

IN THIS ISSUE

A NOTE ON AN ESTIMATOR FOR THE MEAN OF FINITE POPULATIONS

In this paper, the authors have derived an optimal estimator of the mean of a finite population. They have shown that it is more efficient than the usual sample mean as well as Searls' estimator of the mean of infinite populations. Moreover, their estimator is "more optimal" than Searls' estimator in that the optimality is derived over a wide class of estimators.

Saralees Nadarajah is assistant Professor at the department of Mathematics at Illinois State University. His research interests include distribution theory, extreme value theory, branching processes, sampling theory and simulation techniques. Before coming to Illinois, he held positions at the University of California at Riverside, at the University of California in Santa Barbara and at the universities of Sheffield, Plymouth and Nottingham in England. He obtained his Ph.D. in Statistics from the University of Sheffield in 1994.



CHANGE-POINT MODEL FOR A SEQUENCE OF RANDOM VARIABLES WITH INDEPENDENT INCREMENTS

In the second paper the authors consider a sequence of random variables with independent increments which can change their distribution at some unknown instant. They propose a test for detecting a change when these variables follow Poisson distributions and study the behavior of the statistic under both the null hypothesis and the alternative. They use this model with two data sets.



Héctor de Arazoza was born in 1948 in Havana, Cuba. He studied Mathematics at the University of Havana graduating in 1970. He obtained a M.Sc. in Mathematics at McGill University, Canada (1971) and his Ph.D. at the University of Havana (2000). Currently Dr. de Arazoza is with the University of Havana.

AN ALTERNATIVE DERIVATION OF THE NONCENTRAL CHI-SQUARE DISTRIBUTION

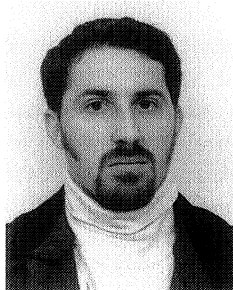
Needless to say, the noncentral chi-square distribution is an important distribution in statistics. Many famous statisticians studied the noncentral chi-square distribution and its properties. Tools for studying the noncentral chi-square distribution include the moment generating function and the characteristic function. Of course, we can also use the distribution theory approach to deal with the noncentral chi-square distribution. The aim of this paper is to point out an application of the distribution theory to the noncentral chi-square distribution.

Dr. Wan-Kai, Pang obtained his PhD degree in Statistics from the University of Southampton (UK) in year 2000. He has been working in the Department of Applied Mathematics in The Hong Kong Polytechnic University for almost twenty years as lecturer in Statistics. His current research interests include computational statistics, Markov chain Monte Carlo methods, statistical reliability and Bayesian modelling.



ADAPTIVE REGRESSION: A SIMULATION STUDY

This paper presents an extension of Harter's adaptive method to the regression model. This simulation study shows that Harter's procedure can be applied to the regression model when sample size is larger than 200. Christophe Beuret and Dr. Valentin Rousson propose another related adaptive method based on hypothesis testing and tabulated critical values for the kurtosis of the residuals of a regression model.



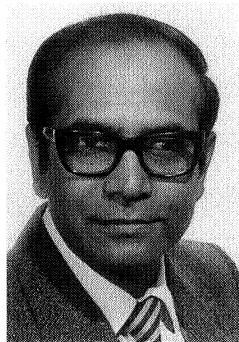
Christophe Beuret was born in Neuchâtel in 1972. He obtained his Bachelor of Science in Economics from the University of Neuchâtel in 1997 and his Master degree in Statistics in 1999. Christophe Beuret is currently a Ph.D Student at the University of Neuchâtel. His current research is on application of method of handling missing data to unemployment data.

A NOTE ON THE BEHRENS-FISHER PROBLEM FOR ELLIPTICALLY SYMMETRIC DISTRIBUTIONS

The paper deals with the Behrens-Fisher problem for the elliptical class of distributions from a Bayesian perspective. A joint posterior distribution of

location parameters is derived and this can be used to make inferences about any function of the location parameters when variance parameters are unknown and not equal.

Lehana Thabane graduated from the National University of Lesotho (NUL) in 1990. He received his MSc. In Statistics at Sheffield University. He completed his PhD in Statistics in 1998 at the University of Western Ontario. Lehana Thabane is currently working as Assistant Professor in the Department of Mathematics, Statistics and Computer Science at St. Francis Xavier University, Canada.



Professor Safiul Haq attended the University of Bangladesh from which he graduated with a BSc. In 1954 and Msc. In 1955. After five years as a Statistician in the Agriculture Department in Pakistan he obtained an MA in 1964 and PhD in Mathematics in 1966 at the University of Toronto, Canada. His research interests include prediction analysis, multivariate analysis, statistical inference using structural, Bayesian or likelihood approaches. He currently holds a position of Professor Emeritus at University of Ontario

CAPTURING HISTORY

A highlight of this issue is a letter from Peter J. Huber dated 1972 in which he writes to Geof Watson about the Princeton Study and his PhD student Theo Gasser (now head of Department of Biostatistics, University of Zurich) and his book on Robustness.

DATA AND STATISTICS

Data and Statistics Contains two sets of data one on health related problem and the other on olympic data. The objective of the first study is to estimate the effect of radio frequency electromagnetic fields on melatonin in human beings and cattle. Since it is well established that melatonin has a 24h rhythm Lerchl and Partch propose a model which they term complex-cosinor-analysis. This model consists in two cosinors of frequency 24h and 12h. The paper presents data from one cow and proposes a solution to the estimation and characterization problem.

In the second data set the authors propose different criteria of ranking nations in Olympic Games. In particular a standardized score produces a ranking which improves the representativeness of countries of modest size.

A Dual formulation of Laplace's Minimax problem

Richard W. Farebrother

11 Castle Road, Bayston Hill, Shrewsbury, United Kingdom

Abstract: In this paper we explain how the minimax or L_∞ line fitting problem may be solved by way of a closely related least absolute deviations or L_1 problem. We illustrate this technique by solving a geodetic problem that Laplace (1786) used as a worked example.

Keywords: duality; L_1 -norm; L_∞ -norm; least absolute deviations; linear programming; minimax absolute deviation.

1 Introduction

In this paper we will show how a result due to Farebrother (1985) may be used to obtain the solution to the minimax or L_∞ line fitting problem by solving a closely related least absolute deviations or L_1 problem.

2 Transformed Dual Formulation of the Problem

Let T be a given $n \times q$ matrix of rank q , and let s be a given $n \times 1$ matrix that is not linearly dependent on the columns of T . then the minimax or L_∞ -norm estimation problem is concerned with the determination of a $q \times 1$ matrix d in such a way that the largest in absolute value of the elements in the $n \times 1$ matrix $g = s - Td$ is as small as possible. Setting this largest value equal to k , we find that this problem may be expressed in the form of a linear programming problem:

$$\text{Minimise } k \tag{2.1}$$

subject to

$$g = s - Td \tag{2.2a}$$

$$\text{and } -kw \leq g \leq kw \tag{2.2b}$$

where w is an $n \times 1$ column of ones. (In a more general context, w could be a given $n \times 1$ matrix of positive numbers.)

Let X be an $n \times p$ matrix of rank $p = n - q - 1$ satisfying $X'T = 0$ and $X's = 0$; let y be an $n \times 1$ matrix of rank one satisfying $y'T = 0$ and $y's = m$ where m is a positive scalar, and let Z be an $n \times q$ matrix of rank q such that the $n \times n$ matrix $[X \ y \ Z]$ is nonsingular, then we may rewrite problem (2.1)-(2.2) as

$$\text{minimise } k \quad (2.3)$$

subject to

$$X'g = 0 \quad (2.4a)$$

$$y'g = m \quad (2.4b)$$

$$Z'g = Z's - Z'Td \quad (2.4c)$$

$$\text{and } -kw \leq g \leq kw \quad (2.4d)$$

that is, as

$$\text{minimise } k \quad (2.5)$$

subject to

$$X'g = 0 \quad (2.6a)$$

$$y'g = m \quad (2.6b)$$

$$\text{and } -kw \leq g \leq kw \quad (2.6c)$$

before forming

$$d = (Z'T)^{-1}Z'(s - g^*) \quad (2.6d)$$

where (2.6d) expresses the optimal value of d in terms of the optimal value of g .

Setting $g = kh$ in problem (2.5)-(2.6) where k is the objective function of the minimax problem, we have

$$\text{maximise } m/k = y'h \quad (2.7)$$

subject to

$$X'h = 0 \quad (2.8a)$$

$$\text{and } -w \leq h \leq w \quad (2.8b)$$

where h is an $n \times 1$ matrix whose elements are unrestricted in sign.

Now, the linear programming dual of this problem is given by

$$\text{minimise } w'e_1 + w'e_2 \quad (2.9)$$

subject to

$$Xb + e_1 - e_2 = y \quad (2.10a)$$

$$\text{and } e_1, e_2 \geq 0 \quad (2.10b)$$

where e_1 and e_2 are $n \times 1$ matrices with nonnegative elements, and where b is a $p \times 1$ matrix whose elements are unrestricted in sign.

Thus, we have established that the minimax problem (2.1)-(2.2) is intimately associated with a transformation of the least absolute deviations or L_1 problem (2.9)-(2.10).

3 An Application of the Dual Technique

The minimax line fitting problem was formally introduced by Laplace (1786) in the context of the following system of four equations:

$$56753 - c - 0.00000d + v_0 = 0 \quad (3.1)$$

$$57037 - c - 0.30142d + v_1 = 0 \quad (3.2)$$

$$57074 - c - 0.576207d + v_2 = 0 \quad (3.3)$$

$$57405 - c - 0.838866d + v_3 = 0 \quad (3.4)$$

where all the unknowns are measured in *toises* (French fathoms), and where, for simplicity, we have replaced Laplace's unknowns z and y by c and d , and his errors x, x', x'' and x''' by $v_0 = -g_0, v_1 = -g_1, v_2 = -g_2$ and $v_3 = -g_3$.

Although these four equations were not explicitly stated by Laplace, the necessary latitude and meridian arc measurements observed in Equador (then part of Peru), the Cape of Good Hope, France, and Lapland are to be found in the text, see Laplace (1786; 1895, p.5). Moreover, on subtracting equation (3.1) from equations (3.2)–(3.4), and renormalising on the coefficient of d , we have the expressions:

$$d = 942.19 + 3.3176(v_1 - v_0) \quad (3.5)$$

$$d = 557.09 + (1.7355(v_2 - v_0)) \quad (3.6)$$

$$d = 777.24 + 1.1921(v_3 - v_0) \quad (3.7)$$

which were given by Laplace (1786; 1895, p.6). Further, on subtracting equation (3.6) from equations (3.5) and (3.7), we obtain two constraints on the error terms:

$$0 = 385.10 - 1.7355v_2 + 3.3176v_1 - 1.5821v_0 \quad (3.8)$$

and

$$0 = 220.15 + 1.1921v_3 - 1.7355v_2 + 0.5434v_0 \quad (3.9)$$

which were also given by Laplace (1786; 1895, p. 6).

If Laplace had continued his analysis in this manner, he would presumably have obtained a homogeneous constraint on the errors by renormalising equations (3.8) and (3.9) on their constant terms before subtracting one expression from the other. For simplicity, we obtain an equivalent result by multiplying equation (3.8) by $220.15/385.10 = 0.5717$ to obtain

$$0 = 220.15 - 0.9921v_2 + 1.8966v_1 - 0.9044v_0 \quad (3.10)$$

and, on subtracting this equation from (3.9), we have

$$0 = 1.1921v_3 - 0.7443v_2 - 1.8966v_1 + 1.4478v_0 \quad (3.11)$$

setting $h_j = g_j/k = -v_j/k$ in equations (3.9) and (3.11), we find that we have to solve the linear programming problem (2.7)–(2.8) or:

$$\text{Maximise} \quad 1.1921h_3 - 1.7355h_2 + 0.5434h_0 \quad (3.12)$$

subject to

$$0 = 1.1921h_3 - 0.7443h_2 - 1.8966h_1 + 1.4478h_0 \quad (3.13a)$$

$$\text{and} \quad -1 \leq h_0, h_1, h_2, h_3 \leq +1 \quad (3.13b)$$

Further, setting $u = e_1 + e_2$ in equations (2.9)–(2.10), we find that the linear programming dual of equations (3.12)–(3.13) takes the form of the least absolute deviations problem:

$$\text{Minimise} \quad |u_0| + |u_1| + |u_2| + |u_3| \quad (3.14)$$

subject to

$$1.1921 = 1.1921b + u_3 \quad (3.15a)$$

$$-1.7355 = -0.7443b + u_2 \quad (3.15b)$$

$$0.000 = -1.8966b + u_1 \quad (3.15c)$$

$$0.5434 = 1.4478b + u_0 \quad (3.15d)$$

where the unknown parameter b and the errors u_0 , u_1 , u_2 and u_3 are unrestricted in sign.

Now, it is well-known that the solution to the l_1 -norm line fitting problem with p unknown parameters is characterised by a set of p zero residuals. In the present context, we have a single parameter and thus a choice of four possible values $1.7355/0.7443 = 2.33$, $1.1921/1.1921 = 1.00$, $0.5434/1.4478 = 0.38$, and $0.0000/1.8966 = 0.00$ for b , where the ratios have been arranged in decreasing order. The denominators of these expressions (taken in this order) define the cumulated sums of b -coefficients: 0.7443, 1.9364, 3.3852, 5.2808. And the third term in this series is the first to exceed one-half the total, $5.2808/2 = 2.6404$. Thus, we may deduce that the ratio corresponding to a zero value for u_0 defines the solution to the L_1 line fitting problem of equations (3.14)–(3.15). Further, we may deduce that the solution to the L_∞ line fitting problem of equations (3.1)–(3.4) is associated with a common absolute value for v_1 , v_2 , and v_3 , (that is, for all errors except v_0).

Clearly, Laplace's characterisation of the optimal solution to the minimax problem may be deduced from Gauss's subsequent characterisation of the optimal solution to the least absolute deviations problem and *vice versa*.

4 An Alternative Primal Solution

Of course, Laplace could not have used this form of analysis more than 150 years before the development of linear programming duality theory, see Grattan-Guinness (1994). Instead, he arbitrarily selected one of the four equations of the type indicated by equations (3.8) and (3.9), and noted, without proof, that the largest in absolute value of the three errors defining the selected equation is minimised when they have a common magnitude and signs that are opposite to the signs of the coefficients in the selected equation.

In the present context, this common magnitude is most easily determined by dividing the absolute value of the constant term by the sum of the absolute values of the coefficients in the selected equation. for example, in the case of equation (3.9), we have

$$220.15/(1.1921 + 1.7355 + 0.5434) = 63.43 \quad (3.16)$$

for the common absolute value of v_0, v_2 , and v_3 . The corresponding value of v_1 may then be determined from equation (3.8). If it is no larger in absolute value than the other three errors, then the computed values correspond to the desired solution. If not, then we have to consider alternative equations of the type (3.8)–(3.9) that, in turn, omit v_0, v_2 , and v_3 . In fact, Laplace (1786; 1895, p.8) found that the minimax solution is given by the equation with v_0 omitted, and obtained a value of $d = 684.4$ *toises* with errors $-v_1 = v_2 = -v_3 = 75.57$ *toises* and $v_0 = 2.04$ *toises*.

In principle, this second (primal) technique may be used when there are any number of equations in any lesser number of unknowns. However, in its original form, this procedure clearly lacked a mechanism for choosing successive subsets of the equations in an appropriate manner. And, for this, or for some other reason, the approach adopted by Laplace in 1786 seems to have been lost to the literature until rediscovered by de la Vallée Poussin in 1911, see Farebrother (1999) for details.

5 Concluding Remarks

In this paper we have shown that the solution to the minimax line fitting problem may be obtained by solving a closely related least absolute deviations problem (and *vice versa*). Unfortunately, although this procedure is of considerable pedagogic interest, it is most unlikely to have any practical value except, possibly, when the minimax fitting problem is only slightly overdetermined; compare the related work of Seneta and Steiger (1984).

References

- [1] Farebrother, R. W. (1985), Unbiased L_1 and L_∞ estimation, *Communications in Statistics (Series A)* 14: 1941–1962.
- [2] Farebrother, R. W. (1999), *Fitting Linear Relationships: A History of the Calculus of Observations 1750–1900*, Springer-Verlag, New York.
- [3] Grattan-Guinness, I. (1994), "A new type of question": On the prehistory of linear and non-linear programming 1770–1940, *The History of Modern Mathematics* 3: 43–89.
- [4] Laplace, P. S. (1786), Mémoire sur la figure de la terre, *Mémoires de l'Académie Royale des Sciences de Paris (pour 1783)*, pp. 17–46. Reprinted in his *Oeuvres Complètes*, Vol. 1, Gauthier-Villars et Fils, Paris, 1895, pp. 3–32.
- [5] Seneta, E. and W. J. Steiger (1984), A new LAD curve-fitting algorithm: Slightly overdetermined equation systems in L_1 , *Discrete Applied Mathematics* 7: 79–91.

A Note on An Estimator for the Mean of Finite Populations

S. Nadarajah

Department of Mathematics, University of South Florida, Tampa, USA

D. Choi

Department of Mathematicla Sciences, Portland State University, Portland, USA

M. Ruiz-Espejo

Faculty of Economics, Universidad National de Educación a Distancia, Madrid, Spain

H.P. Singh and R. Singh

School of Studies in Statistics, Vikram University, Ujjain, India

Abstract: Donald T. Searls proposed an optimal estimator of the mean of infinite populations that is more efficient than the usual sample mean. In this note we generalize his result for finite populations, under simple random sampling without replacement, and establish optimality over a wider class of estimators. We note that our estimator – in addition to being more efficient than the usual sample mean – can be more efficient than Searls' estimator also.

Key Words: Coefficient of variation, Mean squared error, Optimal estimation, Simple random sampling.

1 Introduction

Let $U = \{1, \dots, N\}$ denote a finite population of N distinct and identifiable units, uniquely labeled from 1 to N . Let Y denote the study variable taking positive values, Y_i for the i th unit of the population, U . Assume that Y_i are independent and identically distributed (iid) random variables with $E(Y_i) = \mu$ and $Var(Y_i) = \sigma^2$. This amounts

to assuming that the population values are a random sample from a ‘superpopulation’ with mean μ and standard deviation σ . Consider the population mean

$$M = \frac{1}{N} \sum_{i \in U} Y_i$$

and a class of estimators

$$Z = \sum_{i \in s} t_i Y_i, \quad (1)$$

where t_i are real constants and $s \subset U$ is a simple random sample of size n drawn without replacement. For an infinite population ($N = \infty$), Searls (1964) [5] showed that

$$Z_{opt} = \frac{1}{n + c^2} \sum_{i \in s} Y_i, \quad (2)$$

where $c = \sigma/\mu$, is an optimal estimator of M in that it minimizes the mean squared error within the class

$$Z = w \sum_{i \in s} Y_i.$$

This note generalizes this result for finite N (under simple random sampling) and for the wider class, (1), where every observation is allowed the freedom to be associated with a different coefficient. The proof of the main result (Theorem 1) is sketched in the Appendix.

2 Results

The main theorem derives expressions for t_i for (1) to be optimal. The mean squared error is taken with respect to both the random variables and sampling: E_R denotes the expectation with respect to the random variables and E_S denotes the expectation with respect to sampling. In the literature this is called the superpopulation approach.

2.1 Superpopulation Approach

Theorem 1 *Suppose that Y_i are positive-valued and iid random variables with $E(Y_i) = \mu$ and $\text{Var}(Y_i) = \sigma^2$. Then $Z = \sum_{i \in s} t_i Y_i$ is a uniform minimum mean squared error estimator for M , in the sense that*

$$\begin{aligned} E_R \left\{ E_S (Z - M)^2 \right\} &= \frac{\sigma^2}{N} + \mu^2 - \frac{2n}{N^2} (\sigma^2 + N\mu^2) \sum_{i \in U} t_i \\ &\quad + \frac{n}{N} (\sigma^2 + \mu^2) \sum_{i \in U} t_i^2 \\ &\quad + \frac{n(n-1)}{N(N-1)} \mu^2 \sum_{i \in U} \sum_{j \in U \setminus \{i\}} t_i t_j \end{aligned} \quad (3)$$

attains its global minimum if

$$t_i = \frac{N + c^2}{N(n + c^2)}, \quad i \in U. \quad (4)$$

Remark 1 Since (from (4))

$$t = \sum_{i \in U} t_i = \frac{N + c^2}{n + c^2} \neq \frac{N}{n} \text{ unless } c^2 = 0,$$

the usual sample mean,

$$\bar{Y} = \frac{1}{n} \sum_{i \in s} Y_i,$$

cannot be optimal for M in the sense of Theorem 1 unless $c = 0$.

