Evaluation of links in heroin seizures

L. Dujourdy, G. Barbati, F. Taroni, O. Guéniat, P. Esseiva, F. Anglada, P. Margot

*Institut de Police Scientifique et de Criminologie, University of Lausanne, BCH, 1015 Lausanne-Dorigny, Switzerland
*Institut de Médecine Légale, University of Lausanne, Rue du Bugnon 21, 1005 Lausanne, Switzerland
*Police Cantonale Neuchâteloise, Poudrières 14, 2006 Neuchâtel, Switzerland

Received 25 February 2002; accepted 19 November 2002

Abstract

The evaluation of a link between two heroin seizures using a descriptive method is presented. It is based on the measure of the angles between two chromatograms assimilated to vectors, and interpreted using a continuous approach based on the likelihood ratio of Bayes’ theorem. A complete evaluation model thus avoids the drawbacks of decision thresholds used until now to establish a link. Validation is obtained through tests and simulation methods.

© 2002 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: Drug profiling; Cosine function modelling; Bayes’ theorem; Likelihood ratio; Evidence evaluation; Diacetylmorphine

1. Introduction

The chemical profiling of illicit drug samples seizures is an appropriate analytical tool to detect links between different drug products [1]. Various analyses are used to establish the chemical signature of a drug. Gas chromatography is the main method; it allows a good resolution in impurity separation [2] together with appropriate sensitivity and reproducibility. A proficient method for the separation of impurities, adulterants, diluents and the quantification of the main active substance is necessary and allows to observe links that may lead police operations or reinforce suspected links obtained by separate police investigations.

The evidence helps to objectively back evidence of the activity extent of a trafficker or simply his/her belonging to a criminal network and it thus useful for court. The limits and capacity of a method to establish a link between two samples need distance measuring methods that are an efficient solution [3] by making the most of the impurity profiles obtained by an analysis.

A model was proposed by Keto with exploitation of the correlation coefficient described by the cosine function for pyrograms comparison [4] and has been adapted to the comparison of heroin samples chromatograms. Our model has been validated by research and samples obtained through other mean of investigations [5]. It is based on a continuous likelihood ratio approach widely used in forensic sciences [6].

2. The comparison method

Links are generally established by considering a threshold correlation value $C_s$, which normally results from the calculation of within-source variation determined as correlation values of replicates of the seizure to be compared, i.e. the within-source variation.

Correlation values smaller than $C_s$ usually indicate that two samples can be considered as not linked; above $C_s$, the samples are considered linked. Therefore, a decision is made for this threshold value set at a relative high level (for instance, $C_s = 99.8$) in order to avoid the detection of false positives. This decision suffers from the ‘fall of the cliff’ effect [7] where either two samples are declared undifferentiated or, suddenly, different based on a decision threshold. It is clearly illogical to say that a correlation value of 99.81 is evidence of common origin between two seizures and a test result of 99.79 is inconclusive or should lead to rejection of this previous hypothesis (therefore, the seizures
comes from two different populations). In reality, it could be admitted that the more extreme values of \( C_s \) support strongly that the two seizures comes from the same population, whereas low \( C_s \) values clearly support the alternative hypothesis. The use of a continuous likelihood ratio approach avoids this pitfall and allows to treat intermediate values meaningfully [8].

The current study was carried out with a database called ‘Database Not Linked’ (DBNL), which contained 128 independent seizures [3]. The hypothesis being that there will be a spread of correlation values (they are after all heroin samples!) with few if any very high values.

Fig. 1 shows the number of pairs (8128 in total) compared by their correlation values, all of which were obtained with a automatically using Visual Basic® macros for Excel®. An important number of pairs fall within the correlation values between 85 and 100. The chemical profiles are predominantly close, which makes difficult any estimation on the decision threshold.

3. The distribution curves and their estimations

A theoretical description of the distribution curve of the non-linked seizures database (DBNL) was a first concern and the objective was to estimate its density and therefore know the statistical behaviour of the data.

We have chosen beta distributions because they are very versatile and a variety of uncertainties can be usefully modelled by them. This flexibility encourages its empirical use in a wide range of applications. The beta distributions are among the most frequently employed to model theoretical distributions [9].

The family of beta distributions is composed of all distributions with probability density functions of form:

\[
p_Y(y) = \frac{1}{B(p,q)} (y-a)^{p-1}(b-y)^{q-1}, \quad a \leq y \leq b, \quad (1)
\]

with \( p > 0, \ q > 0 \), and \( B(p,q) = \int_0^1 t^{p-1}(1-t)^{q-1} \, dt \).

