Statistical Bulletin
Reliability & Variation Research

DETROIT RESEARCH INSTITUTE
21900 GREENFIELD ROAD • OAK PARK, MICHIGAN 48237 • (313) 968-1818

LEONARD G. JOHNSON
EDITOR

WANG H. YEE
DIRECTOR

VOLUME 6 BULLETIN 8 FEBRUARY, 1977

A PROBABILISTIC GRAVITY MODEL FOR MEDICAL DIAGNOSIS

Published 8 times a year by the Detroit Research Institute. Subscription Rate: One year $40.00. Quantity rates on request.
Address subscription orders and editorial correspondence to:
Detroit Research Insitute, 21900 Greenfield Road, Oak Park, Michigan 48237

Volume 6

Bulletin 8 February, 1977

INTRODUCTION

In attempting to classify a medical patient into one of several possible disease categories, various types of factors are customarily employed. Among such factors are the following:

(1) Subjective questions with "Yes" or "No" or "Undecided" answers. These can be quantified by using +1 for "Yes", -1 for "No", and 0 (zero) for "Undecided".

(2) Subjective assessments of medical experts.

(3) Objective measurements, such as temperature, blood pressure, blood cell counts, EKG's.

(4) Special stimulus responses, such as reactions to allergens or antigens.

(5) A list of symptoms

(6) Pictorial criteria, such as X-Rays.

Each factor, defined by a quantitative measure of its level, has a particular distribution of possible values, i.e., a range of values, for each particular disease. The multi-dimensional set of all factors under study will then possess a certain type of multi-variate distribution function, which is uniquely defined for each particular disease.

Volume 6
Bulletin 8
February , 1977

GEOMETRIC REPRESENTATION OF DISEASES

Each disease can be represented as a multi-dimensional manifold in the multi-dimensional factor space, with each factor measured along along its own coordinate axis. The multi-dimensional manifold for a particular disease will represent the total range of values possible for all the N factors as they are involved in that disease. Furthermore, the multi-dimensional manifold representing a disease will have a specific centroid whose coordinates are the average values of the N factors as they are involved in that disease. Also, each factor in a disease will have a specific variance peculiar to that factor as it is involved in the population of patients with that disease. Thus, we can picture the N -Dimensional factor space as containing a collection of N-Dimensional manifolds (some of which may be intersecting each other), with each manifold representing a specific disease with a specific centroid and a specific variance for each factor in the N-dimensional factor space. This is simply illustrated for a three dimensional factor space with four diseases in FIGURE 1 .

