

# Confidence intervals by the profile likelihood method, with applications in veterinary epidemiology

H. Stryhn<sup>1</sup>, J. Christensen\*<sup>2</sup>

<sup>1</sup> Atlantic Veterinary College, University of PEI, Charlottetown PE C1A 4P3, Canada

<sup>2</sup> CFIA, 93 Mount Edward Road, Charlottetown PE C1A 5T1, Canada

## Summary

The commonly used Wald-type confidence intervals are based on the approximately normal distribution of parameter estimates valid in many “nice” estimation settings. However, one may come across “difficult” situations where the procedure works poorly or breaks down. In this paper, we review an alternative method based on the likelihood function, illustrating by 2 examples its use and the computations involved.

## Introduction

The standard procedure for computing a confidence interval (CI) for a parameter in a statistical model is by the formula: estimate  $\pm$  percentile  $\times$  SE(estimate), where SE is the standard error. The percentile is selected according to a desired confidence level of the interval and a reference distribution (a t-distribution for regression coefficients in a linear model, otherwise a standard normal distribution). This procedure is commonly referred to as a Wald-type CI. It may work poorly if the distribution of the parameter estimate is markedly skewed or if the (estimated) standard error is useless. A common example of the latter occurs when a regression coefficient in logistic regression corresponds to comparison of 2 categories of which one has either all outcomes negative or all outcomes positive. Examples of the former are variance-covariance parameters in mixed models <sup>1</sup> or estimates close to a boundary in their range, e.g. for estimation of the correlation (covariance) between 2 diagnostic tests <sup>2</sup> or maximum-likelihood estimation of test characteristics in absence of a gold standard <sup>3</sup>.

Besides the ability to recognise when Wald-type CIs fail, the applied researcher is in need of an alternative method. The profile likelihood (or likelihood or likelihood-ratio) method <sup>4</sup> is applicable to all likelihood-based statistical analyses and is generally less sensitive to the difficulties described above.

## Profile likelihood method

The idea is to invert a likelihood-ratio test to obtain a CI for the parameter in question. Consider a statistical model with parameters  $\theta$  and  $\delta$  where  $\theta$  is the parameter of interest and  $\delta$  is the (vector of) additional parameter(s) in the model. Denote by  $L(\theta, \delta)$  the likelihood function, and by  $(\theta^*, \delta^*)$  the maximum likelihood (ML) estimates. The likelihood ratio test statistic ( $G^2$ ) of the hypothesis  $H_0: \theta = \theta_0$  (where  $\theta_0$  is a fixed value) equals the drop in  $2\ln L$  between the “full” model and the reduced model with  $\theta$  fixed at  $\theta_0$ , i.e.  $G^2 = 2(\ln L(\theta^*, \delta^*) - \ln L(\theta_0, \delta_0^*))$ , where  $\delta_0^*$  is the ML estimate of the reduced model. Alternatively, we may express the test statistic in terms of the *profile likelihood function*  $L_1$  for the parameter  $\theta$  which is obtained from the usual likelihood function by maximizing over the parameter  $\delta$ , i.e.  $L_1(\theta) = \max_{\delta} L(\theta, \delta)$ . Then we have  $G^2 = 2(\ln L_1(\theta^*) - \ln L_1(\theta_0))$ . A 95% CI for  $\theta$  consists of those values of

$\theta_0$  for which the test is non-significant at significance level 0.05; this is the case when  $G^2$  does not exceed 3.84 (95%-percentile of the  $\chi^2(1)$  distribution). Thus, the CI consists of the  $\theta_0$ -values for which  $\ln L_1(\theta_0) \geq \ln L_1(\theta^*) - 3.84/2 = \ln L_1(\theta^*) - 1.92$ . For a CI with coverage  $(1-\alpha)*100\%$ , use instead the  $(1-\alpha)$ -percentile of the  $\chi^2(1)$  distribution. Figure 1 depicts the profile likelihood CI graphically for the data of Example 1.

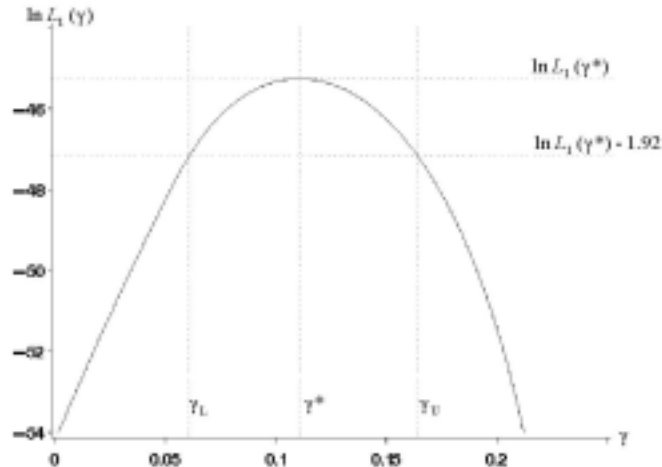


Figure 1. Profile likelihood function and 95% CI:  $(\gamma_L, \gamma_U)$  for the data in Example 1.

### Computation of a profile likelihood CI

For standard analyses, such as logistic regression, it is available as an option in some statistical packages (e.g., SAS (“logistic” and “genmod” procedures) and Stata (“logprof” command)). Also, some packages offer profile likelihood CIs in more general frameworks (e.g., SAS (“model” procedure) and R (“plkhci” function)). The latter is based on a general algorithm<sup>4</sup> which can be implemented in any programming environment with built-in routines for maximization. It is also possible to extend the EM algorithm to provide profile likelihood intervals<sup>5</sup>. However, we outline a simpler approach which in many situations will be relatively easy to use or implement.

