Computing exact 1-sided limits for relative risk

Chris Lloyd, Melbourne Business School
Computing exact 1-sided limits for relative risk

Chris J. Lloyd

Abstract. A simple construction for optimal and exact upper confidence limits subject to an ordering constraint was given by Buehler (1957) and extended by Lloyd and Kabaila (2003). Computing these limits requires exact calculation of certain tail probabilities and optimisation of potentially erratic functions of the nuisance parameter. Naive implementation results in limits that are computationally unreliable and highly burdensome. For two dimensional problems based on likelihood based statistics it is shown how to avoid the computational burden. Problems with the optimisation can be mitigated by applying two quite different computational strategies which tend to break down under different conditions and taking the larger of the two computed values. Exact lower limits can be computed using analogous methods. This paper specifically applies the ideas to limits for the relative risk in a clinical trial, but the method extends simply to arbitrary measures of treatment effect.

Keywords. tight upper limits; Buehler bounds

1 Introduction

One sided confidence limits are part of the standard technology for summarising the results of clinical trial. The most simple clinical trials involve a case group 1, a control group 0 and a so-called endpoint such as survival. The effect of the treatment can be measured in many ways. A general class of measures is $\theta = g(\pi_1) - g(\pi_0)$ when $\pi_i$ is the probability of survival for group $i$ and $g( )$ is an increasing function. The data will be the number the number of survivals $Y_1$ from the $n_1$ given the treatment and

---

1 C.Lloyd
Melbourne Business School, Carlton, 3053, AUSTRALIA
email:c.lloyd@mbs.edu
the number $Y_0$ from the $n_0$ controls. We want an upper bound $u(Y_1, Y_0)$ for $\theta$ with pre-specified confidence.

Let $Y$ denote any discrete random vector taking values in a known countable set $\mathcal{Y}$ with distribution determined by a vector parameter $\pi$. Let $\theta(\pi)$ be a scalar parameter of interest taking values in $\Theta$. Let $C(\pi)$ be the probability that $\theta$ is less than or equal to $S(Y)$ as a function of the parameter $\pi$. A $1 - \alpha$ upper confidence limit for $\theta$ is a statistic $S(Y)$ with the property that $\inf_\pi C(\pi) = 1 - \alpha$. We would like $S(Y)$ to be as small as possible subject to this constraint being satisfied. One natural way to investigate this is to look at the partially minimised coverage as a function of $\theta$ i.e. the function

$$C(\theta') = \inf\{C(\pi) : \theta(\pi) = \theta'\} \quad (1)$$

A typical example of this function is in the left panel of Figure 1. Note that $C(\theta)$ has discontinuous local minima whenever $\theta$ equals an observable value $S(y_j)$ of $S$. For standard approximate upper limits, $C(\theta)$ often falls well below $1 - \alpha$ even for quite large sample sizes. This could apparently be corrected if we increased/decreased the values of $S(y_j)$ at points where $C(\theta)$ is smaller/larger than $1 - \alpha$.

This idea was formalised by Buehler (1957). Starting with an approximate upper limit $S(Y)$ with observed value $s = S(y)$, one solves an optimisation problem to compute an adjusted statistic $u^*_S(s)$. Under very mild regularity conditions on the probability model, Lloyd and Kabaila (2003) proved that $u^*_S(s)$ (Property 1) is a non-decreasing function of $s$, and (Property 2) has coverage function that never falls below $1 - \alpha$. Most importantly, they showed that amongst statistics satisfying Property 1 and 2, $u^*_S(s)$ is as small as possible for all possible $s$. Simply stated, amongst statistics that order the sample space in the same way as $S(Y)$, $u^*_S(s)$ is the smallest possible upper confidence limit. We will henceforth call these tight upper limits. An exact formula is given later in equation (2).

There are various computational issues that arise in computing tight upper limits
and lower limits. In this paper we show how to compute them efficiently when $S(Y)$ has certain natural monotonicity properties. These properties are typically possessed by likelihood based upper limits but are often not possessed by some other standard methods such as Wald-type upper limits. We also utilise two alternative representations of $u^*_S(s)$ which lead to different algorithms that are likely to break down under different conditions. We develop the algorithm specifically for $\theta = \log(\pi_1/\pi_0)$ known as the logged relative risk, but the extension to arbitrary measures of treatment effect $\theta = g(\pi_1) - g(\pi_0)$ seems straightforward.

