School of Electrical and Computer Engineering
Australian Telecommunications Research Institute

# Some Aspects of Signal Processing in Heavy Tailed Noise

Ramon Francis Brcic

This thesis is presented for the Degree of
Doctor of Philosophy
of
Curtin University of Technology

October 2002

# Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgment has been made.

Signature: .....

Date: 5/8/2003

*To Mum and Dad — for everything.*
*Also to Oma & Opa and Baba & Dida.*

# Keywords

Heavy tailed distributions, impulsive noise, stable distributions, goodness-of-fit tests, multiple hypothesis tests, characteristic function, parameter estimation, array processing, source detection, model selection, Akaike's information criterion, minimum description length, sphericity test, resampling, bootstrap, bias estimation, M-estimation, robust statistics, density estimation, signal detection.

# Abstract

This thesis addresses some problems that arise in signal processing when the noise is impulsive and follows a heavy tailed distribution. After reviewing several of the more well known heavy tailed distributions, the common problem of which of these best models the observations is considered. To this end, a test is proposed for the symmetric alpha stable distribution. The test threshold is found using both asymptotic theory and parametric bootstrap resampling. In doing so, some modifications are proposed for Koutrouvelis' estimator of the symmetric alpha stable distribution's parameters that improve performance. In electrical systems impulsive noise is generated externally to the receiver while thermal Gaussian noise is generated internally by the receiver electronics, the resultant noise is an additive combination of these two independent sources. A characteristic function domain estimator for the parameters of the resultant distribution is developed for the case when the impulsive noise is modeled by a symmetric alpha stable distribution. Having concentrated on validation and parameter estimation for the noise model, some problems in signal detection and estimation are considered. Detection of the number of sources impinging on an array is an important first step in many array processing problems for which the development of optimal methods can be complicated even in the Gaussian case. Here, a multiple hypothesis test for the equality of the eigenvalues of the sample array covariance is proposed. The nonparametric bootstrap is used to estimate the distributions of the test statistics, removing the assumption of Gaussianity and offering improved performance for heavy tailed observations. Finally, some robust estimators are proposed for estimating parametric signals in additive noise. These are based on M-estimators but implicitly incorporate an estimate of the noise distribution, enabling the estimator to adapt to the unknown noise distribution. Two estimators are developed, one uses a nonparametric kernel density estimator while the other models the score function of the noise distribution with a linear combination of basis functions.

# Contents

# List of Figures

# List of Tables

# Abbreviations, Acronyms and Symbols

| | |
|---|---|
| $\alpha S$ | alpha stable |
| $\varepsilon-$mix | epsilon mixture |
| AIC | Aikaike information criterion |
| ARE | asymptotic relative efficiency |
| BER | bit error rate |
| cdf | cumulative distribution function |
| cf | characteristic function |
| CDMA | code division multiple access |
| CLT | central limit theorem |
| CMMSE | constrained minimum mean squared error |
| CMVN | complex multivariate normal |
| DOA | direction of arrival |
| ecf | empirical characteristic function |
| edf | empirical distribution function |
| FLOM | fractional lower order moment |
| FWE | family wise error rate |
| GCLT | generalised central limit theorem |
| iid | independent and identically distributed |
| IF | influence function |
| IMSE | integrated mean square error |
| IQR | interquartile range |
| LLS | linear least squares |
| LS | least squares |
| MDL | minimum description length |
| MHT | multiple hypothesis test |
| MISE | mean integrated squared error |
| MLE | maximum likelihood estimate/estimator |

| | |
|---|---|
| MMSE | minimum mean squared error |
| MSE | mean squared error |
| MUD | multiuser detection |
| MVN | multivariate normal |
| N | normal |
| NFR | near far ratio |
| NWLS | nonlinear weighted least squares |
| p-value | probability value |
| pdf | probability density function |
| RV | random variable |
| $S\alpha S$ | symmetric alpha stable |
| $S\alpha SG$ | symmetric alpha stable Gaussian sum |
| SNR | signal to noise ratio |
| SRB | sequentially rejective Bonferroni |
| WCDMA | wideband code division multiple access |
| WLS | weighted least squares |
| $\sim$ | distributed as |
| $\overset{a}{\sim}$ | asymptotically distributed as |
| $\xrightarrow{d}$ | converges in distribution, converges weakly |
| $\approx$ | approximately |
| $(\cdot)^*$ | complex conjugation |
| $(\cdot)^\dagger$ | the pseudoinverse |
| $(\cdot)^{\mathsf{H}}$ | Hermitian transpose |
| $(\cdot)^{\mathsf{T}}$ | transposition |
| $\overset{d}{=}$ | equality in distribution |
| $\lvert\cdot\rvert$ | absolute value |
| $\lVert\cdot\rVert$ | Euclidean norm |
| $\lfloor\cdot\rfloor$ | the integer part |
| $\alpha$ | characteristic exponent of the $\alpha S$ distribution |
| $\alpha S(\alpha, \beta, c, \delta)$ | alpha stable distribution with characteristic exponent $\alpha$, skewness parameter $\beta$, scale parameter $c$ and location parameter $\delta$ |
| $\beta$ | skewness parameter of the $\alpha S$ distribution |
| $\beta(a, b)$ | beta distribution with parameters $a, b$ |
| $\gamma_G$ | dispersion of the Gaussian component of the $S\alpha SG$ distribution |
| $\gamma_{S\alpha S}$ | dispersion of the $S\alpha S$ component of the $S\alpha SG$ distribution |
| $\Gamma(\cdot)$ | Euler's Gamma function |
| $\delta$ | location parameter of the $\alpha S$ distribution |

| | |
|---|---|
| $\varepsilon$ | proportion of outliers in the $\varepsilon$-mixture model |
| $\varepsilon-\text{mix}(\varepsilon, \kappa)$ | $\varepsilon$-mixture distribution with parameters $\varepsilon$ and $\kappa$ |
| $\zeta$ | set level of a test |
| $\hat{\theta}$ | estimate or estimator of a parameter $\theta$ |
| $\kappa$ | ratio of the variance of the impulsive and Gaussian components in the $\varepsilon$-mixture model |
| $\lambda_i$ | population eigenvalues |
| $\mu$ | mean |
| $\sigma^2$ | variance |
| $\phi_X$ | characteristic function of the random variable $X$ |
| $\hat{\phi}_X$ | an estimate of the characteristic function |
| $\chi^2_d$ | Chi-squared distribution with $d$ degrees of freedom |
| $\boldsymbol{A}$ | array steering matrix |
| $(\boldsymbol{A})_{ij}$ | element $(i, j)$ of the matrix $\boldsymbol{A}$ |
| $c$ | scale parameter of the $\alpha S$ distribution |
| $\text{CMVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ | complex multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ |
| $\mathsf{Cov}[X, Y]$ | covariance between $X$ and $Y$ |
| $\text{diag}(\cdot)$ | a diagonal matrix |
| $D_{KS}$ | Kolmogorov-Smirnov statistic |
| $\mathsf{E}[X]$ | expectation of $X$ |
| $f'(\cdot)$ | the first derivative of a function $f(\cdot)$ |
| $f_X(x)$ | probability density function of $X$ |
| $F_X(x)$ | cumulative distribution function of $X$ |
| $\mathsf{H}$ | null hypothesis |
| $\text{I}(\cdot)$ | the indicator function |
| $\mathsf{K}$ | alternative hypothesis |
| $l_i$ | sample eigenvalues |
| $\text{L}(\mu, \sigma^2)$ | Laplace distribution with mean $\mu$ and variance $\sigma^2$ |
| $\text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ | multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ |
| $\text{N}(\mu, \sigma^2)$ | normal distribution with mean $\mu$ and variance $\sigma^2$ |
| $o$ | order notation, of same order as |
| $O$ | order notation, of smaller order than |
| $\mathcal{P}$ | p-value |
| $P_D$ | probability of detection |
| $P_{FA}$ | probability of false alarm |
| $\mathsf{Pr}[E]$ | probability of the event $E$ |

| | |
|---|---|
| $\mathsf{Pr}[E \mid E_0]$ | conditional probability of the event $E$ given the event $E_0$ |
| $Q_{AD}$ | Anderson-Darling statistic |
| $Q_{CM}$ | Cramér-von Mises statistic |
| $\mathbb{R}$ | the set of real numbers |
| $\mathrm{Re}(\cdot)$ | real part of a complex variable |
| $\mathrm{sgn}(\cdot)$ | the signum function |
| $t_d$ | Student's $t$ distribution with $d$ degrees of freedom |
| $\mathrm{TU}(a, b)$ | convolution of three $\mathrm{U}(a, b)$ distributions |
| $\mathrm{U}(a, b)$ | uniform distribution on $[a, b]$ |
| $\mathsf{Var}[X]$ | variance of $X$ |
| w.p.1 | converges with probability one, converges strongly |
| $\mathbb{Z}$ | the set of integers |

# Publications

The following publications have been produced during the period of PhD candidacy.

## Internationally Refereed Journal Articles

1. R. Brcich, A. Zoubir, and P. Pelin. Detection of sources using bootstrap techniques. *IEEE Transactions on Signal Processing*, 50(2):206–15, February 2002.

2. A. Zoubir and R. Brcich. Multiuser detection in heavy tailed noise. *Digital Signal Processing: A Review Journal*, 12(2-3):262–73, April-July 2002.

3. A. Zoubir and R. Brcich. Tolerance intervals for accuracy control of bootstrapped matched filters. *IEEE Signal Processing Letters*, 9(8):247–50, August 2002.

## Internationally Refereed Conference Papers

1. R. Brcich and A. Zoubir. Robust estimation with parametric score function estimation. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 2, pages 1149–52, Orlando, USA, May 2002.

2. A. Zoubir and R. Brcich. Multiuser detection in heavy tailed noise. In *Proceedings of the Defence Applications of Signal Processing (DASP) Workshop*, Adelaide, Australia, September 2001. Held in June 2002.

3. R. Brcich and A. Zoubir. Resampling based techniques for source detection in array processing. In *Proceedings of the $11^{th}$ IEEE Workshop on Statistical Signal Processing*, pages 26–9, Singapore, August 2001.

4. A. Taleb, R. Brcich, and M. Green. Suboptimal robust estimation for signal plus noise models. In *Conference Record of the 34$^{rd}$ Asilomar Conference on Signals, Systems and Computing*, volume 2, pages 837–41, Pacific Grove, USA, October 2000.

5. R. Brcich, P. Pelin, and A. Zoubir. Detection of sources in array processing using the bootstrap. In *Proceedings of the 10$^{th}$ IEEE Workshop on Statistical Signal and Array Processing*, pages 448–52, Pocono Manor, USA, August 2000.

6. P. Pelin, R. Brcich, and A. Zoubir. A bootstrap technique for rank estimation. In *Proceedings of the 10$^{th}$ IEEE Workshop on Statistical Signal and Array Processing*, pages 94–8, Pocono Manor, USA, August 2000.

7. A. Zoubir, R. Brcich, D. Tufts, and E. Real. The bootstrap and tolerance regions. In *Conference Record of the 33$^{rd}$ Asilomar Conference on Signals, Systems and Computing*, volume 1, pages 15–9, Pacific Grove, USA, October 1999.

8. R. Brcich and A. Zoubir. Estimation and detection in a mixture of symmetric alpha stable and Gaussian interference. In *Proceedings of the IEEE Signal Processing Workshop on Higher Order Statistics*, pages 219–22, Ceasarea, Israel, June 1999.

# Acknowledgments

It's hard to believe that once I finish writing this, that will be it, PhD done. Fait accompli. The past five years have been a bit of an adventure and there are a lot of people I'd like to thank for their advice, support and generally making things interesting.

To my supervisor, Prof. Abdelhak Zoubir, Abdelhak, you were one of the most interesting and definitely most dynamic (even without the coffee) lecturers I had at QUT, which is why it didn't take much (any?) persuasion to get me to do a PhD. Your support from start to finish has been indispensable, thanks for keeping me on track. Also to Dr Robert Iskander, I think when I asked you to be my co-supervisor, you said it would be an honour. Well, it's been an honour to work with someone whom I respect both personally and professionally, I've always valued your advice.

Dr Chris Brown, Dr Per Pelin and Dr Anisse Taleb, I've enjoyed working with you all, thanks for providing the inspiration which helped me out of a few dead ends. On another note, some equivalent to Pauli's exclusion principle must exist which Pelle and Anisse obey by mutual agreement, otherwise the universe would have destroyed itself by now. You definitely livened things up.

