Version dated: August 2, 2016

Beyond Brownian motion and the Ornstein-Uhlenbeck process: Stochastic diffusion models for the evolution of quantitative characters.

Simon P. Blomberg

School of Biological Sciences, University of Queensland, St. Lucia, Queensland, 4072, Australia.

E-mail: s.blomberg1@uq.edu.au

Abstract.—Gaussian processes such as Brownian motion and the Ornstein-Uhlenbeck process have been popular models for the evolution of quantitative traits and are widely used in phylogenetic comparative methods. However, they have drawbacks which limit their utility. Here we describe new, non-Gaussian stochastic differential equation (diffusion) models of quantitative trait evolution. We present general methods for deriving new diffusion models, and discuss possible schemes for fitting non-Gaussian evolutionary models to trait data. The theory of stochastic processes provides a mathematical framework for understanding the properties of current, new and future phylogenetic comparative methods. Attention to the mathematical details of models of trait evolution and diversification may help avoid some pitfalls when using stochastic processes to model macroevolution.

(Keywords: Brownian Motion, Ornstein-Uhlenbeck, Stochastic Differential Equations, Diffusions, Continuous Traits, Comparative Methods)

> "Brownian motion is a poor model, and so is Ornstein-Uhlenbeck, but just as democracy is the worst method of organizing a society "except for all the others", so these two models are all we've really got that is tractable. Critics will be admitted to the event, but only if they carry with them another tractable model." - J. Felsenstein, r-sig-phylo email list, 8th April 2008.

The parametric estimation of phylogenies depends on having an appropriate model of character evolution (Posada and Crandal 2001). Molecular systematists are spoiled for choice in this regard. For example, the program jModelTest2 can fit 1624 models of DNA sequence evolution (Darriba et al. 2012). The situation for the comparative analysis of continuous traits is quite different. Here, we have mainly two analytical models in popular use: Brownian motion (BM) and the Ornstein-Uhlenbeck (OU) process. Other models such as "early burst" are also sometimes used (e.g. Blomberg et al. 2003; Ingram et al. 2012) and there have been several extensions to the OU model (see below). There are other approaches to phylogenetic comparative analyses that do not use explicit models of evolution (in terms of being able to write down the appropriate equations). Some non-analytical models can be used to estimate sampling distributions for regression parameters using computer simulation (Garland et al. 1993), and the evolutionary model for continuous traits can also be altered by applying branch-length transformations (e.g. Grafen 1989; Pagel 1999; Freckleton et al. 2002; Blomberg et al. 2003). We do not consider these approaches to phylogenetic comparative analyses here. Instead, we focus on providing an approach to comparative analyses based on the theory of stochastic processes, which unites BM, OU and other processes in a common statistical and probabilistic framework.

Starting with Bachelier (1900), the main application of stochastic processes has been in finance where models have been developed for stock prices, derivatives, options and other financial products. In that domain, the model of Black and Scholes (1973) has been particularly successful (in terms of citations), but research into the theory of stochastic processes is still thriving across a wide range of disciplines, especially the physical sciences (e.g. Uhlenbeck and Ornstein 1930; Einstein 1956; Freund and Pöschel 2000; Gardiner 2009). Although diffusion models are common in epidemiology and other life sciences (Fuchs 2013), applications in evolutionary biology are rare. The Wright-Fisher model and the Moran model in population genetics are well-known exceptions (Fisher 1922; Wright 1931; Feller 1951; Moran 1958; Ewens 2004). Population geneticists have used these stochastic processes to model microevolution. Here we examine the possible uses of stochastic processes in studies of macroevolution, i.e. evolution above the species level (Simpson 1953; Rensch 1959; Stanley 1975; Benton 2015; Serrelli and Gontier 2015), with the aim to provide new models and methods for the phylogenetic comparative analysis of non-Gaussian traits. Such models are necessary because current evolutionary models for quantitative traits can have poor performance (Pennell et al. 2015).

Mathematical Background

In order to fully understand the mathematics of stochastic processes, some background is required. At least, some knowledge of Riemann-Stieltjes integrals, as well as some understanding of measure-theoretic probability

theory is necessary. Introductory books such as Øksendal (2007) or Klebaner (2012) can be helpful. Gardiner (2009) provides an excellent practical approach which largely ignores the measure-theoretic foundations, but concentrates mainly on applications in the physical sciences.

Consider a sample space Ω consisting of discrete, elementary events ω, occurring over time such that:

$$\Omega = \{\omega : \omega = (x_1, x_2, x_3, \ldots, x_T)\}$$

where $x_1, \ldots, x_T$ are values of a (univariate) trait x at time $= 1, \ldots, T$. Our trait may come from a good palaeontological sequence, for example. Our knowledge about the evolution of the trait increases as we observe more and more values of x as T increases. However, in general we only observe one possible x at each time step. There could have been many other outcomes for the trait at each time, so we only have information on the observed values A, a subset of Ω. We therefore know that the "true" state of the trait over all time must be in A and not in that part of Ω that is not A. In set notation, $\Omega \backslash A = \bar{A}$. Now define $\mathcal{F}_t$ as the information available about the trait (ie the trait values) at all times up to time t, ie A. At $t = 0$ we have no information about the trait so $\mathcal{F}_0 = \{\emptyset, \Omega\}$ where $\emptyset$ is the empty set. At $t = 1$ we know x_1, at $t = 2$ we know x_1 and x_2, etc. We do not forget the previously gained information. Therefore each $\mathcal{F}_t$ contains the trait value at time t plus all the $\mathcal{F}_t$ that have gone before. ie. $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \ldots \subset \mathcal{F}_T$. Each $\mathcal{F}_t$ is called a *field* or *algebra*. The collection of all fields is called a *filtration*. ie.

$$\mathbb{F} = \{\mathcal{F}_0, \mathcal{F}_1, \ldots, \mathcal{F}_t, \ldots, \mathcal{F}_T\} \qquad \mathcal{F}_t \subset \mathcal{F}_{t+1}$$

The above explanation of fields is limited to the discrete time case. We would also like to model probability as a continuous function of time. The definition of fields in the continuous time case is more tiresome but can be found in any textbook on measure-theoretic probability (e.g. Pollard 2002). The most important field in measure-theoretic probability theory is the Borel σ-field ($\mathfrak{B}$). A probability (Lebesgue) measure, $\mathbb{P}$ can be defined for a process that generates a σ-field. Hence, a probability space is defined as the triple $(\Omega, \mathcal{F}, \mathbb{P})$. A *random variable* X on $\{\Omega, \mathcal{F}\}$ is a measurable function from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathfrak{B})$, where $\mathfrak{B}$ is the Borel σ-field on the real number line. A *stochastic process* is a collection of random variables $\{X(t)\}$. A stochastic process is said to be *adapted* to a filtration $\mathbb{F}$ if for all t, $X(t)$ is a random variable on $\mathcal{F}_t$. We will only be considering stochastic processes that are adapted to a filtration. This is not really a limitation, as a major property of adapted processes is that they are unable to anticipate the future, which does not appear to be a very limiting assumption for a macroevolutionary model.

Brownian Motion

Brownian motion (BM) is named for the movement of pollen grains suspended in water, as first observed by the botanist Robert Brown in 1837, but it is observed in many other multi-particle settings. The mathematics of BM were first analysed by Bachelier (1900), who anticipated almost all the mathematical results of Einstein's work in 1905 in the context of molecular movement (see Einstein 1956). Wiener (1923) was the first to rigorously

characterise BM as a stochastic process, and hence BM is sometimes also known as the Wiener process. BM was first proposed as a model of character evolution for phylogeny estimation by Felsenstein (1973), who also introduced this model into phylogenetic comparative regression analyses (Felsenstein 1985).

Let $B(t)$ be the trait value of a BM process at time t. BM has the following defining properties (e.g. Klebaner 2012). BM has independent increments. $B(t) - B(s)$ for $t > s$ is independent of the past $B(u)$ where $0 \leq u \leq s$. The increments are also Gaussian. $B(t) - B(s)$ has a Normal (Gaussian) distribution with a mean $\mu = 0$ and variance $\sigma^2 = t - s$. This means we can use all the powerful mathematical machinery appropriate to Gaussian distributions.

Further, the sample paths of a BM process $B(t)$ have the following properties, for almost every sample path (i.e. other than those of Lebesgue measure zero): $B(t)$ is a continuous function of t. Hence, BM can be used to model continuous traits in continuous time. $B(t)$ is not monotone in any time interval, no matter how small the interval. BM paths are jagged at all time scales. Despite being continuous, $B(t)$ is nowhere differentiable. This property makes it difficult to estimate rates of evolution from sample paths, although σ^2 is usually associated with the rate of evolution. The quadratic variation of $B(t) = t$. That is, the variance of $B(t)$ increases linearly with t. There doesn't seem to be any biological reason why the variance of a trait should increase linearly with time. Further, this property implies that there are no bounds to evolution and that traits have no physical limits. This is unlikely to be true for any trait (e.g. McGhee 2015) but see Conway Morris et al. (2015); Vermeij (2015, *ibid.*).

