An application of Stable law regression models to animal movement GPS telemetry data

Dr. Robert Mathenge Mutwiri*
School of Pure and Applied sciences, University of Embu, Embu, Kenya
mutwiri.robert@embuni.ac.ke

Abstract

The potential advantage of stable distribution assumption in modelling ecological disturbances is of animal movement is the central theme of this paper. Studies relating animal movement paths to structured landscape data are particularly lacking despite the obvious importance of such information to understanding animal movement. In this paper we model the heavy tails using the student t regression model, the skewness and the heavy tails with the stable law regression. The new models add substantial flexibility and capabilities, including the ability to incorporate multiple variables. We use a likelihood based approach that utilizes the Fourier Transform technique to evaluate the densities and demonstrate the approach with movement data from five elephant herds (*Africana Loxadonta*). We discuss our results in the context of the current knowledge of animal movement and in particular elephant ecology highlighting potential applications of our approach to the study of wide ranging animals.
Keywords: Stable law regression model; Elephant movement; Habitat types; GPS data; Student-t regression model.

1. Introduction

Regression analysis is one of the most popular methods in ecology and statistics; where most variables of interest such animal movement step lengths are assumed to be normally distributed. However, the normality assumption is not appropriate for many ecological variables, especially animal movement metrics (speed, step lengths) and also, in some cases circular metrics (turn angles) [1]. Animal movement linear metrics are typically heavy tailed and excessively highly peaked around zero. A stable law distribution, whose shape is governed by the stability index parameter α, represent one such alternative. Thus such a stable law distribution is better suited to describing such variables; the normal distribution is a special case of the stable distribution. The four parameter family of stable law distribution is more of a generalization of the central limit theorem than an alternative to normal distribution.
Advances in statistical computation have made it possible to estimate the unconditional stable density as well as incorporate covariates [10]. However, estimates of the stable distribution conditional on a set of explanatory variables in the context of regression framework used by applied researchers poses an overwhelming computational challenge [16]. One of the methods used for evaluating the stable density (the direct numerical integration techniques) is non-trivial and burdensome from a computational perspective [14]. As a consequence, maximum likelihood estimation algorithms based on such an approximations are difficult to implement especially for huge data sets encountered in movement ecology [8]. However, with increasing computational power and efficient algorithms, maximum likelihood estimation and other comparative techniques have been implemented by Nolan and Ojeda [14]. Due to the above mentioned drawbacks, stable distributions are not well explored in movement ecology.
The aim of this paper is to evaluate the application stable regression model to analyse the dependence of animal movement on habitat cover by relaxing the commonly use normal assumption on regression models. In section 2, the stable regression model is developed by assuming the stable distribution assumptions. Finally, section 3 gives an application to elephant movement data, including a comparison between the student t regression model fits.

2. Statistical models

2.1 Stable Paretian regression model

In may practical applications in animal ecology, it is known that animal movement rate can be affected by a number of covariates (explanatory variables) such as the nearest distance to the water point, vegetation cover type, distance to tourist roads, soil topology, seasons, amount of rainfall, temperature and many others [3]. However animal movement data is characterized by skewed and heavy tailed distributions. Thus a model that provides a good fit to movement data will definitely yield more precise estimates of the quantities of interest. Based on the stable distribution assumption, we propose a linear regression type model linking the response y_i and the explanatory variables $X = (x_1, x_2, \dots x_n)$ as

$$y_i = \beta_0 + \sum_{i=1}^{n} \beta_{ij} x_i + \epsilon_i, \quad i = 1, \dots, n \tag{1}$$

where β is a vector of the unknown parameters to be estimated and ϵ_i is the random error term. The notion of stable regression models(SRMs) was developed by [12] for symmetric stable distribution and discussed in detail by McHale and Laycock [13]. In SRMs, the error terms $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ are assumed to be independent identically distributed stable random variables denoted by $\epsilon_i \sim S(\alpha, \beta, \gamma, \mu)$.
DuMouchel [4–6] showed that subject to certain conditions, the maximum likelihood estimates of the parameters of an α of a maximum likelihood estimator. They are asymptotically normal, unbiased and have an asymptotic covariance matrix. McCulloch [12] examines the linear regression model in the context of α-stable distribution paying particular attention to the symmetric case. If we denote the stable density function by $S(\epsilon_i; \alpha, \beta, \gamma, \mu)$ then we may rewrite the density of ϵ_i as

