Generalised Confidence Intervals in Meta Regression

Friedrich, Thomas

Fakultät Statistik, Technische Universität Dortmund, 44221 Dortmund, Germany
friedrich@statistik.tu-dortmund.de

Knapp, Guido

Fakultät Statistik, Technische Universität Dortmund, 44221 Dortmund, Germany
knapp@statistik.tu-dortmund.de

Introduction

The explanation of heterogeneity that occurs when combining different studies is an important issue in meta analysis. Besides including a heterogeneity parameter in the analysis, it is also important to understand the possible causes of heterogeneity. One possibility is to incorporate study-specific covariates in the analysis that account for such between-trial variability. This yields the random effects meta regression model. We examine commonly used tests on the regression coefficients and propose a new method for constructing confidence intervals for the regression coefficients based on principles from generalised inference. The proposed method will be compared by simulation studies with respect to coverage probability and average length.

The random effects meta regression model

Before we start, let us fix some notation. Let \( k \in \mathbb{N} \), \( n = (n_1, \ldots, n_k) \in \mathbb{N}^k \), \( \tau \in \mathbb{R}_{>0} \), \( \sigma = (\sigma_1, \ldots, \sigma_k) \in \mathbb{R}_{>0}^k \) and \( \delta = (\delta_1, \ldots, \delta_k) \in \mathbb{R}_{>0}^k \). For any vector \( c \in \mathbb{R}^k \), let \([c]\) denote the \( k \times k \) matrix having \( c \) on the diagonal, in signs diag\([c]\) = \( c \). If \( c \) is a scalar, i. e. \( c \in \mathbb{R} \), then let \([c]_{k} \) denote the \( k \times k \) matrix having \( c \) on each of its diagonal elements. Hence,

\[
[\tau]_{k} = \begin{pmatrix} \tau & \cdots & \tau \\ \vdots & \ddots & \vdots \\ \tau & \cdots & \tau \end{pmatrix}, \quad [\delta] = \begin{pmatrix} \delta_1 & \cdots & \delta_k \\ \vdots & \ddots & \vdots \\ \delta_1 & \cdots & \delta_k \end{pmatrix}, \quad ([\tau]_{k} + [\delta])^{-1} = \begin{pmatrix} \frac{1}{\tau + \delta_1} & \cdots & \frac{1}{\tau + \delta_k} \\ \vdots & \ddots & \vdots \\ \frac{1}{\tau + \delta_k} & \cdots & \frac{1}{\tau + \delta_1} \end{pmatrix}.
\]

We want to perform a meta analysis on \( k \) studies. We denote the summary response of the \( j \)th study by \( Y_j \). For example, one may think of \( Y_j \) as the mean response of \( n_j \) patients having participated in the \( j \)th study, in signs: \( Y_j = \frac{1}{n_j} \sum_i Y_{ji} \) and \( Y_{ji} \) is the individual response of patient \( i \) in the \( j \)th study. Let \( \theta_j \) denote the expected value and \( \delta_j \) denote the variance of \( Y_j \). The variance \( \delta_j \) will usually be a function of the \( j \)th population variance \( \sigma_j \) and its sample size \( n_j \). For example, in case of \( Y_j \) denoting a mean response, it is \( \delta_j = \frac{\sigma_j^2}{n_j} \). Each study comes with an estimate \( D_j \) of \( \delta_j \). Let \( S_j \) denote the sum of squares of the \( j \)th study, in signs: \( S_j = \sum_i (Y_{ji} - Y_j)^2 \). If \( Y_j \) denotes the mean response, we may define \( D_j := \frac{S_j}{n_j(n_j - 1)} \), since \( E(S_j) = \sigma_j(n_j - 1) \).

Let \( D = (D_1, \ldots, D_k) \), \( S = (S_1, \ldots, S_k) \) and \( Y = (Y_1, \ldots, Y_k) \). Consider the Gaussian-Gaussian hierarchical model \((j = 1, \ldots, k)\):

1. \( Y_j | \theta_j \sim \mathcal{N}(\theta_j, \delta_j) \),
2. \( \theta_j \sim \mathcal{N}(x_j \beta, \tau) \),

where \( x_j \) denotes the \( j \)th row of a \( k \times p \) matrix \( X \) with \( \text{rank}(X) = p < k - 1 \) and \( \beta \in \mathbb{R}^p \). The matrix \( X \) is referred to as the design matrix of the model. The vector \( \beta \) is a vector of parameters and called the vector of regression coefficients. The parameter \( \tau \) stands for the variability between the studies. If we assume that \((Y, \theta_1, \ldots, \theta_k)\) has a multivariate Gaussian distribution, the marginal model for \( Y \)
yields the random effects meta regression model. In matrix notation,

\[(3) \quad Y \sim \mathcal{N}_k (X\beta, [\tau]_k + [\delta]).\]

