Power laws govern the abundance distribution of birds by rank

Sergio Da Silva[abc], Raul Matsushita[def]

[a]Department of Economics, Federal University of Santa Catarina, 88049-970, Florianopolis S.C., Brazil
[b]Graduate Program in Economics, Federal University of Santa Catarina, 88049-970, Florianopolis S.C., Brazil
[c]Graduate Program in Economics, Federal University of Espirito Santo, 29075-910, Vitoria E.S., Brazil
[d]Department of Statistics, University of Brasilia, 70910-900, Brasilia, D.F., Brazil
[e]Graduate Program in Statistics, University of Brasilia, 70910-900, Brasilia D.F., Brazil
[f]Graduate Program in Business Administration, University of Brasilia, 70910-900, Brasilia D.F., Brazil

ORCID: SDS, 0000-0002-7987-4098; RM, 0000-0001-8864-6356.

Abstract
Only a few bird species are abundant. According to new data, a log left-skewed distribution, rather than a lognormal distribution, better adjusts to the abundance distribution of bird species. We look at the rank abundance distribution rather than the species abundance distribution and find three power laws that fit the tails using the same data.

Keywords
Birds, Rank abundance distribution, Species abundance distribution, Power-law distribution.

Data accessibility.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8166167/bin/pnas.2023170118.sd01.xlsx.
Authors' contributions. All authors discussed the ideas presented and contributed to writing the manuscript.
Competing interests. We declare we have no competing interests.
Funding. This work was supported by CNPq, Capes, FAP-DF, and DPI-DPG UnB.

Introduction

One of the few universal patterns in ecology is the distribution of species abundance (Morlon et al. 2009). The fact that some species are common but many more are rare is a crucial aspect of life's diversity (Enquist et al. 2019). The species abundance distribution has a consistent general form, with many rare species coexisting with a few abundant species. The lognormal distribution is widely accepted in the ecology literature for the distribution of species abundance (Chisholm 2007). A lognormal distribution emerges when the number of individuals or another measure of abundance is plotted in a log scale, usually of base 2.

Callaghan et al. (2021) used the most recent influx of citizen science data to estimate the number of individuals (abundance) for 9,700 bird species, accounting for roughly 92 percent of all birds. They merged counts from the app eBird (https://ebird.org/home), which allows users to record bird sightings, with data from 724 well-studied species. They then utilized an algorithm to extrapolate sample estimates. As a result, they discovered a large number of species with small populations confined in niche habitats, as well as a limited number of species spread throughout a large area.

Callaghan et al. (2021) note that few of the species in their data are extremely abundant, and most have low population estimates. However, they also note that there would be about 200 more bird species than anticipated if the abundance of bird species followed a true lognormal distribution. They, therefore, propose a log left-skewed distribution.

Estimates of bird species abundance are critical for ecology, evolutionary biology, and conservation, and progress in quantifying abundance is welcome. Rather than the distribution of bird species abundance, we are interested in the distribution of bird rank abundance using the same data as Callaghan et al. (2021). We show that power laws fit the data well.

Materials and methods

The data is depicted in Figure 1a. There are four undomesticated species with billions of individuals among the 50 billion birds: house sparrows (*Passer domesticus*), European starlings (*Sturnus vulgaris*), ring-billed gulls (*Larus delawarensis*), and barn swallows (*Hirundo rustica*). In comparison, each of the 5,022 species has fewer than 500,000 birds. Figure 1a shows a glimpse of the power-law pattern we discover and immediately draws attention to the group of 1,178 species that appear to deviate from the pattern. The abundance of a species is shaped by a variety of ecological processes (McGill et al. 2007). Many rarer species may have evolved to colonize a single island, for example, but human activity like deforestation can also explain the pattern deviation in Figure 1a. Some species may be rare as a result of our interference. From the standpoint of conservation initiatives, this issue needs to be looked into further.

A power law governs a quantity when the probability of receiving a given value varies inversely with its power (Newman 2005). Thus, a power law is a relationship between two quantities in which a change in one causes a proportional change in the other. This holds true regardless of the starting values. In the dynamics of hierarchical systems that arise in physics, economics, and biology, power laws are common. A power law appears geometrically as a straight line in a plot of log of rank and log of the quantity at hand. After ranking the bird species from top to bottom, three power laws emerge. Figure 1b is the result of plotting the log of rank against the log of bird

abundance. The power laws in Figure 1b can be quantified by calculating their Pareto exponents in relation to the straight line slopes.