$$S(x; \alpha, \beta, \gamma, \mu) = \frac{1}{\gamma} S\Bigg(\frac{y_i - \sum\limits_{j=1}^{k} x_{ij}\beta_j}{\gamma}, \beta, 1, 0 \Bigg),$$

then using the least squares notation, we may write the normal equations as

$$\beta = (XWX)^{-1} X'Wy \tag{2}$$

where $W = \left(\frac{\Phi(\hat{\epsilon}_i)}{\hat{\epsilon}_i}, \frac{\Phi(\hat{\epsilon}_i)}{\hat{\epsilon}_i}, \dots, \frac{\Phi(\hat{\epsilon}_i)}{\hat{\epsilon}_i} \right)$ is the diagonal matrix and $\epsilon_i = y_i x_{im} - \sum\limits_{i=1}^{n} x_i' \hat{\beta}$. Nolan and Ojeda [14] showed that the evaluation of the likelihood function is made possible by using efficient non-linear optimizers. Maximum likelihood algorithm used in this work are provided by Nolan and Ojeda-Revah [14] within the R package *stable 5.1* which can be obtained commercially from `www.RobustAnalysis.com`. Initial values for $\alpha, \beta, \gamma, \mu, \beta_0, \beta_1, \dots, \beta_p$ can be taken from the fit of the stable distribution model.

2.2 Regression model with t errors

We consider the univariate non-linear regression model where the observations $y = (y_1, \dots, y_n)'$ are independent, y_i having a student t distribution with location parameter μ_i, scale parameter σ and v degrees of freedom. We define a linear regression for y_i by

$$y_i = \beta_0 + \sum_{i=1}^{n} \beta_{ij} x_i + \epsilon_i, \quad i = 1, \dots, n \tag{3}$$

where $\epsilon = (\epsilon_1, \dots, \epsilon_n)'$ is the error vector where the components are independent and identically distributed according to the student t distribution with location zero and scale δ and degrees of freedom v [11]. $X = (x_1, \dots, x_n)'$ is the $n \times k$ matrix of explanatory variables. The least squares estimator of β is

$$\beta = (X'X)^{-1} X'Y$$

3.0 Application: Elephant movement data

3.1 Data description

The telemetry data employed in this study was collected by the South African National parks (SANPARKS). In May 2006, 18 African elephants were fitted with GPS -argos telemetry collars (Telenics). Capturing and handling was done according to the University of KwaZulu Natal animal care regulations. 50,000 GPS points were recorded every 30 minutes during the first three years, across a 19,485 km^2 area after collaring and transmitted to SANPARKS via an Argos satellite uplink every day when the elephant was within network range [2, 15]. Telemetry points collected within the first 24 hours after capturing and those with obvious errors were excluded from the data analysis.

3.2 Vegetation cover types

To determine the effects of various habitat types in the pattern of elephant movement, we extracted the vegetation cover types data of before and after the breakpoints. Land cover types and distances to different landscape features within a spatial resolution of 25 m $\times$ 25 m pixels were obtained from the Kruger national park Land cover database. This database is based on the Thematic mapper sensor on Landsat Earth-resource satellites using data frames recorded between 2006 and 2009 (spectral analysis Inc.2009). Dummy variables of vegetation cover types were created and fitted to a regression model assuming stable distributed error terms. The land cover of Kruger national park (KNP) consist of fourteen vegetation cover types.

3.3 Model formulation

The observations of the response variable $y_1, y_2, \ldots, y_n$ represent the movement rate of five elephant herds derived before and after breakpoint home ranges [see 2, for further details]. The covariate vector x_i is the dummy variables representing the vegetation cover type created from the habitat variable. Due to computational complexity of the stable regression model and lack of rich data set with covariates of elephant herds, we shall demonstrate the results of habitat cover types only in this study. The dummy variables created from vegetation cover type are: Y_i Speed of the animal, X_{11} Nearest distance to the river X_{12} Combo, X_{13} thicket, X_{14} Mixed Combretum/Terminalia sericea woodland, X_{15} Combretum/mopane woodland of Timbavati, X_{16} Acacia welwitschii thickets on Karoo sediments, X_{17} Kumana Sandveld, X_{18} Punda Maria Sandveld on Cave Sandstone, X_{19} Sclerocarya birrea subspecies caffra /Acacia nigrescens savanna, X_{20} Dwarf Acacia nigrescens savanna, X_{21} Bangu Rugged Veld, X_{22} Combretum / Acacia nigrescens Rugged Veld, X_{23} Lebombo South. Now we present the results by fitting the model

