Confidence intervals for common problems

D. F. Andrews  
University of Toronto  
Summer 2001

Abstract

Many common parameters may be defined to minimize a loss function. Such parameters may be estimated by minimizing the observed loss. A single general procedure is implemented to compute confidence intervals for such parameters from the signed root of the loss. The intervals are not grounded in any assumed distribution of the data. The procedure requires very little coding beyond that for estimation. Examples are given for multiple logistic regression and correlation. The resulting intervals are compared with BCA and bootstrapped estimating equations.

Key words and phrases: confidence limits, signed root, jackknife, bootstrap

1 Introduction

Statistical inference refers, typically, to statements about parameters based on observations. The data are known as are the computational procedures used to estimate the parameters. About nothing else are we certain. More specifically, the distribution of the population giving rise to the data is unknown. For this reason we are reluctant to use distribution dependent methods such as those based on saddlepoint approximations to the distribution of the likelihood ratio even though these are highly accurate when the distribution is known exactly. Furthermore, as Tukey (1949) noted, practical uncertainties as to the exact distributions involved leave us doubtful of the
practical significance of extreme percentage levels. Extreme tail behaviour cannot be estimated from typical sample sizes. Here we focus on 2.5% points.

Confidence limits are usually based on a statistic \( T(x, \theta) \), a function of the observations and the parameter of interest. The limits \( \theta_i \) are computed from probability statements of the form

\[
P[T(x, \theta_i) < T_\alpha] \approx \alpha
\]

or equivalently, for symmetric distributions,

\[
P[T^2(x, \theta_i) < T^2_\alpha] \approx 2\alpha
\]

The different types of intervals differ in the basis of the approximation.

Student’s \textit{t}-\textit{statistic} is robust. Statements based on it are useful under a broad collection of assumed underlying distributions. Intervals based on the \( t \)-statistic arise from setting \( T(x, \theta) = (\hat{\theta} - \theta)/\hat{\sigma}_\theta \) and the approximation

\[
P[\frac{\hat{\theta} - \theta_i}{\hat{\sigma}_\theta} < t_\alpha] \approx \alpha
\]

where \( t_\alpha \) is a percentile of the \( t \)-distribution.

However if

1. the variance of \( \hat{\theta} \) depends on \( \theta \) or
2. the distribution of \( \hat{\theta} \) is asymmetric or far from Gaussian

the approximation is poor and alternative methods are required. BCA bootstrap intervals (Efron and Tibshirani 1993) address the first concern by estimating the dependence of the variance on \( \theta \) at \( \hat{\theta} \). This procedure gives excellent results if \( \theta \) is close to \( \hat{\theta} \). However, as we shall see, it does not perform well when the confidence interval for \( \theta \) is large. Hu and Kalbfleisch (2000) use a score statistic \( \psi' \) with a bootstrap estimate of its variance at \( \hat{\theta} \) to generate intervals. Here again, the bootstrap estimation of the distribution of the score at \( \hat{\theta} \) leads to poor results if the interval is wide.

Variance stabilizing transformations typically improve both 1 and 2 as they yield almost constant variances and more symmetric distributions for the transformed estimate. Other forms of non-Gaussian distributions are best handled by robust definitions of parameters. The loss based intervals, described below, address the remaining concerns of 1 and 2.
We propose a general procedure for computing confidence limits. The procedure is simply implemented. It is based on old, well understood ideas. The only novelty of this paper is in the implementation and in the comparison with more recent proposals.

2 Parameters, estimates and signed roots.

Here we consider parameters defined to minimize a loss function $\psi(\theta)$

$$\theta = \text{arg min } E[\psi(\theta)].$$

This is a large class of parameters that includes those estimated by least squares, maximum likelihood and generalized linear models. We consider inference for parameters estimated from a sample of independent observations. In particular, we focus on the common case of inference for one parameter of interest in the presence of several nuisance parameters.

