UNIVERSITÀ DEGLI STUDI DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Statistiche
SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE STATISTICHE
CICLO XXV

Combined composite likelihoods

Direttore della Scuola: Prof. Monica Chiogna

Supervisore: Prof. Nicola Sartori

Co-supervisore: Prof. Alessandra Salvan

Dottorando: Euloge Clovis Kenne Pagui

31/01/2013

Abstract

Composite likelihood is a particular pseudo-likelihood built by adequately combining likelihoods based on lower dimensional events. It appears to be a very appealing alternative to the standard likelihood when the latter is too time-consuming to evaluate or unavailable due to a complex, and possibly unknown, structure of dependence in the data. After the brief introduction in the first chapter, Chapter 2 gives notation and basic definitions, but also states a condition for full efficiency of the maximum composite likelihood estimator in exponential families.

The core of the thesis is Chapter 3, where we explore a linear combination of two types of composite likelihood which leads to a new objective function that depends on a constant to be chosen. In particular, this new combined composite likelihood uses both bivariate margins and univariate margins. Exact and asymptotic properties are explored. The exact properties lead to the identification of a possible strategy for finding the range of admissible values for the constant. The resulting estimator enjoys desirable asymptotic properties such as consistency and asymptotic normality. Two examples are analyzed in details, also through simulation studies.

Chapter 4 studies a weighted independence likelihood in a prediction framework. The aim of this chapter is to determine the weights in order to get an improved prediction of a component of interest of the data vector. In particular, the weights are calculated by means of a delete-one approach in a cross-validation procedure. Through simulation studies, situations in which the weighted independence likelihood works well with respect to the standard independence likelihood are highlighted.

Sommario

La verosimiglianza composita è una pseudo-verosimiglianza particolare costruita combinando adeguatamente validi oggetti di verosimiglianza relativi a piccoli sottoinsiemi di dati. Essa appare essere un'attraente alternativa alla verosimiglianza completa quando la sua computazione richiede troppo tempo o quando non può essere trattata a causa della complessa struttura di dipendenza nei dati. Dopo la breve introduzione contenuta nel primo capitolo, verrà introdotta nel secondo capitolo una condizione per la piena efficienza dello stimatore di massima verosimiglianza composita nelle famiglie esponenziali.

Il nucleo della tesi è presentato nel terzo capitolo ed esplora la combinazione lineare di due tipi di verosimiglianza composita in una nuova funzione obiettiva mediante una costante da scegliere. Il primo tipo si basa solo sulle marginali bivariate mentre il secondo sulle marginali univariate. Vengono esplorate sia le proprietà esatte che le proprietà asintotiche. Le proprietà esatte conducono all'identificazione di una possibile strategia per trovare l'intervallo di valori ammissibili per la costante. Lo stimatore risultante gode di desiderabili proprietà asintotiche, come la consistenza e la normalità asintotica. Due esempi sono analizzati nel dettaglio, anche mediante studi di simulazione.

Il quarto capitolo studia verosimiglianze di indipendenza pesate in un contesto di previsione. L'obiettivo è quello di determinare i pesi per ottenere una migliore previsione di un componente di interesse del vettore di dati. Viene considerata una procedura basata su cross-validation per affrontare l'argomento e, attraverso studi di simulazione, vengono evidenziate le situazioni in cui la verosimiglianza di indipendenza pesata funziona meglio rispetto alla verosimiglianza di indipendenza.

Acknowledgements

I would foremost like to thank my supervisors Professor Nicola Sartori and Professor Alessandra Salvan for their support and guidance, for their invaluable advice and especially for their patience and encouragement. Without their help, this thesis would not have been completed. I have learned to face with many statistical problems thanks to their expertise, and this will have a strong impact in my future path.

I would particularly like to thank my uncle Maurice Wamba and my family for all they have done to help me continue my university studies in Italy. Without them, I would have never undertaken a PhD program in Statistics.

I spent some months as visiting PhD student at the university of Toronto. I would like to express my deep gratitude to Professor Nancy Reid for her comments and suggestions and for all of the useful discussions we had.

I thank Professor Bruce Lindsay for his helpful suggestion about a step of a proof in Chapter 2.

I would infinitely like to thank my fiancé Marta Salusti for her encouragements, sweetness and love mainly during tough times of the PhD period.

Many thanks to my best friend Raoul Passo for our friendship, encouragements and support on everything I undertake. Lastly, I thank my friends who I met in Padova and other cities of Italy, Maria Salusti, Wolegha, Tazitouo, Fokou, Nguenang, Nono, Ngonkeu, Temfack, Mbah and Anaclet. All of them, before or during the PhD period, we shared tough and beautiful time.

Contents

Chapter 1

Introduction

The likelihood function was introduced by Fisher (1921, 1922) and, since then, it plays a crucial role in several approaches to statistics, mainly due to the fact that it provides inferential procedures with a number of desirable properties. However, in many statistical problems, the standard likelihood may not be a practical solution, either because of computational burdens or for the inability of specifying the whole joint distribution of the data. An alternative inferential tool with properties similar to those of a proper likelihood is the composite likelihood (Lindsay, 1988).

1.1 Overview

The composite likelihood (Lindsay, 1988) is a particular pseudo-likelihood which may validly replace the standard likelihood when the density involved in the latter is difficult to specify or computationally intractable; see Pace & Salvan (1997, chap. 4) and Molenberghs & Verbeke (2005, chap. 9) for pseudo-likelihoods. The idea behind the composite likelihood is to construct a new objective function by adequately compounding likelihoods based on appropriate events in the sample space.

In recent years, composite likelihood methods have received increasing interest in both theoretical and applied field. Recent advances in the area of composite likelihood have been presented at dedicated workshops held in Warwick (April, 2008) and Banff (April, 2012).

There have been many variations in the composite likelihood formulation to balance the trade-off between efficiency and computational cost. Even the pseudo-likelihood proposed by Besag (1974) for approximate inference in spatial processes is now recognized as a composite conditional likelihood. This pseudo-likelihood is the product of conditional densities of a single observation given its neighbours. Cox & Reid (2004) investigated composite likelihoods constructed from lower dimensional marginal densities, called composite marginal likelihoods. The simplest version, called independence likelihood (Chandler & Bate, 2007) is constructed under a working independence assumption. This pseudolikelihood could be useful when inference is about marginal parameters only. Another example is the pairwise likelihood (Le Cessie & Van Houwelingen, 1994) based on pairs of observations. This pseudo-likelihood could be useful when the parameters related to the correlations are of interest. In Hjort & Varin (2008), the composite likelihood is constructed by compounding likelihoods based on triplets of observations in the context of Markov chain models. See the recent review papers by Varin (2008) and Varin et al. (2011) for more examples and applications of the composite likelihood.

Under mild regularity conditions (Molenberghs & Verbeke, 2005, Chap. 9), the composite score yields an unbiased estimating function, leading to the result that the composite maximum likelihood estimator is consistent and asymptotically normally distributed, with variance given by the inverse of the Godambe information matrix (Godambe, 1960). The composite likelihood ratio statistic has the drawback of a non-standard asymptotic distribution and, for this, adjustments have been proposed in order to recover the standard asymptotic chi-squared distribution (Chandler & Bate, 2007; Pace et al., 2011).

1.2 Main contributions of the thesis

The main research objective of this thesis focuses on the exploration of new forms of composite likelihood. The first one follows a suggestion in Cox & Reid (2004). Precisely, this new objective function is given by a linear combination of independence and pairwise likelihoods through a constant to be chosen, possibly in an optimal way. Particular values of the constant lead to notable composite likelihoods, such as pairwise marginal and conditional likelihoods. This new type of composite likelihood is called in the thesis combined composite likelihood, and particular attention has been paid to the development of its exact and asymptotic properties. Two examples dealing with combined composite likelihood are considered, both dealing with instances of the multivariate normal distribution with structured covariance matrix. In the first example, the parameter is scalar, whereas it is a vector in the second example. In particular, exploiting the exact properties of the combined composite likelihood, we found a condition on the admissible values for the constant for which the combined composite likelihood satisfies the necessary requirements of being a sensible pseudo-likelihood in both the scalar and the multidimensional parameter case. Moreover, for the asymptotic properties, we showed that the combined composite likelihood estimator is still consistent and asymptotically normal and that consistency is not generally guaranteed when the sample size is fixed as the random vector's length goes to infinity. In addition, we also showed that the combined composite likelihood ratio statistic still maintains the drawback of a non-standard asymptotic distribution.

A well-known example of the composite marginal likelihood is the independence likelihood which is constructed by using only the univariate marginal densities, under the working assumption of independence. In certain contexts, it could be appropriate to give different weights to the univariate marginal densities obtaining a weighted independence likelihood. The second research objective is to determine the weights in order to have a good prediction of a component of interest of the data vector, considering the remaining components as auxiliary variables.

In Chapter 2, the concept and the main results related to the likelihood approach are reviewed. The brief introduction to the theory of likelihood is useful for introducing the composite likelihood's idea, which is presented together with its definition and main properties. This chapter also considers conditions for full efficiency of the maximum composite likelihood estimator in exponential families.

Chapter 3 explores in detail the combined composite likelihood approach. Motivation arises in contexts in which there is information on the parameter of interest, both in one and two-dimensional marginal densities. Therefore, instead of using either the independence or the pairwise likelihood for inference about the parameter of interest, we combine them obtaining the combined composite likelihood. We give its formal definition and study its exact and asymptotic properties.

Two examples of the combined composite likelihood, both based on the multivariate normal distributions are considered in detail. We focus on efficiency of the combined composite likelihood estimator by comparing it with the standard maximum likelihood estimator. Both examples seem to suggest the pairwise conditional likelihood, which is a particular case of the combined composite likelihood, as a close to optimal choice.

In Chapter 4 we deal with prediction using the weighted independence likelihood. To determine the weights, we use the delete-one approach in the cross-validation procedure as done by Wang & Zidek (2005) in a weighted likelihood framework. Two illustrative examples are considered together with different simulation scenarios. Situations in which the weighted independence likelihood works well with respect to the standard independence likelihood are highlighted.

Chapter 2

Background on composite likelihood

2.1 Introduction

Composite likelihoods are pseudo-likelihoods built by pooling likelihood components, with each component based on appropriate events in the sample space, such as marginal or conditional events.

The motivation for the use of the composite likelihood as a surrogate of the standard likelihood is two-fold: to reduce the computational complexity to cope with large data set and/or models involving complex interdependencies and to make inference about parameters of interest without making assumptions on the whole joint distribution of the data.

Here we focus mainly on inferential aspects and properties of the composite likelihood. In addition, we introduce the most commonly used versions of composite likelihood which are the composite marginal likelihood and the composite conditional likelihood. Before exploring the methods of inference based on composite likelihoods, we recall some basic definitions and properties of the standard likelihood for regular models.

The next section is devoted to some basic results of inference based on the likelihood function. This theory is helpful for better understanding the properties of the composite likelihood. Section 3 is devoted to the definition and properties of the composite likelihood. In Section 4, we define some quantities related to the composite likelihood. At last, we introduce different efficiency measures of the maximum composite likelihood estimator compared to the one based on the standard likelihood, in situations in which the latter is available. In Section 5, we give conditions for which the maximum composite likelihood estimator is fully efficient in exponential families.

2.2 Likelihood, related quantities and asymptotic properties

Let

$$\mathcal{F} = \{f(y;\theta) : \theta \in \Theta \subseteq \mathbb{R}^d, y \in \mathcal{Y} \subseteq \mathbb{R}^q\} \tag{2.1}$$

be a parametric model, where $f(y;\theta)$ is the probability density function for a random variable Y and θ the parameter of the model. Given n independent observations, $y_1,\dots,y_n$, the function

$$L(\theta) = \prod_{i=1}^{n} f(y_i;\theta), \tag{2.2}$$

considered as a function of θ, is called the likelihood function. In practice, it is often more convenient to work with the logarithm of the likelihood function, called the log-likelihood whose expression is

$$\ell(\theta) = \sum_{i=1}^{n} \log f(y_i;\theta). \tag{2.3}$$

The maximum likelihood estimator, denoted by $\hat{\theta}$, is the value of θ which maximizes $L(\theta)$ or $\ell(\theta)$, that is,

$$\hat{\theta} = \arg\max_{\theta\in\Theta} L(\theta) = \arg\max_{\theta\in\Theta} \ell(\theta). \tag{2.4}$$

The *score* vector, also called Fisher's *score function*, is the gradient of the log-likelihood function, and is denoted by

$$U(\theta) = \frac{\partial \ell(\theta)}{\partial \theta}. \tag{2.5}$$

In regular models (Severini, 2000, §3.4) the maximum likelihood estimator can be found as a solution of the likelihood equation

$$U(\theta) = 0.$$

The Hessian matrix of $\ell(\theta)$, i.e. the matrix of second derivatives of the log-likelihood, is

$$h(\theta) = \frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta^{\mathrm{T}}} = \begin{pmatrix} \frac{\partial^2 \ell(\theta)}{\partial \theta_1^2} & \cdots & \frac{\partial^2 \ell(\theta)}{\partial \theta_1 \partial \theta_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \ell(\theta)}{\partial \theta_d \partial \theta_1} & \cdots & \frac{\partial^2 \ell(\theta)}{\partial \theta_d^2} \end{pmatrix}.$$

The observed information matrix is given by $j(\theta) = -h(\theta)$ and the Fisher information matrix is defined as

$$i(\theta) = \mathrm{Var}_\theta \left\{ U(\theta) \right\} = \mathrm{E}_\theta \left\{ U(\theta) U(\theta)^{\mathrm{T}} \right\}, \tag{2.6}$$

since $\mathrm{E}_\theta \left\{ U(\theta) \right\} = 0$. The Fisher information matrix can also be calculated as the expectation of the observed information matrix, denoted by

$$i(\theta) = \mathrm{E}_\theta \left\{ j(\theta) \right\} = \mathrm{E}_\theta \left\{ U(\theta) U(\theta)^{\mathrm{T}} \right\}. \tag{2.7}$$

Equation (2.7) is known as the second Bartlett identity, while $\mathrm{E}_\theta \left\{ U(\theta) \right\} = 0$ is the first Bartlett identity.

Under suitable regularity conditions, the maximum likelihood estimator $\hat{\theta}$ of θ has the following asymptotic properties

- $\hat{\theta} \xrightarrow{p} \theta$, as $n \longrightarrow \infty$, i.e. $\hat{\theta}$ is a consistent estimator of θ;
- $\hat{\theta} \xrightarrow{d} N_d(\theta, i(\theta)^{-1})$, i.e. $\hat{\theta}$ is asymptotically normally distributed.

Moreover, the quantities

$$W_u(\theta) = U(\theta)^{\mathrm{T}} i(\theta)^{-1} U(\theta)$$
$$W_e(\theta) = (\hat{\theta} - \theta)^{\mathrm{T}} i(\theta)(\hat{\theta} - \theta)$$
$$W(\theta) = 2\{\ell(\hat{\theta}) - \ell(\theta)\}$$

can be used to test $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ and are known as the *score* test, *Wald* test and likelihood ratio test respectively. They are asymptotically distributed as χ^2_d under H_0. The same statistics could be used also for the construction of confidence regions.

2.3 Composite likelihood: definition and properties

Let us consider the parametric statistical model (2.1). Given an observation $y = (y_1, \ldots, y_q)$, the composite likelihood is defined through a set of marginal or conditional events $\{\mathcal{A}_1(y), \ldots, \mathcal{A}_K(y)\}$, usually related to small subsets of the data, with component likelihoods given by $L_k(\theta; y) = L_k(\theta; \mathcal{A}_k(y))$. Therefore, following Lindsay (1988), the composite likelihood obtained by compounding these component likelihoods is defined as

$$cL(\theta; y) = \prod_{k=1}^{K} L_k(\theta; y)^{w_k},$$

where $\{w_1 \ldots, w_K\}$ is a set of non-negative weights. The associated composite log-likelihood is $c\ell(\theta; y) = \sum_{k=1}^{K} w_k \ell_k(\theta; y)$ with $\ell_k(\theta; y) = \log cL(\theta; y)$ and its maximizer is defined as $\hat{\theta}_C = \arg\max_\theta c\ell(\theta; y)$. Under random sampling of size n, the composite log-likelihood becomes $c\ell(\theta) = \sum_{i=1}^{n} c\ell(\theta; y_i)$.

Two important instances of composite likelihood are the composite marginal and the composite conditional likelihood. The composite marginal likelihood (Cox & Reid, 2004) is constructed from low dimensional marginal densities. Two important examples belong to this class: the independence likelihood (Chandler & Bate, 2007) which is constructed by using only the univariate marginal densities, under the working assumption of independence,

$$cL^{\mathrm{I}}(\theta; y) = \prod_{r=1}^{q} f(y_r; \theta)^{w_r}, \;\; r = 1, \ldots, q,$$

and the pairwise likelihood (Le Cessie & Van Houwelingen, 1994), built by using only bivariate marginal densities

$$cL^{\mathrm{P}}(\theta; y) = \prod_{r=1}^{q-1} \prod_{s=r+1}^{q} f(y_r, y_s; \theta)^{w_{rs}}, \;\; r, s = 1, \ldots, q.$$

The second notable instance is the composite conditional likelihood. Here the pseudo-likelihood is obtained by combining only low dimensional conditional densities. Two examples are the full conditional likelihood, whose expression is

$$cL^{\mathrm{FC}}(\theta; y) = \prod_{r=1}^{q} f(y_r \mid y_{(-r)}; \theta)^{w_r},$$

where $y_{(-r)}$ denotes the vector of all the observations with y_r deleted, and the pairwise conditional likelihood which is defined as

$$cL^{\mathrm{PC}}(\theta; y) = \prod_{r=1}^{q} \prod_{s \neq r}^{q} f(y_r \mid y_s; \theta)^{w_{rs}}.$$

See Molenberghs & Verbeke (2005) or Mardia et al. (2008) for the example of the full and pairwise conditional likelihood.

Inference based on the composite likelihood may be justified by the fact that the true value of the parameter, θ_0, is the maximizer of the expected value of the composite log likelihood. That is,

$$\mathrm{E}_{\theta_0}[c\ell(\theta; Y)] < \mathrm{E}_{\theta_0}[c\ell(\theta_0; Y)],$$

for any $\theta \neq \theta_0$. In fact,

$$\begin{aligned}
\mathrm{E}_{\theta_0}[\sum_{k=1}^{K} w_k \log L_k(\theta;Y)] &= \sum_{k=1}^{K} w_k \mathrm{E}_{\theta_0}[\log L_k(\theta;Y)] \\
&= \sum_{k=1}^{K} w_k \mathrm{E}_{\theta_0}[\log L(\theta;A_k(Y))] \\
&< \sum_{k=1}^{K} w_k \mathrm{E}_{\theta_0}[\log L(\theta_0;A_k(Y))],
\end{aligned}$$

for any $\theta \neq \theta^0$ according to the Kullback-Leibler information inequality which holds for any single term in the summation.

As for a genuine likelihood, for any composite log-likelihood, the following exact properties would be desirable:

(a) $\mathrm{E}_{\theta_0}[c\ell(\theta;Y)] \leq \mathrm{E}_{\theta_0}[c\ell(\theta_0;Y)]$ for any $\theta \neq \theta_0$;

(b) $\mathrm{E}_{\theta_0}\left[\frac{\partial}{\partial\theta} c\ell(\theta;Y)\right]\Big|_{\theta=\theta_0} = 0$;

(c) the matrix $\mathrm{E}_{\theta_0}\left[\frac{\partial^2}{\partial\theta\partial\theta^{\mathrm{T}}} c\ell(\theta;Y)\right]\Big|_{\theta=\theta_0}$ is negative definite.

Property (a), is the Kullback-Leibler information, or Wald inequality, which is a key property for ensuring that the maximizer of the composite log likelihood gives a consistent estimator. Property (b), ensures that the composite score satisfies the requirement of being an unbiased estimating function. Finally, the last one ensures the expected Hessian of $c\ell(\theta)$ is negative definite at θ_0 and as a consequence, $c\ell(\theta)$ is on average locally maximized at θ_0.

Composite likelihood methods enjoy many other good properties. Being the composition of low dimensional marginal or conditional distributions, they are easier to evaluate and to maximize. Although a loss of efficiency is expected when one uses a pseudo-likelihood in place of the standard likelihood, the use of the composite marginal likelihood provides an important computational gain. In very few cases the composite likelihood estimator is identical to the full likelihood estimator, and thus fully efficient. This situation is examined by Mardia et al. (2009) who, show that the composite marginal and conditional likelihood estimators are full efficient and identical to the full maximum likelihood estimators in exponential families under a certain closure property. In Section 2.5 we will study this aspects in more detail.

Composite likelihood inference procedures are in general considered as robust, since they only require the specification of lower dimensional conditional or marginal densities . There could be however different types of robustness, such as robustness of consistency, studied in detail by Xu & Reid (2011), and computational robustness (Varin et al., 2011).

2.4 Composite likelihood quantities

The composite score function is defined as the first derivative of the composite log-likelihood. Its expression is

$$cU(\theta; y) = \frac{\partial}{\partial\theta} c\ell(\theta; y) = \sum_{k=1}^{K} w_k \frac{\partial}{\partial\theta} \ell_k(\theta; y).$$

Under random sampling of size n, it becomes

$$cU(\theta) = \sum_{i=1}^{n} cU(\theta; y_i).$$

The composite score function is unbiased, because it is a linear combination of score functions related to proper likelihood, that is $\mathrm{E}_\theta\{cU(\theta; Y\} = 0$. Under the usual regularity conditions (Molenberghs & Verbeke, 2005, chap. 9), the maximum composite likelihood estimator is consistent and asymptotically normal as $n \longrightarrow \infty$,

$$\hat{\theta}_C \dot{\sim} N(\theta, G(\theta)^{-1}) \, ,$$

where $G(\theta) = H(\theta)J(\theta)^{-1}H(\theta)$ is known as the Godambe information or sandwich information, $H(\theta) = \mathrm{E}_\theta\{-\partial cU(\theta)/\partial\theta\}$ is the sensitivity matrix and $J(\theta) = \mathrm{Var}_\theta\{cU(\theta)\}$ is the variability matrix.

The use of the composite likelihood in place of the standard likelihood could lead in general to a loss of efficiency. The composite likelihood can refer to the theory of misspecified models, being constructed under the working assumption of independence among the component likelihoods. As a consequence the composite likelihood does not satisfy the second Bartlett identity, since $H(\theta) \neq J(\theta)$. In Section 2.5 we will show an example in which $H(\theta) \neq J(\theta)$ but we however reach full efficiency, since $i(\theta) = G(\theta)$.

Suppose we are interested in testing the null hypothesis

$$H_0 : \theta = \theta_0,$$

where θ is a d-dimensional vector. As in the standard likelihood setting, we may use one of the following test statistics related to the composite likelihood,

$$\begin{aligned} W_u^{\mathrm{C}}(\theta) &= cU(\theta)^{\mathrm{T}} J(\theta)^{-1} cU(\theta) \\ W_e^{\mathrm{C}}(\theta) &= (\hat{\theta}_C - \theta)^{\mathrm{T}} G(\theta)(\hat{\theta}_C - \theta) \\ W^{\mathrm{C}}(\theta) &= 2\{c\ell(\hat{\theta}_C) - c\ell(\theta)\}. \end{aligned}$$

The Wald (W_e^{C}) and score (W_u^{C}) statistics based on the composite likelihood have the usual χ^2_d distribution. Even in this case, W_e^{C} is not invariant to reparametrization, while W_u^{C} is seen in many examples to be numerically unstable. The composite likelihood ratio statistic may be preferable despite the fact that it has a non-standard asymptotic distribution. Indeed, its asymptotic distribution is a linear combination of independent χ^2_1 distributions,

$$W^{\mathrm{C}}(\theta) \xrightarrow{d} \sum_{i=1}^{d} \lambda_i Z_i^2,$$

where $Z_i^2, i = 1, \ldots, d$ are independent χ^2_1 variables and $\lambda_1(\theta), \ldots, \lambda_d(\theta)$ are the eigenvalues of the matrix $J(\theta)^{-1} H(\theta)$. Adjusted versions of $W^{\mathrm{C}}(\theta)$ can recover the usual χ^2_d asymptotic distribution (Chandler & Bate, 2007; Pace et al., 2011).

Since the composite likelihood could be seen as the likelihood of a misspecified model, the issue here is to know how to measure the loss of efficiency given by its use in place of the standard likelihood. To this end, one way is to compare in terms of efficiency the estimator based on full likelihood with the one based on the composite likelihood. This comparison, in the scalar case, is based on the Asymptotic Relative Efficiency, ARE, which is the ratio between the asymptotic variances of the two estimators. In particular, the expression is given by

$$ARE(\hat{\theta}_C) \doteq \frac{\mathrm{Var}_\theta(\hat{\theta})}{\mathrm{Var}_\theta(\hat{\theta}_C)} = \frac{G(\theta)}{i(\theta)}. \tag{2.8}$$

When the parameter is multidimensional, an overall measure of efficiency (Davison, 2003, p. 113), which is however quite difficult to interpret, can be summarized by

$$ARE(\hat{\theta}_C) = \left(\frac{|G(\theta)|}{|i(\theta)|}\right)^{\frac{1}{d}}. \tag{2.9}$$

On the other hand, if the interest is focused on single r-th component of θ, we may use the appropriate measure (Davison, 2003, p. 113) given by

$$ARE(\hat{\theta}_{C_r}) = \frac{[i(\theta)^{-1}]_{rr}}{[G(\theta)^{-1}]_{rr}}, \tag{2.10}$$

where for instance $[i(\theta)^{-1}]_{rr}$ is the (r,r)-th of the inverse matrix $i(\theta)^{-1}$.

2.5 Full efficiency in exponential families

In some models, the estimator based on a composite likelihood is identical to the maximum likelihood estimator. Mardia et al. (2009) provide an explanation for this, by showing that such identity holds for exponential families that have a certain closure property. However, there are some models for which the maximum composite likelihood estimator is still fully efficient and do not fall into this class. We hereafter propose a new sufficient condition which also includes some of those models.

Theorem 1. Let us consider the parametric statistical model defined in (2.1). If $\ell(\theta) = \theta^{\mathrm{T}} t(y) - K(\theta)$ and $c\ell(\theta) = \psi(\theta)^{\mathrm{T}} t(y) - V(\theta)$ are of canonical exponential family type, with the same sufficient statistic $t(y)$, then

1. $cU(\theta; y) = \psi_\theta(\theta) U(\theta; y)$ where, $\psi_\theta(\theta) = \frac{\partial}{\partial\theta}\psi(\theta)$;
2. $\hat{\theta}_C = \hat{\theta}$;
3. $G(\theta) = i(\theta)$, even though $J(\theta) \neq H(\theta)$.

Proof. The score based on $\ell(\theta)$ becomes

$$U(\theta) = t(y) - K_\theta(\theta)$$

where, $K_\theta(\theta) = \frac{\partial}{\partial\theta} K(\theta)$. While the score based on $c\ell(\theta)$ is

$$cU(\theta) = \psi_\theta(\theta) t(y) - V_\theta(\theta) \tag{2.11}$$

where, $\psi_\theta(\theta) = \frac{\partial}{\partial\theta}\psi(\theta)$ and $V_\theta(\theta) = \frac{\partial}{\partial\theta} V(\theta)$. Since $\mathrm{E}_\theta\left\{U(\theta)\right\} = 0$ and $\mathrm{E}_\theta\left\{cU(\theta)\right\} = 0$, we have that

$$\mathrm{E}_\theta\{t(Y)\} = K_\theta(\theta) \tag{2.12}$$

and

$$\psi_\theta(\theta)\mathrm{E}_\theta\{t(Y)\} = V_\theta(\theta). \tag{2.13}$$

Putting together (2.12) and (2.13), we get

$$\psi_\theta(\theta) K_\theta(\theta) = V_\theta(\theta). \tag{2.14}$$

Then substituting (2.14) in (2.11) gives

$$\begin{aligned} cU(\theta) &= \psi_\theta(\theta) t(y) - \psi_\theta(\theta) K_\theta(\theta) \\ &= \psi_\theta(\theta)\left\{t(y) - K_\theta(\theta)\right\} \\ &= \psi_\theta(\theta) U(\theta). \end{aligned} \tag{2.15}$$

As we can see, the score based on pairwise likelihood is proportional to the one based on the full likelihood. As a result, the estimate based on the composite likelihood coincides with the estimate based on the full likelihood. Thus, the first two results are proved. As last step, we want to show that the Godambe information is identical to the Fisher information. To this end, we will use index notation (Pace & Salvan, 1997, chap. 9). In the following, indices r, s, a, b, with $r, s, a, b = 1, \dots, d$, are used to indicate the components of a vector. For example, θ^r will indicate the generic component of the vector $\theta = (\theta^1, \dots, \theta^d)$. Regarding a $d \times d$ matrix A, A_{rs} will indicate its generic element and A^{rs} the element of its inverse. For simplicity of notation, we will adopt the Einstein summation convention which says that when an index appears two or more times in a product of elements of arrays, then summation over the range of that index is understood. For example, if x and y are column vectors in $\mathbb{R}^d$, the scalar product can be expressed as

$$x \cdot y = x^r y^r = \sum_{r=1}^{d} x^r y^r.$$

Based on the new notations, the likelihood function can be rewritten as

$$\ell(\theta) = \theta^r t^r(y) - K(\theta).$$

Differentiating $\ell(\theta)$ with respect to θ^r, gives the generic element of the score function

$$U_r(\theta) = t^r(y) - K_r(\theta),$$

where $K_r(\theta) = \partial K(\theta)/\partial\theta^r$.

$$\mathrm{E}_\theta\left\{U_r(\theta)\right\} = 0 \Longleftrightarrow \mathrm{E}_\theta\left\{t^r(Y)\right\} = K_r(\theta). \tag{2.16}$$

The generic element of the Fisher information is given by

$$i_{rs}(\theta) = \mathrm{Var}_\theta\left\{U_r(\theta)\right\} = \mathrm{Var}_\theta\left\{t^r(Y)\right\} = \mathrm{E}_\theta\left\{-U_{rs}(\theta)\right\} = K_{rs}(\theta),$$

where $U_{rs}(\theta) = \partial\ell(\theta)/(\partial\theta^r\partial\theta^s)$ and $K_{rs}(\theta) = \partial K(\theta)/(\partial\theta^r\partial\theta^s)$. On the other hand, the composite log-likelihood can be rewritten as

$$c\ell(\theta) = \psi^a(\theta)t^a(y) - V(\theta),$$

and its differentiation with respect to θ^r, gives the generic element

$$cU_r(\theta) = t^a(y)\psi_r^a(\theta) - V_r(\theta), \tag{2.17}$$

where $\psi_r^a(\theta) = \partial\psi^a(\theta)/\partial\theta^r$ and $V_r(\theta) = \partial V(\theta)/\partial\theta^r$.

$$\mathrm{E}_\theta\left\{cU_r(\theta)\right\} = 0 \Longleftrightarrow \mathrm{E}_\theta\left\{t^a(Y)\right\}\psi_r^a(\theta) = V_r(\theta). \tag{2.18}$$

Then, substituting (2.16) in (2.18) gives

$$K_a(\theta)\psi_r^a(\theta) = V_r(\theta), \tag{2.19}$$

and (2.19) in (2.17) lead to

$$\begin{aligned} cU_r(\theta) &= t^a(y)\psi_r^a(\theta) - V_r(\theta) \\ &= t^a(y)\psi_r^a(\theta) - K_a(\theta)\psi_r^a(\theta) \\ &= \psi_r^a(\theta)\left[t^a(y) - K_a(\theta)\right] \\ &= \psi_r^a(\theta)U_a(\theta), \end{aligned}$$

which coincides with (2.15).

We now compute the necessary quantities for the Godambe information. The generic element of the Hessian matrix based on $c\ell(\theta)$ is

$$cU_{rs}(\theta) = \frac{\partial}{\partial\theta_s}cU_r(\theta) = \psi_{rs}^a(\theta)U_a(\theta) + \psi_r^a(\theta)U_{as}(\theta),$$

where, $\psi_{rs}^a(\theta) = \partial\psi^a(\theta)/(\partial\theta^r\partial\theta^s)$. The matrix $H(\theta)$ has elements

$$\begin{aligned} H_{rs}(\theta) &= \mathrm{E}_\theta\left\{-cU_{rs}(\theta)\right\} = -\psi_{rs}^a(\theta)\mathrm{E}_\theta\left\{U_a(\theta)\right\} + \psi_r^a(\theta)\mathrm{E}_\theta\left\{-U_{as}(\theta)\right\} \\ &= \psi_r^a(\theta)i_{as}(\theta). \end{aligned}$$

The variability matrix $J(\theta)$ has elements

$$\begin{aligned} J_{rs}(\theta) &= \mathrm{E}_\theta\left\{cU_r(\theta)cU_s(\theta)\right\} = \mathrm{E}_\theta\left\{\psi_r^a(\theta)U_a(\theta)\psi_s^b(\theta)U_b(\theta)\right\} \\ &= \psi_r^a(\theta)\psi_s^b(\theta)\mathrm{E}_\theta\left\{U_a(\theta)U_b(\theta)\right\} \\ &= \psi_r^a(\theta)\psi_s^b(\theta)i_{ab}(\theta). \end{aligned}$$

The inverse of $J(\theta)$ has elements

$$J^{rs}(\theta) = \Theta_a^s(\theta) i^{ab}(\theta) \Theta_b^r(\theta),$$

where $\Theta_a^s(\theta)\psi_r^a(\theta) = \delta_r^s$ and $\Theta_b^r(\theta)\psi_s^b(\theta) = \delta_s^r$. The symbol δ_r^s takes value 1 when $r = s$ and 0 otherwise. Therefore, the generic element of the Godambe information is

$$\begin{aligned} G_{rs}(\theta) &= H_{rs}(\theta) J^{rs}(\theta) H_{sr}(\theta) \\ &= \psi_r^a(\theta) i_{as}(\theta) \Theta_b^s(\theta) i^{bc}(\theta) \Theta_c^r(\theta) \psi_s^d(\theta) i_{dr}(\theta) \\ &= \delta_c^a i_{as}(\theta) i^{bc}(\theta) i_{dr}(\theta) \delta_b^a \\ &= i_{as}(\theta) i^{ba}(\theta) i_{br}(\theta) \\ &= i_{as}(\theta) \delta_r^a \\ &= i_{rs}(\theta) = i_{sr}(\theta). \end{aligned}$$

In conclusion, the Fisher information coincides with the Godambe information. □

Example 1 *One-way random effects*

This example is consider in Cox & Reid (2004). In this model, it is assumed that Y is q-dimensional multivariate normal with components having mean μ and variances σ^2. The correlation between any two components of the same vector is ρ with $-1/(q-1) < \rho < 1$, which is the necessary condition for the covariance matrix to be positive definite. The interest parameter is then three-dimensional, with $\theta = (\mu, \sigma^2, \rho)$.

