In: MGH Jansen & WH van Schuur (Eds.)
The many faces of multivariate analysis.
Groningen: RION
ISBN 90-6690-257-4

The power function in log-linear models with and without ordered categories

Ab Mooijaart & Peter G. M. van der Heijden *

Summary: In this paper the asymptotic (Pitman) power of tests for log-linear models is investigated. We summarize the statistical literature on power in frequencies in general, and of power in loglinear model tests in particular. Three applications of power in this context are discussed. The first discusses the situation that one is primarily interested in the power instead of the type-I error. The second is that the power of a test increases if the alternative hypotheses become more restricted. The third is to determine the sample size when a specific power is required against some misspecification. We concentrate on alternative hypotheses which assume the 'uniform association' model.

1. Introduction

In loglinear analysis much attention is devoted to model testing. Compared with the attention for the type-I error, or α, much less attention is spent on the power of these tests. The power of a test is defined as the probability that the nul-hypothesis is rejected when it is not true. One reason for giving attention to power in loglinear analysis is that testing in loglinear analysis is quite different from, for example, the test for a difference between two means, as was indicated by Bonett & Bentler (1983), among others. In many instances research is done to prove that some nul-hypothesis should be rejected, for example the H_0 that some mean for males is equal to the mean for females. In the latter case we generally choose α=.05 or even α=.01 in order to be sure to have a small probability to reject H_0 when H_0 is true. In this way we protect ourselves against our inclination to accept the alternative hypothesis. When we compare this with the situation in loglinear analysis, H_0 also specifies a more restrictive model than H_1, but here the objective is most often to find a parsimonious model, i.e. to accept H_0. So the roles of H_0 and H_1 are reversed here. So trying to reach a high power in loglinear analysis corresponds with trying to reach a low α in the test stating that two means are equal.

Much attention is given to power in the context of covariance structure models recently (see, for example, Saris & Stronkhorst, 1984; Satorra & Saris, 1985; Saris, Satorra & Sörbom, 1987). In this paper we study the power in loglinear models. Although applications are scarse, many results concerning the power in loglinear models do already exist for a long time. The power in loglinear models has not been given much attention in the applied literature, though. An exception is Agresti (1984), who motivates the use of loglinear models with ordered categories by claiming that these models have a higher power. In this paper his claim is elaborated. One of our conclusions is that, when a high power is considered to be important, it is advisable to use as much knowledge we have about the data as possible. It is better to test a hypothesis against a restricted alternative (when this alternative cannot be rejected) than to test it against the unrestricted alternative (which is always true). For example, a higher power is obtained when we test for independence conditional on a so-called uniform association model than when we test for independence conditional on the unrestricted alternative.

*University of Leiden, Department of Psychology, Hooigracht 15, 2312KM Leiden, The Netherlands. Tel. (71)273761. Electronic Mail: MENTCI@HLERUL2

We will start now with a two theoretical sections in which we summarize the mathematical research on power in the frequency domain. We will give special attention to the case of ordinary loglinear analysis and loglinear analysis with ordered categories in section 3. In section 4 we will give appplications for the earlier results. Readers who do not want to be bothered by many formulas might as well skip sections 2 and 3 and go directly to section 4.

2. Distribution of Pearson's chi-square statistic

Although more than one formulation of power is possible, most attention has been given to asymptotic Pitman power. Consider a nul hypothesis H_0. Given H_0 many alternative hypotheses can be specified, depending on the type and the largeness of the misspecification under H_0. The concept of Pitman power implies that the set of all possible alternative hypotheses tends to the null hypothesis for increasing sample size. This type of power has shown to be very useful in deriving the power function of a test. We concentrate here on tests for frequency tables.

Under the null hypothesis Pearson's goodness of fit chi-square test is very popular. Mitra (1958) derived the distribution of this test statistic under the non-null distribution. This distribution is a noncentral chi-square distribution, which is specified by the degrees of freedom and the noncentrality parameter. With this noncentrality parameter it is possible to derive the power function of several tests. The derivation of Mitra does not imply any restriction on the set of alternatives. Diamond (1963), among others, generalizes Mitra's theorem to the case where the set of alternatives is restricted (see also Fix, Hodges & Lehmann, 1959; Aitchison, 1962; Meng & Chapman, 1966; Chapman & Nam, 1968). We will start with a discussion of the asymptotic Pitman power. The main theorem we discuss here was given by Mitra (1958). We give this theorem without any proof.

Let there be q *sequences*, each sequence i (i=1,...,q) has a *multinomial* distribution with r_i possible *outcomes*, where $r = \Sigma_i r_i$. Let π_{ij} denote the *probability of outcome* j in sequence i, where $\Sigma_j \pi_{ij} = 1$. Analogously, n_{ij} denotes the *frequency* that is observed for outcome j in sequence i. Let $n_i = \Sigma_j n_{ij}$ denote the number of *trials* in sequence i, where $n = \Sigma_i n_i$. Define $\theta = (\theta_1,...,\theta_s)$ as an s-dimensional vector of unknown parameters, with $s < r - q$. The null-hypothesis is defined as:

$$H_0 : \pi_{ij} = p_{ij}(\theta) \quad , j = 1,...,r_i, \; i=1,...,q. \tag{1}$$

where the $p_{ij}(\theta)$ specify the π_{ij} as functions of θ that can be different for each outcome j in sequence i.