Making the transformation \( X = (Y-a)/(b-a) \), we obtain the probability density function:

\[
p_X(x) = \frac{1}{B(p,q)} x^{p-1}(1-x)^{q-1}, \quad 0 \leq x \leq 1. \quad (3)
\]

This is the standard form of the beta distribution with parameters \( p \) and \( q \), which must be estimated but direct algebraic solutions of the maximum likelihood (ML) equations cannot be obtained for beta distributions.

Estimation of all four parameters \((a, b, p \text{ and } q)\) can be done by equating sample and population values of the first four moments [9]. If the values of \( a \) and \( b \), respectively, the minimum and maximum of the distribution, are known, then only the first and second moments need to be used, giving:

\[
p + q = \frac{[(\bar{x} - a)/(b-a)][1 - (\bar{x} - a)/(b-a)]}{\text{var}(x)/(b-a)^2} - 1,
\]

\[
p = \frac{(\bar{x} - a)^2}{b - a} \left( 1 - \frac{\bar{x} - a}{b - a} \right) \left( \frac{\text{var}(x)}{(b-a)^2} \right)^{-1} \frac{\bar{x} - a}{b - a}, \quad (4)
\]

where \( \bar{x} = \text{mean}(x) \) and \( \text{var}(x) = \text{variance}(x) \) and \( x = \text{value observed} \).
Note that in some cases, the method of moments yield more accurate estimates of \( p \) and \( q \) than does the method of maximum likelihood (ML), possibly due to bias introduced by the computational method used in determining the ML estimators [9]. Let us define \( a = Y_n \) and \( b = Y_n^* \) which are the minimum and the maximum values of the population, respectively (i.e. the total range of variation of the results).

The non-linked samples were split into two populations to which beta distribution studies were applied:

- The population with correlation values between 22 and 95; the figures corresponding to the 10 and 90% quantiles (database number 1).
- The population with values between 95 and 100 (database number 2).

This choice had arose from the important quantity of data (>8000 comparison pairs) and we focused on the selection of the more problematic grey area where the linkage probability increased.

Indeed, the number of samples between 22 and 95 is very high but has a very low link probability whereas the lower numbers between 95 and 100 show an increasingly more likely link probability thus creating the risk of false positives. The distribution part of most interest (correlation values between 95 and 100) is shown in Fig. 2.

Beta distribution parameters \( p \) and \( q \) were calculated for the first database (22 \( \leq C \leq 95 \)) and for the second one (95 \( \leq C \leq 100 \)). These parameters allow to draw the curves corresponding to the two distributions fitting. Beta distribution parameter values for database number 1 are \( p = 1.104 \) and \( q = 0.734 \), whereas for the database number 2 we get \( p = 1.157 \) and \( q = 1.439 \).

Figs. 3 and 4 show empirical and theoretical adjustments by a beta distribution of the first and second databases of not linked sample seizures. We can note the nearly perfect fitting between the theoretical and empirical values. It shows an excellent description of the data.

We have also tested the adjustment using the goodness of fit for the estimated models with the Pearson’s \( \chi^2 \). The Chi-square test examines the frequency distribution of \( n \) observations grouped into \( k \) classes. The observed counts \( c_i \) in each class are compared to the expected counts \( e_i \) from the hypothetical distribution.

Under the null hypothesis that the sample comes from the hypothetical distribution, it has a \( \chi^2 \) distribution with \( k - 1 \) degrees of freedom. When the sample size is very large (as is the case of the not linked database) almost any model yields a highly significant Chi-square value; this is because the Chi-square statistics not only increases as the model fitted increases in distance from the true model, but also increases with the \( n \) dimensions.

In our case, for the database number 1, we obtain a Chi-square value of 3.4952 (d.f. = 1), therefore we can consider that this is not significant at a level \( \alpha = 0.01 \) (P-value = 0.0615). For the database number 2, we obtain a Chi-square value of 0.9006 (d.f. = 2), therefore we can consider that this is not significant at a level \( \alpha = 0.01 \) (P-value = 0.6374). Therefore, the theoretical model (fitting with beta distributions) we proposed can be accepted.

