

SCIENTIFIC REPORTS



OPEN

A measure of the concentration of rare events

Rafael Prieto Curiel & Steven Bishop

Received: 09 May 2016

Accepted: 05 August 2016

Published: 31 August 2016

We introduce here an index, which we call the Rare Event Concentration Coefficient (*RECC*), that is a measure of the dispersion/concentration of events which have a low frequency but tend to have a high level of concentration, such as the number of crimes suffered by a person. The Rare Event Concentration Coefficient is a metric based on a statistical mixture model, with a value closer to zero meaning that events are homogeneously distributed, and a value closer to one meaning that the events have a higher degree of concentration. This measure may be used to compare the concentration of events over different time periods and over different regions. Other traditional approaches for the dispersion/concentration of a variable tend to be blind to structural changes in the pattern of occurrence of rare events. The *RECC* overcomes this issue and we show here two simple applications, first by using the number of burglaries suffered in Netherlands and then by using the number of volcanic eruptions in the world.

In many different practical contexts, being able to determine a measure of the degree of concentration of a variable is particularly useful. For instance, in the case of wealth distribution, the Gini coefficient has been used in many studies in Economical and Political Sciences. The Gini coefficient is a single valued number which works as a summary for the whole distribution and it helps us to determine whether a country is moving into a more egalitarian distribution of income or if its disparity is increasing, thus in a way it is a summary statistic which helps us compare different regions and over different time periods.

Similar comparisons are desirable in alternative contexts, for example, is crime more or less concentrated in certain regions after the introduction of surveillance systems in a city? Are road accidents more spatially dispersed in Paris compared to Frankfurt? Are the number of claims that an insurance company receives from their customers being more concentrated this year? The data for these type of question has two characteristics which make it hard to deal with: it is both rare and also highly concentrated. In all these examples, the majority of the observations are equal to zero, but then, if a particular observation is not zero, then it is likely that the actual number is not small; so for instance, many accidents happen at the same place, leading to an accident black spot.

This phenomenon has been studied in different settings; in the case of crime, for example, although a high number of people or houses, in fact, suffer no crimes, those who suffer crime have an increased risk of suffering subsequent crimes¹ and as a result, the majority of the observations are equal to zero but then, some observations are far away from being zero². Another example comes from the study of human mobility patterns, where it has been studied by tracing the consecutive sightings of nearly half-a-million bank notes and also, by following 100,000 mobile users, that most of the individuals travel only over short distances³, which means that an individual is likely to be found only in a handful of different places and, if a threshold was placed on travel, most journeys would fall below the threshold.

Having a measure of the degree of concentration or dispersion of such events is useful since sometimes interventions (such as a policing strategy in the case of crime, or a road intervention in the case of road accidents) might result in the displacement of such events, rather than a genuine reduction. However, traditional measures of the concentration of a variable (such as the Gini coefficient or the entropy), fail to work as a tool to compare different levels of the concentration or to track structural changes in the way that these events happen, due to their extreme small frequency and their high levels of concentration.

Counting data and its distribution

Here we focus on a counting process, that is, we select a variable that reflects the number of events with a certain property. Let X_i be the number of events that occurred over a fixed time interval, counted over some set $i = 1, 2, \dots, N$, referred to as *individuals*. This could be, for example, the number of burglaries suffered by the i -th household during the period of one year, say, or the number of insurance claims from the i -th customer or postcode. Assuming

University College London, Mathematics Department, London, WC1E 6BT, United Kingdom. Correspondence and requests for materials should be addressed to R.P.C. (email: rafael.prieto.13@ucl.ac.uk)

that having one unit of these events does not affect future probabilities of having any additional units, the number X_i follows a Poisson distribution with rate λ_i . Under this assumption, the number X_i becomes one observation from a Poisson distribution, which means that if X_i is small or even zero, it could be the result of a small rate, but it could also be (with small probability perhaps) the result of a large rate and it was just good luck, or vice-versa in the case that X_i is large. If a person suffered zero crimes last year it does not mean that their rate is zero and they will never suffer crime.

