Closed form modeling of evolutionary rates by exponential Brownian functionals

Nicolas Privault*
Division of Mathematical Sciences
Nanyang Technological University
21 Nanyang Link
Singapore 637371

Stéphane Guindon†
Department of Statistics
The University of Auckland
Auckland, 1010
New Zealand

September 12, 2015

Abstract

Accurate estimation of species divergence times from the analysis of genetic sequences relies on probabilistic models of evolution of the rate of molecular evolution. Importantly, while these models describe the sample paths of the substitution rates along a phylogenetic tree, only the (random) average rate can be estimated on each edge. For mathematical convenience, the stochastic nature of these averages is generally ignored. In this article we derive the probabilistic distribution of the average substitution rate assuming a geometric Brownian motion for the sample paths, and we investigate the corresponding error bounds via numerical simulations. In particular we confirm the validity of the gamma approximation proposed in Guindon [5] for "small" values of the autocorrelation parameter.

Key words: Evolutionary rates, exponential Brownian functionals, geometric Brownian bridge, molecular clocks, phylogenetics.
Mathematics Subject Classification (2010): 92D15; 92D20; 33C10; 62J10; 60J22; 60J27; 60J65, 81S40; 60H30; 60H07.

1 Introduction

The comparison of homologous nucleotide or amino-acid sequences collected from multiple species or populations serves as a basis to infer divergence times [26, 20, 16,

*nprivault@ntu.edu.sg
†s.guindon@auckland.ac.nz

9, 19, 22, 11]. In short, aligned homologous genetic sequences are used to estimate a phylogenetic tree with the length of each branch corresponding to an expected number of substitutions that accumulated at individual sites of the alignment. Fossil data and/or prior knowledge about the overall substitution rate help translate this number of substitutions into calendar time units.

The hierarchical Bayesian estimation framework is well suited to the inference of model parameters in this particular context. The objective is then to estimate the joint posterior probability density of the model parameters, including the phylogenetic tree topology, the pathwise average rate of substitution along edges, the rate of substitution at the nodes of the tree (or the midpoint of each edge, as in Rannala and Yang [18]), the node ages, plus other parameters generally considered as nuisance parameters. The posterior density of a particular combination of parameter values is then proportional to the product of the likelihood, i.e., the probability of the sequence alignment given the model parameters by the prior probability (density) of these values, cf. Felsenstein [3].

In a pioneering work, Thorne, Kishino and Painter [22] implemented such an approach, assuming a simple model describing the variability of substitution rates during the course of evolution. Specifically, in their model, the substitution rate follows a geometric Brownian process, i.e., the logarithm of the rate at the end of a branch is normally distributed with mean set to the logarithm of the rate at the beginning of the branch, and variance proportional to the time elapsed along that edge. Mean reverting stochastic processes such as the Ornstein-Uhlenbeck process or the Cox-Ingersoll-Ross (CIR) process were also used afterwards, cf. respectively Aris-Brosou and Yang [2], and Lepage et al. [13].

While these models define priors on the rate trajectories, i.e., priors on the sample paths of the rate process, only average rates can be estimated from the available data. For instance, given the observed differences between two sequences and the time to their common ancestor, one can only estimate the expected rate at which substitu-

tions accumulated in the corresponding time period. Hence, while rate trajectories lend themselves to statistical modeling, they are not identifiable from the available data. For this reason, it is necessary to derive average rates conditional on the corresponding trajectories.

The approach used to derive these averages is of course of central importance. Thorne, Kishino and Painter [22] and subsequent studies used deterministic functions of the rates at both extremities or on the mid-point of the corresponding branch. This simplification therefore amounts to considering the average substitution rate conditional on the rate at the branch extremities as a constant. However, in the general and more realistic setting where finite time intervals are considered, the average substitution rate is a random variable and should therefore be treated as such. Considering a given branch, a random "trajectory" then describes the fluctuation of rates at any individual site of the alignment. While the trajectories at different sites are distinct, they are all governed by the same stochastic process with single drift and diffusion parameters. Such a model thus authorizes a given branch to evolve rapidly at a first site and slowly at a second, while another branch can evolve slowly at the first site and then fast at the second, in a manner similar to that of the covarion model (Tuffley and Steel [23]).

Lepage et al. [13] were the first to acknowledge and tackle this issue. Assuming a CIR model for the rate trajectories, they derived analytical expressions for the transition probabilities between character states along the phylogeny. Lepage et al. [12] have given an application of the techniques of Lepage et al. [13], however they have ignored the stochasticity of the pathwise average rate by equating it to its expectation under the CIR model.

Guindon [5] suggested that average substitution rates are approximately gamma distributed when rate trajectories are governed by a geometric Brownian process. This approximation is convenient from a computational perspective as it does not incur any extra calculation in the derivation of the likelihood compared to the traditional

"fixed-average" approach. The first and second moments of the gamma distribution conditional on the rate at the branch extremities were approximated by a Taylor series in order to calculate these values in practice.

In this paper, we give closed form integral expressions for the probability density of the (random) average rates of substitution using Bessel functions. We compare those expressions to the gamma distribution approximation of Guindon (2013) and show that it is accurate in a broad range of biologically realistic conditions. We also provide explicit formulas for the first two moments of the average rates and we use them as an alternative to the Taylor series approximation applied in Guindon (2013).

We proceed as follows. After presenting the Bayesian estimation of divergence times and the modeling of substitution rates in Section 2, we present closed form formulas for the conditional distribution of pathwise averages of rates in Section 3. Section 4 is devoted to the gamma approximation of these quantities. Section 5 contains the proofs of our main results, including the computations of conditional mean and variance needed for the approximations.

2 Bayesian estimation of divergence times

In this section, we present the phylogenetic model used to estimate divergence times from the analysis of molecular and fossil data. We then describe how the posterior probability of the parameters of interest can be estimated in a Bayesian framework. Let τ_n be a reconstructed phylogenetic tree [15, 4] with n tips. It is a rooted, binary tree with unique leaf labels and ultrametric edge lengths. Let $\mathbf{R}$ be a vector of $2n-1$ substitution rates at the vertices of τ_n and $\mathbf{X}$ the vector of $2n-2$ average rates along the edges of the tree. Let $\mathbf{T}$ be the vector of $n-1$ (internal) node heights distinct from the tips. For simplicity, we omit other (nuisance) parameters from the phylogenetic model.

The data consist in an alignment of n homologous sequences. Each sequence in the

alignment is of length l and S_{ij} represents the character (i.e., a nucleotide, amino-acid or unknown character) observed in the sequence corresponding to the i-th tip in τ_n at the j-th column in the alignment. Let $\mathbf{S}$ denote the whole alignment. Calibration information is usually incorporated in the analysis through fossil data. This source of information is used to date some of the internal nodes, corresponding to the ancestor of groups of species for which fossil information, denoted as $\mathbf{F}$, is available.

Samples from the posterior probability density of the parameters of interest are obtained using Markov Chain Monte Carlo. It is usually decomposed as follows:

$$p(\tau_n, \mathbf{R}, \mathbf{T}|\mathbf{F}, \mathbf{S}) \propto \int \Pr(\mathbf{S}|\mathbf{X}, \mathbf{T}, \tau_n) p(\mathbf{X}|\mathbf{T}, \mathbf{R}) p(\mathbf{R}|\mathbf{T}, \tau_n) p(\mathbf{T}, \tau_n|\mathbf{F}) d\mathbf{X},$$

where $\Pr(\mathbf{S}|\mathbf{X}, \mathbf{T}, \tau_n)$ is the conditional probability of the alignment given the phylogeny, and is computed efficiently using Felsenstein's [3] peeling algorithm. The quantity $p(\mathbf{R}|\mathbf{T}, \tau_n)$ represents the rate trajectories, i.e., the joint probability density of the rates at every node in the tree given the node heights, whereas $p(\mathbf{T}, \tau_n|\mathbf{F})$ is the joint density of node heights given the fossil information, which can be calculated under a variety of models, the most popular ones being Kingman's coalescent [8] and the birth-death process [21]. The function $p(\mathbf{X}|\mathbf{T}, \mathbf{R})$ represents the conditional density of the (random) average rates given the times elapsed along each branch of the phylogeny and the substitution rates at each node. As indicated previously, the vector $\mathbf{X}$ of substitution rates is usually considered deterministic by equating every value in $\mathbf{X}$ to the arithmetic average of the pairs of values in $\mathbf{R}$ corresponding to each edge extremities, and to zero otherwise.

In this paper we focus on the actual distribution of $\mathbf{X}|\mathbf{T}, \mathbf{R}$ and present an analytical approach to the exact calculation of $\int \Pr(\mathbf{S}|\mathbf{X}, \mathbf{T}, \tau_n) p(\mathbf{X}|\mathbf{T}, \mathbf{R}) d\mathbf{X}$, by assuming that the distribution of the rate of evolution R_t at time t on a given edge given $\mathbf{T}, \tau_n$ is modeled by a geometric Brownian bridge. Precisely we take

$$\log R_t = \log x + \frac{t}{T} \log(y/x) + \sigma U_t = \log x + \frac{t}{T} \log(y/x) + \sigma(W_t - tW_T/T),$$

conditioned to start at $\log x$ and to end at $\log y$, where

$$U_t := W_t - \frac{t}{T} W_T, \qquad t \in [0,T],$$

is the standard Brownian bridge on $[0,T]$ with $U_0 = U_T = 0$, cf. [22], [10], and σ is the rate autocorrelation parameter. The pathwise average substitution rate along a branch of length T (in calendar time units) is given by Λ_T/T, where

$$\Lambda_T = \int_0^T R_s ds. \tag{2.1}$$

The character state X_t (typically a nucleotide or an amino-acid) at time $t \in \mathbb{R}_+$, is modeled as in [5] as the time-changed process

$$X_t = Y_{\Lambda_t}$$

where $(Y_t)_{t\in \mathrm{R}_+}$ is a continuous-time Markov chain with finite state space $\{1,\dots,n\}$ and infinitesimal $n \times n$ generator matrix $\mathbf{Q}$, and Λ_t is the random time change

$$\Lambda_t := \int_0^t R_s ds = R_0 \int_0^t e^{\sigma W_s - p\sigma^2 s/2} ds, \qquad t \in \mathbb{R}_+.$$

We note that the transition probabilities of $(X_t)_{t\in \mathrm{R}_+}$ are given by

$$\begin{aligned}
&\mathbb{P}(X_T = b \mid X_0 = a,\ R_0 = x,\ R_T = y) \\
&\quad = E[\mathbf{1}_{\{X_T = b\}} \mid X_0 = a,\ R_0 = x,\ R_T = y] \\
&\quad = E[\mathbf{1}_{\{Y_{\Lambda_T} = b\}} \mid X_0 = a,\ R_0 = x,\ R_T = y] \\
&\quad = E[E[\mathbf{1}_{\{Y_{\Lambda_T} = b\}} \mid Y_0 = a,\ \Lambda_T,\ R_0 = x,\ R_T = y] \mid Y_0 = a,\ R_0 = x,\ R_T = y] \\
&\quad = E[[e^{\Lambda_T \mathbf{Q}}]_{a,b} \mid Y_0 = a,\ R_0 = x,\ R_T = y] \\
&\quad = E[[e^{\Lambda_T \mathbf{Q}}]_{a,b} \mid R_0 = x,\ R_T = y] \\
&\quad = [h^T_{x,y}(-\mathbf{Q})]_{a,b}, \qquad 1 \le a,b \le n,
\end{aligned}$$

where $P_z := e^{z\mathbf{Q}}$, $z > 0$, is the transition semigroup of the continuous-time chain $(Y_t)_{t\in \mathrm{R}_+}$ and $h^T_{x,y}(\lambda)$ is the conditional Laplace transform

$$h^T_{x,y}(\lambda) := E\left[\exp\left(-\lambda \Lambda_T\right) \Big| R_0 = x,\ R_T = y\right], \qquad \lambda \ge 0, \quad x,y > 0. \tag{2.2}$$

