Rescaling Process by Using the Constrained $L_1$ Norm Method

M. A. K. Al-Mansoob$^1$

Abstract:

The constrained $L_1$ norm method (CL1) was used to rescale ordered variables in a regression problem with a continuous dependent variable. The rescaling process has lead to new scoring systems to the ordered variables that effectively improved the estimation accuracy. The rescaling process may also be adopted for variable selection.

---

$^1$Mathematics Department, Faculty of Science, Sana'a University
1. Introduction

For a data set with a quantitative dependent variable and ordered predictor variables, the following model has often been applied:

\[ y_i = \beta_{0i} + \beta_{1i} \sum_j x_{ij} + \varepsilon_i, \quad i = 1,2,\ldots,n \]  

(1)

Where \( y_i \) is the value of the dependent variable in the \( i \)th individual, \( \beta_{0i} \) and \( \beta_{1i} \) are parameters to be estimated, \( x_{ij} \) is the score of the \( j \)th independent variable on the \( i \)th individual, so that \( \sum_j x_{ij} \) is the total score for the \( i \)th individual and \( \varepsilon_i \) is random error term with mean \( E(\varepsilon_i) = 0 \) [1,2,3]. Often \( \{\varepsilon_i\} \) is assumed to be a sequence of iid \( N(0,\sigma^2) \) random variables. Usually the scores are non-negative integers: sometimes the scores are given the values 0,1,2,3,... .

Model (1) might be inappropriate for any of several reasons. If there are outliers, the normality assumption might be grossly violated, and so the usual least square approach is inappropriate. Here we consider minimum absolute deviations and so give less weight to outliers. This provides maximum likelihood estimates if the error term \( \varepsilon_i \) is assumed to follow a double exponential distribution. Furthermore, the scores might need reallocation subject to certain properties: for our envisaged uses, the scores should be positive, integers and ordered within each independent
variable. Later we will see how such a scoring system can be derived and how the system can be used to reduce the number of independent variables in the estimation procedure. We show how to apply a constrained least absolute deviation algorithm, CLI in the nomenclature adopted.

2. Rescaling Process

The rescaling or the rescaling process starts by recoding the independent variables in model (1). Let

\[ y_i = \alpha_0 + \sum_{j,k} \alpha_{jk} v_{jk} + \varepsilon_i , \quad i = 1,2,\ldots,n \]  

(2)

Where \( y_i \) is the ith value of the dependent variable, \( \alpha_0 \) and \( \alpha_{j,k} \) for all pairs \((i, j)\) are parameters, and \( x_{ij} \) is the jth independent variable on the ith individual.

\[ v_{jk} = \begin{cases} 
1 & \text{if } x_{ij} = k \\
0 & \text{if } x_{ij} \neq k, \text{ and } \alpha_{j0} = 0 \forall j
\end{cases} \]

The new design matrix \( V \) contains the variables that are the components of each characteristic: these new variables \( v_{jk} \) are of binary form and the arrangement for the ith observation is considered as a vector \( v_i \). The next step is to use a suitable multiple regression process with all the new independent variables. These parameter estimates form the new scoring system. Although collinearity or multicollinearity can arise among the new
independent variables, available algorithms do not require the
design matrix to be of full rank. In addition [1,2], the collinearity or
the multicollinearity can be adjusted in the definition of the
variables themselves: for example, the score of the smallest ordered
value of each variable should be set to zero.

2.1 CL1 Formulation

By considering model (2) in matrix notation, the $L_1$ norm
minimizes

$$\left\| y - V \alpha \right\|_1 = \sum_i |y_i - v_i \alpha|$$

(3)

The CL1 algorithm of Barrodale and Roberts [4,5] minimizes (3)
subject to:

$$A \alpha = B$$

(4)

$$C \alpha = D$$

Where $A$ and $C$ are matrices of dimensions $l \times m$ and $k \times m$
respectively. $B$ and $D$ are columns vectors of dimensions $l$ and $k$, respectively. Matrix $C$ accommodates the fact that the scores are
ordered in specific ways that reflect the physical process. Barrodale
and Roberts have mentioned that matrices $V, A$ and $C$ are not
required to be of full rank [4]. This property is important because of
the likely multicollinearity among the independent variables.