We assume also that X_i is independent to X_j for $i \neq j$, which might be a strong assumption for the particular context under consideration and needs to be fully examined before moving to the following step. In the case of crime, for example, the assumption of independence is perhaps valid only for large populations. Now, if there is a way in which we can collect the N individuals into $k \geq 1$ distinct groups, where group j say, has Q_j individuals (or equivalently, has a relative size $q_j = Q_j/N$), which have the same rate λ_j , with $j = 1, 2, \dots, k$. Each one of the N individuals of the whole population belongs to one and only one group, so that $Q_1 + Q_2 + \dots + Q_k = N$ (or written in terms of the relative size $q_1 + q_2 + \dots + q_k = 1$). To avoid ambiguous definitions, we order the groups by their rate in increasing order, so $\lambda_1 < \lambda_2 < \dots < \lambda_k$. This type of model is known as a mixture model⁴. If we consider a random individual from the population, the distribution of X_i might be expressed as

$$q_1 \text{Pois}(\lambda_1) + q_2 \text{Pois}(\lambda_2) + \dots + q_k \text{Pois}(\lambda_k), \quad (1)$$

which means that the individual is allocated into the j -th group (with probability q_j) and then has a Poisson distribution with the corresponding rate λ_j .

The number of groups k is crucial for the mixture model. An easy (but useless) solution is to assign each individual to a different group, however, solutions with larger numbers of groups are less useful since for each additional group, its size and its rate need to be estimated, so this increases the number of parameters of the model. The (non-parametric) maximum likelihood estimator (*mle* or *npml*) helps us compare between models with different number of groups, k , and to pick the best (in some sense) amongst them⁵ since in our case we have no prior information on the number of groups⁶. Other techniques to estimate the number of groups, using bootstrapping for example, are also available⁷. The model can be easily fitted using the statistical package CAMAN (Computer Assisted Analysis of Mixtures)⁸ in R⁹ by considering the observed X_i , with $i = 1, 2, \dots, N$ ¹⁰.

The results obtained are: an estimate of the number of population groups \hat{k} , the corresponding rate for each group $\hat{\lambda}_j$, so that the collection of the rate of each group can be viewed as a vector $\hat{\lambda}$, and the relative size of each population group \hat{q}_j , also expressed as a vector as \hat{q} . A goodness of fit test can help us accept or reject the distribution obtained¹⁰. A similar procedure using a mixture model has been used in different scenarios⁴, such as road accidents, mapping hepatitis B in Berlin¹¹ and many more examples in epidemiology⁵.

Two special cases are interesting from the mixture model. First, if $\hat{k} = 1$ then this means that the best way to explain the observations is simply as a Poisson process with rate $\hat{\lambda}_1$, which is a homogeneous distribution over the whole population. The second case is when $\hat{k} = 2$ and $\hat{\lambda}_1 = 0$, which means that the population can be divided into two groups, the first group has a rate equal to zero while the other group has a non-zero rate, which is a model known also as a Zero-Inflated Poisson Model⁴. Both scenarios, the homogeneous distribution and the Zero-Inflated Poisson Model, might be the result obtained from the mixture model.

The distribution of the rates $(\hat{q}, \hat{\lambda})$ is powerful by itself since we can use these to simulate different observations under that distribution so that we can understand the natural departures from the distribution. However, the Rare Event Concentration Coefficient works as a summary statistic based on that distribution.

