Probabilistic programming for birth-death models of evolution using an alive particle filter with delayed sampling

Jan Kudlicka
Uppsala University
Uppsala, Sweden

Lawrence M. Murray
Uber AI Labs
San Francisco, CA, USA

Fredrik Ronquist
Swedish Museum
of Natural History
Stockholm, Sweden

Thomas B. Schön
Uppsala University
Uppsala, Sweden

SUPPLEMENTARY MATERIAL

A PROOF OF THE UNBIASEDNESS OF THE MARGINAL LIKELIHOOD ESTIMATOR OF THE EXTENDED APF

In this section we prove that the marginal likelihood estimator

$$\widehat{Z} = \prod_{t=1}^{T} \frac{\sum\limits_{n=1}^{N} w_t^{(n)}}{P_t - 1},$$

produced by the extended alive particle filter (APF) for the state space model (Figure 1)

$$\begin{aligned} &x_0 \sim p(x_0), \\ &x_t \sim f_t(x_t|x_{t-1}), \text{ for } t = 1, 2, \ldots, T, \\ &y_t \sim g_t(y_t|x_t), \end{aligned}$$

is unbiased in the sense that $\mathbb{E}[\widehat{Z}] = p(y_{1:T})$.

The structure of our proof is similar to that of Pitt et al. (2012) for the Auxiliary Particle Filter. Let $\mathcal{F}_t = \{x_t^{(n)}, w_t^{(n)}\}_{n=1}^N$ denote the internal state of the particle filter, i.e., the states and weights of all particles, at time t.

Lemma 1.

$$\mathbb{E}\left[\frac{\sum_{n=1}^N w_t^{(n)}}{P_t - 1}\middle|\mathcal{F}_{t-1}\right] = \sum_{n=1}^N \frac{w_{t-1}^{(n)}}{\sum_{m=1}^N w_{t-1}^{(m)}} p\left(y_t\middle|x_{t-1}^{(n)}\right).$$

Proof. In the interest of brevity, we will omit conditioning on $\mathcal{F}_{t-1}$ in the notation. For each particle, the APF constructs a candidate sample x' by drawing a sample from $\{x_{t-1}^{(n)}\}$ with the probabilities proportional to the weights $\{w_{t-1}^{(n)}\}$ and propagating it forward to time t such that

$$x' \sim \sum_{n=1}^N \frac{w_{t-1}^{(n)}}{\sum_{m=1}^N w_{t-1}^{(m)}} f_t\left(x'\middle|x_{t-1}^{(n)}\right).$$

If $g_t(y_t|x') = 0$, the candidate sample is rejected and the procedure is repeated until acceptance (when $g_t(y_t|x') > 0$).

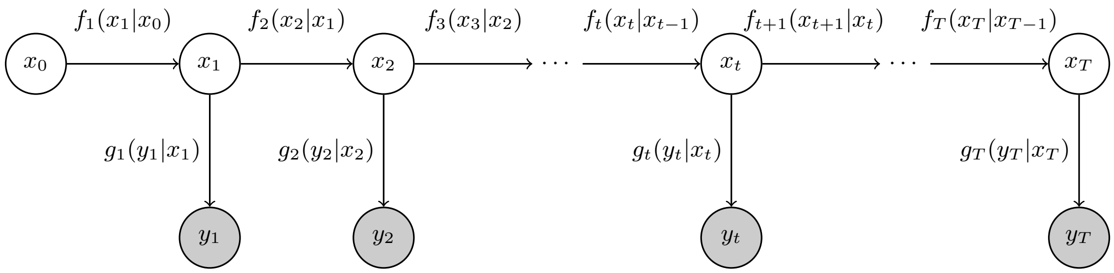

Figure 1: Graphical model of the state space model.

Let $A_t = \{x' : g_t(y_t|x') > 0\}$. The acceptance probability p_{A_t} is then given by

$$p_{A_t} = \int \mathbf{1}_{A_t}(x') \sum_{n=1}^{N} \frac{w_{t-1}^{(n)}}{\sum_{m=1}^{N} w_{t-1}^{(m)}} f_t\left(x' \middle| x_{t-1}^{(n)}\right) dx',$$

where $\mathbf{1}$ denotes the indicator function.