$$y_i = \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \ldots + \beta_p X_{23} \tag{4}$$

where the dependent variable y_i speed of elephants follows the stable law distribution or the Student's t distribution for $i = 1 \ldots, 200$. The dependent variable y_i is the speed of elephant before and after breakpoint home ranges obtained as described in [2] and [15]. The MLEs of the model parameters are calculated using the procedure nlm in R statistical software.

4. Results of the analysis

Table 1 lists the MLEs of the parameters for the SRMs and HTRMs models fitted to the current data. The SRMs model involves four extra parameters which gives it more flexibility to fit the elephant movement data. Due to lack of rich data set of animal movement with covariates we investigate only the effects of habitat types as dummy variables. Most of the variables considered here were selected as drivers of movement rates before and after break point analysis of home ranges [see 15]. The fitted SRM indicates that the dummy variables $X_{12}, X_{13}, X_{14}, X_{18}$ and X_{23} are significant at 5% level of significance. The linear regression intercept was, however, significantly less than 1 indicating the ability of our models to predict the movement of the elephant at moderate speed.

Since we have demonstrated that the residuals are non-Gaussian, we will now compare the stable estimates with those obtained from the heavy tailed regression model with Student's t distributed disturbances. The results of Student's t regression model in Table 1 indicates that the vegetation covers combretum, thicket, mixed combretum, Punda maria Sandveld and Lebombo south the elephants slowed down their movement while at Mopane woodland, Acacia welwisitchii, Kumana sandveld, Sclerocarya birrea subspecies, Dwarf Acacia savanna, Bangu rugged and Combretum Acacia rugged increased the movement rates though not significantly. The results of stable law regression model indicates that the vegetation covers combretum, Thicket, Mixed combretum, Punda Maria Sandveld and Lebombo south significantly reduced the movement rates of elephants while Mopane woodland, Acacia Welwisitchii, Kumana Sandveld, Sclerocarya Birrea subspecies, Dwarf acacia Savanna, Bangu Rugged and Combretum Acacia Rugged increased the movement rates

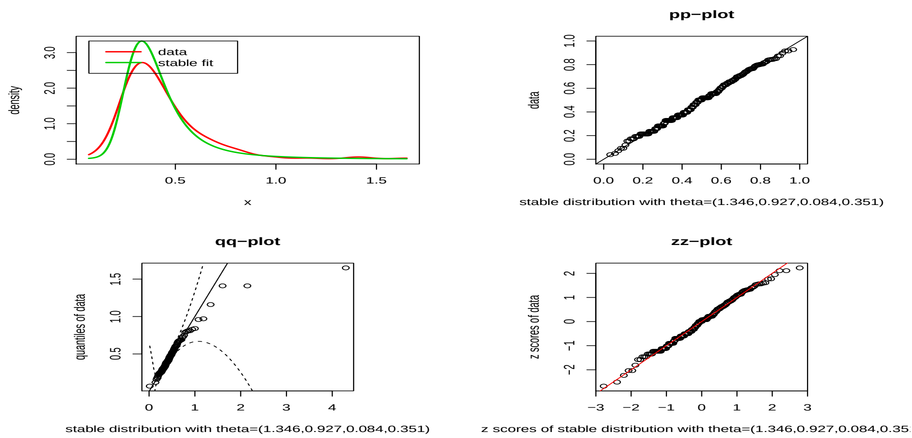
though not significantly.
We note that movement of elephants in the resource poor patches are positive and significant indicating that elephants increased their movement speed when moving in search of food and water while in the resource rich patches they move at a slower speed as they forage. The stability index parameter estimated is 1.31 which is less than 2 with a standard error of 0.0511 indicating that the data is heavy tailed. Clearly we can reject the null hypothesis that the random disturbance follows a Gaussian distribution (the hypothesis $\alpha = 2$) in favour of the alternative that the disturbance follows a non-Gaussian stable distribution with infinite variance. Figure 1, further supports the findings of the fitted model with residuals of the stable distribution plotted along the empirical density of the data. The density plot shows that the empirical distribution has heavier tails and a higher more concentrated peak compared to the Gaussian distribution. These attributes convey the ecological importance of the tails with appropriate statistical assumption. The results of Table 1