An important application is the analysis of the variance with random effects, where it is usually assumed that $Y_{ir} = \mu + a_i + e_{ir}$, $r = 1, \dots, q, i = 1, \dots, n$, with effect a_i which has distribution $N(0, \sigma_a^2)$, and the error e_{ir}, with distribution $N(0, \sigma_e^2)$, and independent of a_i. The statistical model considered above, has $\sigma^2 = \sigma_a^2 + \sigma_e^2$ and $\rho = \sigma_a^2/(\sigma_a^2 + \sigma_e^2)$.

The full log-likelihood based on n independent observations is given by

$$
\begin{aligned}
\ell(\theta) = & -\frac{1}{2\sigma^2(1-\rho)} ss_W - \frac{q}{2\sigma^2\left\{1+\rho(q-1)\right\}}(ss_B + n\overline{y}^2) \\
& + \frac{nq\mu}{\sigma^2\left\{1+\rho(q-1)\right\}}\overline{y} - \frac{nq\mu^2}{2\sigma^2\left\{1+\rho(q-1)\right\}} - \frac{nq}{2}\log\sigma^2 \\
& - \frac{n(q-1)}{2}\log(1-\rho) - \frac{n}{2}\log\left\{1+\rho(q-1)\right\}
\end{aligned}
$$

and the pairwise log-likelihood is

$$
\begin{aligned}
c\ell^{\mathrm{P}}(\theta) = & -\frac{q-1+\rho}{2\sigma^2(1-\rho^2)} ss_W - \frac{q(q-1)}{2\sigma^2(1+\rho)}(ss_B + n\overline{y}^2) + \frac{nq(q-1)\mu}{\sigma^2(1+\rho)}\overline{y} \\
& - \frac{nq(q-1)\mu^2}{2\sigma^2(1+\rho)} - \frac{nq(q-1)}{2}\log\sigma^2 - \frac{nq(q-1)}{4}\log(1-\rho^2),
\end{aligned}
$$

where $ss_W = \sum_{i=1}^{n}\sum_{r=1}^{q}(y_{ir} - \overline{y}_i)^2$ e $ss_B = \sum_{i=1}^{n}(\overline{y}_i - \overline{y})^2$. As we can see, both full and pairwise log-likelihoods are of canonical exponential type with same sufficient statistic, $t(y) = (ss_W, ss_B, \bar{y})$. Therefore, the pairwise likelihood estimator is fully efficient. This result is also valid for pairwise conditional likelihood estimator. We note that this model is not a closed exponential family in the definition of Mardia et al. (2009).

When we moved outside full exponential families, typically, full efficiency can not be attained. For instance, let us consider the parametric statistical model defined in (2.1), with $\ell(\theta)$ and $c\ell(\theta)$ of curved exponential type. Even if they have the same sufficient statistic, full efficiency is not guaranteed.

In the following we start by showing that the estimator based on the composite likelihood may not coincide with the one based on the standard likelihood. This will be done using the geometry of exponential families. The joint density of the data can be written as

$$
p(y;\theta) = h(y)\exp\left[\phi(\theta)^{\mathrm{T}}t(y) - K(\theta)\right]
$$

with $q > d$. The natural parameter space Ω_θ, corresponds to the set of all θ such that the normalizing constant

$$
e^{K(\theta)} = \int h(y)e^{\phi(\theta)^{\mathrm{T}}t(y)}\mu(dy) < +\infty.
$$

Let us denote by $\eta = \eta(\theta) = E_\theta\{t(Y)\}$ the expectation parameter and by Ω_η the corresponding parameter space. The full log-likelihood is given by

$$\ell(\theta) = \phi(\theta)^{\mathrm{T}} t(y) - K(\theta),$$

and the composite log-likelihood is

$$c\ell(\theta) = \alpha(\theta)^{\mathrm{T}} t(y) - Q(\theta).$$

Following the terminology of Efron (1975), the curved exponential family $\mathcal{F}$ is represented by the curves

$$\mathcal{F}^{\ell}_{\Omega_\theta} = \{\phi(\theta) : \theta \in \Theta\}, \quad \mathcal{F}^{c}_{\Omega_\theta} = \{\alpha(\theta) : \theta \in \Theta\} \text{ in } \Omega_\theta$$
$$\mathcal{F}_{\Omega_\eta} = \{\eta = \eta(\theta) : \theta \in \Theta\} \text{ in } \Omega_\eta,$$

where, with the superscripts ℓ and c we refer to the full log-likelihood and the composite log-likelihood respectively. Note that $\mathcal{F}_{\Omega_\eta}$ is the same for $\ell(\theta)$ and $c\ell(\theta)$, because they both have the same sufficient statistic. The score based on the full log-likelihood is given by

$$U(\theta) = t(y)^{\mathrm{T}} \phi_\theta(\theta) - K_\theta(\theta),$$

where subscripts denote differentiation, for instance $K_\theta(\theta) = \partial K(\theta)/\partial\theta$. Since $\mathrm{E}_\theta\{U(\theta)\} = 0 = \mathrm{E}_\theta\{t(Y)\}^{\mathrm{T}}\phi_\theta(\theta) - K_\theta(\theta) \implies K_\theta(\theta) = \eta(\theta)^{\mathrm{T}}\phi_\theta(\theta)$. Therefore, $U(\theta) = \{t(y) - \eta(\theta)\}^{\mathrm{T}}\phi_\theta(\theta)$. The maximum full likelihood estimator, assuming it exists, satisfies

$$U(\hat\theta) = \{t(y) - \eta(\hat\theta)\}^{\mathrm{T}}\phi_\theta(\hat\theta) = 0, \tag{2.20}$$

From the latter equation, it follows that the set of $t(y)$ points having $\hat\theta$ as solution of (2.20) is the straight line orthogonal to $\phi_\theta(\hat\theta)$ and intersecting $\mathcal{F}_{\Omega_\eta}$ at $\eta(\hat\theta)$. In other words, the maximum likelihood estimator of $\eta(\hat\theta)$ is obtained by projecting the data point $t(y)$ onto $\mathcal{F}_{\Omega_\eta}$ orthogonally to $\phi_\theta(\hat\theta)$. We now denote such set of $t(y)$ points by

$$\mathcal{L}_{\hat\theta} = \{t(y) : U(\hat\theta) = 0\}.$$

Following the same steps, we get that the score based on the composite log-likelihood is

$$cU(\theta) = \{t(y) - \eta(\theta)\}^{\mathrm{T}}\alpha_\theta(\theta).$$

The maximum composite likelihood satisfies

$$cU(\hat{\theta}_C) = \{t(y) - \eta(\hat{\theta}_C)\}^{\mathrm{T}}\alpha_\theta(\hat{\theta}_C) = 0. \tag{2.21}$$

It follows that the maximum composite likelihood estimator $\eta(\hat{\theta}_C)$ is obtained by projecting the data point $t(y)$ onto $\mathcal{F}_{\Omega_\eta}$ orthogonally to $\alpha_\theta(\hat{\theta}_C)$. Here, the set of $t(y)$ points having $\hat{\theta}_C$ as solution of (2.21) is denoted by

$$\mathcal{L}_{\hat{\theta}_C} = \{t(y) : cU(\hat{\theta}_C) = 0\}.$$

In Figure 2.1, the curved exponential family $\mathcal{F}$ is represented by $\mathcal{F}_{\Omega_\eta}$ in Ω_η.

Conditions (2.20) and (2.21) will generally lead to different estimates, $\hat{\theta}$ and $\hat{\theta}_C$, as illustrated in Figure 2.1. We note that, we have not proved that the two estimates have to be different. In fact, assuming $\hat{\theta} = \hat{\theta}_C$, (2.20) and (2.21) imply that $\{t(y) - \eta(\hat{\theta})\}^{\mathrm{T}}\{\phi_\theta(\hat{\theta}) - \alpha_\theta(\hat{\theta})\} = 0$ and this could be satisfied even with $\alpha(\theta) \neq \phi(\theta)$. However, we have not found any example of curved exponential family with $\hat{\theta} = \hat{\theta}_C$.

We now show also that in general the Fisher and Godambe information do not coincide. Returning to index notation, we can rewrite the full log-likelihood as

$$\ell(\theta) = t^a(y)\phi^a(\theta) - K(\theta),$$

with the corresponding score vector

$$U_r(\theta) = t^a(y)\phi^a_r(\theta) - K_r(\theta),$$

where, $\phi^a_r(\theta) = \partial\phi^a(\theta)/\partial\theta_r$. Since $\mathrm{E}_\theta\{U_r(\theta)\} = 0 \Longrightarrow K_r(\theta) = \eta^a(\theta)\phi^a_r(\theta)$. Therefore, the generic element of the score becomes

$$U_r(\theta) = \{t^a(y) - \eta^a(\theta)\}\phi^a_r(\theta).$$

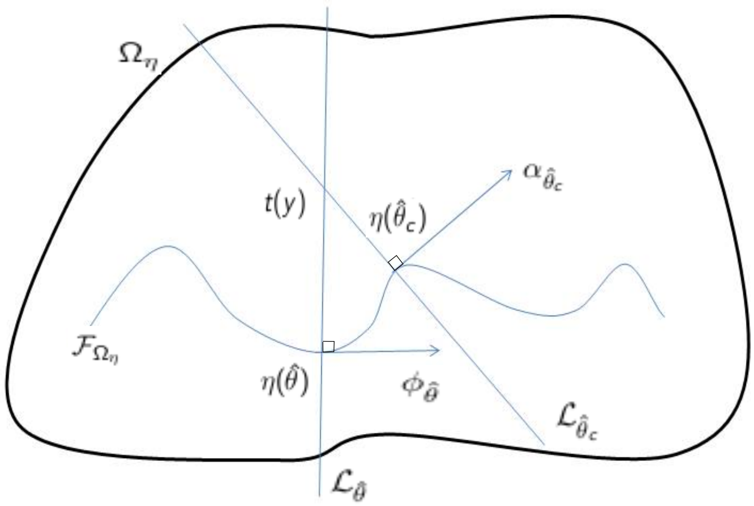

Figure 2.1: The curved exponential family $\mathcal{F}$ is represented by the curve $\mathcal{F}_{\Omega_\eta} = \{\eta = \eta(\theta) : \theta \in \Theta\}$. The maximum likelihood estimator of $\eta(\hat{\theta})$ and the maximum composite likelihood estimator of $\eta(\hat{\theta}_C)$ are obtained by projecting the $t(y)$ point onto $\mathcal{F}_{\Omega_\eta}$ orthogonally to $\phi_{\hat{\theta}}$ and $\alpha_{\hat{\theta}_C}$ respectively.

The generic element of the observed information matrix is

$$j_{rs}(\theta) = -U_{rs}(\theta) = \eta^a_s(\theta)\phi^a_r(\theta) - \{t^a - \eta^a(\theta)\}\phi^a_{rs}(\theta),$$

where, $\eta^a_s(\theta) = \partial\eta^a(\theta)/\partial\theta_s$ and $\phi^a_{rs}(\theta) = \partial^2\phi^a(\theta)/(\partial\theta_r\partial\theta_s)$. The Fisher information has therefore elements,

$$i_{rs}(\theta) = \mathrm{E}_\theta\{j_{rs}(\theta)\} = \eta^a_s(\theta)\phi^a_r(\theta).$$

The composite log-likelihood can be rewritten as

$$c\ell(\theta) = t^a(y)\alpha^a(\theta) - Q(\theta),$$

and the corresponding generic element of its score function is given by

$$cU_r(\theta) = \{t^a(y) - \eta^a(\theta)\}\alpha^a_r(\theta),$$

where, $\alpha_r^a(\theta) = \partial\alpha^a(\theta)/\partial\theta_r$. The matrix $H(\theta)$ has elements

$$H_{rs}(\theta) = \mathrm{E}_\theta\{-cU_{rs}(\theta)\} = \eta_s^a(\theta)\alpha_r^a(\theta)$$

while $J(\theta)$, has elements

$$\begin{aligned} J_{rs}(\theta) &= \mathrm{E}_\theta\left[cU_r(\theta)cU_s(\theta)\right] = \mathrm{E}_\theta\left[\{t^a(Y) - \eta^a(\theta)\}\,\alpha_r^a(\theta)\left\{t^b(Y) - \eta^b(\theta)\right\}\alpha_s^b(\theta)\right] \\ &= \alpha_r^a(\theta)\alpha_s^b(\theta)\mathrm{E}_\theta\left[\{t^a(Y) - \eta^a(\theta)\}\left\{t^b(Y) - \eta^b(\theta)\right\}\right] \\ &= \alpha_r^a(\theta)V_{ab}(\theta)\alpha_s^b(\theta), \end{aligned}$$

where, $V_{ab}(\theta)$ are the elements of the matrix $V = \mathrm{Var}_\theta\{t(Y)\}$. Inverting $J(\theta)$, we obtain

$$J^{rs}(\theta) = \Theta_a^s(\theta)V^{ab}(\theta)\Theta_b^r(\theta),$$

where $\Theta_a^s(\theta)\alpha_r^a(\theta) = \delta_r^s$ and $\Theta_b^r(\theta)\alpha_s^b(\theta) = \delta_s^r$. Hence, the Godambe information matrix has elements

$$\begin{aligned} G_{rs}(\theta) &= H_{rs}(\theta)J^{rs}(\theta)H_{sr}(\theta) = \eta_s^a(\theta)\alpha_r^a(\theta)\Theta_b^s(\theta)V^{bc}(\theta)\Theta_c^r(\theta)\eta_r^d(\theta)\alpha_s^d(\theta) \\ &= \delta_c^a\delta_b^d\eta_s^a(\theta)V^{bc}(\theta)\eta_r^d \\ &= \eta_s^a(\theta)V^{ba}(\theta)\eta_r^b(\theta). \end{aligned}$$

Hence, the Fisher information has a different form than the Godambe information, and therefore they are not guaranteed to be equal.

Example 2 *Equicovariance normal model*
This is the same example as the previous one with mean 0 and an equicovariance matrix with common known variances $\sigma^2 = 1$. The one-dimensional parameter is ρ. The calculation of the full log-likelihood based on n independent observations gives

$$\begin{aligned} \ell(\rho) = &-\frac{n(q-1)}{2}\log(1-\rho) - \frac{n}{2}\log\{1+\rho(q-1)\} - \frac{1}{2(1-\rho)}ss_W \\ &- \frac{1}{2q\{1+\rho(q-1)\}}ss_V \end{aligned}$$

while the pairwise log-likelihood is

$$c\ell^{\mathrm{P}}(\rho) = -\frac{nq(q-1)}{4}\log(1-\rho^2) - \frac{q-1+\rho}{2(1-\rho^2)}ss_W - \frac{(q-1)}{2q(1+\rho)}ss_V,$$

where $ss_W = \sum_{i=1}^{n}\sum_{r=1}^{q}(y_{ir} - \overline{y}_i)^2$ e $ss_V = \sum_{i=1}^{n} y_{i\cdot}^2$ with $y_{i\cdot} = \sum_{r=1}^{q} y_{ir}$, are components of the sufficient statistic. Since the only parameter is ρ, the pairwise likelihood coincides with the pairwise conditional likelihood. Both full and pairwise log-likelihood are of curved exponential type. Cox & Reid (2004, Example 1) show that there is a loss of efficiency when using $c\ell^{\mathrm{P}}(\rho)$ in place of $\ell(\rho)$.

Example 3 *First order Gaussian autoregressive process*
Let $\{Y_t\}$ be a normal autoregressive process of order one with correlation coefficient ρ. The model is defined as

$$Y_t = \rho Y_{t-1} + \epsilon_t,\ t = 1, \ldots, T,$$

where ϵ_t is normal with mean zero and constant variance σ^2. Here it is assumed that the initial value Y_0 is a random variable from $N(0, \sigma^2/(1 - \rho^2))$, and independent of $\epsilon_1, \ldots, \epsilon_T$. The correlation between Y_{t-1} and Y_t is equal to $\sigma^2\rho/(1-\rho^2)$. Given the sample $Y_1, \ldots, Y_T$, the likelihood is given by

$$\ell(\sigma^2, \rho) = \frac{1}{2}\log(1-\rho^2) - \frac{T}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}[S_1 + \rho^2 S_2 - 2\rho S_{12}]$$

and while the pairwise log-likelihood is

$$c\ell^{\mathrm{P}}(\sigma^2, \rho) = -(T-1)\log(\sigma^2) + \frac{T-1}{2}\log(1-\rho^2) - \frac{1}{2\sigma^2}[S_1 + S_2 - 2\rho S_{12}],$$

where $S_1 = \sum_{t=1}^{T} y_t^2$, $S_2 = \sum_{t=2}^{T-1} y_t^2$ and $S_{12} = \sum_{t=1}^{T-1} y_t y_{t+1}$ are components of the sufficient statistic. Even in this case, both $\ell(\sigma^2, \rho)$ and $c\ell^{\mathrm{P}}(\sigma^2, \rho)$ are of curved exponential type, with the same sufficient statistic $t(y) = (S_1, S_2, S_{12})$. We consider a numerical example using data on the luteinizing hormone in $T = 48$ blood samples, taken at 10 minutes intervals from a human female. The data can be found in object *lh* in the standard R distribution (R Core Team, 2012). The estimates of σ^2 and ρ based on $\ell(\sigma^2, \rho)$ are 0.25075 and 0.98077, respectively, while those based on $c\ell^{\mathrm{P}}(\sigma^2, \rho)$ are given by 0.25033 and 0.97904, respectively. As we can see, the maximum likelihood estimates do not coincide with the ones based on the pairwise log-likelihood.

Chapter 3

Combined composite likelihood

3.1 Introduction

For a number of complex statistical models, the composite likelihood may be considered as an useful alternative to the standard likelihood, due to its many appealing features discussed in Section 2.3. Since there are different ways to formulate a composite likelihood, the crucial question that arises here is relative to the choice of the composite likelihood which has more desirable properties.

In contexts in which one may use either the pairwise or the independence likelihood (Varin, 2008; Varin et al., 2011) in place of the standard likelihood, it is possible to combine them adequately obtaining thus a combined composite likelihood. A formulation of a combined composite likelihood, almost not explored in the literature, is suggested in Cox & Reid (2004), and is given by a linear combination of independence and pairwise log-likelihoods resulting in a new objective function, which depends on a constant to be chosen. Of course, this objective function could represent a usefull composite likelihood in situations where there is information on the parameter of interest both in one and two-dimensional marginal densities.

Notable composite likelihoods are particular case of the combined composite likelihood such as the pairwise marginal and conditional likelihood. Hence, this new type of pseudo-likelihood could be seen as a compromise between the pairwise marginal and conditional likelihood. On the other hand, from the form of the combined composite likelihood, it comes out that the independence likelihood is not a particular case.

As for general composite likelihoods, under suitable regularity conditions, the various inferential procedures based on the combined composite likelihood have theoretical properties often similar to those based on the standard likelihood, although expecting a loss of efficiency.

This chapter aims to study the properties and to explore the inferential aspects of the combined composite likelihood. Exact and asymptotic properties are studied. Two examples of the combined composite likelihood, both based on the multivariate normal distribution are considered. Particular attention will be focused on efficiency by comparing the results obtained with the combined composite likelihood with those obtained with the standard likelihood.

This chapter will be organized as follows. Section 3.2 is devoted to the definition of combined composite likelihood and the study of its exact and asymptotic properties. The exact properties lead us to the identification of a possible strategy for finding the range of admissible values for the constant which combines the independence and pairwise likelihood. In Section 3.3, two examples of combined composite likelihood are analyzed in detail. In the first, the parameter is scalar and it is the common partial correlation of a multivariate normal model. An empirical check at consistency of the combined composite likelihood estimator is performed by simulation in the case in which the number of sample units is fixed and the dimension of the observed vector increases. Moreover, several empirical studies of efficiency suggest the conclusion that a good, although not optimal, choice for the value of the constant lead to the pairwise conditional likelihood. The second example deals with a model for microarray data and the overall parameter is four-dimensional. The efficiency of the estimator for each component is considered. Also here, a good choice for the value of the constant leads again to the pairwise conditional likelihood.

3.2 Definitions and properties

Let $Y_i = (Y_{i1}, \ldots, Y_{iq})^{\mathrm{T}}$, $i = 1, \ldots, n$, be a q-dimensional random vector with joint density $f(y_i; \theta)$, where θ is some unknown d-dimensional real parameter. Suppose that the full q-dimensional distribution is not easily tractable, but it is possible to evaluate both univariate and bivariate marginal distributions. Moreover, in the following, we assume that both such marginal distributions depend on θ.

In order to make inference on θ, one may consider the combined composite log likelihood, which is defined as

$$c\ell^a(\theta; y_i) = 2c\ell^{\mathrm{P}}(\theta; y_i) - a(q-1)c\ell^{\mathrm{I}}(\theta; y_i),$$

where, $c\ell^{\mathrm{P}}(\theta; y_i) = \sum_{r=1}^{q-1}\sum_{s=r+1}^{q} \log f(y_{ir}, y_{is}; \theta)$ and $c\ell^{\mathrm{I}}(\theta; y_i) = \sum_{r=1}^{q} \log f(y_{ir}; \theta)$ are the unweighted pairwise and independence log likelihoods for a single observation, respectively. Under random sampling of size n, $c\ell^a(\theta) = \sum_{i=1}^{n} c\ell^a(\theta; y_i), c\ell^{\mathrm{P}}(\theta) = \sum_{i=1}^{n} c\ell^{\mathrm{P}}(\theta; y_i)$ and $c\ell^{\mathrm{I}}(\theta) = \sum_{i=1}^{n} c\ell^{\mathrm{I}}(\theta; y_i)$ will denote the combined composite, pairwise and independence log-likelihood, respectively. Some values of a lead to notable composite likelihoods. In fact, $a = 0$ corresponds to the pairwise likelihood, while $a = 1$ corresponds to the pairwise conditional log likelihood, which is based on all possible conditional distributions of one component given another. Indeed, as an example, for simplicity with $n = 1, a = 1$ and $q = 2$, we obtain

$$\begin{aligned} c\ell^1(\theta) &= 2\log f(y_{11}, y_{12}; \theta) - \log f(y_{11}; \theta) - \log f(y_{12}; \theta) \\ &= \log f(y_{11} \mid y_{12}; \theta) + \log f(y_{12} \mid y_{11}; \theta). \end{aligned}$$

3.2.1 Exact properties

We now consider whether the desirable exact properties of the composite likelihood, defined in Section 2.3, are satisfied by $c\ell_a(\theta)$. Property (a) is definitely satisfied for $a = 0$ and $a = 1$, because such values correspond to well-known composite likelihoods. For other values of a, knowing that

$$\mathrm{E}_{\theta_0}\left\{c\ell^{\mathrm{P}}(\theta)\right\} \leq \mathrm{E}_{\theta_0}\left\{c\ell^{\mathrm{P}}(\theta_0)\right\} \text{ and } \mathrm{E}_{\theta_0}\left\{c\ell^{\mathrm{I}}(\theta)\right\} \leq \mathrm{E}_{\theta_0}\left\{c\ell^{\mathrm{I}}(\theta_0)\right\},$$

and rewriting the property (a) as follows,

$$2\left[\mathrm{E}_{\theta_0}\left\{c\ell^{\mathrm{P}}(\theta_0)\right\} - \mathrm{E}_{\theta_0}\left\{c\ell^{\mathrm{P}}(\theta)\right\}\right] \geq a(q-1)\left[\mathrm{E}_{\theta_0}\left\{c\ell^{\mathrm{I}}(\theta_0)\right\} - \mathrm{E}_{\theta_0}\left\{c\ell^{\mathrm{I}}(\theta)\right\}\right],$$

we see that (a) will be automatically satisfied for any $a \leq 1$, since the latter inequality is satisfied for $a = 1$. On the other hand, for $a > 1$, the property (a) is not guaranteed. Property (b) is always satisfied since $U^a(\theta) = \partial c\ell^a(\theta)/\partial\theta$ is a linear combination of two unbiased terms. Finally, as regards to property (c), with θ scalar, we have that

$$\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}c\ell^a(\theta) = 2\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}c\ell^{\mathrm{P}}(\theta) - a(q-1)\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}c\ell^{\mathrm{I}}(\theta),$$

hence,

$$\mathrm{E}_{\theta_0}\left[\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}c\ell^a(\theta)\right]\bigg|_{\theta=\theta_0} = -2H^{\mathrm{P}}(\theta_0) + a(q-1)H^{\mathrm{I}}(\theta_0),$$

where $H^m(\theta_0) = \mathrm{E}_{\theta_0}\left[-\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}c\ell^m(\theta)\right]\Big|_{\theta=\theta_0}$, m=P,I. Then, the expectation of the second derivative will be negative if

$$-2H^{\mathrm{P}}(\theta_0) + a(q-1)H^{\mathrm{I}}(\theta_0) < 0.$$

Quantities $H^{\mathrm{P}}(\theta_0)$ and $H^{\mathrm{I}}(\theta_0)$ are positive, due to the fact that $c\ell^{\mathrm{P}}(\theta)$ and $c\ell^{\mathrm{I}}(\theta)$ satisfy the Wald inequality and therefore we get that a must satisfy

$$a < \frac{2}{q-1}\frac{H^{\mathrm{P}}(\theta_0)}{H^{\mathrm{I}}(\theta_0)} = A_q(\theta_0)\cdot \tag{3.1}$$

The quantity $A_q(\theta_0)$ is positive and depends only on θ_0, since in practice q is known. For the multiparameter case, it is difficult to derive a condition similar to (3.1). It will be necessary to use one of the properties of positive definite matrices, as for instance through the eigenvalues of $H^a(\theta_0) = \mathrm{E}_{\theta_0}\left[-\frac{\partial^2}{\partial\theta\partial\theta^{\mathrm{T}}}c\ell^a(\theta)\right]\Big|_{\theta=\theta_0}$, which should all be positive, as done in Example 3.3.2 below.

3.2.2 Asymptotic properties

We now turn to the asymptotic properties of the combined composite likelihood quantities, such as consistency, asymptotic distribution of the estimator and of the likelihood ratio statistic. To this end, two scenarios are considered. The first one concerns the case in which q is fixed and $n \longrightarrow \infty$, while the second considers n fixed and $q \longrightarrow \infty$.

i) q fixed and $n \longrightarrow \infty$

We start by looking at the problem of consistency of the maximum combined composite likelihood estimator. To this end, we define

$$
\begin{aligned}
U_{1i}^{\mathrm{P}}(\theta) &= \frac{\partial}{\partial\theta}\sum_{r=1}^{q-1}\sum_{s=r+1}^{q}\log f(y_{ir}, y_{is};\theta), \quad U_1^{\mathrm{P}}(\theta) = \sum_{i=1}^{n} U_{1i}^{\mathrm{P}}(\theta),\\
U_{1i}^{\mathrm{I}}(\theta) &= \frac{\partial}{\partial\theta}\sum_{r=1}^{q}\log f(y_{ir};\theta), \quad U_1^{\mathrm{I}}(\theta) = \sum_{i=1}^{n} U_{1i}^{\mathrm{I}}(\theta),\\
U_{2i}^{\mathrm{P}}(\theta) &= \frac{\partial^2}{\partial\theta\partial\theta^{\mathrm{T}}}\sum_{r=1}^{q-1}\sum_{s=r+1}^{q}\log f(y_{ir}, y_{is};\theta), \quad U_2^{\mathrm{P}}(\theta) = \sum_{i=1}^{n} U_{2i}^{\mathrm{P}}(\theta),\\
U_{2i}^{\mathrm{I}}(\theta) &= \frac{\partial^2}{\partial\theta\partial\theta^{\mathrm{T}}}\sum_{r=1}^{q}\log f(y_{ir};\theta), \quad U_2^{\mathrm{I}}(\theta) = \sum_{i=1}^{n} U_{2i}^{\mathrm{I}}(\theta),\\
U^a(\theta) &= 2U_1^{\mathrm{P}}(\theta) - a(q-1)U_1^{\mathrm{I}}(\theta)\cdot
\end{aligned}
$$

The maximum combined composite likelihood estimator will be denoted by $\hat{\theta}_C^a$. Since $U^a(\hat{\theta}_C^a) = 0$, the expansion of $U^a(\hat{\theta}_C^a)$ around θ can be written

$$
\begin{aligned}
U^a(\hat{\theta}_C^a) &= 0\\
&\doteq 2U_1^{\mathrm{P}}(\theta) - a(q-1)U_1^{\mathrm{I}}(\theta) + \{2U_2^{\mathrm{P}}(\theta) - a(q-1)U_2^{\mathrm{I}}(\theta)\}(\hat{\theta}_C^a - \theta).
\end{aligned}
\tag{3.2}
$$

From equation (3.2), we obtain

$$
(\hat{\theta}_C^a - \theta) \doteq -\left\{2U_2^{\mathrm{P}}(\theta) - a(q-1)U_2^{\mathrm{I}}(\theta)\right\}^{-1}\left\{2U_1^{\mathrm{P}}(\theta) - a(q-1)U_1^{\mathrm{I}}(\theta)\right\}\cdot \tag{3.3}
$$

The first quantity in curly brackets on the right-hand side of equation (3.3) is the inverse of a $p \times p$ matrix where each element is of order $O_P(n)$, this is, n^{-1} times a $p \times p$ matrix with each element of order $O_P(1)$. Instead, the second quantity is a $p \times 1$ vector where each of its elements is of order $O_P(n^{1/2})$. Their product produces a $p \times 1$ vector with elements of order $O_P(n^{-1/2})$. Hence, we get

$$(\hat{\theta}_C^a - \theta) \doteq O_P(n^{-1/2}).$$

This result suggests that the combined composite likelihood estimator is a consistent estimator of θ.

We now focus on the determination of the asymptotic distribution of the combined composite likelihood estimator. Exploiting the above expansion, (3.2) may be written as

$$-\{2U_1^{\mathrm{P}}(\theta) - a(q-1)U_1^{\mathrm{I}}(\theta)\} \doteq n\left\{2\frac{1}{n}U_2^{\mathrm{P}}(\theta) - a(q-1)\frac{1}{n}U_2^{\mathrm{I}}(\theta)\right\}(\hat{\theta}_C^a - \theta)\cdot \quad (3.4)$$

In the following, the subscript one on the matrices H and J will indicate that the quantity is calculated with only one observation. Applying the law of large numbers to each term of the matrices $\frac{1}{n}U_2^{\mathrm{P}}(\theta)$ and $\frac{1}{n}U_2^{\mathrm{I}}(\theta)$, we get

$$\frac{1}{n}U_2^{\mathrm{P}}(\theta) \xrightarrow{p} \mathrm{E}_\theta\{U_{2i}^{\mathrm{P}}(\theta)\} = -H_1^{\mathrm{P}}(\theta),$$
$$\frac{1}{n}U_2^{\mathrm{I}}(\theta) \xrightarrow{p} \mathrm{E}_\theta\{U_{2i}^{\mathrm{I}}(\theta)\} = -H_1^{\mathrm{I}}(\theta).$$

This means that

$$n\left\{2\frac{1}{n}U_2^{\mathrm{P}}(\theta) - a(q-1)\frac{1}{n}U_2^{\mathrm{I}}(\theta)\right\} \xrightarrow{p} -\{2H^{\mathrm{P}}(\theta) - a(q-1)H^{\mathrm{I}}(\theta)\}.$$

Noting that $H^a(\theta) = 2H^{\mathrm{P}}(\theta) - a(q-1)H^{\mathrm{I}}(\theta)$, the above expression becomes

$$n\left\{2\frac{1}{n}U_2^{\mathrm{P}}(\theta) - a(q-1)\frac{1}{n}U_2^{\mathrm{I}}(\theta)\right\} \xrightarrow{p} -H^a(\theta),$$

recalling that for n independent and identically distributed observations $H^a(\theta) = nH_1^a(\theta)$, $H^{\mathrm{P}}(\theta) = nH_1^{\mathrm{P}}(\theta)$ and $H^{\mathrm{I}}(\theta) = nH_1^{\mathrm{I}}(\theta)$. The quantity on the left-hand side of equation (3.4) has zero expectation under θ, since it is the score based on $c\ell_a(\theta)$; hence from the central limit theorem it follows that

$$\{2U_1^{\mathrm{P}}(\theta) - a(q-1)U_1^{\mathrm{I}}(\theta)\} \dot{\sim} N_d(0, J^a(\theta)),$$

where

$$\begin{aligned} J^a(\theta) &= \mathrm{Var}_\theta\{2U_1^{\mathrm{P}}(\theta) - a(q-1)U_1^{\mathrm{I}}(\theta)\} \\ &= 4\mathrm{Var}_\theta\{U_1^{\mathrm{P}}(\theta)\} + a^2(q-1)^2\mathrm{Var}_\theta\{U_1^{\mathrm{I}}(\theta)\} - 4a(q-1)\times \\ &\quad \mathrm{Cov}_\theta\{U_1^{\mathrm{P}}(\theta), U_1^{\mathrm{I}}(\theta)\} \\ &= 4J^{\mathrm{P}}(\theta) + a^2(q-1)^2 J^{\mathrm{I}}(\theta) - 4a(q-1)J^{PI}(\theta), \end{aligned}$$

with $J^{PI}(\theta) = \mathrm{Cov}_\theta\{U_1^{\mathrm{P}}(\theta), U_1^{\mathrm{I}}(\theta)\}$ which represents the covariance between vectors $U_1^{\mathrm{P}}(\theta)$ and $U_1^{\mathrm{I}}(\theta)$ and, it may be calculated as follows

$$\begin{aligned} \mathrm{Cov}_\theta\{U_1^{\mathrm{P}}(\theta), U_1^{\mathrm{I}}(\theta)\} &= \mathrm{E}_\theta\left[\left(U_1^{\mathrm{P}}(\theta) - E_\theta[U_1^{\mathrm{P}}(\theta)]\right)\left(U_1^{\mathrm{I}}(\theta) - E_\theta[U_1^{\mathrm{I}}(\theta)]\right)^{\mathrm{T}}\right] \\ &= \mathrm{E}_\theta\left[U_1^{\mathrm{P}}(\theta) U_1^{\mathrm{I}}(\theta)^{\mathrm{T}}\right]. \end{aligned}$$

Exploiting the above developed quantities, (3.4) may be written as

$$-\{2U_1^{\mathrm{P}}(\theta) - a(q-1)U_1^{\mathrm{I}}(\theta)\} \doteq -H^a(\theta)(\hat{\theta}_C^a - \theta)\cdot$$

Thus, we get

$$(\hat{\theta}_C^a - \theta) \doteq H^a(\theta)^{-1}\{2U_1^{\mathrm{P}}(\theta) - a(q-1)U_1^{\mathrm{I}}(\theta)\}\cdot \qquad (3.5)$$

It follows from (3.5) that

$$(\hat{\theta}_C^a - \theta) \dot{\sim} N_d(0, G^a(\theta)^{-1}),$$

where, $G^a(\theta) = H^a(\theta)J^a(\theta)^{-1}H^a(\theta)$ is the Godambe information.

