1)^{g-k} a_{gk}^{\theta'} e^{-\frac{g(g+\theta'-1)t}{2}} \mathbb{E}_{m,k}\left[\mathcal{D}_{\theta'_1+J,\theta'_2+k-J}(z)\right] \qquad (14)
\end{aligned}$$

where $a_{hm}^{\theta} := \frac{(\theta+2h-1)(\theta+m)_{(h-1)}}{m!(h-m)!}$, with $a_{(n)} := \Gamma(a+n)/\Gamma(a)$ for $a > 0$ and $n \geq -1$, and the expectation is with respect to the pair of random variables (L, J) such that $L \sim \mathrm{Bin}(m, x)$ and $J|L = l \sim \mathrm{BetaBin}(k, \theta_1 + l, \theta_2 + m - l)$.

Define $\lambda := m + k$ and $\gamma := h + g$, then combining (4) with (14) allows us to deduce that

$$\begin{aligned}
\int_0^1 p_{\boldsymbol{\theta}}(t,x,y) p_{\boldsymbol{\theta'}}(s,y,z) dy &= \sum_{\lambda=0}^{\infty}\sum_{\gamma=\lambda}^{\infty} (-1)^{\gamma-\lambda} A_{\gamma,\lambda}^{(x,z,s,t,\boldsymbol{\theta},\boldsymbol{\theta'})} \\
A_{\gamma,\lambda}^{(x,z,s,t,\boldsymbol{\theta},\boldsymbol{\theta'})} &:= \sum_{(m,k,h,g)\in\mathcal{S}_{\gamma,\lambda}} a_{hm}^{\theta} a_{gk}^{\theta'} e^{-\frac{h(h+\theta-1)s}{2}} e^{-\frac{g(g+\theta'-1)t}{2}} \mathbb{E}_{m,k}\left[\mathcal{D}_{\theta'_1+J,\theta'_2+k-J}(z)\right], \\
\mathcal{S}_{\gamma,\lambda} &:= \{(m,k,h,g) : m+k=\lambda,\ \ h+g=\gamma,\ \ h \geq m,\ \ g \geq k\}.
\end{aligned}$$

In what follows, we shall drop the superscript in $A_{\gamma,\lambda}^{(x,z,s,t,\boldsymbol{\theta},\boldsymbol{\theta'})}$ for ease of exposition.

We further define $d_{2\gamma} := \sum_{j=0}^{\gamma} A_{\gamma+j,\gamma-j}$ and $d_{2\gamma+1} := \sum_{j=0}^{\gamma} A_{\gamma+j+1,\gamma-j}$, and observe that (4) can be re-written as the alternating series,

$$\int_0^1 p_{\boldsymbol{\theta}}(t,x,y) p_{\boldsymbol{\theta'}}(s,y,z) dy = \sum_{\lambda=0}^{\infty}\sum_{\gamma=\lambda}^{\infty} (-1)^{\gamma-\lambda} A_{\gamma,\lambda} = \sum_{i=0}^{\infty} (-1)^i d_i.$$

Thus we can in principle adapt Proposition 3 in Jenkins and Spanò (2017) to target this quantity through monotonically converging sequences of upper and lower bounds, namely we want to show that

$$\text{for all } \gamma \text{ large enough, it holds that } d_{2\gamma+2} < d_{2\gamma+1} < d_{2\gamma} \tag{15}$$

Because of a potential change in mutation parameters at time t, adapting their Proposition is not straight forward since we lose the ability to 'group' terms in a convenient way and the argument must be redone as follows.

For a given $\varepsilon \in (0,1)$ fixed, define

$$\begin{aligned}
C_m^{(\boldsymbol{\theta},s)} &:= \inf\left\{i \geq 0 : b_{i+m+1}^{(\boldsymbol{\theta},s)}(m) < b_{i+m}^{(\boldsymbol{\theta},s)}(m)\right\} \\
C^{(\boldsymbol{\theta},s)}(\gamma) &:= \max_{m\in\{0,\dots,\gamma\}}\left\{C_m^{(\boldsymbol{\theta},s)}\right\} \\
D_\varepsilon^{(\boldsymbol{\theta},s)} &:= \inf\left\{k \geq \max\left\{\left(\frac{1}{t} - \frac{\theta+1}{2}\right), 0\right\} : (\theta+2k+1)\, e^{\frac{-(2k+\theta)t}{2}} < 1-\varepsilon\right\} \\
F_m^{(\boldsymbol{\theta},s)} &:= \inf\left\{i \geq m : 2b_{i+m+1}^{(\boldsymbol{\theta},s)}(m) < b_{i+m}^{(\boldsymbol{\theta},s)}(m)\right\} \\
F^{(\boldsymbol{\theta},s)}(\gamma) &:= \max_{m\in\{0,\dots,\gamma\}}\left\{F_m^{(\boldsymbol{\theta},s)}\right\} \\
K^{(x,\boldsymbol{\theta})} &:= \max\left\{\frac{\theta}{\theta_2}, \frac{1+\theta}{1-x}\right\} + \left(1+\frac{1}{x}\right)\max\left\{\frac{x\theta+1}{\theta_1}, 1\right\}.
\end{aligned} \tag{16}$$

Then we have the following result:

Proposition 2. *Let $A_{\gamma,\lambda}$, $d_{2\gamma}$, $d_{2\gamma+1}$, $D_\varepsilon^{(\boldsymbol{\theta},s)}$, $F^{(\boldsymbol{\theta},s)}(\gamma)$, $K^{(x,\boldsymbol{\theta})}$ be defined as above, and let $\varepsilon \in (0,1)$ be some constant. For*

$$\gamma > G^{(x,z,s,t,\boldsymbol{\theta},\boldsymbol{\theta'})} := 2\max\left\{F^{(\boldsymbol{\theta},s)}(\gamma), F^{(\boldsymbol{\theta'},t)}(\gamma), \frac{K^{(x,\boldsymbol{\theta})}}{\varepsilon}, \frac{K^{(z,\boldsymbol{\theta'})}}{\varepsilon}, D_\varepsilon^{(\boldsymbol{\theta},s)}, D_\varepsilon^{(\boldsymbol{\theta'},t)}\right\}$$

we have that $d_{2\gamma+2} < d_{2\gamma+1} < d_{2\gamma}$.

Proof. We start by showing that $d_{2\gamma+1} < d_{2\gamma}$. For this to be the case, we can expand out the $A_{\lambda,\gamma}$ terms to deduce that

$$\begin{aligned}
d_{2\gamma+1} < d_{2\gamma} \iff \sum_{j=0}^{\gamma} A_{\gamma+j+1,\gamma-j} < \sum_{j=0}^{\gamma} A_{\gamma+j,\gamma-j} &\Longleftarrow A_{\gamma+j+1,\gamma-j} < A_{\gamma+j,\gamma-j} \text{ holds } \forall j = 0,\dots,\gamma \\
\iff \sum_{(m,k,h,g)\in\mathcal{S}_{\gamma+j+1,\gamma-j}} & b_h^{\theta,s}(m) b_g^{\theta',t}(k)\mathbb{E}_{m,k}\left[\mathcal{D}_{\theta_1'+J,\theta_2'+k-J}(z)\right] \\
&< \sum_{(m,k,h,g)\in\mathcal{S}_{\gamma+j,\gamma-j}} b_h^{\theta,s}(m) b_g^{\theta',t}(k)\mathbb{E}_{m,k}\left[\mathcal{D}_{\theta_1'+J,\theta_2'+k-J}(z)\right] \text{ holds } \forall j = 0,\dots,\gamma \\
\Longleftarrow \sum_{h+g=\gamma+j+1} b_h^{\theta,s}(m) b_g^{\theta',t}(\gamma-j-m) &< \sum_{h+g=\gamma+j} b_h^{\theta,s}(m) b_g^{\theta',t}(\gamma-j-m) \\
&\text{holds } \forall j = 0,\dots,\gamma,\ \ \forall m = 0,\dots,\gamma-j
\end{aligned} \tag{17}$$

where $b_h^{\theta,s}(m) := a_{hm}^{\theta} e^{-\frac{h(h+\theta-1)s}{2}}$.

Consider for the moment a fixed m and j, and assume that $\gamma + j + 1$ is even (if not similar calculations hold, with suitably altered indexing). We can re-write

$$\begin{aligned}
&\sum_{h+g=\gamma+j+1} b_h^{\theta,s}(m) b_g^{\theta',t}(\gamma - j - m) \\
&\quad = \sum_{h=\lfloor \frac{\gamma+j+1}{2} \rfloor + 2}^{\gamma+j+1} b_{\gamma+j+1-h}^{(\theta,s)}(m) b_h^{(\theta',t)}(\gamma - j - m) + \sum_{h=\lfloor \frac{\gamma+j+1}{2} \rfloor + 1}^{\gamma+j+1} b_h^{(\theta,s)}(m) b_{\gamma+j+1-h}^{(\theta',t)}(\gamma - j - m) \\
&\quad + b_{\lfloor \frac{\gamma+j+1}{2} \rfloor - 1}^{(\theta,s)}(m) b_{\lfloor \frac{\gamma+j+1}{2} \rfloor + 1}^{(\theta',t)}(\gamma - j - m) + b_{\lfloor \frac{\gamma+j+1}{2} \rfloor}^{(\theta,s)}(m) b_{\lfloor \frac{\gamma+j+1}{2} \rfloor}^{(\theta',t)}(\gamma - j - m) \\
&\quad =: A + B + C.
\end{aligned}$$

We deal with each of these three terms in turn.

Term C:

Recall that for any k, m we have that

$$b_{k+1}^{(\theta,s)}(m) / b_k^{(\theta,s)}(m) \leq (\theta + 2k + 1) e^{-\frac{(2k+\theta)s}{2}} \tag{18}$$

(see (26) and (27) in Jenkins and Spanò (2017)), from which we get that $b_{k+1}^{(\theta,s)}(m)/b_k^{(\theta,s)}(m) < 1/2$ for $k > F_m^{(\theta,s)}$. In particular, if γ is such that $\lfloor \frac{\gamma}{2} \rfloor \geq \max\{F^{(\theta',t)}(\gamma), F^{(\theta,s)}(\gamma)\}$, then

$$\begin{aligned}
&b_{\lfloor \frac{\gamma+j+1}{2} \rfloor - 1}^{(\theta,s)}(m) b_{\lfloor \frac{\gamma+j+1}{2} \rfloor + 1}^{(\theta',t)}(\gamma - j - m) + b_{\lfloor \frac{\gamma+j+1}{2} \rfloor}^{(\theta,s)}(m) b_{\lfloor \frac{\gamma+j+1}{2} \rfloor}^{(\theta',t)}(\gamma - j - m) \\
&\quad < \frac{1}{2} b_{\lfloor \frac{\gamma+j+1}{2} \rfloor - 1}^{(\theta,s)}(m) b_{\lfloor \frac{\gamma+j+1}{2} \rfloor}^{(\theta',t)}(\gamma - j - m) + \frac{1}{2} b_{\lfloor \frac{\gamma+j+1}{2} \rfloor - 1}^{(\theta,s)}(m) b_{\lfloor \frac{\gamma+j+1}{2} \rfloor}^{(\theta',t)}(\gamma - j - m) \\
&\quad = b_{\lfloor \frac{\gamma+j+1}{2} \rfloor - 1}^{(\theta,s)}(m) b_{\lfloor \frac{\gamma+j+1}{2} \rfloor}^{(\theta',t)}(\gamma - j - m)
\end{aligned} \tag{19}$$

holds independently of m and j. Recall that if $b_{k+1}^{(\theta,s)}(m) < b_k^{(\theta,s)}(m)/2$ holds for some k, it must also hold for all values greater than k (see display before eq (27) in Jenkins and Spanò (2017)).