## 2 Upper limits for relative risk

The data comprise $y_1$ responses from $n_1$ independent individuals with treatment applied and $y_0$ responses from $n_0$ independent individuals with treatment not applied. The parameter of interest is the relative risk $\psi = \pi_1/\pi_0$ and without loss of generality we take the nuisance parameter to be $\lambda = \pi_0\pi_1$. Then the log-likelihood

$$
\ell(\psi, \lambda; y_0, y_1) = 0.5(y_1 + y_0) \log \psi + 0.5(y_1 + y_0) \log \lambda + (n_1 - y_1) \log \left(1 - \sqrt{\lambda \psi}\right) + (n_0 - y_0) \log \left(1 - \sqrt{\lambda/\psi}\right)
$$

where we note that for fixed $\lambda$, $\psi \in \Psi(\lambda) := [\lambda, \lambda^{-1}]$ and for fixed $\psi$, $\lambda \in \Lambda(\psi) := [0, \min(\psi, \psi^{-1})]$. Henceforth we take the interest parameter to be $\theta = \log \psi$ and use subscript $\theta$ to indicate maximum likelihood (ML) estimation with $\theta$ known, which involves solving a quadratic, see Miettinen and Nurminen (1985).

There are three basic test statistics, each with an asymptotic standard normal distribution. The ML estimator of $\theta$ has asymptotic variance $\sigma^2(\pi_0, \pi_1) = (1 - \pi_0)/(n_0\pi_0) + (1 - \pi_1)/(n_1\pi_1)$ giving rise to the standard error $\hat{\sigma} = \sigma(\hat{\pi}_0, \hat{\pi}_1)$ as well as the restricted ML estimator $\hat{\sigma}_\theta = \sigma(\hat{\pi}_{0\theta}, \hat{\pi}_{1\theta})$. Then we have two Wald-type statistics

$$
W_1(\theta) = \frac{\hat{\theta} - \theta}{\hat{\sigma}}, \quad W_2(\theta) = \frac{\hat{\theta} - \theta}{\hat{\sigma}_\theta}.
$$
It is common to add a small number (for instance $1/2$) to all the counts in order to avoid perverse behaviour at the extremes of the sample space. The third standard statistic is the likelihood root

$$r(\theta) = \text{sign} (\hat{\theta} - \theta) \left[ 2 \left\{ \ell (\hat{\theta}, \hat{\lambda}) - \ell (\theta, \hat{\lambda}) \right\} \right]^{1/2}.$$

For reasons that will emerge later, this is the preferred statistic for generating a tight upper limit. Note that when $y_0 = 0$, the ML estimate of $\theta$ and the upper limit $u_L$ are infinite. Any upper limit that is not infinite when $y_0 = 0$ will have zero coverage for $\lambda$ sufficiently close to zero.

Each of the three test statistics generate an approximate upper confidence limit based on an approximate standard normal distribution. This normal approximation holds for large $(n_0, n_1)$ provided the parameters $\pi_1, \pi_0$ are not on the boundary, so we will not expect accurate results for extreme data sets, especially those close to the "corner" data sets $(0, 0)$ and $(n_0, n_1)$. The upper limit $u_{W1}(y_0, y_1)$ for $\theta$ based on $W_1$ has the explicit form $u_{W1} = \hat{\theta} + q_\alpha \hat{\sigma}$ while the upper limit $u_{W2}(y_0, y_1)$ for $\theta$ based on $W_2$ is the solution for $\theta$ of the equation

$$\theta = \hat{\theta} + q_\alpha \hat{\sigma}.$$

The LR upper limit $u_L(Y_0, Y_1)$ for $\theta$ is the solution for $\theta > \hat{\theta}$ of the equation

$$\ell (\theta, \hat{\lambda}) = \ell (\hat{\theta}, \hat{\lambda}) - q_\alpha^2 / 2.$$

