Bayesian Criteria for Non-Zero Effects Detection under Skew-Normal Search Model

Authors: Sara Sadeghi
– Department of Statistics, University of Isfahan
Isfahan, Iran

Hooshang Talebi
– Department of Statistics, University of Isfahan
Isfahan, Iran
h-talebi@sci.ui.ac.ir

Abstract:

- Shirakura et al. [12] introduced search probability (SP) in order to compare search designs (SD). Afterwards, the SP-based and other related criteria were developed, all for the normal model. In the present study, we considered a general underlying skew-normal (SN) model and obtained new criteria in a simple explicit form using the Bayesian approach. These criteria are design-dependent and hence are able to rank SDs with respect to their search performance.

Key-Words:

- *Bayesian approach; Kullback-Leibler distance; search design; search linear model; skew-normal distribution.*

AMS Subject Classification:

- 62K15

1. INTRODUCTION

At the screening stage of an experiment, a main effect plan (MEP) is employed to estimate the main effects, assuming that all interactions are negligible. MEPs were introduced and implemented after World War II, for more details refer to the pioneering paper of Box and Wilson [6]. MEPs, including saturated resolution III regular and irregular designs, have been widely used in practical industrial experiments. For example, Plackett and Burman [11] introduced an irregular saturated MEP for 2^m factorial experiments, where $m = 4t - 1$, for $t \geq 3$. Nevertheless, there might exist a small number of non-zero lower order interactions, which cause bias in estimating main effects. Enhancing the resolution, i.e., upgrading to resolution IV or V, for instance through fold-over approach to overcome the problem, increases the number of runs and, in turn, the cost of the experiment.

To save the number of runs, Srivastava [14] introduced and suggested using SDs to search for and estimate k unknown non-zero interactions in addition to estimating the main effects. Such a design is known as main effect plus k plan (MEP.k). Several researchers have developed the MEP.k (see Ghosh et al. [8], for a thorough review). For example, Esmailzadeh et al. [7] and Talebi and Jalali [18] constructed MEP.1 for 2^m factorial designs respectively, for odd and even m. Consider search linear model for providing a key condition in planning a general SD and in particular MEP.k. For a vector of observations $\mathbf{y}(N \times 1)$, the search linear model is

$$\mathbf{y} = \mathbf{A}_1\boldsymbol{\xi}_1 + \mathbf{A}_2\boldsymbol{\xi}_2 + \mathbf{e}, \qquad Cov(\mathbf{e}) = \sigma^2\mathbf{I}_N, \tag{1.1}$$

where $\mathbf{A}_i(N \times \nu_i)$ are known design matrices; and $\boldsymbol{\xi}_i(\nu_i \times 1)$ are vectors of effects for $i = 1, 2$; $\mathbf{e}(N \times 1)$ is an error vector; σ^2 is the error variance; and $\mathbf{I}_N$ is the identity matrix of order N. It is known for a fact that k effects in $\boldsymbol{\xi}_2$ are non-zero, but we don't know which ones. Therefore, the plan sets out to search for and identify the non-negligible effects in $\boldsymbol{\xi}_2$ and estimate them in addition to estimating the effects in $\boldsymbol{\xi}_1$. Alternatively, let S be the set of all $\binom{\nu_2}{k}$ models with only one correct model, each including a set of k possible non-zero effects from $\boldsymbol{\xi}_2$ and $\boldsymbol{\xi}_1$. The j-th model, $j = 1, 2, ..., \binom{\nu_2}{k}$, in S is expressed as follows:

$$\mathbf{y} = \mathbf{A}_1\boldsymbol{\xi}_1 + \mathbf{A}_{21}(\boldsymbol{\zeta}_j)\boldsymbol{\zeta}_j + \mathbf{e}, \tag{1.2}$$

where $\boldsymbol{\zeta}_j(k \times 1)$ is a vector of k effects from $\boldsymbol{\xi}_2$ and $\mathbf{A}_{21}(\boldsymbol{\zeta}_j)$ is the $N \times k$ submatrix of $\mathbf{A}_2$ whose columns are corresponding to $\boldsymbol{\zeta}_j$.

To identify the non-zero set of effects in $\boldsymbol{\xi}_2$ for noisy case ($\sigma^2 > 0$), Srivastava [14] suggested choosing the model in (1.2) with the lowest sum of square error (SSE). Moreover, Shirakura et al. [12] studied the stochastic properties of SSE and derived the SP in an explicit form for k=1 under the normal error. SP is design-dependent and hence Shirakura et al. [12] suggested using it for comparing SDs with respect to their search performance. Subsequently, Ghosh and

Teschmacher [9] and Talebi and Esmailzadeh [16] derived the SP-based criteria. Furthermore, Talebi and Esmailzadeh [15] conducted another design-comparison study and derived the KL (Kullback-Leibler) criterion based on Kullback-Leibler distance, which can be used for $k \geq 1$.

All of the above proposed criteria were obtained for models with normal error. However, such models may not adequately fit the data in many practical situations. For example, Arnold and Beaver [2] described a real situation in which the observations followed a non-normal distribution. They termed this situation 'hidden truncation', for which the model is SN. Afterwards, Arnold et al. [3] reported observations related to the hidden truncation. Moreover, Arellano-Valle et al. [1] assumed the SN error to fit a mixed model to a real set of longitudinal data on cholesterol levels collected as a part of the famed Framingham heart study. The above examples revealed the abundance of phenomena with SN models in real situations. The present study was also motivated by a hidden truncation problem, i.e. candidates who want to partake in the PhD Admission Examination of Iranian Universities must have an overall above-average Masters GPA. To deal with this, distributions such as skew-t distribution or mixture of two normal distributions may be proposed. However, based on our findings, such proposed distributions may not lead to an explicit solution. We considered the rival models in (1.2) with the multivariate SN distribution for error and used a Bayesian method to propose a new approach for finding the true model. This led to criteria which will be presented in an explicit form. The Bayesian approach in developing new explicit criteria allowed us to take into account the hierarchical principle in factorial experiments, by which the lower order interactions are more important than the higher orders. It was, therefore, rational to choose an appropriate prior distributional model for the factorial effects in order to deal with this issue. Through this prior distribution, we allocated non-zero probability to the main effects and k possible low order non-zero interactions, while all other interactions came down to zero probability. In this study, which is the first Bayesian research in the context of search design, it was shown that the Bayesian approach could simplify the complexity in deriving the appropriate criteria.

In the next section some useful preliminaries are presented. The new Bayesian search criteria will be proposed in section 3. These criteria are 1- expected Shannon information (ESI) and 2- Bayesian expected Kullback-Leibler (BEKL), which enable us to compare the search performance of any given SD. The calculations are moved to the Appendix in order to enhance the readability of the article.

2. PRELIMINARIES

The primary aim of this study was to acquire criteria for model identification in the context of search linear model. This problem has long been investigated by several researchers for models with normal error. In this study, we considered models with SN error. Thus, a better understanding of the SN distribution can be helpful.