The accepted samples are distributed according to the following distribution:

$$x_t \sim \frac{\mathbf{1}_{A_t}(x_t)}{p_{A_t}} \sum_{n=1}^{N} \frac{w_{t-1}^{(n)}}{\sum_{m=1}^{N} w_{t-1}^{(m)}} f_t\left(x_t \middle| x_{t-1}^{(n)}\right).$$

The expected value of the weight $w_t = g_t(y_t|x_t)$ of an accepted sample is given by

$$\mathbb{E}[w_t] = \int g_t(y_t|x_t) \frac{\mathbf{1}_{A_t}(x_t)}{p_{A_t}} \sum_{n=1}^{N} \frac{w_{t-1}^{(n)}}{\sum_{m=1}^{N} w_{t-1}^{(m)}} f_t\left(x_t \middle| x_{t-1}^{(n)}\right) dx_t.$$

The factor $\mathbf{1}_{A_t}(x_t)$ can be omitted since $\mathbf{1}_{A_t}(x_t) = 0 \Leftrightarrow g_t(y_t|x_t) = 0$, resulting in

$$\begin{aligned}
\mathbb{E}[w_t] &= \int \frac{1}{p_{A_t}} \sum_{n=1}^{N} \frac{w_{t-1}^{(n)}}{\sum_{m=1}^{N} w_{t-1}^{(m)}} f_t\left(x_t \middle| x_{t-1}^{(n)}\right) g_t(y_t|x_t) dx_t \\
&= \frac{1}{p_{A_t}} \sum_{n=1}^{N} \frac{w_{t-1}^{(n)}}{\sum_{m=1}^{N} w_{t-1}^{(m)}} \int f_t\left(x_t \middle| x_{t-1}^{(n)}\right) g_t(y_t|x_t) dx_t \\
&= \frac{1}{p_{A_t}} \sum_{n=1}^{N} \frac{w_{t-1}^{(n)}}{\sum_{m=1}^{N} w_{t-1}^{(m)}} p\left(y_t \middle| x_{t-1}^{(n)}\right).
\end{aligned}$$

The APF repeats drawing new samples until $N+1$ samples have been accepted. The total number of draws of candidate samples at time t, P_t, is itself a random variable distributed according to the negative binomial distribution

$$P(P_t = D) = \binom{D-1}{(N+1)-1} p_{A_t}^{N+1} (1-p_{A_t})^{D-(N+1)}$$

with the support $D \in \{N+1, N+2, N+3, \dots\}$.

Finally, using the fact that $\mathbb{E}[w_t]$ does not depend on the value of P_t,

$$\begin{aligned}
\mathbb{E}\left[\frac{\sum_{n=1}^N w_t^{(n)}}{P_t-1}\right] &= \sum_{D=N+1}^{\infty} \frac{N\mathbb{E}[w_t]}{D-1} P(P_t = D) = \sum_{D=N+1}^{\infty} \frac{N\mathbb{E}[w_t]}{D-1} \binom{D-1}{N} p_{A_t}^{N+1} (1-p_{A_t})^{D-(N+1)} \\
&= N\mathbb{E}[w_t] \sum_{D=N+1}^{\infty} \frac{1}{D-1} \binom{D-1}{N} p_{A_t}^{N+1} (1-p_{A_t})^{D-(N+1)} \\
&= N\mathbb{E}[w_t] \sum_{D=N+1}^{\infty} \frac{1}{D-1} \frac{(D-1)(D-2)!}{N(N-1)!(D-(N+1))!} p_{A_t}^{N+1} (1-p_{A_t})^{D-(N+1)} \\
&= \mathbb{E}[w_t] p_{A_t}^{N+1} \sum_{D=N+1}^{\infty} \binom{D-2}{D-(N+1)} (1-p_{A_t})^{D-(N+1)}
\end{aligned}$$

