Bayesian credible intervals for monitoring liquid blending rates

Dewi Rahardja*, Yan D. Zhao and Xian-Jin Xie
Department of Clinical Sciences and Simmons Cancer Center, UT Southwestern Medical Center, Dallas, TX, USA

Received 28 December 2009
Revised 16 August 2010
Accepted 31 August 2010

Abstract. We consider the problem of constructing confidence intervals (CIs) for the blending coefficient of different liquid, such as the blended underground storage tank (UST) leak data for compliance. For this problem, confidence intervals based on Fieller’s Method have been proposed. This method utilizes a blending coefficient estimator which is a ratio of two correlated normal random variables. However, this method assumes normally distributed random errors in the UST leak model and therefore may be inappropriate for the UST leak data which typically have heavy-tailed empirical distributions. In this paper we develop a Bayesian approach assuming non-normal random errors with the Power Exponential Distribution (PED). A real-data example using Cary blended site data is given to illustrate both the Fieller’s CIs and the Bayesian credible intervals. Monte Carlo simulations are conducted to compare the coverage probability and average width of CIs for both methods. For data with heavy-tailed distributions, the simulations show that both Fieller’s and Bayesian intervals perform adequately in terms of coverage. However, Bayesian intervals perform better in terms of yielding CIs with shorter expected width.

Keywords: Bayesian credible interval, confidence interval, Fieller’s method, liquid blending rate, power exponential distribution

1. Introduction

In many modern service stations, an intermediate grade of gasoline is formed by blending regular unleaded gasoline having an (R + M)/2 rating of 87 with super unleaded having an (R + M)/2 rating of 92 or 93. The latter blend is often used in high performance vehicles due to their need for fuel mixtures with richer octane ratings. As presented in Gill et al. [6] and Gill and Keating [5], the retail gasoline market has ventured into blended sites to reduce the number of on-site underground storage tanks (USTs) and thereby reduce the number (see Fig. 1) that may potentially leak and create environmental problems such as those as encountered in the catastrophic spills in the Charnock Basin, Santa Monica, California.

As mentioned in Gill and Keating [5], as retail-marketers move into blended sites, the blending coefficient looms as a critical factor in determining the integrity of the advertised intermediate blend and finding small leaks through the blender. They still provide three grades of gasoline but the intermediate (or plus) grade is provided by blending the unleaded with the super blend. The blender produces gasoline with ratings between 87 (unleaded) and 92 (super). For example, a 60:40 blend will produce an intermediate grade of 89 (plus). The blending coefficient, $\gamma$, is therefore supposed to be preset to a certain designated amount such as 60%. However, in reality, the true blending coefficient in a particular blending site tends to be off slightly from the designated amount, and thus precise estimation methods are needed for estimating the true blending coefficient and for determining compliance.

*Corresponding author. E-mail: rahardja@gmail.com.
Several methods have been proposed for estimating the blending coefficient in the blended UST leak model. Keating and Mason [7] expanded linear regression methodology to cover blended and manifolded USTs. Gill et al. [6] proposed two regression equations which contain common and different predictors and parameters. Gill and Keating [5] considered the problem of obtaining a confidence interval for the blending coefficient in the blended underground storage tank (UST) leak model with the blending coefficient estimator being the ratio of two correlated normal random variables through Fieller’s method. Their method is applied to the Cary blended site data, and several variations are included to account for outliers (heavy-tailed real data).

Typically UST leak data have an empirical distribution function which is bell-shaped and symmetric but with outliers [5]. Such distributions may not be approximated well by using the normal distribution. Therefore, the Fieller’s method applied by Gill and Keating based on normality assumptions on error terms will likely produce suboptimal CIs. In this article, we propose a Bayesian approach with non-normal random errors through the Power Exponential Distribution (PED). We choose PED because such distribution adequately captures the empirical distribution function of UST leak data. Mathematical forms and examples of distribution functions of PED are presented in Section 2.2.

The remainder of this article is organized as follows. In Section 2 we introduce the blended UST leak model, review the confidence intervals by Fieller’s method [5], and describe our new Bayesian method. In Section 3 we apply the Fieller’s method and the Bayesian method for construction CIs for the blending coefficient using the Cary data described in Gill and Keating [5]. We conducted simulation studies to compare the Fieller’s method and the Bayesian method in terms of the coverage probability and average interval width in Section 4. Some concluding remarks are provided in Section 5.