Following Azzalini [4], who introduced SN distribution, a random variable Y has an SN distribution, denoted by $Y \sim SN(\mu, \sigma, \lambda)$, with location parameter μ; scale parameter σ; and shape parameter λ, if its probability density function (pdf) is

$$f(y) = \frac{2}{\sigma}\phi(\frac{y-\mu}{\sigma})\Phi(\lambda\frac{y-\mu}{\sigma}), \quad y \in R \tag{2.1}$$

where $\phi(.)$ and $\Phi(.)$ are the pdf and cumulative distribution function (cdf) of the standard normal distribution, respectively. The multivariate SN distribution has also been proposed by some researchers. That is, an N-dimensional random vector, $\mathbf{Y}$, follows a multivariate SN distribution $SN_N(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ with location vector $\boldsymbol{\mu} \in R^N$; positive definite dispersion matrix $\boldsymbol{\Sigma}_{\mathbf{N\times N}}$; and skewness vector $\boldsymbol{\lambda} \in R^N$, if its pdf is

$$f(\mathbf{y}) = 2\phi_N(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi_1(\boldsymbol{\lambda}'\boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{y}-\boldsymbol{\mu})), \quad \mathbf{y} \in R^N, \tag{2.2}$$

where $\phi_N(.)$ is the pdf of the $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, (Arellano-Valle et al. [1]). Evidently, the random vector $\mathbf{Y}$ follows $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for $\boldsymbol{\lambda} = \mathbf{0}$. Following Arellano-Valle et al. [1], the random vector $\mathbf{Y} \sim SN_N(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ can be expressed as

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + \boldsymbol{\Sigma}^{\frac{1}{2}}(\boldsymbol{\delta}|T_0| + (\mathbf{I}_N - \boldsymbol{\delta}\boldsymbol{\delta}')^{\frac{1}{2}}\mathbf{T}_1), \tag{2.3}$$

where $\boldsymbol{\delta} = \frac{\boldsymbol{\lambda}}{\sqrt{1+\boldsymbol{\lambda}'\boldsymbol{\lambda}}}$; $T_0 \sim N(0,1)$; $\mathbf{T}_1 \sim N(\mathbf{0}, \mathbf{I}_N)$ is independent of T_0, and $\stackrel{d}{=}$ stands for equality in distribution. In $Z = |T_0|$, Z has a half-normal distribution. It is worth noting that model (2.3) covers bias and correlation among errors in addition to skewness. Now, for hidden truncation problem, the SN distribution is written as follows. Suppose random vector $(X, W_1, W_2, ..., W_N)'$ distributed as $N_{N+1}(\boldsymbol{\theta}, \boldsymbol{\Omega})$, where $\boldsymbol{\theta} = (\mu_x, \boldsymbol{\mu}')'$ and $\boldsymbol{\Omega} = \begin{pmatrix} 1 & \boldsymbol{\delta}' \\ \boldsymbol{\delta} & \mathbf{I}_N \end{pmatrix}$. Let $\mathbf{W} = (W_1, W_2, ..., W_N)'$, then following Azzalini [5]

$$\mathbf{Y} = \mathbf{W}|X > \mu_x \sim SN_N(\boldsymbol{\mu}, \mathbf{I}_N, \boldsymbol{\lambda}), \tag{2.4}$$

where $\boldsymbol{\lambda} = (1-\boldsymbol{\delta}'\boldsymbol{\delta})^{-\frac{1}{2}}\boldsymbol{\delta}$. We calculated some of the existing criteria for detecting non-zero effects under the SN search model. Based on the findings, the calculation of SP for SN model has proven to be very intricate. Furthermore, the expected KL (EKL) criterion, proposed by Talebi and Esmailzadeh [15], for $\mathbf{Y} \sim SN_N(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ led to the integral below:

$$\int 2\phi_N(\mathbf{y}|\boldsymbol{\mu}_0, \boldsymbol{\Sigma})\Phi_1\boldsymbol{\lambda}'\boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{y}-\boldsymbol{\mu}_0))\log\{\frac{\phi_N(\mathbf{y}|\boldsymbol{\mu}_0, \boldsymbol{\Sigma})\Phi_1(\boldsymbol{\lambda}'\boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{y}-\boldsymbol{\mu}_0))}{\phi_N(\mathbf{y}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma})\Phi_1(\boldsymbol{\lambda}'\boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{y}-\boldsymbol{\mu}_i))}\}d\mathbf{y}, \tag{2.5}$$

where for non-zero $\boldsymbol{\zeta}_0$, $\boldsymbol{\mu}_0 = \mathbf{A}_1\boldsymbol{\xi}_1 + \mathbf{A}_{21}(\boldsymbol{\zeta}_0)\boldsymbol{\zeta}_0$ and $\boldsymbol{\mu}_j = \mathbf{A}_1\boldsymbol{\xi}_1 + \mathbf{A}_{21}(\boldsymbol{\zeta}_j)\boldsymbol{\zeta}_j$. This can not be made any simpler, and thus it is hard to be satisfied with (2.5) as a criterion. The desire of finding a very simple and conceivable criterion, consequently, motivated us to look for a different approach.

Lindley [10] defined the expected information about $\boldsymbol{\theta}$ for observation vector $\mathbf{y}$ in an experiment E, prior function $\pi(\boldsymbol{\theta})$, and posterior pdf $\pi(\boldsymbol{\theta}|\mathbf{y})$ as below:

$$I^{\boldsymbol{\theta}}\{E,\pi(\boldsymbol{\theta})\} = \int f(\mathbf{y}) \int \pi(\boldsymbol{\theta}|\mathbf{y}) \log \frac{\pi(\boldsymbol{\theta}|\mathbf{y})}{\pi(\boldsymbol{\theta})} d\boldsymbol{\theta} d\mathbf{y}, \tag{2.6}$$

provided that the integral exists. This is the expected KL distance between prior and posterior distributions, which measures the average overall observations information. Using Bayes' theorem, $I^{\boldsymbol{\theta}}\{E,\pi(\boldsymbol{\theta})\}$ in (2.6) can be written as follows:

$$I^{\boldsymbol{\theta}}\{E,\pi(\boldsymbol{\theta})\} = E_{\boldsymbol{\theta}}\{E_{\mathbf{Y}|\boldsymbol{\theta}}(\log f(\mathbf{y}|\boldsymbol{\theta}))\} - E_{\mathbf{Y}}\{\log f(\mathbf{y})\}. \tag{2.7}$$

The distance in (2.7) will be used for proposing the new criteria in section 3.