One common approach is to use the survival function, $S(x)$, in a Pareto type I model (Jenkins 2017). Those bird species whose abundance exceeds x – that is, one minus the cumulative distribution function $P(x)$ – are given by

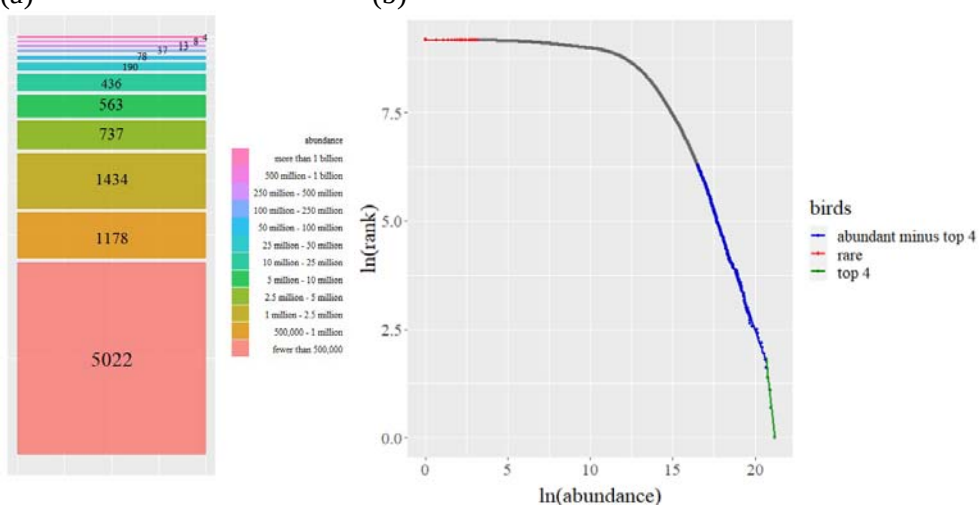
$$S(x) = 1 - P(x) = \left(\frac{x}{x_m}\right)^{-\alpha} \quad , \tag{1}$$

where x_m is the lower bound on abundance, and $x \geq x_m > 0$. The Pareto exponent (tail index) is a shape parameter α that describes the heaviness of the right tail, with smaller values indicating a heavier tail.

Results

We can estimate the three Pareto exponents associated with the rare and abundant species by ordinary least squares regressions of log of the survivor function on log of abundance and a constant term (Jenkins 2017). These result in straight-line intervals of slope $-\alpha$ (Figure 1b).

(a) (b)



(c)

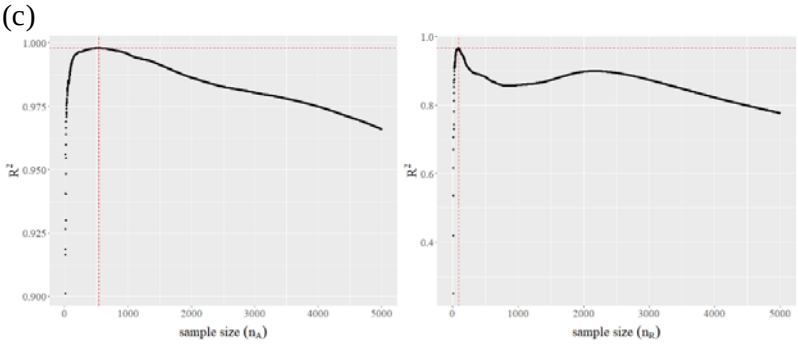


Figure 1 (a) Only a few bird species are abundant. (b) Three power laws emerge when we take ln(rank) versus ln(abundance). (c) R^2 values versus sample size for the most abundant n_A species (apart from the top four). The crossing of the dotted lines indicates the optimal sample size ($n_A = 534$) that provides a maximum $R^2 = 0.998$ (left). R^2 values versus sample size for the rarer n_R species; the optimal sample size is $n_R = 89$ with a maximum $R^2 = 0.966$ (right).

