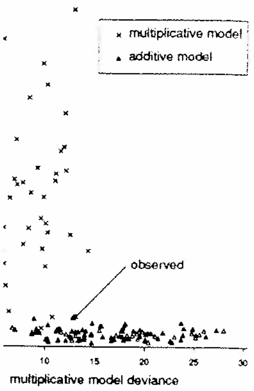
In: L. Fahrmeir, B. Francis, R. Gilchrist, G. Tutz (Eds.) (1992) Advances in GLIM and Statistical Modelling, Proccedings of the GLIM 92 Conference and the 7th International Workshop on Statistical Modelling, Munich, 13-17 July 1992. Berlin: Springer-Verlag (Lecture Notes in Statistics 78)



ary Heart Disease - bootstrap
iance plot

Statistical Modelling in GLIM.

(with Discussion) J. R. Statist.

ord University Press, Oxford.
on of Statistics (A.C. Atkinson

Fourth Berkeley Symposium on

otheses. J.R. Statist. Soc. B.

Proceedings 6th International
P.G.M van der Heijden, eds).

inear models. Computational

nparison of non-nested gener-
lemiology. Appl. Statist., 36.

determine the pattern of en-
32.

A class of models for the simultaneous analysis of square contingency tables

Peter G.M. van der Heijden and Wim Jansen
Department of Sociology, University of Utrecht
Postbus 80.140, 3508 TC Utrecht, the Netherlands

Abstract

A class of models is presented for the analysis of square contingency tables. The models fall in the class of loglinear models or models with logbilinear terms for the association. The models in this class differ in three ways: 1. the association is either assumed to be symmetric or asymmetric 2. the association is assumed to be completely different in each subtable, to have the same form but having different strength, or to be the same and having the same strength 3. for each subtable separately the association that is proposed is full, or has a logbilinear form, or is uniform. An example from research on social mobility will be discussed. The stability of the parameter estimates is studied with the jackknife.

1. Introduction

For the analysis of square contingency tables many models are available in the loglinear framework. More recently this abundance of models is extended by a way of parsimonious modeling of the association that makes use of a logbilinear term. This extension is quite popular now.

In this paper our aim is to apply these ideas, that are developed in the context of two-way tables, for the analysis of *sets* of square contingency tables. We first describe the ideas that we adopt from the context of square two-way contingency tables (section 2). Then we develop these ideas for the analysis of sets of square two-way tables (section 3). Many ideas presented in this section appeared earlier in the literature (see Agresti, 1983, 1990; Becker and Clogg, 1989; Becker, 1989, 1990; Choulakian, 1988; Clogg, 1982; Goodman, 1979, 1985, 1986; Mooijaart, in press). Thus, although the precise formulation and presentation of many of the models is new, this paper should be seen as an application paper.

2. Some models for square two-way contingency tables

Let m_{ij} be the expected frequency for category i (i=1,...,I) of the row variable and category j (i=1,...,J) of the column variable. Since the table is square, I=J. The saturated loglinear model is

$$\log m_{ij} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)} \qquad (1)$$

Some restrictions are necessary to identify this model. We choose the usual ANOVA-type constraints $\Sigma_i u_{1(i)} = \Sigma_j u_{2(j)} = \Sigma_i u_{12(ij)} = \Sigma_j u_{12(ij)} = 0$.

A recent development that stimulated much new research is the proposal by Goodman (1979) and Andersen (1980), to model the association $u_{12(ij)}$ parsimoniously as $u_{12(ij)} = \lambda v_i w_j$ with identifying restrictions $\Sigma_i v_i = \Sigma_j w_j = 0$ and $\Sigma_i v_i^2 = \Sigma_j w_j^2 = 1$. Model (1) with $u_{12(ij)} = \lambda v_i w_j$ is called the RC-association model. The parameters v_i and w_j can be interpreted as scores for the row and column categories, and, due to the identifying restrictions, the parameter λ indicates the association strength. Another idea is to fix the parameters v_i and w_j to some scores, if there is any theoretical reason to do so. If the categories i and j are ordered, then often used possibility is to fix these parameters to the equidistant scores.