The parameters are typically estimated from $n$ independent observations by

$$\hat{\theta} = \text{arg min } AV[\psi_i(\theta)]$$

where $\psi_i(\theta)$ is the loss function minimized over the nuisance parameters and evaluated for the $i^{th}$ observation and where $AV$ denotes the average.

The avoidable loss or deviance function is defined:

$$d(\theta) = AV[\psi(\theta) - \psi(\hat{\theta})].$$

In Section 5 we show that in many important common problems, the distribution of $r(\theta) = \text{sign}(\hat{\theta} - \theta)\sqrt{d(\theta)}$ is approximately symmetric with almost constant variance. This signed root may be used to define confidence intervals since

$$P[T^2(x, \theta) < T^2_{\alpha}] \approx 2\alpha$$

where $T^2(x, \theta) = d(\theta)/\text{\var}(r(\theta))$ and $\text{\var}(r(\theta))$ is the jackknife estimate of variance. The limits are found as roots of the equivalent equation

$$r(\theta)^2 = d(\theta) = 4 \text{\var}(r(\theta))$$

with $T^2_{0.05} = 4$. The intervals may be implemented almost as easily as BCA intervals and often involve much less computation. These intervals which based on minimizing a loss or optimizing a fit are called OPT below.
3 Examples

This section presents some comparisons of the practical intervals with others.

3.1 Correlation

The sample correlation coefficient, $\hat{\rho}$ typically has a skew distribution with a variance that depends on the correlation $\rho$. The coverage of the OPT intervals is compared with that of BCA bootstrap intervals. Non-Gaussian, correlated observations were generated from $\chi^2_d$ distributions. Two independent samples $c_1$ and $c_2$ were generated. The correlated observations $(x_1, x_2)$, with $\rho = \sqrt{2}$, were generated by $x_1 = c_1$, $x_2 = c_1 + c_2$.

The non-coverage of these methods was computed by monte carlo of size 2000. This size is sufficient to estimate non-coverage of 5% with a standard error of 0.5%. The non-coverages are given in Table 1 below for a range of values of the sample size $n$ and the degrees of freedom $d$.

<table>
<thead>
<tr>
<th>n</th>
<th>OPT</th>
<th>BCA</th>
<th>OPT</th>
<th>BCA</th>
<th>OPT</th>
<th>BCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>10.1</td>
<td>11.9</td>
<td>7.5</td>
<td>8.5</td>
<td>6.1</td>
<td>6.7</td>
</tr>
<tr>
<td>64</td>
<td>7.1</td>
<td>9.2</td>
<td>6.5</td>
<td>8.4</td>
<td>5.4</td>
<td>6.5</td>
</tr>
<tr>
<td>128</td>
<td>7.4</td>
<td>9.8</td>
<td>4.8</td>
<td>6.7</td>
<td>5.0</td>
<td>6.0</td>
</tr>
</tbody>
</table>

For small degrees of freedom, the distribution of the data is far from Gaussian. The Gaussian maximum likelihood estimate of correlation leads to intervals with serious undercoverage. For larger degrees of freedom (more Gaussian data), for these realistic sample sizes, the OPT intervals have non-coverage much closer to the nominal 5% than the BCA intervals. Most of this improvement is associated with the form of the statistic used. BCA intervals based on some transformed correlation coefficient would be better. However, the OPT intervals make such a transformation automatically as shown in Section 5. This results from the function used to define and estimate the parameter. Thus we can expect similar favourable comparisons in other cases.
3.2 Logistic regression

Logistic regression is commonly used to estimate relative risk. In many studies involving disease in humans, with modest sample sizes, the resulting intervals are relatively large. It is not uncommon for such problems to involve many covariates with the result that the effective degrees of freedom are the order of 25 per parameter. For example, a typical study of 100 patients will often have at least 3 covariates in addition to a constant term. Such problems are not well served by intervals based on standard errors or $\chi^2$ approximations to likelihood ratios.