As a last step, we consider the asymptotic distribution of the combined composite likelihood ratio statistic, $W^a(\theta)$. An expansion of $c\ell^a(\theta)$ around $\hat{\theta}_C^a$ gives

$$\begin{aligned} c\ell^a(\theta) &= \{2c\ell^{\mathrm{P}}(\hat{\theta}_C^a) - a(q-1)c\ell^{\mathrm{I}}(\hat{\theta}_C^a)\} + (\theta - \hat{\theta}_C^a)^{\mathrm{T}}\Big\{2U_1^{\mathrm{P}}(\hat{\theta}_C^a) \\ &\quad - a(q-1)U_1^{\mathrm{I}}(\hat{\theta}_C^a)\Big\} + \frac{1}{2}(\theta - \hat{\theta}_C^a)^{\mathrm{T}}\Big\{2U_2^{\mathrm{P}}(\hat{\theta}_C^a) \\ &\quad - a(q-1)U_2^{\mathrm{I}}(\hat{\theta}_C^a)\Big\}(\theta - \hat{\theta}_C^a) + R\cdot \end{aligned}$$

where the remainder term R involves a $d \times d \times d$ array of third order partial derivatives. It is straightforward to show that R is of order $O_P(n^{-\frac{1}{2}})$. Replacing $c\ell_a(\theta)$ in $W^a(\theta)$, we obtain after some simplification that

$$\begin{aligned}
W^a(\theta) &= 2\{c\ell^a(\hat{\theta}^a_C) - c\ell^a(\theta)\} \\
&= -(\theta - \hat{\theta}^a_C)^{\mathrm{T}} \left\{2U_2^{\mathrm{P}}(\hat{\theta}^a_C) - a(q-1)U_2^{\mathrm{I}}(\hat{\theta}^a_C)\right\}(\theta - \hat{\theta}^a_C) + O_P(n^{-\frac{1}{2}}) \\
&= -n(\theta - \hat{\theta}^a_C)^{\mathrm{T}} \left\{2\frac{1}{n}U_2^{\mathrm{P}}(\hat{\theta}^a_C) - a(q-1)\frac{1}{n}U_2^{\mathrm{I}}(\hat{\theta}^a_C)\right\}(\theta - \hat{\theta}^a_C) + O_P(n^{-\frac{1}{2}}).
\end{aligned}$$

Applying the law of large numbers to each term of the matrix $2\frac{1}{n}U_2^{\mathrm{P}}(\hat{\theta}^a_C) - a(q-1)\frac{1}{n}U_2^{\mathrm{I}}(\hat{\theta}^a_C)$, and taking into account that $\hat{\theta}^a_C \xrightarrow{P} \theta$, we get

$$\begin{aligned}
W^a(\theta) &= n(\theta - \hat{\theta}^a_C)^{\mathrm{T}}\{2H_1^{\mathrm{P}}(\theta) - a(q-1)H_1^{\mathrm{I}}(\theta)\}(\theta - \hat{\theta}^a_C) + O_P(n^{-1/2}) \\
&= (\hat{\theta}^a_C - \theta)^{\mathrm{T}} H^a(\theta)(\hat{\theta}^a_C - \theta) + O_P(n^{-1/2}),
\end{aligned}$$

where $H^a(\theta) = 2H^{\mathrm{P}}(\theta) - a(q-1)H^{\mathrm{I}}(\theta)$. Since $\hat{\theta}^a_C$ is asymptotically normally distributed with mean θ, covariance matrix $G^a(\theta)^{-1}$ and $H^a(\theta)$ is a $d \times d$ nonnegative definite, matrix, we can apply Theorem 8.5 of Severini (2005, page 245), which states that $Q = (\hat{\theta}^a_C - \theta)^{\mathrm{T}} H^a(\theta)(\hat{\theta}^a_C - \theta)$ has asymptotically cumulant-generating function

$$K_Q(t) = -\frac{1}{2}\sum_{k=1}^{d} \log(1 - 2t\lambda_k),$$

where $\lambda_1, \ldots, \lambda_d$ are the eigenvalues of the matrix

$$G^a(\theta)^{-1}H^a(\theta) = H^a(\theta)^{-1}J^a(\theta).$$

We can easily recognize that $K_Q(t)$ is the cumulant-generating function of a random variable $\sum_{k=1}^{d} \lambda_k X_k$, where $X_1, \ldots, X_d$ are independent χ^2_1 random variables. Therefore,

$$W^a(\theta) \dot{\sim} \sum_{k=1}^{d} \lambda_k X_k.$$

ii) n **fixed and** $q \longrightarrow \infty$

Even in this case, we start by looking at the problem of consistency of the maximum combined composite likelihood estimator. We assume for simplicity that θ is scalar and we consider without loss of generality $n = 1$. For a good understanding of the steps, we define

$$U^{\mathrm{P}}_{rs}(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta}\log f(y_r, y_s; \theta), \quad U^{\mathrm{P}'}_{rs}(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta}U^{\mathrm{P}}_{rs}(\theta),$$
$$U^{\mathrm{I}}_{r}(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta}\log f(y_r; \theta), \quad U^{\mathrm{I}'}_{r}(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta}U^{\mathrm{I}}_{r}(\theta),$$
$$U^{a}(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta}c\ell^{a}(\theta) = 2\sum_{r=1}^{q-1}\sum_{s=r+1}^{q} U^{\mathrm{P}}_{rs}(\theta) - a(q-1)\sum_{r=1}^{q} U^{\mathrm{I}}_{r}(\theta), \quad U^{a'}(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta}U^{a}(\theta).$$

Exploiting the fact that $U^{a}(\hat{\theta}^{a}_{C}) = 0$, we can expand $U^{a}(\hat{\theta}^{a}_{C})$ around θ, to first order and get

$$U^{a}(\theta) + (\hat{\theta}^{a}_{C} - \theta)U^{a'}(\theta) \doteq 0.$$

This is equivalent to

$$\left\{2\sum_{r=1}^{q-1}\sum_{s=r+1}^{q} U^{\mathrm{P}}_{rs}(\theta) - a(q-1)\sum_{r=1}^{q} U^{\mathrm{I}}_{r}(\theta)\right\} + (\hat{\theta}^{a}_{C} - \theta)\left\{2\sum_{r=1}^{q-1}\sum_{s=r+1}^{q} U^{\mathrm{P}'}_{rs}(\theta) \right.$$
$$\left. -a(q-1)\sum_{r=1}^{q} U^{\mathrm{I}'}_{r}(\theta)\right\} \doteq 0\cdot$$

It may also be written as

$$q^{-2}\left\{2\sum_{r=1}^{q-1}\sum_{s=r+1}^{q} U^{\mathrm{P}}_{rs}(\theta) - a(q-1)\sum_{r=1}^{q} U^{\mathrm{I}}_{r}(\theta)\right\} + q^{-2}(\hat{\theta}^{a}_{C} - \theta)\times$$
$$\left\{2\sum_{r=1}^{q-1}\sum_{s=r+1}^{q} U^{\mathrm{P}'}_{rs}(\theta) - a(q-1)\sum_{r=1}^{q} U^{\mathrm{I}'}_{r}(\theta)\right\} \doteq 0\cdot \tag{3.6}$$

Since the second quantity in curly brackets in (3.6) has expectation different from zero, it follows that its order in probability depends on the order of its mean which, multiplied by q^{-2}, is of order $O_P(1)$. The first quantity in curly brackets has zero expectation and this means that its order in probability depends on the order of its variance which is given by

$$q^{-4}\left[4\mathrm{Var}_\theta\left\{\sum_{r=1}^{q-1}\sum_{s=r+1}^{q}U_{rs}^{\mathrm{P}}(\theta)\right\}+a^2(q-1)^2\mathrm{Var}_\theta\left\{\sum_{r=1}^{q}U_r^{\mathrm{I}}(\theta)\right\}-4a(q-1)\times\right.$$
$$\left.\mathrm{Cov}_\theta\left\{\sum_{r=1}^{q-1}\sum_{s=r+1}^{q}U_{rs}^{\mathrm{P}}(\theta),\sum_{r=1}^{q}U_r^{\mathrm{I}}(\theta)\right\}\right]. \tag{3.7}$$

Assuming that each pair of observations has the same bivariate distribution, as in Cox & Reid (2004), the first summand in (3.7) can be calculated as

$$\begin{aligned}\mathrm{Var}_\theta\left\{\sum_{r=1}^{q-1}\sum_{s=r+1}^{q}U_{rs}^{\mathrm{P}}(\theta)\right\} &= N_1\mathrm{Var}_\theta\left\{U_{rs}^{\mathrm{P}}(\theta)\right\}\\ &\quad+2N_2\mathrm{Cov}_\theta\{U_{\mathrm{P}st}(\theta),U_{\mathrm{P}sv}(\theta)\}\\ &\quad+2N_3\mathrm{Cov}_\theta\{U_{\mathrm{P}st}(\theta),U_{\mathrm{P}vw}(\theta)\}\\ &= N_1\mathrm{E}_\theta\left\{U_{rs}^{\mathrm{P}}(\theta)^2\right\}+2N_2\mathrm{E}_\theta\{U_{\mathrm{P}st}(\theta)U_{\mathrm{P}sv}(\theta)\}\\ &\quad+2N_3\mathrm{E}_\theta\{U_{\mathrm{P}st}(\theta)U_{\mathrm{P}vw}(\theta)\},\end{aligned}$$

where, given that we have q indices, $N_1=q(q-1)/2$ is the number of pairs of ordered indices and $N_1(N_1-1)/2$ is the number of pairs of pairs of ordered indices. Then, $N_2=q\binom{q-1}{2}=q(q-1)(q-2)/2$ is the number of pairs of pairs of ordered indices which have one element in common. Finally, $N_3=N_1(N_1-1)/2-q(q-1)(q-2)/2=q(q-1)(q-2)(q-3)/8$ is the number of pairs with no common elements. Hence,

$$\begin{aligned}\mathrm{Var}_\theta\left\{\sum_{r=1}^{q-1}\sum_{s=r+1}^{q}U_{rs}^{\mathrm{P}}(\theta)\right\} &= \frac{q(q-1)}{2}\mathrm{E}_\theta\left\{U_{rs}^{\mathrm{P}}(\theta)^2\right\}+q(q-1)(q-2)\times\\ &\quad\mathrm{E}_\theta\{U_{\mathrm{P}st}(\theta)U_{\mathrm{P}sv}(\theta)\}+\frac{q(q-1)(q-2)(q-3)}{4}\times\\ &\quad\mathrm{E}_\theta\{U_{\mathrm{P}st}(\theta)U_{\mathrm{P}vw}(\theta)\}.\end{aligned}$$

Moreover, we have that

$$\begin{aligned}\mathrm{Cov}_\theta\left\{\sum_{r=1}^{q-1}\sum_{s=r+1}^{q}U_{rs}^{\mathrm{P}}(\theta),\sum_{r=1}^{q}U_r^{\mathrm{I}}(\theta)\right\} &= N_3\mathrm{E}_\theta\left\{U_{\mathrm{P}st}(\theta)U_{\mathrm{I}s}(\theta)\right\}\\ &\quad+N_4\mathrm{E}_\theta\left\{U_{\mathrm{P}st}(\theta)U_{\mathrm{I}v}(\theta)\right\},\end{aligned}$$

where $N_4 = \frac{q(q-1)}{2} \times (q-2)$ corresponds to the number of cases in which the two terms of the covariance have no common elements. Instead, $N_3 = \frac{q(q-1)}{2} \times q - \frac{q(q-1)}{2} \times (q-2) = q(q-1)$ is the number of cases in which the two terms of the covariance have one element in common. Therefore,

$$\mathrm{Cov}_\theta \left\{ \sum_{r=1}^{q-1} \sum_{s=r+1}^{q} U_{rs}^{\mathrm{P}}(\theta), \sum_{r=1}^{q} U_r^{\mathrm{I}}(\theta) \right\} = q(q-1)\mathrm{E}_\theta \left\{ U_{\mathrm{P}st}(\theta) U_{\mathrm{I}s}(\theta) \right\} + \frac{q(q-1)(q-2)}{2} \mathrm{E}_\theta \left\{ U_{\mathrm{P}st}(\theta) U_{\mathrm{I}v}(\theta) \right\}.$$

Finally,

$$\mathrm{Var}_\theta \left\{ \sum_{r=1}^{q} U_r^{\mathrm{I}}(\theta) \right\} = q\mathrm{E}_\theta\{U_{\mathrm{I}s}(\theta)\} + q(q-1)\mathrm{E}_\theta\{U_{\mathrm{I}s}(\theta)U_{\mathrm{I}t}(\theta)\}.$$

Going back to (3.7), we can conclude that the leading term in q is given by

$$\mathrm{E}_\theta\{U_{\mathrm{P}st}(\theta)U_{\mathrm{P}vw}(\theta)\} + a\mathrm{E}_\theta \left\{ U_{\mathrm{P}st}(\theta) U_{\mathrm{I}v}(\theta) \right\} + a^2 \mathrm{E}_\theta\{U_{\mathrm{I}s}(\theta)U_{\mathrm{I}t}(\theta)\}\cdot$$

This means that both the first and second quantities in curly brackets in (3.6) have the same order in probability, that is $O_P(1)$. As a result, it turns out that $(\hat{\theta}_C^a - \theta) = O_P(1)$. This result suggests that the estimating equation will not usually lead to a consistent estimator of θ, which is something to be expected even for ordinary (composite) likelihoods.

3.3 Examples

In this section we highlight properties of estimators based on combined composite likelihood, comparing results with those of estimators based on full likelihood. In particular, the two examples considered deal with multivariate normal distributions with structured covariance matrix. The first one has a scalar parameter while the second one has a multidimensional parameter. In both examples, situations in which a is fixed and q changes and viceversa, are considered.

3.3.1 Common partial correlation model

This example has been suggested in Lindsay et al. (2011). Assume we have n i.i.d. normal variables where each realization $Y_i = (Y_{i1}, \ldots, Y_{iq})^{\mathrm{T}}$, has mean zero and covariance matrix Σ, i.e $Y_i \sim N_q(0, \Sigma)$, where the inverse of Σ is given by

$$\Sigma^{-1} = (1-\beta)\mathrm{I}_q + \beta 1_q 1_q^{\mathrm{T}},$$

where, I_q denotes the q-dimensional identity matrix and 1_q the q-dimensional vector with all elements equal to one. Denoting by Σ_{rs} the element with position (r, s) of the matrix Σ, we have

$$\Sigma_{rs} = \frac{1}{1-\beta}\left[\delta_{rs} - \frac{\beta}{1+(q-1)\beta}\right],$$

where $\delta_{rs} = 1$ when $r = s$ and 0 otherwise. Therefore, the variance of each component of the vector is given by

$$\mathrm{Var}_\beta(Y_{ir}) = \frac{1+\beta(q-2)}{(1-\beta)\{1+\beta(q-1)\}},$$

and the correlation between any two components of the same vector is

$$\mathrm{Cor}_\beta(Y_{ir}, Y_{is}) = -\frac{\beta}{1+\beta(q-2)}.$$

Moreover, the determinant of Σ is given by

$$|\Sigma| = \frac{1}{(1-\beta)^{q-1}\{1+\beta(q-1)\}},$$

and is positive for $\beta > -\frac{1}{q-1}$, which is a necessary condition for Σ to be a covariance matrix.

Full likelihood

Considering a single observation, the density function of the q-dimensional normal distribution is

$$f_{Y_i}(y_i; \beta) = \frac{1}{(2\pi)^{\frac{q}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} y_i^{\mathrm{T}} \Sigma^{-1} y_i\right).$$

Since $y_i^{\mathrm{T}}\Sigma^{-1}y_i = (1-\beta)\sum_{r=1}^q y_{ir}^2 + \beta\left(\sum_{r=1}^q y_{ir}\right)^2$, the likelihood calculated for a single observation is

$$L(\beta) \propto (1-\beta)^{\frac{q-1}{2}}\{1+\beta(q-1)\}^{\frac{1}{2}} \exp\left\{-\frac{1}{2}(1-\beta)\sum_{r=1}^q y_{ir}^2 - \frac{\beta}{2}\left(\sum_{r=1}^q y_{ir}\right)^2\right\}.$$

Therefore, the log likelihood based on n independent observations is given by

$$\ell(\beta) = \frac{n(q-1)}{2}\log(1-\beta) + \frac{n}{2}\log\{1+\beta(q-1)\} - \frac{(1-\beta)}{2}\sum_{i=1}^n\sum_{r=1}^q y_{ir}^2 \\ - \frac{\beta}{2}\sum_{i=1}^n\left(\sum_{r=1}^q y_{ir}\right)^2.$$

Since $\sum_{r=1}^q y_{ir}^2 = \sum_{r=1}^q (y_{ir}-\bar{y}_i)^2 + q\bar{y}_i^2 = \sum_{r=1}^q (y_{ir}-\bar{y}_i)^2 + \left(\sum_{r=1}^q y_{ir}\right)^2/q$, we can also write

$$\begin{aligned}\ell(\beta) &= \frac{n(q-1)}{2}\log(1-\beta) + \frac{n}{2}\log\{1+\beta(q-1)\} - \frac{(1-\beta)}{2}\times \\ &\quad \sum_{i=1}^n\sum_{r=1}^q (y_{ir}-\bar{y}_i)^2 - \frac{\{1+\beta(q-1)\}}{2q}\sum_{i=1}^n\left(\sum_{r=1}^q y_{ir}\right)^2 \\ &= \frac{n(q-1)}{2}\log(1-\beta) + \frac{n}{2}\log\{1+\beta(q-1)\} + \beta\left(\frac{1}{2}\sum_{i=1}^n\sum_{r=1}^q (y_{ir}-\bar{y}_i)^2\right. \\ &\quad \left. -\frac{(q-1)}{2q}\sum_{i=1}^n\left(\sum_{r=1}^q y_{ir}\right)^2\right) \\ &= \frac{n(q-1)}{2}\log(1-\beta) + \frac{n}{2}\log\{1+\beta(q-1)\} + \beta\left(\frac{1}{2}SSW\right. \\ &\quad \left. -\frac{(q-1)}{2q}SSR\right),\end{aligned}$$

where $SSW = \sum_{i=1}^n\sum_{r=1}^q (y_{ir}-\bar{y}_i)^2$ and $SSR = \sum_{i=1}^n y_{i\cdot}^2$, with $y_{i\cdot} = \sum_{r=1}^q y_{ir}$. From the expression of $\ell(\theta)$, we see that the full likelihood function belongs to a one-parameter exponential family with sufficient statistic given by $SSW/2 - (q-1)SSR/(2q)$ and canonical parameter β.

For later use, we study the distribution of SSW and SSR.
We can rewrite $SSW = \sum_{i=1}^{n}\sum_{r=1}^{q}(Y_{ir}-\overline{Y}_i)^2 = \sum_{i=1}^{n} Q_i$, where

$$\begin{aligned} Q_i &= \sum_{r=1}^{q}(Y_{ir}-\overline{Y}_i)^2 \\ &= \sum_{r=1}^{q} Y_{ir}^2 - q\overline{Y}_i^2 \\ &= \sum_{r=1}^{q} Y_{ir}^2 - \frac{1}{q}\sum_{r=1}^{q} Y_{ir}\sum_{r=1}^{q} Y_{ir} \\ &= Y_i^{\mathrm{T}}\left(\mathrm{I}_q - \frac{1}{q}1_q 1_q^{\mathrm{T}}\right)Y_i. \end{aligned}$$

In order to determine the distribution of SSW, we define $X_i = (1-\beta)^{\frac{1}{2}}Y_i$. Therefore, $X_i \sim N_q(0,R)$, where $R = \mathrm{I}_q - \frac{\beta}{1+\beta(q-1)}1_q 1_q^{\mathrm{T}}$, and consequently,

$$Q_i = Y_i^{\mathrm{T}}\left(\mathrm{I}_q - \frac{1}{q}1_q 1_q^{\mathrm{T}}\right)Y_i = (1-\beta)^{-1}X_i^{\mathrm{T}}\left(\mathrm{I}_q - \frac{1}{q}1_q 1_q^{\mathrm{T}}\right)X_i.$$

Noting that the matrix $R\left(\mathrm{I}_q - \frac{1}{q}1_q 1_q^{\mathrm{T}}\right)$ is idempotent, we can conclude, according to Theorem 8.6 of Severini (2005, p. 246) that

$$X_i^{\mathrm{T}}\left(\mathrm{I}_q - \frac{1}{q}1_q 1_q^{\mathrm{T}}\right)X_i \sim \chi^2_m,$$

where $m = \operatorname{tr}\left\{R\left(\mathrm{I}_q - \frac{1}{q}1_q 1_q^{\mathrm{T}}\right)\right\} = \operatorname{tr}\left\{\left(\mathrm{I}_q - \frac{1}{q}1_q 1_q^{\mathrm{T}}\right)\right\} = q-1$. Therefore, exploiting the fact that the observations are independent, we get that

$$SSW \sim \frac{1}{1-\beta}\chi^2_{n(q-1)}.$$

Analogously, we have

$$\begin{aligned} \sum_{r=1}^{q} Y_{ir} = Y_{i\cdot} = 1_q^{\mathrm{T}}Y_i \sim N(0, 1_q^{\mathrm{T}}\Sigma 1_q) \\ \sim N\left(0, \frac{q}{1+\beta(q-1)}\right), \end{aligned}$$

and therefore

$$SSR = \sum_{i=1}^{n} Y_{i\cdot}^2 = \sum_{i=1}^{n}\left[\left\{\frac{q}{1+\beta(q-1)}\right\}^{\frac{1}{2}} Z_i\right]^2 = \left\{\frac{q}{1+\beta(q-1)}\right\}\sum_{i=1}^{n} Z_i^2.$$

This implies, due to the independence of the observations, that

$$SSR \sim \frac{q}{1+\beta(q-1)}\chi^2_n.$$

Taking into account the fact that SSW is a function of $Y_i^{\mathrm{T}}\left(\mathrm{I}_q - \frac{1}{q}1_q1_q^{\mathrm{T}}\right)Y_i$ and SSR a function of $Y_{i\cdot}$, the variables SSW and SSR will be independent if and only if $Y_i^{\mathrm{T}}\left(\mathrm{I}_q - \frac{1}{q}1_q1_q^{\mathrm{T}}\right)Y_i$ and $Y_{i\cdot}$ are independent. Defining $A = \left(\mathrm{I}_q - \frac{1}{q}1_q1_q^{\mathrm{T}}\right)$, through result (viii) of Rao (1973, p. 188), for which the necessary and sufficient condition for independence is that

$$\Sigma A \Sigma 1_q = 0,$$

we have

$$\begin{aligned}
\Sigma A\Sigma 1_q &= \frac{1}{(1-\beta)^2}\left(\mathrm{I}_q - \frac{\beta}{1+\beta(q-1)}1_q1_q^{\mathrm{T}}\right)\left(\mathrm{I}_q - \frac{1}{q}1_q1_q^{\mathrm{T}}\right)\times \\
&\quad \left(\mathrm{I}_q - \frac{\beta}{1+\beta(q-1)}1_q1_q^{\mathrm{T}}\right)1_q \\
&= \frac{1}{(1-\beta)^2}\left(\mathrm{I}_q - \frac{\beta}{1+\beta(q-1)}1_q1_q^{\mathrm{T}}\right)\left(\mathrm{I}_q - \frac{1}{q}1_q1_q^{\mathrm{T}}\right)1_q \\
&= \frac{1}{(1-\beta)^2}\left(\mathrm{I}_q - \frac{1}{q}1_q1_q^{\mathrm{T}}\right)1_q \\
&= \frac{1}{(1-\beta)^2}\left(1_q - \frac{1}{q}q1_q\right) = 0.
\end{aligned}$$

This implies that SSW and SSR are independent.

The score function is given by

$$U(\beta) = -\frac{n(q-1)}{2(1-\beta)} + \frac{n(q-1)}{2\{1+\beta(q-1)\}} + \frac{1}{2}SSW - \frac{q-1}{2q}SSR.$$

The observed information is

$$\begin{aligned}
j(\beta) &= -\frac{d}{d\beta}U(\beta) \\
&= \frac{n(q-1)}{2(1-\beta)^2} + \frac{n(q-1)^2}{2\{1+\beta(q-1)\}^2} \\
&= \frac{nq(q-1)\{1+\beta^2(q-1)\}}{2(1-\beta)^2\{1+\beta(q-1)\}^2}.
\end{aligned}$$

Since the observed information is a constant, we also have that $i(\beta) = \mathrm{E}_\beta\{j(\beta)\} = j(\beta)$.

Combined composite likelihood

We now consider a combined composite likelihood, which might be appropriate since both univariate and bivariate marginal densities depend on the parameter of interest. The pairwise likelihood with all weights equal to 1 is based on the bivariate marginal distributions, that is

$$Y_i = \begin{pmatrix} Y_{ir} \\ Y_{is} \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \frac{1+\beta(q-2)}{(1-\beta)\{1+\beta(q-1)\}} & -\frac{\beta}{(1-\beta)\{1+\beta(q-1)\}} \\ -\frac{\beta}{(1-\beta)\{1+\beta(q-1)\}} & \frac{1+\beta(q-2)}{(1-\beta)\{1+\beta(q-1)\}} \end{bmatrix}\right).$$

Let us define

$$\Sigma_2 = \frac{1}{(1-\beta)\{1+\beta(q-1)\}} \begin{pmatrix} 1+\beta(q-2) & -\beta \\ -\beta & 1+\beta(q-2) \end{pmatrix}.$$

The pairwise likelihood corresponding to only one observation is given by

$$\begin{aligned} CL^{\mathrm{P}}(\beta) &\propto \prod_{r=1}^{q-1} \prod_{s=r+1}^{q} |\Sigma_2|^{-\frac{1}{2}} \exp\{-\frac{1}{2} y_i^{\mathrm{T}} \Sigma_2^{-1} y_i\} \\ &\propto \prod_{r=1}^{q-1} \prod_{s=r+1}^{q} \left[\frac{1+\beta(q-3)}{(1-\beta)^2\{1+\beta(q-1)\}}\right]^{-\frac{1}{2}} \exp\left\{-\frac{(1-\beta)\{1+\beta(q-2)\}}{2\{1+\beta(q-3)\}} \times \right. \\ &\left. \sum_{r=1}^{q-1} \sum_{s=r+1}^{q} (y_{ir}^2 + y_{is}^2) - \frac{\beta(1-\beta)}{2\{1+\beta(q-3)\}} \sum_{r=1}^{q-1} \sum_{s=r+1}^{q} 2 y_{ir} y_{is} \right\}. \end{aligned}$$

The corresponding pairwise log-likelihood calculated for n independent observations is

$$\begin{aligned} c\ell^{\mathrm{P}}(\beta) = -\frac{nq(q-1)}{4} \log\left[\frac{1+\beta(q-3)}{(1-\beta)^2\{1+\beta(q-1)\}}\right] - \frac{(1-\beta)\{1+\beta(q-2)\}}{2\{1+\beta(q-3)\}} \times \\ \sum_{i=1}^{n} \sum_{r=1}^{q-1} \sum_{s=r+1}^{q} (y_{ir}^2 + y_{is}^2) - \frac{\beta(1-\beta)}{2\{1+\beta(q-3)\}} \sum_{i=1}^{n} \sum_{r=1}^{q-1} \sum_{s=r+1}^{q} 2 y_{ir} y_{is}. \end{aligned}$$

Noting that

$$\begin{aligned}\sum_{r=1}^{q-1}\sum_{s=r+1}^{q}(y_{ir}^2+y_{is}^2) &= \sum_{r=1}^{q-1}\sum_{s=r+1}^{q}y_{ir}^2+\sum_{r=1}^{q-1}\sum_{s=r+1}^{q}y_{is}^2\\ &= \sum_{r=1}^{q}(q-r)y_{ir}^2+\sum_{r=1}^{q}(r-1)y_{ir}^2\\ &= (q-1)\sum_{r=1}^{q}y_{ir}^2\end{aligned}$$

and $\sum_{r=1}^{q-1}\sum_{s=r+1}^{q}2y_{ir}y_{is}=\left(\sum_{r=1}^{q}y_{ir}\right)^2-\sum_{r=1}^{q}y_{ir}^2$, the pairwise log-likelihood, after some simplifications, can be rewritten as

$$\begin{aligned}c\ell^{\mathrm{P}}(\beta) = &-\frac{nq(q-1)}{4}\log\left[\frac{1+\beta(q-3)}{(1-\beta)^2\{1+\beta(q-1)\}}\right]\\ &-\frac{(1-\beta)\{q-1+\beta(q^2-2q+1)\}}{2q\{1+\beta(q-3)\}}SSR\\ &-\frac{(1-\beta)\{q-1+\beta(q^2-3q+1)\}}{q\{1+\beta(q-3)\}}SSW\cdot\end{aligned}$$

Therefore, the pairwise likelihood function belongs to a curved exponential family where SSW and SSR are components of the sufficient statistic.

We need the expectation and variance of the variables SSW and SSR to calculate quantities related to the pairwise likelihood. The results are

$$\begin{aligned}\mathrm{E}_\beta(SSW) &= \frac{n(q-1)}{1-\beta}\\ \mathrm{E}_\beta(SSR) &= \frac{nq}{1+\beta(q-1)}\\ \mathrm{Var}_\beta(SSW) &= \frac{2n(q-1)}{(1-\beta)^2}\end{aligned}$$

$$\mathrm{Var}_\beta(SSR) = \frac{2nq^2}{\{1+\beta(q-1)\}^2}.$$

The score based on the pairwise likelihood is

$$
\begin{aligned}
cU^{\mathrm{P}}(\beta) &= \frac{\mathrm{d}}{\mathrm{d}\beta} c\ell^{\mathrm{P}}(\beta) \\
&= -\frac{nq(q-1)\{2\beta(2q-3)+2\beta^2(q^2-4q+3)\}}{4(1-\beta)\{1+\beta(q-1)\}\{1+\beta(q-3)\}} \\
&\quad + \frac{(q-1)\{\beta^2(q^2-4q+3)+2\beta(q-1)-1\}}{2q\{1+\beta(q-3)\}^2} SSR \\
&\quad + \frac{\{\beta^2(q^3-6q^2+10q-3)+\beta(2q^2-6q+2)+1\}}{2\{1+\beta(q-3)\}^2} SSW\cdot
\end{aligned}
$$

The quantities required for the calculation of the Godambe information are

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\beta} cU^{\mathrm{P}}(\beta) &= -\frac{nq(q-1)g_1(\beta,q)}{2(1-\beta)^2\{1+\beta(q-3)\}^2\{1+\beta(q-1)\}^2} + \frac{2(q-2)(q-1)}{\{1+\beta(q-3)\}^3} SSR \\
&\quad + \frac{(q-2)^2}{\{1+\beta(q-3)\}^3} SSW,
\end{aligned}
$$

where

$$
\begin{aligned}
g_1(\beta,q) &= \beta^4(q^4-8q^3+22q^2-24q+9)+\beta^3(4q^3-22q^2+36q-18) \\
&\quad + \beta^2(2q^2-6q+6)+\beta(2q^2-8q+6)+2q-3.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
H^{\mathrm{P}}(\beta) &= \mathrm{E}_\beta\left\{-\frac{\mathrm{d}}{\mathrm{d}\beta} cU^{\mathrm{P}}(\beta)\right\} \\
&= \frac{nq(q-1)g_1(\beta,q)}{2(1-\beta)^2\{1+\beta(q-3)\}^2\{1+\beta(q-1)\}^2} \\
&\quad - \frac{2(q-2)(q-1)}{\{1+\beta(q-3)\}^3}\mathrm{E}_\beta(SSR) - \frac{(q-2)^2}{\{1+\beta(q-3)\}^3}\mathrm{E}_\beta(SSW) \\
&= \frac{nq(q-1)g_1(\beta,q)}{2(1-\beta)^2\{1+\beta(q-3)\}^2\{1+\beta(q-1)\}^2} \\
&\quad - \frac{2nq(q-2)(q-1)}{\{1+\beta(q-1)\}\{1+\beta(q-3)\}^3} - \frac{n(q-1)(q-2)^2}{(1-\beta)\{1+\beta(q-3)\}^3} \\
&= \frac{nq(q-1)g_2(\beta,q)}{2(1-\beta)^2\{1+\beta(q-3)\}^2\{1+\beta(q-1)\}^2},
\end{aligned}
$$

where

$$
\begin{aligned}
g_2(\beta,q) &= \beta^4(q^4-8q^3+22q^2-24q+9)+\beta^3(4q^3-22q^2+36q-18) \\
&\quad + \beta^2(4q^2-12q+10)-2\beta+1.
\end{aligned}
$$

Finally we also have

$$
\begin{aligned}
J^{\mathrm{P}}(\beta) &= \mathrm{Var}_\beta\{cU^{\mathrm{P}}(\beta)\} \\
&= \left[\frac{(q-1)\{\beta^2(q^2-4q+3)+2\beta(q-1)-1\}}{2q\{1+\beta(q-3)\}^2}\right]^2 \mathrm{Var}_\beta(SSR) \\
&\quad + \left[\frac{\{\beta^2(q^3-6q^2+10q-3)+\beta(2q^2-6q+2)+1\}}{2\{1+\beta(q-3)\}^2}\right]^2 \mathrm{Var}_\beta(SSW) \\
&= \frac{n(q-1)^2\{\beta^2(q^2-4q+3)+2\beta(q-1)-1\}^2}{2\{1+\beta(q-3)\}^4\{1+\beta(q-1)\}^2} \\
&\quad + \frac{n(q-1)g_3(\beta,q)}{2\{1+\beta(q-3)\}^4(1-\beta)^2},
\end{aligned}
$$

where, $g_3(\beta,q) = \{\beta^2(q^3-6q^2+10q-3)+\beta(2q^2-6q+2)+1\}^2$.