In the context of square contingency tables an often asked question is whether the association between i and j is symmetric, i.e. whether $u_{12(ij)} = u_{12(ji)}$. Model (1) with this restriction is called the quasi-symmetry model. By imposing the additional constraint $u_{1(i)} = u_{2(j)}$ we find the so-called symmetry model, but this restriction is not our first interest here. A restriction like $u_{12(ij)} = u_{12(ji)}$ leads for the RC association model to $v_i = w_j$.

In modeling square tables attention also often goes out to the diagonal, since for the diagonal cells the row category is identical to the column category. Processes that lead to the off-diagonal association will very often be different from the diagonal association. In order to be able to study the off-diagonal association without bothering about the diagonal association, sometimes diagonal cells are defined as structural zeros, i.e. the likelihood is maximized over the off–diagonal cells only. Another way to reach this effect is to add a separate parameter for each diagonal cell. So in this case such a parameter is $\delta_{ij}u_i$, where $\delta^{ij} = 1$ if i=j, and $\delta^{ij} = 0$ else.

So, concluding, we have sketched some ideas that have been used in the past for the analysis of square two–way tables. One idea is to investigate whether the association is symmetric. A second idea is whether the association can be modelled as RC–association. And a third idea is to give special attention to the diagonals. Many more ideas have been proposed for the analysis of square two–way contingency tables, but these can be implemented in a straightforward way into the ideas for the analysis of *sets* of square two–way tables that we will introduce below (for references, see section 1).

3. A class of models for the analysis of a set of square contingency tables

For three–way tables the situation becomes more complicated, because we can generalize the models discussed above along different lines. Let us start again from the saturated model. Let there be K square contingency tables indexed by k (k=1,...,K). The expected frequency of cell (i,j,k) is denoted as m_{ijk}. The saturated model is for this three–way table is

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{13(ik)} + u_{23(jk)} + u_{12(ij)} + u_{123(ijk)} \qquad (2)$$

Below we will focus on particular ways to restrict (2). Not all possible restrictions will be considered. The restrictions are:

1. the first order interactions $u_{13(ik)}$ and $u_{23(jk)}$ are unconstrained. The interaction $u_{13(ik)}$ takes care of the fact that for row i the margins of the K square tables may be different. This does not really interest us in this paper.

2. in all models we are not interested in the diagonal cells of the square tables. Therefore we would like eliminate the effect of the diagonal cells on parameters that are also used to model effects for the off–diagonal cells.

The focus described in points 1 and 2 is not really essential. They are only introduced to structure the discussion below. In applications they can easily be dropped.

We now discuss our proposals. The models we will discuss differ along three dimensions. First, either symmetry is assumed in each of the K square tables, or symmetry is not assumed. Second, either the association in each of the K square tables can be completely different, or it is only different in strength but otherwise the same, or it is completely identical. Third, either the association in each of the square tables is unrestricted, or it is constrained to follow a constraint similar to $\lambda v_i w_j$, or it is constrained even further by fixing the scores v_i and w_j.

To simplify the discussion, we introduce the models in two groups. One group of proposals pertains to an assumption of asymmetry (i.e. symmetry is not assumed). This group is discussed in section 3.1. The other group of proposals pertains to an assumption of symmetry. This group is discussed in section 3.2. In order to focus attention on the most interesting part of the model, we rewrite (2) as $\log m_{ijk} = h_{123(ijk)} + u_{12(ij)} + u_{123(ijk)}$, where $h_{123(ijk)} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{13(ik)} + u_{23(jk)}$.

3.1. Asymmetric versions

We will start with the most general model for the set of square contingency tables. This model is the saturated model (8). We denote it as model C1:

$$\log m_{ijk} = h_{123(ijk)} + u_{12(ij)} + u_{123(ijk)} \qquad (C1)$$

A first property of this model is that the association is not symmetric (this holds for all the models to follow in section 3.1, so we will not mention this property again but focus instead on other properties). A second property is that the association may be different in each of the K square tables. The part of the association that the K tables have in common is parameterized by the $u_{12(ij)}$. This may be interpreted as the average association over the K tables. The second order interaction $u_{123(ijk)}$ shows how the association for square table k differs from this average association. A third property is that the

association is unconstrained. Basically, C1 comes down to fitting a saturated model to each square table separately.