Term B:

If γ is such that $\lfloor \frac{\gamma}{2} \rfloor > C_m^{(\theta,s)}$, then $b_{h+1}^{(\theta,s)}(m) < b_h^{(\theta,s)}(m)$ for $h \in \{\lfloor \frac{\gamma+j+1}{2} \rfloor, \ldots, \gamma + j + 1\}$. Recall we have fixed m and j here, but we require (17) to hold $\forall m = 0, \ldots, \gamma - j$ and $\forall j = 0, \ldots, \gamma$. Thus if we

choose $\lfloor \frac{\gamma}{2} \rfloor > F^{(\theta,s}(\gamma) > C^{(\theta,s)}(\gamma)$, we get that

$$\sum_{h=\lfloor \frac{\gamma+j+1}{2} \rfloor+1}^{\gamma+j+1} b_h^{(\theta,s)}(m) b_{\gamma+j+1-h}^{(\theta',t)}(\gamma-j-m) < \sum_{h=\lfloor \frac{\gamma+j+1}{2} \rfloor+1}^{\gamma+j+1} b_{h-1}^{(\theta,s)}(m) b_{\gamma+j+1-h}^{(\theta',t)}(\gamma-j-m) \tag{20}$$

$$= \sum_{h=\lfloor \frac{\gamma+j+1}{2} \rfloor}^{\gamma+j} b_h^{(\theta,s)}(m) b_{\gamma+j-h}^{(\theta',t)}(\gamma-j-m) \tag{21}$$

Term A:

We can apply a similar argument to the above, namely: choose γ such that $\lfloor \frac{\gamma}{2} \rfloor > F^{(\theta',t)}(\gamma)$, then $b_{h+1}^{(\theta',t)}(\gamma-j-m) < b_h^{(\theta',t)}(\gamma-j-m)$ holds for all $h \in \{\lfloor \frac{\gamma+j+1}{2} \rfloor + 2, \gamma+j+1\}$, and thus

$$\sum_{h=\lfloor \frac{\gamma+j+1}{2} \rfloor+2}^{\gamma+j+1} b_{\gamma+j+1-h}^{(\theta,s)}(m) b_h^{(\theta',t)}(\gamma-j-m) < \sum_{h=\lfloor \frac{\gamma+j+1}{2} \rfloor+2}^{\gamma+j+1} b_{\gamma+j+1-h}^{(\theta,s)}(m) b_{h-1}^{(\theta',t)}(\gamma-j-m)$$

$$= \sum_{h=\lfloor \frac{\gamma+j+1}{2} \rfloor+1}^{\gamma+j} b_{\gamma+j-h}^{(\theta,s)}(m) b_h^{(\theta',t)}(\gamma-j-m). \tag{22}$$

We point out that since $C_m^{(\theta,s)} = 0$ for all $m > D_0^{(\theta,s)}$ as per Proposition 1(iii) in Jenkins and Spanò (2017), the worst case scenario would be taking $\gamma > 2\max\{D_0^{(\theta,s)}, D_0^{(\theta',t)}\}$ to ensure both (20) and (22) hold.

So if we choose γ such that $\gamma > 2\max\{F^{(\theta,s)}(\gamma), F^{(\theta',t)}(\gamma)\}$, we get that we can combine (19), (20),

and (22) to deduce that

$$
\begin{aligned}
\sum_{h+g=\gamma+j+1} b_h^{\theta,s}(m) b_g^{\theta',t}(\gamma-j-m) &= \sum_{h=0}^{\gamma+j+1} b_h^{(\theta,z)}(m) b_{\gamma+j+1-h}^{(\theta',t)}(\gamma-j-m) \\
&= \sum_{h=0}^{\lfloor\frac{\gamma+j+1}{2}\rfloor-2} b_h^{(\theta,z)}(m) b_{\gamma+j+1-h}^{(\theta',t)}(\gamma-j-m) + \sum_{h=\lfloor\frac{\gamma+j+1}{2}\rfloor+1}^{\gamma+j+1} b_h^{(\theta,s)}(m) b_{\gamma+j+1-h}^{(\theta',t)}(\gamma-j-m) \\
&+ \quad b_{\lfloor\frac{\gamma+j+1}{2}\rfloor-1}^{(\theta,s)}(m) b_{\lfloor\frac{\gamma+j+1}{2}\rfloor+1}^{(\theta',t)}(\gamma-j-m) + b_{\lfloor\frac{\gamma+j+1}{2}\rfloor}^{(\theta,s)}(m) b_{\lfloor\frac{\gamma+j+1}{2}\rfloor}^{(\theta',t)}(\gamma-j-m) \\
&= \sum_{h=\lfloor\frac{\gamma+j+1}{2}\rfloor+2}^{\gamma+j+1} b_{\gamma+j+1-h}^{(\theta,s)}(m) b_h^{(\theta',t)}(\gamma-j-m) + \sum_{h=\lfloor\frac{\gamma+j+1}{2}\rfloor+1}^{\gamma+j+1} b_h^{(\theta,s)}(m) b_{\gamma+j+1-h}^{(\theta',t)}(\gamma-j-m) \\
&+ \quad b_{\lfloor\frac{\gamma+j+1}{2}\rfloor-1}^{(\theta,s)}(m) b_{\lfloor\frac{\gamma+j+1}{2}\rfloor+1}^{(\theta',t)}(\gamma-j-m) + b_{\lfloor\frac{\gamma+j+1}{2}\rfloor}^{(\theta,s)}(m) b_{\lfloor\frac{\gamma+j+1}{2}\rfloor}^{(\theta',t)}(\gamma-j-m) \\
&< \sum_{h=\lfloor\frac{\gamma+j+1}{2}\rfloor+1}^{\gamma+j} b_{\gamma+j-h}^{(\theta,s)}(m) b_h^{(\theta',t)}(\gamma-j-m) + \sum_{h=\lfloor\frac{\gamma+j+1}{2}\rfloor}^{\gamma+j} b_h^{(\theta,s)}(m) b_{\gamma+j-h}^{(\theta',t)}(\gamma-j-m) \\
&\quad + b_{\lfloor\frac{\gamma+j+1}{2}\rfloor-1}^{(\theta,s)}(m) b_{\lfloor\frac{\gamma+j+1}{2}\rfloor}^{(\theta',t)}(\gamma-j-m) \\
&= \sum_{h=0}^{\lfloor\frac{\gamma+j+1}{2}\rfloor-2} b_h^{(\theta,s)}(m) b_{\gamma+j-h}^{(\theta',t)}(\gamma-j-m) + \sum_{h=\lfloor\frac{\gamma+j+1}{2}\rfloor}^{\gamma+j} b_h^{(\theta,s)}(m) b_{\gamma+j-h}^{(\theta',t)}(\gamma-j-m) \\
&\quad + b_{\lfloor\frac{\gamma+j+1}{2}\rfloor-1}^{(\theta,s)}(m) b_{\lfloor\frac{\gamma+j+1}{2}\rfloor}^{(\theta',t)}(\gamma-j-m) \\
&= \sum_{h=0}^{\gamma+j} b_h^{(\theta,s)}(m) b_{\gamma+j-h}^{(\theta',t)}(\gamma-j-m) = \sum_{h+g=\gamma+j} b_h^{\theta,s}(m) b_g^{\theta',t}(\gamma-j-m)
\end{aligned}
$$

and thus $d_{2\gamma+1} < d_{2\gamma}$ holds from (17).

It now remains to show that for sufficiently large γ we have that $d_{2\gamma+2} < d_{2\gamma+1}$. Again we use a similar approach to that in the proof of Proposition 3 in Jenkins and Spanò (2017), namely because we have shown that

$$A_{\gamma+j+1,\gamma-j} < A_{\gamma+j,\gamma-j} \ \ \forall j = 0, \ldots, \gamma,$$

we can increment the first index by 1 and sum over $j = 1, \ldots, \gamma$ to deduce

$$\sum_{j=1}^{\gamma} A_{\gamma+1+j+1,\gamma-j} < \sum_{j=1}^{\gamma} A_{\gamma+1+j,\gamma-j}.$$

Thus it remains to show that

$$A_{\gamma+1,\gamma+1} + A_{\gamma+2,\gamma} < A_{\gamma+1,\gamma},$$

for then it follows that $d_{2\gamma+2} < d_{2\gamma+1}$. Observe that if $\lfloor \frac{\gamma}{2} \rfloor \geq D_\varepsilon^{(\theta,s)}$, then

$$\begin{aligned}\sum_{h=\lfloor \frac{\gamma+2}{2} \rfloor+1}^{\gamma+2} b_h^{(\theta,s)}(m) b_{\gamma+2-h}^{(\theta',t)}(\gamma-m) &< (1-\varepsilon) \sum_{h=\lfloor \frac{\gamma+2}{2} \rfloor+1}^{\gamma+2} b_{h-1}^{(\theta,s)}(m) b_{\gamma+2-h}^{(\theta',t)}(\gamma-m) \\ &= (1-\varepsilon) \sum_{h=\lfloor \frac{\gamma+2}{2} \rfloor}^{\gamma+1} b_h^{(\theta,s)}(m) b_{\gamma+1-h}^{(\theta',t)}(\gamma-m), \end{aligned} \tag{23}$$

following the same reasoning as in the display above equation (34) in Jenkins and Spanò (2017). We can repeat this argument with $\lfloor \frac{\gamma}{2} \rfloor \geq D_\varepsilon^{(\theta',t)}$ to get

$$\sum_{h=\lfloor \frac{\gamma+2}{2} \rfloor+2}^{\gamma+2} b_{\gamma+2-h}^{(\theta,s)}(m) b_h^{(\theta',t)}(\gamma-m) < (1-\varepsilon) \sum_{h=\lfloor \frac{\gamma+2}{2} \rfloor+1}^{\gamma+1} b_{\gamma+1-h}^{(\theta,s)}(m) b_h^{(\theta',t)}(\gamma-m). \tag{24}$$

Finally, for $\lfloor \frac{\gamma}{2} \rfloor \geq \max\{F^{(\theta,s)}(\gamma), F^{(\theta',s)}(\gamma)\}$, using the same reasoning that led to (19), it follows that

$$b_{\lfloor \frac{\gamma+2}{2} \rfloor-1}^{(\theta,s)}(m) b_{\lfloor \frac{\gamma+2}{2} \rfloor+1}^{(\theta',t)}(\gamma-m) + b_{\lfloor \frac{\gamma+2}{2} \rfloor}^{(\theta,s)}(m) b_{\lfloor \frac{\gamma+2}{2} \rfloor}^{(\theta',t)}(\gamma-m) < b_{\lfloor \frac{\gamma+2}{2} \rfloor-1}^{(\theta,s)}(m) b_{\lfloor \frac{\gamma+2}{2} \rfloor}^{(\theta',t)}(\gamma-m), \tag{25}$$

and thus summing (23), (24) and (25) gives

$$A_{\gamma+2,\gamma} < (1-\varepsilon) A_{\gamma+1,\gamma}. \tag{26}$$

In view of (26), it suffices to show that

$$A_{\gamma+1,\gamma+1} < \varepsilon A_{\gamma+1,\gamma} \tag{27}$$

To this end, observe that

$$\begin{aligned} A_{\gamma+1,\gamma+1} &= \sum_{(m,k,h,g)\in\mathcal{S}_{\gamma+1,\gamma+1}} b_h^{(\theta,s)}(m) b_g^{(\theta',t)}(k) \mathbb{E}_{m,k}\left[\mathcal{D}_{\theta_1'+J,\theta_2'+k-J}(z)\right] \\ &= \sum_{m=0}^{\gamma+1} b_m^{(\theta,s)}(m) b_{\gamma+1-m}^{(\theta',t)}(\gamma+1-m) \mathbb{E}_{m,\gamma+1-m}\left[\mathcal{D}_{\theta_1'+J,\theta_2'+\gamma+1-m-J}(z)\right] \end{aligned} \tag{28}$$

since $\gamma+1-h = g \geq k = \gamma+1-m$, and similarly

$$\begin{aligned} A_{\gamma+1,\gamma} &= \sum_{(m,k,h,g)\in\mathcal{S}_{\gamma+1,\gamma}} b_h^{(\theta,s)}(m) b_g^{(\theta',t)}(k) \mathbb{E}_{m,k}\left[\mathcal{D}_{\theta_1'+J,\theta_2'+k-J}(z)\right] \\ &= \sum_{m=0}^{\gamma} \left(b_m^{(\theta,s)}(m) b_{\gamma+1-m}^{(\theta',t)}(\gamma-m) + b_{m+1}^{(\theta,s)}(m) b_{\gamma-m}^{(\theta',t)}(\gamma-m) \right) \mathbb{E}_{m,\gamma-m}\left[\mathcal{D}_{\theta_1'+J,\theta_2'+\gamma-m-J}(z)\right] \end{aligned} \tag{29}$$

since $\gamma+1-h = g \geq k = \gamma-m$.

