Bartlett-type correction terms for tests in intensity regression models.

by Lars Klingberg
Bartlett-type correction terms for tests in intensity regression models:

by

Lars Klingberg

Abstract

For multi-factor intensity regression models both computational and theoretical difficulties arise when one tries to find an exact and useful correction term to adjust the likelihood ratio test for deviations from the asymptotic chi-square distribution. In cases with low intensities, short observation intervals, or small samples, the Bartlett adjustment cannot be recommended as a correction even for simple exponential distribution models, for the convergence of the expansion of the expected likelihood ratio is too slow. In this paper some suggestions for approximate correction terms are discussed. These corrections are constructed as sums of correction terms for simple censored exponential distribution models, with a structure similar to that of the test statistic in analysis of variance. A simulation study is described, and results concerning the likelihood ratio statistic for some simple models are presented. The study indicates that the Bartlett corrected test tends to be too conservative, and that the use of parameter estimates instead of the true parameter values in the correction formula makes the test even more conservative. Finally, we discuss the choice of degrees of freedom for likelihood ratio tests for data sets with practically empty cells. For these cases, we suggest a very careful use of a chi-square statistic with a decreased number of degrees of freedom.
Contents.

1. Introduction.
   1.1 The need for a corrected likelihood ratio test. 3
   1.2 Outline of the paper. 6
2. The Bartlett correction factor. 7
3. Expansions of the likelihood ratio statistic. 10
4. Simplified correction terms.
   4.1 Nuisance parameters replaced by their ML-estimates in the model parametrization. 13
   4.2 Nuisance parameters replaced by their ML-estimates in the likelihood equation. 18
5. Correction terms for models with many nuisance parameters. 20
6. Correction terms for simple models.
   6.1 The Bartlett correction. 24
   6.2 Simulation as a way to find correction values. 24
6.3 An alternative correction term. 26
7. Software to compute and test correction terms.
   7.1 An outline of a simulation study 27
   7.2 Bartlett corrections in the presence of competing risks and time varying intensities. 28
   7.3 A description of the software. 29
   7.4 Some simulation results. 31
   7.5 A few remarks. 33
   7.6 Comments on the direct use of Lawley’s formula. 35
8. Degrees of freedom in the likelihood ratio test in cases with few observations. 36
Acknowledgements. 40
References. 40
1. Introduction

1.1 The need of a corrected likelihood ratio test

In this paper we will consider a situation where life history data are analyzed through ML-methods applied to an intensity regression model with piecewise constant baseline intensity functions.

We have defined a set of states such that each event in the life history of an individual corresponds to a transition between two distinct states. Our interest is directed towards the transition from one specific state of origin to a specific state of destination, and towards the impact that a given set of background factors has on the transition intensity. Each background factor is assumed to have a finite number of levels, and we express the transition intensities as products of background parameters; some of these may correspond to an interactive or joint impact from two or more factors. Instead of providing a general expression for the intensity model we use the following example as an illustration.

Example  A model for life history data.

Suppose that we want to analyze marriage intensities among women with no children. The transitions into the three states at the left are then said to compete with the marriage formation. The set of background factors and the parameters corresponding to their levels could be as follows:

...
Social background \(a_1; a_1^i, i=1,2,\ldots,n_a\)

Religiousness \(a_2; a_2^i, i=1,2,\ldots,n_r\) fixed

Year of birth \(a_3; a_3^i, i=1,2,\ldots,n_b\) covariates

Regional background \(a_4; a_4^i, i=1,2,\ldots,n_y\)

Level of education \(a_5; a_5^i, i=1,2,\ldots,n_e\) time varying

Occupational status \(a_6; a_6^i, i=1,2,\ldots,n_j\) covariates

Age (time factor) \(a_7; a_7(t) = a_7^i\) if \(t e(t_{i-1}, t_i], i=1,\ldots,n_a\)

One model for the marriage intensity is:

\[ \lambda(t) = \theta_0 \cdot a_1 \cdot a_2 \cdot a_3 \cdot a_4 \cdot a_5 \cdot a_6 \cdot a_7(t) \cdot (a_1 a_7)(t) \cdot (a_3 a_7)(t) \]

Here \(\theta_0\) is the baseline intensity level, and the two \((a_1 a_j)\)'s are two interaction factors. Completed with the following restrictions the above parametrization is unique:

\(a_1^i=1\) for each \(i\),

and

\((a_1^j a_k^l)=1\) whenever \(j=1\) or \(l=1\) for all \(i\) and \(k\).

We are interested in developing simple but explanatory models of this kind, and to find such models we use the likelihood ratio test on hypotheses of the type

\[ H_0: a_1 J = 1 \text{ for all } J, \text{ some } i, \]

or

\[ H_0: (a_1 J a_k)^l = 1 \text{ for all } J \text{ and } l, \text{ some } i \text{ and } k, \]

i.e. we choose

\[ w = 2(\hat{l} - \hat{l}_0) \]

as our test statistic, where \(\hat{l}\) and \(\hat{l}_0\) are the maximal values of the log-likelihoods under our two competing models. It is well known that the
likelihood ratio test is asymptotically \( \chi^2 \)-distributed under \( H_0 \), under some very general assumptions. The problem is that this approximation may not be good enough to be useful in situations with small samples.

In the research at the Section of Demography of the University of Stockholm, these methods frequently have been applied to a fertility survey of Swedish women. Small sample problems often appears, though the reason for them varies. Here are a few examples.

1. The grouping of the data in the study relative the levels of the background factors may result in some groups ("cells") into which practically no individual will fit, such as highly educated 17-year-olds.

2. The state of origin of an investigation may sometimes be a state that is highly uncommon for some of the cells. For the Swedish fertility survey of 1981 concerning women born 1936-60, we have few married women among the young and highly educated.

3. The transition we study may just concern a small group of women. For example we may be interested in the demographic behaviour of Swedish single mothers who have never lived in a conjugal union. This is a very small group in the fertility survey.

4. Too many levels on a factor will make the number of observations per level small. We may for example suspect a certain kind of time structure for the intensity function. This would motivate a large number of time intervals if we had data enough.

In many situations, a cell with few observations is of little or no interest to the substantive investigation, and one may as well exclude its content from the study. Unfortunately, this will remove some information concerning other background factors, and interesting trends may become less distinct. An alternative to excluding cells is to merge two or more levels into one. If the original levels have been chosen carefully to distinguish between certain interesting population groups, level combination may weaken a perceived pattern.

Analysis of small data sets (see e.g. Broström, 1986, B. Hoem, 1983, Jensen, 1985, and Schou & Veth, 1980) indicates that the real density
function of the likelihood ratio statistic lies to the right of the corresponding \( \chi^2 \)-distribution; i.e. we tend to reject more \( H_0 \)-hypotheses than our test strategy suggests, and we may base our further research on false assumptions, or else decide to choose a more conservative strategy than is necessary. Small samples problems may also direct the analysis at an earlier stage since we may choose not to analyze data sets we consider to be too small. This, not just because the small sample size will yield significant results just for very obvious trends, but since the researcher may feel that the test statistic cannot be trusted.

1.2 Outline of the paper

In Chapter 2 we recapitulate some facts about the likelihood ratio statistic and the Bartlett correction term. It is not difficult to obtain a correction term for models with just one main factor. A simulation study presented in Chapter 7 stress the already known fact that in many situations the likelihood ratio test is too conservative. Chapter 3 tries to give an explanation for this. For many combinations of sample size, intensity, and censoring time, the Bartlett correction, i.e. a correction using just the first term in the Taylor expansion of the expected value of the likelihood ratio statistic is not enough. In fact it looks as if the convergence of the expansion is quite slow.

Chapter 4 is devoted to a heuristic method for providing correction factors in intensity regression models with large sets of nuisance parameters. The basic idea is that the nuisance parameters, or some of them, are replaced by their ML-estimates. As an approximation to the correction term, we then use the correction for the simplified model with only a restricted number of parameters. Thus, correction terms for simple exponential distribution models are used to compute approximate correction terms for more complex models. We present the basic idea in Section 4.1. In Section 4.2, we motivate an approximate correction term for test of interactions. Arguments analogous to those in Section 4.2 are useful to motivate a use of the correction terms presented in Chapter 4 in multi-factor models. This is shown in Chapter 5.
In Chapter 6, we discuss some correction terms for simple models with realizations from one exponential distribution. We stress that simulations can be a method to find such correction terms. A Bartlett correction where we argue conditionally on the number of observed occurrences is presented in Section 6.3. Computations of correction terms are discussed in Chapter 7. We describe a simulation study, and give a few results. We also comment on the use of Lawley's formula to compute the Bartlett correction. In Chapter 8, finally, we discuss the appropriate number of degrees of freedom for data sets with 0 occurrences in one of the cells. Our conclusion is that to decrease the number of degrees of freedom is something that should be done only if the cells with null occurrences have very little total exposure time.