How accurate are these approximate bounds and how does the tight upper limit perform? The left panel of Figure 1 shows the partially minimised coverage function $C(\theta)$ of the LR upper limit for $n_0 = 10, n_1 = 15$. The points of discontinuities are highlighted and correspond to the unique values of $u_L(y_j)$. Coverage is well below nominal for most parameter values. The very low coverage in the centre of the plot corresponds to $\theta = 0.13523$. This corresponds to the "corner" data set $y^* = (y_0, y_1) =$
Figure 1: **LR and tightened upper limits.** Sample sizes are \( n_0 = 10, n_1 = 15 \) and target coverage is 95%. **Left.** Partially minimised coverage \( C(\theta) \) for \( u_L(y_j) \). **Centre.** \( u^*_L(y_j) \) versus \( u_L(y_j) \). **Right.** Partially minimised coverage \( C(\theta) \) for \( u^*_L(y_j) \).

(10, 15) for which the \( L(y^*) = 0.13523 \) is way too small. The central panel shows the tight upper limits. These rank the sample space in the same order as the original LR upper limits but tend to be larger.

The right panel shows the coverage function of the tight upper limits \( u^*_L(y_j) \) based on the LR upper limit. The coverage is always greater than nominal and equals nominal at the discontinuities. It is apparent that reducing any possible value of \( u^*_L(y_j) \) would move the discontinuity to the left and result in coverage falling below nominal. If you have seen the coverage function of one tight upper limit you have seen them all. From a strict frequentist perspective, their coverage properties are ideal. The issue then is how to calculate them.

### 3 Tight upper and lower limits

Buehler (1957) gave a formula for converting an approximate upper limit \( S(Y) \) into a new upper limit \( u^*_S(Y) \) that was as small as possible subject to satisfying the coverage constraint and also being a non-decreasing function of \( S(Y) \). The calculation is based on calculating the probability \( G(s; \pi) \) of the tail set \( R_S(s) = \{S(Y) \leq s\} \). Note that
since $S(Y)$ is an upper limit, this probability is not a P-value and the tail set is not even statistically meaningful. The tight upper limit is calculated from the probability surface $G(s; \pi)$ according to the formula

$$u_\ast^s(s) = \sup \{\theta(\pi) : G(s; \pi) > \alpha\}$$

(2)

see Buehler (1957, equation 27). Taking the supremum with respect to the nuisance parameter is consistent with testing theory of Bickel and Doksum (1977) who recommend that the quoted size of a test should be maximised over the nuisance parameter. Rohmel and Mansmann (1999) have shown that P-values maximised with respect to nuisance parameters have optimal frequentist properties similar to those of Buehler’s limits.

The adjusted upper limits $u_\ast^s(s)$ are called Buehler bounds in the reliability literature, but we will use the terminology tight upper limits in this article. If the set $\{G(s; \pi) > \alpha\}$ whose supremum is taken is empty, then $u_\ast^s(s)$ should be set equal to $\inf \Theta$ and still enjoys optimality, as shown in Lloyd and Kabaila (2003). Note also that $u_\ast^s(s)$ depends both on the statistic $S$ and its observed value $s$.

**Example ct’d.** Consider the relative risk model described in the previous section. Any sensible approximate upper limit $S(Y_1, Y_0)$ take its smallest value $s_{\text{min}}$ when $y_1 = 0$ and $y_0 = n_0$. In this case, and this case only, we can give a direct formula for the tight upper limit. The set $R_S(s_{\text{min}}) = \{S(Y) \leq s_{\text{min}}\}$ is just $\{Y_1 = 0, Y_0 = n_0\}$ and the probability surface $G(s_{\text{min}}; \pi) = \pi_0^{n_0}(1 - \pi_1)^{n_1}$. We need to find those parameter values $(\pi_0, \pi_1)$ for which $\theta = \log(\pi_1/\pi_0)$ is maximised subject to this probability being larger than $\alpha$. Since the probability is increasing in $\pi_0$, we first set $\pi_0 = 1$. Since $G(s_{\text{min}}; \pi)$ is continuous and decreasing in $\pi_1$, we equate $G(s_{\text{min}}; \pi)$ to $\alpha$ to obtain $\pi_1 = 1 - \alpha^{1/n_1}$. So the maximal solution for $\theta = \log(\pi_1/\pi_0)$ is $u^*(s_{\text{min}}) = \log(1 - \alpha^{1/n_1})$.