The second model, C2, is

$$\log m_{ijk} = h_{123(ijk)} + \delta^{ij} u_{D(i)k} + \lambda_{(k)} v_{i(k)} w_{j(k)} \qquad (C2)$$

In this model we have eliminated the influence of the diagonal cells by introducing the term $\delta^{ij} u_{D(i)k}$ for each distinct square table k. More interesting is that, compared with C1, the associations $u_{12(ij)}$ and $u_{123(ijk)}$ are now replaced by the term $\lambda_{(k)} v_{i(k)} w_{j(k)}$. Whereas in C1 a saturated model is fitted to each square table k separately, C2 can be understood as a model in which in each of the K square tables a separate model of the form $\lambda v_i w_j$ is fitted. Over the K tables the parameters $\lambda_{(k)}$, $v_{i(k)}$, and $w_{j(k)}$ are not in any way restricted (apart from identifying restrictions). So the association may be completely different in each of the square tables, but the association is restricted in each of the square tables in a similar way.

The third model, C3, is

$$\log m_{ijk} = h_{123(ijk)} + \delta^{ij} u_{D(i)k} + \lambda_{(k)} v^*_{i(k)} w^*_{j(k)} \qquad (C3)$$

where the symbols '*' indicate that the parameters $v_{i(k)}$ and $w_{j(k)}$ are fixed to some predetermined scores. The parameters λ_k are still free parameters.

The models C1, C2 and C3 are nested, in the sense that C3 is a special case of C2, and C2 is a special case of C1. We will now discuss three other models that are nested in models C1, C2 and C3.

The fourth model is

$$\log m_{ijk} = h_{123(ijk)} + \delta^{ij} u_{D(i)k} + \lambda_k u_{12(ij)} \qquad (C4)$$

In this model there is first order interaction between i and j for each table k. The basic form of this interaction, denoted by $\{u_{12(ij)}\}$, is identical in each table k, but the parameter λ_k makes that this basic form gets a different strength in each table k.

We have to identify this last term, and we do this, first by setting $u_{12(ii)} = 0$, and, second, by imposing the restriction $\Sigma_k \lambda_k / K = 1$. Thus we can rewrite the term $\lambda_k u_{12(ij)} = u_{12(ij)} + \lambda_k^* u_{12(ij)}$, where $\lambda_k^* = \lambda_k - 1$. This shows that the basic form of the interaction is the first–order interaction described by $u_{12(ij)}$, and the different strengths generate second order interaction described by $\lambda_k^* u_{12(ij)}$.

If $\lambda_k > 1$, then the basic form of the interaction as defined by $u_{12(ij)}$ is larger than average in layer k, and if $0 < \lambda_k < 1$ then it is smaller than average in layer k. If $\lambda_k < 0$, then the form of the interaction in layer k has an opposite form compared to the average form.

Model C4 is a special case of model C1: in C4 the first–order interaction in each table k is different, but it only differs in strength.

In model C5 we get a special case of both model C2 as well as model C4:

$$\log m_{ijk} = h_{123(ijk)} + \delta^{ij} u_{D(i)k} + \lambda_k v_i w_j \qquad (C5)$$

If v_i and w_j are restricted as in (4), no further identifying restrictions have to be made. Similar to C4, the basic interaction is $v_i w_j$, and λ_k defines the strength of this interaction in layer k. Note that, similar to C4, we can rewrite $\lambda_k v_i w_j = \lambda v_i w_j + \lambda_k^* v_i w_j$, where $\lambda_k^* = \Sigma_k \lambda_k / K$, showing that the first order interaction is defined by $\lambda v_i w_j$ and the second order interaction is defined by $\lambda_k^* v_i w_j$.

C5 is a special case of C4 since the interaction $u_{12(ij)}$ is further constraint. C5 is a special case of C2 since there is basic interaction defined by $v_i w_j$ that is used as a building block in each table k.

In model C6 we simply have model C5 with fixed parameters v_i and w_j, denoted by v_i^* and w_j^*:

$$\log m_{ijk} = h_{123(ijk)} + \delta^{ij} u_{D(i)k} + \lambda_k v_i^* w_j^* \qquad (C6)$$

Only the strength parameters λ_k have to be estimated. It will be clear that C6 is a special case of C5 and of C3: it is a special case of C5 because in C6 some parameters are fixed that are free in C5; it is a special case of C3 since in C3 the fixed parameters may be different for each table k, whereas they have to be identical in C6.

