J. R. Statist. Soc. A (2008)
171, *Part* 1, *pp.* 137–157

The analysis of change, Newton's law of gravity and association models

Mark de Rooij

Leiden University, The Netherlands

[Received September 2004. Final revision April 2007]

Summary. Newton's law of gravity states that the force between two objects in the universe is equal to the product of the masses of the two objects divided by the square of the distance between the two objects. In the first part of the paper it is shown that a model with a 'law-of-gravity' interpretation applies well to the analysis of longitudinal categorical data where the number of people changing their behaviour or choice from one category to another is a measure of force and the goal is to obtain estimates of mass for the two categories and an estimate of the distance between them. To provide a better description of the data dynamic masses and dynamic positions are introduced. It is shown that this generalized law of gravity is equivalent to Goodman's RC(M) association model. In the second part of the paper the model is generalized to two kinds of three-way data. The first case is when there are multiple two-way tables and in the second case we have change over three points of time. Conditional and partial association models are related to three-way distance models, like the INDSCAL model, and triadic distance models respectively.

Keywords: Categorical data; Euclidean distance; Gravity model; Longitudinal data; Square tables; Triadic distance

1. Introduction

This paper will be concerned with longitudinal categorical data, i.e. repeated measurements on a number of observational units with the same instrument. The main interest in studying longitudinal data is whether change occurred and, if so, what the nature of the change is. We shall confine ourselves to the case of categorical data. Our questions concern qualitative change, i.e. changes in attitude, opinion, behaviour or any other categorical variable. This is typically different from continuous data where it might be possible to describe change in terms of better or worse; for categorical data descriptions are in terms of 'different' or 'the same'.

Once longitudinal categorical data have been collected they can be represented in transition frequency tables, which are contingency tables where each way corresponds to the categories of a variable measured at a specific time point (we adopt the way mode terminology for the tables of Carroll and Arabie (1980)). The number of time points defines the number of ways of the transition frequency table. Having measured a group of people twice on a categorical variable, a square transition frequency table arises. If measurements are obtained at three time points the data can be gathered in a three-way contingency table, and so forth.

An example of such data is obtained from Upton and Särlvik (1981) who discussed changes in political voting in Sweden. The data are shown in Table 1. There are five political parties: the *Communists* COM; the *Social Democrats* SD; the *Centre Party* C; the *People's Party* P; the *Conservatives* CON. These are the anglicized names following Upton and Särlvik (1981). The rows correspond to the political parties in 1964 (in capital letters); the columns to the political parties in 1970 (lower-case letters).

Address for correspondence: Mark de Rooij, Methodology and Statistics Unit, Institute for Psychological Research, Leiden University, PO Box 9555, 2300 RB Leiden, The Netherlands.
E-mail: rooijm@fsw.leidenuniv.nl

0964–1998/08/171137

Table 1. Swedish voting data representing voting changes from 1964 (rows) to 1970 (columns)†

	com	sd	c	p	con
COM	(22)	27	4	1	0
SD	16	(861)	57	30	8
C	4	26	(248)	14	7
P	8	20	61	(201)	11
CON	0	4	31	32	(140)

†From Upton and Särlvik (1981).

The focus will be on change, i.e. on the off-diagonal entries. The values on the diagonal are within parentheses; for these cells special parameters (which are often called loyalty parameters) will be included in the models to be developed.

The question, once we have such change data, is not *whether* there is association but *what the pattern* of association looks like. We shall propose a model for these data based on Newton's law of gravity, which states that the force between any two objects in the universe is proportional to the masses of the two objects and inversely related to the squared distance between the two objects (Newton's law of gravity will be discussed in more detail in the next section). This deterministic model will be applied to the analysis of change where the objects in the universe are the categories of the variable. The force between two objects is measured by the number of respondents making a transition from one category to another. This number is not accurately measured, however, since a sample is obtained from a population. Therefore, the law of gravity is used as a probabilistic model assuming a multinomial sampling scheme (which is the usual sampling scheme for such data). The force will be modelled by the mass of the two categories and a function of the distance between the two objects.