4. Parameters stability for not linked seizures database

To verify the stability of the parameters estimated for the total not linked databases distribution, we have considered the global distribution for the years 1997–2000, with more than 65,000 observations; from this distribution we have
Fig. 3. Empirical and hypothesized beta simulation. Solid line is the empirical distribution function $\hat{F}_n(x)$ and dotted line is the theoretical distribution function $F_0(x)$ for DBNL number 1.

Fig. 4. Empirical and hypothesized beta simulation. Solid line is the empirical distribution function $\hat{F}_n(x)$ and dotted line is the theoretical distribution function $F_0(x)$ for DBNL number 2.
extracted 65 samples of size 1000, 2000, 3000 and so on until 65,000.

Then, for every new samples we have estimated the parameters for the beta distribution, to analyse if the \( p \) and \( q \) values change with the sample size dimension (dim), two regression models have been built:
\[ p = \alpha + \beta (\text{dim}) + \text{err}, \quad q = \alpha' + \beta' (\text{dim}) + \text{err}. \] (5)
The beta values are not significantly different from 0; therefore, we can confirm that the not linked database distribution is stable with respect to its dimension. It can be safely used as the “relevant population” for any assessment estimate. Its mean values are: \( E(p) = 1.600 \) and \( E(q) = 0.661 \) (\( E \) means expectation).

If the whole of the not linked database is stable, we can suppose that sub-groups are also stable. It is therefore valid for the two databases (corresponding to \( 22 \leq C \leq 95 \) and \( 95 \leq C \leq 100 \)) used in the research.

5. The continuous likelihood ratio for link assessment

The evaluation of link can be determined in similar fashion to glass fragments [10]. In cases involving glass fragments, if fragments are recovered on a suspect’s clothing, the scientist investigates if those could come from a control sample (a broken window). In such a situation, the potential source is represented by the window glass.

With illicit drugs samples, the source is generally unknown (the laboratory where the drug is made is unknown). Therefore, the knowledge of the chemical variability of the drug batch made and put on the market is incomplete (uncertain).

From a practical point of view, the lack of information on the variability of the potential source can be solved by estimating the variation by combining the results obtained from the two seizures in order to extract a set of correlation values.

Our concern was to assess the linking value between two seizures \( \lambda \) and \( \gamma \). To this end, we used the likelihood ratio approach \( P(E|H_1)/P(E|H_2) \), where the hypotheses \( H_1 \) corresponds to: “seizure \( \lambda \) is from the same source as seizure \( \gamma \)”, and the alternative \( H_2 \) is: “seizure \( \lambda \) is from a different source than seizure \( \gamma \)”. The evidence \( E \) is the correlation of analytical results from samples of the two seizures.

\( E \) is represented by the median of the distribution (a mix between the two seizures). We have chosen the median of the distribution, which is more robust than the mean with respect to deviations from the normality assumptions. \( H_1 \) and \( H_2 \) are represented by the densities of seizures and of database (DBNL), respectively. Under \( H_1 \), one seizure is considered at numerator (\( \lambda \) or \( \gamma \)) and under \( H_2 \), it is DBNL at denominator.

Fig. 5 gives an illustration of the concept with a fictive example where seizures are not linked (in dotted line) and linked (in grey). ‘Database Not Linked’ is in black.

Thus, the quotient of values of the probability density functions of the median value (Med) of the blend distribution estimated for the seizures at numerator (5) and for the not linked samples (DBNL) at denominator is computed:
\[ LR = \frac{f_S(\text{Med}|H_1)}{f_{\text{Med|DBNL}}(\text{Med}|H_2)}. \] (6)

Large values of the LR support the link and indicate that the two seizures are coming from the same source; as the values decrease the strength of the link is weakened until values

![Figure 5](image_url)

**Fig. 5.** Illustration with a fictive example of curves resulting in combining two seizures: not linked in dotted line, linked in grey and ‘Database Not Linked’ is in black.
below 1 so that the LR support that the two seizures are not related or that they came from different chemical batches. [Note that the likelihood ratio, also called the Bayes’ factor, is a factor that depends only upon the sample data (so, it depends on the degree of within-source of the seizures examined).]

The main advantage of such an approach is that we can test values without predetermining the outcome of the test according to the selection of the sample size, and there is no need to pre-specify an arbitrary level of significance [11].

6. A case example

Let us illustrate the use of the model with two examples:

1. A and B not linked, informations obtained by external sources (police informations).

2. X and Y linked.

The continuous likelihood ratio (LR) approach compares the within-source and between-source estimates [6]. The within-source variation was carried out by comparing all the seizure samples two-by-two. The result is shown in Fig. 6. The A and B seizures are both composed of 30 samples each (435 pairs per seizure).