Concentration metric

The Rare Event Concentration Coefficient (*RECC*) is defined in terms of the distribution of the rates $(\hat{q}, \hat{\lambda})$ given by

$$RECC = \frac{1}{2 \sum_{i=1}^{\hat{k}} \hat{\lambda}_i \hat{q}_i} \sum_{i=1}^{\hat{k}} \sum_{j=1}^{\hat{k}} \hat{q}_i \hat{q}_j |\hat{\lambda}_i - \hat{\lambda}_j|, \quad (2)$$

which is the Gini coefficient applied to the distribution of the rates. The Lorenz curve¹² and the Gini coefficient¹³ of a distribution are often used as a measure of the concentration or dispersion of a variable, and so here we apply them to the mixture model. It is important to note that it is not the Gini coefficient computed directly from the observations X_i , but rather the Gini coefficient of the distribution of the rates $(\hat{q}, \hat{\lambda})$. A value of the Gini coefficient closer to zero is interpreted as the process being more homogeneously distributed across the population, and a value closer to one means that the process is more concentrated in some population groups.

The Lorenz curve and the corresponding Gini coefficient of the distribution of the individual rates are comparable between different time periods and over different regions, even in the case in which the number of individuals changes from one region to the other, or the total number of events of the process changes. With this simple tool, we can compare the rates of processes in which there is randomness involved, and we can determine a useful metric for the concentration of events which are rare and tend to be highly concentrated.

Relevant scenarios

Two relevant cases might be obtained from the *RECC*. The first scenario, if the *RECC* = 0 then this means that the process is homogeneously distributed across the entire population so that every individual has the same rate $\hat{\lambda}_1$. This scenario might happen even when the individuals have different observations X_i , since here we consider the distribution of the rates and not the actual numbers X_i .

The second scenario is the case when from data obtained is a Zero-Inflated Poisson Model ($\hat{k} = 2$ and $\hat{\lambda}_1 = 0$). In such a case, the Rare Event Concentration Coefficient gives $RECC = \hat{q}_1$, the relative size of the group which has a zero rate.

Demonstration

Here, to demonstrate its use, we consider the number of burglaries suffered per household, obtained from a victimisation survey called the 1993 Police Monitor in Netherlands. The data has been used before as a test bed for analysis and to explore the level of concentration of burglary in that country and it contains a discussion on how data was obtained¹⁴. The number of houses that suffered 0, 1, 2 or more crimes is displayed in Table 1.

Burglaries are highly concentrated. From the population surveyed, 91.9% suffered no burglaries, but then from the houses that in fact suffered a burglary, 21% suffered more than one. The reasons why a particular house suffers more crime than others has been studied before (refs 1,15–17).

The mixture model applied to the number of burglaries suffered in the Netherlands gives as results $\hat{k} = 5$, with sizes and rates displayed in Table 2.

The results of the mixture model applied to this data show that the population can be divided into five groups, the largest one (58.2% of the population) suffers no crime ($\hat{\lambda}_1 = 0$), the second largest one (24.3%) suffers a rate of $\hat{\lambda}_2 = 0.11$, so that among any 9 houses of that group they expect to have experienced a burglary, and so on. There is a group (less than 3.3 out of 10,000 houses) which experiences a rate of almost 8 crimes per year ($\hat{\lambda}_5 = 7.93$).

The Rare Event Concentration Coefficient gives $RECC = 0.7643$ and the Lorenz curve of the observed number of crimes and the estimated rate is displayed in Fig. 1 for comparison purposes. It is relevant to notice that the distribution of the rates λ is much more evenly distributed than the actual crime and this is usually the case since, for example, the mixture model tells us that 24.3% of the population (Group 2) suffers a rate of $\hat{\lambda}_2 = 0.11$, so that within the population from that group we expect only one victim from every 9 houses. In that group, the observed burglaries are highly uneven (for each house victimised there are eight houses not victimised), but the rates are uniformly distributed. The main element here is that the events considered are rare, so most of the observations (nearly 92% in the example from Netherlands) are zero, but it does not mean that their rate is zero.

