UPPSALA UNIVERSITET

Tree–free PCM

Tree–free phylogenetic comparative methods: macroevolutionary dynamics on a branching process

Krzysztof Bartoszek

Department of Mathematics
Uppsala University

Będlewo, 11 June 2015

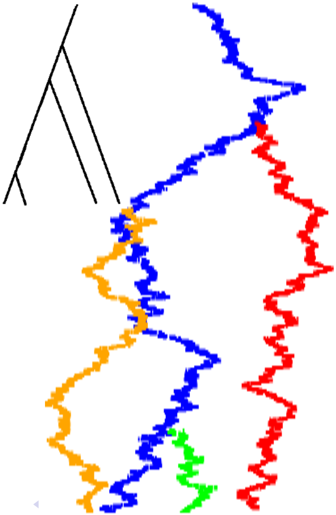
joint work with Serik Sagitov
Mathematical Sciences
CTH, GU

Phylogenetic comparative methods

UPPSALA UNIVERSITET

Introduction:

Euplectes spp. (subfam. Ploceinae = weaverbirds)



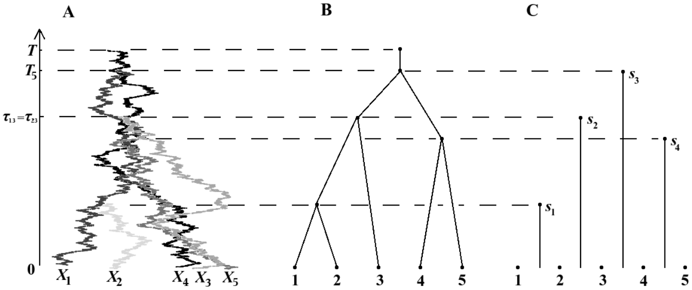
"Tree–free" methods

But what if we do not observe the tree ?

UPPSALA UNIVERSITET

Tree–free PCM

Conditioned branching processes

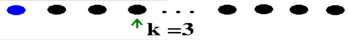


UPPSALA
UNIVERSITET

Tree–free PCM

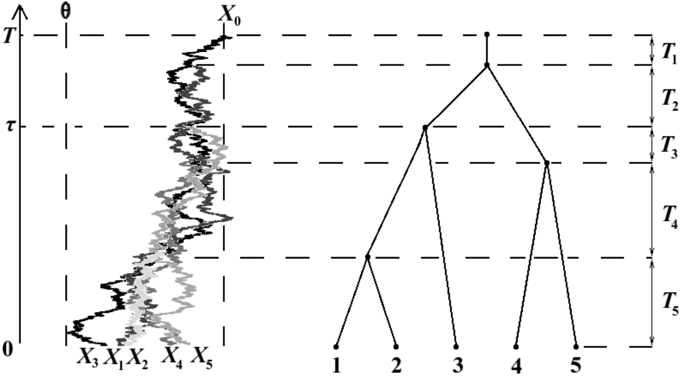
Gernhard (2008), Sagitov and KB (2012)

- Conditional on T all $n-1$ s_is are independent.
- $$q_n(T|n) = n\lambda^n(\lambda-\mu)^2\frac{(1-e^{-(\lambda-\mu)T})^{n-1}e^{-(\lambda-\mu)T}}{(\lambda-\mu e^{-(\lambda-\mu)T})^{n+1}}$$
- $$F(s|T) = \frac{1-\lambda e^{-(\lambda-\mu)s}-\mu e^{-(\lambda-\mu)T}}{(\lambda-\mu e^{-(\lambda-\mu)s})(1-e^{-(\lambda-\mu)T})}\mathbf{1}(s\le T)$$
- $$\begin{aligned}\mathrm{E}\left[T-\tau\right] &= \mathrm{E}\left[T-\int_0^T 1-F^{\kappa}(s|T)\mathrm{d}s\right] \\ &= \sum_{k=1}^{n-1}\frac{n-k}{\binom{n}{2}}\int_0^{\infty}\left(\int_0^T F^k(s|T)\mathrm{d}s\right)q_n(T)\mathrm{d}T\end{aligned}$$