The last three models that we discuss are similar to the triple C1, C2, C3, and the triple C4, C5, C6. Now the interaction between i and j is completely identical in each table k. This means that there is only first–order interaction, and that the second order interaction is absent. Thus we find the models

$$\log m_{ijk} = h_{123(ijk)} + \delta^{ij} u_{D(i)k} + u_{12(ij)} \qquad (C7)$$

$$\log m_{ijk} = h_{123(ijk)} + \delta^{ij} u_{D(i)k} + \lambda v_i w_j \qquad (C8)$$

$$\log m_{ijk} = h_{123(ijk)} + \delta^{ij} u_{D(i)k} + \lambda v_i^* w_j^* \qquad (C9)$$

The relation between these models is displayed in the following diagram:

C1 →	C4 →	C7
↓	↓	↓
C2 →	C5 →	C8
↓	↓	↓
C3 →	C6 →	C9

Arrows indicate that the model pointed at is a more restricted version of the model from which the arrow departs.

In going from row 1 to row 2 to row 3 (e.g. from C1 to C2 to C3), the association between the row and the column variable in each table k is restricted by using less parameters. In going from column 1 to column 2 to column 3 (e.g. from C1 to C4 to C7), the association between the row and column variable in each table k is restricted by making it more similar to the interaction in the other tables.

In the next section we will show that all models can be restricted further by assuming that the association in each table k is symmetric.

3.2. Symmetric versions

In many applications it will be useful to investigate whether the association between i and j is symmetric, i.e. whether the association in cell (i,j) of table k is identical to the association in cell (j,i) of table k. This can be investigated for all tables in section 3.1. It implies that the nine models described there have their symmetric versions. These symmetric versions are

A1, like C1, with $u_{12(ij)} = u_{12(ji)}$ and $u_{123(ijk)} = u_{123(jik)}$

A2, like C2, with $v_{i(k)} = w_{j(k)}$

A3, like C3, with $v_{i(k)}^* = w_{i(k)}^*$

A4 and A7, like C4 and C7, with $u_{12(ij)} = u_{12(ji)}$

A5 and A8, like C5 and C8, with $v_i = w_i$

A6 and A9, like C6 and C9, with $v_i^* = w_j^*$

Model A1 is equivalent to quasi–symmetry defined for each table k separately. Model A2 is a constrained version of quasi–symmetry fitted to each table k separately.

4. Example: a comparison of British and Danish mobility

The example deals with a comparison of a British occupational mobility table (Glass, 1954) with a Danish one (Svalastoga, 1959). An earlier comparison of these data by standardizing margins of each of the tables can be found in Bishop, Fienberg and Holland (1975).

Considering the models C1 to C9 and A1 to A9, we might want to go in a structured way through them. The path we choose is a bit arbitrary, but for this example different paths lead all to the same end result. The objective will be to start with the saturated model C1, and to come as close as possible to A9, because this is the most restrictive model. (For the fixed scores v_i^* and w_j^* we use equidistant scores −2, −1, 0 1 2, and therefore models C9 and A9 are equivalent, and C3, A3, C6 and A6 are equivalent).

Table 1: models fitted to the British and Danish social mobility tables

	Non–symmetric models:				Symmetric models			
Terms	models	Df	G^2	X^2	models	Df	G^2	X^2
$u_{123(ijk)}$	C1	0	0	0	A1	12	11.1	11.0
$\lambda_{(k)} v_{i(k)} w_{j(k)}$	C2	8	9.9	9.8	A2	14	15.0	15.6
$\lambda_{(k)} v_{i(k)}^* w_{j(k)}^*$	C3	20	29.5	34.8	A3	20	29.5	34.8
$\lambda_k u_{12(ij)}$	C4	10	14.1	13.6	A4	15	19.8	19.4
$\lambda_k v_i w_j$	C5	14	18.4	18.3	A5	17	22.4	23.0
$\lambda_k v_i^* w_j^*$	C6	20	29.5	34.8	A6	20	29.5	34.8
$u_{12(ij)}$	C7	11	24.6	23.6	A7	16	29.1	27.8
$\lambda v_i w_j$	C8	15	28.4	27.1	A8	18	30.9	29.6
$\lambda v_i^* w_j^*$	C9	21	39.1	39.4	A9	21	39.1	39.4

