

IN THIS ISSUE

SPECIAL INVITED PAPER

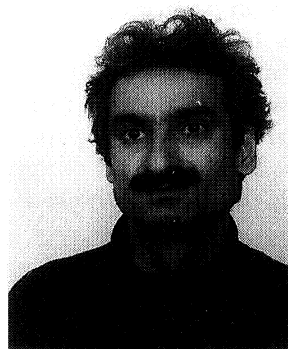
SEVEN INEQUALITIES IN STATISTICAL ESTIMATION THEORY

What motivated the pioneers in our field to do what they did and what kind of mathematical training is needed to do research in statistics? In his letter to *Student*, C.R. Rao writes: "I was mainly motivated by what R.A. Fisher described as the two fundamental problems of statistics, specification and estimation and provided a general solution, especially to the problem of estimation. I tried to resolve some issues raised by Fisher as explained in my paper on seven inequalities. Fortunately, the solutions did not involve complicated mathematics. I hope the paper would be of some interest to students and teachers of statistics". We are pleased to present in this issue a paper written by C.R. Rao for *Student* on Seven Inequalities in Statistical Estimation Theory. An homage is also paid to him in *Capturing History* for his 50 years contribution to the science of statistics.

AN APPLICATION OF THE LIKELIHOOD RATIO TEST TO THE CHANGE POINT PROBLEM

The change point models are becoming extremely important in environmental time series and spatial data analysis. Problems of statistical inference associated with detection of changes have many approaches. Dariush Ghorbanzadeh uses a class of tests based on the likelihood ratio test to detect a change in the parameter of a sequence of independent random variables.

Dariush Ghorbanzadeh was born on July 23, 1959 in Andimeshk, Iran. He obtained Bachelor of Science in pure mathematics from University of Paris VII in 1987 and his Master in applied mathematics in 1988. In 1992, he was awarded Doctoral under the direction of Professor Dominique Piccard. Currently, Dr. Ghorbanzadeh is with the Conservatoire National des Arts et Métiers, in Paris as a research collaborator.



VARIANCE ESTIMATION OF SWISS CONSUMER PRICE INDEX

The Swiss Consumer Price Index is constructed on data issued from several samples. Each of them has a complex survey sample design. In her article, Sandrine König develops a methodology to estimate the variance of the Swiss CPI.



Sandrine König was born on April 20, 1965 in Boudevilliers. She obtained her Bachelor of Science in Political Economics in 1989 and her postgraduate degree in Statistics in 1991 from the University of Neuchâtel. In February 1995, she received her doctoral degree in Economics in the same University. She was for several years a research assistant with the Statistics Group at the University of Neuchâtel. Dr König is working as an economist at the Tax Office of the State of Neuchâtel.

ON SHRINKAGE RIDGE REGRESSION ESTIMATORS FOR LINEAR MODELS WITH MULTIVARIATE t DISTURBANCES

Since its discovery almost 200 years ago, least squares has been the most popular method of regression analysis. Over the last three decades, there has been increasing interest in other methods due to discoveries of deficiencies in the least squares method. Golam Kibria discusses a combination of shrinkage-ridge method to overcome the problem of heavy tailed distribution of errors and restriction on the regression parameters.

Golam Kibria was born on February 1, 1963 in Bangladeshi. He graduated in statistics from the University of Jahangirnagar, Bangladesh in 1986. He obtained a Master of mathematics from Carleton University, Ottawa, Canada in 1993. His research interests are predictive inference, linear model and regression analysis. He is currently a Ph.D. student at the Department of Statistical and Actuarial Sciences of the University of Western Ontario, Canada.



FROM THE EDITOR'S NOTEBOOK

Why calculate billions of decimals of π ? From the editor's notebook we find a remarkable usefulness: to be used as the natural random number generator.

DATA & STATISTICS

Three data sets with some guidelines as to how they might be analysed from their authors' point of view are presented in this issue. Readers are very much welcome to propose alternative methods to *Student*. The floor is open for discussion.

Editorial

Student, Data & Statistics

The best form of applied statistical research is driven by the underlying scientific context. The raw material for this work are careful descriptions of real data and their associated scientific problems. Data sets are a powerful source of inspiration for both experienced professionals and students. Data, and their associated descriptions, foster an appreciation for the field of study and an understanding of the relationship between statistics and scientific method. The analysis of real data sets often provides the background and impetus for the development of new statistical methodology.

For this reason *Student* has been extended to incorporate *Data & Statistics*. As such the Journal provides a forum for research workers, as well students, to present their data and their initial data explorations. Articles give emphasis to careful descriptions of the underlying problems, the set of research questions which need to be addressed, the practical concerns about obtaining more data, and a selective set of references to the appropriate application-specific literature.

The Journal serves an educational role by encouraging articles on historical statistics, in particular, those emphasising the influence of data in the development of the subject area. Moreover, statistical methods are often presented using carefully chosen hypothetical data sets, which mask many technical problems which may arise in practice, such as, incomplete data, censoring, and non-representative samples. Exposing the reality is a focus of interest for the Journal. The Journal is of special interest to the consulting or applied statistician, e.g. industrial, who face challenging data sets.

Neuchâtel, March 1996

The Editor

Seven Inequalities in Statistical Estimation Theory

C. Radhakrishna Rao

Department of Statistics, The Pennsylvania State University, United States

Abstract: When I was learning statistics as a student, the only serious literature available on statistical theory and practice was Fisher's book *Statistical Methods for Research Workers* and his numerous papers on estimation and small sample exact distributions of some test statistics. Fisher was laying the logical foundations of statistics as a separate discipline. He introduced numerous ideas and analytical tools which we still use. Fisher was criticized for lack of rigor in mathematical rendering of his ideas and in proving his results, and for making general statements without specifying the underlying assumptions (Savage, 1976). My understanding of statistics is based on reading each of Fisher's results n times depending on the complexity of the problem involved. I am still reading some of his papers. In this process, I developed some inequalities to interpret and clarify Fisher's ideas. I shall describe them to you with some historical notes.

Key words: Cramér-Rao inequality, Fisher consistency, Fisher-Rao theorem, Fisher's theorem, Inequality on expected harmonic mean, matrix inequality, Rao-Blackwell theorem, second order efficiency.

Notations

1. $f(x_n, \theta), x_n \in \mathcal{X}, \theta \in \Theta$, is a family of probability density functions indexed by a k -vector parameter $\theta = (\theta_1, \dots, \theta_k)'$, where x_n represents the whole sample of n observations.
2. $T_n : \mathcal{X} \rightarrow R^s$ is an s -vector statistic (function of x_n) estimating an s -vector parametric function $m(\theta)$.

3. $g_r(x_n, \theta) = \partial \log f / \partial \theta_r$, $E(g_r g_s) = i_{rs}$, $I = (i_{rs})$ is the $k \times k$ Fisher information matrix.
4. $V(T_n)$ is the variance-covariance matrix of T_n . $V_a(T_n)$ is the variance-covariance matrix of the asymptotic distribution of $\sqrt{n}(T_n - b(\theta))$, for some choice of $b(\theta)$, as $n \rightarrow \infty$.
5. When $k = 1$, we represent the single parameter by θ , and use the same symbols $V(T_n)$ and $V_a(T_n)$ to represent the variance and asymptotic variance of T_n , an estimator of θ .

1 Cramér-Rao inequality (single parameter)

Fisher (1922) proved in the one parameter case and iid observations

$$T_n \rightarrow \theta \text{ in prob. as } n \rightarrow \infty \Rightarrow V_a(T_n) \geq 1/i,$$

where i is information per single observation.

While giving a course on estimation theory in December 1943, one student in the class asked: Does there exist an analogue of Fisher's theorem for $V(T_n)$, where T_n is an unbiased estimate of θ ? I was prepared for an answer in the next class with the following simple proof.

$$\theta = E(T_n) = \int T_n f dv \quad (1)$$

where dv is the volume element in $\mathcal{X}$. Differentiating both sides of (1) w.r.t θ under the integral sign

$$1 = \int T_n \frac{f'}{f} f dv = \text{Cov}(T_n, \frac{f'}{f}) = \text{Cov}(T_n - \theta, \frac{f'}{f}).$$

Then by Cauchy-Schwarz (CS) inequality (see Rao, 1973, p.54),

$$[\text{Cov}(T_n, \frac{f'}{f})]^2 \leq V(T_n) V(f'/f) \quad (2)$$

$$1 \leq V(T_n) I, \text{ or } V(T_n) \geq 1/I,$$

where I is the information on θ in the whole sample x_n (need not be iid observations). The inequality (2) for exact variance is called Cramér-Rao inequality¹ by Neyman and Scott (1948).

¹ Cramér-Rao inequality is included as a "scientific and technical term" in McGraw-Hill Dictionary of Scientific and Technical Terms, (Fifth edition, p.476).

Cramér-Rao inequality [Stat] An inequality that is the basis of a method for determining a lower bound to the variance of an estimate of a parameter.

It is seen by the same argument, that if $E(T_n) = m(\theta)$, a function of θ , then

$$V(T_n) \geq [dm(\theta)/d\theta]^2/I.$$

2 Cramér-Rao inequality (multiparameter case)

If $T_n = (T_{n1}, \dots, T_{nk})'$ is unbiased for $\theta = (\theta_1, \dots, \theta_k)'$, then for any $a = (a_1, \dots, a_k)'$

$$\begin{aligned} E(\sum a_r T_{nr}) &= \int (\sum a_r T_{nr}) f dv = \sum a_r \theta_r \\ \int (\sum a_r T_{nr}) \frac{\partial f}{\partial \theta_s} dv &= a_s = \int (\sum a_r T_{nr}) g_s f dv. \end{aligned}$$

Now for any $b = (b_1, \dots, b_k)'$

$$\int (\sum a_r T_{nr}) (\sum b_s g_s) f dv = \sum a_s b_s.$$

Using the CS inequality,

$$\begin{aligned} V(\sum a_r T_{nr}) V(\sum b_s g_s) &\geq (\sum a_s b_s)^2 \\ (a' V(T_n) a) (b' I b) &\geq (a' b)^2 \quad \forall a, b \\ a' V(T_n) a &\geq \frac{(a' b)^2}{b' I b} \quad \forall a, b. \end{aligned}$$

Choosing $b = I^{-1}a$,

$$a' V(T_n) a \geq a' I^{-1} a \quad \forall a$$

$$\Rightarrow V(T_n) \geq I^{-1} \text{ or } V(T_n) - I^{-1} \text{ is non-negative definite.}$$

This inequality is the matrix version of Cramér-Rao inequality $V - I^{-1} \geq 0$ in the one parameter case. Similarly, it is easy to prove (Rao, 1948):

If T_n is unbiased for $(b_1(\theta), \dots, b_s(\theta))$ and $H = (\partial b_r / \partial \theta_j)$, then

$$V(T_n) - H I^{-1} H' \text{ is non-negative definite.}$$

3 Fisher's inequality

Fisher's inequality, $V_a(T_n) \geq 1/i$ for asymptotic variance of any consistent estimator T_n , referred to in Section 1, is not strictly true as shown by Hodges (see LeCam, 1953) by giving a counter example. Is it true under further assumptions on T_n ?

When I drew Fisher's attention to Hodge's example in 1954, he said that his definition of consistency was in fact a small sample concept and was different from the probability limit. He referred to sampling from a multinomial distribution with $\pi_i(\theta)$, $i = 1, \dots, q$ as the hypothetical probabilities depending on a parameter θ , and p_i , $i = 1, \dots, q$ as the observed proportions in a sample of size n in q cells. He was considering a statistic T_n as an explicit function $T_n = T(p_1, \dots, p_q)$ of the observed proportions only, and called it consistent if $T(\pi_1(\theta), \dots, \pi_q(\theta)) \equiv \theta$. I called this property Fisher consistency in a paper (Rao, 1957, p.26) presented in 1955 at the Rio-de-Janeiro conference of the International Statistical Institute. It is interesting to note that in all the editions of *Statistical Methods for Research Workers* (SMRW) up to the eleventh (1950, p.11) and in his papers on estimation, Fisher did not explicitly state consistency in the above form, but perhaps thought that his definition and the probability definition were the same. For the first time in the twelfth edition of SMRW (1954, p.12), he refers to the definition of consistency in small samples as defined above.

Once we adopt Fisher consistency, Fisher's theorem in the multinomial case follows easily as the little lamb follows Mary (Rao, 1957).

As shown in Rao (1973, p.387), if $T(p_1, \dots, p_q)$ is differentiable, then

$$\sqrt{n}(T - \theta) \sim N(0, V_a) \text{ asymptotically}$$

where

$$V_a = \sum \pi_i \left(\frac{\partial T}{\partial \pi_i} - m \right)^2, \quad m = \sum \pi_i \frac{\partial T}{\partial \pi_i}.$$

Now, we use Fisher consistency

$$T(\pi_1(\theta), \dots, \pi_q(\theta)) \equiv \theta \quad (3)$$

and differentiate both sides of (3) with respect to θ to obtain

$$\left(\frac{\partial T}{\partial \pi_1} - m \right) \frac{d\pi_1}{d\theta} + \dots + \left(\frac{\partial T}{\partial \pi_q} - m \right) \frac{d\pi_q}{d\theta} = 1. \quad (4)$$

The introduction of m in (4) makes no difference. Applying the CS inequality to (4)

$$\left[\sum \pi_i \left(\frac{\partial T}{\partial \pi_i} - m \right)^2 \right] \left[\sum \pi_i \left(\frac{1}{\pi_i} \frac{d\pi_i}{d\theta} \right)^2 \right] \geq 1 \quad (5)$$

$$V_a i \geq 1 \text{ or } V_a \geq 1/i$$

where the second term on the left hand of (5) is Fisher information in a single observation denoted by i . We thus have a rigorous demonstration

of Fisher's theorem. A multivariate version of Fisher's theorem is easily obtained.

The same result was proved in the continuous case in a subsequent paper by Kallianpur and Rao (1955) by considering $T_n = T(F_n)$ with the property $T(F_\theta) \equiv \theta$ where F_n and F_θ are the empirical and true distribution functions respectively. See Rao (1962a, 1962b, 1963, 1965) for further details.

4 Rao-Blackwell theorem

By a sufficient statistic T , Fisher meant that we can use T instead of the whole sample in drawing inferences. I was looking for a simple demonstration of this fact and succeeded in getting a good example. I proved in Rao (1945) the following result.

If $t = (t_1, \dots, t_k)'$ is an unbiased estimate of the vector parameter $\theta = (\theta_1, \dots, \theta_k)'$ and if the distribution admits a sufficient statistic $T = (T_1, \dots, T_s)'$, then there exists $f(T) = (f_1(T), \dots, f_k(T))'$ such that $f(T)$ is unbiased for θ and

$$\text{Cov}(t) - \text{Cov}(f(T)) \text{ is non-negative definite.}$$

The result simply follows by taking

$$f(T) = E(t|T) \tag{6}$$

and observing that

$$\begin{aligned} \text{Cov}(t) &= E(t - \theta)(t - \theta)' \\ &= E[t - f(T) + f(T) - \theta][t - f(T) + f(T) - \theta]' \\ &= E[t - f(T)][t - f(T)]' + E[f(T) - \theta][f(T) - \theta]' \\ &\geq \text{Cov}(f(T)) \end{aligned} \tag{7}$$

since the cross product term vanishes and the first term on the lefthand side of (7) is non-negative definite.

There are two interesting side lights to Rao-Blackwellization, a term coined by Berkson (1955, p.142) to denote the above process.

Berkson was giving a talk at the Biometric conference in Bellagio in 1953, where he used the conditional expectation method (6) to improve the efficiency of an unbiased estimator, referring to the method as Blackwellization based on a paper of Blackwell published two years later than mine. When I pointed out this to Berkson, he said Raoization does not sound good, and he would call the method Rao-Blackwellization when his paper is submitted for publication.

A few years later, L.K. Schmetterer wrote a book on mathematical statistics where he attributed my result on sufficient statistics to Blackwell only. D.V. Lindley, in a review of the book, referred to the result as Blackwell theorem. When I wrote to Lindley that I proved the theorem two years earlier than Blackwell, he replied on the following lines:

“Of course, the result is there in your 1945 paper², but you were not aware of the importance of the result as you have not mentioned it in the introduction to the paper”.

I replied to Lindley saying that it was my first research paper and I was not aware that the introduction is generally written for the benefit of those who do not want to read the paper.

5 A matrix inequality

In developing a unified theory of linear estimation in a mixed linear model, I was led to solve the following problem. What is the minimum of the matrix valued function

$$f(X) = X'VX - F'X - X'F \text{ subject to } G'X = P$$

where X is an $n \times m$ variable matrix, V is $n \times n$ non-negative definite matrix, and F, G and P are given matrices of orders $n \times m$, $n \times k$ and $k \times m$ respectively. It is shown (Rao, 1989) that

$$f(X) - f(X_0) \text{ is a non-negative definite matrix}$$

where (X_0, Y_0) is a solution of the equations

$$\begin{aligned} VX_0 + GY_0 &= F, \\ G'X_0 &= P. \end{aligned}$$

The result is proved by writing $X = X_0 + C$ and showing that $f(X) = f(X_0) + B$ where B is non-negative definite matrix. To find X_0, Y_0 , we need to find a symmetric generalized inverse

$$\begin{pmatrix} V & G \\ G' & 0 \end{pmatrix}^- = \begin{pmatrix} C_1 & C_2 \\ C_2' & -C_4 \end{pmatrix}$$

²Rao's 1945 paper where Cramér-Rao inequalities, Rao-Blackwell theorem and differential geometric approach to statistical inference are first introduced, is included in *Breakthroughs in Statistics: 1890-1990* edited by S. Kotz and N.L. Johnson, Wiley. These results are also mentioned in *Modern Concepts and Theorems of Mathematical Statistics* by E.B. Moukian (Springer-Verlag, 1986).

and take the solution

$$X_0 = C_1 F + C_2 P, Y_0 = C_2' F - C_4 P.$$

For further details and applications reference may be made to Rao (1989).

6 Fisher-Rao theorem: Second order efficiency

During the fifties, inspired by the work on BAN (best asymptotically normal) estimators by J. Neyman, a number of authors made attempts to show that there exist many methods of estimation which are computationally simple and which have the same asymptotic variance as the maximum likelihood (ml) estimate. I was looking for some higher order asymptotic property which distinguishes the ml from the others. I came across Fisher's (1925) seminal work, where he introduced the expression

$$E_2 = \lim_{n \rightarrow \infty} [I_n(\theta) - I(T_n, \theta)] \quad (8)$$

as the asymptotic loss of information in a statistic T_n , where $I_n(\theta)$ is the total information in the sample and $I(T_n, \theta)$ is that contained in the statistic T_n . Fisher computed the limit as

$$E_2 = i(\theta) \left\{ \frac{\mu_{02} - 2\mu_{21} + \mu_{40}}{[i(\theta)]^2} - 1 - \frac{\mu_{11}^2 \mu_{30}^2 - 2\mu_{11} \mu_{30}}{[i(\theta)]^3} \right\} \quad (9)$$

where $i(\theta)$ is Fisher information per unit observation and

$$\mu_{hj} = E \left(\frac{1}{f} \frac{df}{d\theta} \right)^h \left(\frac{1}{f} \frac{d^2 f}{d\theta^2} \right)^j$$

with f as the density function. The result is not, however, strictly true. I tried two possible approaches to evaluate higher order efficiency, one of which is given in Rao (1961) and another in Rao (1963). I shall describe the latter result which was slightly refined by Ghosh and Subramaniam (1974).

Consider a possible expansion for a truncated (at d) mean square loss in an estimator t_n corrected for bias up to terms of $O(1/n)$ in the form

$$E_\theta[\min\{(t_n - \theta)^2, d\}] = \frac{1}{ni} + \frac{\Psi}{n^2} + o\left(\frac{1}{n^2}\right). \quad (10)$$

Then we may define Ψ as a measure of second order efficiency (SOE) using the term first order efficiency for estimators with $1/ni$ as the first term in the expansion (10). I have shown that in the case of sampling from multinomial distributions $\Psi \geq E_2$, the same expression (9) Fisher obtained for the limiting loss in information, and that equality is attained for the ml

estimator. Ghosh and Subramaniam (1974) extended the result to exponential family of distributions. Efron (1975) called the inequality on Ψ as Fisher-Rao theorem and identified the “curious looking expression” E_2 as the curvature of the distribution.

7 An inequality on the expectation of harmonic mean

In trying to understand a result of Fisher on ancillary statistics (see Rao, 1992, p.43), I had to use an inequality of the following type.

If $X > 0$ and $Y > 0$ are random variables, then

$$E\left(\frac{XY}{X+Y}\right) \leq \frac{E(X)E(Y)}{E(X+Y)}. \quad (11)$$

This is easily proved as follows.