The move from Brisbane to Perth would have been harder without the rest of the (post-revolution) CIP gang: Dr Matt Green, Dr Mark Morelande, Dr Hwa-Tung Ong and Dr Chris Brown, who became the core of the CSP group. We got to know each other better than we might have. To those at QUT, Greg, George and everyone else, it was a pleasure.

Now, on to the next generation CSP group. In roughly chronological order: Diva, Luke, Ahdi, Mashury, Amar, Said, Henrik (our Swedish ambassador), Andrew, Julius, Paul, Sonny and the passers through, Florian, Mark, Stefan and Zohra. You guys made the lab an interesting place, there are some disadvantages to having an office.

Thanks to everyone at Curtin, the School of ECE, ATRI and the ATCRC, especially the Swedish contingent who are too numerous to mention but include

Anders, Erik the Viking, Christer, Lisbeth, Jonas, Mattias, Jorgen, Jimmy and Anders D, sometimes the lunchroom felt like a Swedish dependency.

Thank you to all the support staff. At QUT, Annette, thanks for slipping that PhD application through (it really pays to know people in the office of research), Cherie and Denise. At Curtin, Todd, Susan, Ingrid, Desi and Marilyn, who all made the transition much easier. Also thanks to the tireless computer administrators Phil, Iain and Alison.

Thanks to the CRCSS, the School of ECE, ATRI, the ATCRC and the Australian government for scholarship support and holiday/conference travel.

To all the members of Dynamo Chicken Kiev: Kip, Justin, Graeme, Mark (The Enforcer) and the dozen or so others I've already mentioned in one way or the other — we never lost a game, but sometimes we just ran out of time. It was fun.

Tanya Vernon has been involved in the group, on and off, since before I had ever heard of Signal Processing and who I got to know in Perth. Thanks for the many long hours you put into helping all us PhD students. I've got some unforgettable memories of the trips you organised, all good times. I wish you the best in you current endeavor, which I'm sure you'll finish in world record time.

Thanks to the following for keeping me awake during some long nights and days: Powderfinger, The Smashing Pumpkins, Nirvana, Clouds, RHCP, Foo Fighters, Midnight Oil, INXS, REM, The B-52's, Grinspoon, Frente, MGF, The Offspring, Regurgitator, Hole, The Superjesus, Spiderbait, Deadstar and The Pixies.

Finally, to anyone who actually reads this and is contemplating or currently doing a PhD (Tanya B. and Tanya V.), I've only got a few bits of advice. If you are going through hell, keep going. Everyone has bad days. Otherwise, listen to some music or read a good book, I'd recommend: Dialogue Concerning the Two Chief World Systems by Galileo Galilei, Cosmos by Carl Sagan, Surely You're Joking Mr Feynman! and QED The Strange Theory of Light and Matter both by Richard P. Feynman or The Source by James A. Michener.

*The universe is not just stranger than we have previously imagined, it is stranger than we can imagine.*

— *Niels Bohr*

*Keep an open mind, but not so open your brains fall out.*

— *James Oberg*

# Chapter 1

# Introduction

*Do, or do not. There is no 'try'.*

*— Yoda*

For several reasons the Gaussian distribution dominates as a statistical model. Theoretical justification is provided by the central limit theorem (CLT) which states that, subject to some conditions [25], the result of summing the contributions of a large number of random variables with similar behaviour is a Gaussian random variable. A well known example is the Gaussian nature of thermal noise. Procedures based on the assumption of Gaussian noise tend to be structurally simple, the optimal Neyman-Pearson detector for a known signal in Gaussian noise is the matched filter, a linear transformation. Finally, there exists a large body of knowledge regarding the statistical characteristics of Gaussian processes, which expediates the design and analysis of these procedures.

In contrast, non-Gaussian models generally lead to nonlinear structures. This can complicate matters to the extent that although the Gaussian model is not accurate it is preferred for its tractability and the performance loss is accepted. Despite the simplicity of a Gaussian model, performance may be degraded to the point where more accurate, albeit more complicated, non-Gaussian models are necessary. Performance degradation is most severe when the noise is impulsive. Impulsive noise is characterised by larger values which occur too frequently for the Gaussian model to be an accurate description of the process. The heavy tailed distributions used to model impulsive noise are common in the statistical literature where they model data contaminated by outliers.

There now exists a large body of evidence for impulsive noise in a variety of important practical situations such as wireless and submarine communications, switching transients in powerlines [4, 129, 135], ignition noise [80, 161, 162], underwater acoustic signals [174], radar sea clutter [147], infra-red remote sens-

ing [96], synthetic aperature radar returns [12, 45], shot noise [146] and telephone static [165]. Impulsive noise is also prevalent in work environments due to office equipment, fluorescent lighting, elevators, microwaves and electrical machinery [20, 21]. Heavy tailed processes have been used to model heart rate variability [28, 29], seismic activity and climate change [147], computer networks [62, 93, 94] and econometric time series [117, 123].

A number of heavy tailed distributions exist and the question of which best models a set of observations is an important one since the performance of any procedure is dependent on the model faithfully capturing reality. Existing techniques used in testing for specific distributions tend to perform poorly when deciding between heavy tailed distributions and so more powerful tests are needed. Following validation of the model, any defining parameters must be estimated. The first part of this thesis concentrates on the problems of validation and estimation when using symmetric alpha stable ($S\alpha S$) distributions to model impulsive noise.

The prevalence of heavy tailed phenomena has motivated the design of statistical methods with improved performance over those based on Gaussian assumptions and, for the case when none exist, the development of new techniques. This is the focus of the second part of the thesis where specific problems in detection and estimation are addressed.

Detection of the number of sources impinging on an array is an application of the general problem of estimating the dimension of a signal subspace. The associated statistical theory is complicated even for the Gaussian case, while existing solutions are known to be sensitive to departures from Gaussianity. The source detection problem has not been properly addressed for heavy tailed models, one approach is proposed here.

An optimal approach to estimation of a parametric signal in additive noise is maximum likelihood estimation (MLE). Some disadvantages of MLE are its potentially debilitating computational complexity and reliance on a specific model. Robust estimators such as M-estimation seek to reduce dependence on a specific model while still offering certain level of performance with acceptable complexity. Some modifications to existing robust estimators are proposed which improve their performance.

## 1.1 Aims and Objectives

The broad aim of this thesis is to develop techniques for the validation and parameter estimation of models for impulsive noise, and then to design methods

for detection and estimation of signals in impulsive noise. More specifically,

1. Validation and parameter estimation: To develop tests for the validation of $S\alpha S$ distributions. To both improve and develop estimators for the parameters of heavy tailed distributions.

2. Detection: To design methods for the detection of sources in array processing.

3. Estimation: To design robust estimators when the noise is impulsive.

## 1.2 Contributions

The original contributions made in this dissertation include

1. A goodness-of-fit test for $S\alpha S$ distributions based on the stability property.

2. Asymptotically optimal selection of points at which to sample the empirical characteristic function (ecf) for Koutrouvelis' $S\alpha S$ parameter estimator.

3. An improved estimator for the parameters of the $S\alpha S$ Gaussian sum ($S\alpha SG$) distribution, a sum of $S\alpha S$ and Gaussian random variables which models impulsive and thermal noise simultaneously.

4. A method for source detection in array processing when the sample size is small and/or the observations are non-Gaussian.

5. Generic adaptive robust estimators for the parameters of a deterministic signal in additive noise.

## 1.3 Scope and Overview

**Chapter 2.** Reviews several heavy tailed distributions which are used as impulsive noise models throughout the thesis.

**Chapter 3.** A goodness-of-fit hypothesis test is developed for $S\alpha S$ distributions based on the stability property. The null distributions of the test statistics are derived using asymptotic theory and estimated using bootstrap techniques.

**Chapter 4.** An estimator for the parameters of the $S\alpha SG$ distribution model of impulsive noise is proposed based on a nonlinear weighted least squares regression in the characteristic function (cf) domain.

**Chapter 5.** The problem of source detection in array processing is considered. A detector utilising the bootstrap to estimate the null distributions of the test statistics is developed.

**Chapter 6.** An adaptive robust estimator is developed for a parametric signal in impulsive noise. An M-estimation structure is used but with the score function of the noise distribution being estimated from the observations.

**Chapter 7.** Conclusions and future directions.

**Appendix A.** A review of the cf, ecf and associated theory.

**Appendix B.** Optimal sampling of the ecf for estimation of the parameters of $S\alpha S$ distributions by the method of Koutrouvelis.

**Appendix C.** Taylor series expansions of random variables.

**Appendix D.** A review of the bootstrap and similar resampling techniques.

**Appendix E.** A review of multiple hypothesis tests.

# Chapter 2

# Models for Heavy Tailed Distributions

*Research is what I'm doing when*
*I don't know what I'm doing.*
*— Wernher von Braun*

Heavy tailed distributions may be defined as those which possess probability density functions (pdfs) whose tails decay at a rate which is less than that of the Gaussian pdf. Although this definition admits an unlimited number of possible pdfs, several dominate due to their mathematical tractability and capacity to accurately model data. The models may be broadly classified as either statistical-physical or empirical models.

In communications, radar and sonar systems it is generally assumed that there are an unspecified number of independent noise sources transmitting at random in time and space. The mechanisms generating the noise include natural and artificial electromagnetic or acoustic transients and multiple access interference. Statistical-physical models are based solely on these broad assumptions regarding the noise generating mechanisms and generally lead to accurate but complex models whose parameters have a direct physical significance.

Empirical models are favored for their ability to accurately model the data in a mathematically tractable fashion. Their parameters may not have any direct physical significance, but rather describe the level of impulsive behaviour in an arbitrary fashion.

Several of the most common heavy tailed distributions which model impulsive noise are Middleton's models [129], the $S\alpha S$ distribution [135] and the Gaussian mixture distribution [186]. Next, the models are briefly reviewed. Herein, the general assumption of iid noise is made.

## 2.1 Middleton's Models

Middleton developed the statistical-physical Class A, B and C models of noise [19, 129, 130, 132] based on general assumptions regarding the statistics of the noise sources. The models are canonical, that is, they are independent of the specific mechanisms which generate the noise. They model electromagnetic and acoustic noise equally well given the fundamental assumptions hold, these are

1. An unlimited number of sources generate random waveforms independent of each other.

2. The distribution of these sources is random in time and space, being temporally and spatially Poisson distributed independent of the waveform or propagation law.

3. The propagation law is an inverse power one.

Inherent in these models is a Gaussian component which describes thermal noise generated at the receiver.

The three classes, A, B and C are characterised by the relative bandwidths of the noise and the receiver. For Class A noise the signal bandwidth is less than the receiver bandwidth and visa versa for Class B noise. Class C is an intermediate case. The relationship between the noise and receiver bandwidths results in Class B noise producing significant transients at the receiver while for Class A noise the transients are negligible. Consequently, the Class B model is more impulsive.

The pdf of Class A noise is a scale mixture of Gaussian distributions, the instantaneous envelopes of the sources being Rayleigh,

$$f_X(x) = \sum_{n=0}^{\infty} \frac{e^{-A}A^n}{n!} \frac{1}{\sqrt{2\pi\sigma_n^2\sigma^2}} \exp\left(-\frac{x^2}{2\sigma_n^2\sigma^2}\right), \tag{2.1}$$

where $\sigma_n^2 = (n/A + \Gamma_p)/(1 + \Gamma_p)$. The pdf is described by the 3 parameters $A \geq 0$, $\sigma^2 > 0$ and $\Gamma_p \geq 0$ which describe the intensity with which impulsive events occur, the power (variance) of the noise and the ratio of the powers of the Gaussian and non-Gaussian components respectively.

The more general Class B model requires up to 6 parameters and is a generalisation of Hall's model for atmospheric interference when the Gaussian component is ignored [128]. Middleton's Class A model has since been generalised to model the electromagnetic noise received by antenna arrays in communications systems [125].