In view of Lemma 3, we have that for $m \in \{0, \ldots, \lfloor \frac{\gamma+1}{2} \rfloor\}$

$$\frac{b^{(\theta',s)}_{\gamma+1-m}(\gamma+1-m)}{b^{(\theta',s)}_{\gamma+1-m}(\gamma-m)} \frac{\mathbb{E}_{m,\gamma+1-m}\left[\mathcal{D}_{\theta'_1+J,\theta'_2+\gamma+1-m-J}(z)\right]}{\mathbb{E}_{m,\gamma-m}\left[\mathcal{D}_{\theta'_1+J,\theta'_2+\gamma-m-J}(z)\right]} < \frac{1}{k+1}\frac{\theta'+2k}{\theta'+k}K^{(z,\boldsymbol{\theta'})} \tag{30}$$

and that for $m \in \{\lfloor \frac{\gamma+1}{2} \rfloor, \ldots, \gamma\}$

$$\frac{b^{(\theta,t)}_{m+1}(m+1)}{b^{(\theta,t)}_{m+1}(m)} \frac{\mathbb{E}_{m+1,\gamma-m}\left[\mathcal{D}_{\theta'_1+J,\theta'_2+k-J}(z)\right]}{\mathbb{E}_{m,\gamma-m}\left[\mathcal{D}_{\theta'_1+J,\theta'_2+k-J}(z)\right]} < \frac{1}{m+1}\frac{\theta+2m}{\theta+m}K^{(x,\boldsymbol{\theta})}. \tag{31}$$

where we recall that $k = \gamma - m$. Observe that for $m \leq \lfloor \frac{\gamma+1}{2} \rfloor$, we also have that $k = \gamma - m \geq \lfloor \frac{\gamma}{2} \rfloor > K^{(z,\boldsymbol{\theta'})}/\varepsilon$, such that for the RHS of (30) we have

$$\frac{1}{k+1}\frac{\theta'+2k}{\theta'+k}K^{(z,\boldsymbol{\theta'})} \leq \frac{2}{k+1}K^{(z,\boldsymbol{\theta'})} < \varepsilon.$$

Similarly, for (31), $m \geq \lfloor \frac{\gamma+1}{2} \rfloor > K^{(x,\boldsymbol{\theta})}/\varepsilon$, such that

$$\frac{1}{m+1}\frac{\theta+2m}{\theta+m}K^{(x,\boldsymbol{\theta})} < \varepsilon.$$

Thus (27) follows since for $\lfloor \frac{\gamma}{2} \rfloor > \max\{K^{(x,\boldsymbol{\theta})}, K^{(z,\boldsymbol{\theta'})}\}/\varepsilon$, we have that

$$\begin{aligned}
A_{\gamma+1,\gamma+1} &= \sum_{m=0}^{\lfloor \frac{\gamma+1}{2} \rfloor} b^{(\theta,s)}_m(m) b^{(\theta',t)}_{\gamma+1-m}(\gamma+1-m)\mathbb{E}_{m,\gamma+1-m}\left[\mathcal{D}_{\theta'_1+J,\theta'_2+\gamma+1-m-J}(z)\right] \\
&\quad + \sum_{m=\lfloor \frac{\gamma+1}{2} \rfloor+1}^{\gamma+1} b^{(\theta,s)}_m(m) b^{(\theta',t)}_{\gamma+1-m}(\gamma+1-m)\mathbb{E}_{m,\gamma+1-m}\left[\mathcal{D}_{\theta'_1+J,\theta'_2+\gamma+1-m-J}(z)\right] \\
&= \sum_{m=0}^{\lfloor \frac{\gamma+1}{2} \rfloor} b^{(\theta,s)}_m(m) b^{(\theta',t)}_{\gamma+1-m}(\gamma+1-m)\mathbb{E}_{m,\gamma+1-m}\left[\mathcal{D}_{\theta'_1+J,\theta'_2+\gamma+1-m-J}(z)\right] \\
&\quad + \sum_{m=\lfloor \frac{\gamma+1}{2} \rfloor}^{\gamma} b^{(\theta,s)}_{m+1}(m+1) b^{(\theta',t)}_{\gamma-m}(\gamma-m)\mathbb{E}_{m+1,\gamma-m}\left[\mathcal{D}_{\theta'_1+J,\theta'_2+\gamma-m-J}(z)\right] \\
&< \varepsilon\Bigg[\sum_{m=0}^{\lfloor \frac{\gamma+1}{2} \rfloor} b^{(\theta,s)}_m(m) b^{(\theta',t)}_{\gamma+1-m}(\gamma-m)\mathbb{E}_{m,\gamma-m}\left[\mathcal{D}_{\theta'_1+J,\theta'_2+\gamma-m-J}(z)\right] \\
&\quad + \sum_{m=\lfloor \frac{\gamma+1}{2} \rfloor}^{\gamma} b^{(\theta,s)}_{m+1}(m) b^{(\theta',t)}_{\gamma-m}(\gamma-m)\mathbb{E}_{m,\gamma-m}\left[\mathcal{D}_{\theta'_1+J,\theta'_2+\gamma-m-J}(z)\right]\Bigg] \\
&< \varepsilon\Bigg[\sum_{m=0}^{\gamma} b^{(\theta,s)}_m(m) b^{(\theta',t)}_{\gamma+1-m}(\gamma-m)\mathbb{E}_{m,\gamma-m}\left[\mathcal{D}_{\theta'_1+J,\theta'_2+\gamma-m-J}(z)\right] \\
&\quad + \sum_{m=0}^{\gamma} b^{(\theta,s)}_{m+1}(m) b^{(\theta',t)}_{\gamma-m}(\gamma-m)\mathbb{E}_{m,\gamma-m}\left[\mathcal{D}_{\theta'_1+J,\theta'_2+\gamma-m-J}(z)\right]\Bigg] = \varepsilon A_{\gamma+1,\gamma},
\end{aligned}$$

□

Lemma 3. *Let $K^{(x,\boldsymbol{\theta})}, K^{(z,\boldsymbol{\theta'})}$ be defined as above. Then for any $m \in \{0, \dots, \gamma\}$ and $k = \gamma - m$ we have that*

$$\mathbb{E}_{m+1,\gamma-m}\left[\mathcal{D}_{\theta_1'+J,\theta_2'+\gamma-m-J}(z)\right] < K^{(x,\boldsymbol{\theta})}\mathbb{E}_{m,\gamma-m}\left[\mathcal{D}_{\theta_1'+J,\theta_2'+\gamma-m-J}(z)\right] \tag{32}$$

$$\mathbb{E}_{\gamma-k,k+1}\left[\mathcal{D}_{\theta_1'+J,\theta_2'+k+1-J}(z)\right] < K^{(z,\boldsymbol{\theta'})}\mathbb{E}_{\gamma-k,k}\left[\mathcal{D}_{\theta_1'+J,\theta_2'+k-J}(z)\right] \tag{33}$$

Proof. Assume that $l \leq \lfloor mx \rfloor$, and observe that

$$\mathbb{E}_{m+1,\gamma-m}\left[\mathcal{D}_{\theta_1'+J,\theta_2'+k-J}(z)\right] = \sum_{l=0}^{m+1}\sum_{j=0}^{\gamma-m}\mathcal{B}_{m+1,x}(l)\mathcal{DM}_{\theta_1+l,\theta_2+m+1-l;\gamma-m}(j)\mathcal{D}_{\theta'_1+j,\theta'_2+\gamma-m-j}(z).$$

Additionally

$$\begin{aligned}
&\mathcal{B}_{m+1,x}(l)\mathcal{DM}_{\theta_1+l,\theta_2+m+1-l;\gamma-m}(j)\\
&\quad= \binom{m+1}{l}x^l(1-x)^{m+1-l}\binom{\gamma-m}{j}\frac{B(\theta_1+l+j,\theta_2+m+1-l+\gamma-m-j)}{B(\theta_1+l,\theta_2+m+1-l)}\\
&\quad= \frac{m+1}{m+1-l}\frac{\theta_2+\gamma-l-j}{\theta+\gamma}\frac{\theta+m}{\theta_2+m-l}(1-x)\mathcal{B}_{m,x}(l)\mathcal{DM}_{\theta_1+l,\theta_2+m-l;\gamma-m}(j)\\
&\quad=: a(m)b(m)c(m)(1-x)\mathcal{B}_{m,x}(l)\mathcal{DM}_{\theta_1+l,\theta_2+m-l;\gamma-m}(j)
\end{aligned}$$

Trivially, $\theta_2 + \gamma - l - j \leq \theta_2 + \gamma \leq \theta + \gamma$, so $b(m) \leq 1$ independently of m.

For $a(m)$, observe that this function is increasing in m such that for any m it holds that

$$a(m) = \frac{m+1}{m+1-l} \leq \frac{m+1}{m+1-mx} \leq \frac{1}{1-x} =: a(\infty)$$

For $c(m)$, we have that for $m \geq 1$

$$c(m) = \frac{\theta+m}{\theta_2+m-l} \leq \frac{\theta+m}{m(1-x)} \leq \frac{1+\theta}{1-x}.$$

Putting all this together

$$\begin{aligned}
\mathcal{B}_{m+1,x}(l)\mathcal{DM}_{\theta_1+l,\theta_2+m+1-l;\gamma-m}(j) &= a(m)b(m)c(m)(1-x)\mathcal{B}_{m,x}(l)\mathcal{DM}_{\theta_1+l,\theta_2+m-l;\gamma-m}(j)\\
&\leq \max\{\frac{c(0)}{1-x}, \frac{1+\theta}{(1-x)^2}\}(1-x)\mathcal{B}_{m,x}(l)\mathcal{DM}_{\theta_1+l,\theta_2+m-l;\gamma-m}(j)\\
&= \max\left\{\frac{\theta}{\theta_2}, \frac{1+\theta}{1-x}\right\}\mathcal{B}_{m,x}(l)\mathcal{DM}_{\theta_1+l,\theta_2+m-l;\gamma-m}(j) \qquad (34)
\end{aligned}$$

