IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 14, Issue 6 Ver. I (Nov - Dec 2018), PP 01-06
www.iosrjournals.org

Matrix Form of The Bayes Theorem And Diagnostic Tests

María Magdala Pérez-Nimo[1] José A. Camúñez-Ruiz[1]
[1]*Departamento de Economía Aplicada I Universidad de Sevilla Facultad de Ciencias Económicas y Empresariales Avda. Ramón y Cajal 1 41008 Sevilla Spain*
Corresponding Author; María Magdala Pérez-Nimo

Abstract: *In order to solve certain problems in calculating probabilities, such as Markov chains or conditional specification of discrete distributions, the use of matrix and vector treatment of conditioned probabilities and of vectors of marginal probabilities is common. Following these ideas, the present study obtains matrix forms of some elementary results of probability theory, such as the total probability and Bayes theorems. These results and methodology are applied to the matrix study of results of diagnostic tests, allowing an immediate generalization to tests with more than two results. In addition,we propose safety and validity measures of a test based on matrix rules, which in some cases are related to the well-known Youden index.*

Key Word: *Conditioned Probabilities, BayesTheorem, Matrix, Diagnostic Tests*

Date of Submission: 01-11-2018 Date of acceptance: 15-11-2018

I. Introduction

In the Markov chain theory ([1];[2]) as well as in conditional specification problems for finite random variables [3], the matrix treatment of conditional probabilities is common. This line is followed in the present study, and the main results of conditional probabilities are obtained, including a matrix form of the Bayes' theorem. The results obtained, are applied to the study of diagnostic test.

II. Matrix Approach

Let us consider two complete sets of events, $A_1, A_2, ..., A_i, ..., A_n$ and $B_1, B_2, ..., B_j, ..., B_m$, both with non-zero probabilities $P(B_1), P(B_2), ..., P(B_j), ..., P(B_m)$.

We define the $P(A_1), P(A_2), ..., P(A_i), ..., P(A_n)$ vectors.

$$\boldsymbol{\alpha} = \begin{bmatrix} P(A_1) \\ P(A_2) \\ \vdots \\ P(A_n) \end{bmatrix} \qquad \boldsymbol{\beta} = \begin{bmatrix} P(B_1) \\ P(B_2) \\ \vdots \\ P(B_m) \end{bmatrix}$$

And the matrices

$$\mathbf{A} = ((a_{ij})) \qquad a_{ij} = P(A_i \mid B_j), \quad i = 1, 2, ..., n; \quad j = 1, 2, ..., m$$

$$\mathbf{A} = \begin{bmatrix} P(A_1 \mid B_1) & P(A_1 \mid B_2) & \dots & P(A_1 \mid B_m) \\ P(A_2 \mid B_1) & P(A_2 \mid B_2) & \dots & P(A_2 \mid B_m) \\ \vdots & \vdots & \ddots & \vdots \\ P(A_n \mid B_1) & P(A_n \mid B_2) & \dots & P(A_n \mid B_m) \end{bmatrix}$$

$$\mathbf{B} = ((b_{ij})) \qquad b_{ij} = P(B_j \mid A_i), \quad j = 1, 2, ..., m; \quad i = 1, 2, ..., n$$

$$\mathbf{B}=\begin{bmatrix} P(B_1|A_1) & P(B_2|A_1) & \ldots & P(B_m|A_1) \\ P(B_1|A_2) & P(B_2|A_2) & \ldots & P(B_m|A_2) \\ \vdots & \vdots & \ddots & \vdots \\ P(B_1|A_n) & P(B_2|A_n) & \ldots & P(B_m|A_n) \end{bmatrix}$$

Both matrices have $n\times m$ dimension.

We denote $\mathbf{1_n}$ the column vector of n dimension with all its components equal to one, and by $\mathbf{0_n}$ the column vector of n dimension with all components equal to zero. On the other hand, it should be remembered that a matrix is said to be stochastic by rows (columns) if it is non-negative and the sum of the elements in each row (column) is equal to one.

Note that $\mathbf{1_n^t A} = \mathbf{1_m^t}$ and $\mathbf{B1_m} = \mathbf{1_n}$, therefore, A is stochastic by columns and B is stochastic by rows.

