

Forget trees and regimes for the moment. Recall that for the OU process $dX = \alpha_1(\theta_1 - X)dt + \sigma_1 dW_t$ starting at X_0 at time 0, it's value at time T_1 is given by the:

$$X_{T_1} = X_0 e^{-\alpha_1 T_1} + \theta_1 (1 - e^{-\alpha_1 T_1}) + \int_0^{T_1} \sigma_1 e^{\alpha_1(s-T_1)} dW_s \quad (1)$$

Image at time $T = T_i$ this process switches to possibly new values of the constants α_i , θ_i , and σ_i , which we will denote with the subscript i . The process just picks up where it left off:

$$X_{T_2} = X_{T_1} e^{-\alpha_2(T_2-T_1)} + \theta_2 (1 - e^{-\alpha_2(T_2-T_1)}) + \int_{T_1}^{T_2} \sigma_2 e^{\alpha_2(s-T_1)} dW_s \quad (2)$$

And so on. Define the length of an interval as $t_i \equiv (T_i - T_{i-1})$. Then the value of X at the end of the n th interval is given by, using the recursion above,

$$X_{T_n} = X_0 e^{-\sum_{i=1}^n \alpha_i t_i} + \sum_{i=1}^n \left(\theta_i (1 - e^{\alpha_i t_i}) \exp \left(- \sum_{j=i+1}^n \alpha_j t_j \right) \right) + \sum_{i=1}^n \left[\exp \left(- \sum_{j=i+1}^n \alpha_j t_j \right) \int_{T_{i-1}}^{T_i} \sigma_i e^{\alpha_i(s-T_i)} dW_s \right] \quad (3)$$

Let's define a bit more notation to simply this expression before we go on. Inside the sums across the whole time series are these sums from the current position i forward till the end. So let's define that term:

$$\gamma_i = - \sum_{j=i+1}^n \alpha_j t_j \quad (4)$$

Also note that because the parameters are constant over the range inside each integration, we can rewrite those integrals in terms of the interval lengths t_i as well, simplifying the notation a bit further.¹ Now we're have the more manageable expression:

$$X_{T_n} = X_0 e^{\gamma_0} + \sum_{i=1}^n (\theta_i (1 - e^{\alpha_i t_i}) e^{\gamma_i}) + \sum_{i=1}^n \left[e^{\gamma_i} \int_0^{t_i} \sigma_i e^{\alpha_i(s-t_i)} dW_s^i \right] \quad (5)$$

Let's first consider the expectation of this process. Remeber that the Ito integral is a martingale, so it's expectation is zero, while the more friendly sums remain:

$$\mathbb{E}(X_{T_n}) = X_0 e^{\gamma_0} + \sum_{i=1}^n (\theta_i (1 - e^{\alpha_i t_i}) e^{\gamma_i}) \quad (6)$$

¹This technically involves a change of variables, $s - T_{i-1}$. The important thing is just that the integrals aren't over the *same* Brownian motion on each interval. We distinguish them with the added i superscript.

Now let us consider the covariance $\text{Cov}(X_T, X_S) = \text{E}[(X_T - \text{E}(X_T))(X_S - \text{E}(X_S))]$, given that each started at X_0 at time 0. Subtracting out the expected value from our full equation (5) leaves us with the product of the Ito integrals:

$$\text{Cov}(X_T, X_S) = \sum_{i=1}^n \left[e^{\gamma_i} \int_0^{t_i} \sigma_i e^{\alpha_i(u-t_i)} dW_u^i \right] \sum_{j=1}^m \left[e^{\gamma'_j} \int_0^{s_j} \sigma'_j e^{\alpha'_j(v-s_j)} dW_v^j \right] \quad (7)$$

Where we use a prime to with the γ'_j and not γ_i to remind us that those functions are defined as sums starting from i or j and going all the way to different tip indices. This will matter later.

We are now ready for some tree thinking. Let X_T and X_S be the traits of tips in the phylogeny. Consider that the parameters $\alpha_i, \theta_i, \sigma_i$ can only change at branching points (the common simplifying assumption). Consider each edge or branch in the phylogeny traversed by going back to the root from X_T as one of the n intervals in the above equation, where the i th interval has length t_i . Consider each edge going back from the other tip X_S to be one of the m intervals in the above equation, where the j th interval has branch length s_j . We use primes simply to remind us that when a parameter is being indexed by j , it is relative to the history of the m intervals in X_s . Let us begin at the root, and assume that X_S and X_T share a common ancestor some time more recently than that. (If they do not, their history is clearly independent given the root state, and the covariance is 0).

While they share history, for each branch where $i=j$, $t_i = s_j$, they are driven by the same Brownian process, $dW_u = dW_v$. As we factor out the terms in the sum, we can immediately set any cross-term to zero, since it multiplies independent intervals of the path. Similarly, we can set to zero all intervals in the sum after the most common ancestor k , since again, the paths are independent and contribute no covariance.

$$= \sum_{i=1}^k \left[e^{\gamma_i} \int_0^{t_i} \sigma_i e^{\alpha_i(u-t_i)} dW_u \right] \left[e^{\gamma'_i(m)} \int_0^{t_i} \sigma_i e^{\alpha_i(u-t_i)} dW_u \right] \quad (8)$$

Note that we have taken the product along $i = j$, and asserted $t_i = s_i$, $\alpha_i = \alpha'_i$, etc. The only difference in the two terms are the γ_i and γ'_i values – recall though they both start at time i one sums all the way up the history of X_s and the other up X_t . only when we consider the variance of a single tip, $X_s = X_t$, are they the same. Note also that these terms sit outside of the Ito integrals.

We can now use Ito isometry to multiply the identical Ito integrals, returning a normal Riemann-Stiljes integral,

$$\begin{aligned} &= \sum_{i=1}^k \left[e^{\gamma_i} e^{\gamma'_i} \left(\int_0^{t_i} \sigma_i e^{\alpha_i(u-t_i)} dW_u \right)^2 \right] \\ &= \sum_{i=1}^k \left[e^{\gamma_i} e^{\gamma'_i} \int_0^{t_i} \sigma_i^2 e^{2\alpha_i(u-t_i)} du \right] \end{aligned} \quad (9)$$

integrating, we have

$$\text{Cov}(X_T, X_S) = \sum_{i=1}^k \left[e^{\gamma_i} e^{\gamma'_i} \frac{\sigma_i^2}{2\alpha_i} (1 - e^{-2\alpha_i t_i}) \right] \quad (10)$$

Since this is still a Gaussian process Hansen (1997), the probability distribution is completely determined by the mean and covariance equations (6) and (10), with γ_i given in (4).

References

- Hansen, T. F., 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution; international journal of organic evolution* 51 (5), 1341–1351.
URL <http://www.jstor.org/stable/2411186>