Barrodale and Roberts [4] reformulated the problem represented by
(3) and (4) as a linear program:

$$Z = \min \sum_i |e_i^+ + e_i^-| = \min (e_i^+ + e_i^-), \ i = 1, 2, ..., n$$

---

Rescaling Process by Using the Constrained $L_1$
Norm Method
subject to

\[ Y = V(\hat{\alpha}^+ - \hat{\alpha}^-) + \varepsilon^+ - \varepsilon^- \]
\[ H = A(\hat{\alpha}^+ - \hat{\alpha}^-) \]
\[ D = C(\hat{\alpha}^+ - \hat{\alpha}^-) + \varepsilon^{''+} \]
\[ \hat{\alpha}^+, \hat{\alpha}^- \geq 0, \varepsilon^+, \varepsilon^- \geq 0, \varepsilon^{''+} \geq 0 \]

Where \( \varepsilon^+ \) and \( \varepsilon^- \) are \( n \times 1 \) column vectors and \( \varepsilon^{''+} \) is an \( m \times 1 \) vector.

The algorithm uses the two-phase simplex methods (more detailed about these methods can be seen on Barrodale and Roberts and Taha [4,6]). In a particular problem we faced, the parameter estimates had to be ordered as follows:

For matrix \( V \), corresponding to 1\(^{st} \) variable, \( 0 \leq \hat{\alpha}_1 \leq \hat{\alpha}_2 \), and for the remaining variables

\[ V_2 \rightarrow 0 \leq \hat{\alpha}_3 \leq \hat{\alpha}_4 \leq \hat{\alpha}_5 \leq \hat{\alpha}_6 \]
\[ V_3 \rightarrow 0 \leq \hat{\alpha}_7 \leq \hat{\alpha}_8 \leq \hat{\alpha}_9 \]
\[ V_4 \rightarrow 0 \leq \hat{\alpha}_{10} \leq \hat{\alpha}_{11} \leq \hat{\alpha}_{12} \leq \hat{\alpha}_{13} \]

M

---

Rescaling Process by Using the Constrained L\(_1\) Norm Method
The matrix $C$ for this structure takes the form:

$$
\begin{bmatrix}
-1 & 1 \\
1 & -1 & 1 \\
1 & 1 & -1 & 1 \\
& 1 & -1 & 1 & 1 \\
& & 1 & -1 & 1 & 1 \\
& & & 1 & -1 & 1 & 1 \\
& & & & 1 & -1 & 1 & 1 \\
& & & & & 1 & -1 & 1 & 1 \\
& & & & & & 1 & -1 & 1 & 1 \\
& & & & & & & 1 & -1 & 1 & 1 \\
\end{bmatrix}
$$

To use the published algorithm, $C$ is placed under the design matrix $V$ to form a composite matrix. The imposition of these constraints ensures the positiveness and the order within each of the original variables. By equating $A$ to 0, the equality constraints are unnecessary.
2.2 Application

The gestational age (G.A) of a newborn baby is often estimated by observing some physical characteristics, and some characteristics are shown in Table 1. Dubowitz et al (1970) assigned scores to each characteristic so that the higher the score the greater the maturity where, of course, that these characteristics are continuous [7]. Estimating the gestational age then is often done by regressing the total sum (T.S.) of the characteristics scores against the actual gestational ages [8, 9, 10, 11].

The normality assumption to the error term in model (1) appeared to be violated when the least square method was used for two data sets: the first for 200 Sudanese babies and the second for 144 British babies [11].

3. Results and Discussion

Model (2) is fitted to both the Sudanese and the British data (the final characteristic is omitted from the Sudanese study) by using the constrained $L_1 (C11)$ Algorithm of Barrodale and Roberts (1980) [5]. The parameter estimates (Tables 2) could be used for the new scoring systems. Once the new scoring systems have been constructed, the total sum of the scores of the remaining variables is recalculated in model (1) and $\beta_0$ and $\beta_i$ are estimated by minimizing the absolute deviations. Those variables in Tables 2 with parameter estimates zero are omitted. While no integer constraint were imposed initially, the parameter estimates in Table 2 are seen
to be integers which is a desirable property given the likely use of the results. The estimates are integers because the data have been transformed into binary form. These integer-valued estimates led to consider the problem as an exercise in integer programming. In integer programming the problem can be reformulated as follows:

\[ Z = \min |e| \]

subject to

\[ Y - V \hat{a} = e \]
\[ C \hat{a} \leq 0 \]
\[ Y^T V, C \geq 0, \hat{a} \geq 0 \]

with integer elements

The Branch and Bound method was applied, but no improvements in computational time achieved. C1 analysis results are summarized and compared in Table 3. The original and the new scoring systems are used to estimate the gestational age of model (1). In C1 the means of the sum absolute deviations (MSAD) are given for comparison: the procedure reduces the values of MSAD for both data sets and there is less variability about the line in the MSAD sense for the British data then for the Sudanese data.
Conclusion