BM is useful as a simple model of trait evolution. Its simplicity is due to the above properties, as well as to the fact that it has the Markov property. Further, BM is a martingale, which means that the expectation of the process at time $t + s$ is the value of the process at time t. That is, $E(B(t + s)|\mathcal{F}_t) = B(t)$. The Markov and martingale properties simplify the mathematics of working with BM processes. BM lends itself to two evolutionary interpretations. Either it is a model implying no selection and evolution occurs just by random drift, or it can be viewed as a model of very strong selection in a randomly varying environment (see Hansen and Martins 1996). These interpretations cannot be simultaneously correct, and both are likely to be wrong for any real quantitative trait.

The Ornstein-Uhlenbeck Process

The OU process was introduced as an improved model for physical Brownian motion, which incorporates the effect of friction (Uhlenbeck and Ornstein 1930). It also has a long history in evolutionary biology. It can be derived from the consideration of stabilising selection and genetic drift (Lande 1976). Its use in phylogenetic comparative methods has been promoted by many authors (Felsenstein 1988; Hansen and Martins 1996; Hansen 1997; Martins and Hansen 1997; Butler and King 2004; Hansen et al. 2008; Beaulieu et al. 2012). It has the

following form:

$$X(t) = \mu + e^{-\alpha t}\left(X(0) - \mu + \sigma \int_0^t e^{\alpha s} dB(s)\right) \quad (1)$$

where μ is the mean of the process. Note that $X(t)$ in (1) depends on $B(s)$. That is, BM is one building block of the OU process. The biological interpretation of the OU process is controversial. Most authors have interpreted α as the strength of a restraining force corresponding to stabilizing selection, and the sample paths as trajectories of evolution of organisms' traits (e.g. Beaulieu et al. 2012; Butler and King 2004). However, Hansen et al. (2008) interpret the sample paths as paths of an evolutionary optimum itself, subject to an overall central tendency with strength α and stochastic perturbations.

The properties of OU are well known (e.g. Insua et al. 2012; Klebaner 2012). The OU process is a Gaussian process with continuous paths. It has the Markov property and it is stationary, provided the initial distribution is the stationary distribution $N(\mu, \frac{\sigma^2}{2\alpha})$. It is the *only* stochastic process which has all three properties (Gaussian, Markov, stationarity) (Breiman 1968; Klebaner 2012). OU is *not* a martingale. The Gaussian property of both BM and OU makes them relatively simple to work with, for example, (Hansen 1997; Butler and King 2004) used likelihood methods to fit models with different α values on different branches of the phylogeny. Beaulieu et al. (2012) extend this idea by allowing σ to vary with time.

Note that the stochastic integral in (1) is with respect to "white noise", implying that $B(t)$ is differentiable, whereas one of the properties of BM is that it is *not* differentiable. The meaning of such integrals is therefore not straight forward. In fact, it requires a new definition for integration. The definition adopted here is that of Itô (1944, 1946). There are other approaches to stochastic integration, most notably the Stratonovich integral (e.g. Gardiner 2009). Turelli (1977) has discussed situations in which one definition may be preferred over the other. In practice, the Itô integral is the most widely used:

$$\int_{t_0}^{t} f(s) dB_s = \underset{n\to\infty}{\text{ms-lim}} \left\{ \sum_{i=1}^{n} f(t_{i-1})[B(t_i) - B(t_{i-1})] \right\} \quad (2)$$

where ms-lim means the mean square limit (Gardiner 2009, p. 41).

The main drawback of both BM and OU that we wish to highlight is the Gaussian nature of both stochastic processes. While analytically and computationally useful, this assumption limits the application of the models to traits that are Normally-distributed. Of course, one could transform the response variable so that it is then approximately Gaussian, such as using the $logit(x)$probit(x), or $Sin^{-1}\sqrt{x}$ transformations for proportions, or the $\log(x)$ transformation for counts, and then use Gaussian process models (Ives 2015; Warton et al. 2016). However, shoe-horning data using transformations can make interpretation of model outputs more difficult. Instead, we suggest that the direct modelling of non-Gaussian evolutionary processes provides a much more elegant view of the evolutionary process. There is a strong analogy with the development of Generalized Linear

Models, which greatly extended the analysis of non-Gaussian linear models (McCullagh and Nelder 1989). Here we outline a generalized method of constructing new stochastic process models for continuous trait evolution.

Diffusions as Models of Trait Evolution

Consider the stochastic differential equation (SDE):

$$dX_t = b(X_t, t)dt + \sigma(X_t, t)dB_t \tag{3}$$

Such SDEs are termed "diffusion" equations and arise as solutions to the Fokker-Planck equation (Gardiner 2009). The left-hand side of the equation represents a small change in trait variable X_t at time t. The right-hand side has two terms. The first term is the deterministic part of the model. $b(X_t, t)$ is termed the *drift coefficient*. The differential of the first term is dt, which denotes a differential with respect to (continuous) time. Note the difference in usage of the term compared to its use in population genetics, where drift implies a stochastic process. We will retain the traditional mathematical terminology. The second term is stochastic, as the differential is $dB(t)$, "white noise". $\sigma(X_t, t)$ is termed the *diffusion coefficient*. In financial statistics, $\sigma(X_t, t)$ is termed the "volatility"(Mikosch 1998). Note that both b and σ can depend on both X_t and t in some arbitrary way. It is important that the only meaning of (3) is with respect to the Itô definition of the integral (2). Stochastic processes of this type are termed "Itô diffusions."

The Ornstein-Uhlenbeck diffusion process can be defined by the following SDE:

$$dX_t = \alpha(\mu - X_t)dt + \sigma dB_t \tag{4}$$

for α, μ and σ as real, positive constants. Here, α represents the restraining force of stabilising selection. μ represents the mean trait value (at stationarity). The drift coefficient here is a linear function of X_t. The form of the drift is significant, as it is this expression that controls the forcing of the trait X_t back towards μ. OU is thus said to be "mean-reverting": X_t tends to return to μ over time. However, the property of mean reversion is not limited to the OU process. Any simple diffusion model with a drift coefficient of the form in (4) will exhibit mean reversion. It is also clear that (4) is time-homogeneous since neither b nor σ depend on t. The process is also ergodic. That is, given enought time, the time average for any particular species' trait is equal to the average trait value across species (Lebowitz and Penrose 1973). These properties suggest that a stationary distribution exists for this process. Figure 1 shows a sample evolution along a 5-species tree for the OU model.

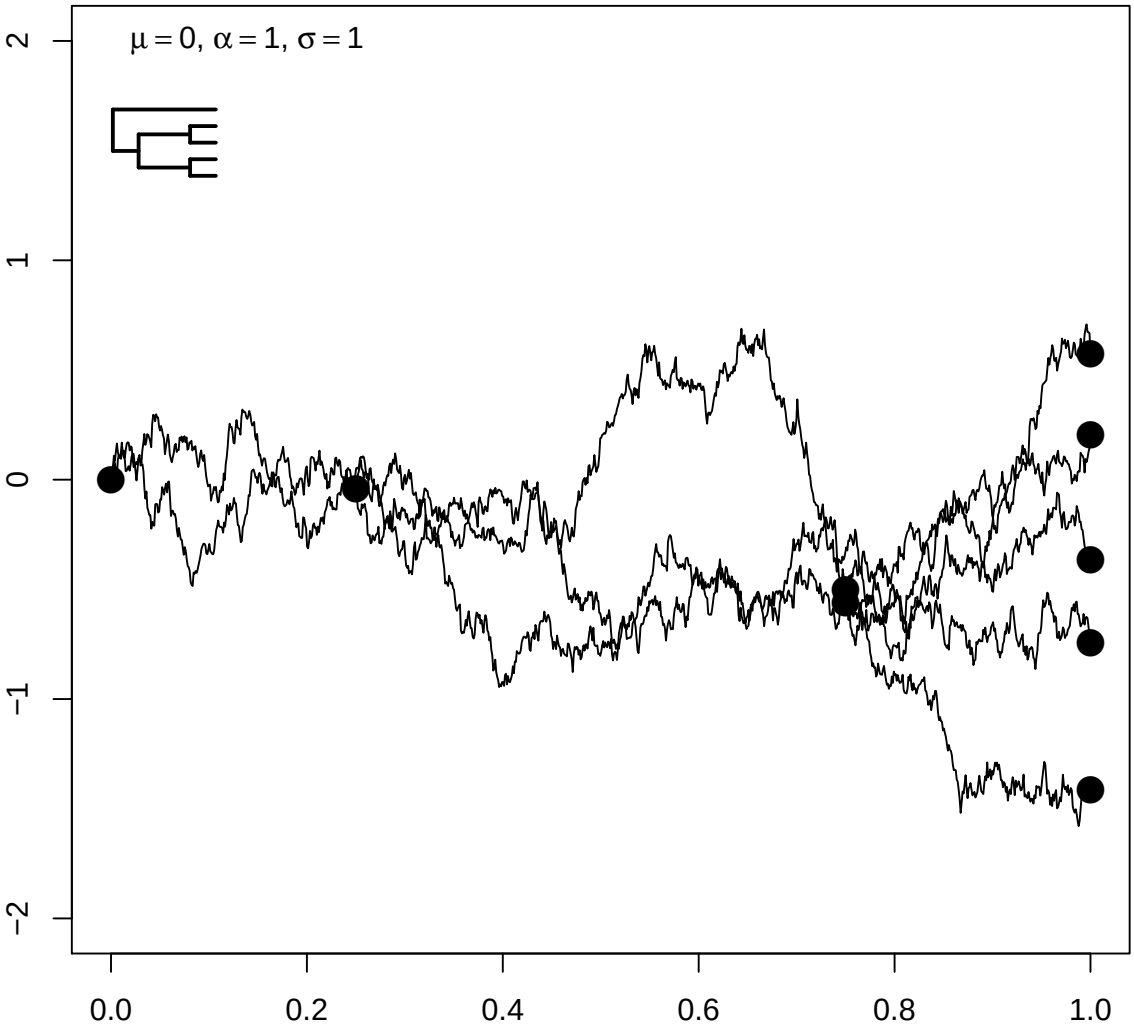

Figure 1: Ornstein-Uhlenbeck evolution along a 5-species tree. μ is the mean of the process, α is the strength of the restraining force, and σ is the diffusion coefficient. Large dots are nodes and tips.