(using the binomial theorem)

$$\begin{aligned}
&= \mathbb{E}[w_t] p_{A_t}^{N+1} p_{A_t}^{-N} = \frac{1}{p_{A_t}} \sum_{n=1}^N \frac{w_{t-1}^{(n)}}{\sum_{m=1}^N w_{t-1}^{(m)}} \, p\left(y_t \middle| x_{t-1}^{(n)}\right) p_{A_t} \\
&= \sum_{n=1}^N \frac{w_{t-1}^{(n)}}{\sum_{m=1}^N w_{t-1}^{(m)}} \, p\left(y_t \middle| x_{t-1}^{(n)}\right).
\end{aligned}$$

□

Lemma 2.

$$\mathbb{E}\left[\frac{\sum_{n=1}^N w_t^{(n)} p\left(y_{t+1:t'} \middle| x_t^{(n)}\right)}{P_t - 1} \middle| \mathcal{F}_{t-1}\right] = \sum_{n=1}^N \frac{w_{t-1}^{(n)}}{\sum_{m=1}^N w_{t-1}^{(m)}} p\left(y_{t:t'} \middle| x_{t-1}^{(n)}\right).$$

Proof. Similar to the proof of Lemma 1 we have that

$$\begin{aligned}
\mathbb{E}[w_t p(y_{t+1:t'}|x_t)] &= \int \frac{1}{p_{A_t}} \sum_{n=1}^N \frac{w_{t-1}^{(n)}}{\sum_{m=1}^N w_{t-1}^{(m)}} f_t\left(x_t \middle| x_{t-1}^{(n)}\right) g_t(y_t|x_t) p(y_{t+1:t'}|x_t) dx_t \\
&= \frac{1}{p_{A_t}} \sum_{n=1}^N \frac{w_{t-1}^{(n)}}{\sum_{m=1}^N w_{t-1}^{(m)}} \int f_t\left(x_t \middle| x_{t-1}^{(n)}\right) g_t(y_t|x_t) p(y_{t+1:t'}|x_t) dx_t \\
&= \frac{1}{p_{A_t}} \sum_{n=1}^N \frac{w_{t-1}^{(n)}}{\sum_{m=1}^N w_{t-1}^{(m)}} \, p\left(y_{t:t'} \middle| x_{t-1}^{(n)}\right)
\end{aligned}$$

and using this result we have that

$$\begin{aligned}
\mathbb{E}\left[\frac{\sum_{n=1}^N w_t^{(n)} p\left(y_{t+1:t'} \middle| x_t^{(n)}\right)}{P_t - 1}\right] &= \sum_{D=N+1}^{\infty} \frac{N\mathbb{E}[w_t p(y_{t+1:t'}|x_t)]}{D-1} \binom{D-1}{N} p_{A_t}^{N+1} (1-p_{A_t})^{D-(N+1)} \\
&= N\mathbb{E}[w_t p(y_{t+1:t'}|x_t)] \sum_{D=N+1}^{\infty} \frac{1}{D-1} \binom{D-1}{N} p_{A_t}^{N+1} (1-p_{A_t})^{D-(N+1)} \\
&= N\mathbb{E}[w_t p(y_{t+1:t'}|x_t)] \frac{p_{A_t}}{N} = N \frac{1}{p_{A_t}} \sum_{n=1}^N \frac{w_{t-1}^{(n)}}{\sum_{m=1}^N w_{t-1}^{(m)}} \, p\left(y_{t:t'} \middle| x_{t-1}^{(n)}\right) \frac{p_{A_t}}{N} \\
&= \sum_{n=1}^N \frac{w_{t-1}^{(n)}}{\sum_{m=1}^N w_{t-1}^{(m)}} \, p\left(y_{t:t'} \middle| x_{t-1}^{(n)}\right).
\end{aligned}$$

□

Lemma 3.

$$\mathbb{E}\left[\prod_{t'=t-h}^{t} \frac{\sum_{n=1}^{N} w_{t'}^{(n)}}{P_{t'}-1}\middle|\mathcal{F}_{t-h-1}\right] = \sum_{n=1}^{N} \frac{w_{t-h-1}^{(n)}}{\sum_{m=1}^{N} w_{t-h-1}^{(m)}} p\left(y_{t-h:t}\middle|x_{t-h-1}^{(n)}\right).$$

Proof. By induction.