By diagonalization of the symmetric matrix $\mathbf{Q} = \mathbf{M}^{-1}\mathbf{D}\mathbf{M}$ where $\mathbf{D} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ and $\lambda_1 = 0, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of $\mathbf{Q}$, this allows us to compute the transition probabilities of $(X_t)_{t\in\mathrm{R}_+}$ as

$$\mathbb{P}(X_T = b \mid X_0 = a,\ R_0 = x,\ R_T = y) = [\mathbf{M}^{-1}\mathbf{D}(x,y)\mathbf{M}]_{a,b}$$

where $\mathbf{D}(x,y)$ is the diagonal matrix

$$\mathbf{D}(x,y) = \mathrm{diag}(1, h^T_{x,y}(-\lambda_2), \ldots, h^T_{x,y}(-\lambda_n)).$$

Note that if $\mathbf{Q}$ has only non-zero entries then $\lambda_1 = 0$ is an eigenvalue of $\mathbf{Q}$ with eigenvector $(1, \ldots, 1)$, while all other eigenvalues $\lambda_2, \ldots, \lambda_n$ are strictly negative, cf. e.g. Theorem 2.1 in Chapter 10 of [7], and correspond to eigenvectors $u^{(2)}, \ldots, u^{(n)} \in \mathbb{R}^n$ which are orthogonal to the invariant distribution $(\pi_1, \ldots, \pi_n)$ of $(Y_t)_{t\in\mathrm{R}_+}$. The matrices $\mathbf{M}$ and $\mathbf{M}^{-1}$ then take the form

$$\mathbf{M} = \begin{bmatrix} \pi_1 & \cdots & \pi_n \\ M_{2,1} & \cdots & M_{2,n} \\ \vdots & \ddots & \vdots \\ M_{n,1} & \cdots & M_{n,n} \end{bmatrix} \quad \text{and} \quad \mathbf{M}^{-1} = \begin{bmatrix} 1 & u_1^{(2)} & \cdots & u_1^{(n)} \\ 1 & u_2^{(2)} & \cdots & u_2^{(n)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & u_n^{(2)} & \cdots & u_n^{(n)} \end{bmatrix}.$$

By (2.2) we have $h^T_{x,y}(\lambda_1) = h^T_{x,y}(0) = 1$ and $\lim_{T\to\infty} h^T_{x,y}(\lambda_k) = 0$, $k = 2, \ldots, n$, when T tends to infinity, hence

$$\begin{aligned}
\lim_{T\to\infty} [\mathbb{P}(X_T = b \mid X_0 = a,\ R_0 = x,\ R_T = y)]_{1\le a,b\le n} &= [\mathbf{M}^{-1}\mathbf{D}(x,y)\mathbf{M}]_{1\le a,b\le n} \\
&= \begin{bmatrix} 1 & u_1^{(2)} & \cdots & u_1^{(n)} \\ 1 & u_2^{(2)} & \cdots & u_2^{(n)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & u_n^{(2)} & \cdots & u_n^{(n)} \end{bmatrix} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \pi_1 & \cdots & \pi_n \\ M_{2,1} & \cdots & M_{2,n} \\ \vdots & \ddots & \vdots \\ M_{n,1} & \cdots & M_{n,n} \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \pi_1 & \cdots & \pi_n \\ M_{2,1} & \cdots & M_{2,n} \\ \vdots & \ddots & \vdots \\ M_{n,1} & \cdots & M_{n,n} \end{bmatrix} \\
&= \begin{bmatrix} \pi_1 & \cdots & \pi_n \\ \pi_1 & \cdots & \pi_n \\ \vdots & \ddots & \vdots \\ \pi_1 & \cdots & \pi_n \end{bmatrix},
\end{aligned}$$

which recovers the limiting distribution of $(Y_t)_{t\in\mathrm{R}_+}$.

Finally we note the scaling property

$$\Lambda_T = \int_0^T e^{\sigma W_t - p\sigma^2 t/2} dt \overset{d}{=} \int_0^T e^{W_{\sigma^2 t} - p\sigma^2 t/2} dt = \frac{1}{\sigma^2} \int_0^{\sigma^2 T} e^{W_t - pt/2} dt, \quad T > 0, \tag{2.3}$$

where $\overset{d}{=}$ denotes equality in distribution, which shows that the probability distribution of the average rate Λ_T/T depends only on the product $\sigma^2 T$.

3 Closed form expressions

In this section we show how the Laplace transform

$$h^T_{x,y}(\lambda) = \int_0^\infty e^{-z\lambda} d\mathbb{P}(\Lambda_T = z \mid R_0 = x,\ R_T = y), \qquad \lambda > 0, \tag{3.1}$$

defined in (2.2) can be computed in closed integral form using Bessel functions, and we provide closed form expressions for the conditional mean and variance of Λ_T. The proofs of the results of this section are deferred to Section 5.

In the next lemma we use the function

$$\theta(v,\tau) := \frac{v e^{\pi^2/(2\tau)}}{\sqrt{2\pi^3\tau}} \int_0^\infty e^{-\xi^2/(2\tau)} e^{-v\cosh\xi} \sinh(\xi) \sin(\pi\xi/\tau)\, d\xi, \qquad v,\tau > 0, \tag{3.2}$$

in order to evaluate the conditional distribution of Λ_T given $R_0 = x$ and $R_T = y$.

Lemma 3.1 *For all $x, y > 0$ we have*

$$\begin{aligned} &\mathbb{P}\left(\Lambda_T \in du \,\middle|\, R_0 = x,\ R_T = y\right) \\ &\quad = \sigma\sqrt{\frac{\pi T}{2}} \exp\left(\frac{(\log(y/x))^2}{2\sigma^2 T} - 2\frac{x+y}{\sigma^2 u}\right) \theta\left(\frac{4\sqrt{xy}}{\sigma^2 u}, \frac{\sigma^2 T}{4}\right) \frac{du}{u}, \qquad u > 0. \end{aligned} \tag{3.3}$$

From Lemma 3.1 we will deduce a closed form integral expression for the conditional Laplace transform $h^T_{x,y}(\lambda) = E\left[\exp\left(-\lambda\Lambda_T\right) \,\middle|\, R_0 = x,\ R_T = y\right]$ of Λ_T. Note that $\theta(v,\tau)$ is difficult to evaluate numerically due to the oscillating behavior of its integrand, cf. [6], [17] for several attempts to the numerical computation of the function $\theta(v,\tau)$. As will be shown subsequently, the numerical evaluation of this integral

expression (3.3) fails for small values of T due to the highly oscillating behavior of its integrand, making the gamma approximation described in [5] preferable in this parameter range.

In the next proposition we use the modified Bessel function of the second kind

$$K_\zeta(z) = \frac{z^\zeta}{2^{\zeta+1}} \int_0^\infty \exp\left(-u - \frac{z^2}{4u}\right) \frac{du}{u^{\zeta+1}}, \qquad \zeta \in \mathbb{R}, \quad z \in \mathbb{C}, \quad \mathcal{R}(z^2) > 0, \quad (3.4)$$

cf. [24] page 183, to compute the Laplace transform (3.1). Note that the expression (3.5) below is more complex than the closed-form expressions available for the CIR model, compare for example with Relation (3.51) of [1] or Relations (23)-(24) of [12].

Proposition 3.2 *For all $x, y, \lambda > 0$ we have*

$$h^T_{x,y}(\lambda) = \quad (3.5)$$

$$\frac{4x\sqrt{\lambda}e^{-\sigma^2 T/8}}{\pi^{3/2}\sigma^2 p(y/x)\sqrt{T}} \int_0^\infty e^{2(\pi^2-\xi^2)/(\sigma^2 T)} \sin\left(\frac{4\pi\xi}{\sigma^2 T}\right) \sinh(\xi) \frac{K_1\left(\sqrt{8\lambda}\sqrt{x+2\sqrt{xy}\cosh\xi+y}/\sigma\right)}{\sqrt{x+2\sqrt{xy}\cosh\xi+y}} d\xi.$$

The numerical evaluation of (3.5) can be done by standard integral discretization methods, however the integral is also oscillating due to the presence of the term $\sin(4\pi\xi/(\sigma^2 T))$, making the numerical scheme unstable for small values of $\sigma^2 T$, as observed above for $\theta(v,\tau)$.

For the next proposition we will need the functions

$$a_T(z) := \frac{1}{\sigma^2 p(z)} \left(\Phi\left(\frac{\log z}{\sqrt{\sigma^2 T}} + \frac{1}{2}\sqrt{\sigma^2 T}\right) - \Phi\left(\frac{\log z}{\sqrt{\sigma^2 T}} - \frac{1}{2}\sqrt{\sigma^2 T}\right)\right),$$

and

$$b_T(z) := \frac{1}{\sigma^2 q(z)} \left(\Phi\left(\frac{\log z}{\sqrt{\sigma^2 T}} + \sqrt{\sigma^2 T}\right) - \Phi\left(\frac{\log z}{\sqrt{\sigma^2 T}} - \sqrt{\sigma^2 T}\right)\right),$$

where

$$p(z) = \frac{1}{\sqrt{2\pi\sigma^2 T}} e^{-\left(\sigma^2 T/2 + \log z\right)^2/(2\sigma^2 T)}, \quad \text{and} \quad q(z) = \frac{1}{\sqrt{2\pi\sigma^2 T}} e^{-\left(\sigma^2 T + \log z\right)^2/(2\sigma^2 T)},$$

$z > 0$, and

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy, \qquad x \in \mathbb{R},$$

is the standard Gaussian cumulative distribution function. In the next Proposition 3.3 we derive the closed form expressions of $E[\Lambda_T \mid R_0 = x,\ R_T = y]$ and $\mathrm{Var}\,[\Lambda_T \mid R_0 = x,\ R_T = y]$.