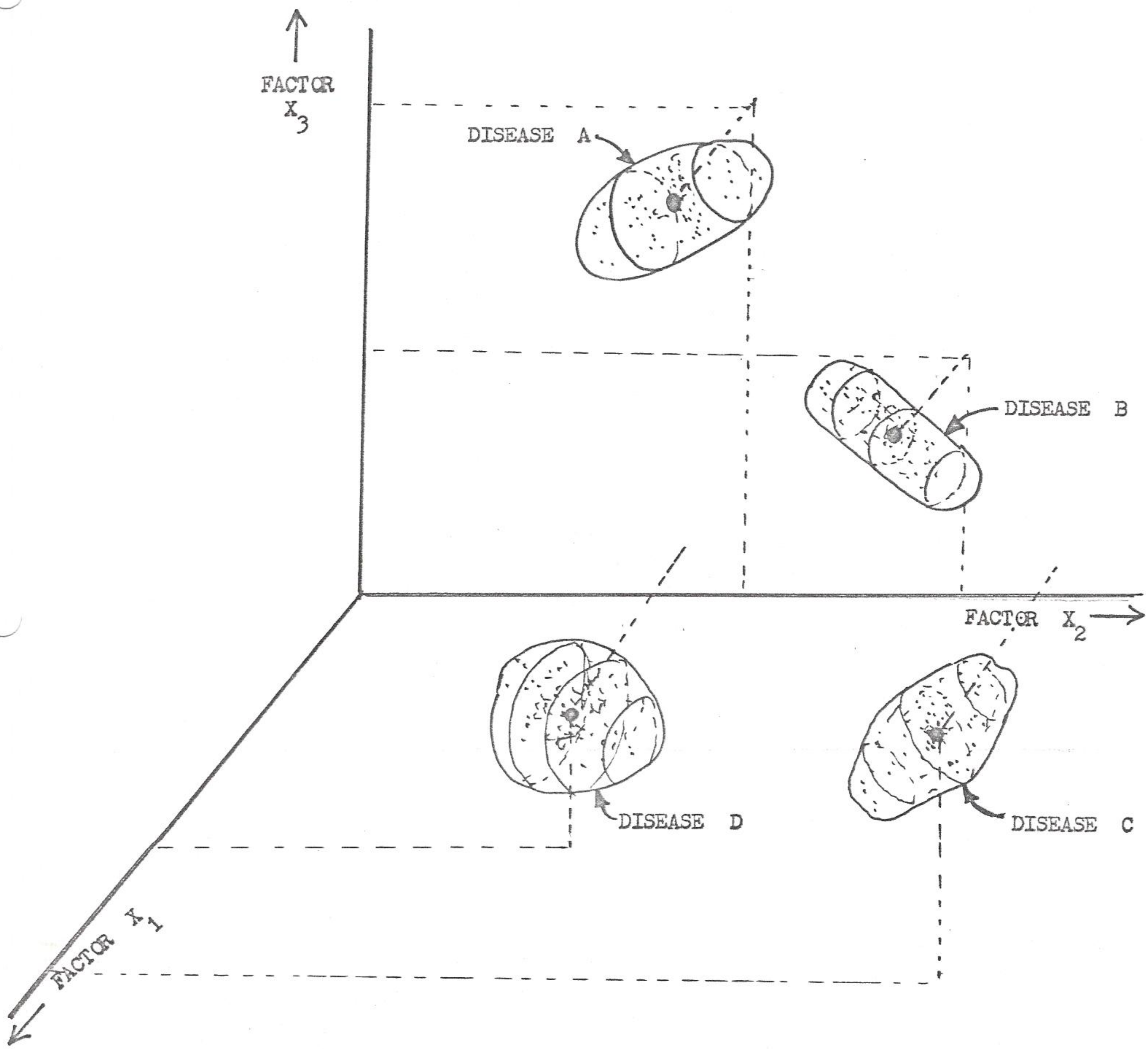
FACTOR
X_3
DISEASE A
DISEASE B
FACTOR X_2
DISEASE D
DISEASE C
FACTOR X_1

FIGURE 1

Volume 6

February, 1977

Bulletin 8

CLASSIFYING A NEW PATIENT INTO A DISEASE CATEGORY

A new medical patient will be represented by a set of N coordinates $(X_1, X_2, X_3, \ldots, X_N)$, where X_i is the measured value of factor i for the patient. Since the factor space is N-dimensional, there are N coordinates. Thus, the coordinate set $(X_1, X_2, X_3, \ldots, X_N)$ for a new patient is geometrically represented as a single point in the factor space containing the several N-dimensional manifolds which represent the several disease possibilities under consideration. The following question then arises: "Which disease does the patient who is represented by the point $(X_1, X_2, X_3, \ldots, X_N)$ have ? " To answer this question we mathematically construct a "GRAVITY MODEL" in which we define the "FORCE OF GRAVITY" exerted by each disease manifold on a UNIT MASS at the point $(X_1, X_2, X_3, \ldots, X_N)$. This "FORCE OF GRAVITY" exerted by a disease manifold on a unit mass at $(X_1, X_2, X_3, \ldots, X_N)$ is represented by the PROBABILITY that the point $(X_1, X_2, X_3, \ldots, X_N)$ could belong to the N-dimensional distribution represented by that disease manifold. Thus, the greatest force of gravity, i.e., the greatest pull on the unit mass at $(X_1, X_2, X_3, \ldots, X_N)$, would be exerted by the disease manifold for which $(X_1, X_2, X_3, \ldots, X_N)$ has the GREATEST BELONGING PROBABILITY.