2. Blended UST leak model and methods

2.1. Normal random error model and Fieller’s method

Previously, Gill and Keating [5] considered confidence interval estimation concerning the blending coefficient parameter $\gamma$ in the blended UST model. This model was first considered in Keating and Mason [7] and subsequently in Gill et al. [6]. The blended UST leak model is given by the following linear equations:

$$ Y_{1j} = \beta_{01} + \beta_1 X_{1j} + \gamma \beta_2 X_{2j} + \epsilon_{1j} $$

$$ Y_{2j} = \beta_{02} + (1 - \gamma) \beta_2 X_{2j} + \beta_3 X_{3j} + \epsilon_{2j} $$

(1)
for \( j = 1, \ldots, n \), where \( Y_{ij} \) is the total volume in gallons dispensed from tank \( i \), \( X_{ij} \) is the volume of gallons dispensed through meter \( i \), and \( \varepsilon_{ij} \) is normal random error associated with the reading for tank \( i \) at the \( j \)th time period. The intercept parameters \( \beta_01 \) and \( \beta_02 \) represent the leak rates based on the sampling rate, and the slope coefficients, \( \beta_1, \beta_2, \) and \( \beta_3 \), are the meter calibrations of the regular unleaded, intermediate, and super unleaded flowmeters (refer to Gill et al. [6, p. 729]). If data are collected daily, intercepts represent daily leak rates whereas they are hourly leak rates if data are collected hourly. Because there are two response variables \( Y_1 \) and \( Y_2 \), this is the two-regression problem with the issue that both linear equations contain some common and some different predictors and parameters. Define the operator \( a^p = [a_2, a_1, \ldots, a_{2d}, a_{2d-1}] \) for a \((2d)\)-dimensional vector \( a = [a_1, a_2, \ldots, a_{2d-1}, a_{2d}] \). These equations can be more concisely expressed as:

\[
Y = X\beta + \varepsilon
\]

where \( Y \) is a \( 2n \times 1 \) vector such that \( Y^T = [Y_{11}, Y_{21}, \ldots, Y_{1n}, Y_{2n}] \), \( X \) is a \( 2n \times 6 \) matrix such that \( X = [1 \ 1^* \ X_1 \ X_2^* X_2^* X_3^*] \), \( X_i^T = [X_{i1}, 0, \ldots, X_{in}, 0] \), \( 1^T = [1, 0, \ldots, 1, 0] \), \( \beta^T = [\beta_{01}, \beta_{02}, \beta_1, \beta_2, \beta_3] \) and \( \varepsilon^T = [\varepsilon_{11}, \varepsilon_{21}, \ldots, \varepsilon_{1n}, \varepsilon_{2n}] \) is assumed to follow a multivariate normal distribution with mean vector \( 0_{2n} \) and covariance matrix \( \sigma_{2n}^2I_{2n} \) so that the errors are independent and identically distributed.

In the above setting, the maximum likelihood estimator (MLE) of \( \beta \) is \( [b_1, b_2, b_3, b_4, b_5, b_6]^T = (X^TX)^{-1}X^TY \) and by the invariance property of the MLE, the MLE of \( \gamma \) is \( b_0/b_6 \). Thus the MLE of \( \gamma \) is the ratio of two correlated normal random variables. Fieller [2,3] proposed a method for obtaining confidence sets for ratios of correlated normal random variables. The above inferential approach of obtaining the estimate of the blending coefficient \( \gamma \) and its confidence interval has been discussed in Gill and Keating [5], and the references therein. They described this method within the context of the blended UST leak model in their paper.

### 2.2. Power exponential distribution (PED) random error

In reality, data usually have heavy-tailed distributions because in most cases they contain some outliers. To capture these outliers in the blended UST leak model described above in Section 2.1, we propose a Bayesian approach to obtain a confidence interval for the blending coefficient, \( \gamma \), in the blended underground storage tank (UST) leak model with non-normal random errors through the Power Exponential Distribution (PED). In general, the Power Exponential Distribution (PED) can be used to model both light and heavy tailed, symmetric and unimodal continuous data sets. However, in reality data mostly have heavy-tailed distributions, and hence normally we use PED model to overcome the common problem in the real data such as the existence of the outliers (i.e. heavy-tailed distribution).