Table 1: Summary of Heavy tailed -t distribution and stable law regression model

	Heavy tailed model			Stable law regression model		
parameter	Estimate	Std.err	p-value	Estimate	Std.err	p-value
α				1.309	0.051	1.000
β				0.857	0.081	1.000
γ				0.091	0.003	1.000
intercept	0.376	0.025	1.000	0.34	0.011	0.000
X_{12}	-0.089	0.054	0.000	-0.086	0.024	0.000
X_{13}	-0.114	0.047	0.000	-0.144	0.02	0.000
X_{14}	-0.068	0.034	0.000	-0.053	0.015	0.000
X_{15}	-0.001	0.044	0.000	0.008	0.019	0.663
X_{16}	0.067	0.037	0.9999	0.033	0.016	0.978
X_{17}	0.228	0.058	1.000	0.165	0.026	1.000
X_{18}	-0.047	0.045	0.0202	-0.08	0.02	0.000
X_{19}	0.008	0.034	0.6772	0.008	0.015	0.500
X_{20}	0.064	0.039	0.9999	0.043	0.017	0.994
X_{21}	0.006	0.042	0.6103	-0.006	0.019	0.382
X_{22}	0.021	0.041	0.8508	0.014	0.018	0.773
X_{23}	-0.038	0.047	0.0559	-0.048	0.020	0.009
loglike	88.00481			84.16449		
AIC	-170.4644			-162.7838		

indicates that the student t regression model better fits the data than the stable law regression model with an AIC of -170.46 and -162.78 respectively.
The empirical analysis shows that the effect of vegetation cover types is to reduce the movement rates of elephant in abundant food patches and increased the movement rate in poor resource areas. This finding is consistent with the descriptive analysis by Duffy et al. [3] who hypothesized that the quality and availability of forage suppresses movement rates of elephants. The findings also support Hopcraft et al. [7] who urged that resource rich vegetation cover reduced the movement rates of animals in Serengeti Game reserve in Kenya. The estimated stable regression model sheds more light on the underlying ecological processes which result to differential habitat use. African elephants have large effects on vegetation and high numbers can lead to extensive habitat modifications. Another implication of the stable regression analysis is that the distribution of movement rates-even when conditioning on the vegetation attributes-has infinite variance. This means that the point predictions are useless because they lack precision especial when the stable parameters α and β are at the boundaries.
Confirming the foraging success and measuring the impact of environmental drivers is one of the challenges

Figure 1: Diagnostic analysis of elephant movement data



facing ecologists today. Thus the finding of this paper provides a direct link of inferring the effects of vegetation cover types on elephant movement speed. However, the stable Paretian model does not permit the conditional distribution of movement rates to be quantified and it can be used to make probability statements that can be useful; for example in optimal foraging theory. A potentially important application of the stable analysis is movement strategies analysis. A strategy that includes both the Lévy stable walks and the Lévy flights are thought to optimize foraging. Kawai and Petrovskii [9] show in movement ecology applications that stable models- because they capture both skewness and heavy tails in movement rates- perform considerably well than models based on power law distribution or the empirical distribution. Further, due to the analytical tractability of the stable distributions, it is possible to use the stable models to construct optimal search strategies for animals within the framework of movement ecology. In animal movement studies, where rare steps in the upper tail of the distribution drive search optimality, it appears promising to use the stable regression models developed above as an input into constructing an optimal search strategies for animals that help understand the relationship between elephant herds and their habitats.

5. Conclusion

We have described the theoretical justification for the use of stable law regression models and t regression models in analysing animal movement data. To be useful in practice, a statistical model of the speed of animal movement should capture asymmetry, the heavy tails implied by the importance of extreme events and allow the speed to be conditioned on a vector of explanatory variables. Recent advances in the statistical theory of non-symmetric density functions and their estimation make it feasible to estimate statistical models based on the stable law and the student's t distribution. It is also possible to estimate t regression models using standard maximum likelihood techniques.