2. The Bartlett correction factor.

We introduce the Bartlett correction factor by considering a likelihood ratio test of a parametric hypothesis \( H_0 \) of the type in Section 1.1. Let \( w \) be the likelihood ratio test statistic, and assume that under \( H_0 \),

\[
w = x^2(r).
\]

Suppose that under \( H_0 \),

\[
Dw = r + B + O(n^{-2})
\]

where \( n \) is the dimension of the observed random vector and \( B \) is \( O(n^{-1}) \). Then

\[
w' = w(1 + B + r)
\]

has an expected value closer to \( r \) and may therefore have a distribution closer to the \( x^2(r) \) distribution than \( w \) has. See Barndorff-Nielsen and Cox (1984) and Jensen (1985). The function \( B \) depends on the values of the model parameters and is called a Bartlett correction. Lawley (1956) expressed a general formula for one version of \( B \). See e.g. Klingberg (1987).
Suppose that we have \( n \) observations from an exponential distribution with a fixed censoring time and intensity parameter \( \theta \). Then the log likelihood may be expressed as

\[
\ell = N \log \theta - \theta T,
\]

and the maximum log likelihood as

\[
\hat{\ell} = \frac{N}{T} \log \frac{N}{T} = \bar{N},
\]

where \( N \) is the number of non-censored individuals and \( T \) is the total exposure time. According to Lawley's formula we have, for the hypothesis \( \theta = \theta_0 \) that

\[
B(\theta_0, \mu) = \frac{1}{6\mu} - \frac{2\mu'\theta_0}{\mu^2} + \frac{\theta_0^2 \mu''}{\mu^2} + \frac{2(\mu'\theta_0)^2}{\mu^3},
\]

where \( \mu \) is the mean value of \( N \), and \( \mu' \) and \( \mu'' \) are derivatives of \( \mu \) with respect to \( \theta \). This formula is valid not just for the example above but whenever (2.1) holds, including for the case with varying censoring times. We note that the censoring time and the intensity are related such that the exponential model above is equivalent to an exponential distribution model with intensity 1 and censoring time \( \theta T \). Thus, using \( \mu^* \) to denote the mean value of \( N \) under the latter model, formula (2.2) may just as well be written as

\[
B(\mu^*) = \frac{1}{6\mu^*} - \frac{2\mu'^*}{\mu^*} + \frac{2\mu'^*}{\mu^*} + \frac{2\mu'^*}{\mu^*},
\]

which is the standardized notation for the Bartlett correction term we will use henceforth.

Now, assume that we have a situation with more than one cell sample, and assume that the log likelihood can be written as

\[
\ell = \sum_i \left( N_i \log \theta_i - \theta_i T_i \right),
\]

where \( N_i \) and \( T_i \) are the number of occurrences and the exposure time in cell \( i \), respectively, and \( \theta_i \) is the corresponding intensity. Then, whether
the \( N_i \)'s are independent or not, we can write the Bartlett correction for
the hypothesis

\[
H_0: \theta_i = \theta_i^0 \text{ for each } i, \text{ and some fixed numbers } \theta_i^0, \quad (2.4.a)
\]
as

\[
\sum B(\mu_i) \quad (2.4.b)
\]

where \( \mu_i \) is the "standardized" mean value function corresponding to cell
1. Note that the structure of the mean value function may depend on param-
eters beside the intensity parameters in question. Now, consider the one-
factor model

\[
\lambda \equiv \theta \theta_1^0,
\]

where the \( i \)'s represent the levels of the factor. Suppose that we want to
test homogeneity, that is

\[
H_0: \alpha_i = 1, \text{ each } i. \quad (2.5.a)
\]

If we use invariance arguments (see McCullagh and Cox, 1981) and use the
\( \lambda_i \)-parameters to compute the Bartlett correction for the hypothesis

\[
H_0: \lambda_i = \lambda_j, \quad i \neq j.
\]

Lawley's formula now yields the Bartlett correction in the following form

\[
B = \sum B(\mu_i) - B(\Sigma \lambda_i). \quad (2.5.b)
\]

A comparison between the Bartlett corrections for the situations
(2.4) and (2.5) in a situation with the same number of degrees of freedom
in the two cases, identical intensity values, and the same number of indi-
viduals in each cell will show that the correction is greater in case
(2.5). Compared to (2.4), we have added one extra cell sample in case
(2.5) because we have a nuisance parameter which adds one further term to
the Bartlett sum. The term subtracted in case (2.5) is less than the one
added, since the \( B \)-function decreases when the sample size increases.

For models with more factors, and thus more nuisance parameters, it
is difficult to compute the Bartlett correction. Whenever we are able to
compute the expected number of occurrences explicitly we can always use Lawley's formula. However the computer time needed for each calculation of the correction will be considerable, and simulation studies of the correction's properties time consuming. We are therefore interested in simple approximations of the correction term. For example, the Bartlett correction for the case (2.4) could be useful even in the somewhat more complicated model (2.5). In the next chapter we will indicate that it in many situations will be insufficient to correct the likelihood ratio statistic with the $n^{-1}$-term in the expansion of $Ew$. This may imply that results that consider the expansion of $Ew$, and which are both exact and useful to us are difficult to obtain.

3. Expansions of the likelihood ratio statistic.

The conservative nature of some of the Bartlett corrected test statistics in the Chapter 7 makes us question the usefulness of our own result in Klingberg (1987, p 11). The rest term in formula (3.5) of that paper may be considerable for many of the cell sample sizes we consider in the appendix.

We therefore evaluate the expected value of the likelihood ratio statistic for the model with a sample from a censored exponential distribution in order to find the term that involves $n^{-2}$. Let us express the expansion as

$$Ew = 1 + \sum_{i=1}^{\infty} \frac{c_i}{n^i}$$

(3.1)

where the $c_i$ are coefficients that depend on the intensity and the censoring time but not on the sample size.

Let $Y, Y_1; i = 1, 2, \ldots, n$, be exponentially distributed with intensity $\theta$, and censored at time $T$. We may write the likelihood ratio statistic as

$$w = 2m[v - \ln(1+v)]$$

where $m$ is the number of occurrences, and $v = y+z$, where, assuming that $Y_1, \ldots, Y_m$ were not censored,
\[ y = \frac{e^m}{\sum_{i=1}^{m} (Y_i - \mu)}, \text{ and } z = e^T \frac{K - ne^{-\theta T}}{\ln(1 - e^{-\theta T})}. \]  

(3.2)

Here \( K \) is the number of censored individuals, and \( \mu \) is the mean value of \( Y_i \), i.e. the mean value of \( Y \) conditioned on the outcome that \( Y \) was not censored. We now evaluate \( w \) in terms of \( v \) to get

\[ w = \sum_{i=2}^{\infty} \frac{(-v)^1}{i(i-1)}. \]  

(3.3)

In order to see how many terms we need to find the \( n^{-2} \)-term, we look at products of the form \( y^k z^l \). These are what we will have to face when we develop formula (3.3). We get

\[ y^k z^l = \frac{e^k}{m^k} \frac{(\sum_{i=1}^{m} Y_i - \mu)^k}{(1 - q)^1}, \]  

where we have put \( p = e^{-\theta T} \) and \( q = 1 - p \). To find the mean value of \( y^k z^l \) we condition on \( K \) and calculate the term of the largest \( m \)-order in the \( k \)-th moment above. We obtain

\[ E(y^k z^l) = E\left(\frac{e^k}{m^k} \frac{(K - np)^1}{(1 - q)^1}\right), \]  

We evaluate \( m^{-1} \) to get

\[ \frac{1}{m^k} = \left( \sum_{j=0}^{\infty} \frac{(K - np)^1}{(mq)^1} \right)^{-j}. \]  

(3.4)

Now \( E(K - np)^1 = O(n^{INT(1/2)}) \), which together with (3.4) yields

\[ E(y^k z^l) = O(n^{INT(1/2)} + INT(1/2) - k - 1)). \]

From this we see that we do not have to go beyond terms of order 6 to find the \( n^{-2} \)-term of the evaluation (3.1). We also see that for each new term we would like to compute in (3.1) we have to calculate two extra terms in (3.3). Computing the first five terms of (3.3) yields
where we have put $s=\theta T$. As $s \to 0$ we find, just as we did for $c_1$ (the Bartlett correction) that the expression becomes simple. We then have

$$
\frac{c_2}{n^2} = \frac{1}{6n^2q^2},
$$

(3.5)

We clearly see that for small $nq$-values the $n^{-2}$-term gives a considerable contribution to the expected value of the likelihood ratio statistic, actually one that is comparable to the $n^{-1}$-term. We have not completed any computation of further terms in the expansion (3.3), but we will provide a method to find further coefficients $c_i$ for small $\theta T$-values. To see the tendency of the correction term when $\theta T \to 0$, it is sufficient to study the result when $T \to 0$. We will compare the expected value of the likelihood ratio statistic given the number of occurrences $m$ with the same statistic for the case of type II censoring (i.e. we end the experiment immediately after we observe the $m$:th occurrence).