For H_0 some regularity conditions are defined (see Mitra). One of these conditions is

$$\mathbf{P}(\theta) = \{\delta p_{ij}/\delta\theta'\} \tag{2}$$

is of full column rank, where $\mathbf{P}(\theta)$ is an (r x s) matrix. This ensures that θ is identified. Another condition is that results only apply for *local* alternatives, i.e. the misspecification under H_0 may not be too large. Let $\hat{\theta}$ be the ML estimator, or some other asymptotically equivalent estimator, then the *test statistic* for H_0 is

$$X^2 = \Sigma_i \Sigma_j (n_{ij} - n_i p_{ij}(\hat{\theta}))^2 / n_i p_{ij}(\hat{\theta}). \tag{3}$$

X^2 defines Pearson's chi-square statistic. It is well known that, when H_0 is true, X^2 is asymptotically (central) chi-squared distributed with r-q-s degrees of freedom.

However, we are interested in the behavior of X^2 when H_0 is *not* true: the test statistic should indicate that H_0 must be rejected when H_0 is not true. We will therefore define the *set of alternative hypotheses*. Let c_{ij} be a constant for outcome j in sequence i, such that $\Sigma_j c_{ij} = 0$ (the constants add up to zero in each sequence). The set of alternative hypotheses, depending on n, can be written as:

$$H_{1,n} : \pi_{ij}(n) = p_{ij}(\theta) + c_{ij} / \sqrt{n} \quad , j = 1,...,r_i, \; i=1,...,q. \tag{4}$$

This is the so-called Pitman power definition of alternative hypotheses. Notice that (4) defines a set of alternative hypotheses which tends to H_0 for n to infinity (compare (1).

Define $\mathbf{D}_{p_i(\theta)}$ as a (r_i x r_i) diagonal matrix with elements $p_{ij}(\theta)$, and collect these matrices into a (r x r) block-diagonal super matrix $\mathbf{D}_{p(\theta)}$. Let $\mathbf{W}_i$ be a (r_i x r_i) diagonal matrix with elements n_i / n, and collect these diagonal matrices $\mathbf{W}_i$ into a (r x r) block-diagonal super matrix **W**. Let $\mathbf{c}_i$ be a (r_i x 1) vector with elements c_{ij}, and collect these $\mathbf{c}_i$ in a (r x 1) super-vector **c**. We can now define

$$\delta = \mathbf{W}\mathbf{D}_{p(\theta)}^{-1/2}\mathbf{c} \tag{5a}$$

$$\mathbf{B} = \mathbf{W}\mathbf{D}_{p(\theta)}^{-1/2}\mathbf{P}(\theta), \tag{5b}$$

and present the following theorem.

Theorem : for n going to infinity, when $H_{1,n}$ is true, X^2 as defined in (3) follows a *non-central* chi-squared distribution with df = r - q - s, and *noncentrality parameter* λ, where

$$\lambda = \delta'(\mathbf{I} - \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}')\delta. \tag{6}$$

This theorem shows us the behavior of X^2 when H_0 is *not* true, and that was what we were looking for. It shows that when H_0 is not true, $c_{ij}=\sqrt{n}(\pi_{ij} - p_{ij}(\theta))$ will increase for a fixed departure from H_0 when n becomes larger (compare (4)), hence δ will increase when n becomes larger, and hence λ will increase. Therefore when H_0 is not true, the distribution of X^2 is translated to the right, which leads to the situation that the probability of finding a value of X^2 that leads to the rejection of H_0 increases. Similar reasoning shows that, for fixed sample size but growing departure from H_0, the probability of finding a value of X^2 that leads to the rejection of H_0 also increases. Concluding, when H_1 is true, the probability of correctly rejecting H_0 (i.e. the power) increases both as a function of growing sample size and as a function of the departure from H_0.

However, this theorem specifies the non-centrality parameter in a *general* situation, far more general than necessary in many practical situations. In fact, the only restriction for the alternative hypotheses is that $\Sigma_j c_{ij} = 0$ (for i = 1,...,q), i.e. constants are added to each element, and these constants must add up to zero for each sequence i. In practical situations the researcher may have some insight in what the alternative hypotheses are. In these situations we can perform a test against *a set of more restricted alternative hypotheses*, and we will show that in these situations the non-centrality parameter defined in (6) depends on less parameters. We also show that this has an effect on the power of the test. See for a detailed treatment of this Diamond (1963). We give here a simple derivation for a test with restricted alternatives, see also Haber (1981, 1984). Define the restricted alternative hypotheses as

$$H_2 : \pi_{ij} = p_{ij}(\theta) + g_{ij}(\theta,\xi) \quad , j = 1,...,r_i, \ i=1,...,q, \tag{7}$$