For the normal distribution $N(\mathbf{0}, \sigma^2\boldsymbol{\Sigma})$ and SN distribution $SN_N(\mathbf{0}, \sigma^2\boldsymbol{\Sigma}, \boldsymbol{\lambda})$ with unknown σ^2, let's take $\mathbf{y}^* = \mathbf{y}/\sigma$ and rewrite model (1.2) as below, which will be used throughout this article,

$$\mathbf{y}^* = \mathbf{A}(\boldsymbol{\xi}_j)\boldsymbol{\xi}_j^* + \mathbf{e}^*, \quad j = 1, 2, ..., \binom{\nu_2}{k}, \tag{2.8}$$

where $\boldsymbol{\xi}_j^* = \frac{1}{\sigma}(\boldsymbol{\xi}_1', \boldsymbol{\zeta}_j')'$; $\mathbf{A}(\boldsymbol{\xi}_j) = [\mathbf{A}_1 : \mathbf{A}_{21}(\boldsymbol{\zeta}_j)]$; and $\mathbf{e}^* = \mathbf{e}/\sigma$. In the Bayesian framework, $\boldsymbol{\xi}_j^*$ is assumed to have the prior distribution $N(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{0}})$, where $\boldsymbol{\Sigma}_0$ is a known $(\nu_1+k)\times(\nu_1+k)$ diagonal matrix. Following Wu and Hamada [19, p.434], by assuming large diagonal elements in $\boldsymbol{\Sigma}_0$, we are assured of the possibility of the presence of non-zero effects in $\boldsymbol{\xi}_j^*$. For a given prior, $\pi(\boldsymbol{\xi}_j^*)$, the event of observing a small interior integral in (2.6) indicates that the data support the existence of the non-negligible effects. Therefore, a small interior integral value in $I^{\boldsymbol{\theta}}\{E,\pi(\boldsymbol{\theta})\}$, presumably confirms the possibility of the presence of non-zero effects in $\boldsymbol{\xi}_j^*$. By this scenario, we suggested calculating the interior integral in (2.6) for all $\binom{\nu_2}{k}$ models in (2.8) and selecting the model with the lowest value as the true model. The following simulation study was performed as the verity performance assessment of the proposed criterion.

The search design D_1 given in the Appendix was used to generate data. Let $\boldsymbol{\xi}_1$ be the vector of the general mean and main effects and let ζ_0 be the two-factor interaction AB. Furthermore, in a hidden truncation model, assume that $\boldsymbol{\delta} = 0.2\mathbf{1}_{12}$, where $\mathbf{1}_{12}$ is a 12×1 vector of 1s, and $\boldsymbol{\Sigma}_0 = 100\mathbf{I}_6$. Based on these parameter values, 1000 data set were simulated from a 12 dimensional SN distribution using "sn" package in R software. The interior integral in (2.6) was calculated for all 6 possible models with any one of the two-factor interactions. The simulation results showed that the interior integral had the lowest value for the true model with AB interaction. We also calculated SSE for all models and found that the same model had the minimum SSE. Moreover, we ran this simulation for the case $k = 2$, by assuming $\boldsymbol{\zeta}_0$ to be (AB AC) and found that the interior integral and SSE were minimal for the chosen model.

Meanwhile, for a given model, Zhang [20] used $I^{\boldsymbol{\theta}}\{E,\pi(\boldsymbol{\theta})\}$ to select the optimum design, i.e. the design which maximizes the expression in (2.6). Due to the design-independence of the prior in denominator, she concluded that maximizing $I^{\boldsymbol{\theta}}\{E, \pi(\boldsymbol{\theta})\}$ comes down to maximizing the following quantity

$$U = \int f(\mathbf{y})\{\int \pi(\boldsymbol{\theta}|\mathbf{y}) \log \pi(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}\} d\mathbf{y}. \tag{2.9}$$

It is worth noting that for any given design, say D, $U(D)$ is the expected Shannon information of the posterior distribution denoted by ESI_D. Zhang [20] achieved an expression for (2.9) in the normal regression model and showed that maximization of $U(D)$ is equivalent to maximizing the determinant of inverted posterior variance of unknown parameter.

Under model uncertainty, when one is faced with a multi-model case, it is logical to calculate (2.9) for all models, opt for the model with the lowest value and then, select a design that has the maximum of such the value. In other words, let $U_i(D)$ be ESI_D in (2.9) for the i-th model, i=1,2,...,$\binom{\nu_2}{k}$, then $MESI_D = min_S U_i(D)$. Evidently, in the context of search design for any given design D, the larger the value of $MESI_D$, the higher the performance of D in searching for non-zero effects. So, for comparing and ranking the SDs with respect to their search performance, $MESI_D$ can be used as a criterion for design comparison. Hence, we present the following definition.

Definition 2.1. Suppose D_1 and D_2 are two SDs with N treatments, D_1 is said to be better than D_2 for identifying the set of non-zero effects if $MESI_{D_1} > MESI_{D_2}$.

3. MAIN RESULTS

3.1. ESI search criterion

In this section, we first introduce ESI_D as a criterion under normality assumption and then give a generalized form of the criterion using the SN model.

Consider the model in (2.8) and assume that $\mathbf{Y}^* \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = \mathbf{A}(\boldsymbol{\xi}_j)\boldsymbol{\xi}_j^*$. Then for foregoing $\pi(\boldsymbol{\xi}_j^*)$, $j = 1, 2, ..., \binom{\nu_2}{k}$, the posterior distribution of $\boldsymbol{\xi}_j^*$ is proportional to $f(\boldsymbol{\xi}_j^*, \mathbf{y}^*)$ given in (6.1) below. After some calculations, as given in the Appendix, the interior integral in U becomes

$$E_{\boldsymbol{\xi}_j^*|\mathbf{y}^*}\{\log \pi(\boldsymbol{\xi}_j^*|\mathbf{y}^*)\} = -\frac{1}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{\xi}}| - \frac{\nu_1 + k}{2}, \tag{3.1}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}$ is a conditional posterior variance of $\boldsymbol{\xi}_j^*$ given $\mathbf{y}^*$. U is obtained from (3.1) by integration with respect to the marginal distribution of $\mathbf{Y}^*$. After removing the redundant terms, U is reduced to a simple form $\psi(D)$ for design D,

$$\psi(D) = \log|\boldsymbol{\Sigma}_{\boldsymbol{\xi}}|^{-1}. \tag{3.2}$$

Note that $|\boldsymbol{\Sigma}_{\mathbf{y}^*}| = |\boldsymbol{\Sigma}_{\boldsymbol{\xi}}|^{-1}|\boldsymbol{\Sigma}_0|$, hence $\psi(D)$ is proportional to $\log|\boldsymbol{\Sigma}_{\mathbf{y}^*}|$. It should also be noted that $\psi(D)$ is design-dependent and written in terms of the hyper parameter $\boldsymbol{\Sigma}_0$. Therefore, for any given design D, $\psi(D)$ is calculable.

Remark 3.1. In $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}$, the expression $\mathbf{A}'(\boldsymbol{\zeta}_j)\boldsymbol{\Sigma}^{-1}\mathbf{A}(\boldsymbol{\zeta}_j)$ is the inverted variance of $(\mathbf{A}'(\boldsymbol{\zeta}_j)\boldsymbol{\Sigma}^{-1}\mathbf{A}(\boldsymbol{\zeta}_j))^{-1}\mathbf{A}(\boldsymbol{\zeta}_j)'\boldsymbol{\Sigma}^{-1}\mathbf{y}^*$, and $\boldsymbol{\Sigma}_0^{-1}$ is the inverted prior variance of $\boldsymbol{\xi}_j^*$ which, in fact, combines prior information with extracted information from the data.