As for the not linked database, the distributions of the two seizures were carried out according to the beta distribution. The following parameters were obtained: $p_A = 3.783$, $q_A = 0.288$ and $p_B = 3.540$, $q_B = 0.382$. The goodness of fit with the Chi-square test accepts the null hypothesis of beta distributions with these specific parameters.

The second example, seizures called X (11 samples corresponding to 55 pairs) and Y (29 samples corresponding to 406 pairs) were chosen. Fig. 7 shows the within-source variation among those seizures. Beta distribution parameters for the two seizures are: $p_X = 0.957$, $q_X = 0.643$ and $p_Y = 1.483$, $q_Y = 0.294$. The goodness of fit with Chi-square test accepts null hypothesis of beta distributions with these specific parameters. Figs. 8 and 9 show the correlation values distributions for the seizure combinations A and B, and X and Y, respectively.

In Table 1 we put a summary of numerical results obtained with the modelling in the two examples. In the A and B seizures, the standardised value of the median Med = 0.456, when B density is introduced to the numerator. The likelihood ratio LR = 0.11. In the X and Y seizures, Med = 0.998, when Y density is put in the numerator. The LR = 243.86.

So, the LR values allow us to conclude that A and B are probably not coming from the same source and X and Y are probably coming from the same source. In the first situation, the LR value supports the alternative hypothesis, in fact a value smaller than 1 has been obtained. In the second situation, the numerator hypothesis is supported because the LR is smaller than 1. It is 244 time more likely to observe a median of 0.998 if X and Y are coming from the same source.

The calculation of the likelihood ratio with the following and slightly modified hypotheses can also prove interesting. $H_1$: “seizure B is from the same source as seizure A”, and under the alternative hypothesis $H_2$: “seizure B is from a different source than seizure A”. Alternatively, for the assessment of seizure X and Y, $H_1$ becomes, “seizure Y is from the same source as seizure X and $H_2$,” “seizure Y is from a different source than seizure X”.

In the A and B seizures, Med = 0.456, LR = 0.07, when density A is used in the numerator. In the X and Y seizures, Med = 0.998, LR = 50.96, when density X is introduced to the numerator.

![Fig. 6. Distribution of pairs number: seizure A in black and seizure B in grey.](image-url)
Fig. 7. Distribution of pairs number: seizure X in black and seizure Y in grey.

Fig. 8. Distribution of pairs number for combination of seizure A with seizure B.

<table>
<thead>
<tr>
<th>Seizure</th>
<th>$p$</th>
<th>$q$</th>
<th>Median value for mix seizures distribution</th>
<th>LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3.783</td>
<td>0.288</td>
<td>0.456</td>
<td>0.11 (B at numerator)</td>
</tr>
<tr>
<td>B</td>
<td>3.540</td>
<td>0.382</td>
<td></td>
<td>0.07 (A at numerator)</td>
</tr>
<tr>
<td>X</td>
<td>0.957</td>
<td>0.643</td>
<td>0.998</td>
<td>243.86 (Y at numerator)</td>
</tr>
<tr>
<td>Y</td>
<td>1.483</td>
<td>0.294</td>
<td></td>
<td>50.96 (X at numerator)</td>
</tr>
</tbody>
</table>
We observe that there is a great variation between the two estimated likelihood ratio for seizure X and seizure Y. In order to explain this phenomenon, we have calculated LRs (with a standardised median equal to 0.998) of many seizures linked with X (seizures D1–D4 were stored in a heroin samples database, used to describe the numerator of the LR). The results of beta parameters and LRs are summarised in Table 2.

The variation of LRs seems to be independent from the size of the seizures (expressed by the number of pairs). The comparison of equal number of samples (D2 and D3) produces different LR values (50.15 and 214.43).

Consequently, we are interested in obtaining not only a point estimate of the LR statistic but also an estimate of its variation to assess a confidence interval for the unknown value of the statistic. Re-sampling techniques such as bootstrap can provide a measure of the standard error, confidence intervals and distributions for the LR statistic [12].

In the bootstrap operation, n new samples are drawn with replacement from the observed data. This operation is repeated 1000 times using the same size of data. The statistic is calculated for each new set of data, yielding a bootstrap distribution. A picture of the shape of the distribution is observed through the generated histogram.