To give the reader some feel for how tight limits differ from approximate limits, consider the outcome $(y_1, y_0) = (50, 75)$ from $(n_1, n_0) = (150, 250)$. The ML estimator
of $\theta$ is 0.1054. Approximate 95% lower and upper limits based on the likelihood ratio are -0.1471 and 0.3501. Exact lower and upper limits are -0.2809 and 0.4882. The exact limits took around 3 seconds each to compute in the package R, using the algorithm described in this paper. This example has not been specially chosen.

Tight lower limits can quite generally be found by reparametrising the model in terms of $\theta' = -\theta$, finding an upper limit for $\theta'$ and then changing sign. For binomial models with $\theta = g(\pi_1) - g(\pi_0)$, reparametrisation is simply achieved by reversing the group labels. Henceforth, we will only discuss tight upper limits.

The statistic $S(Y)$ is called the designated statistic and ,while it can in principle be any statistic, for efficiency reasons it should be chosen to be an approximate upper or lower limit, see Kabaila and Lloyd (2002). Note that tight upper limits based on $S$ are the same as those based on $S + 10$ or any increasing function of $S$. It is only the ordering on the sample space induced by the function $S(\ )$ that matters. Consequently, it makes little difference which approximate upper limit is chosen, since different approximate upper limits order the sample space in a similar way.

While expression (2) is the most elegant formulation, the computations are more conveniently expressed in terms of an interest parameter $\theta(\pi)$ and a nuisance parameter $\lambda(\pi)$. Where there is no confusion, we henceforth suppress dependence of any quantities on the chosen statistic $S(Y)$. The probability of the tail set $R(s) = \{S(Y) \leq s\}$ is now denoted $G(\theta, \lambda; s)$ and the tight upper limit is

$$u^*(s) = \sup \{\theta : G(s; \theta, \lambda) > \alpha\}.$$  

(3)

There are two quite distinct computational difficulties that may arise in computing (3). The first issue is to calculate the probability $G(s; \theta, \lambda)$ of $R(s)$ for a single parameter value $(\theta, \phi)$. This can be hugely difficult for common models. The second problem is that even if computing $G(s; \theta, \lambda)$ is easy, searching the surface for the required solution is not straightforward as it typically does not have any monotonicity or convexity.
properties, especially with respect to the nuisance parameter $\lambda$. Consequently, incorrect solutions may always be found even for extreme search effort.

4 Monotonicity

Harris and Soms (1991) and Kabaila (2005) have investigated various monotonicity conditions on the model, designated statistic and parameter with a view to ensuring a unique solution to (3) and simplifying computations. They first require a condition on the model, namely that each component $Y_j$ of the data $(Y_0, ..., Y_k)$ has a distribution that depends on $\pi_j$ only and is stochastically increasing in $\pi_j$. This is the case for standard discrete data models based on binomial, negative binomial or Poisson variation.

The second condition is on the interest parameter $\theta(\pi_0, ..., \pi_k)$. We focus on parameters that are a monotone (i.e. either non-decreasing or non-increasing) function of the basic parameters $\pi_j$, which includes sums, products, differences and ratios.

The third condition is on the designated statistic $S(Y)$. We require that $S(Y_0, ..., Y_k)$ is a non-increasing or non-decreasing function of $Y_j$ as $\theta$ is a non-increasing or non-decreasing function of $\pi_j$. In this case, $S$ is said to be logically ordered (for $\theta$). As a very simple example, if $\theta = \pi_1 - \pi_0$ then the estimator $(Y_1 + \epsilon)/(n_1 + 2\epsilon) - (Y_0 + \epsilon)/(n_0 + 2\epsilon)$ is logically ordered for any $\epsilon > 0$. Standard Wald upper limits based on this estimator however are not necessarily logically ordered, because the standard error breaks down at extreme data sets. Upper limits based on the LR typically are logically ordered. For instance, it has been shown in Lloyd (2005) that LR based upper and lower limits for $\theta = \pi_1\pi_0$ are logically ordered, and the proof given there can easily be modified to cover a parameter $\theta$ that is a sum of monotone functions of $\pi_j$.