The base step for $h = 0$ was proved in Lemma 1.

In the induction step, let us assume that the equality holds for h and prove it for $h + 1$:

$$\begin{aligned}
\mathbb{E}\left[\prod_{t'=t-h-1}^{t} \frac{\sum_{n=1}^{N} w_{t'}^{(n)}}{P_t'-1}\middle|\mathcal{F}_{t-h-2}\right] &= \mathbb{E}\left[\mathbb{E}\left[\prod_{t'=t-h}^{t} \frac{\sum_{n=1}^{N} w_{t'}^{(n)}}{P_t'-1}\middle|\mathcal{F}_{t-h-1}\right] \frac{\sum_{n=1}^{N} w_{t-h-1}^{(n)}}{P_{t-h-1}-1}\middle|\mathcal{F}_{t-h-2}\right] \\
&\text{(using the induction assumption)} \\
&= \mathbb{E}\left[\sum_{n=1}^{N} \frac{w_{t-h-1}^{(n)}}{\sum_{m=1}^{N} w_{t-h-1}^{(m)}} p\left(y_{t-h:t}\middle|x_{t-h-1}^{(n)}\right) \frac{\sum_{n=1}^{N} w_{t-h-1}^{(n)}}{P_{t-h-1}-1}\middle|\mathcal{F}_{t-h-2}\right] \\
&= \mathbb{E}\left[\sum_{n=1}^{N} \frac{w_{t-h-1}^{(n)}}{P_{t-h-1}-1} p\left(y_{t-h:t}\middle|x_{t-h-1}^{(n)}\right)\middle|\mathcal{F}_{t-h-2}\right] \\
&\text{(using Lemma 2)} \\
&= \sum_{n=1}^{N} \frac{w_{t-h-2}^{(n)}}{\sum_{m=1}^{N} w_{t-h-2}^{(m)}} p\left(y_{t-h-1:t}\middle|x_{t-h-2}^{(n)}\right).
\end{aligned}$$

□

Theorem 1.

$$\mathbb{E}\left[\prod_{t=1}^{T} \frac{\sum\limits_{n=1}^{N} w_t^{(n)}}{P_t - 1}\right] = p(y_{1:T}).$$

Proof. Using Lemma 3 with $t = T, h = T - 1$ and

$$\mathbb{E}\left[\frac{1}{N}\sum_{n=1}^{N} p\left(y_{1:T}\middle|x_0^{(n)}\right)\right] = p(y_{1:T}).$$

□

B GENERATIVE MODEL FOR CRBD

The pseudocode for generating phylogenetic trees using the CRBD model is listed in Algorithm 1.

C RELEVANT CONJUGACY RELATIONSHIPS

C.1 NEGATIVE BINOMIAL AND LOMAX DISTRIBUTION

Negative binomial distribution

Parameters: number of successes $k > 0$ before the experiment is stopped, probability of success $p \in (0, 1)$

Algorithm 1 Pseudocode for generating trees using the CRBD model.

```
function CRBD(τ_orig)
    return (τ_orig, {CRBD'(τ_orig)})

function CRBD'(τ)
    Δ ~ Exponential(λ + μ)
    τ' ← τ − Δ
    if τ' < 0 then
        return (0, ∅)
    e ~ Cat(p_1 = λ/(λ+μ), p_2 = μ/(λ+μ))
    if e = 1 then
        return (τ', {CRBD'(τ'), CRBD'(τ')})
    else
        return (τ', ∅)
```

Probability mass function:

$$f(r|k,p) = \binom{r+k-1}{k-1} p^k (1-p)^r \text{ for } r \in \mathbb{N} \cup \{0\},$$

where r is the number of failures.