(a) If $E(X)/E(Y) = p/q$ is rational, consider $X_1 = X/p$, $Y_1 = Y/q$ so that $E(X_1) = E(Y_1) = \mu > 0$. Since the harmonic mean is not greater than the geometric mean, we have

$$\frac{p+q}{\frac{p}{Y_1} + \frac{q}{X_1}} \leq Y_1^{p/(p+q)} X_1^{q/(p+q)}.$$

Then

$$E\left(\frac{XY}{X+Y}\right) = \frac{pq}{p+q} E\left\{\frac{p+q}{\frac{p}{Y_1} + \frac{q}{X_1}}\right\} \leq \frac{pq}{p+q} E(Y_1^{p/(p+q)} X_1^{q/(p+q)}). \quad (12)$$

Using Hölder's inequality (Rao, 1973, p.55), the last expression in (12) is

$$\begin{aligned} &\leq \frac{pq}{p+q} (EY_1)^{p/(p+q)} (EX_1)^{q/(p+q)} \\ &= \frac{pq}{p+q} \mu = \frac{E(X)E(Y)}{E(X+Y)}. \end{aligned} \quad (13)$$

Thus, (12) and (13) establish (11) in the special case considered in (a).

(b) Now assume that $E(X)/E(Y) = \lambda > 0$ and consider a sequence $p/q \rightarrow \lambda$. Since

$$\frac{\frac{p}{q\lambda}XY}{\frac{p}{q\lambda}X+Y} \leq Y \text{ and } \frac{\frac{p}{q\lambda}XY}{\frac{p}{q\lambda}X+Y} \rightarrow \frac{XY}{X+Y},$$

we have by the dominated convergence theorem

$$E\left\{\frac{\frac{p}{q\lambda}XY}{\frac{p}{q\lambda}X+Y}\right\} \rightarrow E\left\{\frac{XY}{X+Y}\right\}.$$

Since $E(pX/q\lambda) = pE(Y)/q$, we have by (a), considering $pX/q\lambda$ as X ,

$$E\left\{\frac{\frac{p}{q\lambda}XY}{\frac{p}{q\lambda}X + Y}\right\} \leq \frac{\frac{p}{q\lambda}(EX)(EY)}{\frac{p}{q\lambda}(EX) + E(Y)}.$$

Taking the limit, we get (11).

Résumé: Lorsque j'étudiais la statistique, la seule référence sérieuse valable en matière de statistique théorique et pratique était le livre de Fisher *Statistical Methods for Research Workers* ainsi que ses nombreux articles sur l'estimation à partir de petits échantillons de distributions exactes de tests statistiques. Fisher considérait les bases logiques des statistiques comme une discipline à part. Il introduisit bon nombre d'idées et d'outils analytiques toujours utilisés actuellement. Fisher était critiqué pour son manque de rigueur dans l'interprétation mathématique de ses idées et dans la preuve de ses résultats, ainsi que pour établir des modèles généraux sans spécifier les hypothèses sous jacentes (Savage, 1976). Ma compréhension de la statistique est basée sur la lecture de chacun des résultats de Fisher n fois, en fonction de la complexité du problème évoqué. Je lis toujours certains de ses articles. Dans cette démarche, j'ai développé des inégalités pour interpréter et clarifier les idées de Fisher. Je les ai décrites ici accompagnées de notes historiques.

Mots clés: Consistence de Fisher, efficacité du second ordre, inégalité de Cramér-Rao, inégalité de l'espérance de la moyenne harmonique, inégalité de marice, théorème de Fisher, théorème de Fisher-Rao, théorème de Rao-Blackwell.

References

- [1] Berkson, J. (1955), "Maximum likelihood and minimum chi-square estimates of the logistic function", *J. Am. Statist. Assoc.*, **50**, 130-162.
- [2] Efron, B. (1975), "Defining the curvature of a statistical problem, with applications to second order efficiency (with discussion)", *Ann. Statist.*, **3**, 1189-1242.
- [3] Fisher, R.A. (1922), "On the mathematical foundations of theoretical statistics", *Philos. Trans. Roy. Soc. London Series A*, **222**, 309-368.
- [4] Fisher, R.A. (1925), "Theory of statistical estimation", *Proc. Cambridge Philos. Soc.*, **22**, 700-725.
- [5] Ghosh, J.K. and Subramaniam, K. (1974), "Second order efficiency of maximum likelihood estimators", *Sankhyā A*, **36**, 325-358.
- [6] Kallianpur, G. and Rao, C.R. (1955), "On Fisher's lower bound to asymptotic variance of a consistent estimate", *Sankhyā*, **16**, 331-342.

- [7] LeCam, L. (1953), "On some asymptotic properties of maximum likelihood estimates and related Bayes estimates", *University of California Publications in Statistics*, **1**, 299-329.
- [8] Neyman, J. and Scott, E.L. (1948), "Consistent estimates based on partially consistent observations", *Econometrika*, **16**, 1-32.
- [9] Rao, C.R. (1945), "Information and accuracy attainable in the estimation of statistical parameters", *Calcutta Mathematical Bulletin*, **37**, 81-91.
- [10] Rao, C.R. (1946), "Minimum variance and the estimation of several parameters", *Proc. Cambridge Philosophical Society*, **43**, 280-283.
- [11] Rao, C.R. (1948), "Sufficient statistics and minimum variance estimation", *Proc. Cambridge Philosophical Society*, **45**, 215-218.
- [12] Rao, C.R. (1957), "Theory of the method of estimation by minimum chi-square", *Bulletin de l'Institut International de Statistique*, **35**, 25-32.
- [13] Rao, C.R. (1961), "Asymptotic efficiency and limiting information", *Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 531-546, University of California.
- [14] Rao, C.R. (1962a), "Apparent anomalies and irregularities in maximum likelihood estimation (with discussion)", *Sankhyā B*, **24**, 73-102.
- [15] Rao, C.R. (1962b), "Efficient estimates and optimum inference procedures (with discussion)", *J. Roy. Statist. Soc., Series B*, **24**, 46-72.
- [16] Rao, C.R. (1963), "Criteria of estimation in large samples", *Sankhyā A*, **25**, 189-206.
- [17] Rao, C.R. (1965), "Efficiency of an estimator and Fisher's lower bound to asymptotic variance", *Bull. de l'Institut International de Statistique*, **XLI**, 55-63.
- [18] Rao, C.R. (1973), *Linear Statistical Inference and its Applications* (2nd edition), Wiley, New York.
- [19] Rao, C.R. (1989), "A lemma on optimization of a matrix function and a review of unified theory of linear estimation", In *Statistical Data Analysis and Inference* (ed. Y. Dodge), pp. 397-418, North Holland.
- [20] Rao, C.R. (1992), "R.A. Fisher: The founder of modern statistics", *Statistical Science*, **7**, 34-48.
- [21] Savage, L.J. (1976), "On rereading of R.A. Fisher", *Ann. Statist.*, **4**, 441-500.

An Application of the Likelihood Ratio Test to the Change Point Problem

Dariusz Ghorbanzadeh

Department of Mathematics, CNAM, Paris, France

Abstract: We consider a class of tests to detect a change in the parameter of a sequence of independent random variables. The class of tests considered is based on the likelihood ratio appropriately adapted for this change point problem. The limiting distribution of the test statistics, under the null hypothesis and the alternative hypothesis is evaluated by asymptotical techniques according to the LeCam concept of contiguity. Analytical expressions for the asymptotic power function test is provided. The performance of the test, in two particular cases, normal distribution and exponential distribution, is studied by simulations.

Key words: Change point, contiguity, Girsanov theorem, reflection principle.

1 Introduction

Let $X_1, X_2, \dots, X_n$ be a sequence of independent random variables (r.v.'s), each with distribution function $F(., \theta_i)$ having probability density function $f(., \theta_i)$, $\theta_i \in \Theta$, $i = 1, 2, \dots, n$ (the parameter set Θ is an open interval of R). We shall study the problem of the following test: null hypothesis, the r.v.'s $X_1, X_2, \dots, X_n$ are identically distributed with $\theta_i = \theta_0$ for $i = 1, 2, \dots, n$, versus alternative hypothesis, the r.v.'s $X_1, X_2, \dots, X_n$ has a change-point at unknown time k ($1 \leq k \leq n - 1$) (shift of the parameter from $\theta_1^{(n)} = \theta_0 + u_1/\sqrt{n}$ to $\theta_2^{(n)} = \theta_0 + u_2/\sqrt{n}$, $u_1 > u_2$). In the literature, the above problem has often been studied for the normal distribution, many test statistics have been proposed according to the initial mean $E[X_1] = \theta_0$ (usually θ_0 is unknown). The power of the test has been studied analytically or by simulations, for example, the power function of

the tests proposed by Chernoff and Zacks (1964) and of Gardner (1969) can be studied analytically. Besides the power function of maximum likelihood type statistics can be evaluated by numerical methods (see Sen and Srivastava, 1975; Hawkins, 1977; Cobb, 1978; Worsley, 1983 and 1986). Many authors such as Deshayes and Picard (1986) or Kim and Siegmund (1989) have used asymptotical methods to determine the distributions of the test statistics, which correspond to certain processes associated with Brownian motion process (for example, Brownian bridge). In this paper, we detail the problem of test in case of a parametric model. A class of tests based on the likelihood ratio test is proposed. The concept of contiguity is used to evaluate the asymptotical distribution of test statistics under the two hypotheses.

2 The statistical model

Let $X_1, X_2, \dots, X_n$ be a sequence of independent r.v.'s, each with distribution function $F(\cdot, \theta_i)$ having probability density function $f(\cdot, \theta_i)$, $i = 1, 2, \dots, n$. We intend to test the following hypotheses:

$$H_0(\theta_0) : \theta_i = \theta_0 \quad \text{for } i = 1, 2, \dots, n \quad (1)$$

against

$$H_1^{(n)} = \bigcup_{k=1}^{n-1} H_1^{(n)}(u_1, k, u_2)$$

with

$$H_1^{(n)}(u_1, k, u_2) : \begin{cases} \theta_i = \theta_1^{(n)} = \theta_0 + \frac{u_1}{\sqrt{n}} & \text{if } i = 1, 2, \dots, k \\ \theta_i = \theta_2^{(n)} = \theta_0 + \frac{u_2}{\sqrt{n}} & \text{if } i = k+1, \dots, n \end{cases}$$

It is assumed that θ_0 and u_1, u_2 are known.

2.1 The assumptions of regularity

We assume in the following:

- (i) the function $\theta \mapsto f(\cdot, \theta)$ is of $\mathcal{C}^3$ class in a neighbourhood of θ_0 .
- (ii) for all $\theta \in \Theta$ the Fisher information integral is convergent, i.e.

$$I(\theta) = - \int \frac{\partial^2}{\partial \theta^2} \text{Log}(f(x, \theta)) f(x, \theta) dx < \infty.$$

- (iii) the function $I(\theta)$ is continuous in a neighbourhood of θ_0 .
- (iv) the function $\theta \mapsto \frac{\partial^3}{\partial \theta^3} \text{Log}(f(\cdot, \theta))$ is bounded in a neighbourhood of θ_0 .

2.2 The likelihood ratio test

For testing $H_0(\theta_0)$ against $H_1^{(n)}$, we consider the Log-likelihood ratio statistic defined by $\Lambda_n^* = \max_{1 \leq k \leq n-1} \Lambda_n(k/n)$ with

$$\Lambda_n\left(\frac{k}{n}\right) = \sum_{i=1}^k \text{Log} \frac{f(X_i, \theta_1^{(n)})}{f(X_i, \theta_0)} + \sum_{i=k+1}^n \text{Log} \frac{f(X_i, \theta_2^{(n)})}{f(X_i, \theta_0)}. \quad (2)$$

According to a size α ($\alpha \in]0, 1[$), the likelihood ratio test rejects $H_0(\theta_0)$ in favour of $H_1^{(n)}$ whereas $\Lambda_n^* \geq C_\alpha$ where the threshold C_α is determined by the equation $P_{H_0(\theta_0)}(\Lambda_n^* \geq C_\alpha) \leq \alpha$ with a corresponding power of $\beta_n = P_{H_1^{(n)}}(\Lambda_n^* \geq C_\alpha)$.

Definition 1 We define $\{\Lambda_n(t); t \in [0, 1]\}$ as the interpolated linear process of the process $\{\Lambda_n(k/n); k = 1, \dots, n-1\}$.

2.3 Asymptotic behaviour of Λ_n

The Taylor expansion of the right side of (2) gives:

$$\begin{aligned} \Lambda_n(t) &= \frac{u_1}{\sqrt{n}} \sum_{i=1}^{[nt]} \frac{\partial}{\partial \theta} \text{Log} f(X_i, \theta_0) + \frac{u_2}{\sqrt{n}} \sum_{i=[nt]+1}^n \frac{\partial}{\partial \theta} \text{Log} f(X_i, \theta_0) \quad (3) \\ &+ \frac{u_1^2}{2n} \sum_{i=1}^{[nt]} \frac{\partial^2}{\partial \theta^2} \text{Log} f(X_i, \theta_0) + \frac{u_2^2}{2n} \sum_{i=[nt]+1}^n \frac{\partial^2}{\partial \theta^2} \text{Log} f(X_i, \theta_0) + o(1). \end{aligned}$$

The weak convergence in $\mathcal{C}([0, 1])$ will be denoted by $\xrightarrow{\mathcal{C}}$.

Proposition 1 For any $t \in [0, 1]$, under $H_0(\theta_0)$, we have that $\Lambda_n \xrightarrow{\mathcal{C}} \Lambda_0$ and the process Λ_0 is defined by

$$\Lambda_0(t) = I^{\frac{1}{2}}(\theta_0) [(u_1 - u_2)W(t) + u_2W(1)] - \frac{1}{2}[u_1^2 t + u_2^2(1-t)] I(\theta_0) \quad (4)$$

with $\{W(t); t \in [0, 1]\}$ is a standard Brownian motion.

Proof For each $i = 1, 2, \dots, n$, let us consider the r.v.'s $Z_i = \partial \text{Log} f(X_i, \theta_0) / \partial \theta$. It can be easily verified that, under $H_0(\theta_0)$, $Z_1, Z_2, \dots, Z_n$ are independent r.v.'s, satisfying

$$E_{H_0(\theta_0)}[Z_i] = 0 \quad \text{Var}_{H_0(\theta_0)}[Z_i] = I(\theta_0)$$

for each $i = 1, 2, \dots, n$.

By Theorem 12.1 of Billingsley (1968), under the hypothesis $H_0(\theta_0)$, $(1/\sqrt{n}) \sum_{i=1}^{[nt]} Z_i$ and $(1/\sqrt{n}) \sum_{i=[nt]+1}^n Z_i$ converge in distribution to $I^{1/2}(\theta_0)W(t)$ and $I^{1/2}(\theta_0)(W(1)-W(t))$. On the other hand, by the law of large numbers, $(-1/n) \sum_{i=1}^{[nt]} \partial^2 \text{Log} f(X_i, \theta_0)/\partial \theta^2$ and $(-1/n) \sum_{i=[nt]+1}^n \partial^2 \text{Log} f(X_i, \theta_0)/\partial \theta^2$ converge in probability to $tI(\theta_0)$ and $(1-t)I(\theta_0)$. $\square$

We deduce, for any η given,

$$\lim_{n \rightarrow \infty} P\left(\sup_{t \in [0,1]} \Lambda_n(t) \geq \eta\right) = P\left(\sup_{t \in [0,1]} \Lambda_0(t) \geq \eta\right),$$

and the following lemma:

Lemma 1 *For all fixed $s \in [0, 1]$ the probability measures $P_{H_1^{(n)}(u_1, [ns], u_2)}$ are contiguous to $P_{H_0(\theta_0)}$.*

Proof Let,

$$U = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \Sigma_s = \begin{pmatrix} sI(\theta_0) & 0 \\ 0 & (1-s)I(\theta_0) \end{pmatrix} \text{ and } V_s^2 = U^* \Sigma_s U,$$

where U^* indicates the transpose of vector U .

By Proposition 1, under $P_{H_0(\theta_0)}$, Λ_n has an asymptotic normal distribution with mean $(-1/2)V_s^2$ and variance V_s^2 . Hence the contiguity of $P_{H_1^{(n)}(u_1, [ns], u_2)}$ to $P_{H_0(\theta_0)}$ follows directly from *LeCam's First Lemma* (see Lemma 1.2, VI, in Hájek and Šidák, 1967). $\square$

Proposition 2 *For a fixed $\tau_0 \in [0, 1]$, under $H_1^{(n)}(u_1, [n\tau_0], u_2)$, we have that $\Lambda_n \xrightarrow{\mathcal{C}} \Lambda_{\tau_0}$ and the process Λ_{τ_0} is defined by*

$$\Lambda_{\tau_0}(t) = \Lambda_0(t) + [u_1^2(t \wedge \tau_0) + u_2^2(1 - t \vee \tau_0) + u_1 u_2(t \vee \tau_0 - t \wedge \tau_0)]I(\theta_0) \quad (5)$$

with $(A \wedge B) = \min(A, B)$ and $(A \vee B) = \max(A, B)$.

Proof It suffices to show that under $H_1^{(n)}(u_1, [n\tau_0], u_2)$:

- (i) the tightness of Λ_n .
- (ii) all finite dimensional distribution of Λ_n converge to those of Λ_{τ_0} .

Tightness. By Proposition 1, the process Λ_n converges under $H_0(\theta_0)$ to an affine function of the Brownian motion process and therefore it is tight under $P_{H_0(\theta_0)}$. Since by Lemma 1, the probability measures $P_{H_1^{(n)}(u_1, [n\tau_0], u_2)}$

are contiguous to $P_{H_0(\theta_0)}$, it follows from theorem 4.3.4 of Sen (1981) that the process Λ_n is also tight under $H_1^{(n)}(u_1, [n\tau_0], u_2)$.

Finite dimensional distributions. For any $r \geq 1$, let $t_1, t_2, \dots, t_{r+1}$ verify $0 \leq t_1 < t_2 < \dots < t_r \leq 1$ with $\tau_0 = t_{r+1}$ and

$$L_n = \begin{pmatrix} \Lambda_n(t_1) \\ \vdots \\ \Lambda_n(t_r) \end{pmatrix}, K_n = \begin{pmatrix} L_n \\ \Lambda_n(\tau_0) \end{pmatrix} \text{ and } L(\tau_0) = \begin{pmatrix} \Lambda_{\tau_0}(t_1) \\ \vdots \\ \Lambda_{\tau_0}(t_r) \end{pmatrix}.$$

We need to show that under $H_1^{(n)}(u_1, [n\tau_0], u_2)$, L_n converges to $L(\tau_0)$.

Using Proposition 1, it can be verified that under $P_{H_0(\theta_0)}$, K_n converges to $\mathcal{N}(M_0, V_0)$ where M_0 is the mean vector defined by

$$M_0 = \begin{pmatrix} m_1^0 \\ \vdots \\ m_{r+1}^0 \end{pmatrix}$$

with $m_i^0 = (-1/2)[u_1^2 t_i + u_2^2(1 - t_i)]I(\theta_0)$, $i = 1, \dots, r+1$, and V_0 is the covariance matrix defined by $V_0 = (\sigma_{i,j})_{1 \leq i,j \leq r+1}$, the elements $\sigma_{i,j}$ are given by

$$\sigma_{i,j} = [u_1^2(t_i \wedge t_j) + u_2^2(1 - t_i \vee t_j) + u_1 u_2(t_i \vee t_j - t_i \wedge t_j)] I(\theta_0).$$

Since for any fixed $\tau_0 \in [0, 1]$, the probability measures $P_{H_1^{(n)}(u_1, [n\tau_0], u_2)}$ are contiguous to $P_{H_0(\theta_0)}$, it follows at once from *LeCam's Third Lemma* (see Lemma 1.4 ,VI, in Hájek and Šidák, 1967), under $H_1^{(n)}(u_1, [n\tau_0], u_2)$, L_n converges to $\mathcal{N}(M_1, V_1)$ where M_1 is the mean vector defined by

$$M_1 = \begin{pmatrix} m_1^1 \\ \vdots \\ m_{r+1}^1 \end{pmatrix}$$

with $m_i^1 = m_i^0 + \sigma_{i,r+1}$, $i = 1, \dots, r$ and V_1 is the corresponding covariance matrix defined by $V_1 = (\sigma_{i,j})_{1 \leq i,j \leq r}$.

