Benford's Law: Digits within Numbers

Alex Ely Kossovsky, City University of New York

Contrary to common intuition that all digits should occur randomly with equal chances within numbers in real data, empirical examinations consistently show that not all digits are created equal, but rather that low digits such as {1, 2, 3} occur much more frequently than high digits such as {7, 8, 9} in almost all data types, such as those relating to geology, chemistry, astronomy, physics, and engineering, as well as in accounting, financial, econometrics, and demographics data sets. This intriguing digital phenomenon is known as Benford's Law, and it constitutes the only multidisciplinary mathematical pattern occurring throughout all the sciences. Yet, the phenomenon is essentially quantitative and physical, not merely digital and numerical, constituting a scientific reality independent of our arbitrarily invented positional number system; originating from the fact that in extreme generality, nature creates many small quantities but very few big ones, corroborating the motto "**small is beautiful**". Such an unorthodox point of view of Benford's Law is mathematically worked out via the postulate that the generic pattern in how relative quantities are found in nature is such that the frequency of quantitative occurrences is inversely proportional to quantity, leading to what is termed 'The General Law of Relative Quantities', expressed algebraically as **ln((D+d(F-1))/(D+(d-1)(F-1)))/ln(F)**. When real-life data sets are checked against this expression they are found to be in agreement with it, corroborating this rather radical interpretation of the law and endowing it scientific credibility. Classic Benford's Law regarding the first order distribution of our numerical digits, namely **LOG(1+1/d)**, is then demonstrated to be merely a consequence and a special case of this more general law. In addition, the surprising logarithmic status of chains of distributions is explored. Such complex random distributions are constructed with the twist that (the supposedly fixed) parameters themselves vary and are derived from other random distributions.