Research Article

Marianne Jonker* and Aad Van der Vaart

On the Correction of the Asymptotic Distribution of the Likelihood Ratio Statistic If Nuisance Parameters Are Estimated Based on an External Source

Abstract: In practice, nuisance parameters in statistical models are often replaced by estimates based on an external source, for instance if estimates were published before or a second dataset is available. Next these estimates are assumed to be known when the parameter of interest is estimated, a hypothesis is tested or confidence intervals are constructed. By this assumption, the level of the test is, in general, higher than supposed and the coverage of the confidence interval is too low. In this article, we derive the asymptotic distribution of the likelihood ratio statistic if the nuisance parameters are estimated based on a dataset that is independent of the data used for estimating the parameter of interest. This distribution can be used for correctly testing hypotheses and constructing confidence intervals. Four theoretical and practical examples are given as illustration.

Keywords: likelihood ratio test, nuisance parameters, testing, test level, chi-squared distribution

DOI 10.1515/ijb-2013-0063

1 Introduction

Statistical models often contain multiple unknown parameters, some of interest, and some not, referred to as nuisance parameters. The presence of the latter parameters may make estimation, hypothesis testing and the construction of confidence intervals for the parameters of interest more complex, because more parameters must be estimated, possibly of infinite dimension.

Estimates of the nuisance parameters may be available in the literature or can sometimes be obtained from an independent dataset. Inserting these estimates into the model and next performing inference on the parameters of interest may simplify the statistical inference enormously. However, treating the nuisance parameters as known may affect the level of a test or the coverage of a confidence interval, making these higher and lower than nominal, respectively. In this paper, we investigate these effects for likelihood ratio procedures.

For standard statistical models, the likelihood ratio statistic for a finite-dimensional parameter is well known to be asymptotically distributed under the null hypothesis as a chi-squared random variable with degrees of freedom equal to the number of free parameters [1–3]. This result extends to likelihood ratio statistics for the parameter of interest in semiparametric models, as shown in various settings in Thomas and Grunkemeier [4] (right censoring), Murphy and van der Vaart [5] (semiparametric models), Banerjee and Wellner [6] (current status data), Banerjee [7] (monotone missingness), Kosorok [8], with empirical ratio likelihood statistics for functionals [9–11] as an important special case. The likelihood ratio statistic in the latter semiparametric models can be viewed as the likelihood ratio statistic for the parameter of interest in

*Corresponding author: Marianne Jonker, Department of Epidemiology and Biostatistics, VUmc, Postbus 7057 1007 MB, Amsterdam, The Netherlands, E-mail: m.jonker2@vumc.nl

Aad Van der Vaart, Mathematical Institute, Leiden University, Leiden, The Netherlands, E-mail: avdvaart@math.leidenuniv.nl

the model with the nuisance parameter replaced by its maximum likelihood estimator, based on the same data; equivalently as the likelihood ratio statistic based on the profile likelihood Murphy and van der Vaart [12]. With this estimate of the nuisance parameter, the asymptotic chi-squared distribution is retained, much as when testing only a part of the parameter in a finite-dimensional model. In contrast, it is shown below that the distribution changes if the nuisance is estimated using an independent sample of observations.

Naturally, this change will be small or even negligible if the additional sample is much bigger than the sample used to construct the likelihood ratio statistics. An example is estimating the disease-allele frequency in an inbred population, described in Example 4 below (see also Jonker et al. [13] and Teeuw et al. [14]). In this example, only 20 individuals were available for estimating the nuisance parameter, and 80 individuals to estimate the parameter of interest. It turns out that for two different approaches using a likelihood ratio, in the one case the distribution is well approximated by a chi-squared distribution as if the nuisance were known, whereas in the other case a correction is needed. One motivation of the present paper is to gain insight into the fact that not only the relative sample sizes determine the amount of correction needed but also the structure of the model. This is explained below in terms of a relevant Fisher information quantity.

The original motivation for the present paper was a study on the genetic determinants of the onset of migraine, described in Jonker et al. [15]. In this example, a baseline survival distribution was estimated using an additional, much larger dataset. As these data were subject to current status censoring, this estimate was relatively imprecise, of order comparable to the estimation error of the direct data, notwithstanding the size of the dataset. However, in a simulation study, it was found that substitution of the estimator in the likelihood ratio test caused only a minor error in the level of the test. This finding is explained by the detailed derivation of the likelihood ratio statistic in this paper, which reveals that the estimate for the baseline survival function enters (asymptotically) into the likelihood ratio through a smoothing operation, which repairs its inefficiency as an estimate for the survival function itself. See Example 3 below for details.

To our knowledge, the asymptotic distribution of the likelihood ratio statistic in these and other examples was not studied before. On the other hand, several authors have studied plug-in estimators within the context of empirical likelihood for parameters defined by estimating equations [9], among others Qin and Jing [16, 17], Li and Wang [18] and Wang and Jing [19]. In particular, Hjort et al. [20] study a similar problem in the setting of empirical likelihood. Besides the limit distribution, these authors also discuss a bootstrap approximation and allow an increasing number of estimating equations.

In the next section, we describe the theorem and give the proof. Thereafter we consider four examples; two theoretical examples, for a parametric and a semiparametric model, and two practical problems, for a large and a small dataset for estimating the nuisance parameter.

2 The likelihood ratio test

Suppose we observe a random sample $X_1, \dots, X_n$ from a distribution $P_{\theta, \eta}$ with corresponding density $p_{\theta, \eta}$ indexed by two parameters: a vector $\theta \in \Theta \subset \mathbb{R}^d$ and a second parameter η belonging to an arbitrary parameter set $\mathcal{H}$. We let $\tilde{\eta}$ be an estimator for η based on a second sample $Y_1, \dots, Y_m$ of observations that is independent of $X_1, \dots, X_n$ and possesses a distribution depending on η only.

In the following, we will test the two-sided hypothesis $H_0: \theta \in \Theta_0$ which is of the form $H_0: \Sigma \theta \in \{(c_1, \dots, c_{d-k}, x_{d-k+1}, \dots, x_d)^T: x_{d-k+1}, \dots, x_d \in \mathbb{R}\}$ for $c_1, \dots, c_{d-k} \in \mathbb{R}$ given and Σ a given $d \times d$ invertible matrix. If Σ equals the identity matrix and $c_1 = \dots = c_{d-k} = 0$ it is tested whether the first $d - k$ coordinates of θ equal zero. We assume that, $\sqrt{n}(\Theta - \theta_0)$ and $\sqrt{n}(\Theta_0 - \theta_0)$ tend to d and k dimensional spaces. The convergence of the sets $\sqrt{n}(\Theta - \theta_0) = \{\sqrt{n}(\theta - \theta_0): \theta \in \Theta\}$ and $\sqrt{n}(\Theta_0 - \theta_0) = \{\sqrt{n}(\theta - \theta_0): \theta \in \Theta_0\}$ is defined as follows (see van der Vaart [2], chapter 16). A set H_n converges to a set H if H is the set of all limits of all converging sequences h_n with $h_n \in H_n$ for all n and, if $h = \lim_i h_{n_i}$ for $n_1 < n_2 < \dots$ with $h_{n_i} \in H_{n_i}$ for every i , then h is contained in H .

To test the null hypothesis, we use the likelihood ratio statistic in which we replace the parameter η by its estimator $\tilde{\eta}$:

$$\Lambda_{m,n} := 2 \log \frac{\sup_{\theta \in \Theta} \prod_{i=1}^n p_{\theta, \tilde{\eta}}(X_i)}{\sup_{\theta \in \Theta_0} \prod_{i=1}^n p_{\theta, \tilde{\eta}}(X_i)} = 2 \log \frac{\prod_{i=1}^n p_{\hat{\theta}, \tilde{\eta}}(X_i)}{\prod_{i=1}^n p_{\hat{\theta}_0, \tilde{\eta}}(X_i)}.$$

In the theorem below, we will determine the asymptotic distribution of the test-statistic under $\theta = \theta_0$ with $\theta_0 \in \Theta_0 \subset \Theta$.

Theorem

Under the assumptions stated below and under $\theta = \theta_0$ the test-statistic $\Lambda_{m,n}$ converges in distribution to the distribution of a linear combination of $d - k$ independent chi-squared distributed variables each with one degree of freedom. The coefficients in the linear combination equal $1 + \lambda d_i, i = 1, \dots, d - k$ with $\lambda = \lim_{n, m \rightarrow \infty} n/m$ and d_i positive values equal to the eigenvalues of a covariance matrix that can be computed explicitly.

For $k = 0$ (the dimension of the null space is zero), the asymptotic distribution has a nice form and is given directly after the assumptions. For $k > 0$ the values d_i can be found in the proof of the theorem.

Assumptions:

1. Assume that for $l_{\theta, \eta}(x) = \log p_{\theta, \eta}(x)$ the log likelihood function in (θ, η) , its first and second derivative with respect to θ , $\dot{l}_{\theta, \eta}(x) = \frac{\partial}{\partial \theta} l_{\theta, \eta}(x)$ and $\ddot{l}_{\theta, \eta}(x) = \frac{\partial^2}{\partial \theta^2} l_{\theta, \eta}(x)$ exist for all x .
2. Assume that for $\hat{\theta} = \operatorname{argmax}_{\theta \in \Theta} \prod_{i=1}^n p_{\theta, \tilde{\eta}}(X_i)$ it holds that $\hat{\theta} - \theta_0 = O_p(n^{-1/2})$ for θ_0 the true value.
3. Assume that the sequence $-\mathbb{P}_n \ddot{l}_{\theta_0, \tilde{\eta}} = -n^{-1} \sum_{i=1}^n \ddot{l}_{\theta_0, \tilde{\eta}}(X_i)$ converges in probability to $I_{\theta_0, \eta_0} = \operatorname{Var}_{\theta_0, \eta_0} \dot{l}_{\theta_0, \eta_0}(X_i)$ the Fisher information matrix for θ in the model where $\eta = \eta_0$ is known to be the true value.
4. Assume that $\mathbb{G}_n \dot{l}_{\theta_0, \tilde{\eta}} = \sqrt{n}(\mathbb{P}_n - P_{\theta_0, \eta_0}) \dot{l}_{\theta_0, \tilde{\eta}} = \mathbb{G}_n \dot{l}_{\theta_0, \eta_0} + o_p(1)$ converges in distribution to a normal distribution with mean zero and covariance matrix equal to the Fisher information I_{θ_0, η_0} .
5. Assume that $\sqrt{m} P_{\theta_0, \tilde{\eta}} \dot{l}_{\theta_0, \eta_0}$ is asymptotically normal with mean zero and covariance matrix J_{θ_0, η_0} .
6. Assume that $\sqrt{n}(P_{\theta_0, \eta_0} - P_{\theta_0, \tilde{\eta}})(\dot{l}_{\theta_0, \tilde{\eta}} - \dot{l}_{\theta_0, \eta_0})$ converges in probability to zero.
7. Assume that $l_{\theta, \eta}^{(3)}(x) = \frac{\partial}{\partial \theta} \ddot{l}_{\theta, \eta}(x)$ exists $\forall x$ and in a small neighborhood of (θ_0, η_0) and that

$$\sup_{\sqrt{n} \|\hat{\theta} - \theta_0\| \leq M_\theta, m^r \|\tilde{\eta} - \eta_0\|_\infty \leq M_\eta} \left| \frac{1}{n} \sum_{i=1}^n l_{\hat{\theta}, \tilde{\eta}}^{(3)}(X_i) \right|$$

is bounded in probability for m, n to infinity, and for $M_\theta > 0$ and $M_\eta > 0$ constants and so that $m^r \|\tilde{\eta} - \eta_0\|_\infty = O_p(1)$.