Remark 2 The class of estimators, (1), is a special subclass of estimators studied by Godambe for finite populations (see, for example, Godambe (1995) [2]). It would be of interest to generalize the above theorem for this more general class. We leave this to an interested reader.

The resulting optimal estimator

$$Z_{opt} = \sum_{i \in s} t_i Y_i = \frac{1 + \frac{1}{N} c^2}{n + c^2} \sum_{i \in s} Y_i \quad (5)$$

is the generalization of (2) for finite N . This can be used in practice only when the coefficient of variation, $c = \sigma/\mu$, is known in advance. In many situations, the experimenter may have a figure for the coefficient of variation (particularly, in life sciences and biological experiments) from long association with the experimental material or from empirical evidence gathered from repeated experiments, see, for example, Searls (1964) [5], Murthy (1967, p. 96) [4], Khan (1968) [3], Gleser and Healy (1976) [1] and Sen (1978) [6].

$c = \sigma/\mu$	Population size					
	1000	5000	10000	25000	50000	100000
0.50	2.47	0.49	0.25	0.10	0.05	0.02
0.75	5.57	1.11	0.56	0.22	0.11	0.06
1.00	9.89	1.98	0.99	0.40	0.20	0.10
1.25	15.44	3.09	1.55	0.62	0.31	0.15
1.50	22.22	4.45	2.23	0.89	0.45	0.22

Table 1: Efficiency gain of (5) over the usual sample mean (in percentage) when the sample size n is 1% of the population size N .

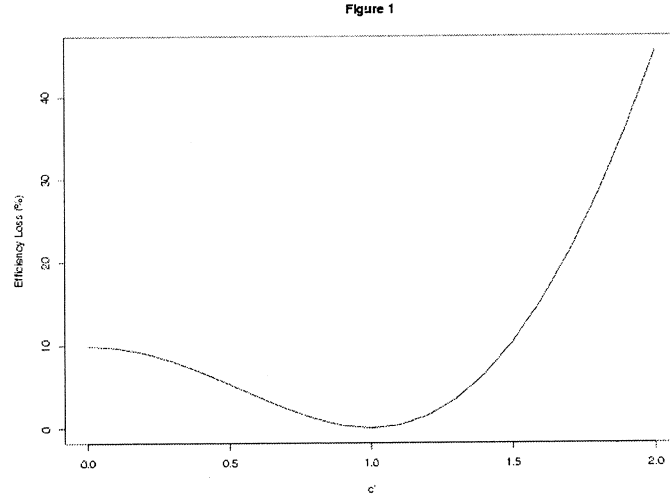


Figure 1: Efficiency loss of (3) over (6) when c in (4) is incorrectly specified by c' . The true value of $c = 1$, $N = 1000$ and n is 1% of N .

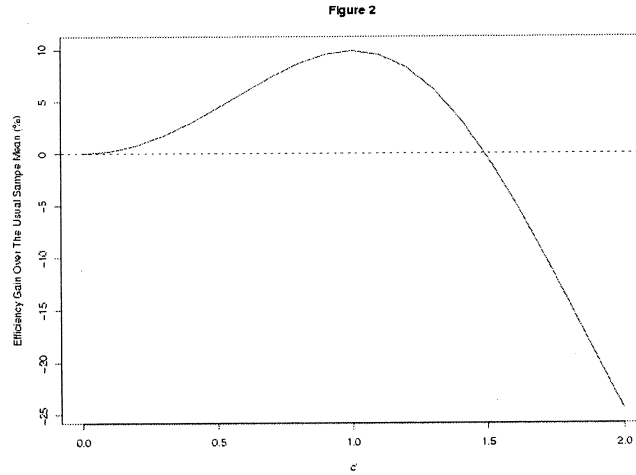


Figure 2: Efficiency gain of (3) over (7) when c in (4) is incorrectly specified by c' . The true value of $c = 1$, $N = 1000$ and n is 1% of N .

Substituting (4) into (3), we get the mean squared error of Z_{opt} as

$$E_R \left\{ E_S (Z_{opt} - M)^2 \right\} = \frac{(N - n)(N + c^2)}{N^2(n + c^2)} \sigma^2. \quad (6)$$

Setting $t_i = 1/n$ into (3), we have

$$E_R \{Var_S(\bar{Y})\} = \frac{N-n}{nN} \sigma^2. \quad (7)$$

Therefore, the relative efficiency of Z_{opt} over $\bar{Y}$ is

$$\frac{N(n+c^2)}{n(N+c^2)} \geq 1.$$

Hence, Z_{opt} is more efficient than the usual sample mean. The efficiency is 1 for $c^2 = 0$ and tends to N/n for c^2 tending to infinity. Table 1 calculates the efficiency gain in percentage for combinations of c and N when n is 1% of N . For example, if c is estimated correctly as $c = 1$ and $N = 1000$ then the efficiency gain of (5) over the usual sample mean is almost 10%. However, if c is incorrectly specified in (4) as c' , say, then the corresponding mean squared error in (3) will be greater than the minimized mean squared error (6). Figure 1 shows the efficiency loss for a range of values of c' when the true value of $c = 1$, $N = 1000$ and n is 1% of N . The efficiency loss for $c' = 0.75$ and $c' = 1.25$ are 1.70% and 2.34%, respectively, which are considerably small. But for $c' = 2$ the efficiency loss is a big 45.41%. Figure 2 checks to see whether (5) can remain more efficient than the usual sample mean even when c is incorrectly specified by c' . In the case considered it is evident that (5) remains more efficient only when $c' \leq 1.5$. The relative efficiency of Z_{opt} over Searls' estimator ($N = \infty$ case) is

$$\frac{N^2}{(N-n)(N+c^2)}. \quad (8)$$

Therefore, Z_{opt} can be more efficient than Searls' estimator if $n/N \geq c^2/(N+c^2)$, which can be expected to be satisfied since usually $n \gg c^2$. The efficiency gain would be of the order $n/(N-n)$ and can be substantial if, for example, the sample fraction n/N is close to one.

In the above we took the superpopulation approach for finding optimality because the population values Y_i were defined as a random sample from a superpopulation. Suppose instead that the population is a collection of N realizations, say $y_1, \dots, y_N$. Then – as described below – we could take a purely sampling based approach if we assume $t_i = w$, $i \in U$.

2.2 Sampling-based Approach

Consider the population mean

$$m = \sum_{i=1}^N y_i / N$$

and the class of estimators

$$z = w \sum_{i \in s} y_i.$$

Setting $t_i = w$ into equations (11)–(13) in appendix, the mean squared error

$$E_S(z-m)^2 = \frac{n \{n(N-1)m^2 + (N-n)\theta^2\}}{N-1} w^2 - 2nm^2w + m^2, \quad (9)$$

where

$$\theta = \sqrt{\sum_{i=1}^N (y_i - m)^2 / N}.$$

It attains its global minimum when

$$w = \frac{N-1}{n(N-1) + (N-n)d^2},$$

where $d = \theta/m$, and so the optimal estimator

$$z_{opt} = \frac{N-1}{n(N-1) + (N-n)d^2} \sum_{i \in s} y_i.$$

Substituting this back into (9), we have

$$E_S(z_{opt} - m)^2 = \frac{(N-n)\theta^2}{n(N-1) + (N-n)d^2}.$$

Setting $w = 1/n$ into (9), we obtain the mean squared error of $\bar{y} = \sum_{i \in s} y_i/n$ as

$$Var_S(\bar{y}) = \frac{N-n}{n(N-1)}\theta^2.$$

Therefore, the relative efficiency of z_{opt} over $\bar{y}$ is

$$1 + \frac{N-n}{n(N-1)}d^2 \geq 1.$$

Hence, as before, z_{opt} is more efficient than the usual sample mean: efficiency is 1 for $d^2 = 0$ and tends to infinity for d^2 tending to infinity. Furthermore, since the relative efficiency of z_{opt} over the $N = \infty$ case is

$$\frac{n(N-1) + (N-n)d^2}{(N-n)(n+d^2)} = \frac{1 - \frac{n}{N} \frac{1+d^2}{n+d^2}}{1 - \frac{n}{N}} \geq 1,$$

z_{opt} is also more efficient than Searls' estimator.

Appendix: Proof of Theorem 1

Let $S = \{s \subset U : \text{card}(s) = n\}$ denote the set of all subsets of U of size n . Every element $s \in S$ has a probability,

$$p(s) = 1/\binom{N}{n},$$

of being drawn without replacement. So, writing the mean squared error of Z as

$$E_S\{(Z - M)^2\} = E_S(Z^2) - 2ME_S(Z) + M^2, \quad (10)$$

we note that

$$\begin{aligned} E_S(Z^2) &= \sum_{s \in S} \left(\sum_{i \in s} t_i Y_i \right)^2 p(s) \\ &= \sum_{s \in S} \left(\sum_{i \in s} \sum_{j \in s} t_i t_j Y_i Y_j \right) \frac{1}{\binom{N}{n}} \end{aligned}$$

$$\begin{aligned}
&= \sum_{i \in U} \sum_{j \in U} t_i t_j Y_i Y_j \text{card}\{s : i, j \in s\} \frac{1}{\binom{N}{n}} \\
&= \sum_{i \in U} t_i^2 Y_i^2 \text{card}\{s : i \in s\} \frac{1}{\binom{N}{n}} \\
&\quad + \sum_{i \in U} \sum_{j \in U \setminus \{i\}} t_i t_j Y_i Y_j \text{card}\{s : i, j \in s\} \frac{1}{\binom{N}{n}} \\
&= \sum_{i \in U} t_i^2 Y_i^2 \frac{\binom{N-1}{n-1}}{\binom{N}{n}} + \sum_{i \in U} \sum_{j \in U \setminus \{i\}} t_i t_j Y_i Y_j \frac{\binom{N-2}{n-2}}{\binom{N}{n}} \\
&= \frac{n}{N} \sum_{i \in U} t_i^2 Y_i^2 + \frac{n(n-1)}{N(N-1)} \sum_{i \in U} \sum_{j \in U \setminus \{i\}} t_i t_j Y_i Y_j. \tag{11}
\end{aligned}$$

Similarly, note that

$$E_S(Z) = \frac{n}{N} \sum_{i \in U} t_i Y_i. \tag{12}$$

From (10), write

$$E_R \{E_S(Z - M)^2\} = E_R \{E_S(Z^2)\} - 2E_R \{ME_S(Z)\} + E_R(M^2), \tag{13}$$

and note using (11)-(12) that

$$E_R \{E_S(Z^2)\} = \frac{n}{N} (\sigma^2 + \mu^2) \sum_{i \in U} t_i^2 + \frac{n(n-1)}{N(N-1)} \mu^2 \sum_{i \in U} \sum_{j \in U \setminus \{i\}} t_i t_j, \tag{14}$$

$$-2E_R \{ME_S(Z)\} = -\frac{2n}{N^2} (\sigma^2 + N\mu^2) \sum_{i \in U} t_i, \tag{15}$$

$$E_R(M^2) = \frac{\sigma^2}{N} + \mu^2. \tag{16}$$

Combining (13)-(16), we obtain (3). To minimize (3), note that the system of equations

$$\frac{\partial E_R \{E_S(Z - M)^2\}}{\partial t_i} = 0, \quad i \in U$$

simplifies to

$$t_i = \frac{1}{\sigma^2 + \mu^2} \left\{ \frac{\sigma^2 + N\mu^2}{N} - \frac{(n-1)\mu^2}{N-1} (t - t_i) \right\}, \quad i \in U,$$

where $t = \sum_{i \in U} t_i$, and further to

$$t_i = \frac{(N-1)(\sigma^2 + N\mu^2) - N(n-1)\mu^2 t}{N(N-1)(\sigma^2 + \mu^2) - N(n-1)\mu^2}, \quad i \in U. \tag{17}$$

Now, we can compute t as

$$t = \sum_{i \in U} t_i = \frac{\sigma^2 + N\mu^2}{\sigma^2 + n\mu^2}.$$

Substituting this back into (17), we get

$$t_i = \frac{(\sigma^2 + N\mu^2) \left\{ N - 1 - \frac{N(n-1)\mu^2}{\sigma^2 + n\mu^2} \right\}}{N(N-1)\sigma^2 + N(N-n)\mu^2} = \frac{\sigma^2 + N\mu^2}{N(\sigma^2 + n\mu^2)}, \quad i \in U,$$

as required. Finally, note that these solutions correspond to an absolute minimum, i.e.

$$\frac{\partial^2 E_R \left\{ E_S (Z - M)^2 \right\}}{\partial t_i^2} = \frac{2n}{N} (\sigma^2 + \mu^2) > 0, \quad i \in U$$

and

$$\frac{\partial^2 E_R \left\{ E_S (Z - M)^2 \right\}}{\partial t_i \partial t_j} = \frac{n(n-1)}{N(N-1)} \mu^2 > 0, \quad i \neq j \in U.$$

The proof is complete.

References

- [1] Gleser, L.J. and Healy, J.D. (1976). Estimating the mean of a normal distribution with known coefficient of variation. *Journal of the American Statistical Association*, **71**, 977-981
- [2] Godambe, V.P. (1995). Estimation of parameters in survey sampling: optimality. *Canadian Journal of Statistics*, **23**, 227-243.
- [3] Khan, R. A. (1968). A note on estimating the mean of a normal distribution with known coefficient of variation. *Journal of the American Statistical Association*, **63**, 1039-1041.
- [4] Murthy, M. N. (1967). *Sampling Theory and Methods*. Calcutta: Statistical Publishing Society.
- [5] Searls, D. T. (1964). The utilization of a known coefficient of variation in estimation procedure. *Journal of the American Statistical Association*, **59**, 1225-1226.
- [6] Sen, A. R. (1978). Estimation of the population mean when the coefficient of variation is known. *Communications in Statistics-Theory and Methods*, **7**, 657-672.

Change-Point Model for a Sequence of Random Variables with Independent Increments

Dariusz Ghorbanzadeh

CNAM-Département Mathématiques, Paris, France

Rachid Lounes

Université René Descartes, Laboratoire de Statistique Médicale, Paris, France

H. de Arazoza

Departemeto Ecuaciones Diferenciales, Universidad de la Habana, Cuba

Abstract: In this paper, we consider a sequence of random variables with independent increments which can change their distribution at some unknown instant. We use this model with two different data sets.

Key Words: change-point, independent increments, Brownian Bridge, Kolmogorov test.

1 Introduction

In the change-point literature, various authors have proposed different tests for the detection of a change-point and in most of them the detection concerns the mean. For the case of Poisson variables we can cite the papers of Raftery & Akman (1986), Simpkin & Downham (1988), Henderson & Matthews (1993), Ghorbanzadeh & Lounes (1996) and West & Ogden (1997).

Raftery & Akman (1986) proposed, a bayesian approach to the retrospective analysis of a Poisson process with a single change-point at an unknown time. They derived

the posterior distribution for the parameters before and after the change time, and derived tests of the existence of a change-point. Simpkin & Downham (1988) have proposed a test of detection for Poisson random variables, based on maximum likelihood ratio statistic; assuming the change-point is fixed, they approximate the distribution of this statistic by a χ^2 distribution. Henderson & Matthews (1993) propose a classical model for Poisson random variables to detect changes in the annual number of cases of haemolytic uraemic syndrome (HUS). They studied not only at most one change but also multi-change. West and Ogden (1997) used the HUS data for the estimation problem of change-points in a sequence of Poisson random variables, they approached the problem by allowing the change to range over a continuous time interval $(0, T)$, and derived a bayesian-based interval estimator. Ghorbanzadeh & Lounes (1996) proposed two methods for the detection of change-point for grouped Poisson random variables. The first method is based on penalized maximum likelihood ratio statistic. The second one uses a bayesian approach on the posterior distribution of change-point parameter.

In this work, we consider a sequence of random variables $X_1, \dots, X_n$ with independent increments (i.e. $X_1, X_2 - X_1, \dots, X_n - X_{n-1}$ independent) with Poisson distributions. The independent increments $X_1, \dots, X_k - X_{k-1}$ follow a Poisson distribution $\mathcal{P}(\lambda)$ and the independent increments $X_{k+1} - X_k, \dots, X_n - X_{n-1}$ follow a Poisson distribution $\mathcal{P}(\lambda + \frac{\delta}{\sqrt{n}})$ where the parameters λ and δ are unknown. We propose the following test problem:

$$\begin{cases} H_0 : & \delta = 0 \\ H_1 : & \delta \neq 0 \end{cases} \quad (1)$$

2 Statistical Model

For testing the existence of the change-point we propose the statistic:

$$R_n(k) = \sqrt{n} \left(\frac{X_n - X_k}{n - k} - \frac{X_k}{k} \right) \quad (2)$$

This model is a particular case considered in Chapter 2 of [3] with a different methodology.

To study the behavior of R_n under the hypotheses H_0 and H_1 we have the following lemma:

Lemma 1 *Under H_0 , we have :*

$$(1) \quad X_i \sim \mathcal{P}(i\lambda)$$

$$(2) \quad \text{Cov}(X_i, X_j) = \min(i, j) : \lambda$$

Proof. Let $Y_1 = X_1$ and for $i = 2, \dots, n$ $Y_i = X_i - X_{i-1}$. From the fact that $X_1, \dots, X_n$ have independent increments, then the variables $Y_1, \dots, Y_n$ are independent and the variables $Y_1, \dots, Y_k \sim \mathcal{P}(\lambda)$ and $Y_{k+1}, \dots, Y_n \sim \mathcal{P}(\lambda + \frac{\delta}{\sqrt{n}})$. In the other

hand, we have :

$$\forall i = 1, \dots, n \quad X_i = \sum_{j=1}^i Y_j \quad (3)$$

and from this we deduce that

$$\text{Cov}(X_i, X_j) = \text{Cov} \left(\sum_{r=1}^{\min(i,j)} Y_r, \sum_{r=1}^{\min(i,j)} Y_r \right) = \min(i, j) \lambda. \quad (4)$$

From (4), we have the following lemma

Lemma 2 Under H_0 , and for some k ,

$$\begin{aligned} \mathbb{E}[R_n(k)] &= 0 \\ \forall k_1 \leq k_2 \quad \text{Cov}(R_n(k_1), R_n(k_2)) &= \frac{\lambda}{\frac{k_2}{n} (1 - \frac{k_1}{n})} \end{aligned} \quad (5)$$

Remark 1 From (4) we have: $\text{Var}[R_n(k)] = \frac{\lambda}{\frac{k}{n} (1 - \frac{k}{n})}$ which is not bounded for k near n . In the rest of our paper, we will penalize the statistic R_n so that the variance remains bounded for all values k (see [6]).

Let φ be a stepwise continuous function on $[0, 1]$ that satisfies $\int_0^1 \left(\frac{\varphi(t)}{\sqrt{t}} \right)^2 dt < \infty$ and

$$R_{n,\varphi}^*(k) = \varphi\left(\frac{k}{n}\right) \varphi\left(1 - \frac{k}{n}\right) R_n(k) \quad (6)$$

From (5) we have $\forall : k_1 \leq k_2$

$$\text{Cov}(R_{n,\varphi}^*(k_1), R_{n,\varphi}^*(k_2)) = \frac{\varphi\left(\frac{k_1}{n}\right) \varphi\left(1 - \frac{k_1}{n}\right) \varphi\left(\frac{k_2}{n}\right) \varphi\left(1 - \frac{k_2}{n}\right) \lambda}{\frac{k_2}{n} (1 - \frac{k_1}{n})} \quad (7)$$

and

$$\text{Var}(R_{n,\varphi}^*(k)) = \left(\frac{\varphi\left(\frac{k}{n}\right) \varphi\left(1 - \frac{k}{n}\right)}{\sqrt{\frac{k}{n} (1 - \frac{k}{n})}} \right)^2 \lambda$$

We note that if we take $\varphi(t) = t$, (6) can be written in the form:

$$\mathbb{K}_n(k) = n^{-3/2} (k X_n - n X_k). \quad (8)$$

For any consistent estimator $\hat{\lambda}_n$ de λ , let

$$\mathbb{K}_n^*(k) = \frac{1}{\sqrt{\hat{\lambda}_n}} : \mathbb{K}_n(k).$$

We define $\{\mathbb{K}_n^*(t); t \in [0, 1]\}$ as the linear interpolated process of the process $\{\mathbb{K}_n^*(k); 1 < k < n\}$.

Theorem 1 Under H_0 , $\mathbb{K}_n^*(t)$ converges in the sense of weak convergence of the finite dimensional distributions to the Brownian Bridge $B(t) = W(t) - tW(1)$ where $\{W(t); t \in [0, 1]\}$ represent the Wiener process.