The scores of ordered variables are usually to distinguish the progressing level of the variables. In analysing data with ordered variables, trivial questions may arise "Is the utilized scoring system efficient?" And if not "Are there any possible ways for rescaling or rescoring the variables?". The proposed method for rescaling ordered variables in this paper is mainly a regression procedure. However, if data of the ordered variables that to be rescaled are showing some erroneous values (outliers), using the ordinary least square method is inappropriate. In a previous work, Al-Mansoob (1998) has found that some outliers are present in the two sets of the data; the Sudanese and the British [11]. Hence, it was necessary to seek other regression procedures that give less weight to the outliers effect; $L_1$ norm method was used here. Since the rescaled values should be positive and ordered within each variable, the algorithm of Barrodale and Roberts (1980), that based on $L_1$ norm procedure, was found suitable [5]. Applying this algorithm on the two sets of data has produced ordered, positive and unexplainable integer values new scoring systems to some variables and eliminates some others. When the remaining variables have been reallocated with the derived scores, the estimation accuracy has improved. Therefore, the proposed rescoring procedure is worthy to be tested and evaluated.
References


Table 1: Dubowitz, Dubowitz and Goldberg Original Scoring System

<table>
<thead>
<tr>
<th>Variables</th>
<th>Original Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oedema</td>
<td>$x_1$</td>
</tr>
<tr>
<td>Skin Texture</td>
<td>$x_2$</td>
</tr>
<tr>
<td>Skin Color</td>
<td>$x_3$</td>
</tr>
<tr>
<td>Skin Opacity</td>
<td>$x_4$</td>
</tr>
<tr>
<td>Lanugo (Over Back)</td>
<td>$x_5$</td>
</tr>
<tr>
<td>Plantar Creases</td>
<td>$x_6$</td>
</tr>
<tr>
<td>Nipple Formation</td>
<td>$x_7$</td>
</tr>
<tr>
<td>Breast Size</td>
<td>$x_8$</td>
</tr>
<tr>
<td>Ear Formation</td>
<td>$x_9$</td>
</tr>
<tr>
<td>Ear Firmness</td>
<td>$x_{10}$</td>
</tr>
<tr>
<td>Genitalia</td>
<td>$x_{11}$</td>
</tr>
<tr>
<td>Skull Hardness</td>
<td>$x_{12}$</td>
</tr>
</tbody>
</table>
Tables 2: The New Scoring Systems under CL1 Approximations

2.1 The Sudanese Babies

<table>
<thead>
<tr>
<th>Variables</th>
<th>New Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oedema $x_1$</td>
<td>0 21 23</td>
</tr>
<tr>
<td>Skin Texture $x_2$</td>
<td>0 5 11 11 11</td>
</tr>
<tr>
<td>Skin Color $x_3$</td>
<td>0 13 14 14</td>
</tr>
<tr>
<td>Skin Opacity $x_4$</td>
<td>0 0 1 1 1</td>
</tr>
<tr>
<td>Lanugo (Over Back) $x_5$</td>
<td>0 0 0 0 0 0</td>
</tr>
<tr>
<td>Plantar Creases $x_6$</td>
<td>0 0 0 0 0 1</td>
</tr>
<tr>
<td>Nipple Formation $x_7$</td>
<td>0 0 0 0</td>
</tr>
<tr>
<td>Breast Size $x_8$</td>
<td>0 13 23 26</td>
</tr>
<tr>
<td>Ear Formation $x_9$</td>
<td>0 0 0 1</td>
</tr>
<tr>
<td>Ear Firmness $x_{10}$</td>
<td>0 0 31 33</td>
</tr>
<tr>
<td>Genitalia $x_{11}$</td>
<td>0 1 2</td>
</tr>
</tbody>
</table>

