bioRxiv preprint doi: https://doi.org/10.1101/2021.12.06.471443; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Power laws from the bird species abundance distribution

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
A recent study found that bird species with fewer individuals are abundant, but large species are rare. We show that this new data strongly suggests a power-law distribution rather than the most accepted log-normal. Moreover, we discuss extinction risk across the bird phylogeny and future conservation efforts by profiting from the hierarchical structure revealed by the new data.

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

Data accessibility.
https://www.pnas.org/highwire/filestream/984963/field_highwire_adjunct_files/1/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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.06.471443; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

The species abundance distribution is one of the few universal patterns in ecology (Morlon et al. 2009). A new study found that bird species with fewer individuals are abundant, but large species are rare (Callaghan et al. 2021). The researchers considered the recent influx of citizen science data to make the number of individuals (abundance) estimates for 9,700 species, about 92 percent of all birds. They combined data for 724 well-studied species with counts from the app eBird (https://ebird.org/home), where people submit bird sightings. Then, they used an algorithm to extrapolate estimates for the sample. As a result, they discovered many species with small populations isolated in niche habitats and relatively few species over a vast territory. The study provides the methodological blueprint for detailed quantifying. Here, we offer an extra quantifying of this interesting nature dynamics and show it belongs to the class of hierarchical system dynamics. In particular, we discover power laws from biggest to most minor bird species abundance.

Materials and methods

Figure 1*a* shows the data (pnas.2023170118.sd01). From 50 billion birds, there are four undomesticated species with billion-plus individuals: house sparrows (*Passer domesticus*), European starlings (*Sturnus vulgaris*), ring-billed gulls (*Larus delawarensis*), and barn swallows (*Hirundo rustica*). By contrast, 5,022 species number fewer than 500,000 birds each. Figure 1a provides a glimpse of the power-law pattern we find and immediately demands attention to the group of 1,178 species departing from the pattern at the naked eye. Many ecological processes structure the abundance of a species (McGill et al. 2007). For example, the numerous rarer species may have evolved to occupy a single island, but human activities such as deforestation can further explain the pattern deviation in Figure 1a. Possibly some species are rare because we meddle with them. This issue is a matter of further investigation from the perspective of conservation efforts.

A power law governs a quantity when the probability of getting a given value varies inversely with a power of that value (Newman 2005). Thus, a power law is a relationship between two quantities where a relative change in one leads to a proportional change in the other. Moreover, this holds regardless of the initial values. Power laws are common in hierarchical system dynamics arising in physics, economics, and biology.

A power law geometrically reveals itself 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. We show this in Figure 1b after taking log of rank versus log of bird abundance. Here, we need to quantify the power laws in Figure 1b by computing their Pareto exponents related to the slopes of the straight lines.

One standard method is to take the survival function $S(x)$ in a Pareto type I model (Jenkins 2017). Those bird species with abundance greater than x – that is, one minus the cumulative distribution function $P(x)$ – are given by

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.06.471443; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

where x_m is the lower bound on abundance, and $x \geq x_m > 0$. The shape parameter α is the Pareto exponent (tail index), describing the heaviness of the right tail, with smaller values meaning 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 generate straight-line intervals of slope $-\alpha$ (Figure 1b).