Given $m$ occurrences, the expected difference between the two likelihood ratio statistic is

$$
\tilde{w} = 2mE_m[\log T_2 - \log T_1],
$$

where $E_m$ denotes the conditional expectation given $m$, $T_1 = \sum Y_1$, and $T_2$ is the total exposure time for the type II censoring experiment. To see that $\tilde{w} \to 0$ when $T \to 0$ we may argue conditionally on the time when the last occurrence took place. Thus for small $\theta T$-values the expected likelihood ratio value given $m$ is approximately the same as in the type II censoring case with $m$ occurrences. We thus have (see e.g. Broström, 1986)

$$
\tilde{w} \approx E [1 + \frac{1}{6m} - \frac{1}{60m^3} + \frac{1}{120m^5} + O(m^{-7})].
$$

(3.6)

An evaluation of this formula to get a sum of the form
yields
\[ d_3 = \frac{19}{60} \quad \text{and} \quad d_4 = \frac{9}{10}, \]

and for small values of \( \theta T \) we have \( c_1 = d_1 \). This is a surprising result considering the conservative property of the uncorrected likelihood ratio test. We had expected at least one negative term to help neutralize the effect of the two positive coefficients \( c_1 \) and \( c_2 \).

We conclude this chapter with a remark about two simulation studies with censored exponential distribution models (Broström, 1986, and Jensen, 1985). In those two simulations the \( nq \)-values from formula (3.5) are much higher than 1, except for a few situations with small \( n \)-values that Broström considers and in which the correction does not work that well. Thus, they do not investigate the cases where our study suggests that distribution of the Bartlett corrected likelihood ratio test still cannot be approximated by that of the chi-square.

4. Simplified correction terms.

In this chapter we will motivate the use of correction terms from simple models to compute approximate corrections of the likelihood ratio test for more complex parametrizations. We will use \( C(\omega) \) to denote a good correction term in the one-parameter case. In many situations the Bartlett factor \( B(\omega) \) is such a correction. Two different methods to extend the use of simple correction terms are presented.
4.1 Nuisance parameters replaced by their ML-estimates in the parametrization of the model.

Suppose we are interested in intensities of the form

\[ \phi_0 \cdot \alpha_1^{(1)} \cdot \alpha_2^{(2)} \cdot \ldots \cdot \alpha_n^{(n)} \cdot \beta^{(J)} \]  

where the \( \alpha \)- and the \( \beta \)-parameters may be either main effects or interaction parameters in a parametrization such that

\[ \alpha_k^{(1)} = 1 \text{ for all } k \text{ and } l_k \in I_k, \]

where \( I_k \) is a set of levels of factor \( k \),

\[ \beta^{(J)} = 1 \text{ for all } j \in J, \]

where \( J \) is a set of levels of the "\( \beta \)-factor". Here, \( \beta \) is our parameter of interest, the others are nuisance parameters. We wish to test

\[ H_0: \beta(J) = 1, \ \forall J, \]

and for this we use the log likelihood ratio test.

The presence of the nuisance parameters make it difficult to compute the Bartlett correction. (See e.g. Klingberg, 1987.) However, if we knew the values of those parameters (or at least of some of them), the Bartlett term would be easier to calculate since the number of parameters involved would be less. For simple models, Lawley's formula (see e.g. Klingberg, 1987) is reduced to a fairly simple expression. This motivates an interest for corrections in those simple situations, and in what impact the presence of nuisance parameters will have on the likelihood ratio statistic.

By means of two examples we will now present one suggestion for an approximation to the correction term for the likelihood ratio test in cases with nuisance parameters.

**Example 4.1**

Model: \( \lambda_{ijk} = \theta_0 \cdot \psi_1 \cdot \psi_j \cdot \xi_k \cdot \gamma_{ij}; \ i = 1, 2; \ j = 1, 2, 3, 4; \ k = 1, 2, 3 \)

\( \psi_1 = \psi_3 = \xi_1 = 1, \ \gamma_{1j} = 1 \text{ if } i = 1 \text{ or } j = 1, \)
with \( H_0: \gamma_{1j} = 1 \) for each \( i \) and \( j \).

To calculate a simplified correction term we behave as if the ML-estimates of the nuisance parameters were their true values. Thus we behave as if

\[
\hat{\theta}_0 = \theta_0^0, \quad \hat{\varphi}_2 = \varphi_2^0, \quad \hat{\psi}_j = \psi_j^0 \quad \text{for } j = 2, 3, 4,
\]

and

\[
\hat{\xi}_k = \xi_k^0 \quad \text{for } k = 2, 3.
\]

That means that so far as the calculation of the correction is concerned we behave as if our model were

\[
\lambda_{1jk} = \theta_0^0 \cdot \varphi_1^0 \cdot \psi_j^0 \cdot \xi_k^0 \cdot \gamma_{11j}, \quad i = 1, 2; \quad j = 1, 2, 3, 4; \quad k = 1, 2, 3,
\]

for known \( \theta_0^0, \varphi_1^0, \psi_j^0, \xi_k^0 \), where we have put \( \varphi_1^0 - \varphi_1^0 - \xi_1^0 = 1 \). Then \( \lambda_{1jk} \) would be completely known for all \( j, k \), and so would \( \lambda_{1jk} \) for all \( i, k \). The only intensities that would not be completely known would be \( \lambda_{2jk} \) for \( j = 2, 3, 4 \) and \( k = 1, 2, 3 \). We could behave, therefore, as if we had a model that involved only 3x3 cells of unknown intensities a priori, and only three unknown parameters (\( \gamma_{22}, \gamma_{23}, \) and \( \gamma_{24} \)). The hypothesis \( H_0 \) would transform to

\[
H_0': \lambda_{2jk} = \theta_0^0 \varphi_2^0 \psi_j^0 \xi_k^0, \quad j = 2, 3, 4; \quad k = 1, 2, 3,
\]

which is fully specified. Thus, the correction term is calculated as if the parameters of interest (although they are interaction parameters in reality) correspond to levels of an unknown main factor in a situation with many preknown baseline intensity levels. Note that the value of the simplified correction term will depend on which cells we have included in our new hypothesis \( H_0' \).

**Example 4.2**

Assume that we are faced with the model of Example 1 but that we choose to replace not all the nuisance parameters by their ML-estimates, only to let

\[
\hat{\xi}_k = \xi_k^0 \quad \text{for } k = 2, 3.
\]

The related test hypothesis now becomes
i.e. we calculate the correction term as if we are analyzing the presence of an interaction factor in a case with three different samples whose intensity relation are known through the $\epsilon_k^0$-values.

In general, our suggested method to test $H_0$ is to use a correction term for a situation where the nuisance parameters (or some of them) are replaced by ML-estimates.

In the first example we see that the parametrization will determine which cells to consider in the simplified situation, and thus the value of the simplified correction. We also note that the more estimates we use, the simpler the obtained expression will become. On the other hand, the divergence from the real correction term may increase in this manner.