Let $\tilde{G}(\theta; s) = \sup_\lambda G(\theta, \lambda; s)$. Kabaila’s (2005) main result is that $\tilde{G}(\theta; s)$ is a non-increasing function of $\theta$. This requires that $S$ be logically ordered and a weak regularity condition on the parameter space being connected. He then shows that
$u^*(s)$ can be obtained as the largest solution of $\bar{G}(\theta; s) = \alpha$ and that a solution always exists. It is stated that this result “...not only simplifies the calculation of $u^*(y)$ but also guarantees the trustworthiness of the calculated result.” The second part of this statement ignores potential difficulties in calculating $\bar{G}(\theta; s)$ so that programming a solution of $\bar{G}(\theta; s) = \alpha$ does not necessarily result in a trustworthy output. Secondly, the monotonicity of $S(y_1, y_0)$ has an additional advantage not mentioned in that paper, namely that $G(\theta, \lambda; s)$ can be efficiently computed without enumeration of the sample space.

5 Calculating $G(\theta, \lambda; s)$

The most general expression for $G(\theta, \lambda; s)$ is a sum of probabilities of all possible outcomes $y \in \mathcal{Y}$ satisfying $S(y) \leq s$. For instance, a logistic regression with 100 binary observations has a sample space whose cardinality $N = 2^{100}$. We need to determine the subset $R(s)$ for which the LR based upper limit is equal or smaller than a given value $s$. This is clearly not feasible. One possible solution is to resort to some kind of simulation, such as importance sampling, to accurately estimate the probability of the set without enumerating it. This is not an approach we pursue in this paper.

For low dimensional sample spaces, the set $R(s)$ may often be represented as an intersection of one dimensional tail sets, provided the underlying test statistic has appropriate monotonicity properties. This is especially easy for two dimensional applications based on likelihood statistics such as arise in clinical trials and will be pursued below. The upshot of the approach adopted is that computing $R(s)$ involves evaluating $S(y_0, y_1)$ at only a small subset of the whole sample space.

Suppose that $n_1 \geq n_0$. Let $R(y_0, s) = \{y_1 : S(y_0, y_1) \leq s\}$. Then

$$R(s) = \{y_0, y_1 : S(y_0, y_1) \leq s\} = \cup_{y_0=0}^{n_0} R(y_0, s).$$

Provided that $S(y_0, y_1)$ is monotone increasing in $y_1$, $R(y_0, s)$ will be of the form
\{0, ..., y_1^*(s, y_0)\} where \(y_1^*(s, y_0)\) is the largest integer value of \(y_0\) such that \(S(y_0, y_1) \leq s\). This value can be found by bisection and the number of operations will be \(O(\log n_1)\). If the set is empty we set \(y_1^*(s, y_0) = -1\). The set \(R(s)\) is then represented by \(n_0\) integers \(\{y_1^*(s, y_0) : y_0 = 0, ..., n_0\}\). So \(R(s)\) can be identified with only \(O(n_0 \log n_1)\) evaluations of the statistic \(S(y_0, y_1)\). Moreover, if \(S(y_0, y_1)\) is monotone decreasing in \(y_0\), the sequence of integers \(y_1^*(s, y_0)\) is non-decreasing which allows the limits for bisection to reduce at each iteration.

If \(R(s)\) is represented in this form then the probability can be represented as

\[
G(\theta, \lambda; s) = \sum_{y_0=0}^{n_0} \Pr(Y_0 = y_0) \Pr(Y_1 \leq y_1^*(s, y_0)).
\]

For standard discrete distributions such as binomial or negative binomial, cumulative probabilities have direct representations and do not require accumulating \(O(n_1)\) primitive probabilities. Thus, once \(R(s)\) has been represented in parsimonious form, \(G(\theta, \lambda; s)\) can be computed in \(O(n_0)\) operations.

When \(n_0 \geq n_1\) there is a similar representation in terms of tail sets \(R(y_1, s) = \{y_0 : S(y_0, y_1) \leq s\}\) which will be of the form \(\{y_0^*(s, y_1), ..., n_1\}\) and this can be achieved in \(O(n_1 \log n_0)\) operations. The probability of the set can be expressed as a sum of \(n_1\) terms. So in summary, representing the tail set \(R(s)\) takes \(O(n_{\min} \log n_{\max})\) operations and then each evaluation of \(G(\theta, \lambda; s)\) takes \(O(n_{\min})\) operations. Especially for experiments where the sample sizes \(n_0, n_1\) are very unbalanced, the number of computations can be quite small.