The remainder of this paper is organized as follows. The next section discusses Newton's law of gravity in more detail. Section 3 describes the analysis of change in terms of Newton's law of gravity and introduces dynamic elements in the law to adapt for different data settings. After introducing the dynamic elements it will be shown that the model is a reparameterization of the RC(M) association model (Goodman, 1979, 1991). The usual identification constraints for this model, however, are not suited to the analysis of change. A new way of identifying the solution will be presented and finally the model will be applied to the data in Table 1. In Section 4 the model will be generalized to the case of multiple two-way tables. The gravity models that are developed are related to conditional association models (Clogg, 1982), but again the usual identification constraints are not suited to the analysis of change. Section 5 treats gravitational models for three time points. These models are related to partial association models (Clogg, 1982). Identification and an application to empirical data will be discussed. We shall conclude with discussion and reflection about the results obtained and show some limitations of the work presented.

2. Newton's law of gravity

One of the major laws of the natural sciences is Newton's law of gravity:

> 'All matter attracts all other matter with a force proportional to the product of their masses and inversely proportional to the square of the distance between them'.

J. R. Statist. Soc. A (2008)
171, *Part* 1, *pp.* 137–157

The analysis of change, Newton's law of gravity and association models

Mark de Rooij

Leiden University, The Netherlands

[Received September 2004. Final revision April 2007]

Summary. Newton's law of gravity states that the force between two objects in the universe is equal to the product of the masses of the two objects divided by the square of the distance between the two objects. In the first part of the paper it is shown that a model with a 'law-of-gravity' interpretation applies well to the analysis of longitudinal categorical data where the number of people changing their behaviour or choice from one category to another is a measure of force and the goal is to obtain estimates of mass for the two categories and an estimate of the distance between them. To provide a better description of the data dynamic masses and dynamic positions are introduced. It is shown that this generalized law of gravity is equivalent to Goodman's RC(M) association model. In the second part of the paper the model is generalized to two kinds of three-way data. The first case is when there are multiple two-way tables and in the second case we have change over three points of time. Conditional and partial association models are related to three-way distance models, like the INDSCAL model, and triadic distance models respectively.

Keywords: Categorical data; Euclidean distance; Gravity model; Longitudinal data; Square tables; Triadic distance

1. Introduction

This paper will be concerned with longitudinal categorical data, i.e. repeated measurements on a number of observational units with the same instrument. The main interest in studying longitudinal data is whether change occurred and, if so, what the nature of the change is. We shall confine ourselves to the case of categorical data. Our questions concern qualitative change, i.e. changes in attitude, opinion, behaviour or any other categorical variable. This is typically different from continuous data where it might be possible to describe change in terms of better or worse; for categorical data descriptions are in terms of 'different' or 'the same'.

Once longitudinal categorical data have been collected they can be represented in transition frequency tables, which are contingency tables where each way corresponds to the categories of a variable measured at a specific time point (we adopt the way mode terminology for the tables of Carroll and Arabie (1980)). The number of time points defines the number of ways of the transition frequency table. Having measured a group of people twice on a categorical variable, a square transition frequency table arises. If measurements are obtained at three time points the data can be gathered in a three-way contingency table, and so forth.

An example of such data is obtained from Upton and Särlvik (1981) who discussed changes in political voting in Sweden. The data are shown in Table 1. There are five political parties: the *Communists* COM; the *Social Democrats* SD; the *Centre Party* C; the *People's Party* P; the

Address for correspondence: Mark de Rooij, Methodology and Statistics Unit, Institute for Psychological Research, Leiden University, PO Box 9555, 2300 RB Leiden, The Netherlands.
E-mail: rooijm@fsw.leidenuniv.nl

 0964–1998/08/171137

J. R. Statist. Soc. A (2008)
171, *Part* 1, *pp.* 137–157

The analysis of change, Newton's law of gravity and association models

Mark de Rooij

Leiden University, The Netherlands

[Received September 2004. Final revision April 2007]

Summary. Newton's law of gravity states that the force between two objects in the universe is equal to the product of the masses of the two objects divided by the square of the distance between the two objects. In the first part of the paper it is shown that a model with a 'law-of-gravity' interpretation applies well to the analysis of longitudinal categorical data where the number of people changing their behaviour or choice from one category to another is a measure of force and the goal is to obtain estimates of mass for the two categories and an estimate of the distance between them. To provide a better description of the data dynamic masses and dynamic positions are introduced. It is shown that this generalized law of gravity is equivalent to Goodman's RC(M) association model. In the second part of the paper the model is generalized to two kinds of three-way data. The first case is when there are multiple two-way tables and in the second case we have change over three points of time. Conditional and partial association models are related to three-way distance models, like the INDSCAL model, and triadic distance models respectively.