If for some reason, the number of households which suffer 6 or more burglaries drops to zero (a change in only 17 out of the 39,849 observations) the $RECC$ would be 0.7092, which means a difference of 0.0551. With the traditional approach to the concentration of a variable, if we compute the Gini coefficient directly to the number of crimes that each household suffered its value is 0.9362 and again, if for some reason the number of households which suffer 6 or more burglaries drops to zero, the Gini coefficient of the number of crimes suffered would be 0.9351, which only means a difference of 0.0011; an almost negligible change.

Naturally, if the number of households that suffer 6 or more crimes drops to zero, the change in terms of the number of crimes might not be significant, but it is relevant in terms of the structural change in the way that crime is suffered, and it is a change that would not be detected by the traditional Gini coefficient. However, our new metric would allow us to detect such change, even in the case when it occurred for such a small population group.

Volcanic Eruptions

Another application of the $RECC$ comes from the study of volcanic eruptions. Information about the location the 1,532 different volcanoes in the world and their eruptions is available¹⁸ and here we consider the number of confirmed eruptions for each volcano between 1966 and 2015 (50 years of confirmed eruptions), giving us a total of 1,746 eruptions.

Are volcanic eruptions a rare and concentrated event? In our context, out of the 1,532 different volcanoes, only 315 (around 21%) had an eruption in the last 50 years, yet, those volcanoes which had an eruption in the past 50 years, had on average 5.5 eruptions, meaning that volcanic eruptions are relatively rare and highly concentrated. The number of volcanoes which had 0, 1, 2 or more eruptions between 1966 and 2015 is displayed in Table 3.

Results of the mixture model applied to the volcanic eruptions gives a total of $\hat{k} = 6$ groups, so that the 1,532 volcanoes are grouped in an optimal way into 6 groups; the first one has an eruption rate of $\hat{\lambda}_1 = 0$ and a relative size $\hat{q}_1 = 49.9\%$, so that nearly half of the volcanoes are not expected to have an eruption. The second group has an eruption rate of $\hat{\lambda}_2 = 0.17$ and a relative size $\hat{q}_2 = 34.2\%$, which means that nearly one-third of the volcanoes expect to have an eruption every 287 years. The group with the highest eruption rate has an eruption rate of $\hat{\lambda}_6 = 36.2$ with a relative size of $\hat{q}_6 = 0.3\%$, meaning that volcanoes within that group expect to have an eruption every 16.6 months. For volcanic eruptions, the $RECC = 0.883$.

The distribution of volcanoes throughout the world is highly similar to the positioning of the major tectonic belts and so many of the major volcanoes are clustered¹⁹, as shown in Fig. 2. For example, the three most active volcanoes during the past 50 years were Etna (with 43 eruptions), in Sicily, Italy; Bezymianny (with 37 eruptions) and Klyuchevskoy (also with 37 eruptions), both in Kamchatka, Russia. Additionally, volcanoes include a variety of cones and craters and some features are destroyed by continuing eruptions¹⁸, which raises the question of how we deal with observations that might be highly correlated? For example, Bezymianny and Klyuchevskoy are located at a distance of 9.7 kilometres and so in that small region, there was a total of 74 volcanic eruptions in the past 50 years.

Clustering volcanoes which are at a distance smaller than 10 kilometres into *volcanic regions* allows us to deal with the problem of correlated observations. By considering volcanic regions, so that Bezymianny and Klyuchevskoy in Kamchatka fall into a single region, instead of the 1,532 volcanoes, we obtain 1,439 regions with volcanoes, and by taking into account the number of eruptions from each region we can compute the mixture model and the corresponding $RECC$. By following this procedure, the number of regions changes, the largest X_i

Number of burglaries	0	1	2	3	4	5	6	7	8	9	10	11+	Total
Number of houses	36,632	2,548	448	134	51	19	8	2	1	2	2	2	39,849
Frequency (%)	91.9	6.4	1.1	0.3	0.1	0	0	0	0	0	0	0	100

Table 1. Observed number of burglaries in Netherlands, 1993.