Proof. We can rewrite $\mathbb{K}_n(t)$ as difference of two independent asymptotically gaussian processes. In the fact we have

$$\mathbb{K}_n(t) = \frac{[nt](n - [nt])}{n^{3/2}} \left(\frac{1}{n - [nt]} \sum_{i=[nt]+1}^n Y_i - \frac{1}{[nt]} \sum_{i=1}^{[nt]} Y_i \right)$$

where $Y_1, \dots, Y_n$ are defined in the proof of lemma 1. By using (7) we have

$$\forall t_1 \leq t_2 \quad \text{cov}(\mathbb{K}_n(t_1), \mathbb{K}_n(t_2)) = \frac{[nt_1]}{n} \left(1 - \frac{[nt_2]}{n}\right) : \lambda \quad (9)$$

Then

$$\lim_{n \rightarrow \infty} \text{Cov}(\mathbb{K}_n^*(t_1), \mathbb{K}_n^*(t_2)) = t_1 (1 - t_2)$$

Now, for all $t_1 \leq t_2$, we have $\text{Cov}(B(t_1), B(t_2)) = t_1 (1 - t_2)$ and also, $\hat{\lambda}_n \xrightarrow[n \rightarrow \infty]{P} \lambda$ ■

We have then the following corollaries.

Corollary 1 Under H_0 , and for all $\eta \geq 0$, we have :

$$\lim_{n \rightarrow \infty} P_{H_0} \left(\sup_{t \in [0,1]} |\mathbb{K}_n^*(t)| \geq \eta \right) = P \left(\sup_{t \in [0,1]} |B(t)| \geq \eta \right) \quad (10)$$

Proof of corollary 1. Using the continuity of the function “sup”, (cf. [2]), on space $\mathcal{C}([0, 1])$.

Corollary 2 Let $\hat{\tau}_n = \text{Arg} \sup_{t \in [0,1]} |\mathbb{K}_n^*(t)|$ and $\tau = \text{Arg} \sup_{t \in [0,1]} |B(t)|$. Under H_0 , we have

$$\hat{\tau}_n \xrightarrow[n \rightarrow \infty]{P} \tau \quad (11)$$

Proof of corollary 2. The proof is the same as the Lemma 2.1 of [5] .

For the behavior under the hypothesis H_1 of $\mathbb{K}_n^*$, we have the following results.

Lemma 3 Under H_1 , and for all $i = 1, \dots, n$ we have:

$$\mathbb{K}[X_i] = \begin{cases} i \lambda & \text{if } 1 \leq i \leq k \\ k \lambda + (i - k) \left(\lambda + \frac{\delta}{\sqrt{n}} \right) & \text{if } k + 1 \leq i \leq n \end{cases} \quad (12)$$

Proof. Using (3), writing, X_i in the following form :

$$X_i = \begin{cases} \sum_{j=1}^i Y_j & \text{if } 1 \leq i \leq k \\ \sum_{j=1}^k Y_j + \sum_{j=k+1}^i Y_j & \text{if } k + 1 \leq i \leq n \end{cases}$$

Lemma 4 Under H_1 , and for all k_1 et k_2 we have :

$$\text{Cov}(X_{k_1}, X_{k_2}) = \begin{cases} k \lambda + (\min(k_1, k_2) - k) \left(\lambda + \frac{\delta}{\sqrt{n}} \right) & \text{if } \min(k_1, k_2) \geq k \\ \min(k_1, k_2) \lambda & \text{otherwise} \end{cases} \quad (13)$$

Proof. It is enough to note that from (3), for $k \leq k_1 \leq k_2$, we have :

$$X_{k_1} = \sum_{j=1}^k Y_j + \sum_{j=k+1}^{k_1} Y_j \quad \text{et} \quad X_{k_2} = \sum_{j=1}^k Y_j + \sum_{j=k+1}^{k_1} Y_j + \sum_{j=k_1+1}^{k_2} Y_j.$$

Theorem 2 Under H_1 , $\mathbb{K}_n^*$ converges in the sense of weak convergence of the finite dimensional distributions to the process B_τ^* defined by:

$$B_\tau^*(t) = (t \wedge \tau) (1 - t \vee \tau) \frac{\delta}{\sqrt{\lambda}} + B(t) \quad (14)$$

where $t \wedge \tau = \inf\{t, \tau\}$ and $t \vee \tau = \sup\{t, \tau\}$.

Proof. The same argument as in the proof of theorem 1 shows that $\mathbb{K}_n(t)$ converge to gaussian process. From (12), for all $t \in [0, 1]$ we have :

$$\mathbb{E}[\mathbb{K}_n(t)] = \begin{cases} \frac{[nt]}{n} (1 - \frac{[n\tau]}{n}) \delta & \text{if } t \leq \tau \\ \frac{[n\tau]}{n} (1 - \frac{[nt]}{n}) \delta & \text{if } t \geq \tau \end{cases} \quad (15)$$

Let $t_1 \leq t_2$. By ((13)), if : $[nt_1] \leq [nt_2] \leq [n\tau]$:

$$\text{Cov}(\mathbb{K}_n(t_1), \mathbb{K}_n(t_2)) = \frac{[nt_1]}{n} (1 - \frac{[nt_2]}{n}) \lambda + \frac{[nt_1]}{n} \frac{[nt_2]}{n} (1 - \frac{[n\tau]}{n}) \frac{\delta}{\sqrt{n}}$$

If : $[nt_1] \leq [n\tau] \leq [nt_2]$

$$\text{Cov}(\mathbb{K}_n(t_1), \mathbb{K}_n(t_2)) = \frac{[nt_1]}{n} (1 - \frac{[nt_2]}{n}) \lambda + \frac{[nt_1]}{n} (1 - \frac{[nt_2]}{n}) \frac{[n\tau]}{n} \frac{\delta}{\sqrt{n}}$$

If : $[n\tau] \leq [nt_1] \leq [nt_2]$

$$\begin{aligned} \text{Cov}(\mathbb{K}_n(t_1), \mathbb{K}_n(t_2)) &= \frac{[nt_1]}{n} (1 - \frac{[nt_2]}{n}) \lambda \\ &+ (1 - \frac{[nt_2]}{n}) \left(\frac{[nt_1]}{n} \frac{[n\tau]}{n} + \frac{[nt_1]}{n} - \frac{[n\tau]}{n} \right) \frac{\delta}{\sqrt{n}} \end{aligned}$$

Hence

$$\lim_{n \rightarrow \infty} \text{Cov}(\mathbb{K}_n(t_1), \mathbb{K}_n(t_2)) = t_1 (1 - t_2) : \lambda \quad \blacksquare$$

3 Building the test

To build the test, we take as the rejection region

$$\left\{ \sup_{t \in [0, 1]} |\mathbb{K}_n^*(t)| \geq \eta \right\}$$

For a fixed α in $[0, 1]$, by (10), the level η_α is given by (see [9]) :

$$P\left(\sup_{t \in [0,1]} |B(t)| \geq \eta_\alpha\right) = 2 \sum_{j=1}^{\infty} (-1)^{j+1} e^{-2j^2 \eta_\alpha^2} \quad (16)$$

It is interesting to note that the proposed test has the same asymptotic critical value as the Kolmogorov test.

By (14) we obtain the power of the test:

$$p(\tau) = P\left(\sup_{t \in [0,1]} |B_\tau^*(t)| \geq \eta_\alpha\right) \quad (17)$$

In the following applications, we estimate the power of the test by building the percentiles table of (14) after replacing δ and τ by their estimate. This construction is based on Monte Carlo simulation of the brownian bridge, for this see [4] .

4 Applications and Discussion

4.1 HSU case

In this section we apply the previous results to the data used by [7] and [11] . The data represent the number of haemolytic uraemic syndrome (HUS) cases in Birmingham and Newcastle from 1970 – 1989.

In [7] the following model was used:

$$Y_i \sim \begin{cases} \mathcal{P}(\lambda_1) & \text{if } i = 1, \dots, k \\ \mathcal{P}(\lambda_2) & \text{if } i = k + 1, \dots, n \end{cases}$$

where Y_i denotes the number of HUS cases for the year i and $\mathcal{P}(\lambda)$ is the Poisson distribution with mean λ . By using the Log-likelihood statistic, for the at most one change point case they found that the change occurs at $k = 11$ for the Birmingham data and $k = 15$ for the Newcastle data. In [11] a Poisson process was used to suppress the hypothesis that the change is an integer. They applied this to the Newcastle data to obtain (maximum likelihood method), the value $\tau = 14.944$ which closely corresponds with the estimate of 15 given by [7] .

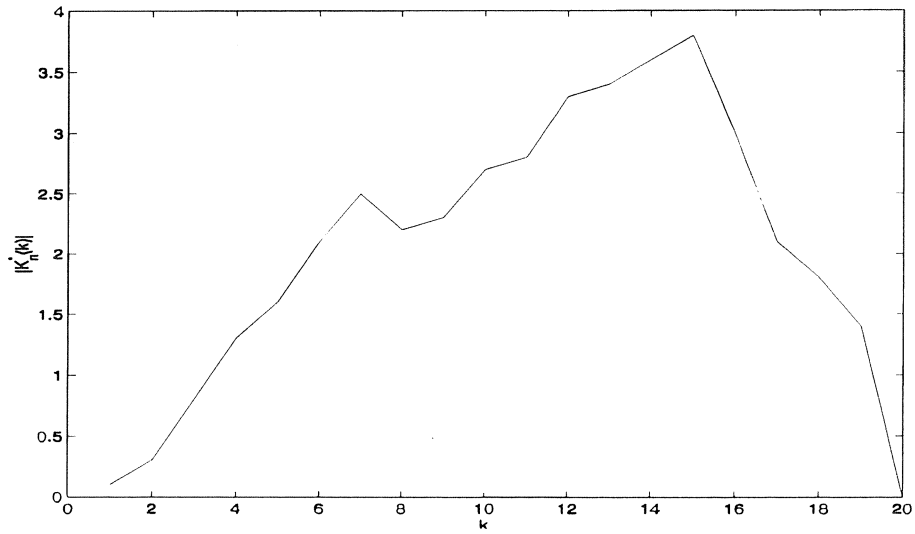
We have applied the model (1) by taking $X_i = Y_1 + \dots + Y_i$, where Y_i is the same as in [7] .

For the Newcastle data, we have $\hat{\lambda}_n = \frac{x_n}{n} = 5$.

The maximum is obtained at $k = 15$, corresponding to the value of the statistic $|\mathbb{K}_n^*(15)| = 3.8$. For $\alpha = 5\%$, critical value is $\eta_\alpha = 1.36$, then we reject H_0 . From (8), we obtain $\hat{\delta} = |\mathbb{K}_n(15)| = 8.50$, which gives an order of 21% for the estimation of the power.

k	y_k	x_k	$ \mathbb{K}_n^*(k) $	k	y_k	x_k	$ \mathbb{K}_n^*(k) $
1	6	6	0.1	11	4	27	2.8
2	1	7	0.3	12	0	27	3.3
3	0	7	0.8	13	4	31	3.4
4	0	7	1.3	14	3	34	3.6
5	2	9	1.6	15	3	37	3.8
6	0	9	2.1	16	13	50	3
7	1	10	2.5	17	14	64	2.1
8	8	18	2.2	18	8	72	1.8
9	4	22	2.3	19	9	81	1.4
10	1	23	2.7	20	19	100	0

Table 1. Newcastle results

Figure 1: Graph of $|\mathbb{K}_n^*(k)|$ for the Newcastle data .

For the Birmingham data, we have $\hat{\lambda}_n = \frac{x_n}{n} = 5.65$.

k	y_k	x_k	$ \mathbb{K}_n^*(k) $	k	y_k	x_k	$ \mathbb{K}_n^*(k) $
1	1	1	0.43744	11	4	18	4.1533
2	5	6	0.49858	12	0	25	4.0263
3	3	9	0.74787	13	4	36	3.523
4	2	11	1.0912	14	3	40	3.6782
5	2	13	1.4346	15	3	47	3.5512
6	1	14	1.872	16	13	57	3.142
7	0	14	2.4035	17	14	73	2.1684
8	0	14	2.935	18	8	89	1.1947
9	2	16	3.2784	19	9	98	0.87957
10	1	17	3.7158	20	19	113	0

Table 2. Birmingham results

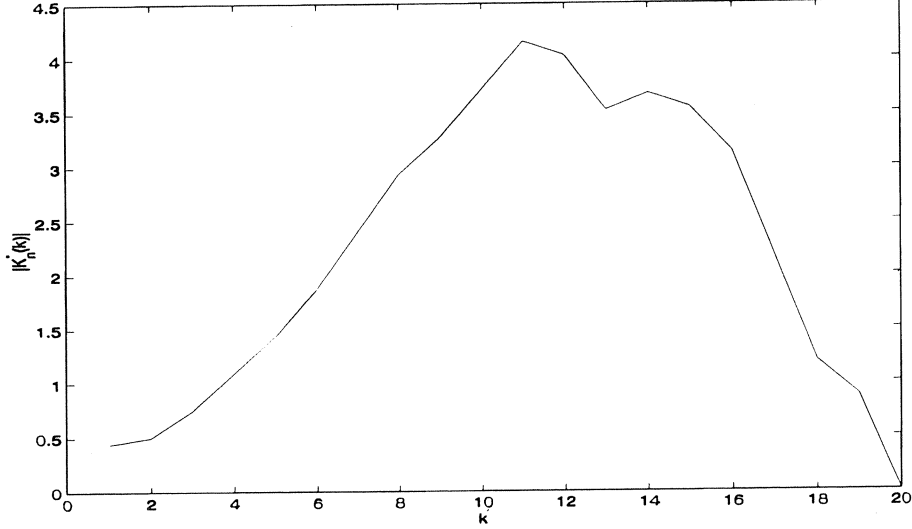


Figure 2: Graph of $|\mathbb{K}_n^*(k)|$ for the Birmingham data .

The maximum is obtained at $k = 11$, corresponding to the value of the statistic $|\mathbb{K}_n^*(11)| = 4.15$. For $\alpha = 5\%$, critical value is $\eta_\alpha = 1.36$, then we reject H_0 . From (8), we obtain $\hat{\delta} = |\mathbb{K}_n(11)| = 9.87$, which gives an order of 41% for the estimation of the power.

We obtain the same point estimations for the AMOC case as in [7] and [11] .

4.2 AIDS case

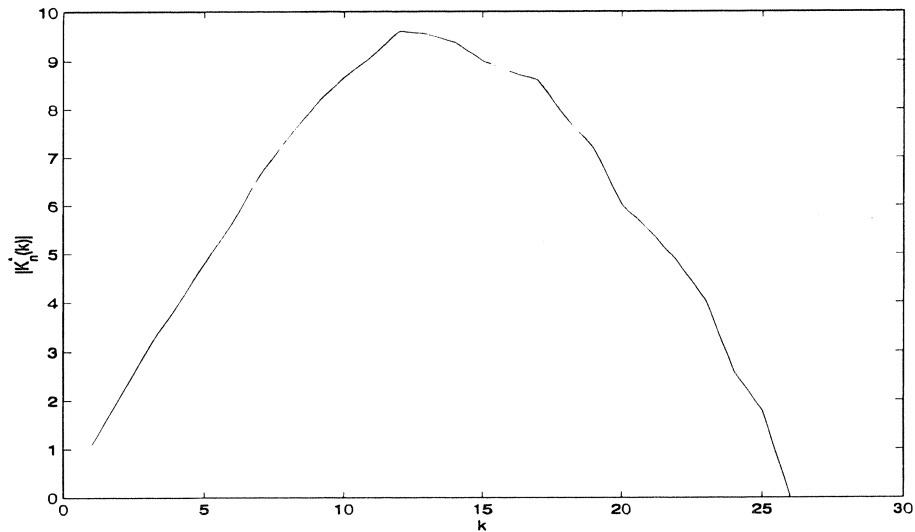
The Cuban case is an interesting one because the epidemiological information is reliable, due to the existence of a centralized system, with a National Reference Center that validates and confirms the new AIDS cases diagnosed, at regular intervals. Also not only AIDS but also HIV infection is under a mandatory report directive since 1985. This means that a high percentage of the HIV infected persons are known to the Health System, and have special medical care that makes the report of the dates of AIDS. In the case of the AIDS epidemic in Cuba, we will look at the number of new AIDS cases reported by semester since the start of the epidemic in 1986 [12] . If we take the new AIDS cases in Cuba by semester, following [1] , the data can be regarded as independent Poisson random variables.

For this data, we have $\hat{\lambda}_n = \frac{x_n}{n} = 32.15$.

The maximum is obtained at $k = 12$, corresponding to the value of the statistic $|\mathbb{K}_n^*(12)| = 9.61$. For $\alpha = 5\%$, critical value is $\eta_\alpha = 1.36$, then we reject H_0 . From (8), we obtain $\hat{\delta} = |\mathbb{K}_n(12)| = 54.49$, which gives an order of 98% for the estimation of the power.

sem-year	k	y_k	x_k	$ \mathbb{K}_n^*(k) $	sem-year	k	y_k	x_k	$ \mathbb{K}_n^*(k) $
1st 1986	1	1	1	1.0775	2nd 1992	14	37	179	9.3781
2nd 1986	2	4	5	2.0512	1st 1993	15	43	222	9.0029
1st 1987	3	3	8	3.0595	2nd 1993	16	39	261	8.7662
2nd 1987	4	8	16	3.8949	1st 1994	17	37	298	8.5985
1st 1988	5	6	22	4.7994	2nd 1994	18	55	353	7.8084
2nd 1988	6	8	30	5.6348	1st 1995	19	50	403	7.1912
1st 1989	7	3	33	6.6431	2nd 1995	20	66	469	6.0206
2nd 1989	8	10	43	7.4093	1st 1996	21	48	517	5.4725
1st 1990	9	12	55	8.1064	2nd 1996	22	51	568	4.8207
2nd 1990	10	16	71	8.6651	1st 1997	23	55	623	4.0306
1st 1991	11	20	91	9.0854	2nd 1997	24	74	697	2.5833
2nd 1991	12	17	108	9.6095	1st 1998	25	55	752	1.7931
1st 1992	13	34	142	9.5457	2nd 1998	26	84	836	0

Table 3. Cuban AIDS results

Figure 3: Graph of $|\mathbb{K}_n^*(k)|$ for the Cuban AIDS data.

We find that there is a change-point that corresponds to the second semester of 1991. This is consistent with what is known of the AIDS epidemic in Cuba. There were some persons infected before 1981-82, but the transmission of the HIV virus started to establish itself in the period 1982-86. Considering that the average incubation period was estimated at 9.2 years, an increase of the number of new AIDS cases was to be expected at some time in the period 1991-92. This is confirmed by our analysis of the data.

Taking into account the power, the model we propose in this paper, is best adapted to the AIDS data than to the HSU data in the sense that in the AIDS data case the jump is larger than in the two other cases.

References

- [1] Bailey, N. (1995). Operational modelling of HIV/AIDS to assist public health control. In *Epidemic Models: Their Structure and Relation to Data* (ed. D. Mollison 404–415), Cambridge University Press.
- [2] Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
- [3] Csörgö, M. and Horváth, L. (1997). *Limit theorems in Change Point Analysis*. J. Wiley, New York.
- [4] Eastwood, B. J. and Eastwood, V. R. (1990). On the distribution of some gaussian functionals related to changepoint problems. *Technical Report 146*, Laboratory for Research in Statistics and Probability, Carleton University, Ottawa.
- [5] Ferger, D. (1994). Change-point estimators in case of small disorders. *J. Statist. Plann. Inference*, 40, 33–49.
- [6] Ghorbanzadeh, D. and Lounes, R. (1996). Deux procédures de détection de rupture pour des observations poissonniennes groupées. *Revue de Statistique Appliquée*, XLIV(3), 47–61.
- [7] Henderson, R. and Matthews, J. N. S. (1993). An Investigation of Changepoints in the Annual number of Cases of Haemolytic Uraemic Syndrome. *Appl. Statist.* 42 3. 461–471.
- [8] Raftery, A.E. and Akman, V.E. (1986). Bayesian analysis of a Poisson process with a change-point. *Biometrika*.73, 85–89.
- [9] Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes With Applications to Statistics*. J. Wiley, New York.
- [10] Simpkin, J.M. and Downham, D.Y. (1988). Testing for a change-point in registry data with an example on Hypospadias. *Statistics in Medicine*. 7 387–393.
- [11] West, R. W. and Ogden, R. T. (1997). Continuous-time estimation of a change-point in a Poisson process. *Journ. Statist. Comput. Simul.* 56 293–302.
- [12] The HIV-AIDS Epidemic in Cuba. Internal Report. Departamento de Epidemiología, Sanatorio Santiago de Las Vegas. Cuba 1999.