In [3] Arnold and Press obtain the theorem 1:

Theorem 1. (Total probability theorem, matrix form).

$$\mathbf{A\beta}=\boldsymbol{\alpha} \qquad\qquad \mathbf{B^t}\boldsymbol{\alpha}=\boldsymbol{\beta}$$

Demonstration: we rely on the total probability theorem in its usual form:

$$\mathbf{A\beta}=\begin{bmatrix} P(A_1|B_1) & P(A_1|B_2) & \ldots & P(A_1|B_m) \\ P(A_2|B_1) & P(A_2|B_2) & \ldots & P(A_2|B_m) \\ \vdots & \vdots & \ddots & \vdots \\ P(A_n|B_1) & P(A_n|B_2) & \ldots & P(A_n|B_m) \end{bmatrix}\begin{bmatrix} P(B_1) \\ P(B_2) \\ \vdots \\ P(B_m) \end{bmatrix}=$$

$$=\begin{bmatrix} \sum_{j=1}^{m} P(A_1|B_j)P(B_j) \\ \sum_{j=1}^{m} P(A_2|B_j)P(B_j) \\ \vdots \\ \sum_{j=1}^{m} P(A_n|B_j)P(B_j) \end{bmatrix}=\begin{bmatrix} P(A_1) \\ P(A_2) \\ \vdots \\ P(A_n) \end{bmatrix}=\boldsymbol{\alpha}$$

Similarly, it is shown that $\mathbf{B^t}\boldsymbol{\alpha}=\boldsymbol{\beta}$.

Theorem 2. The matrices $\mathbf{AB^t}$ and $\mathbf{B^tA}$ have eigenvalue one with $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ eigenvectors associated with that value and, respectively, each matrix.

Demonstration: combining the formulas of the previous theorem we have:

$$\mathbf{A\beta}=\boldsymbol{\alpha} \rightarrow \mathbf{AB^t}\boldsymbol{\alpha}=\boldsymbol{\alpha}$$

$$\mathbf{B^t}\boldsymbol{\alpha}=\boldsymbol{\beta} \rightarrow \mathbf{B^tA}\boldsymbol{\beta}=\boldsymbol{\beta}$$

As a result:

$$\mathbf{AB^t}\boldsymbol{\alpha}=\boldsymbol{\alpha} \rightarrow (\mathbf{AB^t}-\mathbf{I_n})\boldsymbol{\alpha}=\mathbf{0_n} \rightarrow \det(\mathbf{AB^t}-\mathbf{I_n})=0$$

$$\mathbf{B^tA}\boldsymbol{\beta}=\boldsymbol{\beta} \rightarrow (\mathbf{B^tA}-\mathbf{I_m})\boldsymbol{\beta}=\mathbf{0_m} \rightarrow \det(\mathbf{B^tA}-\mathbf{I_m})=0$$

Theorem 3.

$$\boldsymbol{\alpha}^{\mathbf{t}}\mathbf{A}\boldsymbol{\beta}=\|\boldsymbol{\alpha}\|_2^2 \qquad \boldsymbol{\beta}^{\mathbf{t}}\mathbf{B^t}\boldsymbol{\alpha}=\|\boldsymbol{\beta}\|_2^2 \qquad \boldsymbol{\alpha}^{\mathbf{t}}\mathbf{B}\boldsymbol{\beta}=\|\boldsymbol{\beta}\|_2^2$$

Demonstration: Just need to apply Theorem 1.

$$\boldsymbol{\alpha}^{\mathbf{t}}\mathbf{A}\boldsymbol{\beta}=\boldsymbol{\alpha}^{\mathbf{t}}\boldsymbol{\alpha}=\|\boldsymbol{\alpha}\|_2^2$$

$$\boldsymbol{\beta}^{\mathbf{t}}\mathbf{B}^{\mathbf{t}}\boldsymbol{\alpha} = \boldsymbol{\beta}^{\mathbf{t}}\boldsymbol{\beta} = \|\boldsymbol{\beta}\|_2^2$$

On the other hand, the last equality is obtained by transposing the second.
Let **M** be the matrix of the probabilities of the intersections of the two complete sets of events:

$$\mathbf{M} = \begin{bmatrix} P(A_1B_1) & P(A_1B_2) & \dots & P(A_1B_m) \\ P(A_2B_1) & P(A_2B_2) & \dots & P(A_2B_m) \\ \vdots & \vdots & \ddots & \vdots \\ P(A_nB_1) & P(A_nB_2) & \dots & P(A_nB_m) \end{bmatrix}$$

Theorem 4. (Non-conditional total probability theorem).

$$\mathbf{1}_{\mathbf{n}}^{\mathbf{t}}\mathbf{M} = \boldsymbol{\beta}^{\mathbf{t}} \qquad \mathbf{M}^{\mathbf{t}}\mathbf{1}_n = \boldsymbol{\beta} \qquad \mathbf{M}\mathbf{1}_{\mathbf{m}} = \boldsymbol{\alpha}$$

