*Bulletin of the Iranian Mathematical Society Vol.* 37 *No. 4* (*2011*), *pp 269-279.*

# USING KULLBACK-LEIBLER DISTANCE FOR PERFORMANCE EVALUATION OF SEARCH DESIGNS

H. TALEBI* AND N. ESMAILZADEH

Communicated by Ahmad Reza Soltani

Abstract. This paper considers the search problem, introduced by Srivastava [10]. This is a model discrimination problem. In the context of search linear models, discrimination ability of search designs has been studied by several researchers. Some criteria have been developed to measure this capability, however, they are restricted in a sense of being able to work for searching only one possible nonzero effect. In this paper, two criteria are proposed, based on Kullback-Leibler distance. These criteria are able to evaluate the search ability of designs, without any restriction on the number of nonzero effects.

## 1. **Introduction**

Consider the following linear model for a $2^m$ factorial experiment,

$$y = X_1\beta_1 + X_2\beta_2 + e, \quad Var(e) = \sigma^2 I, \tag{1.1}$$

where $y(N \times 1)$ is a vector of observations, $X_i(N \times \nu_i)$ are known design matrices and $\beta_i(\nu_i \times 1)$ are vectors of fixed unknown factorial effects for $i = 1, 2$, $e(N \times 1)$ is an error vector, $\sigma^2$ is the error variance and $I_N$ is the identity matrix of order $N$. We know about $\beta_2$ partially. That is, at most $k$ elements of $\beta_2$ are nonzero but those elements are unknown. We

MSC(2011): Primary: 62k15; Secondary: 62k05

Keywords: Search designs, search linear model, Kullback-Leibler distance, model discrimination.

Received: 28 June 2010, Accepted: 31 October 2010.

*Corresponding author

are interested in identifying and estimating the $k$ nonzero parameters of $\beta_2$ along with the estimation of $\beta_1$. This is so called a search problem and the related design which is able to solve this problem is a search design (SD). Now, let $S$ denotes a set of $\binom{\nu_2}{k}$ partially non-nested reduced models, $M_i$, of (1.1). That is,

$$S = \{M_i : y = X_1\beta_1 + X_{2i}\beta_{2i} + e\}, \tag{1.2}$$

where $\beta_{2i}$ is the i-th possible $k \times 1$ vector of parameters of $\beta_{2i} \in \beta_2$ and $X_{2i}$ are its corresponding columns in $X_2$ for $i = 1, 2, ..., \binom{\nu_2}{k}$. Models in $S$ are common in $\beta_1$ and all differ in $\beta_{2i}$ and only one of them, say $M_0$, is true. One may discriminate between any two models in order to come up with the true model. In the context of search linear model, Srivastava [10] proposed the following condition to solve the discrimination problem

$$\text{rank}(X_1; X_{2i}; X_{2j}) = \nu_1 + 2k, \tag{1.3}$$

for any pairs of $X_{2i}$ and $X_{2j}$ in $X_2$. This condition is necessary and sufficient for the noiseless case, $\sigma^2 = 0$. However, for the noisy case, $\sigma^2 > 0$, (3) is not sufficient but is still necessary. One way to overcome this problem, proposed by Srivastava [10], is to calculate the sum of squared error ($SSE$) for each of $\binom{\nu_2}{k}$ models and choose one with minimum $SSE$ as the true model. Shirakura et al. [9] considered the stochastic properties of $SSE$ for a given search design and defined the search probability (SP) $P[SSE(M_0) < SSE(M_i)|M_0, M_i, \sigma^2]$, for any two models, in order to measure the discrimination capability of the SD. They gave an exact expression of SP for $k = 1$ when errors are normally distributed. Based on the SP, to compare the SDs, some criteria are given in [9, 5] and [11]. All of these criteria are applicable for $k = 1$. Creation and development of criterion to do the task for $k \geq 1$ is our challenge in this research. In general, some advances have been made by researchers to provide criteria for model discrimination through the design optimization. For normally distributed error, Atkinson and Fedorov [1, 2] introduced a criterion to obtain the optimum design through discrimination between two homoscedastic models, which is the so called T-optimality. Uciniski and Bogacka [12] developed it to the heteroscedastic case. Lopez-Fidalgo et al. [8] extended the optimal design criterion, based on KL distance, for non-normal models to discriminate between the true model and its alternative by considering the system of hypotheses