In particular, \(Y_j \sim \mathcal{N} (x_j \beta, \tau + \delta_j)\) for \(j = 1, \ldots, k\). Since \(\mathcal{N}\) is infinitely divisible, we have

\[(4) \quad \frac{S_j}{\sigma_j} \sim \chi^2_{n_j-1}, \text{ for } j = 1, \ldots, k.\]

Here, \(\chi^2\) denotes the distribution of a chi-square random variable with \(\nu\) degrees of freedom. Let \(y\) be the observed value of \(Y\), let \(s\) be the observed value of \(S\) and \(d\) the observed value of \(D\). From (4) follows

\[(5) \quad D_j = \frac{S_j}{n_j(n_j - 1)} = \frac{\sigma_j}{n_j} \cdot \frac{S_j}{\sigma_j(n_j - 1)} = \frac{\delta_j}{K_j(n_j - 1)}, \quad \text{where } K_j^{-1} \sim \chi^2_{n_j-1}.\]

In this paper, we will look at different constructions of confidence intervals for the regression coefficients \(\beta_i, i = 1, \ldots, p\), in the presence of the nuisance parameters \(\tau\) and \(\delta\). For notational brevity, define \(\Omega_{\tau\delta} = (\tau_k^2 + [\delta])^{-1}\). Then \(\mathcal{V}(Y) = \Omega_{\tau\delta}^{-1}\) and \(Y \sim \mathcal{N}_k (X\beta, \Omega_{\tau\delta}^{-1})\). Let

\[
V_{\tau\delta} = (X'\Omega_{\tau\delta}X)^{-1},
B_{\tau\delta} = V_{\tau\delta}X'\Omega_{\tau\delta},
H_{\tau\delta} = XB_{\tau\delta} = X(X'\Omega_{\tau\delta}X)^{-1}X'\Omega_{\tau\delta},
E_{\tau\delta} = I - H_{\tau\delta} = I - X(X'\Omega_{\tau\delta}X)^{-1}X'\Omega_{\tau\delta}.
\]

Then \(V_{\tau\delta}\) is a \(p \times p\) matrix, \(B_{\tau\delta}\) is a \(p \times k\) matrix and \(H_{\tau\delta}\) and \(E_{\tau\delta}\) are \(k \times k\) matrices. In particular, \(B_{\tau\delta}, H_{\tau\delta}\) and \(E_{\tau\delta}\) are linear operators acting on \(\mathbb{R}^k\), the image space of \(Y\). For simplicity write: \(V = V_{\tau\delta}, B = B_{\tau\delta}, H = H_{\tau\delta}, E = E_{\tau\delta}\) and \(\Omega = \Omega_{\tau\delta}\) whenever \(\tau, \delta\) are equal to the true \(\tau, \delta\) in (3).

Since \(E(BY) = \beta\) and \(\mathbb{V}(BY) = V\), this yields an estimator of \(\beta\), namely \(BY\), with

\[BY \sim \mathcal{N}_p (\beta, V).\]

The observed value of \(BY\) is \(By\). In fact, if we assume that \(\tau, \delta\) were known, \(BY\) is the maximum likelihood estimator of \(\beta\).

Now, let \(b_i = (b_{i1}, \ldots, b_{ik})\) denote the \(i\)th row of \(B\). Then \(b_iy\) is an estimate of \(\beta_i\), since \(b_iY \sim \mathcal{N} (\beta_i, V_{ii})\). Say, \(\tau\) and \(\delta\) are consistent estimators of \(\tau\) and \(\delta\). Let \(\hat{V} = V_{\hat{\tau}\hat{\delta}}, \hat{B} = B_{\hat{\tau}\hat{\delta}}, \hat{H} = H_{\hat{\tau}\hat{\delta}}, \hat{E} = E_{\hat{\tau}\hat{\delta}}\) and \(\hat{\Omega} = \Omega_{\hat{\tau}\hat{\delta}}\). To keep notation consistent, we denote the \(i\)th row of \(B\) by \(\hat{b}_i\). Since \(\tau\) and \(\delta\) are consistent, also \(\hat{b}_i \xrightarrow{P} b_i\) and \(\hat{V}_{ii} \xrightarrow{P} V_{ii}\) and, hence, \(\frac{\hat{b}_i Y - b_i}{\sqrt{V_{ii}}} \sim \mathcal{N}(0, 1)\), which yields the approximate \((1 - \gamma)\)-confidence interval for \(\beta_i\) and fixed \(i = 1, \ldots, p\)