An Alternative Derivation of The Noncentral Chi-square Distribution

W.K. Huang and W. K. Pang

Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong

Abstract : The noncentral chi-square distribution is an important distribution in statistics. Unfortunately, understanding the derivation of the noncentral chi-square distribution is not easy. This is because the derivation of the noncentral chi-square distribution heavily depends on a theorem of which the proofs are usually profound and the readers have to spend a lot of time on it. The aim of this paper is to point out this theorem and give an elegant proof of it.

Key Words : Noncentral chi-square distribution; Convolution; uniform convergence.

1 Introduction

The noncentral chi-square distribution is an important sampling distribution. It has been studied by Fisher (1928), Wishart (1932), Patnaik (1949), Tiku (1965), Han (1975) Guenther (1964) and McNolty (1962). In many advanced textbooks in mathematical statistics, we often find the derivation of the noncentral chi-square distribution. In the book by Biswas (1991), the author showed a detailed derivation of the distribution (pp 517-522). In the following we list the major results:

Result I: Let $X \sim N(\mu, \sigma^2)$. The distribution of $\frac{X^2}{\sigma^2}$ follows a noncentral chi-square distribution with one degree of freedom and noncentrality parameter $\lambda = \frac{\mu^2}{\sigma^2}$. (See pp 517-519 of Biswas (1991) for detailed derivation.)

Result II: Let $X_1 \sim N(\mu, \sigma^2)$ and $X_2, \dots, X_n$ be a sample observations of size (n-1) drawn from $N(0, \sigma^2)$. If $X_1, X_2, \dots, X_n$ are independent, then

$$\chi^2 = \frac{X_1^2 + X_2^2 + \dots + X_n^2}{\sigma^2}$$

has a noncentral chi-square distribution with n degrees of freedom and noncentrality parameter $\lambda = \frac{\mu^2}{\sigma^2}$.

Result III: The distribution of $\sum_{i=1}^n \frac{X_i^2}{\sigma_i^2}$ has a noncentral chi-square distribution with n degrees of freedom and noncentrality parameter $\lambda = \sum_{i=1}^n \frac{\mu_i^2}{\sigma_i^2}$ where $X_i \sim N(\mu_i, \sigma_i^2)$, $1 \leq i \leq n$, and all X_i are independent. (See pp 519-521 of Biswas (1991) for detailed derivation.)

The proof of result I is easy. Result II is only a special case of result III, but result III can also be regarded as a consequence of result II. In the book by Biswas (1991), result III is proved using result II. Of course, we can use the moment generating function to deal with the noncentral chi-square distribution as shown in the book by Graybill (1976) or we can use the characteristic function to deal with it as shown in the book by Biswas (1991). Then we can prove results concerning independent random variables having noncentral chi-square distributions without using their density functions.

Many different proofs of result II are rather cumbersome and profound. Here we present an elegant alternative derivation.

2 Alternative Derivation

Result II is an immediate consequence of the following theorem.

Theorem 1

Suppose U and V are independent random variables, $U \sim \chi_{1, \lambda}^2$, and $V \sim \chi_n^2$, then the random variable $Z = U + V$ follows a noncentral chi-square distribution with $n+1$ degrees of freedom and noncentrality parameter λ .

Proof.

Theorem 1 is in fact a simple consequence of Theorem 2 below.

Theorem 2

Suppose U and V are independent random variables where $U \sim \chi_{k, \lambda}^2$, and $V \sim \chi_{n, \delta}^2$, then $U + V \sim \chi_{k+n, \lambda+\delta}^2$.

Proof.

Let $\mathbf{Y}$ be a $k \times 1$ multivariate normal random vector with mean μ and covariance matrix $\mathbf{I}$ where $\mu' = (\sqrt{\lambda}, 0, \dots, 0)$. Let $\mathbf{Z}$ be a $n \times 1$ multivariate normal random vector with mean η and covariance matrix $\mathbf{I}$ where $\eta' = (\sqrt{\delta}, 0, \dots, 0)$. Assume $\mathbf{Y}$ and $\mathbf{Z}$ are independent. Then $\mathbf{Y}'\mathbf{Y} \sim \chi_{k, \mu'\mu}^2$ and $\mathbf{Z}'\mathbf{Z} \sim \chi_{n, \eta'\eta}^2$ (see pp 125-126 of Graybill (1976)). That is, $\mathbf{Y}'\mathbf{Y} \sim \chi_{k, \lambda}^2$ and $\mathbf{Z}'\mathbf{Z} \sim \chi_{n, \delta}^2$.

If we consider the random vector $\begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix}$ which has the multivariate normal distribution with mean $\begin{pmatrix} \mu \\ \eta \end{pmatrix}$ and covariance matrix $\mathbf{I}$, then we have $\mathbf{Y}'\mathbf{Y} + \mathbf{Z}'\mathbf{Z} \sim \chi_{k+n, \mu'\mu + \eta'\eta}^2$. That is, $\mathbf{Y}'\mathbf{Y} + \mathbf{Z}'\mathbf{Z} \sim \chi_{k+n, \lambda+\delta}^2$. Since U and V have the same dis-

tribution as $\mathbf{Y}'\mathbf{Y}$ and $\mathbf{Z}'\mathbf{Z}$, we conclude that $U + V$ and $\mathbf{Y}'\mathbf{Y} + \mathbf{Z}'\mathbf{Z}$ have the same distribution so that $U + V \sim \chi_{k+n, \lambda+\delta}^2$. Theorem 2 follows immediately.

We can now justify our assertion that Theorem 1 is a simple consequence of Theorem 2 by putting $k = 1$ and $\delta = 0$ so that $U \sim \chi_{1, \lambda}^2$, $V \sim \chi_n^2$ and $U + V \sim \chi_{n+1, \lambda}^2$.

For those who like to construct density functions directly, we would like to point out that the density function of $U + V$ can be obtained using the density function of U given by

$$f(u) = \sum_{r=0}^{\infty} \frac{\lambda^r e^{-\lambda}}{r!} \frac{e^{-\frac{u}{2}} u^{r-\frac{1}{2}}}{\Gamma(r + \frac{1}{2}) 2^{r+\frac{1}{2}}},$$

where $u > 0$, the density function of V , and the formula

$$f_{U+V}(a) = \int_{-\infty}^{\infty} f_U(a-v) f_V(v) dv$$

where $a > 0$.

Résumé : La distribution Chi-deux non centrée est une distribution importante en statistiques. Malheureusement, comprendre comment on déduit cette distribution n'est pas facile, puisque cela dépend d'un théorème dont la preuve est ardue et les lecteurs doivent y consacrer beaucoup de temps. L'objectif de cet article est de discuter de ce théorème et d'en présenter une preuve élégante.

Mots clés : Distribution Chi-deux non centrée; Convolution; uniforme convergence.

Acknowledgement: We are grateful to an anonymous referee for helpful comments and suggestions.

References

- [1] Biswas, J. H. (1991). *Topics in Statistical Methodology*. Wiley Eastern Limited, New Delhi.
- [2] Fisher, R. A. (1928). The general sampling distribution of the multiple correlation coefficient. *Proc. Roy. Soc., A*, **121**, 654-673.
- [3] Han, C. P. (1975). Some relationships between non-central chi-square and normal distributions. *Biometrika*, **62**, 213-214.
- [4] Graybill, F. A. (1976). *Theory and Application of the Linear Model*. Duxbury Press.
- [5] Guenther, W. C. (1964). Another derivation of the non-central chi-square distribution. *J. Amer. Statist. Ass.*, **59**, 957-960.
- [6] McNolty, F. (1962). A contour integral derivation of the non-central χ^2 -distribution. *Annals of Math. Statist.*, **33**, 796-800.
- [7] Patnaik, R. B. (1949). The non-central χ^2 and F -distributions and their applica-

- tions. *Biometrika*, **36**, 202-232.
- [8] Tiku, M. L. (1965). Laguerre series forms of non-central χ^2 and F -distributions. *Biometrika*, **52**, 415-426.
- [9] Wishart, J. (1932). A note on the distribution of the correlation ratio. *Biometrika*, **24**, 441-447.

Adaptive Regression : A Simulation Study

Christophe A. Beuret

Statistics Group, University of Neuchâtel, Switzerland

Valentin Rousson

Department of Biostatistics, University of Zürich, Switzerland

Abstract : The purpose of adaptive methods is to select from a class of estimates the one which is the most appropriate for a given sample, according to the underlying distribution. Harter's adaptive method was proposed for the location model and uses sample kurtosis as selection statistics. For large samples, the method allows to select the maximum likelihood estimate (MLE) of the mean for populations distributed as Uniform, Normal or Laplace. In this paper, we extend Harter's method to the linear regression model. According to the kurtosis of some initial residuals, our method selects either the least squares estimate (MLE when the errors are Normal) or the LAD estimate (MLE when the errors are Laplace). In all cases considered, a simulation study shows that our method provides 95% of good results when the sample size is larger than 200. Moreover, the method is not sensitive to the choice of the initial residuals. We also propose another related adaptive procedure based on hypothesis testing. In particular, we tabulate critical values to test whether the errors in a regression model are distributed as Normal or Laplace.

Key Words : Kurtosis, adaptive method, maximum likelihood estimates, least squares, least absolute deviation.

1 Introduction

Suppose that a given sample was drawn from one of the distributions $F_1, F_2, \dots, F_k$. For each of these distributions F_j , we would have an optimal estimate t_j ($j = 1, \dots, k$) for a parameter T of interest. In practice, however, we do not know the true distribution, and hence we do not know which estimate to use. The purpose of adaptive methods is to select the appropriate estimate based on the information contained in the sample.

Harter's adaptive method (1974) [3] was proposed for the location model (see also Harter et al., 1979 [4]). The goal was to estimate the mean μ from data

$$y_i = \mu + \varepsilon_i \quad (1)$$

($i = 1, \dots, n$), where the ε_i were independent and identically distributed with mean zero, but where the distribution of the ε_i (and hence the y_i) was not known. Harter's procedure allows to select the maximum likelihood estimate (MLE) of μ when the distribution is Uniform, Normal or Laplace. To achieve this, Harter's method uses the sample kurtosis

$$k = \frac{n \sum (y_i - \bar{y})^4}{\left(\sum (y_i - \bar{y})^2 \right)^2}$$

where

$$\bar{y} = \frac{\sum y_i}{n}.$$

For symmetric distributions, recall that kurtosis may be interpreted as a scale-invariant measure of tails, such that the higher the kurtosis, the longer the tails of the distribution (see e.g. Darlington, 1970 [1] ; Hogg, 1972 [5]). In particular, kurtosis is equal to 1.8, 3 and 6, for Uniform, Normal and Laplace distributions, respectively. The method of Harter works as follows:

- (1) If $k < 2.2$, consider the sample to be distributed Uniform and use the midrange estimate

$$\hat{\mu} = \frac{x_{(1)} + x_{(n)}}{2},$$

where $x_{(1)}$ and $x_{(n)}$ denote the smallest and the largest observation in the sample, respectively.

- (2) If $2.2 \leq k \leq 3.8$, consider the sample to be distributed Normal and use the sample mean as estimate of μ .
- (3) If $k > 3.8$, consider the sample to be distributed Laplace and use the sample median as estimate of μ .

The limits 2.2 and 3.8 were determined by simulation and were chosen such that the proportion of Normal samples with $k < 2.2$ was equal to the proportion of Uniform samples with $k > 2.2$, and such that the proportion of Normal samples with $k > 3.8$ was equal to the proportion of Laplace samples with $k < 3.8$ (in average for various sample sizes).

In this paper, we propose to extend Harter's procedure to the linear regression model. Our goal is to estimate a p -dimensional vector of parameters β in the model

$$y_i = X\beta + \varepsilon_i, \quad (2)$$

where X is a $n \times p$ matrix of covariates. If X contains just one column with all entries equal to 1, then (2) is equivalent to (1). Recall that MLEs for Uniform, Normal and Laplace errors are given by the minimax, the least squares (LS), and the least absolute

deviations (LAD) estimate, respectively (see e.g. Jackson, 1924 [6]). Our method works as follows:

- (1) We first use one of these estimates of β (e.g. the least squares estimate) in order to get initial residuals.
- (2) We calculate the kurtosis k of these initial residuals.
- (3) As final estimate of β , we select the minimax estimate if $k < 2.2$, the least squares estimate if $2.2 < k < 3.8$ or the LAD estimate if $k > 3.8$.

Note that our method does not depend on the variance of the errors since kurtosis is scale-invariant. Also, as we shall see in the next section, the choice of the initial estimate does not affect much the performance of the method. As a consequence, we propose to calculate the initial residuals using the least squares estimate. This has the advantage to make our method similar to Harter's criterion for the location model (since the kurtosis k of the y_i is calculated with respect to the sample mean $\bar{y}$ which is the least squares estimate of μ in the location model).

In Section 2, we investigate the performance of our method under various situations via a simulation study. In Section 3, we propose another related adaptive procedure based on hypothesis testing. In Section 4, we consider the case where the errors are "between" a Normal and a Laplace distribution. Finally, some concluding remarks take place in Section 5.

2 Simulations

We simulated data from three models, namely

$$\begin{aligned} y_{1i} &= 3 + \varepsilon_i \\ y_{2i} &= 3 + 5x_{1i} + \varepsilon_i \\ y_{3i} &= 3 + 5x_{1i} - 6x_{2i} + \varepsilon_i, \end{aligned}$$

where the errors ε_i were Uniform, Normal or Laplace distributed with mean zero. The first model had only $p = 1$ parameter to be estimated and corresponded to a location model, whereas the next two models were regression models with $p = 2$ and $p = 3$ parameters, respectively.

For each of the three models and each of the three error distributions considered, we generated 10'000 replications of size $n = 10, 20, 50, 100, 200$ and 500 . For a given sample size n , we used the same values of x_{i1} and x_{i2} ($i = 1, \dots, n$) for each of the 10'000 replications, which were roughly uniformly distributed between -1 and 1 . On each of these samples, we applied our method described in Section 1, using the least squares residuals as initial residuals. Table 1 provides the proportion of times that our method selected the correct estimate (i.e. the MLE), as well as the proportion of

times that our method selected the least squares estimate (i.e. the MLE for Normal errors) for each case considered.

Table 1: Proportion of times that our method selected the MLE and the least squares estimate from 10'000 samples

Error distribution	n	MLE			LS estimate		
		$p = 1$	$p = 2$	$p = 3$	$p = 1$	$p = 2$	$p = 3$
Uniform	10	0.7406	0.6566	0.6033	0.2441	0.3232	0.3734
	20	0.8291	0.7531	0.6811	0.1693	0.2433	0.3126
	50	0.9580	0.9271	0.8916	0.0420	0.0729	0.1084
	100	0.9943	0.9900	0.9841	0.0057	0.0100	0.0159
	200	0.9999	0.9999	0.9998	0.0001	0.0001	0.0002
	500	1.0000	1.0000	1.0000	0	0	0
Normal	10	0.4888	0.4851	0.4885	Same as left		
	20	0.6591	0.6695	0.6680			
	50	0.8573	0.8571	0.8571			
	100	0.9447	0.9441	0.9456			
	200	0.9824	0.9823	0.9830			
	500	0.9977	0.9977	0.9977			
Laplace	10	0.2153	0.1745	0.1139	0.5448	0.5382	0.5457
	20	0.4186	0.3825	0.3432	0.5324	0.5557	0.5840
	50	0.6709	0.6465	0.6276	0.3280	0.3523	0.3710
	100	0.8419	0.8330	0.8227	0.1581	0.1670	0.1773
	200	0.9589	0.9559	0.9541	0.0411	0.0441	0.0459
	500	0.9988	0.9987	0.9987	0.0012	0.0013	0.0013

For Uniform and Laplace errors, we see that the performance of our adaptive procedure decreased when the number of parameters in the model increased. For Normal errors, however, the performance of our method did not depend on p . For $n = 100$ and $p = 3$, our method selected the correct estimate in about 98%, 95% and 82% of the cases for Uniform, Normal and Laplace errors, respectively. In the latter case, we should increase the sample size to $n = 200$ in order to get 95% of correct results.

We next investigated the influence of the choice of the initial residuals on the performance of our method. For this, we used the following algorithm (for a given sample):

- (1) We first calculated the kurtosis k_0 of the y_i .
- (2) We selected the estimate E_0 according to Harter's criterion based on k_0 .
- (3) We calculated the kurtosis k_1 of the residuals obtained by performing E_0 .
- (4) We selected the estimate E_1 according to Harter's criterion based on k_1 .
- (5) We calculated the kurtosis k_2 of the residuals obtained by performing E_1 .

(6) We selected the estimate E_2 according to Harter's criterion based on k_2 .

We wanted to know the percentage of contradictions which arises between E_0 and E_1 and between E_1 and E_2 . For this, we generated 1000 samples of size 50, 100, 200 and 500 according to a regression model (2) with 2 parameters, and with either Normal or Laplace errors. We did not consider Uniform errors since the corresponding MLE requires a complicated algorithm. Thus, in Steps 2, 4 and 6 of our algorithm, we selected the least squares estimate when the kurtosis was smaller than 3.8, and the LAD estimate when the kurtosis was larger than 3.8. Table 2 summarizes the results obtained.

Table 2 : Number of contradictions out of 1000 samples

Error distribution	n	Number of contradictions between	
		E_0 and E_1	E_1 and E_2
Normal	50	74	4
	100	39	1
	200	15	1
	500	0	0
Laplace	50	643	2
	100	818	3
	200	955	1
	500	1000	0

For Normal errors, we see that the number of contradictions between E_0 and E_1 decreased with the sample size. For Laplace errors, however, the opposite was true. In particular, for $n = 500$, the kurtosis of the y_i was always smaller than 3.8, such that we had 100% of contradictions between E_0 and E_1 . This shows that the residuals calculated with respect to the constant fit $\bar{y}$ were not adequate as initial residuals. The reason for this was that the kurtosis k_0 of such residuals was not a good estimate of the kurtosis of the true errors.

On the other hand, for both Normal and Laplace errors, the number of contradictions between E_1 and E_2 was negligible (less than 0.5% in each case). This indicates that our method will provide mostly the same outcome whether we use LS residuals or LAD residuals as initial residuals.

3 Adaptive procedure based on hypothesis testing

In this section, we propose an adaptive procedure based on hypothesis testing. We are interested in two different (although related) tests. The first one tests the null hypothesis

$$H_0 : \varepsilon_i \sim \text{a Normal distribution}$$

versus the alternative

$$H_1 : \varepsilon_i \sim \text{a more leptokurtic distribution.}$$

The second one tests the null hypothesis

$$H'_0 : \varepsilon_i \sim \text{a Laplace distribution}$$

versus the alternative

$$H'_1 : \varepsilon_i \sim \text{a more platikurtic distribution.}$$

Then, our strategy will be the following. If H_0 is accepted while H'_0 is rejected, the errors will be considered as Normal and we shall use the least squares estimate for the vector of parameters β in the regression model (2). Similarly, if H'_0 is accepted while H_0 is rejected, we shall consider the errors to be distributed as Laplace and we shall use the LAD estimate.

Compared to Harter's procedure, this method has the advantage of being related to a confidence level (such as a 95% confidence level). A serious disadvantage, however, is that we may have situations where both hypotheses are accepted, or where both hypotheses are rejected. In the former case, this means that the sample is not large enough to allow to distinguish between a Normal and a Laplace distribution. In the latter case, this means that the error distribution is anywhere "between" a Normal and a Laplace distribution.

We propose to test both hypotheses H_0 and H'_0 using the kurtosis k of some appropriate residuals, e.g. the least squares residuals. The null hypothesis H_0 will be rejected if k is too large. Thus, we should determine the values $k_{n,p,1-\alpha}^N$ for which

$$\Pr\{k > k_{n,p,1-\alpha}^N\} = \alpha$$

where k is the kurtosis of the least squares residuals calculated from a sample of size n drawn from a regression model (2) with p parameters and Normal errors. Therefore, we shall reject H_0 at the α -level of significance if $k > k_{n,p,1-\alpha}^N$.

Similarly, the null hypothesis H'_0 will be rejected if the kurtosis k of the least squares residuals is too small. Thus, we should determine the values $k_{n,p,\alpha}^L$ for which

$$\Pr\{k < k_{n,p,\alpha}^L\} = \alpha$$

where k is the kurtosis of the least squares residuals calculated from a sample of size n drawn from a regression model (2) with p parameters and Laplace errors. Therefore, we shall reject H'_0 at the α -level of significance if $k < k_{n,p,\alpha}^L$.