Now, it is assumed that vector $\mathbf{Y}^*$ in the model (2.8) is distributed as a multivariate SN, $SN_N(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$. Calculation of $E_{\boldsymbol{\xi}_j^*|\mathbf{y}^*}\{\log\pi(\boldsymbol{\xi}_j^*|\mathbf{y}^*)\}$ for SN distribution is not simple due to complexity of such distribution. To simplify the problem, we used expression (2.3) for $\mathbf{Y}^*$ and apply the conditional distribution below:

$$\mathbf{Y}^*|Z = z \sim N(\boldsymbol{\mu} + z\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\delta}, \mathbf{G}), \tag{3.3}$$

where $\mathbf{G} = \boldsymbol{\Sigma}^{\frac{1}{2}}(\mathbf{I}_N - \boldsymbol{\delta}\boldsymbol{\delta}')\boldsymbol{\Sigma}^{\frac{1}{2}}$. Following Sorensen and Gianola [13], we use the distribution of $\mathbf{Y}^*$ condition on the latent variable Z, in writing the posterior distribution as given in (6.2). Insert the unobserved random variable Z in the parameters vector, i.e. $\boldsymbol{\theta}_j' = (\boldsymbol{\xi}_j^{*'}, Z)$, $j = 1, 2, ..., \binom{\nu_2}{k}$, and take the prior distributions $N(\mathbf{0}, \boldsymbol{\Sigma}_0)$ for $\boldsymbol{\xi}_j^*$. The joint posterior distribution of $\boldsymbol{\theta}_j$ is proportional to $f(\boldsymbol{\theta}_j, \mathbf{y}^*)$ in (6.2).

The Shannon information criterion is

$$E_{\boldsymbol{\theta}_j|\mathbf{y}^*}\{\log\pi(\boldsymbol{\theta}_j|\mathbf{y}^*)\} = E_{Z|\mathbf{y}^*}\{E_{\boldsymbol{\xi}_j^*|z,\mathbf{y}^*}(\log\pi(\boldsymbol{\theta}_j|\mathbf{y}^*))\}. \tag{3.4}$$

More calculations and details are given in the Appendix, based on which, the conditional expectation in (3.4) is simplified to the reduced form below:

$$\begin{aligned} E_{\boldsymbol{\theta}_j|\mathbf{y}^*}\{\log\pi(\boldsymbol{\theta}_j|\mathbf{y}^*)\} = &- \frac{1}{2}\log|2\pi\boldsymbol{\Sigma}_{\boldsymbol{\xi}}| - \frac{\nu_1 + k}{2} \\ &- \frac{1}{2}\log(2\pi\sigma_z^2) - \frac{1}{2} + \frac{z^*\phi(\frac{z^*}{\sigma_z})}{2\sigma_z\Phi(\frac{z^*}{\sigma_z})} - \log(\Phi(\frac{z^*}{\sigma_z})), \end{aligned}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}$ and σ_z^2 are conditional posterior variance of $\boldsymbol{\xi}_j^*$ given $(z, \mathbf{y}^*)$ and conditional posterior variance of Z given $\mathbf{y}^*$, respectively. z^* is conditional posterior mean of Z given $\mathbf{y}^*$. More details on these can be found in the Appendix.

Meanwhile, the expected value of (2.9) is computed with respect to the marginal distribution of $\mathbf{Y}^*$ given in the Appendix, i.e. $SN_N(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{y}^*}, \boldsymbol{\gamma}_{\mathbf{y}^*})$. It gives,

$$\begin{aligned} U = &- \frac{\nu_1 + k + 1}{2}\log(2\pi) - \frac{\nu_1 + k + 1}{2} - \frac{1}{2}\log\{|\boldsymbol{\Sigma}_{\boldsymbol{\xi}}|(\sigma_z^2)\} \\ &+ \frac{1}{2}E_T\{T\frac{\phi(T)}{\Phi(T)} - 2\log[\Phi(T)]\}, \end{aligned} \tag{3.5}$$

where $T = \frac{z^*}{\sigma_z}$ with $T \sim SN(0, \sigma_t^2, \sigma_t)$; $\sigma_t^2 = \frac{\boldsymbol{\delta}'\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{M}'\boldsymbol{\Sigma}_{\boldsymbol{y}^*}\mathbf{M}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\delta}}{1+\boldsymbol{\delta}'\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{M}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\delta}}$; and $\mathbf{M}$ is given in the Appendix. ESI_D in (3.5) can be written as the following design-dependent criterion and then the minimum of such the criterion over all models in S be maximized over SDs to come up with the superior design.

$$\psi(D, \boldsymbol{\lambda}) = \log\{|\boldsymbol{\Sigma}_{\boldsymbol{\xi}}|^{-1}(\sigma_z^2)^{-1}\} + E_T\{T\frac{\phi(T)}{\Phi(T)} - 2\log[\Phi(T)]\}. \tag{3.6}$$

It should also be noted that $|\boldsymbol{\Sigma}_{\mathbf{y}^*}| = |\mathbf{G}||\boldsymbol{\Sigma}_{\boldsymbol{\xi}}|^{-1}|\boldsymbol{\Sigma}_0|(\sigma_z^2)^{-1}$, therefore

$$\psi(D, \boldsymbol{\lambda}) \propto \log|\boldsymbol{\Sigma}_{\mathbf{y}^*}| + E_T\{T\frac{\phi(T)}{\Phi(T)} - 2\log[\Phi(T)]\}. \tag{3.7}$$

The subsequent remarks present more details on $\psi(D, \boldsymbol{\lambda})$.

Remark 3.2. Generally, $\boldsymbol{\lambda}$ is an $N \times 1$ unknown vector. Lacking a specific knowledge on $\boldsymbol{\lambda}$ may lead one to follow the Bayesian approach for choosing a prior distribution such as uniform on a sphere.

Remark 3.3. Similar to Remark 3.1, the term $\mathbf{A}'(\boldsymbol{\xi}_j)\mathbf{G}^{-1}\mathbf{A}(\boldsymbol{\xi}_j)$ in $|\boldsymbol{\Sigma}_{\boldsymbol{\xi}}|^{-1}$ is the inverted variance of $(\mathbf{A}'(\boldsymbol{\xi}_j)\mathbf{G}^{-1}\mathbf{A}(\boldsymbol{\xi}_j))^{-1}\mathbf{A}(\boldsymbol{\xi}_j)'(\mathbf{VG})^{-\frac{1}{2}}\mathbf{y}^*$ where $\mathbf{V} = \boldsymbol{\Sigma}^{\frac{1}{2}}(\mathbf{I}_N - \frac{2}{\pi}\boldsymbol{\delta}\boldsymbol{\delta}')\boldsymbol{\Sigma}^{\frac{1}{2}}$.