Using the form (4.1) for the intensities, the symbols $w$ and $w_r$ for the two log-likelihood ratios (the original one and the restricted one), and $\theta_0$ and $\theta_0^0$ for the real values and the estimates of the nuisance parameters (incorporating $\theta_0$); our heuristic argument for our interest in the simplified situation suggests that

$$P_{\theta_0} \{ wzc \} = P_{\theta_0^0} \{ w_rzc \}.$$  \hspace{1cm} (4.2)

If (4.2) is true, we trust that the two corresponding corrections will be approximately equal.

The correction term $\hat{C}$ for the simplified situation is our suggestion for an approximation to the original one, $C$. If we choose to estimate all nuisance parameters (as we did in Example 1) the function $\hat{C}$ can be expressed analogously to formula (2.4.b) as

$$\sum_{i \in I_C} C(\mu_i^0),$$  \hspace{1cm} (4.3)

where $\mu_i^0$ is the mean value function for the occurrences in cell $i$ after we have standardized with the ML-estimates, computed as if those were the
true values, and \( I \) is the set of baseline levels of the test factor. If instead we choose not to use the estimate of the main factor parameters involved in the test hypothesis, but to only use the other ML-estimates, and we use the common parametrisation from the examples, \( \hat{c} \) can be expressed with the help of formula (2.5.b). We write
\[
\hat{c} = \sum_{i \in I^C} C(\mu_i^0) + C(\sum_{j \in I} \mu_j^0) - C(\sum_{k \in IUC} \mu_k^0).
\]
(4.4)

In cases where the parameters of interest are interaction parameters, as pointed out in Example 2, we also have the option of estimating neither the baseline intensity level nor the parameters corresponding to levels of the two main factors involved in the interaction. The true structure of the correction term for a model with just two main factors where we test if there is an interaction between them cannot be found with the methods indicated in Chapter 2, since we do not have explicit formulas for the intensity estimates under the hypothesis. Instead we will provide an approximation which has the structure one would expect intuitively, and which in fact is the result when we apply the technique of Section 4.2 to a test of interaction. We illustrate this approximation using a model with intensities on the form
\[
\lambda_{ij} = \Theta_0 \psi_1 \psi_j \psi_{1j}, \quad \psi_1 = \psi_1 = \psi_{11} = \psi_{1j} = 1, \text{ for each } i, j.
\]
(4.5)
where \( i = 1, 2, \ldots, l_1 \), and \( j = 1, 2, \ldots, l_2 \), and for which we test \( H_0: \psi_{1j} = 1 \) for each \( i, j \). The following formula for the correction term can be motivated heuristically.
\[
\sum_{i \in I^C} C(\mu_{i,j}^0) = \sum_{i \in I} C(\Sigma_{i,j}^0) = \sum_{j \in I} C(\Sigma_{i,j}^0) + C(\Sigma_{i,j}^0),
\]
(4.6)
where the two indices correspond to the levels of the two factors.
4.2 Nuisance parameters replaced by ML-estimates in the likelihood equation.

In this section we will motivate our notion that formula (4.6) holds approximately for model (4.5). Thus, we again consider this model and the hypothesis concerning the interaction factor quantified by the $\gamma_{1j}$.

We introduce the notation $m_{1j}$ and $Y_{1j}$ for the occurrences and exposures for the $(1,j)$-cell. The restricted ML-estimates are given from the following system of equations:

\[
\begin{align*}
\hat{\theta}_i &= \frac{\Sigma_j m_{1j} \psi_j}{\Sigma_j \hat{\theta}_0 Y_{1j}}, \\
\hat{\psi}_j &= \frac{\Sigma_i m_{1j}}{\Sigma_i \hat{\psi}_0 Y_{1j}}, \quad \text{and} \\
\hat{\theta}_0 &= \frac{\Sigma_{ij} m_{1ij}}{\Sigma_{ij} \hat{\psi}_0 Y_{1j}}.
\end{align*}
\]

Without restrictions, we have

\[
\hat{\theta}_{ij} = \frac{m_{ij}}{Y_{ij}}.
\]

We now insert the above expressions in the log-likelihood ratio to obtain

\[
w = 2(\hat{1} - \hat{1}),
\]

where

\[
\hat{1} = \Sigma_{1j} \frac{m_{1j}}{Y_{1j}} \log \frac{m_{1j}}{Y_{1j}}.
\]

and

\[
\hat{1} = \Sigma_{1j} \left( \frac{\Sigma_i m_{11} \psi_{11}}{\Sigma_i \psi_1 Y_{11}} + \frac{\Sigma_k m_{k1}}{\Sigma_k \psi_{k1} Y_{k1}} - \log \frac{\Sigma_k m_{k1}}{\Sigma_k \psi_{k1} Y_{k1}} + 1 \right).
\] (4.7)

To compute $E_1$, we just note that we in this case have a full model, which implicate that all cell intensity estimates can be calculated explicitly. Thus, for each cell the $n^{-1}$-term in the expected likelihood ratio expansion can be expressed by formula (2.2). We obtain

\[
\Sigma \hat{E}_1 = \Sigma (1 + C(\mu_{1j})).
\]

We are not able to use the same methods to evaluate $E_1$, since we do not have explicit expressions for the estimates of $\lambda_{1j}$. The trick to
obtain formula (4.6) is to simply replace \( \tilde{\varphi}_1 \) and \( \tilde{\psi}_j \) in formula (4.7) with the true values \( \varphi_1^0 \) and \( \psi_j^0 \). We will then calculate the mean value of something less than the maximum likelihood value, but not much less if

\[
\varphi_1^* = \frac{\Sigma_1 m_{1l}}{\Sigma_{\psi_1 y_{1l}}} \quad \text{and} \quad \psi_j^* = \frac{\Sigma_k m_{kj}}{\Sigma_{\psi_k y_{kj}}}
\]

are reasonable approximations of \( \tilde{\varphi}_1 \tilde{\psi}_0 \) and \( \tilde{\psi}_j \tilde{\psi}_0 \). We now obtain

\[
2E\tilde{I} = \Sigma f_i + \Sigma g_j - h
\]

where

\[
f_i = 2E[m_i (\log \frac{m_i}{\Sigma_1 \psi_{1l}} - 1)], \quad g_j = 2E[m_j (\log \frac{m_j}{\Sigma_k \psi_{kj}} - 1)]
\]

and

\[
h = 2E[m_\ldots (\log \frac{m_\ldots}{\Sigma_k \psi_{k1} \psi_{1l}} - 1)]
\]

Here we have introduced dots to indicate summation. We now note that the above expressions can be regarded as expected values of maximum likelihood values for samples from exponential distributions with varying censoring times. For \( f_i \) we have \( n_i \) realizations from an exponential distribution with intensity \( \varphi_1 \psi_0 \). A number \( n_j \) of those are censored at time \( \psi_j T \).

Thus,

\[
2E\tilde{I} = l_1 + l_2 - 1 + \Sigma_{i} C(\mu_i) + \Sigma_{j} C(\mu_j) - C(\mu_\ldots)
\]

which motivates our interest in formula (4.6).

Remark: Arguments similar to those in this section will provide a structure for a correction term when testing interactions of higher order than two.
5. Correction terms for models with many nuisance parameters.

We will now motivate the idea that the correction terms presented in Chapter 4 may be useful when we test models with a large set of nuisance factors. The technique we will use is the same as in Section 4.2. We will not provide a general statement, but will instead consider an example in which we have added two extra main factors and one interaction to model (4.3).

Example 5.1.