Note that Snedecor and Cochran (1980) [8] tabulated critical values of $k_{n,1,1-\alpha}^N$ for $\alpha = 0.01$ and $\alpha = 0.05$, which were based on Pearson's (1954) [7] tables. Values of $k_{n,p,0.95}^N$ and $k_{n,p,0.05}^L$ for $p = 1, 2, 3$ and various sample sizes n between 10 and 1000 are given in Table 3. These values were obtained from 10'000 replications.

Values $k_{n,1,0.95}^N$ and $k_{n,1,0.05}^L$ are plotted in Figure 1 in function of n (using '+' and 'o' symbols, respectively). An horizontal line at the value 3.8 has been added to the plot to allow a comparison with Harter's procedure. Our adaptive procedure works here as follows:

- (1) If $k < k_{n,p,0.05}^L$ and $k < k_{n,p,0.95}^N$, we consider the errors to be Normal and we select the LS estimate.

- (2) If $k_{n,p,0.95}^N < k$ and $k_{n,p,0.05}^L < k$, we consider the errors to be Laplace and we select the LAD estimate.
- (3) If $k_{n,p,0.05}^L < k < k_{n,p,0.95}^N$, both the Normal and the Laplace hypotheses are accepted and we do not know which estimate to use of the LS and the LAD.
- (4) If $k_{n,p,0.95}^N < k < k_{n,p,0.05}^L$, both the Normal and the Laplace hypotheses are rejected and we should use another kind of estimate than the LS and the LAD.

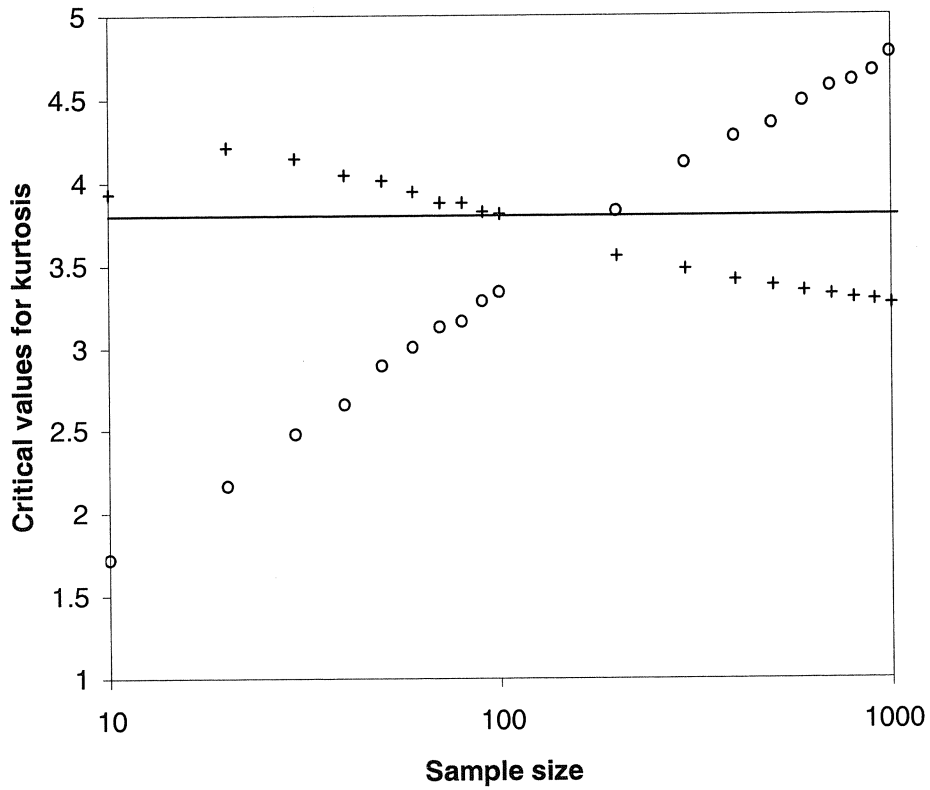


Figure 1: Critical values $k_{n,1,0.95}^N$ (+) and $k_{n,1,0.05}^L$ (o) in function of sample size. The horizontal line refers to the value 3.8 used by Harter's criterion.

We note that $k_{n,1,0.95}^N = k_{n,1,0.05}^L \approx 3.62$ for a sample size n between 100 and 200 (using linear interpolation we found $n = 162$). For $p = 2$ and $p = 3$, we had approximately the same result ($n = 167$ and $k = 3.65$ for $p = 2$ and $n = 168$ and $k = 3.69$ for $p = 3$). As this value $k = 3.62$ is very close to the value 3.8 used by Harter's procedure, this means that the adaptive procedure described in this section will be mostly equivalent to Harter's criterion, if we are in one of the two favourable cases 1 or 2 listed here above. For sample sizes smaller than 162, the third case listed here above may happen, whereas the fourth case will never happen. Similarly, for

sample sizes larger than 162, the fourth case listed here above may happen, whereas the third case will never happen. Note that a sample size $n = 162$ is very close to the value $n = 200$ for which Harter's criterion is known to perform well (see our simulations results in Section 2, as well as Sposito et al., 1983 [9] ; Sposito, 1990 [10]).

Table 3 : Critical values of sample kurtosis under Normal and Laplace errors

	$p = 1$		$p = 2$		$p = 3$	
n	$k_{n,p,0.95}^N$	$k_{n,p,0.05}^L$	$k_{n,p,0.95}^N$	$k_{n,p,0.05}^L$	$k_{n,p,0.95}^N$	$k_{n,p,0.05}^L$
10	3.931	1.728	3.843	1.666	3.952	1.624
20	4.214	2.170	4.121	2.141	4.268	2.055
30	4.137	2.485	4.164	2.456	4.065	2.403
40	4.043	2.655	4.038	2.649	4.017	2.595
50	4.003	2.897	4.000	2.821	3.961	2.819
60	3.937	3.009	3.954	2.974	3.880	2.928
70	3.879	3.123	3.922	3.068	3.869	2.996
80	3.873	3.167	3.833	3.168	3.817	3.160
90	3.820	3.282	3.794	3.248	3.829	3.264
100	3.803	3.339	3.747	3.333	3.671	3.386
200	3.550	3.834	3.596	3.800	3.560	3.839
300	3.474	4.119	3.469	4.072	3.509	4.127
400	3.407	4.276	3.404	4.250	3.429	4.208
500	3.375	4.355	3.367	4.374	3.406	4.414
600	3.339	4.484	3.340	4.485	3.337	4.460
700	3.317	4.582	3.312	4.562	3.328	4.544
800	3.296	4.611	3.289	4.608	3.275	4.606
900	3.289	4.661	3.273	4.687	3.280	4.659
1000	3.261	4.780	3.252	4.778	3.258	4.716

4 Linear combination of LS and LAD

In this section, we consider regression models (2) where the law of the errors ε_i is a convolution of a Normal and a Laplace distribution. Thus, we consider errors

$$\varepsilon_i = (1 - \gamma)e_{1i} + \gamma e_{2i} \quad (3)$$

with $0 \leq \gamma \leq 1$ and where the e_{1i} and e_{2i} are mutually independent and distributed as Normal and Laplace, respectively (both with mean zero). For $\gamma = 0$ we get Normal errors, whereas $\gamma = 1$ defines Laplace errors. For $0 < \gamma < 1$, we get an error distribution which is “between” Normal and Laplace. Such an error distribution would be a possible candidate in cases where the hypotheses of Normal and Laplace errors are both rejected (see the last section).

For such an error distribution (3), one may conjecture that a linear combination of the LS estimate $\hat{\beta}_{LS}$ and the LAD estimate $\hat{\beta}_{LAD}$

$$\hat{\beta}(\delta) = (1 - \delta)\hat{\beta}_{LS} + \delta\hat{\beta}_{LAD},$$

for some $0 \leq \delta \leq 1$ which depends on γ , might be an adequate estimate of β (for example with $\delta = \gamma$). If such a conjecture were true, one may think of using the kurtosis k of the least squares residuals in order to estimate γ (and hence to select the appropriate δ). Note that estimates $\hat{\beta}(\delta)$ were proposed and studied by Dodge and Jurečková (2000) [2] .

When the sample size is “large enough”, we saw in Section 2 that the percentage of times π that k was larger than 3.8 was close to zero for Normal errors (i.e. when $\gamma = 0$), and close to one for Laplace errors (i.e. when $\gamma = 1$). We are here interested in the proportion π when $0 < \gamma < 1$. If π would be e.g. linearly related to γ , one could then estimate π using an appropriate bootstrap procedure and use it to estimate γ (as long as a “large enough” sample would be available).

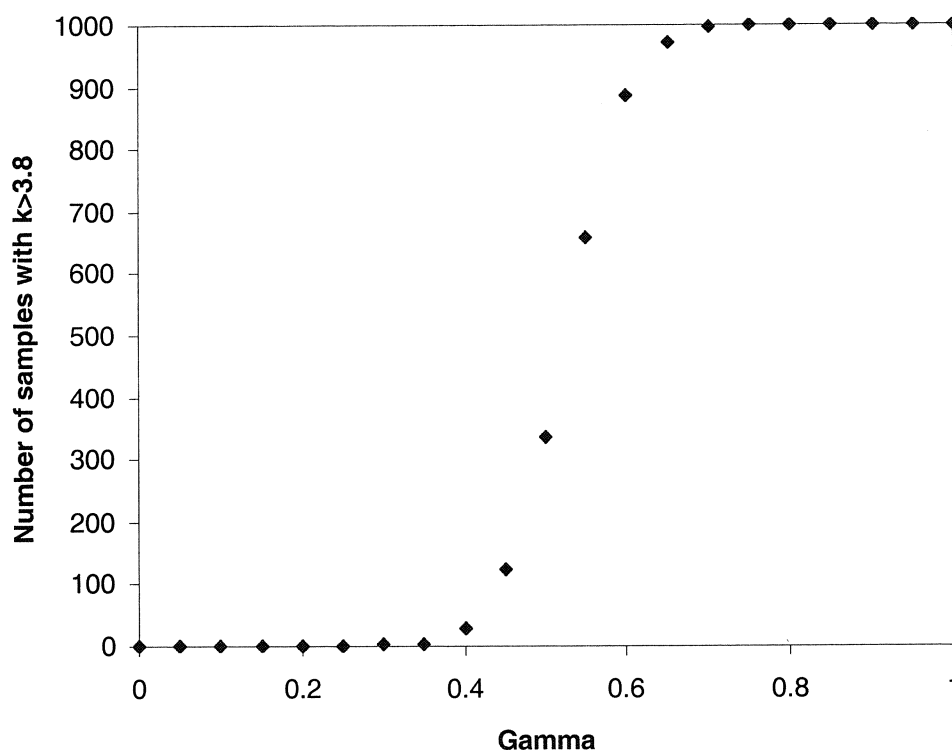


Figure 2: Number of times (out of 1000 samples of size 1000) that the kurtosis of the residuals was larger than 3.8 when the errors were distributed as (3) with various values of γ .

In order to know how this percentage π is related to γ , we performed a small simulation study. We generated 1000 samples of size $n = 1000$ (this should be “large enough”) according to a regression model (2) with 2 parameters, with errors distributed as (3), and with various values of γ between 0 and 1. In each case, we noted the

percentage of times π that the kurtosis k of the least squares residuals was larger than 3.8. Results are reported in Table 4 and illustrated in Figure 2.

We see that π was close to zero for any value of $\gamma < 0.4$, and close to one for any value of $\gamma > 0.6$. Thus, the value of π provided very few information on γ , such that the idea suggested in the last paragraph is not applicable.

Table 4 : Relation between γ and the number of times that $k > 3.8$ out of 1000 samples

γ	$k > 3.8$	γ	$k > 3.8$	γ	$k > 3.8$
0.00	0	0.35	4	0.70	998
0.05	0	0.40	29	0.75	1000
0.10	0	0.45	122	0.80	1000
0.15	0	0.50	336	0.85	1000
0.20	0	0.55	659	0.90	1000
0.25	1	0.60	886	0.95	1000
0.30	2	0.65	970	1.00	1000

5 Conclusions

In this paper, we extended Harter's criterion from the location model to a linear regression model with more than one parameter. Our simulations have shown that the method performed similarly well whatever the number of parameters in the model when the errors were Normal. For Uniform and Laplace errors, the performance of Harter's adaptive procedure decreased when the number of parameters increased. For a model with three parameters, the method performed well in each case considered when the sample size was larger than 200. We then proposed another adaptive procedure based on hypothesis testing which was roughly equivalent to Harter's procedure in the favourable cases. However, this method may reveal cases where none of the LS and the LAD estimates are adequate. In these cases, a solution could be to use a linear combination of the LS and the LAD estimate. Unfortunately, our attempt to select such an estimate using the sample kurtosis did not succeed.

Acknowledgments: We would like to thank Professor Yadolah Dodge who initiated this work as Christophe Beuret's Master Thesis and encouraged us to write this paper.

References

- [1] Darlington, R.B. (1970). Is Kurtosis Really Peakedness ? *The American Statistician*, **24**, 19–22
- [2] Dodge, Y. and Jurečková, J. (2000). *Adaptive linear regression*. Springer, New York.
- [3] Harter, H.L. (1974). The Method of Least Squares and Some Alternatives. *Inter-*

- national Statistical Review*, **42**, 147–174, 235–264, 282 ; **43** (1975) 1–44, 125–190, 269–278 ; **44** (1976) 113–159.
- [4] Harter, H.L., Moore, A.H. and Curry, T.F. (1979). Adaptive Robust Estimation of Location and Scale Parameters of Symmetric Populations. *Communications in Statistics*, **A8** (15), 1473–1491.
 - [5] Hogg, R.V. (1972). More Light on Kurtosis and Related Statistics. *Journal of the American Statistical Association*, **72**, No 338, 422–424.
 - [6] Jackson, D. (1924). On the Method of Least m th Powers for a Set of Simultaneous Equations. *Annals of Mathematics*, serie 2, **25**, 185–192
 - [7] Pearson, E.S. and Hartley, H.O. (1954). *Biometrika Tables for Statisticians*, Vol. 1, Cambridge Univ. Press.
 - [8] Snedecor, G.W. and Cochran, W.G. (1980). *Statistical Methods*. 7th ed., The Iowa Univ. Press, Ames Iowa.
 - [9] Sposito, V.A., Hand, M.L. and Skarpness, B. (1983). On the efficiency of using the sample kurtosis in selecting optimal L_p estimators. *Communications in Statistics*, **12**, 265–272.
 - [10] Sposito, V.A. (1990). Some Properties of L_p -Estimators, In *Robust regression : analysis and applications*. Editor K.D. Lawrence, pp. 23–58, Dekker, New York.

A Note on the Behrens-Fisher Problem for Elliptically Symmetric Distributions

L. Thabane

St. Francis Xavier University, Department of Mathematics, Statistics & Computer Science, Nova Scotia, CANADA

M. Safiul Haq

The University of Western Ontario, Department of Statistical & Actuarial Sciences, Ontario, CANADA

Abstract: In this note, we derive the posterior distribution of the location parameters for dependent (but uncorrelated) samples from a class of elliptically symmetric distributions and show that the result is the same for all members of this class including the normal case.

Key Words: Behrens-Fisher problem, Bayesian approach, elliptical class, multivariate-t distributions.

1 Introduction

A p -dimensional random variable $\mathbf{X} = (X_1, \dots, X_p)'$ is said to have a distribution belonging to the class of *elliptically symmetric distributions* with location parameter $\boldsymbol{\mu}$ and scale matrix Σ (positive definite) if the probability density function (pdf) of $\mathbf{X}$ has the form

$$f(\mathbf{x}) = C_1 |\Sigma|^{-\frac{1}{2}} q((\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})), \quad -\infty < \mathbf{x} < \infty, \quad -\infty < \boldsymbol{\mu} < \infty, \quad (1)$$

where C_1 is the normalizing constant and q is a function on $[0, \infty)$. We shall denote the family (1) by $E_p(\boldsymbol{\mu}, \Sigma)$. This family includes among others the multivariate normal, multivariate-t, compound multivariate normal and Pearson type VII distributions. The family ((1)) is commonly viewed as a generalization of the multivariate normal distribution. In this paper, we follow the Bayesian approach to derive the solution of the Behrens-Fisher problem for the class of elliptically symmetric distributions, and

we show that the original solution obtained under the normality assumption remains valid for any member of this class.

The Behrens-Fisher problem involves comparing the means (or location parameters) of two populations with unknown and unequal variances (or scale parameters) on the basis of independent samples. As the name suggests, the problem was first studied by Behrens (1929) [1] and Fisher (1935) [4]. In the multivariate (p dimensional) setup the problem can be formulated as follows: suppose we have two independent samples $\mathbf{y}_1 = (\mathbf{y}_{11}, \dots, \mathbf{y}_{1n_1})$ and $\mathbf{y}_2 = (\mathbf{y}_{21}, \dots, \mathbf{y}_{2n_2})$ from p -dimensional normal populations $N_p(\boldsymbol{\mu}_1, \Sigma_1)$ and $N_p(\boldsymbol{\mu}_2, \Sigma_2)$, respectively. The means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are unknown and the covariance matrices Σ_1 and Σ_2 are also unknown and not equal. The sample sizes n_1 and n_2 may be equal or unequal. The problem involves making inferences about the difference $\boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$. Equivalently, the problem may be stated in terms of testing the hypothesis

$$\begin{aligned} H_0 : \boldsymbol{\mu}_1 &= \boldsymbol{\mu}_2 = \boldsymbol{\mu} \text{ versus} \\ H_1 : \boldsymbol{\mu}_1 &\neq \boldsymbol{\mu}_2. \end{aligned} \quad (2)$$

This problem is commonly known as the *multivariate Behrens-Fisher problem*.

The object of this note is to derive the solution of the problem assuming that the samples come from the distributions of the form (1). The discussion is entirely within the Bayesian framework. In section 2, we consider a k -sample generalization of the Behrens-Fisher problem under the model (1).

1.1 The Probability Model For Samples

Following Sutradhar (1990) [8] we can express the joint density of n p -dimensional observations from an elliptical model $E_p(\boldsymbol{\mu}, \Sigma)$ as

$$f(\mathbf{y}_1, \dots, \mathbf{y}_n | \boldsymbol{\mu}, \Sigma) = C_1^* |\Sigma|^{-\frac{n}{2}} q \left(\sum_{j=1}^n (\mathbf{y}_j - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{y}_j - \boldsymbol{\mu}) \right), \quad (3)$$

where C_1^* is the normalizing constant. Applying successive integrations and using Lemma 2.1 of Sutradhar and Ali (1989) [9], one can easily show that the marginal distribution of $\mathbf{Y}_j$ is of the form (1). That is, for $j = 1, 2, \dots, n$, $\mathbf{Y}_j \sim E_p(\boldsymbol{\mu}, \Sigma)$. It is to be noted that in this scenario, the $\mathbf{Y}_j$'s are pair-wise uncorrelated, but not necessarily independent.

The above model can be extended easily to k uncorrelated samples from different elliptical distributions $E_p(\boldsymbol{\mu}_i, \Sigma_i)$, $i = 1, 2, \dots, k$ as follows: Suppose we let $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_k)$ represent the k samples under consideration, where $\mathbf{Y}_i = (\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{in_i})$ with $\mathbf{Y}_{ij}$ as a p -dimensional column vector, $j = 1, 2, \dots, n_i$ and $i = 1, 2, \dots, k$. Let $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_k)$, $\Sigma = (\Sigma_1, \dots, \Sigma_k)$ and $\Sigma^{-1} = (\Sigma_1^{-1}, \dots, \Sigma_k^{-1})$. Analogous to (3), we assume that the density of $\mathbf{Y}$ is given by

$$f(\mathbf{y} | \boldsymbol{\mu}, \Sigma) = C_2 \prod_{i=1}^k |\Sigma_i|^{-\frac{n_i}{2}} q \left(\sum_{i=1}^k \sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i)' \Sigma_i^{-1} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i) \right), \quad (4)$$

where C_2 is the normalizing constant. From the above argument, it follows immediately that

$$\mathbf{Y}_{ij} \sim E_p(\boldsymbol{\mu}_i, \Sigma_i), \quad j = 1, 2, \dots, n_i; \quad i = 1, 2, \dots, k,$$

and that the samples are pair-wise uncorrelated, but not necessarily independent.

2 K-Sample Generalization of the Behrens-Fisher Problem

We shall consider the problem in a general form: Suppose, we have k uncorrelated samples whose joint density is given by (4). Consider the problem of testing the null hypothesis

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \dots = \boldsymbol{\mu}_k, \quad \text{versus}$$

$$H_1 : \text{At least two are different}$$

where the scale matrices Σ_i 's are assumed to be unequal.