New Evolutionary Models

The key to the construction of new models for evolution is the solution of the Fokker-Planck (Kolmogorov Forward) equation (Risken 1996). In one dimension it takes the form:

$$\frac{\partial f(x,t)}{\partial t} = -\frac{\partial}{\partial x}[b(x,t)f(x,t)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[\sigma^2(x,t)f(x,t)] \tag{5}$$

(5) governs the time evolution of the underlying probability law $f(x,t)$. It is a partial differential equation in x and t. Note that it is *not* stochastic. If the stochastic process is time-homogeneous, (5) can be written as:

$$\frac{\mathrm{d}}{\mathrm{d}x}[b(x)f(x)] - \frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}x^2}[\sigma^2(x)f(x)] = 0 \tag{6}$$

Solving for $f(x)$ gives the following formula for the construction of the stationary distribution (Appendix 1):

$$f(x) = \frac{C}{\sigma^2(x)}\exp\left(\int_{x_0}^{x}\frac{2b(y)}{\sigma^2(y)}\mathrm{d}y\right) \tag{7}$$

where C is a constant of integration found by solving $\int f(x)\mathrm{d}x = 1$. (7) is sometimes known as Wright's equation (Wright 1938; Cobb 1998).

Consider the following diffusion equations:

$$dX_t = \alpha(\mu - X_t)dt + \sqrt{\epsilon X_t}dB_t \tag{8}$$

$$dX_t = \alpha(\mu - X_t)dt + \sqrt{\epsilon X_t(1-X_t)}dB_t \tag{9}$$

The drift terms in (8) and (9) are of the same form as in (4). Hence, these processes are both mean-reverting, and will be driven by a central tendency towards μ, with a restraining force α. The difference between these two processes and OU is in the diffusion coefficient. With a mean-reverting process, the form of the diffusion coefficient determines the distribution of the stationary distribution. The stationary distributions for each process described by (8) and (9) are derived in Appendix 2. While the notation for calculating with diffusion models is powerful and elegant, stochastic processes come alive when visualised using simulation. We provide example plots of paths mapped onto a phylogeny with five species (Figs.2 and 3). (8) has as its stationary distribution:

$$f(x|\mu,\delta) = \left(\frac{x}{\delta}\right)^{-1+\frac{\mu}{\delta}}\frac{e^{-\frac{x}{\delta}}}{\Gamma(\frac{\mu}{\delta})},\ \delta = \frac{\epsilon}{\alpha}.$$

Γ is the Gamma function. That is, $f(x|\mu,\delta)$ is a density of a Gamma distribution with mean $= \mu$, mode $= \mu - \delta$, and variance$= \delta\mu$: $x \sim \text{Gamma}(\frac{\mu}{\delta}, \frac{1}{\delta})$. In fact, (8) is the Cox, Ingersoll and Ross (CIR) model commonly used in finance (Cox et al. 1985). See Figure 2.

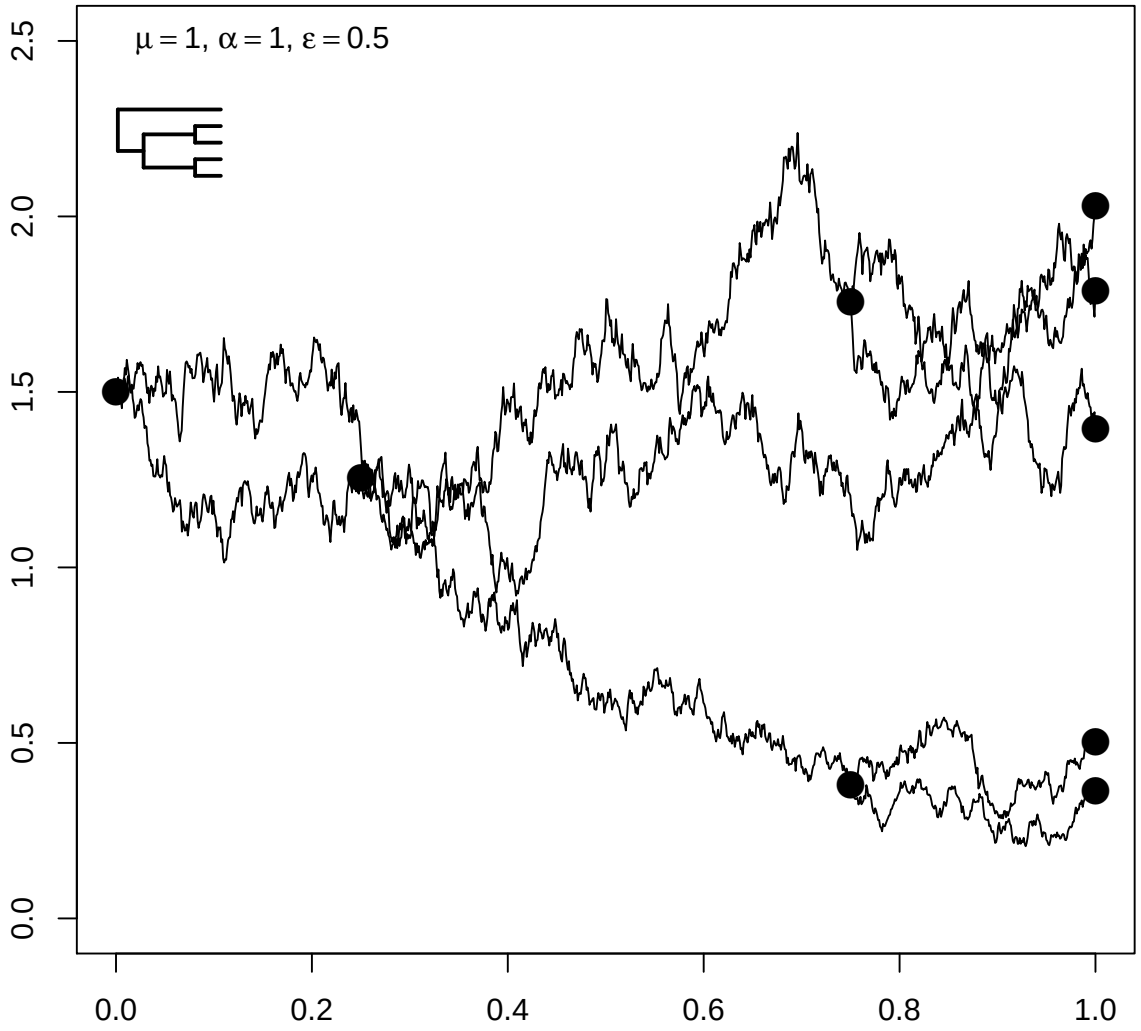

Figure 2: Cox-Ingersoll-Ross evolution along a 5-species tree. Large dots are nodes and tips. μ is the mean of the process, α is the strength of the restraining force, and ϵ is the scaling constant for the diffusion coefficient. Large dots are nodes and tips.

The stationary distribution of the process described by(9) is:

$$f(x) = \frac{1}{\mathrm{B}(\frac{\mu}{\delta}, \frac{(1-\mu)}{\delta})} x^{\frac{\mu}{\delta}-1}(1-x)^{\frac{(1-\mu)}{\delta}-1},\ \delta = \frac{\epsilon}{\alpha}$$

B is the Beta function. That is, $f(x|\mu,\delta)$ is the density of a Beta distribution with $x \sim Beta(\frac{\mu}{\delta}, \frac{(1-\mu)}{\delta})$ (Fig 2). The analysis of (8) and (9) and several other examples have been provided by Cobb (1998). It is interesting that in both cases, the substitution $\delta = \frac{\epsilon}{\alpha}$ was necessary in order to correctly recognise the distributions as Gamma or Beta. This suggests that the separate estimation of ϵ and α is difficult if estimation is based solely on the stationary distribution. The same stationary distributions occur for arbitrary α and ϵ, so long as their ratio (δ) remains constant. A sample evolutionary path from this Beta process is presented in Figure 3.

