Reply to Chen: Under specified assumptions, adequate random samples of skewed distributions obey Taylor's law

Chen (1) simulated random samples of beta, lognormal, and Poisson distributions with varying parameters following the method that we (2) used for fixed parameters. Chen claimed that the relationship between the supporting rate of Taylor's law (TL) and skewness is "complex and nonmonotonic," as some random samples of some skewed distributions did not support TL. Chen did not consider two crucial issues: satisfying the assumptions of our theorem (2) and estimating the quantities in it adequately.

For a beta distribution with parameters α and β, Chen fixed $\beta = 0.5$ and considered α as small as 0.01. It is known analytically (3) that values of α near zero give means and variances close to zero. A zero mean or a zero variance is contrary to assumptions of our theorem (2). When mean and variance are near zero, sample estimates of log(mean) and log(variance) and parameter estimates of TL become unstable. Thus, Chen's observation that the supporting rate of TL fell toward zero (figure 1*A* in ref. 1) as α approached zero did not contradict our theorem. By contrast, when we fixed $\beta = 0.5$ and varied $\alpha \in [0.3,0.7]$ to avoid mean near zero and variance near zero in accordance with our assumptions, random samples of beta distributions yielded a high supporting rate of TL when $\alpha < \beta$ (positively skewed) and when $\alpha > \beta$ (negatively skewed), and a low supporting rate of TL when $\alpha \approx \beta$ (skewness near zero) (Fig. 1*A*), confirming our theorem (2). With $\beta = 0.5$, the minimal mean and minimal variance using our $\alpha \in [0.3,0.7]$ were 0.375 and 0.1105, respectively, roughly 19 and 9 times the minimal mean and minimal variance using Chen's $\alpha \in [0.01,0.5]$.

For a lognormal distribution, it is known analytically (3) that the skewness grows in proportion to $\exp(3\sigma^2/2)$ asymptotically for large σ. Chen fixed the lognormal parameter $\mu = 1$ and varied $\sigma \in [0.001,10]$. The sample skewness computed by Chen severely underestimated the true population skewness as σ increased beyond about 2 (figure 1*D* in ref. 1 and Fig. 1*H*). Clark and Perry (4) and we (2) observed that lognormal sample variances underestimated lognormal population variances, with adverse effects on estimates of TL parameters. Because the underestimation of population variance (Fig. 1*G*) was more severe than the underestimation of population mean (Fig. 1*F*) when σ was large, it distorted the true log(mean)–log(variance) relationship and generated artificial nonlinearity (Fig. 1*E*), leading to a low supporting rate of TL (figure 1*C* in ref. 1). When we restricted σ to [0.001,1.2], a range where estimates of mean (Fig. 1*F*), variance (Fig. 1*G*), and skewness (Fig. 1 *D* and *H*) were relatively accurate, the results confirmed our theory (Fig. 1*C*). This range of σ was found in many scientific fields (table 2 in ref. 5, where $s^* = \exp(\sigma) \in [1.001,3.32]$ if $\sigma \in [0.001,1.2]$).

In conclusion, Chen's examples using beta and lognormal distributions did not disprove our theorem (2). Rather, they violated its assumptions or estimated the quantities in it inadequately.

ACKNOWLEDGMENTS. J.E.C. was partially supported by US National Science Foundation Grant DMS-1225529.

***Joel E. Cohen*[a,1] *and Meng Xu*[a,b]**
[a]*Laboratory of Populations, Rockefeller University and Columbia University, New York, NY 10065; and* [b]*Department of Mathematics and Physics, University of New Haven, West Haven, CT 06516*

1 Chen Y (2015) Random sampling of skewed distributions does not necessarily imply Taylor's law. *Proc Natl Acad Sci USA*, 10.1073/pnas.1507266112.
2 Cohen JE, Xu M (2015) Random sampling of skewed distributions implies Taylor's power law of fluctuation scaling. *Proc Natl Acad Sci USA*, 10.1073/pnas.1503824112.
3 Forbes C, Evans M, Hastings N, Peacock B (2010) *Statistical Distributions* (Wiley, Hoboken, NJ), 4th Ed.
4 Clark SJ, Perry JN (1994) Small sample estimation for Taylor's power law. *Environ Ecol Stat* 1(4):287–302.
5 Limpert E, Stahel WA, Abbt M (2001) Log-normal distributions across the sciences: Keys and clues. *Bioscience* 51(5):341–352.

Author contributions: J.E.C. and M.X. performed research and wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. Email: cohen@rockefeller.edu.

