DOI: 10.1685/CSC06097

Communications to SIMAI Congress, ISSN 1827-9015, Vol. 1 (2006)

Algebra and Statistic

Vincenzo Iorfida MIUR c.da Muzio n.10, 88046 Lamezia Terme vincenzo.iorfida@tin.it

Gaetana Restuccia Departement of Mathematics, University of Messina C.da Papardo, salita Sperone, 31, 98166 Messina, Italy grest@dipmat.unime.it

1 Introduction

Statistical methods are applied to answer questions in many of the major business disciplines including accounting, finance, marketing, economy, production and, of course, general management. The relationships between sample data and population values are based on the definitions:

DEFINITION 1.1. (Population): the total group of objects being studied or investigated.

DEFINITION 1.2. (Sample): a group of objects that is selected from population and from which information is gathered.

DEFINITION 1.3. (Statistical inference): The methods involved in drawing conclusions about population based upon information collected from a sample.

Algebraic methods can be used where algebraic functions are utilized in our process. On the other and, algebraic functions are very important:

1) \overline{X} : the sample mean;

2) σ^2 : the sample variance.

 $\overline{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$, where *n* is the sample size and X_1, X_2, \dots, X_n are the random variables of the random process. Hence \overline{X} is a linear function of *n* variables X_1, X_2, \dots, X_n , i.e. a linear polynomial of $\mathbb{R}[X_1, X_2, \dots, X_n]$, \mathbb{R} the field of real numbers. Moreover $\sigma^2 = \frac{1}{n}[(X_1 - \mu)^2 + \dots + (X_n - \mu)^2]$, where μ is the population mean (value being estimate) is a polynomial of $\mathbb{R}[X_1, X_2, \dots, X_n]$ of degree two. We call \overline{X} and σ^2 statistics.

In general we can have other statistics in our process, but these are the most studied and important. Methods of computational algebra can help us to study interesting questions.

DEFINITION 1.4. Let X_1, X_2, \ldots, X_n be random variables and $\underline{X} = (X_1, X_2, \ldots, X_n)$ be the correspondent sample. Let (a_1, \ldots, a_n) and (b_1, \ldots, b_n) be two sample data. Let $(a_1, \ldots, a_n) < (b_1, \ldots, b_n)$ if $(a_1, \ldots, a_n) < (b_1, \ldots, b_n)$ for a total order in \mathbb{R}^n (lexicographic order, degreelexicographic order, ...).

As a consequence we can order all sample data of the sample X. From now on , the space of sample data will be totally ordered.

DEFINITION 1.5. Let $s(X_1, X_2, \ldots, X_n) \in R[X_1, X_2, \ldots, X_n]$ a statistic. An order on the random variables X_1, X_2, \ldots, X_n is called compatible with the statistic function

V. Iorfida et al

 $s(X_1, X_2, ..., X_n)$ if $(a_1, ..., a_n) < (b_1, ..., b_n)$ implies $s(a_1, ..., a_n) \le s(b_1, ..., b_n)$ for every sample data.

EXAMPLE 1.1. The sample mean X is compatible with respect to the degree lexicographic order and not with respect to the lexicographic order: (1, 4, 5) < (2, 2, 3) for lex, but 1 + 4 + 5 > 2 + 2 + 3.

EXAMPLE 1.2. The sample variance σ^2 is not compatible with respect to the degree lexicographic order, neither with respect to the lexicographic order: (1, 2, 3) < (2, 4, 5) for degree lex order and for lex order, but in both cases, if the population mean $\mu = 3$, $(1-3)^2 + (2-3)^2 + (3-3)^2 > (2-3)^2 + (4-3)^2 + (5-3)^2$. If $\mu = 0$, σ^2 is compatible w.r.t. the degree lex order.

2 Applications

- 1. Hypothesis tests on means.
- 2. Stochastic matrices.

1. One of the most common types of hypothesis tests deals with the testing of means. EXAMPLE 2.1. State the null hypothesis (H_0) , which is the statement that we test. $(H_0): \mu = a, \mu > a, \mu < a,$ that is the population mean is equal to some specified value or least some specified value of a, or at most some specified value of a. Since the sample mean is compatible with the order degreelex, this implies the rejection region of (H_0) or the fail to reject region of (H_0) can be easily computed accordingly our order <.