Furthermore, for each $i = 1, \dots, r$,

$$\sigma_{i,r+1} = [u_1^2(t_i \wedge \tau_0) + u_2^2(1 - t_i \vee \tau_0) + u_1 u_2(t_i \vee \tau_0 - t_i \wedge \tau_0)] I(\theta_0),$$

since $\tau_0 = t_{r+1}$. Therefore,

$$\begin{aligned} m_i^1 &= [u_1^2(t_i \wedge \tau_0) + u_2^2(1 - t_i \vee \tau_0) + u_1 u_2(t_i \vee \tau_0 - t_i \wedge \tau_0)] I(\theta_0) \\ &\quad - (1/2)[u_1^2 t_i + u_2^2(1 - t_i)] I(\theta_0), \end{aligned}$$

which is just the mean of $\Lambda_{\tau_0}(t_i)$, $i = 1, \dots, r$.

Finally, comparison of variance structure of the process Λ_n with that of the process Λ_{τ_0} , completes the proof. $\square$

2.4 Asymptotic level and power function

First we shall give through asymptotic evaluation of the probabilities of crossing the frontier the affine Brownian functions, which will allow the calculation of the asymptotic level and power function.

Proposition 3 *Let $\{W(t); t \in [0, T]\}$ be a standard Brownian motion and $\{Z(t); t \in [0, T]\}$ be a process defined by $Z(t) = \alpha W(t) + \beta W(T) + \alpha_0 t + \beta_0$. For any $\eta \geq \beta_0$, we have*

$$P\left(\sup_{t \in [0, T]} Z(t) \geq \eta\right) = 1 - \Phi(a) + \int_{-\infty}^a \phi(\xi) \Phi\left(\frac{|\beta|}{\alpha}(\xi - a) + \frac{\alpha_0 \sqrt{T}}{\alpha}\right) d\xi \\ + \exp\left\{\frac{2\alpha_0(\eta - \beta_0) + 2(\alpha_0 \beta \sqrt{T})^2}{\alpha^2}\right\} \int_{-\infty}^b \phi(\xi) \Phi\left(\frac{|\beta|}{\alpha}(\xi - b) - \frac{\alpha_0 \sqrt{T}}{\alpha}\right) d\xi$$

where $a = (\eta - \beta_0)/(|\beta| \sqrt{T})$ and $b = a + (2\alpha_0 |\beta| \sqrt{T})/\alpha^2$ with Φ beeing the standard Normal c.d.f. with density ϕ .

This result can be obtained, on evaluating the joint distribution of $(W(T), \sup_{t \in [0, T]} W(t))$ by the reflection principle, then using the Girsanov theorem.

Let us show first the following lemma.

Lemma 2 *Let $\{W(t); t \in [0, T]\}$ be a standard Brownian motion. Then the probability density function of $\sup_{t \in [0, T]} \{W(t) + \mu t + \gamma\}$ is given by*

$$\left\{\frac{2}{\sqrt{T}} \phi\left(\frac{y - \mu T - \gamma}{\sqrt{T}}\right) - 2\mu e^{2\mu(y - \gamma)} (1 - \Phi\left(\frac{y - \mu T - \gamma}{\sqrt{T}}\right))\right\} \mathbf{1}_{\{y \geq \gamma\}} \quad (6)$$

with Φ denoting the standard Normal c.d.f. with density ϕ .

Proof By the reflection principle, (see Bhattacharya and Waymire, 1990, p. 432), the joint density function of $(W(T), \sup_{t \in [0, T]} W(t))$ is given by

$$\frac{2(2y - x)}{T^{\frac{3}{2}}} \phi\left(\frac{2y - x}{\sqrt{T}}\right) \mathbf{1}_{\{x \leq y, y \geq 0\}}.$$

Let $\mathcal{F}_T = \sigma\{W(s); 0 \leq s \leq T\}$, by the Girsanov theorem (see Williams, 1979), we have

$$\frac{dP_\mu}{dP_0} \Big|_{\mathcal{F}_T} = \exp\left\{\mu \int_0^T dW(s) - \frac{\mu^2}{2} \int_0^T ds\right\}$$

if P_μ is the distribution of the process on $\mathcal{C}([0, 1]) : t \mapsto W(t) + \mu t$.

Hence, the joint density function of $(W(T), \sup_{t \in [0, T]} \{W(t) + \mu t\})$ is given by

$$\frac{2}{\sqrt{2\pi}} \frac{(2y-x)}{T^{\frac{3}{2}}} \exp\left\{-\frac{(2y-x)^2}{2T} + \mu x - \frac{\mu^2 T}{2}\right\} \mathbf{1}_{\{x \leq y, y \geq 0\}}. \quad (7)$$

For the proof of Proposition 3, we notice that

$$\begin{aligned} & P\left(\sup_{t \in [0, T]} Z(t) \geq \eta \mid W(T) = z\right) \\ &= P\left(\sup_{t \in [0, T]} Z(t) - \beta W(T) \geq \eta - \beta z \mid W(T) = z\right) \\ &= P\left(\sup_{t \in [0, T]} \{\alpha W(t) + \alpha_0 t + \beta_0\} \geq \eta - \beta z\right). \end{aligned}$$

Using the Lemma 2, we obtain the result. $\square$

Let $\delta = (u_1 - u_2) I^{1/2}(\theta_0)$ and $\Delta = (u_1 + u_2) I^{1/2}(\theta_0)$.

• *Asymptotic level:* for any $\eta \geq -(\Delta - \delta)^2/8$, we have

$$\begin{aligned} P\left(\sup_{t \in [0, 1]} \Lambda_0(t) \geq \eta\right) &= \Phi(-a) + \int_{-\infty}^a \phi(\xi) \Phi\left(\frac{|\Delta - \delta|}{2\delta}(\xi - a) - \frac{\Delta}{2}\right) d\xi \\ &+ \exp\left\{-\frac{\Delta(8\eta + (\Delta - \delta)^2(1 - \Delta))}{8\delta}\right\} \int_{-\infty}^b \phi(\xi) \Phi\left(\frac{|\Delta - \delta|}{2\delta}(\xi - b) + \frac{\Delta}{2}\right) d\xi \quad (8) \end{aligned}$$

where $a = (8\eta + (\Delta - \delta)^2)/(4|\Delta - \delta|)$ and $b = a - (\Delta|\Delta - \delta|)/\delta$.

In the particular case $\Delta = 0$, we have

$$P\left(\sup_{t \in [0, 1]} \Lambda_0(t) \geq \eta\right) = 2 \left[\Phi\left(-\frac{8\eta + \delta^2}{4\delta}\right) + \frac{8\eta + \delta^2}{4\delta} \phi\left(\frac{8\eta + \delta^2}{4\delta}\right) \right]. \quad (9)$$

• *Asymptotic power function:* The process Λ_{τ_0} , given by (5), is written as

$$\Lambda_{\tau_0}(t) = \begin{cases} \delta W(t) + \frac{\Delta - \delta}{2} W(1) + \frac{\delta^2}{2} t + \frac{(\Delta - \delta)^2 + 4\delta(\Delta - \delta)\tau_0}{8} & \text{if } t < \tau_0 \\ \delta W(t) + \frac{\Delta - \delta}{2} W(1) - \frac{\delta^2}{2} t + \frac{(\Delta - \delta)^2 + 4\delta(\Delta + \delta)\tau_0}{8} & \text{if } t > \tau_0 \end{cases}$$

Let $M = \sup_{t \in [0, 1]} \Lambda_{\tau_0}(t)$, $M_1 = \sup_{t \in [0, \tau_0]} \Lambda_{\tau_0}(t)$ and $M_2 = \sup_{t \in [\tau_0, 1]} \Lambda_{\tau_0}(t)$ then, for any $\eta \geq -(\Delta - \delta)^2/8$, we have

$$P(M \geq \eta) = P(M_1 \geq \eta) + P(M_2 \geq \eta) - Q(\eta, \tau_0)$$

with

$$Q(\eta, \tau_0) = \int P(M_1 \geq \eta, M_2 \geq \eta \mid W(\tau_0) = z_1, W(1) = z_2) dP(z_1, z_2).$$

The first two terms of that last expression are obtained using Lemma 2. For the third term, we must notice that conditional as $(W(\tau_0), W(1))$ the variables $M_1 - \delta W(\tau_0) - ((\Delta - \delta)/2)W(1)$ and $M_2 - \delta W(\tau_0) - ((\Delta - \delta)/2)W(1)$ are independent. $\square$

3 Simulations

We have simulated 1000-samples of distributions $\mathcal{N}(0, 1)$ and $Exp(2)$ for the following hypotheses testing problem:

- Normal distribution

$$H_0 : X_i \sim \mathcal{N}(0, 1) \quad for \quad i = 1, 2, \dots, n$$

$$H_1^{(n)}(u_1, k) : \begin{cases} X_i \sim \mathcal{N}(\frac{u_1}{\sqrt{n}}, 1) & for \quad i = 1, 2, \dots, k \\ X_i \sim \mathcal{N}(-\frac{u_1}{\sqrt{n}}, 1) & for \quad i = k + 1, \dots, n \end{cases}$$

- Exponential distribution

$$H_0 : X_i \sim Exp(2) \quad for \quad i = 1, 2, \dots, n$$

$$H_1^{(n)}(u_1, k) : \begin{cases} X_i \sim Exp(2 + \frac{u_1}{\sqrt{n}}) & for \quad i = 1, 2, \dots, k \\ X_i \sim Exp(2 - \frac{u_1}{\sqrt{n}}) & for \quad i = k + 1, \dots, n \end{cases}$$

with $n=450$ and $k=150$.

For every sample we have calculated the corresponding statistics, given by (2). Figure 1 presents the proportions of the detections according to $\delta = 2u_1 I$, where I denotes the Fisher information. The graphics show that the proportions of the detections increase with the amplitude δ , (δ amplitude of the change point), on the other hand, the test is more powerful in the case of the normal distribution.

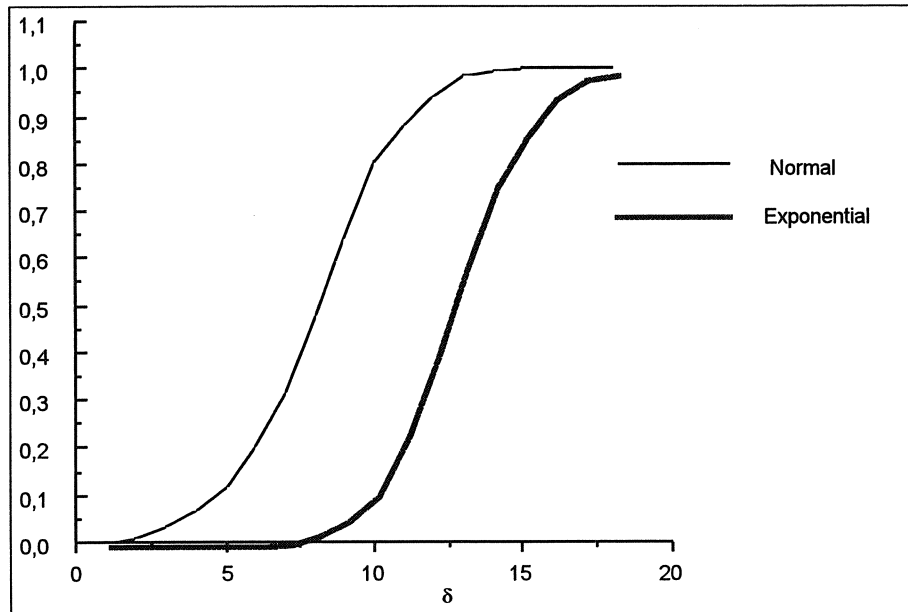


Figure 1: Proportions of detections for normal and exponential cases.

Résumé: Nous considérons une catégorie de tests permettant de détecter un changement dans le paramètre d'une séquence de variables aléatoires indépendantes. La catégorie des tests considérés est basée sur le ratio de vraisemblance approprié pour ce problème de changement. La distribution limitée des tests statistiques sous l'hypothèse nulle et alternative est évaluée par des techniques asymptotiques suivant le concept de contiguïté de LeCam. Des formulations analytiques de la fonction de puissance asymptotique sont définies. La performance du test, dans deux cas particuliers, distribution normale et exponentielle, est étudiée par simulation.

Mots clés: Contiguïté, point de changement, principe de réflexion, théorème de Girsanov.

References

- [1] Bhattacharya, R.N. and Waymire, E.C. (1990), *Stochastic Processes with Application*, Wiley, New York.
- [2] Billingsley, P. (1968), *Convergence of Probability Measures*, Wiley, New York.
- [3] Chernoff, H. and Zacks, S. (1964), "Estimating the current mean of a

- normal distribution which is subjected to change in time", *Ann. Math. Statist.*, **35**, 999-1018.
- [4] Cobb, G.W. (1978), "The problem of the Nile : Conditional solution to a changepoint problem", *Biometrika*, **65**(2), 243-251.
 - [5] Deshayes, J. and Picard, D. (1986), "Off line statistical analysis of change-point models using non parametric and likelihood methods", *Lecture Notes in Control and Information Science*, N° 77, Springer-Verlag, Berlin, 103-168.
 - [6] Gardner, L.A. JR. (1969), "On detecting changes in the mean of normal variates", *Ann. Math. Statist.*, **40**, 116-126.
 - [7] Hájek, J. and Šidák, Z. (1967), *Theory of Rank Tests*, Academic Press, New York.
 - [8] Hawkins, D.M. (1977), "Testing a sequence of observations for a shift in location", *J. Amer. Statist. Assoc.*, **72**, 180-186.
 - [9] Kim, H.J. and Siegmund, D. (1981), "The likelihood ratio test for a change point in simple linear regression", *Biometrika*, **57**, 409-423.
 - [10] Sen, P.K. (1981), *Sequential Nonparametric, Invariance Principal and Statistical Inference*, John Wiley, New York.
 - [11] Sen, A. and Srivastava, M.S. (1975), "On tests for detecting change in mean", *Ann. Statist.*, **3**, 90-108.
 - [12] Williams, D. (1979), *Diffusions, Markov processes and Martingales*, Vol 1, Wiley, New York.
 - [13] Worsley, K.J. (1983), "The power of likelihood ratios and cumulative sum tests for a change in binomial probability", *Biometrika*, **70**, 91-104.
 - [14] Worsley, K.J. (1986), "Confidence regions and tests for a change-point in a sequence of exponential family random variables", *Biometrika*, **73**, 455-464.

Estimation de la Variance de l'Indice Suisse des Prix à la Consommation

Sandrine König

Département des finances et des affaires sociales, Etat de Neuchâtel, Suisse

Résumé: L'indice suisse des prix à la consommation est calculé avec des données provenant de cinq échantillons différents (ménages, communes, points de vente, articles et habitations). Comme la construction de cet indice est complexe, il n'est pas possible d'estimer sa variance à l'aide des méthodes traditionnelles. Cet article propose néanmoins une méthodologie basée sur la méthode des répliques qui permet d'estimer la variance de cet indice.

Mots clés: Echantillonnage, estimation de la variance, indice de Laspeyres, indice suisse des prix à la consommation, méthode des répliques.

1 Introduction

L'indice suisse des prix à la consommation (ISPC) est calculé mensuellement par l'Office fédéral de la statistique (OFS). Sa méthode de construction et de calcul est exposée notamment dans OFIAMT (1985) et dans OFS (1990, 1993) et est résumée ci-dessous.

L'ISPC entre les périodes t et 0 , que l'on note ici par I^t , est calculé en utilisant la formule de Laspeyres et

$$I^t = \sum_j w_j \cdot I_j^t$$

où I_j^t est l'indice du produit j entre les périodes t et 0 , et où w_j est le "coefficient budgétaire" représentant la part du budget des ménages consacrée au produit j à la période 0 .

Pour estimer les coefficients budgétaires, on échantillonne un certain nombre de ménages suisses. Pour estimer les indices des différents pro-

duits, on échantillonne tout d'abord des aires géographiques, par exemple des communes, puis à l'intérieur des aires sélectionnées, on échantillonne différents points de vente, par exemple Migros et Coop, et enfin, dans ces points de vente sélectionnés, on échantillonne des articles représentant le produit en question. Par ailleurs, un échantillon d'habitations est requis pour estimer le niveau moyen des loyers (on pourra ainsi estimer I_j^t , où j est le produit "loyer").

On utilisera alors la formule suivante:

$$I_j^t = \sum_k a_k \cdot I_{jk}^t$$

où I_{jk}^t est l'indice du produit j au point de vente k avec

$$I_{jk}^t = \sum_l g_l \cdot I_{jkl}^t$$

où I_{jkl}^t est l'indice du produit j au point de vente k dans la commune l avec

$$I_{jkl}^t = 100 \cdot \frac{x_{jkl}^t}{x_{jkl}^0}$$

où x_{jkl}^t et x_{jkl}^0 sont les valeurs du produit j au point de vente k dans la commune l , respectivement aux périodes t et 0 , estimées à partir de l'échantillon des articles.

Les a_k et les g_l sont des coefficients de pondération relatifs à l'importance respectivement des points de vente et des communes. Par exemple, au niveau des communes, on a un coefficient g_l de 0.2 pour Zürich, 0.1 pour Bâle, 0.1 pour Berne, 0.1 pour Genève et 0.5 pour l'ensemble des autres communes de Suisse.

Comme la construction de l'ISPC est complexe, sa variance ne peut pas être estimée en utilisant les formules analytiques classiques telles que celles développées par Cochran (1963) et par Hansen et al. (1953). Qui plus est, les données utilisées ne proviennent pas toutes d'échantillons aléatoires. Cet article propose néanmoins une méthodologie qui permet d'estimer la variance de l'ISPC ainsi que la variance d'un ratio de deux ISPC consécutifs, en utilisant la méthode des répliques ainsi que les séries de Taylor. Notons toutefois que cette proposition a été développée avant la dernière révision complète de l'ISPC qui s'est achevée en mai 1993. C'est pourquoi, on a considéré ici la méthode de construction et de calcul de l'ISPC qui était utilisée avant mai 1993. Cependant la méthodologie proposée peut aussi être adaptée à l'ISPC révisé.

2 Méthode des répliques et séries de Taylor

Soient n observations $x_1, \dots, x_n$ provenant d'une population de moyenne μ . Dans le cas d'un échantillon aléatoire simple, l'estimateur usuel de μ est $\bar{x} = \sum x_i/n$ et la variance de cet estimateur est estimée par

$$v(\bar{x}) = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n \cdot (n-1)}.$$

Considérons maintenant un estimateur $\hat{\theta}$ d'un paramètre d'intérêt quelconque θ obtenu à partir d'un échantillon aléatoire. Avec la méthode des répliques, le principe de base pour estimer la variance de $\hat{\theta}$ est le suivant. Premièrement, on divise aléatoirement l'échantillon initial en B groupes appelés répliques. Puis, on calcule des estimateurs $\hat{\theta}_\alpha$ de θ dans chaque réplique. On définit ensuite $\bar{\theta}$ par

$$\bar{\theta} = \sum_{\alpha=1}^B \frac{\hat{\theta}_\alpha}{B}.$$

La variance de $\hat{\theta}$ est alors estimée par

$$v(\hat{\theta}) = \sum_{\alpha=1}^B \frac{(\hat{\theta}_\alpha - \bar{\theta})^2}{B \cdot (B-1)}.$$

Au paragraphe 4, cette méthode est appliquée pour estimer la variance de l'ISPC.

Pour estimer la variance d'un ratio de deux variables aléatoires Y et X , on peut utiliser l'approximation suivante obtenue à l'aide de séries de Taylor (voir Wolter, 1985, p.236):

$$v\left(\frac{Y}{X}\right) = \left(\frac{Y}{X}\right)^2 \cdot \left(\frac{v(Y)}{Y^2} + \frac{v(X)}{X^2} - 2\frac{c(Y, X)}{XY}\right)$$

où $v(Y)$, $v(X)$ et $c(Y, X)$ sont respectivement des estimateurs de la variance de Y , de la variance de X et de la covariance entre Y et X . Au paragraphe 5, cette formule est utilisée pour estimer la variance du ratio de deux ISPC consécutifs.

3 Choix des répliques

Afin de pouvoir utiliser la méthode des répliques exposée ci-dessus, on construira des répliques à l'intérieur de l'échantillon des communes. On distinguera en fait deux types de communes. Le premier type sera constitué

des communes dont on supposera qu'elles ne représentent qu'elles-mêmes. On appellera de telles communes des unités primaires d'échantillonnage autoreprésentatives (UPE-AR). Ce groupe de communes sera par exemple composé de Bâle, Berne, Genève et Zurich qui ont chacune un coefficient de pondération qui leur est propre dans le calcul de l'ISPC, comme on l'a vu dans l'introduction. En règle générale, on retiendra dans ce type de communes celles qui sont retenues d'office dans un échantillon aléatoire de communes, c'est-à-dire les grandes villes. Le second type englobera les communes qui représentent non seulement elles-mêmes mais aussi les autres communes. On parlera alors d'unités primaires d'échantillonnage non autoreprésentatives (UPE-NAR). Ce groupe de communes sera composé des communes restantes.