Summing from $j = 0, \ldots, \gamma - m$ and $l = 0, \ldots, \lfloor mx \rfloor$ we get

$$\sum_{l=0}^{\lfloor mx \rfloor} \sum_{j=0}^{\gamma-m} \mathcal{B}_{m+1,x}(l)\mathcal{DM}_{\theta_1+l,\theta_2+m+1-l;\gamma-m}(j)\mathcal{D}_{\theta'_1+j,\theta'_2+\gamma-m-j}(z)$$
$$< \max\left\{\frac{\theta}{\theta_2}, \frac{1+\theta}{1-x}\right\} \sum_{l=0}^{\lfloor mx \rfloor} \sum_{j=0}^{\gamma-j} \mathcal{B}_{m,x}(l)\mathcal{DM}_{\theta_1+l,\theta_2+m-l;\gamma-m}(j)\mathcal{D}_{\theta'_1+j,\theta'_2+\gamma-m-j}(z). \quad (35)$$

Suppose now that $l \geq \lfloor mx \rfloor$, then

$$\mathcal{B}_{m+1,x}(l+1)\mathcal{DM}_{\theta_1+l+1,\theta_2+m+1-(l+1);\gamma-m}(j) = \frac{m+1}{l+1}\frac{\theta+m}{\theta_1+l}\frac{\theta_1+l+j}{\theta+\gamma} x\mathcal{B}_{m,x}(l)\mathcal{DM}_{\theta_1+l,\theta_2+m-l;\gamma-m}(j)$$
$$=: d(l,m)e(m)x\mathcal{B}_{m,x}(l)\mathcal{DM}_{\theta_1+l,\theta_2+m-l;\gamma-m}(j)$$

For $e(m)$ we again have that trivially $\theta_1 + l + j \leq \theta_1 + \gamma \leq \theta + \gamma$, so $e(m) \leq 1$ independently of m, whilst for $d(l,m)$, we have that if $\lfloor mx \rfloor > 1$, then

$$d(l,m) \leq h(mx-1,m) = \frac{m+1}{mx}\frac{\theta+m}{\theta_1+mx-1} \leq \left(1+\frac{1}{x}\right)\frac{\theta+m}{\theta_1+mx-1} =: \tilde{d}(m)$$

and (as in the display in the middle of p.27 in Jenkins and Spanò (2017)), since the sign of $\tilde{d}'(m)$ is independent of m, it follows that

$$d(l,m) \leq \max\left\{\tilde{d}(x^{-1}), \tilde{d}(\infty)\right\} \leq \left(1+\frac{1}{x}\right)\max\left\{\frac{x\theta+1}{x\theta_1}, \frac{1}{x}\right\}.$$

On the other hand, if $\lfloor mx \rfloor \leq 1$, then

$$d(l,m) \leq d(0,m) = (m+1)\frac{\theta+m}{\theta_1} \leq \left(1+\frac{1}{x}\right)\frac{x\theta+1}{x\theta_1}$$

and thus it follows that independently of whether $\lfloor mx \rfloor > 1$ or $\lfloor mx \rfloor \leq 1$,

$$\mathcal{B}_{m+1,x}(l+1)\mathcal{DM}_{\theta_1+l+1,\theta_2+m+1-(l+1);\gamma-m}(j) < \left(1+\frac{1}{x}\right)\max\left\{\frac{x\theta+1}{x\theta_1}, \frac{1}{x}\right\} x\mathcal{B}_{m,x}(l)\mathcal{DM}_{\theta_1+l,\theta_2+m-l;\gamma-m}(j)$$
$$= \left(1+\frac{1}{x}\right)\max\left\{\frac{x\theta+1}{\theta_1}, 1\right\}\mathcal{B}_{m,x}(l)\mathcal{DM}_{\theta_1+l,\theta_2+m-l;\gamma-m}(j).$$

Summing over $j = 0, \ldots, \gamma - m$ and $l = \lfloor mx \rfloor + 1, \ldots, m+1$ we get that

$$\sum_{l=\lfloor mx \rfloor+1}^{m+1} \sum_{j=0}^{\gamma-m} \mathcal{B}_{m+1,x}(l)\mathcal{DM}_{\theta_1+l,\theta_2+m+1-l;\gamma-m}(j)\mathcal{D}_{\theta'_1+j,\theta'_2+\gamma-m-j}(z)$$
$$< \left(1+\frac{1}{x}\right)\max\left\{\frac{x\theta+1}{\theta_1}, 1\right\} \sum_{l=\lfloor mx \rfloor}^{m} \sum_{j=0}^{\gamma-m} \mathcal{B}_{m,x}(l)\mathcal{DM}_{\theta_1+l,\theta_2+m-l;\gamma-m}(j)\mathcal{D}_{\theta'_1+j,\theta'_2+\gamma-m-j}(z) \quad (36)$$

Summing (35) and (36) gives

$$\begin{aligned}\mathbb{E}_{m+1,\gamma-m}\left[\mathcal{D}_{\theta_1'+J,\theta_2'+k-J}(z)\right] &= \sum_{l=0}^{\lfloor mx\rfloor}\sum_{j=0}^{\gamma-m}\mathcal{B}_{m+1,x}(l)\mathcal{DM}_{\theta_1+l,\theta_2+m+1-l;\gamma-m}(j)\mathcal{D}_{\theta'{}_1+j,\theta'{}_2+\gamma-m-j}(z)\\ &\quad+\sum_{l=\lfloor mx\rfloor+1}^{m+1}\sum_{j=0}^{\gamma-m}\mathcal{B}_{m+1,x}(l)\mathcal{DM}_{\theta_1+l,\theta_2+m+1-l;\gamma-m}(j)\mathcal{D}_{\theta'{}_1+j,\theta'{}_2+\gamma-m-j}(z)\\ &< K^{(x,\boldsymbol{\theta})}\mathbb{E}_{m,\gamma-m}\left[\mathcal{D}_{\theta_1'+J,\theta_2'+k-J}(z)\right]\end{aligned}$$

and thus (32) holds.

For (33), recall that $k=\gamma-m$, so that

$$\begin{aligned}\mathbb{E}_{m,\gamma+1-m}\left[\mathcal{D}_{\theta_1'+J,\theta_2'+\gamma+1-m-J}(z)\right] &= \mathbb{E}_{\gamma-k,k+1}\left[\mathcal{D}_{\theta_1'+J,\theta_2'+k+1-J}(z)\right]\\ &= \sum_{l=0}^{\gamma-k}\sum_{j=0}^{k+1}\mathcal{B}_{\gamma-k,x}(l)\mathcal{DM}_{\theta_1+l,\theta_2+\gamma-k-l;k+1}(j)\mathcal{D}_{\theta'{}_1+j,\theta'{}_2+k+1-j}(z).\end{aligned}$$

Suppose $j\le\lfloor kz\rfloor$, then

$$\begin{aligned}&\mathcal{DM}_{\theta_1+l,\theta_2+\gamma-k-l;k+1}(j)\mathcal{D}_{\theta'{}_1+j,\theta'{}_2+k+1-j}(z)\\ &\quad=\binom{k+1}{j}\frac{B(\theta_1+l+j,\theta_2+\gamma-k-l+k+1-j)}{B(\theta_1+l,\theta_2+\gamma-k-l)}\frac{z^{\theta'{}_1+j-1}(1-z)^{\theta'{}_2+k+1-j-1}}{B(\theta'{}_1+j,\theta'{}_2+k+1-j)}\\ &\quad=\frac{k+1}{k+1-j}\frac{\theta_2+\gamma-l-j}{\theta+\gamma}\frac{\theta'+k}{\theta_2'+k-j}(1-z)\mathcal{DM}_{\theta_1+l,\theta_2+\gamma-k-l;k}(j)\mathcal{D}_{\theta'{}_1+j,\theta'{}_2+k-j}(z)\\ &\quad=:a(k)b(k)c(k)(1-z)\mathcal{DM}_{\theta_1+l,\theta_2+\gamma-k-l;k}(j)\mathcal{D}_{\theta'{}_1+j,\theta'{}_2+k-j}(z).\end{aligned}$$

As done previously, we see that $a(k)$ is increasing in k and furthermore

$$a(k)=\frac{k+1}{k+1-j}\le\frac{k+1}{k+1-kz}\le\frac{1}{1-z}=a(\infty).$$

Also we have that $\theta_2+\gamma-l-j\le\theta_2+\gamma<\theta+\gamma$, so that $b(k)\le 1$ independently of k.

For $c(k)$ we once more have that for $k\ge 1$

$$c(k)=\frac{\theta'+k}{\theta'{}_2+k-j}\le\frac{\theta'+k}{k(1-z)}\le\frac{1+\theta'}{1-z},$$

allowing us to deduce that

$$\begin{aligned}
&\mathcal{DM}_{\theta_1+l,\theta_2+\gamma-k-l;k+1}(j)\mathcal{D}_{\theta'_1+j,\theta'_2+k+1-j}(z)\\
&\quad< \max\left\{\frac{\theta'}{\theta'_2(1-z)},\frac{1+\theta'}{(1-z)^2}\right\}(1-z)\mathcal{DM}_{\theta_1+l,\theta_2+\gamma-k-l;k}(j)\mathcal{D}_{\theta'_1+j,\theta'_2+k-j}(z)\\
&\quad= \max\left\{\frac{\theta'}{\theta'_2},\frac{1+\theta'}{(1-z)}\right\}\mathcal{DM}_{\theta_1+l,\theta_2+\gamma-k-l;k}(j)\mathcal{D}_{\theta'_1+j,\theta'_2+k-j}(z). \qquad (37)
\end{aligned}$$

Thus, summing (37) over $j=0,\dots,\lfloor kz\rfloor$, and over $l=0,\dots,\gamma-k$ we get

$$\begin{aligned}
&\sum_{l=0}^{\gamma-k}\sum_{j=0}^{\lfloor kz\rfloor}\mathcal{B}_{\gamma-k,x}(l)\mathcal{DM}_{\theta_1+l,\theta_2+\gamma-k-l;k+1}(j)\mathcal{D}_{\theta'_1+j,\theta'_2+k+1-j}(z)\\
&\qquad< \max\left\{\frac{\theta'}{\theta'_2},\frac{1+\theta'}{(1-z)}\right\}\sum_{l=0}^{\gamma-k}\sum_{j=0}^{\lfloor kz\rfloor}\mathcal{B}_{\gamma-k,x}(l)\mathcal{DM}_{\theta_1+l,\theta_2+\gamma-k-l;k}(j)\mathcal{D}_{\theta'_1+j,\theta'_2+k-j}(z)
\end{aligned}$$

Now if $j\geq\lfloor kz\rfloor$, we have that

$$\begin{aligned}
&\mathcal{DM}_{\theta_1+l,\theta_2+\gamma-k-l;k+1}(j+1)\mathcal{D}_{\theta'_1+j+1,\theta'_2+k+1-(j+1)}(z)\\
&\quad=\binom{k+1}{j+1}\frac{B(\theta_1+l+j+1,\theta_2+\gamma-k-l+k+1-(j+1))}{B(\theta_1+l,\theta_2+\gamma-k-l)}\frac{z^{\theta'_1+j+1-1}(1-z)^{\theta'_2+k+1-(j+1)-1}}{B(\theta'_1+j+1,\theta'_2+k+1-(j+1))}\\
&\quad=\frac{k+1}{j+1}\frac{\theta'+k}{\theta'_1+j}\frac{\theta_1+l+j}{\theta+\gamma}z\mathcal{DM}_{\theta_1+l,\theta_2+\gamma-k-l;k}(j)\mathcal{D}_{\theta'_1+j,\theta'_2+k-j}(z)\\
&\quad=: f(j,k)g(k)z\mathcal{DM}_{\theta_1+l,\theta_2+\gamma-k-l;k}(j)\mathcal{D}_{\theta'_1+j,\theta'_2+k-j}(z)
\end{aligned}$$

For $g(k)$, recall that $j\leq k$ and $l\leq\gamma-k$, so trivially $\theta_1+l+j\leq\theta_1+\gamma\leq\theta+\gamma$ and thus $g(k)\leq 1$ independently of k.