## 2.2 Symmetric Alpha Stable Distributions

$S\alpha S$ distributions were first encountered by Cauchy. He considered whether a pdf would result if the Gaussian cf, $\exp(-\omega^2)$, was generalised to $\exp(-|\omega|^\alpha)$. The only solution obtained in closed form was for $\alpha = 1$ and is now known as the Cauchy distribution. The inverse Fourier transform of this cf is a proper pdf for $0 < \alpha \leq 2$ and is the $S\alpha S$ distribution [90]. Since then, the theory of $S\alpha S$ distributions has been extended to alpha stable ($\alpha S$) distributions, which may be nonsymmetric, for multivariate random variables and processes [33, 52, 98, 152, 154, 156]. Due to the general lack of a closed form expression for the pdf of $\alpha S$ distributions but the existence of a simple expression for the cf, much of the theory has been developed in the cf domain [114]. The theory of $\alpha S$ distributions is considered in more detail in Chapter 3.

These heavy tailed distributions possess several interesting properties, two of which are highly relevant to the use of $\alpha S$ distributions in modelling impulsive noise. First, except for the specific case of Gaussian distributions, they possess infinite variance. Some argue that infinite variance is an unrealistic assumption, since in practice the variance will always be bounded. By the same reasoning the Gaussian model can also be rejected since in practice unbounded random variables do not occur. Also, the sample variance of a finite realisation of $\alpha S$ random variables will always be bounded. Second, the generalised central limit theorem (GCLT) states that regardless of the existence of the variance, the limiting distribution of a sum of independent and identically distributed (iid) random variables is $\alpha S$. This encompasses the CLT where the limiting distribution of a sum of iid random variables with finite variance is Gaussian. Just as the CLT is a powerful motivation for the use of a Gaussian model, so too does the GCLT make $\alpha S$ distributions an appealing model.

Characterised by two parameters, $S\alpha S$ distributions can offer greater mathematical tractability over Middleton's more complex statistical-physical models. In particular, $S\alpha S$ distributions approximate Middleton's Class B model if the Gaussian component is ignored. As the Class B model requires 6 parameters, the parsimonious $S\alpha S$ model may be preferred [132]. The $S\alpha S$ distribution can be regarded as a hybrid between statistical-physical and empirical models.

Practical applications of $S\alpha S$ distributions go back to the work by Holtsmark who showed that the gravitational field resulting from an infinite collection of Poisson distributed sources in space is $S\alpha S$ [35]. Later, $\alpha S$ distributions were suggested as a model for econometric time series by Mandelbrot [118, 117] and

others [47, 48, 57, 123]. Further evidence for $\alpha S$ distributions came when telephone noise was shown to be non-Gaussian and well modelled by $\alpha S$ distributions [165].

Motivation for the $\alpha S$ model is not limited to empirical evidence. In a similar approach to Middleton's it has been shown that spatially and temporally distributed Poisson scatterers result in $S\alpha S$ noise at the front end of a receiver [135]. The same was shown to be true in a multiple access system where the noise sources are other users sharing the channel and of backscattered echo in radar systems [82, 84].

Some more recent applications of $\alpha S$ distributions have been in modelling submarine communications, switching transients in power-lines, automobile ignition, seismic activity, ocean waves, radar sea clutter, file transfer size and packet inter-arrival times in computer networks, heart rate variability, synthetic aperature radar data, infra-red remote sensing data, shot noise, Cerenkov radiation and ultrasound imaging of tissue [4, 12, 28, 29, 45, 62, 93, 94, 96, 146, 147].

## 2.3 Gaussian Mixtures

Gaussian mixtures possess a pdf which is a weighted sum of Gaussian pdfs. As the ratio of each components' variance increases with respect to the smallest, the heavier tailed the distribution becomes. The model is quite general and for a suitably large number of components can approximate a wide variety of symmetric zero mean random variables such as the Laplace distribution [9] and the $S\alpha S$ distribution [110]. The approximation becomes equality for a continuous scale mixture of Gaussian distributions, so that $S\alpha S$ and Laplace distributions can be considered to be conditionally Gaussian [109, 156, 9]. The definition of Middleton's Class A model directly shows it to be a Gaussian mixture distribution comprised of an infinite number of components.

A large number of components, while accurate, is not particularly parsimonious and tractability will suffer. In [22, 108, 107] a 2 to 4 term Gaussian mixture was suggested as being sufficient in most problems where the parameters were estimated using the expectation-maximisation algorithm. A popular compromise is the 2 component $\varepsilon$-mixture ($\varepsilon-$mix) distribution,

$$f_X(x) = (1-\varepsilon) f_G(x;\sigma^2) + \varepsilon f_G(x;\kappa\sigma^2), \tag{2.2}$$

where $0 < \varepsilon < 1$, $\kappa > 1$ and $f_G(x;\sigma^2)$ is a Gaussian pdf with zero mean and variance $\sigma^2$. The component with the larger variance models the impulsive noise,

impulsive events occurring with probability $\varepsilon$. The $\varepsilon-$mix distribution can be regarded as an approximation to Middleton's Class A model where all the components comprising the non-Gaussian noise are lumped together into a single component of the Gaussian mixture [186]. Aside from this theoretical motivation, the $\varepsilon-$mix distribution is an empirical model for impulsive noise. It is also one of the simplest models, in terms of mathematical tractability, which includes the effects of Gaussian noise.

## 2.4 Generalised Distributions

Distributions such as the generalised Gaussian, generalised Cauchy and generalised Laplace have not been as extensively studied for the purpose of modelling impulsive noise [36, 99, 172, 133] as the aforementioned Middleton's models, $S\alpha S$ distributions or Gaussian mixtures.

The generalised Gaussian distribution has been considered as an alternative to the $S\alpha S$ distribution [173] and as a model for multiple access interference in direct sequence spread spectrum systems [172], while in [174] an asymmetric generalised Gaussian distribution was shown to accurately model underwater acoustic data. The tails of the generalised Gaussian distribution decay exponentially, meaning the distribution is not as heavy tailed as one with algebraically decaying tails, such as the $S\alpha S$ or Cauchy distribution. This exponential decay is evident from the generalised Gaussian pdf,

$$f_X(x) = \frac{\nu}{2a\Gamma(1/\nu)} \exp\Big(-\Big|\frac{x}{a}\Big|^{\nu}\Big) \tag{2.3}$$

and is controlled by the shape parameter $0 < \nu \leq 2$. The scale parameter $a > 0$ is related to the variance of the distribution as $\sigma^2 = a^2\Gamma(3/\nu)/\Gamma(1/\nu)$ where $\Gamma(\cdot)$ is Euler's Gamma function. For $\nu = 2$ the distribution is Gaussian and for $\nu = 1$ it is the Laplace or double exponential. Note that the form of the generalised Gaussian pdf, $\exp(-|x|^{\nu})$, is the same as that of the $S\alpha S$ cf, $\exp(-|\omega|^{\alpha})$. Use of the generalised Gaussian distribution as an impulsive noise model is motivated by the simple form of the pdf, the existence of all moments and the reduction to a Gaussian distribution for $\nu = 2$.

In comparison, the generalised Cauchy distribution possesses algebraically decaying tails [99], except for a limiting case where it approaches the generalised Gaussian distribution [133]. Also note that the generalised Cauchy and $S\alpha S$ distributions coincide for the case of a Cauchy distribution. One reason for using

the generalised Cauchy distribution is that it broadens some existing models for impulsive noise [133, 126].

The generalised Laplace distribution has not appeared a great deal in the literature except for the special case of the Laplace distribution. It can be shown to have a limiting Gaussian distribution and hence coincides with the generalised Gaussian distribution in this and the Laplace case [36].

## 2.5 Examples

To gain an intuitive understanding into the nature of impulsive noise, realisations from several heavy tailed distributions are shown in Figure 2.1 with their associated pdfs appearing in Figure 2.2. The distributions have been scaled to have unit variance and for the $S\alpha S$ distribution, a unit scale. Note the common characteristic impulses which tend to dominate the sample. Their relative intensity is related to the tail thickness. For example, the $S\alpha S$ distribution, having the thickest tails, produces significant impulses.

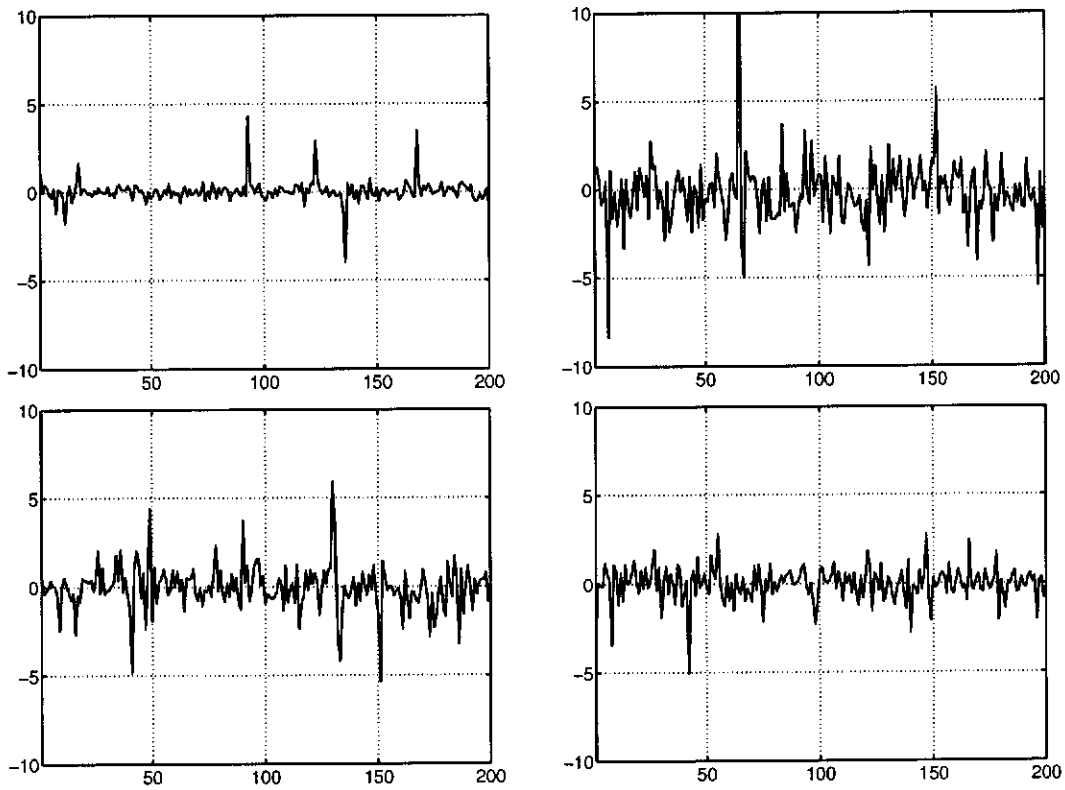


Figure 2.1: Realisations from several heavy tailed distributions. Middleton's Class A model with $A = 0.1$, $\Gamma_p = 0.1$ (top left), the $S\alpha S$ distribution with $\alpha = 1.8$ (top right), the $\varepsilon-$mix distribution with $\varepsilon = 0.1$, $\kappa = 10$ (bottom left) and the generalised Gaussian distribution with $\nu = 1$.

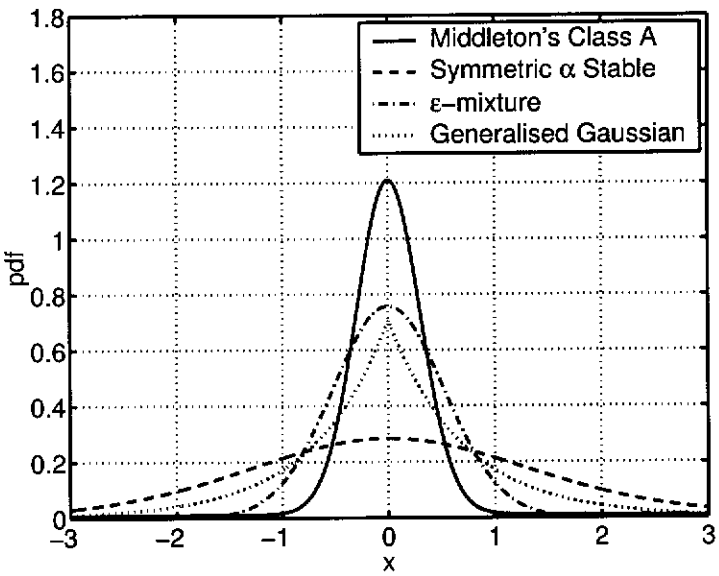


Figure 2.2: Pdfs of the heavy tailed distributions referred to in Figure 2.1.