On procédera de manière différente selon qu'il s'agit des UPE-AR ou des UPE-NAR. En particulier, il faudra disposer d'au moins deux répliques par UPE-AR. Pour cela, on devra avoir au moins deux échantillons différents de points de vente et d'articles à l'intérieur de chaque UPE-AR. Ces échantillons constitueront les répliques et on calculera pour chacune d'entre elles un ISPC. Dans le cas des UPE-NAR, on considèrera chaque commune comme une réplique des autres communes. Il serait cependant aussi possible d'envisager plusieurs groupes d'UPE-NAR, par exemple un groupe réunissant les communes de taille moyenne et un autre groupe celles de petite taille. La Table 1 nous donne un exemple de constitution de répliques à partir d'un échantillon de 48 communes.

	Communes	Nombre d'UPE	Nombre de répliques
UPE-AR	Bâle	1	2
	Berne	1	2
	Genève	1	2
	Zurich	1	4
UPE-NAR	Autres communes	44	44
Total		48	54

Table 1: Constitution des répliques à partir des communes

4 Variance de l'ISPC

Voyons à présent comment on estime la variance de l'ISPC. La méthode

consiste tout d'abord à calculer un indice I_l^t dans chaque commune l , puis à estimer les variances de ces indices communaux à l'aide de la méthode des répliques. On posera finalement

$$v(I^t) = \sum_l g_l \cdot v(I_l^t).$$

Pour le calcul des I_l^t , on procédera dans chaque commune l comme s'il s'agissait de la Suisse toute entière. On posera donc

$$I_l^t = \sum_j \sum_k w_j \cdot a_k \cdot I_{jkl}^t$$

où I_{jkl}^t est l'indice du produit j au point de vente k dans la commune l .

Pour estimer la variance de I_l^t pour une commune l de type UPE-AR, on calculera encore un indice I_α^t dans chaque réplique α de la même manière que ci-dessus, à savoir

$$I_\alpha^t = \sum_j \sum_k w_j \cdot a_k \cdot I_{jk\alpha}^t$$

où $I_{jk\alpha}^t$ est l'indice du produit j au point de vente k dans la réplique α . On posera ensuite

$$v(I_l^t) = \sum_{\alpha=1}^{B_l} \frac{(I_\alpha^t - I_l^t)^2}{B_l \cdot (B_l - 1)}$$

où B_l est le nombre de répliques pour la commune l , par exemple $B_l = 2$ pour Bâle, Berne et Genève et $B_l = 4$ pour Zürich si on suit les chiffres de la Table 1.

Pour estimer la variance de I_l^t pour une commune l de type UPE-NAR, on calculera la moyenne $\bar{I}_l^t$ des I_l^t de toutes les communes de ce type, à savoir

$$\bar{I}_l^t = \frac{1}{B} \sum_{l=1}^B I_l^t$$

où B est le nombre de communes de type UPE-NAR, par exemple $B = 44$. On posera ensuite

$$v(I_l^t) = \sum_{l=1}^B \frac{(I_l^t - \bar{I}_l^t)^2}{B \cdot (B - 1)}.$$

5 Variance d'un ratio de deux ISPC

Intéressons-nous à présent à un ratio de deux ISPC consécutifs

$$\Delta I^t = \frac{I^t - I^{t-1}}{I^{t-1}} = \frac{I^t}{I^{t-1}} - 1.$$

En utilisant la formule pour estimer la variance d'un ratio de deux variables aléatoires vue au paragraphe 2, on peut obtenir une estimation de la variance de ΔI^t :

$$\begin{aligned} v(\Delta I^t) &= v\left(\frac{I^t}{I^{t-1}}\right) \\ &= \left(\frac{I^t}{I^{t-1}}\right)^2 \cdot \left(\frac{v(I^t)}{(I^t)^2} + \frac{v(I^{t-1})}{(I^{t-1})^2} - 2\frac{c(I^t, I^{t-1})}{I^{t-1} \cdot I^t}\right) \\ &= \frac{1}{(I^{t-1})^2} \cdot \left(v(I^t) + \left(\frac{I^t}{I^{t-1}}\right)^2 v(I^{t-1}) - 2\frac{I^t}{I^{t-1}} c(I^t, I^{t-1})\right) \end{aligned}$$

où $v(I^t)$ et $v(I^{t-1})$ sont calculés comme dans le paragraphe précédent, et où $c(I^t, I^{t-1})$, l'estimation de la covariance entre l'ISPC à la période t et l'ISPC à la période $t-1$, est calculée en utilisant la méthode des répliques pour la covariance. C'est-à-dire que l'on calculera tout d'abord pour chaque commune l de type UPE-AR

$$c(I_l^t, I_l^{t-1}) = \sum_{\alpha=1}^{B_l} \frac{(I_\alpha^t - I_l^t)(I_\alpha^{t-1} - I_l^{t-1})}{B_l \cdot (B_l - 1)}$$

et pour chaque commune l de type UPE-NAR

$$c(I_l^t, I_l^{t-1}) = \sum_{i=1}^B \frac{(I_l^t - \bar{I}_l^t)(I_l^{t-1} - \bar{I}_l^{t-1})}{B \cdot (B - 1)}.$$

On posera ensuite

$$c(I^t, I^{t-1}) = \sum_l g_l \cdot c(I_l^t, I_l^{t-1}).$$

6 Conclusion

L'estimation de la variance de l'ISPC nous permettra de calculer la précision de cet indice, en particulier de calculer un intervalle de confiance qui contiendra l'ISPC réel avec une grande probabilité. L'estimation de la variance

d'un ratio de deux ISPC consécutifs nous permettra de tester si une variation de l'ISPC observée entre deux périodes est vraiment significative ou si elle est seulement due aux aléas de l'échantillonnage.

Ces calculs nous paraissent essentiels car l'ISPC est l'un des indicateurs économiques les plus importants. Il sert notamment à mesurer l'inflation, à déflater certains indicateurs économiques ou encore à indexer les loyers, les rentes et les salaires.

L'estimation de la variance des indices des prix à la consommation n'en est d'ailleurs qu'à ces premiers pas au niveau international; quelques pays comme la Suède (Dalén, 1991), Les Pays-Bas (Balk, 1991), ou l'Italie (Biggeri and Giommi, 1987) se sont intéressés à ce problème. Remarquons que les Etats-Unis (voir BLS, 1988) sont, à l'heure actuelle, le seul pays où la construction de l'indice des prix à la consommation repose sur un ensemble d'échantillons entièrement aléatoires et où l'on estime officiellement sa variance. Au niveau international, notamment chez Turvey (1989), le cas américain est considéré comme le modèle de référence.

Remerciements: Nous remercions le Fonds national suisse de la recherche scientifique (projet no 12-27837.89) pour son soutien financier.

Abstract: The Swiss Consumer Price Index is calculated with data issued from five different samples (households, municipalities, points of purchase, item and housing). As its construction is very complex, it is impossible to estimate its variance with traditional methods. In this paper, we propose however a methodology based on the method of random groups that allows to estimate the variance of this index.

Key words: Laspeyres price index, method of random groups, sampling, Swiss consumer price index, variance estimation.

Références

- [1] Balk, B.M. (1991), "Estimating the Precision of a Consumer Price Index; Some Experiences from the Netherlands", *Proc. of the 48th ses. of the ISI*, book 1, 51-52, Cairo.
- [2] Biggeri, L. and Giommi, A. (1987), "On the Accuracy and Precision of the Consumer Price Indices. Methods and Applications to Evaluate the Influence of the Sampling of Households", *Proc. of the 46th ses. of the ISI*, book 3, 137-154, Tokyo.
- [3] BLS. (1988), "BLS Handbook of Methods", Bulletin 2285, Bureau of Labor Statistics, U.S. Department of Labor, U.S.A.

- [4] Cochran, W.G. (1963), *Sampling Techniques*, John Wiley, New York.
- [5] Dalén, J. (1991), "The Impact of Sampling and Non-Sampling Errors on the Swedish CPI", *Proc. of the 48th ses. of the ISI*, book 1, 151-152, Cairo.
- [6] Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1960), *Sample Survey Methods and Theory*, Vol. I and II, John Wiley, New York.
- [7] OFIAMT (1985), "Nouvelle méthode de calcul de l'indice suisse des prix à la consommation. Base décembre 1982=100.97", *La Vie économique*, numéro spécial, Office fédéral de l'industrie, des arts et métiers et du travail, Berne.
- [8] OFS (1990), "La révision de l'indice suisse des prix à la consommation. Concept de base", Rapport technique, Berne.
- [9] OFS (1993), "Révision de l'indice suisse des prix à la consommation. Conception du nouvel indice suisse des prix à la consommation", Rapport technique, Berne.
- [10] Turvey, R. et al. (1989), "Consumer Prices Indices. An ILO Manual", Bureau International du Travail, Geneva.
- [11] Wolter, K.M. (1985), *Introduction to Variance Estimation*, Springer-Verlag, New York.

On Shrinkage Ridge Regression Estimators for Restricted Linear Models with Multivariate t Disturbances

B.M. Golam Kibria

University of Western Ontario, London, Canada

Abstract: This paper combines the idea of shrinkage estimation and ridge regression methodology, when it is suspected that the regression coefficients may be restricted to a subspace and the regression disturbances follow a multivariate t distribution. We propose three estimators namely the Unrestricted Ridge Regression Estimator, the Restricted Ridge Regression Estimator and finally the Shrinkage Ridge Regression Estimator. The bias and the mean square errors (MSE) of the proposed estimators are derived under the null and alternative hypotheses. By studying the MSE criterion, the regions of optimality of the estimators are determined. The conditions of superiority of the proposed estimator over the traditional estimators are also discussed.

Key words: Dominance, mean square error, multivariate t distribution, non central Chi-square, ridge regression, shrinkage estimator, superiority.

1 Introduction

Consider the linear regression model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \quad (1)$$

where $\mathbf{Y}$ is a $n \times 1$ vector of observations, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression coefficients, $\mathbf{X}$ is a $n \times p$ known design matrix of rank p and $\mathbf{e}$ is an $n \times 1$ error vector which follows a multivariate student t distribution with ν degrees of freedom, mean vector $\mathbf{0}$ and variance covariance matrix, $(\nu\sigma^2/(\nu-2))\mathbf{I}_n$.

In regression analysis the distribution of the errors is usually assumed

to be normal. However, in practice there may arise some cases where the error distribution has heavier tails than that of normal distribution. One can tackle this problem by using the well known t -distribution, as it has heavier tails than that of normal distribution. As $\nu \rightarrow \infty$, the probability density function of student t approaches to normal density and for $\nu = 1$, it becomes Cauchy. Therefore the normal and the Cauchy error distributions may be considered as special cases of this model.

Our primary interest is to estimate the regression coefficient β when the error distribution is multivariate t and it is *a priori* suspected that β may be restricted to the subspace $\mathbf{H}\beta = \mathbf{h}$, where $\mathbf{H}$ is a $q \times p$ known matrix of rank $q(< p)$ and $\mathbf{h}$ is an $q \times 1$ vector of known constants. The usual least squares estimator (LSE) of β depends on the characteristics of the matrix $\mathbf{C} = \mathbf{X}'\mathbf{X}$. If the $\mathbf{C}$ matrix is ill-conditioned, then the LSE produces unduly large sampling variances. To overcome this problem, Hoerl and Kennard (1970a and b) suggested to use $\mathbf{C}(k) = \mathbf{X}'\mathbf{X} + k\mathbf{I}_p$, ($k \geq 0$) rather than $\mathbf{C}$ in the estimation of β .

Study of shrinkage estimators is pioneered by James and Stein (1961) followed by Mittelhammer (1984), Saleh and Chowdhury (1992) and Sen and Saleh (1985) among others.

In Section 2 we propose the ridge regression estimators namely the Unrestricted Ridge Regression Estimator, the Restricted Ridge Regression Estimator and the Shrinkage Ridge Regression Estimator. Section 3 contains the bias and the MSE expressions of these estimators. In Section 4 we present the performance of the ridge regression estimators. Concluding remarks are presented in Section 5.

2 The proposed estimators

We propose the following ridge-type estimators of β , when $\mathbf{H}\beta = \mathbf{h}$ is *a priori* suspected under the presence of multicollinearity in the observed data.

1. *Unrestricted Ridge Regression Estimator* (URRRE): the following ridge regression estimator has been considered by Hoerl and Kennard (1970a),

$$\tilde{\beta}_n(k) = \mathbf{W}\tilde{\beta}_n, \quad \mathbf{W} = [\mathbf{I}_p + k\mathbf{C}^{-1}]^{-1}, \quad (2)$$

where $\tilde{\beta}_n = \mathbf{C}^{-1}\mathbf{X}'\mathbf{Y}$ is the unrestricted least squares estimator of β and $k \geq 0$ is the biasing parameter.

2. *Restricted Ridge Regression Estimator* (RRRE):

$$\hat{\beta}_n(k) = \mathbf{W}\hat{\beta}_n, \quad (3)$$

where $\hat{\beta}_n = \tilde{\beta}_n - \mathbf{C}^{-1}\mathbf{H}'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1}(\mathbf{H}\tilde{\beta}_n - \mathbf{h})$ is the restricted least squares estimator. This estimator is considered by Sarkar (1992).

It is well known that the RRRE performs better than the URRRE, when the restrictions hold but as long as the parameter β moves away from the subspace $\mathbf{H}\beta = \mathbf{h}$, the RRRE becomes biased and completely inefficient while the performance of the URRRE remains stable. As a result, one may combine the URRRE and RRRE to obtain a better performance of the estimators in presence of the uncertain prior information $\mathbf{H}\beta = \mathbf{h}$, which leads to the shrinkage ridge regression estimator.

3. Shrinkage Ridge Regression Estimator (SRRE):

$$\hat{\beta}_n^S(k) = \mathbf{W}\hat{\beta}_n^S, \quad (4)$$

where $\hat{\beta}_n^S = \hat{\beta}_n + \{1 - u L_n^{-1}\}(\tilde{\beta}_n - \hat{\beta}_n)$ is the usual shrinkage least squares estimator (SLSE) as in Saleh and Han (1990), where

$$u = \frac{(q-2)(n-p)}{q(n-p+2)}, (u < 1) \quad q \geq 3.$$

Here, L_n is the test-statistic for testing the null-hypothesis $H_0: \mathbf{H}\beta = \mathbf{h}$ against $H_1: \mathbf{H}\beta \neq \mathbf{h}$ and is given by

$$L_n = \frac{(\mathbf{H}\tilde{\beta}_n - \mathbf{h})'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1}(\mathbf{H}\tilde{\beta}_n - \mathbf{h})}{q\tilde{\sigma}_e^2}, \quad (5)$$

where $\tilde{\sigma}_e^2 = (\mathbf{Y} - \mathbf{X}\tilde{\beta}_n)'(\mathbf{Y} - \mathbf{X}\tilde{\beta}_n)/(n-p)$ is the unrestricted unbiased estimator of σ_e^2 .

Following Box and Tiao (1973, p-117), we can say that the test-statistic L_n follows a central F -distribution with $(q, n-p)$ degrees of freedom under H_0 . However, when H_0 does not hold then Giles (1991) and Sutradhar (1988) have shown that the test statistic L_n follow a non-central F distribution with non centrality parameter

$$\delta = (\mathbf{H}\beta - \mathbf{h})'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1}(\mathbf{H}\beta - \mathbf{h})/\sigma^2.$$

For the power of the F -test, readers are referred to Zellner (1976) and Sutradhar (1988) among others.

The proposed estimators have been considered by the researchers when only the error distributions are normal. In this paper we consider all of the aforementioned estimators to study the performance of the $\hat{\beta}_n^S(k)$ over the estimators $\hat{\beta}_n^S$, $\tilde{\beta}_n(k)$ and $\hat{\beta}_n(k)$, when the regression disturbances follow a multivariate t distribution.

3 Biases and MSE expressions

The expressions for the bias and the MSE for the proposed estimators are routinely followed from Judge and Bock (1978, Chapter 10), Giles (1991) and Saleh and Chowdhury (1992).

For the URRRE, we have the following expressions:

$$Bias = E(\tilde{\beta}_n(k) - \beta) = B_1(k) = -k\mathbf{C}^{-1}(k)\beta, \quad (6)$$

$$MSE(\tilde{\beta}_n(k)) = M_1(k, \nu) = \frac{\nu}{\nu - 2} \sigma^2 tr[\mathbf{W}\mathbf{C}^{-1}\mathbf{W}'] + k^2 \beta' \mathbf{C}^{-2}(k) \beta. \quad (7)$$

For the RRRE, we have the following expressions:

$$B_2(k) = -\mathbf{W}\eta - k\mathbf{C}^{-1}(k)\beta, \quad (8)$$

where $\eta = \mathbf{C}^{-1}\mathbf{H}'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1}(\mathbf{H}\beta - \mathbf{h})$,

$$M_2(k, \nu) = \frac{\nu}{\nu - 2} \sigma^2 \left\{ tr(\mathbf{W}\mathbf{C}^{-1}\mathbf{W}') - tr(\mathbf{W}\mathbf{A}\mathbf{W}') \right\} \\ + 2k\eta' \mathbf{W}' \mathbf{C}^{-1}(k) \beta + \eta' \mathbf{W}' \mathbf{W} \eta + k^2 \beta' \mathbf{C}^{-2}(k) \beta, \quad (9)$$

where $\mathbf{A} = \mathbf{C}^{-1}\mathbf{H}'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1}\mathbf{H}\mathbf{C}^{-1}$.

Finally for the SRRE, we have the following expressions:

$$B_3(k, \delta) = -u\mathbf{W}\eta E^{(1)}(\chi_{q+2}^{-2}(\delta, \nu)) - k\mathbf{C}^{-1}(k)\beta, \quad (10)$$

$$M_3(k, \delta, \nu) = \frac{\nu}{\nu - 2} \sigma^2 \left[tr(\mathbf{W}\mathbf{C}^{-1}\mathbf{W}') - qu tr(\mathbf{W}\mathbf{A}\mathbf{W}') \right. \\ \left. \left\{ E^{(1)}(\chi_{q+2}^{-2}(\delta, \nu)) + \delta E^{(1)}(\chi_{q+4}^{-4}(\delta, \nu)) \right\} \right] + qu(q+2)(\eta' \mathbf{W}' \mathbf{W} \eta) \\ E^{(2)}(\chi_{q+4}^{-4}(\delta, \nu)) + 2k\eta' \mathbf{W}' \mathbf{C}^{-1}(k) \beta E^{(1)}(\chi_{q+2}^{-2}(\delta, \nu)) + k^2 \beta' \mathbf{C}^{-2}(k) \beta, \quad (11)$$

where

$$E^{(1)}(\chi_{q+2}^{-2}(\delta, \nu)) = \sum_{i=0}^{\infty} \frac{\Gamma(\frac{\nu}{2} + i - 1)(\delta/\nu)^i (1 + \frac{\delta}{\nu})^{-(\frac{\nu}{2} + i - 1)} (q + 2i)^{-1}}{\Gamma(i + 1) \Gamma(\frac{\nu}{2} - 1)}, \quad (12)$$

$$E^{(2)}(\chi_{q+4}^{-4}(\delta, \nu)) = \sum_{i=0}^{\infty} \frac{\Gamma(\frac{\nu}{2} + i)(\delta/\nu)^i (1 + \frac{\delta}{\nu})^{-(\frac{\nu}{2} + i)} (q + s + 2i - 2)^{-1} (q + s + 2i - 4)^{-1}}{\Gamma(i + 1) \Gamma(\frac{\nu}{2})}, \quad (13)$$

and $E^{(\cdot)}(n, \delta)$ is the expected value of a non-central chi-square variate with n degrees of freedom and non centrality parameter δ .

4 Performance of the estimators

In this Section we will compare the SRRE with the SLSE, RRRE and URRRE by the MSE criterion.

4.1 Comparison between $\hat{\beta}_n^S$ and $\hat{\beta}_n^S(k)$.

Here we will consider two cases:

Case 1. Under $H_0 : \mathbf{H}\beta = \mathbf{h}$ the MSE of SRRE is given by

$$M_3(k, 0, \nu) = \frac{\nu}{\nu - 2} \sigma^2 \text{tr}(\mathbf{W}\mathbf{C}^{**}\mathbf{W}') + k^2 \beta' \mathbf{C}^{-2}(k) \beta, \quad (14)$$

where $\mathbf{C}^{**} = \mathbf{C}^{-1} - qu\{E^{(1)}(\chi_{q+2}^{-2}(\delta, \nu)) + \delta E^{(1)}(\chi_{q+4}^{-4}(\delta, \nu))\}\mathbf{A}$.

