N. Papadatos

Introduction The Problem On the probability that a randomly chosen pmf on $\{0, 1, ..., n\}$ 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

◆□▶ ◆□▶ ▲□▶ ▲□▶ □ のQ@

The Problem

Sums of Independent Indicators

N. Papadatos

Introduction

The Problem

Solution

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

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

belongs to the simplex

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

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

Definition: Proper pmf

Sums of Independent Indicators

N. Papadatos

Introduction

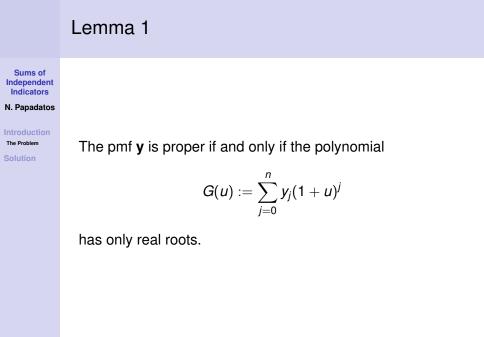
The Problem

Solution

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

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

▲□▶ ▲□▶ ▲ 三▶ ▲ 三▶ - 三■ - のへぐ



< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

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 .

▲□▶▲□▶▲□▶▲□▶ □ のQ@

Solution

Sums of Independent Indicators

N. Papadatos

Introduction

The Problem

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,

$$y_j = y_j(p_1, p_2, ..., p_n), (j = 1, ..., n)$$

with the exact formula

$$y_{j}(p_{1}, p_{2}, \dots, p_{n}) = p_{1}p_{2} \cdots p_{n} \sum_{1 \leq i_{1} < i_{2} < \dots < i_{j} \leq n} \frac{(1-p_{i_{1}})(1-p_{i_{2}})\cdots(1-p_{i_{j}})}{p_{i_{1}}p_{i_{2}}\cdots p_{i_{j}}}.$$

▲□▶▲□▶▲□▶▲□▶ □ のQ@

N. Papadatos

Introduction The Problem

Solution

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 $\Delta_n := {\mathbf{p} := (p_1, \dots, p_n) : 0 \le p_1 \le p_2 \le \cdots \le p_n \le 1}.$ Then, $\mathbf{y} : \Delta_n \to \prod_n$

is a bijection. In other words,

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

◆□▶ ◆□▶ ▲□▶ ▲□▶ □ のQ@

The co-area formula

Sums of Independent Indicators

N. Papadatos

Introduction

The Problem

Solution

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

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

▲□▶ ▲□▶ ▲ 三▶ ▲ 三▶ - 三 - のへぐ

is the Jacobian matrix.

N. Papadatos

Introduction

The Problem

Solution

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

Sums of Independent Indicators

N. Papadatos

Introduction The Problem Solution The probability P_n is given by

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

Equivalently, from symmetry reasons,

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

▲□▶ ▲□▶ ▲ 三▶ ▲ 三▶ - 三 - のへぐ

N. Papadatos

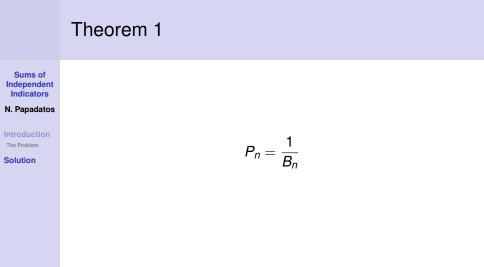
Introduction

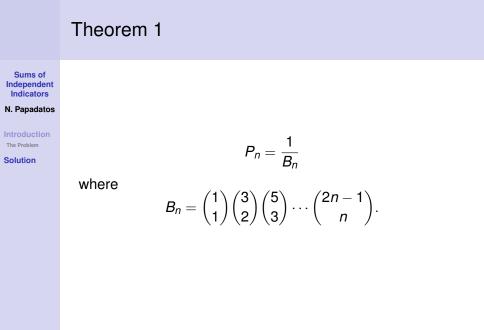
The Problem

Solution

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

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>





・ ロ ト ・ 信 ト ・ 言 ト ・ 言 ・ の へ ()

Theorem 1

Sums of Independent Indicators

N. Papadatos

Introduction The Problem

Solution

$$P_n=rac{1}{B_n}$$

where

$$B_n = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \end{pmatrix} \begin{pmatrix} 5 \\ 3 \end{pmatrix} \cdots \begin{pmatrix} 2n-1 \\ n \end{pmatrix}.$$

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

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

Selberg's Integral

Sums of Independent Indicators N. Papadatos

Introduction The Problem The proof is based on Selberg's Integral,

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

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

N. Papadatos

Introductio

The Problem

Solution

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

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

N. Papadatos

Introduction The Problem

Solution

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.

N. Papadatos

Introduction

The Problem

Solution

Thank you !!!

▲□▶ ▲□▶ ▲□▶ ▲□▶ ▲□ ● のへぐ