3.2.1 the simplest case, the binomial parameter $p$

We consider first the simplest logistic regression problem, one which involves a common proportion: $\theta = p$. Here we use the likelihood to define the loss function $\psi(p) = -(y \log(p/(1-p)) + \log(1 - p))$.

Since the sample space is small, $0 \leq y \leq n$, the properties of of intervals may be computed numerically, without monte carlo. The cases $y = 0$ and $y = n$ present special problems and are unlikely to be used as a basis for confidence intervals. Conditioning on the event $0 < y < n$, we compute the non-coverage and expected lengths of the likelihood ration interval based on its $\chi^2$ approximation, the OPT interval, the estimating function score interval, Hu and Kalbfleisch (2000), denoted here by score and the BCA interval.

Table 2 presents the non-coverage for the case $n = 25$ and $0.1 \leq p \leq 0.9$
Table 2. Non-coverage in % of intervals for a binomial parameter \( p, n = 25 \)

<table>
<thead>
<tr>
<th>( p )</th>
<th>Likelihood</th>
<th>OPT</th>
<th>Score</th>
<th>BCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>0.2</td>
</tr>
<tr>
<td>0.2</td>
<td>1.7</td>
<td>4.1</td>
<td>1.7</td>
<td>10.0</td>
</tr>
<tr>
<td>0.3</td>
<td>2.6</td>
<td>5.1</td>
<td>0.6</td>
<td>9.6</td>
</tr>
<tr>
<td>0.4</td>
<td>2.3</td>
<td>6.4</td>
<td>1.1</td>
<td>8.7</td>
</tr>
<tr>
<td>0.5</td>
<td>4.3</td>
<td>4.3</td>
<td>1.3</td>
<td>6.0</td>
</tr>
<tr>
<td>0.6</td>
<td>2.3</td>
<td>6.4</td>
<td>4.0</td>
<td>8.7</td>
</tr>
<tr>
<td>0.7</td>
<td>2.6</td>
<td>5.1</td>
<td>0.6</td>
<td>7.7</td>
</tr>
<tr>
<td>0.8</td>
<td>1.7</td>
<td>4.1</td>
<td>1.7</td>
<td>34.1</td>
</tr>
<tr>
<td>0.9</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>85.1</td>
</tr>
</tbody>
</table>

The OPT intervals are consistently closer to the nominal non-coverage of 5%. Although the bootstrap samples sizes used for the score and BCA intervals were both 2000, the vertical asymmetry of their columns indicates that there is considerable remaining monte carlo variation associated with the bootstrap estimation.

Table 3. Expected length in % of intervals for a binomial parameter \( p, n = 25 \)

<table>
<thead>
<tr>
<th>( p )</th>
<th>Likelihood</th>
<th>OPT</th>
<th>Score</th>
<th>BCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>27</td>
<td>24</td>
<td>80</td>
<td>21</td>
</tr>
<tr>
<td>0.2</td>
<td>33</td>
<td>31</td>
<td>59</td>
<td>29</td>
</tr>
<tr>
<td>0.3</td>
<td>37</td>
<td>35</td>
<td>44</td>
<td>35</td>
</tr>
<tr>
<td>0.4</td>
<td>39</td>
<td>38</td>
<td>42</td>
<td>38</td>
</tr>
<tr>
<td>0.5</td>
<td>40</td>
<td>39</td>
<td>43</td>
<td>39</td>
</tr>
<tr>
<td>0.6</td>
<td>39</td>
<td>38</td>
<td>43</td>
<td>38</td>
</tr>
<tr>
<td>0.7</td>
<td>37</td>
<td>35</td>
<td>44</td>
<td>35</td>
</tr>
<tr>
<td>0.8</td>
<td>33</td>
<td>31</td>
<td>57</td>
<td>28</td>
</tr>
<tr>
<td>0.9</td>
<td>27</td>
<td>24</td>
<td>79</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 3 presents the expected lengths of the confidence intervals for a range of values of \( p \). The OPT intervals have shorter expected lengths that the other methods in almost all cases.
3.2.2 logistic linear regression