with ξ a vector of order (t x 1), with $t < r - s - q$. So, compared to H_0, the alternative hypothesis H_2 specifies the situation that t *extra* parameters are needed to model the proportions, and these extra parameters are collected in ξ. Let $\Sigma_j\, g_{ij}(\theta,\xi) = 0$ (this is comparable with $\Sigma_j c_{ij}=0$ in (4)). H_2 can be more restrictive than $H_{1,n}$, where a constant is added to each function, and assume that there is a ξ^0 for which $g_{ij}(\theta,\xi^0) = 0$, for all i and j. Given that the alternative model specified in (7) is true, we can write the null hypothesis H_0 as

$$H_0 : \xi = \xi^0. \tag{8}$$

This H_0 specifies that H_2 can be restricted further. Notice that this H_0 is different from H_0 in (1), where H_0 was tested against the *unrestricted* alternative specified in (4). When we rewrite the restricted alternative hypothesis (7) in terms of the Pitman power, (7) becomes

$$H_{2,n} : \pi_{ij}(n) = p_{ij}(\theta) + g_{ij}(\theta,\xi(n)) \quad , j = 1,...,r_i, \ i=1,...,q, \tag{9}$$

where

$$\xi(n) = \xi^0 + \gamma/\sqrt{n} \tag{10}$$

and γ is a (t x 1) vector. Comparable to (4), the set of (restricted) alternative hypotheses in (9) tends to H_0 if n goes to infinity. If we define

$$G = \{\delta g_{ij}/\delta\xi\}, \tag{11}$$

then, by using the first two terms of the Taylor expansion of π_{ij}, (9) can be rewritten as

$$\pi(n) = p(\theta) + G(\xi(n) - \xi^0) + O(\|\xi(n) - \xi^0\|^2)$$

$$\equiv p(\theta) + G\gamma/\sqrt{n}. \tag{12}$$

where $\pi(n)$ is a column vector with values $\pi_{ij}(n)$, and $p(\theta)$ is a column vector with values $p_{ij}(\theta)$. Following results of Aitchison (1962), we have the following theorem.

Theorem: for n going to infinity, when the restricted alternative hypothesis $H_{2,n}$ defined in (9) is true, and we test H_0 as defined in (8), X^2 is asymptotically noncentral chi-square distributed with t degrees of freedom and noncentrality parameter

$$\lambda = \gamma' \mathbf{A}'(\mathbf{I} - \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}')\mathbf{A}\gamma, \tag{13}$$

where **A** is a (r x t) matrix defined as $\mathbf{W}\mathbf{D}_{p(\theta)}^{-1/2}\mathbf{G}$, where **G** is evaluated at $\xi = \xi^0$. By writing $\mathbf{c} = \mathbf{G}\gamma$ (compare (4) with (12)) the noncentrality parameter λ in (13) can be written in the same form as the noncentrality parameter in the unrestricted case (6).

We now find the following important result:

Corrollary: if **c** in the unrestricted case is identical to $\mathbf{G}\gamma$ in the restricted case, so that the λ in (6) is equal to the λ in (13), the test of H_0 against the restricted alternative (13) will have a higher power than the test against the unrestricted alternative (4).

This follows from the fact that for both cases the value found for λ will be identical in both cases, but the distribution of the Pearson chi-square test for the restricted alternative has a smaller number of degrees of freedom (since we assumed that $t < r - s - q$) than the test for the unrestricted alternative. For this context Das Gupta and Perlman (1974) proved that, for fixed noncentrality parameter, the probability of rejecting H_0 at a fixed α-level increases if the number of degrees of freedom decreases. So the power is larger in the test against the restricted alternative. For practical situations this implies that, if possible (i.e., if a restrictive alternative can be found that fits adequately), tests against restrictive alternatives should be preferred above tests against unrestrictive alternatives when a higher power is preferred above a lower power.

In the next section we will discuss power for loglinear model tests, and derive simple formulae for the non-centrality parameter.

3. The power in log-linear models

In the models we discuss here it is assumed that there is just one sequence of observations, i.e. q = 1. In general, log-linear analysis can be written as

$$\mathbf{p}(\theta) = \{\exp[\mathbf{X}\theta]\}/\mathbf{1}'\exp[\mathbf{X}\theta], \tag{14}$$

where **X** is some design matrix. (In this notation, **X** does not contain a column with ones for the average, instead it is divided by the sum $\mathbf{1}'\exp[\mathbf{X}\theta]$ so that the probabilities add up to one). θ is some vector consisting of model parameters, which may be interpreted as main effects and interaction effects, similar to the situation in analysis of variance. The term $\exp[\mathbf{X}\theta]$ is a (r x 1) vector with its i-th element equal to $\exp[\mathbf{x}_i'\theta]$, where $\mathbf{x}_i$ is row i of matrix **X**. "**1**" is a (r x 1) column vector consisting of unit elements. The following equations are easy to verify. We can rewrite (2), (5a) and (5b) as

$$\mathbf{P}(\theta) = (\mathbf{D}_{\mathbf{p}(\theta)} - \mathbf{p}(\theta)\mathbf{p}(\theta)')\mathbf{X}, \tag{15}$$

$$\delta = \mathbf{D}_{\mathbf{p}(\theta)}^{-1/2}\mathbf{c}, \tag{16}$$

$$\mathbf{B} = \mathbf{D}_{\mathbf{p}(\theta)}^{-1/2}(\mathbf{D}_{\mathbf{p}(\theta)} - \mathbf{p}(\theta)\mathbf{p}(\theta)')\mathbf{X}, \tag{17}$$