Zheng et al. [11] expressed the probability density function (pdf) of PED \((\mu, \sigma, \beta)\) as follows, where \( \beta \) is called the shape parameter and its impact on the shape of the pdf of PED is illustrated in Fig. 2.

\[
f(y; \mu, \sigma, \beta) = \frac{\sigma \Gamma \left( \frac{1}{2} \left( \frac{1}{\beta} + 1 \right) \right)}{\Gamma \left( \frac{1}{2} \right)} \exp \left( -\frac{1}{2} \frac{|y - \mu|^{2\beta}}{\sigma^{2\beta}} \right), -\infty < \mu < \infty, \sigma > 0, 0 < \beta \leq \infty. \tag{2}
\]

Note that in particular, one obtain the normal distribution for \( \beta = 1 \). We will use Zheng’s notation on PED in this paper. In the PED literature, there are different ways of expressing the pdf of PED, as shown in the following two examples. Tadikamalla [10] expressed the pdf of PED as follow.

\[
f(x) = \exp \left( -|x|^\alpha \right) / [2 \Gamma(1 + 1/\alpha)], -\infty < x < \infty, \alpha \geq 1.
\]

In addition, Mineo [9] expressed the pdf of PED as follows.

\[
f(x) = \frac{1}{2\sigma_p^{1/p} \Gamma(1 + 1/p)} \exp \left( -\frac{|x - \mu|^{\alpha}}{p \sigma_p^\alpha} \right), -\infty < x < \infty, -\infty < \mu < \infty, \sigma_p > 0, \ p > 0.
\]

In fact, by changing the shape parameter \( p \), the PED describes both leptokurtic \((0 < p < 2)\) and platikurtic \((p > 2)\) distributions; in particular the PED becomes the Laplace distribution (or double exponential distribution) for \( p = 1 \), the normal distribution for \( p = 2 \) and the Uniform distribution for \( p \rightarrow \infty \) (in notation of Mineo [9]). For illustration, using the R package (normalp) by Mineo [9], four different PED distributions can be plotted with \( p = 1, 2, 4, 100 \) (the last one will be the uniform distribution) as seen below. Note that the above Mineo [9] notation of \( p = 1, 2, 4, 100 \) are equivalent to \( \beta = 0.5, 1, 2, 50 \) in Zheng et al. [11] notation.
2.3. Bayesian approach with PED errors

We propose to use Bayesian approach to analyze the blended UST leak data. The sampling distributions are based on Eq. (1), i.e.,

\[ Y_{1j} | \beta_{01}, \beta_{02}, \beta_1, \beta_2, \beta_3, \gamma, \beta, \sigma \sim PED(\beta_{01} + \beta_1 X_{1j} + \gamma \beta_2 X_{2j}, \sigma, \beta) \]

\[ Y_{2j} | \beta_{01}, \beta_{02}, \beta_1, \beta_2, \beta_3, \gamma, \beta, \sigma \sim PED(\beta_{02} + (1 - \gamma) \beta_2 X_{2j} + \beta_3 X_{3j}, \sigma, \beta) \]

(3)

Where \( Y_{1j} \) and \( Y_{2j} \) are independent, \( j = 1, \ldots, n \). We put non-informative flat prior distributions on all the parameters due to lack of prior information on these parameters,

\[ \beta \sim \text{Gamma}(1.001, 0.001) \]

\[ \beta_{01} \sim N(0, 10^{-6}) \]

\[ \beta_{02} \sim N(0, 10^{-6}) \]

\[ \beta_1 \sim N(0, 10^{-6}) \]

\[ \beta_2 \sim N(0, 10^{-6}) \]

\[ \beta_3 \sim N(0, 10^{-6}) \]

\[ \gamma \sim N(0, 10^{-6}) \]

\[ \sigma \sim \text{Unif}(0, 10) \]

Note that in the above notation the second parameter for the normal distribution is the precision parameter. For the gamma distribution, the shape parameter is 1.001 and the rate parameter is 0.001.