Remark 3.4. For $\boldsymbol{\lambda} \rightarrow \mathbf{0}$ (Normality error case) random variable T is degenerated at zero. Therefore, the second term in (3.7) disappears and $\psi(D, \boldsymbol{\lambda})$ remains with its first term. It is similar to what is given in (3.2) for normal case. In the hidden truncation model, if for every i=1,2,...,N, $\delta_i \rightarrow \mathbf{0}$, then $\mathbf{Y}^* \sim N(\boldsymbol{\mu}, \mathbf{I}_N)$ and $\psi(D)$ is simplified to (3.2) with $\boldsymbol{\Sigma} = \mathbf{I}$.

Remark 3.5. For the special case of identical skewness, i.e. $\lambda = \lambda\mathbf{1}_N$, $\lambda \in R$, σ_t and $\mathbf{G}^{-1} = \boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{I}_N + \lambda^2\mathbf{1}_N\mathbf{1}'_N)\boldsymbol{\Sigma}^{-\frac{1}{2}}$ are symmetric in λ. Therefore, $\psi(D, \lambda)$ is symmetric in λ. It should also be noted that for a hidden truncation problem with $\boldsymbol{\delta} = \delta\mathbf{1}_N$, $\psi(D, \delta)$ is symmetric in δ.

3.2. BEKL search criterion

In what follows, we obtain the expected KL distance, $I^{\boldsymbol{\theta}}\{E, \pi(\boldsymbol{\theta})\}$, under normal and SN distributions for error. It should be noted that by keeping the prior distribution in expected information (2.6) the results in this section will be different from the findings in section 3.1, which were obtained from U in (2.9).

Consider model (2.8), and for more understanding, first assume that $\mathbf{Y}^* \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Now, for $\boldsymbol{\xi}_j^* \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{0})$, $j = 1, 2, ..., \binom{\nu_2}{k}$, compute $E_{\mathbf{Y}^*|\boldsymbol{\xi}_j^*}(\log f(\mathbf{y}^*|\boldsymbol{\xi}_j^*))$

and $E_{\mathbf{Y}^*}\{\log f(\mathbf{y}^*)\}$ to reach $I^{\boldsymbol{\theta_j}}\{E,\pi(\boldsymbol{\theta_j})\}$ given in (2.7). From marginal distribution of $\mathbf{Y}^*$, which is given in the Appendix, we have

$$E_{\mathbf{Y}^*}\{\log f(\mathbf{y}^*)\} = -\frac{N}{2}(\log(2\pi)+1) - \frac{1}{2}\log|\boldsymbol{\Sigma}_{\mathbf{y}^*}|.$$

Clearly, $E_{\mathbf{Y}^*|\boldsymbol{\xi}_j^*}(\log f(\mathbf{y}^*|\boldsymbol{\xi}_j^*)) = -\frac{N}{2}(\log(2\pi)+1)$, hence $I^{\boldsymbol{\theta_j}}\{E,\pi(\boldsymbol{\theta_j})\}$ is

$$I^{\boldsymbol{\theta}_j}\{E,\pi(\boldsymbol{\theta}_j)\} = \frac{1}{2}\log|\boldsymbol{\Sigma}_{\mathbf{y}^*}|. \tag{3.8}$$

As can be seen in (3.8), in order to minimize $I^{\boldsymbol{\theta_j}}\{E,\pi(\boldsymbol{\theta_j})\}$, it is enough to minimize the simple form $|\boldsymbol{\Sigma}_{\mathbf{y}^*}|$ over all possible $\binom{\nu_2}{k}$ models.

Now, suppose $\mathbf{Y}^* \sim \mathbf{SN_N}(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda})$. Let's add the unobserved random variable Z to the parameters vector to get $\boldsymbol{\theta}'_j = (\boldsymbol{\xi}_j^{*'}, Z)$, $j = 1,2,...,\binom{\nu_2}{k}$. By assuming the prior distribution for the vector $\boldsymbol{\xi}_j^*$, as given herein, and noting that $\mathbf{Y}^*$ can be written as (2.3), we have

$$E_{\mathbf{Y}^*|\boldsymbol{\theta}_j}(\log f(\mathbf{y}^*|\boldsymbol{\theta}_j)) = -\frac{N}{2}(\log(2\pi)+1) - \frac{1}{2}\log|\mathbf{G}|,$$

and

$$E_{\mathbf{Y}^*}(\log f(\mathbf{y}^*)) = \log 2 - \frac{N}{2}(\log(2\pi)+1) - \frac{1}{2}\log|\boldsymbol{\Sigma}_{\mathbf{y}^*}| + E_T\{\log[\Phi(T)]\}.$$

Therefore, $I^{\boldsymbol{\theta_j}}\{E,\pi(\boldsymbol{\theta_j})\}$ provides the following:

$$I^{\boldsymbol{\theta}_j}\{E,\pi(\boldsymbol{\theta}_j)\} = -\log 2 - \frac{1}{2}\log|\mathbf{G}| + \frac{1}{2}\log|\boldsymbol{\Sigma}_{\mathbf{y}^*}| - E_T\{\log[\Phi(T)]\}. \tag{3.9}$$

Evidently, minimizing $I^{\boldsymbol{\theta_j}}\{E,\pi(\boldsymbol{\theta}_j)\}$ in (3.9) is equivalent to minimizing $\Phi(D,\boldsymbol{\lambda}) = \log|\boldsymbol{\Sigma}_{\mathbf{y}^*}| - 2E_T\{\log(\Phi(T))\}$ over the set of all possible models in S, known as the BEKL criterion. Note that the ESI in (3.7) has an extra term $E_T(T\frac{\phi(T)}{\Phi(T)})$ in comparing to the BEKL. That is, although the prior distribution is design-independent, keeping such the prior in (2.6) leads to a simple and more flexible criterion.

The proposed BEKL measure, which is primarily proposed for model discrimination, can also be used to compare search performance of SDs. In doing so, first for each of the SDs the minimum of the BEKL ($MBEKL_D$) is obtained over the set of all models. Then, the design with a larger $MBEKL_D$ is considered to be the desired one. Therefore, definition 2.1 is valid for designs D_1 and D_2 with respect to $MBEKL_D$-criterion if $MBEKL_{D_1} > MBEKL_{D_2}$.

4. IMPLEMENTATION

In this section, we assess the performance of the two proposed Bayesian criteria through comparing and ranking rival SDs. To do so, we use $MBEKL_D$

and $MESI_D$ under SN distribution for error. We compare search performance of three 12-run search designs D_1, D_2 and D_3, as given in the Appendix, for a 2^4 factorial experiment. Design D_1 is a balanced array of full strength, Design D_2 is the projection of a 12-run Plackett-Burman design onto its 4 columns, and Design D_3 is a non-repeated run orthogonal main effect plan. These designs have already been compared by Ghosh and Teschmacher [9] and Talebi and Esmailzadeh [17], under normality.