Proposition 3.3 *We have*

$$\begin{aligned} E[\Lambda_T \mid R_0 = x,\ R_T = y] &= x a_T(y/x) \\ &= \frac{x}{\sigma^2 p(y/x)} \left(\Phi\left(\frac{\log(y/x)}{\sqrt{\sigma^2 T}} + \frac{1}{2}\sqrt{\sigma^2 T} \right) - \Phi\left(\frac{\log(y/x)}{\sqrt{\sigma^2 T}} - \frac{1}{2}\sqrt{\sigma^2 T} \right) \right), \end{aligned} \tag{3.6}$$

and

$$E[(\Lambda_T)^2 \mid R_0 = x,\ R_T = y] = \frac{2x}{\sigma^2}(x b_T(y/x) - (x+y) a_T(y/x)), \qquad x, y > 0. \tag{3.7}$$

From Proposition 3.3 we also obtain the equivalences as T tends to infinity

$$E\left[\Lambda_T \middle| R_0 = x,\ R_T = y \right] \simeq \frac{1}{\sigma}\sqrt{2\pi T x y} e^{\sigma^2 T/8}, \qquad [T \to \infty], \tag{3.8}$$

and

$$E\left[\Lambda_T^2 \middle| R_0 = x,\ R_T = y \right] \simeq \frac{2}{\sigma^3} x y \sqrt{2\pi T} e^{\sigma^2 T/2}, \qquad [T \to \infty], \tag{3.9}$$

hence both quantities tend to infinity as T goes to infinity.

In addition we can also use the short time asymptotics for $E[\Lambda_T \mid R_0 = x,\ R_T = y]$ and $\mathrm{Var}\,[\Lambda_T \mid R_0 = x,\ R_T = y]$ of Proposition 5.1 in Section 5 to derive the Laplace transform approximation

$$\begin{aligned} h_{x,y}^T(\lambda) &= E\left[\exp\left(-\lambda \Lambda_T\right) \middle| R_0 = x,\ R_T = y \right] \\ &\simeq 1 - \lambda E[\Lambda_T \mid R_0 = x,\ R_T = y] + \frac{\lambda^2}{2} E[(\Lambda_T)^2 \mid R_0 = x,\ R_T = y] \\ &\simeq 1 - \frac{\lambda T (y-x)}{\log(y/x)} - \lambda \sigma^2 T^2 \left(\frac{x+y}{2(\log(y/x))^2} - \frac{y-x}{(\log(y/x))^3} \right) + \frac{T^2 \lambda^2}{2} \frac{(y-x)^2}{(\log(y/x))^2} + o(T^2), \end{aligned} \tag{3.10}$$

in small time T.

4 Gamma approximation

In order to approximate the distribution of Λ_T we will use the gamma probability density function

$$f(x) = \frac{1}{\eta^\beta \Gamma(\beta)} x^{\beta-1} e^{-x/\eta}, \qquad x > 0, \tag{4.1}$$

with mean and variance

$$E[X] = \beta\eta, \qquad \mathrm{Var}\,[X] = \beta\eta^2,$$

where $\Gamma(a) = \displaystyle\int_0^\infty t^{a-1}e^{-t}dt$, $a > 0$, is the gamma function. The shape parameter $\beta > 0$ and the scale parameter $\eta > 0$ can be estimated from the mean and variance of X as

$$\eta = \frac{\mathrm{Var}\,[X]}{E[X]}, \qquad \beta = \frac{(E[X])^2}{\mathrm{Var}\,[X]} = \frac{E[X]}{\eta}. \tag{4.2}$$

The next proposition is a consequence of (4.2) and Proposition 3.3.

Proposition 4.1 *For any $x, y > 0$, the gamma random variable with scale parameter*

$$\eta_T(x,y) := \frac{2}{\sigma^2}\left(x\frac{b_T(y/x)}{a_T(y/x)} - x - y\right) - xa_T(y/x), \tag{4.3}$$

and shape parameter

$$\beta_T(x,y) := x\frac{a_T(y/x)}{\eta_T(x,y)} \tag{4.4}$$

has same first and second moments as Λ_T given $\{R_0 = x,\ R_T = y\}$.

Based on Proposition 4.1 we will use the approximation

$$\mathbb{P}\left(\Lambda_T \in du \,\middle|\, R_0 = x,\ R_T = y\right) \simeq \frac{e^{-u/\eta_T(x,y)}}{\eta_T(x,y)} \frac{(u/\eta_T(x,y))^{-1+\beta_T(x,y)}}{\Gamma\left(\beta_T(x,y)\right)} du, \qquad u > 0. \tag{4.5}$$

Note that from (3.8)-(3.9) we have the equivalences

$$\eta_T(x,y) \simeq \frac{2}{\sigma^2}\sqrt{xy}e^{3\sigma^2 T/8}, \qquad [T \to \infty], \tag{4.6}$$

and

$$\beta_T(x,y) \simeq \sqrt{\frac{\pi\sigma^2 T}{2}}e^{-\sigma^2 T/4}, \qquad [T \to \infty], \tag{4.7}$$

as T becomes large. On the other hand, by Proposition 5.1 in Section 5 and (4.2) we also have the short time limits

$$\lim_{T\to 0} \eta_T(x,y) = 0 \quad \text{and} \quad \lim_{T\to 0} \beta_T(x,y) = +\infty. \tag{4.8}$$

Figure 1 represents $\eta_T(x,y)$ as a function of T, together with its large time approximation (4.6) when $\sigma = 1$ and $x = 0.0001$, $y = 0.001$.

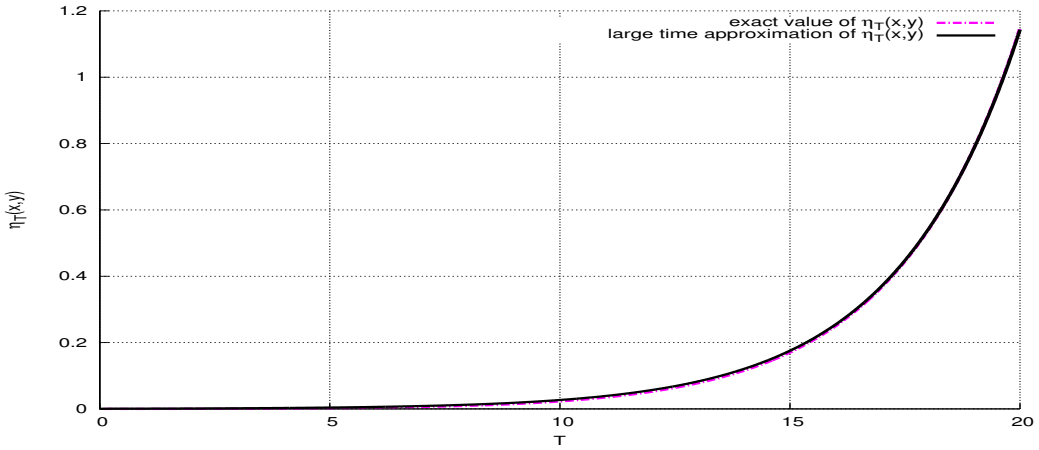

Figure 1: Graph of $T \mapsto \eta_T(x,y)$ in (4.3) and its large time approximation (4.6).

In Figure 2 we plot $\beta_T(x,y)$ as a function of T, together with its large time approximation (4.7) with $\sigma = 1$.

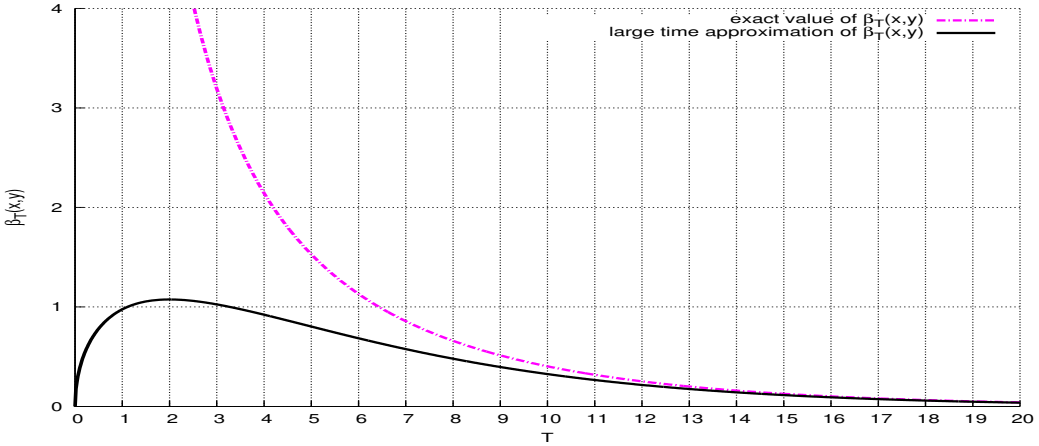

Figure 2: Graph of $T \mapsto \beta_T(x,y)$ in (4.4) and its large time approximation (4.7).

Figure 2 shows in particular that we should require $\sigma^2 T \leq 6$ in order to have $\beta_T(x,y) > 1$ so that the approximating gamma probability density with parameter

$\beta_T(x,y)$ does not tend to infinity at 0, as observed in the graphs of actual densities in Figures 4 to 6 below. Similarly, the condition $\beta_T(x,y) > 2$, i.e. $\sigma^2 T \leq 4$, would be required in order for the derivative of the gamma probability density to vanish at the origin. Note also that in order to have $\mathbb{P}(R_T/R_0 \geq 2)$ in the interval $[10^{-3}, 0.25]$ we need to choose $\sigma^2 T$ within $[0.05, 1.05]$. For this range of parameter values, the expected value of the (random) average rate as approximated using the gamma distribution is reasonably close to the corresponding exact values. Based on these remarks, when $T = 1$ we will retain the range $[0, 2]$ for the values of σ within the domain of validity of the gamma approximation in the numerical simulations described in the following sections.

In Figures 3-6 we compare the density approximation (4.5) with the closed form expression (3.3) for $T = 1$ with $x = 0.0001$, $y = 0.001$ and $\sigma = 0.57$, $\sigma = 1$ and $\sigma = 2$, respectively. It turns out that the numerical implementation of the integral expression (3.6) fails for values of $\sigma^2 T < 0.5$ due to instabilities in the oscillating integral (3.2) of $\theta(v,\tau)$, while the performance of the gamma approximation (4.5) is better for σ lower than 1.

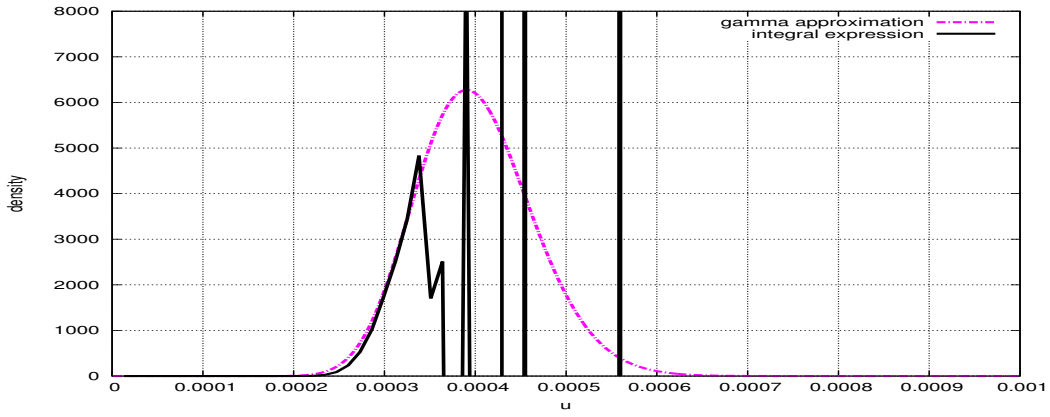

Figure 3: The integral expression (3.3) *vs* its gamma approximation (4.5) with $\sigma = 0.57$ and $\beta^1_{x,y} = 38.84$.