The plot of lrXbootstrap distribution (corresponding to seizure X) in Fig. 10 reveals a skewed distribution, skewed toward the right (that is, towards greater values). The distribution is thus not normal, because of this right tail.

The plot of lrYbootstrap distribution (corresponding to seizure Y) in Fig. 11 reveals a distinctly skewed distribution, skewed toward the left (that is, toward smaller values). The distribution is not normal, because of this longer left tail.

The standard error of the lrXbootstrap is 27.9, and for lrYbootstrap is 23.4 (the mean of lrXbootstrap t is 57.0 and lrYbootstrap mean is 237.8). Standard errors are often used to assign approximate confidence intervals. In the present case, distributions are not Gaussian and asymmetric; it is better to compute the bias-corrected percentile intervals, obtained from the bootstrap distribution [13] (see Appendix A):

\[ lrXbootstrap \rightarrow [13.6, 99.6], \]
\[ lrYbootstrap \rightarrow [207.3, 275.5]. \]

In conclusion, differences between LRs calculated from seizure X and from seizure Y is not due to the size of the seizures but to the within-source distribution of seizures. For seizure X distribution of within-source is nearly bimodal and this affects variability of LR values which are lower than for the seizure Y.

A third situation may appear where the median obtained by comparing two seizures is 0.95. In this case, the value may belong to the two primarily selected databases. A semi-parametric estimation such as Kernel’s density [6] is used preferably then. Indeed, for an interval centred on that particular threshold, it allows a reasonable estimation of

![Fig. 9. Distribution of pairs number for combination of seizure X with seizure Y.](image)

Table 2
The \( p, q \) parameter values and LR values for different seizures linked with X

<table>
<thead>
<tr>
<th>Seizure</th>
<th>Number of pairs</th>
<th>( p )</th>
<th>( q )</th>
<th>LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>134</td>
<td>1.089</td>
<td>0.832</td>
<td>22.57</td>
</tr>
<tr>
<td>D2</td>
<td>15</td>
<td>0.505</td>
<td>0.549</td>
<td>50.15</td>
</tr>
<tr>
<td>D3</td>
<td>15</td>
<td>0.710</td>
<td>0.236</td>
<td>214.43</td>
</tr>
<tr>
<td>D4</td>
<td>87</td>
<td>0.874</td>
<td>0.669</td>
<td>41.90</td>
</tr>
</tbody>
</table>
7. Evaluation of the LR performance

The concept of between-source comparisons as a way of investigating the LR technique [13–15] has been used. The first step is to estimate the performance of the method in cases in which the null hypothesis ($H_1$) is true (the two seizures came from the same source).

The results can be plotted as follows: the X-axis is graduated in term of ascending values of likelihood ratio and the Y-axis indicates the number of existing case when a LR value is exceeded. Each chart comprises two curves, the first accounts for the evolution of the likelihood ratio estimated when the assumption $H_1$ is confirmed (the two seizures come from the same seizure) and the second when it is $H_2$ which is confirmed (the two seizures did not come from the same source). Finally, the discriminating power of the technique can be measured by analysing the proportion of equals LR value between the two curves. This is an indication of the performance of the linking of the method.

We have estimated the bootstrap distribution of the LR under the $H_1$ hypothesis in the two cases considered (seizure $X$ or seizure $Y$ at the numerator). Fig. 12 shows the LR distribution under such a condition for the seizure $X$. When seizure $X$ is used for the numerator, in all cases the LR will be greater than 5.48, and in nearly all cases it will exceed 25. In half of the cases it will exceed 50 and in a quarter of cases it will even exceed 75. When the alternative hypothesis is true, the LR bootstrap distribution has been measured with seizure $A$ or with seizure $B$ at numerator.

Fig. 13 shows the LR distribution under $H_2$ hypothesis for the seizure $A$ compare to $B$. With seizure $A$ used for the numerator, in nearly all the cases (98%), the likelihood ratio will be smaller than 0.3. In 66% of cases, it will be smaller than 0.2 and in no case, it will be smaller than 0.05. So it is clear that in these bootstrap simulations based on a 1000 generated samples there is no possibility of a fortuitous match, in the sense that under the $H_2$ hypothesis the LR values are much greater than under the alternative hypothesis.