Conditioned Yule tree

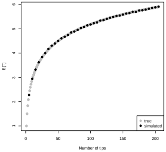
$$T_i \sim exp(\lambda i)$$



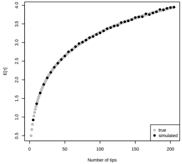
UPPSALA UNIVERSITET

Tree–free PCM

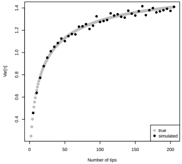
Conditioned Yule tree

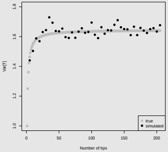


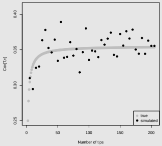
$$H_{n,k} = \sum_{i=1}^{n} \frac{1}{i^k} \quad b_{n,x} = \prod_{i=1}^{n} \frac{i}{i+x}$$



- $\mathrm{E}\left[T\right] = H_{n,1}/\lambda \sim \ln n/\lambda$
- $\mathrm{Var}\left[T\right] = H_{n,2}/\lambda^2 \to \frac{\pi^2}{6\lambda^2}$
- $\mathrm{E}\left[\tau\right] = \frac{n+1}{\lambda(n-1)} H_{n,1} - \frac{2n}{\lambda(n-1)} \sim \ln n/\lambda$
- $\mathrm{Var}\left[\tau\right] = \frac{(n^2-1)H_{n,2} - 2(n+1)H_{n,1}^2 + 4(n+1)H_{n,1} - 4n}{\lambda^2(n-1)^2} \to \frac{\pi^2}{6\lambda^2}$
- $\mathrm{Cov}\left[T,\tau\right] = \frac{1}{\lambda^2(n-1)}(2n-(n+1)H_{n,2}) \to \frac{1}{\lambda^2}(2-\frac{\pi^2}{6})$
- $\mathrm{E}\left[e^{-\gamma T}\right] = b_{n,\gamma}, \quad \mathrm{E}\left[e^{-\gamma \tau}\right] = \frac{2-(n+1)(\gamma+1)b_{n,\gamma}}{(n-1)(\gamma-1)}$







Application: Interspecies correlation ($\lambda = 1$)

How similar do we expect two random species to be?

$$\rho_n = \frac{\mathrm{Cov}\left[X_1, X_2\right]}{\mathrm{Var}\left[X\right]}$$

Brownian motion

$$\mathrm{d}X(t) = \sigma \mathrm{d}B(t) \quad X(0) = X_0$$

$$\mathrm{E}\left[X(t)\right] = X_0 \qquad \mathrm{Var}\left[X(t)\right] = \sigma^2 t$$

$$\rho_n \sim 2(\ln n)^{-1}$$

Ornstein–Uhlenbeck process

$$\mathrm{d}X(t) = -\alpha(X(t) - \theta)\mathrm{d}t + \sigma \mathrm{d}B(t) \quad X(0) = X_0$$

$$\mathrm{E}\left[X(t)\right] = e^{-\alpha t} X_0 + (1 - e^{-\alpha t})\theta \quad \mathrm{Var}\left[X(t)\right] = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t})$$

$$\rho_n \sim \begin{cases} C n^{-2\alpha} & 0 < \alpha < 0.5 \\ 2n^{-1}\ln n & \alpha = 0.5 \\ C n^{-1} & \alpha > 0.5 \end{cases}$$

Application: Classical estimators ($\lambda = 1$)

$$\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i \qquad S_n^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

	$\mathrm{E}\left[\bar{X}_n\right]$	$\mathrm{Var}\left[\bar{X}_n\right]$	$\mathrm{E}\left[S_n^2\right]$	$\mathrm{Var}\left[S_n^2\right]$
$\alpha = 0$	X_0	$2\sigma^2$	$\sigma^2 \ln n + O(1)$	$(\frac{\pi^2}{6} + 1)\sigma^4$
$\alpha \in (0, \frac{1}{2})$	$\theta + O(n^{-\alpha})$	$c_\alpha n^{-2\alpha}$	$\frac{\sigma^2}{2\alpha} + O(n^{-2\alpha})$	$O(n^{-2\alpha})$
$\alpha = \frac{1}{2}$	$\theta + O(n^{-\alpha})$	$2\frac{\ln n}{n}$	$\frac{\sigma^2}{2\alpha} + O(\frac{\ln n}{n})$	$O(\frac{\ln n}{n})$
$\alpha > \frac{1}{2}$	$\theta + O(n^{-\alpha})$	$c_\alpha n^{-1}$	$\frac{\sigma^2}{2\alpha} + O(n^{-2\alpha})$	$O(n^{-1})$