To determine the group of the most abundant n_A species (aside from the top four), we run linear regressions varying n_A from 10 to 5000 and record the corresponding values of the coefficient of determination R^2. Then, we do the same for the n_R rarer species, also ranging n_R from 10 to 5000. Figure 1c shows the results. For the most abundant species, model (1) fits well for the $n_A = 534$ most abundant species, providing $R^2 = 0.998$. As for the rarest species, the best fit presents $R^2 = 0.966$ for the $n_R = 89$ rarest species.

Table 1 shows the Pareto exponents for the three subsamples. The value of α very close to zero for rare species suggests a heavier tail. Thus, these species are in a high uncertainty zone because there is no expected value for the number of rare species *X*. As a result, such species are in a non-equilibrium state in which they can be on the verge of extinction (*X* = 0) or becoming increasingly rare. Apart from the top four, the distribution of the most abundant species has an expected value *X* for abundance. The distribution, however, lacks a clear variability (as $\alpha < 2$), implying that this group is still vulnerable to abrupt hierarchical rank adjustments. Finally, the top four have the shortest tails. This finding is consistent with the literature, which shows that the risk of extinction varies across the avian phylogeny (Purvis 2008).

Table 1 Finding the Pareto exponents

Group	*Number of species*	R^2	*Pareto exponent* α	$\ln x_m$
Top four	4	0.972	3.485 (±0.422)	73.90 (± 8.85)
Abundant minus top four	546	0.998	1.107 (±0.002)	24.62 (± 0.04)
Rare	89	0.966	0.0027 (±0.00005)	9.18 (± 0.0001)

Figure 2 depicts a hypothetical evolution of the values of the Pareto exponent for the top four species as the *k* most abundant species are gradually removed. Because no abundant species are removed in the benchmark of Figure 1b, *k* = 0. The impact on the value of the Pareto exponent of removing the most abundant species is represented by *k* = 1, and the impact on the remaining top four is shown in Figure 2. The Pareto exponents are then computed after removing the two most abundant species (*k* = 2), the three most abundant (*k* = 3), and so on. Figure 2 shows a Loess smooth curve fitting with a blue line. The shaded area represents a 95% confidence interval for α.

The results of the exercise show a pattern of interest in the impact of extinction on the remaining (top four) species. The Pareto exponents, in particular, become unstable. The alphas are oscillating between light-tailed ($\alpha > 2$) and heavy-tailed regimes. Of note, when $1 < \alpha \leq 2$ the Pareto has no variance and when $\alpha \leq 1$ it has no mean either. As a result, the extinction of the big four significantly increases the overall uncertainty for the other species.