To draw posterior distribution based on this Bayesian PED model, we used WinBUGS software. This software draws all parameters from the joint posterior distribution using the Gibbs Sampling method [8]. For this particular Bayesian PED model, we suggest to use a single chain of 10000 iterations and thinning of 50%. After a posterior draw of each parameter is obtained, we use the lower and upper \((\alpha/2)\)th quantiles as the limits for a \((1 - \alpha)100\%\) credible interval. These suggestions were used in the example (Section 3) and simulation study (Section 4).
Table 1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>β₀₁</th>
<th>β₀₂</th>
<th>β₁</th>
<th>β₂</th>
<th>β₃</th>
<th>γ</th>
<th>σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate [7]</td>
<td>1.359</td>
<td>1.620</td>
<td>0.999</td>
<td>1.004</td>
<td>0.996</td>
<td>0.650</td>
<td>3.789</td>
</tr>
<tr>
<td>Bayesian Normal model</td>
<td>1.373</td>
<td>1.627</td>
<td>0.999</td>
<td>1.005</td>
<td>0.996</td>
<td>0.6504</td>
<td>3.822</td>
</tr>
<tr>
<td>standard error for Normal model</td>
<td>1.0</td>
<td>1.052</td>
<td>0.0023</td>
<td>0.0085</td>
<td>0.0061</td>
<td>0.0043</td>
<td>0.257</td>
</tr>
<tr>
<td>Bayesian PED model (β = 0.4917)</td>
<td>1.49</td>
<td>1.744</td>
<td>0.997</td>
<td>1.006</td>
<td>1.0</td>
<td>0.6554</td>
<td>1.302</td>
</tr>
<tr>
<td>standard error for PED model</td>
<td>0.629</td>
<td>0.783</td>
<td>0.0020</td>
<td>0.0067</td>
<td>0.0061</td>
<td>0.0030</td>
<td>0.516</td>
</tr>
</tbody>
</table>

3. An example case

Keating and Mason [7] provide the Cary real data of sample size \( n = 54 \), which consist of five variables: \( Y_1 \), \( Y_2 \), \( X_1 \), \( X_2 \), \( X_3 \) as described in expression (1). Using all the Cary real data, their analysis produces the estimate of the blending coefficient \( \gamma \) to be 0.6503487 and the 95% Confidence Interval for \( \gamma \) is (0.6424092, 0.6582959). This analysis produces estimates that are roughly 0.65035 ± 0.00794. The estimates of the other parameters are given in Table 1. The approximate Fieller’s [1–3] large sample confidence interval produced by Gill and Keating [5] is 0.65035 ± 0.00785, which is considered to be better than the CI by Keating and Mason [7].

We applied our Bayesian method to the same Cary data. In the implementation of the Bayesian method, we used normal error as well as PED error. The parameter estimates and their standard errors are summarized in Table 1. As expected, the Keating and Mason model and the Bayesian Normal model produced similar parameter values. However, the Bayesian PED model produced an estimate of \( \beta \) of 0.4917, indicating the Cary data is not normally distributed. As a result, the standard error estimates from the Bayesian PED model are considerably smaller than the Keating and Mason model for all the linear model parameters. The parameter estimates of the Bayesian PED model differ slightly with those by the Keating and Mason method.

4. A simulation study

In this section we conduct a simulation study to compare the performance of CIs by Fieller’s method and the Bayesian method. The performance of both methods is evaluated in terms of coverage probability and average interval width. In the simulation we consider 95% CIs.

Our simulation is set up as follows.

1. The response variables \( Y_{1j} \) and \( Y_{2j} \) are generated according to the model in Eq. (3).
2. The values for six of the model parameters (\( \beta_{01} \), \( \beta_{02} \), \( \beta_1 \), \( \beta_2 \), \( \beta_3 \), and \( \sigma \)) are set as in first row of Table 1, which are obtained using the original Keating and Mason [7] estimates.
3. For the covariates \( X_{1j} \), we first calculated the sample mean (\( \mu_1 \)) and sample variance (\( s_1^2 \)) from the observed \( \{X_{1j}, j = 1, \ldots, 54\} \) in the Cary data, then we generate \( X_{1j} \), in our simulations from \( N(\mu_1, s_1^2) \). Similarly, \( X_{2j} \) and \( X_{3j} \) were generated.
4. For the blending coefficients, we choose three values: \( \gamma = 0.5, 0.65, \) and 0.8, which cover the most possible values of the blending coefficient.
5. For the parameter used in the PED model, we choose \( \beta = 0.5, 0.65, \) and 0.85. The choice of \( \beta \) covers a range of distributions with heavy tails.
6. The PED errors in the simulations were generated using the Mineo [9] normalp R statistical software package.