In another way, the LR values were estimated when the median varies and $p$, $q$ beta parameters remain constant. This corresponds to a situation where a seizure $z$ is compared with $x$ other seizures. Using $LR = \frac{f_{z}(\text{Med}[H_1])}{f_{\text{database}}(\text{Med}[H_2])}$ assuming $H_1$ is true, the seizures are linked and the denominator corresponds to database number 2 (where correlation values are comprised between 95 and 100), then the distribution of LRs from this experiment is summarised in Fig. 14.

In all cases the likelihood ratio will be greater than 24.5, in nearly all cases it will exceed 30. In slightly over 54% of
Fig. 11. Picture of the shape of the distribution of LR calculated with seizure Y at numerator.

Fig. 12. LR distribution under null hypothesis ($H_0$) built with bootstrap distribution of LR, seizure X.
Fig. 13. LR distribution under alternative hypothesis (H₂) built with bootstrap distribution of LR, seizure A.

Fig. 14. LR distribution under null hypothesis (H₀) built with, at numerator, p, q beta parameters of seizure X and median values comprised between 95 and 100. At denominator, p, q beta parameters of database number 2 (95 ≤ C ≤ 100).
In other words, adopting a ‘classical statistical’ terminology, the probability to refuse $H_1$ (null hypothesis) while $H_1$ is true, corresponding to type I error, is null and the probability to accept $H_1$ while $H_1$ is false, corresponding to type II error, is also null. There is no overlapping because scales are completely different. A fortuitous match should never occur and this demonstrates the discriminating power of the method.

8. Validation using real samples

The method based on the measure of the square cosine of the angle between two vectors was validated using more than 5000 heroin samples corresponding to police seizures. It is also applied to the seizures of cocaine (approximately 3000 samples).

The established chemical links were confirmed by results coming from police investigations. Reliable chemical links allows the scientist to obtain a complementary and objective knowledge on the width and nature of the traffic.

9. Conclusion

The cosine function method represents a valuable tool to carry out heroin seizures comparisons and measure links between them. However, the selection of a threshold value to determine if two samples are linked (come from the same source) or not poses a clear problem of continuity in establishing those links.

The distributions of sample pair numbers according to correlation values obtained by this method have been adjusted using beta distributions. Various populations were chosen. The numerical adjustments resulting from these distributions allowed the use of continuous approach based on likelihood ratios. Its continuous extension solves the problem of an arbitrary fixed threshold.

The existence of a within-source variability phenomenon among seizures leads to high variations in the likelihood ratio values. However, description of the results using Tippett plots demonstrates that there is no fortuitous match between the LR values under the $H_1$ hypothesis (seizures are coming from the same source) and alternative hypothesis (seizures are not coming from the same source). These high variations of LR have no real influence for the problem considered in this study. Such a method was validated by using many samples of street.

Acknowledgements

We would like to thank Mrs. Anne Tricot for her technical support.

Appendix A

The bias-corrected percentile method is more appropriate when the statistics of interest (LR distribution called LR) is not normally distributed, because this method adjusts, for the...
bias in \( L_\mathbf{R} \), the mean value of the bootstrap distribution, i.e. the bootstrap estimate of \( L_\mathbf{R} \). With such an approach we obtain the following.

Let \( \hat{z}_0 \equiv \Phi^{-1}\{P(L_\mathbf{R})\} \), where \( \Phi(\cdot) \) denotes the cumulative distribution function of the standard normal distribution, so that \( \Phi(z_0) = 1 - \alpha \). We obtain the cumulative distribution of \( L_\mathbf{R}^*_h \) as:

\[
P(t) \equiv \Phi\{L_\mathbf{R}^*_h < t\} / B;
\]

where \( B \) is the number of bootstrap replications.

We may call \( \hat{z}_0 \) the median bias of \( L_\mathbf{R}^*_h \) (asymmetric\(^1\) distribution of bootstrap replicates) because it is the discrepancy between the median of \( L_\mathbf{R}^*_h \) and \( L_\mathbf{R} \) in normal units. Clearly \( \hat{z}_0 \) is zero if the distribution of \( L_\mathbf{R}^*_h \) is symmetrical. We then express the \( 1 - 2\alpha \) confidence interval CI as:

\[
L_\mathbf{R} \in \left[ \Phi^{-1}\{\Phi(2\hat{z}_0 - z_0)\}, \Phi^{-1}\{\Phi(2\hat{z}_0 + z_0)\} \right].
\]

References


\(^1\)The underlying source of the skewness is in the non-linear nature of the LR statistic.