where $\mathbf{D}_{\mathbf{p}(\theta)}$ is a diagonal matrix having as elements the elements of vector $\mathbf{p}(\theta)$. Notice that the asymptotic covariance matrix of $\hat{\theta}$ is

$$\mathrm{cov}(\hat{\theta}) = \mathbf{P}(\theta)'\mathbf{D}_{\mathbf{p}(\theta)}^{-1}\mathbf{P}(\theta)/n = (\mathbf{B}'\mathbf{B})^{-1}/n = (\mathbf{X}'(\mathbf{D}_{\mathbf{p}(\theta)} - \mathbf{p}(\theta)\mathbf{p}(\theta)')\mathbf{X})^{-1}/n \tag{18}$$

where **B** is evaluated at $\hat{\theta}$. Utilizing the restrictions $\mathbf{1}'\mathbf{c} = 0$, and therefore $\mathbf{1}'\mathbf{p}(\theta) = 1$, we find for the noncentrality parameter

$$\lambda = \mathbf{c}'\mathbf{D}_{\mathbf{p}(\theta)}^{-1}\mathbf{c} - \mathbf{c}'\mathbf{X}[\mathbf{X}'(\mathbf{D}_{\mathbf{p}(\theta)} - \mathbf{p}(\theta)\mathbf{p}(\theta)')\mathbf{X}]^{-1}\mathbf{X}'\mathbf{c}. \tag{19}$$

The first term on the right hand side of this equation can be written as

$$\mathbf{c}'\mathbf{D}_{\mathbf{p}(\theta)}^{-1}\mathbf{c} = n(\pi(n) - \mathbf{p}(\theta))'\mathbf{D}_{\mathbf{p}(\theta)}^{-1}(\pi(n) - \mathbf{p}(\theta)), \tag{20}$$

which is the formula often given for the noncentrality parameter for the distribution of the X^2-statistic under the alternative hypothesis (see, e.g., Bishop, Fienberg and Holland, 1975). However, this is only true when the second term of the right hand side of (18) is zero. This second term will be zero when $\mathbf{c}'\mathbf{X} = \mathbf{0}'$. So the question is when is $\mathbf{c}'\mathbf{X} = \mathbf{0}'$?

Let's first consider the class of hierarchical loglinear models in which information about ordered categories is not explicitly modeled. For these models $\mathbf{p}'\mathbf{X} = \pi_{(0)}'\mathbf{X}$, i.e given a hierarchical loglinear model specified by **X**, a vector of estimates of expected proportions $\pi_{(0)}$ is found for which the margins specified by $\pi_{(0)}'\mathbf{X}$ are equal to the corresponding margins in the observed proportions, that are collected in the vector **p**. When we define $\mathbf{c} = \sqrt{n}(\mathbf{p} - \pi_{(0)})$ (compare (4)), we find that $\mathbf{c}'\mathbf{X} = \sqrt{n}(\mathbf{p} - \pi_{(0)})'\mathbf{X} = 0$. We conclude that, when H_0 is some restrictive loglinear model, and the alternative hypothesis is the unrestricted alternative (the observed proportions), $\lambda = \mathbf{c}'\mathbf{D}_{\mathbf{p}}^{-1}\mathbf{c}$.

Now consider the same restrictive loglinear model with matrix **X** that specifies its margins as the model under H_0, but let the alternative hypothesis be that a *restricted* hierarchical loglinear model is true that has the model under H_0 as a special case (i.e. the alternative model has a matrix $\tilde{\mathbf{X}}=(\mathbf{X},\mathbf{X}^{(1)})$ with the first columns identical to those of **X**, plus some extra columns collected in $\mathbf{X}^{(1)}$), and let the estimates of expected probabilities be collected in $\pi_{(1)}$. This implies that for the model under H_1 $\pi_{(0)}'\mathbf{X} = \pi_{(1)}'\mathbf{X}$, i.e. the minimal sufficient margins under H_0 can be derived from the minimal sufficient margins under H_1. Hence we find that $\mathbf{c}'\mathbf{X} = \sqrt{n}(\pi_{(1)} - \pi_{(0)})'\mathbf{X} = 0$, and we conclude that, when the alternative model is a *restricted* hierarchical loglinear model that has the model under H_0 as a special case, then $\lambda = \mathbf{c}'\mathbf{D}_{\pi}^{-1}\mathbf{c}$. Untill sofar, comparable results can be found in Fenech and Westfall (1988).