\[(6) \quad \left[\hat{b}_i Y + z_\gamma \cdot \sqrt{V_{ii}}, \hat{b}_i Y + z_{1-\gamma} \cdot \sqrt{V_{ii}}\right]\]

for any \(\gamma \in (0, 1]\). Here, \(z_\gamma\) denotes the \(\gamma\)-quantile of the standard Gaussian distribution \(\mathcal{N}(0, 1)\).

The confidence interval in (6) is based on approximations of the nuisance parameters \(\tau\) and \(\delta\). If we can assume that \(D\) is consistent for \(\delta\), we simply use \(\hat{\delta} := D\). In the case of \(\tau\), we are spoilt for choice. For an extensive list of possible estimators and a study of their performances see [5].

**Generalised confidence intervals for the regression coefficients**

The concepts of generalised inference were developed by Weerahandi, see [9]. His interest was in constructing exact confidence sets for all parameters of interest, in our case \(\beta\), in the presence of
nuisance parameters, here $\tau$ and $\delta$. To construct such exact confidence sets, it is necessary to find a (generalised) pivotal quantity for $\beta$ which distribution is free of the nuisance parameters $\tau$ and $\delta$.

First, though, we will need to construct pivotal quantities for $\tau$ and $\delta$.

A pivotal quantity for $\delta$ is quickly constructed. Recall that under model (3) the estimator $D_j$ in each study is distributed via (5). Let

$$
(7) \quad \tilde{D}_j := \delta_j \cdot \frac{d_j}{D_j} = \delta_j \cdot \frac{K_j(n_j - 1)}{\delta_j} = d_j(n_j - 1)K_j, \quad \text{where } K_j^{-1} \sim \chi^2_{n_j - 1}.
$$

The distribution of $\tilde{D}_j$ is independent of unknowns. The observed value of $\tilde{D}_j$ is $\tilde{d}_j = \delta_j$ which is monotone in $\delta_j$. Hence, $\tilde{D}_j$ is a pivotal quantity for $\delta_j$.

Unfortunately, the construction of a pivotal quantity for $\tau$ is not as straight forward as for $\delta$. The following construction is based on a quadratic form used to construct confidence intervals for monotone in $\delta$.

Since $E\tau_j$ is monotone decreasing in $\delta$, let $\tilde{\eta}$ be a pivotal quantity for $\tau_j$. Note that for all $\delta_j$, we want to consider a generalised pivotal quantity from this quadratic.

$$
(8) \quad Q_\delta(\tau) := (E\tau Y')\Omega_{\tau\delta}(E\tau Y) = Y' (E\tau\delta\Omega_{\tau\delta}E\tau Y) Y = \left\|Y(E\tau Y)^{-\frac{1}{2}} \cdot E\tau Y\right\|^2.
$$

Let $q_\delta$ denote the observed version of $Q_\delta$. That is $q_\delta(\tau) = y'(E\tau\delta\Omega_{\tau\delta}E\tau Y)$. As the above notation already suggests, we want to consider $Q_\delta$ as a function in $\tau$. Recall that $\Omega_{\tau\delta}$ denotes a diagonal matrix. Note that for all $j = 1, \ldots, k$ each $(\text{diag } \Omega_{\tau\delta})_j$ is strictly monotone decreasing in $\tau$. Moreover, $\text{diag } \Omega_{\tau\delta}$ converges uniformly to $0$, in signs: $\lim_{\tau} \|\text{diag } \Omega_{\tau\delta}\|_\infty = 0$. Hence, $Q_\delta(\tau)$ is strictly monotone decreasing in $\tau$ with $\lim_{\tau} Q_\delta(\tau) = 0$. This enables us to define an inverse function to $Q_\delta$, namely

$$
P_\delta(\eta) = \begin{cases} Q_\delta^{-1}(\eta) : & 0 < \eta < Q_\delta(0) \\ 0 : & \text{otherwise} \end{cases} \quad \quad p_\delta(\eta) = \begin{cases} q_\delta^{-1}(\eta) : & 0 < \eta < q_\delta(0) \\ 0 : & \text{otherwise} \end{cases}
$$