Following Geisser and Cornfield (1963) [6], and Geisser (1964) [5] among others, we assume the non-informative prior,

$$p(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1}) \propto \prod_{i=1}^k |\Sigma_i|^{-\frac{p+1}{2}}. \quad (5)$$

The joint posterior density of $(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1})$ is obtained by combining (4) and (5):

$$\begin{aligned} p(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1} | \mathbf{y}) &= C_3 \prod_{i=1}^k |\Sigma_i^{-1}|^{\frac{n_i-p-1}{2}} q \left(\sum_{i=1}^k \sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i)' \Sigma_i^{-1} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i) \right) \\ &= C_3 \prod_{i=1}^k |\Sigma_i^{-1}|^{\frac{n_i-p-1}{2}} q \left(\sum_{i=1}^k \text{tr} \Sigma_i^{-1} V_i \right), \end{aligned} \quad (6)$$

where

$$\begin{aligned} V_i &= (n_i - p)S_i + n_i(\boldsymbol{\mu}_i - \bar{\mathbf{y}}_i)(\boldsymbol{\mu}_i - \bar{\mathbf{y}}_i)', \quad \bar{\mathbf{y}}_i = \sum_{j=1}^{n_i} \mathbf{y}_{ij} / n_i \\ (n_i - p)S_i &= \sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)(\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)', \quad n_i > p \quad \text{and} \end{aligned}$$

C_3 is the normalizing constant so that (6) is a density with respect to $(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1})$.

From (6), the marginal posterior density of $\boldsymbol{\mu}$ is obtained as

$$\begin{aligned} p(\boldsymbol{\mu} | \mathbf{y}) &= \int_{\Sigma_1^{-1} > 0} \dots \int_{\Sigma_k^{-1} > 0} p(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1} | \mathbf{y}) d\boldsymbol{\Sigma}^{-1} \\ &= C_3 \int_{\Sigma_1^{-1} > 0} \dots \int_{\Sigma_k^{-1} > 0} \prod_{i=1}^k |\Sigma_i^{-1}|^{\frac{n_i-p-1}{2}} q \left(\sum_{i=1}^k \text{tr} \Sigma_i^{-1} V_i \right) \prod_{i=1}^k d\Sigma_i^{-1}. \end{aligned} \quad (7)$$

To perform, the integration with respect to Σ_i^{-1} in (7), we use theorem 2.1 of Sutradhar and Ali (1989) [9]. From equation (2.2) of the theorem, we have that for any Wishart matrix A based on a random sample of size n from model (1)

$$\int_{A>0} |A|^{\frac{n-p-1}{2}} q(tr \Sigma^{-1} A) dA \propto |\Sigma|^{\frac{n}{2}}. \quad (8)$$

Using (8) and equation 2.1 of theorem 2.1 of Sutradhar and Ali (1989) [9], we obtain from (7) that for $i = 1, 2, \dots, k$,

$$\int_{\Sigma_i^{-1}>0} |\Sigma_i^{-1}|^{\frac{n_i-p-1}{2}} q \left(tr \Sigma_i^{-1} V_i + \sum_{j \neq i}^k tr \Sigma_j^{-1} V_j \right) d\Sigma_i^{-1} \propto |V_i|^{-\frac{n_i}{2}} q_{k-1} \left(\sum_{j \neq i}^k tr \Sigma_j^{-1} V_j \right), \quad (9)$$

where $q_{k-1}(\cdot)$ is determined by the form of q and the number of terms in the sum $\sum_{j \neq i}^k tr \Sigma_j^{-1} V_j$. Continuing with the integration term by term in (7), we obtain

$$p(\boldsymbol{\mu}|\mathbf{y}) = C_4 \prod_{i=1}^k |V_i|^{-\frac{n_i}{2}} = C_4 \prod_{i=1}^k \left| (n_i - p) S_i + n_i (\boldsymbol{\mu}_i - \bar{\mathbf{y}}_i) (\boldsymbol{\mu}_i - \bar{\mathbf{y}}_i)' \right|^{-\frac{n_i}{2}}, \quad (10)$$

where C_4 is the normalizing constant. Note that (10) can be written as

$$p(\boldsymbol{\mu}|\mathbf{y}) = \prod_{i=1}^k p(\boldsymbol{\mu}_i|\mathbf{y}_i), \quad (11)$$

where

$$p(\boldsymbol{\mu}_i|\mathbf{y}_i) = \frac{\Gamma(\frac{n_i}{2}) |\frac{S_i}{n_i}|^{-\frac{1}{2}}}{\Gamma(\frac{n_i-p}{2}) [\pi(n_i-p)]^{\frac{p}{2}}} \left\{ 1 + \frac{n_i}{n_i-p} (\boldsymbol{\mu}_i - \bar{\mathbf{y}}_i)' S_i^{-1} (\boldsymbol{\mu}_i - \bar{\mathbf{y}}_i) \right\}^{-\frac{n_i}{2}}.$$

That is, $p(\boldsymbol{\mu}|\mathbf{y})$ is the product of k independent p -dimensional multivariate-t posterior densities each with $n_i - p$ degrees of freedom, location parameter $\bar{\mathbf{y}}_i$ and scale matrix S_i/n_i , which we shall denote by $t_p(\bar{\mathbf{y}}_i, \frac{S_i}{n_i}, n_i - p)$, $i = 1, 2, \dots, k$. We note here that (11) is the exact solution obtained under the assumption of normality for the k independent samples, assuming a vague prior for the parameters (cf. Dickey (1968) [3]). One can very easily verify this result for the various members of the family (1). The properties of this distribution have been studied by Dickey (1967 [2], 1968 [3]). In particular, theorem 2 of Dickey (1968) [3] gives the distribution of any linear function

$$\boldsymbol{\delta} = \sum_{j=1}^k B_j \boldsymbol{\mu}_j.$$

For the special case $k = 2, p = 1$, equation (11) gives a solution to the Behrens-Fisher problem for two samples. It is to be noted that the original solution for the problem as given by Behrens (1929) [1] (see also Fisher (1935) [4] for the fiducial solution), is presented as the pdf of the difference $\delta = \mu_1 - \mu_2$ given the observed data.

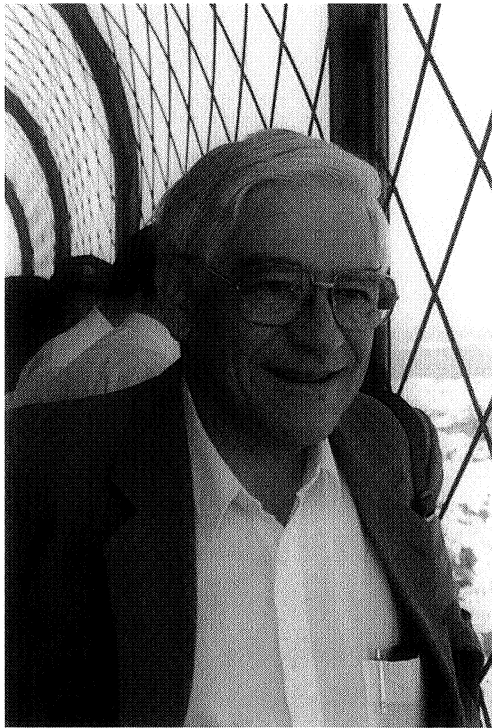
It is to be noted that we have considered the case of dependent, but uncorrelated samples from the elliptical class. The independent samples case is often quite difficult to handle; see for example Komuzi (1994) [7]. However as mentioned earlier the posterior distribution of the location parameters is independent.

Acknowledgements: This research was partially funded by the Natural Sciences and Engineering Research Council of Canada. The first author is grateful to the Canadian Commonwealth and Fellowship Committee for funding the research and his studies at the University of Western Ontario in a form of scholarship.

References

- [1] Behrens, W.V. (1929). Ein Beitrag zur Fehlerberechnung bei wenigen Beobachtungen. *Landwirtschaftliche Jahrensberichte*, **68**, 807–837.
- [2] Dickey, J.M. (1967). Matric-variate Generalizations of the Multivariate-t distribution and the inverted Multivariate-t distribution, *Ann. Math. Stat.* **38**, 511–518.
- [3] Dickey, J.M. (1968). Three Multidimensional-Integral Identities with Bayesian Applications, *Ann. Math. Stat.* **39**, No. 5, 1615–1627.
- [4] Fisher, R.A. (1935). The fiducial argument in statistical inference. *Ann. of Eugenics*, **6**, 391–398.
- [5] Geisser, S. (1964). Posterior odds for multivariate normal classification. *JRSS B*(1), 69–76.
- [6] Geisser, S. and Cornfield, J. (1963). Posterior distributions for multivariate normal parameters. *JRSS B*(25), 368–376.
- [7] Komuzi, H. (1994). Testing Equality of the Means in two Independent Multivariate-t Distributions, *Commun. Statist.* **23**(1), 215–227.
- [8] Sutradhar, B.C. (1990). Discrimination of Observations into one of two t Populations, *Biometrics*, **46**, 827–835.
- [9] Sutradhar, B.C. and Ali, M.M. (1989). A Generalization of the Wishart Distribution for the Elliptical Model and Its Moments for the Multivariate t Model, *Journal of Multi. Anal.*, **29**, 155–162.

Peter J. Huber: Founder of Robust Statistics



Professor Huber was born in Wohlen, Switzerland, on March 25, 1934. He obtained his Ph.D. from the Swiss Federal Institute of Technology in Zurich where he was awarded the silver medal for the scientific quality of his thesis. In 1964 he was appointed Professor of Mathematical Statistics at the Swiss Federal Institute of Technology in Zurich. He later went to the United States' most prestigious universities in the capacity of visiting professor at Princeton, Yale, Cornell, Harvard and Berkeley. In 1977, he was appointed Professor of Statistics at Harvard University, then Professor of Applied Mathematics at the Massachusetts Institute of Technology, and finally Professor of Statistics at the University of Bayreuth in Germany.

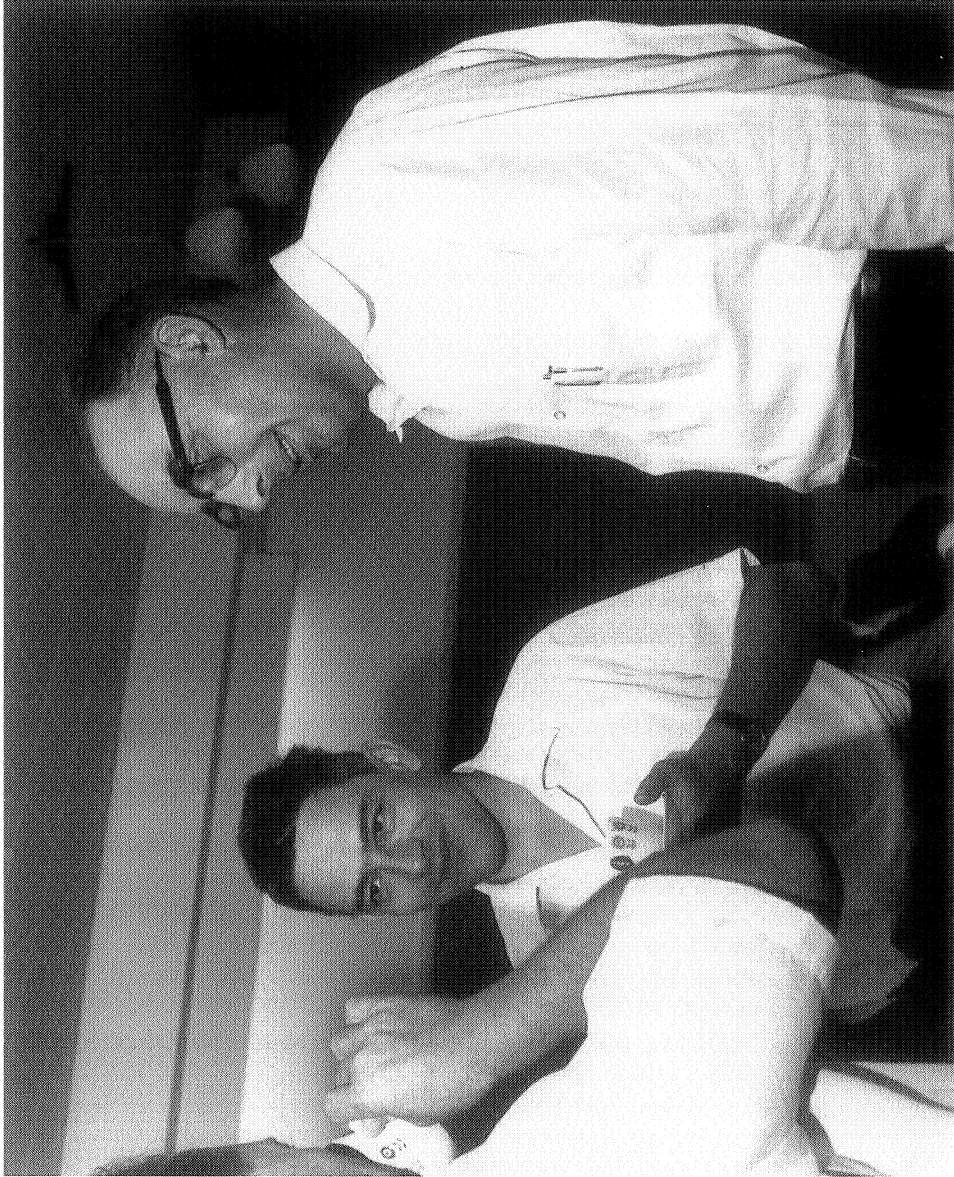
Peter Huber is an internationally renowned statistician. He is a member of the American Academy of Arts and Sciences. Among his numerous achievements to date, Dr. Huber received the Humboldt Award in 1988; he was named Wald Lecturer at the IMS Annual Meeting in 1972, and was awarded the NSF Senior Foreign Scientist Fellowship in 1970-71. He is considered to be the founder of robust statistics. His article in 1964 entitled "Robust Estimation of a Location Parameter" served as an inspiration for many scientists and is an important reference in this field of statistics. He is certainly the person most often cited in the statistical literature in the last 30 years in robust statistics. Professor Peter Huber is the author of an impressive list of books and articles in various fields of statistics. The Interactive Scientific Processor (ISP)

software he developed was a revolution in statistical software and had a strong impact on progress of statistics.

We wish to express our sincere gratitude to Peter J. Huber for providing the journal *Student* with a page of his handwriting, a letter to Geof Watson, and many historical pictures.

Selected list of his contributions:

- [1] Robust Estimation of a Location Parameter. *Ann. Math. Statist.* **35** (1964) 73–101.
- [2] A Robust Version of the Probability Ratio Test. *Ann. Math. Statist.* **36** (1965) 1753–1758.
- [3] Robust Confidence Limits. *Z.f. Wahrscheinlichkeitstheorie.* **10** (1968) 269–278.
- [4] Robust Estimation. *Mathematical Centre Tracts.* **27** (1968) 1–25, Amsterdam.
- [5] Studentizing Robust Estimates. In: *Nonparametric Techniques in Statistical Inference*, ed. M. L. Puri, Bloomington (1970) 453–463.
- [6] Robust Statistics: A Review (The A. Wald Lecture). *Ann. Math. Statist.* **43** (1972) 1041–1067.
- [7] *Robust Estimates of Location: Survey and Advances*. Princeton University Press, Princeton, New Jersey (1972), with Andrews, Bickel, Hampel, Rogers, Tukey
- [8] Robust Regression: Asymptotics, Conjecture and Monte Carlo (Wald Memorial Lecture). *Ann. Statist.* **1** (1973) 799–821.
- [9] Some Mathematical Problems Arising in Robust Statistics. *Proceedings of the International Congress of Mathematicians*, Vancouver (1974) 821–824.
- [10] Robust Methods of Estimation of Regression Coefficients. *Math. Operationsforsch. Statist. Ser. Statistics* **8** No. 1 (1977) 41–53.
- [11] Robust Covariances. In: *Statistical Decision Theory and Related Topics II*, eds. S. S. Gupta and D. S. Moore, Academic Press (1977) 165–191.
- [12] *Robust Statistical Procedures*. Society for Industrial and Applied Mathematics, Philadelphia (1977).
- [13] Robust Smoothing. In: *Robustness in Statistics*, eds. R. L. Launer and G. N. Wilkinson, Academic Press (1979) 33–47.
- [14] Robustness, Computing and Data Analysis. In: *Interactive Statistics*, ed. D. McNeil, North-Holland (1979) 85–90.
- [15] *Robust Statistics*. Wiley, New York (1981).
- [16] Current Issues in Robust Statistics. In: *Proc. Symp. on Recent Advances in Statistics*. Academic Press, New York (1983) 183–196.
- [17] The Place of the L1-norm in robust estimation. *Computational Statistics and Data Analysis* **5** (1987) 255–262.
- [18] Issues in Computational Data Analysis. In: *Computational Statistics*, eds. Y. Dodge and J. Whittaker. (1992) Vol. **2**, 3–13.
- [19] Robustness: Where are we now ? (1995) *Student*, **1**, 75–86.



Peter J. Huber with Lucien Le Cam at Montreal 1968.



Montreal 1968: Sam Karlin, Constance Van Eeden, Mark Kac, Peter J. Huber, Lucien Le Cam and Jacques Neven



 Eidg. Technische Hochschule
 ZÜRICH
 Lehrstuhl für mathematische Statistik

8006 Zürich,
 Tel. 061 265 22 11 32 62 11

8 May 1972

Dear Geof,

Really, I should have written to you
 a long time ago and thanked you
 for the most interesting year we have
 spent at Princeton! Last week, Theo
 Gasser had his Ph.D. - examination, it
 appears that the stimuli he received
 at Princeton were rather decisive for
 him to get ahead with his thesis.

Frank and I just started a seminar
 on robustness; we shall first discuss
 the Princeton M.C. - project and then
 go into robust regression, covariance
 and correlation. By the way, I now
 have a draft for some 2/3 of my
 robustness book; I hope to have drafted
 the rest by the end of this summer,
 so that I could begin with rewriting
 the whole.

Here in Zurich not much
 seems to have changed, except that
 the bureaucracy blossoms at the ETH!

Peter J. Huber's handwriting: the letter he wrote to Geof Watson after the Princeton robustness study.

How are things in Princeton? Is
the department growing into the
direction you want it to grow?
Are you still living in the same
house?

Best wishes, also from Eff.
Peter



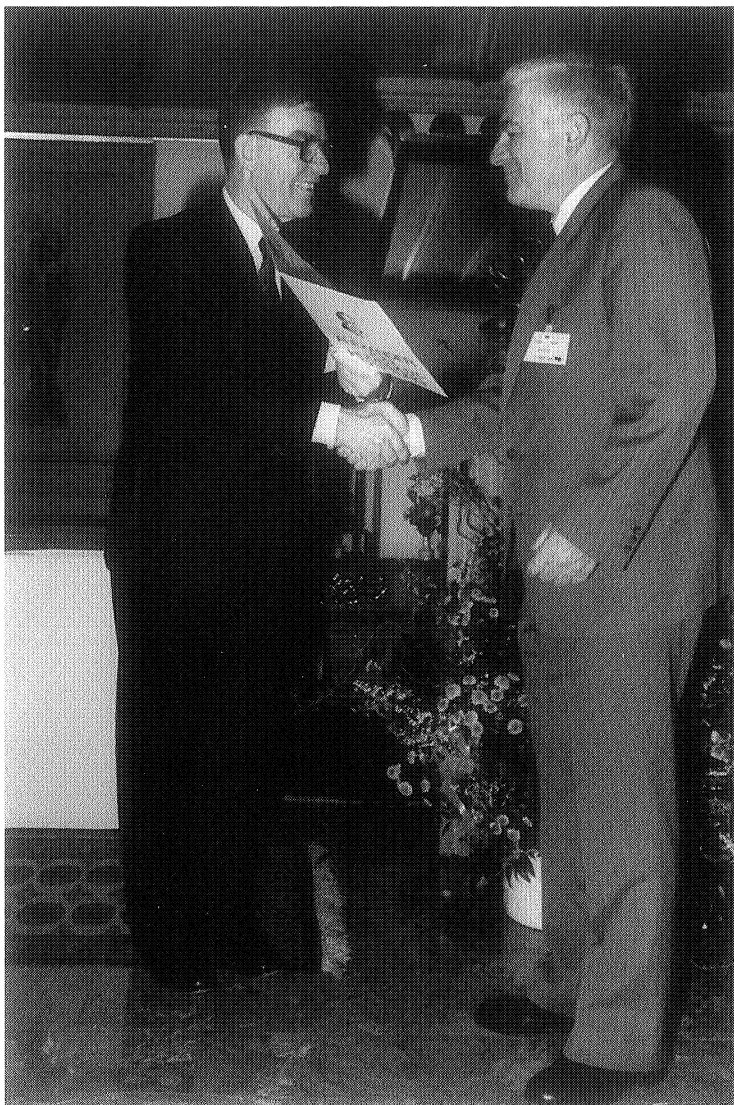
At the occasion of F. Mosteller 70th birthday: Ha-Shwa Lo, Peter J. Huber, Art Dempster and Herman Chernoff, Harvard, 1986.