After these results for the class of hierarchical loglinear models that do not model information about the order of categories, we will now give some results for the class of hierarchical loglinear models that *do* model information about the order of categories, and we restrict attention to so-called *association models*. Let's consider association models for a two way table. The gerenal loglinear model can be written as $\log \pi_{ij} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)}$. In association models the interaction term $u_{12(ij)}$ is modelled as

$$u_{12(ij)} = \beta\,(w_i - \bar{w})(v_j - \bar{v}) \tag{21}$$

We consider two cases: the *uniform association model* where w_i and v_j are fixed equally spaced scores, e.g. 1, 2, 3 for three categories, $(w_i - \bar{w})$ and $(v_j - \bar{v})$ set these fixed scores in average from zero. So in the uniform association model only the parameter β has to be estimated. Apart from the minimal sufficient statistic $\pi_{i+} = p_{i+}$ and $\pi_{+j} = p_{+j}$, we have the minimal sufficient statistics

$$\Sigma_i\Sigma_j\,\pi_{ij}w_iv_j = \Sigma_i\Sigma_j\,p_{ij}w_iv_j. \tag{22}$$

The uniform association model can be fitted using a matrix $\mathbf{X}^{(2)}$, the first columns of which take care of the fitting of the row and column margins, and the last column takes care of condition (22). We collect the estimates of expected probabilities in the column vector $\pi_{(2)}$.

A less restrictive version of (21) is the *row effects association model*, in which not only β has to be estimated, but also u_i. In the row effect association model the minimal sufficient statistic (22) simplifies to

$$\Sigma_j \pi_{ij} v_j = \Sigma_j p_{ij} v_j \qquad (23)$$

and this model can be fitted using a matrix $\mathbf{X}^{(3)}$, in which the first columns fit the marginal row and column proportions to the estimates of expected probabilities, and the last columns take care of condition (23). For more details we refer to Agresti (1984). We collect the estimates of expected probabilities in the column vector $\pi_{(3)}$.

Let H_0 be either the uniform association model, or the row effect association model, and let the alternative hypothesis be the unrestricted alternative. Under the uniform association model $\pi'_{(2)}\mathbf{X}^{(2)} = \mathbf{p}'\mathbf{X}^{(2)}$, due to condition (22), and similarly, under the row effect association model $\pi'_{(3)}\mathbf{X}^{(3)} = \mathbf{p}'\mathbf{X}^{(3)}$, and therefore $\mathbf{c}'\mathbf{X}^{(2)} = 0$ and $\mathbf{c}'\mathbf{X}^{(3)} = 0$, and we conclude that when H_0 is a uniform association model or a row effect assocition model, and the alternative hypothesis is the row effect association model, then $\lambda = \mathbf{c}'\mathbf{D}_\pi^{-1}\mathbf{c}$.

Another case is that H_0 is the independence model, specified by $\mathbf{X}^{(0)}$ with estimates of expected probabilities $\pi_{(0)}$, and that H_1 is either the uniform association model or the row effects association model. In these cases $\pi'_{(0)}\mathbf{X}^{(0)} = \pi'_{(2)}\mathbf{X}^{(0)} = \pi'_{(3)}\mathbf{X}^{(0)}$, and therefore we conclude that in these cases $\lambda = \mathbf{c}'\mathbf{D}_\pi^{-1}\mathbf{c}$. As a last case, let H_0 be the uniform association model and H_1 be the row effect association model. In that case $\pi'_{(2)}\mathbf{X}^{(2)} = \pi'_{(3)}\mathbf{X}^{(2)}$ since (23) is a special case of (22), and therefore $\lambda = \mathbf{c}'\mathbf{D}_\pi^{-1}\mathbf{c}$ in this case also.

The results just discussed for loglinear models that use information of ordered categories is more general than the cases discussed here. For example, $\mathbf{c}'\mathbf{X} = \mathbf{0}$ also in case of association models with fixed scores that are not equally spaced, or column effect assocition models, and association parameters in higher-way tables. In cases where both u_i and v_j are estimated, the model is not loglinear anymore, but logmultiplicative. In this case, when H_1 is the model in which β, u_i and v_j parameters have to estimated, the test for H_0: $\beta=0$ does not follow asymptotically a chi-squared

distribution but instead the distribution of the first eigenvalue of a Wishart matrix. For details, and results on the power of this test, we refer to Haberman (1981) and Schriever (1983).

We come to the following general conclusion: when the minimal sufficient statistics for the model under H_0 can be derived from the minimal sufficient statistics under H_1, then $\mathbf{c}'\mathbf{X} = \mathbf{0}$, and therefore $\lambda = \mathbf{c}'\mathbf{D}_\pi^{-1}\mathbf{c}$ in these cases. Notice that λ is in fact the same as the value of the Pearson chi-squared statistic X^2. This shows that X^2 has two applications: when H_0 is true, X^2 gives us the test statistic value that can be compared with the central chi-squared distribution; when H_1 is true, X^2 gives us the non-centrality parameter λ for the non-central chi-squared distribution. This λ can be used to find the power from specific power tables, the entries of which are the number of degrees of freedom, α, the non-centrality, and the power. See, for example, Haynam, Govindarajulu and Leone (1973).

4. Applications

We will discuss three types of applications of power in the loglinear analysis context. They are, firstly, the concern for type-I errors as compared to the power, secondly, using power as an argument to test for restricted alternatives instead of unrestricted alternatives, and thirdly, using power in a design of a study, i.e. to set up a sample size. Before we do this, we mention here without proof that the results given before for the Pearson chi-square test also apply for the likelihood ratio chi-square test, since both tests have the same limiting distribution under Pitman type set up.

4.1. Power versus type-I error

In loglinear analysis tests are often performed in the hope that H_0 is *not* rejected, instead of the reverse, namely that the hope is that H_0 is rejected. In this latter case it is usually best to choose a small α, so that you are protected against the type I error to reject H_0 while it is true. However, when the hope is that H_0 is true, it is more important to have a small type II error β, i.e. a large power $(1-\beta)$. A large power is necessary when you want to conclude from non-rejection of H_0 that H_0 is true. This is one instance that a large power is more important than a small type I error (compare Bradley, 1976; Bonett & Bentler, 1983).

Another situation in which power is important is described by Saris & Stronkhorst (1984) in the context of covariance structure analysis, but the same reasoning applies here. Saris & Stronkhorst make a distinction between having a high or low power, and having a large or small misspecification. When the misspecification is large, and the power to detect it is high, so the X^2 will be too high and H_0 will be rejected. When the misspecification is small, and the power to detect the misspecification is low, the X^2 will be low, and H_0 will be accepted. These two situations will be satisfactory: usually you are willing to accept a small misspecification, but you rather include extra parameters in the model when the misspecification is large, and the X^2 guides you into the right direction. The other two possibilities are less satisfactory: consider that the misspecification is small but the power is high, leading to a significant X^2. The test tells you to reject H_0 and to include the extra parameter(s), but the extra parameter(s) are not substantively interesting, so in practice you would rather stick to H_0 for reasons of parsimony. The other unsatisfactory situation is that the misspecification is large, but the power is low, leading to a non-significant chi-square. In this situation you would rather reject H_0, i.e. include the extra parameter(s). These examples show that in general you are not only interested in controlling α, but also in controlling β.

It is not very difficult to calculate the power of a test, when we know the misspecification. One approach is sketched here, a more systematic one will be given in section 4.2. In section 3 we saw that when the alternative model is true, X^2 follows a non-central chi-square distribution with the same number of degrees of freedom as the test under H_0 but with non-centrality parameter $\lambda = c'D_{\pi}^{-1}c$. In section 3 we saw that this λ is just the value of X^2 when we would calculate X^2 from the estimates of expected frequencies under H_0 and H_1, and it is easy to find the power that is associated with a specific α from special power tables (given by, for example, Saris & Stronkhorst, 1984). To give an example: let H_0 be independence for some two-way cross-table of order 3x4, and let the alternative model be the unrestricted alternative. We find X^2 is 7.6, which plays a double role: firstly, when H_0 is true, X^2 is asymptotically distributed as a central chisquared distribution with 6 degrees of freedom; secondly, when H_1 is true, X^2 is asymptotically distributed as a *non*-central chi-squared with 6 degrees of freedom and non-centrality parameter λ. If we are willing to assume that the observed data represent precisely the true situation, so that $X^2 = \lambda = 7.6$, we can look up the power in a power table for α=.05, df=6 and λ=7.6, and we find a power of .50.

4.2 Larger power of tests against restricted alternatives

At the end of section 2 we found the result that a test of H_0 against a restricted alternative that is true has a larger power than a test of H_0 against the unrestricted alternative. The idea is that X^2 for H_0 against the restricted alternative is approximately

the same as the X^2 for H_0 against the unrestricted alternative (since the estimates of expected frequencies under the restricted alternative are approximately the same as the

Table 1: Dumping severity data

Operation	*Dumping severity*			
	none	slight	moderate	total
A	61	28	7	96
B	68	23	13	104
C	58	40	12	110
D	53	38	16	107

estimates in the unrestricted alternative), but the test for H_0 against the restricted alternative has less degrees of freedom, and leads therefore more often to the rejection of H_0. This point was raised by Agresti (1984) as a plea for the use of loglinear models with ordered alternatives. He presented the following convincing example (Agresti uses the likelihood ratio statistic G^2 instead of X^2 in his example, but for our purpose this makes no difference since asymptotically they are equivalent). Consider the data in table 1. When we test for independence against the unrestricted alternative, we find G^2=10.88 with df=6, which is not significant at α=.05. This is the first test for independence. However, we find that uniform association model gives a good fit: G^2=4.59 with df=5. If independence is tested given that the uniform association model is true, we find G^2=6.29 with df=1, which is highly significant, showing that the data are not independent. The first test was just not powerful enough to pick this up. We will now study the power of both tests for independence more closely.

Goodman (1981) shows that, if the data are drawn from a discretized bivariate normal distribution, and we analyze the data with the RC-model, then β is related to the correlation ρ as $\beta=\rho/(1-\rho^2)$, when u_i and v_j are normalized as $\Sigma_i p_{i+} w_i = 0 = \Sigma_j p_{+j} v_j$ and $\Sigma_i p_{i+} w_i^2 = 0 = \Sigma_j p_{+j} v_j^2$. As an illustration we will derive the power of both independence tests for α=.05 for distinct values of ρ. We can do this by calculating the non-centrality parameter λ for distinct values of ρ, and we can find λ by calculating X^2 from estimates of expected frequencies under independence against estimates of expected frequencies under a misspecified model. In order to do this first calculate a matrix of values $u_{12(ij)} = \beta(w_i-\bar{w})(v_j-\bar{v})$ with, for example, β=.1 (this corresponds approximately with to ρ = .1). As a second step take the anti-logarithms of this table, and fit the observed margins of table 1 to this table of values $\exp(u_{12(ij)})$, so that this table has the same margins as the table under the alternative hypothesis. An iterative procedure should be used that does not affect the interaction in the table in the sense that the odds ratios remain the same (see Bishop et al., 1975). Then, when we have this table with misspecified values, we calculate X^2 for this table against the estimates of expected frequencies under independence, and the value found for X^2=4.2 is equal to