for $\eta > 0$. The observed value of $P_\delta(\eta)$ is $p_\delta(\eta)$. Let us now investigate the distribution of $Q_\delta(\tau)$. Since $E = E\tau\delta$ is a linear operator, $EY$ is multivariate Gaussian distributed with $E(Y) = 0$ and $\mathbb{V}(EY) = \Omega - X'(X\Omega X)^{-1}X'$. Hence, $\text{diag } \Omega_{\tau\delta}$. We define $Q_\delta(\tau)$ is strictly monotone decreasing in $\tau$ with $\lim_{\tau} Q_\delta(\tau) = 0$. This enables us to define an inverse function to $Q_\delta$, namely

$$
(9) \quad L_i := \tilde{b}_iy - \frac{b_iY - \beta_i}{\sqrt{V_{ii}}} = \tilde{b}_iy - N \cdot \sqrt{X'(Q_k^{-1})_i + (K_k^{-1})_i X^{-1}}
$$

where $K = (K_1, \ldots, K_p)$, $K_j^{-1} \sim \chi^2_{n_j - 1}$, $Q \sim \chi^2_{k-p}$ and $N \sim \mathcal{N}(0, 1)$. Thus, the distribution of $L_i$ only depends on the observed data and is free of the nuisance parameters $\tau$ and $\delta$. The observed value of $L_i$ is $\beta_i$ which is monotone in $\beta_i$. Hence, $L_i$ is a generalised pivotal quantity for $\beta_i$. Let $L_i$ denote the distribution of $L_i$ and $I_{i, \gamma}$ its $\gamma$-quantile for any $\gamma \in (0, 1]$. Then for any fixed $\gamma \in (0, 1]$ and fixed

$$
(10) \quad [l_{i, \gamma}, l_{i, 1-\gamma}]
$$
is an exact \((1 − \gamma)\)-confidence interval for \(\beta_i\).

The question remains how to obtain the necessary quantiles \(l_{i,\gamma}\). For this, we generate random variables \(K_1, \ldots, K_k, Q\) and \(N\) with \(K_j^{-1} \sim \chi^2_{n_j-1}\) and \(Q \sim \chi^2_{k-p}\) and \(N \sim \mathcal{N}(0, I)\) and define \(L_i\) such as in (9). Then \(L_i\) has the distribution \(\mathcal{L}_i\). Let \((L_{il})_{l \in \mathbb{N}}\) be a sequence of independent draws from \(\mathcal{L}_i\) and let \(l_{i,\gamma,m}\) denote the empirical \(\gamma\)-quantile of \((L_{il})_{l \leq m}\). Then \(l_{i,\gamma,m} \xrightarrow{P} l_{i,\gamma}\) for \(m \to \infty\).

So, instead of obtaining approximate confidence intervals for \(\beta_i\) such as in (6), we obtain exact confidence interval by approximating the quantiles \(l_{i,\gamma}\). By doing so, we have shifted the approximation problem from a statistical one – that depends on sample size – to a numerical one – that may be solved up to any required precision.

**Performance study of the confidence intervals based on a real world example**

We studied the performance of the above confidence intervals with respect to coverage probability and average length. For comparison, we turned to a data set that has already been discussed in the literature, e.g. in [1, 5]. This data set is indeed quite predestined for such an analysis. It consists of a combination of 13 clinical trials which evaluated the efficacy of the Bacillus Calmette-Guérin vaccine for the prevention of tuberculosis. The distance of a trial to the equator may serve as a potential influential covariate here. One can think of this distance as a surrogate for the presence of environmental mycobacteria that provide a certain level of natural immunity against tuberculosis.

The data set is put together in Table 1. It can also be found in the metafor package of the statistical software environment R, see [7, 8]. Table 1 and Figure 1 show how unbalanced the data are when it comes to study size. Study sizes range from a minimum of 262 in trial A to 176782 in trial H. But the clinical trials not only differ in total sample size. Within each study the proportion of vaccinated and non-vaccinated subjects is quite different. Whereas most trials are relatively balanced between the two groups, Figure 1 shows that in some trials the vaccinated group is considerably larger.
than the non-vaccinated, see trials K and G for example. All this makes the data difficult to access for methods relying on good statistical approximations.

Following [1] and [5], we use the logarithm of the relative risk as outcome measure in the analysis. The model of choice is (3) with one covariate. The parameters of interest are the regression coefficients \( \beta_0, \beta_1 \) for the intercept and slope of the regression line respectively. In the analysis and in the simulation study we choose the restricted maximum-likelihood estimator \( \hat{\tau} \) for \( \tau \), see [6]. Besides the above confidence intervals (6) and (10), we have also included an adjustment to (6) by Knapp and Hartung in our analysis, see [5]. We used the statistical software environment R for analysis and simulation study, see [7].