# Chapter 3

# Testing for Symmetric Alpha Stable Distributions

*Do not worry about your problems*
*with mathematics, I assure you*
*mine are far greater.*
*— Albert Einstein*

The symmetric alpha stable distribution is a well-known statistical model for heavy tailed phenomena encountered in communications, radar, biomedicine and finance. Theoretical motivation for the symmetric alpha stable model is provided by results such as the generalised central limit theorem and the stability property which transcend similar results for Gaussian distributions. Aside from theoretical arguments, heavy tailed phenomena have been shown to be well modelled by symmetric alpha stable distributions in many cases. The empirical evidence is largely qualitative in nature, such as visual comparisons of the amplitude probability distributions of impulsive noise and symmetric alpha stable distributions. As human judgement is subjective, quantitative statistical tests are preferred. Here, a hypothesis test for the goodness-of-fit to symmetric alpha stable distributions is developed based on the stability property.

## 3.1 Introduction

While a model may be well motivated theoretically, sufficient empirical evidence should exist to support its use, as statistical procedures developed under the incorrect model may perform poorly. Much effort has been devoted to the analysis of signals modelled using $\alpha S$ distributions, even though they may not always provide the best fit to the data [16, 88]. Empirical evidence for $\alpha S$ distributions

tends to be based on qualitative subjective tests or exploratory analysis, hence quantitative statistical tests are needed.

Current techniques include tests for infinite variance and examination of the estimated pdf's tails [135]. Tests for infinite variance exclude the Gaussian case, which is $S\alpha S$, but admit all other distributions with infinite variance. Both require visual inspection, leading to subjective decisions. A statistical test for $\alpha S$ distributions has been proposed based on the maximum distance between the ecf (empirical characteristic function) and the parametric cf evaluated at estimates of the distributional parameters [29]. This is an adaptation of a well known cf based goodness-of-fit test [197].

In this Chapter statistical tests for $S\alpha S$ distribution are developed based on their unique stability property, the approach being to determine whether the observations possess this property. Both asymptotic theory and the parametric bootstrap are used to obtain the distribution of the test statistic under the null hypothesis.

This Chapter begins by defining the $\alpha S$ distribution and reviewing some of the properties required later for development of the test in Section 3.2. Next, in Section 3.3, generic goodness-of-fit tests and specific tests for heavy tailed and $\alpha S$ distributions are reviewed. The proposed test is then developed in Section 3.4 and is compared with some existing goodness-of-fit tests in Section 3.5.

## 3.2 Theory of Alpha Stable Distributions

In this section the $\alpha S$ distribution is defined and some of its important properties are reviewed.

### 3.2.1 Definitions of Alpha Stable Distributions

Several equivalent definitions for $\alpha S$ distributions exist. A common thread linking these definitions is the invariance of $\alpha S$ pdfs to convolution, so that sums of iid $\alpha S$ random variables are also $\alpha S$.

**Definition 1** *A random variable $X$ follows an $\alpha S$ distribution if for any $A, B > 0$, there is a $C > 0$ and a $D \in \mathbb{R}$ such that*

$$AX_1 + BX_2 \stackrel{d}{=} CX + D, \tag{3.1}$$

*where $X_1$, $X_2$ are independent copies of $X$ and $\stackrel{d}{=}$ denotes equality in distribution [156].*

**Definition 2** *A random variable $X$ follows an $\alpha S$ distribution if for any $n \geq 2$, there is a $C_n > 0$ and a $D_n \in \mathbb{R}$ such that*

$$X_1 + X_2 + \cdots + X_n \stackrel{d}{=} C_n X + D_n, \tag{3.2}$$

*where $X_1, X_2, \cdots, X_n$ are independent copies of $X$. $C_n$ can be shown to be $n^{1/\alpha}$ [156].*

The term independent copy implies that if a random variable $X$ is distributed as $F_X$ and another random variable $Y$ is said to be an independent copy of $X$, then the distribution of $Y$, $F_Y$, is identical to $F_X$.

A random variable is strictly $\alpha S$ if $D = 0$ in (3.1), or if $D_n = 0$ in (3.2). A random variable is $S\alpha S$ if $X$ and $-X$ have identical distributions, $S\alpha S$ random variables are then a special case of strictly $\alpha S$ random variables.

These two definitions demonstrate the stability property of $\alpha S$ distributions. They imply that sums of iid $\alpha S$ random variables also follow the same $\alpha S$ distribution to within some scale ($C$ in (3.1), $C_n$ in (3.2)) and shift ($D$ in (3.1), $D_n$ in (3.2)).

**Definition 3** *A random variable $X$ follows an $\alpha S$ distribution if it has a domain of attraction such that there exists a sequence of iid random variables $Y_1, Y_2, \ldots$ and sequences $d_n > 0$, $a_n \in \mathbb{R}$ fulfilling*

$$\frac{Y_1 + Y_2 + \cdots + Y_n}{d_n} + a_n \stackrel{d}{\rightarrow} X, \tag{3.3}$$

*where $\stackrel{d}{\rightarrow}$ denotes convergence in distribution [156].*

An equivalent interpretation of this definition is the so called GCLT which states that the limiting distribution of a normalised sum of iid random variables is $\alpha S$, if a limiting distribution exists [25, 52, 63, 156]. The GCLT generalises the CLT for sums of iid random variables when the variance does not exist, reducing to the CLT when the variance does exist.

**Definition 4** *A random variable $X$ follows an $\alpha S$ distribution if there are parameters $0 \leq \alpha \leq 2$, $-1 \leq \beta \leq 1$, $c > 0$ and $\delta \in \mathbb{R}$ such that its cf is of the form*

$$\phi_X(\omega) = \exp(\jmath\delta\omega - |c\omega|^\alpha (1 + \jmath\beta \mathrm{sgn}(\omega)\, g(\omega))), \tag{3.4}$$

*where*

$$g(\omega) = \begin{cases} -\tan\left(\frac{\pi\alpha}{2}\right) & \alpha \neq 1 \\ \frac{2}{\pi}\log|\omega| & \alpha = 1 \end{cases} \tag{3.5}$$

*and $\mathrm{sgn}(\cdot)$ denotes the signum function [65, 156].*

An $\alpha S$ distribution is uniquely defined by the four parameters used in the above definition. The meaning of each of the parameters is as follows

$\alpha$: The characteristic exponent is a measure of the degree of heavy tailedness of the distribution, smaller values corresponding to heavier tails. The asymptotic rate of decay in the tails of the pdf is proportional to $|x|^{-\alpha-1}$.

$\beta$: The skewness parameter is positive when the distribution is skewed to the right and negative when skewed to the left. When $\beta = \pm 1$ the distribution is zero over part of the real line.

$c$: The scale parameter is a measure of dispersion or spread. When $\alpha = 2$ the variance, $\sigma^2$, exists and $c^2 = \sigma^2/2$.

$\delta$: The location parameter shifts the distribution along the real line. When $1 < \alpha \leq 2$ the mean, $\mu$, exists and $\mu = \delta$.

An $\alpha S$ distribution is standardised if $c = 1$ and $\delta = 0$, in which case it is entirely determined by the shape parameters $\alpha$ and $\beta$, excluding the special case $\alpha = 1$, $\beta \neq 0$. If $X$ follows a standard $\alpha S$ distribution, then $cX + \delta$ has an $\alpha S$ distribution with the same shape parameters, but with scale $c$ and location $\delta$, this is easily proven using the $\alpha S$ cf. Of interest here is the $S\alpha S$ distribution for which $\beta = 0$, $\delta = 0$ and the cf is

$$\phi_X(\omega) = \exp(-|c\omega|^{\alpha}). \tag{3.6}$$

## 3.2.2 Properties of Alpha Stable Distributions

### 3.2.2.1 Probability Density and Cumulative Distribution Functions

The pdf of an $\alpha S$ distribution is unimodal and continuous. In all but three cases the pdf cannot be expressed in closed form but exists as a definite integral obtained from an inverse Fourier transform of the cf. The three exceptions are the Gaussian ($\alpha = 2$, $\beta = 0$), Cauchy ($\alpha = 1$, $\beta = 0$) and Lévy ($\alpha = 0.5$, $\beta = 1$) distributions. Convergent and asymptotic series representations for $\alpha S$ distributions also exist.

Much effort has been devoted to developing fast algorithms for calculation of the pdf and cumulative distribution function (cdf), though a numerically expensive inversion of the cf is the only general approach which can be taken. Efficient methods for numerical integration coupled with an alternative representation of

the cf for improved accuracy over parts of the parameter space was considered in [139, 136, 138]. Further reductions were obtained by using spline interpolation over a grid covering the support of the distribution and the parameter space. In [135, 182] series expansions are used when they are known to be accurate, while a fast Fourier transform inversion of the cf was used otherwise. An alternative approach was taken in [124] where the $S\alpha S$ distribution was modeled as a linear combination of Cauchy and Gaussian distributions with the residue being approximated by splines, this gave accurate results for $0.84 < \alpha \leq 2$.

The numerical difficulties associated with evaluating the $\alpha S$ pdf and cdf have limited the widespread application of common statistical techniques, such as MLE [113], to $\alpha S$ distributions and has motivated investigation into alternatives.

#### 3.2.2.2 Tail Behaviour

Non-Gaussian ($\alpha < 2$) $\alpha S$ distributions are heavy tailed, with an asymptotically algebraic rate of decay [156],

**Property 1** *If* $X \sim \alpha S(\alpha, \beta, c, \delta)$ *with* $0 < \alpha < 2$ *then,*

$$\lim_{x \to \infty} \mathsf{Pr}[X > x] = (1+\beta) C_\alpha c^\alpha x^{-\alpha} \tag{3.7}$$

$$\lim_{x \to \infty} \mathsf{Pr}[X < -x] = (1-\beta) C_\alpha c^\alpha x^{-\alpha}, \tag{3.8}$$

*where*

$$C_\alpha = \begin{cases} \frac{1-\alpha}{2\Gamma(2-\alpha)\cos\left(\frac{\pi\alpha}{2}\right)} & \alpha \neq 1, \\ \frac{1}{\pi} & \alpha = 1. \end{cases} \tag{3.9}$$

*The Gaussian* ($\alpha = 2$) *distribution decays exponentially, if* $X \sim \alpha S(2, 0, c, \delta)$ *then*

$$\lim_{|x| \to \infty} \mathsf{Pr}[X > |x|] = \frac{1}{2\sqrt{\pi} c x} \exp\left(-\frac{x^2}{4c^2}\right). \tag{3.10}$$

#### 3.2.2.3 Fractional Lower Order Moments

For an $\alpha S$ distribution with $\alpha < 2$, the variance does not exist, in addition, the mean does not exist for $\alpha \leq 1$. This limits the use of the method of moments and has motivated the use of fractional lower order moments (FLOMs) in their place. The FLOM of order $p$ is defined as

$$\mathsf{E}[|X|^p], \tag{3.11}$$

where in general $0 < p \leq 2$, although negative values of $p$ are sometimes used.

**Property 2** *The FLOMs of an $\alpha S$ distribution are finite for $0 < p < \alpha$ [135, 156]. For a $S\alpha S$ distribution the FLOMs exist for $-1 < \alpha < p$ [116, 135].*

The conditions required for existence of the FLOMs of an $\alpha S$ distribution can be obtained by way of Property 1.

## 3.3 Goodness-of-Fit Tests

The goodness-of-fit problem can be stated as follows. Given $N$ iid random variables $X_1, \ldots, X_N$, determine whether they are distributed as $F_X(x; \boldsymbol{\theta})$ where the distributional parameters contained in $\boldsymbol{\theta}$ may be unknown.

An objective answer to this question is obtained by formulating the problem as a statistical hypothesis test [112]. The hypothesis testing framework for the goodness-of-fit problem is comprised of the null hypothesis $\mathsf{H}$, that the $X_n$ are distributed as $F_X(x; \boldsymbol{\theta})$, versus the alternative hypothesis $\mathsf{K}$, that they are not. Let the unknown distribution of the $X_n$ be $G_X(x)$, then the test is succinctly described as

$$\begin{aligned} \mathsf{H} &: G_X(x) = F_X(x; \boldsymbol{\theta}) \\ \mathsf{K} &: G_X(x) \neq F_X(x; \boldsymbol{\theta}). \end{aligned} \tag{3.12}$$

Note that while the distribution of the $X_n$ is not equal to $F_X(x; \boldsymbol{\theta})$ under the alternative hypothesis, it is otherwise unspecified. Futhermore, when testing for a class of distributions parameterised by $\boldsymbol{\theta}$, estimates for $\boldsymbol{\theta}$ are generally required as their exact values are unspecified. In general then, the goodness-of-fit problem consists of null and alternative hypotheses which are both composite. Before proceeding, a short digression is made to briefly review the theory of hypothesis testing and define some commonly used terms.

Two general philosophies of hypothesis testing exist, Bayesian and Neyman-Pearson [131, 112, 157]. Bayesian tests seek to minimise the cost of making a decision. The necessity of assigning costs to correct and incorrect decisions have limited the use of Bayesian tests as these costs are difficult to assign and interpret in practice. For these reasons, Neyman-Pearson tests are the most common and are used exclusively herein.

A Neyman-Pearson test seeks to maximise the probability of correctly accepting the alternative hypothesis (or equivalently, correctly rejecting the null hypothesis) subject to a constraint on the probability of incorrectly accepting the alternative hypothesis.

The former probability is known as the power of the test or the probability of detection, depending on the particular application, and is denoted $P_D$. A test is said to be powerful if its power is large compared to other tests. The latter probability is known as the false alarm rate or the probability of false alarm and is denoted $P_{FA}$. The constraint placed on $P_{FA}$ is $P_{FA} \leq \zeta$ where $\zeta$ is set a priori and is known as the set level of the test. A test which meets this constraint is said to maintain the set level, if $P_{FA} < \zeta$ the test is said to be conservative. The decision is based on the test statistic $T(X_1, \ldots, X_N)$, a function of the observations.

In a binary test where there are two hypotheses $\mathsf{H}$ and $\mathsf{K}$, every possible value of $T(X_1, \ldots, X_N)$ is assigned by the test to one of two mutually exclusive regions, the rejection or critical region $\mathcal{R}$ and the acceptance region $\mathcal{A}$. If $T(X_1, \ldots, X_N) \in \mathcal{R}$ then $\mathsf{K}$ is accepted, alternatively, if $T(X_1, \ldots, X_N) \in \mathcal{A}$ then $\mathsf{H}$ is accepted. The regions $\mathcal{R}$ and $\mathcal{A}$ are related to the probabilities $P_D$ and $P_{FA}$ as follows,

$$\mathsf{Pr}[T(X_1, \ldots, X_N) \in \mathcal{R} \,|\, \mathsf{H}] = P_{FA} \leq \zeta \tag{3.13}$$

$$\mathsf{Pr}[T(X_1, \ldots, X_N) \in \mathcal{R} \,|\, \mathsf{K}] = P_D, \tag{3.14}$$

where the probabilities are conditioned on which hypothesis is true. The boundary between these two regions is know as the threshold of the test or the critical value. A test is optimal in the Neyman-Pearson sense if it is the most powerful among all tests which maintain the set level. The power of a test can always be increased at the cost of increasing the set level, thereby increasing $P_{FA}$. Typically $\zeta$ is set to an acceptably small value, say 0.05.

In order to formulate an optimal test, the distribution of the observations is required under the null and alternative hypotheses. In the goodness-of-fit problem it is only required that the distribution of the observations, $G_X(x)$, is not equal to that under the null hypothesis, $F_X(x; \boldsymbol{\theta})$, but is otherwise unspecified. Hence, an optimal test cannot be formulated. Instead, tests must be compared in terms of their power, given they maintain the set level.

Returning now to the question of testing for goodness-of-fit, a survey of the literature reveals that several techniques exist which attempt to solve this problem. Generic methods include graphical approaches, chi-squared ($\chi^2$) and empirical distribution function (edf) tests, each of which has many variants [38]. Other tests make use of a property particular to the null distribution [117]. More recently, techniques based on the ecf have been proposed [197]. A brief overview of these techniques follows with special regard to their application in testing for $\alpha S$ distributions.

### 3.3.1 Graphical Tests

Probability plots are generic graphical tests useful in exploratory work. They plot a transformation of the ecdf (empirical cumulative distribution function) such that under the hypothesised distribution a straight line is obtained. Regression tests or human judgement are then used to measure the fit.

From (3.12) a plot of $G_X^{-1}(F_X(x;\boldsymbol{\theta}))$ versus $x$ is linear under the null hypothesis $G_X(x) = F_X(x;\boldsymbol{\theta})$, as is a plot of $G_X(x)$ versus $F_X(x;\boldsymbol{\theta})$. The former, a QQ (quantile-quantile) plot, plots quantiles under the null hypothesis versus empirical quantiles. The latter, a PP (percentile-percentile) plot, plots percentiles under the null hypothesis versus empirical percentiles. Empirical estimates for $G_X(x)$ may be obtained from the ecdf,

$$\hat{G}_X(x) = \frac{1}{N}\sum_{n=1}^{N} \mathrm{I}(X_n \leq x)\,, \tag{3.15}$$

where $\mathrm{I}(\cdot)$ is the indicator function.

QQ plots are not recommended for heavy tailed null distributions as their variance may grow very large away from the centre of the distribution. Subjective judgements are more easily made using PP plots as their variance reaches a bounded maximum at the $50^{\text{th}}$ percentile and decreases moving out towards the extreme percentiles, if necessary variance stabilised PP plots may be used [139, 137].

In Figure 3.1 QQ and PP plots are shown under the null hypothesis where $N = 1000$ observations were drawn from a standard $S\alpha S$ distribution with $\alpha = 1$. For the QQ plot it is difficult to judge visually whether the observations follow a straight line except near the centre of the distribution due to the aforementioned high variance in the extreme quantiles. In contrast, the PP plot appears to favour the null hypothesis as the observations clearly follow a straight line across all percentiles. This highlights the unsuitability of QQ plots when testing for a heavy tailed null hypothesis.

Other graphical techniques compare the fit of the ecdf to the null cdf either subjectively, or using statistical methods such as the Kolmogorov-Smirnov test. At least for $\alpha S$ distributions subjective tests of this sort are not recommended as the ecdf changes rapidly near the mode. Plots of the smoothed empirical pdf (epdf) versus the pdf have been suggested instead [137, 139].

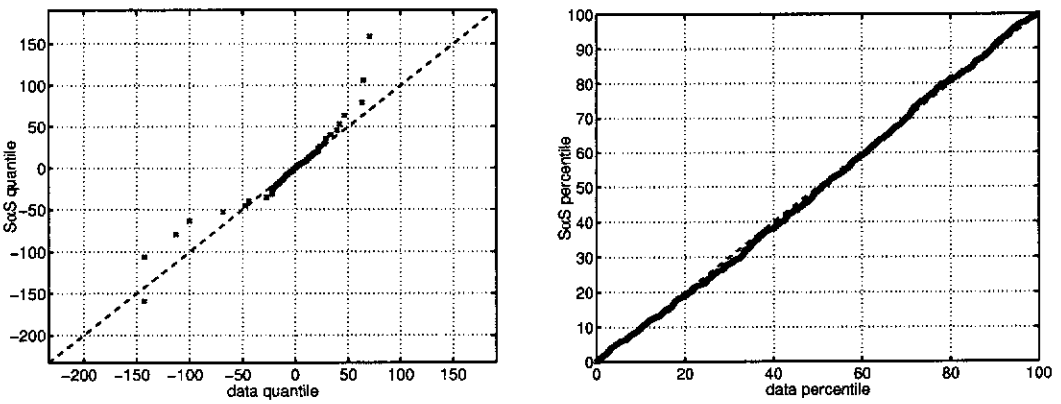


Figure 3.1: Testing for $S\alpha S$ distributions using QQ plots (left) and PP plots (right) where $N = 1000$ observations were drawn from a standard $S\alpha S$ distribution with $\alpha = 1$.

### 3.3.2 Chi-Squared Tests

$\chi^2$ tests group the observations into bins and then compare observed with expected counts under the null distribution. The test statistic is the squared difference between counts, suitably normalised such that the asymptotic distribution is $\chi^2$. In the literature, corrections to the classical $\chi^2$ statistic can be found for the case where the distributional parameters must be estimated, as can guidelines for selecting the bins [38]. In general $\chi^2$ tests are not as powerful as edf tests such as the Kolmogorov-Smirnov test.

### 3.3.3 Empirical Distribution Function Tests

Edf tests are generic procedures which measure the distance between the ecdf and the cdf under the null hypothesis [38]. In this sense, probability plots and $\chi^2$ tests can be viewed as special cases of edf tests, but are generally not regarded as such.

Edf tests may be classified according to the nature of their test statistic, the most common falling into two classes, those based on supremum statistics and those based on quadratic statistics.

Supremum statistics measure the vertical difference between the ecdf $\hat{G}_X(x)$ and the null distribution $F_X(x;\boldsymbol{\theta})$,

$$D_{KS}^{+} = \sup_x \left( \hat{G}_X(x) - F(x;\boldsymbol{\theta}) \right) \tag{3.16}$$

$$D_{KS}^{-} = \sup_x \left( F(x;\boldsymbol{\theta}) - \hat{G}_X(x) \right), \tag{3.17}$$

the well known Kolmogorov-Smirnov test being based on the statistic

$$D_{KS} = \sup_x \left| \hat{G}_X(x) - F(x;\boldsymbol{\theta}) \right| = \max(D_{KS}^+, D_{KS}^-). \tag{3.18}$$

Quadratic statistics, also known as the Cramér-von Mises family of statistics, measure the weighted integrated squared error between the ecdf and the cdf under the null hypothesis,

$$Q = N \int_{-\infty}^{\infty} \left( \hat{G}_X(x) - F(x;\boldsymbol{\theta}) \right)^2 w(x)\, dF(x;\boldsymbol{\theta}), \tag{3.19}$$

where $w(x)$ is a weighting function. Various quadratic statistics are obtained through selection of $w(x)$. These include the Cramér-von Mises statistic, $Q_{CM}$, where $w(x) = 1$ and the Anderson-Darling statistic, $Q_{AD}$, where $w(x) = (F(x;\boldsymbol{\theta})(1 - F(x;\boldsymbol{\theta})))^{-1}$ [38].

Evaluation of the test statistics are simplified by using the probability integral transform, $Z_n = F_X(X_n;\boldsymbol{\theta})$, which transforms the statistics to a uniform distribution on $[0,1]$, $\mathrm{U}(0,1)$, under the null hypothesis. The general problem then becomes one of testing for a uniform distribution.

The asymptotic null distributions of the edf statistics $D_{KS}$, $Q_{CM}$ and $Q_{AD}$ are known when the null distribution is fully specified, while corrections are available for finite samples. If distributional parameters are estimated, the asymptotic distributions are incorrect since the $Z_n$ are not $\mathrm{U}(0,1)$ distributed under the null hypothesis. In general, the null distributions of the test statistics will then depend on quantities such as $F_X(x;\boldsymbol{\theta})$, the true parameters, the estimator, and the sample size. Critical values must then be computed through Monte Carlo simulation [38].

The power of edf tests is dependent on the statistic used and the null distribution. In general the Kolmogorov-Smirnov statistic is less powerful than the Cramér-von Mises or Anderson-Darling statistics. Compared to the Cramér-von Mises statistic, the Anderson-Darling statistic gives more weight to extreme observations through $w(x)$ and tends to be more powerful at detecting departures in the tails [38].

### 3.3.4 Property Based Tests

#### 3.3.4.1 Testing for the Tail Behaviour

One of the first tests for $\alpha S$ distributions was motivated by the need to determine whether they could model econometric data [117, 47]. The test utilised the

tail behaviour of $\alpha S$ distributions, as defined in Property 1, with a subjective judgement being made on whether the observations followed a power law in the tails. To demonstrate, consider the case $\beta = 0$. From Property 1,

$$\log(\mathsf{Pr}[X > x]) = a_0 + a_1 \log(x) \tag{3.20}$$

for $\alpha < 2$ as $x \to \infty$, where $a_0 = \log(C_\alpha c^\alpha)$ and $a_1 = -\alpha$. Under the null hypothesis a plot of the estimated cdf tails should appear linear on a logarithmic scale. This technique has been used for exploratory work, though one could carry out a statistical analysis by ascertaining the fit of the model (3.20) through an $F$ test based on the residuals of a linear regression. It should be noted that the data in such a regression is heteroscedastic and correlated. A question that arises when using such a test is how to define the tail region of the data. How far out in the tails this property becomes apparent has never been answered satisfactorily, though very large data sets may be required. Furthermore, the most extreme samples often have to be excluded as they do not follow (3.20) closely due to their large, and in some cases infinite, variance.