Example 4.1. Let in model (1.1), $\boldsymbol{\xi}_1$ be the vector of the general mean and main effects, and $\boldsymbol{\xi}_2$ be two- and three-factor interactions, while assuming that four-factor interaction is negligible. Furthermore, it is assumed that $\boldsymbol{\xi}_2$ includes two non-zero effects at the most. D_1, D_2, and D_3 are MEP.1. They are also MEP.2 plans, when $\boldsymbol{\xi}_2$ includes only two-factor interactions, assuming higher-order interactions are all zero. We were interested in studying scores of 12 EEIU volunteers with a GPA over than mean, i.e. $\mathbf{Y} = \mathbf{W}|X > \mu_x$, where W_is, i=1,2,...,12, are the scores and X is the GPA. Consider model (2.8) for the vector of observations $\mathbf{Y}$ and assume $(X, \mathbf{W}')' \sim N_{13}(\boldsymbol{\theta}, \boldsymbol{\Omega})$, where $\boldsymbol{\theta} = (\mu_x, \boldsymbol{\mu}')'$, $\boldsymbol{\mu} = \mathbf{A}(\boldsymbol{\xi}_j)\boldsymbol{\xi}_j^*$, and $\boldsymbol{\Omega} = \begin{pmatrix} 1 & \boldsymbol{\delta}' \\ \boldsymbol{\delta} & \mathbf{I}_{12} \end{pmatrix}$. In this case, $\mathbf{Y}$ satisfies the conditional distribution of (2.4). Data were collected through 3 possible designs D_1, D_2, and D_3. For $\boldsymbol{\delta} = \delta\mathbf{1}_{12}$, let $\boldsymbol{\sigma}_{D,\delta} = [\sigma_{t_{\zeta_1}}, \sigma_{t_{\zeta_2}}, ..., \sigma_{t_{\zeta_l}}]'$, where $l = \binom{\nu_2}{k}$, and $\sigma_{t_{\zeta_j}}$ denotes σ_t for the j-th model. Matlab software was used to calculate amount of the criterion. It was learned that $\boldsymbol{\sigma}_{D,\delta} = c_\delta \mathbf{1}_l$, for D_1, D_2 and D_3, where c_δ is scalar and depends on δ for all models. It is also true that $\boldsymbol{\sigma}_{D_1,\delta} = \boldsymbol{\sigma}_{D_2,\delta} = \boldsymbol{\sigma}_{D_3,\delta}$, which means that the value of σ_t depends neither on the model nor on the design. Consequently, in order to compare designs D_1, D_2, and D_3, for a fixed value of δ, the second expression for both criteria is canceled out and, therefore, both ESI_D and $BEKL_D$ become the same. This is true for the following design comparison and hence there is no difference in computing either of the criteria. For $k = 1$, once again we considered the prior distribution $N(\mathbf{0}, \boldsymbol{\Sigma}_0)$ for $\boldsymbol{\xi}_j^*$ in which $\boldsymbol{\Sigma}_0$ is a 6×6 diagonal matrix, with large diagonal elements of 100. The comparisons showed that D_2 is better than both D_1 and D_3, and D_1 is better than D_3. This result is the same as what was obtained using the compound criteria proposed by Talebi and Esmailzadeh [17]. For instance, when $\delta = 0.2$, values of criterion are 42.6251, 42.6738, and 42.4026 for D_1, D_2, and D_3, respectively, while the EKL values for these Designs are the same and equal to 10.667. This shows that the EKL is unable to discriminate search abilities of D_1, D_2, and D_3.

Example 4.2. In continuation of Example 4.1, let $\boldsymbol{\xi}_2$ be the vector of two-factor interactions only, and assume that three- and four-factor interactions are all zero. For $k = 1$, results showed that D_3 has the same search ability as D_1, and they are better than D_2, based on the present criteria. For example, when $\delta = 0.2$, criterion value for D_1 and D_3, is 42.6895 and for D_2 is 42.6738. For $k = 2$, assume that $\boldsymbol{\xi}_j^*$ is distributed as $N(\mathbf{0}, \boldsymbol{\Sigma}_0)$ in which $\boldsymbol{\Sigma}_0$ is a 7×7 diagonal matrix, with diagonal elements of 100. When $\delta = 0.2$, criterion value for D_1 and D_3 is 49.376, and for D_2 is 49.3115.

5. DISCUSSION

Findings in section 4 reveal that both criteria, $MESI_D$ and $MBEKL_D$ increase as δ increases; this means as $\delta(\geq 0)$ gets larger, the capability of SD enhances in identifying the non-zero effects, which has been ignored by the former criteria. The proposed criteria are also applicable for $k > 1$. So, an important advantage of the present criteria is their flexibility with respect to distributional model and the number of non-zero effects in $\boldsymbol{\xi}_2$. This study generalizes the previously-obtained results for the normal model by utilizing the SN distribution, where normal distribution is its special case. It is notable that unlike SP, $MESI_D$ and $MBEKL_D$ do not depend on an unknown parameter. This allows us to come up with numerical values for the criteria. Furthermore, the results presented in section 4 showed that $MESI_D$ and $MBEKL_D$ criteria have a higher discriminating power than the EKL, obtained by Talebi and Esmailzadeh [15].

6. APPENDIX

6.1. Conditional posterior distributions for normal distribution error

For $\mathbf{e}^* \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, $f(\boldsymbol{\xi}_j^*, \mathbf{y}^*)$ can be written as

$$\begin{aligned} f(\boldsymbol{\xi}_j^*, \mathbf{y}^*) &= f(\mathbf{y}^*|\boldsymbol{\xi}_j^*)\pi(\boldsymbol{\xi}_j^*) \\ &= (2\pi)^{-\frac{N+\nu_1+k}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\mathbf{y}^* - \mathbf{A}(\boldsymbol{\xi}_j)\boldsymbol{\xi}_j^*)'\boldsymbol{\Sigma}^{-1}(\mathbf{y}^* - \mathbf{A}(\boldsymbol{\xi}_j)\boldsymbol{\xi}_j^*)\} \\ &\quad \times |\boldsymbol{\Sigma}_0|^{-\frac{1}{2}} \exp\{-\frac{1}{2}\boldsymbol{\xi}_j^{*\prime}\boldsymbol{\Sigma_0}^{-1}\boldsymbol{\xi}_j^*\}, \end{aligned} \tag{6.1}$$

where $|.|$ stands for determinant. Using the joint distribution in (6.1) together with some other calculations, it can be shown that the conditional posterior distributions of parameters are as follows:

$$\boldsymbol{\xi}_j^*|\mathbf{y}^* \sim N(\boldsymbol{\mu}_{\boldsymbol{\xi}_j^*}, (\mathbf{A}'(\boldsymbol{\xi}_j)\boldsymbol{\Sigma}^{-1}\mathbf{A}(\boldsymbol{\xi}_j) + \boldsymbol{\Sigma}_0^{-1})^{-1}),$$

where $\boldsymbol{\mu}_{\boldsymbol{\xi}_j^*} = (\mathbf{A}'(\boldsymbol{\xi}_j)\boldsymbol{\Sigma}^{-1}\mathbf{A}(\boldsymbol{\xi}_j) + \boldsymbol{\Sigma}_0^{-1})^{-1}\mathbf{A}'(\boldsymbol{\xi}_j)\boldsymbol{\Sigma}^{-1}\mathbf{y}^*$. The logarithm of the joint posterior distribution is