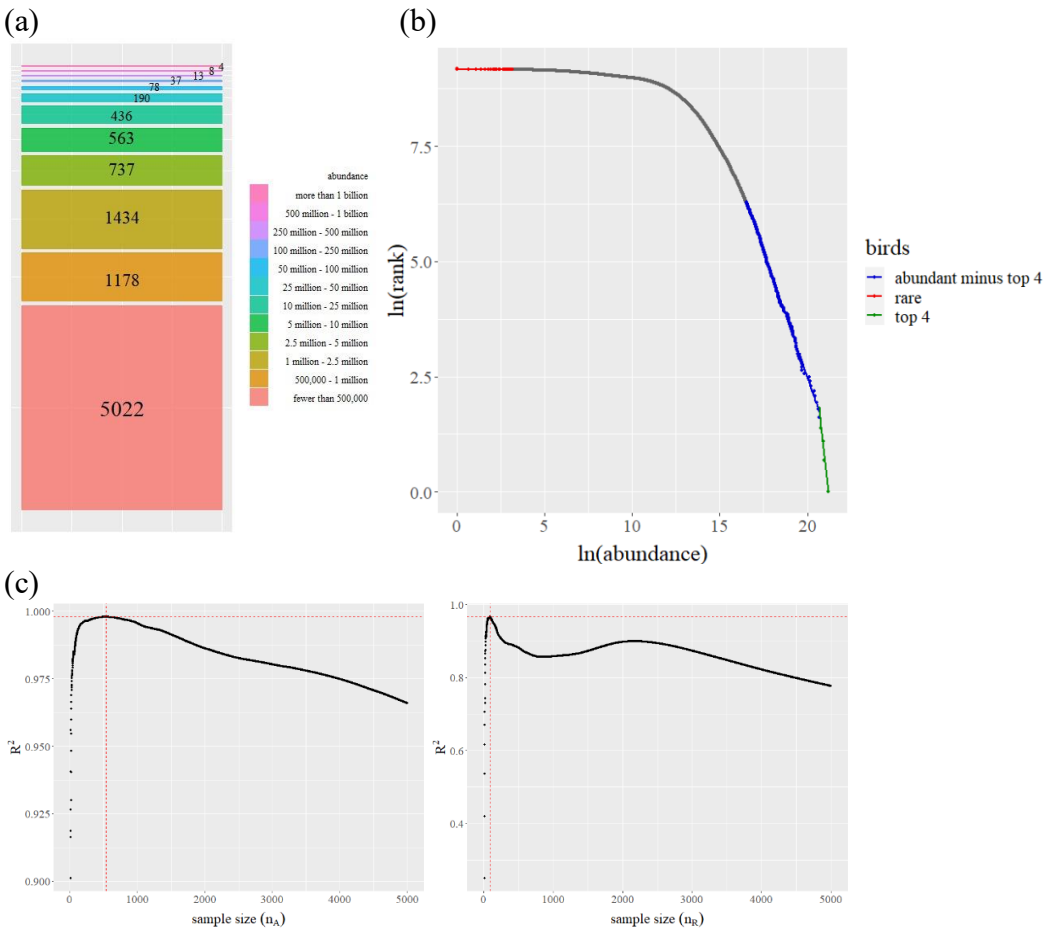

Figure 1 (a) The rarer a bird species, the more numerous. (b) Three power laws emerge when we take ln(rank) versus ln(abundance). (c) R^2 values versus sample size for the most abundant n_A species (apart from the top four). The crossing of the dotted lines indicates the optimal sample size ($n_A = 534$) 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).

In particular, to determine the group of the most abundant n_A species (apart 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.06.471443; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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*. Therefore, such species are in a non-equilibrium state where they are heading either toward extinction ($X = 0$) or becoming increasingly less rare. On the other hand, the distribution of the most abundant species apart from the top four has an expected value *X* for abundance. However, the distribution does not have a defined variability (as $\alpha < 2$), which means this group is still subject to abrupt hierarchical position shifts. Finally, the group of top four has the lightest tail. The literature already points to the fact that extinction risk is unequally spread across the bird phylogeny (Purvis 2008), but we provide detailed information of interest for future conservation efforts.

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)

Discussion

Abundance estimates of bird species are critical for ecology, evolutionary biology, and conservation, and improvements in quantifying abundance are welcome. A log-normal is mainly accepted for abundance distributions in the ecology literature (Chisholm 2007), but the new data we analyze here strongly suggests a more appropriate power-law distribution. This insight comes from the hierarchical structure revealed by the data. This article contributes to show that power laws summarize the finding that the rarer a bird species, the more numerous. This circumstance means that all the results from the literature on hierarchical system dynamics can now come to the fore to enrich the discussion.

Figure 2 shows a hypothetical evolution of the values of the Pareto exponent for the top four species as we increasingly remove the *k* most abundant species. In the benchmark of Figure 1*b*, no abundant species are removed, and so $k = 0$. Removing the most abundant species is represented by $k = 1$, and the impact on the value of the Pareto exponent of the remaining top four is displayed in Figure 2. Then, we compute the Pareto exponents after progressively removing the two most abundant species ($k = 2$), the three most abundant ($k = 3$), and so on. The blue line in Figure 2 is a Loess smooth curve fitting. The shaded region represents a 95 percent confidence interval for α.