The independence likelihood, with all weights equal to 1, is based on the univariate marginal distributions, that is

$$
Y_{ir} \sim N\left(0, \frac{1+\beta(q-2)}{(1-\beta)\{1+\beta(q-1)\}}\right) \quad r=1,\ldots,q.
$$

The independence likelihood for one observation is

$$
\begin{aligned}
CL^{\mathrm{I}}(\beta) &= \prod_{r=1}^{q} f(y_{ir};\beta) \\
&\propto \left[\frac{1+\beta(q-2)}{(1-\beta)\{1+\beta(q-1)\}}\right]^{-\frac{q}{2}} \exp\left\{-\frac{(1-\beta)\{1+\beta(q-1)\}}{2\{1+\beta(q-2)\}}\sum_{r=1}^{q} y_{ir}^2\right\}.
\end{aligned}
$$

Hence, the independence log likelihood calculated for n independent observations is

$$
c\ell^{\mathrm{I}}(\beta) = -\frac{nq}{2}\log\left[\frac{1+\beta(q-2)}{(1-\beta)\{1+\beta(q-1)\}}\right] - \frac{(1-\beta)\{1+\beta(q-1)\}}{2\{1+\beta(q-2)\}}\sum_{i=1}^{n}\sum_{r=1}^{q} y_{ir}^2.
$$

As with the pairwise, substituting $\sum_{i=1}^{n}\sum_{r=1}^{q} y_{ir}^2 = SSW + \frac{1}{q}SSR$, we finally get

$$
\begin{aligned}
c\ell^{\mathrm{I}}(\beta) = &-\frac{nq}{2}\log\left[\frac{1+\beta(q-2)}{(1-\beta)\{1+\beta(q-1)\}}\right] - \frac{(1-\beta)\{1+\beta(q-1)\}}{2\{1+\beta(q-2)\}}SSW \\
&- \frac{(1-\beta)\{1+\beta(q-1)\}}{2q\{1+\beta(q-2)\}}SSR.
\end{aligned}
$$

Even the independence likelihood belongs to a curved exponential family with the components of the sufficient statistic given by SSW and $SSR.$ Therefore, the combined composite likelihood is also a curved exponential family.

Now, we calculate some quantities of the interest related to the independence likelihood. We start with the score function and the second derivative of $c\ell^{\mathrm{I}}(\beta)$ which are given by

$$\begin{aligned} cU^{\mathrm{I}}(\beta) &= \frac{\mathrm{d}}{\mathrm{d}\beta}c\ell^{\mathrm{I}}(\beta) \\ &= -\frac{nq\beta(q-1)\{2+\beta(q-2)\}}{2(1-\beta)\{1+\beta(q-2)\}\{1+\beta(q-1)\}} \\ &\quad + \frac{\beta(q-1)\{2+\beta(q-2)\}}{2\{1+\beta(q-2)\}^2}SSW + \frac{\beta(q-1)\{2+\beta(q-2)\}}{2q\{1+\beta(q-2)\}^2}SSR\cdot \end{aligned}$$

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\beta}cU^{\mathrm{I}}(\beta) &= -\frac{nq(q-1)g_4(\beta,q)}{2(1-\beta)^2\{1+\beta(q-2)\}^2\{1+\beta(q-1)\}^2} + \frac{(q-1)}{\{1+\beta(q-2)\}^3}SSW \\ &\quad + \frac{(q-1)}{q\{1+\beta(q-2)\}^3}SSR, \end{aligned}$$

where $g_4(\beta,q) = \{\beta^4(q^3-5q^2+8q-4)+\beta^3(4q^2-12q+8)+\beta^2(2q-2)+\beta(2q-4)+2\}$. Therefore

$$\begin{aligned} H^{\mathrm{I}}(\beta) &= \mathrm{E}_\beta\left\{-\frac{\mathrm{d}}{\mathrm{d}\beta}c\ell^{\mathrm{I}}(\beta)\right\} \\ &= \frac{nq(q-1)g_4(\beta,q)}{2(1-\beta)^2\{1+\beta(q-2)\}^2\{1+\beta(q-1)\}^2} \\ &\quad - \frac{(q-1)}{\{1+\beta(q-2)\}^3}\mathrm{E}_\beta(SSW) - \frac{(q-1)}{q\{1+\beta(q-2)\}^3}\mathrm{E}_\beta(SSR) \\ &= \frac{nq\beta^2(q-1)^2\{2+\beta(q-2)\}^2}{2(1-\beta)^2\{1+\beta(q-2)\}^2\{1+\beta(q-1)\}^2}. \end{aligned}$$

Now, let us calculate the Godambe information as defined in Subsection 3.2.2. In order to simplify the expression of $J^a(\beta) = \mathrm{Var}_\beta\{cU^a(\beta)\}$, where $cU^a(\beta) = \mathrm{d}c\ell^a(\beta)/\mathrm{d}\beta$, we can rewrite $cU^{\mathrm{P}}(\beta)$ and $cU^{\mathrm{I}}(\beta)$ as follows

$$cU^{\mathrm{P}}(\beta) = k_1 + k_2SSR + k_3SSW,\ \ cU^{\mathrm{I}}(\beta) = e_1 + e_2SSR + e_3SSW,$$

where the coefficients $k_1, k_2, k_3, e_1, e_2, e_3$ are given by

$$
\begin{aligned}
k_1 &= -\frac{nq(q-1)\{2\beta(2q-3)+2\beta^2(q^2-4q+3)\}}{4(1-\beta)\{1+\beta(q-1)\}\{1+\beta(q-3)\}} \\
k_2 &= \frac{(q-1)\{\beta^2(q^2-4q+3)+2\beta(q-1)-1\}}{2q\{1+\beta(q-3)\}^2} \\
k_3 &= \frac{\{\beta^2(q^3-6q^2+10q-3)+\beta(2q^2-6q+2)+1\}}{2\{1+\beta(q-3)\}^2} \\
e_1 &= -\frac{nq\beta(q-1)\{2+\beta(q-2)\}}{2(1-\beta)\{1+\beta(q-2)\}\{1+\beta(q-1)\}} \\
e_2 &= \frac{\beta(q-1)\{2+\beta(q-2)\}}{2q\{1+\beta(q-2)\}^2} \\
e_3 &= \frac{\beta(q-1)\{2+\beta(q-2)\}}{2\{1+\beta(q-2)\}^2}.
\end{aligned}
$$

Therefore, we have that

$$
\begin{aligned}
cU^a(\beta) &= 2cU^{\mathrm{P}}(\beta) - a(q-1)cU^{\mathrm{I}}(\beta) \\
&= 2(k_1 + k_2 SSR + k_3 SSW) - a(q-1)\{e_1 + e_2 SSR + e_3 SSW\} \\
&= 2k_1 - a(q-1)e_1 + \{2k_2 - a(q-1)e_2\}SSR + \{2k_3 - a(q-1)e_3\}SSW,
\end{aligned}
$$

and it follows that

$$
\begin{aligned}
J^a(\beta) &= \{2k_2 - a(q-1)e_2\}^2 \mathrm{Var}_\beta(SSR) + \{2k_3 - a(q-1)e_3\}^2 \mathrm{Var}_\beta(SSW) \\
H^a(\beta) &= 2H^{\mathrm{P}}(\beta) - a(q-1)H^{\mathrm{I}}(\beta).
\end{aligned}
$$

Using these two quantities, we get the Godambe information

$$
G^a(\beta) = H^a(\beta)J^a(\beta)^{-1}H^a(\beta) = \frac{H^a(\beta)^2}{J^a(\beta)}.
$$

Now, we consider the problem of finding the admissible values for the constant a for which the combined composite likelihood satisfies the properties given in Section 2.2. In the present context, condition (3.1) leads to the following constraint

$$
a \le A_q(\beta) = \frac{2\{1+\beta(q-2)\}^2 c(q,\beta)}{\beta^2(q-1)^2\{1+\beta(q-3)\}^2\{2+\beta(q-2)\}^2},
$$

where,

$$c(q,\beta) = \beta^4(q^4 - 8q^3 + 22q^2 - 24q + 9) + \beta^3(4q^3 - 22q^2 + 36q - 18) + \beta^2(4q^2 - 12q + 10) - 2\beta + 1.$$

We note here that the threshold $A_q(\beta)$ depends on both β and q. Since β is not known in advance, the aim is to determine a value for the threshold which does not depend on the parameter β. On the other hand, once faced with a real problem, the dependence on q is no longer an issue.

Thus, a conservative choice would be to find the minimum value of $A_q(\beta)$ with respect to β, in the range $(-\frac{1}{q-1}, 1)$. We note in this example that q is generic. Hence, we wish to find an upper bound for the constant a which is a valid for any q. To this end, we will study graphically the behavior of $\min_\beta A_q(\beta)$ with increasing q. Figure 3.1 displays the behavior of $\min_\beta A_q(\beta)$ as a function of q and therefore, the plot leads us to the conclusion that the upper bound for the constant a corresponds to the lower bound of $\min_\beta A_q(\beta)$, which is equal to 1 in this case. In other words, the constant a assumes values less than or equal to 1 independently of q. As a result, for $a \leq 1$, the combined composite likelihood for this particular model satisfies the requirement of being a sensible pseudo-likelihood.

As an illustration, if we consider the values for the constant a which are not admissible, that is, values greater than 1, the combined composite likelihood estimator may lose some of its fundamental properties and hence, may not be suitable for inference on β. Figure 3.2 displays plots of the combined composite likelihood in a simulated dataset for four different values of a and highlights the cases in which the combined composite likelihood is not useful.

We now perform a numerical assessment of consistency of the combined composite likelihood estimator. Here, we only focus attention on the case where n is fixed and q increases. In the opposite case, i.e. q fixed and n increases, we have shown in Subsection 2.2.2 that the consistency of $\hat{\beta}^a$ is guaranteed. To this end, we ran a simulation experiment, with $n = 1$, $q = 3, 10, 10, 1000, 5000$ and $a \in \{-10, -5, -1, 0, 0.5, 1, 2, 3, 5\}$. For

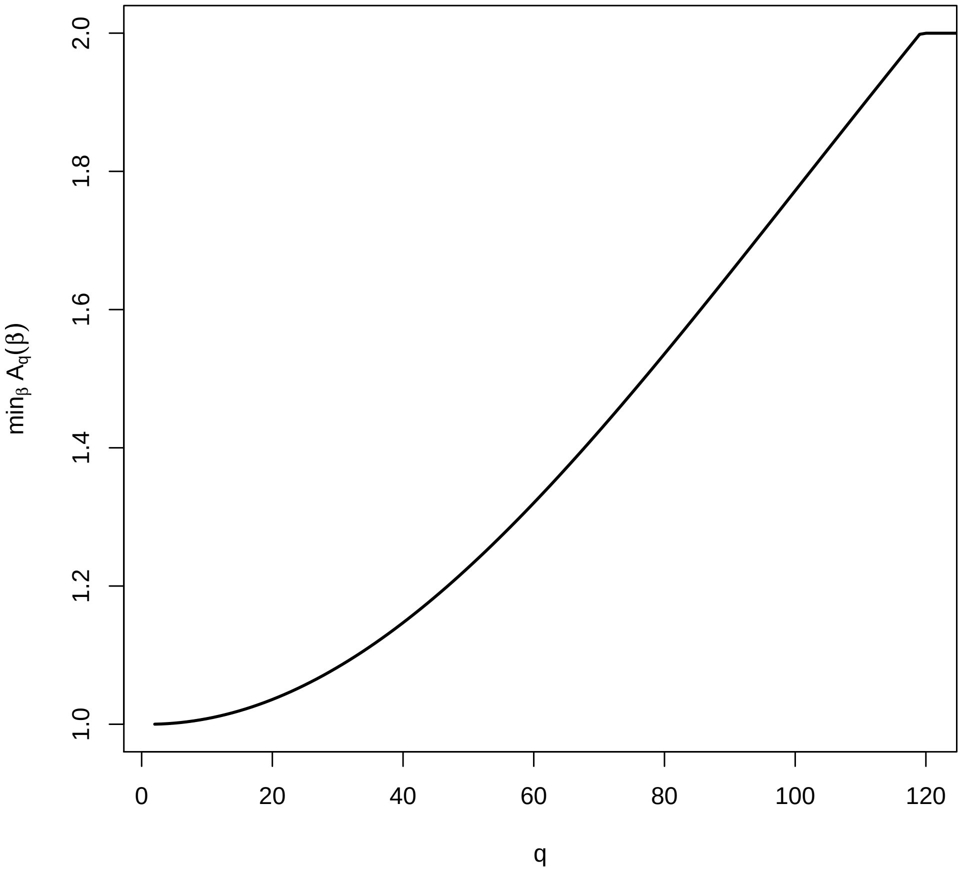

Figure 3.1: Common partial correlation model. Behavior of $\min_{\beta} A_q(\beta)$ as a function of q.

each combination, we performed 2000 iterations. The true value of β was fixed to 0.7. Table 3.1 reports the mean squared errors (MSE) and as we see, for the admissible values of a, they decrease toward zero as q increases. This seems to suggest, although empirically, consistency of $\hat{\beta}^a$. Instead, as expected, for values of a greater than one, the MSE of the estimator does not decrease to zero.

We now compare in terms of efficiency the estimator based on full likelihood with the one based on the combined composite likelihood. This

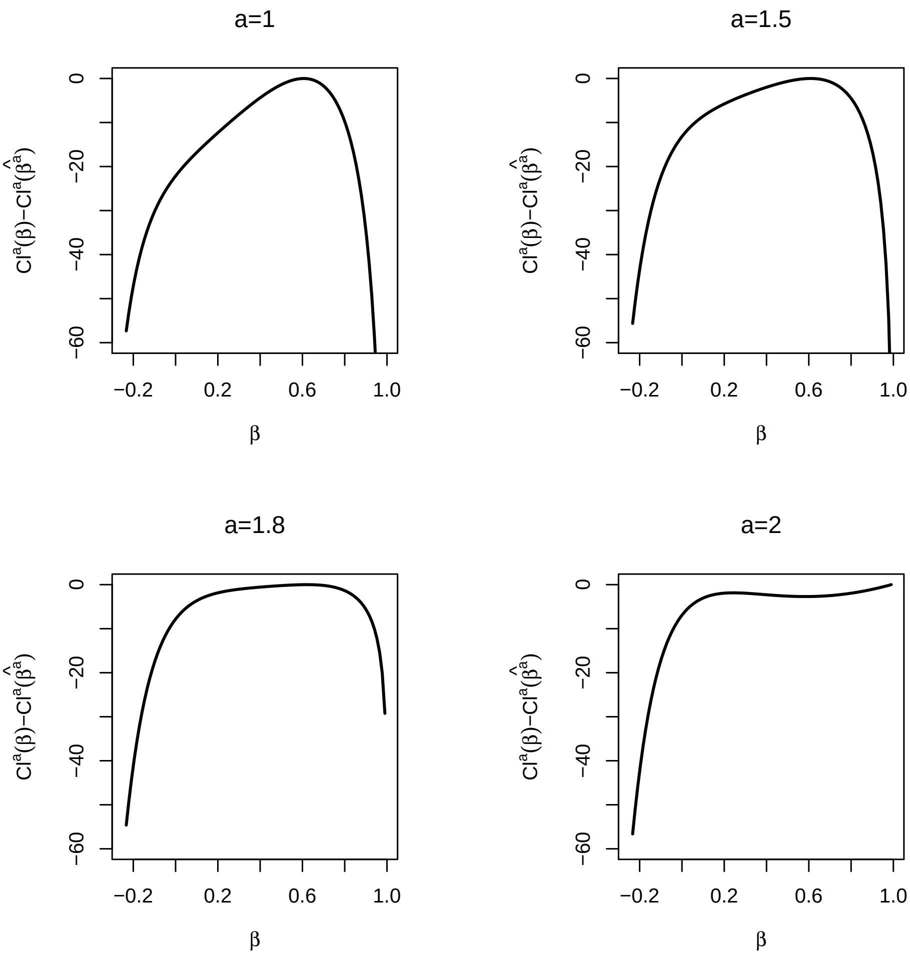

Figure 3.2: Common partial correlation model. Combined composite likelihood for β, for a simulated sample with $n = 10, q = 4$ and true parameter $\beta_0 = 0.7$, in cases $a = 1, 1.5, 1.82, 2$.

comparison is based on the asymptotic relative efficiency defined in (2.8). Figure 3.3 displays the relative efficiencies in different situations. From the plots, we note that for fixed values of a, as q increases, there is a loss of efficiency of $\hat{\beta}^a$ with respect to $\hat{\beta}$. This result is more or less expected when one uses a pseudo-likelihood in place of the full likelihood. Instead, the plots in Figure 3.4 seem to suggest that for fixed values of q the efficiency improves for increasing values of a although, for large values of q,

Table 3.1: Common partial correlation model. Mean squared error of $\hat{\beta}^a$, when $n = 1, a = -10, -5, -1, 0, 0.5, 1, 2, 3, 5$ and $q = 3, 10, 100, 1000, 5000$. Results are obtained with 2000 simulated samples, and $\beta_0 = 0.7$.

		a								
		-10	-5	-1	0	0.5	1	2	3	5
q	3	0.18119	0.17503	0.17326	0.16978	0.16759	0.17749	0.20214	0.24283	0.27699
	10	0.04473	0.04733	0.04584	0.04642	0.04271	0.04265	0.30427	0.40254	0.41615
	100	0.00205	0.00204	0.0019	0.00198	0.00206	0.002	0.45859	0.48957	0.48991
	1000	0.00017	0.00018	0.00017	0.00017	0.00018	0.00018	0.47504	0.48991	0.48991
	5000	0.000035	0.000034	0.000035	0.000035	0.000036	0.000036	0.47800	0.48879	0.48879

different choices of a seem to lead to the same results. See also Table 3.1. Therefore, $a = 1$ seems to be the best choice among the admissible values. In other words, for this example the combined composite likelihood which works better is the pairwise conditional likelihood.

3.3.2 A model for microarray data

Let us consider here a model for a microarray data. An application of this particular model is suggested in Roverato & Di Lascio (2011). As in the previous example, the aim is to compare the properties of the full maximum likelihood estimator with the one based on combined composite likelihood, focusing the attention on the loss of efficiency. For the goals of this study, we do not take into account the biological aspects.

Here, $X_V = (X_1, \ldots, X_p, X_{tf})$ is a vector of random variables and it is assumed to have a multivariate normal distribution. In addition, $X_1, \ldots, X_p$ are (mutually) conditionally independent given X_{tf}. In particular, the model

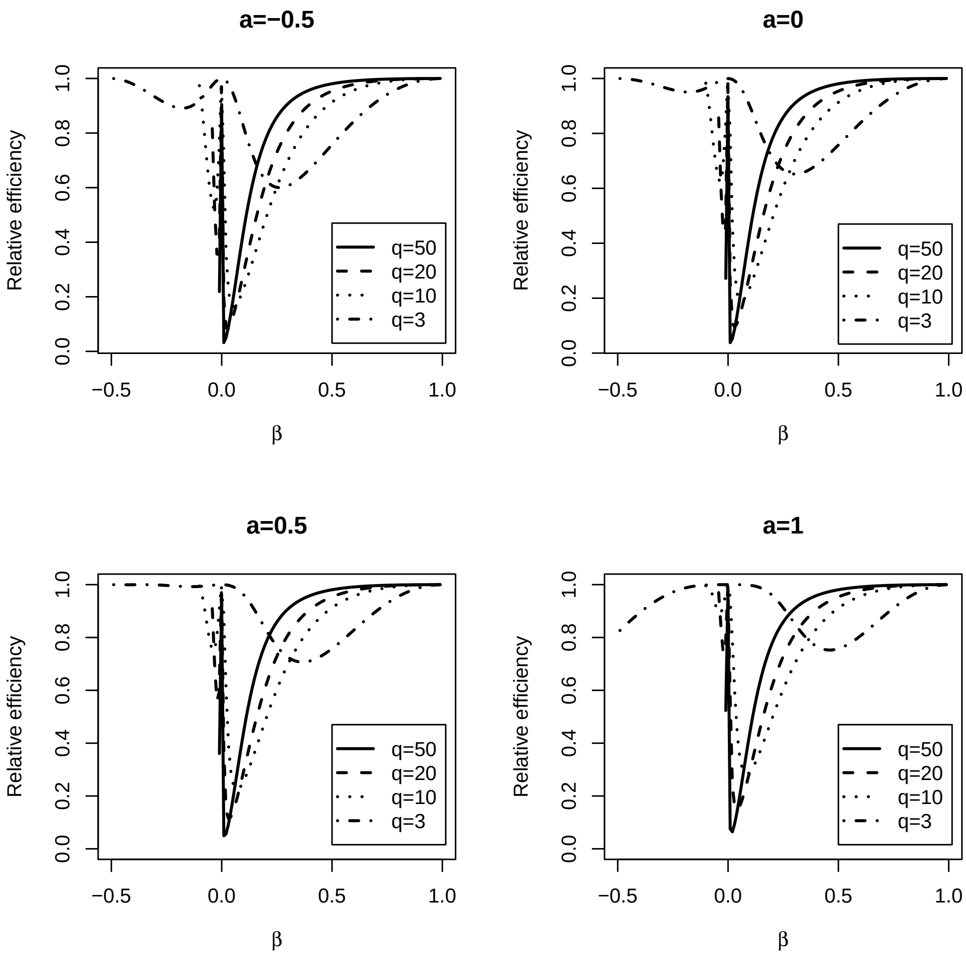

Figure 3.3: Common partial correlation model. Comparisons of the relative efficiency for fixed a and different values of q.

is characterized as follows:

$$\begin{aligned}
\mathrm{E}(X_{tf}) &= \mu_{tf} \\
\mathrm{Var}(X_{tf}) &= \sigma^2_{tf} \\
\mathrm{E}(X_r \mid X_{tf} = x_{tf}) &= \beta_{0,r} + \beta_{1,r} x_{tf}, \ r = 1, \ldots, p, \\
\mathrm{Var}(X_r \mid X_{tf} = x_{tf}) &= \sigma^2_r, \ r = 1, \ldots, p,
\end{aligned}$$

where $\mu_{tf}, \sigma^2_{tf}, \beta_{0,r}, \beta_{1,r}$ and $\sigma^2_r, r = 1, \ldots, p$, are parameters to be estimated.

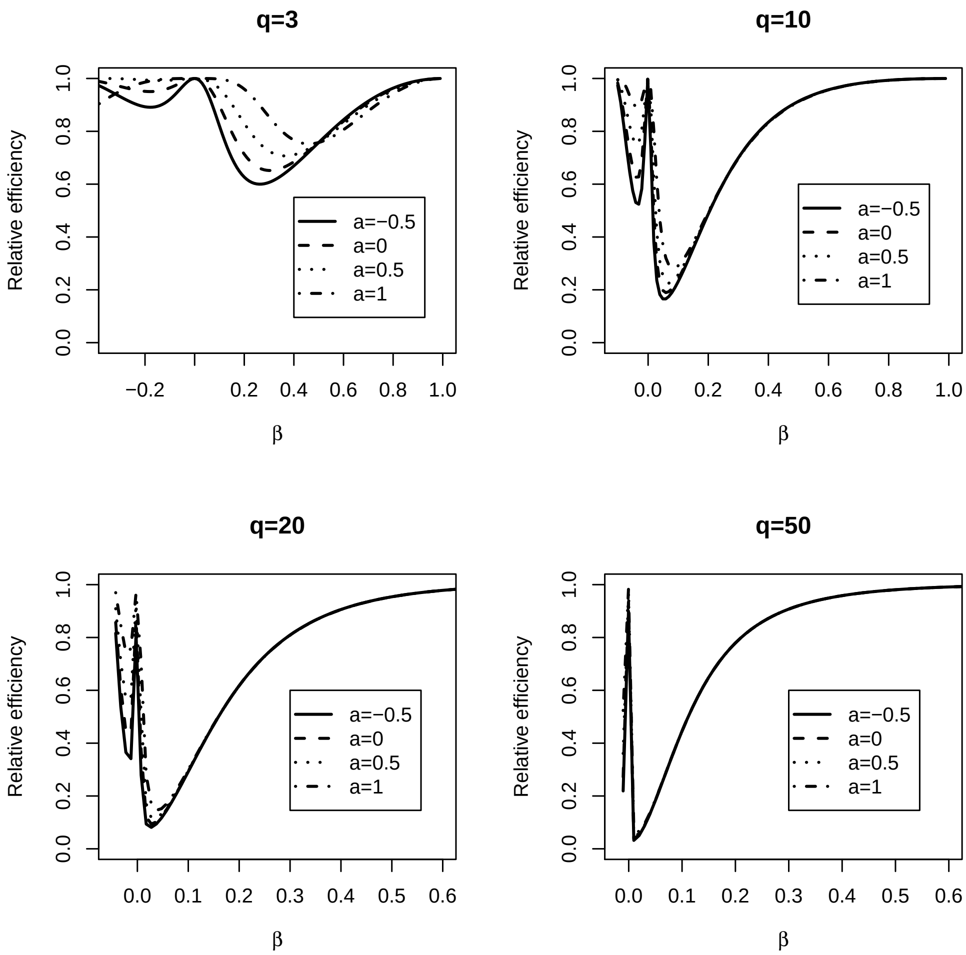

Figure 3.4: Common partial correlation model. Comparisons of the relative efficiency for fixed q and different values of a.

The components of the correlation matrix $\Sigma = \mathrm{Cor}(X_V)$ are given by

$$\mathrm{Cor}(X_r, X_{tf}) = \rho_r, \ \mathrm{Cor}(X_r, X_s) = \rho_r \rho_s,$$

where

$$\rho_r = \mathrm{sgn}(\beta_{1,r}) \left(\frac{\eta_r \sigma_{tf}^2}{\eta_r \sigma_{tf}^2 + 1} \right)^{\frac{1}{2}},$$

with $\mathrm{sgn}(\cdot)$ denoting the sign function and $\eta_r = \beta_{1,r}^2 / \sigma_r^2$ the signal-to-noise ratio for the regression of X_r on X_{tf}. After some algebra, we obtain that

$\text{Cov}(X_r, X_{tf}) = \beta_{1,r}\sigma^2_{tf}$ and $\text{Cov}(X_r, X_s) = \beta_{1,r}\beta_{1,s}\sigma^2_{tf}$. It is also of interest the submodel which assumes that the linear regressions of X_r on X_{tf} have all the same signal-to-noise ratio $\eta_r = \eta$, this means that either $\beta_{1r} \propto \sigma_r$ or $\beta_{1r} = \beta$ and $\sigma_r = \sigma$. In the following, we consider the latter case, with the further simplification of $\beta_{0r} = 0$, for $r = 1, \ldots, p$. Therefore, the resulting model will have a four dimensional unknown parameter, $\theta = (\beta_1, \sigma^2, \mu_{tf}, \sigma^2_{tf})$.

The determinant of Σ is

$$| \Sigma |= (\sigma^2)^p \sigma^2_{tf},$$

and the corresponding determinant of the principal minor of order p is

$$| M_{p\times p} |= (\sigma^2)^p \left[1 + p\frac{\beta_1^2 \sigma^2_{tf}}{\sigma^2}\right].$$

As we can see, the determinants of all principal minors are positive and therefore, the covariance matrix Σ is positive definite for any θ.