Demonstration: Simply operate on the first member of each equation,

$$\mathbf{1}_{\mathbf{n}}^{\mathbf{t}}\mathbf{M} = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} P(A_1B_1) & P(A_1B_2) & \dots & P(A_1B_m) \\ P(A_2B_1) & P(A_2B_2) & \dots & P(A_2B_m) \\ \vdots & \vdots & \ddots & \vdots \\ P(A_nB_1) & P(A_nB_2) & \dots & P(A_nB_m) \end{bmatrix} =$$

$$= \begin{bmatrix} \sum_{i=1}^{n} P(A_iB_1) & \sum_{i=1}^{n} P(A_iB_2) & \dots & \sum_{i=1}^{n} P(A_iB_m) \end{bmatrix} = \boldsymbol{\beta}^{\mathbf{t}}$$

The second equality is obtained by transposition of the first. The third equality is demonstrated latter is tested as the first one.□ Let $\mathrm{M}_n(\mathbb{R})$ be the set of square matrices of *n* dimension. Let us now define the application $D_n : \mathbb{R}^n \to \mathrm{M}_n(\mathbb{R})$, which associates to each vector $\mathbf{v} \in \mathbb{R}^n$ the square matrix whose diagonal is $\mathbf{v}$ and the rest of elements are null. If there is no room for confusion we will omit the *n* subscript, that is, $D_n = D$. In this way, we obtain:

$$\mathbf{D}(\boldsymbol{\alpha}) = \begin{bmatrix} P(A_1) & 0 & \dots & 0 \\ 0 & P(A_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & P(A_n) \end{bmatrix} \qquad \mathbf{D}(\boldsymbol{\beta}) = \begin{bmatrix} P(B_1) & 0 & \dots & 0 \\ 0 & P(B_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & P(B_m) \end{bmatrix}$$

Theorem 5.

$$\mathbf{AD}(\boldsymbol{\beta}) = \mathbf{M} = \mathbf{D}(\boldsymbol{\alpha})\mathbf{B}$$

Demonstration:

$$\mathbf{AD}(\boldsymbol{\beta}) = \begin{bmatrix} P(A_1 \mid B_1) & P(A_1 \mid B_2) & \dots & P(A_1 \mid B_m) \\ P(A_2 \mid B_1) & P(A_2 \mid B_2) & \dots & P(A_2 \mid B_m) \\ \vdots & \vdots & \ddots & \vdots \\ P(A_n \mid B_1) & P(A_n \mid B_2) & \dots & P(A_n \mid B_m) \end{bmatrix} \begin{bmatrix} P(B_1) & 0 & \dots & 0 \\ 0 & P(B_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & P(B_m) \end{bmatrix} =$$

$$= \begin{bmatrix} P(A_1 | B_1)P(B_1) & P(A_1 | B_2)P(B_2) & \dots & P(A_1 | B_m)P(B_m) \\ P(A_2 | B_1)P(B_1) & P(A_2 | B_2)P(B_2) & \dots & P(A_2 | B_m)P(B_m) \\ \vdots & \vdots & \ddots & \vdots \\ P(A_n | B_1)P(B_1) & P(A_n | B_2)P(B_2) & \dots & P(A_n | B_m)P(B_m) \end{bmatrix} =$$

$$= \begin{bmatrix} P(A_1B_1) & P(A_1B_2) & \dots & P(A_1B_m) \\ P(A_2B_1) & P(A_2B_2) & \dots & P(A_2B_m) \\ \vdots & \vdots & \ddots & \vdots \\ P(A_nB_1) & P(A_nB_2) & \dots & P(A_nB_m) \end{bmatrix} = \mathbf{M}$$

The demonstration of the other equality is analogous.

Theorem 6.(Bayes' theorem, matrix form).

$$\mathbf{A} = \mathbf{D(\alpha)BD(\beta)^{-1}} \qquad \mathbf{B} = \mathbf{D(\alpha)^{-1}AD(\beta)}$$

Its proof is an immediate consequence of Theorem 5.

3.- Diagnostic tests.

Consider a clinical test for the diagnosis of a certain disease (see, [4]). Suppose, as is usual in this context, that each patient can be sick (event D) or healthy and that the test can only present positive results (event R) if it detects the disease, and negative results otherwise. Define the vectors:

Disease prevalence vector: $\boldsymbol{\alpha} = \begin{bmatrix} P(D) \\ P(\bar{D}) \end{bmatrix}$.

