Here we derive the theoretical value distribution for a gene in uniformly-sampled probability vectors, and compare it with empirical results:

Since the entries of p add up to 1, we can compute $P(p_i = k)$ by evaluating its equivalent $P\left(\sum_{j\neq i} p_j = 1-k\right)$. Let S_1 be the solution space spanned by the condition

$$\sum_{j\neq i} p_j = 1-k$$

or equivalently,

$$\|p_{-i}\|_1 = 1-k, \quad p_{-i} = \{p_0, \ldots, p_{i-1}, p_{i+1}, \ldots\, p_n\}$$

In this formulation, we see that S_1 forms the surface of a taxicab (L_1) sphere. From [1], we know that the volume of S_1 is proportional to the radius raised by the degree of freedom. Since p_i is an independent variable, the degree of freedom is one less $|p|-1$, yielding

$$V_n(S_1) \propto (1-k)^{|p|-2} \tag{1}$$

Since the probability density is uniform across all vector solutions, we know that the probability mass of S_1 as directly proportional to its volume defined in Equation (1),

$$P(p_i = k) = \lambda(1-k)^{|p|-2}$$

where λ is a normalizing constant,

$$\int_0^1 \lambda(1-k)^{|p|-2} dk = 1$$

$$\lambda\left[\frac{1}{|p|-1}(1-k)\right]_1^0 = 1$$

$$\lambda = |p|-1$$

$$P(p_i = k) = (|p|-1)(1-k)^{|p|-2} \tag{2}$$

Equation (2) thus gives us the theoretical distribution of a gene in uniformly-sampled probability vectors.

References

[1] K. P. Thompson, "The nature of length, area, and volume in taxicab geometry," *arXiv preprint arXiv:1101.2922*, 2011.