The patient whose N factors are $(X_1, X_2, X_3, \ldots, X_N)$ would thus be diagnosed to have that disease for which the belonging probability of the point $(X_1, X_2, X_3, \ldots, X_N)$ is the greatest. The belonging probability of $(X_1, X_2, X_3, \ldots, X_N)$ to any particular disease can be calculated by using the centroid and factor variances of the particular disease manifold, together with the "DISTANCE" from the disease manifold's centroid to the point $(X_1, X_2, X_3, \ldots, X_N)$ representing the patient. Using the centroid and factor variances of each disease manifold, it is possible to calculate the PROBABILITY OF BEING AS FAR AWAY AS $(X_1, X_2, X_3, \ldots, X_N)$ from each disease centroid. That disease yielding the greatest probability of containing $(X_1, X_2, X_3, \ldots, X_N)$ in its manifold will then be selected as the patient's disease according to this "PROBABILISTIC GRAVITY MODEL" . In some cases it may be that the patient has none of the diseases in our selected list (when all the probabilities are very small) , while in other cases the patient may have several of the selected diseases simultaneously.

CALCULATING THE STANDARDIZED GRAVITATIONAL DISTANCE FROM A PATIENT'S POINT $(X_1, X_2, X_3, \ldots, X_N)$ TO THE CENTROID OF A DISEASE MANIFOLD

Consider the geometric representation of a disease manifold and a patient as shown in FIGURE 2 .