An example is shown in Figure 3.2 (left) where under the null hypothesis $N = 1000$ observations were drawn from a standard $S\alpha S$ distribution with $\alpha = 1$. Observations between the $90^{\text{th}}$ and $99^{\text{th}}$ percentiles were taken as representative of the upper tail. Against this is a linear regression for which the estimated slope was $-1.03$, very near the theoretical slope of $-\alpha = -1$.

#### 3.3.4.2 Testing for Infinite Variance

The lack of a finite variance for non-Gaussian $\alpha S$ distributions leads to a test based on the running sample variance, $Y_n$, of the observations,

$$Y_n = \frac{1}{n} \sum_{i=1}^{n} \left( X_i - \frac{1}{n} \sum_{j=1}^{n} X_j \right)^2 . \tag{3.21}$$

For distributions with finite variance, $Y_n$ will converge to the population variance by the law of large numbers as $n \to \infty$, but will diverge for distributions with infinite variance.

As an example, Figure 3.2 (right) shows the running variance for the same scenario as in Figure 3.2 (left). Note how the running variance continues to increase in sharp jumps. This is due to the occurrence of an observation sufficiently large in magnitude to dominate the sample, for impulsive noise the probability of such an occurrence is non-negligible even for large sample sizes.

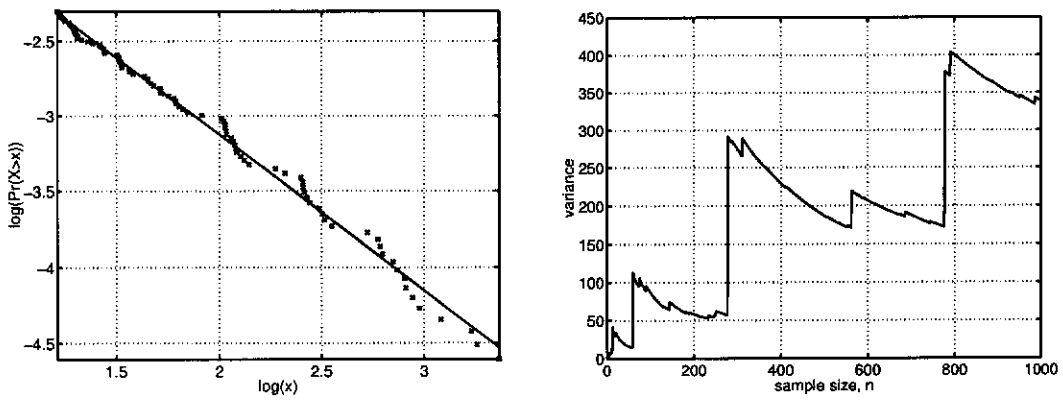


Figure 3.2: Testing for $\alpha S$ distributions using the known tail behaviour (left) and the running variance (right) where $N = 1000$ observations were drawn from a standard $S\alpha S$ distribution with $\alpha = 1$.

### 3.3.5 Techniques based on the Empirical Characteristic Function

Goodness-of-fit tests based on the ecf are primarily motivated by the mathematical tractability of the $\alpha S$ cf compared to the pdf or cdf. Furthermore, the cf retains the information contained in the latter representations while the ecf possesses favourable properties such as strong consistency and asymptotic normality (the terms 'Gaussian' and 'normal' are used interchangeably throughout this thesis), see Appendix A for a brief review of cf and ecf theory.

The goodness-of-fit is based on the degree with which the ecf, $\hat{\phi}_X(\omega)$, matches the cf, $\phi_X(\omega)$, under the null hypothesis. In this respect ecf techniques are similar in nature to edf tests and hence similar measures of distance are used. For example, Kolmogorov-Smirnov type measures can be developed,

$$\sup_{\omega} \left| \hat{\phi}_X(\omega) - \phi_X(\omega) \right| \tag{3.22}$$

or Cramér-von Mises type measures,

$$\int_{-\infty}^{\infty} \left| \hat{\phi}(\omega) - \phi_X(\omega) \right|^2 v(\omega)\, d\omega, \tag{3.23}$$

where $v(\omega)$ is a weighting function [38]. A Kolmogorov-Smirnov type statistic was used to test for the Gaussian distribution in [11, 197], while a quadratic statistic was used in [69].

A supremum based test for $\alpha S$ distributions was proposed in [28, 29, 200] where the bootstrap was used to estimate critical values, avoiding the extensive Monte Carlo simulations otherwise required.

Goodness-of-fit tests conceptually similar to the Cramér-von Mises tests have been developed which measure the quadratic error between the ecf and the cf, which is known under the null hypothesis, for several values of $\omega$. By utilising the asymptotic Gaussianity of the ecf and its known covariance structure under the null hypothesis, the quadratic statistic was shown to be asymptotically $\chi^2$ under the null hypothesis. The case of a simple null hypothesis was considered in [100], this was generalised to the case when parameters must be estimated under the null hypothesis in [104].

A comparison between all these approaches appears in Table 3.1. Simpler, less powerful tests tend to admit non-$\alpha S$ distributions under the null hypothesis, while more powerful tests require calculation of the cdf and critical values. For $\alpha S$ distributions calculation of the cdf is a computationally expensive task. Determining critical values for the edf and ecf tests by Monte Carlo simulation is very computationally expensive, but less so for the ecf test if the bootstrap is used.

Table 3.1: Advantages and disadvantages of the various goodness-of-fit tests.

| Test type | Disadvantages | Advantages |
|---|---|---|
| Tail test | Admits all distributions whose tails decay algebraically<br>Excludes the $\alpha S$ Gaussian distribution<br>Subjective | Simple |
| Running variance | Admits all distributions with infinite variance<br>Excludes the $\alpha S$ Gaussian distribution<br>Subjective | Simple |
| Probability plots | Requires cdf and estimators for shape parameters<br>QQ plot not recommended for a heavy tailed null<br>Subjective | Moderately simple |
| $\chi^2$ tests | Loses information when grouping observations into bins, selection of bins<br>Requires cdf and estimators for shape parameters | Quantitative |
| edf tests | Critical values must be found by Monte Carlo<br>Requires cdf and estimators for shape parameters | Powerful |
| ecf tests | Critical values must be found<br>Requires estimators for shape parameters | Powerful |

## 3.4 Testing for Symmetric Alpha Stable Distributions using the Stability Property

### 3.4.1 Proposed Test

The stability property as stated in Definition 2 has been exploited in testing for $\alpha S$ distributions. The concept is to split the sample into a number of non-overlapping segments and then to estimate the characteristic exponent when the segments are summed. Under the null hypothesis of stability the characteristic exponent is invariant to how many segments are taken, given statistical variations caused by a finite sample.

In the statistical literature this stability test has been suggested as an exploratory technique [48, 64, 118] where the characteristic exponents from sums of 2 to 10 segments were compared subjectively [50, 141, 144, 171]. Recently, there has been some effort towards developing a statistically rigorous stability test. In [105] a stability test was proposed where the test statistic was asymptotically $\chi^2$ distributed, contingent on the condition that the imaginary part of the cf was nonzero. This condition is not met by $S\alpha S$ distributions which are of concern here. In [143] it was tested whether there was a linear trend in the estimates of $\alpha$ as the number of segments summed increased, thresholds were calculated using Monte Carlo analysis. In both [105] and [143] the number of segments required was chosen heuristically. Here, the stability test is presented in a hypothesis testing framework giving a quantitative statistical test, but first, the question of how many segments should be used is addressed.

When using the stability test the number of segments to be compared must be chosen. While an optimal choice might maximise the power of the test, the unspecified nature of the alternative hypothesis makes this decision difficult. Guidelines for the maximum number of segments are considered later, however, the minimum number of segments necessary is determined from the following result [196],

**Property 3** *It is not necessary to ensure that (3.2) holds for all $n \geq 2$, rather, a necessary and sufficient condition for $X$ to have an $\alpha S$ distribution is if*

$$\begin{aligned} X_1 + X_2 &\stackrel{d}{=} C_2 X + D_2 \\ X_1 + X_2 + X_3 &\stackrel{d}{=} C_3 X + D_3, \end{aligned} \tag{3.24}$$

*where $C_2, C_3 > 0$, $D_2, D_3 \in \mathbb{R}$ and $X_1$, $X_2$ and $X_3$ are independent copies of $X$. That is, to confirm Definition 2 only for $n = 2, 3$.*

The essence of this property is that by repeated application of (3.24) it is possible to arrive at (3.2) for the general case $n \geq 2$. That (3.24) is necessary implies it must be fulfilled, while sufficiency guarantees that there are no other criteria which must be satisfied. The property is then complete within itself as a definition for $\alpha S$ distributions.

Since the characteristic exponents from sums of only 2 and 3 non-overlapping segments are compared to confirm stability, define

$$\begin{aligned} Z_2 &= X_1 + X_2 \\ Z_3 &= X_1 + X_2 + X_3 \end{aligned} \tag{3.25}$$

and let $\alpha_X$, $\alpha_{Z_2}$ and $\alpha_{Z_3}$ be the characteristic exponents of $X$, $Z_2$ and $Z_3$ respectively. Then it follows that $\alpha_X = \alpha_{Z_2}$ and $\alpha_X = \alpha_{Z_3}$ given that $X$ is $S\alpha S$. To test for $\alpha S$ distributions, a similar comparison between the skewness parameters must be made as well.

In a hypothesis testing framework the test for $S\alpha S$ distributions is set up as,

$$\begin{aligned} \mathsf{H} &: \alpha_{Z_2} = \alpha_X \cap \alpha_{Z_3} = \alpha_X \\ \mathsf{K} &: \alpha_{Z_2} \neq \alpha_X \cup \alpha_{Z_3} \neq \alpha_X. \end{aligned} \tag{3.26}$$

Under the null hypothesis $X$ has a $S\alpha S$ distribution while under the alternative hypothesis the distribution is not $S\alpha S$ but otherwise unspecified. The hypotheses suggest the test statistics $T_2 = \alpha_{Z_2} - \alpha_X$ and $T_3 = \alpha_{Z_3} - \alpha_X$, so that the two hypotheses $T_2 = 0$ and $T_3 = 0$ comprise the global null hypothesis. To maintain a global level of significance when the global null hypothesis is comprised of more than a single hypothesis a multiple hypothesis test (MHT) procedure is used. MHTs can also closely maintain the global level when the hypotheses comprising the global null hypothesis are dependent. A review on the theory of MHTs, detailing the specific procedures used here, appears in Appendix E.

#### 3.4.1.1 A Note on Multiple Hypothesis Test Procedures for the Stability Test

For the stability test the global null hypothesis consists of only two individual hypotheses, $\mathsf{H}_1 : T_2 = \alpha_{Z_2} - \alpha_X = 0$ and $\mathsf{H}_2 : T_3 = \alpha_{Z_3} - \alpha_X = 0$, which simplifies the MHTs reviewed in Appendix E. Let the p-values corresponding to $\mathsf{H}_1$ and $\mathsf{H}_2$ be $\mathcal{P}_1$ and $\mathcal{P}_2$ respectively and denote the ordered p-values as $\mathcal{P}_{(1)} \leq \mathcal{P}_{(2)}$.

The Bonferroni and Holm procedures accept the global null hypothesis if $\mathcal{P}_{(1)} > \zeta/2$ and reject it otherwise. The Hochberg and Simes procedures accept

the global null hypothesis if $\mathcal{P}_{(1)} > \zeta/2$ and $\mathcal{P}_{(2)} > \zeta$, rejecting it otherwise. It can be seen that the later test must be less conservative due to the extra condition $\mathcal{P}_{(2)} > \zeta$. Hommel's procedure is the most conservative of all, accepting the global null hypothesis if $\mathcal{P}_{(1)} > \zeta/3$ and $\mathcal{P}_{(2)} > 2\zeta/3$, otherwise rejecting it.

The procedures of Hochberg and Simes are then equally most powerful, followed by those of Bonferroni and Holm and lastly Hommel. As Hommel's procedure was found to be significantly more conservative, it is not considered any further, only the procedures of Bonferroni and Hochberg are used.