The graph of Figure 6 confirms that the gamma approximation can no longer be used to fit the actual probability density when $\beta^1_{x,y} < 1$.

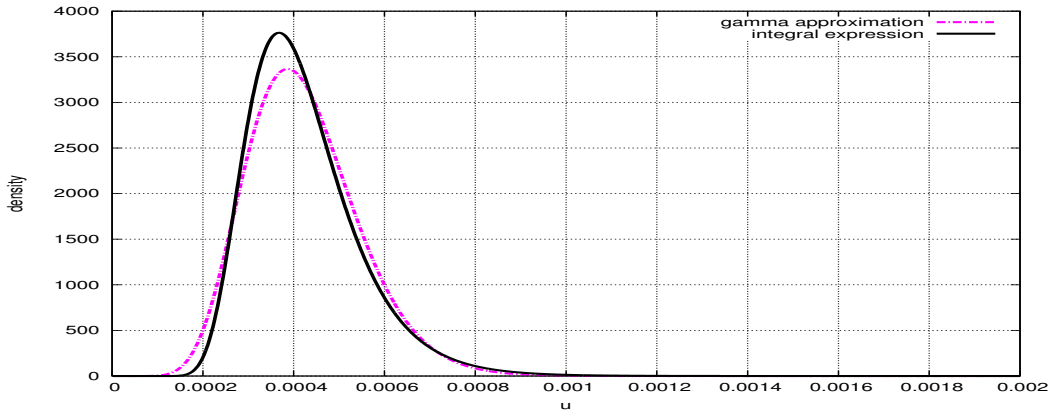

Figure 4: The integral expression (3.3) *vs* its gamma approximation (4.5) with $\sigma = 1$ and $\beta^1_{x,y} = 11.80$.

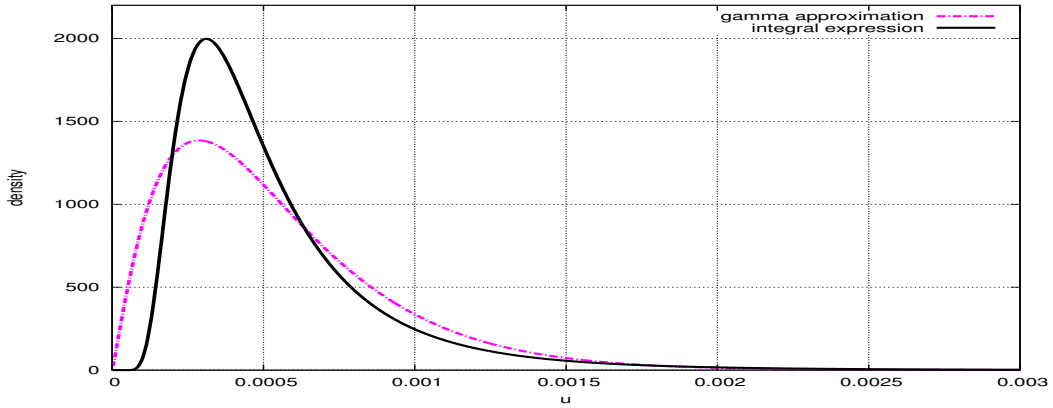

Figure 5: The integral expression (3.3) *vs* its gamma approximation (4.5) with $\sigma = 2$ and $\beta^1_{x,y} = 2.144$.

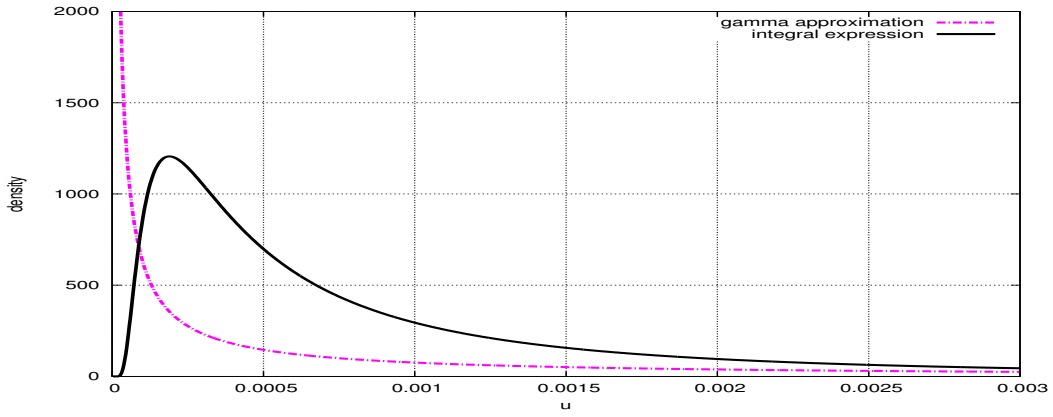

Figure 6: The integral expression (3.3) *vs* its gamma approximation (4.5) with $\sigma = 4$ and $\beta^1_{x,y} = 0.1014$.

Transition probabilities

From Proposition 4.1 the Laplace transform of Λ_T also admits the gamma approximation

$$h^T_{x,y}(\lambda) = E\left[\exp\left(-\lambda\Lambda_T\right) \ \Big| R_0 = x,\ R_T = y\right] \simeq \left(1+\lambda\eta_T(x,y)\right)^{-\beta_T(x,y)}, \qquad (4.9)$$

$\lambda \geq 0$, $x, y > 0$, hence the transition probability approximation

$$[\mathbb{P}(X_T = b \mid X_0 = a,\ R_0 = x,\ R_T = y)]_{1\leq a,b\leq n} \simeq M^{-1}D(x,y)M$$

where $D(x,y)$ is the diagonal matrix

$$D(x,y) = \mathrm{diag}\left(1, \left(1+\lambda_2\eta_T(x,y)\right)^{-\beta_T(x,y)}, \ldots, \left(1+\lambda_n\eta_T(x,y)\right)^{-\beta_T(x,y)}\right).$$

From (4.6)-(4.7) we find that

$$\begin{aligned}
h^T_{x,y}(\lambda) &= E\left[\exp\left(-\lambda\Lambda_T\right) \ \Big| R_0 = x,\ R_T = y\right] \\
&\simeq \left(1+\frac{2\lambda}{\sigma^2}\sqrt{xy}e^{3\sigma^2T/8}\right)^{-\sigma\sqrt{\pi T/2}e^{-\sigma^2T/4}} \\
&\simeq \exp\left(-\sigma\sqrt{\frac{\pi T}{2}}e^{-\sigma^2T/4}\log\left(1+\frac{2\lambda}{\sigma^2}\sqrt{xy}e^{3\sigma^2T/8}\right)\right) \\
&\simeq \exp\left(-\frac{3}{8}(\sigma^2T)^{3/2}\sqrt{\frac{\pi}{2}}e^{-\sigma^2T/4}\right),
\end{aligned}$$

which shows that the gamma approximation (4.9) tends to 1 as T tends to infinity, although the exact value of $h^T_{x,y}(\lambda)$ in (3.5) tends to 0 when $\lambda > 0$, as can be checked in Figure 7.

For numerical simulations of the transition probabilities $\mathbb{P}(X_t = b \mid X_0 = a,\ R_0 = x,\ R_T = y)$ we consider the Jukes-Cantor model

$$\mathbf{Q} = \begin{bmatrix} -1+1/n & 1/n & \cdots & 1/n \\ 1/n & -1+1/n & \cdots & 1/n \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 1/n & 1/n & \cdots & -1+1/n \end{bmatrix}$$

with limiting distribution $(\pi_1, \ldots, \pi_n) = (1/n, \ldots, 1/n)$ and

$$\mathbf{M} = \begin{bmatrix} 1/n & 1/n & 1/n & \cdots & 1/n \\ -1/n & -1/n & -1/n & \cdots & 1-1/n \\ \vdots & \vdots & \vdots & \iddots & \vdots \\ -1/n & -1/n & 1-1/n & \cdots & -1/n \\ -1/n & 1-1/n & -1/n & \cdots & -1/n \end{bmatrix} \quad \text{and} \quad \mathbf{M}^{-1} = \begin{bmatrix} 1 & -1 & -1 & \cdots & -1 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 1 & 0 & \ddots & \ddots & 0 \\ 1 & 0 & 0 & \cdots & 1 \end{bmatrix},$$

with eigenvalues $(\lambda_1, \lambda_2, \ldots, \lambda_n) = (0, -1, \ldots, -1)$, which yields

$$[\mathbb{P}(X_t = b \mid X_0 = a,\ R_0 = x,\ R_T = y)]_{1 \le a,b \le n}$$
$$= \begin{bmatrix} 1/n + (1-1/n)h^T_{x,y}(1) & (1-h^T_{x,y}(1))/n & \cdots & (1-h^T_{x,y}(1))/n \\ (1-h^T_{x,y}(1))/n & 1/n + (1-1/n)h^T_{x,y}(1) & \ddots & (1-h^T_{x,y}(1))/n \\ \vdots & \vdots & \ddots & \vdots \\ (1-h^T_{x,y}(1))/n & (1-h^T_{x,y}(1))/n & \cdots & 1/n + (1-1/n)h^T_{x,y}(1) \end{bmatrix}.$$

Figure 7 compares the gamma approximation (4.9) in the Jukes-Cantor model with $n = 4$ with the integral expression (3.5) evaluated by a standard discretization.

We check in Figure 7 that the numerical evaluation of (3.5) fails with divergent infinities for "small" values of $\sigma^2 T \le 0.5$, due to the oscillating behavior of the integral, while the gamma approximation performs well until $T \simeq 50$, i.e. $\sigma^2 T \simeq 0.5^2 \times 50 = 12.5$ or $\sigma \simeq 3.5$ when $T = 1$. A Monte Carlo simulation of $h^T_{x,y}(\lambda)$ is plotted together with the other estimates for verification, including the small time estimate (3.10).

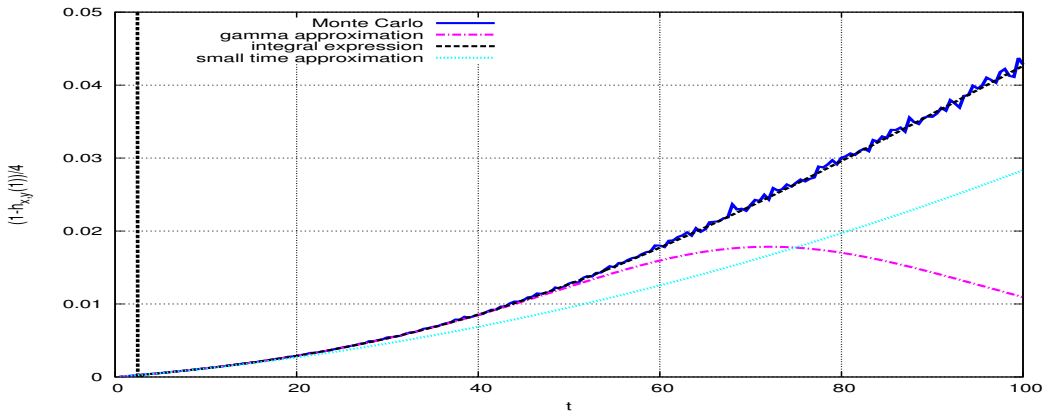

Figure 7: The integral expression (3.5) *vs* the gamma approximation (4.9) with $\sigma = 0.5$.