Lomax distribution

Parameters: scale $\lambda > 0$, shape $\alpha > 0$

Probability density function:

$$f(\Delta|\lambda,\alpha) = \frac{\alpha}{\lambda}\left(1 + \frac{\Delta}{\lambda}\right)^{-(\alpha+1)} \text{ for } \Delta \geq 0$$

C.2 CONJUGACY RELATIONSHIPS

Gamma-Poisson mixture

Prior distribution: $\nu \sim \text{Gamma}(k,\theta)$ with the probability density function

$$f(\nu|k,\theta) = \frac{1}{\Gamma(k)\theta^k}\nu^{k-1}e^{-\nu/\theta} \text{ for } \nu > 0$$

Likelihood: $n \sim \text{Poisson}(\nu\Delta)$ with the probability mass function

$$f(n|\lambda) = \frac{\lambda^n}{n!}e^{-\lambda} \text{ for } n \in \mathbb{N} \cup \{0\},$$

where $\lambda = \nu\Delta$.

Prior predictive distribution ($k \in \mathbb{N}$):

$$\begin{aligned} f(n|k,\theta) &= \int_0^\infty \frac{1}{\Gamma(k)\theta^k}\nu^{k-1}e^{-\nu/\theta} \times \frac{(\nu\Delta)^n}{n!}e^{-\nu\Delta}d\nu = \frac{\Delta^n}{n!(k-1)!\theta^k}\int_0^\infty \nu^{n+k-1}e^{-\nu(1/\theta+\Delta)}d\nu \\ &= \frac{\Delta^n}{n!(k-1)!\theta^k}\left(\frac{1}{\theta}+\Delta\right)^{-(n+k)}(n+k-1)! = \binom{n+k-1}{k-1}\left(\frac{1}{1+\Delta\theta}\right)^k\left(1-\frac{1}{1+\Delta\theta}\right)^n \end{aligned}$$

$$n|k,\theta \sim \text{NegativeBinomial}\left(k, \frac{1}{1+\Delta\theta}\right)$$

Posterior distribution:

$$f(\nu|n) \propto \frac{1}{\Gamma(k)\theta^k}\nu^{k-1}e^{-\nu/\theta} \times \frac{(\nu\Delta)^n}{n!}e^{-\nu\Delta} \propto \nu^{k+n-1}e^{-\nu(1/\theta+\Delta)} = \nu^{(k+n)-1}e^{-\nu/\left(\frac{\theta}{1+\Delta\theta}\right)}$$

$$\nu|n \sim \mathrm{Gamma}\left(k+n, \frac{\theta}{1+\Delta\theta}\right)$$

Gamma-exponential mixture

Prior distribution: $\nu \sim \mathrm{Gamma}(k, \theta)$

Likelihood: $\Delta \sim \mathrm{Exponential}(\nu)$ with the probability density function

$$f(\Delta|\nu) = \nu e^{-\nu\Delta} \text{ for } \Delta \geq 0$$

Prior predictive distribution:

$$f(\Delta|k,\theta) = \int_0^\infty \frac{1}{\Gamma(k)\theta^k}\nu^{k-1}e^{-\nu/\theta} \times \nu e^{-\nu\Delta}d\nu = \frac{1}{\Gamma(k)\theta^k}\int_0^\infty \nu^k e^{-\nu(1/\theta+\Delta)}d\nu$$

$$= \frac{1}{\Gamma(k)\theta^k}\left(\frac{1}{\theta+\Delta}\right)^{-(k+1)}\Gamma(k+1) = \frac{k}{\theta^k}\left(\frac{1}{\theta}+\Delta\right)^{-(k+1)} = k\theta(1+\Delta\theta)^{-(k+1)}$$

$$\Delta|k,\theta \sim \mathrm{Lomax}\left(\frac{1}{\theta}, k\right)$$

Posterior distribution:

$$f(\nu|\Delta) \propto \frac{1}{\Gamma(k)\theta^k}\nu^{k-1}e^{-\nu/\theta} \times \nu e^{-\nu\Delta} \propto \nu^k e^{-\nu(1/\theta+\Delta)} = \nu^{(k+1)-1}e^{-\nu/\left(\frac{\theta}{1+\Delta\theta}\right)}$$

$$\nu|\Delta \sim \mathrm{Gamma}\left(k+1, \frac{\theta}{1+\Delta\theta}\right)$$

D SOURCE CODE

Birch is available at

`https://birch-lang.org/`

The source code for the CRBD and BiSSE models is available at

`https://github.com/kudlicka/paper-2019-probabilistic`

References

M. K. Pitt, R. dos Santos Silva, P. Giordani, and R. Kohn. On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. *Journal of Econometrics*, 171(2):134–151, 2012.