### 6 Searching \(G(\theta, \lambda; s)\) for \(u^*(s)\)

In this section we give two quite different strategies for solving (3) and show how either of these strategies could potentially fail to find the optimum. Luckily, the two strategies tend to breakdown under different conditions and will always produce an upper limit smaller than \(u^*(s)\) when they do. So taking the maximum of computed values using
both strategies should give a robust solution. These two algorithms correspond to two
different representations of (3). Simply stated, to solve (3) you can maximise with
respect to $\lambda$ first and then find the solution for $\theta$, or you can find the solution for fixed
$\lambda$ and then maximise over different choices of $\lambda$.

**Algorithm 1.** Recall the earlier notation $\bar{G}(\theta; s)$ for $\sup_\lambda G(\theta, \lambda; s)$. Under weak regu-
lation conditions

$$u^*(s) = \sup \{\theta : \bar{G}(s; \theta) = \alpha\}.$$  

(5)

Under the monotonicity conditions in the previous section, $\bar{G}(\theta; s)$ is monotone in $\theta$
and so $u^*(s)$ is computed by finding the unique root of a monotone one-dimensional
function. The only potential difficulty then is that $G(\theta, \lambda; s)$ is an erratic function
of $\lambda$ with multiple local maxima so that $\bar{G}(\theta; s)$ is difficult to compute reliably. This
maximisation must be computed many times in order to find the root of $\bar{G}(\theta; s) - \alpha$.

**Example ct’d.** Returning to the problem of finding an upper bound for the logarithm
of the risk ratio of two binomials, consider the case where $n_0 = 25, n_1 = 30$ and we
observe $y = (y_0, y_1) = (9, 23)$. It is easy to calculate \( \log(\hat{\theta}, \hat{\lambda}) = (0.756, -0.276) \)
and the approximate 95% upper limit for $\theta$ based on the LR statistic is $s = 1.288$. The tail
set $R(s)$ for computing the tight upper limits comprises 589 of the 806 points in the
sample space. It can be parsimoniously represented by the 26 key integers $y^*_1(s; y_0)$
which are -1, 0, 1, 3, 6, 9, 12, 15, 19, 23, 26, 30(15). The Buehler upper limit turns
out to equal $u^*_L = 1.798$.

The left hand side of Figure 2 displays $\bar{G}(s; \theta)$ as a function of $\theta$ as a solid line,
computed by applying the R-function `optimise` to $B = 100$ separate sub-intervals
of the unit interval for $\lambda$. Note the discontinuous derivative, which can occur when the
location of the supremum with respect to $\lambda$ switches with changing $\theta$. According to (5),
the tight upper limit is the value of $\theta$ for which this plot equals $\alpha = 0.05$, which gives
1.798. The dotted line shows the computed value of $G(s; \theta)$ when $B = 5$ sub-intervals
are used for the $\lambda$-maximisation.

Clearly, the R-function `optimimise` is missing the maximum of $G(s; \theta, \lambda)$ which suggests that it must be an erratic function of $\lambda$. The right panel shows $G(\theta, \lambda; s)$ as a function of $\lambda$ for one particular value $\theta = 1.35$. There are multiple local maxima as well as maxima at the endpoints. This is not at all atypical though more regular plots are also common. When the function is irregular, it is very easy to miss the maximum unless the optimiser is embedded within an intensive grid-search. Notice also that the computed value $u^*$ will always be less than the correct value if the maximum is missed.