Table 2: Power for α=.05 for various misspecifications

ρ	λ	df=1	df=6
-.22	20.8	1.00	0.95
-.18	13.8	0.96	0.81
-.14	8.3	0.82	0.55
-.10	4.2	0.54	0.28
-.06	1.5	0.23	0.12
-.02	.2	0.07	0.06
0.00	.0	0.05	0.05
0.02	.2	0.07	0.06
0.06	1.5	0.23	0.12
0.10	4.2	0.54	0.28
0.14	8.3	0.82	0.55
0.18	13.8	0.96	0.81
0.22	20.4	1.00	0.95

the non-centrality parameter λ for the alternative hypothesis that ρ = .1. We now look up the power in a table for the non-central chi-squared distribution with the non-centrality parameter λ=4.2 and α=.5 with df=1 and df=6, and we find a power of .53 for df=1 (the test of independence against the restricted alternative) and a power of .28 for df=6 (the test of independence against the unrestricted alternative). Due to the result of das Gupta and Perlman (1974), the power for df=1 must be higher, and for our example the power is approximately 2 times as high for a misspecification of ρ=.1. In table 2 we show the powers for other values of ρ. For misspecifications larger than .22, the power is 1 for the conditional test. For a misspecification of 0, the power is equal to α=.05.

This illustration clearly suggests the advantage of using stepwise testing procedures in loglinear analysis (see also Aitchison, 1962). A general procedure is this. Let us be interested in the presence or absence of the interaction (or the interaction in a specific form). Find the following two models: let model 1 be the *least* parsimonious model that does *not* include the interaction, let model 2 be equal to model 1 *plus* the interaction (or the interaction in specific form) we are interested in. The idea is that model 1 is the model that gives the best description of the data without including the interaction of interested, i.e. by controlling for the other variables the two variables of interest are independent. By definition, model 1 is a conditional independence model, i.e. variable 1 is independent of variable 2 given the other variables. Model 2 is model 1

but specifies that 1 and 2 are directly related, and not indirectly via the other variables. Now first test model 2. If model 2 fits adequately, the test of model 1 given model 2 tests whether the interaction between 1 and 2 is absent. If it is absent, the relation between 1 and 2 can be explained by the relation of 1 and 2 with the other variables. If it is present, we know that 1 and 2 are directly related. The test of model 1 given model 2 is more powerful test for the presence of the interaction between variables 1 and 2 than the test of model 1 against the unrestricted alternative. As a (verbal) example for four-way data, let us be interested in the interaction between variables 1 and 2. Then model 1 is [134][234] (i.e. 1 and 2 are independent in each level of variables 3 and 4 jointly), model 2 is [134][234][12], and the conditional test for model 1 given model 2 is more powerful than the test of model 1 against the unrestricted alternative. Notice that a prerequisite for the conditional test is that model 2 fits adequately. However, if model 2 does not fit adequately, obviously, interactions have to be included in the model in which variables 1 and 2 are both included, so in this case we know that the interaction between 1 and 2 cannot be neglected.

It is emphasized that for conditional tests it is necessary that the least restrictive model fits adequately (see for example Agresti, 1984, in the context of loglinear models with ordinal variables). This is necessary since the conditional test is only more powerful in detecting departures from H_0 that are specified by the alternative hypothesis. A second reason is that X^2 is only asymptotically chi-squared distributed if the alternative hypothesis is true. This might also explain findings of Oler (1987), who showed, using simulation studies, that the model search procedure forward selection is less powerful than the model search procedure backward elimination. The practical conclusion must therefore be that, if is is necessary to perform exploratory model search procedures (for example since there is not enough theory to be able to say that, for example, the relation between variable 1 and 2 is of most interest), backward selection should be preferred, since with forward inclusion the probability of incorrectly excluding parameters is larger.

4.3 Sample size determination

A third way to use the power in practical situations is for sample size determination. In order to determine the power, you must be able to specify a critical effectsize, the desired power level, and some extra information about the data (compare also Lachin, 1977; Guenther, 1977). We will illustrate this by means of an example.

Consider that we are designing a study, and we will collect data that can be coded into a 5x3 cross table. We have a rough guess about the row proportions and the column proportions for our population from other research (e.g. from opinion polls), the row proportions for our population are 1/15, 2/15, 3/15, 4/15 and 5/15, and the column proportions are 8/15, 5/15 and 2/15. For some reason we also happen to know that, if there is dependence between the two variables, it will take the form of a uniform association model. For a desired α=.05, what sample size must we choose in order to detect ρ=.25 or larger (i.e. to reject H_0 correctly) with a power of .95?

