Supplementary online material to

Comment on: "Lévy walks evolve through interaction between movement and environmental complexity"

Vincent A.A. Jansen, Alla Mashanova and Sergei Petrovskii

Details of models used

We used the following distributions to model mussel step length distribution: exponential, power law and truncated power law distributions as in [1] and hyperexponential functions, which are exponential mixtures (with 2, 3 or 4 exponential distributions). Since the power law and truncated power law distributions require specifying the lower truncation value (and since a data set is naturally lower truncated by the resolution of observations) we truncated all the models at the smallest value in the data set to make model comparison fair.

Exponential distribution

The probability density function for the exponential distribution with the left truncation is

$$P(X = x) = \lambda e^{-\lambda(x - x_{min})}$$

from which the maximum likelihood estimate of the parameter λ can be calculated as [2]

$$\lambda_{best} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} X_i - x_{min}}$$

and the inverse cumulative function is

$$P(X \geq x) = e^{-\lambda(x - x_{min})}.$$

It is readily seen that the mean value of the step length is

$$E(X) = x_{min} + \lambda^{-1}.$$

λ^{-1} thus quantifies the spatial scale of the corresponding random walk 'on top' of the truncation distance x_{min}.

Power law distribution

The probability density function for the power law distribution with left truncation is

$$P(X = x) = \frac{\mu - 1}{x_{min}^{1-\mu}} x^{-\mu}$$

with $\mu > 1$, from which the maximum likelihood estimate of the parameter μ can be calculated as [2]

$$\mu_{best} = 1 + \frac{1}{\frac{1}{n}\sum_{i=1}^{n} \log X_i - \log x_{min}}$$

and the inverse cumulative function is

$$P(X \geq x) = \left(\frac{x}{x_{min}}\right)^{1-\mu}.$$

Truncated power law distribution

The probability density function for the power law distribution with both the left and right truncation is

$$P(X = x) = \frac{\mu - 1}{x_{min}^{1-\mu} - x_{max}^{1-\mu}} x^{-\mu}.$$

The log-likelihood function is

$$L = n(\log(\mu - 1) - \log(x_{min}^{1-\mu} - x_{max}^{1-\mu})) - \mu \sum_{i=1}^{n} \log X_i.$$

It can be seen from inspection of the log-likelihood function that if x_{max} is not known and is taken as a parameter the likelihood is maximized at the smallest x_{max} possible which would naturally be the maximum value observed in the data. We chose this is as the upper truncation point in order to give the truncated power law as much statistical support as possible. We did not try to fit any other values for x_{max}.

We maximized the log-likelihood numerically to estimate μ. The inverse cumulative function for the truncated power law distribution is

$$P(X \geq x) = 1 - \frac{x_{min}^{1-\mu} - x^{1-\mu}}{x_{min}^{1-\mu} - x_{max}^{1-\mu}}.$$

Composite Brownian walk as a mixture of 2 exponential distributions

The probability density function is

$$P(X = x) = p\lambda_1 e^{-\lambda_1(x-x_{min})} + (1-p)\lambda_2 e^{-\lambda_2(x-x_{min})}.$$

To find the maximum likelihood estimate of the parameters we maximized the log-likelihood function $L = \sum_{i=1}^{n} \log(P(X_i = x_i))$ numerically. The inverse cumulative function is

$$P(X \geq x) = pe^{-\lambda_1(x-x_{min})} + (1-p)e^{-\lambda_2(x-x_{min})}.$$

Composite Brownian walk as a hyperexponential distribution

The probability density function for a mixture of k exponentials can be described by the hyperexponential function as

$$P(X = x) = \sum_{j=1}^{k} p_j \lambda_j e^{-\lambda_j(x-x_{min})}$$

with $p_k = 1 - \sum_{j=1}^{k-1} p_j$. The log-likelihood function $L = \sum_{i=1}^{n} \log(P(X_i = x_i))$ can be maximized numerically in the same way as for a mixture of 2 exponentials. To ensure $\sum_{j=1}^{k} p_j = 1$ during fitting process we used the log-ratio transformation $y_{j\in\{1,2...k-1\}} = \log(p_j/p_k)$ with the inverse transform $p_{j\in\{1,2...k-1\}} = e^{y_j}/(1 + \sum_{i=1}^{k-1} e^{y_i})$ [3]. The inverse cumulative function is

$$P(X \geq x) = \sum_{j=1}^{k} p_j e^{-\lambda_j(x-x_{min})}.$$

Note that the average value of the hyperexponential distribution is given by a linear combination of the averages of the contributing exponentials:

$E(X) = \sum_{j=1}^{k} p_j(x_{min} + \lambda_j^{-1})$. For the mixture of 3 exponentials that had the best fit we found the averages had the following contributions. When we used the full data set with $x_{min} = 0.21095$mm we found $E(X) = 0.034 * 14.51 + 0.099 * 1.56 + 0.867 * 0.30$ whereas we found for the truncated data set with $x_{min} = 0.02236$mm that $E(X) = 0.063 * 14.10 + 0.21 * 1.41 + 0.73 * 0.44$. Note how the contributing averages (respectively 14.51mm, 1.56mm and 0.30mm versus 14.10mm, 1.41mm and 0.44mm) are similar which shows that the value of the contributing averages is relatively robust against truncation.

Concerns regarding the analysis and presentation of the data in [1]

- The calculation of the upper truncation point as a maximum likelihood estimation is incorrectly executed. The routine used in [1] is based on a formula from [2] for finding the maximum likelihood estimate for the exponent μ, but it seems that in the analysis in [1] the value of the derivative with respect to μ is used to calculate the likelihood. Therefore the routine for the calculation of the upper values does not return the maximum likelihood estimate of the upper truncation point. It is easy to show that the maximum likelihood value for the upper truncation point is the highest observed value, and any deviation from this is therefore incorrect. As the value of the truncation point in [1] appears to differ between different parts of the analysis, and the value is not given in the paper nor in supplementary material, it is not possible to judge how this have affected the results.

- Judging from the R file provided by M. de Jager, and by the form of the curve, the curve plotted in Fig 1B in [1] appears to be a Rayleigh distribution, not an exponential distribution. Fig 1B in [1] therefore appears not to depict what the text in the paper describes. Moreover, the formula for the exponential distribution in the online supplementary material omits the lower truncation point and is unsuitable for use as presented. This last point is inconsequential for [1] as it appears that the exponential distribution has not been used in the analysis in [1] at all.

- Figure 1B in [1] omits a number of data points on the right of the figure. This is not documented or explained.

- De Jager et al. [1] reports that Figure 1B is based on 12,401 data points. Instead, Figures 1A and 1B in [1] appear to depict far fewer points. The data file we received contained 4696 data points. It appears this has been used for Figure 1B in [1].

- When scrutinizing the data that we received and that were apparently used in [1], we noticed that many data points were doubled or quadrupled. It transpired that the data file had been corrupted. We received a data file with uncorrupted data from M. de Jager which contained 3584 data points. When we truncated these data by discarding values below 0.21095 mm (as was done in [1]) in order to obtain data comparable to Fig 1B in [1] only 2029 data points remained.

- The application of AIC in [1] is problematic as it is not applied to a maximum likelihood. In [1] the parameters are estimated using a maximum likelihood estimate as given in [2], but the AIC is calculated based on a different likelihood, i.e. based on the fit of the cumulative distribution. This is not necessarily a maximum likelihood.

- The goodness of fit measure used in [1] appears to assume that the departure from the cumulative distribution is normally distributed with constant variance. This not generally the case for a cumulative distribution. The error is usually not symmetric and the variance is not constant. We therefore question the appropriateness of this method [3].

- The calculation of the AIC weights in [1] is incorrect. It appears from the R file which we received from M. de Jager that the weights are calculated from normalized AIC measures, as opposed to the normalized exponentials of the AIC measures. Therefore the AIC weights in Table 1 in [1] are incorrect.

We reanalyzed both the corrupted and uncorrupted data, using a standard model selection based on AIC, as described in [2]. In all cases the truncated power law formed the best description if compared to the random walk only. Analysis of the data using a broader range of models conclusively shows the the movement pattern in this data set is not a Lévy walk as the hyper-exponential distribution with 3 components gives a much better fit of the data than the power law.

References

[1] De Jager, M. *et al.* (2011) Lévy walks evolve through interaction between movement and environmental complexity. *Science* 33: 1551.

[2] Edwards A.M. *et al.* (2007) Revisiting Lévy flight search patterns of wandering albatrosses, bumblebees and deer. *Nature* 449: 1044.

[3] Bolker B.M. (2008) Ecological models and data in R. Princeton University Press, Princeton.