$$\begin{aligned} &H_0 : M = M_i, \\ &H_1 : M = M_0. \end{aligned} \tag{1.4}$$

In this paper, we proposed two new criteria, based on the Kullback-Liebler (KL) distance, in the context of search linear model (1.1). The KL distance was proposed by Kullback and Liebler [7] to measure the distance between density functions, $f_1$ and $f_2$, under two candidate models.

Now, let $f_1(y, X_{20}, \beta_{20}, \sigma)$ and $f_2(y, X_{2i}, \beta_{2i}, \sigma)$ be density functions of observations under true and alternative models, respectively. The KL distance is given by

$$(1.5)\ I_i(f_1, f_2) = \int f_1(y, X_{20}, \beta_{20}, \sigma) \log\{\frac{f_1(y, X_{20}, \beta_{20}, \sigma)}{f_2(y, X_{2i}, \beta_{2i}, \sigma)}\}dy,$$

where $dy = dy_1...dy_N$ and $y' = (y_1, ..., y_N)$. Suppose the components of the error vector $e$ are independently distributed as $N(0, \sigma^2)$. The expression (1.5) is reduced to the following explicit form:

$$(1.6) \quad I_i(f_1, f_2) = \frac{1}{2\sigma^2}(X_{20}\beta_{20} - X_{2i}\beta_{2i})'(X_{20}\beta_{20} - X_{2i}\beta_{2i}).$$

It means, under the normality assumption the KL distance is simplified to a quantity which measures the distance between means of two models. The minimum discrimination distance function, minimizes (1.6) with respect to all vector $\beta_{2i}$. The minimum value of (1.6) occurs at $\beta_{2i} = (X'_{2i}X_{2i})^{-1}X'_{2i}X_{20}\beta_{20}$ and is given by

$$(1.7) \qquad I_i(\rho, X_{20}) = \min_{\beta_{2i}} I_i(f_1, f_2) = \frac{1}{2}\rho' X'_{20}(I - H_i)X_{20}\rho,$$

where $\rho = \frac{1}{\sigma}\beta_{20}$ and $H_i = X_{2i}(X'_{2i}X_{2i})^{-1}X'_{2i}$. This is the T-optimality criterion obtained by Atkinson and Fedorov [1], [2] and is the so-called non-centrality parameter of the likelihood ratio test for lack of fit of postulating $M_i$ when the true model is $M_0$. [8]

In next two sections, we will use (1.7) to propose new dual-job KL-criterion for discriminating models and evaluate the search capability of SDs for $k = 1$ and 2. The expected KL-criterion (EKL) is given in Section 4. It will allow us to evaluate the search performance of design through the discrimination job for all $k \geq 1$.

## 2. **Kullback-Leibler search criterion**

Minimum discrimination distance given in (1.7) provides a measure to discriminate between two candidate models, $M_0$ and $M_i$. Now, in view of searching the true model consider the $\binom{\nu_2}{k}$ rival models of the set $S$ given in (1.2). Any model in $S$ can be the true model, $M_0$, given by,

$$M_0 : y = X_1\beta_1 + X_{20}\beta_{20} + e. \tag{2.1}$$

Let $S_0$ denotes the set of all $\binom{\nu_2}{k}$-1 possible alternative $M_i(\neq M_0)$, of which at least one effect in $\beta_{2i}$ is not of $\beta_{20}$. To discriminate the true model based on (1.7), we consider a two-stage procedure and propose the Kullback-Leibler search criterion (KL-criterion), for a given search design $T$, by