Full likelihood

The resulting model can be summarized as follows: $X_{tf} \sim N(\mu_{tf}, \sigma^2_{tf})$ and $X_r \mid X_{tf} \sim N(\beta_1 x_{tf}, \sigma^2), r = 1, \ldots, p$. Considering a single observation, the density function of X_{iV} can be written as

$$\begin{aligned}
f(x_{iV}) &= f(x_{i1}, \ldots, x_{ip}, x_{itf}) \\
&= f(x_{i1} \mid x_{itf}) f(x_{i2} \mid x_{itf}) \ldots f(x_{ip} \mid x_{itf}) f(x_{itf}) \\
&= \prod_{r=1}^{p} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\left(x_{ir} - \beta_1 x_{itf}\right)^2\right\} \frac{1}{\sqrt{2\pi\sigma^2_{tf}}} \times \\
&\quad \exp\left\{-\frac{1}{2\sigma^2_{tf}}\left(x_{itf} - \mu_{tf}\right)^2\right\} \\
&= \left(2\pi\sigma^2\right)^{-p/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{r=1}^{p}\left(x_{ir} - \beta_1 x_{itf}\right)^2\right\} \left(2\pi\sigma^2_{tf}\right)^{-1/2} \times \\
&\quad \exp\left\{-\frac{1}{2\sigma^2_{tf}}\left(x_{itf} - \mu_{tf}\right)^2\right\}.
\end{aligned}$$

The full likelihood based on n independent observations is given by

$$L(\theta) \propto \left(\sigma^2\right)^{-\frac{np}{2}} \left(\sigma^2_{tf}\right)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{r=1}^{p}\left(x_{ir}-\beta_1 x_{itf}\right)^2\right\} \times$$
$$\exp\left\{-\frac{1}{2\sigma^2_{tf}}\sum_{i=1}^{n}\left(x_{itf}-\mu_{tf}\right)^2\right\},$$

and the corresponding log likelihood is

$$\begin{aligned}\ell(\theta) &= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{r=1}^{p}\left(x_{ir}-\beta_1 x_{itf}\right)^2 - \frac{1}{2\sigma^2_{tf}}\sum_{i=1}^{n}\left(x_{itf}-\mu_{tf}\right)^2 - \frac{np}{2}\log(\sigma^2)\\ &= -\frac{n}{2}\log(\sigma^2_{tf}) - \frac{1}{2\sigma^2}SS_1(\beta_1) - \frac{1}{2\sigma^2_{tf}}SS_2(\mu_{tf}) - \frac{np}{2}\log(\sigma^2) - \frac{n}{2}\log(\sigma^2_{tf}),\end{aligned}$$

where $SS_1(\beta_1) = \sum_{i=1}^{n}\sum_{r=1}^{p}\left(x_{ir}-\beta_1 x_{itf}\right)^2$ and $SS_2(\mu_{tf}) = \sum_{i=1}^{n}\left(x_{itf}-\mu_{tf}\right)^2$. The score function has components

$$\begin{aligned}\frac{\partial\ell(\theta)}{\partial\beta_1} &= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{r=1}^{p}2\left(x_{ir}-\beta_1 x_{itf}\right)\left(-x_{itf}\right) = \frac{1}{\sigma^2}\sum_{i=1}^{n}\sum_{r=1}^{p}\left(x_{ir}-\beta_1 x_{itf}\right)x_{itf}\\ \frac{\partial\ell(\theta)}{\partial\mu_{tf}} &= \frac{1}{\sigma^2_{tf}}\sum_{i=1}^{n}\left(x_{itf}-\mu_{tf}\right)\\ \frac{\partial\ell(\theta)}{\partial\sigma^2} &= \frac{1}{2(\sigma^2)^2}SS_1(\beta_1) - \frac{np}{2\sigma^2}\\ \frac{\partial\ell(\theta)}{\partial\sigma^2_{tf}} &= \frac{1}{2(\sigma^2_{tf})^2}SS_2(\mu_{tf}) - \frac{n}{2\sigma^2_{tf}}.\end{aligned}$$

The likelihood equations can be solved explicitly producing the following estimates:

$$\begin{aligned}\hat\beta_1 &= \frac{1}{p}\frac{\sum_{i=1}^{n}\left(x_{itf}\sum_{r=1}^{p}x_{ir}\right)}{\sum_{i=1}^{n}\left(x_{itf}\right)^2}\\ \hat\mu_{tf} &= \frac{\sum_{i=1}^{n}x_{itf}}{n}\\ \hat\sigma^2 &= \frac{SS_1(\hat\beta_1)}{np}\\ \hat\sigma^2_{tf} &= \frac{SS_2(\hat\mu_{tf})}{n}.\end{aligned}$$

The observed information matrix

$$j(\beta_1, \mu_{tf}, \sigma^2, \sigma_{tf}^2) = \begin{pmatrix} j_{\beta_1\beta_1} & j_{\beta_1\mu_{tf}} & j_{\beta_1\sigma^2} & j_{\beta_1\sigma_{tf}^2} \\ j_{\beta_1\mu_{tf}} & j_{\mu_{tf}\mu_{tf}} & j_{\mu_{tf}\sigma^2} & j_{\mu_{tf}\sigma_{tf}^2} \\ j_{\beta_1\sigma^2} & j_{\mu_{tf}\sigma^2} & j_{\sigma^2\sigma^2} & j_{\sigma^2\sigma_{tf}^2} \\ j_{\beta_1\sigma_{tf}^2} & j_{\mu_{tf}\sigma_{tf}^2} & j_{\sigma^2\sigma_{tf}^2} & j_{\sigma_{tf}^2\sigma_{tf}^2} \end{pmatrix}$$

has elements

$$\begin{aligned}
j_{\beta_1\beta_1} &= -\frac{\partial^2 \ell(\theta)}{\partial \beta_1^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}\sum_{r=1}^{p}\left(x_{itf}\right)^2, \\
j_{\beta_1\mu_{tf}} &= -\frac{\partial^2 \ell(\theta)}{\partial \beta_1 \partial \mu_{tf}} = 0, \\
j_{\beta_1\sigma^2} &= -\frac{\partial^2 \ell(\theta)}{\partial \beta_1 \partial \sigma^2} = \frac{1}{(\sigma^2)^2}\sum_{i=1}^{n}\sum_{r=1}^{p}\left(x_{ir} - \beta_1 x_{itf}\right) x_{itf}, \\
j_{\beta_1\sigma_{tf}^2} &= -\frac{\partial^2 \ell(\theta)}{\partial \beta_1 \partial \sigma_{tf}^2} = 0, \\
j_{\mu_{tf}\mu_{tf}} &= -\frac{\partial^2 \ell(\theta)}{\partial \mu_{tf}^2} = \frac{n}{\sigma_{tf}^2}, \\
j_{\mu_{tf}\sigma^2} &= -\frac{\partial^2 \ell(\theta)}{\partial \mu_{tf} \partial \sigma^2} = 0, \\
j_{\mu_{tf}\sigma_{tf}^2} &= -\frac{\partial^2 \ell(\theta)}{\partial \mu_{tf} \partial \sigma_{tf}^2} = \frac{1}{(\sigma_{tf}^2)^2}\sum_{i=1}^{n}\left(x_{itf} - \mu_{tf}\right), \\
j_{\sigma^2\sigma^2} &= -\frac{\partial^2 \ell(\theta)}{\partial (\sigma^2)^2} = \frac{1}{(\sigma^2)^3}\sum_{i=1}^{n}\sum_{r=1}^{p}\left(x_{ir} - \beta_1 x_{itf}\right)^2 - \frac{np}{2(\sigma^2)^2}, \\
j_{\sigma^2\sigma_{tf}^2} &= -\frac{\partial^2 \ell(\theta)}{\partial \sigma^2 \sigma_{tf}^2} = 0, \\
j_{\sigma_{tf}^2\sigma_{tf}^2} &= -\frac{\partial^2 \ell(\theta)}{\partial (\sigma_{tf}^2)^2} = \frac{1}{(\sigma_{tf}^2)^3}\sum_{i=1}^{n}\left(x_{itf} - \mu_{tf}\right)^2 - \frac{n}{2(\sigma_{tf}^2)^2}.
\end{aligned}$$

The Fisher information matrix is

$$i(\theta) = \mathrm{E}_\theta\{j(\theta)\} = \begin{pmatrix} i_{\beta_1\beta_1} & i_{\beta_1\mu_{tf}} & i_{\beta_1\sigma^2} & i_{\beta_1\sigma_{tf}^2} \\ i_{\beta_1\mu_{tf}} & i_{\mu_{tf}\mu_{tf}} & i_{\mu_{tf}\sigma^2} & i_{\mu_{tf}\sigma_{tf}^2} \\ i_{\beta_1\sigma^2} & i_{\mu_{tf}\sigma^2} & i_{\sigma^2\sigma^2} & i_{\sigma^2\sigma_{tf}^2} \\ i_{\beta_1\sigma_{tf}^2} & i_{\mu_{tf}\sigma_{tf}^2} & i_{\sigma^2\sigma_{tf}^2} & i_{\sigma_{tf}^2\sigma_{tf}^2} \end{pmatrix}$$

with elements

$$
\begin{aligned}
i_{\beta_1\beta_1} &= \mathrm{E}_\theta\{j_{\beta_1\beta_1}\} = \frac{np\mathrm{E}_\theta\left\{(X_{itf})^2\right\}}{\sigma^2} = \frac{np\left\{\sigma^2_{tf}+\mu^2_{tf}\right\}}{\sigma^2},\\
i_{\beta_1\mu_{tf}} &= \mathrm{E}_\theta\{j_{\beta_1\mu_{tf}}\} = 0,\\
i_{\beta_1\sigma^2} &= \mathrm{E}_\theta\{j_{\beta_1\sigma^2}\} = \frac{1}{(\sigma^2)^2}\left\{\sum_{i=1}^{n}\sum_{r=1}^{p}\mathrm{E}_\theta\left(X_{itf}X_{ir}\right) - \beta_1 p\sum_{i=1}^{n}\mathrm{E}_\theta\left(X_{itf}\right)^2\right\}\\
&= \frac{1}{(\sigma^2)^2}\left\{np\left(\beta_1\sigma^2_{tf}+\beta_1\mu^2_{tf}\right) - np\beta_1\left(\sigma^2_{tf}+\mu^2_{tf}\right)\right\} = 0,\\
i_{\beta_1\sigma^2_{tf}} &= \mathrm{E}_\theta\{j_{\beta_1\sigma^2_{tf}}\} = 0,\\
i_{\mu_{tf}\mu_{tf}} &= \mathrm{E}_\theta\{j_{\mu_{tf}\mu_{tf}}\} = \frac{n}{\sigma^2_{tf}},\\
i_{\mu_{tf}\sigma^2} &= \mathrm{E}_\theta\{j_{\mu_{tf}\sigma^2}\} = 0,\\
i_{\mu_{tf}\sigma^2_{tf}} &= \mathrm{E}_\theta\{j_{\mu_{tf}\sigma^2_{tf}}\} = \frac{1}{(\sigma^2_{tf})^2}\sum_{i=1}^{n}\mathrm{E}_\theta\left\{(X_{itf}-\mu_{tf})\right\} = 0,\\
i_{\sigma^2\sigma^2} &= \mathrm{E}_\theta\{j_{\sigma^2\sigma^2}\} = \frac{1}{(\sigma^2)^3}\sum_{i=1}^{n}\sum_{r=1}^{p}\mathrm{E}_\theta\left\{(X_{ir}-\beta_1X_{itf})^2\right\} - \frac{np}{2(\sigma^2)^2},\\
&= \frac{np\sigma^2}{(\sigma^2)^3} - \frac{np}{2(\sigma^2)^2} = \frac{np}{2(\sigma^2)^2},\\
i_{\sigma^2\sigma^2_{tf}} &= \mathrm{E}_\theta\{j_{\sigma^2\sigma^2_{tf}}\} = 0,\\
i_{\sigma^2_{tf}\sigma^2_{tf}} &= \mathrm{E}_\theta\{j_{\sigma^2_{tf}\sigma^2_{tf}}\} = \frac{1}{(\sigma^2_{tf})^3}\sum_{i=1}^{n}\mathrm{E}_\theta\left\{(X_{itf}-\mu_{tf})^2\right\} - \frac{n}{2(\sigma^2_{tf})^2},\\
&= \frac{n\sigma^2_{tf}}{(\sigma^2_{tf})^3} - \frac{n}{2(\sigma^2_{tf})^2} = \frac{n}{2(\sigma^2_{tf})^2}.
\end{aligned}
$$

Combined composite likelihood

We now consider the combined composite likelihood which might be appropriate since both univariate and bivariate marginal densities depend on all parameters of interest, except for the parameter μ_{tf}. Indeed, only the univariate marginal densities depend on μ_{tf}.

For this model, the pairwise likelihood is based on the following two bivariate marginal distributions

$$\begin{pmatrix} X_{ir} \\ X_{is} \end{pmatrix} \sim N_2 \left(\begin{bmatrix} \beta_1\mu_{tf} \\ \beta_1\mu_{tf} \end{bmatrix}, \begin{bmatrix} \sigma^2 + \beta_1^2\sigma_{tf}^2 & \beta_1^2\sigma_{tf}^2 \\ \beta_1^2\sigma_{tf}^2 & \sigma^2 + \beta_1^2\sigma_{tf}^2 \end{bmatrix} \right), \quad r \neq s,$$

$$\begin{pmatrix} X_{ir} \\ X_{itf} \end{pmatrix} \sim N_2 \left(\begin{bmatrix} \beta_1\mu_{tf} \\ \mu_{tf} \end{bmatrix}, \begin{bmatrix} \sigma^2 + \beta_1^2\sigma_{tf}^2 & \beta_1\sigma_{tf}^2 \\ \beta_1\sigma_{tf}^2 & \sigma_{tf}^2 \end{bmatrix} \right),$$

and thus the pairwise likelihood based on n independent observations is given by

$$CL^{\mathrm{P}}(\theta) = \prod_{i=1}^{n}\prod_{r=1}^{p-1}\prod_{s=r+1}^{p} f(x_{ir}, x_{is}; \theta) \prod_{i=1}^{n}\prod_{r=1}^{p} f(x_{ir}, x_{itf}; \theta),$$

and the corresponding pairwise log likelihood, after some algebra, is

$$\begin{aligned} c\ell^{\mathrm{P}}(\theta) = &-\frac{\{p\sigma^2 + \beta_1^2\sigma_{tf}^2(p+2)\}}{2\sigma^2(\sigma^2 + 2\beta_1^2\sigma_{tf}^2)} SS_1 + \frac{\beta_1\mu_{tf}(p-1)}{\sigma^2 + 2\beta_1^2\sigma_{tf}^2} SS_2 + \frac{\beta_1^2\sigma_{tf}^2}{2\sigma^2(\sigma^2 + 2\beta_1^2\sigma_{tf}^2)} \\ &\times SS_3 - \frac{(\sigma^2 + \beta_1^2\sigma_{tf}^2)p}{2\sigma_{tf}^2\sigma^2} SS_4 + \frac{\mu_{tf}p}{\sigma_{tf}^2} SS_5 + \frac{\beta_1}{\sigma^2} SS_6 - \frac{np(p-1)}{4} \times \\ &\log\left\{\sigma^2(\sigma^2 + 2\beta_1^2\sigma_{tf}^2)\right\} - \frac{np}{2}\log(\sigma_{tf}^2\sigma^2) - \frac{np\mu_{tf}^2\{\sigma^2 + \beta_1^2\sigma_{tf}^2(p+1)\}}{2\sigma_{tf}^2(\sigma^2 + 2\beta_1^2\sigma_{tf}^2)}, \end{aligned}$$

where,

$$\begin{aligned} &SS_1 = \sum_{i=1}^{n}\sum_{r=1}^{p} x_{ir}^2,\ SS_2 = \sum_{i=1}^{n}\sum_{r=1}^{p} x_{ir},\ SS_3 = \sum_{i=1}^{n}\left(\sum_{r=1}^{p} x_{ir}\right)^2, \\ &SS_4 = \sum_{i=1}^{n} x_{itf}^2,\ SS_5 = \sum_{i=1}^{n} x_{itf},\ SS_6 = \sum_{i=1}^{n}\sum_{r=1}^{p} x_{ir}x_{itf}. \end{aligned} \tag{3.8}$$

The independence likelihood, with all weights equal to 1, is based on the following univariate marginal distributions

$$\begin{aligned} X_{itf} &\sim N(\mu_{tf}, \sigma_{tf}^2), \\ X_{ir} &\sim N(\beta_1\mu_{tf}, \sigma^2 + \beta_1^2\sigma_{tf}^2),\ r = 1, \dots, p, \end{aligned}$$

and is given by

$$
\begin{aligned}
c\ell^{\mathrm{I}}(\theta) &= \prod_{i=1}^{n}\prod_{r=1}^{p} f(x_{ir};\beta_1,\mu_{tf},\sigma^2,\sigma^2_{tf})\prod_{i=1}^{n} f(x_{itf};\mu_{tf},\sigma^2_{tf})\\
&= \prod_{i=1}^{n}\prod_{r=1}^{p}\frac{1}{\sqrt{2\pi(\sigma^2+\beta_1^2\sigma^2_{tf})}}\exp\left\{-\frac{1}{2(\sigma^2+\beta_1^2\sigma^2_{tf})}(x_{ir}-\beta_1\mu_{tf})^2\right\}\\
&\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2_{tf}}}\exp\left\{-\frac{1}{2\sigma^2_{tf}}(x_{itf}-\mu_{tf})^2\right\}\\
&= -\frac{1}{2(\sigma^2+\beta_1^2\sigma^2_{tf})}SS_1+\frac{\beta_1\mu_{tf}}{\sigma^2+\beta_1^2\sigma^2_{tf}}SS_2-\frac{1}{2\sigma^2_{tf}}SS_4+\frac{\mu_{tf}}{\sigma^2_{tf}}SS_5\\
&-\frac{n\mu^2_{tf}\{\sigma^2+\beta_1^2\sigma^2_{tf}(p+1)\}}{2\sigma^2_{tf}(\sigma^2+\beta_1^2\sigma^2_{tf})}-\frac{np}{2}\log(\sigma^2+\beta_1^2\sigma^2_{tf})-\frac{n}{2}\log\sigma^2_{tf}.
\end{aligned}
$$

The resulting combined composite likelihood for this model is therefore

$$
\begin{aligned}
c\ell^{a}(\theta) = &-\frac{\{2\beta_1^4(p+2)\}(\sigma^2_{tf})^2+\{(4-2a)\beta_1^2p+4\beta_1^2\}\sigma^2\sigma^2_{tf}+(2-a)p(\sigma^2)^2}{2\sigma^2(\sigma^2+\beta_1^2\sigma^2_{tf})(\sigma^2+2\beta_1^2\sigma^2_{tf})}SS_1\\
&-\frac{\beta_1\mu_{tf}[\{(2a-2)\beta_1^2p+2\beta_1^2\}\sigma^2_{tf}+\{(a-2)p+2\}\sigma^2]}{(\sigma^2+\beta_1^2\sigma^2_{tf})(\sigma^2+2\beta_1^2\sigma^2_{tf})}SS_2\\
&+\frac{\beta_1^2\sigma^2_{tf}}{\sigma^2(\sigma^2+2\beta_1^2\sigma^2_{tf})}SS_3-\frac{p(2\beta_1^2\sigma^2_{tf}+(2-a)\sigma^2)}{2\sigma^2\sigma^2_{tf}}SS_4-\frac{(a-2)\mu_{tf}p}{\sigma^2_{tf}}SS_5\\
&+\frac{2\beta_1}{\sigma^2}SS_6-\frac{np}{2}\{(p-1)\log\{\sigma^2(\sigma^2+2\beta_1^2\sigma^2_{tf})\}-ap\log(\sigma^2+\beta_1^2\sigma^2_{tf})\}\\
&+\frac{np\mu^2_{tf}\{(2a-2)\beta_1^2\sigma^2_{tf}+(a-2)\sigma^2\}\{(\beta_1^2p+\beta_1^2)\sigma^2_{tf}+\sigma^2\}}{2\sigma^2_{tf}(\sigma^2+\beta_1^2\sigma^2_{tf})(\sigma^2+2\beta_1^2\sigma^2_{tf})}\\
&-\frac{np}{2}(2\log(\sigma^2\sigma^2_{tf})-a\log\sigma^2_{tf})\cdot
\end{aligned}
\tag{3.9}
$$

For compactness of notation, we rewrite $c\ell_a(\theta)$ as

$$
c\ell^{a}(\theta) = h_1SS_1+h_2SS_2+h_3SS_3+h_4SS_4+h_5SS_5+h_6SS_6+h_7,
$$

where h_1 is the factor that multiplies SS_1 in (3.9); and so on. For the calculations and simplifications of some quantities related to the $c\ell_a(\theta)$ we have used the Computer Algebra System Maxima (version 5.20.1); see http://maxima.sourceforge.net/ for information about the software. For the calculation of the matrix $H^a(\theta)$, it is necessary to calculate the expected values of the statistics defined in (3.8), that are

$$\begin{aligned}
\mathrm{E}_\theta(SS_1) &= np\mathrm{E}_\theta(X_{ir}^2) = np(\sigma^2 + \beta_1^2\sigma_{tf}^2 + \beta_1^2\mu_{tf}^2),\\
\mathrm{E}_\theta(SS_2) &= np\mathrm{E}_\theta(X_{ir}) = np\beta_1\mu_{tf},\\
\mathrm{E}_\theta(SS_3) &= n\mathrm{E}_\theta\left\{\left(\sum_{r=1}^p X_{ir}\right)^2\right\} = n\left[\mathrm{Var}_\theta\left(\sum_{r=1}^p X_{ir}\right) + \left\{\mathrm{E}_\theta\left(\sum_{r=1}^p X_{ir}\right)\right\}^2\right],\\
&= \beta_1\mu_{tf}(3\beta_1^2\sigma_{tf}^2 + 3\sigma^2 + \beta_1^2\mu_{tf}^2),\\
\mathrm{E}_\theta(SS_4) &= n\mathrm{E}_\theta(X_{itf}^2) = n(\sigma_{tf}^2 + \mu_{tf}^2),\\
\mathrm{E}_\theta(SS_5) &= n\mathrm{E}_\theta(X_{itf}) = n\mu_{tf},\\
\mathrm{E}_\theta(SS_6) &= np\mathrm{E}_\theta(X_{ir}X_{itf}) = np\{\mathrm{Cov}_\theta(X_{ir}, X_{itf}) + \mathrm{E}_\theta(X_{ir})\mathrm{E}_\theta(X_{itf})\},\\
&= np(\beta_1\sigma_{tf}^2 + \beta_1\mu_{tf}^2).
\end{aligned}$$

The elements of the matrix

$$H^a(\theta) = \begin{pmatrix}
H^a_{\beta_1\beta_1} & H^a_{\beta_1\mu_{tf}} & H^a_{\beta_1\sigma^2} & H^a_{\beta_1\sigma^2_{tf}}\\
H^a_{\beta_1\mu_{tf}} & H^a_{\mu_{tf}\mu_{tf}} & H^a_{\mu_{tf}\sigma^2} & H^a_{\mu_{tf}\sigma^2_{tf}}\\
H^a_{\beta_1\sigma^2} & H^a_{\mu_{tf}\sigma^2} & H^a_{\sigma^2\sigma^2} & H^a_{\sigma^2\sigma^2_{tf}}\\
H^a_{\beta_1\sigma^2_{tf}} & H^a_{\mu_{tf}\sigma^2_{tf}} & H^a_{\sigma^2\sigma^2_{tf}} & H^a_{\sigma^2_{tf}\sigma^2_{tf}}
\end{pmatrix}$$

are reported in Section A.1 of the Appendix A. In order to calculate the terms of the matrix $J^a(\theta)$, it is necessary to know the variance of each random variable defined in (3.8) and also the correlation between each pair of variables. As regards the calculation of the mixed moments or moments of order greater than two, we used the moment generating function of the multivariate normal. Thus it follows that

$$\begin{aligned}
\mathrm{Var}_\theta(SS_1) &= n\left\{\sum_{r=1}^p \mathrm{Var}_\theta(X_{ir}^2) + \sum_{r=1}^p\sum_{\substack{s=1\\ s\neq r}}^p \mathrm{Cov}_\theta(X_{ir}^2, X_{is}^2)\right\}\\
&= n\left\{p\mathrm{Var}_\theta(X_{ir}^2) + p(p-1)\mathrm{Cov}_\theta(X_{ir}^2, X_{is}^2)\right\},
\end{aligned}$$

$$
\begin{aligned}
&= n\left[p\left[\mathrm{E}_\theta(X_{ir}^4) - \left\{\mathrm{E}_\theta(X_{ir}^2)\right\}^2\right] + p(p-1)\left\{\mathrm{E}_\theta(X_{ir}^2X_{is}^2)\right.\right.\\
&\quad \left.\left.-\mathrm{E}_\theta(X_{ir}^2)\mathrm{E}_\theta(X_{is}^2)\right\}\right]\\
&= 2np\left[\beta_1^4 p(\sigma_{tf}^2)^2 + 2\beta_1^2\sigma^2\sigma_{tf}^2 + 2\beta_1^4\mu_{tf}^2 p\sigma_{tf}^2 + (\sigma^2)^2 + 2\beta_1^2\mu_{tf}^2\sigma^2\right],\\
\mathrm{Var}_\theta(SS_2) &= n\left\{\sum_{r=1}^p \mathrm{Var}_\theta(X_{ir}) + \sum_{r=1}^p\sum_{\substack{s=1\\ s\neq r}}^p \mathrm{Cov}_\theta(X_{ir}, X_{is})\right\}\\
&= n(\beta_1^2p^2\sigma_{tf}^2 + p\sigma^2),\\
\mathrm{Var}_\theta(SS_3) &= n\mathrm{Var}_\theta\left\{\left(\sum_{r=1}^p X_{ir}\right)^2\right\}\\
&= n\left[\mathrm{E}_\theta\left\{\left(\sum_{r=1}^p X_{ir}\right)^4\right\} - \left(\mathrm{E}_\theta\left\{\left(\sum_{r=1}^p X_{ir}\right)^2\right\}\right)^2\right]\\
&= 2np^2(p\beta_1^2\sigma_{tf}^2 + \sigma^2)(p\beta_1^2\sigma_{tf}^2 + \sigma^2 + 2p\beta_1^2\mu_{tf}^2),\\
\mathrm{Var}_\theta(SS_4) &= n\mathrm{Var}_\theta(X_{itf}^2) = n\left[\mathrm{E}_\theta(X_{itf}^4) - \left\{\mathrm{E}_\theta(X_{itf}^2)\right\}^2\right]\\
&= 2n\sigma_{tf}^2(\sigma_{tf}^2 + 2\mu_{tf}^2),\\
\mathrm{Var}_\theta(SS_5) &= n\mathrm{Var}_\theta(X_{itf}) = n\sigma_{tf}^2,\\
\mathrm{Var}_\theta(SS_6) &= n\left\{\mathrm{Var}_\theta\left(\sum_{r=1}^p X_{ir}X_{itf}\right)\right\}\\
&= np(2\beta_1^2p(\sigma_{tf}^2)^2 + \sigma^2\sigma_{tf}^2 + 4\beta_1^2\mu_{tf}^2p\sigma_{tf}^2 + \mu_{tf}^2\sigma^2).
\end{aligned}
$$

We now calculate the covariance between each pair of variables. Thus we have that

$$\mathrm{Cov}_\theta(SS_1, SS_2) = \mathrm{E}_\theta(SS_1SS_2) - \mathrm{E}_\theta(SS_1)\mathrm{E}_\theta(SS_2).$$

Since the corresponding expected values are known, we only need to calculate the expected value of SS_1SS_2, which can be formally rewritten as

$$
\begin{aligned}
SS_1SS_2 &= \sum_{i=1}^n\sum_{r=1}^p X_{ir}^2 \sum_{j=1}^n\sum_{s=1}^p X_{js} = \sum_{i=1}^n T_{i1}\sum_{j=1}^n T_{j2}\\
&= \sum_{i=1}^n T_{i1}T_{i2} + \sum_{i=1}^n\sum_{\substack{j=1\\ j\neq i}}^n T_{i1}T_{j2},
\end{aligned}
$$

where $T_{i1}T_{i2}$ can in turn be rewritten as

$$T_{i1}T_{i2} = \sum_{r=1}^{p} X_{ir}^2 \sum_{s=1}^{p} X_{is} = \sum_{r=1}^{p} X_{ir}^3 + \sum_{r=1}^{p} \sum_{\substack{s=1 \\ s \neq r}}^{p} X_{ir}^2 X_{is}.$$

Considering the expected values, it follows that

$$\begin{aligned}
\mathrm{E}_\theta(T_{i1}T_{i2}) &= p\mathrm{E}_\theta(X_{ir}^3) + p(p-1)\mathrm{E}_\theta(X_{ir}^2 X_{is}), \\
\mathrm{E}_\theta(SS_1, SS_2) &= n\{p\mathrm{E}_\theta(X_{ir}^3) + p(p-1)\mathrm{E}_\theta(X_{ir}^2 X_{is})\} + n(n-1)\mathrm{E}_\theta(T_{i1})\mathrm{E}_\theta(T_{j2}) \\
&= n\{p\mathrm{E}_\theta(X_{ir}^3) + p(p-1)\mathrm{E}_\theta(X_{ir}^2 X_{is})\} + \frac{(n-1)}{n}\mathrm{E}_\theta(SS_1)\mathrm{E}_\theta(SS_2).
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\mathrm{Cov}_\theta(SS_1, SS_2) &= n\{p\mathrm{E}_\theta(X_{ir}^3) + p(p-1)\mathrm{E}_\theta(X_{ir}^2 X_{is})\} + \frac{(n-1)}{n}\mathrm{E}_\theta(SS_1)\mathrm{E}_\theta(SS_2) \\
&\quad - \mathrm{E}_\theta(SS_1)\mathrm{E}_\theta(SS_2), \\
&= 2np\beta_1\mu_{tf}(\beta_1^2 p\sigma_{tf}^2 + \sigma^2).
\end{aligned}$$

Proceeding in the same way for the calculation of the remaining covariances, we get the following results

$$\begin{aligned}
\mathrm{Cov}_\theta(SS_1, SS_3) &= n\{p\mathrm{E}_\theta(X_{ir}^4) + p(p-1)\mathrm{E}_\theta(X_{ir}^2 X_{is}^2) + 2p(p-1)\mathrm{E}_\theta(X_{ir}^3 X_{is}) \\
&\quad + p(p-1)(p-2)\mathrm{E}_\theta(X_{ir}^2 X_{is} X_{iv})\} + \frac{(n-1)}{n}\mathrm{E}_\theta(SS_1)\mathrm{E}_\theta(SS_3) \\
&\quad - \mathrm{E}_\theta(SS_1)\mathrm{E}_\theta(SS_3) \\
&= 2np(\beta_1^2 p\sigma_{tf}^2 + \sigma^2)(\beta_1^2 p\sigma_{tf}^2 + \sigma^2 + 2\beta_1^2\mu_{tf}^2 p), \\
\mathrm{Cov}_\theta(SS_1, SS_4) &= 2np\beta_1^2\sigma_{tf}^2(\sigma_{tf}^2 + 2\mu_{tf}^2), \\
\mathrm{Cov}_\theta(SS_1, SS_5) &= 2np\beta_1^2\mu_{tf}\sigma_{tf}^2, \\
\mathrm{Cov}_\theta(SS_1, SS_6) &= 2np\beta_1(\beta_1^2 p(\sigma_{tf}^2)^2 + \sigma^2\sigma_{tf}^2 + 2\beta_1^2\mu_{tf}^2 p\sigma_{tf}^2 + \mu_{tf}^2\sigma^2), \\
\mathrm{Cov}_\theta(SS_2, SS_3) &= 2np^2\beta_1\mu_{tf}(\beta_1^2 p\sigma_{tf}^2 + \sigma^2), \\
\mathrm{Cov}_\theta(SS_2, SS_4) &= 2np\beta_1\mu_{tf}\sigma_{tf}^2, \\
\mathrm{Cov}_\theta(SS_2, SS_5) &= np\beta_1\sigma_{tf}^2, \\
\mathrm{Cov}_\theta(SS_2, SS_6) &= np\mu_{tf}(2\beta_1^2 p\sigma_{tf}^2 + \sigma^2), \\
\mathrm{Cov}_\theta(SS_3, SS_4) &= 2np^2\beta_1^2\sigma_{tf}^2(\sigma_{tf}^2 + 2\mu_{tf}^2), \\
\mathrm{Cov}_\theta(SS_3, SS_5) &= 2np^2\beta_1^2\mu_{tf}\sigma_{tf}^2,
\end{aligned}$$

$$\begin{aligned}
\mathrm{Cov}_\theta(SS_3, SS_6) &= 2np^2\beta_1(\beta_1^2 p(\sigma_{tf}^2)^2 + \sigma^2\sigma_{tf}^2 + 2\beta_1^2\mu_{tf}^2 p\sigma_{tf}^2 + \mu_{tf}^2\sigma^2),\\
\mathrm{Cov}_\theta(SS_4, SS_5) &= 2n\mu_{tf}\sigma_{tf}^2,\\
\mathrm{Cov}_\theta(SS_4, SS_6) &= 2np\beta_1\sigma_{tf}^2(\sigma_{tf}^2 + 2\mu_{tf}^2),\\
\mathrm{Cov}_\theta(SS_5, SS_6) &= 2np\beta_1\mu_{tf}\sigma_{tf}^2.
\end{aligned}$$

Using the above quantities we can compute the elements of the matrix

$$J^a(\theta) = \mathrm{Var}\left\{\frac{\partial c\ell^a(\theta)}{\partial\theta}\right\} = \begin{pmatrix} J^a_{\beta_1\beta_1} & J^a_{\beta_1\mu_{tf}} & J^a_{\beta_1\sigma^2} & J^a_{\beta_1\sigma^2_{tf}} \\ J^a_{\beta_1\mu_{tf}} & J^a_{\mu_{tf}\mu_{tf}} & J^a_{\mu_{tf}\sigma^2} & J^a_{\mu_{tf}\sigma^2_{tf}} \\ J^a_{\beta_1\sigma^2} & J^a_{\mu_{tf}\sigma^2} & J^a_{\sigma^2\sigma^2} & J^a_{\sigma^2\sigma^2_{tf}} \\ J^a_{\beta_1\sigma^2_{tf}} & J^a_{\mu_{tf}\sigma^2_{tf}} & J^a_{\sigma^2\sigma^2_{tf}} & J^a_{\sigma^2_{tf}\sigma^2_{tf}} \end{pmatrix}$$

which are reported in Section A.2 of the Appendix A. Finally, the Godambe information will be given by

$$G^a(\theta) = H^a(\theta)J^a(\theta)^{-1}H^a(\theta).$$

Since in this example the parameter is a vector, we need a condition similar to the one defined in (3.1), in order to determine the values of a for which the eigenvalues of $H^a(\theta)$ are all positive. The eigenvalues are required to be positive so that the matrix $H^a(\theta)$ is positive definite. Several combinations of values of θ lead to the conclusion that $a \le 1$. In other words, for those values, the combined composite satisfies the requirements of being a proper pseudo-likelihood with the desirable properties defined in Subsection 3.2.1. This result is independent of the length of the random vector.

Next we focus attention on the properties of the maximum combined composite likelihood estimator. We first perform a numerical assessment of the consistency of the combined composite likelihood and the full likelihood estimators. We consider two scenarios: in the first one n is fixed and p increases, while p is fixed and n increases in the second case. We start with the first case, running a simulation experiment, with $n = 10, q = 3, 10, 100, 1000, 5000$ and $a \in \{-5, -1, 0, 0.5, 1, 2, 5\}$. For each combination,

Table 3.2: Model for microarray data. Mean squared error of $\hat{\theta}_C^a = (\hat{\beta}_1^a, \hat{\mu}_{tf}^a, \hat{\sigma}^{2a}, \hat{\sigma}_{tf}^{2a})$, when $n = 10, a = -5, -1, 0, 0.5, 1, 2, 5$ and $p = 3, 10, 100, 1000, 5000$. Results are obtained with 2000 simulated samples, and with the true parameter value $\theta_0 = (-1, 0, 1, 1)$.

		a						
	p	-5	-1	0	0.5	1	2	5
	3	0.0592	0.0440	0.0432	0.0430	0.0433	1.716867×10^{5}	3.093910×10^{5}
	10	0.0228	0.0139	0.0139	0.0140	0.0140	2.618219×10^{5}	3.422988×10^{5}
$\boldsymbol{MSE}(\hat{\beta}_1^a)$	100	0.0047	0.0012	0.0012	0.0012	0.0012	2.343502×10^{5}	3.696692×10^{5}
	1000	0.0210	0.0229	0.0146	0.0093	0.0034	1.833506×10^{5}	3.053121×10^{5}
	5000	0.2529	0.2157	0.1890	0.1221	0.0652	1.968528×10^{5}	2.720468×10^{5}
	3	0.1085	0.1056	0.1029	0.0999	0.0980	7.502026×10^{4}	4.030207×10^{4}
	10	0.1061	0.1059	0.1055	0.1050	0.1034	5.749886×10^{4}	4.107768×10^{4}
$\boldsymbol{MSE}(\hat{\mu}_{tf}^a)$	100	0.1054	0.1052	0.1051	0.1051	0.1050	4.781126×10^{4}	7.102014×10^{4}
	1000	0.1190	0.1292	0.1154	0.1199	0.1106	3.808045×10^{4}	4.680990×10^{4}
	5000	0.2203	0.2247	0.2228	0.2140	0.1943	3.841907×10^{4}	3.908486×10^{4}
	3	0.0703	0.0679	0.0679	0.0680	0.0681	4.245588×10^{103}	1.660055×10^{302}
	10	0.0227	0.0223	0.0222	0.0222	0.0222	4.163829×10^{62}	7.005268×10^{301}
$\boldsymbol{MSE}(\hat{\sigma}^{2a})$	100	0.0020	0.0020	0.0020	0.0020	0.0020	5.147217×10^{24}	7.396962×10^{302}
	1000	0.0002	0.0002	0.0002	0.0002	0.0002	2.096144×10^{5}	4.821654×10^{301}
	5000	0.0000	0.0000	0.0000	0.0000	0.0000	1.290243×10^{4}	1.295549×10^{300}
	3	0.1714	0.1713	0.1716	0.1721	0.1752	1.311488×10^{305}	8.153252×10^{298}
	10	0.1874	0.1877	0.1878	0.1878	0.1881	4.013616×10^{302}	3.367751×10^{298}
$\boldsymbol{MSE}(\hat{\sigma}_{tf}^{2a})$	100	0.2128	0.2096	0.2095	0.2096	0.2095	1.151755×10^{294}	3.419950×10^{292}
	1000	0.4412	0.3913	0.3241	0.4060	0.2629	2.082348×10^{292}	1.384276×10^{291}
	5000	2.2101	2.3075	2.2067	2.5153	1.6164	5.781598×10^{291}	1.107073×10^{291}

we performed 2000 iterations. The results correspond to $\theta_0 = (-1, 0, 1, 1)$ and $\theta_0 = (-2, 1, 3, 2)$ as the true parameter values. Table 3.2 and Table 3.3 report the mean squared errors (MSE) based on the combined composite likelihood and the full likelihood, respectively. The results in both tables are obtained by means of the first set of true parameter values, while the results in Table 3.4 and Table 3.5 are obtained with the second set of true parameter values. As we can see from Table 3.2 and Table 3.4, for admissible values of a, only the mean squared errors of $\hat{\sigma}^{2a}$ decrease toward zero as p increases. This seems to suggest, although empirically, that the

Table 3.3: Model for microarray data. Mean squared error of $\hat{\theta} = (\hat{\beta}_1, \hat{\mu}_{tf}, \hat{\sigma}^2, \hat{\sigma}^2_{tf})$, when $n = 10$ and $p = 3, 10, 100, 1000, 5000$. Results are obtained with 2000 simulated samples, and with the true parameter value $\theta_0 = (-1, 0, 1, 1)$.