Model: \[ \lambda_{ijkl} = \theta_0 \cdot \phi_i \cdot \psi_j \cdot \xi_k \cdot \delta_l \cdot y_{ij} \cdot \eta_{il}; \]

\[ i = 1, \ldots, 11; \quad j = 1, \ldots, 12; \quad k = 1, \ldots, 13; \quad l = 1, \ldots, 14; \]

\[ \phi_i = \psi_j = \xi_k = \delta_l = 1, \quad y_{ij} = 1 \text{ if } i = 1 \text{ or } j = 1, \]

and \[ \eta_{ii} = 1 \text{ if } i = 1 \text{ or } l = 1. \]

with \[ H_0: \quad y_{ij} = 1 \text{ for each } i \text{ and } j. \]

Without the restriction \[ H_0 \] the normal equations are

\[ \hat{\theta}_0 = \frac{\sum_{ijkl} \phi_i \psi_j \xi_k \delta_l y_{ij} \eta_{il} \hat{y}_{ijkl}}{m}, \]

\[ \hat{\phi}_i = \frac{\sum_{jkl} \hat{\theta}_0 \psi_j \xi_k \delta_l y_{ij} \eta_{il} \hat{y}_{ijkl}}{m}, \]

\[ \hat{\psi}_j = \frac{\sum_{ikl} \hat{\theta}_0 \phi_i \xi_k \delta_l y_{ij} \eta_{il} \hat{y}_{ijkl}}{m}, \]

\[ \hat{\xi}_k = \frac{\sum_{ijl} \hat{\theta}_0 \phi_i \psi_j \delta_l y_{ij} \eta_{il} \hat{y}_{ijkl}}{m}, \]

\[ \hat{\delta}_l = \frac{\sum_{ijkl} \hat{\theta}_0 \phi_i \psi_j \xi_k y_{ij} \eta_{il} \hat{y}_{ijkl}}{m}. \]
\[ \hat{\eta}_{1l} = \frac{m_{1l}}{\sum_{l} \hat{\theta}_{0l} \hat{\psi}_{l} \hat{\epsilon}_{kl} \hat{\delta}_{l1l} Y_{1jkl}} \]

where \( Y_{1jkl} \) is the exposure time for the level combination \((i, j, k, l)\).

Under \( H_0 \), the system of equations is the same but with the test parameters excluded. Write the logged maximum likelihood under the unrestricted model as

\[
\hat{l} = m_{1l} \log \hat{\theta}_0 + \sum_{l} m_{1l} \log \hat{\psi}_l + \sum_{j} m_{1l} \log \hat{\psi}_j + \sum_{k} m_{1l} \log \hat{\epsilon}_k + \sum_{i} m_{1l} \log \hat{\delta}_i
\]

Under \( H_0 \), the system of equations is the same but with the test parameters excluded. Write the logged maximum likelihood under the unrestricted model as

\[
\hat{l} = m_{1l} \log \hat{\theta}_0 + \sum_{l} m_{1l} \log \hat{\psi}_l + \sum_{j} m_{1l} \log \hat{\psi}_j + \sum_{k} m_{1l} \log \hat{\epsilon}_k + \sum_{i} m_{1l} \log \hat{\delta}_i
\]

Now replace the maximum likelihood estimates in (5.3) with approximations obtained by substituting the ML-estimates with the true values on the right side of the equality sign in (5.2). This yields

\[
\hat{l} = m_{1l} \log \hat{\theta}_0 + \sum_{l} m_{1l} \log \hat{\psi}_l + \sum_{j} m_{1l} \log \hat{\psi}_j + \sum_{k} m_{1l} \log \hat{\epsilon}_k + \sum_{i} m_{1l} \log \hat{\delta}_i
\]
To keep the formula compact we have not indicated the true value of the parameters with the superscript 0. We now replace the log $\theta_j$, the log $\phi_i$, and the log $\psi_j$ by their ML-estimates, to obtain

$$\hat{1} = -(\sum_{l} \log \frac{m_{ij}}{\Sigma_{l} k_{i} n_{11} Y_{1,l} j k l}) + (\sum_{j} \log \frac{m_{ij}}{\Sigma_{j} k_{j} n_{11} Y_{1,j} l j k l}) + (\sum_{k} \log \frac{m_{ij}}{\Sigma_{k} k_{i} n_{11} Y_{1,j} l j k l}) + (\sum_{l} \log \frac{m_{ij}}{\Sigma_{l} k_{i} n_{11} Y_{1,j} l j k l}) + (\sum_{l} \log \frac{m_{ij}}{\Sigma_{l} k_{i} n_{11} Y_{1,j} l j k l})$$

(5.5)
We see that all the terms in (5.5) can be written in the form of (2.1.b), in the same manner as for the random variable in formula (4.9). For the restricted model we obtain a similar expression, but without the term associated with the test factor in (5.4). Taking the expectation under $H_0$ of twice the difference between these two terms gives us formula (4.6).

Remark 1: The structure of formula (5.4) can also be used to suggest an extended use of formula (4.4) in situations where the redundancy of a main factor is tested in the presence of nuisance parameters.

Remark 2: For models of the type (4.1), the arguments hold irrespective of what model we consider. This implies that all the correction terms obtained by replacing parameters with ML-estimates in the parametrization of the intensities are also covered by the result in this chapter. In fact, in situations as in Example 4.2 where we only replace parameters not involved in the parametrization of the test hypothesis, we will get the same correction term whether we make the substitution in the model parametrization or in the likelihood equation.

Remark 3: How well formula (4.6) approximates the real correction term will of course depend on how well approximations such as those in (4.8) work as substitutes for the ML-estimates in formula (5.4). It is possible that the more nuisance parameters the model has, the less accurate formula (4.6) will be. At least in situations where the unrestricted model is completely parametrized, the correction term is less than the real correction, since, generally, the estimates used for the restricted model are not the ML-estimates. This implies that we will not get too conservative a test.
6. Correction terms for simple models.

In this paper we have been dealing with two different problems which both must be solved if we want to apply our ideas in order to use the corrected likelihood ratio test successfully:

1. Finding correction terms for simple censored exponential distribution models.

2. Finding a method where the results from fairly simple models can be used to correct test statistic for more complicated parameterizations.

The paper has focused primarily on the second problem, but in this chapter we will discuss some correction terms for the likelihood ratio test when we have n realizations from a censored exponential distribution with intensity parameter \( \theta \) and censoring time \( T \).

6.1 The Bartlett correction.

In Chapter 3 we proved that the convergence of the Taylor expansion of \( D_w \) is slow when the values of \( n \cdot (1-e^{-\theta T}) \) is small (close to 1 or less). We also found that the \( n^{-2} \)-term in the expansion could be computed without much difficulty. A simulation study would reveal the ranges in which the Bartlett correction and the two term correction will be useful.

For large values of \( \theta T \) we can approximate the distribution with the uncensored exponential distribution for which the first terms of the Taylor expansion are fairly simple to compute. For this model, we have

\[
D_w = 1 + \frac{1}{6n} - \frac{1}{60n^3} + \frac{1}{126n^5} + O(n^{-7}).
\]  

(6.1)

6.2 Simulations as a way to find correction values.

With the above difficulties of finding a suitable correction term for certain parameter combinations, it might be better to use correction terms obtained through simulation. We simulate realizations from a censored exponential distribution to obtain estimates of the expected values of the likelihood ratio statistic for a set of \( n \)- and \( \theta T \)-values. We then use
regression methods to approximate the correction term as a function of \( n \) and \( \theta T \). We would like to stress that this simulation study can be carried out either by using different kernels for each combination of \( n \) and \( \theta T \), or by letting the same kernel initiate all the models, and just change kernel after each completed simulation lap.

An alternative to correction terms in which we multiply the statistic with a factor, would be to correct the test statistic's expected value and variance separately. Let \( w \) be a LR-statistic with \( m \) degrees of freedom. Then

\[
Dw = m + c, \quad \text{and} \quad Vw = 2m + d
\]

for some numbers \( c \) and \( d \). We wish to find correction terms \( a \) and \( b \) such that

\[
\frac{w-a}{b} = m, \quad \text{and} \quad \frac{w-a}{b} = 2m.
\]

This yields

\[
a = c + m \left( 1 - \frac{2m+d}{2m} \right), \quad \text{and} \quad b = \frac{2m+d}{2m}.
\]

We will now illustrate how a multiplicative correction term such that the expected value of the corrected test statistic is exactly the assumed number of degrees of freedom will correct the variance; it all comes down to the relation between \( c \) and \( d \).

By simple computation, we see that

\[
V \left( \frac{w}{1+(c/m)} \right) = 2m \quad \text{if and only if} \quad d = 2m \left( \frac{2c}{m} + \frac{c^2}{m^2} \right).
\]

If

\[
d < 2m \left( \frac{2c}{m} + \frac{c^2}{m^2} \right)
\]

(6.2)
the corrected test statistic will have a lower variance than expected. In the simulation study we compare the estimated variance for the LR-statistic, which gives us an estimate of $d$, with the variance for the LR-statistic if the estimate of the expected value of the LR-statistic was in fact $Ew$. (This is an approximation of the term on the right side of (6.2).)

### 6.3 An alternative correction term.

In this section we will provide a result from Broström (1986) about the Bartlett correction when we consider the likelihood ratio statistic conditioned on the number of observed occurrences, and use type II censoring i.e. the study ends after a pre-determined number of occurrences.