$$\begin{aligned} \log\{\pi(\boldsymbol{\xi}_j^*|\mathbf{y}^*)\} &= -\frac{1}{2}\log|2\pi(\mathbf{A}'(\boldsymbol{\xi}_j)\boldsymbol{\Sigma}^{-1}\mathbf{A}(\boldsymbol{\xi}_j) + \boldsymbol{\Sigma}_0^{-1})^{-1}| \\ &- \frac{1}{2}(\boldsymbol{\xi}_j^* - \boldsymbol{\mu}_{\boldsymbol{\xi}_j^*})'(\mathbf{A}'(\boldsymbol{\xi}_j)\boldsymbol{\Sigma}^{-1}\mathbf{A}(\boldsymbol{\xi}_j) + \boldsymbol{\Sigma}_0^{-1})(\boldsymbol{\xi}_j^* - \boldsymbol{\mu}_{\boldsymbol{\xi}_j^*}) \end{aligned}$$

Note that,

$$(\boldsymbol{\xi}_j^* - \boldsymbol{\mu}_{\boldsymbol{\xi}_j^*})'(\mathbf{A}'(\boldsymbol{\xi}_j)\boldsymbol{\Sigma}^{-1}\mathbf{A}(\boldsymbol{\xi}_j) + \boldsymbol{\Sigma}_0^{-1})(\boldsymbol{\xi}_j^* - \boldsymbol{\mu}_{\boldsymbol{\xi}_j^*})|\mathbf{y}^* \sim \chi^2_{\nu_1+k}$$

where $\chi^2_{\nu_1+k}$ is chi-squared distribution with $\nu_1 + k$ degrees of freedom. Marginal distribution of $\mathbf{Y}^*$ is obtained from the joint distribution in (6.1). It can be easily shown that $\mathbf{Y}^*$ is distributed as $N(\mathbf{0}, \mathbf{\Sigma_{y^*}})$, where

$$\mathbf{\Sigma_{y^*}} = \{\mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\mathbf{A}(\boldsymbol{\xi}_j)\}(\mathbf{A}'(\boldsymbol{\xi}_j)\mathbf{\Sigma}^{-1}\mathbf{A}(\boldsymbol{\xi}_j) + \mathbf{\Sigma}_0^{-1})^{-1}\mathbf{A}'(\boldsymbol{\xi}_j)\mathbf{\Sigma}^{-1}\}^{-1}.$$

6.2. Conditional posterior distributions for SN distribution error

For $\mathbf{Y}^* \sim SN_N(\boldsymbol{\mu}, \mathbf{\Sigma}, \boldsymbol{\lambda})$, the joint density of vector $(\boldsymbol{\theta}_j, \mathbf{y}^*)$ is

$$\begin{aligned} f(\boldsymbol{\theta}_j, \mathbf{y}^*) &= f(\mathbf{y}^*|\boldsymbol{\theta}_j) f(z|\boldsymbol{\xi}_j^*)\pi(\boldsymbol{\xi}_j^*) \\ &= |\mathbf{G}|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\mathbf{y}^* - \mathbf{A}(\boldsymbol{\xi}_j)\boldsymbol{\xi}_j^* - z\mathbf{\Sigma}^{\frac{1}{2}}\boldsymbol{\delta})'\mathbf{G}^{-1}(\mathbf{y}^* - \mathbf{A}(\boldsymbol{\xi}_j)\boldsymbol{\xi}_j^* - z\mathbf{\Sigma}^{\frac{1}{2}}\boldsymbol{\delta})\} \\ &\times\ 2(2\pi)^{-\frac{N+\nu_1+k+1}{2}} |\mathbf{\Sigma}_0|^{-\frac{1}{2}} \exp\{-\frac{1}{2}[z^2 + \boldsymbol{\xi}_j^{*\prime}\mathbf{\Sigma}_0^{-1}\boldsymbol{\xi}_j^*]\}. \end{aligned} \tag{6.2}$$

From (6.2), the conditional posterior distributions of unknown parameters are obtained as:

$\boldsymbol{\xi}_j^*|z, \mathbf{y}^* \sim N(\boldsymbol{\mu}_{\boldsymbol{\xi}_j^*}, \mathbf{\Sigma}_{\boldsymbol{\xi}})$, and $Z|\mathbf{y}^* \sim N(z^*, \sigma_z^2)I(Z > 0)$, in which the conditional posterior distribution of $Z|\mathbf{y}^*$ is truncated normal at zero with the following pdf: $\pi(Z|\mathbf{y}^*) = \phi(\frac{z-z^*}{\sigma_z})/(\sigma_z\Phi(\frac{z^*}{\sigma_z}))$ and

$$\boldsymbol{\mu}_{\boldsymbol{\xi}_j^*} = \mathbf{\Sigma}_{\boldsymbol{\xi}}\mathbf{A}'(\boldsymbol{\xi}_j)\mathbf{G}^{-1}(\mathbf{y}^* - z\mathbf{\Sigma}^{\frac{1}{2}}\boldsymbol{\delta}),\ \mathbf{\Sigma}_{\boldsymbol{\xi}} = (\mathbf{A}'(\boldsymbol{\xi}_j)\mathbf{G}^{-1}\mathbf{A}(\boldsymbol{\xi}_j) + \mathbf{\Sigma}_0^{-1})^{-1},$$

$$z^* = \sigma_z^2\mathbf{y}^{*\prime}\mathbf{M}\mathbf{\Sigma}^{\frac{1}{2}}\boldsymbol{\delta},\ \sigma_z^2 = (1 + \boldsymbol{\delta}'\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{M}\mathbf{\Sigma}^{\frac{1}{2}}\boldsymbol{\delta})^{-1},$$

$$\mathbf{M} = \mathbf{G}^{-1} + \mathbf{G}^{-1}\mathbf{A}(\boldsymbol{\xi}_j)[(\mathbf{\Sigma}_{\boldsymbol{\xi}}\mathbf{A}'(\boldsymbol{\xi}_j)\mathbf{G}^{-1}\mathbf{A}(\boldsymbol{\xi}_j) + \mathbf{I}_{\nu_1+k})^{-1} - \mathbf{I}_{\nu_1+k}](\mathbf{A}'(\boldsymbol{\xi}_j)\mathbf{G}^{-1}\mathbf{A}(\boldsymbol{\xi}_j))^{-1}\mathbf{A}'(\boldsymbol{\xi}_j)\mathbf{G}^{-1}.$$