The exercise results reveal a pattern of interest for the impact of the extinction of the most abundant species on the remaining (top four) species. Notably, the Pareto exponents become unstable. The alphas oscillate 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. Therefore, the impact of the big four's extinction dramatically increases the overall uncertainty for the other species.

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.06.471443; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

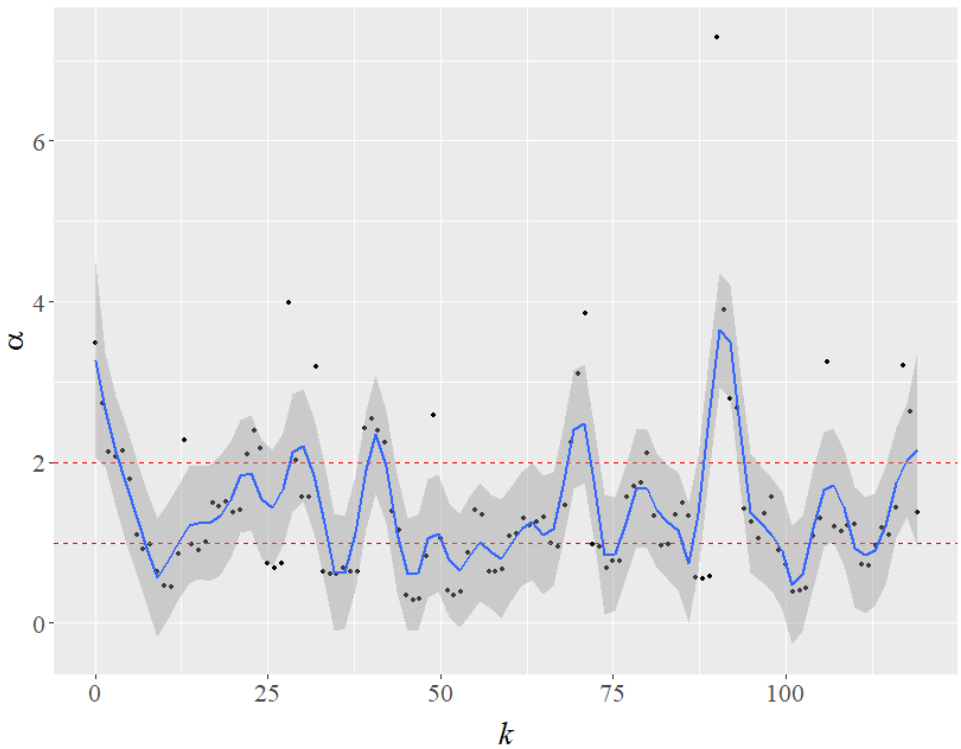

Figure 2 Changes in the value of the Pareto exponent (α) as we increasingly simulate the impact of the extinction of the most abundant species (as we increase *k*) on the other remaining four large species.

Concluding remarks

Callaghan et al. (2021) correctly observe that many species in their data have minimal population estimates, and relatively few species are very abundant. There are about 200 more species than expected if bird species abundance followed a truly log-normal distribution. Nevertheless, they settle for a log left-skewed distribution (Callaghan et al. 2021) rather than the power-law distribution we consider here. The flexibility of the log left-skewed distribution accommodates the data well, but the Pareto seems more compelling for our piecewise approach. Nevertheless, we do not dismiss the log left-skewed distribution; we consider the Pareto limit because of our focus on the tail behavior.

One weakness of the data is that it is hard to believe that many species have a Nearctic distribution within the ten most abundant, including the third more numerous ring-billed gull. This limitation prevents reliable estimates using the eBird database for most species that do not occur widely in the Americas. Nevertheless, we believe more reliable estimates will not dismiss our discovery of the hierarchical nature of the bird species abundance distribution because power laws reveal themselves even if data are scant.

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)

bioRxiv preprint doi: https://doi.org/10.1101/2021.12.06.471443; this version posted December 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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)

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)