Table 2: Different 95% confidence intervals for \( \beta_0 \) and \( \beta_1 \) respectively. The exact confidence intervals are based on 1000 draws from \( L_i \), \( i = 1, 2 \) respectively.

<table>
<thead>
<tr>
<th>for ( \beta_0 )</th>
<th>lower bound</th>
<th>upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approximate</td>
<td>-0.24</td>
<td>0.74</td>
</tr>
<tr>
<td>Adjusted</td>
<td>-0.37</td>
<td>0.88</td>
</tr>
<tr>
<td>Exact</td>
<td>-0.77</td>
<td>0.98</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>for ( \beta_1 )</th>
<th>lower bound</th>
<th>upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approximate</td>
<td>-0.04</td>
<td>-0.01</td>
</tr>
<tr>
<td>Adjusted</td>
<td>-0.05</td>
<td>-0.01</td>
</tr>
<tr>
<td>Exact</td>
<td>-0.05</td>
<td>-0.00</td>
</tr>
</tbody>
</table>

Figure 2: Different 95% confidence intervals for \( \beta_0 \) and \( \beta_1 \) respectively. The dotted lines correspond to the point estimates \( \hat{b}_0 \) and \( \hat{b}_1 \) respectively. The exact confidence interval is based on 1000 draws from \( L_i \), \( i = 1, 2 \).

Table 3: Performance of 95% confidence intervals for \( \beta_0, \beta_1 \) with respect to coverage probability and expected length based on 1000 random draws of \( (Y, D) \) via (3) and (5). The generalised confidence intervals for each \( (Y, D) \) were based on 120 draws from \( L_i \), \( i = 0, 1 \) respectively.

<table>
<thead>
<tr>
<th>( \beta_0 )</th>
<th>Approximate</th>
<th>Adjusted</th>
<th>Exact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>0.8453</td>
<td>0.9386</td>
<td>1.3840</td>
</tr>
<tr>
<td>Covers</td>
<td>0.8720</td>
<td>0.8990</td>
<td>0.9640</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \beta_1 )</th>
<th>Approximate</th>
<th>Adjusted</th>
<th>Exact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>0.0246</td>
<td>0.0273</td>
<td>0.0382</td>
</tr>
<tr>
<td>Covers</td>
<td>0.9020</td>
<td>0.9150</td>
<td>0.9580</td>
</tr>
</tbody>
</table>

Applying the methods to the data in Table 1 yield the results in Table 2 and Figure 2. The confidence intervals based on stochastic approximation are shorter than the exact one. However, as we shall see in the simulation study, these intervals are in fact underestimating the variability of \( \hat{\beta} \) and the two confidence intervals based on (6) claim way more confidence than they can actually guarantee.

For the simulation, we fixed a set of parameters \( \beta \) and \( \delta \) and generated responses \( Y \) and \( D \) from (3) and (5) respectively. We choose \( \beta = \hat{\beta} \) and \( \delta = \hat{\delta} \), where \( \hat{\beta} \) and \( \hat{\delta} \) correspond to the point
estimates based on the data set from Table 1. For each simulated pair \((Y, D)\) we logged the lengths of each of the 95\% confidence intervals and whether or not the interval was covering the true parameter \(\beta\). The results of the simulation are put together in Table 3.

As we can see, the mean length of the approximate and the adjusted confidence intervals are smaller than the exact ones. However, the cost of their short mean length bears their coverage probability. The approximate confidence interval as defined in (6) is underestimating the true variability of \(\hat{\beta}\), which leads to a short but poor performing confidence interval. The adjustment introduced in [5] accounts for the variability of \(\hat{\tau}\) by introducing a correction factor to its estimated standard derivation. This makes the expected length of the confidence interval larger but brings its coverage probability closer to the requested significance bound. In both cases, though, the approximate confidence intervals do not fulfil the requested confidence bound. The exact confidence intervals, we have developed in this paper, are in fact the only ones to adhere the requested confidence bounds.

Conclusion

We have developed a new method for constructing confidence intervals for the regression coefficients in the random effects meta regression model. The method we have developed here is based on principles from generalised inference. In particular, the proposed intervals are exact. In simulation studies, we were able to show that the coverage probability of our proposed intervals not only meet predefined confidence bounds but also outperform classic approaches based on stochastic approximation.

References