Keywords: Categorical data; Euclidean distance; Gravity model; Longitudinal data; Square tables; Triadic distance

1. Introduction

This paper will be concerned with longitudinal categorical data, i.e. repeated measurements on a number of observational units with the same instrument. The main interest in studying longitudinal data is whether change occurred and, if so, what the nature of the change is. We shall confine ourselves to the case of categorical data. Our questions concern qualitative change, i.e. changes in attitude, opinion, behaviour or any other categorical variable. This is typically different from continuous data where it might be possible to describe change in terms of better or worse; for categorical data descriptions are in terms of 'different' or 'the same'.

Once longitudinal categorical data have been collected they can be represented in transition frequency tables, which are contingency tables where each way corresponds to the categories of a variable measured at a specific time point (we adopt the way mode terminology for the tables of Carroll and Arabie (1980)). The number of time points defines the number of ways of the transition frequency table. Having measured a group of people twice on a categorical variable, a square transition frequency table arises. If measurements are obtained at three time points the data can be gathered in a three-way contingency table, and so forth.

An example of such data is obtained from Upton and Särlvik (1981) who discussed changes in political voting in Sweden. The data are shown in Table 1. There are five political parties: the *Communists* COM; the *Social Democrats* SD; the *Centre Party* C; the *People's Party* P; the

Address for correspondence: Mark de Rooij, Methodology and Statistics Unit, Institute for Psychological Research, Leiden University, PO Box 9555, 2300 RB Leiden, The Netherlands.
E-mail: rooijm@fsw.leidenuniv.nl

© 2008 Royal Statistical Society 0964–1998/08/171137

Table 1. Swedish voting data representing voting changes from 1964 (rows) to 1970 (columns)†

	com	sd	c	p	con
COM	(22)	27	4	1	0
SD	16	(861)	57	30	8
C	4	26	(248)	14	7
P	8	20	61	(201)	11
CON	0	4	31	32	(140)

†From Upton and Särlvik (1981).

Conservatives CON. These are the anglicized names following Upton and Särlvik (1981). The rows correspond to the political parties in 1964 (in capital letters); the columns to the political parties in 1970 (lower-case letters).

The focus will be on change, i.e. on the off-diagonal entries. The values on the diagonal are within parentheses; for these cells special parameters (which are often called loyalty parameters) will be included in the models to be developed.

The question, once we have such change data, is not *whether* there is association but *what the pattern* of association looks like. We shall propose a model for these data based on Newton's law of gravity, which states that the force between any two objects in the universe is proportional to the masses of the two objects and inversely related to the squared distance between the two objects (Newton's law of gravity will be discussed in more detail in the next section). This deterministic model will be applied to the analysis of change where the objects in the universe are the categories of the variable. The force between two objects is measured by the number of respondents making a transition from one category to another. This number is not accurately measured, however, since a sample is obtained from a population. Therefore, the law of gravity is used as a probabilistic model assuming a multinomial sampling scheme (which is the usual sampling scheme for such data). The force will be modelled by the mass of the two categories and a function of the distance between the two objects.

The remainder of this paper is organized as follows. The next section discusses Newton's law of gravity in more detail. Section 3 describes the analysis of change in terms of Newton's law of gravity and introduces dynamic elements in the law to adapt for different data settings. After introducing the dynamic elements it will be shown that the model is a reparameterization of the RC(M) association model (Goodman, 1979, 1991). The usual identification constraints for this model, however, are not suited to the analysis of change. A new way of identifying the solution will be presented and finally the model will be applied to the data in Table 1. In Section 4 the model will be generalized to the case of multiple two-way tables. The gravity models that are developed are related to conditional association models (Clogg, 1982), but again the usual identification constraints are not suited to the analysis of change. Section 5 treats gravitational models for three time points. These models are related to partial association models (Clogg, 1982). Identification and an application to empirical data will be discussed. We shall conclude with discussion and reflection about the results obtained and show some limitations of the work presented.

2. Newton's law of gravity

One of the major laws of the natural sciences is Newton's law of gravity:

> 'All matter attracts all other matter with a force proportional to the product of their masses and inversely proportional to the square of the distance between them'.