Let $Y_i \sim \exp(\theta_0)$ censored at time $T_i$, where $i=1,2,\ldots,n$. Given the observed occurrences $m$, we have for the likelihood ratio statistic

$$Ew = \sum_{j=1}^{k} \frac{1}{m} + O(m^{-3}).$$

(6.3)

where, the rest term can be calculated just as in the case of the uncensored exponential distribution. Formula (6.3) can be extended to hold for one-factor models, and we may argue analogously to Chapter 4 to find an approximate correction term for models with interactions. Even though the properties of this correction term are proved for another type of experiment, we may use it as a conditional correction term, where we argue as if the occurrences were fixed. The intensity parameters are not expressed explicitly in formula (6.3), thus the use of the methods of Chapter 4 in order to obtain a correction term for more complex parametrizations would not result in a need for the ML-estimates.
7. Software to compute and test correction terms.

7.1 An outline of a simulation study.

We have written a computer program that calculates correction terms for models with interaction parameters, given the Bartlett correction for the censored exponential distribution as a "correction kernel". We have simulated censored exponential distributions to estimate the expected value of the likelihood ratio statistic and the size of the rejection area, which is supposed to be 5% according to the asymptotic distribution of the test statistic. The program that calculates the correction term is written to be easy to use for users familiar with the program package for event history analysis at the Section of Demography, University of Stockholm (SUDA).

It is not difficult to change this program so that it will consider other "correction kernels" as well. The correction term it calculates has the structure of the correction for a test with a number of all samples (see formula (4.3)), that are not necessarily independent, for we have taken into consideration that the tested factor may be a time factor. The focus on this structure is a limitation, since the intuitively most interesting corrections are those where we have kept the structure typical for the test hypothesis (formula (4.4) in the main factor case, formula (4.6) in the interaction case). At the time when the study was made we were not sure about the limited utility of the Bartlett correction. We directed our interest to data sets with few observations in some of the cells. The correction terms that used all the observations (not just observations from some chosen cells) turned out to provide extremely conservative tests because the Bartlett corrected test of the intensity in a single cell is very conservative if the number of observations is small. Given the Bartlett correction as kernel, the best possible correction term would be obtained with a choice of basic levels corresponding to the cells with the smallest content.
7.2 Bartlett corrections in the presence of competing risks and time varying intensities.

We will show how we calculate the contribution to the Bartlett correction from a single cell in a model where we have competing risks, time varying factors, and a time variable. According to Klingberg (1987, page 21), the log likelihood can be written in the form (2.3), and thus the Bartlett correction is expressed by (2.2). In Klingberg (1987) we introduced the name "episode" for a single period in an individual's life under observation during which our specific intensity of interest is constant.

Remark: Note that with this definition the number of episodes will depend on the model. If we do not want to calculate a new set of episodes for each likelihood ratio test, we must relate our episodes to a set of main factors and levels that includes every single factor level we could consider in our study.

The important thing is that no factor changes level during an episode. The episodes were in fact invented as substitutes for individuals when computing Bartlett corrections for models with time varying intensities. Assume that we have a test factor A with a set of parameters $\alpha_1, \alpha_2, \ldots, \alpha_k$ associated with the k levels of A. The Bartlett correction contribution from level i can be written on the form (2.2) with $\theta_0$ replaced by $\alpha_i$, and with $\mu$ as the total expected number of occurrences for the episodes with level i on factor A. This expected value will of course be a sum of contributions from each single episode for which the level of factor A is i. The contribution from episode j can be written as

$$EN_j = \prod_{1: r(1) = r(j), s(1) < s(j)} e^{-\theta_1 T_i (1 - e^{-\theta_j T_j})},$$

where $N_j$ is the occurrences during episode j, the index 1 denotes other episodes, $r(1)$ is the individual associated with episode 1, $s(1)$ is the time when episode 1 ends, $\theta_1$ is the intensity during episode 1, and $T_1$ is the length of episode 1.
Given a starting point \( T_0 \), the length of an episode can be written as \( T_s - T_0 \) where \( T_s = \min(T_a, T_b, T_c) \), and \( T_a \) is the time of the next level change for any of the time varying factors (including the time factor), \( T_b \) is the time for the end of the observation interval, and \( T_c \) is the time for the next transition caused by a competing risk. In many cases a small error in the computation of the episodes expected occurrences will result from the fact that we may not know \( T_s \) for those episodes in which our transition of interest will take place.

7.3 A description of the software.

We have developed a set of computer programs that will provide long term results for the correction term and the likelihood ratio statistic when they are used together. Each computer program produces data to be used by the next program sequence. A command file initiates the simulation study. When one lap is completed, useless data is deleted, and a new simulation starts. The input parameters are the true values of the parameters, a description of the model, the sample size, and the test hypothesis. A program that simulates realizations from a censored exponential distribution creates a life history for each individual in the sample. The structure of the life histories is suited for the software used at SUDA to compute ML-estimates and likelihood ratio values. Thus, we use this software to produce the input for our correction computing program. This latter program also needs to know the length and the factor levels for each episode, an information we provide separately. When we have computed the correction value and the value of the corrected LR-statistic, we store the result temporarily to update our averages. This marks the end of one simulation lap. Below we illustrate this procedure.
7.4 Some simulation results.

In this simulation study, four different types of parametric models are considered, namely,

1. \( \lambda_{ij} = \theta_{ij} \psi_{ij} \); \( i=1,2; j=1,2,3,4; H_0: \psi_{ij}=1 \\forall i,j \).
2. \( \lambda_{ij} = \theta_{ij} \psi_{ij} \); \( i=1,2; j=1,2,3,4; H_0: \psi_{ij}=1 \\forall j \).
3. \( \lambda_i = \theta_i \); \( i=1,2,3; H_0: \theta_i = \theta_0 \); some constant \( \theta_0 \).
4. \( \lambda_i = \theta_i \); \( i=1; H_0: \theta = \theta_0 \); some constant \( \theta_0 \).

For each model we will consider a case where all intensities are equal, \( \lambda_{ij} = e^{-4.5} \), and the censoring time is constant at \( C=90 \). The sample sizes will vary, however. Below we show the different sample sizes for each one of the models.

<table>
<thead>
<tr>
<th>Model 1 &amp; 2</th>
<th>1=1</th>
<th>1=2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>j=1</strong></td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td><strong>j=2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td><strong>j=3</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>e</td>
<td>f</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td><strong>j=4</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>b</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>c</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>d</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>e</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>f</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model 3</th>
<th>1=1</th>
<th>1=2</th>
<th>1=3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a</td>
<td>b</td>
<td>c</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>5</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model 4</th>
<th>1=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>6</td>
</tr>
<tr>
<td>b</td>
<td>10</td>
</tr>
<tr>
<td>c</td>
<td>20</td>
</tr>
</tbody>
</table>

The results are given in Table 7.1. All corrections are computed with the true value of the intensities replaced by the ML-estimates under \( H_0 \).
Table 7.1