If $n/m \rightarrow 0$, $\lambda = 0$ and $\Lambda_{m,n}$ converges in distribution to a chi-squared distribution with $d - k$ degrees of freedom. So, if $m \gg n$ the distribution of the test-statistic is hardly affected if the estimated nuisance parameters are seen as the true values and the level of the test is close to the supposed level under the assumption of known nuisance parameters. However, also if $m < n$ and the values of d_i are close to zero, the distribution of the likelihood ratio statistic may be close to the chi-squared distribution with $d - k$ degrees of freedom (see Example 4).

In the theorem, the exact asymptotic distribution is not given. However, if the dimension of the null space Θ_0 is zero (Θ_0 consists of a single point, $k = 0$) the asymptotic distribution of the likelihood ratio statistic has a nice form: the likelihood ratio statistic $\Lambda_{m,n}$ is asymptotically distributed as $\sum_{i=1}^d (1 + \lambda d_i) Z_i^2$ with $\lambda = \lim_{m, n} n/m$ and $d_1, \dots, d_d$ the eigenvalues of the covariance matrix $I_{\theta_0, \eta_0}^{-1/2} J_{\theta_0, \eta_0} I_{\theta_0, \eta_0}^{-1/2}$ and $Z_1^2, \dots, Z_d^2$ independent random variables with a chi-squared one distribution (the matrices I_{θ_0, η_0} and J_{θ_0, η_0} were defined in the assumptions 3 and 5). If $d = 1$ then, $I_{\theta_0, \eta_0}^{-1/2} J_{\theta_0, \eta_0} I_{\theta_0, \eta_0}^{-1/2} = J_{\theta_0, \eta_0} I_{\theta_0, \eta_0}^{-1}$ and the asymptotic distribution simplifies to $(1 + \lambda J_{\theta_0, \eta_0} I_{\theta_0, \eta_0}^{-1}) Z^2$; a chi-squared one distribution multiplied with a correction factor bigger than one.

For a statistical model probably the most difficult assumption to check is the asymptotic normality of $\sqrt{m}P_{\theta_0, \tilde{\eta}} \dot{l}_{\theta_0, \eta_0}$ in case $\tilde{\eta}$ is a non-parametric estimator of η which converges at a lower rate than $m^{1/2}$ (assumption 5). For example, if $\tilde{\eta}$ is the non-parametric maximum likelihood estimator (NPMLE) of a distribution function based on interval censored data, we know that $\tilde{\eta}$ converges pointwise at rate $n^{1/3}$, but, under some assumptions, some functionals of it, for instance $\int x(\tilde{\eta}(x) - \eta(x))dx$, converge at rate $n^{1/2}$ to normal distributions (theorem 5.1 in Huang and Wellner [21]). So, even if the convergence rate of $\tilde{\eta}$ is lower than $m^{1/2}$, it may be possible to prove that $\sqrt{m}P_{\theta_0, \tilde{\eta}} \dot{l}_{\theta_0, \eta_0}$ is asymptotically normal.

Proof

In order to test the hypothesis $H_0 : \theta \in \Theta_0$ with test-statistic $\Lambda_{m,n}$, we have to compute the (asymptotic) distribution of $\Lambda_{m,n}$ under the null hypothesis. Remember that θ_0 is a point in Θ and Θ_0 so that $\sqrt{n}(\Theta - \theta_0)$ and $\sqrt{n}(\Theta_0 - \theta_0)$ tend to d and k dimensional spaces. Now write

$$\begin{aligned} \Lambda_{m,n} &:= 2 \log \frac{\sup_{\theta \in \Theta} \prod_{i=1}^n p_{\theta, \tilde{\eta}}(X_i)}{\sup_{\theta \in \Theta_0} \prod_{i=1}^n p_{\theta, \tilde{\eta}}(X_i)} \\ &= 2 \log \frac{\sup_{\theta \in \Theta} \prod_{i=1}^n p_{\theta, \tilde{\eta}}(X_i)}{\prod_{i=1}^n p_{\theta_0, \tilde{\eta}}(X_i)} + 2 \log \frac{\prod_{i=1}^n p_{\theta_0, \tilde{\eta}}(X_i)}{\sup_{\theta \in \Theta_0} \prod_{i=1}^n p_{\theta, \tilde{\eta}}(X_i)} \\ &= 2 \sum_{i=1}^n \log \frac{p_{\hat{\theta}, \tilde{\eta}}(X_i)}{p_{\theta_0, \tilde{\eta}}(X_i)} - 2 \sum_{i=1}^n \log \frac{p_{\hat{\theta}_0, \tilde{\eta}}(X_i)}{p_{\theta_0, \tilde{\eta}}(X_i)} \end{aligned} \quad (1)$$

for $\hat{\theta}$ the maximum likelihood estimator of θ and $\hat{\theta}_0$ the restricted maximum likelihood; restricted to the parameter space Θ_0 after inserting $\tilde{\eta}$ into the likelihood. Under assumption 2 and under $\theta_0 \in \Theta_0$ the maxima over $\theta \in \Theta$ and $\theta \in \Theta_0$ are in $1/\sqrt{n}$ -shrinking neighborhoods of θ_0 . That means that $\hat{\theta} = \theta_0 + \hat{h}/\sqrt{n}$ for some value $\hat{h} \in \sqrt{n}(\Theta - \theta_0)$ and $\hat{\theta}_0 = \theta_0 + \hat{h}/\sqrt{n}$ for $\hat{h} \in \sqrt{n}(\Theta_0 - \theta_0)$. Instead of maximizing the likelihood $\prod_{i=1}^n p_{\theta, \tilde{\eta}}$ in the test-statistic with respect to $\theta \in \Theta$ or $\theta \in \Theta_0$, the product $\prod_{i=1}^n p_{\theta_0 + h/\sqrt{n}, \tilde{\eta}}$ can be maximized with respect to $h \in H_n = \sqrt{n}(\Theta - \theta_0)$ or $h \in H_{n,0} = \sqrt{n}(\Theta_0 - \theta_0)$. This yields the likelihood ratio statistic

$$\Lambda_{m,n} = 2 \sup_{h \in H_n} \sum_{i=1}^n \log \frac{p_{\theta_0 + h/\sqrt{n}, \tilde{\eta}}(X_i)}{p_{\theta_0, \tilde{\eta}}(X_i)} - 2 \sup_{h \in H_{n,0}} \sum_{i=1}^n \log \frac{p_{\theta_0 + h/\sqrt{n}, \tilde{\eta}}(X_i)}{p_{\theta_0, \tilde{\eta}}(X_i)}.$$

By a Taylor expansion and by assumption 7

$$\begin{aligned} \sum_{i=1}^n \log \frac{p_{\theta_0 + h/\sqrt{n}, \tilde{\eta}}(X_i)}{p_{\theta_0, \tilde{\eta}}(X_i)} &= \sum_{i=1}^n \log p_{\theta_0 + h/\sqrt{n}, \tilde{\eta}}(X_i) - \sum_{i=1}^n \log p_{\theta_0, \tilde{\eta}}(X_i) \\ &= h^T \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{l}_{\theta_0, \tilde{\eta}}(X_i) - \frac{1}{2n} h^T \sum_{i=1}^n \ddot{l}_{\theta_0, \tilde{\eta}}(X_i) h + O_p\left(\frac{1}{\sqrt{n}}\right) \\ &= h^T \mathbb{G}_n \dot{l}_{\theta_0, \tilde{\eta}} + h^T \sqrt{n} P_{\theta_0, \eta_0} \dot{l}_{\theta_0, \tilde{\eta}} - \frac{1}{2} h^T \mathbb{P}_n \ddot{l}_{\theta_0, \tilde{\eta}} h + O_p(1/\sqrt{n}), \end{aligned} \quad (2)$$

with

$$\mathbb{G}_n \dot{l}_{\theta_0, \tilde{\eta}} = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \dot{l}_{\theta_0, \tilde{\eta}}(X_i) - P_{\theta_0, \eta_0} \dot{l}_{\theta_0, \tilde{\eta}} \right)$$

$$\mathbb{P}_n \ddot{l}_{\theta_0, \tilde{\eta}}(X_i) = \frac{1}{n} \sum_{i=1}^n \ddot{l}_{\theta_0, \tilde{\eta}}(X_i).$$

The term in eq. (2) should be maximized with respect to h .

We consider the three terms in eq. (2) separately. By assumption 4 the sequence $\mathbb{G}_n \dot{l}_{\theta_0, \tilde{\eta}}$ converges in distribution to $N(0, I_{\theta_0, \eta_0})$ and $-\mathbb{P}_n \ddot{l}_{\theta_0, \tilde{\eta}}(X_i)$ converges in probability to I_{θ_0, η_0} by assumption 3. Left is the second term in eq. (2). For any η in its parameter space $\mathcal{H}$, $P_{\theta_0, \eta} \dot{l}_{\theta_0, \eta} = 0$. Write

$$\begin{aligned} \sqrt{n} P_{\theta_0, \eta_0} \dot{l}_{\theta_0, \tilde{\eta}} &= \sqrt{n} (P_{\theta_0, \eta_0} - P_{\theta_0, \tilde{\eta}}) \dot{l}_{\theta_0, \tilde{\eta}} \\ &= \sqrt{n} (P_{\theta_0, \eta_0} - P_{\theta_0, \tilde{\eta}}) \dot{l}_{\theta_0, \eta_0} + \sqrt{n} (P_{\theta_0, \eta_0} - P_{\theta_0, \tilde{\eta}}) (\dot{l}_{\theta_0, \tilde{\eta}} - \dot{l}_{\theta_0, \eta_0}). \end{aligned}$$

By assumption 6 the last term at the right-hand side converges in probability to zero. Thus

$$\sqrt{n} P_{\theta_0, \eta_0} \dot{l}_{\theta_0, \tilde{\eta}} = -\sqrt{n} P_{\theta_0, \tilde{\eta}} \dot{l}_{\theta_0, \eta_0} + o_p(1) = -\frac{\sqrt{n}}{\sqrt{m}} \sqrt{m} P_{\theta_0, \tilde{\eta}} \dot{l}_{\theta_0, \eta_0} + o_p(1),$$

what is, by assumption 5, asymptotically normal with mean zero and covariance matrix $\lambda J_{\theta_0, \eta_0}$ with λ equal to the limit of n/m if $m, n \rightarrow \infty$.

Combining the previous results, by assumption 4 and by independence of the samples $X_1, \dots, X_n$ and $Y_1, \dots, Y_m$ the sum

$$\mathbb{G}_n \dot{l}_{\theta_0, \tilde{\eta}} + \sqrt{n} P_{\theta_0, \eta_0} \dot{l}_{\theta_0, \tilde{\eta}} = \mathbb{G}_n \dot{l}_{\theta_0, \eta_0} + \sqrt{n} P_{\theta_0, \eta_0} \dot{l}_{\theta_0, \tilde{\eta}} + o_p(1)$$

converges to the normal distribution $N(0, I_{\theta_0, \eta_0} + \lambda J_{\theta_0, \eta_0})$. The same computations hold for the second term of eq. (1).