Application: weak convergence

Theorem

Let $\delta = \frac{X_0 - \theta}{\sqrt{\sigma_a^2/2\alpha}}$ *be a normalized difference between the ancestral and optimal values. Consider the normalized sample mean* $\overline{Y}_n = \frac{\overline{X}_n - \theta}{\sqrt{\sigma_a^2/2\alpha}}$ *of the Yule-Ornstein-Uhlenbeck process with* $\overline{Y}_0 = \delta$*. As* $n \to \infty$ *the process* $\overline{Y}_n$ *has the following limit behaviour.*
(i) If $\alpha > 0.5$*, then* $\sqrt{n} \cdot \overline{Y}_n$ *is asymptotically normally distributed with zero mean and variance* $\frac{2\alpha+1}{2\alpha-1}$*.*
(ii) If $\alpha = 0.5$*, then* $\sqrt{n/\ln n} \cdot \overline{Y}_n$ *is asymptotically normally distributed with zero mean and variance* 2*.*
(iii) If $\alpha < 0.5$*, then* $n^{\alpha} \cdot \overline{Y}_n$ *converges a.s. and in* L^2 *to a random variable* $Y_{\alpha,\delta}$ *with* $\mathrm{E}\left[Y_{\alpha,\delta}\right] = \delta\Gamma(1+\alpha)$ *and* $\mathrm{E}\left[Y_{\alpha,\delta}^2\right] = \left(\delta^2 + \frac{4\alpha}{1-2\alpha}\right)\Gamma(1+2\alpha)$*.*

Application: confidence interval

$$\mathrm{E}\left[S_n^2\right] \to \frac{\sigma_a^2}{2\alpha} \qquad \mathrm{Var}\left[S_n^2\right] \to 0$$

allows us to write confidence intervals

for $\alpha > 0.5$ $\quad \overline{X}_n \pm z_{1-x/2} \cdot \sqrt{S_n^2} \cdot \frac{1}{\sqrt{n}} \cdot \sqrt{\frac{2\alpha+1}{2\alpha-1}},$

for $\alpha = 0.5$ $\quad \overline{X}_n \pm z_{1-x/2} \cdot \sqrt{S_n^2} \cdot \frac{\sqrt{2\ln n}}{\sqrt{n}},$

for $\alpha < 0.5$ $\quad (\overline{X}_n - q_{x/2} \cdot \sqrt{S_n^2} \cdot n^{-\alpha},$

$\quad \overline{X}_n + q_{1-x/2} \cdot \sqrt{S_n^2} \cdot n^{-\alpha}).$

Thank you !!!!

- S. Andersson, G. Jones, P. Mostad, B. Oxelman, S. Sagitov, M. Prager (GU, CTH)
- S. Glemin, I. Kaj, M. Lascoux (UU)
- V. Mitov, T. Stadler (ETH Zürich)
- T.F. Hansen, K. Voje (Oslo Univ.)
- P. Liò, H. Xiao (Cambridge Univ.)
- J. Pienaar (Alabama Univ.)

UPPSALA
UNIVERSITET

Tree–free PCM

References

- K.B., J. Pienaar, P. Mostad, S. Andersson, T.F. Hansen. **A phylogenetic comparative method for studying multivariate adaptation.** *J. Theor. Biol.*, 314:204–215, 2012
- S.S., K.B. **Interspecies correlation for neutrally evolving traits.** *J. Theor. Biol.*, 309:11–19, 2012
- K.B., G. Jones, B. Oxelman, S.S. **Time to a single hybridization event in a group of species with unknown ancestral history.** *J. Theor. Biol.*, 322:1–6, 2013
- K.B., S.S. **Phylogenetic confidence intervals for the optimal trait value.** *J. Appl. Probab.*, 52, 2015
- K.B. **Quantifying the effects of anagenetic and cladogenetic evolution.** *Math. Biosc.*, 254:42–57, 2014
- K.B., S.S. **A consistent estimator of the evolutionary rate.** *J. Theor. Biol.*, 371:69–78, 2015