It contains, among other elements, the prevalence of the disease *P(D)*

Test results vector: $\boldsymbol{\beta} = \begin{bmatrix} P(R) \\ P(\bar{R}) \end{bmatrix}$.

Containing the probabilities of positive and negative results, $P(R)$ and $P(\bar{R})$.

We also define the matrices:

Test security matrix: $\mathbf{A} = \begin{bmatrix} P(D | R) & P(D | \bar{R}) \\ P(\bar{D} | R) & P(\bar{D} | \bar{R}) \end{bmatrix}$

It contains, among other elements, the predictive values of the test, $P(D | R)$ (positive) and $P(\bar{D} | \bar{R})$ (negative).

Test validity matrix: $\mathbf{B} = \begin{bmatrix} P(R | D) & P(\bar{R} | D) \\ P(R | \bar{D}) & P(\bar{R} | \bar{D}) \end{bmatrix}$.

It contains, among other elements, $P(R | D)$ sensitivity and $P(\bar{R} | \bar{D})$ specificity.

Test Matrix: $\mathbf{M} = \begin{bmatrix} P(D \cap R) & P(D \cap \bar{R}) \\ P(\bar{D} \cap R) & P(\bar{D} \cap \bar{R}) \end{bmatrix}$.

Containing the true positives $P(D \cap R)$, true negatives $P(\bar{D} \cap \bar{R})$, false positives $P(\bar{D} \cap R)$ and false negatives $P(D \cap \bar{R})$.

The relation between the test matrix and the prevalence and results vectors is then obtained in matrix form $\mathbf{M1_2} = \boldsymbol{\alpha}$ and $\mathbf{M^t1_2} = \boldsymbol{\beta}$: as well as the relation between all the elements of a diagnostic test: $\mathbf{AD(\beta)} = \mathbf{M} = \mathbf{D(\alpha)B}$.

On the other hand, applying the forms of the total probability theorem and the Bayes' theorem, we obtain $\mathbf{A} = \mathbf{D(B^t\alpha)^{-1}BD(\alpha)}$. This is the matrix version of well-known formulas of the type

$$TPV = \frac{prevalence\,x\,sensibility}{prevalence\,x\,sensibility + \left(1 - prevalence\right)\left(1 - specificity\right)}$$

,

which correlates prevalence, sensitivity and specificity with predictive values.
Note that the use of matrices allows an easy generalization to tests with several levels of diagnosis or to diseases with different typologies. In addition, it allows definingtest security measures and indexes in terms of distances.
The safety of a test usually requires the study of its predictive values. A test is completely safe if $P(D \mid R) = P(\bar{D} \mid \bar{R}) = 1$.That is, both predictive values are worth one. For these tests, the safety matrix A defined above is the unit . In this case, a safety measure for a test with the safety matrix **A**, can be defined as a certain distance from the matrix **A** to the identity $MS(\mathbf{A}) = d(\mathbf{I}, \mathbf{A}) = \| \mathbf{I} - \mathbf{A} \|$, for some matrix norm (See, for instance, [5]). For a completely unsafe test these measures are worth zero. For a completely useless test, that is, $P(D \mid R) = P(\bar{D} \mid \bar{R}) = 0$, the safety matrix is $\mathbf{J} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and ***MS(J)***must be maximal. Its value will depend on the norm used. Normalizing and transferring the *MS* measure,security indexes -*SI*- can be obtained, so that if the maximum value of *MS* is μ then $SI = 1 - \frac{MS}{\mu}$,taking values between 0 and 1. The test is completely safe if it takes the value 1 and completely useless when it takes the value 0. Let's look at some specific cases.

i.-**Row norm.**

$$MS_\infty(\mathbf{A}) = \| \begin{bmatrix} 1 - P(D \mid R) & -P(D \mid \bar{R}) \\ -P(\bar{D} \mid R) & 1 - P(\bar{D} \mid \bar{R}) \end{bmatrix} \|_\infty =$$
$$= máx\left\{ |1 - P(D \mid R)| + |-P(D \mid \bar{R})|, |-P(\bar{D} \mid R)| + |1 - P(\bar{D} \mid \bar{R})| \right\}$$

$$MS_\infty(\mathbf{A}) = P(\bar{D} \mid R) + P(D \mid \bar{R})$$
$$0 \le MS_\infty(\mathbf{A}) \le 2$$