<table>
<thead>
<tr>
<th>Case</th>
<th>#sim.</th>
<th>% Rejections</th>
<th>\hat{\beta}_W</th>
<th>\hat{\sigma}_W^2</th>
<th>\tilde{\hat{\beta}}_W</th>
<th>\tilde{\hat{\sigma}}_W^2</th>
<th>\tilde{\hat{\beta}}</th>
<th>\tilde{\hat{\sigma}}_W</th>
<th>V*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>508</td>
<td>8.27 4.13</td>
<td>3.55 8.53</td>
<td>2.64 5.47</td>
<td>0.251 0.183</td>
<td>8.40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1b</td>
<td>508</td>
<td>6.50 4.53</td>
<td>3.13 7.56</td>
<td>2.78 5.97</td>
<td>0.126 0.043</td>
<td>6.53</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1c</td>
<td>508</td>
<td>8.07 4.92</td>
<td>3.23 7.77</td>
<td>2.99 6.63</td>
<td>0.083 0.077</td>
<td>6.96</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1d</td>
<td>508</td>
<td>10.83 5.91</td>
<td>4.06 8.30</td>
<td>3.60 6.56</td>
<td>0.143 0.352</td>
<td>10.99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1e</td>
<td>508</td>
<td>5.31 3.74</td>
<td>3.18 6.22</td>
<td>2.81 4.87</td>
<td>0.133 0.060</td>
<td>6.74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2a</td>
<td>996</td>
<td>5.92 4.52</td>
<td>3.26 6.64</td>
<td>2.87 5.22</td>
<td>0.126 0.087</td>
<td>7.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2b</td>
<td>1006</td>
<td>5.77 4.97</td>
<td>3.15 6.22</td>
<td>2.96 5.50</td>
<td>0.062 0.049</td>
<td>6.62</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2c</td>
<td>1006</td>
<td>5.77 5.07</td>
<td>3.13 6.75</td>
<td>2.99 6.19</td>
<td>0.042 0.043</td>
<td>6.53</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2d</td>
<td>1000</td>
<td>7.30 6.30</td>
<td>3.47 7.56</td>
<td>3.25 6.66</td>
<td>0.064 0.157</td>
<td>8.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2e</td>
<td>999</td>
<td>5.71 4.20</td>
<td>3.15 6.57</td>
<td>2.95 5.71</td>
<td>0.066 0.048</td>
<td>6.62</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3a</td>
<td>1022</td>
<td>9.40 1.47</td>
<td>3.56 8.35</td>
<td>2.45 3.83</td>
<td>0.460 0.186</td>
<td>8.45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3b</td>
<td>1022</td>
<td>6.95 4.40</td>
<td>3.24 7.90</td>
<td>2.57 4.87</td>
<td>0.267 0.081</td>
<td>7.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3c</td>
<td>1020</td>
<td>5.87 4.40</td>
<td>3.20 7.12</td>
<td>2.84 5.53</td>
<td>0.132 0.065</td>
<td>6.83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3d</td>
<td>1022</td>
<td>4.60 3.52</td>
<td>3.04 5.72</td>
<td>2.85 5.05</td>
<td>0.063 0.012</td>
<td>6.16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3e</td>
<td>1022</td>
<td>6.46 3.23</td>
<td>3.29 7.06</td>
<td>2.69 4.62</td>
<td>0.227 0.097</td>
<td>7.22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3f</td>
<td>1022</td>
<td>5.48 4.11</td>
<td>3.16 7.12</td>
<td>2.73 5.27</td>
<td>0.157 0.052</td>
<td>6.66</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4a</td>
<td>1024</td>
<td>5.37 4.98</td>
<td>1.05 2.04</td>
<td>0.97 1.69</td>
<td>0.082 0.053</td>
<td>2.21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4b</td>
<td>1027</td>
<td>6.62 6.04</td>
<td>1.08 2.20</td>
<td>1.03 1.95</td>
<td>0.050 0.083</td>
<td>2.33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4c</td>
<td>1026</td>
<td>5.26 4.87</td>
<td>1.07 2.35</td>
<td>1.05 2.24</td>
<td>0.022 0.069</td>
<td>2.29</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Above, \( \tilde{\hat{\beta}} \) is the mean value of the Bartlett approximation estimates, and \( \tilde{\hat{\sigma}}_W \) is the estimate of the error term derived from our estimate \( \hat{\beta}_W \) (the average of the LR-statistics) of \( \beta_W \). We also give the average of the corrected likelihood ratio statistics, and the estimated variances. We have used the notation \( \tilde{\hat{\beta}}_W \) for the \( \tilde{\hat{\beta}} \)-corrected test statistic. The last variable \( \tilde{\hat{\sigma}}_W \) denotes the true variance of the LR-statistic if \( \beta_W = \hat{\beta}_W \), and the mean value corrected statistic has the same variance as the asymptotic \( \chi^2 \)-distribution.

For the two simple models we compared the Bartlett correction with the intensities replaced by estimates of the "true" Bartlett correction. We present the results in Table 7.2.
Table 7.2

<table>
<thead>
<tr>
<th>Case</th>
<th>True Value</th>
<th>ML-Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%rej. Vw'</td>
<td>%rej. Vw'</td>
</tr>
<tr>
<td>3a</td>
<td>2.15</td>
<td>1.47</td>
</tr>
<tr>
<td>3b</td>
<td>4.40</td>
<td>4.40</td>
</tr>
<tr>
<td>3c</td>
<td>4.51</td>
<td>4.41</td>
</tr>
<tr>
<td>3d</td>
<td>3.52</td>
<td>3.52</td>
</tr>
<tr>
<td>3e</td>
<td>3.33</td>
<td>3.23</td>
</tr>
<tr>
<td>3f</td>
<td>4.21</td>
<td>4.11</td>
</tr>
<tr>
<td>4a</td>
<td>4.98</td>
<td>4.98</td>
</tr>
<tr>
<td>4b</td>
<td>6.13</td>
<td>6.04</td>
</tr>
<tr>
<td>4c</td>
<td>4.67</td>
<td>4.87</td>
</tr>
</tbody>
</table>

The number of simulations for the study was chosen so that we would be able to analyze many different models within a limited time schedule, and still get sufficiently strong results. Each single simulation required between 20 and 30 seconds of CPU-time, and we figured that approximately 1000 simulations per case would satisfy our needs. Table 7.3 contains 95%-confidence intervals for some parameter combinations valid in Table 7.1 and 7.2.

Table 7.3

<table>
<thead>
<tr>
<th>#sim.</th>
<th>Dw</th>
<th>Vw</th>
<th>%rej.</th>
<th>95%-confidence intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>3</td>
<td>6</td>
<td>0.05</td>
<td>(2.79, 3.21), (5.31, 6.81), (3.09, 6.91)</td>
</tr>
<tr>
<td>500</td>
<td>4</td>
<td>8</td>
<td>0.08</td>
<td>(3.75, 4.25), (7.09, 9.08), (5.62, 10.38)</td>
</tr>
<tr>
<td>1000</td>
<td>1</td>
<td>2</td>
<td>0.03</td>
<td>(0.91, 1.09), (1.83, 2.19), (1.94, 4.06)</td>
</tr>
<tr>
<td>1000</td>
<td>3</td>
<td>6</td>
<td>0.06</td>
<td>(2.85, 3.15), (5.50, 6.56), (3.65, 6.35)</td>
</tr>
<tr>
<td>1000</td>
<td>4</td>
<td>8</td>
<td>0.07</td>
<td>(3.82, 4.18), (7.33, 8.75), (6.42, 8.56)</td>
</tr>
</tbody>
</table>

7.5 A few remarks.

A. In every case except for case 3d, the likelihood ratio test rejects the hypothesis too often. The unexpectedly conservative behaviour of the test in this model seems to be more the effect of a small deviation among the LR-realizations than a decreased average level. Eight of the 19 rejection percentages lie outside the 95%-confidence interval for the rejection rate 0.05. For the estimates of Dw and Vw are those numbers are 11 and 14, respectively.
E. The number of rejections for the likelihood ratio test does not follow the expected pattern of rejection percentages closer to 5 for larger sample sizes very closely. For model 1 and model 2 we have fewer rejections in case e than in cases b and c where the samples are larger. For model 4, case b is the one with the largest rejection percentages. The results for model 3 are more regular in this sense; the only surprise is the high percentage in case f. One explanation is probably the relatively small number of simulations in the study. The standard error is 0.019 for 500 simulations and 0.014 for 1000 simulations.

F. Except for simulation 1d and 2d, where two samples contained only one individual, and simulation 4a, the corrected test rejected less often than the supposed 5%. For the two first models we have actually chosen cells in order to maximize the correction value. For model 2 the real correction value is definitely lower (see argument immediately below formula (2.5.b)). For model 1 the situation is somewhat more complicated. By the same argument as for model 2, the correction term is smaller than the Bartlett correction approximation on the form (4.6). This approximation will, however, give an over-estimate of the full correction term. Perhaps the same relation holds between the Bartlett correction and the approximation.

This conservative attitude of the corrected test is a more severe objection to the use of Bartlett corrected tests than the theoretical results derived from Jensen (1986) suggested. In fact, for many of the samples the corrected test is not more accurate than the uncorrected one.

D. It looks as if the rejection percentages increases with the complexity in the model, though this has to be examined further.

E. A comparison between $\hat{\sigma}^2_w$ and $\hat{\lambda}$ yields that $\hat{\lambda}$ is the greater 11 times out of 19. This could mean that the fully mean value corrected likelihood ratio test will have a variance lower than that of the $x^2$-distribution, and the corrected test would become too conservative.
F. In every single case where we compared (i) the Bartlett correction with the parameters replaced by their ML-estimates with (ii) the real correction values, the former gave equal or lower rejection percentages, and lower LR-averages and -deviations as well. It seems as if this substitution results in an even more strongly conservative behaviour of the test.