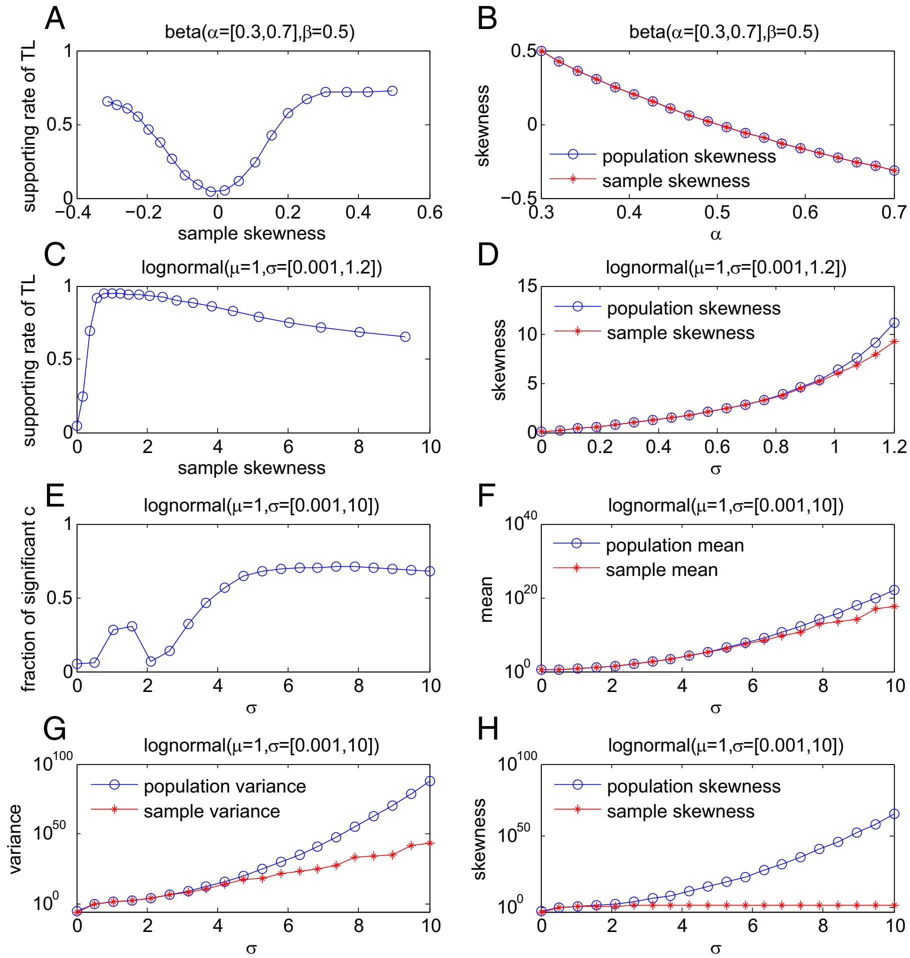

Fig. 1. (*A*) Supporting rate of TL against sample skewness of beta ($\alpha \in [0.3,0.7]$, $\beta = 0.5$). Supporting rate of TL increased as sample skewness deviated from zero. (*B*) Population skewness and sample skewness of beta ($\alpha \in [0.3,0.7]$, $\beta = 0.5$) against α. (*C*) Supporting rate of TL against sample skewness of lognormal ($\mu = 1$, $\sigma = [0.001,1.2]$). Except when sample skewness was close to zero, TL was not rejected in at least 60% of samples. (*D*) Population skewness and sample skewness of lognormal ($\mu = 1$, $\sigma = [0.001,1.2]$) against σ. (*E*) Fraction of 10,000 quadratic regressions with significant c ($P < 0.05$) for each pair of parameters of lognormal ($\mu = 1$, $\sigma = [0.001,10]$) against σ, where c was defined in ref. 1. The fraction was >~0.6 when $\sigma \geq \sim 5$. (*F*) Population mean and sample mean of lognormal ($\mu = 1$, $\sigma = [0.001,10]$) against σ on semilog scale. (*G*) Population variance and sample variance of lognormal ($\mu = 1$, $\sigma = [0.001,10]$) against σ on semilog scale. (*H*) Population skewness and sample skewness of lognormal ($\mu = 1$, $\sigma = [0.001,10]$) against σ on semilog scale. In *F–H*, the sample statistics severely underestimated the corresponding population parameters when σ was large. For each distribution and pair of parameters, 10,000 random copies of a 100×100 matrix were simulated. Sample mean and sample variance of lognormal distributions were calculated using all 10,000 elements in each simulated 100×100 matrix, then averaged over the 10,000 random copies of the matrix. Sample skewness and supporting rate of TL were computed using the method in ref. 1.