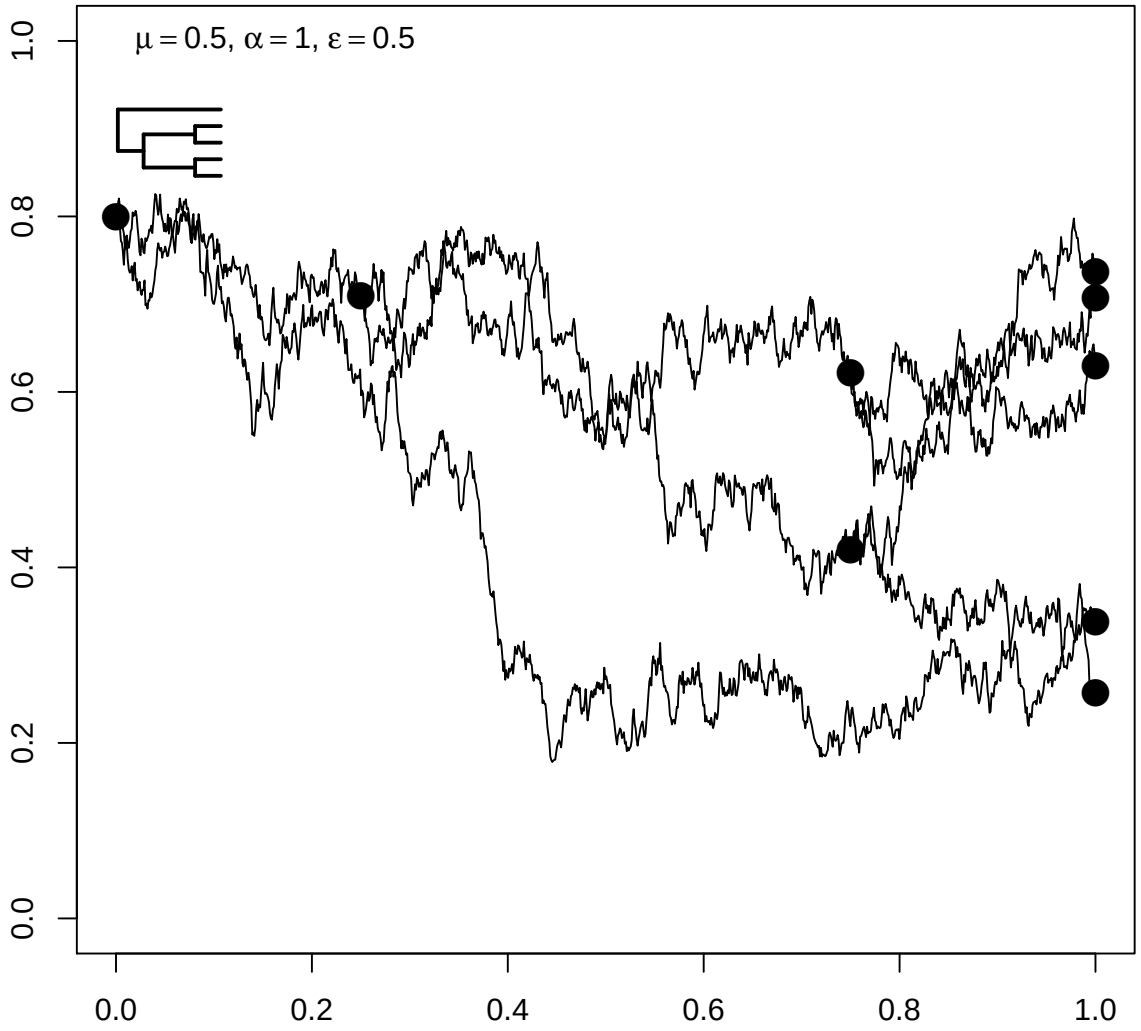

Figure 3: Beta evolution along a 5-species tree. μ is the mean of the process, α is the strength of the restraining force, and ϵ is the scaling constant for the diffusion coefficient. Large dots are nodes and tips.

Stochastic Differential Equations from Stationary Distributions

Reversing the procedure, that is deriving an SDE given a stationary distribution, is more difficult since the correspondence between SDEs and their stationary distribution (if it exists) is not unique. The problem has been addressed by Cai and Lin (1996). Extra information is needed, specifically the form of the spectral density of the process, which affects the structure of the drift coefficient in the SDE. Unfortunately for models of trait evolution, we rarely have detailed information on the evolutionary trajectory of a trait (ie the true historical realisation of the process) and hence we cannot analyse the spectral density of the trajectory in order to infer

a good model for the drift coefficient. We need to make extra assumptions. Fortunately, if we assume that the spectral density is of the low-pass filter type:

$$\Phi_{XX}(\omega) = \frac{\alpha\delta^2}{\pi(\omega^2 + \alpha^2)} \tag{10}$$

where Φ_{XX} is the spectral density at frequency ω, δ^2 is the mean-square value of the process $X(t)$, then the drift coefficient will be of the mean-reverting OU type in (4), with α in (10) being identical to α in (4). The low-pass filter assumption implies that the drift coefficient is determined mainly by the low frequency (long wavelength) characteristics of the evolutionary trajectory. That is, the form of the drift is mainly determined by long-lasting, slow deviations from μ and short-term (high-frequency) excursions are less important. To our knowledge, this assumption has never been made explicit in the literature on the application of the OU model in phylogenetic comparative methods.

Calculation of the diffusion coefficient comes directly from the application of the time-homogeneous Fokker-Planck equation (6), except instead of solving for $f(x)$, we now solve for $\sigma(x)$ (Cai and Lin 1996). The expression for $\sigma^2(x)$ becomes:

$$\sigma^2(x) = -\frac{2\alpha}{f(x)} \int^x y f(y) \mathrm{d}y.$$

Transition Distributions

The stationary distribution is not the only distribution associated with a Markov diffusion process. The transition, or conditional, distribution is important for simulation and likelihood calculations (Iacus 2008). It can be found as a solution to the Fokker-Planck equation (Klebaner 2012) and is defined as:

$$P(y, t, x, s) = P(X(t) \leq y | X(s) = x). \tag{11}$$

Equation (11) defines the probability distribution of y, the value of X occurring at time t, given that the process X has reached x at time s, where $s < t$. Unfortunately, for most processes the transition distribution is unknown or intractable. For Gaussian processes, the conditional density is usually straightforward. Brownian motion has as its transition distribution the Normal distribution with mean $\mu = \mathbb{E}(X_t | X_s = x) = x$, by the martingale property. The conditional variance of BM is $\mathbb{V}\mathrm{ar}(X_t | X_s = x) = \sigma^2 t$. That is, the variance is independent of the trait value and depends only on t.

For the OU process (4) and $t > s \geq 0$ the transition density is Gaussian with mean $\mathbb{E}(X_t | X_s = x) = x\mathrm{e}^{-\alpha(t-s)} + \mu(1 - \mathrm{e}^{-\alpha(t-s)})$ and variance $\mathbb{V}\mathrm{ar}(X_t | X_s = x) = \frac{\sigma^2}{2\alpha}(1 - \mathrm{e}^{-2\alpha(t-s)})$. The complexity of transition distributions increases quickly with the complexity of the corresponding SDE. Equation (8), the CIR model, has the the following transition density (Cox et al. 1985):

$$f(s, x, t, y) = c \left(\frac{u}{v}\right)^{\frac{\nu}{2}} \exp(-(u + v)) I_\nu(2\sqrt{uv})$$

for $t > s \geq 0$ where

$$c = \frac{2\alpha}{\epsilon^2(1 - \mathrm{e}^{-\alpha(t-s)})}, \; u = cx^{-\alpha(t-s)}, \; v = cy, \; \nu = \frac{2\alpha\mu}{\epsilon^2} - 1.$$

I_ν is the modified Bessel function of the first kind of order ν:

$$I_\nu(z) = \sum_{k=0}^{\infty} \left(\frac{z}{2}\right)^{2k+\nu} \frac{1}{k\,!\,\Gamma(k + \nu + 1)}$$

where $z \in \mathbb{R}^+$ and $\Gamma(\cdot)$ is the Gamma function. The expectation and variance of this distribution are:

$$\mathbb{E}(X_t | X_s = x) = 2\frac{\alpha\mu}{c\epsilon^2} + x\mathrm{e}^{-\alpha(t-s)}$$
$$\mathbb{V}\mathrm{ar}(X_t | X_s = x) = \frac{2}{c}\left(\frac{\alpha\mu}{c\epsilon^2} + x\mathrm{e}^{-\alpha(t-s)}\right)$$

respectively. The transition density of equation (9) is even more complicated and involves infinite sums of hypergeometric functions (Abundo 1997). Transition densities are of extreme importance for phylogenetic comparative methods, as they determine the structure of the evolutionary covariance matrix and the relationship between branch lengths and covariances. Hence, approaches such as PGLS (Grafen 1989; Martins and Hansen 1997; Blomberg et al. 2012) are intimately dependent on knowing transition densities. However, even quite simple models like the non-Gaussian models presented here are likely to present formidable problems with calculating evolutionary covariances, as the formulae for the transition densities for these models are extremely difficult to work with, if they are known at all.

Discussion

Fitting Models to Data

To be useful, theory must be confronted with data (Hilborn and Mangel 2013). The evolutionary models discussed here (of which BM and OU are special cases) therefore require methods to fit them to comparative data in order to estimate parameters and test hypotheses about those parameters. For BM, OU and other Gaussian processes, we can use the machinery developed for Normal distributions. In particular, there are simple relationships between branch lengths on a phylogeny and covariances for linear, Gaussian processes (Hansen and Martins 1996). There are no such relationships for non-linear processes that have stationary and/or transition distributions that are non-Gaussian so we cannot use inference methods that rely on expected covariances among species. Methods have recently been proposed for the calculation of likelihoods for continuous characters on a tree, if the transition density of the evolutionary model is known (Hiscott et al. 2015). However, often the transition density is not known in closed form, or not known at all. If there was no phylogenetic dependence (that is, a star phylogeny), we could estimate model parameters, as the existence of a stationary distribution implies that at any time point independent samples will follow the stationary distribution. Unfortunately in the case of comparative data with "phylogenetic signal", the data are not independent. We cannot make use of this result.