The addition of nuisance parameters usually complicates inference. Here we consider a simple problem with one parameter of interest and one nuisance parameter. The situation modelled is of two populations, of sizes 40 and 10 respectively. The parameter of interest is the relative risk associated with being in the second population. If the underlying proportions are \( \text{logit}(p_1) = -2 \) and \( \text{logit}(p_2) = -1 \) the relative risk associated with the second population is \( \exp(\theta) \) with \( \theta = 1 \). This parameter may be simply estimated using generalized linear models for the binomial family.

Although calculation of the OPT and BCA intervals is fast some steps were taken to reduce the calculation to compute coverage. We first restricted the calculation to those outcomes with probability \( > 0.0001 \). This reduced the size of the sample space from \( 41 \times 11 = 451 \) to 148, a set with probability 0.9965. Some outcomes are so consistent with the true value of \( \theta \) that any reasonable interval will cover it. Other intervals are so discrepant that no reasonable interval will cover \( \theta \). The 57 outcomes with \( -2 \log - \text{likelihood} \) between 2 and 8 were selected and \( C(y) \), an indicator variable of coverage of each type of interval was computed. The probability of this set of difficult outcomes is 0.16. Since the probability of the outcomes with \( -2 \log - \text{likelihood} > 8 \) is negligible, we will call \( E[1 - C(y)] \) the non-coverage of the interval. In addition to the methods used above, we include the coverage of intervals based on the estimated standard error of the parameter denoted by SE. The non-coverages are listed in Table 4.

**Table 4.** Non-coverage for a relative risk parameter

<table>
<thead>
<tr>
<th>Method</th>
<th>OPT</th>
<th>BCA</th>
<th>Score</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Coverage</td>
<td>0.0352</td>
<td>0.0105</td>
<td>0.0027</td>
<td>0.0284</td>
</tr>
</tbody>
</table>

There are quite large differences in the coverage of the methods. Extremely conservative intervals give a misleading representation of the information about the parameter. Some of the extremely conservative behaviour of the bootstrap methods arises for the singular nature of some bootstrap samples. This, of course, could be handled with special programming. However, most users have expertise in other areas and need general procedures which do not require this coding.

The coverage of the OPT interval is closest to the nominal 5%.
4 Implementation

The computation of OPT intervals is based on the elegant simplicity of the implementation of BCA intervals. There are two components:

- a general function `optint` which calls
- a loss specific function

The loss specific function is the link between the general function and the particular procedures used to estimate the parameter. It returns, on demand, one of

- the parameter estimate, $\hat{\theta}$
- an approximate standard error of the estimate
- the deviance associated with a particular $\theta$ or
- bounds for the parameter space

In some cases involving iterated estimates, it may be useful to have it return a starting value as a linear predictor. Examples of this function are given in the appendix.

The general function `optint` uses the loss specific function to compute the roots of equation (1).

5 Theoretical basis

The distribution of $r(\theta)$, the signed root of $d(\theta) - d(\hat{\theta})$ the avoidable loss is almost independent of $\theta$. This approximate invariance improves the accuracy of estimation based on the empirical distribution with parameter $\hat{\theta}$.

This approximate invariance is illustrated by the bivariate correlation. Here we use the more conventional $\theta = \rho$, estimated using Gaussian maximum likelihood. The signed root $r(\rho) = 2 \log(1 - \hat{\rho}) - \log(1 - \hat{\rho}^2) - \log(1 - \rho^2)$ is a transformation of $\hat{\rho}$ which approximates a variance stabilizing transformation namely Fisher’s transformation of the correlation coefficient $t(\hat{\rho}) = log((1 + \hat{\rho})/(1 - \hat{\rho}))$. 
Figure 1: The signed root $r(0.7)$ and Fisher’s $a + b \ t(\hat{\rho})$ where $a$ and $b$ are chosen so the statistics are matched at $\hat{\rho} = 0$ and $\hat{\rho} = 0.7$, plotted against $\hat{\rho}$ on the horizontal axis.