The following is similar to the derivation of the asymptotic distribution of the likelihood ratio statistic, if there is no nuisance parameter (van der Vaart [2], Theorem 16.7). The differentiability of $\log p_{\theta_0 + h/\sqrt{n}, \tilde{\eta}}(X_i)$ with respect to h implies that

$$\mathbb{Z}_n(h) = \sum_{i=1}^n \log \frac{p_{\theta_0 + h/\sqrt{n}, \tilde{\eta}}(X_i)}{p_{\theta_0, \tilde{\eta}}(X_i)} - \left(h^T \mathbb{G}_n \dot{l}_{\theta_0, \tilde{\eta}} + \sqrt{n} h^T P_{\theta_0, \eta_0} \dot{l}_{\theta_0, \tilde{\eta}} - \frac{1}{2} h^T I_{\theta_0, \eta_0} h \right)$$

converges in probability to zero for every h . By assumption 7 (the remainder term of the Taylor expansion converges uniformly in probability to zero) and the fact that

$$\sup_{\|h\| \leq M} |h^T I_{\theta_0, \eta_0} h - h^T \mathbb{P}_n \ddot{l}_{\theta_0, \tilde{\eta}} h| \xrightarrow{P} 0 \quad \forall M > 0,$$

there is also uniform convergence of $\mathbb{Z}$ on a bounded set:

$$\sup_{\|h\| \leq M} |\mathbb{Z}_n(h)| \xrightarrow{P} 0 \quad \forall M > 0.$$

The convergence in the previous display also holds if $M = M_n$ increases to infinity sufficiently slowly. Since $\hat{\theta}$ and $\hat{\theta}_0$ are both $\sqrt{n}$ consistent, both estimators are within a ball of radius $M_n/\sqrt{n}$ around θ_0 with probability tending to 1. Then $\hat{h}$ is within a ball of radius M_n around zero with probability tending to 1. Consequently, the limit distribution of $\Lambda_{m,n}$ does not change if H_n or $H_{n,0}$ are replaced by the intersection of these sets and the ball around zero with radius M_n . These intersected sets also converge to H and H_0 , respectively. By uniform convergence we conclude that for $G_{n, \tilde{\eta}} = \mathbb{G}_n \dot{l}_{\theta_0, \tilde{\eta}} + \sqrt{n} P_{\theta_0, \eta_0} \dot{l}_{\theta_0, \tilde{\eta}}$

$$\begin{aligned}
\Lambda_{m,n} &= 2 \sup_{h \in H_n} \left(h^T \mathbb{G}_n \dot{l}_{\theta_0, \tilde{\eta}} + \sqrt{n} h^T P_{\theta_0, \eta_0} \dot{l}_{\theta_0, \tilde{\eta}} - \frac{1}{2} h^T I_{\theta_0, \eta_0} h \right) \\
&\quad - 2 \sup_{h \in H_{n,0}} \left(h^T \mathbb{G}_n \dot{l}_{\theta_0, \tilde{\eta}} + \sqrt{n} h^T P_{\theta_0, \eta_0} \dot{l}_{\theta_0, \tilde{\eta}} - \frac{1}{2} h^T I_{\theta_0, \eta_0} h \right) + o_p(1) \\
&= \inf_{h \in H_{n,0}} (-2h^T G_{n, \tilde{\eta}} + h^T I_{\theta_0, \eta_0} h) - \inf_{h \in H_n} (-2h^T G_{n, \tilde{\eta}} + h^T I_{\theta_0, \eta_0} h) + o_p(1) \\
&= \inf_{h \in H_{n,0}} \left(I_{\theta_0, \eta_0}^{-1/2} G_{n, \tilde{\eta}}^T G_{n, \tilde{\eta}} I_{\theta_0, \eta_0}^{-1/2} - 2h^T G_{n, \tilde{\eta}} + h^T I_{\theta_0, \eta_0} h \right) \\
&\quad - \inf_{h \in H_n} \left(I_{\theta_0, \eta_0}^{-1/2} G_{n, \tilde{\eta}}^T G_{n, \tilde{\eta}} I_{\theta_0, \eta_0}^{-1/2} - 2h^T G_{n, \tilde{\eta}} + h^T I_{\theta_0, \eta_0} h \right) + o_p(1) \\
&= \| I_{\theta_0, \eta_0}^{-1/2} G_{n, \tilde{\eta}} - I_{\theta_0, \eta_0}^{1/2} H_{0,n} \|^2 - \| I_{\theta_0, \eta_0}^{-1/2} G_{n, \tilde{\eta}} - I_{\theta_0, \eta_0}^{1/2} H_n \|^2 + o_p(1) \\
&= \| I_{\theta_0, \eta_0}^{-1/2} G - I_{\theta_0, \eta_0}^{1/2} H_0 \|^2 - \| I_{\theta_0, \eta_0}^{-1/2} G - I_{\theta_0, \eta_0}^{1/2} H \|^2 + o_p(1)
\end{aligned} \tag{3}$$

where the last equality follows from Lemma 7.13 in van der Vaart [2] with H and H_0 the limit sets of H_n and $H_{n,0}$ as explained at the beginning of this section and G distributed as $N(0, I_{\theta_0, \eta_0} + \lambda J_{\theta_0, \eta_0})$, the limit distribution of $G_{n, \tilde{\eta}}$ (as was seen before).

Define $Z := I_{\theta_0, \eta_0}^{-1/2} G$, then Z possesses a normal distribution with mean zero and covariance matrix

$$Q_{\theta_0, \eta_0} := I + \lambda I_{\theta_0, \eta_0}^{-1/2} J_{\theta_0, \eta_0} I_{\theta_0, \eta_0}^{-1/2},$$

with I the $d \times d$ identity matrix. Furthermore, since θ_0 is an inner point of Θ , the set H is the full space in $\mathbb{R}^k$ and the second term in eq. (3) equals zero, and

$$\Lambda_{m,n} \rightsquigarrow \| Z - I_{\theta_0, \eta_0}^{1/2} H_0 \|^2 =: \Lambda$$

for m, n to infinity.

Suppose that $k = 0$, then H_0 equals the singleton $\{0\}$ in a d -dimensional space and the asymptotic distribution of $\Lambda_{m,n}$ is equal to the law of $\|Z\|^2$ with Z normally distributed with mean zero and covariance matrix Q_{θ_0, η_0} . The covariance matrix Q_{θ_0, η_0} is a symmetric positive definite matrix and can therefore be factorized as $Q_{\theta_0, \eta_0}^{1/2} = U^T D^{1/2} U$ with U an orthonormal matrix and D a diagonal matrix with the eigenvalues of Q_{θ_0, η_0} on the diagonal. Then $Q_{\theta_0, \eta_0} = Q_{\theta_0, \eta_0}^{1/2} Q_{\theta_0, \eta_0}^{1/2}$ with $Q_{\theta_0, \eta_0}^{1/2} = U^T D^{1/2} U$ and

$$\|Z\|^2 = \|Q_{\theta_0, \eta_0}^{1/2} Q_{\theta_0, \eta_0}^{-1/2} Z\|^2 = \|U^T D^{1/2} U (Q_{\theta_0, \eta_0}^{-1/2} Z)\|^2.$$

Now, $Q_{\theta_0, \eta_0}^{-1/2} Z$ has a d dimensional normal distribution with mean zero and covariance matrix the identity matrix. Since U is an orthonormal matrix and only rotates the vector and thus $\|Ux\| = \|x\|$, the last term in the previous display equals $\|D^{1/2} U (Q_{\theta_0, \eta_0}^{-1/2} Z)\|^2$. Furthermore, $U (Q_{\theta_0, \eta_0}^{-1/2} Z)$ and $Q_{\theta_0, \eta_0}^{-1/2} Z$ are equally distributed by symmetry of the standard normal distribution. Consequently

$$\|D^{1/2} U (Q_{\theta_0, \eta_0}^{-1/2} Z)\|^2 = \|D^{1/2} Q_{\theta_0, \eta_0}^{-1/2} Z\|^2 = \sum_{i=1}^d e_i Z_i^2,$$

with $e_1, \dots, e_d$ the diagonal elements of the matrix D and $Z_1, \dots, Z_d$ independent standard normal distributed random variables. So, Z_i^2 has a chi-squared distribution with one degree of freedom and $\Lambda_{m,n}$ converges in distribution to a linear combination of independent chi-squared random variables (all with one degree of freedom): $\sum_{i=1}^d e_i Z_i^2$. The eigenvalues $e_1, \dots, e_d$ of the matrix Q_{θ_0, η_0} equal $e_i = 1 + \lambda d_i$, $i = 1, \dots, d$ with $d_1, \dots, d_d$ the eigenvalues of the covariance matrix $I_{\theta_0, \eta_0}^{-1/2} J_{\theta_0, \eta_0} I_{\theta_0, \eta_0}^{-1/2}$.

Now, let $k > 0$. Remember that the asymptotic distribution of $\Lambda_{m,n}$ equals the law of $\|Z - I_{\theta_0, \eta_0}^{1/2} H_0\|^2$ for H_0 a subspace of dimension k . If P is the orthogonal projection onto the orthocomplement of $I_{\theta_0, \eta_0}^{1/2} H_0$, which is a $d - k$ dimensional linear subspace of $\mathbb{R}^d$, then

$$\|Z - I_{\theta_0, \eta_0}^{1/2} H_0\|^2 = \|PZ\|^2 = \|W\|^2, \quad W \sim N(0, PQ_{\theta_0, \eta_0} P^T).$$

In the situation that $k = 0$ the covariance matrix $Q_{\theta_0, \eta_0}^{1/2}$ of Z was factorized in order to rewrite the quadratic norm $\|Z\|^2$ into a sum. The same can now be done for the covariance matrix $PQ_{\theta_0, \eta_0} P^T$ of W . It follows that $\|W\|^2$ equals the sum $\sum_{i=1}^d e_i Z_i^2$ with $Z_1^2, \dots, Z_d^2$ chi-squared one distributed and $e_1, \dots, e_d$ equal to the eigenvalues of $PQ_{\theta_0, \eta_0} P^T$. We still have to compute these eigenvalues; we will show that k eigenvalues equal zero and that the remaining $d - k$ eigenvalues are strictly positive and equal to $e_i = 1 + \lambda d_i$, $i = 1, \dots, d - k$ with $d_1, \dots, d_{d-k}$ the strictly positive eigenvalues of $PS_{\theta_0, \eta_0} P^T$ with $S_{\theta_0, \eta_0} := I_{\theta_0, \eta_0}^{-1/2} J_{\theta_0, \eta_0} I_{\theta_0, \eta_0}^{-1/2}$. Then, the sum reduces to $\sum_{i=1}^{d-k} (1 + \lambda d_i) Z_i^2$.