We can derive proportions under independence as the product of the margins, and we can derive proportions under the uniform association model using the same procedure as the one used in section 4.2. Using these proportions we find a chi-squared X_p^2 based on proportions (i.e. a sample size of 1), and so $nX_p^2 = X^2 = \lambda$. We find X_p^2 = .058. We know from power tables that for α=.05 and a power of .95, for df=1, we need a non-centrality parameter λ=13.0. So we need a sample size of 13/.058 = 224 to reach a power of .95. When we would not have used the information that the interaction takes the form of uniform association, we would have the same non-centrality parameter divided by n, but need 8 degrees of freedom to test independence. The power tables for df=8 show that λ = 23.0 for a power of .95. It follows that we would need approximately 23/13 times as much observations to have a power of .95 when the misspecification would be ρ = .25. So using this information about the form of the interaction reduces the costs of the study considerably.

However, in many studies we do not have the information to calculate the power. For example, we would not know that the interaction takes the form of uniform association, or we would not have an estimate for the margins of our table.

5. Conclusions

In this paper we have summarized some of the theoretical results on power in loglinear model tests using both ordered and non-ordered variables. We have also indicated some applications, and shown how power estimates can be derived. We think that in practical situations the advice to use restricted alternatives will be most easy to apply. In order to be able to determine the sample size much should be know from earlier research, and power calculations cannot be routinely performed using standard computer programs.

6. References

Agresti, A. (1984). *Analysis of Ordinal Categorical Data.* Wiley, New York.

Aitchison, J. (1962). Large-sample restricted tests. *Journal of the Royal Statistical Society*, **B24**, 234-250.

Bishop, Y.M.M., Fienberg, S.E. & Holland, P.W. (1975). *Discrete Multivariate Analysis: Theory and Practice.* Cambridge, MA.

Bonett, D.G. & Bentler, P.M. (1983). Goodness-of-fit procedures for the evaluation and selection of loglinear models. *Psychological bulletin*, 93, 149-166.

Chapman, D.G. & Nam, J. (1968). Asymptotic power of chi-square tests for linear trends in proportions. *Biometrics*, **24**, 315-327.

Das Gupta, S., & Perlman, M.D. (1974). Power of the non-central F-test: effect of additional variates on Hotelling's T^2-test. *Journal of American Statistical Association*, **69**, 174-180.

Diamond, E.L. (1963). The limiting power of categorical data chi-square tests analogous to normal analysis of variance. *Annals of Mathematical Statistics*, **35**, 1432-1441.

Fenech, A. P. & Westfall, P.H. (1988). The power function of conditional log-linear model tests. *Journal of the American Statistical Association*, **83**, 198-203.

Fix, E., Hodges, J.L. & Lehmann, E.L. (1959). The restricted chi-square test. In: U. Grenander, ed., *Probability and Statistics - the H. Cramér Volume.* Wiley, New York, pp. 92-107.

Goodman, L.A. (1981). Association models and the bivariate normal for contingency tables with ordered categories. *Biometrika*, **68**, 347-355.

Guenther, W.C. (1977). Power and sample size for approximate chi-square tests. *The American Statistician*, **31**, 83-85.

Haber, M. (1981). On the asymptotic power and relative efficiency of the frequency X^2 test. *Journal of Statistical Planning and Inference*, **5**, 299-308.

Haber, M. (1984). The large-sample power of the X^2 tests for multidimensional contingency tables. *Metrika*, **31**, 195-202.

Haberman, S.J. (1981). Tests for independence in two-way contingency tables based on canonical correlation and on linear-by-linear interaction. *Annals of Statistics*, **9**, 1178-1186.

Haynam, G.E., Govindarajulu, Z. & Leone, F.C. (1973). Tables of the cumulative noncentral chi-square distribution. In: H.L. Harter & D.B. Owen, eds., *Selected Tables in Mathematical Statistics*, Vol. 1, pp. 1-78. Providence: Mathematical Statistical Society.

Lachin, J.M. (1977). Sample size determination for r x c comparitive trials. *Biometrika*, 33, 315-324.

Meng, R.C. & Chapman, D.G. (1966). The power of chi-squares tests for contingency tables. *Journal of the American Statistical Association*, **61**, 965-975.

Mitra, S.K. (1958). On the limiting power function of the frequency chi-square test. *Annals of Mathematical Statistics*, **29**, 1221-1233.

Oler, J. (1985). Noncentrality parameters in chi-squared goodness-of-fit analyses with an application to log-linear procedures. *Journal of the American Statistical Association*, **80**, 181-189.

Saris, W.E., Stronkhorst, L.H. (1984). *Causal modelling in non-experimental research: An introduction to the LISREL approach*. Amsterdam: Sociometric Research Foundation.

Saris, W.E. Satorra, A, Sörbom, D. (1987). The detection and correction of specification errors in structural equation models. *Sociological Methodology 1987* (Ed. C.C. Clogg), Washington.

Satorra, A., Saris, W.E. (1985). Power of the likelihood ratio tests in covariance structure analysis. *Psychometrika*, **50**, 83-90.

Schriever, B.F. (1985). *Order dependence*. Amsterdam: Centre for Mathematics and Informatics.