Figure 1 displays the signed root $r(\theta)$ based on the Gaussian deviance function for the bivariate Gaussian with interest parameter $\theta = \rho = 0.7$, the correlation coefficient. The figure also shows $a + b \ t(r)$ where the constants $a$ and $b$ were chosen to match the statistics at $\hat{\rho} = 0$ and $\hat{\rho} = 0.7$. The two statistics are almost identical: the signed root approximates the variance stabilizing transformation. The similarity of these to functions is a mathematical result; no distributions or variances were used to produce the plot.

This variance stabilizing characteristic holds quite generally. This is easily seen by from a Taylor expansion of the equations defining the parameter and its estimate.

The parameter is defined by

$$E[\psi'(\theta)] = 0.$$  \hfill (2)

Its estimate satisfies the equation

$$AV[\psi'(\hat{\theta})] = 0.$$
The Taylor expansion of $\psi'$ in the above gives
\[
AV[\psi'(\theta)] + (\theta - \hat{\theta})AV[\psi''(\theta)] + (\theta - \hat{\theta})^2 AV[\psi'''(\theta)]/2 + (\theta - \hat{\theta})^3 R_1(\theta) = 0,
\]
where $R_1(\theta)$ denotes a remainder term of order $O_P(1)$.

The Taylor expansion of the avoidable loss about $\theta = \hat{\theta}$ involves derivatives of $\psi(\hat{\theta})$. If these derivatives, in turn, are expanded about $\hat{\theta} = \theta$ the resulting expression is
\[
2(AV[\psi(\theta)] - AV[\psi(\hat{\theta})]) = (\theta - \hat{\theta})^2 AV[\psi''(\theta)](1 + (\theta - \hat{\theta})R_2(\theta)).
\]
The signed root therefore satisfies
\[
\sqrt{2} \ r(\theta) = (\theta - \hat{\theta})\sqrt{AV[\psi''(\theta)]} + (\theta - \hat{\theta})^2 R_3(\theta).
\]
Its expectation has order
\[
E[r(\theta)] = 0 + E[O_P((\theta - \hat{\theta})^2)]
\]
independent of $\theta$.

The more remarkable fact is that its mean square error also is almost independent of $\theta$. To see this note
\[
2 \ MSE[r(\theta)] = 2 \ E[r(\theta)^2] = E[(\theta - \hat{\theta})^2 AV[\psi''(\theta)] (1 + (\theta - \hat{\theta})R_2(\theta))]
\]
an expression of order $E[O_P((\theta - \hat{\theta})^2)]$. Taking the derivative of this with respect to $\theta$ gives
\[
2 \frac{\partial MSE[r(\theta)]}{\partial \theta} = 2 \ E[2(\theta - \hat{\theta})AV[\psi''(\theta)] + (\theta - \hat{\theta})^2 AV[\psi'''(\theta)] + O_P((\hat{\theta} - \theta)^3)]
\]
from (2) and (3). Thus the signed root $r(\theta)$ has a mean square error of order $E[O_P((\theta - \hat{\theta})^2)]$ while its derivative is of smaller order, $E[O_P((\hat{\theta} - \theta)^3)]$. The signed root, even without standardization, has almost constant mean square error. It is important to note that this result follows primarily from the form of the estimate (3) and not from any assumed distribution.
5.1 symmetry

Asymmetry is often associated with variances dependent on the parameter of interest. This association is exploited by BCA intervals which estimate the dependence of variance from the third moment of the distribution of the statistic. If the statistic has constant variance, the skewness is typically small and no adjustment is required.