THEOREM 2.1. Let T be any hypothesis test involving means and suppose that the null hypothesis (H_0) is the statement that we test. Consider the case $(H_0) : \mu = a$, i.e. the population mean is equal to some specified value. Suppose that the fail to reject region A of T is given in terms of inequality on the statistic sample mean \overline{X} ,

$$A(\alpha) = \{(a_1, \dots, a_n) \in \mathbb{R}^n / \overline{X} < c\}$$

where α is the level of T and $c \in \mathbb{R}$. Suppose that we have a sample data (b_1, \ldots, b_n) such that $b_1 + \cdots + b_n < cn$. Then all $(a_1, \ldots, a_n) \in \mathbb{R}^n$ and such that $(a_1, \ldots, a_n) < (b_1, \ldots, b_n)$ for degree lex order belong to $A(\alpha)$. The assertion depends from the compatibility of the statistic sample mean \overline{X} with the degree lex order, hence $(a_1, \ldots, a_n) < (b_1, \ldots, b_n)$ implies $a_1 + \cdots + a_n \leq b_1 + \cdots + b_n < cn$.

2. The theory of circuits is a very interesting argument of linear algebra. Our interest is its application to study stochastic processes in statistic. Consider a stochastic matrix

$$\left(\begin{array}{ccccc} a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{23} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{array}\right)$$

such that $\sum_{j} a_{ij} = 1$.

EXAMPLE 2.2. Let p and q be two rational numbers such that p+q=1. Consider the

 5×6 - matrix

$$\left(\begin{array}{cccccc} p & q & 0 & 0 & 0 & 0 & 0 \\ 0 & p & q & 0 & 0 & 0 & 0 \\ 0 & 0 & p & q & 0 & 0 & 0 \\ 0 & 0 & 0 & p & q & 0 & 0 \\ 0 & 0 & 0 & 0 & p & q & 0 \end{array}\right)$$

Such a matrix describes a stochastic process that is called "random promenade along the integers" ([1]) and it represents the transition matrix M_j at the step j = 2. It is in echelon form and we have five circuits in the sense of [3]: $f_1 = pX_1 + qX_2$; $f_2 = pX_2 + qX_3$; $f_3 = pX_3 + qX_4$; $f_4 = pX_4 + qX_5$; $f_5 = pX_5 + qX_6$.

 M_j is a submatrix of the matrix

$$\left(\begin{array}{cccccc} 0 & q & 0 & 0 & 0 & 0 \\ p & 0 & q & 0 & 0 & 0 \\ 0 & p & 0 & q & 0 & 0 \\ 0 & 0 & p & 0 & q & 0 \\ 0 & 0 & 0 & p & 0 & q \\ 0 & 0 & 0 & 0 & p & 0 \end{array}\right)$$

called the transition matrix of the "random promenade along the integers".

DEFINITION 2.1. A state defined by a circuit will be called a circuital state.

DEFINITION 2.2. A stochastic matrix in the echelon form will be called strong stochastic.

If we reduce for rows a stochastic matrix , we have a matrix that in general is not stochastic.

We can consider matrices of indeterminates. By Hilbert Nullstellensatz theorem in computational form ([2],chap2), we can solve the problem to find stochastic states that are circuital, by using a Computer algebra system(CoCoA).

EXAMPLE 2.3. Consider the matrix of indeterminates

$$A = \begin{pmatrix} X_{11} & X_{12} & 0 & 0 & 0 \\ 0 & X_{22} & X_{23} & 0 & 0 \\ 0 & 0 & X_{33} & X_{34} & 0 \end{pmatrix}$$

Let $B = \mathbb{C}[X_{11}, X_{12}, X_{22}, X_{23}, X_{33}, X_{34}, Y_1, Y_2, Y_3, Y_4]$. We can utilize the theory of Gröbner bases to determine the solutions of the system

$$\begin{cases} X_{11} + X_{12} = 1 \\ X_{22} + X_{23} = 1 \\ X_{33} + X_{34} = 1 \\ X_{11}Y_1 + X_{12}Y_2 = 0 \\ X_{22}Y_2 + X_{23}Y_3 = 0 \\ X_{33}Y_3 + X_{34}Y_4 = 0 \end{cases}$$

V. Iorfida et al

The resolution of the system is equivalent to find the first sygyzy $\mathbb{C}[X_{11}, X_{12}, X_{22}, X_{23}, X_{33}, X_{34}]$ -module of the module that has A as a representation matrix. Moreover the solutions must satisfy $X_{11} + X_{12} = 1$, $X_{22} + X_{23} = 1$, $X_{33} + X_{34} = 1$.

References

- 1. R. Benevento, Calcolo delle probabilità Centro Editoriale Librario , *Università degli studi della Calabria*, 1992.
- W.Adams, P. Loustaunau, An Introduction to Gröbner Bases, Graduate Studies in Mathematics, Volume 3, Amer. Math. Soc., 1994.
- 3. M. Stoka, Calcolo delle probabilità e statistica, Mathematics, Levrotto Bella, 1986.
- 4. B.Sturmfels, Gröbner bases and Convex polytopes, Univ.Lect.Series, Vol.8, Amer.Math.Soc., 1995.