Let $\mathbf{Q}$ be the orthogonal matrix with eigenvectors of $\mathbf{C}$ so that

$$\mathbf{Q}'\mathbf{C}\mathbf{Q} = \mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p).$$

Then, since $\mathbf{C}$ is symmetric we can write

$$\mathbf{W}\mathbf{A}\mathbf{W}' = \mathbf{Q}[\mathbf{\Lambda} + k\mathbf{I}_p]^{-1}\mathbf{\Lambda}\mathbf{A}^*\mathbf{\Lambda}[\mathbf{\Lambda} + k\mathbf{I}_p]^{-1}\mathbf{Q}', \quad (15)$$

where $\mathbf{Q}'\mathbf{A}\mathbf{Q} = \mathbf{A}^*$. Now without loss of generality we assume that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$, then we can write,

$$\text{tr}(\mathbf{W}\mathbf{C}^{-1}\mathbf{W}') = \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^2} \quad \text{and} \quad \text{tr}(\mathbf{W}\mathbf{A}\mathbf{W}') = \sum_{i=1}^p \frac{\lambda_i^2 a_{ii}^*}{(\lambda_i + k)^2}, \quad (16)$$

where $a_{ii}^* \geq 0$ is the i^{th} diagonal element of $\mathbf{A}^*$. Also,

$$\beta' \mathbf{C}^{-2}(k) \beta = \sum_{i=1}^p \frac{\alpha_i^2}{(\lambda_i + k)^2}, \quad \text{where} \quad \alpha = \mathbf{Q}'\beta. \quad (17)$$

$$\eta' \mathbf{W}' \mathbf{W} \eta = \sum_{i=1}^p \frac{\lambda_i^2 \eta_i^{*2}}{(\lambda_i + k)^2} \quad \text{and} \quad \eta' \mathbf{W}' \mathbf{C}^{-1}(k) \beta = \sum_{i=1}^p \frac{\alpha_i \lambda_i \eta_i^*}{(\lambda_i + k)^2}, \quad (18)$$

where $\eta^* = \eta' \mathbf{Q}$. Then using (16) to (18) in (14), we have

$$M_3(k, 0, \nu) = \frac{\nu}{\nu - 2} \sigma^2 \sum_{i=1}^p \frac{\lambda_i [1 - u \lambda_i a_{ii}^*]}{(\lambda_i + k)^2} + k^2 \sum_{i=1}^p \frac{\alpha_i^2}{(\lambda_i + k)^2}. \quad (19)$$

The first term is a continuous monotonically decreasing function and the second term is also a continuous monotonically increasing function of k .

Thus differentiating (19) with respect to k , we obtain a sufficient condition for $\partial M_3(k, 0, \nu)/\partial k$ to be negative is that $0 < k < k_1(\lambda, \nu)$, where $k_1(\lambda, \nu) = (\nu/(\nu-2))\sigma^2 / \max_{1 \leq i \leq p} [\alpha_i^2 \{1 - u\lambda_i a_{ii}^*\}^{-1}]$. Thus we can write the following theorem.

Theorem 1 *There exists a $k > 0$ in the region $0 < k < k_1(\lambda, \nu)$, such that $\hat{\beta}_n^S(k)$ has smaller MSE value than $\hat{\beta}_n^S$ under H_0 .*

Case 2. When the null hypothesis does not hold, i.e. $\delta \neq 0$, we consider the differences of the MSE's as,

$$\begin{aligned} M_3(0, \delta, \nu) - M_3(k, \delta, \nu) = \\ \frac{\nu}{\nu-2} \sigma^2 \text{tr}[\mathbf{C}^* - \mathbf{W}\mathbf{C}^{**}\mathbf{W}'] + \eta'[\mathbf{I}_p - \mathbf{W}'\mathbf{W}]\eta uq(q+2) \\ E^{(2)}(\chi_{q+4}^{-4}(\delta, \nu)) - 2k\eta'\mathbf{W}'\mathbf{C}^{-1}(k)\beta E^{(1)}(\chi_{q+2}^{-2}(\delta, \nu)) - k^2\beta'\mathbf{C}^{-2}(k)\beta, \end{aligned} \quad (20)$$

where $\mathbf{C}^* = \mathbf{C}^{-1} - u\mathbf{A}$.

The right hand side of (20) is ≥ 0 , according as

$$\sigma^{-2}\eta'(\mathbf{I}_p - \mathbf{W}'\mathbf{W})\eta \geq \frac{f_1(k, \delta, \nu)}{uq(q+2)E^{(2)}(\chi_{q+4}^{-4}(\delta, \nu))}, \quad (21)$$

where

$$\begin{aligned} f_1(k, \delta, \nu) = \frac{\nu}{\nu-2} \text{tr}[\mathbf{W}\mathbf{C}^{**}\mathbf{W}' - \mathbf{C}^{**}] \\ + 2k\sigma^{-2}\eta'\mathbf{W}'\mathbf{C}^{-1}(k)\beta E^{(1)}(\chi_{q+2}^{-2}(\delta, \nu)) + k^2\sigma^{-2}\beta'\mathbf{C}^{-2}(k)\beta. \end{aligned} \quad (22)$$

Thus using the result of Puri and Sen (1971, p. 122) it can easily be shown that $\hat{\beta}_n^S(k)$ is superior to $\hat{\beta}_n^S$, whenever, $\delta \geq \delta_1(k, \nu) > 0$, where

$$\delta_1(k, \nu) = \frac{f_1(k, \delta, \nu)}{\xi_1 uq(q+2)E^{(2)}(\chi_{q+4}^{-4}(\delta, \nu))} \quad (23)$$

and ξ_1 is the largest characteristic root of $[(\mathbf{I}_p - \mathbf{W}'\mathbf{W})\mathbf{C}^{-1}]$.

Now, we consider the variation of the MSE of $\hat{\beta}_n^S(k)$ as a function of k . Using the results (16) to (18), the MSE expression $M_3(k, \delta, \nu)$ can be written in terms of the eigen values as

$$M_3(k, \delta, \nu) =$$

$$\sum_{i=1}^p \frac{1}{(\lambda_i + k)^2} \left[\frac{\nu}{\nu - 2} \sigma^2 \lambda_i \left\{ 1 - qu \lambda_i a_{ii}^* \left(E^{(1)}(\chi_{q+2}^{-2}(\delta, \nu)) + \delta E^{(1)}(\chi_{q+4}^{-4}(\delta, \nu)) \right) \right\} \right. \\ \left. + k^2 \alpha_i^2 + 2k \alpha_i \lambda_i \eta_i^* E^{(1)}(\chi_{q+2}^{-2}(\delta, \nu)) + qu \lambda_i^2 \eta_i^{*2} (q+2) E^{(2)}(\chi_{q+4}^{-4}(\delta, \nu)) \right]. \quad (24)$$

Then differentiating (24) with respect to k , we obtain the sufficient condition on k so that the derivative of $M_3(k, \delta, \nu)$ is negative as given by the following theorem.

Theorem 2 Suppose $k_2(\delta, \nu)$ is defined by

$$k_2(\delta, \nu) = \frac{f_2(\delta, \nu)}{\max_{1 \leq i < p} [\lambda_i \alpha_i^2 - \alpha_i \lambda_i \eta_i^* E^{(1)}(\chi_{q+2}^{-2}(\delta, \nu))]}, \quad (25)$$

where

$$f_2(\delta, \nu) = \min_{1 \leq i \leq p} \left[\frac{\nu}{\nu - 2} \sigma^2 \lambda_i \left\{ 1 - qu \lambda_i a_{ii}^* \left(E^{(1)}(\chi_{q+2}^{-2}(\delta, \nu)) + \delta E^{(1)}(\chi_{q+4}^{-4}(\delta, \nu)) \right) \right\} \right. \\ \left. + uq(q+2) \lambda_i^2 \eta_i^{*2} E^{(2)}(\chi_{q+4}^{-4}(\delta, \nu)) - \alpha_i \eta_i^* \lambda_i^2 E^{(1)}(\chi_{q+2}^{-2}(\delta, \nu)) \right]. \quad (26)$$

Then the $\hat{\beta}_n^S(k)$ has the smaller MSE value than that of $\hat{\beta}_n^S$ for $0 < k < k_2(\delta, \nu)$.

Now combining this theorem and the discussion on $\delta_1(k, \nu)$, we have the following theorem.

Theorem 3 A sufficient condition for the superiority of $\hat{\beta}_n^S(k)$ over $\hat{\beta}_n^S$ is that whenever $(\delta, k) \in (\delta_1(k, \nu), \infty] \times (0, k_2(\delta, \nu)]$, where $\delta_1(k, \nu)$ and $k_2(\delta, \nu)$ are defined in (23) and (25) respectively.

4.2 Comparison between $\hat{\beta}_n^S(k)$ and $\hat{\beta}_n(k)$.

Under H_1 , the difference between MSE's is,

$$M_3(k, \delta, \nu) - M_2(k, \nu) = \frac{\nu}{\nu - 2} \sigma^2 \text{tr}(\mathbf{WAW}') \left\{ 1 - qu \{ E^{(1)}(\chi_{q+2}^{-2}(\delta, \nu)) \right. \\ \left. + \delta E^{(1)}(\chi_{q+4}^{-4}(\delta, \nu)) \} \right\} + \eta \mathbf{W}' \mathbf{W} \eta \left\{ qu(q+2) E^{(2)}(\chi_{q+4}^{-4}(\delta, \nu)) - 1 \right\} \\ + 2k \eta \mathbf{W}' \mathbf{C}^{-1}(k) \beta \left\{ E^{(1)}(\chi_{q+2}^{-2}(\delta, \nu)) - 1 \right\}. \quad (27)$$

Again using the result of Puri and Sen (1971, p. 122) the expression (27) is ≥ 0 whenever

$$\delta_2(k, \nu) \leq \frac{f_3(k, \delta, \nu)}{\xi_2 \left[1 - qu(q+2)E^{(2)}(\chi_{q+4}^{-4}(\delta, \nu)) \right]}, \quad (28)$$

where ξ_2 is the largest characteristic root of $[\mathbf{W}'\mathbf{W}\mathbf{C}^{-1}]$ and

$$\begin{aligned} f_3(k, \delta, \nu) = & \frac{\nu}{\nu-2} \text{tr}(\mathbf{W}\mathbf{A}\mathbf{W}') \left[1 - qu \{ E^{(1)}(\chi_{q+2}^{-2}(\delta, \nu)) + \delta E^{(1)}(\chi_{q+4}^{-4}(\delta, \nu)) \} \right] \\ & - 2k\sigma^{-2} \boldsymbol{\eta}' \mathbf{W}' \mathbf{C}^{-1}(k) \boldsymbol{\beta} \left(1 - E^{(1)}(\chi_{q+2}^{-2}(\delta, \nu)) \right). \end{aligned} \quad (29)$$

Thus $\hat{\beta}_n(k)$ is superior to $\hat{\beta}_n^S(k)$ whenever, $\delta \in (0, \delta_2(k, \nu))$, where $\delta_2(k, \nu)$ is given by (28).

Again using the results (16) to (18) in (27), the eigen values yields

$$\begin{aligned} M_3(k, \delta, \nu) - M_2(k, \nu) = & \sum_{i=1}^p \frac{1}{(\lambda_i + k)^2} \left[\frac{\nu}{\nu-2} \sigma^2 \lambda_i^2 a_{ii}^* \left\{ 1 - qu \left(E^{(1)}(\chi_{q+2}^{-2}(\delta, \nu)) + \delta E^{(1)}(\chi_{q+4}^{-4}(\delta, \nu)) \right) \right\} \right. \\ & \left. - \eta_i^{*2} \lambda_i^2 \left\{ 1 - qu(q+2)E^{(2)}(\chi_{q+4}^{-4}(\delta, \nu)) \right\} - 2k\alpha_i \lambda_i \eta_{i*} \left\{ 1 - E^{(1)}(\chi_{q+2}^{-2}(\delta, \nu)) \right\} \right]. \end{aligned} \quad (30)$$

Thus from the above expression, we can define the following theorem.

Theorem 4 *A sufficient condition for SRRE to have MSE values less than or equal to that of RRRE is that there exists a value of k such that $k > k_3(\delta, \nu)$, where $k_3(\delta, \nu)$ is given by*

$$k_3(\delta, \nu) = \frac{f_4(\delta, \nu)}{\min_{1 \leq i \leq p} \left\{ 2\alpha_i \lambda_i \eta_i^* \left(1 - E^{(1)}(\chi_{q+2}^{-2}(\delta, \nu)) \right) \right\}} \quad (31)$$

and

$$\begin{aligned} f_4(\delta, \nu) = & \max_{1 \leq i \leq p} \left[\frac{\nu}{\nu-2} \sigma^2 \lambda_i a_{ii}^* \left\{ 1 - qu \left(E^{(1)}(\chi_{q+2}^{-2}(\delta, \nu)) \right. \right. \right. \\ & \left. \left. + \delta E^{(1)}(\chi_{q+4}^{-4}(\delta, \nu)) \right) \right\} - \lambda_i^2 \eta_i^{*2} \left(1 - qu(q+2)E^{(2)}(\chi_{q+4}^{-4}(\delta, \nu)) \right) \right]. \end{aligned} \quad (32)$$

Now combining the above theorem and the discussion on $\delta_2(k, \nu)$, we have the following theorem.

Theorem 5 *A sufficient condition for SRRE to have MSE values less than or equal to that of RRRE is that*

$$(\delta, k) \in [\delta_2(k, \nu), \infty) \times (k_3(\lambda, \nu), \infty),$$

where $k_3(\lambda, \nu)$ is defined in (31).

4.3 Comparison between $\hat{\beta}_n^S(k)$ and $\tilde{\beta}_n(k)$.

Under H_1 , the difference between the MSE's is

$$\begin{aligned} M_3(k, \delta, \nu) - M_1(k, \nu) = \\ qu\eta' \mathbf{W}' \mathbf{W} \eta (q+2) E^{(2)}(\chi_{q+4}^{-4}(\delta, \nu)) - \frac{\nu}{\nu-2} \sigma^2 qu \text{tr}(\mathbf{W} \mathbf{A} \mathbf{W}') \{E^{(1)}(\chi_{q+2}^{-2}(\delta, \nu)) \\ + \delta E^{(1)}(\chi_{q+4}^{-4}(\delta, \nu))\} + 2k\eta' \mathbf{W}' \mathbf{C}^{-1}(k) \beta E^{(1)}(\chi_{q+2}^{-2}(\delta, \nu)). \end{aligned} \quad (33)$$

The expression (33) is ≤ 0 , whenever

$$\delta_3(k, \nu) \leq \frac{f_5(k, \delta, \nu)}{\xi_2[qu(q+2)E^{(2)}(\chi_{q+4}^{-4}(\delta, \nu))]}, \quad (34)$$

where

$$\begin{aligned} f_5(k, \delta, \nu) = \frac{\nu}{\nu-2} qu \text{tr}(\mathbf{W} \mathbf{A} \mathbf{W}') [\{E^{(1)}(\chi_{q+2}^{-2}(\delta, \nu)) + \delta E^{(1)}(\chi_{q+4}^{-4}(\delta, \nu))\} \\ - 2k\sigma^{-2} \eta' \mathbf{W}' \mathbf{C}^{-1}(k) \beta E^{(1)}(\chi_{q+2}^{-2}(\delta, \nu))]. \end{aligned} \quad (35)$$

Therefore, $\hat{\beta}_n^S(k)$ is superior to $\tilde{\beta}_n(k)$ whenever, $\delta \in (0, \delta_3(k, \nu))$.

Similarly, rewriting equation (33) in terms of the eigen values we obtain

$$\begin{aligned} M_3(k, \delta, \nu) - M_1(k, \nu) = \\ - \sum_{i=1}^p \frac{1}{(\lambda_i + k)^2} \left[\frac{\nu}{\nu-2} \sigma^2 \lambda_i^2 a_{ii}^* qu \left(E^{(1)}(\chi_{q+2}^{-2}(\delta, \nu)) + \delta E^{(1)}(\chi_{q+4}^{-4}(\delta, \nu)) \right) \right. \\ \left. - \eta_i^{*2} \lambda_i^2 qu(q+2) E^{(2)}(\chi_{q+4}^{-4}(\delta, \nu)) - 2k\alpha_i \lambda_i \eta_{i*} E^{(1)}(\chi_{q+2}^{-2}(\delta, \nu)) \right]. \end{aligned} \quad (36)$$

Thus from the above expression, we can define the following theorem.

Theorem 6 *A sufficient condition for the SRRE to have MSE values less than or equal to that of the URRRE is that there exists a value of k such that $k > k_4(\delta, \nu)$, where $k_4(\delta, \nu)$ is given by*

$$k_4(\delta, \nu) = \frac{f_6(\delta, \nu)}{\min_{1 \leq i \leq p} \left\{ 2\alpha_i \lambda_i \eta_i^* \left(E^{(1)}(\chi_{q+2}^{-2}(\delta, \nu)) \right) \right\}} \quad (37)$$

and

$$f_6(\delta, \nu) = \max_{1 \leq i \leq p} \left[\frac{\nu}{\nu - 2} \sigma^2 \lambda_i^2 a_{ii}^* q u \left(E^{(1)}(\chi_{q+2}^{-2}(\delta, \nu)) + \delta E^{(1)}(\chi_{q+4}^{-4}(\delta, \nu)) \right) - \lambda_i^2 \eta_i^{*2} q u (q + 2) E^{(2)}(\chi_{q+4}^{-4}(\delta, \nu)) \right]. \quad (38)$$

Now combining the above theorem and discussion on $\delta_3(k, \nu)$, we have the following theorem.

Theorem 7 *A sufficient condition for the SRRE to have MSE values less than or equal to that of the URRRE is that*

$$(\delta, k) \in [0, \delta_3(k, \nu)) \times (k_4(\lambda, \nu), \infty),$$

where $\delta_3(k, \nu)$ and $k_4(\lambda, \nu)$ are defined in (34) and (37) respectively.

For the estimated value of k readers are referred to Hoerl and Kennard (1970b), Hoerl, Kennard and Baldwin (1975) and Hemmerle (1975) among others.

5 Concluding remarks

This paper combined the shrinkage estimator with the ridge regression approach to estimate the regression parameter in a linear regression model, when we suspect that the parameter space may belong to a subspace and the error distribution is multivariate t . The MSE criterion has effectively determined the sufficient conditions for the superiority of the SRRE over LSE, RRRE and URRRE on some subset of $R_1^+ \times R_1^+$ space generally using the non-null case. We find that there exist values of (δ, k) such that the proposed estimator (SRRE) dominates the other estimators. Under H_0 , we could find the exact dominance picture of the estimators. However, under H_1 , there is no unique solution for the superiority of one estimator over the others. We may expect that the analysis has been done in this paper will be useful even when the underlying distributions are normal or Cauchy.

Acknowledgments: The author would like to express his gratitude to the Government of Canada for the Canadian Commonwealth Scholarship at The University of Western Ontario. He wishes to thank Professor M. A. Matin for his encouragement and Dr. Hans Nyquist for his valuable comments and suggestions regarding the preparation of this paper. He is very grateful to his brothers, Mr. Abdur Rakib Bhuiyan and Mr. B. M. Golam Kabir for their inspiration and financial support throughout his studies in Bangladesh.

Résumé: Cet article propose de combiner l'estimateur "shrinkage" avec la "ridge régression" dans les cas où les coefficients de la régression sont contenus dans un certain sous-espace et lorsque les erreurs suivent une distribution t multivariée. Nous proposons trois estimateurs: le "Unrestricted Ridge Regression Estimator", le "Restricted Ridge Regression Estimator" et finalement le "Shrinkage Ridge Regression Estimator". Le biais et le "Mean Square Errors" (MSE) des estimateurs proposés sont calculés sous les hypothèses nulles et alternatives. Le critère du MSE détermine les régions d'optimalité de ces estimateurs. Les conditions sous lesquelles l'estimateur proposé est supérieur aux estimateurs traditionnels sont également discutées.

Mots clés: Carré des écarts à la moyenne, chi-deux non central, distribution t multivariée, dominance, estimateur shrinkage, régression ridge, supériorité.

