Applied Mathematics. 2011; 1(2): 65-68
DOI: 10.5923/j.am.20110102.11

Optimal Designs in a Simple Linear Regression with Skew-Normal Distribution for Error Term

Habib Jafari[1,*], Reza Hashemi[2]

[1,2]Department of Statistics, Razi University, Kermanshah, Iran

Abstract The locally D-optimal design was derived for simple linear regression with the error term of Skew-Normal distribution. In this paper, to obtain a D-optimal design, the locally D-optimal criterion was considered, because of depending the information matrix on unknown parameters.

Keywords Information Matrix, Skew-Normal Distribution, Locally D-Optimal Criterion, Locally A-Optimal Criterion, Locally D- And A-Optimal Design

1. Introduction

There are many papers which discuss the optimal design for simple linear regression when the error terms have normal distribution. In this paper, the Skew-Normal distribution was considered for error term. The central role of general random variables in probability and statistics is well-known and can be traced to the simplicity of the functional forms, basic symmetry properties of the probability density function (pdf) and cumulative function (cdf) of the standard normal random variable, *Z*;

$$\phi(z) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{z^2}{2}\right), \Phi(z) = \int_{\infty}^{z} \phi(x)dx \quad (1)$$

Now, the following function can be considered:

$$f_\lambda(x) = 2\phi(x)\Phi(\lambda x) \quad (2)$$

is a bona fide pdf of a random variable *X* which inherits a few features of the normal random variables. Some of these features happen to be the ones which make the normal distribution the darling of statistical inference.

The class of distribution (2) was introduced in**[2]** and Christened Skew-Normal distribution with the skewness parameter λ, in symbol X~ SN (λ). The right-tails of these distributions are virtually indistinguishable for $\lambda > 2$; thus, in this paper, only optimal design was discussed for $\in [-2, +2]$. Some properties of this kind of distribution can be found in**[7].**

In this paper, a simple linear regression model was considered as well as $y = \beta_0 + \beta_1 x + \varepsilon$, where the error term had skew-normal distribution with- parameter λ; meaning that, $\varepsilon \sim SN(\lambda)$. In this situation, *Y* has also skew-normal distribution with the following pdf;

$$f_Y(y; \boldsymbol{\theta}) = 2\phi(y - \beta_0 - \beta_1 x)\Phi\big(\lambda(y - \beta_0 - \beta_1 x)\big); x \in \boldsymbol{\mathcal{X}} \quad (3)$$

where $\boldsymbol{\theta} = (\beta_0, \beta_1, \lambda)^T$**.** As was already written, is obtaining of this paper was to obtain the locally *D*-optimal design of this model based on the unknown parameter vector $\boldsymbol{\theta}$**.**

In this paper, there was concentration on the criterion dependence on the variance of parameter estimator. As is known, the variance of parameter estimator (ML) is inversely proportional to the information matrix**[2].** Thus, there have been searched designs maximizing the information on the estimates as represented in the Fisher information matrix in $\mathbf{I}(\boldsymbol{\theta}; \xi)$, where ξ denotes a design.

The outline of the paper is as follows. In Section 2, the information matrix, the locally *D*-optimal criterion which is a function of the information matrix and the locally *D*-optimal design for model (3) are introduced. At last, conclusion is made in Section 3.

2. Locally D-optimal Design

To obtain the *D*-optimal deign, the information matrix should be known, which was calculated using the derivative degree two of the log-likelihood function. In this paper, the information matrix was obtained based on the following log-likelihood function according to model (3);

$$\ell(\boldsymbol{\theta}; x) = \ln(2) + \ln\big(\phi(y - \beta_0 - \beta_1 x)\big) + \ln\Big(\Phi\big(\lambda(y - \beta_0 - \beta_1 x)\big)\Big) \quad (4)$$

At first, since the information matrix should be calculated for one observation, (for one observation) was obtained by;

$$\mathbf{I}(\boldsymbol{\theta}; x) = -E\left(\frac{\partial^2 \ell(\boldsymbol{\theta}; x)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right) = \begin{pmatrix} I_{\beta_0\beta_0} & I_{\beta_0\beta_1} & I_{\beta_0\lambda} \\ & I_{\beta_1\beta_1} & I_{\beta_1\lambda} \\ & & I_{\lambda\lambda} \end{pmatrix} \quad (5)$$

where the elements of the symmetry information matrix

* Corresponding author:
jafari_habib@yahoo.com (Habib Jafari)
Published online at http://journal.sapub.org/am

(5) were as follows;

- $I_{\beta_0\beta_0} = -E\left(\frac{\partial^2 \ell(\boldsymbol{\theta};x)}{\partial \beta_0^2}\right) = 1 - \frac{1}{2}\lambda^3 a_1 + \lambda^2 b_0,$
- $I_{\beta_1\beta_1} = -E\left(\frac{\partial^2 \ell(\boldsymbol{\theta};x)}{\partial \beta_1^2}\right) = x^2(1 - \lambda^3 a_1 + \lambda^2 b_0),$
- $I_{\lambda\lambda} = -E\left(\frac{\partial^2 \ell(\boldsymbol{\theta};x)}{\partial \lambda^2}\right) = (-\lambda a_3 + b_2),$
- $I_{\beta_0\beta_1} = -E\left(\frac{\partial^2 \ell(\boldsymbol{\theta};x)}{\partial \beta_0 \partial \beta_1}\right) = x(1 - \lambda^3 a_1 + \lambda^2 b_0),$
- $I_{\beta_0\lambda} = -E\left(\frac{\partial^2 \ell(\boldsymbol{\theta};x)}{\partial \beta_0 \partial \lambda}\right) = (-\lambda^2 a_2 + a_0 + \lambda b_1),$
- $I_{\beta_1\lambda} = -E\left(\frac{\partial^2 \ell(\boldsymbol{\theta};x)}{\partial \beta_1 \partial \lambda}\right) = x(-\lambda^2 a_2 + a_0 + \lambda b_1),$

such that;

$$a_{k_1} = a_{k_1}(\beta_0, \beta_1, \lambda, x) = = E\left((-Y + \beta_0 + \beta_1 x)^{k_1} . \frac{\phi(\lambda(Y - \beta_0 - \beta_1 x))}{\Phi(\lambda(Y - \beta_0 - \beta_1 x))}\right); \quad k_1 = 0,1,2,3$$

$$b_{k_2} = b_{k_2}(\beta_0, \beta_1, \lambda, x) = = E\left((-Y + \beta_0 + \beta_1 x)^{k_2} . \left(\frac{\phi(\lambda(Y - \beta_0 - \beta_1 x))}{\Phi(\lambda(Y - \beta_0 - \beta_1 x))}\right)^2\right); \quad k_2 = 0,1,2$$

See Appendix A1.

Especially, suppose $\boldsymbol{\mathcal{X}} = [-1, +1]$ as design space and $\beta_0, \beta_1 \in [-5, +5], \lambda \in [-2, +2]$. Now, for all the values of (β_0, β_1, x), the following can be written;

- Two amounts of a_1 and a_3 are equal to zero and $a_0 (> 0)$ increase for $\lambda \in [-2, 0]$ and decrease for $\lambda \in [0, +2]$. There exists a similar position for $a_2 (> 0)$, where;

$$a_0(\beta_0, \beta_1, 0, x) = \max_{\lambda \in [-2,+2]} a_0(\beta_0, \beta_1, \lambda, x) = 0.39892,$$
$$a_2(\beta_0, \beta_1, 0, x) = \max_{\lambda \in [-2,+2]} a_2(\beta_0, \beta_1, \lambda, x) = 0.39892.$$

- $b_0 (> 0)$ increases for $\lambda \in [-2, 0]$ and decreases for $\lambda \in [0, +2]$. There exists a similar position for $b_2 (> 0)$, where;

$$b_0(\beta_0, \beta_1, 0, x) = \max_{\lambda \in [-2,+2]} b_0(\beta_0, \beta_1, \lambda, x) = 0.63662$$
$$b_2(\beta_0, \beta_1, 0, x) = \max_{\lambda \in [-2,+2]} b_2(\beta_0, \beta_1, \lambda, x) = 0.63662$$

In this case, b_1 increases as λ increases and $b_1(\beta_0, \beta_1, 0, x) = 0.0000$. After proposing the two above properties (items 1 and 2), an optimal design should be obtained for model (3). As is known, there are many optimality criteria for obtaining an optimal design such that D- and A-optimality criteria which are functions of the information matrix (5) and shown by the following notations[1];

$$\psi_D(\boldsymbol{\theta}; \xi) = -\ln\left(det(\mathbf{I}(\boldsymbol{\theta}; \xi))\right) \;\&$$
$$\psi_A(\boldsymbol{\theta}; \xi) = tr(\mathbf{I}^{-1}(\boldsymbol{\theta}; \xi)) \tag{7}$$

where ξ denotes a design with two components; the first components are some values of design space $\boldsymbol{\mathcal{X}}$ and the weight of them are the second components, so that design ξ can be defined as follows;

$$\xi = \begin{pmatrix} x_1 & x_2 & \dots & x_m \\ w_1 & w_2 & \dots & w_m \end{pmatrix} \in \Xi \tag{8}$$

where $\Xi = \{(x, w) | 0 \le w_i \le 1; \sum_{i=1}^{m} w_i = 1, x \in \boldsymbol{\mathcal{X}}\}$, $\mathbf{I}(\boldsymbol{\theta}; \xi) = \sum_{i=1}^{m} w_i . \mathbf{I}(\boldsymbol{\theta}; x_i)$ and $p \le m \le \frac{p(p+1)}{2}$ (p denotes the number of parameters)[5]. The design is called the saturated design when m=p.

In Table 1, an optimal design is shown for model (3) based on the information matrix (5). In this case, for different amounts of parameter $\lambda \in [-2, +2]$ and every value of β_0, β_1 in the interval [-5, +5] and also the equivalence Theorem[6] (***See Appendix A2***), the following can be written[9];

$$tr\left(\mathbf{I}(\boldsymbol{\theta}; x)\mathbf{I}^{-1}(\boldsymbol{\theta}; \xi^*)\right) = 2 + x^2;\ x \in \boldsymbol{\mathcal{X}} = [-1, +1], \forall \boldsymbol{\theta} \tag{10}$$

Table 1. Locally D- and A-optimal design for some values of $\boldsymbol{\lambda}$ and any values of the other parameters

$\boldsymbol{\lambda}$	$\xi = \begin{pmatrix} \boldsymbol{x_1} & \boldsymbol{x_2} & \boldsymbol{x_3} \\ \boldsymbol{\alpha_1} & \boldsymbol{\alpha_2} & \boldsymbol{\alpha_3} \end{pmatrix}$	$\boldsymbol{det}(\mathbf{I}(\xi^*))$	$\boldsymbol{tr}(\mathbf{I}^{-1}(\xi^*))$
0.00	$\xi^* = \begin{pmatrix} \mathbf{-1.0} & \mathbf{0.0} & \mathbf{+1.0} \\ \mathbf{0.5} & \mathbf{0.0} & \mathbf{0.5} \end{pmatrix}$	0.00001	218048.37666
-0.10 (+0.10)		0.00798	206.80126
-0.20 (+0.20)		0.03072	55.51984
-0.50 (+0.50)		0.14759	13.60039
-0.80 (+0.80)		0.25024	9.74240
-0.90 (+0.90)		0.27128	9.60189
-0.95 (+0.95)		0.27919	9.64598
-0.97 (+0.97)		0.28200	9.67855
-0.99 (+0.99)		0.28449	9.72339
-1.00 (+1.00)		0.28567	9.74842
-1.10 (+1.10)		0.29419	10.12550
-1.15 (+1.15)		0.29660	10.38839
-1.18 (+1.18)		0.29747	10.56942
-1.19 (+1.19)		0.29772	10.63339
-1.20 (+1.20)		0.29788	10.69999
-1.30 (+1.30)		0.29772	11.45680
-1.40 (+1.40)		0.29459	12.38894
-1.50 (+1.50)		0.28927	13.49405
-1.53 (+1.53)		0.28726	13.85854
-1.54 (+1.54)		0.28669	13.98369
-1.55 (+1.55)		0.28602	14.10990
-1.57 (+1.57)		0.28462	14.36935
-2.00 (+2.00)		0.24804	21.72277

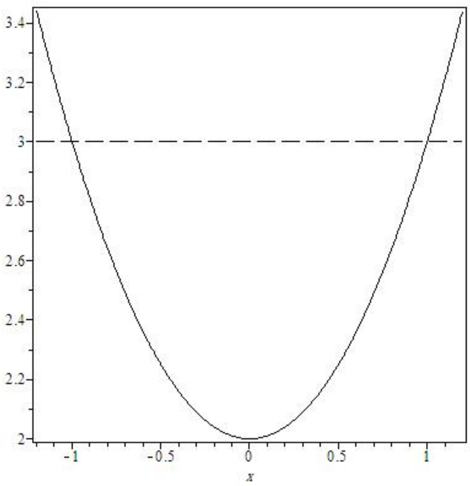

Figure 1. $tr\left(\mathrm{I}(\theta_0;x)\mathrm{I}^{-1}(\theta_0;\xi^*)\right)$

According to Equation (10), Figure 1 shows that in the two points $x=-1.0,+1.0$, $tr\left(\mathbf{I}(\boldsymbol{\theta};x)\mathbf{I}^{-1}(\boldsymbol{\theta};\xi^*)\right)$ is equal to 3; but, for $x=0.0$, this quantity is less than 3 (the number of parameters). Then, it can be said that the following locally *D*-optimal design exists with two support points -1, +1;

$$\xi^*=\begin{pmatrix}-1.0 & +1.0\\ 0.5 & 0.5\end{pmatrix}.$$

In this case, it can be also seen that the maximum of the determinant of the information matrix exists when $\lambda=-1.2, 1.2$ in Table 1.

3. Conclusions

In this paper, the Skew-normal distribution was considered for error term in simple linear regression. In this kind of model, there are three parameters, two of which are related to the regression model and one is the parameter of Skew-normal distribution. Then, based on these three parameters and *Caratheodory*'s theorem[3], a design with three support points assumed. To obtain an optimal design, the *D*-optimal criterion was considered. In this case, due to the dependence of the information matrix on unknown parameters, the locally *D*-optimal design was obtained[8].

In this situation, there was only one locally *D*-optimal design for every value of the parameters $\beta_0,\beta_1\in[-5,+5]$ and for different values of $\lambda\in[-2,+2]$. This result is shown in Table 1, where $\lambda=1.2$ maximizes the determinate of the information matrix. Also, it was shown that locally *A*-optimal design was the same as locally *D*-optimal design, where $\lambda=0.99$ minimizes $tr\left(\mathbf{I}^{-1}(\xi^*)\right)$ (Table 1).

Appendix A1

To calculate the elements of the information matrix (5), the derivatives of the log-likelihood function (4) with respect to three parameters β_0,β_1 and λ is needed as follows;

$$\frac{\partial^2\ell(\theta;x)}{\partial\beta_0^2}=-1+\lambda^3(-y+\beta_0+\beta_1x).\frac{\phi\left(\lambda(y-\beta_0-\beta_1x)\right)}{2\Phi\left(\lambda(y-\beta_0-\beta_1x)\right)}-\lambda^2.\left(\frac{\phi\left(\lambda(y-\beta_0-\beta_1x)\right)}{\Phi\left(\lambda(y-\beta_0-\beta_1x)\right)}\right)^2$$

$$\frac{\partial^2\ell(\theta;x)}{\partial\beta_1^2}=-x^2+\lambda^3x^2(-y+\beta_0+\beta_1x).\frac{\phi\left(\lambda(y-\beta_0-\beta_1x)\right)}{\Phi\left(\lambda(y-\beta_0-\beta_1x)\right)}-\lambda^2x^2.\left(\frac{\phi\left(\lambda(y-\beta_0-\beta_1x)\right)}{\Phi\left(\lambda(y-\beta_0-\beta_1x)\right)}\right)^2$$

$$\frac{\partial^2\ell(\theta;x)}{\partial\lambda^2}=\lambda(-y+\beta_0+\beta_1x)^3.\frac{\phi\left(\lambda(y-\beta_0-\beta_1x)\right)}{\Phi\left(\lambda(y-\beta_0-\beta_1x)\right)}-(-y+\beta_0+\beta_1x)^2.\left(\frac{\phi\left(\lambda(y-\beta_0-\beta_1x)\right)}{\Phi\left(\lambda(y-\beta_0-\beta_1x)\right)}\right)^2$$

$$\frac{\partial^2\ell(\theta;x)}{\partial\beta_0\partial\beta_1}=-x+\lambda^3x(-y+\beta_0+\beta_1x).\frac{\phi\left(\lambda(y-\beta_0-\beta_1x)\right)}{\Phi\left(\lambda(y-\beta_0-\beta_1x)\right)}-\lambda^2x.\left(\frac{\phi\left(\lambda(y-\beta_0-\beta_1x)\right)}{\Phi\left(\lambda(y-\beta_0-\beta_1x)\right)}\right)^2$$

$$\frac{\partial^2\ell(\theta;x)}{\partial\beta_0\partial\lambda}=(\lambda^2(-y+\beta_0+\beta_1x)^2-1).\frac{\phi\left(\lambda(y-\beta_0-\beta_1x)\right)}{\Phi\left(\lambda(y-\beta_0-\beta_1x)\right)}-\lambda(-y+\beta_0+\beta_1x).\left(\frac{\phi\left(\lambda(y-\beta_0-\beta_1x)\right)}{\Phi\left(\lambda(y-\beta_0-\beta_1x)\right)}\right)^2$$

$$\frac{\partial^2\ell(\theta;x)}{\partial\beta_1\partial\lambda}=(\lambda^2x(-y+\beta_0+\beta_1x)^2-x).\frac{\phi\left(\lambda(y-\beta_0-\beta_1x)\right)}{\Phi\left(\lambda(y-\beta_0-\beta_1x)\right)}-\lambda x(-y+\beta_0+\beta_1x).\left(\frac{\phi\left(\lambda(y-\beta_0-\beta_1x)\right)}{\Phi\left(\lambda(y-\beta_0-\beta_1x)\right)}\right)^2$$

Appendix A2

Theorem (Equivalence Theorem)[9]: Based on the design $\boldsymbol{\xi}$ the following three items are equivalent:

- $\boldsymbol{\xi}^*$ is the locally D-optimal design if: $\boldsymbol{\xi}^*=\mathbf{arg\,min}_{\xi\in\Xi}\,\boldsymbol{\psi_D}(\boldsymbol{\theta_0};\boldsymbol{\xi})$, where $\boldsymbol{\theta_0}$ denotes the true value of parameters.
- $\boldsymbol{tr}\left(\mathbf{I}(\boldsymbol{\theta_0};\boldsymbol{x})\mathbf{I^{-1}}(\boldsymbol{\theta_0};\boldsymbol{\xi}^*)\right)\le\boldsymbol{p};\forall\boldsymbol{x}\in\boldsymbol{\mathcal{X}}$ (G-optimality criterion), *p* denotes the number of parameters.
- $\boldsymbol{tr}\left(\mathbf{I}(\boldsymbol{\theta_0};\boldsymbol{x}^*)\mathbf{I^{-1}}(\boldsymbol{\theta_0};\boldsymbol{\xi}^*)\right)=\boldsymbol{p}$, where $\boldsymbol{x}^*$ is the support points.

REFERENCES

[1] A. Atkinson, A. Donev and R. Tobias. (2007), Optimum experimental designs with SAS, Oxford Uni. Press

[2] A. Azzalini. (1985). A class of distribution which includes the normal ones, *Scand. J. Stat., 12, 171-178*

[3] V. V. Fedorov. (1972). *Theory of optimal experiments*. Aca-

demic Press, New York

[4] W. Hector and S. Hugo. (2010). Information matrix for generalized Skew-normal distribution. *Proyecciones Journal of Mathematics, 29, 2, pp. 83-92*

[5] J. Kiefer and J. Wolfowitz (1959). Optimum designs in regression problems (I). *Annals of Mathematical Statistics, Vol.30, No.2, 271-294*

[6] J. Kiefer and J. Wolfowitz (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics, Vol. 12, 363-366*

[7] M. Pourahmadi (preprint). Construction of Skew-Normal random variables: Are they linear combination of normal half-normal?

[8] C. Stefano, R. Walter and V. Laura (Preprint). Bayesian inference on the scalar Skew- normal distribution

[9] L. V. White (1973). An extension of general equivalence theorem to nonlinear models. *Biometrika, Vol. 60, No.2, 345-348*