### 3.4.2 Development of the Test Statistics

To evaluate the test statistics and implement the stability test, several independent realisations from the distribution which generated the observations are required. To obtain these independent realisations the observations are separated into segments. There are then two ways the test statistics can be obtained, consider the statistic $T_2$,

1. Separate the observations into three equi-length segments, then arbitrarily assign the first segment to $X$ and the others to $Z_2$.

2. Let all the observations represent $X$ and obtain $Z_2$ by separating the observations into two equal segments.

The first approach ensures $X$, $X_1$ and $X_2$ are iid. The second introduces dependence between $X$ and $X_1$, $X_2$. The resulting dependence between $\hat{\alpha}_X$ and $\hat{\alpha}_{Z_2}$ confounds theoretical analysis and may reduce the power of the test since the estimates will be correlated, reducing their difference under the alternative hypothesis. However, the reduced variance of $\hat{\alpha}_X$ counteracts this and, as will be shown, results in a more powerful test. The same comments apply to $T_3$.

Let $X_{mM}(l) = X((m-1)L + l)$, $l = 1, \ldots, L$, denote segment $m$ after the iid observations $X(n)$, $n = 1, \ldots, N$, have been split into $M$ segments of length $L = \lfloor N/M \rfloor$, where $\lfloor \cdot \rfloor$ denotes the integer part. The separation of the observations into the $M$ segments is shown in Figure 3.3. Define $Z_{kM}$ as the elementwise sum

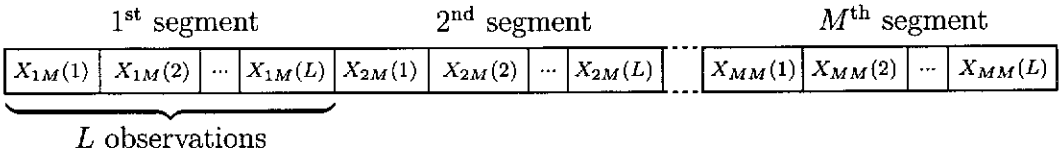


Figure 3.3: Separation of the observations into the $M$ segments.

of $k$ segments from $M$. Although arbitrary, the following ordering is used

$$\begin{aligned} Z_{22}(l) &= X_{12}(l) + X_{22}(l) \\ Z_{13}(l) &= X_{13}(l) \\ Z_{23}(l) &= X_{23}(l) + X_{33}(l) \\ Z_{33}(l) &= X_{13}(l) + X_{23}(l) + X_{33}(l) \\ Z_{14}(l) &= X_{14}(l) \\ Z_{34}(l) &= X_{24}(l) + X_{34}(l) + X_{44}(l), \qquad l = 1, \ldots, L, \end{aligned} \tag{3.27}$$

giving the four test statistics

$$\begin{aligned} T_{22} &= \alpha_{Z_{22}} - \alpha_X \\ T_{23} &= \alpha_{Z_{23}} - \alpha_{Z_{13}} \\ T_{33} &= \alpha_{Z_{33}} - \alpha_X \\ T_{34} &= \alpha_{Z_{34}} - \alpha_{Z_{14}}. \end{aligned} \tag{3.28}$$

The characteristic exponents forming the test statistic must be estimated, for which a variety of estimators exist. An early approach utilised the algebraic tail behaviour of $\alpha S$ distributions [47], but was left undeveloped in favour of superior estimators including those based on ML [23, 27, 39, 42, 140], Bayesian analysis [32], FLOMs [26, 40, 115, 116], sample fractiles or order statistics [50, 119, 120, 122, 181, 184] and cf domain methods [10, 70, 144, 145, 153, 176]. Koutrouvelis' method [101] further developed regression based cf domain estimators [10, 153] and offers an excellent compromise between computational cost and performance in comparison to most other methods. While refinements suggested in [30, 97, 102] improve performance over some regions of the parameter space, this mainly occurs for skewed $\alpha S$ distributions. As only $S\alpha S$ distributions are of concern here Koutrouvelis' original method is used exclusively.

Critical values of the test will be found in two ways. The first follows the traditional approach of finding the asymptotic distributions of the statistics, in doing so the asymptotic distributions of Koutrouvelis' estimator are derived. For finite samples asymptotic results are likely to be inaccurate, unfortunately the finite sample distributions are non-trivial, making an analytical approach prohibitively complex. Instead, a computational method known as the bootstrap is used to estimate the finite sample distributions.

### 3.4.3 Evaluation of Critical Values

#### 3.4.3.1 Asymptotic Theory

Koutrouvelis' estimator is based on the cf domain representation for a $S\alpha S$ random variable,

$$\phi_X(\omega) = \exp(-|c\omega|^\alpha)\,. \tag{3.29}$$

Applying the transformation $\log(-\log|\cdot|^2)$ and using the ecf $\hat{\phi}_X(\omega)$ as an estimate for $\phi_X(\omega)$, $\alpha$ and $c$ are estimated through a linear regression over several points, $\omega_1, \ldots, \omega_K$, in the cf domain,

$$\log\left(-\log\left|\hat{\phi}_X(\omega_k)\right|^2\right) = \log(2c^\alpha) + \alpha\log|\omega_k|\,, \qquad k = 1, \ldots, K. \tag{3.30}$$

Since $\hat{\phi}_X(\omega)$ converges to $\phi_X(\omega)$ with probability one (w.p.1) and the parameters of an $\alpha S$ distribution are unique, both $\hat{\alpha}$ and $\hat{c}$ are strongly consistent.

Koutrouvelis' estimator was developed for $\alpha S$ random variables while only $S\alpha S$ random variables are of concern here. As the $S\alpha S$ cf is real, $\mathrm{Re}(\hat{\phi}_X(\omega))^2$ is used in place of $|\hat{\phi}_X(\omega)|^2$ herein. For finite samples the imaginary part of the ecf is nonzero, giving $|\hat{\phi}_X(\omega)|^2$ an asymptotic variance, as $\omega \to \infty$, twice that of $\mathrm{Re}(\hat{\phi}_X(\omega))^2$. Clearly, taking the real part of the ecf should improve the performance of the estimator and this is the approach taken. This requires some minor modifications to Koutrouvelis' original procedure in order to optimise performance, details and a demonstrated improvement of this estimator are contained in Appendix B. Use of only the real part of the ecf also makes it possible to derive the asymptotic distributions of Koutrouvelis' parameter estimator for the case of $S\alpha S$ distributions. To clarify, in [103], the asymptotic distributions of Koutrouvelis' parameter estimator for $\alpha S$ distributions, using $|\hat{\phi}_X(\omega)|^2$, were derived under the assumption that the imaginary part of the cf is nonzero. The condition is necessary because only then is $|\hat{\phi}_X(\omega)|^2$ asymptotically Gaussian. If the cf has a zero imaginary part, as for $S\alpha S$ distributions, then $|\hat{\phi}_X(\omega)|^2$ is asymptotically non-Gaussian distributed and the derivation in [103] becomes invalid.

Continuing the asymptotic analysis using the regression model

$$\log\left(-\log\mathrm{Re}\left(\hat{\phi}_X(\omega_k)\right)^2\right) = \log(2c^\alpha) + \alpha\log|\omega_k|\,, \qquad k = 1, \ldots, K, \tag{3.31}$$

let $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_K)^\mathsf{T}$, $\hat{\boldsymbol{\phi}}_X = (\hat{\phi}_X(\omega_1), \ldots, \hat{\phi}_X(\omega_K))^\mathsf{T}$ and $\mathbf{1}_K$ be a $K$-length vector of ones. Then the regression is represented by

$$\boldsymbol{u} = \boldsymbol{V}\boldsymbol{\theta} + \boldsymbol{\varepsilon} \tag{3.32}$$

where $\boldsymbol{u} = \log(-\log \mathrm{Re}(\hat{\boldsymbol{\phi}}_X)^2)$, $\boldsymbol{V} = (\mathbf{1}_K, \log|\boldsymbol{\omega}|)$, $\boldsymbol{\theta} = (\log(2c^\alpha), \alpha)^\mathsf{T}$ and $\boldsymbol{\varepsilon}$ is a $K$ length vector of correlated disturbances. The least squares solution for $\hat{\boldsymbol{\theta}}$ is

$$\hat{\boldsymbol{\theta}} = \boldsymbol{V}^\dagger \boldsymbol{u}, \tag{3.33}$$

where $\boldsymbol{V}^\dagger = (\boldsymbol{V}^\mathsf{T}\boldsymbol{V})^{-1}\boldsymbol{V}^\mathsf{T}$ is the pseudoinverse of $\boldsymbol{V}$.

To derive the asymptotic distribution of $\hat{\boldsymbol{\theta}}$, the asymptotic distribution of $\log(-\log \mathrm{Re}(\hat{\boldsymbol{\phi}}_X)^2)$ is required. The following theorem regarding functions of asymptotically multivariate normal (MVN) random vectors is needed ([159] Theorem 3.3A),

**Theorem 1** *Given that* $\boldsymbol{X} = (X_1, \ldots, X_K)^\mathsf{T} \overset{a}{\sim} \mathrm{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, *where* $\overset{a}{\sim}$ *stands for asymptotically distributed as, such that* $\boldsymbol{\Sigma} \to 0$ *as* $N \to \infty$, *let* $\boldsymbol{g}(\boldsymbol{x}) = (g_1(\boldsymbol{x}), \ldots, g_M(\boldsymbol{x}))^\mathsf{T}$ *where every* $g_m$ *is a vector valued function with nonzero derivative at* $\boldsymbol{x} = \boldsymbol{\mu}$. *Define the Jacobian of the transformation at* $\boldsymbol{x} = \boldsymbol{\mu}$ *as*

$$(\boldsymbol{D})_{km} = \left.\frac{\partial g_m}{\partial x_k}\right|_{\boldsymbol{x}=\boldsymbol{\mu}}. \tag{3.34}$$

*Then*

$$\boldsymbol{g}(\boldsymbol{x}) \overset{a}{\sim} \mathrm{MVN}(\boldsymbol{g}(\boldsymbol{\mu}), \boldsymbol{D}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{D}). \tag{3.35}$$

From Appendix A the real part of the ecf is asymptotically MVN,

$$\mathrm{Re}\left(\hat{\boldsymbol{\phi}}_X\right) \overset{a}{\sim} \mathrm{MVN}\left(\boldsymbol{\mu}_{\mathrm{Re}(\boldsymbol{\phi})}, \boldsymbol{R}_{\mathrm{Re}(\boldsymbol{\phi})}\right) \tag{3.36}$$

with $\boldsymbol{R}_{\mathrm{Re}(\boldsymbol{\phi})} \to 0$ as $N \to \infty$.

Define $g_m(\boldsymbol{x}) = g(x_m)$ where $g(x) = \log(-\log x^2)$. The partial derivatives $g'(x) = 1/(x \log|x|)$ are nonzero at $\left(\boldsymbol{\mu}_{\mathrm{Re}(\boldsymbol{\phi})}\right)_k$ for a $S\alpha S$ cf evaluated at $0 < \omega_k < \infty$ and it follows from the above theorem that

$$\log\left(-\log \mathrm{Re}\left(\hat{\boldsymbol{\phi}}_X\right)^2\right) \overset{a}{\sim} \mathrm{MVN}(\boldsymbol{\mu}_g, \boldsymbol{R}_g), \tag{3.37}$$

where $\boldsymbol{\mu}_g = \boldsymbol{g}\left(\boldsymbol{\mu}_{\mathrm{Re}(\boldsymbol{\phi})}\right)$, $\boldsymbol{R}_g = \boldsymbol{D}_g{}^\mathsf{T}\boldsymbol{R}_{\mathrm{Re}(\boldsymbol{\phi})}\boldsymbol{D}_g$ and $\boldsymbol{D}_g$ is a diagonal matrix with $(\boldsymbol{D}_g)_{kk} = g'\left(\left(\boldsymbol{\mu}_{\mathrm{Re}(\boldsymbol{\phi})}\right)_k\right)$.

From (3.33) it can be seen that $\hat{\boldsymbol{\theta}}$ will be asymptotically MVN as it is a linear transformation of the asymptotically MVN $\boldsymbol{u}$,

$$\hat{\boldsymbol{\theta}} \overset{a}{\sim} \mathrm{MVN}(\boldsymbol{\mu}_\theta, \boldsymbol{R}_\theta), \tag{3.38}$$

where $\boldsymbol{\mu}_\theta = \boldsymbol{V}^\dagger\boldsymbol{\mu}_g$ and $\boldsymbol{R}_\theta = \boldsymbol{V}^\dagger\boldsymbol{R}_g\boldsymbol{V}^{\dagger\mathsf{T}}$.