Starting in C1, we might go to C2. In C2 the rows as well as the columns of each of the two tables are scaled in an optimal way, which may be different for each of the tables. This models fits nicely (G^2 = 9.9, df = 8). A natural question is then to ask whether the scalings for the Danish table are the same as those for the British table. This is the case: model C5 is not significant (G^2 = 18.4, df = 14), and the difference between models C5 and C8 is not significant either. We then might want to know whether the row scalings are identical to the column scalings. This corresponds with model A5, and A5 cannot be rejected (G^2 = 22.4, df = 17; the difference between C5 and A5 is not significant either, G^2= 4.0, df = 3). Subsequently we would like to know whether the scalings are equidistant (model A6). This test is unclear: the likelihood ratio chi–square is not significant at p=.05, G^2 = 29.5 (df = 20), but the Pearson chi–square is significant at p=.05: X^2 = 34.8. Therefore we rather stick to model A5. As a final test we test whether the strength of the association as parameterized by λ_k is the same in each of the tables, i.e. whether it can be replaced by λ. This is not the case: model A8 is significant, G^2 = 30.9 (df = 18), which is significant at p = .05, and the difference between A5 and A8 is also significant. Therefore we choose model A5 as our final model.

The parameter estimates for $v_i = w_i$ are −2.28, −1.84, −.57, .50, .94 for levels 1 to 5 respectively. the parameter estimates are identified by imposing the restriction that $\Sigma_i p_i \hat{v}_i = 0$ and $\Sigma_i p_i \hat{v}_i^2 = 1$, where $p_i \equiv (p_{i++}+p_{+i+})/2$. For i≠j we find by $\hat{v}_i \hat{w}_j$ the off–diagonal association for cells (i,j) and (j,i) that both tables have in common. This shows that it is relatively easy to go from levels 1 to 2, and 4 to 5, but it is much harder to go from 2 to 3, and 3 to 4. Making more than two occupational steps is even harder. The association strength parameter estimates for λ_k are .356, .474, showing that the association is much stronger in Denmark than in Britain. For more details concerning parameter interpretation, we refer to Agresti (1990).

In order to study the stability of the parameter estimates, we carried out a jackknife study (compare van der Burg and de Leeuw, 1988). We have written a special purpose program for fitting the model that uses the uni–dimensional Newton algorithm. This procedure converges very slowly (the likelihood is rather flat), and therefore we let the algorithm stop when the increase in the likelihood was smaller than 10^{-12}. The program was written in APL68000, which has 16 digit accuracy. We think that the numerical instability due to APL is small, but that the numerical instability caused by the algorithm can be large. The jackknife provides estimates of the population parameters, the bias, and approximations of the standard errors. The results are shown in table 2. We find the 95 % confidence intervals (+/– 1.96SE) rather large, given the sample size of 5891. The parameter estimates for λ_k are biased upwards.

Table 2: Jackknife results: column 1: sample estimates; column 2: population estimates; column 3: bias; column 4: standard errors.

λ_1	.3570	.3488	.0082	.0289
λ_2	.4744	.4617	.0127	.0391
v_1	2.2752	2.2901	–.0149	.2935
v_2	1.8460	1.8548	–.0088	.1274
v_3	.5707	.5777	–.0070	.0792
v_4	–.5044	–.5067	.0023	.0974
v_5	–.9393	–.9480	.0087	.1220

References

Agresti, A. (1983) A survey of strategies for modeling cross–classifications having ordinal variables. *Journal of the American Statistical Association, 78*, 184–198.

Agresti, A. (1990) *Categorical data analysis*. New York: John Wiley & Sons.

Andersen, E.B. (1980) *Discrete statistical models with social science applications*. Amsterdam: North–Holland Publ. Cy.

Becker, M.P. (1989). Models for the analysis of association in multivariate contingency tables. *Journal of the American Statistical Association, 84*, 1014–1019.

Becker, M.P. (1990) Quasisymmetric models for the analysis of square contingency tables, *Journal of the Royal Statistical Society, Series B*, 26, 35–50.

Becker, M.P. & C.C. Clogg (1989) Analysis of sets of two–way contingency tables using association models, *Journal of the American Statistical Association*, 83, 142–156.