Group	1	2	3	4	5
Crime rate $\hat{\lambda}_i$	0	0.11	0.30	1.79	7.93
Relative size \hat{q}_i (%)	58.2	24.3	15.8	1.7	0
Number of houses	23,188	9,679	6,282	688	12

Table 2. Results of the mixture model applied to the burglaries suffered in the Netherlands. $RECC = 0.7643$.

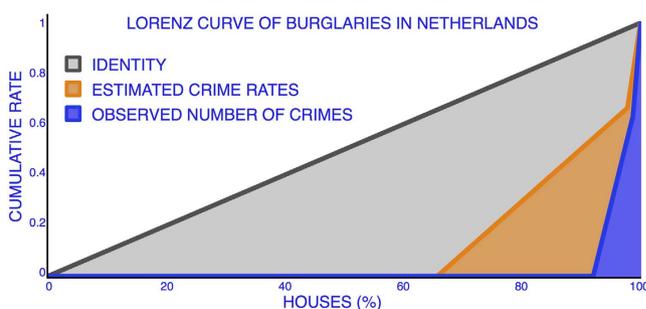


Figure 1. Lorenz curve of the individual crime rates and the number of crimes for Netherlands in 1993.

changes (from 43 eruptions of Mount Etna to 74 eruptions in the Kamchatka region) and the mixture model also changes. However, when we consider the 10 kilometre regions, the $RECC$ changes from 0.883 to 0.879 and even to a value of 0.870 when clustering volcanoes into the considerably large regions with a radius of 20 kilometres.

By grouping observations which have a potential statistical dependence based on a physical attribute, such as nearby volcanoes or crimes separated in space within 200 metres²⁰, or by taking into account burglaries that occur in the same block or neighbourhood, we obtain groups/regions for which the assumption of independence is fairly reasonable. Thus, the $RECC$ is considerably stable when correlated observations are grouped based on a physical attribute.

Human Mobility Patterns

Another area in which the $RECC$ might be useful is in the study of Human Mobility Patterns. As it was studied, heterogeneity suggests that individuals might move following a Lévy flight³. Different research scenarios have been used, for example by following a large number of mobile users and by recording their position each they interact with his or her mobile or recording the position every given time period. The phone towers divide the region into a Voronoi lattice and the data set provides the closest tower to a user so that the location is only recorded by the nearest tower routing which provides the communication service.

Now, the number of times that a particular mobile user is recorded inside a tower vicinity gives us an ideal setting for our study. It is reported³, for example, that from 186 measurements taken from a user, he or she was found inside only in 12 different tower vicinities, from which 96 (51.6%) and 67 (36.0%) occasions happened in the two most preferred locations. The pattern of that person shows that nearly 90% of his or her time is spent in two locations and their neighbouring regions, most likely his or her house and his or her office. In a similar study, some users were found who visit a much higher number of different vicinities²¹, and the frequency in which they move through different vicinities let us determine their mobility patterns.

By counting the number of times that a user is recorded in different tower vicinities, allows us to compare different mobility patterns that users might have. The $RECC$ of the tower vicinities counts of different users gives us a way to compare their levels of mobility and for example, a smaller $RECC$ implies that a user has a higher degree of mobility that a person who has a larger $RECC$. A larger $RECC$ indicates that the person tends to move on a day-to-day basis only through a small number of neighbours of their home city. In terms of human mobility, the $RECC$ takes in account the (potentially) highly concentrated nature of the regions in which a person moves, but also a random component which might motivate a person to visit places which he or she does not regularly visit.

Discussion

The Rare Event Concentration Coefficient $RECC$ based on the mixture model help us compare the concentration rate of events which are not frequent and tend to be highly concentrated by taking into account the random

Eruptions	0	1 to 8	9 to 16	17 to 24	25+
Volcanoes	1,217	249	39	20	7
(%)	79.4	16.3	2.5	1.3	0.5

Table 3. Volcanic eruptions in the world between 1966 and 2015.