References

- [1] Box, J. E. P. and Tiao, G. C. (1973), *Bayesian Inference in Statistical Analysis*, Adison Wesley, London.
- [2] Giles, A. J. (1991), "Pretesting for Linear Restrictions in a Regression Model with Spherically Symmetric Distributions", *J. Econometrics*, **50**, 377-398.
- [3] Hemmerle, W.J. (1975), "An Explicit Solution for Generalized Ridge Regression", *Technometrics*, **17**, 309-314.
- [4] Hoerl, A.E. and Kennard, R.W. (1970a), "Ridge Regression: Biased Estimation for Non-orthogonal Problems", *Technometrics*, **12**, 55-67.
- [5] Hoerl, A.E. and Kennard, R.W. (1970b), "Ridge Regression: Application to Non-orthogonal Problems", *Technometrics*, **12**, 69-82.
- [6] Hoerl, A.E., Kennard, R.W. and Baldwin, K.F. (1975), "Ridge Regression: Some Simulations", *Comm. Statist. A*, **4**, 105-123.
- [7] James, W. and Stein, C. (1961), "Estimation with Quadratic Loss", *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 361-379.
- [8] Judge, G.G. and Bock, M.E. (1978), *The Statistical Implications of Pre-test and Stein-rule Estimators in Econometrics*, North-Holland Publishing Company, Amsterdam.
- [9] Mittelhammer, R. C. (1984), "Restricted Least Squares, Pretest, OLS and Stein Rule Estimators", *Journal of Econometrics*, **25**, 151-164.
- [10] Puri, M. L. and Sen, P. K. (1971), *Nonparametric Methods in Multivariate Analysis*, John Wiley & Sons, New York.

- [11] Saleh, A.K.M. d.E. and Chowdhury, J.A. (1992), *Shrinkage Estimation of Regression Parameters when the Errors are Non-normal*, Unpublished manuscript, Department of Mathematics and Statistics, Carleton University, Ottawa, Canada.
- [12] Saleh, A.K.Md.E. and Han, C-P. (1990), "Shrinkage Estimation in Regression Analysis", *Estadistica*, **42**, 40-63.
- [13] Sarkar, N. (1992), "A new Estimator Combining the Ridge Regression and the Restricted Least Squares Method of Estimation", *Comm. Statist. A*, **21**, 1987-2000.
- [14] Sen, P. K. and Saleh, A. K. Md. E. (1985), "On Some Shrinkage Estimators of Multivariate Location", *The Annals of Statistics*, **13**, 272-281.
- [15] Sutradhar B. C. (1988), "Testing Linear Hypothesis with t-error Variable", *Sankhya B*, **50**, 175-180.
- [16] Zellner, A (1976), "Bayesian and Non-Bayesian Analysis of the Regression Model with Multivariate t -error Terms", *J. Am. Statist. Assoc.*, **71**, 400-405.

From the Editor's Notebook

Les nombres vivants¹

1 Introduction

Les mathématiciens ont introduit au cours de leur histoire plusieurs ensembles de nombres avec successivement les nombres naturels, les nombres entiers relatifs, les nombres rationnels et les nombres réels. A chaque fois, le nouvel ensemble est défini à partir du précédent, et son introduction permet de définir des concepts mathématiques nouveaux, ainsi que de résoudre des problèmes irrésolubles jusque là. Nous proposons ici un rapide résumé de ces ensembles fondamentaux avant d'introduire à notre tour un nouvel ensemble de nombres, construit à partir des nombres réels, celui des nombres vivants.

1.1 Les nombres naturels

$$\mathbf{N} = \{0, 1, 2, 3, 4, \dots\}$$

Les nombres naturels nous sont donnés par Dieu disait Kronecker. Ils sont ce qui différencie les hommes des animaux disait Platon. Ils sont en tout les cas la matière première des mathématiques. A partir des nombres naturels, on peut définir l'addition et la multiplication, opérations qui sont à la base de tout calcul arithmétique et qui existent depuis la nuit des temps. Notons que le mot "calcul" vient du latin "calculi", qui signifie petit caillou, probablement parce que les bergers de Sumer utilisaient de petites pierres pour compter leurs moutons. Notons encore que ce sont les Indiens qui introduisirent le zéro dans cet ensemble.

1.2 Les nombres entiers relatifs

$$\mathbf{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\} = \{\pm n : n \in \mathbf{N}\}$$

Par définition, les entiers relatifs sont des nombres naturels accompagnés

¹Cette note est le résumé d'une conférence donnée par Yadolah Dodge le 6 novembre 1995 dans le cadre du mois de la science organisé par la SENS à Neuchâtel.

d'un signe "+" (généralement omis) ou d'un signe "-". Ils ont permis de définir la soustraction, autre opération élémentaire du calcul arithmétique.

1.3 Les nombres rationnels

$$\mathbf{Q} = \left\{ \frac{p}{q} : p, q \in \mathbf{Z} \right\}$$

Par définition, les nombres rationnels sont des rapports entre deux entiers relatifs. Ils ont permis de définir la division, quatrième opération élémentaire du calcul arithmétique.

1.4 Les nombres réels

On a longtemps cru que tout ce que l'on mesure sur Terre pouvait être exprimé par un nombre rationnel. Or ceci n'est pas le cas, et de loin! On s'est aperçu en effet que $\mathbf{Q}$ n'est pas complet, et on l'a complété en introduisant les nombres réels:

$$\mathbf{R} = \left\{ \lim_{n \rightarrow \infty} q_n : (q_n) \subset \mathbf{Q} \right\}$$

Un nombre réel est par définition la limite d'une suite de Cauchy (q_n) de nombres rationnels (une suite de Cauchy étant une suite de nombres où au bout d'un certain rang la distance entre deux nombres devient arbitrairement petite). Parmi les nombres réels, on distingue les nombres algébriques (rationnels ou irrationnels) et les nombres transcendants.

1.4.1 Les nombres algébriques

Les nombres algébriques sont par définition des nombres qui sont solution d'une équation polynomiale à coefficients dans $\mathbf{Q}$. Par exemple, le polynôme $x^2 - 2$ s'annule pour les valeurs de $x = \pm\sqrt{2}$. Ces deux nombres $+\sqrt{2}$ et $-\sqrt{2}$ sont donc des nombres algébriques. Les nombres algébriques nous permettent de définir la racine carrée (ainsi que d'autres racines). D'autre part, géométriquement parlant, ils nous permettent par exemple d'exprimer par un nombre la longueur de la diagonale d'un carré.

1.4.2 Les nombres transcendants

Les nombres transcendants sont des nombres réels non algébriques. Les plus connus sont e et π . Par définition, le premier est la limite de la suite (q_n) constituées des nombres rationnels $q_n = (1 + 1/n)^n$ $n = 1, 2, \dots$, alors que le second est le rapport entre la circonférence d'un cercle et son diamètre. Ces nombres s'expriment de la manière suivante:

$$\begin{aligned} e &= 2.718281828459... \\ \pi &= 3.141592653589... \end{aligned}$$

Remarquons que ces suites de décimales ne sont pas connues entièrement pour la simple raison qu'elles sont infinies et qu'elles ne sont pas périodiques, contrairement à ce que l'on a pour les nombres rationnels.

1.5 Les nombres vivants

Qu'existe-t-il après les nombres réels? Les mathématiciens ont introduit les nombres complexes qui sont des couples de nombres réels et qui ont permis de définir des solutions pour n'importe quelle équation polynomiale; ils ont encore introduit les quaternions qui sont des couples de couples de nombres réels. Dans cet article, nous proposons une autre extension possible des nombres réels: les nombres vivants.

2 Définition des nombres vivants

Dans Freedman, Pisani and Purves (1978, p.88), on trouve la citation suivante: "If the same thing is measured several times, in an ideal world the same result would be obtained each time. In practice, the results are found to differ." Ceci constitue le point de vue essentiel du statisticien. Il fait une distinction claire et nette entre un nombre, qui ne contient pas d'information, et une donnée (du latin "datum" qui veut dire "fait") qui en contient. Il considère les données comme étant contaminées par des erreurs de tout genre (erreurs de mesures, erreurs de transmission, changement des conditions initiales,...). Son schéma de pensée est le suivant:

$$\text{donnée observée} = \text{valeur exacte} + \text{erreur}$$

Nous le réécrivons comme suit:

$$\text{nombre vivant} = \text{nombre réel} + \text{nombre aléatoire}$$

Les nombres vivants que nous définissons ainsi ont donc un rapport direct avec la statistique. Ce ne sont pas de simples nombres réels abstraits mais des nombres concrets qui représentent quelque chose. Ils sont vivants de par leur composante aléatoire.

3 Définition de l'aléatoire

Pour définir les nombres vivants, il nous faut donc préciser ce qu'on entend par nombre aléatoire. Qu'est-ce qu'un nombre aléatoire ou une suite de nombres aléatoires? La question n'est pas facile. Dans Gardner (1977, p.163), on lit: "Is there any objective, mathematical way to define a completely disordered series? Apparently there is not."

Knuth (1981) dans son chapitre intitulé “What is randomness” ne donne pas moins de six définitions différentes qui nous aident à cerner ce concept. La plus simple (mais pas totalement satisfaisante) est la suivante: une suite infinie de nombres entre 0 et 9 est dite aléatoire si chaque chiffre 0, 1, ..., 9 apparaît dans la suite avec une fréquence de $1/10$, chaque bloc de deux chiffres 00, 01, 02, ..., 98, 99 apparaît dans la suite avec une fréquence de $1/100$, et ainsi de suite chaque bloc de n chiffres 00..00, 00..01, ..., 99..98, 99..99 apparaissant dans la suite avec une fréquence de $1/10^n$.

Il est d’autant plus difficile de définir ce qu’est une suite de nombres aléatoires, qu’il est peut-être impossible d’en exhiber ne serait-ce qu’une seule! En comparaison, les suites de nombres non aléatoires sont légion. En voici quelques unes:

- 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...
- 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, ...
- 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, ...

Le problème de générer des suites de nombres aléatoires est d’ailleurs toujours d’actualité en statistique. A une certaine époque, ces suites étaient obtenues manuellement, à l’aide de roulettes ou d’urnes contenant des boules numérotées. Le problème de ces méthodes est qu’elles sont fastidieuses et surtout qu’elles ne sont pas reproductibles (mais n’est ce pas justement la définition de l’aléatoire que de n’être pas reproductible?). On a ainsi développé des générateurs de nombres dits “pseudo-aléatoires” qui sont des algorithmes qui utilisent des fonctions mathématiques, et dont les itérations successives produisent les suites de nombres désirées. Ces méthodes sont évidemment reproductibles: il suffit d’appliquer l’algorithme une nouvelle fois en partant du même point de départ.

Il se trouve hélas, que la plupart de ces générateurs ne sont pas satisfaisants, c’est-à-dire que les suites qu’ils produisent ont des structures (pas toujours perceptibles au premier regard) qui trahissent le manque d’aléatoire. Toute une série de tests statistiques ont été développés pour déceler si un générateur est bon ou non. Et encore, si un générateur passe avec succès tous ces tests, il n’est pas exclu (et même fort probable) qu’un jour, on découvre un test auquel il échoue! Autrement dit, ces tests peuvent prouver qu’un générateur est mauvais, mais ils ne pourront jamais prouver qu’il est vraiment bon! Sombres perspectives en vérité...

Alors, l’aléatoire ne serait-il qu’un idéal non concrétisable? Peut-être pas! La solution à notre problème pourrait bien se trouver du côté des décimales de π .

4 L'histoire de π et de ses décimales

Le nombre π , dit-on, contient l'histoire de l'humanité. Ce nombre a de tout temps été d'un intérêt particulier pour les mathématiciens. Les Babyloniens, les Egyptiens, les Chinois, les Grecs, les Indiens et les Persans ont tous tenté de le calculer avec un maximum de précision. De 2 décimales à l'époque d'Archimède (3ème siècle avant notre ère) à plus de 6 milliards ($3 \cdot 2^{31}$) de nos jours, record réalisé par le japonais Kanada en septembre 1995, en passant par les 707 décimales (dont 180 fausses!) calculées à la main pendant 29 ans par William Shanks au 19ème siècle, le nombre de décimales exactes connues a augmenté de siècle en siècle.

Ce qui intriguait les mathématiciens et les géomètres en particulier, c'est l'irrationalité de ce nombre. On aurait pu croire que ce rapport géométrique mythique, qui fait intervenir avec le cercle "la figure la plus parfaite de l'univers", soit un nombre au moins rationnel. On peut voir en effet dans le désir effréné de calculer ces décimales (on parle de la course aux décimales de π) un désir de découvrir quelquepart dans cette suite infinie, une certaine régularité, un cycle qui reviendrait périodiquement et qui indiquerait que π est le rapport de deux nombres finis. Il y eut çà et là quelques espoirs de voir apparaître une structure. Par exemple, de la 762ème à la 767ème décimale de π , on ne trouve que des 9. Hélas, la 768ème n'est qu'un 8, ce qui ne fit qu'accroître le mystère de ce nombre. Certains ont même fait des paris. Le docteur Matrix, un fameux numérologue, a prédit en 1970 que la millionième décimale de π serait un 5. Hélas pour lui, il s'est trompé d'une décimale. La 999'999ème décimale est bien un 5, mais la millionième décimale est un 1.

Les grands mathématiciens occidentaux, parmi eux Viète, Newton, Euler, Legendre et Gauss, se sont longuement interrogés sur la nature aléatoire de ce nombre (étant donné une décimale de π , il n'y a aucun moyen de deviner la suivante). Lambert a prouvé en 1771 que π est irrationnel et Lindemann en 1882 que π est transcendant. Depuis lors, tous les tests statistiques dont nous avons parlé ont été appliqués à la suite de ses décimales mais n'ont jamais pu infirmer son caractère aléatoire. On a bien cru un moment donné qu'il y avait une déficience en 7 mais c'était dû aux erreurs de calcul de William Shanks. C'est donc avec grands regrets que les mathématiciens ont dû abandonner tout espoir de découvrir quelque régularité dans π . "On a soumis jusqu'ici la suite des décimales de π à tous les tests statistiques qui pouvaient en montrer le caractère aléatoire. Ceci est un peu déconcertant pour ceux qui pensent qu'il devrait exister un rapport un peu moins irrégulier entre le diamètre et le périmètre d'une courbe aussi belle que le cercle mais la plupart des mathématiciens

pensent qu'on ne trouvera jamais la moindre régularité ni aucun ordre dans le développement décimal de π ." (Gardner, 1966, p.87).

Mais ne dit-on pas que le malheur des uns fait le bonheur des autres? Cette absence totale de structure qui chagrine le mathématicien pourrait bien faire le bonheur du statisticien, toujours à la recherche de l'aléatoire. Ainsi les décimales de π pourraient être utilisées pour générer des nombres aléatoires (voir Dodge, 1996) et comme exemple de suite de nombres aléatoires que nous avons tant de peine à trouver, nous pourrions donner:

- 1, 4, 1, 5, 9, 2, 6, 5, 3, 5, 8, 9, 7, 9, 3, 2, 3, 8, 4, 6, ...

5 Conclusion

En utilisant les décimales de π pour définir l'aléatoire, il nous est possible de donner une définition originale des nombres vivants. Notre nouvel ensemble peut s'exprimer de la manière suivante:

$$V = \{r + a : r \in \mathbf{R}, a = \text{une fonction des décimales de } \pi\}$$

Les nombres vivants nous permettent entre autre de résoudre un problème apparemment sans solution: celui constitué par un système d'équations linéairement indépendantes avec plus d'équations que d'inconnues (et ceci à l'aide d'une régression linéaire qui utilise justement le concept de nombre vivant).

Notons enfin que si on avait défini les nombres entiers relatifs à partir des nombres naturels, les nombres rationnels à partir des nombres entiers relatifs, et les nombres réels à partir des nombres rationnels, on a de même défini les nombres vivants à partir des nombres réels (π étant un nombre réel). L'ensemble des nombres vivants est bien ainsi un descendant légitime des ensembles de nombres fondamentaux introduits au paragraphe 1.

Références

- [1] Dodge, Y. (1996), "The Natural Random Number Generator", *IMS Bulletin*, **25**, 4-5.
- [2] Freedman, D., Pisani, R. and Purves, R. (1978), *Statistics*, Norton, New York.
- [3] Gardner, M. (1966), *New Mathematical Diversions from Scientific American*, Simon and Schuster.
- [4] Gardner, M. (1977), "Random Numbers" in *Mathematical Carnival*, Vintage Books Edition.
- [5] Knuth, D. (1981), *The Art of Computer Programming*, vol.2, 2nd Edition, Addison-Wesley, Reading, MA.

C. Radhakrishna Rao

Growing Strong at the Age of Seventy Five

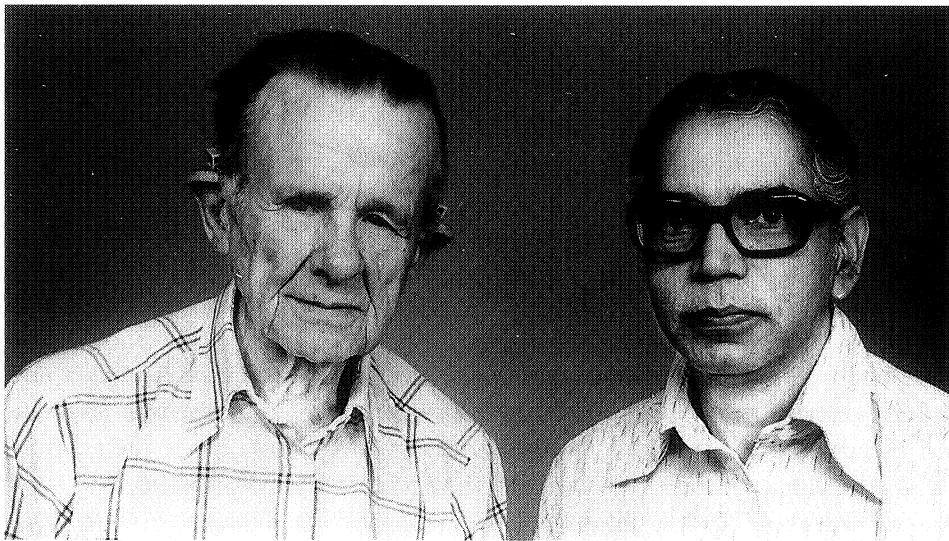


Calyampudi Radhakrishna Rao was born on September 10, 1920, in Hadagali, Karnataka State, India. He received an M.A. with a first rank in Mathematics from Andhra University, India, in 1940, an M.A. in statistics from Calcutta University in 1943. In 1948 C.R. Rao received his Ph.D. degree from Cambridge University under Sir R.A. Fisher, with Wishart as the examiner. Among his many accomplishments, he is a Fellow of the Royal Society of London, American Academy of Arts and Science, American Association for the Advancement of Science, Indian National Science Academy and a Founder Fellow of the Third World Academy of Science.

In 1945, C.R. Rao published "Information and accuracy attainable in the estimation of statistical parameters" in *Calcutta Mathematical Bulletin* (Vol. 37, 81-91). Fifty years after he informs *Student* that:

"Although the paper was completed in 1943, due to delays in publication and suspension of many journals including *Sankhyā* during the war period, the results could not be published earlier. I sent a reprint of my paper to Professor J. Neyman who congratulated me and named the information lower bound to variance as Cramér-Rao inequality."

He adds: “There is an amusing anecdote concerning the C.R. inequality. On one of my trips to Tehran from Delhi, my suit case was lost. The airline said that they would trace it and deliver it to my hotel room. Next day, I received a call from a hostess of the airline: “Good news, Mr. Cramer Rao, we found your bag”. When I said I am C.R. Rao, she said that she had learnt of my full name from my Iranian friend who kept on asking whether Cramer Rao’s suit case was found. Apparently, the Iranian student who was asked to look after me has read about the Cramér Rao inequality and thought that the initials C.R. in my name stood for Cramer.”



with Harold Cramér in 1978

C.R. Rao, the holder of the Eberly Family Chair in Statistics, Director of the Center for Multivariate Analysis at the Pennsylvania State University and Adjunct Professor, University of Pittsburgh was elected in 1995 as a member of the most prestigious scientific body in the USA, the National Academy of Sciences. It is a rare honour given to a scientist whose contributions have had a great impact on the expansion of human knowledge.

Bradley Efron writes in a recent article entitled “the Statistical Century”, published in the Royal Statistical Society News, Vol. 22, January 1995:

“Karl Pearson’s famous chi-square paper appeared in the spring of 1900, an auspicious beginning to a wonderful century for the field of statistics. The first half of the century was the golden age of statistical theory, during which our field grew from ad hoc origins, similar to the current state of computer science, into a firmly grounded mathematical science. Men of

the intellectual calibre of Fisher, Neyman, Pearson, Kolmogorov, Hotelling, Wald, Cramér and Rao were needed to bring statistical theory to maturity."



with Sir R.A. Fisher in 1956

During the international conference on Statistical Data Analysis and Inference held in his honour at Neuchâtel in 1989, C.R. Rao was awarded an Honorary Doctorate by the University of Neuchâtel. Since then, he received Honorary Doctorate degrees from the Indian Statistical Institute, India, Colorado State University, USA, University of Poznan, Poland, University of Hyderabad, India, Slovak Academy of Sciences, Slovakia, University of Barcelona, Spain and University of Guelph, Canada which is his seventeenth honorary degree.