Simulation is currently the only option when the transition density of the process is intractable. Simulation based methods for estimating parameters for stochastic processes are widely available, largely based on theory developed for use in statistical finance (Iacus 2008). For example, stock prices may be observed every fraction of a second, resulting in a large amount of high-frequency data with which to make inferences. Methods to deal with missing data in the high-frequency setting have been developed (e.g Roberts and Stramer 2001). In this context the formidable problem is that data in comparative studies are only observed at the tips of the phylogeny. Rarely, internal branches may be calibrated with fossils. The entire evolutionary history of the trait for each species is thus missing and unknown. This is worse than "low-frequency" data (addressed by Fuchs 2013): it is almost "no-frequency" data. The simulation of the entire evolutionary history, except for the tip and fossil data, is necessary. However, we may be able to combine simulated and real data using data augmentation in a Bayesian framework which might permit the approximate estimation of model parameters (Tanner and Wong 1987; Papaspiliopoulos et al. 2013). An MCMC scheme that alternates between the update of simulated paths, and the sampling of parameters via data augmentation appears to be the most promising method (Fuchs 2013). Such an approach would require the use of bridge processes (e.g. Beskos et al. 2008; Bladt and Sørensen 2010; Lin et al. 2010) both to ensure that observed trait values are always part of the simulated trait evolutionary history and to iteratively update small sections of the simulated trait history at each iteration of an MCMC procedure. Acceptance rates during MCMC are higher when only small parts of the tree are updated at a time (Elerian 1999; Elerian et al. 2001; Roberts and Stramer 2001; Kalogeropoulos 2007). Methods exist for the updating of bridge processes in MCMC algorithms (Beskos et al. 2006, 2008, 2013).

The notion of using fossil phenotypes and dates to fix points in the trait-time space is attractive, but may contain grave difficulties. Cladistic criticisms of the use of fossils to establish ancestor-decendent relationships have never been refuted (Engelmann and Wiley 1977; Patterson 1981). The recent development of "tip dating" methods may avoid such criticism (Ronquist et al. 2012; O'Reilly et al. 2015). Instead we may have to be content to build quantitative trait models that incorporate ancestor-descendent relationships as ancilliary hypotheses, recognising that tests of such hypotheses may be impossible for any real dataset. However, simulation studies may be valuable in assessing the sensitivity of trait model parameter estimation to fossil placement (as an ancestor or as a sister taxon). It may be that inferring a fossil as a direct ancestor rather than as a close sister taxon will make little difference to parameter estimates for models of quantitative trait evolution. However, this has yet to be established.

The Lamperti transformation may be used to improve the simulation of trait trajectories by transforming to a unit diffusion coefficient (Lamperti 1962; Burnecki et al. 1997; Møller and Madsen 2010; Fuchs 2013). Consider equation (3). The Lamperti transformation is $Y = (Y_t)_{t \geq 0}$ where:

$$Y_t = g(X_t) = \int^{X_t} \frac{\mathrm{d}u}{\sigma(u)}$$

Provided the transformation $g(\cdot)$ exists and is invertible, Y fulfils the diffusion equation:

$$\mathrm{d}Y_t = \left(\frac{b(g^{-1}(Y_t), t)}{\sigma(g^{-1}(Y_t))} - \frac{1}{2}\frac{\partial \sigma}{\partial x}(g^{-1}(Y_t)) \right) \mathrm{d}t + \mathrm{d}B_t,$$

with $Y_{t_0} = g(x_0)$.

Transforming the model to remove any dependence of the diffusion coefficient on $X(t)$ and on t makes the transformed process "more Gaussian" but at the cost of increasing the complexity of the drift coefficient (Iacus 2008). However, there are grave difficulties even with fitting Gaussian models, where the transition density is known in closed form. OU has significant problems (Cooper et al. 2015), including problems with the identifiability of parameters (Ho and Ané 2014). Many simulated likelihood methods have been proposed for fitting models where the transition density is unknown (Brandt and Santa-Clara 2002; Durham and Gallant 2002; Sørensen 2004; Cano et al. 2006; Hurn et al. 2007; Kalogeropoulos 2007), including phylogenetic comparative methods (Kutsukake and Innan 2012). These methods often include a discretised, "locally Gaussian" approximation method such as the Euler scheme or the Milstein scheme (Elerian 1998; Iacus 2008). Bayesian simulation methods for parameter estimation in non-Gaussian stochastic process models of evolution is a current topic of research.

Stationarity

The notions of stationarity and stationary distributions have been central to this study. In the absence of an excellent fossil record of trait evolution for most traits and most taxa, it seems to be a necessary assumption for evolutionary stochastic process models that are more complicated than BM. Indeed, the success of the OU process in evolutionary studies is almost as much based on its stationarity as its Gaussian properties. Several authors have constructed non-stationary evolutionary models based on OU (e.g. Bartoszek 2012; Beaulieu et al. 2012; Jhwueng and Maroulas 2014). Non-stationarity can arise because of time dependence in the drift coefficient, time dependence in the diffusion coefficient, or both. For mean-reverting processes, the mean of the process μ and/or the strength of the restraining force α may be time dependent (Beaulieu et al. 2012). σ might vary with time smoothly over the tree (Bartoszek 2012).

Aside from the problem of overparameterisation (Bartoszek 2012), different parameters on different clades of the tree imply at least a short period of non-stationarity as species evolve from an ancestral evolutionary regime to the new conditions. Currently, OU based models assume immediate stationarity after the change in evolutionary regime (e.g. Beaulieu et al. 2012). If the old regime is almost the same as the new conditions, then stationarity in the new conditions may be achieved relatively quickly. However, if the old regime is very different from the new one, the length of the non-stationary period may be considerable and the underlying "instantaneous" stationary model will be wrong. Only fossil evidence can help in this regard because fossils can provide fixed points in the morphospace-time that can anchor the model, and provide evidence of non-stationary trait evolution or stasis. Of course, if the ancestral and derived stationary distributions are very similar so that stationarity is achieved quickly, it will be difficult to tell these two scenarios apart.

Model Extensions

An obvious extension of univariate stochastic processes is to re-cast them in a multivariate or multidimensional framework. There has been some research into multivariate phylogenetic comparative methods, including several software packages, largely based on BM, OU, and early-burst models (Zheng et al. 2009; Klingenberg 2011; Bartoszek 2011; Bartoszek et al. 2012; Klingenberg and Marugn-Lobn 2013; Adams 2014a,b,c; Clavel et al. 2015). Certainly, multivariate diffusions are necessary to understand the correlation among characters (Bartoszek et al. 2012). However, the properties of univariate diffusion models do not always carry over to the multivariate setting. In particular, there are well-known differences between the recurrence and transience properties of Brownian motion in multiple dimensions (Mörters and Peres 2010). The analysis of the properties of multivariate diffusion models for phylogenetically-correlated data is a topic of current research (Sherratt and Blomberg, in prep.).

A further extension of diffusion models is to the case where evolution is not strictly continuous, but consists of continuous evolution punctuated by "jumps" using Lèvy processes Landis et al. (2012). Lèvy processes are stochastic processes with independent, stationary increments. They can be thought of as consisting of three superimposed processes:

$$X_t = \sigma B_t + J_t + M_t$$

where B_t is a BM (possibly with drift), J_t is a compound Poisson point process, and M_t is a (square-integrable) martingale with jumps. Hence, simple BM is a special case of a Lèvy process with no discrete jumps. Note that OU is not a Lèvy process. Landis et al. (2012) estimate parameters for a Lèvy process fitted to data for body mass and brain volume in primates, and found evidence for some jumps in each trait, rejecting a simple BM model. The application of Lèvy processes to phylogenetic comparative data is promising, but given the difficulties and complexities of fitting Itô diffusions, it may pay to be wary of hidden pitfalls. Certainly the *post hoc* identification of jumps may not be of much use without a working hypothesis for *why* we may expect jumps at certain nodes on the tree, and models already exist for postulating *a priori* different rates of evolution in different parts of the tree (e.g. Butler and King 2004; O'Meara et al. 2006). It may be difficult to choose between "jump" models and models that estimate rapid changes of evolutionary rate (large differences in σ) for particular clades (e.g. Alfaro et al. 2009; Rabosky et al. 2013, 2014; Shi and Rabosky 2015). One may also object to "jump" models on theoretical grounds. Itô diffusions are continuous processes (although not differentiable), and as such represent the dictum, "*Natura non facit saltus*". If we are to allow jumps in evolutionary history, we should be able to provide a mechanistic (genetic) explanation of how and why jumps occur, and how to distinguish jumps from rapid, continuous evolution.