	$MSE(\hat{\beta}_1)$	$MSE(\hat{\mu}_{tf})$	$MSE(\hat{\sigma}^2)$	$MSE(\hat{\sigma}^2_{tf})$
$p = 3$	0.0415	0.0961	0.0647	0.1714
$p = 10$	0.0135	0.1010	0.0202	0.1880
$p = 100$	0.0012	0.1035	0.0020	0.2093
$p = 1000$	0.0001	0.1032	0.0002	0.2342
$p = 5000$	0.0000	0.1090	0.0000	0.4086

combined composite likelihood estimator is not consistent. Instead, for values of a greater than one, some fundamental properties of a pseudo-likelihood may not be satisfied by the combined composite likelihood and we see that the estimates based on it are extremely far from the true values. Here, it is important to note that, empirically, not only the combined composite likelihood estimator is not consistent, but also the one based on the full likelihood is not consistent. In fact, Table 3.3 and Table 3.5 show that only the MSE of $\hat{\beta}_1$ and $\hat{\sigma}^2$ decrease toward zero as p increases. Hence, the lack of consistency is not due to the use of the combined combined likelihood in place of the standard likelihood, but rather it may be related to the structure of the model.

In the second scenario we ran a simulation experiment with $p = 9, n = 10, 100, 500, 1000, 5000$ and $a \in \{-5, -1, 0, 0.5, 1, 2, 5\}$. Table 3.6 and Table 3.8 report the mean squared errors based on the combined composite likelihood obtained with $\theta_0 = (-1, 0, 1, 1)$ and $\theta_0 = (-2, 1, 3, 2)$, respectively. We notice that for admissible values of a, the MSE decreases toward zero as n increases. This seems to suggest, although empirically, the consistency of the combined composite likelihood estimator. For val-

Table 3.4: Model for microarray data. Mean squared error of $\hat{\theta}_C^a = (\hat{\beta}_1^a, \hat{\mu}_{tf}^a, \hat{\sigma}^{2a}, \hat{\sigma}_{tf}^{2a})$, when $n = 10, a = -5, -1, 0, 0.5, 1, 2, 5$ and $p = 3, 10, 100, 1000, 5000$. Results are obtained with 2000 simulated samples, and with the true parameter value $\theta_0 = (-2, 1, 3, 2)$.

		$\boldsymbol{a}$						
	$\boldsymbol{p}$	-5	-1	0	0.5	1	2	5
	3	0.0424	0.0420	0.0419	0.0419	0.0420	5.278710×10^{4}	1.570991×10^{5}
	10	0.0126	0.0125	0.0125	0.0125	0.0125	1.709610×10^{5}	1.656774×10^{5}
$\boldsymbol{MSE}(\hat{\beta}_1^a)$	100	0.0015	0.0012	0.0012	0.0012	0.0012	1.572923×10^{5}	2.157054×10^{5}
	1000	0.0716	0.0312	0.0204	0.0172	0.0098	1.444416×10^{5}	2.099350×10^{5}
	5000	0.6107	0.4775	0.3180	0.2842	0.2296	1.143040×10^{5}	1.544233×10^{5}
	3	0.2074	0.2050	0.2024	0.1991	0.2121	5.895803×10^{4}	3.177277×10^{4}
	10	0.2121	0.2119	0.2114	0.2108	0.2081	5.063908×10^{4}	2.192632×10^{4}
$\boldsymbol{MSE}(\hat{\mu}_{tf}^a)$	100	0.1988	0.2004	0.2004	0.2004	0.2004	2.888373×10^{4}	2.868215×10^{4}
	1000	0.2618	0.2201	0.2152	0.2149	0.2114	3.081039×10^{3}	1.072239×10^{4}
	5000	0.8730	0.7566	0.6918	0.6246	0.5360	3.618577×10^{3}	6.011371×10^{3}
	3	0.6507	0.6281	0.6288	0.6318	0.6391	1.653355×10^{180}	8.636679×10^{301}
	10	0.1904	0.1891	0.1892	0.1895	0.1903	2.417218×10^{156}	2.489486×10^{300}
$\boldsymbol{MSE}(\hat{\sigma}^{2a})$	100	0.0186	0.0185	0.0185	0.0185	0.0184	8.378486×10^{79}	4.098673×10^{302}
	1000	0.0020	0.0019	0.0019	0.0019	0.0018	3.187088×10^{64}	3.967151×10^{301}
	5000	0.0005	0.0005	0.0005	0.0005	0.0004	1.982706×10^{60}	7.388832×10^{300}
	3	0.7161	0.7153	0.7134	0.7109	0.7451	7.350199×10^{305}	7.164529×10^{301}
	10	0.7544	0.7619	0.7624	0.7503	0.7549	9.085017×10^{304}	3.115345×10^{301}
$\boldsymbol{MSE}(\hat{\sigma}_{tf}^{2a})$	100	0.7835	0.7870	0.7862	0.7859	0.7826	6.782670×10^{294}	1.213578×10^{294}
	1000	1.2892	0.9880	0.9420	0.9378	0.7877	2.220682×10^{293}	4.607355×10^{292}
	5000	4.4799	3.8167	3.4749	3.2181	2.6509	5.204166×10^{293}	2.012821×10^{292}

Table 3.5: Model for microarray data. Mean squared error of $\hat{\theta} = (\hat{\beta}_1, \hat{\mu}_{tf}, \hat{\sigma}^2, \hat{\sigma}^2_{tf})$, when $n = 10$ and $p = 3, 10, 100, 1000, 5000$. Results are obtained with 2000 simulated samples, and with the true parameter value $\theta_0 = (-2, 1, 3, 2)$.

	$\boldsymbol{MSE(\hat{\beta}_1)}$	$\boldsymbol{MSE(\hat{\mu}_{tf})}$	$\boldsymbol{MSE(\hat{\sigma}^2)}$	$\boldsymbol{MSE(\hat{\sigma}^2_{tf})}$
$p = 3$	0.0416	0.1958	0.5925	0.7067
$p = 10$	0.0124	0.2067	0.1764	0.7445
$p = 100$	0.0012	0.2009	0.0182	0.7895
$p = 1000$	0.0001	0.2452	0.0018	1.0832
$p = 5000$	0.0000	0.4275	0.0004	3.4067

ues of a greater than one, the combined composite likelihood estimates are very far from the true values. On the other hand, Table 3.7 and Table 3.9 report the analogous quantities based on the full likelihood. The MSE decreases toward zero which confirms empirically the consistency of the maximum likelihood estimator. In this situation, the consistency of the estimator based on both combined composite and full likelihood was expected from large-sample theory.

We now look at the efficiency of the combined composite likelihood estimator ($\hat{\theta}^a_C$) with respect to the full likelihood estimator ($\hat{\theta}$). Since the parameter is multidimensional, we suppose that it is of interest to compare the efficiency of the estimators for the single component β_1, being the regression parameter, we use the appropriate measure defined in (2.10).

Figure 3.5 displays the efficiency of the estimator of β_1 based on combined composite likelihood compared to the one based on the full likelihood for fixed a as p increases. As expected, using a combined composite likelihood in place of the standard likelihood, the four plots illustrate a loss of efficiency as p increases. In order to choose a value of a for which the efficiency is high, we present Figure 3.6 which shows the behavior of

Table 3.6: Model for microarray data. Mean squared error of $\hat{\theta}_C^a = (\hat{\beta}_1^a, \hat{\mu}_{tf}^a, \hat{\sigma}^{2a}, \hat{\sigma}_{tf}^{2a})$, when $p = 9, a = -5, -1, 0, 0.5, 1, 2, 5$ and $n = 10, 100, 500, 1000, 5000$. Results are obtained with 2000 simulated samples, and with the true parameter value $\theta_0 = (-1, 0, 1, 1)$.

		a						
	n	-5	-1	0	0.5	1	2	5
$MSE(\hat{\beta}_1^a)$	10	0.0224	0.0144	0.0144	0.0144	0.0145	3.759398×10^{5}	4.472782×10^{5}
	100	0.0011	0.0011	0.0011	0.0011	0.0011	6.274038×10^{5}	2.562903×10^{5}
	500	0.0002	0.0002	0.0002	0.0002	0.0002	8.194322×10^{5}	2.596649×10^{5}
	1000	0.0001	0.0001	0.0001	0.0001	0.0001	8.822953×10^{5}	2.609885×10^{5}
	5000	0.0000	0.0000	0.0000	0.0000	0.0000	1.323419×10^{6}	2.707208×10^{5}
$MSE(\hat{\mu}_{tf}^a)$	10	0.1097	0.1094	0.1088	0.1081	0.1061	6.286917×10^{4}	5.163544×10^{4}
	100	0.0102	0.0101	0.0101	0.0100	0.0098	2.098034×10^{5}	1.261197×10^{4}
	500	0.0022	0.0022	0.0022	0.0022	0.0021	2.741952×10^{5}	7.834310×10^{3}
	1000	0.0010	0.0010	0.0010	0.0010	0.0010	2.817887×10^{5}	6.783551×10^{3}
	5000	0.0002	0.0002	0.0002	0.0002	0.0002	4.572647×10^{5}	4.673932×10^{3}
$MSE(\hat{\sigma}^{2a})$	10	0.0247	0.0239	0.0239	0.0239	0.0239	1.397317×10^{65}	1.045596×10^{302}
	100	0.0023	0.0024	0.0024	0.0024	0.0024	1.941055×10^{118}	4.634396×10^{123}
	500	0.0005	0.0005	0.0005	0.0005	0.0005	3.818256×10^{121}	6.211956×10^{105}
	1000	0.0002	0.0002	0.0002	0.0002	0.0002	7.612732×10^{66}	2.968534×10^{104}
	5000	0.0000	0.0000	0.0000	0.0000	0.0000	4.819678×10^{67}	5.418265×10^{105}
$MSE(\hat{\sigma}_{tf}^{2a})$	10	0.1835	0.1835	0.1836	0.1838	0.1842	5.840136×10^{299}	4.808050×10^{294}
	100	0.0196	0.0196	0.0196	0.0196	0.0196	2.742888×10^{299}	3.540877×10^{293}
	500	0.0041	0.0041	0.0041	0.0041	0.0041	4.778824×10^{302}	1.624803×10^{291}
	1000	0.0021	0.0021	0.0021	0.0021	0.0021	8.562701×10^{302}	6.307040×10^{290}
	5000	0.0004	0.0004	0.0004	0.0004	0.0004	6.211489×10^{301}	2.687328×10^{287}

Table 3.7: Model for microarray data. Mean squared error of $\hat{\theta} = (\hat{\beta}_1, \hat{\mu}_{tf}, \hat{\sigma}^2, \hat{\sigma}^2_{tf})$, when $p = 9$ and $n = 10, 100, 500, 1000, 5000$. Results are obtained with 2000 simulated samples, and with the true parameter value $\theta_0 = (-1, 0, 1, 1)$.

	$\boldsymbol{MSE(\hat{\beta}_1)}$	$\boldsymbol{MSE(\hat{\mu}_{tf})}$	$\boldsymbol{MSE(\hat{\sigma}^2)}$	$\boldsymbol{MSE(\hat{\sigma}^2_{tf})}$
$n = 10$	0.0139	0.1032	0.0220	0.1832
$n = 100$	0.0011	0.0096	0.0022	0.0195
$n = 500$	0.0002	0.0021	0.0005	0.0041
$n = 1000$	0.0001	0.0010	0.0002	0.0021
$n = 5000$	0.0000	0.0002	0.0000	0.0004

the efficiency for different values of a when p is fixed. As we see in each plot, except for a small range of values of β_1, the efficiency improves as a increases. Hence, $a = 1$ seems to be the best choice among the admissible values.

Figure 3.7 displays the efficiency of the estimator of μ_{tf} based on combined composite likelihood compared to the one based on the full likelihood. In each plot, a is fixed and p increases. As we see from the plots, the efficiency improves as p increases. This result is due to the fact that only the one-dimensional marginal distributions depend on μ_{tf} and hence, there is more information as p increases. Even when p is fixed, the efficiency improves as a increases. This result is highlighted by plots in Figure 3.8. In addition, it seems that $a = 1$ is the best choice among the admissible values.

Figure 3.9 displays the efficiency of the estimator of σ^2 based on combined composite likelihood compared to the one based on the full likelihood. Each plot shows the behavior of the efficiency for a fixed as p increases. This is one of the situations in which one could not draw conclusions on the behavior of efficiency as p increases because the curves of

Table 3.8: Model for microarray data. Mean squared error of $\hat{\theta}^a_C = (\hat{\beta}^a_1, \hat{\mu}^a_{tf}, \hat{\sigma}^{2a}, \hat{\sigma}^{2a}_{tf})$, when $p = 9, a = -5, -1, 0, 0.5, 1, 2, 5$ and $n = 10, 100, 500, 1000, 5000$. Results are obtained with 2000 simulated samples, and with the true parameter value $\theta_0 = (-2, 1, 3, 2)$.

		$\boldsymbol{a}$						
	$\boldsymbol{n}$	-5	-1	0	0.5	1	2	5
$\boldsymbol{MSE}(\hat{\beta}^a_1)$	10	0.0144	0.0143	0.0143	0.0142	0.0142	6.593243×10^{104}	9.280760×10^{151}
	100	0.0011	0.0011	0.0011	0.0011	0.0011	2.995195×10^{104}	1.689540×10^{104}
	500	0.0002	0.0002	0.0002	0.0002	0.0002	1.846842×10^{104}	8.140274×10^{103}
	1000	0.0001	0.0001	0.0001	0.0001	0.0001	1.610371×10^{104}	7.525042×10^{103}
	5000	0.0000	0.0000	0.0000	0.0000	0.0000	9.306610×10^{103}	2.966837×10^{103}
$\boldsymbol{MSE}(\hat{\mu}^a_{tf})$	10	0.2110	0.2105	0.2099	0.2092	0.2071	3.465411×10^{159}	2.600087×10^{157}
	100	0.0196	0.0195	0.0195	0.0194	0.0192	5.174811×10^{159}	4.316354×10^{156}
	500	0.0042	0.0042	0.0042	0.0042	0.0041	8.848096×10^{158}	1.804273×10^{155}
	1000	0.0020	0.0020	0.0020	0.0020	0.0019	3.915642×10^{133}	4.469000×10^{120}
	5000	0.0004	0.0004	0.0004	0.0004	0.0004	3.744291×10^{159}	1.256356×10^{155}
$\boldsymbol{MSE}(\hat{\sigma}^{2a})$	10	0.2186	0.2164	0.2165	0.2167	0.2171	1.659716×10^{103}	1.625001×10^{100}
	100	0.0213	0.0213	0.0213	0.0213	0.0213	1.004332×10^{102}	8.915114×10^{100}
	500	0.0044	0.0044	0.0044	0.0044	0.0044	4.855651×10^{102}	2.188066×10^{100}
	1000	0.0021	0.0021	0.0021	0.0021	0.0021	4.302051×10^{103}	4.802324×10^{99}
	5000	0.0004	0.0004	0.0004	0.0004	0.0004	1.131754×10^{103}	1.048383×10^{99}
$\boldsymbol{MSE}(\hat{\sigma}^{2a}_{tf})$	10	0.7415	0.7432	0.7437	0.7435	0.7428	3.223752×10^{71}	1.287497×10^{27}
	100	0.0794	0.0796	0.0797	0.0796	0.0791	3.850310×10^{43}	2.876550×10^{25}
	500	0.0166	0.0167	0.0167	0.0166	0.0165	3.523973×10^{48}	2.801829×10^{21}
	1000	0.0082	0.0082	0.0082	0.0082	0.0082	4.900425×10^{68}	2.114810×10^{21}
	5000	0.0016	0.0016	0.0016	0.0016	0.0016	2.164125×10^{46}	2.947112×10^{20}

Table 3.9: Model for microarray data. Mean squared error of $\hat{\theta} = (\hat{\beta}_1, \hat{\mu}_{tf}, \hat{\sigma}^2, \hat{\sigma}^2_{tf})$, when $p = 9$ and $n = 10, 100, 500, 1000, 5000$. Results are obtained with 2000 simulated samples, and with the true parameter value $\theta_0 = (-2, 1, 3, 2)$.

	$MSE(\hat{\beta}_1)$	$MSE(\hat{\mu}_{tf})$	$MSE(\hat{\sigma}^2)$	$MSE(\hat{\sigma}^2_{tf})$
$n = 10$	0.01403	0.20632	0.19786	0.73268
$n = 100$	0.00113	0.01918	0.01963	0.07818
$n = 500$	0.00023	0.00414	0.00407	0.01638
$n = 1000$	0.00011	0.00194	0.00196	0.00821
$n = 5000$	0.00002	0.00042	0.00041	0.00154

efficiency corresponding to different values of p, intersect in some values of β. However, the plots in Figure 3.10 suggest that for fixed p, the efficiency gets better as a increases except for smaller values of σ^2. Also in this case, $a = 1$ seems to be the best choice.

Figures 3.11 and 3.12 display the efficiency of the estimator of σ^2_{tf} based on combined composite likelihood compared to the one based on the full likelihood. From this graphical analysis, it follows that overall one may consider $a = 1$ as the best choice.

For this model, the efficiency is particularly high in all cases. Therefore, the combined composite likelihood may be considered as a good alternative to the full likelihood. In both examples we propose as the best choice to use $a = 1$, which corresponds to the pairwise conditional likelihood.

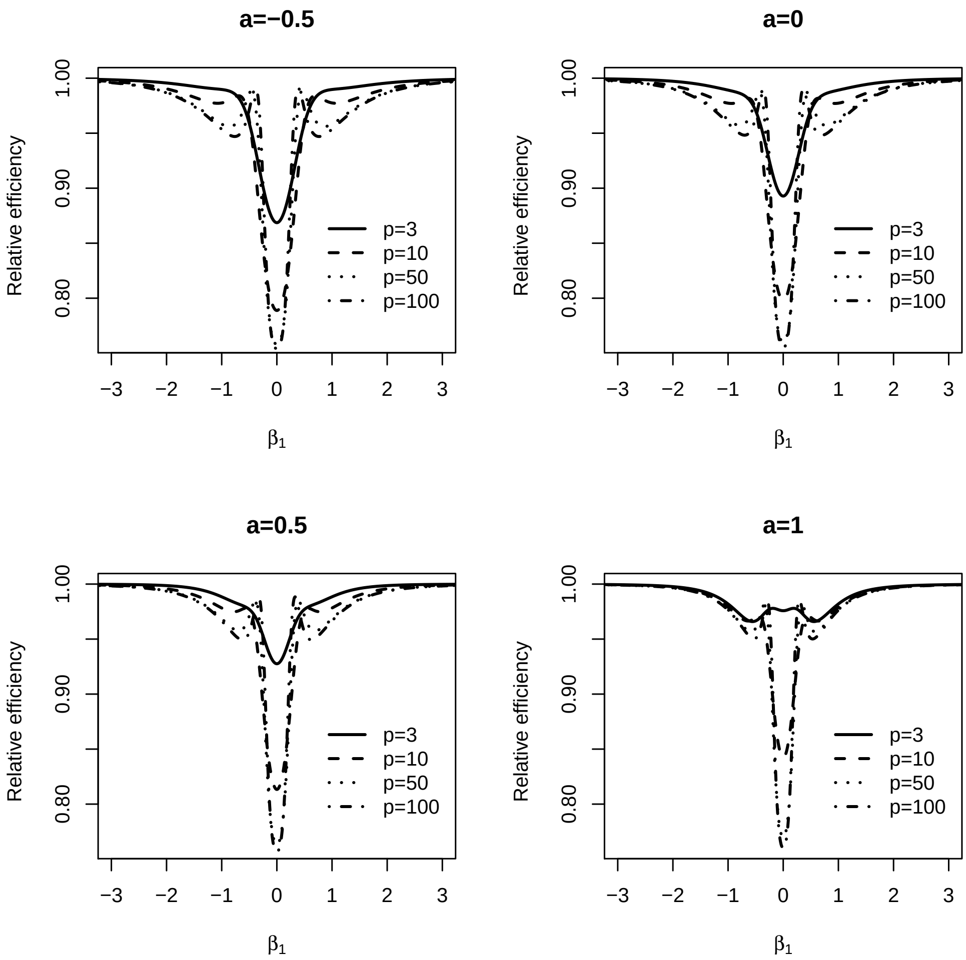

Figure 3.5: Model for microarray data. Comparisons of efficiency of $\hat{\beta}_1^a$ relative to $\hat{\beta}_1$ for fixed a, different values of p and fixed values for the other parameters.

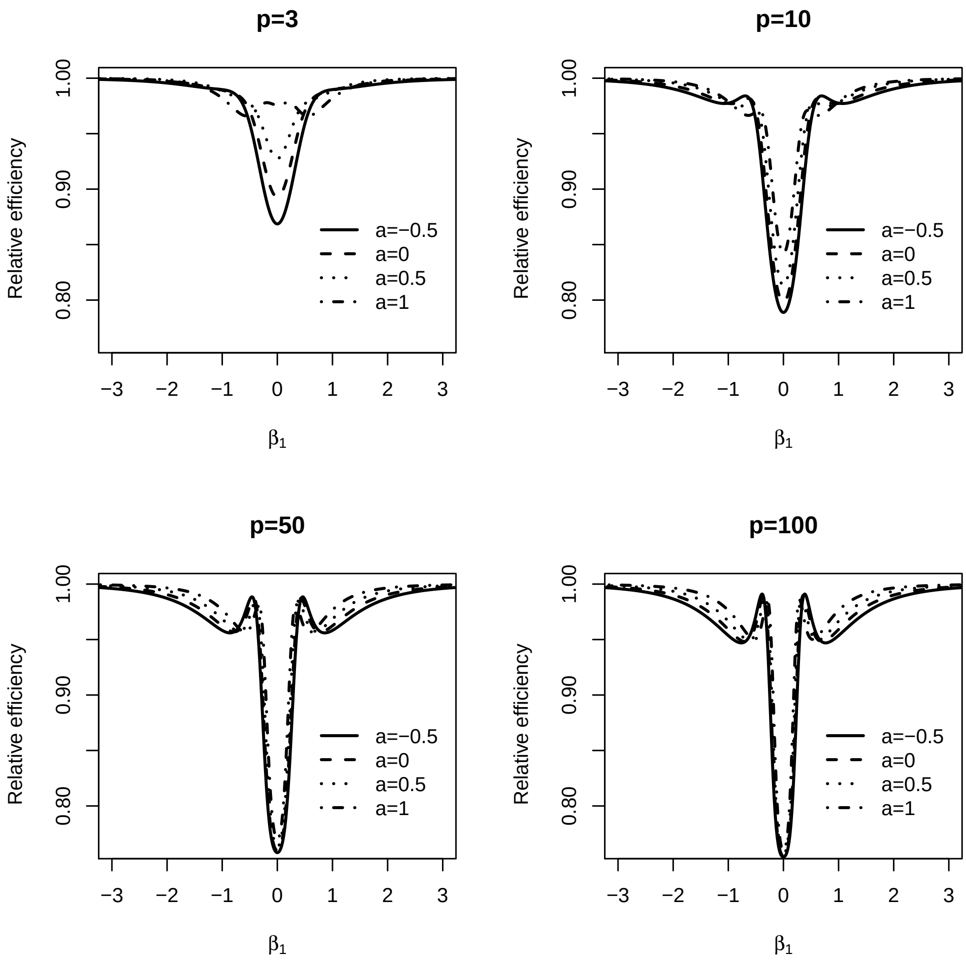

Figure 3.6: Model for microarray data. Comparisons of efficiency of $\hat{\beta}_1^a$ relative to $\hat{\beta}_1$ for fixed p, different values of a and fixed values for the other parameters.

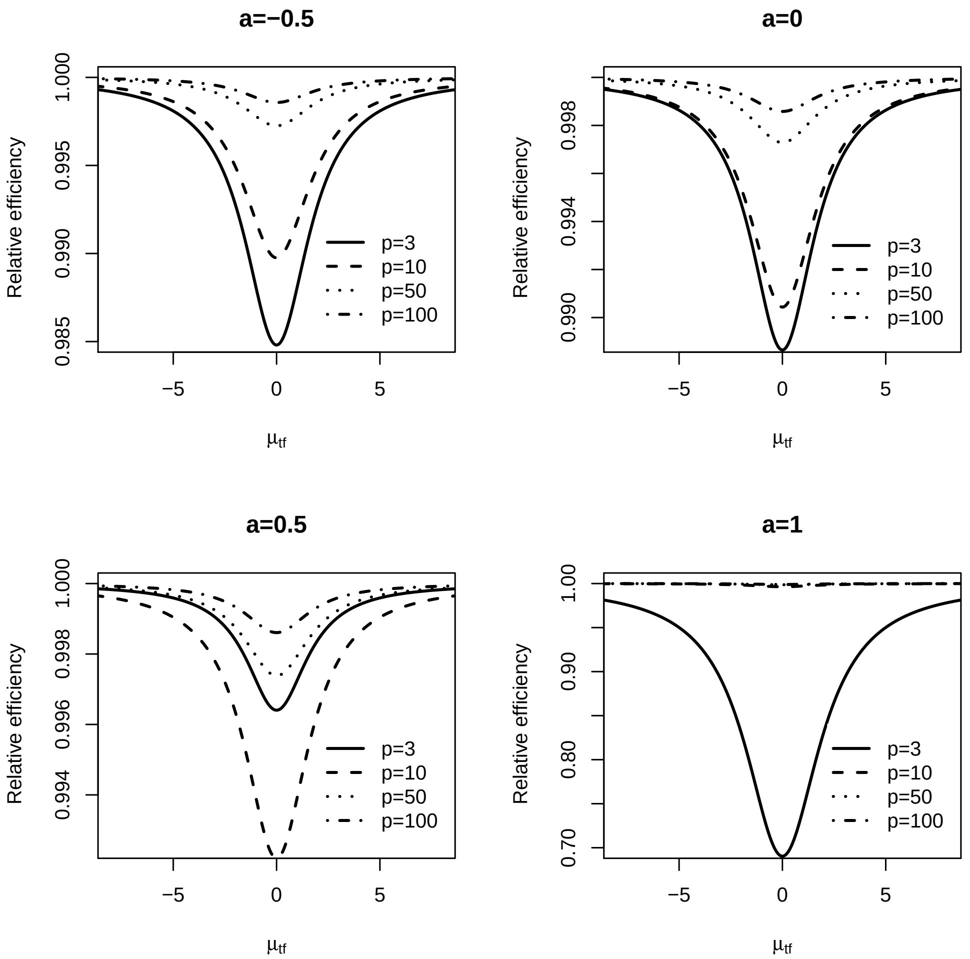

Figure 3.7: Model for microarray data. Comparisons of efficiency of $\hat{\mu}_{tf}^{a}$ relative to $\hat{\mu}_{tf}$ for fixed a, different values of p and fixed values for the other parameters.

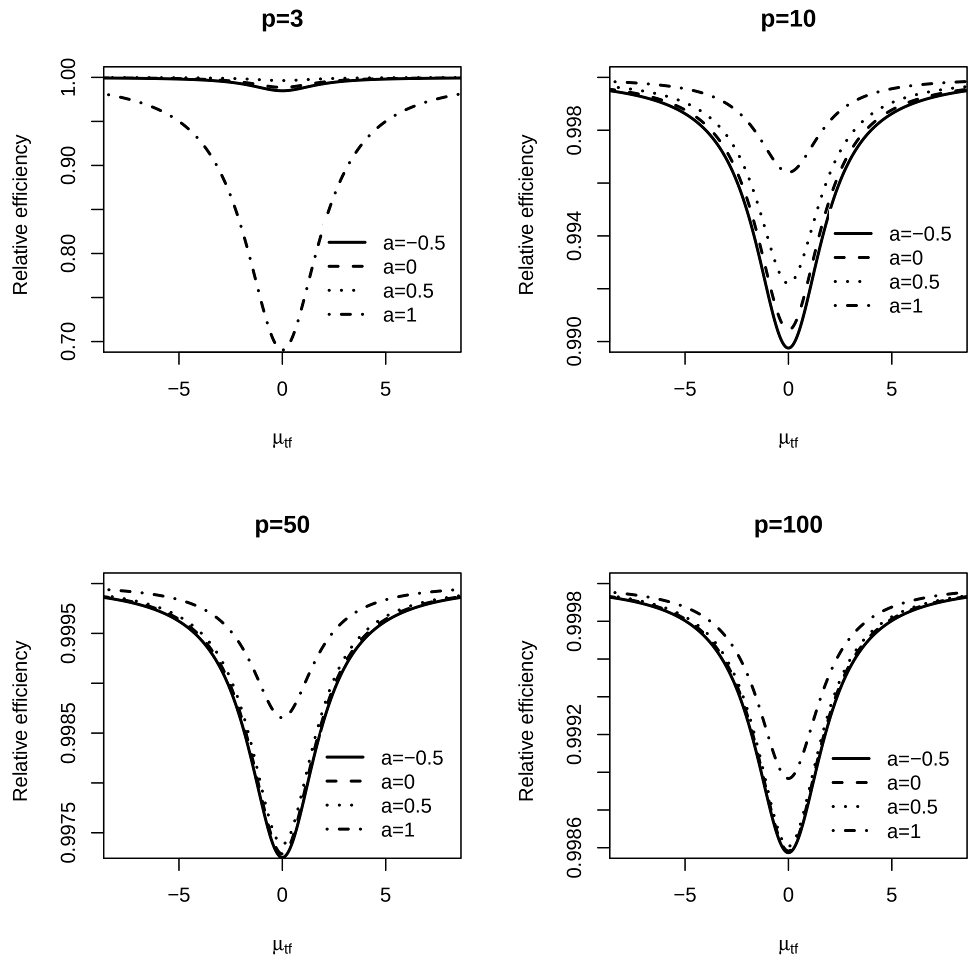

Figure 3.8: Model for microarray data. Comparisons of efficiency of $\hat{\mu}^a_{tf}$ relative to $\hat{\mu}_{tf}$ for fixed p, different values of a and fixed values for the other parameters.

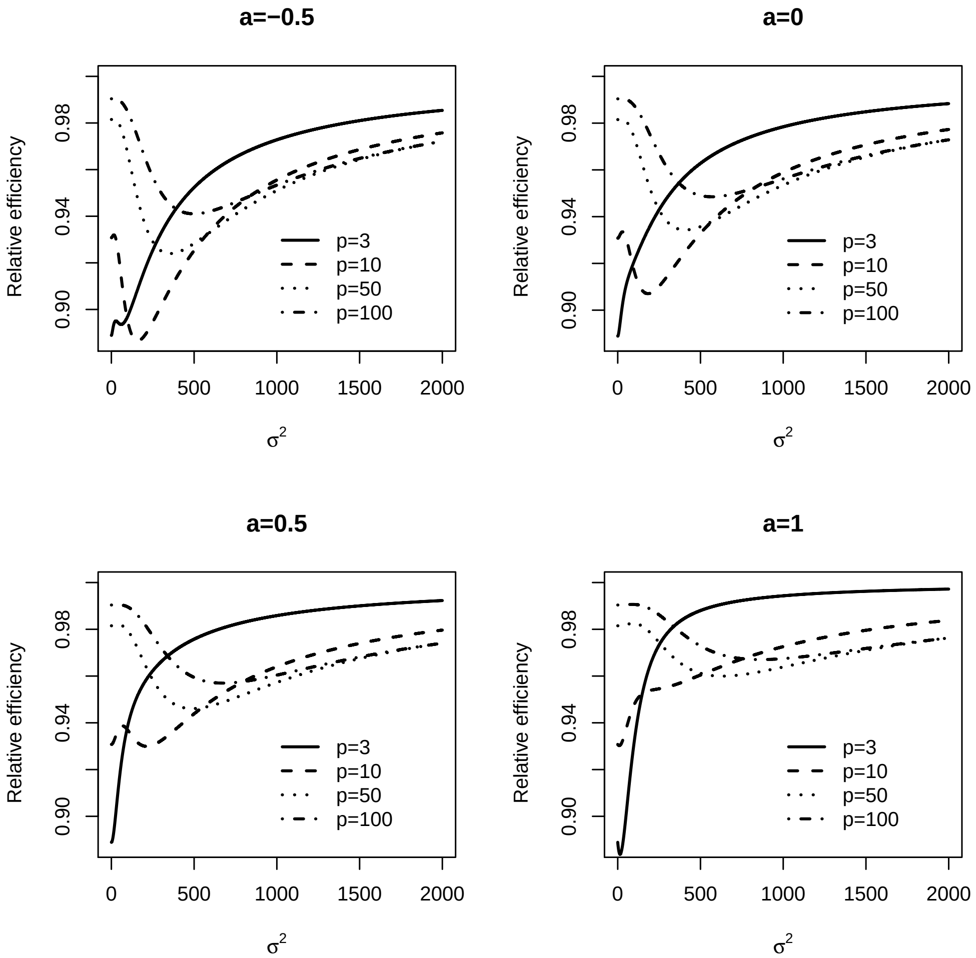

Figure 3.9: Model for microarray data. Comparisons of efficiency of $\hat{\sigma}^{2a}$ relative to $\hat{\sigma}^2$ for fixed a, different values of p and fixed values for the other parameters.

