On the probability that a randomly chosen pmf on $\{0, 1, \ldots, n\}$ is represented as a sum of independent 0-1 indicators

Nickos Papadatos
National and Kapodistrian University of Athens

5th October 2024
In Memory of Professor Ch.A. Charalambides

The Problem

Let S be a random variable with values in $\{0, 1, \ldots, n\}$.
Assume that S has pmf $y_j = \mathbb{P}(S = j)$.
Then,

$$\mathbf{y} := (y_1, \ldots, y_n)$$

belongs to the simplex

$$S_n := \{(y_1, \ldots, y_n) : y_1 \geq 0, \ldots, y_n \geq 0, y_1 + \cdots + y_n \leq 1\}.$$

Definition: Proper pmf

Let $S \sim \mathbf{y}$. The pmf $\mathbf{y}$ is called **proper** if there exist $0-1$ **independent** indicators $I_1, \ldots, I_n$ such that

$$\mathbb{P}(I_1 + \cdots + I_n = j) = \mathbb{P}(S = j) = y_j, \quad j = 1, \ldots, n.$$

Lemma 1

The pmf **y** is proper if and only if the polynomial

$$G(u) := \sum_{j=0}^{n} y_j (1+u)^j$$

has only real roots.

A randomized analogue

If **y** is chosen **randomly** in the simplex S_n, what is the probability that **y** is proper?
Here, as usually, ”randomly” means that **y** has uniform distribution on the simplex S_n.

Solution

Since the area of S_n equals $1/n!$, we wish to calculate the probability

$$P_n := \frac{Area(\Pi_n)}{Area(S_n)} = n! Area(\Pi_n)$$

where Π_n is the subset of S_n that contains all the proper pmfs.

On the other hand it is obvious to see that, when **y** is proper,

$$y_j = y_j(p_1, p_2, \ldots, p_n), \quad (j = 1, \ldots, n)$$

with the exact formula

$$y_j(p_1, p_2, \ldots, p_n)$$
$$= p_1 p_2 \cdots p_n \sum_{1 \le i_1 < i_2 < \cdots < i_j \le n} \frac{(1-p_{i_1})(1-p_{i_2})\cdots(1-p_{i_j})}{p_{i_1} p_{i_2} \cdots p_{i_j}}.$$

Without loss of generality we shall assume that $0 < p_1 < \cdots < p_n < 1$ (recall $p_j = \mathbb{P}(I_j = 1) = 1 - \mathbb{P}(I_j = 0)$), because the function $\mathbf{y} = \mathbf{y}(\mathbf{p})$ is a permutation invariant function on p's. Define
$\Delta_n := \{\mathbf{p} := (p_1, \ldots, p_n) : 0 \le p_1 \le p_2 \le \cdots \le p_n \le 1\}$.
Then,

$$\mathbf{y} : \Delta_n \to \Pi_n$$

is a bijection.
In other words,

$$\Pi_n = \mathbf{y}(\Delta_n).$$

The co-area formula

$$Area(\Pi_n) = \int_{\Pi_n} d\mathbf{y} = \int_{\mathbf{y}(\Delta_n)} d\mathbf{y}.$$

The last integral equals (according to the co-area formula) to

$$\int_{\Delta_n} |\det J(\mathbf{p})| d\mathbf{p}$$

where

$$J(\mathbf{p}) = \frac{\partial \mathbf{y}(\mathbf{p})}{\partial \mathbf{p}}$$

is the Jacobian matrix.

After a lengthy calculation we found

$$\det J(\mathbf{p}) = \prod_{1\le i<j\le n} (p_i - p_j)$$

(this is a Vandermonde type determinant!)

Result 1

The probability P_n is given by

$$P_n = n! \int_{\Delta_n} \prod_{1 \le i < j \le n} (p_j - p_i) d\mathbf{p}.$$

Equivalently, from symmetry reasons,

$$P_n = \int_{(0,1)^n} \prod_{1 \le i < j \le n} |x_j - x_i| d\mathbf{x}.$$

From this formula we can calculate (by hand!) $P_1 = 1$, $P_2 = 1/3$, $P_3 = 1/30$ and (perhaps...) $P_4 = 1/1050$.

Theorem 1

$$P_n = \frac{1}{B_n}$$

Theorem 1

$$P_n = \frac{1}{B_n}$$

where

$$B_n = \binom{1}{1}\binom{3}{2}\binom{5}{3}\cdots\binom{2n-1}{n}.$$

Theorem 1

$$P_n = \frac{1}{B_n}$$

where

$$B_n = \binom{1}{1}\binom{3}{2}\binom{5}{3}\cdots\binom{2n-1}{n}.$$

The first 7 values of the sequence B_n are $1, 3, 30, 1050, 132300, 61122600, 104886381600$.

Selberg's Integral

The proof is based on Selberg's Integral,

$$I(\alpha,\beta,\gamma) := \int_{(0,1)^n} \prod_{i=1}^{n} x_i^{\alpha-1}(1-x_i)^{\beta-1} \prod_{1\le i<j\le n} |x_j - x_i|^{2\gamma}\, d\mathbf{x},$$

the value of which is

$$I(\alpha,\beta,\gamma) = \prod_{j=0}^{n-1} \frac{\Gamma(\alpha+j\gamma)\Gamma(\beta+j\gamma)\Gamma(1+(j+1)\gamma)}{\Gamma(\alpha+\beta+(n+j-1)\gamma)\Gamma(1+\gamma)};$$

this formula holds provided
$\Re(\alpha)>0, \Re(\beta)>0, \Re(\gamma)>-\min\left\{\frac{1}{n}, \frac{\Re(\alpha)}{n-1}, \frac{\Re(\beta)}{n-1}\right\}.$

In our case we set $\alpha = \beta = 1$, $\gamma = 1/2$, so that

$$P_n = I(1,1,1/2) = \frac{2^n}{\pi^{n/2}} \prod_{j=0}^{n-1} \frac{\Gamma(1+j/2)^2 \Gamma(1+(j+1)/2)}{\Gamma(2+(n+j-1)/2)}.$$

In our case we set $\alpha = \beta = 1$, $\gamma = 1/2$, so that

$$P_n = I(1,1,1/2) = \frac{2^n}{\pi^{n/2}} \prod_{j=0}^{n-1} \frac{\Gamma(1+j/2)^2 \Gamma(1+(j+1)/2)}{\Gamma(2+(n+j-1)/2)}.$$

A final induction argument give the formula $P_n = 1/B_n$ with B_n as in the Theorem.

Thank you !!!