The logarithm of $\pi(\boldsymbol{\theta}_j|\mathbf{y}^*)$ (the joint posterior distribution of $\boldsymbol{\theta}_j$) can be written as

$$\begin{aligned} \log\pi(\boldsymbol{\theta}_j|\mathbf{y}^*) &= \log\{\pi(\boldsymbol{\xi}_j|z, \mathbf{y}^*)\} + \log\{\pi(Z|\mathbf{y}^*)\} \\ &= -\frac{1}{2}\log|2\pi\mathbf{\Sigma}_{\boldsymbol{\xi}}| - \frac{1}{2}(\boldsymbol{\xi}_j^* - \boldsymbol{\mu}_{\boldsymbol{\xi}_j^*})'\mathbf{\Sigma}_{\boldsymbol{\xi}}^{-1}(\boldsymbol{\xi}_j^* - \boldsymbol{\mu}_{\boldsymbol{\xi}_j^*}) \\ &- \frac{1}{2}\log(2\pi\sigma_z^2) - \frac{1}{2}(\frac{Z - z^*}{\sigma_z})^2 - \log(\Phi(\frac{z^*}{\sigma_z})). \end{aligned}$$

It should be noted that,

$$(\boldsymbol{\xi}_j^* - \boldsymbol{\mu}_{\boldsymbol{\xi}_j^*})'\mathbf{\Sigma}_{\boldsymbol{\xi}}^{-1}(\boldsymbol{\xi}_j^* - \boldsymbol{\mu}_{\boldsymbol{\xi}_j^*}) \sim \chi^2_{\nu_1+k},$$

and

$$E_{Z|\mathbf{y}^*}(Z - z^*)^2 = \sigma_z^2 - \sigma_z z^* \frac{\phi(\frac{z^*}{\sigma_z})}{\Phi(\frac{z^*}{\sigma_z})}.$$

From (6.2) the marginal distribution of $\mathbf{Y}^*$ is distributed as $SN_N(\mathbf{0}, \mathbf{\Sigma_{y^*}}, \boldsymbol{\gamma}_{\mathbf{y}^*})$, in which $\mathbf{\Sigma_{y^*}} = \{\mathbf{M} - \sigma_{\mathbf{z}}^{\mathbf{2}}\mathbf{M}\mathbf{\Sigma}^{\frac{1}{2}}\boldsymbol{\delta}\boldsymbol{\delta}'\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{M}'\}^{-1}$ and $\boldsymbol{\gamma}_{\mathbf{y}^*} = \mathbf{\Sigma_{y^*}}^{\frac{1}{2}}\frac{\mathbf{M}\mathbf{\Sigma}^{\frac{1}{2}}\boldsymbol{\delta}}{\sqrt{1+\boldsymbol{\delta}'\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{M}\mathbf{\Sigma}^{\frac{1}{2}}\boldsymbol{\delta}}}$.

6.3. Search designs D_1, D_2 and D_3 with 12 runs and 4 factors.

D_1				D_2				D_3			
A	B	C	D	A	B	C	D	A	B	C	D
+	+	+	+	+	-	+	-	+	+	+	+
-	-	-	-	+	+	-	+	+	-	+	+
-	-	-	+	-	+	+	-	-	-	+	+
-	-	+	-	+	-	+	+	-	+	-	+
-	+	-	-	+	+	-	+	+	-	-	+
+	-	-	-	+	+	+	-	-	-	-	-
-	-	+	+	-	+	+	+	-	+	+	-
-	+	-	+	-	-	+	+	+	-	+	-
+	-	-	+	-	-	-	+	-	-	+	-
-	+	+	-	+	-	-	-	+	+	-	-
+	-	+	-	-	+	-	-	+	-	-	-
+	+	-	-	-	-	-	-	-	-	-	-

REFERENCES

[1] Arellano-Valle, R.B.; Bolfarine, H. and Lachos, V.H. (2005). Skew-normal linear mixed models, *Journal of Data Science*, **3**, 415–438.

[2] Arnold, B.C. and Beaver, R.J. (2000). Hidden truncation models, *Sankhya: The Indian Journal of Statistics, Series A*, **62**, 23–35.

[3] Arnold, B.C.; Beaver, R.J.; Azzalini, A.; Balakrishnan, N.; Bhaumik, A.; Dey, D.K.; Cuadras, C.M. and Sarabia, J.M. (2002). Skewed multivariate models related to hidden truncation and/or selective reporting, *Test*, **11**, 1, 7–54.

[4] Azzalini, A. (1985). A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics*, **12**, 171–178.

[5] Azzalini, A. (2005). The skew-normal distribution and related multivariate families, *Scandinavian Journal of Statistics*, **32**, 2, 159–188.

[6] Box, G.E.P. and Wilson, K.B. (1951). On the experimental attainment of optimum conditions, *Journal of the Royal Statistics Society, Series B, Methodology*, **13**, 1–45.

[7] Esmailzadeh, N.; Talebi, H.; Momihara, K. and Jimbo, M. (2011). A new series of main effects plus one plan for 2^m factorial experiments with $m = 4\lambda \pm 1$ and $2m$ runs, *Journal of Statistical Planning and Inference*, **141**, 1567–1574.

[8] Ghosh, S.; Shirakura, T. and Srivastava, J.N. (2007). Model identification using search linear models and search designs, *Entropy, Search, Complexity*, 85–112.

[9] Ghosh, S. and Teschmacher, L. (2002). Comparisons of search designs using search probabilities, *Journal of Statistical Planning and Inference*, **104**, 439–458.

[10] LINDLEY, D.V. (1956). On a measure of the information provided by an experiment, *Ann. Math. Statist.*, **27**, 976–1005.

[11] PLACKETT, R.L. and BURMAN, J.P. (1946). The design of optimum multifactorial experiments, *Biometrika*, **33**, 305–325.

[12] SHIRAKURA, T.; TAKAHASHI, T. and SRIVASTAVA, J.N. (1996). Searching probabilities for nonzero effects in search designs for the noisy case, *The Annals of Statistics*, **24**, 2560–2568.

[13] SORENSEN, D.A. and GIANOLA, D. (2002). *Likelihood, Bayesian and MCMC methods in quantitative genetics*, Springer, New York.

[14] SRIVASTAVA, J.N. (1975). Designs for searching non-negligible effects, *A Survey of Statistical Design and Linear Models*, (J.N. Srivastava, Ed.), 507–519, North-Holland, Amsterdam.

[15] TALEBI, H. and ESMAILZADEH, N. (2011). Using Kullbak-Leibler distance for performance evaluation of search designs, *Bulletin of the Iranian Mathematical Society*, **37**, 4, 269–279.

[16] TALEBI, H. and ESMAILZADEH, N. (2011). Weighted searching probability for classes of equivalent search designs Comparison, *Comm. Statist. Theory Methods*, **40**, 4, 635–647.

[17] TALEBI, H. and ESMAILZADEH, N. (2015). Comparing search and estimation performances of designs based on compound criterion, *Statistics*, **49**, 3, 629–637.

[18] TALEBI, H. and JALALI, E. (2014). A series of single array 2^m factorial search designs for even m, *Australian and New Zealand Journal of statistics*, **56**, 4, 395–405.

[19] WU, C.J. and HAMADA, M.S. (2008). *Experiments: planning, analysis, and parameter design optimization*, John Wiley and Sons.

[20] ZHANG, Y. (2006). *Bayesian D-Optimal Design for Generalized Linear Models*, PhD Thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.