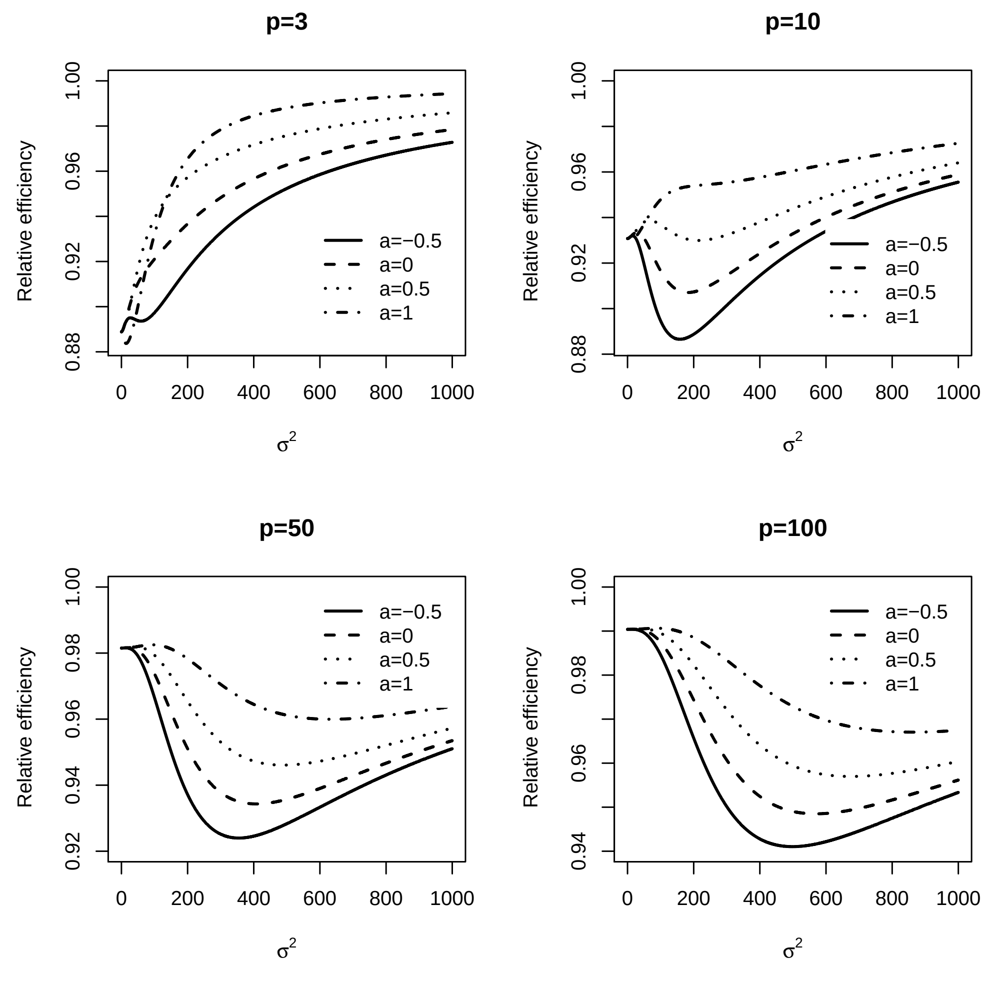

Figure 3.10: Model for microarray data. Comparisons of efficiency of $\hat{\sigma}^{2a}$ relative to $\hat{\sigma}^2$ for fixed p, different values of a and fixed values for the other parameters.

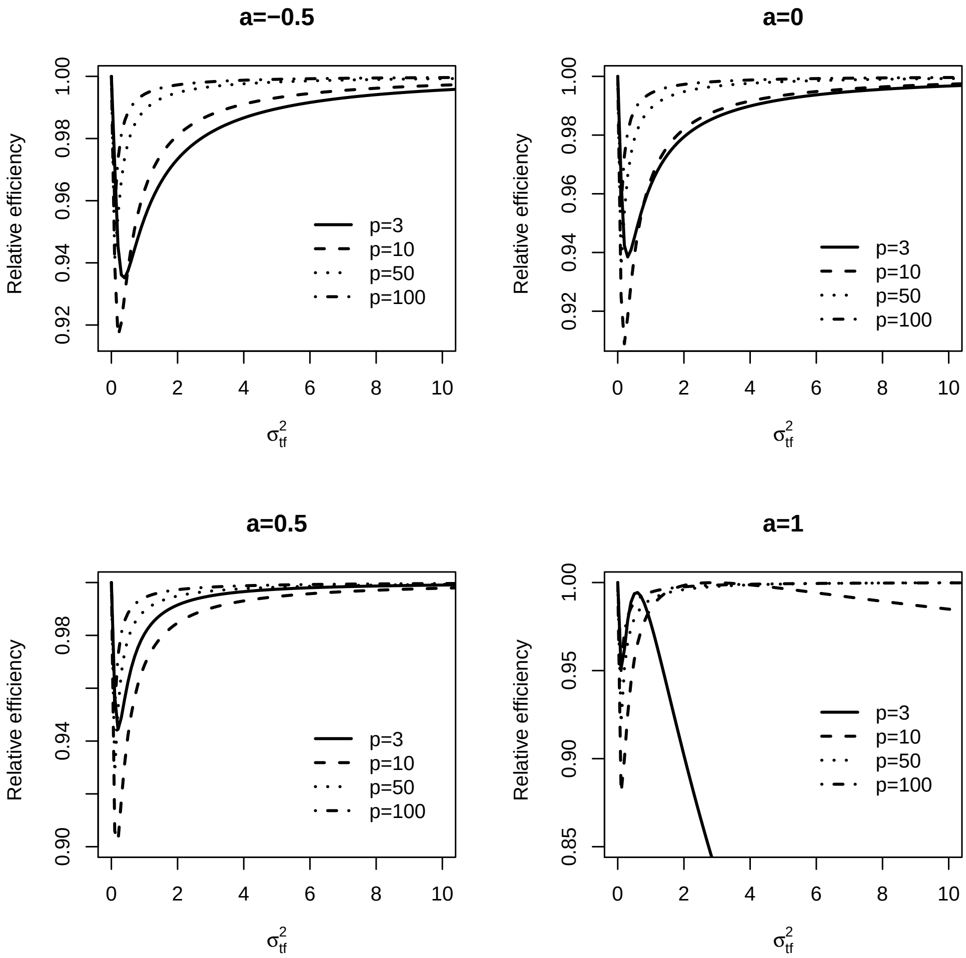

Figure 3.11: Model for microarray data. Comparisons of efficiency of $\hat{\sigma}^{2a}_{tf}$ relative to $\hat{\sigma}^2_{tf}$ for fixed a, different values of p and fixed values for the other parameters.

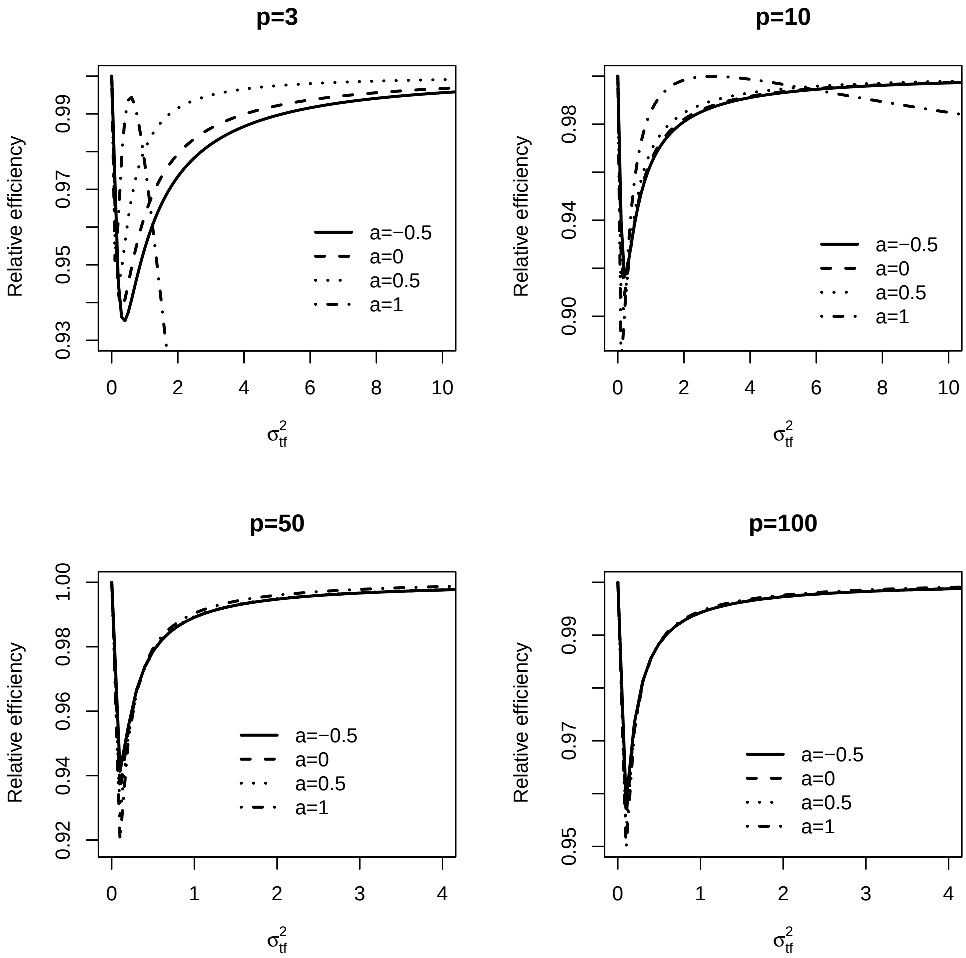

Figure 3.12: Model for microarray data. Comparisons of efficiency of $\hat{\sigma}^{2a}_{tf}$ relative to $\hat{\sigma}^2_{tf}$ for fixed p, different values of a and fixed values for the other parameters.

Chapter 4

Weighted independence likelihood and prediction

4.1 Introduction and motivations

We consider multivariate problems where the ordinary likelihood is unknown or too time-consuming to compute and where the prediction of the future observations of a subset of variables could be of interest. The use of an appropriate pseudo-likelihood in a prediction framework could be a possible solution to this problem. In this situation, it may be useful a well-known example of composite marginal likelihood, the independence likelihood (Chandler & Bate, 2007), which is constructed by using only the univariate marginal densities under the working assumption of independence. Since in the construction of the independence likelihood it is assumed the independence between variables when they are actually dependent, it might be appropriate to give different weights to the univariate marginal densities, obtaining a weighted independence likelihood. In this way, we seek to improve prediction on the variables of interest.

The focus here is on determining the set of weights in order to have a good prediction of the variables of interest, considering the remaining components as auxiliary variables, and also to find suitable definition of weighted independence likelihood. We use a cross-validation procedure based on a particular empirical measure as a criterion for determining the weights.

The present chapter is a first attempt to enter this topic and provide some directions for future works. In particular, Section 4.2 defines the weighted independence likelihood and proposes a criterion for choosing the weights, while Section 4.3 gives two simple examples with preliminary simulation results.

4.2 Weighted independence likelihoods

We want to choose the weights that lead to a good prediction of the variables of interest. Before taking into account the criterion, we describe the forms of the weighted independence likelihood that we have in mind.

Let $\mathcal{F} = \{f(y;\theta) : \theta \in \Theta \subseteq \mathbb{R}^d, y \in \mathcal{Y} \subseteq \mathbb{R}^q\}$ be a statistical parametric model. Let $Y_1, \dots, Y_q$ be dependent random vectors, where $Y_r = (Y_{1r}, Y_{2r}, \dots, Y_{nr})^{\mathrm{T}}, r = 1, \dots, q$, with probability density functions $f_1(\cdot;\theta), \dots, f_q(\cdot;\theta)$. We assume that $Y_{1r}, Y_{2r}, \dots, Y_{nr}$ are independent and identically distributed random variables, for any given $r = 1, \dots, q$. Moreover, we also suppose without loss of generality that the prediction of $Y_{(n+1)1}$ is of interest. In this situation, it could be interesting to use a weighted independence log-likelihood of the form

$$c\ell_1^{\mathrm{I}}(\theta; y) = \ell(\theta; y_1) + \sum_{i=2}^{q} w_i \ell(\theta; y_i), \quad 0 < w_i \leq 1,\ i = 2, \dots, q. \tag{4.1}$$

As we can see in this first form, the likelihood contribution of the variable Y_1 has weight equal to 1, while the remaining components could be down-weighted. In this way, we implicitly seek to give more importance to the variable Y_1 being of interest the prediction of its future observation, while assuming that the remaining variables could still provide useful information on θ.

A different proposal is a classic convex combination of the weights, with which we get the weighted independence log-likelihood

$$c\ell_2^{\mathrm{I}}(\theta; y) = \sum_{i=1}^{q} w_i \ell(\theta; y_i), \text{where} \sum_{i=1}^{q} w_i = 1. \tag{4.2}$$

Returning to the primary aim, a good prediction of $Y_{(n+1)1}$ with minimal error is given by $\phi(\theta) = \mathrm{E}_\theta\{Y_{(n+1)1} \mid Y_{n1}, Y_{(n-1)1}, \ldots, Y_{11}\} = \mathrm{E}_\theta\{Y_{11}\}$ due to the independence assumption. As in Wang & Zidek (2005) in a weighted likelihood framework, we use the delete-one approach in a cross validation procedure based on the empirical measure

$$D(w) = \sum_{i=1}^{n} \left(y_{i1} - \phi(\tilde{\theta}^{(-i)}) \right)^2,$$

where $w = (w_1, \ldots, w_q)$ is the set of weights and $\tilde{\theta}^{(-i)}$ is the estimate of θ without using $y_{i1}, y_{i2}, \ldots, y_{iq}$. In $D(w)$, we predict $Y_{i1}, i = 1, \ldots, n$, by $\phi(\tilde{\theta}^{(-i)})$, the estimator of $\phi(\theta)$ based either on one of the two forms of the weighted independence likelihood defined in (4.1) and (4.2), or on the unweighted independence likelihood where all weights are equal. The optimum weights are obtained minimizing $D(w)$. We can easily see from the form of $D(w)$ that the weights are therefore data-dependent.

4.3 Examples and simulation results

In this section we present two examples in order to evaluate the performance on the prediction of the variable of interest using a weighted independence likelihood in place of the standard likelihood. Moreover, we also compare the results with the classical unweighted independence likelihood. In particular, the first example deals with a bivariate Poisson model and the second one with a bivariate normal model. In both examples, the parameter is scalar.

4.3.1 Bivariate Poisson model

Let $X_i \sim \text{Poisson}(\theta_i)\,, i = 1, 2$ and $X_0 \sim \text{Poisson}(\theta)$, be independent random variables. Consider the random variables $Y_1 = X_1 + X_0$ and $Y_2 = X_2 + X_0$, then (Y_1, Y_2) is a bivariate Poisson, $(Y_1, Y_2) \sim BP(\theta_1, \theta_2, \theta)$, with joint distribution function given by

$$P(Y_1 = y_1, Y_2 = y_2) = e^{-(\theta_1+\theta_2+\theta)} \frac{\theta_1^{y_1}}{y_1!} \frac{\theta_2^{y_2}}{y_2!} \sum_{i=1}^{\min(y_1,y_2)} \binom{y_1}{i} \binom{y_2}{i} i! \left(\frac{\theta}{\theta_1\theta_2}\right)^i .$$

The marginal distributions are Poisson, i.e.

$$Y_1 \sim \text{Poisson}(\theta + \theta_1)$$
$$Y_2 \sim \text{Poisson}(\theta + \theta_2)$$

and $\text{Cov}(X, Y) = \theta$, $\text{Cor}(X, Y) = \theta/\sqrt{(\theta_1 + \theta)(\theta_2 + \theta)}$, which is always positive.

In the following, θ_1 and θ_2 are considered as fixed and hence, the parameter of interest is θ. The calculation of the full likelihood based on n independent observations is

$$\ell(\theta) = -n\theta + \sum_{i=1}^{n} \log \left\{ \sum_{j=1}^{\min(y_{i1},y_{i2})} \binom{y_{i1}}{j} \binom{y_{i2}}{j} j! \left(\frac{\theta}{\theta_1\theta_2}\right)^j \right\},$$

while the weighted independence likelihoods defined in (4.1) and (4.2) are respectively

$$c\ell_1^{\text{I}}(\theta) = -n\theta(1 + w_2) + \sum_{i=1}^{n} y_{i1}\log(\theta_1 + \theta) + w_2 \sum_{i=1}^{n} y_{i2}\log(\theta_2 + \theta),$$
$$c\ell_2^{\text{I}}(\theta) = -n\theta + w_1 \sum_{i=1}^{n} y_{i1}\log(\theta_1 + \theta) + (1 - w_1) \sum_{i=1}^{n} y_{i2}\log(\theta_2 + \theta).$$

We now perform a numerical assessment of the mean squared prediction error, whose expression is given by

$$MSPE = \text{E}_{\theta_0} \left\{ \left(Y_{(n+1)1} - \phi(\tilde{\tilde{\theta}}) \right)^2 \right\},$$

Table 4.1: Bivariate Poisson model. Comparison of the mean squared prediction error based on the full likelihood ($MSPE_F$), the weighted independence likelihood ($MSPE_W$) and the unweighted independence likelihood ($MSPE_U$), when $n = 10, 50, 100, 200$. Results are obtained with 2000 simulated samples, and with the true parameter values $\theta_0 = (10, 1, 9)$, $\theta_0 = (10, 9, 1)$ and $\theta_0 = (10, 10, 10)$. The weighted independence likelihood estimator is based on $c\ell_1^{\mathrm{I}}(\theta)$.

	$\theta_0 = (10, 1, 9)$			$\theta_0 = (10, 9, 1)$			$\theta_0 = (10, 10, 10)$		
	$MSPE_F$	$MSPE_W$	$MSPE_U$	$MSPE_F$	$MSPE_W$	$MSPE_U$	$MSPE_F$	$MSPE_W$	$MSPE_U$
n= 10	12.0367	12.1020	12.1440	19.9311	20.1091	19.9224	20.9556	21.1209	20.9820
n= 50	11.3624	11.3659	11.4006	19.4642	19.4866	19.4633	20.5201	20.5470	20.5262
n=100	10.8148	10.8204	10.8280	18.5581	18.5715	18.5642	19.3630	19.3621	19.3584
n=200	10.5974	10.5956	10.6061	17.8972	17.9083	17.9055	19.8898	19.9128	19.8853

where $\tilde{\tilde{\theta}}$ is the estimate of θ based on the weighted independence likelihood corresponding to the optimal set of weights, or on the unweighted independence likelihood, or on the full likelihood, when the latter is available. To this end, we ran a simulation experiment, with $n = 10, 50, 100, 200$. For each combination, we performed 2000 iterations. Indicating by $\theta_0 = (\theta, \theta_1, \theta_2)$, the results correspond to $\theta_0 = (10, 1, 9)$, $\theta_0 = (10, 9, 1)$ and $\theta_0 = (10, 10, 10)$ as the true parameter values. Table 4.1 reports the mean squared prediction errors based on $c\ell_1^{\mathrm{I}}(\theta)$ compared to the ones based on both full likelihood and unweighted independence likelihood. As we can see from the table, for small sample sizes, in terms of prediction it appears that the weighted independence likelihood is preferable to the unweighted independence likelihood when $\theta_1 < \theta_2$. In the other two situations, either $\theta_1 > \theta_2$ or $\theta_1 = \theta_2$, it seems there is no gain in using the weighted independence likelihood in place of the unweighted independence likelihood. As expected, the full likelihood works better with respect to the two pseudo-likelihoods.

Table 4.2 reports the mean squared prediction errors based on $c\ell_2^{\mathrm{I}}(\theta)$, full likelihood and unweighted independence likelihood. As we can see,

Table 4.2: Bivariate Poisson model. Comparison of the mean squared prediction error based on the full likelihood ($MSPE_F$), the weighted independence likelihood ($MSPE_W$) and the unweighted independence likelihood ($MSPE_U$), when $n = 10, 50, 100, 200$. Results are obtained with 2000 simulated samples, and with the true parameter values $\theta_0 = (10, 1, 9)$, $\theta_0 = (10, 9, 1)$ and $\theta_0 = (10, 10, 10)$. The weighted independence likelihood estimator is based on $c\ell_2^{\mathrm{I}}(\theta)$.

	$\theta_0 = (10, 1, 9)$			$\theta_0 = (10, 9, 1)$			$\theta_0 = (10, 10, 10)$		
	$MSPE_F$	$MSPE_W$	$MSPE_U$	$MSPE_F$	$MSPE_W$	$MSPE_U$	$MSPE_F$	$MSPE_W$	$MSPE_U$
n= 10	12.0367	12.1663	12.1440	19.9311	20.1282	19.9224	20.9556	21.2277	20.9820
n= 50	11.3624	11.3752	11.4006	19.4642	19.4807	19.4633	20.5201	20.5840	20.5262
n=100	10.8148	10.8223	10.8280	18.5581	18.5621	18.5642	19.3630	19.3678	19.3584
n=200	10.5974	10.5917	10.6061	17.8972	17.8999	17.9055	19.8898	19.9144	19.8853

for small sample sizes, it still seems that the weighted independence likelihood is preferable to the unweighted independence likelihood when $\theta_1 < \theta_2$. While, when $\theta_1 > \theta_2$, the weighted independence likelihood seems to work better for some values of n. In the last case, with $\theta_1 = \theta_2$, the unweighted independence likelihood seems to be preferable.

4.3.2 Bivariate normal model

The random vector $(Y_1, Y_2)^{\mathrm{T}}$ follows a bivariate normal model with mean vector $(\mu, \mu)^{\mathrm{T}}$ and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_{y_1}^2 & \sigma_{y_1}\sigma_{y_2}\rho \\ \sigma_{y_1}\sigma_{y_2}\rho & \sigma_{y_2}^2 \end{pmatrix}.$$

The parameter constraint for which Σ is positive definite is $\rho \in (-1, 1)$. The unknown parameter is μ and the other parameters are considered as fixed. The calculation of the full likelihood based on n independent observations gives

$$\ell(\mu) = -\frac{1}{(1-\rho^2)}\left[\frac{1}{2\sigma_{y_1}^2}\sum_{i=1}^{n}(y_{i1}-\mu)^2 - \frac{\rho}{\sigma_{y_1}\sigma_{y_2}}\sum_{i=1}^{n}(y_{i1}-\mu)(y_{i2}-\mu)\right.$$
$$\left.+\frac{1}{2\sigma_{y_2}^2}\sum_{i=1}^{n}(y_{i2}-\mu)^2\right].$$

The maximum full likelihood estimator is given by

$$\hat{\mu} = \frac{\bar{y}_1\sigma_{y_2}^2 + \bar{y}_2\sigma_{y_1}^2 - \rho\sigma_{y_1}\sigma_{y_2}(\bar{y}_1+\bar{y}_2)}{\sigma_{y_1}^2+\sigma_{y_2}^2-2\rho\sigma_{y_1}\sigma_{y_2}},$$

where $\bar{y}_1 = \sum_{i=1}^{n} y_{i1}/n$ and $\bar{y}_2 = \sum_{i=1}^{n} y_{i2}/n$. In this case, the calculation of weighted independence likelihood defined in (4.1) and (4.2) gives

$$c\ell_1^{\mathrm{I}}(\mu) = -\frac{1}{2\sigma_{y_1}^2}\sum_{i=1}^{n}(y_{i1}-\mu)^2 - \frac{w_2}{2\sigma_{y_2}^2}\sum_{i=1}^{n}(y_{i2}-\mu)^2$$
$$c\ell_2^{\mathrm{I}}(\mu) = -\frac{w_1}{2\sigma_{y_1}^2}\sum_{i=1}^{n}(y_{i1}-\mu)^2 - \frac{(1-w_1)}{2\sigma_{y_2}^2}\sum_{i=1}^{n}(y_{i2}-\mu)^2,$$

respectively, with corresponding maximum weighted likelihood estimators given by

$$\tilde{\mu}_1 = \frac{\bar{y}_1\sigma_{y_2}^2 + w_2\bar{y}_2\sigma_{y_1}^2}{\sigma_{y_1}^2 + w_2\sigma_{y_1}^2},$$
$$\tilde{\mu}_2 = \frac{w_1\bar{y}_1\sigma_{y_2}^2 + (1-w_1)\bar{y}_2\sigma_{y_1}^2}{w_1\sigma_{y_2}^2 + (1-w_1)\sigma_{y_1}^2}.$$

As in the previous example, we ran a simulation experiment based on the mean squared prediction errors with $n = 10, 50, 100, 200$. For each combination, we performed 2000 iterations. Table 4.3 reports the mean squared prediction errors based on the full likelihood, weighted independence likelihood ($c\ell_1^{\mathrm{I}}(\mu)$) and unweighted independence likelihood. It seems that there is some gain in using the weighted independence likelihood in place of the unweighted independence likelihood when $\sigma_{y_1}^2 < \sigma_{y_2}^2$. While, when $\sigma_{y_1}^2 > \sigma_{y_2}^2$, the unweighted independence likelihood seems preferable. Table 4.4 reports the mean squared prediction errors based

Table 4.3: Bivariate normal model. Comparison of the mean squared prediction error based on the full likelihood ($MSPE_F$), the weighted independence likelihood ($MSPE_W$) and the unweighted independence likelihood ($MSPE_U$), when $n = 10, 50, 100, 200$. Results are obtained with 2000 simulated samples, and with the true parameter values $\theta_0 = (1, 5, 0.9)$, $\theta_0 = (5, 1, 0.9)$. The weighted independence likelihood estimator is based on $c\ell_1^{\mathrm{I}}(\theta)$.

	$\theta_0 = (1, 5, 0.9)$			$\theta_0 = (5, 1, 0.9)$		
	$MSPE_F$	$MSPE_W$	$MSPE_U$	$MSPE_F$	$MSPE_W$	$MSPE_U$
n= 10	1.0817	1.1511	1.1637	5.1834	5.4480	5.2752
n= 50	1.0711	1.0825	1.0849	5.2483	5.3326	5.2895
n=100	1.0124	1.0187	1.0197	5.0440	5.0664	5.0487
n=200	0.9756	0.9775	0.9779	4.8665	4.8759	4.8684

Table 4.4: Bivariate normal model. Comparison of the mean squared prediction error based on the full likelihood ($MSPE_F$), the weighted independence likelihood ($MSPE_W$) and the unweighted independence likelihood ($MSPE_U$), when $n = 10, 50, 100, 200$. Results are obtained with 2000 simulated samples, and with the true parameter values $\theta_0 = (1, 5, 0.9)$, $\theta_0 = (5, 1, 0.9)$. The weighted independence likelihood estimator is based on $c\ell_2^{\mathrm{I}}(\theta)$.

	$\theta_0 = (1, 5, 0.9)$			$\theta_0 = (5, 1, 0.9)$		
	$MSPE_F$	$MSPE_W$	$MSPE_U$	$MSPE_F$	$MSPE_W$	$MSPE_U$
n= 10	1.0817	1.1649	1.1637	5.1834	5.2683	5.2752
n= 50	1.0711	1.0848	1.0849	5.2483	5.2903	5.2895
n=100	1.0124	1.0188	1.0197	5.0440	5.0476	5.0487
n=200	0.9756	0.9777	0.9779	4.8665	4.8691	4.8684

on the full likelihood, weighted independence likelihood ($c\ell_2^{\mathrm{I}}(\mu)$) and unweighted independence likelihood with the same setting of Table 4.3. Overall it seems that the weighted independence likelihood is again preferable to the unweighted independence likelihood when $\sigma_{y_1}^2 < \sigma_{y_2}^2$. While, there is little indication when $\sigma_{y_1}^2 > \sigma_{y_2}^2$.

4.4 Discussion

This chapter sets out only preliminary results about the weighted independence likelihood in a prediction framework, when the full likelihood is not available. It could be a useful approach when we have no idea about the dependence structure in the data, but we know the marginal distributions.

The weighted independence likelihood with less weight on the likelihood components relative to the auxiliary variables seems to lead to some gain when the variability in the auxiliary variables is greater than that in the variable of interest.

Future work will consider different examples, different approaches for the choice of the optimal weights, and possibly new definitions of weighted independence likelihood.

Conclusions

In many multivariate problems, the standard likelihood may be unfeasible or too time consuming to compute. In these situations, composite likelihood is a very appealing alternative to the standard likelihood, being the composition of likelihoods based on lower-dimensional margins.

Chapter 3 studied the combined composite likelihood which is a new form of composite likelihood constructed as linear combination between the pairwise and the independence likelihood through a constant to be chosen. Identification of a possible strategy for finding the range of admissible values for the constant which combines the independence and pairwise likelihood was achieved by means of exact properties. The resulting combined composite likelihood estimator is still asymptotically consistent and normally distributed. After all, the inferential procedures based on the combined composite likelihood have theoretical properties similar to those based on a more conventional composite likelihood. As for any pseudo-likelihood, the methods based on the combined composite likelihood lead usually to a loss of efficiency because it is no longer valid the information identity. The combined composite likelihood could lead to better inference with respect to the pairwise and independence likelihood despite the difficulty of choosing the optimal values of the constant among the admissible ones. Both examples considered seem to suggest the pairwise conditional likelihood, which is a particular case of the combined composite likelihood, as a close to optimal choice.

Chapter 4 dealt with the weighted independence likelihood in a prediction framework. The weights were chosen by the delete-one approach in a cross validation procedure and determined by minimizing the empirical predictive discrepancy measure proposed in Wang & Zidek (2005). Although this part is still under development, preliminary simulation studies based on the two simple examples, seem to suggest that there is little gain in using the weighted independence likelihood instead of the unweighted independence likelihood. Situations in which such gain is present is characterized by a larger variability in the auxiliary variables. The future work will consider different examples, different approaches for the choice of the optimal weights, and possibly new definitions of weighted independence likelihood.