By normalizing, we obtain the associated safety index,

$$SI_\infty(\mathbf{A}) = 1 - \frac{P(\bar{D} \mid R) + P(D \mid \bar{R})}{2}$$

ii.-**Column norm.**

$$MS_1(\mathbf{A}) = \| \begin{bmatrix} 1 - P(D \mid R) & -P(D \mid \bar{R}) \\ -P(\bar{D} \mid R) & 1 - P(\bar{D} \mid \bar{R}) \end{bmatrix} \|_1 =$$
$$= máx\left\{ |1 - P(D \mid R)| + |-P(\bar{D} \mid R)|, |-P(D \mid \bar{R})| + |1 - P(\bar{D} \mid \bar{R})| \right\}$$
$$MS_1(\mathbf{A}) = 2máx\left\{ P(\bar{D} \mid R), P(D \mid \bar{R}) \right\}$$
$$0 \le MS_1(\mathbf{A}) \le 2$$

By normalizing we obtain the associated safety index,

$$SI_1(\mathbf{A}) = 1 - \max\left\{ P(\bar{D} \mid R), P(D \mid \bar{R}) \right\}$$

iii.-**Euclidean or spectral norm.**

$$MS_s(\mathbf{A}) = \| \mathbf{I} - \mathbf{A} \|_s = \| \begin{bmatrix} 1 - P(D \mid R) & -P(D \mid \bar{R}) \\ -P(\bar{D} \mid R) & 1 - P(\bar{D} \mid \bar{R}) \end{bmatrix} \|_s$$

which is the square root of the greatest singular value. The singular values of that matrix are 0 and $2\left(P^2(\bar{D}\mid R)+P^2(D\mid\bar{R})\right)$ thus:

$$MS_s(\mathbf{A})=\|\mathbf{I}-\mathbf{A}\|_s=\sqrt{2\left(P^2(\bar{D}\mid R)+P^2(D\mid\bar{R})\right)}$$

$$0\le MS_s(\mathbf{A})\le 2$$

The associated safety index is $SI_s(\mathbf{A})=1-\sqrt{\frac{P^2(\bar{D}\mid R)+P^2(D\mid\bar{R})}{2}}$.

The same reasoning may be used with the validity matrix **B**and define test validity measures and indexes. It must be taken into account that for sensitivity and specificity equal to one, we have that **B** = **I**, $MV(\mathbf{B})=d(\mathbf{I},\mathbf{B})=\|\mathbf{I}-\mathbf{B}\|$ and the maximum validity is found for zero MV, while the maximum validity is given in the case of **B** = **J**.

Norm	***MV***	***Index***
Row	$P(R\mid\bar{D})+P(\bar{R}\mid D)$	$1-\frac{P(R\mid\bar{D})+P(\bar{R}\mid D)}{2}$
Column	$2máx\left\{P(R\mid\bar{D}),P(\bar{R}\mid D)\right\}$	$1-máx\left\{P(R\mid\bar{D}),P(\bar{R}\mid D)\right\}$
Spectral	$\sqrt{2\left(P^2(R\mid\bar{D})+P^2(\bar{R}\mid D)\right)}$	$1-\sqrt{\frac{P^2(R\mid\bar{D})+P^2(\bar{R}\mid D)}{2}}$

MV is bounded between 0 and 2 in the three cases studied.

Note that in the case of the row norm, the validity index $VI_\infty=\frac{1+Y}{2}$,where *Y* is the well-known Youden index ([6]).

In addition, this way of dealing with conditional probabilities can be used in teaching as a nexus between subjects containing elementary theory of probability and those containing elements of linear algebra, which are usually separated.

References.

[1]. Berman A. Plemmons R.J. (1994) Nonnegative Matrices in the Mathematical Sciences. SIAM. Piladelphia

[2]. Isaacson D.L. y Madsen R.W. (1976) Markov Chain: Theory and aplications. John Wiley and Sons. New York.

[3]. Arnold B.C. Press S. J. (1989) Compatible conditional distributions. Journal of the American Statistical Association. 84, 152-156

[4].] Gordis, L. (2014) Epidemiology. 5ª edición. Elsevier Saunders Philadelphia

[5]. Schott J.R.(1997) Matrix Analysis for Statistics John Wiley and sons New York.

[6]. Youden, W.J. (1950). Index for rating diagnostic tests. Cancer 3: 32–35.

María Magdala Pérez-Nimo. " Matrix Form Of The Bayes Theorem And Diagnostic Tests." IOSR Journal of Mathematics (IOSR-JM) 14.6 (2018): 01-06.