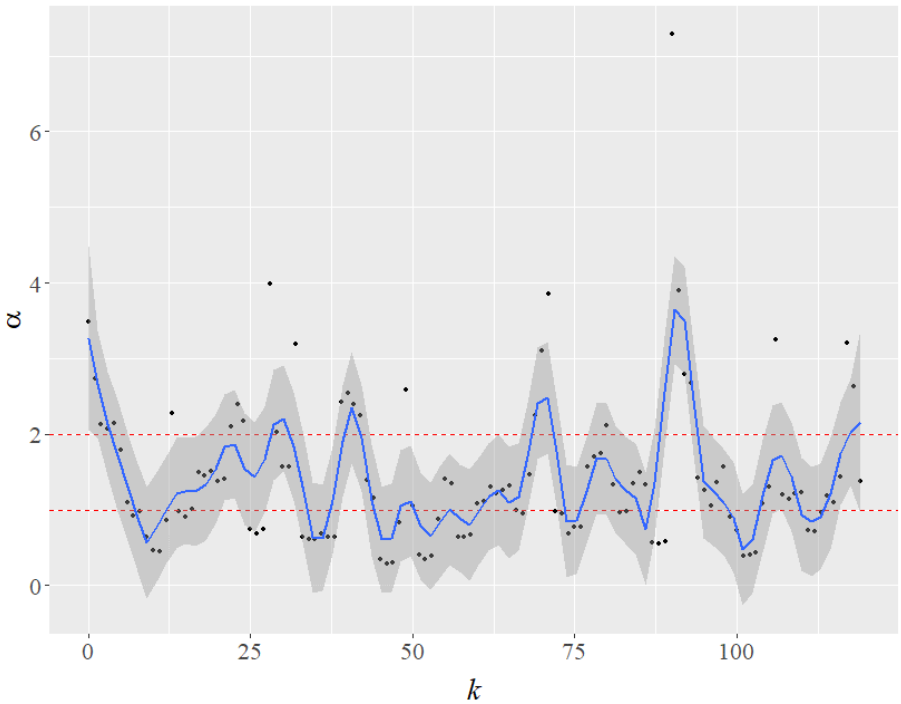


Figure 2 Changes in the value of the Pareto exponent (α) as we simulate the impact of the extinction of the most abundant species (as *k* increases) on the other remaining four large species.

Concluding remarks

Only a few bird species are abundant. Using new data, Callaghan et al. (2021) propose that a log left-skewed distribution, rather than a lognormal distribution, adjusts better to the abundance distribution of bird species. Using the same data, we consider the rank abundance distribution rather than the species abundance distribution. Three power laws are found to fit the tails. Because power laws emerge even when data is scarce, our finding is not overly dependent on data quality.

The fact that many species, like the third-most abundant ring-billed gull, have a Nearctic distribution among the top ten most abundant species may be one limitation of the Callaghan et al.'s data. Because of this supposed limitation, the eBird database could not reliably estimate the abundance of most species that do not occur widely in the Americas. However, this criticism ignores the expert ornithologist knowledge that went into compiling the dataset, and we cannot know any better.

References

Callaghan, C.T., Nakagawa, S. & Cornwell, W.K. (2021). Global abundance estimates for 9,700 bird species. *Proc. Natl Acad. Sci. USA*, 118, e2023170118. (doi:10.1073/pnas.2023170118)

Chisholm, R.A. (2007). Sampling species abundance distributions: Resolving the veil-line debate. *J. Theor. Biol.*, 247, 600-607. (doi:10.1016/j.jtbi.2007.04.006)

Enquist, B.J., Feng, X., Boyle, B., Maitner, B., Newman, E.A., Jorgensen, P.M., Roehrdanz, P.R., Thiers, B.M., Burger, J.R., Corlett, R.T., Couvreur, T.L.P., Dauby, G., Donoghue, J.C., Foden, W., Lovett, J.C., Marquet, P.A., Merow, C.,

Midgley, G., Morueta-Holme, N., Neves, D.M., Oliveira-Filho, A.T., Kraft, N.J.B., Park, D.S., Peet, R.K., Pillet, M., Serra-Diaz, J.M., Sandel, B., Schildhauer, M., Simova, I., Violle, C., Wieringa, J.J., Wiser, S.K., Hannah, L., Svenning, J.C., McGill, B.J. (2019) The commonness of rarity: Global and future distribution of rarity across land plants. *Sci. Adv.*, 5, eaaz0414. (doi:10.1126/sciadv.aaz0)

Jenkins, S.P. (2017). Pareto models, top incomes and recent trends in UK income inequality. *Economica*, 84, 261-289. (doi:10.1111/ecca.12217)

McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K., Dornelas, M., Enquist, B.J., Green, J.L., He, F., Hurlbert, A.H., Magurran, A.E., Marquet, P.A., Maurer, B.A., Ostling, A., Soykan, C.U., Ugland, K.I. & White, E.P. (2007). Species abundance distributions: Moving beyond single prediction theories to integration within an ecological framework. *Ecol. Lett.*, 10, 995-1015. (doi:10.1111/j.1461-0248.2007.01094.x)

Morlon, H., White, E.P., Etienne, R.S., Green, J.L., Ostling, A., Alonso, D., Enquist, B.J., He, F., Hurlbert, A., Magurran, A.E., Maurer, B.A., McGill, B.J., Olff, H., Storch, D. & Zillio T. (2009). Taking species abundance distributions beyond individuals. *Ecol. Lett.*, 12, 488-501. (doi:10.1111/j.1461-0248.2009.01318.x)

Newman, M.E. (2005). Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.*, 46, 323-351. (doi:10.1080/00107510500052444)

Purvis, A. (2008). Phylogenetic approaches to the study of extinction. *Annu. Rev. Ecol. Evol. S.*, 39, 301-319. (doi:10.1146/annurev-ecolsys-063008-102010)