The derivation of the eigenvalues is as follows. The covariance matrix $PQ_{\theta_0, \eta_0} P^T = PP^T + \lambda PS_{\theta_0, \eta_0} P^T = P + \lambda PS_{\theta_0, \eta_0} P^T$. It vanishes on $I_{\theta_0, \eta_0}^{1/2} H_0$ and acts as the strictly positive-definite operator $I + \lambda PS_{\theta_0, \eta_0} P^T$ on the orthocomplement of $I_{\theta_0, \eta_0}^{1/2} H_0$. It follows that $PQ_{\theta_0, \eta_0} P^T$ has 0 as an eigenvalue of multiplicity k (the dimension of $I_{\theta_0, \eta_0}^{1/2} H_0$) and has $d - k$ strictly positive eigenvalues of the form $e_i = 1 + \lambda d_i$, where d_i is a positive eigenvalue of $PS_{\theta_0, \eta_0} P^T$. $\square$

If $d = 1$ then $I_{\theta_0, \eta_0}^{-1/2} J_{\theta_0, \eta_0} I_{\theta_0, \eta_0}^{-1/2} = J_{\theta_0, \eta_0} I_{\theta_0, \eta_0}^{-1}$ and the asymptotic distribution of the likelihood ratio statistic simplifies to $(1 + \lambda J_{\theta_0, \eta_0} I_{\theta_0, \eta_0}^{-1}) Z^2$ where Z^2 follows the chi-squared distribution with one degree of freedom. The matrix J_{θ_0, η_0} is the asymptotic covariance matrix of $\sqrt{m} P_{\theta_0, \eta} \dot{l}_{\theta_0, \eta_0}$ with $P_{\theta_0, \eta} \dot{l}_{\theta_0, \eta_0}$ the “plug in” estimator for the parameter $\eta \mapsto P_{\theta_0, \eta} \dot{l}_{\theta_0, \eta_0}$ based on the sample $Y_1, \dots, Y_m$. The Fisher information matrix I_{θ_0, η_0} is the asymptotic covariance matrix of the alternative estimator $\mathbb{P}_n \dot{l}_{\theta_0, \eta_0}$ based on $X_1, \dots, X_n$. So, the term $J_{\theta_0, \eta_0} I_{\theta_0, \eta_0}^{-1}$ measures the relative efficiency at $\eta = \eta_0$ for estimating $P_{\theta_0, \eta} \dot{l}_{\theta_0, \eta_0}$ using one observation Y compared to one observation X . Then $J_{\theta_0, \eta_0}/m$ and $I_{\theta_0, \eta_0}/n$ are the variances of $P_{\theta_0, \eta} \dot{l}_{\theta_0, \eta_0}$ and $\mathbb{P}_n \dot{l}_{\theta_0, \eta_0}$ based on the samples $Y_1, \dots, Y_m$ and $X_1, \dots, X_n$, respectively. The fraction $(J_{\theta_0, \eta_0}/m)(I_{\theta_0, \eta_0}/n)^{-1} = (n/m)(J_{\theta_0, \eta_0}/I_{\theta_0, \eta_0})$ measures the relative efficiency of the two samples for estimating $P_{\theta_0, \eta} \dot{l}_{\theta_0, \eta_0}$. This fraction is approximately equal to $\lambda J_{\theta_0, \eta_0} I_{\theta_0, \eta_0}^{-1}$ as in the asymptotic distribution of the likelihood ratio statistic. So, not only the sample sizes of the two samples are important but also the estimation precision of the estimators (the asymptotic variances). This is illustrated in Example 4.

Next we consider the asymptotic distribution of the likelihood ratio test-statistic for the situation that Θ is a half-space and Θ_0 its boundary. Suppose we want to test the null hypothesis $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta/\Theta_0$. Let $\theta_0 \in \Theta_0$ and define, as before, $H_n = \sqrt{n}(\Theta - \theta_0)$ and $H_{n,0} = \sqrt{n}(\Theta_0 - \theta_0)$. Assume that H_n and $H_{n,0}$ tend to a half space H in $\mathbb{R}^d$ and its boundary H_0 , respectively, like Θ_0 and Θ .

Under the assumptions mentioned earlier, $\Lambda_{m,n}$ converges, under $\theta = \theta_0$, in distribution to a mixture of a point mass at 0 and the distribution of a linear combination of $d - k$ independent chi-squared distributed random variables, each with one degree of freedom. The coefficients of the linear combination are as described before.

The proof is very similar up to the following. Like in the previous proof

$$\Lambda = \|I_{\theta_0, \eta_0}^{-1/2} G - I_{\theta_0, \eta_0}^{1/2} H_0\|^2 - \|I_{\theta_0, \eta_0}^{-1/2} G - I_{\theta_0, \eta_0}^{1/2} H\|^2,$$

but now the term $\|I_{\theta_0, \eta_0}^{-1/2} G - I_{\theta_0, \eta_0}^{1/2} H\|^2$ is not equal to zero with probability one. The sets $H' := I_{\theta_0, \eta_0}^{1/2} H$ and $H'_0 := I_{\theta_0, \eta_0}^{1/2} H_0$ are a halfspace and its boundary, because this is true for H and H_0 by assumption, and the boundary contains 0, because $\theta_0 \in \Theta_0$. If $Z := I_{\theta_0, \eta_0}^{-1/2} G$ falls in the halfspace H' , the second term in Λ

vanishes, and Λ equals the square distance of Z to the boundary of the halfspace H' . If Z falls outside H' , then Λ equals zero, because in that case the distance of Z to H' is equal to the distance to its boundary, as the projection of Z onto H' is on the boundary of H' . Thus Λ is equal to zero with probability that Z falls outside H' . If Z falls into H' , the right term vanishes and $\Lambda = \|I_{\theta_0, \eta_0}^{-1/2} G - I_{\theta_0, \eta_0}^{1/2} H_0\|^2$. This term can be considered as before.

Similar reasoning holds for other hypotheses, like the situation for the one-sided hypothesis.

3 Examples

In the following we will give some examples for illustration of the theorem.

3.1 Example 1: Theoretical, one-dimension

Let $X_1, \dots, X_n$ be a sample from the gamma distribution with shape parameter α and rate λ . Furthermore, let $Y_1, \dots, Y_m$ be a sample from the exponential distribution with parameter $1/\alpha$. The two samples are independent. Suppose the null hypothesis $H_0 : \lambda = \lambda_0$ is tested versus $H_0 : \lambda \neq \lambda_0$ with $\lambda_0 > 0$ a given value. The likelihood ratio test-statistic is used with the parameter α replaced by its maximum likelihood estimator based on the second sample: $\tilde{\alpha} = \bar{Y}$. Then the likelihood ratio statistic is given by:

$$\Lambda_{m,n} = 2 \left(\log \left(\frac{\hat{\lambda}}{\lambda_0} \right)^{n\tilde{\alpha}} \exp \left((\lambda_0 - \hat{\lambda}) \sum_{i=1}^n X_i \right) \right) = 2n\bar{Y} \log \left(\frac{\hat{\lambda}}{\lambda_0} \right) + 2(\lambda_0 - \hat{\lambda}) \sum_{i=1}^n X_i,$$

since $\hat{\lambda} = \tilde{\alpha}/\bar{X} = \bar{Y}/\bar{X}$. This estimator is asymptotically consistent since $\bar{X}$ converges in probability to α/λ if $n \rightarrow \infty$, $\bar{Y}$ to α if $n \rightarrow \infty$ and the two samples are independent. By a Taylor series of the likelihood ratio statistic with respect to $\hat{\lambda}$ in a neighborhood of λ_0 , the likelihood ratio statistic $\Lambda_{m,n}$ approximately equals

$$\begin{aligned} & 2n\bar{Y} \left(\left(\frac{\hat{\lambda}}{\lambda_0} - 1 \right) - \frac{1}{2} \left(\frac{\hat{\lambda}}{\lambda_0} - 1 \right)^2 \right) + 2(\lambda_0 - \hat{\lambda}) \sum_{i=1}^n X_i + o_P(1) \\ &= 2n \frac{\bar{Y}}{\lambda_0} (\hat{\lambda} - \lambda_0) - n \frac{\bar{Y}}{\lambda_0^2} (\hat{\lambda} - \lambda_0)^2 + 2n(\lambda_0 - \hat{\lambda})\bar{X} + o_P(1) \\ &= \frac{\alpha}{\lambda_0^2} n \left(\frac{\bar{Y}}{\bar{X}} - \lambda_0 \right)^2 + o_P(1) = n \left(\frac{\alpha}{m} + \frac{1}{n} \right) \frac{(\bar{Y} - \lambda_0 \bar{X})^2}{\frac{\alpha^2}{m} + \frac{\alpha}{n}} + o_P(1). \end{aligned}$$

By the central limit theorem and by the independence between the two samples $(\bar{Y} - \lambda_0 \bar{X})/\sqrt{\alpha^2/m + \alpha/n}$ is asymptotically standard normal distributed with mean zero. The test-statistic $\Lambda_{m,n}$ is therefore asymptotically distributed as $1 + \alpha v$ times a chi-squared 1 distributed random variable, with v equal to the limit $\lim_{n,m \rightarrow \infty} n/m$.

Since $I_{\alpha,\lambda} = \alpha/\lambda^2$ and $J_{\alpha,\lambda} = \alpha^2/\lambda^2$, $I_{\alpha,\lambda}^{-1} J_{\alpha,\lambda} = \alpha$, like it should be.

Suppose that $n = 100$ and $m = 1,000$, then $v = 0.1$, and that $\alpha = 10$. Then, the asymptotic distribution of the likelihood ratio statistic is distributed as $1 + \alpha v = 1 + 10 * 0.1 = 2$ times a chi-squared distributed random variable with one degree of freedom instead of 1 times this chi-squared distribution. Note that this asymptotic distribution is independent of the value of λ . The value of α determines the shape of the exponential distribution and of the gamma distribution. If α is small the mean and especially the variance of the exponential distribution are large and an estimate of α based on the sample $Y_1, \dots, Y_m$ is imprecise. Therefore the multiplication term is quite large.

3.2 Example 2: Theoretical, Semi-parametric Cox-model

Suppose we wish to estimate a distribution function F with $F(t) = 1 - S_0(t)^{\exp \beta}$ at the interval $[0, \tau]$ for $\tau > 0$ and $S_0(\tau) > 0$. The function S_0 is an unknown survival function and β is an unknown parameter in $\mathbb{R}$, the parameter of interest. Let $T_1, \dots, T_n$ be a sample from F . We observe $\min\{T_1, \tau\}, \dots, \min\{T_n, \tau\}$ and define Δ_i as the indicator function that indicates whether T_i was before or after τ : $\Delta_i := 1_{T_i \leq \tau}$.

The nuisance parameter S_0 is estimated by the empirical distribution function of a sample $Y_1, \dots, Y_m$ from a distribution with survival curve S_0 . The aim is to estimate β and to test the hypothesis $H_0 : \beta = \beta_0$ versus $H_0 : \beta \neq \beta_0$.

The interpretation of this problem is as follows. Suppose the survival function of a particular group in society, S_0 , is known (i.e. estimated based on an independent large sample); for instance males or non-smokers. We want to estimate the survival curve for a different group (women or smokers). We assume that the survival curve for this group equals $S_0^{\exp \beta}$ for S_0 the survival curve for the first group and β an unknown value in $\mathbb{R}$. The aim is to test whether $\beta = \beta_0$ differs from zero (In particular $\beta_0 = 0$ is of interest.). In order to test β , we use the likelihood ratio test where the survival curve S_0 in the likelihood ratio statistic is estimated based on the first sample, as described before. To compute the asymptotic distribution of the test-statistic under the null hypothesis, the seven assumptions have to be checked.

All assumptions are checked in the appendix and are satisfied. By the theorem, the likelihood ratio statistic is asymptotically distributed as $(1 + \lambda J_{\beta_0, S_0} I_{\beta_0, S_0}^{-1})$ times a chi-squared distributed random variable with one degree of freedom with $I_{\beta_0, S_0} = 1 - S_0(\tau)$ and

$$J_{\beta_0, S_0} = - \int_0^\tau (\exp(\beta_0)(\log S_0(\tau) - \log S_0(t)) - 1)^2 S_0(t)^{2 \exp \beta_0 - 2} \exp(2\beta_0) dS_0(t).$$

This integral can be computed explicitly, but the formula found is complicated and not very interesting on itself. If $\beta_0 = 0$, the value of $J_{\beta_0=0, S_0}$ reduces to $(1 - S_0(\tau) + \log^2 S_0(\tau))$ and

$$\begin{aligned} J_{\beta_0=0, S_0} I_{\beta_0=0, S_0}^{-1} &= (1 - S_0(\tau) + \log^2 S_0(\tau)) / (1 - S_0(\tau)) \\ &= 1 + \log^2 S_0(\tau) / (1 - S_0(\tau)). \end{aligned}$$

If $S_0(\tau) = 0.5$, $J_{\beta_0=0, S_0} I_{\beta_0=0, S_0}^{-1} = 1.96$ and if $S_0(\tau) = 0.8$, $J_{\beta_0=0, S_0} I_{\beta_0=0, S_0}^{-1} = 1.25$. Suppose $\lambda = 0.1$ and $S_0(\tau) = 0.5$, then the asymptotic distribution is 1.2 times a chi-squared distribution with one degree of freedom. Ignoring the fact that S_0 is unknown and was estimated by an independent sample yields a test with level 0.074 instead of 0.05.