In Figure 8, which is plotted in large time and also includes the small time estimate (3.10), we note that the integral expression (3.5) yields the expected convergence of $(1-h^T_{x,y}(\lambda))/4$ to the limiting probability $1/4$ (when $n=4$) as T tends to infinity.

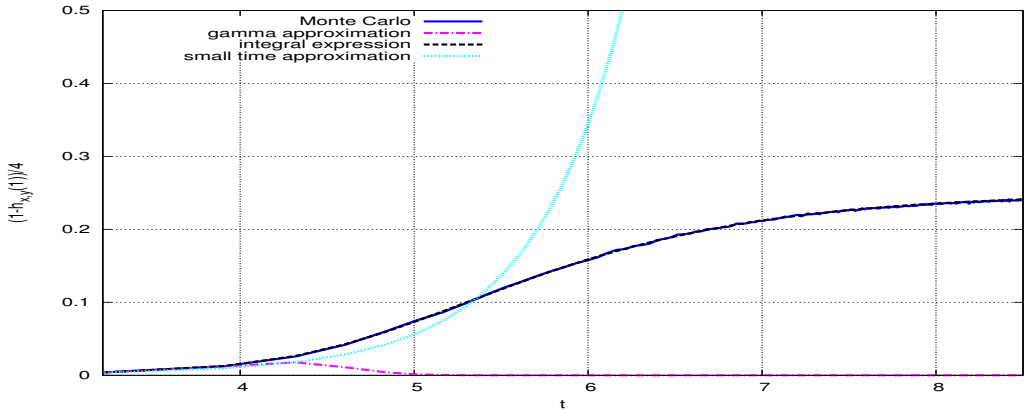

Figure 8: The integral expression (3.5) *vs* the gamma approximation (4.9) in logarithmic time scale with $\sigma = 0.5$.

5 Proofs

Proof of Lemma 3.1. By [25], Proposition 2, the joint probability density distribution of

$$\left(\int_0^T R_0 e^{\sigma W_s - p\sigma^2 s/2} ds,\ W_T - p\sigma T/2 \right)$$

can be written as

$$\begin{aligned} &\mathbb{P}\left(\int_0^T e^{\sigma W_s - p\sigma^2 s/2} ds \in du, W_T - p\sigma T/2 \in dy \right) \\ &\qquad = \frac{\sigma}{2} e^{-p\sigma y/2 - p^2\sigma^2 T/8} \exp\left(-2\frac{1+e^{\sigma y}}{\sigma^2 u} \right) \theta\left(\frac{4e^{\sigma y/2}}{\sigma^2 u}, \frac{\sigma^2 T}{4} \right) \frac{du}{u} dy \\ &\qquad = e^{-p\sigma y/2 - p^2\sigma^2 T/8} \mathbb{P}\left(\int_0^T e^{\sigma W_s} ds \in du, W_T \in dy \right), \end{aligned}$$

$y \in \mathbb{R}$, $u, T > 0$, cf. also [14] for a review of related results, and [1] for extensions to certain diffusion processes. The conclusion follows from the relation

$$\mathbb{P}\left(\int_0^T e^{\sigma W_s - p\sigma^2 s/2} ds \in du,\ e^{\sigma W_T - p\sigma^2 T/2} \in dy \right)$$

$$= \frac{1}{2} e^{-p^2\sigma^2 T/8} y^{-1-p/2} \exp\left(-2\frac{1+y}{\sigma^2 u}\right) \theta\left(\frac{4\sqrt{y}}{\sigma^2 u}, \frac{\sigma^2 T}{4}\right) \frac{du}{u} dy,$$

$u > 0$, $T > 0$, combined with the lognormal distribution

$$d\mathbb{P}(e^{\sigma W_T - p\sigma^2 T/2} = y) = \frac{1}{y\sqrt{2\pi\sigma^2 T}} e^{-(p\sigma^2 T/2 + \log y)^2/(2\sigma^2 T)}.$$

□

Proof of Proposition 3.2. By scaling it suffices to do the proof for $x = 1$ and $\sigma = 1$. By the Fubini theorem we have

$$\int_0^\infty e^{-u\lambda} \exp\left(-2\frac{1+y}{u}\right) \theta\left(\frac{4\sqrt{y}}{u}, \frac{T}{4}\right) \frac{du}{u} \tag{5.1}$$
$$= \frac{4e^{2\pi^2/T}\sqrt{y}}{\sqrt{\pi^3 T/2}} \int_0^\infty e^{-2\xi^2/T} \sin\left(\frac{4\pi\xi}{T}\right) \sinh(\xi) \int_0^\infty \exp\left(-\lambda u - 2\frac{1 + 2\sqrt{y}\cosh\xi + y}{u}\right) \frac{du}{u^2} d\xi,$$

since the above integrand belongs to $L^1(\mathbb{R}_+^2)$ as it is bounded by

$$(\xi, u) \mapsto e^{2(\pi^2 - \xi^2)/T} \sinh(\xi) \exp\left(-\lambda u - 2\frac{1+y}{u}\right).$$

Next we have

$$\int_0^\infty \exp\left(-\lambda u - 2\frac{1 + 2\sqrt{y}\cosh\xi + y}{u}\right) \frac{du}{u^2} = \sqrt{2\lambda} \frac{K_1\left(\sqrt{8\lambda}\sqrt{1 + 2\sqrt{y}\cosh\xi + y}\right)}{\sqrt{1 + 2\sqrt{y}\cosh\xi + y}},$$

where we used the identity (3.4) above. Hence we find

$$E\left[\exp\left(-\lambda\Lambda_T\right) \;\middle|\; R_T = y\right] = \int_0^\infty e^{-u\lambda} \mathbb{P}\left(\Lambda_T \in du \;\middle|\; R_T = y\right)$$
$$= \sqrt{\frac{\pi T}{2}} e^{(\log y)^2/(2T)} \int_0^\infty e^{-u\lambda} \exp\left(-2\frac{1+y}{u}\right) \theta\left(\frac{4\sqrt{y}}{u}, T/4\right) \frac{du}{u} dy$$
$$= \frac{4\sqrt{2\lambda y}}{\pi} e^{(\log y)^2/(2T) + 2\pi^2/T} \int_0^\infty e^{-2\xi^2/T} \sin\left(\frac{4\pi\xi}{T}\right) \sinh(\xi) \frac{K_1\left(\sqrt{8\lambda}\sqrt{1 + 2\sqrt{y}\cosh\xi + y}\right)}{\sqrt{1 + 2\sqrt{y}\cosh\xi + y}} d\xi.$$

When $\sigma \neq 1$ we conclude the proof by the scaling argument

$$E\left[\exp\left(-\lambda\Lambda_T\right) \;\middle|\; R_0 = x,\ R_T = y\right] = E\left[\exp\left(-\frac{\lambda}{\sigma^2}\int_0^{\sigma^2 T} e^{W_t - pt/2} dt\right) \;\middle|\; R_0 = x,\ R_T = y\right],$$

$y > 0$, that follows from (2.3). □

Next we prove Proposition 3.3 which contains the closed form expressions of $E[\Lambda_T \mid R_T = z]$ and $\mathrm{Var}\,[\Lambda_T \mid R_T = z]$.

Proof of Proposition 3.3. By scaling it suffices to do the proof for $\sigma = 1$. Under conditioning we write

$$R_t = e^{\sigma(W_t - tW_T/T) + t(\log z)/T}, \qquad t \in [0,T],$$

hence

$$\begin{aligned}
E[\Lambda_T \mid R_T = z] &= \int_0^T E[R_t \mid R_T = z]dt \\
&= \int_0^T e^{t(\log z)/T + t(T-t)/(2T)}dt \\
&= \int_0^T e^{t((\log z)/T + 1/2) - t^2/(2T)}dt \\
&= e^{(T/2+\log z)^2/(2T)} \int_0^T e^{-(t-T/2-\log z)^2/(2T)}dt \\
&= e^{(T/2+\log z)^2/(2T)} \int_{-T/2-\log z}^{T/2-\log z} e^{-x^2/(2T)}dx \\
&= \sqrt{T} e^{(T/2+\log z)^2/(2T)} \int_{-(\log z)/\sqrt{T}-\sqrt{T}/2}^{-(\log z)/\sqrt{T}+\sqrt{T}/2} e^{-y^2/2}dy \\
&= \sqrt{2\pi T} e^{(T/2+\log z)^2/(2T)} \left(\Phi\left(-\frac{\log z}{\sqrt{T}} + \frac{1}{2}\sqrt{T} \right) - \Phi\left(-\frac{\log z}{\sqrt{T}} - \frac{1}{2}\sqrt{T} \right) \right),
\end{aligned}$$

from which we conclude by (2.3). Next we have

$$\begin{aligned}
E[(\Lambda_T)^2 \mid R_T = z] &= E\left[\left(\int_0^t e^{W_t - tW_T/T + t(\log z)/T} dt \right)^2 \right] \\
&= 2\int_0^T \int_0^t e^{(s+t)(\log z)/T} E[e^{W_s - sW_T/T + W_t - tW_T/T}]dsdt \\
&= 2\int_0^T \int_0^t e^{(s+t)(\log z)/T + (t(T-t) + s(T-s) + 2s(T-t))/(2T)}dsdt, \\
&= 2e^{(3T/2+\log z)^2/(2T)} \int_0^T e^{-t} \int_0^t e^{-(3T/2-s-t+\log z)^2/(2T)}dsdt \\
&= 2e^{(3T/2+\log z)^2/(2T)} \sqrt{T} \int_0^T e^{-t} \int_{-(\log z)/\sqrt{T} - \frac{3}{2}\sqrt{T} + t/\sqrt{T}}^{-(\log z)/\sqrt{T} - \frac{3}{2}\sqrt{T} + 2t/\sqrt{T}} e^{-y^2/2}dydt \\
&= 2\sqrt{2\pi T} e^{(3T/2+\log z)^2/(2T)} \int_0^T e^{-t} \Phi\left(-\frac{\log z}{\sqrt{T}} - \frac{3}{2}\sqrt{T} + \frac{2t}{\sqrt{T}} \right) dt
\end{aligned}$$

$$-2\sqrt{2\pi T}e^{(3T/2+\log z)^2/(2T)}\int_0^T e^{-t}\Phi\left(-\frac{\log z}{\sqrt{T}}-\frac{3}{2}\sqrt{T}+\frac{t}{\sqrt{T}}\right)dt.$$