The common cases of asymmetric distributions arise for estimates of the binomial proportion, the correlation coefficient or of parameters like variances that must be positive. In these cases, the parameter space is bounded. This asymmetry is reduced by simple transformations: logit, hyperbolic tangent and logarithm, respectively. While these transformations also stabilize the variance, the above rational for their use is much more elementary and does not rest on any particular distribution.

The signed root \( r(\theta) \) is typically unbounded. It therefore does not suffer from the asymmetry that the bounds induce. This happens in the examples investigated to date. We are still searching for a yet more elementary explanation of this phenomenon.

6 Discussion

Intervals based on BCA bootstrap or estimating equation scores have excellent asymptotic properties (Hu and Kalbfleisch 2000). These properties are not particularly relevant if the relative range of the confidence interval is large. The properties for samples sizes met in application are a more relevant basis for comparison. In many practical problems, the practical intervals have better properties. This performance is based in part on

- BCA bootstrap methods compute an acceleration parameter, the parameter that influences the asymmetry of the interval, only at \( \theta = \hat{\theta} \). If this parameter is influential in determining the interval, and if it depends heavily on \( \theta \), as it does for the correlation coefficient where \( \partial \sigma^2_\theta / \partial \rho \approx -4\rho(1 - \rho^2)/n \), the resulting intervals may behave poorly.

- The estimating function score methods use bootstrap methods to estimate the distribution of a score statistic \( \psi'(\theta) \) at \( \theta = \hat{\theta} \). If the distribution of \( \psi'(\theta) \) depends heavily on \( \theta \), the intervals will perform poorly.
• The OPT intervals exploit more known information about the estimate, information not used by the standard bootstrap methods. In particular, properties of the distribution of $\psi(\theta)$ are estimated at the confidence limits $\theta = \theta_i$.

• In many common problems the use of a variance stabilizing transformation results in better intervals. Simple Taylor expansions show that the signed root approximates the variance stabilizing transformation of the estimate. If the loss function is a deviance, the asymptotic variance of the signed root is $1 + O(n^{-1})$, independent of $\theta$.

• The intervals described here involve considerable amounts (seconds) of calculation. The OPT intervals typically require about 4 jackknife calculations for each limit. Thus they require the order of 10 $n$ evaluations of $\hat{\theta}$ in all. For many problems ($n < 200$), this is less than the 2000 bootstrap evaluations suggested by Efron and Tibshirani (1993) and much less than the 10,000 suggested by Hu and Kalbfleisch (2000).

The OPT intervals use a jackknife estimate of the variance of $r(\theta)$. Most applications of confidence intervals are for parameters defined by smooth loss functions. The only exception we have met in practice is that of the median; and is met all too rarely. We do not recommend the OPT intervals for computing confidence intervals of quantiles. The procedure could be modified to use bootstrap estimation of the variance in these cases. However, more direct, non-parametric methods are more appropriate for these parameters.

The OPT intervals are simple to implement. The only code required is that which links the general procedure to the specific estimation procedures used for the problem. Even this function is quite general. For example, the function used for linear logistic regression can be used for logistic regression with an arbitrary matrix of predictors. Only one such function is required for all linear logistic regression problems. This function is much simpler than the procedures required for estimating function scores.

ACKNOWLEDGEMENT

I am grateful to Professors Keith Knight and Rob Tibshirani for many helpful discussions. This research was supported in part by the Natural Sciences and Engineering Research Council of Canada.
REFERENCES


7 APPENDIX

This appendix presents the R code for the general function and for the loss specific functions used in the examples.