Evolutionary models for phylogenetic comparative analyses

Scientific models may be developed with several different motivations. The scientist may build models to make a decision (e.g. to reject a null hypothesis), summarise evidence (e.g. calculate the likelihood of observing

the data, given a model) or quantify their beliefs (using Bayes Theorem). Another important property of a model is its predictive ability, and predictive models have long been the favourite approach in the physical sciences: models predict future observations which then test the validity of the model. Biologists, and especially evolutionary biologists, have never put much faith in predictive models (Hillis 1993). So many factors affect the evolution of organisms, and over such a long timespan, that one is tempted to give up hope of developing mathematical models that have any predictive value in the real world. And it is true that it would be foolish to make predictions of where in the phylomorphospace *sensu* Sidlauskas (2008) species will evolve to in some future deep time. We have no hope of making the necessary observations. Although we may not be able to predict the precise evolutionary trajectory of any particular species, we can perhaps predict (or postdict) the probability distribution of traits across species. Given the traits from a newly discovered species (fossil or extant), we can predict that the new trait values fit well within the distribution of the known species' trait values, where the parameters of the distribution are estimated from extant species. If the values for the new species' traits are more extreme so that they fall into the tails of the stationary distribution, we may reject our model of evolution for that set of species and traits. This "grey box" approach to model identification (Kristensen et al. 2004) gives up the possibility of knowledge of the microevolutionary processes leading to species diversification and trait evolution, and replaces it with a tractable stochastic process that summarises the evolution of the statistical distribution of trait values over deep time. Given the quality of most comparative data sets, this may be the best that can be achieved.

The modelling approach and the new models described here involve a considerable amount of mathematical sophistication in their derivation and in the analysis of their properties. Computational skill is necessary in developing algorithms to fit the models to data. Critics may object that the approach outlined here is too complex or unnecessary, given the quality of data in most phylogenetic comparative analyses. However, diffusions are already the most popular model for phylogenetic comparative studies, in the form of BM and OU. The present author hopes simply to widen horizons and provide a unifying framework. It is true that, "All models are wrong, but some are useful" (Box 1976). Nevertheless, mathematics (and its sister taxon, computation) are the best tools we have in order to precisely describe both the nature of macroevolutionary phenomena and our assumptions about them. A small amount of precise mathematics can sometimes cut through imprecise verbal arguments. For example, the microevolutionary genetic theory developed by Fisher, Haldane and Wright effectively silenced the arguments between naturalists and Mendelians on the importance of natural selection and the nature of genetic variation, leading to the Evolutionary Synthesis (Mayr and Provine 1998). A mathematical theory of macroevolution which unites stochastic models of trait evolution with models of phylogenesis, speciation and extinction may allow us to better statistically model the course of phenotypic evolution (e.g. Maddison et al. 2007; FitzJohn 2010; Goldberg et al. 2011), although estimating parameters for these models may be difficult without fossil trait data. Recent applications of trait-mediated diversification models based only on extant trait data may be misleading (Rabosky and Goldberg 2015). A more sophisticated understanding of the mathematics of diffusions and other stochastic processes may allow the critical appraisal

of macroevolutionary models for biological phenomena in deep time.

Conclusion

Currently popular models of trait evolution rely heavily on Gaussian processes and their useful mathematical properties. However, non-Gaussian models are possible and may have some advantages over Gaussian models in certain situations where the data are likely to be non-Normal. The present study describes new, non-Gaussian models of trait evolution, together with methods for building new models, and a discussion of the mathematical and computational difficulties in working with diffusion models in a more generalised setting. Several new avenues for investigation are suggested. In particular, the role of fossils in improving the identifiability of models and the extension of models to multivariate trait space seem especially timely. These areas are not without challenges. Including fossils as ancestors, rather than as sister taxa has been a difficult problem for many years, as the early cladists were well aware. The extension of univariate models to multivariate trait space is likely to be more difficult than expected, as even the simplest evolutionary model, BM, has different properties in multiple dimensions. Another important research direction is to establish the expected covariances for traits in terms of the transition distributions for non-Gaussian models. This is likely to be difficult but would pay off immensely, allowing the the construction of a new Generalized Phylogenetic Model, by analogy with Generalized Linear Models. Nevertheless, research into the application of stochastic process (diffusion) models to the evolution of quantitative traits appears to hold great promise. Critics may now be admitted to the event.

Funding

This work was supported by the University of Queensland and the Australian Research Council (DP140101915).

Acknowledgments

Thanks to A. R. Ives, S. Rathnayake, E. Sherratt for insightful comments on previous drafts of this manuscript. J. S. Keogh kindly allowed me time and office space at the Division of Ecology, Evolution and Genetics at the Research School of Biology, Australian National University, to complete the work.

*

References

Abundo, M. 1997. On some properties of one-dimensional diffusion processes on an interval. Probability and Mathematical Statistics 17:277–310.

Adams, D. C. 2014a. A generalized K statistic for estimating phylogenetic signal from shape and other high-dimensional multivariate data. Systematic Biology 63:685–697.

Adams, D. C. 2014b. A method for assessing phylogenetic least squares models for shape and other high-dimensional multivariate data. Evolution 68:2675–2688.

Adams, D. C. 2014c. Quantifying and comparing phylogenetic evolutionary rates for shape and other high-dimensional phenotypic data. Systematic Biology 63:166–177.

Alfaro, M. E., F. Santini, C. Brock, H. Alamillo, A. Dornburg, D. L. Rabosky, G. Carnevale, and L. J. Harmon. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. Proceedings of the National Academy of Sciences 106:13410–13414.

Bachelier, L. 1900. Theorie De La Speculation. Ph.D. thesis Ecole Normale Superieure.

Bartoszek, K. 2011. Multivariate Aspects of Phylogenetic Comparative Methods. Ph.D. thesis Chalmers University of Technology and University of Gothenburg.

Bartoszek, K. 2012. The Laplace motion in phylogenetic comparative methods. arXiv 403.1562 [q-bio.PE].

Bartoszek, K., J. Pienaar, P. Mostad, S. Andersson, and T. F. Hansen. 2012. A phylogenetic comparative method for studying multivariate adaptation. Journal of Theoretical Biology 314:204 – 215.

Beaulieu, J. M., D.-C. Jhwueng, C. Boettiger, and B. C. O'Meara. 2012. Modeling stabilizing selection: Expanding the Ornstein-Uhlenbeck model of adaptive evolution. Evolution 66:2369–2383.

Benton, M. J. 2015. Exploring macroevolution using modern and fossil data. Proceedings of the Royal Society of London B: Biological Sciences 282.

Beskos, A., K. Kalogeropoulos, and E. Pazos. 2013. Advanced MCMC methods for sampling on diffusion pathspace. Stochastic Processes and their Applications 123:1415 – 1453.

Beskos, A., O. Papaspiliopoulos, and G. O. Roberts. 2006. Retrospective exact simulation of diffusion sample paths with applications. Bernoulli 12:1077–1098.

Beskos, A., G. Roberts, A. Stuart, and J. Voss. 2008. MCMC methods for diffusion bridges. Stochastics and Dynamics 08:319–350.

Black, F. and M. Scholes. 1973. The pricing of options and corporate liabilities. The Journal of Political Economy 81:637–654.

Bladt, M. and M. Sørensen. 2010. Simple simulation of diffusion bridges with application to likelihood inference for diffusions. Tech. rep. School of Economics and Management, Aarhus University Bartholins Allé 10, Building 1322, DK-8000 Aarhus C Denmark cREATES Research Paper 2010-32.

Blomberg, S. P., T. Garland, and A. R. Ives. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution 57:717–745.

Blomberg, S. P., J. G. Lefevre, J. A. Wells, and M. Waterhouse. 2012. Independent contrasts and PGLS regression estimators are equivalent. Systematic Biology Page syr118.

Box, G. E. P. 1976. Science and statistics. Journal of the American Statistical Association 71:791–799.

Brandt, M. W. and P. Santa-Clara. 2002. Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. Journal of Financial Economics 63:161–210.

Breiman, L. 1968. Probability. Addison-Wesley, Reading, MA.

Burnecki, K., M. Maejima, and A. Weron. 1997. The lamperti transformation for self-similar processes. HSC Research Reports HSC/97/02 Hugo Steinhaus Center, Wroclaw University of Technology.

Butler, M. A. and A. A. King. 2004. Phylogenetic comparative analysis: A modeling approach for adaptive evolution. The American Naturalist 164:683–695.

Cai, G. Q. and Y. K. Lin. 1996. Generation of non-gaussian stationary stochastic processes. Physical Review E 54:299–303.

Cano, J., M. Kessler, and D. Salmern. 2006. Approximation of the posterior density for diffusion processes. Statistics & Probability Letters 76:39 – 44.

Clavel, J., G. Escarguel, and G. Merceron. 2015. mvmorph: an R package for fitting multivariate evolutionary models to morphometric data. Methods in Ecology and Evolution 6:1311–1319.

Cobb, L. 1998. Mathematical Frontiers of the Social and Policy Sciences chap. 2 Stochastic Differential Equations for the Social Sciences. Westview Press.

Conway Morris, S., J. F. Hoyal Cuthill, and S. Gerber. 2015. Hunting Darwin's snark: which maps shall we use? Interface Focus 5.

Cooper, N., G. H. Thomas, C. Venditti, A. Meade, and R. P. Freckleton. 2015. A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies. Biological Journal of the Linnean Society .

Cox, J. C., J. E. Ingersoll, and S. A. Ross. 1985. A theory of the term structure of interest rates. Econometrica 53:385–407.

Darriba, D., G. Taboada, R. Doallo, and D. Posada. 2012. jmodeltest2: more models, new heuristics and parallel computing. Nature Methods 9:772.

Durham, G. B. and A. R. Gallant. 2002. Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. Journal of Business & Economic Statistics 20:297–316.