By integration by parts we have

$$\begin{aligned}
\int_0^T e^{-t}\Phi\left(-\frac{\log z}{\sqrt{T}}-\frac{3}{2}\sqrt{T}+\frac{2t}{\sqrt{T}}\right)dt &= \sqrt{\frac{2}{\pi T}}\int_0^T e^{-t-(3T/2-2t+\log z)^2/(2T)}dt \\
&\quad +\Phi\left(-\frac{\log z}{\sqrt{T}}-\frac{3}{2}\sqrt{T}\right)-e^{-T}\Phi\left(-\frac{\log z}{\sqrt{T}}+\frac{1}{2}\sqrt{T}\right) \\
&= \sqrt{\frac{2}{\pi T}}\int_0^T e^{-(2t-T-\log z)^2/(2T)-(3T/2+\log z)^2/(2T)+(T+\log z)^2/(2T)}dt \\
&\quad +\Phi\left(-\frac{\log z}{\sqrt{T}}-\frac{3}{2}\sqrt{T}\right)-e^{-T}\Phi\left(-\frac{\log z}{\sqrt{T}}+\frac{1}{2}\sqrt{T}\right) \\
&= -e^{-T}\Phi\left(-\frac{\log z}{\sqrt{T}}+\frac{1}{2}\sqrt{T}\right)+\Phi\left(-\frac{\log z}{\sqrt{T}}-\frac{3}{2}\sqrt{T}\right) \\
&\quad +\frac{e^{-5T/8}}{\sqrt{z}}\left(\Phi\left(-\frac{\log z}{\sqrt{T}}+\sqrt{T}\right)-\Phi\left(-\frac{\log z}{\sqrt{T}}-\sqrt{T}\right)\right),
\end{aligned}$$

and similarly,

$$\begin{aligned}
&\int_0^T e^{-t}\Phi\left(-\frac{\log z}{\sqrt{T}}-\frac{3}{2}\sqrt{T}+\frac{t}{\sqrt{T}}\right)dt \\
&= -e^{-T}\Phi\left(-\frac{\log z}{\sqrt{T}}-\frac{1}{2}\sqrt{T}\right)+\Phi\left(-\frac{\log z}{\sqrt{T}}-\frac{3}{2}\sqrt{T}\right)+\frac{1}{\sqrt{2\pi T}}\int_0^T e^{-t-(3T/2-t+\log z)^2/(2T)}dt \\
&= -e^{-T}\Phi\left(-\frac{\log z}{\sqrt{T}}-\frac{1}{2}\sqrt{T}\right)+\Phi\left(-\frac{\log z}{\sqrt{T}}-\frac{3}{2}\sqrt{T}\right) \\
&\quad +\frac{e^{-T}}{z}\left(\Phi\left(-\frac{\log z}{\sqrt{T}}+\frac{1}{2}\sqrt{T}\right)-\Phi\left(-\frac{\log z}{\sqrt{T}}-\frac{1}{2}\sqrt{T}\right)\right).
\end{aligned}$$

Consequently we have

$$\begin{aligned}
E[(\Lambda_T)^2] &= 2\sqrt{2\pi T}\left(\Phi\left(\frac{\log z}{\sqrt{T}}+\sqrt{T}\right)-\Phi\left(\frac{\log z}{\sqrt{T}}-\sqrt{T}\right)\right)e^{(3T/2+\log z)^2/(2T)-5T/8-(\log z)/2} \\
&\quad -2\sqrt{2\pi T}\left(\Phi\left(\frac{\log z}{\sqrt{T}}+\frac{1}{2}\sqrt{T}\right)-\Phi\left(\frac{\log z}{\sqrt{T}}-\frac{1}{2}\sqrt{T}\right)\right)\left(e^{(3T/2+\log z)^2/(2T)-T}+e^{(T/2+\log z)^2/(2T)}\right) \\
&= 2\sqrt{2\pi T}\left(\Phi\left(\frac{\log z}{\sqrt{T}}+\sqrt{T}\right)-\Phi\left(\frac{\log z}{\sqrt{T}}-\sqrt{T}\right)\right)e^{(T+\log z)^2/(2T)} \\
&\quad -2(1+z)\sqrt{2\pi T}e^{(T/2+\log z)^2/(2T)}\left(\Phi\left(\frac{\log z}{\sqrt{T}}+\frac{1}{2}\sqrt{T}\right)-\Phi\left(\frac{\log z}{\sqrt{T}}-\frac{1}{2}\sqrt{T}\right)\right),
\end{aligned}$$

which yields (3.7) by (2.3). □

We close this section with the proof of the small time equivalents on the conditional mean and variance of Λ_T used for (4.8) above. Note that the Taylor series mean and variance approximations of [5] can also serve as alternative estimates of $E[\Lambda_T \mid R_0 = x,\ R_T = y]$ and $\mathrm{Var}[\Lambda_T \mid R_0 = x,\ R_T = y]$ in small time.

Proposition 5.1 *As T tends to zero we have*

$$E[\Lambda_T \mid R_0 = x,\ R_T = y] = T\frac{y-x}{\log(y/x)} + \sigma^2 T^2 \left(\frac{x+y}{2(\log(y/x))^2} - \frac{y-x}{(\log(y/x))^3} \right) + o(T^2), \tag{5.2}$$

and $\mathrm{Var}[\Lambda_T \mid R_0 = x,\ R_T = y] = o(T^2)$.

Proof. By a scaling argument it suffices to take $x = 1$. We have

$$\begin{aligned}
a_T(z) &= \frac{\sqrt{T}}{\sigma} e^{(\sqrt{\sigma^2 T}/2 + (\log z)/\sqrt{\sigma^2 T})^2/2} \int_{(\log z)/\sqrt{\sigma^2 T} - \sqrt{\sigma^2 T}/2}^{(\log z)/\sqrt{\sigma^2 T} + \sqrt{\sigma^2 T}/2} e^{-y^2/2} dy \\
&= \frac{T}{2} e^{(\sqrt{\sigma^2 T}/2 + (\log z)/\sqrt{\sigma^2 T})^2/2} \int_{-1}^{1} e^{-(\frac{\log z}{\sqrt{\sigma^2 T}} + \frac{y}{2}\sqrt{\sigma^2 T})^2/2} dy \\
&= \frac{T}{2} \int_{-1}^{1} e^{\sigma^2 T/8 + (\log z)/2 - y^2\sigma^2 T/8 - (y \log z)/2} dy \\
&= \frac{T\sqrt{z}}{2} e^{\sigma^2 T/8} \int_{-1}^{1} e^{-y^2\sigma^2 T/8 - (y \log z)/2} dy \\
&\simeq \frac{T\sqrt{z}}{2} \left(1 + \sigma^2 T/8\right) \int_{-1}^{1} e^{-(y \log z)/2} \left(1 - y^2\sigma^2 T/8\right) dy + o(T^2) \\
&= \frac{T\sqrt{z}}{2} \left(1 + \sigma^2 T/8\right) \int_{-1}^{1} e^{-(y \log z)/2} dy - \frac{\sigma^2 T^2}{16} e^{(\log z)/2} \int_{-1}^{1} e^{-(y \log z)/2} y^2 dy + o(T^2) \\
&= T\sqrt{z} \left(1 + \sigma^2 T/8\right) \frac{e^{(\log z)/2} - e^{-(\log z)/2}}{\log z} - \frac{\sigma^2 T^2}{16} \sqrt{z} \int_{-1}^{1} e^{-(y \log z)/2} y^2 dy + o(T^2) \\
&= T\sqrt{z} \left(1 + \sigma^2 T/8\right) \frac{\sqrt{z} - 1/\sqrt{z}}{\log z} - \frac{\sigma^2 T^2}{16} \sqrt{z} \int_{-1}^{1} e^{-(y \log z)/2} y^2 dy + o(T^2) \\
&= \left(T + \sigma^2 T^2/8\right) \frac{z-1}{\log z} - \frac{\sigma^2 T^2}{16} \sqrt{z} \int_{-1}^{1} e^{-(y \log z)/2} y^2 dy + o(T^2).
\end{aligned}$$

Now we have

$$\begin{aligned}
\sqrt{z} \int_{-1}^{1} e^{-(y \log z)/2} y^2 dy &= \sqrt{z} \left(\int_{0}^{1} e^{-(y \log z)/2} y^2 dy + \int_{-1}^{0} e^{-(y \log z)/2} y^2 dy \right) \\
&= \sqrt{z} \left(\int_{0}^{1} e^{-(y \log z)/2} y^2 dy + \int_{0}^{1} e^{-(y \log z^{-1})/2} y^2 dy \right)
\end{aligned}$$

$$
\begin{aligned}
&= \sqrt{z}\left(\frac{8}{(\log z)^3}\int_0^{(\log z)/2} e^{-y}y^2dy - \frac{8}{(\log z)^3}\int_0^{-(\log z)/2} e^{-y}y^2dy\right)\\
&= \frac{8}{(\log z)^3}\sqrt{z}\left(\int_0^{(\log z)/2} e^{-y}y^2dy - \int_0^{-(\log z)/2} e^{-y}y^2dy\right)\\
&= \frac{8\sqrt{z}}{(\log z)^3}\left(2 - z^{-1/2}(((\log z)/2)^2 + \log z + 2) - (2 - z^{1/2}(((\log z)/2)^2 - \log z + 2))\right)\\
&= -\frac{8}{(\log z)^3}\left(((\log z)/2)^2 + \log z + 2 - z((\log z)/2)^2 + z\log z - 2z\right)\\
&= -\frac{8}{(\log z)^3}\left((1-z)((\log z)/2)^2 + (1+z)\log z + 2(1-z)\right),
\end{aligned}
$$

which also shows that

$$
z\int_{-1}^{1} e^{-y\log z}y^2dy = -\frac{1}{(\log z)^3}\left((1-z^2)(\log z)^2 + 2(1+z^2)\log z + 2(1-z^2)\right).
$$

Hence

$$
\begin{aligned}
a_T(z) &= \left(T + \sigma^2T^2/8\right)\frac{z-1}{\log z} - \frac{\sigma^2T^2}{16}\sqrt{z}\int_{-1}^{1} e^{-(y\log z)/2}y^2dy + o(T^2)\\
&= T\frac{z-1}{\log z} + \frac{\sigma^2T^2}{2(\log z)^3}\left((1+z)\log z + 2(1-z)\right) + o(T^2),
\end{aligned}
$$

which yields (5.2). Next we have

$$
\begin{aligned}
b_T(z) &= \frac{\sqrt{T}}{\sigma}e^{\left(\sqrt{\sigma^2T} + (\log z)/\sqrt{\sigma^2T}\right)^2/2}\int_{(\log z)/\sqrt{\sigma^2T} - \sqrt{\sigma^2T}}^{(\log z)/\sqrt{\sigma^2T} + \sqrt{\sigma^2T}} e^{-y^2/2}dy\\
&= Te^{\left(\sqrt{\sigma^2T} + (\log z)/\sqrt{\sigma^2T}\right)^2/2}\int_{-1}^{1} e^{-\left((\log z)/\sqrt{\sigma^2T} + y\sqrt{\sigma^2T}\right)^2/2}dy\\
&= zTe^{\sigma^2T/2}\int_{-1}^{1} e^{-y^2\sigma^2T/2 - y\log z}dy\\
&\simeq zT\left(1 + \frac{\sigma^2T}{2}\right)\int_{-1}^{1}(1 - y^2\sigma^2T/2)e^{-y\log z}dy + o(T^2)\\
&= zT\left(1 + \frac{\sigma^2T}{2}\right)\int_{-1}^{1} e^{-y\log z}dy - \frac{\sigma^2T^2z}{2}\int_{-1}^{1} y^2e^{-y\log z}dy + o(T^2)\\
&= T\left(1 + \frac{\sigma^2T}{2}\right)\frac{z^2-1}{\log z} - \frac{\sigma^2T^2z}{2}\int_{-1}^{1} y^2e^{-y\log z}dy + o(T^2)\\
&= T\left(1 + \frac{\sigma^2T}{2}\right)\frac{z^2-1}{\log z} + \frac{\sigma^2T^2}{2(\log z)^3}\left((1-z^2)(\log z)^2 + 2(1+z^2)\log z + 2(1-z^2)\right) + o(T^2)\\
&= T\frac{z^2-1}{\log z} + \frac{\sigma^2T^2}{(\log z)^3}\left((1+z^2)\log z + 1 - z^2\right) + o(T^2),
\end{aligned}
$$

hence

$$E[(\Lambda_T)^2 \mid R_0 = x,\ R_T = y] = \frac{2x}{\sigma^2}(xb_T(y/x) - (x+y)a_T(y/x)) \simeq T^2\frac{(y-x)^2}{(\log(y/x))^2} + o(T^2),$$

$x, y > 0$. □

Note that (5.2) above can be recovered intuitively at the first order as

$$E[\Lambda_T \mid R_0 = x,\ R_T = y] = x\int_0^T e^{(t/T)\log(y/x)}dt + o(T) = T\frac{y-x}{\log z} + o(T).$$

Conclusion

In this article we give closed form integral expressions for the probability distribution of the pathwise average of evolutionary rates, assuming a bridged geometric Brownian model with diffusion parameter σ governing the time-dependent fluctuation of the rate itself. The average substitution rate along a branch of a phylogeny whose length corresponds to T calendar time units is then obtained by integrating over the Brownian trajectories conditional on the rates at times 0 and T. Using these expressions we have assessed the validity of the gamma approximation of the average rate distribution using a distribution fit based on closed form expressions of the first two moments.