ROBUST ESTIMATION OF A LOCATION PARAMETER¹BY PETER J. HUBER²*University of California, Berkeley*

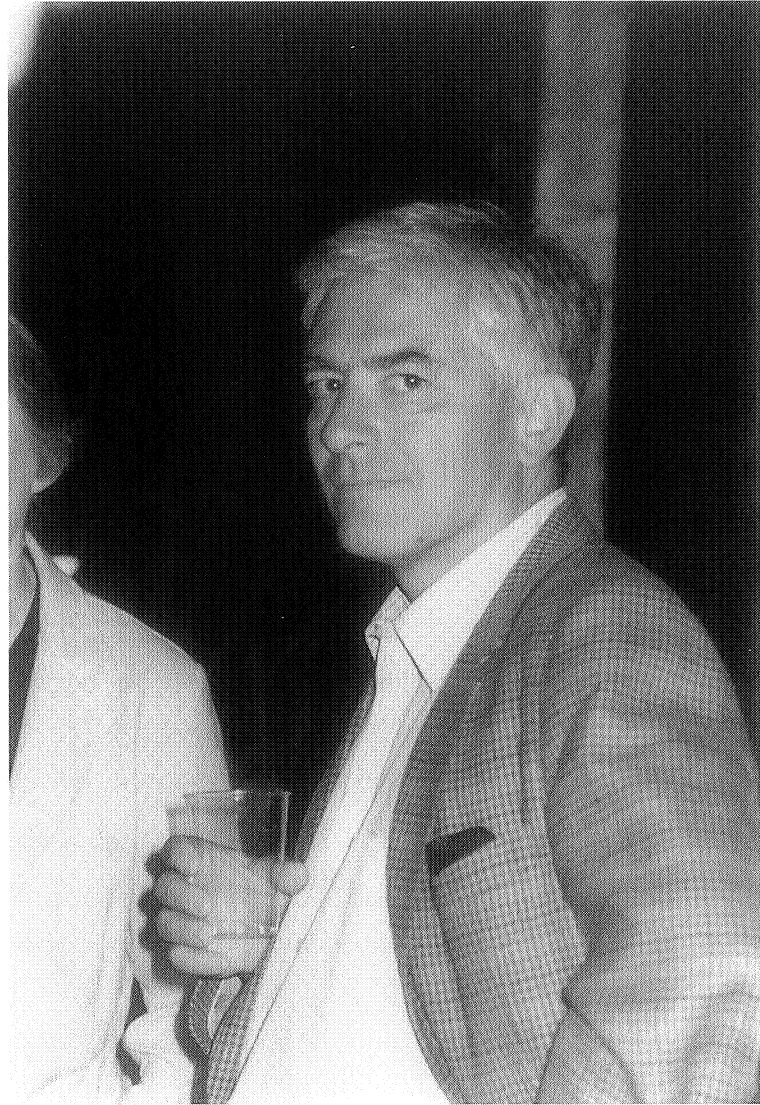
1. Introduction and summary. This paper contains a new approach toward a theory of robust estimation; it treats in detail the asymptotic theory of estimating a location parameter for contaminated normal distributions, and exhibits estimators—intermediaries between sample mean and sample median—that are asymptotically most robust (in a sense to be specified) among all translation invariant estimators. For the general background, see Tukey (1960) (p. 448 ff.)

Let $x_1, \dots, x_n$ be independent random variables with common distribution function $F(t - \xi)$. The problem is to estimate the location parameter ξ , but with the complication that the prototype distribution $F(t)$ is only approximately known. I shall primarily be concerned with the model of indeterminacy $F = (1 - \epsilon)\Phi + \epsilon H$, where $0 \leq \epsilon < 1$ is a known number, $\Phi(t) = (2\pi)^{-1} \int_{-\infty}^t \exp(-\frac{1}{2}s^2) ds$ is the standard normal cumulative and H is an unknown contaminating distribution. This model arises for instance if the observations are assumed to be normal with variance 1, but a fraction ϵ of them is affected by gross errors. Later on, I shall also consider other models of indeterminacy $\sup_t |F(t) - \Phi(t)| \leq \epsilon$.

Some inconvenience is caused by the fact that location and scale are not uniquely determined: in general, for fixed ϵ , there will not be ξ and σ such that $\sup_t |F(t) - \Phi((t - \xi)/\sigma)| \leq \epsilon$, and even in the contaminated case. Although this inherent and unavoidable indeterminacy ϵ is small and is rather irrelevant for practical estimation, it presents problems for the theory, especially for optimality. One may either (i) restrict attention to symmetric distributions, in which case the location of the center of symmetry may define the parameter, or (ii) define the parameter by its asymptotic behavior. The first parameters have been used in the literature, but the second, by its asymptotic behavior, is more natural, and it is more convenient.



Peter Huber receiving the Humboldt-Preis from R. Lüst, President of the Humboldt Foundation. Tegernsee, 1990.



Berkeley, 1983



Peter Huber giving a talk at the occasion of F. Hampel's 60th birthday, 2001.

Description of the circadian melatonin cycle in cattle by the complex cosinor analysis and L_1 -norm

E.S. Altpeter and Th. Abelin

Department of Social- and Preventive Medicine, University of Berne, Switzerland

This paper describes a method to characterize the circadian cycle of melatonin in cattle. The data come from a unique observational study conducted during the shut-down of the short-wave radio-transmitter at Schwarzenburg on March 29th, 1998, at 2:00am. During 3 days before and one week after the shutdown, salivary samples from cattle were sampled and tested for melatonin in order to estimate the effect of radiofrequency electromagnetic fields on melatonin.

Keywords: Complex cosinor analysis, L_1 -norm, melatonin

1 Introduction

A very controversial issue in environmental epidemiology is the impact of radio-frequency electromagnetic fields (RF-EMF) on ill health in humans as well as in animals. The present data come from an observational study in cattle which investigates the influence of RF-EMF on salivary melatonin.

Melatonin is a neurohormone. It is produced in the central nervous system and secreted to the blood by the pineal gland. Melatonin is sleep inducing, is a radical scavenger protecting the body against cancer. It is the time setter of the body clock. Disruptions of the body clock explain the jet lag and insomnia in shift workers - health problems, which may be managed by the timely supplement of oral melatonin. Melatonin is suppressed by day light, the way how the body clock is adapted to the environment. Melatonin is metabolized in the liver and excreted by the kidneys and the saliva. It is well established that changes in melatonin metabolism and excretion are basically not influencing neither the blood nor the salivary concentration in healthy subjects, so changes in melatonin concentration are mainly due to changes in melatonin production and secretion by the pineal gland.

The present study was conducted to investigate the influence of RF-EMF on melatonin cycle characteristics. The purpose was to sort out among other matters of interest are: 1.) whether there was no effect on melatonin metabolism, 2.) whether melatonin production was suppressed and/or 3.) whether the melatonin cycle was phase shifted. For this purpose two characteristics of the cycle were selected: the area-under-the-curve (*AUC*), which is a proxy for melatonin production, and the acrophase (peak time of the melatonin cycle), which is a proxy for the phase. It is impossible to measure these characteristics directly for the following reasons: automatic salivary sampling devices were unavailable and the money for melatonin testing were limited. Therefore the samples had to be taken sequentially between the different animals. The salivary samples had to be centrifuged and stored at -20°C . If less than 1.5 ml saliva had been taken, the sampling procedure had to be repeated. For these reasons sampling time points within one animal were not constant, although a sampling frame of 4 hours was targeted. Together with a constant relative measurement error of melatonin by about 10% and the fact that concentrations of less than 0.8pg/ml and more than 250pg/ml cannot be measured by the laboratory, the following problems had to be addressed: 1.) variance inhomogeneity, 2.) unequally spaced time series, 3.) outlying observations, 4.) censored observation, 5.) missing values, and 6.) directly unobservable melatonin characteristics. This paper describes an approach to estimate the *AUC* and the acrophase in salivary samples. One data set (time series) out of 25 is presented in the paper.

2 The Data

The complete data set of one cattle is presented in Table 1. In the first column is the date, in the second the sampling time (local time, before the time shift it represents winter time, afterwards summer time that means that the subtraction of 1 hour has to be performed before modelling the data), in the third column is the code for the sample, in the fourth is the salivary melatonin concentration in pg/ml, in the last column is an indicator whether the transmitter is on or off. The data can be obtained directly on <http://www.unine.ch/statistics/student/data/data.htm>.

Table 1: Data from one exposed cattle during the shut-down of the Schwarzenburg Transmitter station, Switzerland, in 1998

Date	Time	Sample	Melatonin concentration	Transmitter's action
23.03.1998	12:58	D	leer	A-on
23.03.1998	12:58	D	leer	A-on
23.03.1998	15:15	D2	8.98	A-on
23.03.1998	15:15	D2	leer	A-on
23.03.1998	17:05	E	12.96	A-on
23.03.1998	17:05	E	14.54	A-on
23.03.1998	21:01	F	leer	A-on
23.03.1998	21:01	F	leer	A-on
23.03.1998	21:56	F2	33.11	A-on
23.03.1998	21:56	F2	31.5	A-on
24.03.1998	01:55	A	27.11	A-on
24.03.1998	01:55	A	leer	A-on
24.03.1998	04:13	B	leer	A-on
24.03.1998	04:13	B	leer	A-on
24.03.1998	08:14	C	leer	A-on
24.03.1998	08:14	C	leer	A-on
24.03.1998	09:30	C2	89.8	A-on
24.03.1998	09:30	C2	74.14	A-on
24.03.1998	13:25	D	14.96	A-on
24.03.1998	13:25	D	13.58	A-on
24.03.1998	16:55	E	32.27	A-on
24.03.1998	16:55	E	32.1	A-on
24.03.1998	21:08	F	23.21	A-on
24.03.1998	21:08	F	22.72	A-on
24.03.1998	23:45	A	14.75	A-on
24.03.1998	23:45	A	14.65	A-on
25.03.1998	04:25	B	41.88	A-on
25.03.1998	04:25	B	46.12	A-on
25.03.1998	07:05	C	14.52	A-on
25.03.1998	07:05	C	14.85	A-on
25.03.1998	13:12	D	leer	A-on
25.03.1998	13:12	D	leer	A-on
25.03.1998	14:24	D2	leer	A-on
25.03.1998	14:24	D2	leer	A-on
25.03.1998	17:03	E	12.4	A-on
25.03.1998	17:03	E	9.2	A-on
25.03.1998	20:57	F	leer	A-on
25.03.1998	20:57	F	leer	A-on
25.03.1998	21:23	F2	7.76	A-on
25.03.1998	21:23	F2	8.44	A-on
26.03.1998	00:03	A	36.53	A-on
26.03.1998	00:03	A	leer	A-on
26.03.1998	04:08	B	25.47	A-on
26.03.1998	04:08	B	22.19	A-on
26.03.1998	08:15	C	97.2	A-on
26.03.1998	08:15	C	91	A-on

Date	Time	Sample	Melatonin concentration	Transmitter's action
26.03.1998	13:04	D	16.89	A-on
26.03.1998	13:04	D	17.25	A-on
26.03.1998	17:16	E	3.89	A-on
26.03.1998	17:16	E	4.26	A-on
26.03.1998	21:10	F	8.44	A-on
26.03.1998	21:10	F	8.34	A-on
27.03.1998	04:20	B	16.14	A-on
27.03.1998	04:20	B	18.89	A-on
27.03.1998	08:00	C	44.45	A-on
27.03.1998	08:00	C	45.37	A-on
27.03.1998	13:20	D	18.77	A-on
27.03.1998	13:20	D	18.18	A-on
27.03.1998	17:16	E	11.56	A-on
27.03.1998	17:16	E	12.55	A-on
27.03.1998	21:07	F	33.56	A-on
27.03.1998	21:07	F	leer	A-on
27.03.1998	21:41	F2	36.45	A-on
27.03.1998	21:41	F2	32.06	A-on
27.03.1998	23:05	A	14.42	A-on
27.03.1998	23:05	A	15.44	A-on
28.03.1998	00:03	A	32.33	A-on
28.03.1998	00:03	A	28.74	A-on
28.03.1998	04:20	B	19.55	A-on
28.03.1998	04:20	B	20.25	A-on
28.03.1998	08:00	C	> 250	A-on
28.03.1998	08:00	C	>250	A-on
28.03.1998	11:58	D	46.52	A-on
28.03.1998	11:58	D	45.61	A-on
28.03.1998	16:04	E	7.05	A-on
28.03.1998	16:04	E	8.65	A-on
28.03.1998	19:27	F	9.37	A-on
28.03.1998	19:27	F	7.96	A-on
29.03.1998	01:05	A	15.09	A-on
29.03.1998	01:05	A	13.51	A-on
29.03.1998	05:00	B	leer	A-off
29.03.1998	05:00	B	leer	A-off
29.03.1998	08:52	C	54.28	A-off
29.03.1998	08:52	C	leer	A-off
29.03.1998	12:00	D	20.59	A-off
29.03.1998	12:00	D	leer	A-off
29.03.1998	15:34	E	21.31	A-off
29.03.1998	15:34	E	22.76	A-off
29.03.1998	19:42	F	30.92	A-off
29.03.1998	19:42	F	30.16	A-off
29.03.1998	23:42	A	43.57	A-off
29.03.1998	23:42	A	43.09	A-off
30.03.1998	04:15	B	37.19	A-off
30.03.1998	04:15	B	41.04	A-off
30.03.1998	07:05	C	>250	A-off
30.03.1998	07:05	C	>250	A-off
30.03.1998	13:11	D	19.44	A-off
30.03.1998	13:11	D	18.44	A-off

Date	Time	Sample	Melatonin concentration	Transmitter's action
30.03.1998	16:57	E	6.94	A-off
30.03.1998	16:57	E	8.18	A-off
30.03.1998	20:58	F	105.4	A-off
30.03.1998	20:58	F	80.11	A-off
30.03.1998	23:43	A	30.42	A-off
30.03.1998	23:43	A	27.23	A-off
31.03.1998	04:02	B	31.54	A-off
31.03.1998	04:02	B	28.76	A-off
31.03.1998	08:24	C	29.26	A-off
31.03.1998	08:24	C	23.91	A-off
31.03.1998	12:50	D	10.36	A-off
31.03.1998	12:50	D	10.99	A-off
31.03.1998	16:48	E	18.03	A-off
31.03.1998	16:48	E	16.94	A-off
31.03.1998	20:33	F	10.24	A-off
31.03.1998	20:33	F	9.5	A-off
31.03.1998	23:44	A	55.47	A-off
31.03.1998	23:44	A	57.04	A-off
01.04.1998	04:20	B	22.9	A-off
01.04.1998	04:20	B	26.22	A-off
01.04.1998	08:05	C	6.81	A-off
01.04.1998	08:05	C	7.62	A-off
01.04.1998	12:33	D	7.94	A-off
01.04.1998	12:33	D	8.18	A-off
01.04.1998	15:50	E	3.94	A-off
01.04.1998	15:50	E	3.51	A-off
01.04.1998	20:08	F	6.97	A-off
01.04.1998	20:08	F	7.36	A-off
01.04.1998	23:40	A	3.01	A-off
01.04.1998	23:40	A	2.98	A-off
02.04.1998	03:56	B	13.07	A-off
02.04.1998	03:56	B	12.9	A-off
02.04.1998	07:58	C	>250	A-off
02.04.1998	07:58	C	>250	A-off
02.04.1998	11:53	D	21.26	A-off
02.04.1998	11:53	D	19.58	A-off
02.04.1998	15:38	E	6.28	A-off
02.04.1998	15:38	E	5.37	A-off
02.04.1998	19:58	F	26.17	A-off
02.04.1998	19:58	F	28.2	A-off
03.04.1998	00:08	A	11.12	A-off
03.04.1998	00:08	A	12.2	A-off
03.04.1998	04:27	B	9.14	A-off
03.04.1998	04:27	B	9.16	A-off
03.04.1998	08:28	C	64.85	A-off
03.04.1998	08:28	C	70.05	A-off
03.04.1998	12:04	D	6.11	A-off
03.04.1998	12:04	D	6.87	A-off

3 The complex cosinor model and a proposal of the solution to the addressed problems

Lerchl and Partsch [1] described a model for the melatonin profile which they term complex cosinor analysis (CCA). The CCA takes advantage of the knowledge that melatonin has a very constant cycle of 24hours. Since the cycle is not symmetric they added a frequency of 12hours. They validated the method in human blood samples. The model is specified by

$$y = \beta_0 + \beta_1 \sin \frac{2\pi}{24}t + \beta_2 \cos \frac{2\pi}{24}t + \beta_3 \sin \frac{2\pi}{12}t + \beta_4 \cos \frac{2\pi}{12}t \quad (1)$$

with y the melatonin concentration and time point t in hours. Given $i = 1, \dots, n$ observations per animal, the model is linear with respect to the parameters β_j , $j = 0, \dots, 4$, and can be estimated by any algorithm that solves linear equations by minimizing the remaining random error. The problem of variance homogeneity respectively inhomogeneity could be mastered by taking natural logarithms. The problem of unequally spaced time series and missing values is solved by the CCA model itself. The problem of outlying observations is reduced by selecting L_1 -norm and by taking natural logarithm of the melatonin concentrations. The CCA model in its revised form is

$$\ln y = \beta_0 + \beta_1 \sin \frac{2\pi}{24}t + \beta_2 \cos \frac{2\pi}{24}t + \beta_3 \sin \frac{2\pi}{12}t + \beta_4 \cos \frac{2\pi}{12}t \quad (2)$$

The CCA can be estimated with 5 time points. Basically, 6 time points per day are available for the estimation, and the CCA-model could be estimated daily. But the results would be very unstable. Therefor the study period was divided into the 3 time frames of about equal length, one period before the shutdown and 2 afterwards. For every time frame the parameters β_j were estimated.

The problem of censoring due to the limited detection range of melatonin concentrations between 0.8 pg/ml and 250 pg/ml remains. One solution would have been to insert half respectively the double of the detection range, e.g. 0.4 pg/ml respectively 500 pg/ml. This procedure has large influence on the CCA-model parameters as well as on the estimates of the AUC and the acrophase (see above). Therefor, the decision has been taken to omit measurements outside the range of 0.8 pg/ml and 250 pg/ml.

The AUC was estimated for every of the 3 time frames $k = 1, 2, 3$ as follows

$$AUC_k = \int_0^{24} \exp(\beta_{0k} + \beta_{1k} \sin \frac{2\pi}{24}t + \beta_{2k} \cos \frac{2\pi}{24}t + \beta_{3k} \sin \frac{2\pi}{12}t + \beta_{4k} \cos \frac{2\pi}{12}t) dt. \quad (3)$$

This led to 3 estimates of the AUC per individual.

The acrophase was estimated for every of the 3 time frames $k = 1, 2, 3$ by

$$t_k = \max_t (\beta_{0k} + \beta_{1k} \sin \frac{2\pi}{24}t + \beta_{2k} \cos \frac{2\pi}{24}t + \beta_{3k} \sin \frac{2\pi}{12}t + \beta_{4k} \cos \frac{2\pi}{12}t), \quad (4)$$

where t_k is the estimator of the acrophase. This led to 3 estimates of the acrophase per individual. The 2 unobservable characteristics of the melatonin cycle are estimated.

In a second step the AUC and the acrophase were related to milk production and RF-EMF exposure in 13 exposed and 12 unexposed cattle. A similar observational study in 21 men and 33 women - more or less exposed - was performed during the

same time with a similar statistical approach. For this second step the AUC was transformed by the natural logarithm whereas the acrophase was used untransformed. In this last step the differences between the values before shutdown and afterwards were calculated leading to relative estimates for the melatonin suppression and to absolute changes in acrophase.

4 Results

By applying the L_1 and L_2 fit we found the estimated $\hat{\beta}_{01}, \dots, \hat{\beta}_{41}, \widehat{AUC}_1, \hat{t}_1$. The results are presented in Table 2. In Figure 1 presents similar results.