Some of the statistical techniques that bear his name are the well known Cramér-Rao inequality, Rao-Blackwell Theorem, Fisher-Rao Theorem, Rao's Score Statistic and U-test, Rao's orthogonal arrays and Generalized Inverse of Matrices. He is the author or co-author of 12 books and more than 250 research papers. His book on *Statistics and Truth: Putting Chance to Work* dealing with the logical foundations and applications of statistics, first published in English, is now translated into Japanese, Polish, Spanish and German Languages.

Dr. Rao is the recipient of numerous medals including the Guy medal in Silver of the Royal Statistical Society, U.K, J.C. Bose Gold medal of the Bose Institute, India, M.N. Saha Medal of the Indian National Science Academy and S.S Wilks Medal of the American Statistical Association.



with Blackwell in 1976

Dr. Rao is still active in research and dissemination of statistical knowledge. He is on the editorial boards of numerous journals and is the chief editor of the series, Handbook of Statistics. His current research interests are multivariate shape analysis, imaging, signal detection and graphical representation of categorical data.



with the L_1 -norm distance from Mrs. Rao in 1987

C.R. RAO, Sc. D., F.R.S.

Eberly Professor of Statistics
Director, Center for Multivariate Analysis
The Pennsylvania State University

Statistics Department
326 Classroom Building
University Park, PA 16802-2111

Adjunct Professor
University of Pittsburgh

December 20, 1995
Fax: 41-38-211730

Professor Y. Dodge
Groupe de Statistique
Pierre-à-Mazal 7
CH-2000 Neuchâtel
Switzerland

Dear Professor Dodge,

I am glad that you are editing the new journal *Student*. Statistics, unlike other disciplines, has no subject matter of its own. It is a method or reasoning used in solving problems in other disciplines. What constitutes research in statistics as a separate discipline is the development of methodology for solving problems in new situations. It is important for graduate students to understand the exact role of statistics in solving problems of real life and how research in statistics is motivated by particular applications. A journal like *Student* with papers emphasizing statistical applications rather than the hard-core mathematics, sometimes needed to develop a statistical method, will be extremely valuable in the education and training of statisticians.

It will be educative for the students to know what motivated the pioneers in statistics to do what they did and what kind of mathematical training is needed to do research in statistics. Keeping in view the aim of your journal, I thought I should contribute an article describing what I did in an early stage of my career as a research worker in statistics. I was mainly motivated by what R.A. Fisher described as the two fundamental problems of statistics, specification and estimation and provided a general solution, especially to the problem of estimation. I tried to resolve some issues raised by Fisher as explained in my paper on seven inequalities. Fortunately, the solutions did not involve complicated mathematics. I hope the paper would be of some interest to students and teachers of statistics.

Yours sincerely,

C. R. Rao
C.R. Rao

P.S. I am sending the letter by e-mail as well as Fax.

Tel: (814)865-3194 (office)
(814)234-6209 (residence)

E-mail: CRR1@psuvm.psu.edu

Fax: (814)863-7114 (office)
(814)234-0372 (residence)

SEVEN INEQUALITIES IN STATISTICAL ESTIMATION THEORY

by

C. Radhakrishna Rao

Eberly Professor of Statistics

Pennsylvania State University

Abstract: When I was learning statistics as a student, the only literature available on statistical theory and practice was Fisher's book Statistical Methods for Research Workers and his numerous papers on estimation and small sample exact distributions of some test statistics. Fisher was laying the logical foundations of statistics as a separate discipline. He introduced numerous ideas and analytical tools which we still use. Fisher was criticized for lack of rigor in mathematical rendering of his ideas and in proving his results, and for making general statements without specifying the underlying assumptions (Savage, 1976). My understanding of statistics is based on reading each of Fisher's results n times depending on the complexity of the problem involved. I am still reading some of his papers. In this process, I developed some inequalities to interpret and clarify Fisher's ideas. I shall describe them to you with some historical notes.

KEY WORDS: Cramér-Rao inequality, Fisher consistency, Fisher's theorem, Fisher-Rao theorem, Inequality on expected harmonic mean, Matrix inequality, Rao-Blackwell theorem, Second order efficiency.

1

... and he still writes for statistics

A General Practitioner's Obstetric Diary

Stephen Senn

Department of Statistical Science & Department of Epidemiology and Public Health, University College London, United Kingdom

Abstract: A data set is presented which illustrates various very different aspects of the obstetric practice of a general practitioner in 1916.

Key words: Birth-date, censoring, density estimation, medical statistics, outliers, prediction errors.

1 The source of the data

Sometime between 1976 and 1978, in a second-hand bookshop I purchased a Medical Diary and Visiting List for the year 1916 produced by the Burroughs Wellcome company. A section entitled Obstetric Engagements had been maintained by an unknown person, presumably a general practitioner (GP) and owner of the diary. Some of the entries in the diary form the data of this paper. The data themselves have not been reported before but a partial analysis of various aspects has been published previously by me (see Senn, 1979) The previous analysis was undertaken without aid of the computer and it by no means exhausted all the possibilities of the data.

The original data are recorded as hand-written entries in ink in a pro-forma provided by the diary on 4 double pages. Each pair of pages has 11 pre-printed columns and 31 rows. Each row records an obstetric engagement, that is to say information relevant to the circumstances attending a birth for which the GP was engaged. The 11 column headings are: no, name and address, when engaged, when expected, when attended, sex, no of labour, age, fee, when paid and remarks.

In summary, the data reported in this paper relate to obstetric visits during one year of the First World War carried out almost certainly in Britain, most probably by a GP.

2 The data

From the description above, it would seem, therefore, that there is room to record data consisting of $4 \times 31 = 124$ observations $\times$ 10 variables (if the first column is excluded as only an observation identifier). In actual fact, the doctor recorded information for more than 124 obstetric engagements by using the margins at the bottom of the diary. On the other hand he did not always fill in all columns for the engagements. In particular he only recorded the age of the mother on 4 occasions. The fact that the minimum age recorded is 35 suggests that he perhaps only thought it worth noting the age for older mothers. Furthermore, towards the end of the diary, the date of birth is frequently missing but not the fee. This suggests first, that the fee was agreed beforehand and independently of the work involved in attending the birth, and second that the GP did not record the actual date of birth of those births occurring in 1917 in the diary for 1916, even where his services were engaged in 1916. This is a censoring process which may have to be reflected in certain analyses. Not all missing dates, however, are missing due to this cause. Other explanations may be envisaged: miscarriage or change of address being two obvious ones.

The data in Table 1 and 2 are based on 5 columns of the diary. The identifier in the left-hand column has been added by me since, although the diary had space to record this number, the doctor never did so. In addition I have recorded various additional information to do with difficulties in transcribing data. In particular the three date columns have each had a flag column added which permits attention to be drawn to a problematical value. A problematical value is flagged by a ? symbol. A comment field immediately following the 3 date fields and their respective flags permits an explanation of the problem. Where the comment *imprecise date* is given it means that the month but not the day was recorded in the diary. In such cases, the 15th of the month has been entered in the table by me. Missing values are indicated by blanks. Sex is recorded in the diary as either F or M. All births appear to be singleton with two exceptions. One entry has MM in the original diary (and also in Table 1 here) and this is presumably a case of male twins and another entry has FF, presumably a case of female twins. For patient 43, the entry in this column is F? just as produced in Table 1 here.

Another point worth noting is that in another part of the same diary, in which other general engagements are noted, a radically different handwriting appears between the dates of Wednesday June 28 and Wednesday July 19 inclusive. (This is also detectable in the obstetric diary itself but it is less obvious.) Also, the dates in the diary are nearly all in month/day form.

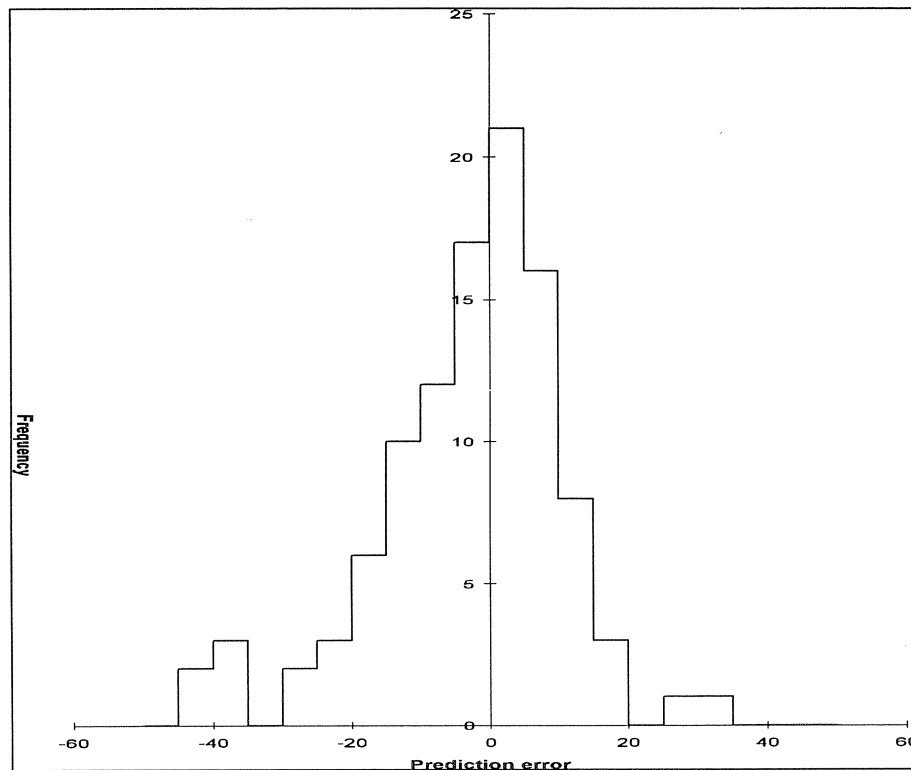


Figure 1: Birth date prediction error

However, a birth attended on 29 June is in day/month form suggesting a different recorder. It appears that a locum must have been engaged for this period, presumably to cover a summer holiday. It is noticeable that no obstetric engagements were made during these three weeks. This point is taken up below.

There are also one or two curious records which are difficult to explain. For example, patients 116 to 118 engaged the services of the GP in the first half of September 1916, as did patients 121 to 124. However *when engaged* is missing for patient 119 but the birth was on July 29 and patient 120 engaged the GP on August 1. There are other occasional indications that the diary was not always maintained chronologically.

Further comments regarding these data will be found in Senn (1979).

3 Suggested analyses

Several different analyses might be made of prediction error, where this

is defined as difference between *when attended* and *when expected*. Figure 1 gives a histogram of prediction error and a curious pattern emerges. It might be interesting to see how various methods of outlier detection perform on these data. Density estimation might also be an interesting application as would methods for mixtures of distributions. The prediction error could also be related to various other original or derived variables, for example engagement interval, where the latter is defined as the difference between *when expected* and *when engaged*. Alternatively the data relating to the three dates could be looked at in terms of survival analysis which raises interesting issues about appropriate time origins.

The different handwriting in the three week period beginning 28 June and the engagement of a locum raises the possibility of a hypothesis test that there is a change in the frequency of new engagements during this period. As already noted above there were none in the three week period in question.

As noted in Senn (1979) birth order is not the same as completed family size. It might be interesting to consider, however, what, if anything, might be inferred about the distribution of completed family size from the data.

Other possible analyses will no doubt suggest themselves to the reader.

Address for correspondence: Email: s.senn@ucl.ac.uk. Fax: +44 (0)171 813 0280.

Reference

- [1] Senn, S.J. (1979), "A Sixty Year Old Medical Record", *Medical Record and Health Care Information Journal*, **20**, 528-531.

No.	When Engaged	When Expected	When Attended	Notes	Sex	Labour
1	aug.27.1915	jan.25.1916				4
2	sep.01.1915	feb.04.1916	feb.10.1916		F	1
3	sep.07.1915	feb.27.1916	mar.06.1916		M	3
4	sep.13.1915	jan.28.1916	jan.21.1916		F	2
5	sep.14.1915	mar.01.1916	mar.20.1916		M	1
6	sep.30.1915	dec.25.1915	jan.26.1916		M	1
7	oct.06.1915	mar.07.1916	mar.26.1916		M	3
8	oct.06.1915	mai.26.1916	mai.23.1916		F	2
9	oct.11.1915	mar.20.1916	apr.03.1916		FF	4
10	oct.14.1915	jan.21.1916	jan.15.1916?	approx date		3
11	oct.23.1915	mar.31.1916	apr.12.1916		F	4
12	oct.10.1915	mai.15.1916?	mai.25.1916	approx date	F	2
13	oct.11.1915	jan.07.1916	jan.10.1916			2
14	oct.27.1915	jan.23.1916	jan.16.1916		M	4
15	nov.16.1915	mar.01.1916?	mar.10.1916	approx date	F	1
16	nov.17.1915	jun.21.1916	jun.23.1916		M	2
17	nov.10.1915	apr.28.1916	apr.19.1916			1
18	dec.17.1915	mar.05.1916	mar.06.1916			1
19	dec.20.1915	feb.10.1916	jan.03.1916		F	2
20	dec.31.1915	apr.20.1916	mai.04.1916		F	5
21	dec.31.1915	mar.17.1916	feb.27.1916		F	2
22			jan.01.1916		M	5
23	jan.01.1916	jul.12.1916	jul.13.1916		F	2
24	jan.01.1916	jun.19.1916	jun.15.1916?	approx date	MM	4
25	jan.03.1916	mai.28.1916	mai.30.1916		M	1
26	jan.04.1916	mai.02.1916	mai.10.1916		F	1
27	jan.06.1916	apr.26.1916	apr.12.1916		M	2
28	jan.11.1916	jun.22.1916	jul.01.1916		F	2
29	jan.12.1916	apr.23.1916	apr.26.1916		F	4
30	jan.12.1916	jul.15.1916	jul.05.1916		M	1
31	jan.18.1916	mar.12.1916				8
32	jan.19.1916	mai.23.1916	mai.31.1916		F	2
33	jan.22.1916	jul.24.1916	jul.13.1916			2
34	jan.22.1916	mar.26.1916	mar.29.1916		M	1
35	jan.22.1916	jun.15.1916	jun.20.1916		F	4
36	jan.24.1916	jul.07.1916	jun.29.1916		F	4
37	jan.25.1916	apr.24.1916	apr.27.1916		M	2
38	jan.31.1916	mar.18.1916	mar.12.1916		M	1
39	feb.05.1916	mai.07.1916	apr.12.1916		M	2
40	feb.06.1916	feb.29.1916	feb.24.1916		M	4
41	feb.24.1916	apr.07.1916	apr.15.1916		F	4
42	feb.29.1916	jun.03.1916	mai.22.1916		F	2
43	feb.29.1916?	feb.10.1916?	feb.13.1916?	suspect dates	F?	3
44	feb.15.1916	jun.01.1916				1
45	feb.29.1916	apr.29.1916	apr.10.1916		M	3
46	mar.06.1916	jul.21.1916	aug.01.1916		M	1
47	mar.10.1916	apr.29.1916				5
48	mar.14.1916	jun.19.1916	jul.04.1916		M	2
49	mar.16.1916	jul.15.1916	jul.16.1916			1
50	mar.18.1916	mai.15.1916	mai.06.1916		M	2
51	mar.20.1916	mai.25.1916	mai.24.1916		M	3
52	mar.21.1916	mai.13.1916				1
53	mar.22.1916	jul.28.1916	aug.01.1916		M	1
54	mar.24.1916	aug.12.1916	aug.13.1916		F	3
55	mar.24.1916	jun.22.1916	jul.02.1916			1
56	mar.29.1916	mai.16.1916	mai.23.1916		M	3
57	mar.29.1916	jun.14.1916	mai.20.1916		M	3
58	apr.05.1916	jul.14.1916	aug.03.1916		M	5
59	apr.05.1916	jun.28.1916	jun.27.1916		M	1
60	apr.05.1916	oct.10.1916				2
61	apr.07.1916	jul.26.1916	jul.28.1916		M	1
62	apr.11.1916	oct.24.1916	oct.21.1916		M	4
63	oct.02.1916	jan.21.1917				2
64	apr.15.1916	oct.01.1916	sep.24.1916		F	1
65	apr.18.1916	sep.09.1916	sep.06.1916		M	1
66	apr.27.1916	apr.27.1916	apr.27.1916		M	10
67	apr.27.1916	aug.07.1916	aug.15.1916		M	4
68	apr.29.1916	jun.30.1916	jul.15.1916		M	13
69	mai.06.1916	sep.25.1916	nov.04.1916		M	2
70	mai.06.1916	jul.30.1916	jul.13.1916		M	1

Table 1: GP's Visiting Diary for 1916

No.	When Engaged	When Expected	When Attended	Notes	Sex	Labour
71	mai.09.1916	aug.03.1916			M	1
72			mai.01.1916			1
73	mai.19.1916	jun.03.1916	mai.31.1916		F	4
74	mai.22.1916	jul.26.1916	jul.29.1916		F	4
75	mai.25.1916	sep.02.1916	sep.08.1916		F	1
76	mai.27.1916	oct.12.1916	oct.05.1916		M	2
77	mai.27.1916	sep.04.1916	sep.12.1916		F	3
78	mai.29.1916	sep.20.1916	sep.27.1916		M	3
79	mai.31.1916	aug.06.1916	jul.17.1916		M	3
80	jun.05.1916	aug.20.1916	aug.22.1916		M	2
81	jun.07.1916	oct.17.1916	oct.08.1916		F	2
82		mai.20.1916	mai.24.1916		F	2
83		jun.20.1916	jun.08.1916		F	3
84	jun.14.1916	sep.01.1916	jul.25.1916			5
85			jun.23.1916		M	4
86	jun.21.1916	nov.26.1916	nov.13.1916		M	2
87	jun.24.1916	jul.15.1916	jul.08.1916		M	1
88		jul.15.1916?	jul.12.1916	approx date	F	3
89	jul.24.1916	oct.14.1916	sep.26.1916		M	4
90	jul.24.1916	oct.07.1916?	oct.14.1916	illegible date	M	1
91	jul.26.1916	oct.22.1916	oct.22.1916		M	4
92	jul.28.1916	oct.23.1916	oct.03.1916		M	3
93	jul.29.1916	oct.10.1916	aug.29.1916		F	1
94	jul.31.1916	oct.17.1916	oct.30.1916		F	1
95	oct.02.1916	nov.30.1916	nov.30.1916		M	1
96	aug.01.1916	jan.11.1917				1
97	aug.03.1916	sep.20.1916	aug.10.1916		M	1
98	aug.05.1916	nov.04.1916				2
99	aug.09.1916	25.aug.16	aug.26.1916		F	3
100	aug.09.1916	aug.23.1916	aug.20.1916		M	1
101	aug.09.1916	aug.25.1916	aug.11.1916		F	1
102	aug.16.1916	oct.29.1916	oct.26.1916		F	1
103	aug.17.1916	feb.20.1917				1
104	aug.17.1916	sep.23.1916	sep.07.1916		M	1
105	aug.19.1916	dec.10.1916	dec.13.1916		M	2
106		aug.10.1916	aug.22.1916		F	4
107	aug.25.1916	sep.25.1916	oct.01.1916		M	2
108	aug.27.1916	oct.14.1916	oct.01.1916		M	2
109	aug.01.1916	jan.08.1917				2
110	jul.24.1916	feb.10.1917				1
111	aug.04.1916	jan.20.1917				3
112	aug.21.1916	jan.22.1917				2
113	aug.28.1916	oct.15.1916				4
114	aug.28.1916	dec.20.1916			F	2
115	aug.28.1916	sep.15.1916	sep.02.1916		F	1
116	sep.05.1916	nov.30.1916	nov.07.1916		M	1
117	sep.05.1916	dec.25.1916				1
118	sep.07.1916	apr.06.1917				2
119		jul.30.1916	jul.29.1916			3
120	aug.01.1916	aug.05.1916	aug.05.1916		M	8
121	sep.12.1916	nov.08.1916	nov.09.1916		M	1
122	sep.13.1916	nov.15.1916	oct.30.1916		M	2
123	sep.13.1916	sep.25.1916	sep.25.1916		F	4
124	sep.14.1916	jan.24.1917				1
125	sep.18.1916	nov.15.1916	nov.02.1916		F	2
126	sep.18.1916	sep.20.1916	sep.20.1916		F	3
127	sep.22.1916	dec.14.1916	dec.19.1916		F	2

Table 2: GP's Visiting Diary for 1916 (continue)

Gassing During the Growth of Bacteriological Infection

Brian Francis and Joe Whittaker

Center for Applied Statistics, Lancaster University, United Kingdom

Mathematics Department, Lancaster University, United Kingdom

Abstract: A data set is presented which exhibits gassing during the cultivation of a bacteriological infection. A method of analysis based on a stochastic process is proposed.