For $f(j,k)$, if $\lfloor kz\rfloor>1$ we get that

$$f(j,k)\leq f(kz-1,k)=\frac{k+1}{kz}\frac{\theta'+k}{\theta'_1+kz-1}\leq\left(1+\frac{1}{z}\right)\frac{\theta'+k}{\theta'_1+kz-1}\leq\left(1+\frac{1}{z}\right)\max\left\{\frac{z\theta'+1}{z\theta'_1},\frac{1}{z}\right\}$$

whilst if $\lfloor kz\rfloor\leq 1$ it follows that

$$f(j,k)\leq f(0,k)=(k+1)\frac{\theta'+k}{\theta'_1}\leq\left(1+\frac{1}{z}\right)\frac{z\theta'+1}{z\theta'_1}$$

and thus regardless of the value of $\lfloor kz\rfloor$ we have that

$$f(j,k)\leq\left(1+\frac{1}{z}\right)\max\left\{\frac{z\theta'+1}{z\theta'_1},\frac{1}{z}\right\}.$$

Thus

$$\begin{aligned}&\mathcal{DM}_{\theta_1+l,\theta_2+\gamma-k-l;k+1}(j+1)\mathcal{D}_{\theta'_1+j+1,\theta'_2+k+1-(j+1)}(z)\\&\quad<\left(1+\frac{1}{z}\right)\max\left\{\frac{z\theta'+1}{\theta'_1},1\right\}\mathcal{DM}_{\theta_1+l,\theta_2+\gamma-k-l;k}(j)\mathcal{D}_{\theta'_1+j,\theta'_2+k-j}(z)\end{aligned}$$

Summing over $j=\lfloor kz\rfloor+1,\dots,k+1$ and $l=0,\dots,\gamma-k$

$$\begin{aligned}&\sum_{l=0}^{\gamma-k}\sum_{j=\lfloor kz\rfloor+1}^{k+1}\mathcal{B}_{\gamma-k,x}(l)\mathcal{DM}_{\theta_1+l,\theta_2+\gamma-k-l;k+1}(j)\mathcal{D}_{\theta'_1+j,\theta'_2+k+1-j}(z)\\&<\left(1+\frac{1}{z}\right)\max\left\{\frac{z\theta'+1}{\theta'_1},1\right\}\sum_{l=0}^{\gamma-k}\sum_{j=\lfloor kz\rfloor}^{k}\mathcal{B}_{\gamma-k,x}(l)\mathcal{DM}_{\theta_1+l,\theta_2+\gamma-k-l;k}(j)\mathcal{D}_{\theta'_1+j,\theta'_2+k-j}(z).\end{aligned}\tag{38}$$

Summing (37) and (38) gives

$$\begin{aligned}\mathbb{E}_{\gamma-k,k+1}\left[\mathcal{D}_{\theta'_1+J,\theta'_2+k+1-J}(z)\right]&=\sum_{l=0}^{\gamma-k}\sum_{j=0}^{\lfloor kz\rfloor}\mathcal{B}_{\gamma-k,x}(l)\mathcal{DM}_{\theta_1+l,\theta_2+\gamma-k-l;k+1}(j)\mathcal{D}_{\theta'_1+j,\theta'_2+k+1-j}(z)\\&\quad+\sum_{l=0}^{\gamma-k}\sum_{j=\lfloor kz\rfloor+1}^{k+1}\mathcal{B}_{\gamma-k,x}(l)\mathcal{DM}_{\theta_1+l,\theta_2+\gamma-k-l;k+1}(j)\mathcal{D}_{\theta'_1+j,\theta'_2+k+1-j}(z)\\&<K^{(z,\boldsymbol{\theta'})}\mathbb{E}_{\gamma-k,k}\left[\mathcal{D}_{\theta'_1+J,\theta'_2+k-J}(z)\right]\end{aligned}$$

□

We can now use use Proposition 1 in Jenkins and Spanò (2017) and Proposition 2 above to deduce the following

Proposition 4. *Define* $F_{m,k,l,j}^{(x,z,s,t,\boldsymbol{\theta},\boldsymbol{\theta'})}:=\max\{C_m^{(\theta,s)},C_k^{(\theta',t)},G^{(x,z,s,t,\boldsymbol{\theta},\boldsymbol{\theta'})}\}$ *and*

$$e_{m,k,l,j}(v)=\mathcal{B}_{m,x}(l)\mathcal{D}_{\theta'_1+k,\theta'_2+k-j}(z)\mathcal{DM}_{\theta_1+l,\theta_2+m-l;k}(j)\frac{\left[\sum_{i=0}^{v}(-1)^ib_{m+i}^{(\theta,s)}(m)\right]\left[\sum_{i=0}^{v}(-1)^ib_{k+i}^{(\theta',t)}(k)\right]}{\sum_{i=0}^{v+1}(-1)^id_i}.$$

Then for $2v\geq F_{m,k,l,j}^{(x,z,s,t,\boldsymbol{\theta},\boldsymbol{\theta'})}$ *it follows that*

$$e_{m,k,l,j}(2v+1)<e_{m,k,l,j}(2v+3)<p_{m,k,l,j}^{(x,z,s,t,\boldsymbol{\theta}),\boldsymbol{\theta'}}<e_{m,k,l,j}(2v+2)<e_{m,k,l,j}(2v).$$

Proof. Follows immediately by applying Proposition 1 from Jenkins and Spanò (2017) to the numerator and Proposition 2 to the denominator. □

3 Simulating draws from the transition densities (5)–(11) and (13)

For the remaining edge cases (i.e. where at least one of x,z is on the boundary $\{0,1\}$), we can extend the same arguments made in Proposition 2 with some slight modifications to obtain eventually monotonic

sequences of upper and lower bounds on the corresponding denominator. Sampling from the corresponding transition density then follows as usual. In particular, observe that if we can obtain such lower and upper bounds for the quantities

$$\sum_{m,k=0}^{\infty} q_m^{\theta}(s) q_k^{\theta'}(t) \frac{B(\theta_1, \theta_2 + m + k)}{B(\theta_1, \theta_2 + m) B(\theta_1', \theta_2' + k)}, \tag{39}$$

$$\sum_{m,k=0}^{\infty} \sum_{j=0}^{k} q_m^{\theta}(s) q_k^{\theta'}(t) \mathcal{DM}_{\theta_1, \theta_2+m; k}(j) \mathcal{D}_{\theta_1'+j, \theta_2'+k-j}(z), \tag{40}$$

$$\sum_{m,k=0}^{\infty} \sum_{l=0}^{m} q_m^{\theta}(s) q_k^{\theta'}(t) \mathcal{B}_{m,x}(l) \frac{B(\theta_1 + l, \theta_2 + m - l + k)}{B(\theta_1 + l, \theta_2 + m - l) B(\theta_1', \theta_2' + k)}, \tag{41}$$

then the result follows for the remaining transition densities by considering suitably modified arguments.

We point out that in all three of the above cases, the exact same arguments in the proof of Proposition 2 can be used, with the only difference being that Lemma 3 needs to be suitably modified. To this end, it suffices to show that

$$\frac{B(\theta_1, \theta_2 + m + \gamma + 1 - m)}{B(\theta_1, \theta_2 + m) B(\theta_1', \theta_2' + \gamma + 1 - m)} < \delta_1 \frac{B(\theta_1, \theta_2 + m + \gamma - m)}{B(\theta_1, \theta_2 + m) B(\theta_1', \theta_2' + \gamma - m)}, \tag{42}$$

$$\frac{B(\theta_1, \theta_2 + m + 1 + \gamma - m)}{B(\theta_1, \theta_2 + m + 1) B(\theta_1', \theta_2' + \gamma - m)} < \delta_2 \frac{B(\theta_1, \theta_2 + m + \gamma - m)}{B(\theta_1, \theta_2 + m) B(\theta_1', \theta_2' + \gamma - m)}, \tag{43}$$

$$\sum_{j=0}^{\gamma-m} \mathcal{DM}_{\theta_1, \theta_2+m+1; \gamma-m}(j) \mathcal{D}_{\theta_1'+j, \theta_2'+\gamma-m-j}(z) < \delta_3 \sum_{j=0}^{\gamma-m} \mathcal{DM}_{\theta_1, \theta_2+m; \gamma-m}(j) \mathcal{D}_{\theta_1'+j, \theta_2'+\gamma-m-j}(z), \tag{44}$$

$$\sum_{j=0}^{\gamma+1-m} \mathcal{DM}_{\theta_1, \theta_2+m; \gamma+1-m}(j) \mathcal{D}_{\theta_1'+j, \theta_2'+\gamma+1-m-j}(z) < \delta_4 \sum_{j=0}^{\gamma-m} \mathcal{DM}_{\theta_1, \theta_2+m; \gamma-m}(j) \mathcal{D}_{\theta_1'+j, \theta_2'+\gamma-m-j}(z), \tag{45}$$

$$\sum_{l=0}^{m+1} \mathcal{B}_{m+1,x}(l) \frac{B(\theta_1 + l, \theta_2 + m + 1 - l + \gamma - m)}{B(\theta_1 + l, \theta_2 + m + 1 - l) B(\theta_1', \theta_2' + \gamma + 1 - d)} < \delta_5 \sum_{l=0}^{m} \mathcal{B}_{m,x}(l) \frac{B(\theta_1 + l, \theta_2 + m - l + \gamma - m)}{B(\theta_1 + l, \theta_2 + m - l) B(\theta_1', \theta_2' + \gamma - m)}, \tag{46}$$

$$\sum_{l=0}^{m} \mathcal{B}_{m,x}(l) \frac{B(\theta_1 + l, \theta_2 + m - l + \gamma + 1 - m)}{B(\theta_1 + l, \theta_2 + m - l) B(\theta_1', \theta_2' + \gamma + 1 - m)} < \delta_6 \sum_{l=0}^{m} \mathcal{B}_{m,x}(l) \frac{B(\theta_1 + l, \theta_2 + m - l + \gamma - m)}{B(\theta_1 + l, \theta_2 + m - l) B(\theta_1', \theta_2' + \gamma - m)}, \tag{47}$$

for an appropriate choice of δ_i, with $i \in \{1, \ldots, 6\}$. We start with (42) and (43) for which we observe that

$$\begin{aligned} \frac{B(\theta_1, \theta_2 + m + \gamma + 1 - m)}{B(\theta_1, \theta_2 + m) B(\theta_1', \theta_2' + \gamma + 1 - m)} &= \frac{\theta_2 + \gamma}{\theta + \gamma} \frac{\theta' + \gamma - m}{\theta_2' + \gamma - m} \frac{B(\theta_1, \theta_2 + m + \gamma - m)}{B(\theta_1, \theta_2 + m) B(\theta_1', \theta_2' + \gamma - m)} \\ &\leq \frac{\theta_2}{\theta} \frac{\theta'}{\theta_2'} \frac{B(\theta_1, \theta_2 + m + \gamma - m)}{B(\theta_1, \theta_2 + m) B(\theta_1', \theta_2' + \gamma - m)}, \end{aligned}$$

so $\delta_1 = \frac{\theta_2 \theta'}{\theta_2' \theta}$, and

$$\begin{aligned} \frac{B(\theta_1, \theta_2 + m + 1 + \gamma - m)}{B(\theta_1, \theta_2 + m + 1) B(\theta_1', \theta_2' + \gamma - m)} &= \frac{\theta_2 + \gamma}{\theta + \gamma} \frac{\theta + m}{\theta_2 + m} \frac{B(\theta_1, \theta_2 + m + \gamma - m)}{B(\theta_1, \theta_2 + m) B(\theta_1', \theta_2' + \gamma - m)} \\ &\leq \frac{B(\theta_1, \theta_2 + m + \gamma - m)}{B(\theta_1, \theta_2 + m) B(\theta_1', \theta_2' + \gamma - m)}, \end{aligned}$$