3.3 Example 3: Practical, large m

In Jonker et al. [15] the aim is to test the genetic contribution to the age at which people experience their first migraine attack (heritability) and to find locations on the chromosomes that show linkage with the genes that influence this age at onset of migraine. The data came from a longitudinal study of Dutch twins and their family members. A detailed description of the data and the model used is given in Jonker et al. [15]. The parameters of interest describe the amount of heritability and linkage; they were finite dimensional. The (non-parametric) nuisance parameter was the survival function for age at onset of migraine for males and females separately. A one-sided hypothesis for heritability and linkage was tested with the likelihood ratio test, based on (genetic and survival) data of 258 dizygotic twins. The survival functions in the test-statistic were replaced by their NPMEs based on a set of interval-censored age at onset data of 4,791 Dutch males and 6,796 Dutch females; for them no genetic data were available. A simulation study was performed to find the distribution of the likelihood ratio test-statistic under the null hypothesis. Based on the results, it was concluded that this

asymptotic distribution was close to a 50–50 mixture distribution of zero and a chi-squared distribution with one degree of freedom; this would be the limit distribution if the survival curves were known.

As the size of the additional sample ($m \approx 6,796 + 4,791$) is considerably larger than the sample size involved in constructing the likelihood ratio statistic ($n = 258$), this result may seem as expected. However, it is well known that the NPMLEs of the unknown survival function possess a $m^{1/3}$ -rate of convergence in this example, due to the severe, current status type of censoring (cf. Groeneboom and Wellner [22]). Thus at first it may seem more appropriate to compare $m^{1/3} \approx 22.6$ to $\sqrt{n} \approx 16$ in this example. This comparison would suggest a serious bias in neglecting the fact that the survival functions were estimated, contrary to what was found in the simulation study.

However, the result of the present paper shows that the bias in the level of the likelihood ratio statistic is driven by the rate of estimation of the functional $\eta \mapsto P_{\theta_0, \eta} \dot{\ell}_{\theta_0, \eta_0}$ by the plug-in estimator $P_{\theta_0, \hat{\eta}} \dot{\ell}_{\theta_0, \eta_0}$, rather than by the rate of estimation of the nuisance parameter η itself (see assumption 5 of the main theorem and the discussion following its proof). The apparent insensitivity of the likelihood ratio test is explained by the fact that this particular functional is estimable at $\sqrt{m}$ -rate, rather than $m^{1/3}$ -rate, making the relevant ratio of sample sizes $m/n \approx 44.9$.

3.4 Example 4: Practical, small m

Suppose we are interested in a disease X caused by mutations at a single disease gene. In total there are K different disease alleles. An individual with two mutated, possibly different, alleles is affected. Let q be the frequency of any of the K mutations in the total population and $a_1, a_2, \dots, a_K$ the relative allele frequencies among the disease alleles, so $\sum_{j=1}^K a_j = 1$. Consequently, the disease alleles have allele frequencies $qa_1, qa_2, \dots, qa_K$ in the total population. For F the inbreeding coefficient of a child with consanguineous parents, the probability the child has disease X , is given by $P(X|F) = Fq + (1 - F)q^2$. In this formula, the first term on the right-hand side, Fq represents the fraction of affected children with the alleles identical by descent (IBD, the alleles come from a common ancestor). The second term, $(1 - F)q^2$, is the fraction whose alleles are not IBD. For individuals with $F = 0$, the alleles are independent and this probability simplifies to $P(X|F = 0) = q^2$.

Based on a sample of n children with disease X , most of them with consanguineous parents (so $F > 0$), the aim is to estimate the frequency q . For all these children, the inbreeding coefficient is known and the alleles are observed. So, from this information it is known whether a child is homozygous (i.e. the alleles are the same) or heterozygous (the alleles are different), but not whether the alleles are identical by descent. ten Kate et al. [23] derived an expression for q in terms of the relative allele frequencies, the inbreeding coefficient and the conditional probability an individual is homozygous given he is affected and has inbreeding coefficient F . In case of a sample of n individuals with equal inbreeding coefficient $F > 0$, q can be estimated by replacing the relative allele frequencies in the expression by the sample frequencies and the conditional probability of homozygosity by the fraction of individuals who are homozygous. However, in case of a sample with varying inbreeding coefficients this estimation method does not hold; replacing the inbreeding coefficient in the expression by the average of the inbreeding coefficient in the sample, the estimator becomes biased (see Jonker et al. [13]). As an alternative, we consider two maximum likelihood estimators. The first estimator is based only on a part of the data, whether an individual is homozygous or heterozygous, whereas the second estimator uses all available information: the observed alleles.

Case A: homo- and heterozygosity data

Of a sample of n individuals with disease X , we observe whether the individuals are homozygous or heterozygous. Define the indicator function Δ that indicates whether an individual is homozygous or heterozygous at the disease gene: $\Delta = 1$ if the individual is homozygous and $\Delta = 0$ if he is heterozygous. The conditional probabilities that the individual is homozygous, respectively, heterozygous at the disease gene, equal

$$\begin{aligned}
 P(\Delta = 1|X, F) &= \frac{Fq + (1-F)q^2 \sum_{j=1}^K a_j^2}{Fq + (1-F)q^2} = \frac{F + (1-F)q \sum_{j=1}^K a_j^2}{F + (1-F)q} \\
 P(\Delta = 0|X, F) &= \frac{(1-F)q^2(1 - \sum_{j=1}^K a_j^2)}{Fq + (1-F)q^2} = \frac{(1-F)q(1 - \sum_{j=1}^K a_j^2)}{F + (1-F)q}
 \end{aligned} \tag{4}$$

Based on these probabilities, it directly follows that the likelihood equals

$$\prod_{i=1}^n \frac{\left(F_i + (1-F_i)q \sum_{j=1}^K a_j^2 \right)^{\Delta_i} \left((1-F_i)q \left(1 - \sum_{j=1}^K a_j^2 \right) \right)^{1-\Delta_i}}{F_i + (1-F_i)q} \tag{5}$$

where the i in the subscript refers to the i th individual.

Case B: Allele information

Suppose we also use the allele information for every individual in the sample. The variable Z_i indicates which pair of alleles was observed; $Z = (j, k)$ means that for the individual the alleles j and k were observed. Then, the corresponding probabilities equal

$$\begin{aligned}
 P(Z = (j, j)|X, F) &= \frac{Fqa_j + (1-F)q^2a_j^2}{Fq + (1-F)q^2} = \frac{Fa_j + (1-F)qa_j^2}{F + (1-F)q} \\
 P(Z = (j, k)|X, F) &= \frac{(1-F)q^22a_ja_k}{Fq + (1-F)q^2} = \frac{(1-F)q2a_ja_k}{F + (1-F)q}
 \end{aligned}$$

from which the likelihood directly follows as the product over the terms for all individuals.

Neither likelihoods admits an explicit expression for the maximum likelihood estimator of q . The maximum likelihood estimates can be found by numerical maximization of the log likelihood with respect to q and all allele frequencies or the sum $\sum a_j^2$. Numerical maximization for multi-dimensional parameters is often unreliable and it makes sense to estimate the relative allele frequencies beforehand, to insert them into the likelihood and next maximize the likelihood with respect to q .

Estimation of the relative allele frequencies can be done in different ways. Sometimes reliable estimates are available from the literature or previous research. Often this is not the case and the observed data have to be used. A part of the individuals has inbreeding coefficient equal to zero. The observations of these individuals do contain information on the relative allele frequencies, but not on the value of q . This can be easily seen from the likelihood functions; their terms do not depend on q . So, we can split the data. One part with $F = 0$ to estimate the allele frequencies and one part, with $F > 0$, to estimate q by maximizing the likelihood after inserting estimates of the allele frequencies.

After estimating q , confidence intervals have to be constructed. This can be done based on the likelihood ratio statistic: the confidence interval contains all values q_0 for which the null hypothesis $H_0 : q = q_0$ is not rejected; i.e. for which the value of the likelihood ratio statistic is lower than the $(1 - \alpha)$ -quantile of the (asymptotic) distribution of the likelihood ratio statistic. If the data of the individuals are independent and identically distributed it follows from Theorem at the beginning of this paper that the asymptotic distribution equals $(1 + \lambda\nu)$ times a chi-squared distribution with one degree of freedom, for ν equal to an expression computed in Appendix B. In our example, the data of the individuals are independent, but not identically distributed because the inbreeding coefficient may vary among the affected individuals. In the following, we first assume that all inbreeding coefficients equal and we will later extend the results for different inbreeding coefficients. In Appendix B, the computations to find ν are given for the

first maximum likelihood estimator. For the second estimator, the computations are analogous, but slightly more complex because the likelihood has a more complex form; only the results are given in Appendix B.

Simulation study

In order to get more insight into the amount of correction that is needed, we consider a situation that is of practical relevance. We have 80 individuals with inbreeding coefficient equal to $1/32$ and 20 children with inbreeding coefficient equal to zero. The relative allele frequencies will be estimated based on the observed alleles of the 20 individuals (so 40 observations, every individual contributes two alleles) and q will be estimated based on data of the 80 individuals. We assumed there were four disease alleles with relative allele frequencies 0.55, 0.15, 0.1 and 0.2. The frequency q was first taken equal to 0.1 and in a second simulation equal to 0.01. For the first maximum likelihood estimator, the theoretical value of the multiplication factor $1 + (n/m)\nu$ equals 1.7 and for the second estimator 1.045. So the correction for the second maximum likelihood estimator is small and ignoring the fact that the relative allele frequencies were estimated beforehand does hardly affect the level of the test and, thus, the level of the confidence interval. Ignoring the fact that the relative allele frequencies were estimated yields confidence intervals of level 0.87 and 0.94 for the two estimation methods, instead of 0.95.

We sampled 1000 times 100 ($= 80 + 20$) individuals from the population as described above, computed the estimates and the value of the likelihood ratio statistic (under the null hypothesis). The parameter ν in the correction factor depends on the unknown parameters in the model. For every sample, we estimated ν by inserting the estimates of these parameters. We plotted QQ-plots of these likelihood ratio values against the quantiles of the chi-squared distribution with one degree of freedom. Next, we plotted QQ-plots for the values of the likelihood ratio test-statistic divided by its estimated correction term against the quantiles of the chi-squared one distribution. The plots are shown in Figure 1. From the plots, it can be clearly seen that the likelihood ratio statistic for the first maximum likelihood estimator has a thicker tail than the chi-square distribution with one degree of freedom. Correction is needed. For the second estimator, no clear difference between the distribution of the likelihood ratio statistic and the chi-square distribution is visible.