Appendix for Chapter 3

A.1 Elements of the matrix $H^a(\theta)$

The elements of the matrix $H^a(\theta)$ defined in Section 3.3.2 are

$$\begin{aligned}
H^a_{\beta_1\beta_1} &= \mathrm{E}_\theta\left\{-\frac{\partial^2 c\ell_a(\theta)}{\partial\beta_1^2}\right\} \\
&= -\left(\frac{\partial^2 h_1}{\partial\beta_1^2}\mathrm{E}_\theta(SS_1) + \frac{\partial^2 h_2}{\partial\beta_1^2}\mathrm{E}_\theta(SS_2) + \frac{\partial^2 h_3}{\partial\beta_1^2}\mathrm{E}_\theta(SS_3) + \frac{\partial^2 h_4}{\partial\beta_1^2}\mathrm{E}_\theta(SS_4) + \frac{\partial^2 h_5}{\partial\beta_1^2}\mathrm{E}_\theta(SS_5)\right. \\
&\quad \left. + \frac{\partial^2 h_6}{\partial\beta_1^2}\mathrm{E}_\theta(SS_6) + \frac{\partial^2 h_7}{\partial\beta_1^2}\right) \\
&= \frac{c_1(n,p,\theta,a)}{\sigma^2(\sigma^2+\beta_1^2\sigma_{tf}^2)^2(\sigma^2+2\beta_1^2\sigma_{tf}^2)^2}, \\
c_1(n,p,\theta,a) &= np\left[8\beta_1^8(\sigma_{tf}^2)^5 + [\{(8-8a)\beta_1^6 p + 16\beta_1^6\}\sigma^2 + 8\beta_1^8\mu_{tf}^2](\sigma_{tf}^2)^4\right. \\
&\quad + [\{(16-8a)\beta_1^4 p + 10\beta_1^4\}(\sigma^2)^2 + \{(4-4a)\beta_1^6\mu_{tf}^2 p + 20\beta_1^6\mu_{tf}^2\}\sigma^2](\sigma_{tf}^2)^3 \\
&\quad + [\{(8-2a)\beta_1^2 p + 4\beta_1^2\}(\sigma^2)^3 + \{(10-8a)\beta_1^4\mu_{tf}^2 p + 16\beta_1^4\mu_{tf}^2\}(\sigma^2)^2](\sigma_{tf}^2)^2 \\
&\quad \left. + \{2(\sigma^2)^4 + ((8-5a)\beta_1^2\mu_{tf}^2 p + 4\beta_1^2\mu_{tf}^2)(\sigma^2)^3\}(\sigma_{tf}^2) + (2-a)\mu_{tf}^2 p(\sigma^2)^4\right], \\
H^a_{\beta_1\mu_{tf}} &= \mathrm{E}_\theta\left\{-\frac{\partial^2 c\ell_a(\theta)}{\partial\beta_1\partial\mu_{tf}}\right\} \\
&= -\frac{np\beta_1\mu_{tf}[\{(2a-2)\beta_1^2 p + 2\beta_1^2\}\sigma_{tf}^2 + \{(a-2)p+2\}\sigma^2]}{(\beta_1^2\sigma_{tf}^2+\sigma^2)(2\beta_1^2\sigma_{tf}^2+\sigma^2)}, \\
H^a_{\beta_1\sigma^2} &= \mathrm{E}_\theta\left\{-\frac{\partial^2 c\ell_a(\theta)}{\partial\beta_1\partial\sigma^2}\right\} \\
&= -\frac{c_2(n,p,\theta,a)}{(\beta_1^2\sigma_{tf}^2+\sigma^2)^2(2\beta_1^2\sigma_{tf}^2+\sigma^2)^2}, \\
c_2(n,p,\theta,a) &= np\beta_1\sigma_{tf}^2\left[\{(4a-2)\beta_1^4 p + 2\beta_1^4\}(\sigma_{tf}^2)^2 + \{(4a-4)\beta_1^2 p + 4\beta_1^2\}\sigma^2\sigma_{tf}^2\right. \\
&\quad \left. + \{(a-2)p+2\}(\sigma^2)^2\right],
\end{aligned}$$

$$
\begin{aligned}
H^a_{\beta_1\sigma^2_{tf}} &= \mathrm{E}_\theta\left\{-\frac{\partial^2 c\ell_a(\theta)}{\partial\beta_1\partial\sigma^2_{tf}}\right\}\\
&= -\frac{c_3(n,p,\theta,a)}{(\beta_1^2\sigma^2_{tf}+\sigma^2)^2(2\beta_1^2\sigma^2_{tf}+\sigma^2)^2},\\
c_3(n,p,\theta,a) &= np\beta_1^3\sigma^2_{tf}\left[\{(4a-4)\beta_1^4p+4\beta_1^4\}(\sigma^2_{tf})^2+\{(4a-8)\beta_1^2p+8\beta_1^2\}\sigma^2\sigma^2_{tf}\right.\\
&\quad\left.+\{(a-4)p+4\}(\sigma^2)^2\right],\\
H^a_{\mu_{tf}\mu_{tf}} &= \mathrm{E}_\theta\left\{-\frac{\partial^2 c\ell_a(\theta)}{\partial\mu^2_{tf}}\right\}\\
&= -\frac{np\{(2a-2)\beta_1^2\sigma^2_{tf}+(a-2)\sigma^2\}\{(\beta_1^2p+\beta_1^2)\sigma^2_{tf}+\sigma^2\}}{\sigma^2_{tf}(\beta_1^2\sigma^2_{tf}+\sigma^2)(2\beta_1^2\sigma^2_{tf}+\sigma^2)}.\\
H^a_{\mu_{tf}\sigma^2} &= \mathrm{E}_\theta\left\{-\frac{\partial^2 c\ell_a(\theta)}{\partial\mu_{tf}\partial\sigma^2}\right\}=0.\\
H^a_{\mu_{tf}\sigma^2_{tf}} &= \mathrm{E}_\theta\left\{-\frac{\partial^2 c\ell_a(\theta)}{\partial\mu_{tf}\partial\sigma^2_{tf}}\right\}=0.\\
H^a_{\sigma^2\sigma^2} &= \mathrm{E}_\theta\left\{-\frac{\partial^2 c\ell_a(\theta)}{\partial(\sigma^2)^2}\right\}\\
&= \frac{c_4(n,p,\theta,a)}{2(\sigma^2)^2(\beta_1^2\sigma^2_{tf}+\sigma^2)^2(2\beta_1^2\sigma^2_{tf}+\sigma^2)^2},\\
c_4(n,p,\theta,a) &= np\left[(4\beta_1^8p+4\beta_1^8)(\sigma^2_{tf})^4+(12\beta_1^6p+12\beta_1^6)\sigma^2(\sigma^2_{tf})^3\right.\\
&\quad+\{(14-4a)\beta_1^4p+12\beta_1^4\}(\sigma^2)^2(\sigma^2_{tf})^2+\{(8-4a)\beta_1^2p+4\beta_1^2\}(\sigma^2)^3\sigma^2_{tf}\\
&\quad\left.+(2-a)p(\sigma^2)^4\right],\\
H^a_{\sigma^2\sigma^2_{tf}} &= \mathrm{E}_\theta\left\{-\frac{\partial^2 c\ell_a(\theta)}{\partial\sigma^2\partial\sigma^2_{tf}}\right\}\\
&= -\frac{c_5(n,p,\theta,a)}{2(\beta_1^2\sigma^2_{tf}+\sigma^2)^2(2\beta_1^2\sigma^2_{tf}+\sigma^2)^2},\\
c_5(n,p,\theta,a) &= np\beta_1^2\left[\{(4a-2)\beta_1^4p+2\beta_1^4\}(\sigma^2_{tf})^2+\{(4a-4)\beta_1^2p+4\beta_1^2\}\sigma^2\sigma^2_{tf}\right.\\
&\quad\left.+\{(a-2)p+2\}(\sigma^2)^2\right],\\
H^a_{\sigma^2_{tf}\sigma^2_{tf}} &= \mathrm{E}_\theta\left\{-\frac{\partial^2 c\ell_a(\theta)}{\partial(\sigma^2_{tf})^2}\right\}\\
&= -\frac{c_6(n,p,\theta,a)}{2(\sigma^2_{tf})^2(\beta_1^2\sigma^2_{tf}+\sigma^2)^2(2\beta_1^2\sigma^2_{tf}+\sigma^2)^2},\\
c_6(n,p,\theta,a) &= np\left[\{(4a-4)\beta_1^8p+(4a-4)\beta_1^8\}(\sigma^2_{tf})^4+\{(4a-8)\beta_1^6p\right.\\
&\quad+(12a-16)\beta_1^6\}(\sigma^2)(\sigma^2_{tf})^3+\{(a-4)\beta_1^4p+(13a-22)\beta_1^4\}(\sigma^2)^2(\sigma^2_{tf})^2\\
&\quad\left.+(6a-12)\beta_1^2(\sigma^2)^3(\sigma^2_{tf})+(a-2)(\sigma^2)^4\right].
\end{aligned}
$$

A.2 Elements of the matrix $J^a(\theta)$

The elements of matrix $J^a(\theta)$ defined in Section 3.3.2 are

$$\begin{aligned}
J^a_{\beta_1\beta_1} &= \left(\frac{\partial h_1}{\partial \beta_1}\right)^2 \mathrm{Var}_\theta(SS_1) + \left(\frac{\partial h_2}{\partial \beta_1}\right)^2 \mathrm{Var}_\theta(SS_2) + \left(\frac{\partial h_3}{\partial \beta_1}\right)^2 \mathrm{Var}_\theta(SS_3) \\
&+ \left(\frac{\partial h_4}{\partial \beta_1}\right)^2 \mathrm{Var}_\theta(SS_4) + \left(\frac{\partial h_5}{\partial \beta_1}\right)^2 \mathrm{Var}_\theta(SS_5) + \left(\frac{\partial h_6}{\partial \beta_1}\right)^2 \mathrm{Var}_\theta(SS_6) \\
&+ 2\left(\frac{\partial h_1}{\partial \beta_1}\frac{\partial h_2}{\partial \beta_1}\right)\mathrm{Cov}_\theta(SS_1, SS_2) + 2\left(\frac{\partial h_1}{\partial \beta_1}\frac{\partial h_3}{\partial \beta_1}\right)\mathrm{Cov}_\theta(SS_1, SS_3) \\
&+ 2\left(\frac{\partial h_1}{\partial \beta_1}\frac{\partial h_4}{\partial \beta_1}\right)\mathrm{Cov}_\theta(SS_1, SS_4) + 2\left(\frac{\partial h_1}{\partial \beta_1}\frac{\partial h_5}{\partial \beta_1}\right)\mathrm{Cov}_\theta(SS_1, SS_5) \\
&+ 2\left(\frac{\partial h_1}{\partial \beta_1}\frac{\partial h_6}{\partial \beta_1}\right)\mathrm{Cov}_\theta(SS_1, SS_6) + 2\left(\frac{\partial h_2}{\partial \beta_1}\frac{\partial h_3}{\partial \beta_1}\right)\mathrm{Cov}_\theta(SS_2, SS_3) \\
&+ 2\left(\frac{\partial h_2}{\partial \beta_1}\frac{\partial h_4}{\partial \beta_1}\right)\mathrm{Cov}_\theta(SS_2, SS_4) + 2\left(\frac{\partial h_2}{\partial \beta_1}\frac{\partial h_5}{\partial \beta_1}\right)\mathrm{Cov}_\theta(SS_2, SS_5) \\
&+ 2\left(\frac{\partial h_2}{\partial \beta_1}\frac{\partial h_6}{\partial \beta_1}\right)\mathrm{Cov}_\theta(SS_2, SS_6) + 2\left(\frac{\partial h_3}{\partial \beta_1}\frac{\partial h_4}{\partial \beta_1}\right)\mathrm{Cov}_\theta(SS_3, SS_4) \\
&+ 2\left(\frac{\partial h_3}{\partial \beta_1}\frac{\partial h_5}{\partial \beta_1}\right)\mathrm{Cov}_\theta(SS_3, SS_5) + 2\left(\frac{\partial h_3}{\partial \beta_1}\frac{\partial h_6}{\partial \beta_1}\right)\mathrm{Cov}_\theta(SS_3, SS_6) \\
&+ 2\left(\frac{\partial h_4}{\partial \beta_1}\frac{\partial h_5}{\partial \beta_1}\right)\mathrm{Cov}_\theta(SS_4, SS_5) + 2\left(\frac{\partial h_4}{\partial \beta_1}\frac{\partial h_6}{\partial \beta_1}\right)\mathrm{Cov}_\theta(SS_4, SS_6) \\
&+ 2\left(\frac{\partial h_5}{\partial \beta_1}\frac{\partial h_6}{\partial \beta_1}\right)\mathrm{Cov}_\theta(SS_5, SS_6) \\
&= \frac{c_7(n,p,\theta,a)}{\sigma^2(\beta_1^2\sigma_{tf}^2+\sigma^2)^4(2\beta_1^2\sigma_{tf}^2+\sigma^2)^4},
\end{aligned}$$

$$\begin{aligned}
c_7(n,p,\theta,a) &= np\,\big[64\beta_1^{16}(\sigma_{tf}^2)^9 + (((32a^2-64a+32)\beta_1^{14}p^3 + (64a-64)\beta_1^{14}p^2 \\
&+ (160-128a)\beta_1^{14}p + 256\beta_1^{14})\sigma^2 + 64\beta_1^{16}\mu_{tf}^2)(\sigma_{tf}^2)^8 \\
&+ (((64a^2-192a+128)\beta_1^{12}p^3 + (64a^2+64a-192)\beta_1^{12}p^2 \\
&+ (640-384a)\beta_1^{12}p + 416\beta_1^{12})(\sigma^2)^2 + ((16a^2-32a+16)\beta_1^{14}\mu_{tf}^2p^3 \\
&+ (32a-32)\beta_1^{14}\mu_{tf}^2p^2 + (80-64a)\beta_1^{14}\mu_{tf}^2p + 320\beta_1^{14}\mu_{tf}^2)\sigma^2)(\sigma_{tf}^2)^7 \\
&+ (((48a^2-208a+192)\beta_1^{10}p^3 + (160a^2-208a-120)\beta_1^{10}p^2 \\
&+ (984-416a)\beta_1^{10}p + 384\beta_1^{10})(\sigma^2)^3 + ((64a^2-144a+80)\beta_1^{12}\mu_{tf}^2p^3 \\
&+ (16a^2+112a-144)\beta_1^{12}\mu_{tf}^2p^2 + (400-288a)\beta_1^{12}\mu_{tf}^2p \\
&+ 656\beta_1^{12}\mu_{tf}^2)(\sigma^2)^2)(\sigma_{tf}^2)^6 + (((16a^2-96a+128)\beta_1^8p^3 \\
&+ (160a^2-416a+160)\beta_1^8p^2 + (736-192a)\beta_1^8p + 260\beta_1^8)(\sigma^2)^4 \\
&+ ((104a^2-264a+164)\beta_1^{10}\mu_{tf}^2p^3 + (64a^2+120a-248)\beta_1^{10}\mu_{tf}^2p^2 \\
&+ (820-528a)\beta_1^{10}\mu_{tf}^2p + 704\beta_1^{10}\mu_{tf}^2)(\sigma^2)^3)(\sigma_{tf}^2)^5 \\
&+ (((2a^2-16a+32)\beta_1^6p^3 + (80a^2-280a+240)\beta_1^6p^2 + (304-32a)\beta_1^6p
\end{aligned}$$

$$
\begin{aligned}
&+ 144\beta_1^6)(\sigma^2)^5 + ((88a^2 - 252a + 176)\beta_1^8\mu_{tf}^2 p^3 \\
&+ (104a^2 - 12a - 188)\beta_1^8\mu_{tf}^2 p^2 + (880 - 504a)\beta_1^8\mu_{tf}^2 p \\
&+ 416\beta_1^8\mu_{tf}^2)(\sigma^2)^4)(\sigma_{tf}^2)^4 + (((20a^2 - 80a + 96)\beta_1^4 p^2 + 96\beta_1^4 p + 56\beta_1^4)(\sigma^2)^6 \\
&+ ((41a^2 - 132a + 104)\beta_1^6\mu_{tf}^2 p^3 + (88a^2 - 120a - 32)\beta_1^6\mu_{tf}^2 p^2 \\
&+ (520 - 264a)\beta_1^6\mu_{tf}^2 p + 128\beta_1^6\mu_{tf}^2)(\sigma^2)^5)(\sigma_{tf}^2)^3 \\
&+ (((2a^2 - 8a + 8)\beta_1^2 p^2 + 24\beta_1^2 p + 16\beta_1^2)(\sigma^2)^7 \\
&+ ((10a^2 - 36a + 32)\beta_1^4\mu_{tf}^2 p^3 + (41a^2 - 96a + 40)\beta_1^4\mu_{tf}^2 p^2 \\
&+ (160 - 72a)\beta_1^4\mu_{tf}^2 p + 16\beta_1^4\mu_{tf}^2)(\sigma^2)^6)(\sigma_{tf}^2)^2 \\
&+ (4(\sigma^2)^8 + ((a^2 - 4a + 4)\beta_1^2\mu_{tf}^2 p^3 + (10a^2 - 32a + 24)\beta_1^2\mu_{tf}^2 p^2 \\
&+ (20 - 8a)\beta_1^2\mu_{tf}^2 p)(\sigma^2)^7)(\sigma_{tf}^2) + (a^2 - 4a + 4)\mu_{tf}^2 p^2(\sigma^2)^8\big],
\end{aligned}
$$

$$
J^a_{\mu_{tf}\mu_{tf}} = \frac{c_8(n,p,\theta,a)}{\sigma_{tf}^2(\beta_1^2\sigma_{tf}^2 + \sigma^2)^2(2\beta_1^2\sigma_{tf}^2 + \sigma^2)^2},
$$

$$
\begin{aligned}
c_8(n,p,\theta,a) = np\,\big[&((4a^2 - 8a + 4)\beta_1^8 p^3 + (8a^2 - 16a + 8)\beta_1^8 p^2 \\
&+ (4a^2 - 8a + 4)\beta_1^8 p)(\sigma_{tf}^2)^4 + ((4a^2 - 12a + 8)\beta_1^6 p^3 \\
&+ (20a^2 - 48a + 28)\beta_1^6 p^2 + (12a^2 - 20a + 8)\beta_1^6 p \\
&+ 4\beta_1^6)\sigma^2(\sigma_{tf}^2)^3 + ((a^2 - 4a + 4)\beta_1^4 p^3 + (14a^2 - 44a + 32)\beta_1^4 p^2 \\
&+ (13a^2 - 24a + 8)\beta_1^4 p + 8\beta_1^4)(\sigma^2)^2(\sigma_{tf}^2)^2 + ((3a^2 - 12a + 12)\beta_1^2 p^2 \\
&+ (6a^2 - 16a + 8)\beta_1^2 p + 4\beta_1^2)(\sigma^2)^3(\sigma_{tf}^2) + (a^2 - 4a + 4)p(\sigma^2)^4\big],
\end{aligned}
$$

$$
J^a_{\sigma^2\sigma^2} = \frac{c_9(n,p,\theta,a)}{2(\sigma^2)^2(\beta_1^2\sigma_{tf}^2 + \sigma^2)^4(2\beta_1^2\sigma_{tf}^2 + \sigma^2)^4},
$$

$$
\begin{aligned}
c_9(n,p,\theta,a) = np\,\big[&(16\beta_1^{16}p^2 + 48\beta_1^{16}p)(\sigma_{tf}^2)^8 + (96\beta_1^{14}p^2 + 288\beta_1^{14}p)\sigma^2(\sigma_{tf}^2)^7 \\
&+ ((16a^2 - 16a + 4)\beta_1^{12}p^3 + (248 - 16a)\beta_1^{12}p^2 \\
&+ (740 - 32a)\beta_1^{12}p)(\sigma^2)^2(\sigma_{tf}^2)^6 + ((32a^2 - 48a + 16)\beta_1^{10}p^3 \\
&+ (32a^2 - 112a + 376)\beta_1^{10}p^2 + (1040 - 96a)\beta_1^{10}p + 8\beta_1^{10})(\sigma^2)^3(\sigma_{tf}^2)^5 \\
&+ ((24a^2 - 52a + 24)\beta_1^8 p^3 + (80a^2 - 260a + 388)\beta_1^8 p^2 \\
&+ (840 - 104a)\beta_1^8 p + 32\beta_1^8)(\sigma^2)^4(\sigma_{tf}^2)^4 + ((8a^2 - 24a + 16)\beta_1^6 p^3 \\
&+ (80a^2 - 280a + 288)\beta_1^6 p^2 + (368 - 48a)\beta_1^6 p + 48\beta_1^6)(\sigma^2)^5(\sigma_{tf}^2)^3 \\
&+ ((a^2 - 4a + 4)\beta_1^4 p^3 + (40a^2 - 152a + 144)\beta_1^4 p^2 \\
&+ (68 - 8a)\beta_1^4 p + 32\beta_1^4)(\sigma^2)^6(\sigma_{tf}^2)^2 + ((10a^2 - 40a + 40)\beta_1^2 p^2 \\
&+ 8\beta_1^2)(\sigma^2)^7\sigma_{tf}^2 + (a^2 - 4a + 4)p^2(\sigma^2)^8\big],
\end{aligned}
$$

$$
J^a_{\sigma_{tf}^2\sigma_{tf}^2} = \frac{c_{10}(n,p,\theta,a)}{2(\sigma_{tf}^2)^2(\beta_1^2\sigma_{tf}^2 + \sigma^2)^4(2\beta_1^2\sigma_{tf}^2 + \sigma^2)^4},
$$

$$
\begin{aligned}
c_{10}(n,p,\theta,a) &= np\left[((16a^2-32a+16)\beta_1^{16}p^3+(32a^2-64a+32)\beta_1^{16}p^2\right.\\
&+(16a^2-32a+16)\beta_1^{16}p)(\sigma_{tf}^2)^8+((32a^2-96a+64)\beta_1^{14}p^3\\
&+(160a^2-384a+224)\beta_1^{14}p^2+(96a^2-160a+64)\beta_1^{14}p\\
&+32\beta_1^{14})\sigma^2(\sigma_{tf}^2)^7+((24a^2-104a+96)\beta_1^{12}p^3\\
&+(288a^2-848a+596)\beta_1^{12}p^2+(248a^2-456a+172)\beta_1^{12}p\\
&+128\beta_1^{12})(\sigma^2)^2(\sigma_{tf}^2)^6+((8a^2-48a+64)\beta_1^{10}p^3\\
&+(256a^2-912a+784)\beta_1^{10}p^2+(360a^2-832a+400)\beta_1^{10}p\\
&+192\beta_1^{10})(\sigma^2)^3(\sigma_{tf}^2)^5+((a^2-8a+16)\beta_1^{8}p^3\\
&+(122a^2-512a+536)\beta_1^{8}p^2+(321a^2-928a+604)\beta_1^{8}p\\
&+128\beta_1^{8})(\sigma^2)^4(\sigma_{tf}^2)^4+((30a^2-144a+176)\beta_1^{6}p^2\\
&+(180a^2-616a+512)\beta_1^{6}p+32\beta_1^{6})(\sigma^2)^5(\sigma_{tf}^2)^3\\
&+((3a^2-16a+20)\beta_1^{4}p^2+(62a^2-236a+228)\beta_1^{4}p)(\sigma^2)^6(\sigma_{tf}^2)^2\\
&\left.+(12a^2-48a+48)\beta_1^{2}p(\sigma^2)^7\sigma_{tf}^2+(a^2-4a+4)p(\sigma^2)^8\right],\\
J^a_{\beta_1\mu_{tf}} &= \frac{c_{11}(n,p,\theta,a)}{(\beta_1^2\sigma_{tf}^2+\sigma^2)^2(2\beta_1^2\sigma_{tf}^2+\sigma^2)^2},\\
c_{11}(n,p,\theta,a) &= \left[\beta_1\mu_{tf}np(2a\beta_1^2p\sigma_{tf}^2-2\beta_1^2p\sigma_{tf}^2+2\beta_1^2\sigma_{tf}^2+ap\sigma^2-2p\sigma^2+2\sigma^2)(2a\beta_1^4p^2(\sigma_{tf}^2)^2\right.\\
&-2\beta_1^4p^2(\sigma_{tf}^2)^2+2a\beta_1^4p(\sigma_{tf}^2)^2-2\beta_1^4p(\sigma_{tf}^2)^2-4\beta_1^4(\sigma_{tf}^2)^2+a\beta_1^2p^2\sigma^2\sigma_{tf}^2\\
&\left.-2\beta_1^2p^2\sigma^2\sigma_{tf}^2+5a\beta_1^2p\sigma^2\sigma_{tf}^2-6\beta_1^2p\sigma^2\sigma_{tf}^2-4\beta_1^2\sigma^2\sigma_{tf}^2+2ap(\sigma^2)^2-4p(\sigma^2)^2)\right],\\
J^a_{\beta_1\sigma^2} &= \frac{c_{12}(n,p,\theta,a)}{(\beta_1^2\sigma_{tf}^2+\sigma^2)^4(2\beta_1^2\sigma_{tf}^2+\sigma^2)^4},\\
c_{12}(n,p,\theta,a) &= np\beta_1\sigma_{tf}^2\left[((16a^2-24a+8)\beta_1^{12}p^3+(8a-8)\beta_1^{12}p^2+(16-48a)\beta_1^{12}p\right.\\
&-16\beta_1^{12})(\sigma_{tf}^2)^6+((32a^2-72a+32)\beta_1^{10}p^3+(32a^2-40a-8)\beta_1^{10}p^2\\
&+(40-144a)\beta_1^{10}p-64\beta_1^{10})\sigma^2(\sigma_{tf}^2)^5+((24a^2-78a+48)\beta_1^{8}p^3\\
&+(80a^2-182a+52)\beta_1^{8}p^2-156a\beta_1^{8}p-100\beta_1^{8})(\sigma^2)^2(\sigma_{tf}^2)^4\\
&+((8a^2-36a+32)\beta_1^{6}p^3+(80a^2-244a+128)\beta_1^{6}p^2\\
&+(-72a-80)\beta_1^{6}p-80\beta_1^{6})(\sigma^2)^3(\sigma_{tf}^2)^3+((a^2-6a+8)\beta_1^{4}p^3\\
&+(40a^2-146a+112)\beta_1^{4}p^2+(-12a-80)\beta_1^{4}p-40\beta_1^{4})(\sigma^2)^4(\sigma_{tf}^2)^2\\
&+((10a^2-40a+40)\beta_1^{2}p^2-24\beta_1^{2}p-16\beta_1^{2})(\sigma^2)^5(\sigma_{tf}^2)\\
&\left.+((a^2-4a+4)p^2-4)(\sigma^2)^6\right],\\
J^a_{\beta_1\sigma_{tf}^2} &= \frac{c_{13}(n,p,\theta,a)}{(\beta_1^2\sigma_{tf}^2+\sigma^2)^4(2\beta_1^2\sigma_{tf}^2+\sigma^2)^4},
\end{aligned}
$$

$$
\begin{aligned}
c_{13}(n,p,\theta,a) &= np\beta_1^3\sigma_{tf}^2\left[((16a^2-32a+16)\beta_1^{12}p^3+(16a^2-16a)\beta_1^{12}p^2+(16-16a)\beta_1^{12}p\right.\\
&\quad -32\beta_1^{12})(\sigma_{tf}^2)^6+((32a^2-96a+64)\beta_1^{10}p^3+(96a^2-176a+64)\beta_1^{10}p^2\\
&\quad +16a\beta_1^{10}p-128\beta_1^{10})(\sigma^2)(\sigma_{tf}^2)^5+((24a^2-104a+96)\beta_1^8p^3\\
&\quad +(184a^2-476a+268)\beta_1^8p^2+(164a-164)\beta_1^8p-200\beta_1^8)(\sigma^2)^2(\sigma_{tf}^2)^4\\
&\quad +((8a^2-48a+64)\beta_1^6p^3+(168a^2-560a+432)\beta_1^6p^2+(256a-336)\beta_1^6p\\
&\quad -160\beta_1^6)(\sigma^2)^3(\sigma_{tf}^2)^3+((a^2-8a+16)\beta_1^4p^3+(81a^2-326a+328)\beta_1^4p^2\\
&\quad +(170a-264)\beta_1^4p-80\beta_1^4)(\sigma^2)^4(\sigma_{tf}^2)^2+((20a^2-92a+112)\beta_1^2p^2\\
&\quad +(52a-80)\beta_1^2p-32\beta_1^2)(\sigma^2)^5(\sigma_{tf}^2)+((2a^2-10a+12)p^2\\
&\quad \left.+(6a-4)p-8)(\sigma^2)^6\right],\\
J^a_{\mu_{tf}\sigma^2} &= 0,\\
J^a_{\mu_{tf}\sigma^2_{tf}} &= 0,\\
J^a_{\sigma^2\sigma^2_{tf}} &= \frac{c_{14}(n,p,\theta,a)}{2(\beta_1^2\sigma_{tf}^2+\sigma^2)^4(2\beta_1^2\sigma_{tf}^2+\sigma^2)^4},\\
c_{14}(n,p,\theta,a) &= np\beta_1^2\left[((16a^2-24a+8)\beta_1^{12}p^3+(16a^2-32a+8)\beta_1^{12}p^2\right.\\
&\quad +(-8a-16)\beta_1^{12}p)(\sigma_{tf}^2)^6+((32a^2-72a+32)\beta_1^{10}p^3\\
&\quad +(96a^2-208a+72)\beta_1^{10}p^2+(24a-120)\beta_1^{10}p+16\beta_1^{10})(\sigma^2)(\sigma_{tf}^2)^5\\
&\quad +((24a^2-78a+48)\beta_1^8p^3+(184a^2-472a+216)\beta_1^8p^2+(134a-328)\beta_1^8p\\
&\quad +64\beta_1^8)(\sigma^2)^2(\sigma_{tf}^2)^4+((8a^2-36a+32)\beta_1^6p^3+(168a^2-508a+304)\beta_1^6p^2\\
&\quad +(192a-432)\beta_1^6p+96\beta_1^6)(\sigma^2)^3(\sigma_{tf}^2)^3+((a^2-6a+8)\beta_1^4p^3\\
&\quad +(81a^2-280a+216)\beta_1^4p^2+(122a-288)\beta_1^4p+64\beta_1^4)(\sigma^2)^4(\sigma_{tf}^2)^2\\
&\quad +((20a^2-76a+72)\beta_1^2p^2+(36a-88)\beta_1^2p+16\beta_1^2)(\sigma^2)^5(\sigma_{tf}^2)\\
&\quad \left.+((2a^2-8a+8)p^2+(4a-8)p)(\sigma^2)^6\right].
\end{aligned}
$$

Bibliography

BESAG, J. (1974). Spatial interaction and the analysis of lattice systems. *Journal of the Royal Statistical Society* 34 192–236.

CHANDLER, R. & BATE, S. (2007). Inference for clustered data using the independence loglikelihood. *Biometrika* 94 167–183.

COX, D. R. & REID, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika* 91 729–737.

DAVISON, A. C. (2003). *Statistical Models*. Cambridge: Cambridge University Press.

EFRON, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency). *The Annals of Statistics* 3 1189–1242.

FISHER, R. A. (1921). On the probable error of a coefficient of correlation deduced from a small sample. *Metron* 1 3–32.

FISHER, R. A. (1922). On the mathematical foundations of theoretical statistics. *The Royal Society* 222 309–368.

GODAMBE, V. P. (1960). An optimum property of regular maximum likelihood equation. *Annals of Mathematical Statistics* 31 1208–1211.

HJORT, N. L. & VARIN, C. (2008). Ml, cl and pl in markov chain models. *Scand. J. Stat.* 35 64–82.

LE CESSIE, S. & VAN HOUWELINGEN, J. (1994). Logistic regression for correlated binary data. *Applied Statistics* 43 95–108.

LINDSAY, B. (1988). Composite likelihood methods. *Comtemporary Mathematics* 80 221–240.

LINDSAY, B. G., YI, G. Y. & SUN, J. (2011). Issues and strategies in the selection of composite likelihoods. *Statistica Sinica* 21 71–105.

MARDIA, K. V., HUGHES, G., TAYLOR, C. C. & SINGH, H. (2008). A multivariate von mises distribution with applications to bioinformatics. *Canadian Journal of Statistics* 36 99–109.

MARDIA, K. V., KENT, J. T., HUGHES, G. & TAYLOR, C. C. (2009). Maximum likelihood estimation using composite likelihoods for closed exponential families. *Biometrika* 96 975–982.

MOLENBERGHS, G. & VERBEKE, G. (2005). *Models for discrete longitudinal data*. Springer Verlag.

PACE, L. & SALVAN, A. (1997). *Principles of statistical inference: from a Neo-Fisherian perspective*, vol. 4. Singapore: World Scientific Pub Co Inc.

PACE, L., SALVAN, A. & SARTORI, N. (2011). Adjusting composite likelihood ratio statistics. *Statistica Sinica* 21 129–148.

R CORE TEAM (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL `http://www.R-project.org/`.

RAO, C. R. (1973). *Linear Statistical Inference and its Applications*. New York: Wiley.

ROVERATO, A. & DI LASCIO, F. M. (2011). Wilks' λ dissimilarity measures for gene clustering: An approach based on the identification of transcription modules. *Biometrics* 67 1236–1248.

SEVERINI, T. A. (2000). *Likelihood Methods in Statistics*. Oxford: Oxford University Press.

SEVERINI, T. A. (2005). *Elements of Distribution Theory*. Cambridge: Cambridge University Press.

VARIN, C. (2008). On composite marginal likelihoods. *Advances in Statistical Analysis* 92 1–28.

VARIN, C., REID, N. & FIRTH, D. (2011). An overview of composite likelihood methods. *Statistica Sinica* 21 5–42.

WANG, X. & ZIDEK, J. V. (2005). Selecting likelihood weights by cross-validation. *The Annals of Statistics* 33 463–500.

XU, X. & REID, N. (2011). On the robustness of maximum composite likelihood estimate. *Journal of Statistical Planning and Inference* 141 3047–3054.

Euloge Clovis Kenne Pagui

CURRICULUM VITAE

Contact Information

University of Padova
Department of Statistics
via Cesare Battisti, 241-243
35121 Padova. Italy.

Tel. +39 049 827 4174
e-mail: kenne@stat.unipd.it

Current Position

---

expected completion: January 2013
PhD Student in Statistical Sciences, University of Padova.
Thesis title: Combined composite likelihoods
Supervisor: Prof. Nicola Sartori
Co-supervisor: Prof. Alessandra Salvan.

Research interests

- Likelihood methods
- Pseudo-likelihood methods
- Composite likelihood methods
- Asymptotic properties of pseudo-likelihood quantities

Education

October 2007 – Octover 2009

MSc in Statistics and Computer Science.
University of Padova, Faculty of Statistics
Title of dissertation: “Pairwise likelihood in multivariate normal models ”
Supervisor: Prof. Alessandra Salvan
Final mark: 109/110

October 2004 – July 2007

BSc in Statistics and computer’s tecnologies.
University of Padova, Faculty of Statistics
Title of dissertation: “Comparison between semi-parametric statistical models: An application to foot- ball championships”
Supervisor: Prof. Stuart Coles
Final mark: 107/110.

Visiting periods

April 2012 – July 2012

University of Toronto,
Toronto, Canada .
Supervisor: Prof. Nancy Reid

Computer skills

- R, LATEX, MS OFFICE, MS WINDOWS (**expert level**)
- SPSS, C, LINUS (**intermediate level**)
- HTML, SQL, PHP, C++, SAS (**basic level**)

Language skills

French: native language; Italian: fluent; English: good;

Conference presentations

Kenne Pagui Euloge Clovis. Combined composite likelihoods. (invited) *Approximate likelihood methods for high-dimensional dependencies* , Venice, Italy, October 8, 2012.

Kenne Pagui Euloge Clovis. Composite likelihood in multivariate normal models. (invited) *Approximate likelihood methods for high-dimensional dependencies* , Verona, Italy, May 2011.

References

Prof. Nicola Sartori
Department of Statistical Sciences, University of Padova, Italy
Via Cesare Battisti, 241/243
Phone: (+39) 049 8274127
e-mail: sartori@stat.unipd.it

Prof. Alessandra Salvan
Department of Statistical Sciences, University of Padova, Italy
Via Cesare Battisti, 241/243
Phone: +39 49 827 4166
e-mail: salvan@stat.unipd.it

Prof. Nancy Reid
Department of statistics, University of Toronto, Canada
100 st George st, Toronto, Ontario M5S 3G3, CANADA
Phone: (416) 978-5046
e-mail: reid@utstat.utoronto.ca