and thus $\delta_2 = 1$. Similar calculations give that

$$\mathcal{DM}_{\theta_1,\theta_2+m+1;\gamma-m}(j)\mathcal{D}_{\theta_1'+j,\theta_2'+\gamma-m-j}(z) \leq \frac{\theta_2+\gamma-j}{\theta+\gamma-j}\frac{\theta+m}{\theta_2+m}\mathcal{DM}_{\theta_1,\theta_2+m;\gamma-m}(j)\mathcal{D}_{\theta_1'+j,\theta_2'+\gamma-m-j}(z)$$

and thus summing over $j = 0, \ldots, \gamma - m$ gives that (44) holds with $\delta_3 = 1$. Further,

$$\mathcal{B}_{m,x}(l)\frac{B(\theta_1+l,\theta_2+m-l+\gamma+1-m)}{B(\theta_1+l,\theta_2+m-l)B(\theta_1',\theta_2'+\gamma+1-m)} \leq \frac{\theta'}{\theta_2'}\frac{\theta_2}{\theta}\mathcal{B}_{m,x}(l)\frac{B(\theta_1+l,\theta_2+m-l+\gamma-m)}{B(\theta_1+l,\theta_2+m-l)B(\theta_1',\theta_2'+\gamma-m)}$$

such that (47) holds with $\delta_6 = \delta_1$. The remaining bounds are slightly more involved but we can adopt the same approach as in Lemma 3, namely for (45) let $j \leq \lfloor kz \rfloor$, and then

$$\begin{aligned}
&\mathcal{DM}_{\theta_1,\theta_2+m;\gamma+1-m}(j)\mathcal{D}_{\theta_1'+j,\theta_2'+\gamma+1-m-j}(z) \\
&\qquad = \frac{k+1}{k+1-j}\frac{\theta_2+\gamma-j}{\theta+\gamma}\frac{\theta'+k}{\theta_2'+k-j}(1-z)\mathcal{DM}_{\theta_1,\theta_2+m;\gamma-m}(j)\mathcal{D}_{\theta_1'+j,\theta_2'+\gamma-m-j}(z) \\
&\qquad \leq \max\left\{\frac{\theta'}{\theta_2'}, \frac{1+\theta'}{1-z}\right\}\mathcal{DM}_{\theta_1,\theta_2+m;\gamma-m}(j)\mathcal{D}_{\theta_1'+j,\theta_2'+\gamma-m-j}(z)
\end{aligned}$$

using the exact same arguments as in Lemma 3, and thus we can sum over $j = 0, \ldots, \lfloor kz \rfloor$ to get

$$\begin{aligned}
&\sum_{j=0}^{\lfloor kz \rfloor}\mathcal{DM}_{\theta_1,\theta_2+m;\gamma+1-m}(j)\mathcal{D}_{\theta_1'+j,\theta_2'+\gamma+1-m-j}(z) \\
&\qquad < \max\left\{\frac{\theta'}{\theta_2'}, \frac{1+\theta'}{1-z}\right\}\sum_{j=0}^{\lfloor kz \rfloor}\mathcal{DM}_{\theta_1,\theta_2+m;\gamma-m}(j)\mathcal{D}_{\theta_1'+j,\theta_2'+\gamma-m-j}(z). \qquad (48)
\end{aligned}$$

For $j \geq \lfloor kz \rfloor$ we get that

$$\begin{aligned}
&\mathcal{DM}_{\theta_1,\theta_2+m;\gamma+1-m}(j+1)\mathcal{D}_{\theta_1'+j+1,\theta_2'+\gamma+1-m-j}(z) \\
&\qquad = \frac{k+1}{j+1}\frac{\theta_1+j}{\theta+\gamma}\frac{\theta'+k}{\theta_1'+j}z\mathcal{DM}_{\theta_1,\theta_2+m;\gamma-m}(j)\mathcal{D}_{\theta_1'+j,\theta_2'+\gamma-m-j}(z)
\end{aligned}$$

and again the same bounds as in Lemma 3 apply, such that summing over $j = \lfloor kz \rfloor + 1, \ldots, \gamma + 1 - m$ gives

$$\begin{aligned}
&\sum_{j=\lfloor kz \rfloor+1}^{\gamma+1-m}\mathcal{DM}_{\theta_1,\theta_2+m;\gamma+1-m}(j)\mathcal{D}_{\theta_1'+j,\theta_2'+\gamma+1-m-j}(z) \\
&\qquad < \left(1+\frac{1}{z}\right)\max\left\{\frac{z\theta'+1}{\theta_1'}, 1\right\}\sum_{j=\lfloor kz \rfloor}^{\gamma-m}\mathcal{DM}_{\theta_1,\theta_2+m;\gamma+1-m}(j)\mathcal{D}_{\theta_1'+j+1,\theta_2'+\gamma+1-m-j}(z), \qquad (49)
\end{aligned}$$

and summing (48) and (49) gives (45) with $\delta_4 = K^{(z,\boldsymbol{\theta'})}$.

We can repeat similar computations for (46), splitting over the case $l \leq \lfloor mx \rfloor$ and $l \geq \lfloor mx \rfloor$ to get the desired inequality with $\delta_5 = K^{(x,\boldsymbol{\theta})}$.

4 Small time approximations

For small times s or t, the quantities $q_m^\theta(s)$ and $q_k^{\theta'}(t)$ can become numerically unstable. To overcome this we make use of Gaussian approximations first developed in Griffiths (1984). We consider two specific cases:

1. When exactly one of s and t is less than some threshold δ
2. When both s and t are smaller than some threshold δ

In case 2, we employ Gaussian approximations for both $q_m^\theta(s)$ and $q_k^{\theta'}(t)$ much in the same way as done in Sant *et al.* (2023). Case 1 however requires a separate approach as we observe that the decomposition introduced earlier making use of $A_{\gamma,\lambda}$ relies on the joint decomposition of $q_m^\theta(s)$ and $q_k^{\theta'}(t)$, and thus if we approximate one by a Gaussian whilst we expand the other, we cannot recover the same scheme as in Proposition 1. Instead, assume for the moment that $s < \delta < t$, then we compute the denominator by using the same decomposition as used in Proposition 4 in Jenkins and Spanò (2017), but where we now approximate the summation over m, h in (14) as follows

$$\sum_{m=0}^{\infty}\sum_{h=m}^{\infty}(-1)^{h-m}a_{hm}^\theta e^{-\frac{h(h+\theta-1)s}{2}}\mathbb{E}_{m,k}\left[\mathcal{D}_{\theta_1'+J,\theta_2'+k-J}(z)\right] \approx \sum_{m\in\mathbb{M}}\tilde{q}_m^\theta(s)\mathbb{E}_{m,k}\left[\mathcal{D}_{\theta_1'+J,\theta_2'+k-J}(z)\right]$$

where $\tilde{q}_m^\theta(s)$ denotes the weight associated with the integer m under the discretised normal distribution used to approximate $q_m^\theta(s)$, and $\mathbb{M}$ denotes the support of $\{\tilde{q}_m^\theta(s) : m \in \mathbb{N}\}$ truncated to be within 5 standard deviations of the corresponding mean. The default setting in `EWF` is $\delta = 0.1$ and can be tweaked by the user, however `EWF` features a robust safeguarding mechanism which automatically detects numerical instabilities (e.g. by detecting negative values for probabilities or lower bounds that exceed upper bounds) irrespective of choice of δ, falling back to Gaussian approximations whenever necessary.

5 Output verification

As in Section 7 of the Supplementary Information of Sant *et al.* (2023), we validate our implementation by comparing histograms of the draws obtained through the exact simulation algorithm to the empirical transition density evaluated by truncating the infinite series expressions for the corresponding bridge transition density. We consider two diffusion bridges, which we shall denote $B1$ and $B2$, with four epochs each. We summarise the mutation parameters, observation times and values, as well as sampling times we consider for both bridges in Table 1 and Figure 3, with the latter featuring an overlay of 100 candidate bridge paths for illustrative purposes.

At each sampling time, we generate 10,000 draws from the corresponding law of the bridge process, and assess the performance of our implementation by comparing the resulting histogram to the empirical transition density, as well as QQ plots. Additionally we conduct a Kolmogorov–Smirnov goodness of fit

$B1$	Left Endpoint	Left $\boldsymbol{\theta}$	Sampling Time	Right Endpoint	Right $\boldsymbol{\theta}$
Interval 1	(0, 0.2)	$\boldsymbol{\theta}_1 = (1.2, 3.6)$	0.2	(0.4, 0.4)	$\boldsymbol{\theta}_2 = (0.6, 0.9)$
Interval 2	(0.4, 0.4)	$\boldsymbol{\theta}_2 = (0.6, 0.9)$	0.5	(0.55, 0.5)	$\boldsymbol{\theta}_3 = (0.4, 0.4)$
Interval 3	(0.55, 0.5)	$\boldsymbol{\theta}_3 = (0.4, 0.4)$	0.6	(0.75, 0.3)	$\boldsymbol{\theta}_4 = (2.1, 0.8)$
Interval 4	(0.75, 0.3)	$\boldsymbol{\theta}_4 = (2.1, 0.8)$	0.95	(1, 0.2)	$\boldsymbol{\theta}_5 = (0.1, 0.1)$
$B2$	Left Endpoint	Left $\boldsymbol{\theta}$	Sampling Time	Right Endpoint	Right $\boldsymbol{\theta}$
Interval 1	(0, 0.95)	$\boldsymbol{\theta}_1 = (1.3, 2.6)$	0.15	(0.3, 0.7)	$\boldsymbol{\theta}_2 = (0.2, 0.5)$
Interval 2	(0.3, 0.7)	$\boldsymbol{\theta}_2 = (0.2, 0.5)$	0.38	(0.6, 0.3)	$\boldsymbol{\theta}_3 = (0.7, 1.4)$
Interval 3	(0.6, 0.3)	$\boldsymbol{\theta}_3 = (0.7, 1.4)$	0.85	(0.9, 0.4)	$\boldsymbol{\theta}_4 = (0.2, 0.5)$
Interval 4	(0.9, 0.4)	$\boldsymbol{\theta}_4 = (0.2, 0.5)$	0.96	(1, 0)	$\boldsymbol{\theta}_5 = (1.8, 0.6)$

Table 1: Sampling setup for bridges $B1$ and $B2$ where endpoints are reported as (t, x) where t is the time in diffusion units and x is the corresponding value of the diffusion bridge.

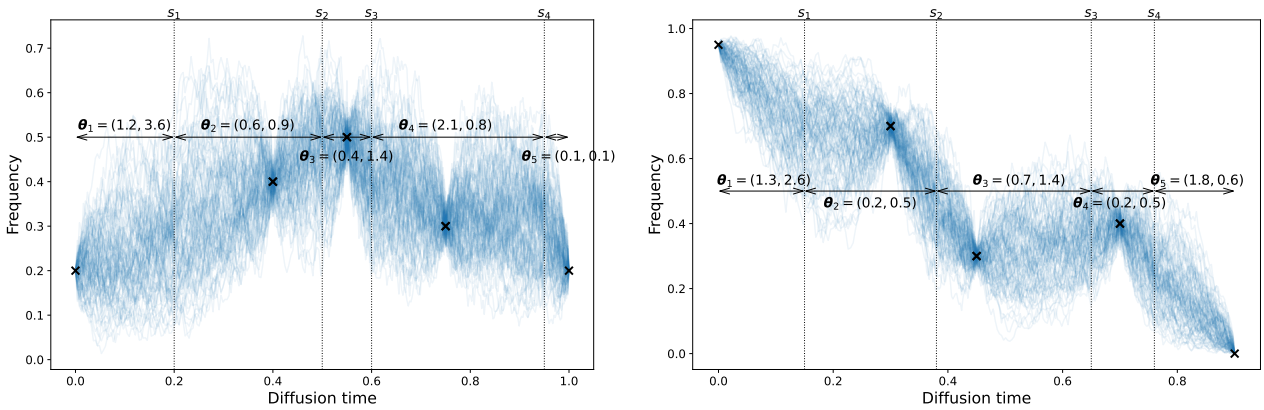

Figure 3: Sampling setups for bridges $B1$ (left) and $B2$ (right) with observations (black crosses), sampling times (vertical dotted lines) corresponding to times when the mutation parameter changes, and overlayed sample paths (blue). The mutation parameters adopted over each interval is annotated.

test, where we compute the empirical cumulative density function through the empirical density function computed as a truncation of the corresponding infinite series expressions (2). As can be observed in Figures 4 and 5, our method targets the correct distributions in all cases considered.

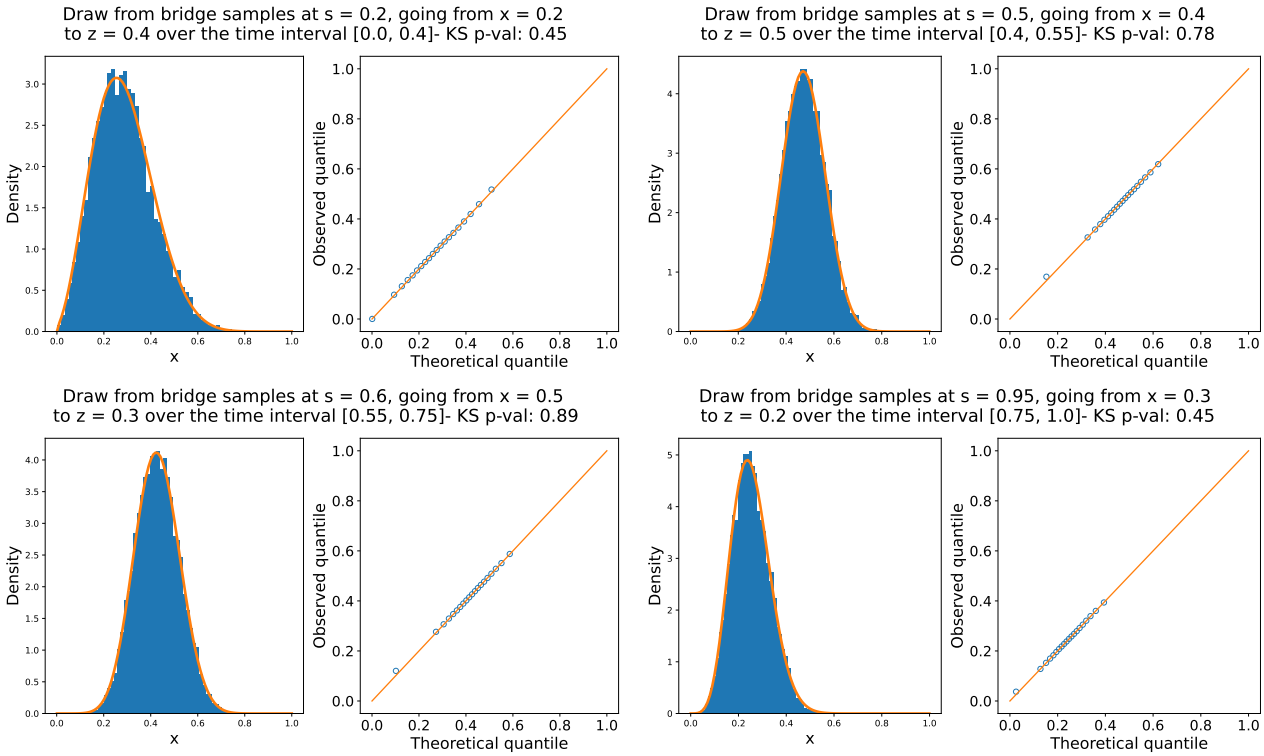

Figure 4: The output generated for bridge $B1$. Each panel corresponds to a different sampling time: $s = 0.2$ (top left), $s = 0.5$ (top right), $s = 0.6$ (bottom left), and $s = 0.95$ (bottom right). In each panel, we display the resulting histogram overlaid with the truncated density in orange on the left, whilst on the right we report the corresponding QQ plot. The p-value obtained from the Kolmogorov–Smirnov test is reported in the title of each panel.

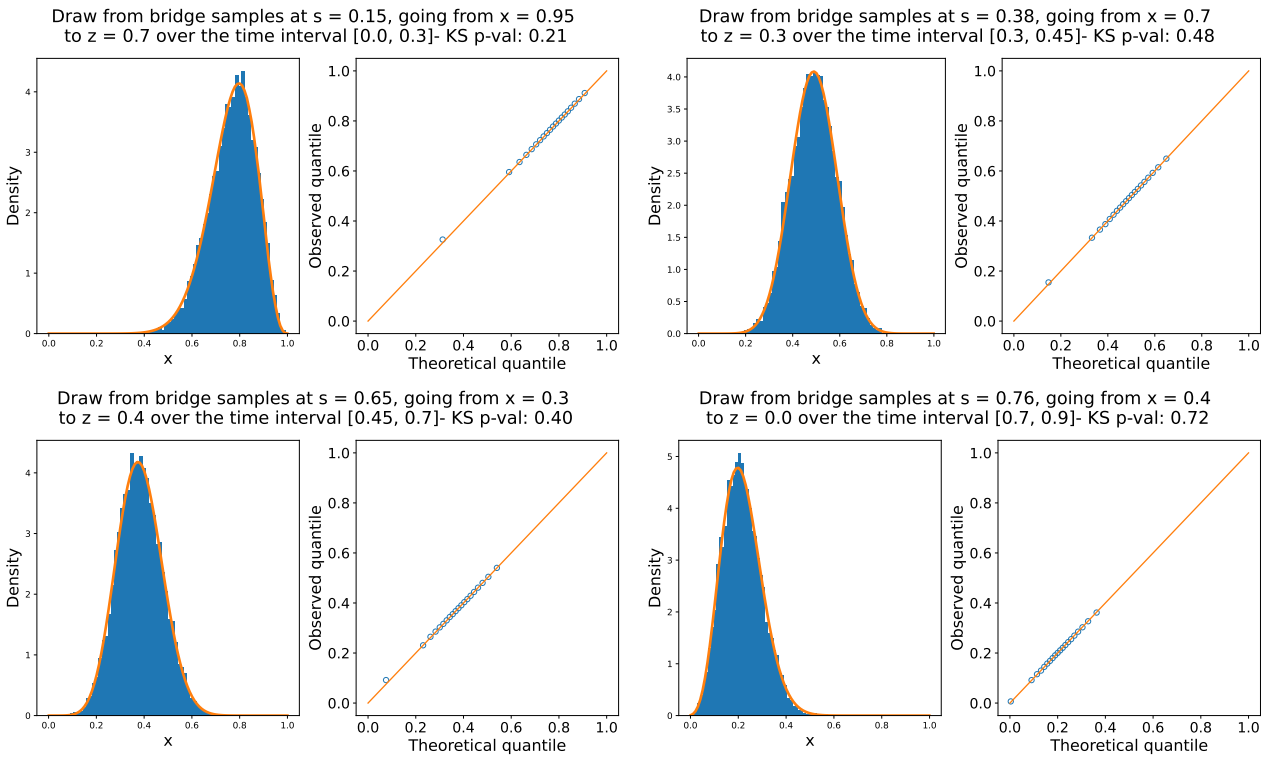

Figure 5: The output generated for bridge $B2$. Each panel corresponds to a different sampling time: $s = 0.15$ (top left), $s = 0.38$ (top right), $s = 0.65$ (bottom left), and $s = 0.76$ (bottom right). In each panel, we display the resulting histogram overlaid with the truncated density in orange on the left, whilst on the right we report the corresponding QQ plot. The p-value obtained from the Kolmogorov–Smirnov test is reported in the title of each panel.

6 Non-neutral draws

In this section we give some more details on how a rejection sampler can be used to generate exact draws from the law of a non-neutral Wright–Fisher diffusion bridge. The crux is viewing the Radon–Nikodym derivative between the laws of neutral and non-neutral Wright–Fisher diffusions as (proportional to) the probability of an event that can be simulated; see (Jenkins and Spanò, 2017, Section 5) for further details.

Suppose now that apart from $\boldsymbol{\theta}$ being the mutation parameter over $[0, s)$, we also have that σ and $\boldsymbol{\eta}$ are the selection parameters, whilst over $[s, s+t)$ we have $\boldsymbol{\theta'}, \sigma'$ and $\boldsymbol{\eta}$ (observe that $\boldsymbol{\eta}$ remains the same over the interval $[0, s+t)$). Then denote by $\mathbb{WF}^{(\boldsymbol{\theta},\boldsymbol{\theta'},s,t,x)}_{\sigma,\sigma',\boldsymbol{\eta}}$ the corresponding path law of a Wright–Fisher diffusion started from x. By considering the marginal distribution of this diffusion at time $s+t$ we get the decomposition

$$d\mathbb{WF}^{(\boldsymbol{\theta},\boldsymbol{\theta'},s,t,x)}_{\sigma,\sigma',\boldsymbol{\eta}}(X^{s+t}) = d\mathbb{WF}^{(\boldsymbol{\theta},\boldsymbol{\theta'},s,t,x,z)}_{\sigma,\sigma',\boldsymbol{\eta}}(X^{s+t}) p^{\boldsymbol{\theta},\boldsymbol{\theta'}}_{\sigma,\sigma',\boldsymbol{\eta}}(x, dz, s+t)$$

where $\mathbb{WF}^{(\boldsymbol{\theta},\boldsymbol{\theta'},s,t,x,z)}_{\sigma,\sigma',\boldsymbol{\eta}}(X^{s+t})$ denotes the law of a Wright–Fisher diffusion *bridge* going from x at time 0 to z at time $s+t$ and having mutation parameters $\boldsymbol{\theta}$ over $[0, s)$ and $\boldsymbol{\theta'}$ over $[s, s+t)$, as well as selection parameters σ over $[0, s)$ and σ' over $[s, s+t)$. Furthermore, the Girsanov transform gives us that the Radon–Nikodym derivative between the laws of a neutral and non-neutral Wright–Fisher diffusion started from the same point x is

$$\frac{d\mathbb{WF}^{(\boldsymbol{\theta},\boldsymbol{\theta'},s,t,x)}_{\sigma,\sigma',\boldsymbol{\eta}}}{d\mathbb{WF}^{(\boldsymbol{\theta},\boldsymbol{\theta'},s,t,x)}_{0,0,\mathbf{0}}}\left(X^{s+t}\right) \propto e^{-\int_0^{s+t}\left(\varphi^{\boldsymbol{\theta}(u)}_{\sigma(u),\boldsymbol{\eta}}(X_u) - \varphi^{\boldsymbol{\theta}(u),-}_{\sigma(u),\boldsymbol{\eta}}\right)du}$$

where

$$\varphi^{\boldsymbol{\theta}(u)}_{\sigma(u),\boldsymbol{\eta}}(x) = \frac{\sigma(u)}{4}\Bigg[\left(-\theta_2(u)\,x + \theta_1(u)\,(1-x)\right)\eta(x) + x(1-x)\left(\frac{\sigma(u)}{2}\eta^2(x) + \eta'(x)\right)\Bigg],$$

is a polynomial in x and thus can be bounded such that

$$\varphi^{\boldsymbol{\theta}(u)}_{\sigma(u),\boldsymbol{\eta}}(x) \in [\varphi^{\boldsymbol{\theta}(u),-}_{\sigma(u),\boldsymbol{\eta}}, \varphi^{\boldsymbol{\theta}(u),+}_{\sigma(u),\boldsymbol{\eta}}],$$

see (WF3) in (Jenkins and Spanò, 2017, Section 5.2). Thus we have that

$$\begin{aligned}\frac{d\mathbb{WF}^{(\boldsymbol{\theta},\boldsymbol{\theta}',s,t,x,z)}_{\sigma,\sigma',\boldsymbol{\eta}}}{d\mathbb{WF}^{(\boldsymbol{\theta},\boldsymbol{\theta}',s,t,x,z)}_{0,0,\mathbf{0}}}\left(X^{s+t}\right) &= \frac{p^{\boldsymbol{\theta},\boldsymbol{\theta}'}_{0,0,\mathbf{0}}(x,dz,s+t)}{p^{\boldsymbol{\theta},\boldsymbol{\theta}'}_{\sigma,\sigma',\boldsymbol{\eta}}(x,dz,s+t)}\frac{d\mathbb{WF}^{(\boldsymbol{\theta},\boldsymbol{\theta}',s,t,x)}_{\sigma,\sigma',\boldsymbol{\eta}}}{d\mathbb{WF}^{(\boldsymbol{\theta},\boldsymbol{\theta}',s,t,x)}_{0,0,\mathbf{0}}}\left(X^{s+t}\right)\\ &\propto e^{-\int_0^{s+t}\left(\varphi^{\boldsymbol{\theta}(u)}_{\sigma(u),\boldsymbol{\eta}}(X_u)-\varphi^{\boldsymbol{\theta}(u),-}_{\sigma(u),\boldsymbol{\eta}}\right)du}\\ &= e^{-\int_0^{s}\left(\varphi^{\boldsymbol{\theta}}_{\sigma,\boldsymbol{\eta}}(X_u)-\varphi^{\boldsymbol{\theta},-}_{\sigma,\boldsymbol{\eta}}\right)du}e^{-\int_s^{s+t}\left(\varphi^{\boldsymbol{\theta}'}_{\sigma',\boldsymbol{\eta}}(X_u)-\varphi^{\boldsymbol{\theta}',-}_{\sigma',\boldsymbol{\eta}}\right)du}\end{aligned}$$

from which it follows that non-neutral bridges can be obtained by generating neutral candidate paths, and subsequently running a rejection step making use of two Poisson point processes, namely

- $\Xi_1 \sim \mathrm{PPP}(1)$ on $[0,s]\times[0,\varphi^{\boldsymbol{\theta},+}_{\sigma,\boldsymbol{\eta}}-\varphi^{\boldsymbol{\theta},-}_{\sigma,\boldsymbol{\eta}}]$
- $\Xi_2 \sim \mathrm{PPP}(1)$ on $[s,s+t]\times[0,\varphi^{\boldsymbol{\theta}',+}_{\sigma',\boldsymbol{\eta}}-\varphi^{\boldsymbol{\theta}',-}_{\sigma',\boldsymbol{\eta}}]$

If the corresponding Poisson points all lie in the epigraph of $u\mapsto\varphi^{\boldsymbol{\theta}(u)}_{\sigma(u),\boldsymbol{\eta}}(X_u)-\varphi^{\boldsymbol{\theta}(u),-}_{\sigma(u),\boldsymbol{\eta}}$, then the draw comes from the law of the non-neutral process as desired.

7 Demography for Figure 2

We use the demography inferred in Der Sarkissian *et al.* (2015) (as provided in Schraiber *et al.* (2016), specifically as found at `https://github.com/Schraiber/selection/blob/master/horse.all.pop`). In that file, time is given in diffusion units measured backward from the present, so $t=0$ denotes the present and all observation times satisfy $t<0$. Effective population sizes are scaled so that the most recent epoch has $N_e=1$. Because our oldest reading corresponds to 20,000 years before present, we restrict our attention to the first two epochs of this demography; the second epoch extends well beyond our oldest observation and features an exponentially shrinking effective population size. Throughout, we index observation times as $t_1,t_2,\ldots$ with $0>t_1>t_2>\ldots$, so in particular $t_1>t_2$. `EWF` allows for piecewise constant demographies, and thus we need to truncate this demography and compute an appropriate effective population size with which to approximate the exponential curve over the shorter time interval we are interested in.

To this end, we write the effective population size as a function of time in diffusion units, namely

$$N_e(t)=x_1e^{-k(t_1-t)},$$

where $k=(t_2-t_1)^{-1}(\ln(x_2)-\ln(x_1))$ and we denote the effective population sizes in the first and second epoch as x_1,x_2 respectively, whilst t_1,t_2 denote the oldest time of the epoch (i.e. the time, in diffusion time units, at which there is a change in the population size). We shall henceforth reserve the notation t to denote time being measured in diffusion time scale, whilst τ is used to denote time as measured in years. Within the epoch having as left endpoint t_2 and right endpoint t_1, standard population genetics

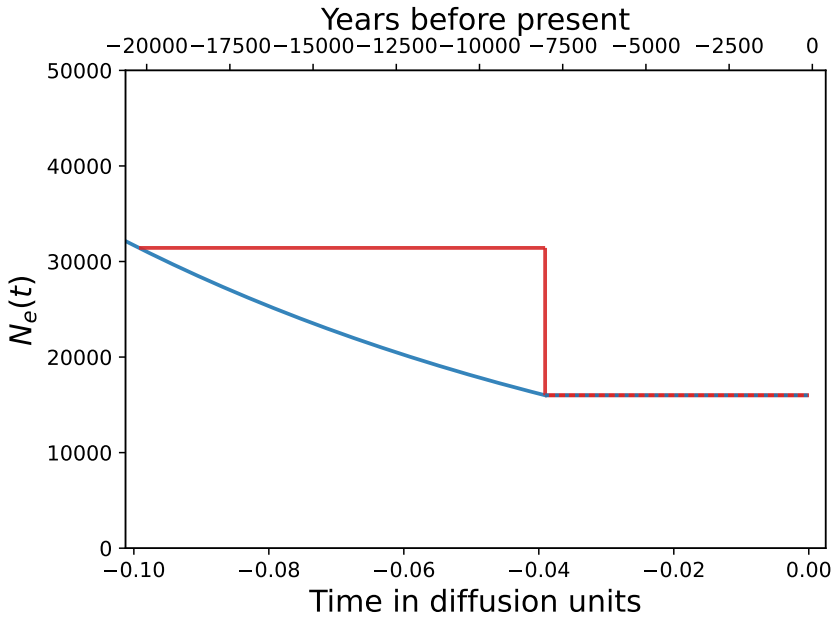

Figure 6: Plot of the demography inferred in Der Sarkissian *et al.* (2015) as provided in Schraiber *et al.* (2016) (blue), together with the piecewise linear approximation we make use of in Figure 2 (red).

theory gives us that $d\tau = 2gN_e(t)dt$, where g is the generation time. Integrating and re-arranging using the above expression for the effective population size to get that for $t \in [t_2, t_1]$ and $\tau \in [\tau_2, \tau_1]$

$$t = t_1 + \frac{1}{k} \ln \left(\frac{k(\tau - \tau_1)}{2gx_1} + 1 \right).$$

For the epoch we are interested in we have $t_1 = -0.0390625$, $t_2 = -0.2109375$, $\tau_1 = 2gN_0t_1$, $x_1 = N_0$, $x_2 = 6.875N_0$, and $\tau = -20,000$, which allows us to compute both t and $N_e(t)$ corresponding to the oldest observation recorded. Taking as initial population size $N_0 = 16,000$ and and generation gap $g = 5$ gives us the following demography, graphically summarised in Figure 6, and which was used to generate Figure 2 of the main text.

$$N_e(t) = \begin{cases} 16,000 & t \in (-0.0390625, 0] \\ 31,423 & t \leq -0.0390625 \end{cases}$$

References

Akbari, A. *et al.* (2024). Pervasive findings of directional selection realize the promise of ancient dna to elucidate human adaptation. *bioRxiv*.

Bollback, J. P. *et al.* (2008). Estimation of $2N_es$ from temporal allele frequency data. *Genetics*, **179**(1), 497–502.

Der Sarkissian, C. *et al.* (2015). Evolutionary genomics and conservation of the endangered Przewalski's horse. *Current Biology*, **25**(19), 2577–2583.

Ethier, S. N. and Kurtz, T. G. (1986). *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, New York. Characterization and convergence.

Fages, A. *et al.* (2019). Tracking five millennia of horse management with extensive ancient genome time series. *Cell*, **177**, 1419 – 1435.e31.

Ferrer-Admetlla, A. *et al.* (2016). An approximate Markov model for the Wright-–Fisher diffusion and its application to time series data. *Genetics*, **203**, 831 – 846.

Griffiths, R. C. (1979). A transition density expansion for a multi-allele diffusion model. *Adv. in Appl. Probab.*, **11**(2), 310–325.

Griffiths, R. C. (1984). Asymptotic line-of-descent distributions. *J. Math. Biol.*, **21**(1), 67–75.

He, Z. *et al.* (2020). Estimation of natural selection and allele age from time series allele frequency data using a novel likelihood-based approach. *Genetics*, **216**(2), 463–480.

Jenkins, P. A. and Spanò, D. (2017). Exact simulation of the Wright–Fisher diffusion. *Ann. Appl. Probab.*, **27**(3), 1478–1509.

Karlin, S. and Taylor, H. M. (1981). *A second course in stochastic processes*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, New York.

Lee, Y.-S. *et al.* (2014). Estimating effective population size of thoroughbred horses using linkage disequilibrium and theta ($4n\mu$) value. *Livestock Science*, **168**, 32–37.

Loog, L. *et al.* (2017). Inferring allele frequency trajectories from ancient DNA indicates that selection on a chicken gene coincided with changes in medieval husbandry practices. *Molecular Biology and Evolution*, **34**(8), 1981–1990.

Ludwig, A. *et al.* (2009). Coat color variation at the beginning of horse domestication. *Science*, **324**(5926), 485–485.

Malaspinas, A.-S. *et al.* (2012). Estimating allele age and selection coefficient from time-serial data. *Genetics*, **192**(2), 599–607.

Mallick, S. *et al.* (2024). The Allen Ancient DNA Resource (AADR) a curated compendium of ancient human genomes. *Scientific Data*, **11**(1), 182.

Sant, J. *et al.* (2023). EWF: simulating exact paths of the Wright–Fisher diffusion. *Bioinformatics*, **39**(1), btad017.

Schraiber, J. G. *et al.* (2016). Bayesian inference of natural selection from allele frequency time series. *Genetics*, **203**(1), 493–511.

Sohail, M. S. *et al.* (2021). MPL resolves genetic linkage in fitness inference from complex evolutionary histories. *Nature biotechnology*, **39**(4), 472–479.

Steinrücken, M. *et al.* (2015). SpectralTDF: transition densities of diffusion processes with time-varying selection parameters, mutation rates and effective population sizes. *Bioinformatics*, **32**(5), 795–797.

Tavaré, S. (1984). Lines-of-descent and genealogical processes, and their applications in population genetics models. *Adv. in Appl. Probab.*, **16**, 27–27.

Wutke, S. *et al.* (2016). Spotted phenotypes in horses lost attractiveness in the middle ages. *Scientific Reports*, **6**.

Živković, D. *et al.* (2015). Transition densities and sample frequency spectra of diffusion processes with selection and variable population size. *Genetics*, **200**(2), 601–617.