Due to the oscillating behaviour of the function (3.2) involved in the expressions of the exact probability density (3.3) and its Laplace transform (3.5), we observe numerical precision issues when rate fluctuations are mild and/or the time intervals considered are short ($\sigma^2 T < 0.5$ for both the conditional density (3.3) and the Laplace transform (3.5)). In this situation, our results indicate that the distribution of the average rate is well approximated by a gamma distribution, which is not hampered by numerical precision issues and therefore provides a relevant alternative to the exact approach.

When $\sigma^2 T > 4$, corresponding to stronger rate fluctuations and/or longer periods of time, the gamma approximation is no longer accurate, as can be seen in Figure 6. In such situations, the approximating distribution of the average rate along a branch is gamma with a small shape parameter, i.e. with a large proportion of small average

rates and few very large values, which is far from the actual density function. As a result the limiting distribution of the transition probabilities between character states (i.e., nucleotide, amino-acids or codons) in large time does not match the expected limiting distribution, which is clearly not satisfactory, as illustrated in Figures 7 and 8.

Our experience with the analysis of real data sets suggests that $\sigma^2 T$ is generally smaller than 4. The gamma approximation is therefore likely to perform well in practice. To the best of our knowledge, the software PhyTime, part of the PhyML package (see http://code.google.com/p/phyml) is the only one implementing the gamma approximation studied here. It is relatively straightforward to verify that the condition $\sigma^2 T < 4$ is met from the output generated by this computer program. Future release of this software will also implement the exact Laplace transform derived in Equation (3.5) as an alternative to the gamma approximation for large values of $\sigma^2 T$.

Acknowledgements

We thank two anynomous referees for useful comments.

References

[1] C. Albanese and S. Lawi. Laplace transforms for integrals of Markov processes. *Markov Process. Related Fields*, 11(4):677–724, 2005.

[2] S. Aris-Brosou and Z. Yang. Effects of models of rate evolution on estimation of divergence dates with special reference to the metazoan 18S ribosomal RNA phylogeny. *Syst. Biol.*, 51:703–714, 2002.

[3] J. Felsenstein. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.*, 17:368–376, 1981.

[4] T. Gernhard. The conditioned reconstructed process. *J. Theor. Biol.*, 253(4):769–778, 2008.

[5] S. Guindon. From trajectories to averages: an improved description of the heterogeneity of substitution rates along lineages. *Syst. Biol.*, 62(1):22–34, 2013.

[6] K. Ishiyama. Methods for evaluating density functions of exponential functionals represented as integrals of geometric Brownian motion. *Methodol. Comput. Appl. Probab.*, 7(3):271–283, 2005.

[7] S. Karlin and H.M Taylor. *A Second Course in Stochastic Processes*. Academic Press Inc., New York, 1981.

[8] J. F. C. Kingman. The coalescent. *Stochastic Process. Appl.*, 13:235–248, 1982.

[9] H. Kishino and M. Hasegawa. Converting distance to time: application to human evolution. *Meth. Enzymol.*, 183:550–570, 1989.

[10] H. Kishino, J.L. Thorne, and W.J. Bruno. Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Mol. Biol. Evol.*, 18(3):352–361, 2001.

[11] S. Kumar. Molecular clocks: four decades of evolution. *Nat. Rev. Genet.*, 6(8):654–662, 2005.

[12] T. Lepage, D. Bryant, H. Philippe, and N. Lartillot. A general comparison of relaxed molecular clock models. *Mol. Biol. Evol.*, 24:2669–2680, 2007.

[13] T. Lepage, S. Lawi, P. Tupper, and D. Bryant. Continuous and tractable models for the variation of evolutionary rates. *Math. Biosci.*, 199:216–233, 2006.

[14] H. Matsumoto and M. Yor. Exponential functionals of Brownian motion. I. Probability laws at fixed time. *Probab. Surv.*, 2:312–347 (electronic), 2005.

[15] S. Nee, R. May, and P. Harvey. The reconstructed evolutionary process. *Phil. Trans. R. Soc. B*, 344(1309):305–311, 1994.

[16] T. Ohta and M. Kimura. On the constancy of the evolutionary rate of cistrons. *J. Mol. Evol.*, 1(1):18–25, 1971.

[17] N. Privault and W.T. Uy. Monte Carlo computation of the Laplace transform of exponential Brownian functionals. *Methodol. Comput. Appl. Probab.*, 15(3):511–524, 2013.

[18] B. Rannala and Z. Yang. Inferring speciation times under an episodic molecular clock. *Syst. Biol.*, 56:453–466, Jun 2007.

[19] M. Sanderson. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol. Biol. Evol.*, 14:1218–1231, 1997.

[20] V. Sarich and A. Wilson. Immunological time scale for hominid evolution. *Science*, 158:1200–1203, 1967.

[21] E. A. Thompson. *Human evolutionary trees*. CUP Archive, 1975.

[22] J. Thorne, H. Kishino, and I. Painter. Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.*, 15:1647–1657, 1998.

[23] C. Tuffley and M. Steel. Modeling the covarion hypothesis of nucleotide substitution. *Math. Biosci.*, 147(1):63–91, 1998.

[24] G. N. Watson. *A treatise on the theory of Bessel functions*. Cambridge University Press, Cambridge, 1995. Reprint of the second (1944) edition.

[25] M. Yor. On some exponential functionals of Brownian motion. *Adv. in Appl. Probab.*, 24(3):509–531, 1992.

[26] E. Zuckerkandl and L. Pauling. Molecular disease, evolution, and genic heterogeneity. In M. Kasha and B. Pullman, editors, *Horizons in Biochemistry*, pages 189–225. Elsevier, Amsterdam, 1962.

6 Addendum - error bounds in the Wasserstein distance

In this section we present some bounds for the error generated by the gamma approximation, based on the Malliavin calculus and the Stein method. Although no numerical estimates are deduced, this provides a link with recent research on Wasserstein type distance estimates between probability distributions based on the Malliavin calculus, cf. [NP09], [PT13] and references therein. Recall that letting

$$I_1(f) = \int_0^T f(t)dW_t$$

denote the first order integral of $f \in L^2([0,T])$ with respect to Brownian motion, the Malliavin gradient is the operator D_t defined as

$$D_t F = \sum_{k=1}^n f_k(t) \frac{\partial g}{\partial x_k}(I_1(f_1), \ldots, I_1(f_n)), \qquad t \in [0,T],$$

where the random variable F has the form $F = g(I_1(f_1), \ldots, I_1(f_n))$, the function g is in the space $\mathcal{C}^1([0,T]^n)$ of continuously differentiable functions on $[0,T]^n$, and $f_1, \ldots, f_n \in L^2([0,T])$, $n \geq 1$. We denote by $(\mathcal{F}_t)_{t\in[0,T]}$ the filtration generated by the Brownian motion $(W_t)_{t\in[0,T]}$ built on the Wiener space W as the coordinate process $W_t(\omega) = \omega(t)$, $\omega \in W$, cf. Chapter 1 of [Üst95]. Recall also that the inverse $(-L)^{-1}$ of the Ornstein-Uhlenbeck operator $-L$ can be defined as

$$(-L)^{-1}(F(\omega) - E[F]) = \int_0^1 E[F(a\omega + \sqrt{1-a^2}\tilde{\omega}) \mid \omega]da$$

for $F \in L^2(\Omega)$ with $E[F] = 0$, where $\tilde{\omega}$ denotes an independent copy of the ω on the Wiener space, cf. § 1.2 of [Üst95], [NP09], and § 5.3 of [Pri09] for details.

We consider the Wasserstein type distance

$$d(X,Y) := \sup_{h\in\mathcal{H}} |\mathrm{E}[h(X)] - \mathrm{E}[h(Y)]|, \tag{6.1}$$

where

$$\mathcal{H} := \{h \in \mathcal{C}_b^2(\mathbb{R}) : \ \max\{\|h\|_\infty, \|h'\|_\infty, \|h''\|_\infty\} \leq 1\}.$$

From Theorem 3.1 of [NP09] the distance between the law of Λ_T and the gamma-$(\beta(z), \eta(z))$ distribution $\Gamma_{\beta,\eta}$ can be bounded as

$$d(\Lambda_T, \Gamma_z) \leq K\sqrt{E[(2\eta(z)\Lambda_T - \langle D\Lambda_T, D(-L)^{-1}(\Lambda_T - E[\Lambda_T])\rangle)^2]},$$

for $K > 0$ a constant, while by Corollary 3.4 of [PT13] we have

$$d(\Lambda_T, \Gamma_z) \leq K\sqrt{E[(2\eta(z)\Lambda_T - \langle D.\Lambda_T, E[D.\Lambda_T \mid \mathcal{F}_.]\rangle)^2]}. \tag{6.2}$$

Next, in Propositions 6.1 and 6.2 below we show how the quantities

$$\langle D\Lambda_T, -DL^{-1}(\Lambda_T - E[\Lambda_T])\rangle = -\int_0^T D_t\Lambda_T D_t L^{-1}(\Lambda_T - E[\Lambda_T])dt$$

and

$$\langle D.\Lambda_T, E[D.\Lambda_T \mid \mathcal{F}_.]\rangle = \int_0^T D_t\Lambda_T E[D_t\Lambda_T \mid \mathcal{F}_t]dt$$

appearing in (6.2) can be computed.