For simplicity, we consider only the lower bound of the CI (the upper bound is similar) and assume the profile likelihood function to be an increasing function to the left of its maximum. As a start, compute the ML estimates  $(\theta^*, \delta^*)$  and the corresponding log-likelihood value. Then proceed by the following steps:

1. Compute a “reasonable” lower bound  $\theta'$  for the lower confidence limit (e.g.,  $\theta^* - 5SE(\theta^*)$ , or 0.0001 if  $\theta$ -values are restricted to be  $>0$ ),
2. Define a grid of values ranging from  $\theta'$  to  $\theta^*$  (e.g., 100 equidistant points),
3. For each grid value  $\theta_i$ , compute the profile log-likelihood value  $\ln L_1(\theta_i)$  by maximizing the  $\ln L(\theta_i, \delta)$  over  $\delta$ -values (a standard analysis allowing  $\theta$  to be fixed at  $\theta_i$  may apply),
4. Take as the lower bound ( $\theta_L$ ) of the 95% CI the smallest  $\theta_i$ -value for which it holds that  $\ln L_1(\theta_i) \geq \ln L_1(\theta^*) - 1.92$ ,
5. If necessary, refine or extend the grid of values around  $\theta_L$  to obtain greater accuracy.

For repeated computations one may replace the crude search over a grid of values by a systematic search procedure (e.g. bisection of the interval from  $\theta'$  to  $\theta^*$ ).

### Example 1: Covariance (correlation) between two diagnostic tests

Results of diagnostic tests may be correlated (dependent) even for true positive samples, intuitively so if they target related quantities or operate in similar ways, and the direct way of expressing such (conditional) dependence between tests with dichotomous outcomes is by the covariance between tests results in a population of true positives<sup>2</sup>. We focus here on the dependence among positives, the sensitivity covariance; correlation among true negatives can be treated similarly. From the 2x2-table of probabilities of test results ( $p_{11}, p_{10}, p_{01}, p_{00}$ ), where e.g.  $p_{10}$  is the probability of testing positive on the first test and negative on the second test, the covariance is computed as  $\gamma = p_{11} - Se_1 Se_2$ ; the sensitivities are  $Se_1 = p_{11} + p_{10}$  and  $Se_2 = p_{11} + p_{01}$ . These formulae apply also to observed proportions from a table ( $n_{11}, n_{10}, n_{01}, n_{00}$ ) of counts and give in this case the ML parameter estimates. A profile likelihood CI for  $\gamma$  can be computed by the procedure outlined above; the log-likelihood function is given by:

$$\ln L = n_{11} \ln[Se_1 Se_2 + \gamma] + n_{10} \ln[Se_1(1 - Se_2) - \gamma] + n_{01} \ln[(1 - Se_1)Se_2 - \gamma] + n_{00} \ln[(1 - Se_1)(1 - Se_2) + \gamma]$$

Due to the inherent restrictions on the parameters (cell probabilities are between 0 and 1), the maximization of  $\ln L$  for fixed  $\gamma$  involves non-linear constraints<sup>2</sup>. As a data example, from observed counts (26,0,11,9) we compute  $Se_1 = 26/46 = 0.565$ ,  $Se_2 = 37/46 = 0.804$ , and  $\gamma = 26/46 - 0.565 * 0.804 = 0.111$ . Fig. 1 shows the profile log-likelihood function with its maximal value of  $-45.25$ , and the 95% CI for  $\gamma$ : (0.061, 0.164) comprising the values of  $\gamma$  where the function exceeds  $-45.25 - 1.92 = -47.17$ .

### Example 2: EC<sub>50</sub> (effective concentration; to obtain reaction in 50% of a population) in a dose-response relation with natural responsiveness

In a study of the sensitivity of sea lice from farms of Atlantic salmon to chemotherapeutants, a probit regression model with natural responsiveness was used for the number of affected (moribond or dead) sea lice in bioassays involving different doses of the agent. The modelling equation for the probability of sea lice being affected was  $\text{probit}(p) = p_0 + (1 - p_0)(\alpha + \beta \text{dose})$ , and the EC<sub>50</sub> is computed as  $-\alpha/\beta$ . Inference about EC<sub>50</sub> is potentially difficult, and several methods exist for computing a CI<sup>6</sup>. The rewriting  $\alpha + \beta \text{dose} = \beta(\text{dose} - EC_{50})$  shows that the profile likelihood equals the usual likelihood for a model with no intercept and the dose centered by the EC<sub>50</sub>. Therefore, step 3 in the algorithm above could be carried out by a standard routine for probit regression. As a numerical illustration, for the control group and doses given as the natural logarithms of (1,3,10,30,100,300), the number of affected lice were: (2/24, 1/22, 2/21, 4/21, 1/17, 9/18, 27/27). The estimated natural responsiveness was  $p_0^* = 0.095$  (SE=0.029), and the intercept and slope were  $\alpha^* = -29.56$  and  $\beta^* = 6.391$ , respectively, with huge and useless Ses. Further, EC<sub>50</sub>=4.625 on logarithmic scale and  $e^{4.625} = 102$  on original scale. The 95% profile likelihood CI (on original scale) was (79.7, 133.7).

### References

1. Longford N (1999), *Multilevel Modelling Newsletter* **11**:10-13.
2. Gardner IA, Stryhn H, Lind P, Collins MT (2000), *Prev. Vet. Med.* **45**:107-122.
3. Hoorfar J, Nielsen EM, Stryhn H, Andersen S (1999), *J. Microb. Meth.* **38**:101-106.
4. Venzon DJ, Moolgavkar AH (1988), *Appl. Statist.* **37**:87-94.
5. Dong KK, Taylor JMG (1995), *J. Amer. Statist. Assoc.* **90**:708-716.
6. Williams DA (1986), *Biometrics* **42**:641-645.