Table 2: L_1 and L_2 estimates

Estimate	L_1	L_2
$\hat{\beta}_{01}$	3.12	3.08
$\hat{\beta}_{11}$	0.59	0.58
$\hat{\beta}_{21}$	-0.27	-0.24
$\hat{\beta}_{31}$	-0.60	-0.32
$\hat{\beta}_{41}$	0.12	0.19
$\widehat{AUC}_1$	671.85	599.19
$\hat{t}_{11}$	9:01	9:17

The AUC estimated by L_2 is 11% lower than the one by L_1 . The acrophase is about -16 minutes phase advanced comparing L_1 to L_2

5 Discussion

The present paper proposes a solution to the identification of unobservable melatonin characteristics such as AUC and acrophase in mammals. The proposal solves the problem of variance heterogeneity, unequal time series and reduces the problem of outlying observations. The problem of censoring by the detection limits of the laboratory test is addressed in a rather rude way - by ommission. Especially, if this problem could be solved in the framework of L_1 estimation, a more adequate analysis of the data would become available. For this very reason, a sample of the time series is presented here, and the authors invite the statistical community for better solutions or better models than the CCA.

The second step of the analysis is a proposal to the problem of unobservable, but invariant confounders in observational follow-up studies with a change in exposure during the follow-up period. By taking differences of the outcomes, e.g. AUC and acrophase, the unobserved confounding variables are cancelling out. It remains the effect of changing conditions during the follow-up period. This sounds complicated, and in the present situation it is in fact. During the study period the short-wave radio-transmitter was shut-down. At that same time point, time shifted from winter

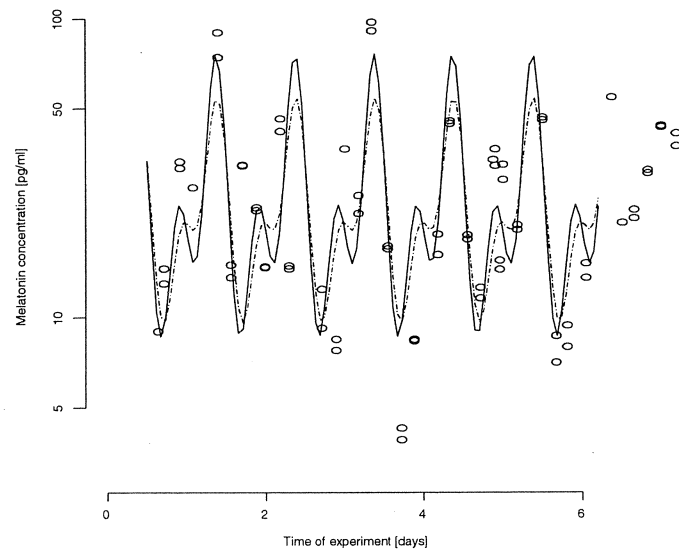


Figure 1. The dash-dotted line is the L_2 fit, the full line is the L_1 fit.

to summer time in Western Europe. However, if a dose-response between the outcomes AUC or acrophase could be shown, this could not be attributed to the time shift but to the changing exposure itself. But all inference in the second step depends on the first one.

Several technical problems beyond the control of the epidemiologists complicated the analysis of the data: in the exposed stable sampling started on monday noon in the unexposed stable on wendsday noon before the shutdown of the transmitter. The devices for measuring temperature and humidity didnot function properly. The measurement of light in the 2 stables were not comparable. RF-EMF exposure was only continously monitored in the exposed stable. The unexposed stable, however, was confirmed to be unexposed to RF-EMF by spot measurements and was selected to be far away from RF-EMF sources. Although these deficiencies the study will contribute to the debate on the influence of RF-EMF exposure on human and animals health.

References

- [1] Lerchl A, Partsch C.J. (1994). Reliable analysis of individual human melatonin profiles by complex cosinor analysis. *J. Pineal Res*, 16(85): 90.

A More Rational Ranking System in Olympic Games

Yadolah Dodge and Giuseppe Melfi

Statistics Group, University of Neuchâtel, Switzerland

Abstract: We analyze the final results of the 2002 Salt Lake City and the 1998 Nagano Winter Olympic Games as well as the 2000 Sydney Summer Olympic Games. Instead of ranking countries by total number of winning medals, we propose a new ranking by standardizing a weighted sum of the number of medals by population size.

Key Words: Olympic data, ranking.

1 Introduction

The 2002 Salt Lake City Winter Olympic Games are now closed in the USA, and, despite of de Coubertin's maxime 'The important thing is to participate', all newspapers in the whole planet exhibited only ranking of nations with the count of gold, silver and bronze medals, and celebrated as modern heroes, their athletes. The final olympic table ranks nations by either the total number of medals obtained, or only by the number of gold medals first, then silver and bronze, or by giving a unique score with special weights to gold, silver and bronze.

The aim of this note is an attempt to present a more rational ranking system. We see that if we slightly modify the ranking criteria, small nations in population size will have a better ranking as compare to the actual system.

2 Current ranking

There are basically three ways of ranking the winning nations in any Olympic games. The first one consists of using a criterion for ranking nations in Olympic games is the lexicographic order of medals. Nation A is better placed than B if A has more gold medals than B . When A and B have the same number of gold medals, A is better

placed than B if A has more silver medals, and so on. This is the most common method of ranking nations in the final table. The media often use this method of ranking. In this method a nation with a unique gold medal is ranked better than a nation with an arbitrarily large number of silver and bronze medals.

A second method, consists of ranking nations by total number of winning medals. The official web site of the Salt Lake City Olympic Games [3] employs this method. The problem with this way of ranking is that a nation may have 20 bronze medals and gets a better ranking than a nation with 19 gold medals.

A third method of ranking the winning nations which is more rational and takes into account the drawbacks of the above two methods can be described as follows. Let P be defined as

$$P = 3 \times (\text{No. of gold medals}) + 2 \times (\text{No. of silver medals}) + \text{No. of bronze medals}.$$

The ranking of nations is determined by decreasing values of the score P . This method is used for example in [2], where is applied to produce a ranking for each edition of Olympic games. Table 1 and 2 respectively show the ranking of nations for the Nagano 1998 and Salt Lake City 2002 Winter Olympic Games using the third methods, according to the score P .

	Nation	Gold	Silver	Bronze	Total	P
1	Germany	12	9	8	29	62
2	Norway	10	10	5	25	55
3	Russia	9	6	3	18	42
4	Canada	6	5	4	15	32
5	USA	6	3	4	13	28
5	Austria	3	5	9	17	28
7	Netherlands	5	4	2	11	25
8	Japan	5	1	4	10	21
9	Italy	2	6	2	10	20
9	Finland	2	4	6	12	20
11	China	0	6	2	8	14
12	South Korea	3	1	2	6	13
12	Switzerland	2	2	3	7	13
12	France	2	1	5	8	13
15	Czech Republic	1	1	1	3	6
16	Sweden	0	2	1	3	5
17	Bulgaria	1	0	0	1	3
18	Ukraine	0	1	0	1	2
18	Denmark	0	1	0	1	2
18	Belarus	0	0	2	2	2
18	Kazakhstan	0	0	2	2	2
22	United Kingdom	0	0	1	1	1
22	Belgium	0	0	1	1	1
22	Australia	0	0	1	1	1

Table 1. Ranking of winning nations according to the score P . Nagano 1998.

To compare the three methods described above, take Austria as an example. For the Salt Lake City Winter Olympic Games, with the method one, sorting by gold medals first, Austria is ranked in the 12th position, while with the second method, ranking by total number of medals, it is ranked in the 5th position and with the third method, according to the P score, it is ranked in the 6th position.

The third method which is based on the P score has also a drawback. Take for example the Nagano Winter Olympic Games. From Table 1 we see that if we use the P score, China with no gold medal gets better ranking than France with two gold

medals. This happens because the number of silver medals of China are superior than that of France. Hence, China is in 11th place, and France is in the 12th place .

	Nation	Gold	Silver	Bronze	Total	P
1	Germany	12	16	7	35	75
2	USA	10	13	11	34	67
3	Norway	11	7	6	24	53
4	Russia	6	6	4	16	34
5	Canada	6	3	8	17	32
6	Austria	2	4	10	16	24
6	Italy	4	4	4	12	24
6	France	4	5	2	11	24
9	Switzerland	3	2	6	11	19
9	Netherlands	3	5	0	8	19
11	Finland	4	2	1	7	17
12	China	2	2	4	8	14
13	Croatia	3	1	0	4	11
14	South Korea	2	2	0	4	10
15	Sweden	0	2	4	6	8
16	Estonia	1	1	1	3	6
16	Australia	2	0	0	2	6
16	Spain	2	0	0	2	6
19	United Kingdom	1	0	2	3	5
20	Bulgaria	0	1	2	3	4
20	Czech Republic	1	0	1	2	4
22	Japan	0	1	1	2	3
22	Poland	0	1	1	2	3
24	Belarus	0	0	1	1	1
24	Slovenia	0	0	1	1	1

Table 2. Ranking of winning nations according to the score P . Salt Lake City 2002.

Now suppose, in addition to the weights furnished by the third method, we add another element of importance, the population size of a winning nation, into consideration. If a country of a small size, for example 7 or 11 millions, win a bronze medal, it should be ranked better than a country with say 350 millions with the same bronze medal. We think it is more rational to bring the population size into account in ranking the nations for the final olympic results. This new weight, the population size, is an attempt to make a new adjustment for ranking nations.

3 Ranking: A proposition

Let Q be defined as $Q = P \cdot 10^6 / \text{Population}$. Here we considered the data of populations for 1998 one can find in [1]. Table 3 shows the Nagano 1998 and Salt Lake City Winter Olympic Games ranking according to the Q score.

The result of this scoring shows that highest positions are occupied by countries of a medium size where winter sports are very popular. Moreover, as we all know, the winter sports are rather expensive and not all nations, specially the southern nations, can participate in such events.

The same idea can be applied to the Summer Olympic Games. For example, we can apply the Q score method to the Sydney 2000 Summer Olympic Games where more countries participated.

Nagano1998				Salt Lake City 2002			
Rank	Nation	Pop.(10 ⁶)	Q	Rank	Nation	Pop.(10 ⁶)	Q
1	Norway	4.42	12.4434	1	Norway	4.42	11.9910
2	Finland	5.15	3.8835	2	Estonia	1.42	4.2254
3	Austria	8.13	3.4440	3	Finland	5.15	3.3010
4	Switzerland	7.26	1.7906	4	Austria	8.13	2.9520
5	Netherlands	15.73	1.5893	5	Switzerland	7.26	2.6171
6	Canada	30.68	1.0430	6	Croatia	4.67	2.3555
7	Germany	82.08	0.7554	7	Netherlands	15.73	1.2079
8	Czech Republic	10.29	0.5831	8	Canada	30.68	1.0430
9	Sweden	8.89	0.5624	9	Germany	82.08	0.9137
10	Denmark	5.33	0.3752	10	Sweden	8.89	0.8999
11	Bulgaria	8.24	0.3641	11	Slovenia	1.97	0.5076
12	Italy	56.78	0.3522	12	Bulgaria	8.24	0.4854
13	Russia	146.86	0.2860	13	Italy	56.78	0.4227
14	South Korea	46.42	0.2801	14	France	58.81	0.4081
15	France	58.81	0.2211	15	Czech republic	10.29	0.3887
16	Belarus	10.41	0.1921	16	Australia	18.61	0.3224
17	Japan	125.93	0.1668	17	USA	270.31	0.2479
18	Kazakhstan	16.85	0.1187	18	Russia	146.86	0.2315
19	USA	270.31	0.1036	19	South Korea	46.42	0.2154
20	Belgium	10.18	0.0982	20	Spain	39.13	0.1533
21	Australia	18.61	0.0537	21	Belarus	10.41	0.0961
22	Ukraine	50.13	0.0399	22	United Kingdom	57.72	0.0866
23	United Kingdom	57.72	0.0173	23	Poland	38.61	0.0777
24	China	1236.92	0.0113	24	Japan	125.93	0.0238
				25	China	1236.92	0.0113

Table 3. Ranking nations according to Q .

A summary of results for the Summer Olympic Games held in Sydney 2000 is given in Table 4, and the ranking according to Q score method is shown in Table 5.

	Nation	Gold	Silver	Bronze	Total	P	Pop.(10 ⁶)	Q
1	USA	39	25	32	96	199	270.31	0.7362
2	Russia	32	27	29	88	179	146.86	1.2188
3	China	28	16	15	59	131	1236.91	0.1059
4	Australia	16	25	17	58	115	18.61	6.1795
5	Germany	14	17	25	56	101	82.07	1.2307
6	France	13	14	11	38	78	58.80	1.3265
7	Italy	13	8	13	34	68	56.78	1.1976
8	Cuba	11	11	7	29	62	11.04	5.6159
9	United Kingdom	11	10	7	28	60	57.72	1.0395
10	Netherlands	12	9	4	25	58	15.73	3.6872
11	Romania	11	6	9	26	54	22.40	2.4107
12	South Korea	8	9	11	28	53	46.42	1.1417
13	Ukraine	3	10	10	23	39	50.13	0.7780
14	Hungary	7	6	3	16	36	10.21	3.5260
15	Japan	5	8	5	18	36	125.93	0.2859
16	Poland	6	5	3	14	31	38.61	0.8029
17	Bulgaria	5	6	2	13	29	8.24	3.5194
18	Greece	4	6	3	13	27	10.66	2.5328
19	Belarus	3	3	11	17	26	10.41	2.4976
20	Sweden	4	5	3	12	25	8.89	2.8121
21	Canada	3	3	8	14	23	30.68	0.7497
22	Norway	4	3	3	10	21	4.42	4.7511
23	Spain	3	3	5	11	20	39.13	0.5111
24	Brazil	0	6	6	12	18	169.81	0.1060
25	Ethiopia	4	1	3	8	17	58.39	0.2911
26	Kazakhstan	3	4	0	7	17	16.85	1.0089
27	Switzerland	1	6	2	9	17	7.26	2.3416
28	Czech Republic	2	3	3	8	15	10.29	1.4577
29	Kenya	2	3	2	7	14	28.34	0.4940
30	Denmark	2	3	1	6	13	5.33	2.4390
31	Indonesia	1	3	2	6	11	212.94	0.0517
32	Jamaica	0	4	3	7	11	2.64	4.1667
33	Turkey	3	0	1	4	10	64.57	0.1549

	Nation	Gold	Silver	Bronze	Total	P	Pop.(10 ⁶)	Q
34	Iran	3	0	1	4	10	68.96	0.1450
35	Slovakia	1	3	1	5	10	5.39	1.8553
36	Mexico	1	2	3	6	10	98.55	0.1015
37	Finland	2	1	1	4	9	5.15	1.7476
38	Lithuania	2	0	3	5	9	3.60	2.5000
39	Austria	2	1	0	3	8	8.13	0.9840
40	Algeria	1	1	3	5	8	30.48	0.2625
41	Azerbaijan	2	0	1	3	7	7.86	0.8906
42	Uzbekistan	1	1	2	4	7	23.78	0.2944
43	Belgium	0	2	3	5	7	10.18	0.6876
44	South Africa	0	2	3	5	7	42.84	0.1634
45	Slovenia	2	0	0	2	6	1.97	3.0457
46	Serbia and Montenegro	1	1	1	3	6	11.21	0.5352
47	Latvia	1	1	1	3	6	2.39	2.5105
48	New Zealand	1	0	3	4	6	3.63	1.6529
49	Nigeria	0	3	0	3	6	110.53	0.0543
50	Argentina	0	2	2	4	6	36.27	0.1654
51	Taiwan	0	1	4	5	6	21.91	0.2738
52	Morocco	0	1	4	5	6	29.11	0.2061
53	Georgia	0	0	6	6	6	5.11	1.1742
54	Bahamas	1	1	0	2	5	0.28	17.8571
55	Thailand	1	0	2	3	5	60.03	0.0833
56	Estonia	1	0	2	3	5	1.42	3.5211
57	North Korea	0	1	3	4	5	22.18	0.2254
58	Croatia	1	0	1	2	4	4.67	0.8565
59	Mozambic	1	0	0	1	3	18.64	0.1609
60	Colombia	1	0	0	1	3	38.58	0.0778
61	Cameroon	1	0	0	1	3	15.03	0.1996
62	Saudi Arabia	0	1	1	2	3	20.79	0.1443
63	Moldova	0	1	1	2	3	4.46	0.6726
64	Trinidad and Tobago	0	1	1	2	3	1.12	2.6786
65	Vietnam	0	1	0	1	2	76.24	0.0262
66	Uruguay	0	1	0	1	2	3.29	0.6079
67	Ireland	0	1	0	1	2	3.62	0.5525
68	Portugal	0	0	2	2	2	9.93	0.2014
69	Costa Rica	0	0	2	2	2	3.61	0.5540
70	Barbados	0	0	1	1	1	0.26	3.8462
71	Kyrgyzstan	0	0	1	1	1	4.52	0.2212
72	Armenia	0	0	1	1	1	3.42	0.2924
73	Qatar	0	0	1	1	1	0.70	1.4286
74	Sri Lanka	0	0	1	1	1	18.93	0.0528
75	Kuwait	0	0	1	1	1	1.91	0.5236
76	Israel	0	0	1	1	1	5.64	0.1773
77	Iceland	0	0	1	1	1	0.27	3.7037
78	India	0	0	1	1	1	983.38	0.0010
79	Chile	0	0	1	1	1	14.68	0.0681

Table 4. Data of Summer Olympic Games, Sydney 2000.

The results of the Table 5 show an unexpected homogeneity. In best positions we have several Caribbean countries, as well as the host country, Australia. European countries of different sizes get lower ranking. Except for Russia, large countries such as USA, China, Japan, Brazil, have a modest ranking, and they have a ranking comparable to that of most emergent countries.

The ranking proposed in this paper according to Q score method standardizes the number of medals by population size. The score Q can be considered as a measure of the popularity of certain disciplines, namely Olympic sports. Of course, one can combine many other socio-economic and geographic factors, the number of participating athletes for each nation, and so on to come up with a more accurate and optimal measure of ranking. A nice problem!

Rank	Nation	Q	Rank	Nation	Q
1	Bahamas	17.8571	41	USA	0.7362
2	Australia	6.1795	42	Belgium	0.6876
3	Cuba	5.6159	43	Moldova	0.6726
4	Norway	4.7511	44	Uruguay	0.6079
5	Jamaica	4.1667	45	Costa Rica	0.5540
6	Barbados	3.8462	46	Ireland	0.5525
7	Iceland	3.7037	47	Serbia and Montenegro	0.5352
8	Netherlands	3.6872	48	Kuwait	0.5236
9	Hungary	3.5260	49	Spain	0.5111
10	Estonia	3.5211	50	Kenya	0.4940
11	Bulgaria	3.5194	51	Uzbekistan	0.2944
12	Slovenia	3.0457	52	Armenia	0.2924
13	Sweden	2.8121	53	Ethiopia	0.2911
14	Trinidad and Tobago	2.6786	54	Japan	0.2859
15	Greece	2.5328	55	Taiwan	0.2738
16	Latvia	2.5105	56	Algeria	0.2625
17	Lithuania	2.5000	57	North Korea	0.2254
18	Belarus	2.4976	58	Kyrgyzstan	0.2212
19	Denmark	2.4390	59	Morocco	0.2061
20	Romania	2.4107	60	Portugal	0.2014
21	Switzerland	2.3416	61	Cameroon	0.1996
22	Slovakia	1.8553	62	Israel	0.1773
23	Finland	1.7476	63	Argentina	0.1654
24	New Zealand	1.6529	64	South Africa	0.1634
25	Czech Republic	1.4577	65	Mozambic	0.1609
26	Qatar	1.4286	66	Turkey	0.1549
27	France	1.3265	67	Iran	0.1450
28	Germany	1.2307	68	Saudi Arabia	0.1443
29	Russia	1.2188	69	Brazil	0.1060
30	Italy	1.1976	70	China	0.1059
31	Georgia	1.1742	71	Mexico	0.1015
32	South Korea	1.1417	72	Thailand	0.0833
33	United Kingdom	1.0395	73	Colombia	0.0778
34	Kazakhstan	1.0089	74	Chile	0.0681
35	Austria	0.9840	75	Nigeria	0.0543
36	Azerbaijan	0.8906	76	Sri Lanka	0.0528
37	Croatia	0.8565	77	Indonesia	0.0517
38	Poland	0.8029	78	Vietnam	0.0262
39	Ukraine	0.7780	79	India	0.0010
40	Canada	0.7497			

Table 5. Ranking according to Q score.

References

- [1] U.S. Bureau of the Census, Report WP/98, *World Population Profile: 1998*, U.S. Government Printing Office, Washington, DC, 1999.
- [2] S.J. Darmoni, URL: <http://www.darmoni.net/>.
- [3] URL: <http://www.saltlake2002.com/>.