ON THE DISTRIBUTION OF FIRST SIGNIFICANT DIGITS[1]

By Roger S. Pinkham

Rutgers—The State University

Introduction. It has been noticed by astute observers that well used tables of logarithms are invariably dirtier at the front than at the back. Upon reflection one is led to inquire whether there are more physical constants with low order first significant digits than high. Actual counts by Benford [2] show that not only is this the case but that it seems to be an empirical truth that whenever one has a large body of physical data, Farmer's Almanac, Census Reports, Chemical Rubber Handbook, etc., the proportion of these data with first significant digit n or less is approximately $\log_{10}(n + 1)$. Any reader formerly unaware of this "peculiarity" will find an actual sampling experiment wondrously tantalizing. Thus, for example, approximately 0.7 of the physical constants in the Chemical Rubber Handbook begin with 4 or less ($\log_{10}(4 + 1) = 0.699$). This is to be contrasted with the widespread intuitive evaluation $\frac{4}{9}$ths.

At least two books call attention to this peculiarity, Furlan [6] and Wallis [18], but to my knowledge there are only five published papers on the subject, Benford [2], Furry et al [7], [9], Gini [8], and Herzel [11]. The first consists of excellent empirical verifications and a discussion of the implied distribution of 2nd, 3rd, $\cdots$ significant digits. The second and third put forth the thesis that the distribution of significant digits should not depend markedly on the underlying distribution, and the authors present numerical evaluations for a range of underlying distributions in support of their contention. The fourth maintains that explanation is to be sought in empiric considerations. The fifth considers three different urn models; each yields a distribution of initial digits which the author compares with $\log_{10}(n + 1)$.

This paper is a theoretical discussion of why and to what extent this so called "abnormal law" must hold. The flavor of the results is, I think, conveyed in the following remarks.

(i) The only distribution for first significant digits which is invariant under scale change of the underlying distribution is $\log_{10}(n + 1)$. Contrary to suspicion this is a non-trivial mathematical result, for the variable n is discrete.

(ii) Suppose one has a horizontal circular disc of unit circumference which is pivoted at the center. Let the disc be given a random angular displacement θ where $-\infty < \theta < \infty$. If the final position of the disc mod one is called φ, i.e.,

$$\varphi \equiv \theta \bmod(1), \qquad\qquad 0 \leq \varphi < 1,$$