Next we lowered the value of q to 0.01. Then the theoretical correction terms equal 1.14 and 1.028 for the two methods and the levels of the uncorrected confidence interval have level 0.93 and 0.95, instead of

Figure 1 QQ-plots of the two maximum likelihood methods (first method: upper row, second method: lower row), uncorrected (left) and corrected (right). The line is $y = x$

0.95. Again we plotted QQ-plots as described above. This time the corrections are hardly visible in the QQ-plots; the values follow more or less the line $y = x$ (QQ-plots not shown).

In Appendix B, we computed the theoretical value of the correction term $1 + (n/m)v$. Based on this formula, it can easily be seen in which situations correction is needed or not.

If we take m much smaller than n the correction term is much bigger than one for both maximum likelihood estimators and the deviance from the chi-squared distribution is clearly visible from the QQ-plots; the tail of the distribution of the likelihood ratio statistic is much thicker than the tail of the chi-squared one distribution. However, in practice, if m , the number of individuals with zero-inbreeding-coefficient is small, or the number of different alleles is large (which is often the case), one would choose to use all data, including the data of the individuals with a positive inbreeding coefficient, to estimate the allele frequencies. Furthermore, in practice, not all individuals have equal inbreeding coefficients. Then, no theory concerning the asymptotic distribution of the likelihood ratio statistic is known. We therefore performed a simulation study. For different populations (different distributions of F), we sampled 100 individuals from the population, estimated q and computed the likelihood ratio statistic under the null hypothesis. We repeated this 1000 times. QQ-plots of the likelihood ratio static values against the quantiles of the chi-squared one distribution showed that the likelihood ratio statistic follows this distribution quite well. We considered only a few combinations of parameter values. In case correction is needed, we cannot use the correction term as in the previous simulation study, because now the inbreeding coefficient varies among the individuals. Instead, we propose to compute the correction term per individual and next take the mean of all correction terms found.

More details on the estimation method and the application to real data are given in Jonker et al. [13] and Teeuw et al. [14].

4 Discussion

In this paper, we computed the asymptotic distribution of the likelihood ratio statistic for testing a finite dimensional parameter in which the nuisance parameter was replaced by an estimator based on an independent dataset. We showed that ignoring the fact that the nuisance parameter was estimated yields a test with an increased level and a confidence interval with a coverage that is too low. If the dataset used for estimating the nuisance parameter is much bigger than the dataset available for estimating the parameter of interest, the asymptotic distribution is close to the chi-squared distribution which holds if the nuisance parameter is indeed known.

If for a statistical model, it is too difficult to compute the asymptotic distribution explicitly and the asymptotic chi-squared distribution is used for testing or the construction of a confidence interval, one should keep in mind that the true level of the test is (slightly) higher and the coverage of the interval (slightly) lower than supposed, depending on the sample sizes for estimating the parameters and the statistical efficiency.

Acknowledgments: We are very grateful to two reviewers for their constructive comments.

Appendix A: computations for Example 3

In this appendix, we check all assumptions of the theorem for the situation described in Example 2, the simplified Cox model. The aim is to test the hypothesis $H_0 : \beta = \beta_0$ versus $H_1 : \beta \neq \beta_0$ with the likelihood ratio test. To simplify notation, we define $T^* := \min\{T, \tau\}$. Then $S_0(T^*) = S_0(T)^\Delta S_0(\tau)^{1-\Delta}$.

The likelihood for the sample $(T_1^*, \Delta_1), \dots, (T_n^*, \Delta_n) = (t_1^*, \delta_1), \dots, (t_n^*, \delta_n)$ is given by

$$\begin{aligned}
\prod_{i=1}^n p_{\beta, S_0}(t_i^*, \delta_i) &= \prod_{i=1}^n \left\{ \frac{d}{dt} F_{\beta, S_0}(t) \Big|_{t=t_i} \right\}^{\delta_i} \left\{ S_0(\tau)^{\exp \beta} \right\}^{1-\delta_i} \\
&= \prod_{i=1}^n \left\{ -\frac{d}{dt} S_0(t)^{\exp \beta} \Big|_{t=t_i} \right\}^{\delta_i} \left\{ S_0(\tau)^{\exp \beta} \right\}^{1-\delta_i} \\
&= \prod_{i=1}^n \left\{ -\exp(\beta) S_0(t_i)^{\exp \beta - 1} \frac{d}{dt} S_0(t) \Big|_{t=t_i} \right\}^{\delta_i} \left\{ S_0(\tau)^{\exp \beta} \right\}^{1-\delta_i}
\end{aligned}$$

This gives a log likelihood, a first and second derivative, and a Fisher information (all in (β_0, S_0)):

$$\begin{aligned}
l_{\beta_0, S_0}(t^*, \delta) &= \delta \left\{ \beta_0 + (\exp(\beta_0) - 1) \log S_0(t) + \log \left(-\frac{d}{dt} S_0(t_i) \right) \right\} \\
&\quad + (1 - \delta) \exp(\beta_0) \log S_0(\tau) \\
\dot{l}_{\beta_0, S_0}(t^*, \delta) &= \delta(1 + \exp(\beta_0) \log S_0(t)) + (1 - \delta) \exp(\beta_0) \log S_0(\tau) \\
\ddot{l}_{\beta_0, S_0}(t^*, \delta) &= \delta \exp(\beta_0) \log S_0(t) + (1 - \delta) \exp(\beta_0) \log S_0(\tau) \\
&= \exp(\beta_0) \log S_0(T^*) \\
I_{\beta_0, S_0} &= -P_{\beta_0, S_0} \ddot{l}_{\beta_0, S_0}(T^*, \Delta) \\
&= -(\exp(\beta_0) \log(S_0(\tau)) S_0(\tau) + \int_0^\tau \exp(\beta_0) \log S_0(t) dS_0(t)^{\exp \beta_0} \\
&= -\int_0^{\tau-} \exp(\beta_0) S_0(t)^{\exp \beta_0 - 1} \frac{d}{dt} S_0(t) dt \\
&= -S_0(t)^{\exp \beta_0} \Big|_0^\tau = 1 - S_0(\tau)^{\exp \beta_0}.
\end{aligned}$$

The Fisher information was obtained by partial integration. The estimator $\hat{\beta}$ is the value that solves the equation

$$\sum \dot{l}_{\beta_0, \tilde{S}_0}(T_i^*, \Delta_i) = \sum_{i=1}^n \Delta_i \left\{ 1 + \exp(\hat{\beta}) \log \tilde{S}_0(T_i) \right\} + (1 - \Delta_i) \exp(\hat{\beta}) \log S_0(\tau) = 0$$

with $\tilde{S}_0$ the estimator of S_0 that is based on the sample $Y_1, \dots, Y_m$. This yields

$$\hat{\beta} = \log \frac{-\sum_{i=1}^n \Delta_i}{\sum_{i=1}^n \{ \Delta_i \log \tilde{S}_0(T_i) + (1 - \Delta_i) \log \tilde{S}_0(\tau) \}} = \log \frac{-\sum_{i=1}^n \Delta_i}{\sum_{i=1}^n \log \tilde{S}_0(T_i)}.$$

Before checking the assumptions, we first consider the asymptotic behavior of the sequence $\sqrt{n}(\mathbb{P}_n \log \tilde{S}_0(T^*) - P_{\beta_0, S_0} \log \tilde{S}_0(T^*))$. For every m large, the function $\log \tilde{S}_0(t^*)$ belongs (almost surely) to the set of bounded monotone non-increasing functions at the interval $[0, \tau]$ and constant at $[\tau, \infty)$. This set is a Donsker class. Moreover,

$$P_{\beta_0, S_0} (\log \tilde{S}_0(t^*) - \log S_0(t^*))^2 \leq \| \log \tilde{S}_0 - \log S_0 \|_{[0, \tau]}^2 \rightarrow 0$$

in probability. From lemma 19.24 of van der Vaart [2], we conclude that

$$\begin{aligned}
&\sqrt{n}(\mathbb{P}_n \log \tilde{S}_0(T^*) - P_{\beta_0, S_0} \log \tilde{S}_0(T^*)) \\
&= \sqrt{n}(\mathbb{P}_n \log S_0(T^*) - P_{\beta_0, S_0} \log S_0(T^*)) + o_p(1) \rightsquigarrow \mathbb{G}
\end{aligned} \tag{6}$$

for $\mathbb{G}$ a Gaussian process.

We now consider the assumptions one by one. Assumption 1 holds; $\dot{l}_{\beta_0, S_0}$ and $\ddot{l}_{\beta_0, S_0}$ do exist. To check assumption 2, we have to show that $\sqrt{n}(\hat{\beta} - \beta_0)$ is bounded in probability. By the Delta method, it is sufficient to show that the sequence $\sqrt{n}(-\exp(-\hat{\beta}) + \exp(-\beta_0))$ is bounded in probability. Since,

$$P_{\beta_0, S_0} \log S_0(T^*) = P_{\beta_0, S_0} \exp(-\beta_0) \ddot{l}_{\beta_0, S_0}(T^*, \Delta) = -\exp(-\beta_0)(1 - S_0(\tau)^{\exp \beta_0}).$$

we can write

$$\exp(-\beta_0) = \frac{-P_{\beta_0, S_0} \log S_0(T^*)}{1 - S_0(\tau)^{\exp \beta_0}}$$

and thus by using the expression of $\hat{\beta}$ we find

$$\begin{aligned} & \sqrt{n}(-\exp(-\hat{\beta}) + \exp(-\beta_0)) \\ &= \sqrt{n} \left(\frac{n}{\sum_{i=1}^n \Delta_i} \mathbb{P}_n \log \tilde{S}_0(T^*) + \frac{-P_{\beta_0, S_0} \log S_0(T^*)}{1 - S_0(\tau)^{\exp \beta_0}} \right) \\ &= \sqrt{n} \left(\frac{n}{\sum_{i=1}^n \Delta_i} - \left(1 - S_0(\tau)^{\exp \beta_0}\right)^{-1} \right) \mathbb{P}_n \log \tilde{S}_0(T^*) \\ & \quad + \left(1 - S_0(\tau)^{\exp \beta_0}\right)^{-1} \sqrt{n} (\mathbb{P}_n \log \tilde{S}_0(T^*) - P_{\beta_0, S_0} \log S_0(T^*)). \end{aligned}$$

Note that $E_{\beta_0, S_0} n^{-1} \sum_{i=1}^n \Delta_i = F(\tau) = 1 - S_0(\tau)^{\exp \beta_0}$. Then, the term

$$\sqrt{n} \left(\frac{n}{\sum_{i=1}^n \Delta_i} - \left(1 - S_0(\tau)^{\exp \beta_0}\right)^{-1} \right)$$

is asymptotically normal, and consequently bounded in probability, by the central limit theorem and the Delta method. The term $\mathbb{P}_n \log \tilde{S}_0(T^*)$ is also bounded in probability, since $S_0(\tau)$ is bounded away from zero by assumption. The second term equals (up to the constant $((1 - S_0(\tau)^{\exp \beta_0})^{-1})$

$$\sqrt{n} (\mathbb{P}_n \log \tilde{S}_0(T^*) - P_{\beta_0, S_0} \log \tilde{S}_0(T^*)) + \sqrt{n} P_{\beta_0, S_0} (\log \tilde{S}_0(T^*) - \log S_0(T^*)).$$

The first term converges in distribution to a Gaussian process, as was seen before, and is therefore bounded in probability. The second term equals $g(\sqrt{n}(\log \tilde{S}_0(T^*) - \log S_0(T^*)))$ with g the function defined as $g(z) = \int z(t) dP_{\beta_0, S_0}(t)$ that maps $D[0, \tau]$ into $\mathbb{R}$. This function is continuous with respect to the supremum norm. By the continuous mapping theorem, it suffices to show that the sequence $\sqrt{n}(\log \tilde{S}_0(T^*) - \log S_0(T^*))$ converges in distribution. This is true by the Delta method and the fact that $\sqrt{n}(\tilde{S}_0 - S_0)$ converges weakly (Donsker theorem).