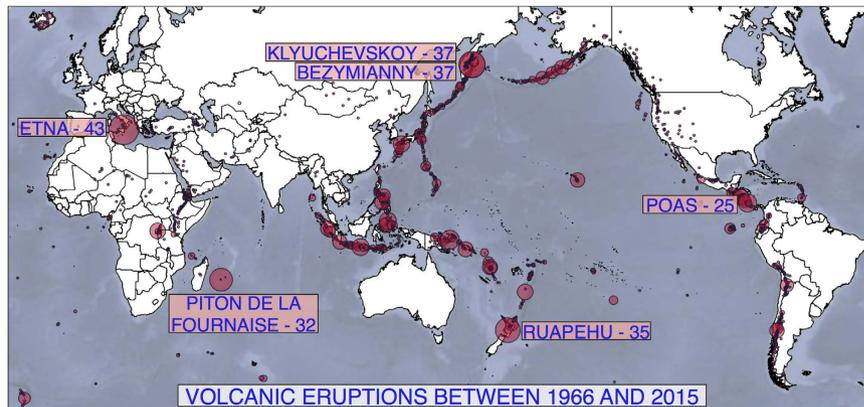


Figure 2. Distribution of the volcanic eruptions between 1966 and 2015. The size represents the number of eruptions. Figure made with Natural Earth, free vector and raster map data naturalearthdata.com (downloaded in August 2015), data from the Global Volcanism Program¹⁹ (downloaded in June 2016) and R (refs 9, 22 and 23).

nature of such events. Other measurements which are traditionally used for the concentration/dispersion are meaningless since they do not detect structural changes of the process or they cannot be used to compare different regions or time intervals.

Using the example of the number of burglaries in Netherlands, we see that the Gini coefficient directly applied to the number of crimes does not change significantly when the distribution changes, which is precisely its main weakness.

The Rare Event Concentration Coefficient *RECC* is easy to compute and provides a summary statistic which is comparable and also, measurements which help us detect structural changes in the dispersion of rare and highly-concentrated events, such as crime, road accidents or human mobility.

Extensions. The *RECC* is designed for rare events, so, in general we observe many zeros which do not ensure that the rate of the individuals is zero, so if a person, for example, suffered zero crimes last year it does not mean that their rate is equal to zero, and so we assume the simplest possible model, which is a mixture model based on a Poisson distribution.

If events are not as rare, then it is possible to estimate the individual rates using a different technique than the mixture model, which considers information at its full and that tries to mimic the underlying pattern. For example, we can consider the rates at which underground stations serve their users, which is best modelled taking into account the hour and the day of the week, or even the number of mobile users within the nearest routing tower vicinity, which might be modelled taking into account the time and space. More sophisticated models for a counting process can also be considered, for example, gang shootings may incite retaliation from rival gangs, and an earthquake increases the chances of a second earthquake, causing in both cases a self-exciting process²⁰. In that case, since it is not a rare event, estimating the individual rates, either as a function of time, space and/or past events, gives a much better approximation to reality. Thus, the Event Concentration Coefficient (*ECC*) can be constructed simply by computing the Gini coefficient of the individual rates, even in the case in which they were estimated using a different model. The resulting metric provides, as in the case of the *RECC* a number between zero and one which reflects the level of concentration of such events.