G. The fact that we correct the likelihood ratio test with something stochastic will change the rejection region, and will actually create a new test with properties similar to that of the likelihood ratio test, but at least slightly different. It may be enlightning to study the differences in these two rejection regions, and to test the ability of the two statistics to detect various types of deviations from the hypothesis.

7.6 Comments on the direct use of Lawley's formula.

The Bartlett correction term as it was derived by Lawley (1956) can be computed exactly whenever the mean values in his formula can be calculated analytically. This is possible for example in cases where we study transitions out of a transient state. Lawley expressed the Bartlett correction as a difference between two sums which are both written on the form

\[ \sum \lambda^{|S|} \sum_{rstu} + \sum \lambda^{|S|} \sum_{rstuvw} \left( \sum_{j} (j)^{(1)} (j)^{(1)} (j)^{(1)} (j)^{(1)} \right) \]

where the indices run over the set of parameters, the matrix \([\lambda^{|S|}]\) is the inverted information matrix, and \(i, g, j, k\) are functions of means of the likelihood derivatives, and derivatives of these mean values. The formula (7.1) suggests that the number of operations needed to complete the computation can be very large. In a situation with 15 parameters in the \(H_0\)-model and a test about an additional 10 parameters, the computer will have to handle more than \(5.5 \cdot 10^9\) operations if we compute (7.1) just by term by term addition and multiplication. A study of the likelihood derivatives (see Klingberg, 1987, Chapter 6) will yield that many of the contributions to the sums are zero. Thus, if we take this into account in our program, the actual number of operations will be considerably less.
A problem in which it seems harder to reduce the number of operations is that of computing the functions \( f, g, h, j, \) and \( k \). For a data set with 10,000 episodes, five main factors with numbers of levels \((3, 3, 3, 5, 7)\), and a test concerning an interaction with 15 levels, at least 38,000 functions has to be computed in advance. Each episode will provide a contribution to approximately 790 of those. Each contribution means an extra \( 4 \times 10 \) operations. We end up with a minimum of \( 3 \times 10^7 \) operations just to compute the expected values of the likelihood derivatives (and the derivatives of the expected values). The total number of operations for our program Raby that computes the simplified correction term was less than \( 1.3 \times 10^6 \), and that program needed about 10 CPU-seconds to compute a correction term for models much more simpler than those in this comparison.

The computation of the exact Bartlett correction term also involves the problem of inverting two information matrices. Those two matrices have similar structures; maybe there is a way to compute them simultaneously.

Thus, although it is rather complicated to find out exactly how many operations a program that computes Bartlett corrections has to make, it seems as a program based on Lawley’s formula for intensity regression models will work rather slowly. A study of the structure of the correction term for the intensity regression model in particular may reduce the number further, but the number of operations will all the same be many times the number needed to compute the simplified correction. Simulation studies to test the properties of the method will always be difficult to perform.

8. Degrees of freedom in the likelihood ratio test in cases with few occurrences.

In this section we discuss the contribution to the likelihood ratio test from a cell with no occurrences. Given no occurrences in a cell, the impact it has on the likelihood ratio is through its total exposure time only. The estimate of certain intensities will decrease if we add the observations from an almost empty cell to a data set for which we have already estimated the parameters and computed the likelihood ratio. For a
model with no nuisance parameters, the only change will be a decrease in 
the estimate of the baseline intensity under $H_0$. If the exposure time is 
very small in the additional cell, the likelihood ratio value will barely 
change at all, and it may be better to exclude the observations in the 
cell from the study. We now intend to look closer at such a strategy.

Assume an intensity regression model and a hypothesis $H_0$ such that 
for the likelihood ratio we have 
\[ w \sim x^2(k). \]
For the cells corresponding to the levels of the test factor, we assume 
that all except one cell contain such large sets of data that 0-occurrences are practically impossible. Suppose that the total exposure time is 
much smaller in the remaining cell, $R$, and that this may induce zero 
ocurrences there. We obtain an alternative to the standard likelihood ratio test if we treat $w$ as if 
\[ w \sim x^2(k-1) \]
whenever there are zero occurrences in cell $R$.

This will of course result in more rejections than for the original 
likelihood ratio test. The number of additional rejections depends on the 
distribution of the number of occurrences in cell $R$. If 0 occurrences is 
something slightly unusual, the decreasing rejection level may put many of 
the realizations that did not result in rejections before above the rejection level. These new rejections will have one thing in common. They all 
are associated with realizations in which we had 0-occurrences. Thus, the 
test will react stronger on deviations from the baseline intensity for 
this level than before. The new test may therefore be slightly skewed.

We may use model 2d of Section 7.4 as an example. The original like-
lihood ratio test rejected 7.3% of the realizations. A decrease of the 
degrees of freedom in the test whenever we do not observe any occurrences 
will increase the rejection percentage to 10.2. In this case, zero occur-
rences was something slightly unexpected since we observe the individuals during such a long time (C=90). Here, two individuals are enough content for the cell not to be excluded from the analysis. For a lower censoring time we might have been better off if we had used the lower d.f. number.

The expected likelihood ratio value is small in cases with very few observations. For a single realization of a variable from a censored exponential distribution with intensity \( \theta \) and censoring time \( C \) we have, when we test if \( \theta=\theta_0 \)

\[
w = N \cdot \log \frac{e_0^C}{T} - N - (N \cdot \log e_0 - T e_0),
\]

where \( N \) is the number of occurrences, and \( T \) is the exposure time. The variable \( N \) is either 0 or 1. We may write \( Ew \) as

\[
Ew = \theta_0 e^{e_0 C} \int_0^C \left( \log - 1 - \log e_0 + te_0 \right) e^{e_0 t} dt
\]

under \( H_0 \). Computing the expected value \( Ew \) yields

\[
Ew = \theta_0 \int_0^C \left( \log - 1 - \log e_0 + te_0 \right) e^{e_0 t} dt.
\]

Thus, we see that

\[
\theta_0 e^{e_0 C} (C \log C - C) - \log \theta_0 (1 - e^{e_0 C}) \leq \theta_0 e^{e_0 C} (C \log C - C)
\]

\[
\leq \theta_0 e^{e_0 C} (C \log C - C) - \log \theta_0 (1 - e^{e_0 C}).
\]

The mean value \( Ew \) is obviously small and stable for small \( C \). This suggests that \( Ew \) is less than the corresponding degrees of freedom for very sparse data, which was not our conclusion about the models in Section 7.1.

The following calculations are supposed to illustrate the degree of impact an additional cell with sparse data has on the likelihood ratio. Assume that we have a one factor model with \( k \) levels, and intensity \( \theta_i \) for level \( i \). We test
We now add observations from a level (k+1) to our study, and make the corresponding correction of $H_0$. The total exposure time of our new observations is very small, and we are interested in the impact on the LR-statistic from level (k+1). The new contribution can be written as

$$w_\delta = 2[k \cdot \log_\frac{N}{c} + N \cdot \log_\frac{c}{T} - (N+k) \cdot \log_\frac{N+k}{T+c}]$$

To clarify this formula, we write

$$w_\delta = 2[k \cdot \log_\frac{N}{c} - k \cdot \log_\frac{c}{T} - (N+k) \cdot (\log(1+c) - \log(1+\frac{c}{N}))],$$

and then replace the $\log(1+c)$ by $c$, which is a good approximation when $c$ is small, as in our case with a small contribution from the additional cell. We obtain

$$w_\delta \approx 2[k \cdot \log_\frac{N}{c} - k \cdot \log_\frac{c}{T} - (N+k) \cdot (\frac{k}{N} + \frac{c}{T})].$$

When $k=0$, we have

$$w_\delta \approx 2^{-c} \frac{N}{T}.$$
Acknowledgements.

Discussions with Britta and Jan M. Hoem have been valuable for my work with this paper. I also wish to thank Britta for her program which simulated life histories, and Jan for extensive editorial advice. This paper has been written at The Center for Demography and Ecology (CDE), University of Wisconsin, Madison. I am grateful to CDE for letting me use their computer facilities for my research. Finally, I would like to acknowledge the National Institute of Child Health and Human Development Center Grant No. HD 05876, and the grant from the William and Flora Hewlett Foundation for research and training.

References.