Rescaling Process by Using the Constrained L1 Norm Method
2.1 The British Babies

<table>
<thead>
<tr>
<th>Variables</th>
<th>New Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oedema $x_1$</td>
<td>0 0 0</td>
</tr>
<tr>
<td>Skin Texture $x_2$</td>
<td>0 0 0 3 18</td>
</tr>
<tr>
<td>Skin Color $x_3$</td>
<td>0 0 9 11</td>
</tr>
<tr>
<td>Skin Opacity $x_4$</td>
<td>0 0 6 6 7</td>
</tr>
<tr>
<td>Lanugo (Over Back) $x_5$</td>
<td>0 0 2 4 4</td>
</tr>
<tr>
<td>Plantar Creases $x_6$</td>
<td>0 0 0 1 1</td>
</tr>
<tr>
<td>Nipple Formation $x_7$</td>
<td>0 0 0 0</td>
</tr>
<tr>
<td>Breast Size $x_8$</td>
<td>0 0 8 8</td>
</tr>
<tr>
<td>Ear Formation $x_9$</td>
<td>0 3 3 3</td>
</tr>
<tr>
<td>Ear Firmness $x_{10}$</td>
<td>0 0 0 1</td>
</tr>
<tr>
<td>Genitalia $x_{11}$</td>
<td>0 0 3</td>
</tr>
<tr>
<td>Skull Hardness $x_{12}$</td>
<td>0 0 21 21 21</td>
</tr>
</tbody>
</table>

Table 3: Fitting the Data According to Model (1) after Reallocating the Scores for CL1 Analysis

<table>
<thead>
<tr>
<th>Scoring System Used</th>
<th>Country</th>
<th>No. of Variables</th>
<th>Equation Used to Estimate G.A.</th>
<th>MSAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>Sudan</td>
<td>11</td>
<td>$215.57 + 2.21 \times T.S.$</td>
<td>12.08</td>
</tr>
<tr>
<td>Original</td>
<td>Britain</td>
<td>12</td>
<td>$248.67 + 1.20 \times T.S.$</td>
<td>07.19</td>
</tr>
<tr>
<td>New</td>
<td>Sudan</td>
<td>9</td>
<td>$164.00 + 1.00 \times T.S.$</td>
<td>09.84</td>
</tr>
<tr>
<td>New</td>
<td>Britain</td>
<td>10</td>
<td>$225.00 + 1.00 \times T.S.$</td>
<td>06.00</td>
</tr>
</tbody>
</table>

Rescaling Process by Using the Constrained L1 Norm Method
عملية إعادة تدرج متغيرات مربحة باستخدام طريقة الانحرافات المطلقة المقدرة

محمود المنصوب

خلاصة:

لتقييم متغير كمياً متصل بناءً على مجموع فيس متغيرات مربحة مستقلة
فقد لا يكون مرهج الادخار المحتملي البسيط مناسبًا لعدة أسباب أهمها وجود فيس شائدة
(Residuals) في البيانات، والذي يعني إنشاء شروط التوزيع الطبيعي للبيانات. الأمر الذي هو أحد أهم الشروط الأساسية لاستخدام طريقة المربعات الصغرى أو أقل الربعات في التقدير. ونلاحظ
مثل هذا الوضع هناك طرق تقييم مختلفة تقليل تأثير القيم الشاهدة أحتر من شاهدة أفراد الانحرافات
LAD (Least Absolute Deviation) أو L1-Norm المطلقة والمصطلح عليها
MAD (Minimum Absolute Deviation) أو LAV (Least Absolute Variation) أو LAD (Least Absolute Deviation)
والذي يشيع استخدامها في الكثير من مجالات التحليل العددي

شفاف مجموعة الكلي المتغيرات المتصلة المستقطبة بما يأتي من أن نظام التدرج في هذه المتغيرات غير مناسبًا
والمثال على ذلك الأمر كذلك فلبيًا العام والكتابة أن تدرج أفضل هذه المتغيرات مع الاحتفاظ بنفس خصائص
الدرج خان القطرة مما، ويجري كذلك مرتبة داخل كل معامل باستخدام طريقة أقل الانحرافات المتقلة
(Constrained L1-Norm Method) أو اعتماداً على أن تكون المنحنية تدرج جيدة لمتغيرات مستقلة
مرتبة من خلال أنظمة التدرج المتقدمة حيث عند استخدامها -الجديدة على نفس البيانات أمكن الحصول على نتائج
تقدير أكثر دقة

عملية إعادة تدرج هذه المعرفة أثرت على طريقة بسيطة لتحديد عدد المتغيرات المستقلة وهي خاصة مزعج فيها
خاصة عندما يكون عددها كبيرًا وكذلك إذا ما ساعد في زيادة دقة التقدير وهذا ما حدث في المثال المصاحب.

كلية العلوم
جامعة صنعاء

Abstract:

الخلاصة: