Sankhyā : The Indian Journal of Statistics
2002, Volume 64, Series A, Pt. 1, pp 86-94

ILL-CONDITIONED DESIGN AND UNIDENTIFIABILITY

By SOMESH DAS GUPTA
Indian Statistical Institute, Kolkata
and
ABHIJIT DASGUPTA
University of Washington

SUMMARY. Suppose $Y_i \sim N(\alpha + \beta x_i, \sigma^2),\ i = 1, \ldots, n$ are independent random variables. For testing $\alpha = \alpha_0,\ \beta = \beta_0$, it is shown that the power of the likelihood ratio test (LRT) may be smaller than the power of the corresponding test under "x_i's are all equal", when $\sum_i (x_i - \bar{x})^2$ is close to 0. A similar phenomenon occurs for the confidence set for (α, β). It is shown that the transition to $\sum_i (x_i - \bar{x})^2 = 0$ is smooth for Bayesian inference.

1. **Introduction**

Identifiability of parameters and design of experiments are intimately related. As a matter of fact, it can be generally said that the lack of identifiability of parameters or some parametric functions accrue from the experimental design used.

We shall now consider a linear model in which an ill-conditioned design leads to lack of identifiability of some parameters.

Consider the following linear regression model:

$$y_i = \alpha + \beta x_i + e_i \quad i = 1, \ldots, n$$

where the e_i's are iid $N(0, \sigma^2)$. If the x_i's are all equal neither α nor β is identifiable. As a matter of fact, the variance of the least-squares estimate of β tends to ∞ as $\sum_1^n (x_i - \bar{x})^2$ tends to 0, where $\bar{x} = n^{-1} \sum_1^n x_i$. We shall consider the following two situations:

Paper received November 2000; revised July 2001.
AMS (1991) *subject classification.* Primary 62J05; secondary 62K99, 62F15.
Keywords and phrases. Ill-conditioned design, linear model, unidentifiability, power function, confidence interval, Bayes decision, HPD region.

$$\text{Model } I \quad : \quad \sum_{i=1}^{n}(x_i-\bar{x})^2 \neq 0$$
$$\text{Model } II \quad : \quad \sum_{i=1}^{n}(x_i-\bar{x})^2 = 0$$

We shall show that we are "better off" to consider Model II instead of Model I if $\sum_{i=1}^{n}(x_i-\bar{x})^2$ is sufficiently small. The term "better off" is clarified in the following theorem.

THEOREM. *(a) For testing* $H_0: \ \alpha=\alpha_0, \ \beta=\beta_0$, *the likelihood ratio test of size* δ $(0<\delta<1)$ *is more powerful under Model II than that under Model I, when* $\sum_{i=1}^{n}(x_i-\bar{x})^2$ *is sufficiently close to 0, and* $\alpha+\beta\bar{x}\neq\alpha_0+\beta_0\bar{x}$.
(b) The $(1-\delta)$*-level confidence region for* α *and* β *based on the likelihood-ratio test is more accurate (in the sense of Lehmann, i.e., having less probability of covering wrong parameter values) under Model II than that under Model I, if* $\sum_{i=1}^{n}(x_i-\bar{x})^2$ *is sufficiently close to 0, provided the values of* $\alpha+\beta\bar{x}$ *coresponding to wrong and correct values of* (α,β) *differ.*

Some heuristic rules have been suggested to choose between Model I and Model II. It is shown that the power of the likelihood-ratio test changes abruptly as the Model I changes to Model II. In the last section the problem is studied from a Bayesian framework and it is shown that the effect of the transition from Model I to Model II is 'smooth' on the posterior distribution and the corresponding HPD region.

2. Proof of the Theorem

We need the following lemma to prove the above theorem.
LEMMA 2.1 For $\Delta^2>0, \ s>0, \ k>0, \ 0<\delta<1$,

$$P\left[\chi^2_{k+s}(\Delta^2)\geq\chi^2_{k+s,\ \delta}\right]<P\left[\chi^2_k(\Delta^2)\geq\chi^2_{k,\ \delta}\right]$$

where $\chi^2_k(\Delta^2)$ denotes the non-central chi-square variate with noncentrality parameter Δ^2 and d.f. k, and $\chi^2_{k,\ \delta}$ is the upper $100\delta\%$ point of the χ^2_k distribution.

PROOF. Note that the pdf of $\chi^2_k(\Delta^2)$ can be expressed as

$$E\{f_{k+2\theta}(x)\}$$

where f_k is the pdf of χ^2_k, and the expectation is taken with respect to θ, distributed as Poisson $(\Delta^2/2)$.

Now consider independent random variables U and V such that (for fixed θ)

$$U \sim \chi^2_{k+2\theta} \quad V \sim \chi^2_s$$

For testing $\theta = 0$ against $\theta > 0$ at level δ based on U and V, the critical region of the unique UMP test is given by

$$U \geq \chi^2_{k,\delta}$$

The critical region

$$U + V \geq \chi^2_{k+s,\delta}$$

also has size δ. Hence, for $\theta > 0$,

$$P\left[\chi^2_{k+s+2\theta} \geq \chi^2_{k+s,\delta}\right] < P\left[\chi^2_{k+2\theta} \geq \chi^2_{k,\delta}\right]$$

We get the desired result after taking expectation of the above with respect to θ.

Note. The above result has been indicated in Das Gupta and Perlman (1974, Remark 4.1) without any proof.

PROOF OF THE THEOREM

(a) We first assume that σ^2 is known.

For testing $H_0 : \alpha = \alpha_0,\ \beta = \beta_0$ against H_1 : not H_0, the critical region of the likelihood-ratio test at level δ under Model I is given by

$$n\{\bar{y} - (\alpha_0 + \beta_o \bar{x})\}^2 + \sum_{i=1}^{n} (x_i - \bar{x})^2 (b - \beta_0)^2 \geq \chi^2_{2,\delta}\sigma^2$$

where $\bar{y} = n^{-1} \sum_1^n y_i$ and b is the least squares estimate of β. The power of this test at (α, β) is given by

$$\pi_I \equiv P\left[\chi^2_2\left(\frac{n}{\sigma^2}\left((\alpha + \beta\bar{x}) - (\alpha_0 + \beta_0\bar{x})\right)^2 + \frac{(\beta - \beta_0)^2}{\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x})^2\right) \geq \chi^2_{2,\delta}\right] \tag{1}$$

Now note that π_I tends to

$$\pi_I^* \equiv P\left[\chi^2_2\left(\frac{n}{\sigma^2}\left((\alpha + \beta\bar{x}) - (\alpha_0 + \beta_0\bar{x})\right)^2\right) \geq \chi^2_{2,\delta}\right] \tag{2}$$

as $\sum_{i=1}^{n}(x_i - \bar{x})^2 \to 0$.

The critical region of the likelihood-ratio test at level δ for testing H_0 against H_1 and $\alpha + \beta\bar{x} \neq \alpha_0 + \beta_0\bar{x}$, under Model II, is given by

$$n\left\{\bar{y} - (\alpha_0 + \beta_0\bar{x})\right\}^2 \geq \chi^2_{1,\delta}\sigma^2$$

The power of this test at (α, β) is given by

$$\pi_{II} \equiv P\left[\chi^2_1\left(\frac{n}{\sigma^2}\left((\alpha + \beta\bar{x}) - (\alpha_0 + \beta_0\bar{x})\right)^2\right) \geq \chi^2_{1,\delta}\right] \quad (3)$$

From Lemma 2.1, we get

$$\pi^*_I < \pi_{II}$$

Hence, for sufficiently small $\sum_{i=1}^n (x_i - \bar{x})^2$,

$$\pi_I < \pi_{II}$$

Now we consider the case when σ^2 is unknown. Then $\chi^2_2(\Delta^2)$ and $\chi^2_{2,\delta}$ are respectively replaced by $f_{2,n-2}(\Delta^2)$ and $f_{2,n-2;\delta}$ in both π_I and π^*_I, where $f_{s,k}(\Delta^2)$ denotes the non-central $f_{s,k}$ variate with non-centrality parameter Δ^2 (defined as the ratio of independent $\chi^2_s(\Delta^2)$ and χ^2_k variates), and $f_{s,k;\delta}$ is the upper $100\delta\%$ point of the central $f_{s,k}$ distribution. Correspondingly, the π's are replaced by $\tilde{\pi}$'s. The desired result can now be obtained following the method of proof for known σ^2 and invoking the result in Das Gupta and Perlman (1974, Theorem 2.1) instead of Lemma 2.1.

(b) The desired result on the confidence intervals follows from (a) above, by invoking the well-known duality between tests and confidence sets; see Lehmann (1986, p.89-90).

Note. The above results demonstrate some of the consequences of the situation which leads to unidentifiability of the parameters as the design becomes ill-conditioned.

Two issues emerge from the above development. Although it would be better to use Model II instead of Model I if $\sum_{i=1}^n (x_i - \bar{x})^2$ is less than λ^2, where λ^2 is sufficiently small, the value of λ^2 depends on the unknown parameters. Is there any way to resolve this issue?

Secondly, as Model I approaches Model II (i.e., as $\sum_{i=1}^n (x_i - \bar{x})^2 \to 0$), the power of the likelihood-ratio test changes abruptly from π^*_I to π_{II} (or from $\tilde{\pi}^*_I$ to $\tilde{\pi}_{II}$). Is there any way to make this transition smooth?

To address these issues, let us consider the power function of the two tests again, assuming σ^2 is known (for simplicity). Let

$$\Delta^2_1 \quad = \quad \frac{n}{\sigma^2}\left\{(\alpha + \beta\bar{x}) - (\alpha_0 + \beta_0\bar{x})\right\}^2, \quad (4)$$

$$\Delta_2^2 = \frac{(\beta-\beta_0)^2}{\sigma^2}\sum_{i=1}^n (x_i-\bar{x})^2, \tag{5}$$

$$d^2 = \sum_{i=1}^n (x_i-\bar{x})^2/\sigma^2 \tag{6}$$

Write

$$\begin{aligned} \pi_I &= \pi_I(\Delta_1^2+\Delta_2^2) \\ \pi_I^* &= \pi_I^*(\Delta_1^2) \\ \pi_{II} &= \pi_{II}(\Delta_1^2) \end{aligned}$$

It follows from the monotone likelihood ratio property of the non-central chi-square distribution that both $\pi_I(\Delta^2)$ and $\pi_{II}(\Delta^2)$ monotonically increase from δ to 1 as Δ^2 increases from 0 to ∞. Moreover,

$$\pi_I(\Delta^2) < \pi_{II}(\Delta^2), \quad \Delta^2 > 0$$

as shown by Lemma 2.1. Hence, for given $\Delta_1^2 > 0$, there exists a function g such that

$$\pi_I(\Delta_1^2+\Delta_2^2) < \pi_I(\Delta_1^2+g(\Delta_1^2)) = \pi_{II}(\Delta_1^2)$$

for $\Delta_2^2 < g(\Delta_1^2)$. Following Das Gupta and Perlman (1974), it can be shown that g is a strictly increasing function. The function g can be computed from the table of non-central chi-square distribution.

Thus Model II would yield more power than that under Model I, if

$$(\beta-\beta_0)^2\sum_{i=1}^n (x_i-\bar{x})^2/\sigma^2 < g(\Delta_1^2) \tag{7}$$

where Δ_1^2 is given in (4). If Δ_1^2 is large, the power of both the tests would be close to 1; hence, the choice between Model I and Model II effectively arises when Δ_1^2 is of moderate to small value. On the other hand, even if $(\beta-\beta_0)^2$ is large, the power of the test under Model I may fail to be larger than that under Model II if $\sum_{i=1}^n (x_i-\bar{x})^2$ is sufficiently small.

If a preliminary sample is available, then it is possible to get an estimate $\hat{\Delta}_1^2$ of Δ_1^2. Then one may consider a heuristic rule which suggests to use Model II if (7) holds with Δ_1^2 replaced by $\hat{\Delta}_1^2$ and β replaced by b, its least squares estimate.

One may also consider a randomized rule to make the transition from Model I to Model II smooth. Suppose φ is the probability of using Model II

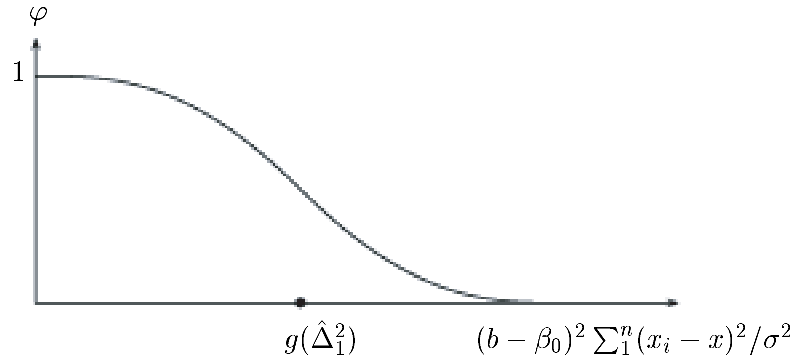

Figure 1: A possible randomized rule for choosing Model II over Model I

and $1-\varphi$ is the probability of using Model I. Then, allowing for the standard error of $\hat{\Delta}_1^2$, one may consider φ described in Figure 1.

A Bayesian analysis of the above problem is given in the next section.

3. Ill-conditioned design and unidentifiability: a Bayesian treatment

We shall now consider the above example from a Bayesian viewpoint. The model considered is the following

$$y_i = \alpha + \beta x_i + e_i, \tag{8}$$

($i = 1, \ldots, n$) where the e_i's are iid $N(0, \sigma^2)$, α and β being unknown. For the present discussion we shall assume that σ^2 is known (for simplicity).

If the x_i's are all equal, both α and β are unidentifiable, the identifying parametric function being $\alpha + \beta\bar{x}$. It can be easily seen that, under Model II, the conditional posterior distribution of β given $\alpha + \beta\bar{x}$ is the same as the corresponding conditional prior distribution. It is expected that, under Model I, the same phenomenon will occur in the limit as $\sum_{i=1}^n (x_i - \bar{x})^2 \to 0$. We shall examine the change in the conditional posterior distribution of β given $\alpha + \beta\bar{x}$, under Model I, as $\sum_1^n (x_i - \bar{x})^2$ approaches 0.

For this study, we consider the prior distribution of (α, β) as

$$N_2\left[\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \Gamma = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix}\right] \tag{9}$$

We consider Model I, namely $\sum_{i=1}^{n}(x_i - \bar{x})^2 \neq 0$. It is sufficient to consider the posterior distribution of (α, β) given $\bar{y}$ and b, the least-squares estimate of β, since $(\bar{y}, b)$ are sufficient for these parameters. For simplicity, we shall consider the posterior distribution of (γ, β), where $\gamma = \alpha + \beta\bar{x}$, noting that (γ, β) has one-to-one correspondence with (α, β).

Given γ and β, the random variables $\bar{y}$ and b are independently distributed as

$$\bar{y} \sim N(\gamma, \sigma^2/n), \quad b \sim N(\beta, \sigma^2/\sum_{i=1}^{n}(x_i - \bar{x})^2) \tag{10}$$

Using generic notation for density functions, the posterior distribution of γ and β can be expressed as

$$p(\bar{y}|\gamma)p(b|\beta)\pi(\gamma)\pi(\beta|\gamma)/p_m(\bar{y}, b) \tag{11}$$

where π stands for prior distribution and p_m denotes the marginal density of $(\bar{y}, b)$. Now

$$\begin{aligned} p_m(\bar{y}, b) &= \int p(\bar{y}|\gamma)\pi(\gamma)\, d\gamma \left[\int p(b|\beta)\pi(\beta|\gamma)\, d\beta\right] \\ &= \int p(\bar{y}|\gamma)\pi(\gamma)\tilde{p}(b|\gamma)\, d\gamma, \end{aligned} \tag{12}$$

where

$$\tilde{p}(b|\gamma) = \int p(b|\beta)\pi(\beta|\gamma)\, d\beta \tag{13}$$

Then the posterior pdf of γ is

$$\tilde{\pi}(\gamma|\bar{y}, b) = p(\bar{y}|\gamma)\pi(\gamma)\tilde{p}(b|\gamma)/p_m(\bar{y}, b), \tag{14}$$

and the conditional posterior pdf of β, given γ, is

$$\tilde{\pi}(\beta|\bar{y}, \beta, \gamma) = p(b|\beta)\pi(\beta|\gamma)/\tilde{p}(b|\gamma) \tag{15}$$

The prior conditional distribution of β given γ is

$$N(a_1\gamma + a_2, \tau^2) \tag{16}$$

where

$$\begin{aligned} a_1 &= \tfrac{\gamma_{12} + \gamma_{22}\bar{x}}{\gamma_{11} + 2\gamma_{12}\bar{x} + \gamma_{22}\bar{x}^2} \\ a_2 &= \mu_\beta - a_1\mu_\gamma \\ \mu_\gamma &= \mu_\alpha + \bar{x}\mu_\beta \\ \tau^2 &= \gamma_{22} - \tfrac{(\gamma_{12} + \gamma_{22}\bar{x})^2}{\gamma_{11} + 2\gamma_{12}\bar{x} + \gamma_{22}\bar{x}^2} \end{aligned} \tag{17}$$

From (15) we find that the conditional posterior distribution of β given γ, is

$$N\left(\frac{1}{d^2\tau^2+1}(a_1\gamma+a_2)+\frac{d^2\tau^2}{d^2\tau^2+1}b, \frac{\tau^2}{d^2\tau^2+1}\right) \tag{18}$$

where

$$d^2=\sum_{i=1}^{n}(x_i-\bar{x})^2/\sigma^2 \tag{19}$$

Note now that the distribution in (18) tends to the distribution in (16) as $d^2 \to 0$. The conditional posterior mean of β, given γ, is a convex combination of the conditional prior mean of β, given γ, and b; the role of b can be judged by its factor $d^2\tau^2/(1+d^2\tau^2)$, which tends to 0 as $d^2 \to 0$. The conditional posterior variance of β, given γ, is smaller than the conditional prior variance of β, given γ; however, as $d^2 \to 0$, these conditional variances become equal in the limit. As a whole, it appears that (18) approaches (16) "smoothly" as $\sum_1^n(x_i-\bar{x})^2$ approaches 0.

Using (14) we find that the posterior distribution of γ under Model I is normal with mean

$$\left[\frac{n}{\sigma^2}+\frac{1}{\sigma_\gamma^2}+\frac{d^2a_1^2}{d^2\tau^2+1}\right]^{-1}\left[\frac{n}{\sigma^2}\bar{y}+\frac{\mu_\gamma}{\sigma_\gamma^2}+\frac{d^2a_1(b-a_2)}{d^2\tau^2+1}\right] \tag{20}$$

and variance

$$\left[\frac{n}{\sigma^2}+\frac{1}{\sigma_\gamma^2}+\frac{d^2a_1^2}{d^2\tau^2+1}\right]^{-1} \tag{21}$$

where

$$\sigma_\gamma^2=\gamma_{11}+2\bar{x}\gamma_{12}+\bar{x}^2\gamma_{22} \tag{22}$$

It can be seen that the above distribution tends to the posterior distribution of γ under Model II as $\sum_1^n(x_i-\bar{x})^2 \to 0$. Furthermore, it may be noted that the posterior variance of γ is smaller under Model I than that under Model II. The influence of b in the posterior distribution of γ under Model I diminishes to 0 as $\sum_1^n(x_i-\bar{x})^2$ approaches 0.

We shall now examine the change of the HPD region of (γ,β) under Model I to that under Model II as $\sum_{i=1}^n(x_i-\bar{x})^2 \to 0$.

After some calculations, it can be found that the posterior distribution of (γ,β) under Model I is normal with mean

$$\begin{pmatrix}\mu_\gamma\\ \mu_\beta\end{pmatrix}+\Delta(D+\Delta)^{-1}\left[\begin{pmatrix}\bar{y}\\ b\end{pmatrix}-\begin{pmatrix}\mu_\gamma\\ \mu_\beta\end{pmatrix}\right] \tag{23}$$

and covariance matrix

$$\Delta - \Delta(D+\Delta)^{-1}\Delta \tag{24}$$

where

$$\begin{aligned} \Delta &= \begin{pmatrix} \gamma_{11} + 2\bar{x}\gamma_{12} + \bar{x}^2\gamma_{22} & \gamma_{12} + \bar{x}\gamma_{22} \\ \gamma_{12} + \bar{x}\gamma_{22} & \gamma_{22} \end{pmatrix} \\ D &= \text{diag } (\sigma^2/n, 1/d^2) \end{aligned}$$

The posterior distribution of (γ, β) under Model II is normal with mean

$$\begin{pmatrix} \mu_\gamma \\ \mu_\beta \end{pmatrix} + \begin{pmatrix} \sigma_\gamma^2 \\ \gamma_{12} + \bar{x}\gamma_{22} \end{pmatrix} \frac{\bar{y} - \mu_\gamma}{\sigma^2/n + \sigma_\gamma^2}, \tag{25}$$

and covariance matrix

$$\Delta - \frac{1}{\sigma^2/n + \sigma_\gamma^2} \begin{pmatrix} \sigma_\gamma^2 \\ \gamma_{12} + \bar{x}\gamma_{22} \end{pmatrix} (\sigma_\gamma^2 \quad \gamma_{12} + \bar{x}\gamma_{22}) \tag{26}$$

Now note that

$$(D+\Delta)^{-1} = D^{-1/2}(I + D^{-1/2}\Delta D^{-1/2})^{-1} D^{-1/2} \tag{27}$$

and

$$D^{-1/2} \to \begin{pmatrix} \sqrt{n}/\sigma & 0 \\ 0 & 0 \end{pmatrix} \tag{28}$$

as $\sum_{i=1}^n (x_i - \bar{x})^2 \to 0$. Now it is easy to see that the HPD region under Model I "smoothly" changes to that under Model II as $\sum_{i=1}^n (x_i - \bar{x})^2 \to 0$.

References

Das Gupta S. and Perlman, M.D..(1974). Power of the noncentral F-test: Effect of additional variates on Hotelling's T^2-test. *J. Amer. Statist. Assoc.*, **69**, 174-180.

Lehmann, E.L. (1986) *Testing Statistical Hypothesis.* Wiley, New York, 2nd edition.

Somesh Das Gupta
BC 249, Bidhan Nagar
Kolkata, India 700 064
E-mail: dasguptasomesh@yahoo.com

Abhijit Dasgupta
Div. of Cancer Epi. and Genetics
National Cancer Institute
Bethesda, MD 20892-7244
U.S.A.
E-mail: dasgupab@mail.nih.gov