Einstein, A. 1956. Investigations on the Theory of the Brownian Movement. Dover, Mineola, NY.

Elerian, O. 1998. A note on the existence of a closed form conditional transition density for the Milstein scheme. Economics Series Working Papers 1998-W18 University of Oxford, Department of Economics.

Elerian, O. 1999. Simulation Estimation of Continuous-time Models with Applications to Finance. Nuffield College theses University of Oxford.

Elerian, O., S. Chib, and N. Shephard. 2001. Likelihood inference for discretely observed nonlinear diffusions. Econometrica 69:959–93.

Engelmann, G. and E. Wiley. 1977. The place of ancestor-descendant relationships in phylogeny reconstruction. Systematic Biology 26:1–11.

Ewens, W. J. 2004. Mathematical population genetics / Warren J. Ewens. 2nd ed. Springer-Verlag Berlin ; New York.

Feller, W. 1951. Diffusion processes in genetics. Pages 227–246 *in* Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability University of California Press, Berkeley, Calif.

Felsenstein, J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. Am. J. Hum. Genet. 25:471–492.

Felsenstein, J. 1985. Phylogenies and the comparative method. The American Naturalist 125:1–15.

Felsenstein, J. 1988. Phylogenies and quantitative characters. Annual Review of Ecology and Systematics 19:445–471.

Fisher, R. A. 1922. On the dominance ratio. Proc. Roy. Soc. Edin. 42:321–341.

FitzJohn, R. G. 2010. Quantitative traits and diversification. Systematic Biology 59:619–633.

Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative data: a test and review of the evidence. The American Naturalist 160:712–726.

Freund, J. A. and T. Pöschel, eds. 2000. Stochastic Processes in Physics, Chemistry, and Biology. No. 0075-8450 in Lecture Notes in Physics Springer, Berlin Heidelberg 557 pp.

Fuchs, C. 2013. Inference for Diffusion Processes: With Applications in Life Sciences. Springer Science & Business Media.

Gardiner, C. 2009. Stochastic Methods. A Handbook for the Natural and Social Sciences. Springer Series in Synergetics 4th ed. Springer.

Garland, T., A. W. Dickerman, C. Janis, and J. A. Jones. 1993. Phylogenetic analysis of covariance by computer simulation. Systematic Biology 42:265–292.

Goldberg, E. E., L. T. Lancaster, and R. H. Ree. 2011. Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. Systematic Biology 60:451–465.

Grafen, A. 1989. The phylogenetic regression. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 326:119–157.

Hansen, T. F. 1997. Selection and the comparative analysis of adaptation. Evolution 51:1341–1351.

Hansen, T. F. and E. P. Martins. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. Evolution 50:1404–1417.

Hansen, T. F., J. Pienaar, and S. H. Orzack. 2008. A comparative method for studying adaptation to a randomly evolving environment. Evolution 62:1965–77.

Hilborn, R. and M. Mangel. 2013. The Ecological Detective. Monographs in Population Biology Princeton University Press.

Hillis, W. D. 1993. Why physicists like models and why biologists should. Current Biology 3:79–81.

Hiscott, G., C. Fox, M. Parry, and D. Bryant. 2015. Efficient recycled algorithms for quantitative trait models on phylogenies. ArXiv e-prints .

Ho, L. S. T. and C. Ané. 2014. Intrinsic inference difficulties for trait evolution with Ornstein-Uhlenbeck models. Methods in Ecology and Evolution 5:1133–1146.

Hurn, A. S., J. I. Jeisman, and K. A. Lindsay. 2007. Seeing the wood for the trees: A critical evaluation of methods to estimate the parameters of stochastic differential equations. Journal of Financial Econometrics 5:390–455.

Iacus, S. M. 2008. Simulation and Inference for Stochastic Differential Equations: With R Examples (Springer Series in Statistics). 1st ed. Springer Publishing Company, Incorporated.

Ingram, T., L. J. Harmon, and J. B. Shurin. 2012. When should we expect early bursts of trait evolution in comparative data? predictions from an evolutionary food web model. Journal of Evolutionary Biology 25:1902–1910.

Insua, D. R., F. Ruggeri, and M. P. Wiper. 2012. Bayesian Analysis of Stochastic Process Models. Wiley, Chichester, UK.

Itô, K. 1944. Stochastic integral. Proceedings of the Imperial Academy 20:519–524.

Itô, K. 1946. On a stochastic integral equation. Proceedings of the Japan Academy 22:32–35.

Ives, A. R. 2015. For testing the significance of regression coefficients, go ahead and log-transform count data. Methods in Ecology and Evolution 6:828–835.

Jhwueng, D.-C. and V. Maroulas. 2014. Phylogenetic Ornstein-Uhlenbeck regression curves. Statistics & Probability Letters 89:110 – 117.

Kalogeropoulos, K. 2007. Likelihood-based inference for a class of multivariate diffusions with unobserved paths. Journal of Statistical Planning and Inference 137:3092 – 3102 special Issue: Bayesian Inference for Stochastic Processes.

Klebaner, F. C. 2012. Introduction to Stochastic Calculus with Applications. 3rd ed. Imperial College Press, London.

Klingenberg, C. P. 2011. MorphoJ: an integrated software package for geometric morphometrics. Molecular Ecology Resources 11:353–357.

Klingenberg, C. P. and J. Marugn-Lobn. 2013. Evolutionary covariation in geometric morphometric data: Analyzing integration, modularity, and allometry in a phylogenetic context. Systematic Biology 62:591–610.

Kristensen, N. R., H. Madsen, and S. B. Jørgensen. 2004. Parameter estimation in stochastic grey-box models. Automatica 40:225–237.

Kutsukake, N. and H. Innan. 2012. Simulation-based likelihood approach for evolutionary models of phenotypic traits on phylogeny. Evolution 67:355367.

Lamperti, J. 1962. Semi-stable stochastic processes. Transactions of the American Mathematical Society 104:62–78.

Lande, R. 1976. Natural selection and random genetic drift in phenotypic evolution. Evolution 30:314–334.

Landis, M. J., J. G. Schraiber, and M. Liang. 2012. Phylogenetic analysis using Lèvy processes: Finding jumps in the evolution of continuous traits. Systematic Biology 62:193–204.

Lebowitz, J. L. and O. Penrose. 1973. Modern ergodic theory. Physics Today Pages 23–29.

Lin, M., R. Chen, and P. Mykland. 2010. On generating Monte Carlo samples of continuousdiffusion bridges. Journal of the American Statistical Association 105:820–838.

Maddison, W. P., P. E. Midford, and S. P. Otto. 2007. Estimating a binary character's effect on speciation and extinction. Systematic Biology 56:701–710.

Martins, E. P. and T. F. Hansen. 1997. Phylogenies and the comparative method: A general approach to incorporating phylogenetic information into the analysis of interspecific data. The American Naturalist 149:646–667.

Mayr, E. and W. B. Provine. 1998. The evolutionary synthesis: perspectives on the unification of biology. Harvard University Press.

McCullagh, P. and J. A. Nelder. 1989. Generalized linear models vol. 37. CRC press.

McGhee, G. R. 2015. Limits in the evolution of biological form: a theoretical morphologic perspective. Interface Focus 5.

Mikosch, T. 1998. Elementary Stochastic Calculus with Finance in View vol. 6 of *Advanced Series on Statistical Science & Applied Probability*. World Scientific, Hackensack, NJ.

Møller, J. K. and H. Madsen. 2010. From State Dependent Diffusion to Constant Diffusion in Stochastic Differential Equations by the Lamperti Transform. IMM-Technical Report-2010-16 DTU Informatics, Building 321.

Moran, P. A. P. 1958. Random processes in genetics. Mathematical Proceedings of the Cambridge Philosophical Society 54:60–71.

Mörters, P. and Y. Peres. 2010. Brownian motion vol. 30. Cambridge University Press.

Øksendal, B. 2007. Stochastic Differential Equations. An Introduction with Applications. 6th ed. Springer, Berlin.

O'Meara, B. C., C. Ane, M. J. Sanderson, and P. C. Wainwright. 2006. Testing for different rates of continuous trait evolution using likelihood. Evolution 60:922–933.

O'Reilly, J. E., M. dos Reis, and P. C. J. Donoghue. 2015. Dating tips for divergence-time estimation. Trends in Genetics 31:637–650.

Pagel, M. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.

Papaspiliopoulos, O., G. O. Roberts, and O. Stramer. 2013. Data augmentation for diffusions. Journal of Computational and Graphical Statistics 22:665–688.

Patterson, C. 1981. Significance of fossils in determining evolutionary relationships. Annual Review of Ecology and Systematics 12:195–223.

Pennell, M. W., R. G. FitzJohn, W. K. Cornwell, and L. J. Harmon. 2015. Model adequacy and the macroevolution of angiosperm functional traits. The American Naturalist 186:E33–E50.

Pollard, D. 2002. A User's Guide to Measure Theoretical Probability. Cambridge Univ Press.

Posada, D. and K. A. Crandal. 2001. Selecting the best-fit model of nucleotide substitution. Systematic Biology 50:580–601.

Rabosky, D. L., S. C. Donnellan, M. Grundler, and I. J. Lovette. 2014. Analysis and visualization of complex macroevolutionary dynamics: An example from australian scincid lizards. Systematic Biology 63:610–627.