Despite several studies detailing analogous statistical approaches, application of such models to GPS tracking is limited due to computational difficulties [8] and lack of adequate data rich in covariates in ecology. The t regression model is particularly appealing in ecology where the data are characterized by heavy tails and where we are interested in conditional distributions. Unlike some other distributions in the Lévy stable family- t models does not account for infinite variance and is not in the domain of attraction of sum of independent and identically distributed random variables. However, the t model is intuitively appealing in that it extends the normal distribution model by permitting tails to be heavy and symmetric. Also, the t model is computationally straightforward and estimable using standard statistical software.

Our empirical application demonstrates the importance of modelling explicitly the asymmetries and heavy tails that characterize animal movement linear metrics (step length or speed) if one is to make the accurate probability statements required to manage the environmental fluctuations. Typically, elephant movement is not predictable as it is difficult to determine analytically when a step starts and ends. However quantifying the distribution of the movement rate conditional on specific-environmental variables is one way to describe the effects of the drivers on the elephant movement. The stable regression models appears to be a useful tool for quantifying this relationship and it may have an important and practical application in assessing the value of artificial incentives in wildlife management especially on private Game ranches in South Africa.

References

[1] Bartumeus, F. (2007). Lévy processes in animal movement: an evolutionary hypothesis. *Fractals*, 15(02):151–162.

[2] Birkett, P. J., Vanak, A. T., Muggeo, V. M. R., Ferreira, S. M., and Slotow, R. (2012). Animal perception of seasonal thresholds: Changes in elephant movement in relation to rainfall patterns. *PLoS ONE*, 7(6):e38363.

[3] Duffy, K. J., Dai, X., Shannon, G., Slotow, R., and Page, B. (2011). Movement patterns of african elephants (*Loxodonta africana*) in different habitat types. *South African Journal of Wildlife Research*, 41(1):21–28.

[4] DuMouchel, W. H. (1971). *Stable distributions in statistical inference*. University Microfilms.

[5] DuMouchel, W. H. (1975). Stable distributions in statistical inference: 2. information from stably distributed samples. *Journal of the American Statistical Association*, 70(350):386–393.

[6] DuMouchel, W. H. et al. (1973). On the asymptotic normality of the maximum-likelihood estimate when sampling from a stable distribution. *The Annals of Statistics*, 1(5):948–957.

[7] Hopcraft, J. G. C., Morales, J. M., Beyer, H. L., Haydon, D. T., Borner, M., Sinclair, A. R., and Olff, H. (2007). Serengeti wildebeest and zebra migrations are affected differently by food resources and predation risks.

[8] Kawai, R. (2012). Continuous-time modeling of random searches: statistical properties and inference. *Journal of Physics A: Mathematical and Theoretical*, 45(23):235004.

[9] Kawai, R. and Petrovskii, S. (2012). Multi-scale properties of random walk models of animal movement: lessons from statistical inference. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 468(2141):1428–1451.

[10] Lambert, P. and Lindsey, J. K. (1999). Analysing financial returns by using regression models based on non-symmetric stable distributions. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 48(3):409–424.

[11] Lange, K. L., Little, R. J. A., and Taylor, J. M. G. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, 84(408):pp. 881–896.

[12] McCulloch, J. H. (1998). Linear regression with stable disturbances. *Adler et al.(1998)*, pages 359–376.

[13] McHale, I. G. and Laycock, P. J. (2006). Applications of a general stable law regression model. *Journal of Applied Statistics*, 33(10):1075–1084.

[14] Nolan, J. P. and Ojeda-Revah, D. (2013). Linear and nonlinear regression with stable errors. *Journal of Econometrics*, 172(2):186–194.

[15] Vanak, A., Thaker, M., and Slotow, R. (2010). Do fences create an edge-effect on the movement patterns of a highly mobile mega-herbivore? *Biological Conservation*, pages 2631–2637.

[16] Walls, W. D. (2005). Modelling heavy tails and skewness in film returns. *Applied Financial Economics*, 15(17):1181–1188.