Key words: Bacteriological infection, change-point detection, growth curve, state space model.

1 The scientific context and the data

A common laboratory procedure for detecting contamination in a biological sample for example, salmonella in a sample of egg-yolk, is to ‘grow’ the sample in a specific culture or medium. A sequence of electronic measurements is made on the medium during the period of cultivation, and typically, if the sample is uncontaminated, the measurements stay roughly constant, while otherwise, after a period of quiescence, the measurements respond to the growth of the organism. From this data a decision as to presence or absence of the infection in the original sample is made, and the statistical problem is to detect a change-point in a sequence of noisy measurements. The ability to make the measurement process automatic has led to many applications within the food processing industry and elsewhere, for instance see Whittaker and Frühwirth-Schnatter (1993) who discuss these issues and analyse the data reproduced here.

2 The data

Two independent replicates from a single biological sample measured at 316 time points are given in Tables 1 and 2 and are shown in Figure 1. The measurements come from a micro-biological testing station in Lancashire and were made in 1990. Quite clearly the samples are infected and Whittaker and Frühwirth-Schnatter (1993) discuss state space models to detect the change-point that occurs around time 50 (= 5 hours). However the noise structure of these replicates differ visibly. Due to the build up and bursting of gas bubbles on the electrode immersed in the sample, the second replicate exhibits characteristic fluctuations at the beginning of the series and during the period between 5 and 20 hours after the major growth phase.

Replicate I-1										
0.0	-2.5	-4.9	-6.7	-9.7	-12.9	3.3	3.4	2.9	3.4	0.5
-0.4	-2.2	-5.0	4.0	4.3	4.1	3.7	4.0	4.4	3.9	4.1
4.0	4.7	4.4	4.6	4.6	5.0	5.0	4.6	4.8	5.4	5.3
5.8	5.8	6.5	6.5	6.8	8.3	11.1	13.6	15.4	18.7	21.3
24.4	28.3	34.8	42.7	50.0	56.8	64.0	73.6	83.8	97.9	107.7
122.9	132.9	142.9	144.7	164.0	183.6	194.9	201.3	211.1	222.0	223.7
237.7	241.4	247.7	261.4	270.6	289.4	292.0	302.6	308.8	314.5	317.4
321.7	323.2	314.7	313.7	334.3	323.5	320.6	339.2	331.0	329.5	320.8
340.4	332.7	331.7	322.3	344.9	339.6	338.9	331.5	325.7	347.3	342.2
332.4	335.1	326.4	350.6	348.0	342.2	343.0	330.0	327.9	340.6	341.0
339.6	340.3	335.5	332.7	356.5	354.9	350.8	348.4	348.2	343.9	366.1
365.9	362.6	360.3	354.9	351.8	371.3	371.3	369.5	367.1	362.7	359.8
355.5	376.2	376.0	372.5	369.9	365.7	362.7	358.4	352.2	375.0	369.5
360.5	368.5	355.8	354.8	352.5	349.6	373.4	373.9	373.2	370.8	369.9
367.3	363.4	362.4	358.2	378.7	379.4	379.4	377.8	375.5	374.3	370.9
367.6	364.3	361.4	357.7	381.1	381.3	381.1	380.4	378.7	375.9	375.0
372.3	369.2	368.3	365.5	361.9	383.8	383.9	384.3	384.6	384.6	385.4
385.7	385.5	386.1	386.8	386.8	387.3	386.6	387.3	387.8	388.7	388.7
388.9	388.7	389.4	389.9	390.3	389.4	390.5	391.4	391.5	391.7	392.2
392.8	393.1	393.0	394.2	394.0	394.5	395.1	395.4	394.9	396.1	396.9
397.2	396.7	389.6	390.5	392.2	393.1	392.8	394.5	396.3	397.9	399.7
401.3	403.3	403.3	402.9	401.3	400.6	400.4	401.5	401.7	403.3	403.4
404.3	405.8	406.6	407.0	409.0	409.5	410.2	411.8	412.9	413.1	414.7
415.8	416.7	418.1	418.8	420.3	420.8	421.9	423.2	423.9	425.0	426.2
426.8	428.6	429.5	430.8	431.8	433.1	434.2	435.3	436.4	437.6	438.6
440.4	441.1	442.9	443.7	445.7	447.1	448.2	449.7	450.6	452.4	453.4
454.8	456.3	458.0	459.6	461.1	461.6	463.5	465.0	466.6	467.2	469.6
471.4	472.6	474.2	475.5	477.2	478.9	479.8	482.0	483.2	485.0	486.7
488.6	489.3	491.6	493.5	494.6	495.5	497.8	499.7			

Table 1: Conductance in μ -Siemens for 316 measurements at 6 min intervals on a biological sample (Serie I-1)

Replicate I-2										
2.0	2.5	2.6	3.1	2.9	2.9	2.8	2.6	2.3	3.9	3.7
3.1	3.6	3.4	3.1	3.4	3.4	3.1	2.0	3.3	2.9	3.4
3.1	3.6	2.8	3.6	3.4	3.4	3.7	3.4	4.0	4.3	4.6
4.6	4.9	5.4	6.5	8.5	10.5	11.6	14.0	15.4	18.1	22.1
27.6	30.2	33.3	37.9	44.3	50.7	63.2	73.9	82.0	90.7	104.4
116.9	122.5	133.5	146.8	169.5	174.3	194.3	217.1	228.2	231.5	254.3
248.0	261.6	266.3	284.8	290.5	294.5	309.9	320.9	327.0	330.4	334.3
341.3	346.2	346.2	351.4	355.0	356.8	361.5	364.2	368.9	371.8	374.3
375.8	377.2	377.6	378.7	379.9	380.8	382.2	383.7	383.9	384.5	385.6
386.6	388.3	388.7	389.3	390.7	392.0	393.2	394.2	394.7	395.3	397.1
398.4	400.0	400.6	402.5	402.9	404.5	405.4	407.0	408.8	410.7	411.3
412.7	413.8	415.4	416.4	417.6	418.9	420.1	421.9	423.1	424.7	424.8
426.0	426.2	427.0	426.8	427.4	428.6	428.8	428.8	429.2	430.6	430.0
430.8	431.6	432.0	432.0	432.6	431.6	433.1	433.1	433.3	433.3	433.5
434.7	434.5	434.9	434.5	434.5	435.7	436.1	435.9	436.7	436.7	437.5
437.3	437.9	437.3	438.7	438.9	439.1	439.1	438.3	439.1	439.1	439.1
440.1	440.5	440.5	440.3	441.9	441.9	440.1	440.5	441.1	441.1	441.3
441.3	442.5	443.1	444.1	445.1	444.9	446.1	446.7	447.3	448.7	448.3
448.5	449.5	450.5	451.5	451.7	452.1	452.9	453.5	453.7	453.9	454.5
455.1	455.9	456.5	457.1	458.1	458.1	458.7	459.3	459.5	460.3	460.7
461.3	461.7	462.1	463.1	463.1	464.3	464.9	465.7	466.5	466.9	466.9
468.4	469.2	470.2	470.2	470.6	473.0	473.8	474.4	475.8	476.3	478.1
478.7	479.3	479.7	478.1	483.2	483.6	484.0	485.6	486.0	487.2	488.9
489.9	490.9	491.7	494.0	495.2	496.0	496.9	498.5	499.9	501.2	503.0
504.1	505.3	506.3	508.4	509.6	510.7	512.5	513.8	514.4	516.0	517.7
519.2	520.6	522.3	523.7	525.0	526.4	527.7	529.4	530.4	532.1	532.3
533.7	534.8	536.9	537.9	539.8	541.5	542.4	543.6	544.9	546.8	547.8
549.1	551.2	552.7	554.2	554.8	557.8	558.8	560.1	562.0	563.7	565.4
567.3	569.3	571.0	572.0	573.7	575.4	577.2	579.3			

Table 2: Conductance in μ -Siemens for 316 measurements at 6 min intervals on a biological sample (Serie I-2)

3 Proposed analysis

A simple but crude way to model this behaviour is to substantially increase the variance of the noise term in analysis of the data. However since the scientific reason for this stochastic variation is roughly understood it would be desirable to characterise it more specifically. A first step would be to model the data as a realisation of a stochastic process. A appropriate process might be one formed by the sum of the ordinary growth process without gassing together with a non-homogenous Poisson process whose jumps indicate the times at which the bubble burst and the size of the bubble.

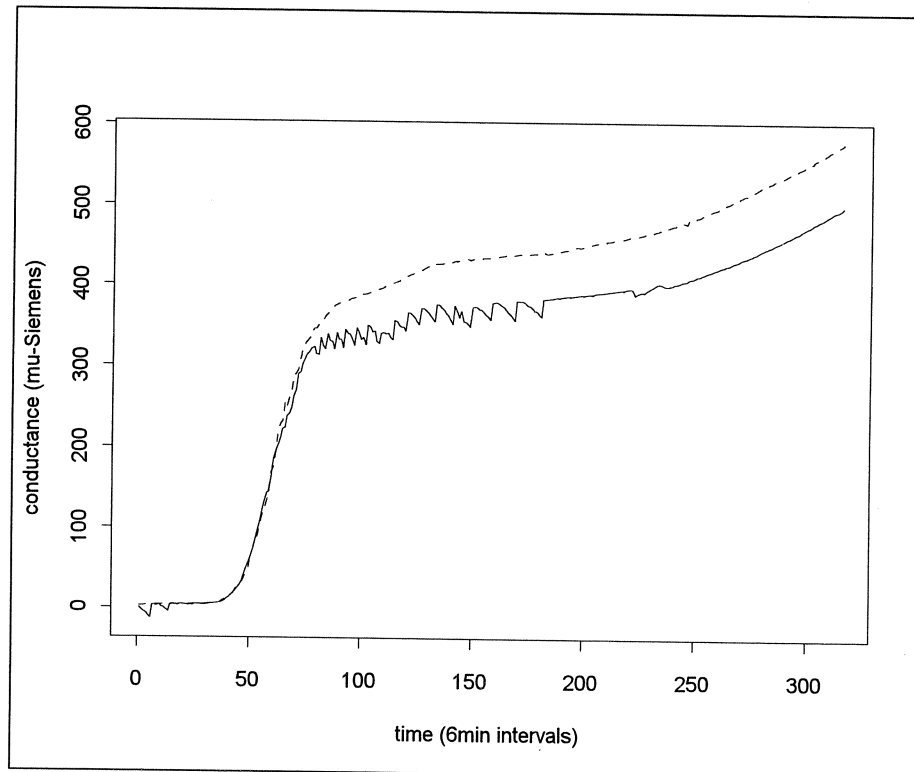


Figure 1: Conductance (in μ -Siemens) plotted at 6min time intervals for Samples I-1 and I-2

Address for correspondence: Email: joe.whittaker@lancaster.ac.uk.
Fax: +44 (0)524 841701.

Reference

- [1] Whittaker, J. and Frühwirth-Schnatter, S. (1994), "A dynamic change-point model for detecting the onset of growth in bacteriological infections", *Appl. Statist.*, **43**, 4, 625-640.

Verbal Memory Following Temporal Lobectomy

Shelley Channon and David J. Hand

Department of Psychology, University College London & Camden & Islington, Community Health Services NHS Trust, United Kingdom

Department of Statistics, The Open University, United Kingdom

Abstract: This data set was collected to study verbal memory in subjects who had undergone unilateral temporal lobectomy for the relief of intractable epilepsy. Three word lists were constructed consisting of uncategorised words and categorised words, presented in randomised and clustered order. It was hypothesised that the left temporal lobectomy subjects should be impaired relative to healthy controls. Since they might be expected to show particular difficulties in remembering material which lacked intrinsic structure, their deficits should be apparent on the uncategorised word list, and on the categorised list presented in randomised order.

Key words: Binomial distributions, counts, epilepsy, memory, neurology, proportions, psychology, repeated counts, two groups.

1 Introduction

The data reported below were collected during 1987. They form part of the findings of a study, which was designed to investigate the effects of structure and clustering on verbal memory in subjects who had undergone unilateral lobectomy for the relief of intractable epilepsy. It has long been established that in right-handed people, the left (dominant) hemisphere of the brain predominantly processes verbal material, whereas the right (nondominant) hemisphere is specialised for the processing of nonverbal material. The temporal lobes and limbic system are known to be important anatomical sites involved in memory function, the left temporal lobe dealing

primarily with verbal material and the right with nonverbal material. The study of subjects with temporal lobe dysfunction can elucidate specific aspects of the role of the temporal lobe structures in memory function. Subjects who have undergone surgery for epilepsy provide a particularly useful opportunity to study the role of specific areas of brain function, since epilepsy surgery generally uses standard operating techniques which remove comparable areas of brain tissue in a series of subjects. The samples of patients who have undergone unilateral (i.e. to one side or other of the brain) temporal lobectomy (removal of a portion of the temporal lobe) therefore consist of a relatively homogeneous group.

Studies of healthy control subjects have shown that the efficiency of memory is increased by the use of strategies which impose order on the to-be-remembered material and integrate it into a structure, rather than relying on simple rote rehearsal (see e.g. Katona, 1940). In the absence of formal instructions about learning strategies, people frequently organise material into categories (see e.g. Bousfield, 1953), even when no obvious relations between the words exist (see e.g. Tulving, 1962). Thus, words which belong to particular semantic categories such as birds, fruit, flowers, and so on, will be more easily remembered than unrelated words. The extent to which categorised words are clustered together is also likely to affect recall. Thus, it is easier to remember the list 'robin, sparrow, duck, seagull, cherry, banana, apple, orange' than to remember the same words in randomised order. The impairments in verbal memory shown by subjects with left temporal lobe dysfunction may be partially attributable to difficulties in initiating or carrying out organisational strategies to aid recall. The present study was designed to investigate this.

2 Method

Three word lists were constructed. The uncategorised list (UC) consisted of 16 unrelated words presented in random order. The clustered categorised list (CC) consisted of 16 words comprising 4 words from each of 4 categories, which listed all the words from the first category, then all those from the second category, and so on. The randomised categorised list (RC) also consisted of 16 words, 4 from each of 4 categories, this time presented in randomised order. The 3 word lists were matched in terms of word frequency, number of syllables, and length.

The word lists were presented on a computer screen, and subjects were instructed to remember as many as they could, but were not given any training or specific information as to the types of strategies they might use to achieve this. The three lists were presented in counterbalanced order

Group	subjects	Immediate recall			Delayed recall		
		UC	RC	CC	UC	RC	CC
Left temporal lobectomy	1	9	5	6	1	5	6
	2	6	6	10	3	3	2
	3	4	2	8	1	1	2
	4	2	6	1	0	0	1
	5	10	9	12	1	7	8
	6	7	9	14	3	6	14
	7	6	6	10	1	6	10
	8	7	7	8	2	0	5
	9	4	4	7	0	0	1
	10	6	6	5	2	1	4
	11	4	9	8	2	1	6
	12	6	6	8	4	1	6
Healthy controls	13	6	11	13	2	10	10
	14	9	11	16	8	13	15
	15	8	7	14	5	10	13
	16	11	12	16	10	10	12
	17	11	15	16	10	16	15
	18	7	6	10	5	6	6
	19	8	10	12	2	11	9
	20	9	11	16	2	14	14
	21	8	10	13	2	11	8
	22	9	10	14	6	4	12
	23	5	4	8	2	5	6
	24	7	8	7	2	7	6

Table 1: Number of words recalled by the left temporal lobectomy and control subjects on the 3 word lists

within each group. Recall was tested both immediately after presentation. Subjects then carried out a recognition test which involved discriminating target words from distractor words; this was designed to investigate further hypotheses about memory function, and is beyond the scope of this paper. In between each word list, they carried out brief distractor tasks, to minimise confusion between the lists. Recall of all three lists was tested again after a 30 minute time delay, since some studies have suggested that temporal lobe dysfunction is often characterised by reasonable retention immediately after presentation of to-be-remembered material, but rapid forgetting over time. These data are a subset of those described by Channon, Daum and Polkey (1989).

It was predicted that left temporal lobectomy subjects would be impaired relative to healthy controls. Furthermore, it was expected that the deficits shown by left temporal lobectomy subjects would be greater on the

randomised categorised list compared to the clustered categorised list.

3 Results

Twelve subjects in each group participated in the study, and the groups did not differ significantly in age or IQ. The number of words recalled for each list for individual subjects in each group is shown in Table 1. A graphical display may be seen in Figure 1.

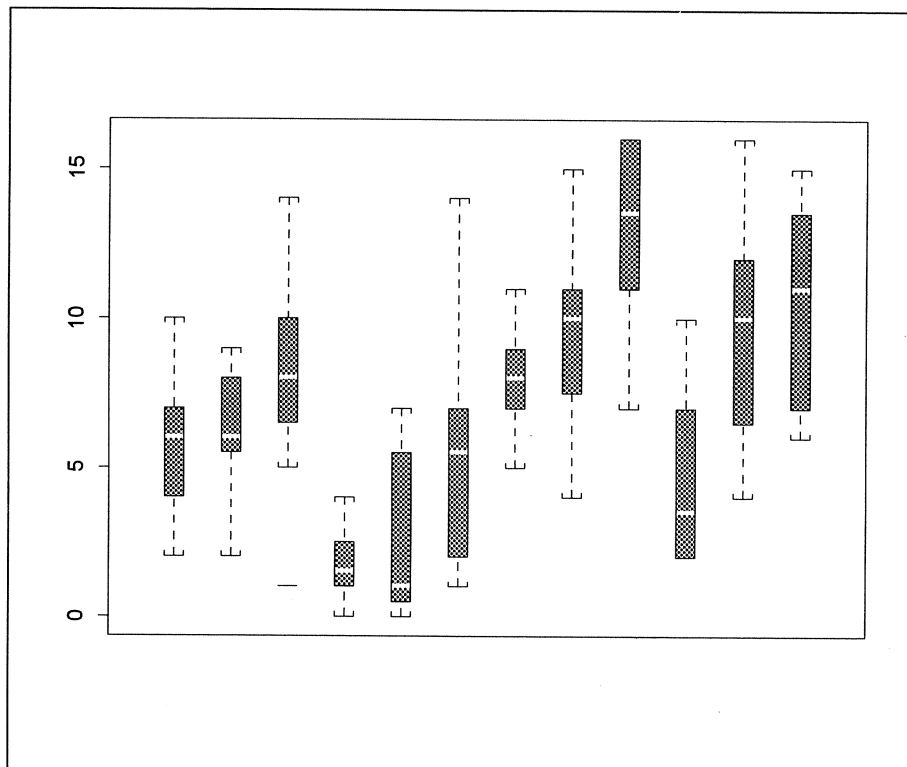


Figure 1: Boxplots of the data, in order, from the left: lobectomy, immediate recall, UC, RC, CC; lobectomy, delayed recall, UC, RC, CC; controls, immediate recall, UC, RC, CC; controls, delayed recall, UC, RC, CC

Inspection of the data without statistical analysis casts some light on the questions addressed by the study. Firstly, the findings suggest that the left temporal lobectomy subjects tended to recall fewer words than the healthy control subjects. Looking at this question in relation to the effects of the three types of word lists, differences between the groups can also be seen. The highest scores for both groups were for recall of the CC list, where

the categorised words had already been organised in a manner likely to promote ease of retrieval. As expected, the healthy control subjects tended to perform better on the RC list than on the RR list, whereas there was less indication of such a tendency for the left temporal lobectomy subjects. Comparison of immediate and delayed recall shows that a number of the healthy control subjects produced slightly more words from the RC list on delayed recall than on immediate recall. This was specific to the RC list, and did not occur for any of the left temporal lobectomy subjects, ruling out any general explanation in terms of the design of the study (such as presentation of the recognition list). This may reflect natural random variation, or slight differences in the strategies used by the healthy controls at immediate and delayed recall, with the latter reflecting a more efficient semantic strategy to code and retrieve these words, such as cueing themselves with category labels. However, subjects were not asked to report on the strategies they used, and such hypotheses can therefore only remain speculative.

Address for correspondence: Dr. S. Channon, Psychology Department, Philips House, UCL, Gower Street, London WC1 6BT, UK.

References

- [1] Bousfield, W.A. (1953), "The occurrence of clustering in recall of randomly arranged associates", *J. of General Psychology*, **49**, 229-240.
- [2] Channon, S., Daum, I. and Polkey, C.E. (1989), "The effect of categorisation on verbal memory after temporal lobectomy", *Neuropsychologia*, **27**, 777-785.
- [3] Katona, G. (1940), *Organising and Remembering*, New York, Columbia University Press.
- [4] Tulving, E. (1962), "Subjective organization in free recall of 'unrelated' words", *Psychological Review*, **69**, 344-354.