The problem specific function to compute confidence intervals for the correlation coefficient is

```r
gausscor <- function(x,ftype,interest=0,linstart=0){
  # familiar notation
  rho <- interest; r <- cor(x)[1,2]; n <- nrow(x);
  # estimates
  if(ftype == "E") return(r)
  # loss or deviance
  if(ftype == "D") return(-n*(log(1-r^2) - 2*log(1 - r*rho) +
                          log(1-rho^2)))
  # approximate variance of estimate
  if(ftype == "V") return(1/n)
  # limits of parameter
  if(ftype == "L") return(c(-0.999,0.999))
}
```

The general function is defined by two procedures.

```r
critstat <- function(t1=t1 ,x=x,critf=critf,critval = 4,start=F){
  est <-critf(x,"E");
  linstarte <- 0;
  if(start) linstarte <- critf(x,"S") - est*x[2,];
  deve <- critf(x,"D",interest=est,linstart = linstarte);
  devt <- critf(x,"D",interest=t1,linstart = linstarte);
  if(!is.null(nrow(x))) n <- nrow(x)
  else n <- length(x)
  pseudo <-rep(0,n);
  for(i in 1:n){
    if(!is.null(nrow(x))) xi <- x[-i,]
    else xi <- x[-i];
    if(start) linstarti <- linstarte[-i]
    else linstarti <- 0
    esti <- critf(xi,"E")

    if(abs(esti - devt) > critval) {
      esti <- 0
      devt <- critf(x,"
```
devie <- critf(xi,"D",interest = esti,lstart = lstarti);
devit <- critf(xi,"D",interest = t1,lstart = lstarti);
pseudo[i]<-(n-1)*sign(est -t1)*(sqrt(max(devit - devie,0)))
}
return(devt - deve -critval* var(pseudo)/(n))
}

optint<- function(x,f,critvalp = 4, startp = F){
#initialize for root finding
    est <- f(x,"E");
    vest <- f(x,"V",interest=est);
    elims <- f(x,"L");
    roots <-c(max( est - 2*sqrt(vest),elims[1]),
        min( est + 2*sqrt(vest),elims[2]));
    iters <- c(0,0)
    estlims <-c(max( est - 4*sqrt(vest),elims[1]),
        min( est + 4*sqrt(vest),elims[2]));
    if(!is.null(nrow(x))) n <- nrow(x)
    else n <- length(x)
    ce <- critstat( est,x,critf =f,critval=critvalp,start=startp)
    cu <- critstat( estlims[2],x,critf =f,critval=critvalp,start=startp)
    cl <- critstat( estlims[1],x,critf=f,critval=critvalp,start=startp)
    #find roots
    if((cu > 0) && (ce < 0)) {
        uroot <- unirroot(critstat,c(est,estlims[2]),x=x,critf=f,
            critval=critvalp,start=startp,tol = sqrt(vest)/10);
        roots[2]<-uroot$root
        iters[2] <- uroot$iter
    }
    if((cl > 0) && (ce < 0)) {
        uroot1 <- unirroot(critstat,c(estlims[1],est),x=x,critf=f,
            critval=critvalp,start=startp,tol = sqrt(vest)/10)
        roots[1]<-uroot1$root
        iters[1] <- uroot1$iter
    }
    list(roots=roots, iters = iters)
}
The loss specific function for the binomial parameter is

```r
binf <- function(x, ftype, interest=0, linstart = 0) {
    # binomial notation
    p <- interest;
    n <- length(x);
    phat <- mean(x);
    # estimates
    if(ftype == "E") return(mean(x))
    # loss or deviance; take care of singular cases
    if(ftype == "D"){
        if(p*(1-p) == 0) return(0)
        else{
            phat <- mean(x);
            if((phat == 0) || (phat == 1)) {
                return(-2 * (phat*log(p/(1-p)) + log(1-p)) )
            }
            else {
                return(-2* (phat*log(p/(1-p)) + log(1-p) -
                           phat*log(phat/(1-phat)) ) - log(1-phat))
            }
        }
    }
    # approximate variance of estimate
    if(ftype == "V") return(max(phat * (1-phat)/n, 4/(n^2)))
    # limits of estimate
    if(ftype == "L") return(c(0.000001, 0.999999))
}
```