References

- Grove, L., Farrell, G., Farrington, D. & Johnson, S. *Preventing Repeat Victimization: A Systematic Review* (The Swedish National Council for Crime Prevention, Stockholm, Sweden, 2012), first edition edn.
- Johnson, S. A brief history of the analysis of crime concentration. *European Journal of Applied Mathematics* **21**, 349–370 (2010).
- Gonzalez, M. C., Hidalgo, C. A. & Barabási, A.-L. Understanding individual human mobility patterns. *Nature* **453**, 779–782 (2008).
- Böhning, D. Zero-inflated poisson models and ca man: A tutorial collection of evidence. *Biometrical Journal* **40**, 833–843 (1998).
- Böhning, D., Dietz, E. & Schlattmann, P. Recent developments in computer-assisted analysis of mixtures. *Biometrics* 525–536 (1998).
- McLachlan, G. & Peel, D. *Finite mixture models* (John Wiley & Sons, 2004).
- Schlattmann, P. On bootstrapping the number of components in finite mixtures of Poisson distributions. *Statistics and Computing* **15**, 179–188 (2005).
- Schlattmann, P., Hoehne, J. & Verba, M. *CAMAN: Finite Mixture Models and Meta-Analysis Tools - Based on C.A.MAN* (2015). URL <http://CRAN.R-project.org/package=CAMAN>. R package version 0.73.

9. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2014). URL <http://www.R-project.org/>.
10. Böhning, D., Schlattmann, P. & Lindsay, B. Computer-assisted analysis of mixtures (CA MAN): Statistical algorithms. *Biometrics* 283–303 (1992).
11. Schlattmann, P. & Böhning, D. Mixture models and disease mapping. *Statistics in Medicine* 12, 1943–1950 (1993).
12. Marsh, C. & Elliott, J. *Exploring data: an introduction to data analysis for social scientists* (Polity, 2008).
13. Dorfman, R. A formula for the Gini coefficient. *The Review of Economics and Statistics* 146–149 (1979).
14. Tseloni, A., Wittebrood, K., Farrell, G. & Pease, K. Burglary victimization in England and Wales, the United States and the Netherlands a cross-national comparative test of routine activities and lifestyle theories. *British Journal of Criminology* 44, 66–91 (2004).
15. Pease, K. *Repeat victimisation: Taking stock* (Home Office Police Research Group London, 1998).
16. Farrell, G. Crime concentration theory. *Crime Prevention & Community Safety* 17, 233–248 (2015).
17. Farrell, G. & Pease, K. Once bitten, twice bitten: Repeat victimization and its implications for crime prevention. *Police Research Group, Crime Prevention Unit Paper* 46, 1–32 (1993).
18. Global Volcanism Program. *Volcanoes of the World*, v. 4.5.0. Venzke, E. (ed). Smithsonian Institution. URL <http://dx.doi.org/10.5479/si.GVP.VOTW4-2013>. Accessed on June 2016 (2013).
19. Peterson, D. W. Volcanoes: tectonic setting and impact on society. *Active Tectonics, Geophysics Study Committee, National Research Council*. National Academy Press, Washington DC, pages 231–246 (1986).
20. Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P. & Tita, G. E. Self-exciting point process modeling of crime. *Journal of the American Statistical Association* (2012).
21. Song, C., Qu, Z., Blumm, N. & Barabási, A.-L. Limits of predictability in human mobility. *Science* 327, 1018–1021 (2010).
22. Hijmans, R. J. *Raster: Geographic Data Analysis and Modeling* (2016). URL <https://CRAN.R-project.org/package=raster>. R package version 2.5–8.
23. Loecher, M. & Ropkins, K. RgoogleMaps and loa: Unleashing R graphics power on map tiles. *Journal of Statistical Software* 63, 1–18 (2015). URL <http://www.jstatsoft.org/v63/i04/>.

Acknowledgements

RPC acknowledges the support the government of Mexico via a Conacyt Scholar. SRB acknowledges the support by the Cimplex project funded by the European Commission in the area “FET Proactive: Global Systems Science” (GSS), as a Research and Innovation Action, under the H2020 Framework programme, Grant agreement number 641191.

Author Contributions

Both authors worked on the design of the metrics and writing the manuscript.

Additional Information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Prieto Curiel, R. and Bishop, S. A measure of the concentration of rare events. *Sci. Rep.* 6, 32369; doi: 10.1038/srep32369 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016