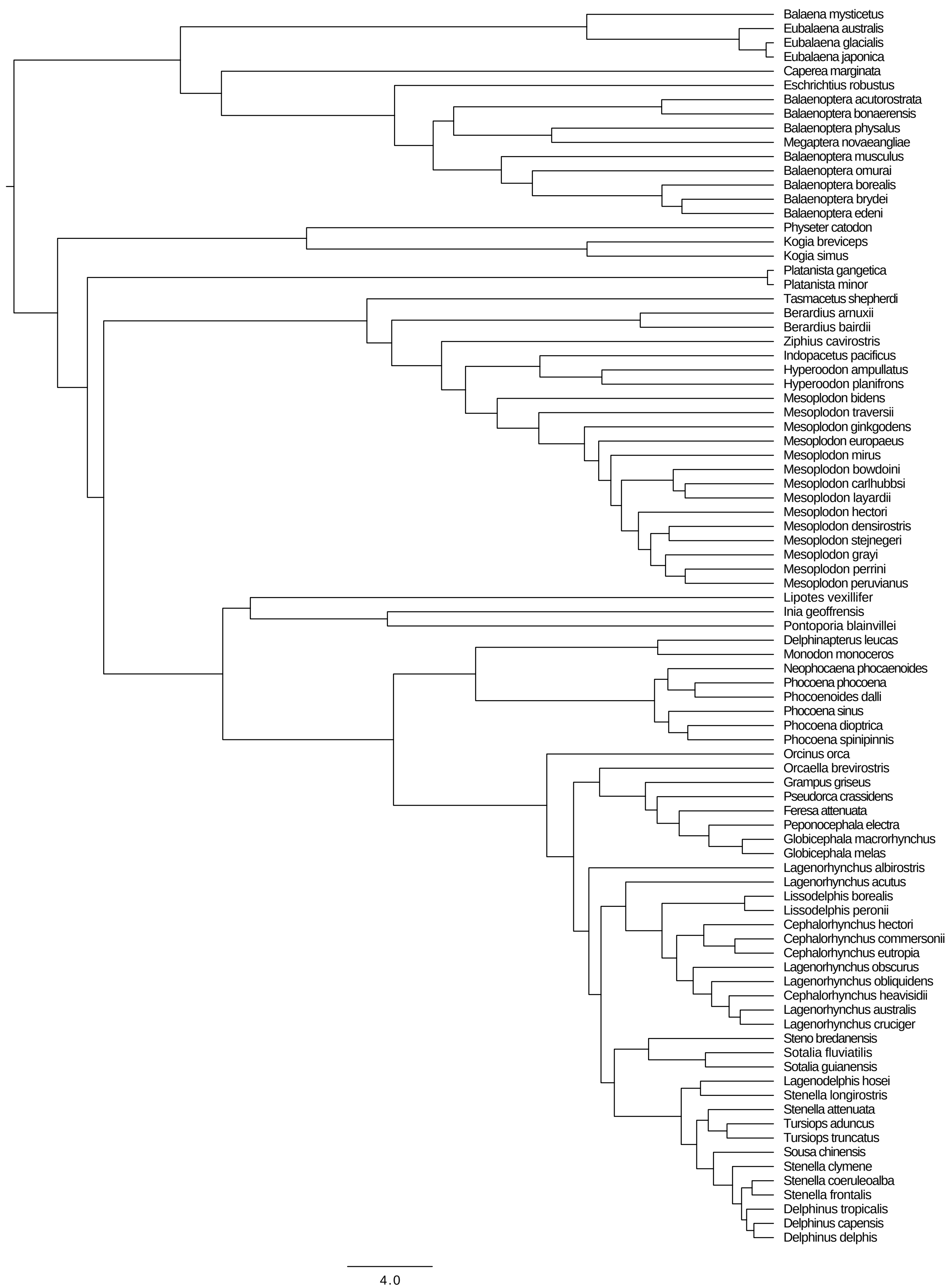

Figure 2: Phylogeny of cetaceans (whales, dolphins and porpoises).