Benford's Law and Scale Invariance

November 27, 2025

In this article we explore a motivation for the Benford's law[1] starting from an assumption of scale-invariance. There are similar derivations from an assumption of scale-invariance[2] and it is also possible to arrive at Benford's law from more minimal assumptions[3].

The Benford Distribution

The Benford distribution, widely observed in datasets, deviates from a uniform distribution and the probability is concentrated more in the lower digits. For base Ω with alphabet $i \in \{0, 1, ...\Omega - 1\}$ the Benford distribution for the first digit is

$$\mathcal{P}(i|B, \Omega) = \log_{\Omega}\left(1 + \frac{1}{i}\right) \qquad i \in \{1, 2, ...\Omega - 1\}.$$

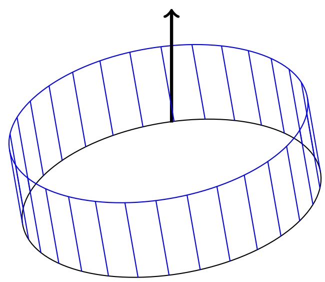
It is easily checked that it is a normalised probability distribution and that the probability weight is skewed to the left.

Uniform Distribution and Symmetry

The assumption of a uniform distribution is made when there is no reason to prefer one outcome over the other. For example, each of the outcomes $\{1, 2, ...6\}$ for a die is assigned equal probability as there is no reason a priori to choose one over the other. Another interpretation is that there is a certain symmetry in the problem and all outcomes of the experiment are equivalent under this symmetry.

If one considers an idealization of a roulette wheel, a circle, which is rotated a random amount a uniform probability distribution on the circle results. Again, in this case a translational symmetry could motivate a uniform distribution on the circle.

1. "Benford's law," https://en.wikipedia.org/wiki/Benford's_law.
2. Eric W. Weisstein, "Benford's Law," *From MathWorld–A Wolfram Web Resource*, https://mathworld.wolfram.com/BenfordsLaw.html.
3. Theodore P. Hill, "The First Digit Phenomenon," *American Scientist* 86, no. 4 (July 1998): 358, https://doi.org/10.1511/1998.4.358.



In the case of physical constants or numbers drawn randomly from a distribution that varies over a large range of scales one could propose a different kind of symmetry, that of scale-invariance, so that the distribution should not change by rescaling the units in which the numbers representing the distribution are measured. An argument could be that the arbitrary definition of a unit one imposes on a physical constant should not affect its probability distribution.

For example, the arbitrary choice to use feet or meters to measure length should not affect the probability distribution of length for a physical constant. As any physical entity, would be expected to have a finite size and therefore to have a decaying probability distribution with respect to length, scale-invariance is a better motivated symmetry to use instead of translational invariance. Indeed, in any situation, it makes more sense to first specify the symmetry and then declare the distribution to be uniform over that symmetry.

Scale Invariance

The expected symmetry is that over transformations:

$$x \to kx,$$

where k is a scaling factor.

Remark. *We only consider scalar quantities like length, volume or mass. For higher order tensorial quantities different symmetries of the tensor structure would have to be taken into account as well.*

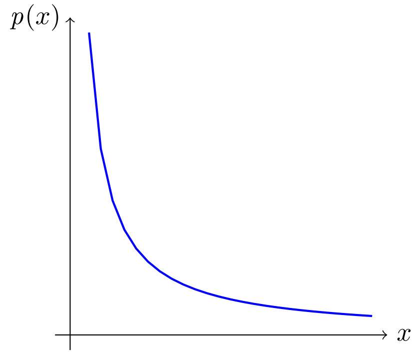
And therefore we require for $y = kx$,

$$\begin{aligned} p(y)dy &\overset{!}{=} p(x)dx \\ \implies p(kx)kdx &\overset{!}{=} p(x)dx \\ \implies p(kx) &\overset{!}{=} \frac{p(x)}{k}. \end{aligned}$$

Fixing x at 1 it can be seen that the probability distribution required for scale-invariance is

$$p(x) = \frac{\mathsf{C}}{x},$$

where C is some constant. Note that the distribution is not normalizable for $x \in (0, \infty)$.



The scale invariance also holds for intervals of the physical quantity. So

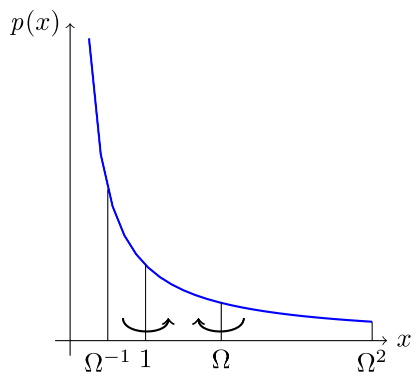
$$\mathcal{P}(a \leq x < b) = \int_a^b p(x)dx = \mathsf{C}\ln\left(\frac{b}{a}\right)$$

which means that even intervals of the physical constant are scale invariant

$$\mathcal{P}(a \leq x < b) = \mathcal{P}(ka \leq kx < kb)! \tag{1}$$

Benford Distribution

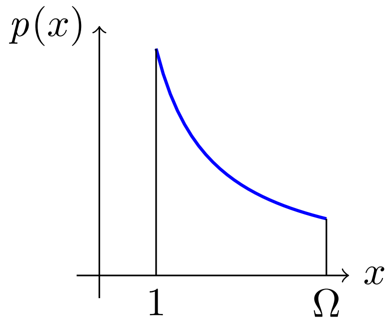
If what is being considered now is the first digit, it would lie between $\{1, ...\Omega-1\}$. In this case, each of the chucks of the probability distribution over the different scales, $\{[\Omega^k, \Omega^{k+1})\}$ are rescaled to the interval between $[1, \Omega)$; this is justified through equation (1). It is similar to expressing the value of the quantity in scientific notation, $165 = 1.65 \times 10^2$, $0.32 = 3.2 \times 10^{-1}$ and so on.



The resulting probability distribution is normalisable:

$$\int_1^\Omega p(x)dx = \mathsf{C}\ln(\Omega)$$

so that $\mathsf{C} = 1/\ln(\Omega)$.



Now, if binning is introduced to consider sets of values with a given first digit, the Benford distribution results.

$$
\begin{aligned}
\mathcal{P}(1 \leq x < 2) &= \mathsf{C} \ln \left(\frac{2}{1}\right) \\
\mathcal{P}(2 \leq x < 3) &= \mathsf{C} \ln \left(\frac{3}{2}\right) ... \\
\mathcal{P}(\Omega - 1 \leq x < \Omega) &= \mathsf{C} \ln \left(\frac{\Omega}{\Omega - 1}\right)
\end{aligned}
$$

and therefore

$$\mathcal{P}(i|B, \Omega) = \log_{\Omega} \left(1 + \frac{1}{i}\right)$$

results.