$$KL_T(\rho) = \min_{S} \min_{S_0} I_i(\rho, X_{20}). \tag{2.2}$$

Note that $KL_T(\rho)$ is a real valued function of vector $\rho$, which depends on design through $X_{20}$ and $X_{2i}$. From T-optimality property given in (1.7), it is most desirable that $KL_T(\rho)$ be a large value. This means that a search design $T$ with larger value of $KL_T(\rho)$ has a larger capability in discriminating the true model among all competing models. Therefore, having (2.2) we are able to compare and rank the SDs with respect to their discrimination capacity. In the following, we will employ $KL_T(\rho)$ in more details for evaluating search ability and ranking search designs. Before going into more details, we give the following definition.

**Definition 2.1.** *Let $T_1$ and $T_2$ be two search designs with N treatments. Then $T_1$ is said to be better than $T_2$ for discriminating the true model if $KL_{T_1}(\rho) > KL_{T_2}(\rho)$.*

Indeed, by this definition, we choose a design that has the better discrimination capability in the worst case. The following lemma will be useful in the next section.

**Lemma 2.2.** *The minimum discrimination distance given in* (1.7) *is bounded above by $\frac{1}{2}\rho' X_{20}' X_{20}\rho$.*

*Proof.* The proof is easily obtained from the positive semi-definite property of the matrix $X_{20}' H_i X_{20}$. □

## 3. Implementation

In this section, we employ KL-criterion in (2.2) for comparing some search designs.

3.1. **Case $k = 1$.** Consider a SD with $N$ treatments and let $k = 1$. It can be easily seen that $I_i(\rho, X_{20}) = I_i(-\rho, X_{20})$, i.e., $I_i$ is symmetric in $\rho$. This will allow us to focus on $\rho \geq 0$ for comparing search designs. It is also clear from Lemma 2.2 that $I_i(\rho, X_{20}) \leq N\rho^2$. Note that, in comparing SDs for this case the $\rho$ will be canceled out and hence the actual value it takes immaterial. We put $\rho = 1$ in using Definition 2.1 for design comparison. Now, we can establish the following proposition.

**Proposition 3.1.** *Consider a search design $T$ with $N$ runs. The following properties of $KL_T(\rho)$ are true:*
*a)* $KL_T(\rho) = KL_T(-\rho)$.
*b) for* $\rho = 1, \;\; 0 \leq \frac{1}{N} KL_T(1) \leq 1$.

The proof is obtained from the above discussion.

Based on Proposition 3.1, we use (2.2) to measure the search ability of designs $D_1$, $D_2$ and $D_3$, given in Table 1 and then rank them. Design

TABLE 1. $D_1$, $D_2$, $D_3$ and $D_4$ with 12 runs

| $D_1$ | | | | | $D_2$ | | | | | $D_3$ | | | | | $D_4$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| + | + | + | + | + | + | - | - | - | - | + | + | + | + | + | + | + | + | + |
| - | - | - | + | - | - | + | - | - | - | + | - | - | - | - | - | - | - | - |
| + | - | - | - | + | - | - | + | - | - | - | + | - | - | - | - | - | - | + |
| + | + | - | - | - | - | - | - | + | - | - | - | + | - | - | - | - | + | - |
| + | + | + | - | - | - | - | - | - | + | - | - | - | + | - | - | + | - | - |
| - | + | + | + | - | + | - | + | + | + | - | - | - | - | + | + | - | - | - |
| + | - | + | + | + | - | + | + | + | - | - | + | - | + | + | - | - | + | + |
| - | + | - | + | + | + | + | + | - | - | + | - | + | + | + | - | + | - | + |
| - | - | + | - | + | + | + | - | - | + | - | + | + | + | - | + | - | - | + |
| + | - | - | + | - | + | - | - | + | - | + | + | + | - | - | - | + | + | - |
| - | + | - | - | + | - | - | + | - | + | + | + | - | - | + | + | - | + | - |
| - | - | + | - | - | - | + | - | + | + | + | - | - | + | - | + | + | - | - |

$D_1$ is a projection of Plackett-Burman design with 12 runs ($PB_{12}$) onto 5 factors. $D_2$ is given in [4] and $D_3$ in [11]. In model (1.1), we take $\beta_1$ as the vector of the general mean and main effects, and $\beta_2$ is restricted to 2- and 3- factor interactions. We also assume that 4 and higher order interactions are negligible. Designs $D_1$, $D_2$ and $D_3$ satisfy condition (1.3) for $k = 1$. The values of $KL_{D_i}(1)$ and $\frac{1}{N} KL_{D_i}(1)$ are given in Table 2 which shows that

$$KL_{D_1} > KL_{D_2} > KL_{D_3}.$$

It means, $D_1$ is superior to $D_2$ and $D_2$ is superior to $D_3$ in identifying the true model. The second row in Table 2 shows the discrimination capacity of these designs compared to its upper bound 1, given in Proposition 3.1(b).

TABLE 2. Comparisons between designs $D_1$, $D_2$ and $D_3$

| | $D_1$ | $D_2$ | $D_3$ |
|---|---|---|---|
| $KL_{D_i}(1)$ | 10.667 | 9.0000 | 6.6667 |
| $\frac{1}{N}KL_{D_i}(1)$ | 0.8889 | 0.7500 | 0.5556 |

3.2. **Case $k=2$.** As already mentioned, none of the criteria given for measuring and comparing the search ability of designs, in the context of search linear models, are applicable for the case $k>1$. An exception is given by Chatterjee et al. [3], in which, they consider $k=2$ for comparing two-level supersaturated designs, using the SP given in [9]. The KL-criterion is applicable for $k \geq 2$. Here, we apply $KL_T(\rho)$ for designs $D_4$ and $D_5$, in order to measure their capability in searching and finding the superior design. Design $D_4$, presented in Table 1 is given in [5]. $D_5$ is any one of the designs that is chosen from the class of isomorphic projection designs of $PB_{12}$ onto any four factors. These two designs satisfy the search condition (1.3) for $k=2$, where $\beta_i$, $i=1,2$ are the same as those considered in Section 3.1. The values of $KL_{D_i}(\rho)$, for various values of $\rho=(\rho_1,\rho_2)$ are shown in Table 3. Results show that $D_4$ is superior in comparison to $D_5$. For $-2 \leq \rho_i \leq 2$, $i=1,2$, the plots of $KL_T(\rho)$ are presented in Figure 1, (a) and (b) for $D_4$ and $D_5$, respectively. The superimposed contour of these plots, at a given

TABLE 3. Comparisons between designs $D_4$ and $D_5$

| $\rho_2$ | -2 | -1.5 | -1 | -0.5 | 0.5 | 1 | 1.5 | 2 |
|---|---|---|---|---|---|---|---|---|
| $\rho_1$ | | | | $D_4$ | | | | |
| -2 | 40.2286 | 22.6286 | 10.0571 | 2.5143 | 2.5143 | 10.0571 | 22.6286 | 40.2286 |
| -1.5 | 22.6286 | 22.6286 | 10.0571 | 2.5143 | 2.5143 | 10.0571 | 22.6286 | 22.6286 |
| -1 | 10.0571 | 10.0571 | 10.0571 | 2.5143 | 2.5143 | 10.0571 | 10.0571 | 10.0571 |
| -0.5 | 2.5143 | 2.5143 | 2.5143 | 2.5143 | 2.5143 | 2.5143 | 2.5143 | 2.5143 |
| | | | | $D_5$ | | | | |
| -2 | 37.3333 | 21 | 9.3333 | 2.3333 | 2.3333 | 9.3333 | 21 | 37.3333 |
| -1.5 | 21 | 21 | 9.3333 | 2.3333 | 2.3333 | 9.3333 | 21 | 21 |
| -1 | 9.3333 | 9.3333 | 9.3333 | 2.3333 | 2.3333 | 9.3333 | 9.3333 | 9.3333 |
| -0.5 | 2.3333 | 2.3333 | 2.3333 | 2.3333 | 2.3333 | 2.3333 | 2.3333 | 2.3333 |

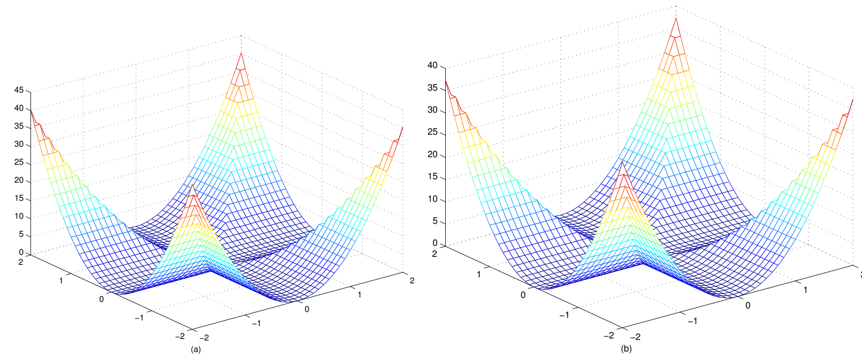


FIGURE 1. Plot of KL for $D_4$ and $D_5$

$KL_T(\rho)$ level, reveals the superiority of the design. That is, a design with interior contour is concordant with a higher $KL_T(\rho)$ at a given $\rho$. The contour plot for levels 4 and 8 are shown in Figure 2.

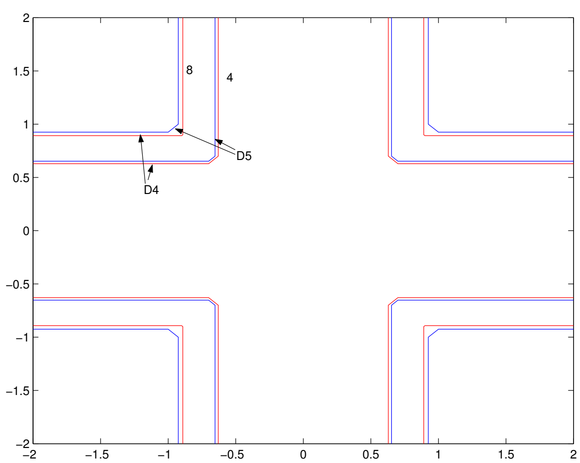


FIGURE 2. Contour plot of KL for $D_4$ and $D_5$

## 4. Expected KL-criterion

In design comparison, the effect size, $\rho$ is canceled out from KL-criterion for $k = 1$. However, there is a serious problem due to the dependence of KL-criterion on $\rho$ for $k > 1$. Therefore, the comparison of search designs need to be done for many given different values of $\rho$. This is a tedious and almost impossible task. In order to overcome this problem, we develop a new criterion by averaging (1.7) over $\rho$, through a weight real valued function $f(\rho)$. That is, consider $\rho$ as a $k \times 1$ random vector with the multivariate probability density function ($pdf$) $f(\rho)$ as a weight function. We propose the following quantity

$$EKL_i(X_{20}) = \int I_i(\rho, X_{20}) f(\rho) d\rho. \tag{4.1}$$

Note that $EKL_i(X_{20})$ depends on the design, through $X_{20}$ and $H_i$ given in (1.7), but is independent of $\rho$. To simplify (4.1), let rewrite the quadratic form $I_i(\rho, X_{20})$ in (1.7) as follow:

$$I_i(\rho, X_{20}) = \sum_{i=1}^{k} \rho_i^2 h_{ii} + \sum\sum_{i \neq j} \rho_i \rho_j h_{ij}, \tag{4.2}$$

where $h_{ij}$ is the $(i, j)$-th, element of the matrix $H_i(X_{20}) = X'_{20}(I - H_i)X_{20}$; $i, j = 1, 2, ..., k$. For i.i.d. $\rho_i$'s, the $EKL_i(X_{2i})$ is reduced to an explicit form of

$$EKL_i(X_{20}) = Trace(H_i(X_{20}))var(\rho_.) + \sum_i \sum_j h_{ij} E^2(\rho_.), \tag{4.3}$$

where $E(\rho_.)$ and $var(\rho_.)$ are common mean and variance of $\rho_i$'s, respectively. For a given design $T$, the new modified criterion, say expected Kullback-Leibler (EKL), is defined as

$$EKL_T = \min_{S} \min_{S_0} EKL_i(X_{20}). \tag{4.4}$$

Larger value of $EKL_T$ supports higher ability of design $T$ in identifying the true model among a set of candidate models. One may use $EKL_T$ to compare and rank the search designs. Let us to have the following definition:

**Definition 4.1.** *Let $T_1$ and $T_2$ be two search designs with N treatments. Then $T_1$ is said to be better than $T_2$ for discriminating the true model if $EKL_{T_1}(\rho) > EKL_{T_2}(\rho)$.*

Note that the non-negative quantity $I_i(\rho, X_{20})$ is a convex function in $\rho$, due to positive semi-definiteness of matrix $H_i(X_{20})$. The comparison of designs for large absolute values of $\rho_i$'s makes no sense, since KL-criterion shows high discrimination ability independent of designs. Thus, one may prefer a design that allocate a high discrimination measure to small significant true effects. Naturally, we may choose a weight function $f(\rho)$ such that allocates larger weights to the lower values of true effects. The following weight function may work adequately,

$$f_i(\rho_i) = \begin{cases} \frac{1}{2}g(-\rho_i) & if \quad \rho_i < 0 \\ \frac{1}{2}g(\rho_i) & if \quad \rho_i > 0 \end{cases} \tag{4.5}$$

where $g(\rho_i)$ is *pdf* of a Gamma random variable with mean $\upsilon\lambda$. For weight function (4.5), the quantity (4.3) reduces to $EKL_i(X_{20}) = (\upsilon\lambda^2 + \upsilon^2\lambda^2)Trace(H_i(X_{20}))$. Then, to rank the competing designs, one may cancel out the fixed coefficient $(\upsilon\lambda^2 + \upsilon^2\lambda^2)$ and make comparison using the term $Trace(H_i(X_{20}))$. That is, the $EKL_T$ in (4.4) reduces to

$$EKL_T = \min_S \min_{S_0} Trace(H_i(X_{20})). \tag{4.6}$$

Clearly, in (4.6), $EKL_T \leq kN$. Now, we use (4.6) to compare search designs $D_4$ and $D_5$ for $k = 2$ once more. The values of (4.6) for designs $D_4$ and $D_5$ are 10.0571 and 9.3333, respectively. Then, by Definition 4.1, $D_4$ is better than $D_5$ for discriminating the true model.

## 5. Discussion

Designing an efficient experiment is very important for making inference on the model parameters. Most of the time, designing is done on a known model. However, sometimes the underlying model is unknown. In this situation, the challenge is to find a design to discriminate between rival models efficiently. Search design, introduced by Srivastava [10], is able to solve this problem for partially non-nested unknown models. For a noisy case, $\sigma^2 > 0$, this gets harder. Based on comparison of SSE of models for discrimination, Shirakiura et al. [9] developed the search probability (SP) to evaluate the searching ability of proposed search designs. This criterion and the others proposed by researchers are based on SP, and are able to measure searching capability of designs only for the case $k = 1$. In this paper, we managed to overcome this restriction problem. Following [8] in presenting the KL-optimality criterion to obtain the optimal design, we developed two new criteria, based on the Kullabck-Leibler distance in the context of search linear model in (1.1). We denoted these by KL and EKL criteria. These criteria coincide with T-optimality criterion, which is used to obtain optimum designs for discriminating between models under normality assumption of errors, introduced in [1] and [2]. T-optimality criterion function is the so-called non-centrality parameter of the likelihood test for the system of hypothesis given in (1.4). It is known, that the power function of test is a non-decreasing function of non-centrality parameter. That is, a design with higher KL-criterion and EKL results in the higher of the power for testing two alternative models. This means that a superior design resulted from KL-criterion and EKL is concordant with a higher performance design, through power value criterion given in [6]. Note that, in Section 4, we considered the influence of nuisance parameter, $\sigma^2$, on the performance evaluation and comparison of designs, through the *pdf* $f(\rho)$. That is, comparing the candidate designs at lower values of $\rho_i$s, i.e., higher values of noise, makes more sense. To confirm the reliability of our new criteria, we employed them to calculate and evaluate the search capability of some search designs.

## Acknowledgments

The authors would like to thank anonymous referee for his valuable comments and an associate editor for his attention and care.

## References

[1] A. C. Atkinson and V. V. Fedorov, The designs of experiments for discriminating between two rival models, *Biometrika* **62** (1975a) 57-70.

[2] A. C. Atkinson and V. V. Fedorov, Optimal design: experiments for discriminating between several models, *Biometrika* **62** (1975b) 289-303.

[3] K. Chatterjee, A. Sarkar and D. K. J. Lin, Supersaturated designs with high searching probability, *J. Statist. Plann. Inference* **138** (2008) 272-277.

[4] N. Esmailzadeh, H. Talebi, K. Momihara and M. Jimbo, A new series of main effect plus one plan for $2^m$ factorial experiments with $m = 4\lambda \pm 1$ and $2m$ runs, *J. Statist. Plann. Inference* **141** (2011), no. 4, 1567-1574.

[5] S. Ghosh and L. Teschmacher, Comparisons of search designs using search probabilities, *J. Statist. Plann. Inference* **104** (2002) 439-458.

[6] S. Ghosh, H. Deng and Y. Luan, Effects of noise in performance comparisons of designs for model identification and discrimination, *J. Statist. Plann. Inference* **137** (2007) 3871-3881.

[7] S. Kullback and R. A. Leibler, On information and sufficiency, *Ann. Math. Statist.* **22** (1951) 79-86.

[8] J. Lopez-Fidalgo, C. Tommasi and P.C. Trandafir, An optimal experimental design criterion for discrimination between non-normal models, *J. R. Statist. Soc. B* **69** (2007) 231-242.

[9] T. Shirakura, T. Takahashi and J. N. Srivastava, Searching probabilities for case, *Ann. Statist.* **24** (1996), no. 6, 2560-2568.

[10] J. N. Srivastava, Designs for searching non-negligible effects, In: Srivastava, J.N. (Ed.), A Survey of Statistical Design and Linear Models. North-Holland, Elsevier, Amsterdam, (1975) 507-519.

[11] H. Talebi and N. Esmailzadeh, Weighted searching probability for classes of equivalent search designs Comparison, *Comm. Statist. Theory Methods* **40:4** (2011) 635-647.

[12] D. Ucinski and B. Bogacka, T-optimum designs for multiresponse heteroscedastic models, In Proc. mODa 7-Advances in Model-oriented Design and Analysis (eds A. Di Bucchianico, H. Luter and H. P. Wynn), Heidelberg: Physica (2004) 191-199.

**Hooshang Talebi**

Department of Statistics, University of Isfahan, P.O. Box 81746, Isfahan, Iran

Email: `h_talebi@sci.ui.ac.ir`

**Nabaz Esmailzadeh**

Department of Statistics, University of Kurdistan, P.O. Box 66177, Sanandaj, Iran

Email: `N.Esmailzadeh@uok.ac.ir`