Bishop, Y.M.M., S.E. Fienberg & P.W. Holland (1975) *Discrete multivariate analysis. Theory and practice*. Cambridge, Mass.: MIT Press.

Choulakian, V. (1988) Exploratory analysis of contingency tables by loglinear formulation and generalizations of correspondence analysis. *Psychometrika, 53*, 235–250.

Clogg, C.C. (1982) Some models for the analysis of association in multiway cross–classifications having ordered categories, *Journal of the American Statistical Association*, 77, 803–815.

Glass. D.V. (ed.) (1954) *Social mobility in Britain*. London: Routledge and Kegan Paul.

Goodman, L.A. (1979) Simple models for the analysis of association in cross–classifications having ordered categories, *Journal of the American Statistical Association*, 74, 537–552.

Goodman, L.A. (1985) The analysis of cross–classified data having ordered and/or unordered categories: association models, correlation models, and asymmetry models for contingency tables with or without missing entries, *The Annals of Statistics*, 13, 10–69.

Goodman, L.A. (1986) Some useful extensions of the usual correspondence analysis approach and the usual log–linear models approach in the analysis of contingency tables. *International statistical review, 54*, 243–309.

Mooijaart, A. (in press). Three factor interaction models by log–trilinear terms in three–way contingency tables. *Statistica Applicata, Italian Journal of Applied Statistics*.

Svalastoga, K. (1959) *Prestige, class and social mobility*. Copenhagen: Gyldendal.

van der Burg, E., and de Leeuw, J. (1988). Use of the multinomial jackknife and bootstrap in generalized non–linear canonical correlation analysis. *Applied stochastic models and data analysis, 4*, 159–172.

Estimation of the parameters of the bilinear association model using the Gibbs sampler

By LUIS MILAN and JOE WHITTAKER

University of Lancaster, England

Summary: The purpose of this paper is to show the results of the application of Gibbs sampling technique to the RC association model in order to obtain estimates of the probability density functions for each parameter. A brief description of the *Gibbs sampling* technique is presented. The application of *Gibbs sampling* to produce estimates of the association model parameters is discussed. The conditional probability density function used in the Gibbs iterations is presented. We also show how conditional independence can be used to reduce the computing time. The technique is then applied to an example and the resulting density estimates of the parameters of the association model are presented in the form of graphics.

Keywords: Monte Carlo techniques, Gibbs sampling, contingency table, association model.

1 Introduction

The bilinear association model proposed by Goodman (1985, 1986) is

$$Y_{ij} = \exp(\lambda + \lambda_i^R + \lambda_j^C + \sum_{m=1}^{M} \phi_m \mu_{im} \nu_{jm}) \tag{1}$$

for $i = 1, 2, \ldots, I$ and $j = 1, 2, \ldots, J$ where $\{Y_{ij}\}$ is an I by J contingency table subject to the following set of constraints to identify the parameters:

$$\sum_{i=1}^{I} \lambda_i^R = 0 \; ; \; \sum_{j=1}^{J} \lambda_j^C = 0$$
$$\sum_{i=1}^{I} \mu_{im} = \sum_{j=1}^{J} \nu_{jm} = 0 \; ; \; \sum_{i=1}^{I} \mu_{im}^2 = \sum_{j=1}^{J} \nu_{jm}^2 = 1$$
$$\sum_{i=1}^{I} \mu_{im}\mu_{im'} = 0 \; ; \; \sum_{j=1}^{J} \nu_{jm}\nu_{jm'} = 0 \tag{2}$$

for $m, m' = 1, \cdots, M$ and $m \neq m'$. In the notation used here the parameters without an index means the set of parameters partitioned by that index. So $\lambda^R = \{\lambda_i^R;$ for $i = 1, \ldots, I\}$ and $\nu_j = \{\nu_{jm};$ for $m = 1, \ldots, M\}$. The vector with all parameters is $\Theta = \{\lambda, \lambda^R, \lambda^C, \phi, \mu, \nu\}$.

2 Gibbs sampler

The *Gibbs sampling algorithm* was introduced by Geman and Geman (1984) in the context of image analysis. It is a part of a wider class of algorithms for stochastic simulation based in Markovian random fields. These methods have been widely used in statistical physics after been introduced by Metropolis *et. al.* (1953) through the *Metropolis algorithm*. A generalization of the *Metropolis*