Rabosky, D. L. and E. E. Goldberg. 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. Systematic Biology 64:340–355.

Rabosky, D. L., F. Santini, J. Eastman, S. A. Smith, B. Sidlauskas, J. Chang, and M. E. Alfaro. 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. Nat Commun 4 article.

Rensch, B. 1959. Evolution above the species level. Methuen.

561 Risken, H. 1996. The Fokker-Planck Equation vol. 18 of *Springer Series in Synergetics*. 2nd edition ed. Springer
562 Berlin Heidelberg.

563 Roberts, G. O. and O. Stramer. 2001. On inference for partially observed nonlinear diffusion models using the
564 metropolis-hastings algorithm. Biometrika 88:603–621.

565 Ronquist, F., S. Klopfstein, L. Vilhelmsen, S. Schulmeister, D. L. Murray, and A. P. Rasnitsyn. 2012. A total-
566 evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. Systematic
567 Biology 61:973–999.

568 Serrelli, E. and N. Gontier. 2015. Macroevolution: Explanation, Interpretation and Evidence. Springer.

569 Shi, J. J. and D. L. Rabosky. 2015. Speciation dynamics during the global radiation of extant bats. Evolution
570 69:1528–1545.

571 Sidlauskas, B. 2008. Continuous and arrested morphological diversification in sister clades of characiform fishes:
572 A phylomorphospace approach. Evolution 62:3135–3156.

573 Simpson, G. G. 1953. The Major Features of Evolution. Columbia Univ. Press, New York.

574 Sørensen, H. 2004. Parametric inference for diffusion processes observed at discrete points in time: A survey.
575 International Statistical Review / Revue Internationale de Statistique 72:337–354.

576 Stanley, S. M. 1975. A theory of evolution above the species level. Proceedings of the National Academy of
577 Sciences of the United States of America 72:646–650.

578 Tanner, M. and W. Wong. 1987. The calculation of posterior distributions by Data Augmentation (with discus-
579 sion). Journal of the American Statistical Association 82:528–550.

580 Turelli, M. 1977. Random environments and stochastic calculus. Theoretical Population Biology 12:140–178.

581 Uhlenbeck, G. E. and L. S. Ornstein. 1930. On the theory of Brownian motion. Physical Review 36:823–841.

582 Vermeij, G. J. 2015. Forbidden phenotypes and the limits of evolution. Interface Focus 5.

583 Warton, D. I., M. Lyons, J. Stoklosa, and A. R. Ives. 2016. Three points to consider when choosing a LM or
584 GLM test for count data. Methods in Ecology and Evolution .

585 Wiener, N. 1923. Differential space. Journal of Mathematical Physics 2:131–174.

586 Wright, S. 1931. Evolution in mendelian populations. Genetics 16:97–159.

587 Wright, S. 1938. The distribution of gene frequencies under irreversible mutation. Proceedings of the National
588 Academy of Sciences of the United States of America 24:253–259.

589 Zheng, L., A. R. Ives, T. Garland, B. R. Larget, Y. Yu, and K. Cao. 2009. New multivariate tests for phylogenetic
590 signal and trait correlations applied to ecophysiological phenotypes of nine Manglietia species. Functional
591 Ecology 23:1059–1069.

Appendix 1: Derivation of Wright's Equation

Consider the Fokker-Planck equation for an Itô diffusion X_t (5). Alternatively, (5) can be re-written as (Risken 1996):

$$\begin{aligned} \frac{\partial}{\partial t} f(x,t) &= L_{FP} f(X_t,t), \\ L_{FP} &= -\frac{\partial}{\partial x} b(X_t,t) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\sigma(X_t,t) \end{aligned} \tag{A1.1}$$

Further, equations (A1.1) can be written as:

$$\begin{aligned} \frac{\partial f(X_t,t)}{\partial t} + \frac{\partial S}{\partial x} &= 0, \\ S(X_t,t) &= \left[b(X_t,t) - \frac{1}{2}\frac{\partial}{\partial x}\sigma(X_t,t) \right] f(X_t,t) \end{aligned} \tag{A1.2}$$

$S(X_t,t)$ can be interpreted as a probability flow. For natural boundary conditions $\min x = -\infty$ and $\max x = \infty$, and assuming time-homogeneity, $S(X_t,t) = S(X_t) = 0$. Letting $x = X_t$ we have the following first-order linear differential equation:

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}x}\left[\sigma^2(x) f(x)\right] - b(x) f(x) = 0$$

Let $m(x) = \sigma^2(x) f(x)$, implying $f(x) = \frac{m(x)}{\sigma^2(x)}$ then

$$\frac{\mathrm{d}m(x)}{\mathrm{d}x} - 2\frac{b(x)m(x)}{\sigma^2(x)} = 0 \tag{A1.3}$$

Equation (A1.3) can be solved using the method of integrating factors. Let $I = \mathrm{e}^{-2\int^x \frac{b(y)}{\sigma^2(y)}\mathrm{d}y}$. Multiplying both sides of equation (A1.3) by I:

$$\mathrm{e}^{-2\int^x \frac{b(y)}{\sigma^2(y)}\mathrm{d}y}\frac{\mathrm{d}m}{\mathrm{d}x} - 2\frac{b(x)m(x)}{\sigma^2(x)}\mathrm{e}^{-2\int^x \frac{b(y)}{\sigma^2(y)}\mathrm{d}y} = 0$$

Integrating both sides and using the product rule on the LHS,

$$\mathrm{e}^{-2\int^x \frac{b(y)}{\sigma^2(y)}\mathrm{d}y} m(x) = C \tag{A1.4}$$

where C is a constant of integration. Substituting $m(x) = \sigma^2(x) f(x)$ and rearranging, we have:

$$f(x) = \frac{C}{\sigma^2(x)}\mathrm{e}^{2\int^x \frac{b(y)}{\sigma^2(y)}\mathrm{d}y} \tag{A1.5}$$

which is Wright's formula.

Appendix 2: Derivation of Stationary Distributions

CIR model

Let $b(x) = \alpha(\mu - x)$, $\sigma = \sqrt{\epsilon x}$. Substituting into Wright's formula (7):

$$\begin{aligned} f(x) &= \frac{C}{\epsilon x}\exp\left[\int_{-\infty}^{x} \frac{2\alpha(\mu - s)}{\epsilon s}\mathrm{d}s\right] \\ &= \frac{C}{\epsilon x}\exp\left[\frac{\mu\log(x) - x}{\epsilon/\alpha}\right] \end{aligned}$$

Let $\delta = \frac{\epsilon}{\alpha}$. Then:

$$\begin{aligned} f(x) &= Cx^{-1}x^{\frac{\mu}{\delta}}\mathrm{e}^{\frac{-x}{\delta}} \\ &= Cx^{\frac{\mu}{\delta}-1}\mathrm{e}^{\frac{-x}{\delta}} \end{aligned} \tag{A2.1}$$

Equation (A2.1) can be recognised as the kernel of a Gamma density, with shape $\frac{\mu}{\delta}$ and scale δ, and with normalising constant C. Therefore,

$$C = \frac{1}{\Gamma(\frac{\mu}{\alpha})\delta^{\frac{\mu}{\delta}}}$$

where $\Gamma(\cdot)$ is the Gamma function. i.e.

$$f(x|\mu,\delta) = \frac{1}{\Gamma(\frac{\mu}{\alpha})\delta^{\frac{\mu}{\delta}}}x^{\frac{\mu}{\delta}-1}\mathrm{e}^{\frac{-x}{\delta}}$$

or $x|\mu,\delta \sim \text{Gamma}(\frac{\mu}{\delta},\delta)$.

Beta model

Let $b(x) = \alpha(\mu - x)$, $\sigma = \sqrt{\epsilon x(1-x)}$. Substituting into Wright's formula (7):

$$\begin{aligned} f(x) &= \frac{C}{\epsilon x(1-x)}\exp\left[\int_{-\infty}^{x}\frac{2\alpha(\mu-s)}{\epsilon s(1-s)}\mathrm{d}s\right] \\ &= \frac{C}{\epsilon x(1-x)}\mathrm{e}^{(\frac{\alpha}{\epsilon}\mu\log x + \frac{\alpha}{\epsilon}(1-\mu)\log(1-x))} \\ &= \frac{C}{\epsilon x(1-x)}x^{\frac{\alpha\mu}{\epsilon}}(1-x)^{\frac{\alpha(1-\mu)}{\epsilon}} \end{aligned}$$

Setting $\delta = \frac{\epsilon}{\alpha}$ and simplifying further, we have:

$$f(x) = Cx^{\frac{\mu}{\delta}-1}(1-x)^{\frac{1-\mu}{\delta}-1} \tag{A2.2}$$

Equation (A2.2) is the kernel of a Beta distribution with shape parameters $\frac{\mu}{\delta}$ and $\frac{(1-\mu)}{\delta}$, and normalising constant C. Hence, the density can be written as:

$$f(x|\mu,\delta) = \frac{1}{\mathrm{B}(\frac{\mu}{\delta},\frac{(1-\mu)}{\delta})}x^{\frac{\mu}{\delta}-1}(1-x)^{\frac{(1-\mu)}{\delta}-1}$$

where $\mathrm{B}(\cdot,\cdot)$ is the Beta function. More succinctly, $x|\mu,\delta \sim \text{Beta}(\frac{\mu}{\delta},\frac{(1-\mu)}{\delta})$.