Algorithm 2. The second expression for $u^*(s)$ is simple in conception but requires a little more notation. Let $\Theta(\lambda') = \{\theta(\pi) : \lambda(\pi) = \lambda'\}$ be the parameter space of $\theta$ when $\lambda$ is fixed. Define

$$u^*(s, \lambda) = \sup \{ \theta \in \Theta(\lambda) : G(s; \theta, \lambda) = \alpha \}$$

(6)

if the set is non-empty. When the set is empty let $u^*_S(s, \lambda) = \inf \Theta$, which applies the
formula of Lloyd and Kabaila (2003) who also showed that

\[ u^*(s) = \sup_{\lambda} u^*(s, \lambda). \]  

(7)

It is shown in the appendix that \( G(\theta, \lambda; s) \) is non-increasing in \( \theta \) for each \( \lambda \) under the same condition for which \( \bar{G}(\theta; s) \) is non-increasing in \( \theta \). Hence \( u(s; \lambda) \) is very easily computed for each fixed \( \lambda \). We then need to maximise \( u(s; \lambda) \). While \( u(s; \lambda) \) need not be a particularly regular function of \( \lambda \), it tends to be much more regular than \( G(s; \theta, \lambda) \) as we will see in the example below. Moreover, there is only a single maximisation to perform whereas the first algorithm requires a potentially difficult maximisation with respect to \( \lambda \) for many values of \( \theta \). Lloyd and Kabaila (2003) gave extremely weak conditions for the equivalence of (3), (5) and (7). Continuity of the model with respect to \( (\theta, \lambda) \) is more than sufficient to ensure these weak regularity conditions.

**Example ct’d.** Recall that \( u^*_S(s; \lambda) \) is the largest solution of \( G(\theta, \lambda; s) = \alpha \) and that this function is non-increasing and has continuous derivatives with respect to \( \theta \), unlike \( \bar{G}(\theta) \). The left hand plot in Figure 3 shows \( G(\theta, \lambda; s) \) as a function of \( \theta \) for \( \lambda = 0.01 \). The solution of equating this to \( \alpha = 0.05 \) is \( u(s; 0.01) = 1.419 \). The right plot shows \( u^*(s, \lambda) \) as a function of \( \lambda \). The supremum of this plot is \( u^* = 1.798 \). Some care is needed in finding the supremum of this function. Again, failure to find the correct supremum will result in a computed values that is smaller than the desired solution.

This second method is typically quicker computationally and is more often the larger (and therefore more correct) of the two computations. In the right plot, the maximum occurs well to the left of the boundary. For extreme data sets, the maximum can occur at a discontinuous boundary point which will present problems for many standard optimisers. In this case, the first approach can be used.
Figure 3: Tight upper limits from observing $y_1/n_1 = 23/30$ and $y_0/n_0 = 9/25$. Left. $G(\theta, \lambda)$ versus $\theta$ for $\lambda = 0.01$ with $u^*(s, 0.01) = 1.419$. Right. $u^*(s, \lambda)$ versus $\log \lambda$. The parameter space boundary given by $\log \lambda < \theta$ and $\theta < -\log \lambda$ are marked as dashed lines. The supremum $u^* = 1.798$.

7 Conclusion

Confidence limits, both approximate or exact, should have monotonicity properties analogous to those of the parameter of interest. This is not only a logical requirement but simplifies computation of exact limits based on them. Firstly, the tail probability can be computed without searching the entire sample space. Secondly, this probability has certain monotonicity properties which facilitate computing the optimum point that defines the exact limit. We have an algorithm whose order of computation is equal to the minimum of the two samples sizes. While actual speed will obviously be dependent on the search effort setting and the machine used, typical elapsed time in seconds is roughly $\min(n_0, n_1)/50$ on a 2.4Gh desktop PC with 4Gb of RAM using the search parameter $B = 20$.

The importance of exact limits is that they have guaranteed coverage and are as
tight as possible subject to the coverage constraint. The table below describes the
distribution of the difference between exact upper limits $u^*_L(y_j)$ and approximate upper
limits $L(y_j)$ for the logged relative risk based on the likelihood ratio. The quantiles
are calculated across the whole sample space, after removing cases where both upper
limits are infinite. Exact upper limits are usually substantially larger than approximate
upper limits. For instance, the median difference is around 0.3 which means that exact
limits for the relative risk are typically 30% larger than approximate limits based on
the LR. Expressed another way, approximate upper limits are liberal, even for quite
large sample sizes, and need to be adjusted upwards by a non-trivial amount. What is
equally interesting is that the typical size of the adjustment does not appear to reduce
with sample size.

Table 1: Difference between exact upper limit for $\theta = \log(\pi_1/\pi_0)$ and approximate
upper limit based on likelihood ratio.