For each configuration of the parameters, we generate 1000 datasets and each dataset has a sample size of \( n = 50 \). Our simulation results are displayed in Table 2 as follows. For each simulation scenario and blending rate \( \gamma \), Table 2 displays the coverage probability, average width, and the percent improvement in average width for Bayesian method compared with the Fieller method.

Based on our simulation study and for the parameter configurations considered here, both Fieller and Bayesian intervals perform adequately in terms of coverage probability, but Bayesian interval perform better in terms of yielding at least nominal coverage with shorter expected interval width. In addition, the average interval width improved (to be shorter) by about 0.73% to 14.6% for the Bayesian CIs as compared with the Fieller CIs.
Table 2

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>$\beta$</th>
<th>CI methods</th>
<th>Coverage probability</th>
<th>Average width</th>
<th>Average width percent improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>Fieller</td>
<td>0.943</td>
<td>0.01778</td>
<td>14.60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bayesian</td>
<td>0.951</td>
<td>0.01518</td>
<td></td>
</tr>
<tr>
<td>0.65</td>
<td>0.5</td>
<td>Fieller</td>
<td>0.955</td>
<td>0.01065</td>
<td>5.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bayesian</td>
<td>0.948</td>
<td>0.01012</td>
<td></td>
</tr>
<tr>
<td>0.85</td>
<td>0.5</td>
<td>Fieller</td>
<td>0.952</td>
<td>0.00743</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bayesian</td>
<td>0.934</td>
<td>0.00738</td>
<td></td>
</tr>
<tr>
<td>0.65</td>
<td>0.5</td>
<td>Fieller</td>
<td>0.939</td>
<td>0.01856</td>
<td>14.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bayesian</td>
<td>0.947</td>
<td>0.01578</td>
<td></td>
</tr>
<tr>
<td>0.65</td>
<td>0.5</td>
<td>Fieller</td>
<td>0.956</td>
<td>0.01112</td>
<td>5.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bayesian</td>
<td>0.960</td>
<td>0.01055</td>
<td></td>
</tr>
<tr>
<td>0.85</td>
<td>0.5</td>
<td>Fieller</td>
<td>0.956</td>
<td>0.00776</td>
<td>1.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bayesian</td>
<td>0.923</td>
<td>0.00767</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.5</td>
<td>Fieller</td>
<td>0.943</td>
<td>0.02072</td>
<td>15.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bayesian</td>
<td>0.940</td>
<td>0.01753</td>
<td></td>
</tr>
<tr>
<td>0.65</td>
<td>0.5</td>
<td>Fieller</td>
<td>0.962</td>
<td>0.01242</td>
<td>6.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bayesian</td>
<td>0.941</td>
<td>0.01160</td>
<td></td>
</tr>
<tr>
<td>0.85</td>
<td>0.5</td>
<td>Fieller</td>
<td>0.949</td>
<td>0.00866</td>
<td>1.46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bayesian</td>
<td>0.923</td>
<td>0.00853</td>
<td></td>
</tr>
</tbody>
</table>

5. Conclusion

It is well-known that UST leak data have symmetric but heavy-tailed empirical distributions. Therefore, the existing Fieller method for constructing CIs for the blending parameter is suboptimal. In this paper we propose a Bayesian model with errors having the PED. This PED can capture the UST leak data better than the normal distribution. Simulations have shown that the Bayesian CIs have shorter interval widths than the Fieller CIs.

Clearly this Bayesian model with PED errors can be applied to other data with symmetric but heavy-tailed distributions. When the data have normal or light-tailed distribution, we recommend the use of Fieller method as this method performs well and is easier for implementation than the Bayesian method.

Acknowledgement

We thank two anonymous referees for their insightful comments and suggestions which have greatly improved the presentation of this article.

References