Proposition 6.1 *For all $T > 0$ we have*

$$\langle D.\Lambda_T, E[D.\Lambda_T \mid \mathcal{F}_.]\rangle = \int_0^T \int_0^T e^{\sigma W_s - p\sigma^2(s+t)/2} \int_0^{s\wedge t} e^{\sigma W_u + \sigma^2(T-u)/2} dudsdt. \tag{6.3}$$

Proof. We have

$$D_t e^{I_1(f)} = f(t)e^{I_1(f)},$$

and

$$\begin{aligned} E[D_t e^{I_1(g)} \mid \mathcal{F}_t] &= g(t)E[e^{I_1(g)} \mid \mathcal{F}_t] \\ &= g(t)e^{\int_0^T g^2(s)ds/2} E[e^{I_1(g) - \int_0^T g^2(s)ds/2} \mid \mathcal{F}_t] \\ &= g(t)e^{\int_0^T g^2(s)ds/2} e^{\int_0^t g(s)dW_s - \int_0^t g^2(s)ds/2} \\ &= g(t)e^{\int_0^t g(s)dW_s + \int_t^T g^2(s)ds/2}. \end{aligned}$$

This yields

$$\langle D.e^{I_1(f)}, E[D.e^{I_1(g)} \mid \mathcal{F}_.]\rangle = e^{I_1(f)} \int_0^T f(u)g(u)e^{\int_0^u g(s)dW_s + \int_u^T g^2(s)ds/2} du, \tag{6.4}$$

which yields (6.3) by taking $f = \sigma\mathbf{1}_{[0,s]}$ and $g = \sigma\mathbf{1}_{[0,u]}$. □

Under conditioning we write

$$R_t = x(y/x)^{t/T} e^{\sigma U_t} = x(y/x)^{t/T} e^{\sigma(W_t - tW_T/T)},$$

where

$$U_t := W_t - \frac{t}{T} W_T, \qquad t \in [0,T],$$

is a standard Brownian bridge with $U_0 = U_T = 0$. For $a \leq b$ we have

$$\mathrm{Cov}(U_a, U_b) = \frac{a\,(t-b)}{t}, \tag{6.5}$$

which yields

$$E[V_s] = v_0 + \frac{(v_t - v_0)}{t} s,$$

and

$$\mathrm{Var}\,[V_s] = \frac{\sigma^2 s\,(t-s)}{t},$$

where $V_s := \log R_s$. Taking

$$f = \sigma\left((1-s/T)\mathbf{1}_{[0,s]} - (s/T)\mathbf{1}_{[s,T]}\right) \quad \text{and} \quad g = \sigma\left((1-t/T)\mathbf{1}_{[0,t]} - (t/T)\mathbf{1}_{[t,T]}\right)$$

in (6.4) yields

$$\begin{aligned}
&\langle D_\cdot e^{\sigma(W_s - (s/T)W_T)}, E[D_\cdot e^{\sigma(W_t - tW_T/T)} \mid \mathcal{F}_\cdot]\rangle \\
&= e^{\sigma(W_s-(s/T)W_T)} \int_0^T f(u)g(u) e^{\int_0^u g(s)dW_s + \int_u^T g^2(s)ds/2} du \\
&= e^{\sigma(W_s-(s/T)W_T)} \int_0^T \left(\left(1-\frac{s}{T}\right)\mathbf{1}_{[0,s]}(u) - (s/T)\mathbf{1}_{[s,T]}(u)\right) \\
&\quad \times \left(\left(1-\frac{t}{T}\right)\mathbf{1}_{[0,t]}(u) - \frac{t}{T}\mathbf{1}_{[t,T]}(u)\right) e^{\int_0^u g(s)dW_s + \int_u^T g^2(s)ds/2} du \\
&= e^{\sigma(W_s-(s/T)W_T)}(1-s/T)(1-t/T) \int_0^{s\wedge t} e^{(1-t/T)W_u + (1-t/T)^2(t-u)/2 + (T-t)(t/T)^2/2} du \\
&\quad + e^{\sigma(W_s-(s/T)W_T)}(s/T)(t/T) \int_{s\vee t}^T e^{(t/T)(W_T - W_u) + (t/T)^2(T-u)/2} du \\
&\quad - \mathbf{1}_{\{s<t\}} e^{\sigma(W_s-(s/T)W_T)}(1-s/T)(t/T) \int_s^t e^{(1-t/T)W_u + (1-t/T)^2(t-u)/2} du \\
&\quad - \mathbf{1}_{\{s>t\}} e^{\sigma(W_s-(s/T)W_T)}(1-t/T)(s/T) \int_t^s e^{(t/T)W_u + (t/T)^2(T-u)/2} du.
\end{aligned}$$

As for (6.2) we have the following result in the unconditional case.

Proposition 6.2 *For all $T > 0$ we have*

$$\langle D.\Lambda_T, E[D.\Lambda_T \mid \mathcal{F}_.]\rangle = \int_0^T \int_0^T \frac{s \wedge t}{t} e^{-p(s+t)\sigma^2/2} e^{\sigma W_s + \sigma^2 t/2}$$
$$\times \left(1 - e^{\sigma W_t - \sigma^2 t/2} + \sqrt{\frac{2\pi}{t}} W_t e^{W_t^2/(2t)} \left(\Phi\left(W_t/\sqrt{t}\right) - \Phi\left(W_t/\sqrt{t} - \sigma\sqrt{t}\right)\right)\right) dsdt.$$

Proof. Letting $g \in L^2([0,T])$ and $\nu^2 = \int_0^T g^2(s)ds > 0$, we find

$$\begin{aligned}
(-L)^{-1}(e^{I_1(g)} - E[e^{I_1(g)}]) &= \int_0^1 E[e^{aI_1(g) + \sqrt{1-a^2} I_1'(g)} \mid I_1(g)]da \\
&= \int_0^1 e^{aI_1(g) + (1-a^2)\nu^2/2} da \\
&= e^{\nu^2/2} \int_0^1 e^{aI_1(g) - a^2\nu^2/2} da \\
&= e^{\nu^2/2} e^{(I_1(g))^2/\nu^2/2} \int_0^1 e^{-(I_1(g)/\nu - a\nu)^2/2} da \\
&= \frac{1}{\nu} e^{\nu^2/2} e^{(I_1(g))^2/\nu^2/2} \int_{-I_1(g)/\nu}^{\nu - I_1(g)/\nu} e^{-a^2/2} da \\
&= \frac{\sqrt{2\pi}}{\nu} e^{\nu^2/2} e^{(I_1(g))^2/\nu^2/2} \left(\Phi\left(I_1(g)/\nu\right) - \Phi\left(I_1(g)/\nu - \nu\right)\right),
\end{aligned}$$

and

$$\begin{aligned}
D_t(-L)^{-1}(e^{I_1(g)} - E[e^{I_1(g)}]) &= \frac{\sqrt{2\pi}}{\nu} e^{\nu^2/2} D_t \left(e^{(I_1(g))^2/\nu^2/2} \left(\Phi\left(I_1(g)/\nu\right) - \Phi\left(I_1(g)/\nu - \nu\right)\right)\right) \\
&= \frac{\sqrt{2\pi}}{\nu} e^{\nu^2/2} e^{(I_1(g))^2/\nu^2/2} D_t \left(\Phi\left(I_1(g)/\nu\right) - \Phi\left(I_1(g)/\nu - \nu\right)\right) \\
&\quad + \frac{\sqrt{2\pi}}{\nu} e^{\nu^2/2} \left(\Phi\left(I_1(g)/\nu\right) - \Phi\left(I_1(g)/\nu - \nu\right)\right) D_t e^{(I_1(g))^2/\nu^2/2} \\
&= \frac{1}{\nu^2} g(t) e^{\nu^2/2} e^{(I_1(g))^2/\nu^2/2} \left(e^{-(I_1(g)/\nu)^2/2} - e^{-(I_1(g)/\nu - \nu)^2/2}\right) \\
&\quad + \frac{\sqrt{2\pi}}{\nu^3} e^{\nu^2/2} e^{(I_1(g))^2/\nu^2/2} \left(\Phi\left(I_1(g)/\nu\right) - \Phi\left(I_1(g)/\nu - \nu\right)\right) g(t) I_1(g) \\
&= \frac{1}{\nu^2} g(t) e^{\nu^2/2} \left(1 - e^{I_1(g) - \nu^2/2}\right) \\
&\quad + \frac{\sqrt{2\pi}}{\nu^3} e^{\nu^2/2} g(t) I_1(g) e^{(I_1(g))^2/\nu^2/2} \left(\Phi\left(I_1(g)/\nu\right) - \Phi\left(I_1(g)/\nu - \nu\right)\right) \\
&= \frac{g(t)}{\nu^2} \left(e^{\nu^2/2} - e^{I_1(g)}\right) + \frac{\sqrt{2\pi}}{\nu^3} e^{\nu^2/2} g(t) I_1(g) e^{(I_1(g))^2/\nu^2/2} \left(\Phi\left(I_1(g)/\nu\right) - \Phi\left(I_1(g)/\nu - \nu\right)\right).
\end{aligned}$$

Hence, for any $f \in L^2([0,T])$ we have

$$\begin{aligned}
&\langle De^{I_1(f)}, D(-L)^{-1}(e^{I_1(g)} - E[e^{I_1(g)}])\rangle \\
&= e^{I_1(f)+\nu^2/2}\frac{\langle f,g\rangle}{\nu^2}\left(1 - e^{I_1(g)-\nu^2/2} + \sqrt{2\pi}\frac{I_1(g)}{\nu}e^{(I_1(g))^2/(2\nu^2)}\left(\Phi\left(I_1(g)/\nu\right) - \Phi\left(I_1(g)/\nu - \nu\right)\right)\right),
\end{aligned}$$

which, taking $f = \sigma\mathbf{1}_{[0,s]}$, $g = \sigma\mathbf{1}_{[0,u]}$, and $\nu^2 = \int_0^T g^2(s)ds = \sigma^2 t$, gives

$$\begin{aligned}
&\langle De^{\sigma W_s - p\sigma^2 s/2}, D(-L)^{-1}(e^{\sigma W_t - p\sigma^2 t/2} - E[e^{\sigma W_t - p\sigma^2 t/2}])\rangle \qquad (6.6)\\
&= \frac{s\wedge t}{t}e^{-p(s+t)\sigma^2/2}e^{\sigma W_s + \sigma^2 t/2} \\
&\quad \times\left(1 - e^{\sigma W_t - \sigma^2 t/2} + \sqrt{\frac{2\pi}{t}}W_t e^{W_t^2/(2t)}\left(\Phi\left(W_t/\sqrt{t}\right) - \Phi\left(W_t/\sqrt{t} - \sigma\sqrt{t}\right)\right)\right),
\end{aligned}$$

and yields (6.6) by integration in $(s,t) \in [0,T]\times[0,T]$. □

References

[NP09] I. Nourdin and G. Peccati. Stein's method on Wiener chaos. *Probab. Theory Related Fields*, 145(1-2):75–118, 2009.

[Pri09] N. Privault. *Stochastic Analysis in Discrete and Continuous Settings*, volume 1982 of *Lecture Notes in Math.* Springer-Verlag, Berlin, 309 pp., 2009.

[PT13] N. Privault and G.L. Torrisi. Probability approximation by Clark-Ocone covariance representation. *Electron. J. Probab.*, 18:1–25, 2013.

[Üst95] A. S. Üstünel. *An introduction to analysis on Wiener space*, volume 1610 of *Lecture Notes in Mathematics*. Springer Verlag, 1995.