For assumption 3 we have to show that $-\mathbb{P}_n \ddot{l}_{\beta_0, \tilde{S}_0}$ converges in probability to I_{β_0, S_0} . In order to see this, write

$$-\mathbb{P}_n \ddot{l}_{\beta_0, \tilde{S}_0}(T^*, \Delta) = -\left(\mathbb{P}_n \ddot{l}_{\beta_0, \tilde{S}_0}(T^*, \Delta) - P_{\beta_0, S_0} \ddot{l}_{\beta_0, \tilde{S}_0}(T^*, \Delta) \right) - P_{\beta_0, S_0} \ddot{l}_{\beta_0, \tilde{S}_0}(T^*, \Delta).$$

Since $\mathbb{P}_n \log \tilde{S}_0(T^*) - P_{\beta_0, S_0} \log \tilde{S}_0(T^*)$ converges in probability to zero by eq. (6), it directly follows that the first term at the right-hand side in the previous display converges in probability to zero. The second term equals $P_{\beta_0, S_0} \exp(\beta_0) \log \tilde{S}_0(T^*)$ and converges in probability to $\exp(\beta_0) P_{\beta_0, S_0} \log S_0(T^*)$ which equals

$$\exp(\beta_0) \exp(-\beta_0)(1 - S_0(\tau)^{\exp \beta_0}) = 1 - S_0(\tau)^{\exp \beta_0}$$

(this was seen when checking assumption 2).

Assumption 4 follows since $\sqrt{n}(\mathbb{P}_n \log \tilde{S}_0(T^*) - P_{\beta_0, S_0} \log \tilde{S}_0(T^*))$ converges in distribution to a Gaussian process.

For assumption 5, we have to show that $\sqrt{m}P_{\beta_0, \tilde{S}_0} \dot{l}_{\beta_0, S_0}(T^*, \Delta)$ converges weakly to a normal distribution.

$$\begin{aligned} \sqrt{m}P_{\beta_0, \tilde{S}_0} \dot{l}_{\beta_0, S_0}(T^*, \Delta) &= \sqrt{m}(P_{\beta_0, \tilde{S}_0} - P_{\beta_0, S_0}) \dot{l}_{\beta_0, S_0}(T^*, \Delta) \\ &= \exp(\beta_0) \log S_0(\tau) \sqrt{m}(\tilde{S}_0^{\exp \beta_0}(\tau) - S_0^{\exp \beta_0}(\tau)) \\ &\quad - \sqrt{m} \int_0^\tau 1 + \exp(\beta_0) \log S_0(t) d(\tilde{S}_0(t)^{\exp \beta_0} - S_0(t)^{\exp \beta_0}) \end{aligned} \quad (7)$$

$$= \sqrt{m} \int_0^\tau (\exp(\beta_0) \log S_0(\tau) - \exp(\beta_0) \log S_0(t) - 1) d(\tilde{S}_0(t)^{\exp \beta_0} - S_0(t)^{\exp \beta_0}).$$

Note that $\sqrt{m}(\tilde{S}_0^{\exp \beta_0} - S_0^{\exp \beta_0}) \approx \sqrt{m}(\tilde{S}_0 - S_0) \exp(\beta_0) S_0^{\exp \beta_0 - 1}$ by a Taylor expansion. The sequence $\sqrt{m}(\tilde{S}_0 - S_0)$ is asymptotically equal to a Gaussian process with mean zero and covariance functions $S_0(t_i \vee t_j) - S_0(t_i)S_0(t_j)$ for $t_i, t_j \in [0, \tau]$. The sequence $\sqrt{m}(\tilde{S}_0 - S_0) \exp(\beta_0) S_0^{\exp \beta_0 - 1}$ is therefore asymptotically equal to a Gaussian process with mean zero and covariance function

$$(S_0(t_i \vee t_j) - S_0(t_i)S_0(t_j)) \exp(2\beta_0) S_0(t_i)^{\exp \beta_0 - 1} S_0(t_j)^{\exp \beta_0 - 1}.$$

for $t_i, t_j \in [0, \tau]$. This Gaussian process is denoted as X_t .

By partial integration, it can be seen that the map g on $D[0, \tau] \rightarrow \mathbb{R}$ defined as $g(z) = \int_0^\tau (\exp(\beta_0) \log S_0(\tau) - \exp(\beta_0) \log S_0(t) - 1) dz(t)$ is continuous with respect to the supremum norm. By use of the continuous mapping theorem, the sequence at the right-hand side in eq. (7) converges weakly to

$$\int_0^\tau (\exp(\beta_0) \log S_0(\tau) - \exp(\beta_0) \log S_0(t) - 1) dX_t. \quad (8)$$

Take M large, $\Delta t = \tau/M$ and define

$$Y^{(M)} = \sum_{k=0}^{M-1} (\exp(\beta_0)(\log S_0(\tau) - \log S_0(t_k)) - 1) \Delta X_k$$

where $\Delta X_k = X_{t_{k+1}} - X_{t_k}$, and $t_k = k\Delta t$. Then, for M growing to infinity, $Y^{(M)}$ converges to the integral in eq. (8). Since $Y^{(M)}$ is a sum of independent Gaussian random variables, it also has a normal distribution with mean zero and variance

$$\begin{aligned} \text{Var } Y^{(M)} &= \sum_{k=0}^{M-1} \text{Var} \{(\exp(\beta_0)(\log S_0(\tau) - \log S_0(t_k)) - 1) \Delta X_k\} \\ &= \sum_{k=0}^{M-1} (\exp(\beta_0)(\log S_0(\tau) - \log S_0(t_k)) - 1)^2 \text{Var} (\Delta X_k) \end{aligned} \quad (9)$$

with

$$\begin{aligned} \text{Var} (\Delta X_k) &= \text{Var} (X_{t_{k+1}} - X_{t_k}) \\ &= \text{Var} X_{t_{k+1}} + \text{Var} X_{t_k} - 2\text{Cov} (X_{t_{k+1}}, X_{t_k}) \end{aligned}$$

with $\text{Var} X_t$ and $\text{Cov}(X_t, X_s)$ as given above. For $t > s$:

$$\begin{aligned}
& \exp(-2\beta_0)(\text{Var}X_t + \text{Var}X_s - 2\text{Cov}(X_t, X_s)) \\
&= (S_0(t) - S_0(s))^2 S_0(t)^{2\exp\beta_0-2} + (S_0(s) - S_0(t))^2 S_0(s)^{2\exp\beta_0-2} \\
&\quad - 2(S_0(t \vee s) - S_0(t)S_0(s))S_0(t)^{\exp\beta_0-1}S_0(s)^{\exp\beta_0-1} \\
&= (1 - S_0(t))S_0(t)^{2\exp\beta_0-1} + (1 - S_0(s))S_0(s)^{2\exp\beta_0-1} \\
&\quad - 2(1 - S_0(s))S_0(t)^{\exp\beta_0}S_0(s)^{\exp\beta_0-1} \\
&= S_0(t)^{2\exp\beta_0-1} - S_0(t)^{2\exp\beta_0} + S_0(s)^{2\exp\beta_0-1} - S_0(s)^{2\exp\beta_0} \\
&\quad - 2S_0(t)^{\exp\beta_0}S_0(s)^{\exp\beta_0-1} + 2S_0(t)^{\exp\beta_0}S_0(s)^{\exp\beta_0} \\
&= S_0(t)^{2\exp\beta_0-1} - S_0(t)^{\exp\beta_0}S_0(s)^{\exp\beta_0-1} - S_0(t)^{2\exp\beta_0} + S_0(t)^{\exp\beta_0}S_0(s)^{\exp\beta_0} \\
&\quad + S_0(s)^{2\exp\beta_0-1} - S_0(t)^{\exp\beta_0}S_0(s)^{\exp\beta_0-1} - S_0(s)^{2\exp\beta_0} + S_0(t)^{\exp\beta_0}S_0(s)^{\exp\beta_0} \\
&= S_0(t)^{\exp\beta_0} \left(S_0(t)^{\exp\beta_0-1} - S_0(s)^{\exp\beta_0-1} \right) - S_0(t)^{\exp\beta_0} \left(S_0(t)^{\exp\beta_0} - S_0(s)^{\exp\beta_0} \right) \\
&\quad - S_0(s)^{\exp\beta_0-1} \left(S_0(t)^{\exp\beta_0} - S_0(s)^{\exp\beta_0} \right) + S_0(s)^{\exp\beta_0} \left(S_0(t)^{\exp\beta_0} - S_0(s)^{\exp\beta_0} \right) \\
&= S_0(t)^{\exp\beta_0} \left(S_0(t)^{\exp\beta_0-1} - S_0(s)^{\exp\beta_0-1} \right) - \left(S_0(t)^{\exp\beta_0} - S_0(s)^{\exp\beta_0} \right)^2 \\
&\quad - S_0(s)^{\exp\beta_0-1} \left(S_0(t)^{\exp\beta_0} - S_0(s)^{\exp\beta_0} \right)
\end{aligned}$$

So,

$$\begin{aligned}
& \text{Var}X_t + \text{Var}X_s - 2\text{Cov}(X_t, X_s) \\
&= \exp(2\beta_0) \left(S_0(t)^{\exp\beta_0} \left(S_0(t)^{\exp\beta_0-1} - S_0(s)^{\exp\beta_0-1} \right) - \left(S_0(t)^{\exp\beta_0} - S_0(s)^{\exp\beta_0} \right)^2 \right. \\
&\quad \left. - S_0(s)^{\exp\beta_0-1} \left(S_0(t)^{\exp\beta_0} - S_0(s)^{\exp\beta_0} \right) \right)
\end{aligned}$$

Filling in this expression into the sum in eq. (9), letting M go to infinity and straightforward computations yield

$$-\int_0^\tau (\exp(\beta_0)(\log S_0(\tau) - \log S_0(t)) - 1)^2 S_0(t)^{2\exp\beta_0-2} \exp(2\beta_0) dS_0(t).$$

So, $\sqrt{m}P_{\beta, \tilde{S}_0} \dot{l}_{\beta_0, S_0}(T\Delta)$ converges weakly to a normal distribution with mean zero and variance equal to the integral in the previous display. This integral can be computed explicitly, but yields a complicated formula which is not interesting on itself. Taking $\beta = 0$ makes the integral easier to compute; this yields $1 - S_0(\tau) + \log^2 S_0(\tau)$.