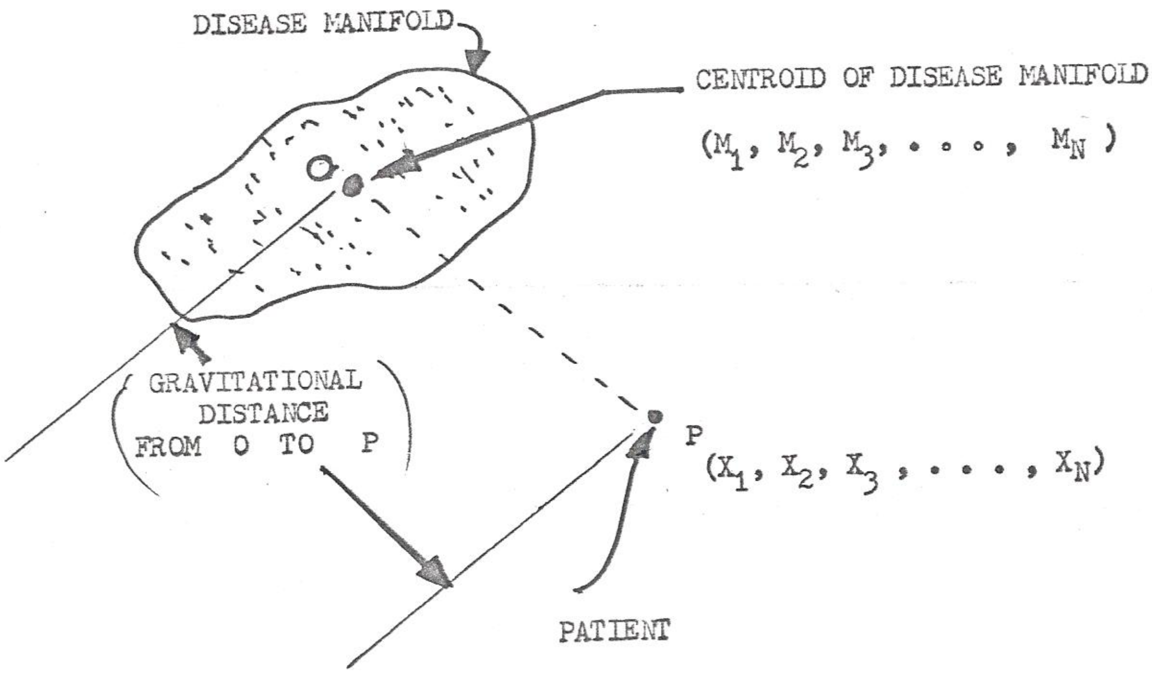

FIGURE 2

Let the centroid (Point O) of the disease manifold be $(M_1, M_2, M_3, \ldots, M_N)$,

where M_1 = Ave. of Factor X_1 for the disease
M_2 = Ave. of Factor X_2 for the disease
$\vdots$
M_N = Ave. of Factor X_N for the disease

Furthermore, let

σ_1^2 = Variance of Factor X_1 for the disease
σ_2^2 = Variance of Factor X_2 for the disease
$\vdots$
σ_N^2 = Variance of Factor X_N for the disease

Now, let

$\hat{Z}$ = Standardized Gravitational Distance between $(M_1, M_2, \ldots, M_N)$ and $(X_1, X_2, \ldots, X_N)$.

Then define

$$\hat{Z} = \sqrt{\sum_{i=1}^{N} \left(\frac{X_i - M_i}{\sigma_i}\right)^2}$$

If the disease manifold is represented as a multivariate normal distribution of N independent factors, then the probability $\hat{P}$ of a patient being at least as far away from the diseases manifold's centroid as $(X_1, X_2, \ldots, X_N)$ is

$$\hat{P} = 1 - V_{N+1}(\hat{Z})$$

where

$V_{N+1}(\hat{Z})$ = (N+1)-DIMENSIONAL VOLUME under an N-VARIATE NORMAL DISTRIBUTION to a radius

$$\rho = \sqrt{\sum_{i=1}^{N} (X_i - M_i)^2}$$

in the N-Dimensional space spanned by the N factors.

Since $V_{N+1}(\hat{Z})$ is a positive number between 0 and 1 , it follows that the probability $\hat{P}$ is a maximum for that disease for which $\hat{Z}$ is a minimum in the collection of diseases being considered.

Volume 6 , Bulletin 8 February , 1977

METRIC CRITERION FOR DISEASE SELECTION FOR A PATIENT

The PROBABILISTIC GRAVITATIONAL FORCE between the disease manifold of FIGURE 2 and a unit mass at the patient's point P is

$$\hat{P} = 1 - V_{N+1}(\hat{Z})$$

where

$$\hat{Z} = \sqrt{\sum_{i=1}^{N} \left(\frac{X_i - M_i}{\sigma_i} \right)^2}$$

is the STANDARDIZED GRAVITATIONAL DISTANCE between the disease manifold's centroid and the patient's point. (See foot note below.)

Since, as stated earlier, the maximum $\hat{P}$ corresponds to the minimum $\hat{Z}$, it follows that a minimum $\hat{Z}$ (i.e., a minimum standardized gravitational distance) can be used as our criterion for the disease selected to correspond to the patient located at $(X_1, X_2, \ldots, X_N)$. This type of diagnosis (selecting the disease with the minimum $\hat{Z}$) is what is known as "METRIC DIAGNOSIS" . It amounts to the same thing as "PROBABILISTIC GRAVITATIONAL FORCE DIAGNOSIS".

The advantage of using $\hat{Z}$ is that we need not get involved with the mathematical complications of the (N+1)-Dimensional Volume under an N-Dimensional Distribution Function, knowing full well that a smaller $\hat{Z}$ must always produce a greater "GRAVITATIONAL PULL" on a unit mass at $(X_1, X_2, \ldots, X_N)$.

NOTE: In general, $\hat{Z}$ turns out to be the SQUARE ROOT OF TWICE THE POLAR ENTROPY for independent factors.