<table>
<thead>
<tr>
<th>$n_1$</th>
<th>$n_0$</th>
<th>min</th>
<th>$q_{.25}$</th>
<th>$q_{.50}$</th>
<th>$q_{.75}$</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>20</td>
<td>-0.005</td>
<td>0.118</td>
<td>0.272</td>
<td>0.443</td>
<td>0.893</td>
</tr>
<tr>
<td>20</td>
<td>30</td>
<td>-0.011</td>
<td>0.150</td>
<td>0.298</td>
<td>0.434</td>
<td>0.886</td>
</tr>
<tr>
<td>20</td>
<td>75</td>
<td>-0.020</td>
<td>0.268</td>
<td>0.373</td>
<td>0.447</td>
<td>0.872</td>
</tr>
<tr>
<td>30</td>
<td>20</td>
<td>-0.003</td>
<td>0.085</td>
<td>0.229</td>
<td>0.441</td>
<td>0.895</td>
</tr>
<tr>
<td>30</td>
<td>30</td>
<td>-0.014</td>
<td>0.120</td>
<td>0.273</td>
<td>0.437</td>
<td>0.887</td>
</tr>
<tr>
<td>30</td>
<td>75</td>
<td>-0.021</td>
<td>0.228</td>
<td>0.350</td>
<td>0.442</td>
<td>0.873</td>
</tr>
<tr>
<td>75</td>
<td>20</td>
<td>0.003</td>
<td>0.046</td>
<td>0.196</td>
<td>0.454</td>
<td>0.896</td>
</tr>
<tr>
<td>75</td>
<td>30</td>
<td>-0.004</td>
<td>0.043</td>
<td>0.198</td>
<td>0.453</td>
<td>0.888</td>
</tr>
<tr>
<td>75</td>
<td>75</td>
<td>-0.012</td>
<td>0.122</td>
<td>0.279</td>
<td>0.430</td>
<td>0.875</td>
</tr>
</tbody>
</table>
References.


Appendix. Montonicity of $G(\theta, \lambda; s)$

In concluding that $\bar{G}(\theta, s)$ is non-decreasing in $\theta$, Kabaila (2005) states and extra regularity condition (called condition C) that is required in addition to the three monotonicity conditions mentioned in section 4. Let $\pi$ denote the vector of basic parameters. We say $\pi_2 \geq \pi_1$ if the $\geq$ relationship holds for each element. Kabaila also notes that without loss of generality we can assume that all the monotone relations are non-decreasing, by suitable redefinition of parameters and data. His condition C is as follows. Let $\theta_1 \in \Theta$ and let $\pi_1$ satisfy $\theta(\pi_1) = \theta_1$. Let $\theta_2 \in \Theta$ and suppose $\theta < \theta_1$. Then there must be a set of parameters $\pi_2$ such that $\pi_2 \leq \pi_1$ and $\theta(\pi_2) = \theta_2$. At first sight, this would seem to automatically hold since $\theta$ is monotone increasing. However, if there a gaps in the parameter space then there may not be a suitable solution. Condition C is really a condition that the parameter space has no holes in it.

Suppose now that $\lambda$ is fixed, so that there is no free nuisance parameter and $\bar{G}(\theta, s) = G(\theta, \lambda; s)$. Let $\theta_1 \in \Theta(\lambda)$ and $\theta(\pi_1) = \theta_1$. Let $\theta_2 \in \Theta(\lambda)$ and suppose $\theta < \theta_1$. Suppose that there exists no set of parameter $\pi_2$ as mentioned above. Then there are two cases. Either all solutions $\pi_2$ of $\theta(\pi_2) = \theta_2$ are such that $\pi_2 \geq \pi_1$. This contradicts the monotonicity of $\theta$ as a function of $\pi$. Alternatively, there are no solutions of $\theta(\pi_2) = \theta_2$. This contradicts the assumption that $\theta_2 \in \Theta(\lambda)$. We may conclude that $G(\theta, \lambda; s)$ is a non-decreasing function of $\theta$ over the set $\Theta(\lambda)$ without condition C. Of course, from the point of view of computing roots of $G(\theta, \lambda; s)$, it is desirable that $\Theta(\lambda)$ should be a closed connected set.