For assumption 6, we have to show that $\sqrt{n}(P_{\beta_0, S_0} - P_{\beta_0, \tilde{S}_0})(\dot{l}_{\beta_0, \tilde{S}_0} - \dot{l}_{\beta_0, S_0})$ converges in probability to zero.

$$\begin{aligned}
& \sqrt{n}(P_{\beta_0, S_0} - P_{\beta_0, \tilde{S}_0})(\dot{l}_{\beta_0, \tilde{S}_0} - \dot{l}_{\beta_0, S_0}) \\
&= \exp(\beta_0) \sqrt{n}(P_{\beta_0, S_0} - P_{\beta_0, \tilde{S}_0})(\log \tilde{S}_0(T^*) - \log S_0(T^*))
\end{aligned}$$

$$\begin{aligned}
&= -\exp(\beta_0)\sqrt{n}(\log \tilde{S}_0(\tau) - \log S_0(\tau))\left(S_0(\tau)^{\exp \beta_0} - \tilde{S}_0(\tau)^{\exp \beta_0}\right) \\
&\quad - \exp(\beta_0)\sqrt{n} \int_0^\tau (\log \tilde{S}_0(t) - \log S_0(t))d\left(S_0(t)^{\exp \beta_0} - \tilde{S}_0(t)^{\exp \beta_0}\right) \\
&= o_P(1) - \exp(\beta_0)\sqrt{n} \int_0^\tau (\log \tilde{S}_0(t) - \log S_0(t))d\left(S_0(t)^{\exp \beta_0} - \tilde{S}_0(t)^{\exp \beta_0}\right) \\
&= o_P(1) - \exp(\beta_0)\sqrt{n} \int_0^\tau \frac{\tilde{S}_0(t) - S_0(t)}{S_0(t)}d\left(S_0(t)^{\exp \beta_0} - \tilde{S}_0(t)^{\exp \beta_0}\right),
\end{aligned}$$

where the last equality follows from a Taylor expansion of $\log \tilde{S}_0$ around S_0 . The function $(g, h) \mapsto \int g(t)dh(t)$ is a continuous map from the set of functions defined on $[0, \tau] \times [0, \tau]$ which are of bounded variation into $\mathbb{R}$ with respect to the supremum norm. Application of the continuous mapping theorem and the fact that $\sqrt{m}(\tilde{S}_0 - S_0)$ and $\sqrt{m}(\tilde{S}_0^{\exp \beta_0} - S_0^{\exp \beta_0})$ converge in distribution to a Gaussian process yields weak convergence of $m \int (\log \tilde{S}_0(t) - \log S_0(t))d(S_0(t)^{\exp \beta_0} - \tilde{S}_0(t)^{\exp \beta_0})$. Since $m = O(n)$, the last term in the previous display converges in distribution to zero.

Assumption 7 follows directly from the fact that $\frac{d}{d\beta} \ddot{l}_{\beta, S}(t^*, \delta) = \ddot{l}_{\beta, S}(t^*, \delta)$ is continuous and differentiable with respect to β . Then $\frac{1}{n} \sum_{i=1}^n \frac{d}{d\beta} \ddot{l}_{\beta, S}(T_i^*, \Delta_i)$ is bounded for (β, S) is a small neighborhood of (β_0, S_0) .

Appendix B: computations for Example 4

In this appendix, we give the computations for the correction term $1 + (n/m)v$ for the maximum likelihood estimators A as described in Example 4. The computations for estimator B are analogous; only the results are given.

We have to compute $v = I_{q_0, a}^{-1} J_{q_0, a}$ for $I_{q_0, a}^{-1}$ equal to the inverse of the Fisher information for estimating q_0 and $a = \sum_{j=1}^K a_j^2$ under the assumption that $(a_1, \dots, a_K)$ is known and $J_{q_0, a}$ is the asymptotic variance of $\sqrt{m}P_{q_0, \tilde{a}} \dot{l}_{q_0, a}$ with $\dot{l}_{q_0, a}$ the derivative of the log likelihood and $\tilde{a} = \sum_{j=1}^K \tilde{a}_j^2$ the estimated value of a based on the data of individuals with $F = 0$.

Straightforward computations show that

$$I_{q_0, a}^{-1} = \frac{(1-a)(1-F)F^2}{q_0(F + (1-F)q_0)^2(F + (1-F)q_0a)},$$

and

$$\sqrt{m}P_{q_0, \tilde{a}} \dot{l}_{q_0, a} = \frac{(1-F)F}{(F + (1-F)q_0)(F + (1-F)q_0a)} \sqrt{m} \left(\sum_{j=1}^K \tilde{a}_j^2 - \sum_{j=1}^K a_j^2 \right)$$

for $a = \sum_{j=1}^K a_j^2$ and $\tilde{a} = \sum_{j=1}^K \tilde{a}_j^2$. We denote the asymptotic variance of $\sqrt{m}(\tilde{a} - a) = \sqrt{m}(\sum_{j=1}^K \tilde{a}_j^2 - \sum_{j=1}^K a_j^2)$ by σ^2 , then the asymptotic variance of $\sqrt{m}P_{q_0, \tilde{a}} \dot{l}_{q_0, a}$ equals

$$J_{q_0, a} = \frac{(1-F)^2 F^2 \sigma^2}{(F + (1-F)q_0)^2 (F + (1-F)q_0a)^2}$$

and, thus,

$$J_{q_0, a} I_{q_0, a}^{-1} = \frac{q_0(1-F)\sigma^2}{(F + (1-F)q_0a)(1-a)}$$

The value of σ^2 can be found with help of the Delta method (see for instance van der Vaart [2]). The second dataset consists of m observations. Define $Z_1, \dots, Z_K$ as the number of observed alleles $1, 2, \dots, K$. Then $\sum Z_i = 2m$ because every individual contributes two alleles. The allele frequency a_i is estimated by $Z_i/(2m)$. Known is that $\sqrt{m}((\tilde{a}_1, \dots, \tilde{a}_{K-1})^T - (a_1, \dots, a_{K-1})^T)$ is asymptotically normal with mean zero and covariance matrix Σ with diagonal $(a_1(1-a_1)/2, \dots, a_{K-1}(1-a_{K-1})/2)$ and $\Sigma_{(ij)}$ for $i \neq j$ equal to $-a_i a_j / 2$ (note that $a_K = 1 - \sum_{j=1}^{K-1} a_j$).

Define the function $g: \{x \in [0, 1]^{K-1} \mid \sum_{j=1}^{K-1} x_j < 1\} \rightarrow \mathbb{R}$ as $g(x) = \sum_{j=1}^K x_j^2$ with $x_K = 1 - \sum_{i=1}^{K-1} x_i$. Then $\sqrt{m}(\sum_{j=1}^K \tilde{a}_j^2 - \sum_{j=1}^K a_j^2) = \sqrt{m}(g(\tilde{a}_1, \dots, \tilde{a}_{K-1}) - g(a_1, \dots, a_{K-1}))$. The asymptotic distribution of $\sqrt{m}(\sum_{j=1}^K \tilde{a}_j^2 - \sum_{j=1}^K a_j^2)$ now follows from the Delta method; it equals a normal distribution with mean zero and variance equal to

$$\sigma^2 = g'(a_1, \dots, a_{K-1})^T \Sigma g'(a_1, \dots, a_{K-1})$$

$$= \sum_{i=1}^{K-1} 2(a_i - a_K)^2 a_i(1-a_i) - \sum_{i=1}^{K-1} \sum_{j=1, j>i}^{K-1} 4a_i a_j (a_i - a_K)(a_j - a_K),$$

where $g'(a_1, \dots, a_{K-1}) = (2(a_1 - a_K), \dots, 2(a_{K-1} - a_K))^T$ is the derivative of g in $(a_1, \dots, a_{K-1})$.

The computations for v for maximum likelihood estimator B are very similar to those for estimator A, but the notation is slightly more complex. We only give the results.

Define the indicator function $\Gamma_{j,k}$ as 1 if $Z = (j, k)$ and otherwise $\Gamma_{j,k}$ equals 0. Then the likelihood for one individual can be written as

$$\prod_{j=1}^K \left(\frac{F a_j + (1-F) q a_j^2}{F + (1-F) q} \right)^{\Delta \Gamma_{jj}} \prod_{j,k=1, j \neq k}^K \left(\frac{(1-F) q 2 a_j a_k}{F + (1-F) q} \right)^{(1-\Delta) \Gamma_{jk}}.$$

Straightforward computations show that

$$I_{q_0, a}^{-1} J_{q_0, a} = \frac{q_0(1-F)\sigma^2}{\sum_{j=1}^K \frac{a_j(1-a_j)}{F + (1-F)q_0 a_j}}.$$

with σ^2 the asymptotic variance of $\sum_{j=1}^K \frac{\tilde{a}_j \sqrt{m}(a_j - \tilde{a}_j)}{F + (1-F)q_0 a_j}$. An expression for σ^2 can be determined with help of Slutsky's lemma and the Delta method (see for instance van der Vaart [2]).

References

1. Cox DR, Hinkley DV. Theoretical statistics. London: Chapman and Hall, 1974.
2. van der Vaart AW. Asymptotic statistics. Cambridge: Cambridge University Press, 1998.
3. Wilks SS. The large-sample distribution of the likelihood ratio for testing composite hypotheses. Ann Math Stat 1938;19:60–2.
4. Thomas DR, Grunkemeier GL. Confidence interval estimation of survival probabilities for censored data. J Am Stat Assoc 1975;70:865–71.
5. Murphy SA, van der Vaart AW. Semiparametric likelihood ratio inference. Ann Stat 1997;25:1471–509.
6. Banerjee M, Wellner JA. Likelihood ratio tests for monotone functions. Ann Stat 2001;29:1699–731.
7. Banerjee M. Likelihood based inference for monotone response models. Ann Stat 2007;35:931–56.
8. Kosorok MR. Introduction to empirical processes and semiparametric inference. Series in statistics. New York: Springer, 2008.
9. Owen AB. Empirical likelihood ratio confidence intervals for a single functional. Biometrika 1988;75:237–49.
10. Owen AB. Empirical likelihood. Monographs on statistics and applied probability 92. New York: Chapman and Hall, 2001.

11. Qin J, Lawless J. Empirical likelihood and general estimating equations. *Ann Stat* 1994;22:300–25.
12. Murphy SA, van der Vaart AW. On profile likelihood. *J Am Stat Assoc* 2000;95:449–85.
13. Jonker MA, Teeuw M, ten Kate LP, Cornel MC. Estimating total pathogenic allele frequency of recessive disorders when parents are consanguineous 1: Methodological aspects. Unpublished manuscript, 2014.
14. Teeuw ME, Kelmami W, Jonker MA, Sefiani A, Laarabi F-Z, Hama I, et al. Estimating total pathogenic allele frequency of recessive disorders when parents are consanguineous 2: Case in Tunisia and FMF in Morocco. Unpublished manuscript, 2014.
15. Jonker MA, Bhulai S, Ligthart RSL, Posthuma D, Boomsma DI, van der Vaart AW. Gamma frailty model for linkage analysis with application to interval censored migraine data. *Biostatistics* 2009;10:187–200.
16. Qin G, Jing BY. Censored partial linear models and empirical likelihood. *J Multivariate Anal* 2001;78:37–61.
17. Qin G, Jing BY. Empirical likelihood for censored linear regression. *Scand J Stat* 2001;28:661–73.
18. Li G, Wang QH. Empirical likelihood regression analysis for right censored data. *Stat Sin* 2003;13:51–68.
19. Wang QH, Jing BY. Empirical likelihood for a class of functionals of survival distribution with censored data. *Ann Inst Stat Math* 2001;53:517–27.
20. Hjort NL, McKeague IW, Keilegom IV. Extending the scope of empirical likelihood. *Ann Stat* 2009;37:1079–111.
21. Huang J, Wellner JA. Asymptotic normality of the NPMLE of linear functionals for interval censored data, case 1. *Stat Neerlandica* 1995;49:153–63.
22. Groeneboom P, Wellner JA. Information bounds and nonparametric maximum likelihood, DMV seminar, vol. 19. Basel: Birkhäuser Verlag, 1992.
23. ten Kate LP, Teeuw M, Henneman L, Cornel MC. Autosomal recessive disease in children of consanguineous parents: inferences from the proportion of compound heterozygotes. *J Community Genet* 2010;1:37–40.