CALCULATING THE ACTUAL PROBABILISTIC GRAVITATIONAL FORCE FOR NORMALLY DISTRIBUTED INDEPENDENT FACTORS

The actual probability that a patient as far out as $(X_1, X_2, \ldots, X_N)$, or farther, belongs to a disease manifold with centroid at $(M_1, M_2, \ldots, M_N)$ and with factor variances of $(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)$ is

$$\hat{P} = 1 - \frac{\left[\text{INCOMPLETE } \Gamma\left(\frac{N}{2}\right)\right]_0^{\frac{\hat{Z}^2}{\sqrt{2N}}}}{\Gamma\left(\frac{N}{2}\right)}$$

where

$$\hat{Z}^2 = \left(\frac{X_1 - M_1}{\sigma_1}\right)^2 + \left(\frac{X_2 - M_2}{\sigma_2}\right)^2 + \ldots + \left(\frac{X_N - M_N}{\sigma_N}\right)^2$$

GENERAL FORMULATION WITH CORRELATIONS

Any Non-Normal factors in the N -Space must be NORMALIZED in order to make normal probability theory valid. Furthermore, suppose the complete CORRELATION MATRIX of the NORMALIZED FACTORS is

$$R \equiv \begin{pmatrix} 1 & \rho_{21} & \cdots & \rho_{N1} \\ \rho_{12} & 1 & \cdots & \rho_{N2} \\ \rho_{13} & \rho_{23} & \cdots & \rho_{N3} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1N} & \rho_{2N} & \cdots & 1 \end{pmatrix}$$

Then, the INVERSE CORRELATION MATRIX will be

$$R^{-1} \equiv \begin{pmatrix} \frac{C_{11}}{\Delta} & \frac{C_{12}}{\Delta} & \cdots & \frac{C_{1N}}{\Delta} \\ \frac{C_{12}}{\Delta} & \frac{C_{22}}{\Delta} & \cdots & \frac{C_{2N}}{\Delta} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{C_{1N}}{\Delta} & \frac{C_{2N}}{\Delta} & \cdots & \frac{C_{NN}}{\Delta} \end{pmatrix}$$

where Δ is the value of the DETERMINANT OF R , and the C_{ij} 's are CO-FACTORS.

We define the POLAR ENTROPY of a patient's vector $(X_1, X_2, \ldots, X_N)$ to be $\hat{\mathcal{E}} = \frac{1}{2} W' R^{-1} W$, where W' is the row vector $(Z_1, Z_2, \ldots, Z_N)$ and W is the column vector $\begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_N \end{bmatrix}$, and $Z_i = \frac{X_i - M_i}{\sigma_i}$.

In more conventional symbolism, the POLAR ENTROPY of point $(X_1, X_2, \ldots, X_N)$ can be expressed as

$$\hat{\mathcal{E}} = \frac{\sum_{i=1}^{N} C_{ii} Z_i^2 + 2 \sum_{i<j}^{\text{ALL DISTINCT PAIRS } (i,j)} C_{ij} Z_i Z_j}{2 \Delta}$$

Then , the BELONGING PROBABILITY of a vector with as much POLAR ENTROPY (or more) as $(X_1, X_2, \ldots, X_N)$ to a disease manifold with centroid at $(M_1, M_2, \ldots, M_N)$ and factor variances $(\sigma_1^2, \sigma_2^2, \ldots \sigma_N^2)$ is given by

$$\hat{P} = 1 - \frac{\left[\text{INCOMPLETE } \Gamma\left(\frac{N}{2}\right)\right]_0^{\hat{\mathcal{E}}}}{\Gamma\left(\frac{N}{2}\right)}$$

$$= \frac{1}{\Gamma\left(\frac{N}{2}\right)} \int_{\hat{\mathcal{E}}}^{\infty} X^{\frac{N}{2} - 1} e^{-X} dX$$

This amazingly simple and general formula is not just an approximation, but is theoretically exact !