A final transformation of $\hat{\boldsymbol{\theta}}$ yields the asymptotic joint distribution of $\hat{c}$ and $\hat{\alpha}$,

$$\hat{c} = g_1(\hat{\boldsymbol{\theta}}) = (e^{\hat{\theta}_1}/2)^{1/\hat{\theta}_2} \tag{3.39}$$

$$\hat{\alpha} = g_2(\hat{\boldsymbol{\theta}}) = \hat{\theta}_2. \tag{3.40}$$

Giving the Jacobian of the transformation at $\boldsymbol{\mu}_\theta$ as

$$\boldsymbol{D} = \begin{pmatrix} \frac{1}{(\mu_\theta)_2}(e^{(\mu_\theta)_1}/2)^{1/(\mu_\theta)_2} & 0 \\ -\frac{1}{(\mu_\theta)_2}(e^{(\mu_\theta)_1}/2)^{1/(\mu_\theta)_2}\log(e^{(\mu_\theta)_1}/2)^{1/(\mu_\theta)_2} & 1 \end{pmatrix}, \tag{3.41}$$

so that

$$\begin{pmatrix} \hat{c} \\ \hat{\alpha} \end{pmatrix} \overset{a}{\sim} \mathrm{MVN}(\boldsymbol{\mu}_{c\alpha}, \boldsymbol{R}_{c\alpha}) \tag{3.42}$$

with $\boldsymbol{\mu}_{c\alpha} = \left((e^{(\mu_\theta)_1}/2)^{1/(\mu_\theta)_2}, (\mu_\theta)_2\right)^\mathsf{T}$ and $\boldsymbol{R}_{c\alpha} = \boldsymbol{D}^\mathsf{T}\boldsymbol{R}_\theta\boldsymbol{D}$.

The estimates $\hat{\alpha}$ and $\hat{c}$ are asymptotically normal. The estimator is asymptotically unbiased and for reasonable sample sizes of $N > 200$ the bias is negligible.

The asymptotic distributions of the test statistics under the null hypothesis can now be found. Take the test statistic $\hat{T}_2 = \hat{\alpha}_{Z_2} - \hat{\alpha}_X$ where the segments from which $\hat{\alpha}_X$ and $\hat{\alpha}_{Z_2}$ were estimated do not overlap, $\hat{\alpha}_X$ and $\hat{\alpha}_{Z_2}$ are then independent with distributions

$$\hat{\alpha}_X \overset{a}{\sim} \mathrm{N}(\mu_{\alpha_X}(\alpha, c, \boldsymbol{\omega}), \sigma^2_{\alpha_X}(\alpha, c, \boldsymbol{\omega})), \tag{3.43}$$

$$\hat{\alpha}_{Z_2} \overset{a}{\sim} \mathrm{N}(\mu_{\alpha_{Z_2}}(\alpha, c, \boldsymbol{\omega}), \sigma^2_{\alpha_{Z_2}}(\alpha, c, \boldsymbol{\omega})). \tag{3.44}$$

Since the estimator is asymptotically unbiased,

$$\hat{T}_2 \overset{a}{\sim} \mathrm{N}\left(0, \sigma^2_{\alpha_X} + \sigma^2_{\alpha_{Z_2}}\right). \tag{3.45}$$

Likewise, the asymptotic distribution of the test statistic $\hat{T}_3 = \hat{\alpha}_{Z_3} - \hat{\alpha}_X$ for nonoverlapping segments is

$$\hat{T}_3 \overset{a}{\sim} \mathrm{N}\left(0, \sigma^2_{\alpha_X} + \sigma^2_{\alpha_{Z_3}}\right). \tag{3.46}$$

The hypothesis test is then carried out by testing $T_2 = 0 \cap T_3 = 0$ using a MHT procedure.

#### 3.4.3.2 Bootstrap Estimator

The bootstrap is a general statistical method which allows the sampling distribution of a descriptive statistic to be estimated. It replaces complicated, and

in many cases intractable, theoretical analysis with computational power. Given that access to increasingly powerful computers is growing, so too is the use of the bootstrap for estimation of confidence intervals and the related problem of hypothesis testing. Though the bootstrap makes heavy use of computation, the validity of its results in terms of such measures as consistency are backed by a substantial amount of theory. References and further details on the bootstrap necessary for its use herein are contained in Appendix D.

The motivation for use of the bootstrap is twofold. First, the bootstrap more closely controls the set level of a test for finite samples as compared to when thresholds are derived asymptotically. Second, the bootstrap may account for dependence between the estimates of the characteristic exponents when they are calculated from overlapping samples. This follows from the plug-in-principle of the bootstrap espoused in [43, 148]. In essence this principle states that the statistical relationships between the sample statistics are generally mirrored in the bootstrap statistics since the mechanism by which they are calculated from a sample is the same in both cases. The mechanism here is comprised of the method by which the sample is separated into segments and the estimator employed in estimating the characteristic exponents of these segments.

When the observations arise from a distribution within the domain of attraction of a non-Gaussian $\alpha S$ law, standard nonparametric bootstrap resampling fails to give good estimates for the sampling distribution [148]. Instead, the parametric bootstrap is used as it is known to behave correctly [43].

The parametric bootstrap assumes that observations are easily generated under the null hypothesis, given any necessary parameters. Unknown parameters may be estimated from the observations. Observations generated under the null then comprise the bootstrap resamples and the statistic of interest is recalculated for each of these bootstrap resamples. A sufficient number of bootstrap statistics computed in this fashion form the bootstrap estimate for the distribution of the statistic under the null hypothesis. For the stability test estimates for $\alpha$ and $c$ are found using Koutrouvelis' procedure, while $S\alpha S$ random variables are generated using the computationally efficient method of Chambers, Mallows and Stuck [34]. The parametric bootstrap procedure is shown in Figure 3.4.

The parametric bootstrap has been successfully employed in carrying out hypothesis tests on the parameters of an $\alpha S$ distribution and in a cf domain goodness-of-fit test for $\alpha S$ distributions [28, 29, 200]. In both cases the parametric bootstrap was used to determine the null distributions of a test statistic where, under the null hypothesis, the observations were $\alpha S$.

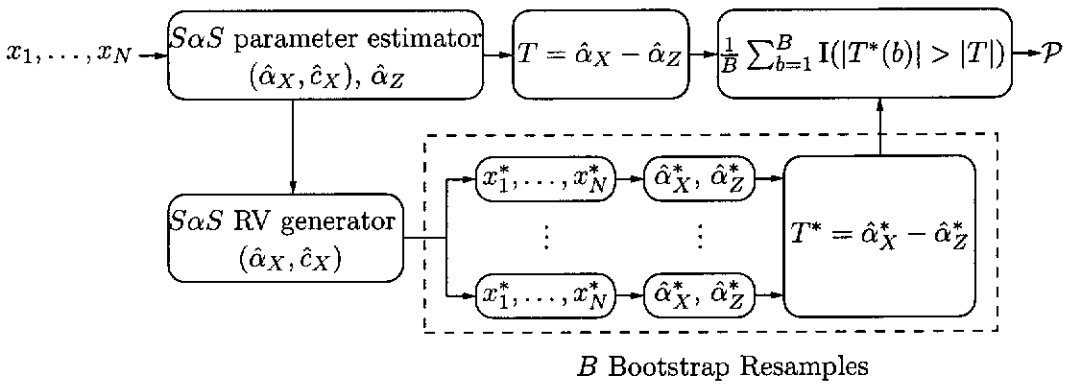


Figure 3.4: The parametric bootstrap procedure for the stability test.

The procedure as outlined in Figure 3.4 uses a non-pivotal test statistic. Pivoting gives a test statistic that is asymptotically independent of any unknown parameters and results in a bootstrap procedure which more closely maintains the set level, see Appendix D for more details. A pivotal statistic was found by normalising the non-pivotal statistic by an estimate of the standard error. Though a nested bootstrap procedure may be used for this purpose, Monte Carlo estimates found offline were used to reduce computational complexity.

## 3.5 Experiments

Sample sizes of approximately 500 are necessary for the asymptotically derived thresholds to be accurate, since the observations are separated into at most $M = 4$ segments, sample sizes of $N = 2000$ were used.

Results are shown for each of the individual test statistics $T_{22}, T_{33}, T_{23}, T_{34}$ and for the Bonferroni and Hochberg MHTs formed from $T_{22}, T_{33}$ and $T_{23}, T_{34}$. The probabilities of false alarm ($P_{FA}$) and detection ($P_D$) were found from 1000 independent Monte Carlo realisations.

### 3.5.1 Maintenance of the Set Level

#### 3.5.1.1 Stability Test with Asymptotic Distributions

Figure 3.5 shows the $P_{FA}$ for the stability test when the asymptotic distributions of the test statistics are used.

When MHTs are not used and the characteristic exponents are estimated from non-overlapping segments, as for the test statistics $T_{23}$ and $T_{34}$, the set level

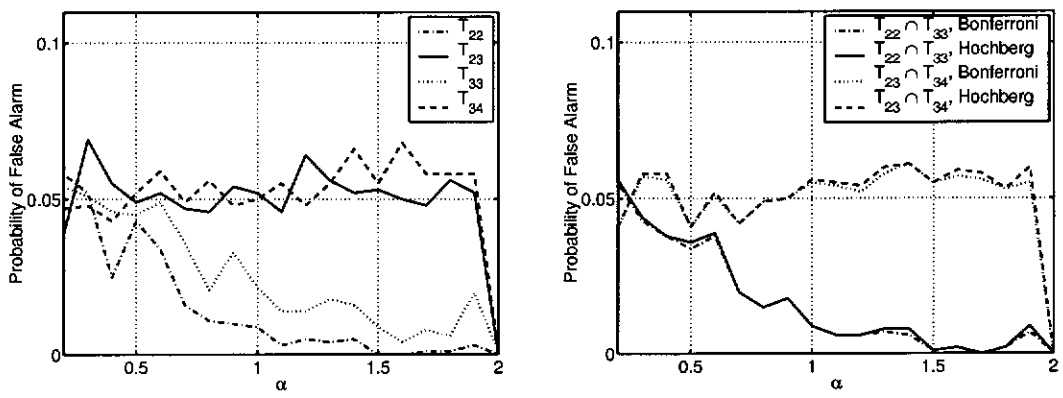


Figure 3.5: Probability of false alarm for the stability test using asymptotic statistics. No MHTs (left) and with MHTs (right).

is closely maintained. An exception occurs when $\alpha$ is near 2 and the $P_{FA}$ is markedly less than the set level. The cause is a combination of finite sample size and the rapid change in the variance of $\hat{\alpha}$ for $\alpha$ near 2. When $\hat{\alpha}$ is less than the true $\alpha \approx 2$, the estimate of $\mathsf{Var}[\hat{\alpha}]$ is much larger than the true variance, making the test more conservative. The effect is reduced with increasing sample size and asymptotically the set level is maintained at $\alpha = 2$.

When the characteristic exponents are estimated from overlapping segments, as for the test statistics $T_{22}$ and $T_{33}$, the set level is not maintained, the tests being far too conservative. Take $T_{22} = \alpha_{Z_{22}} - \alpha_X$, which has variance $\mathsf{Var}[T_{22}] = \mathsf{Var}[\alpha_{Z_{22}}] + \mathsf{Var}[\alpha_X] - 2\mathsf{Cov}[\alpha_{Z_{22}}, \alpha_X]$. Under the null hypothesis there is a high degree of positive correlation between the estimates of these two characteristic exponents. The asymptotic estimate for the variance, which does not account for the covariance between the two characteristic exponents, is much larger than the true value leading to a conservative test.

When MHTs are used the results are very similar to when they are not. MHTs based on non-overlapping segments maintain the set level except for $\alpha$ near 2, while MHTs based on overlapping segments are far too conservative. The type of MHT, Bonferroni or Hochberg, has very little effect on the $P_{FA}$, though as expected Hochberg's procedure is consistently less conservative than Bonferroni's procedure.

#### 3.5.1.2 Stability Test with Bootstrap Distributions

Figure 3.6 shows the $P_{FA}$ for the stability test when the bootstrap distributions of non-pivotal test statistics are used, Figure 3.7 shows the same for pivotal

statistics.

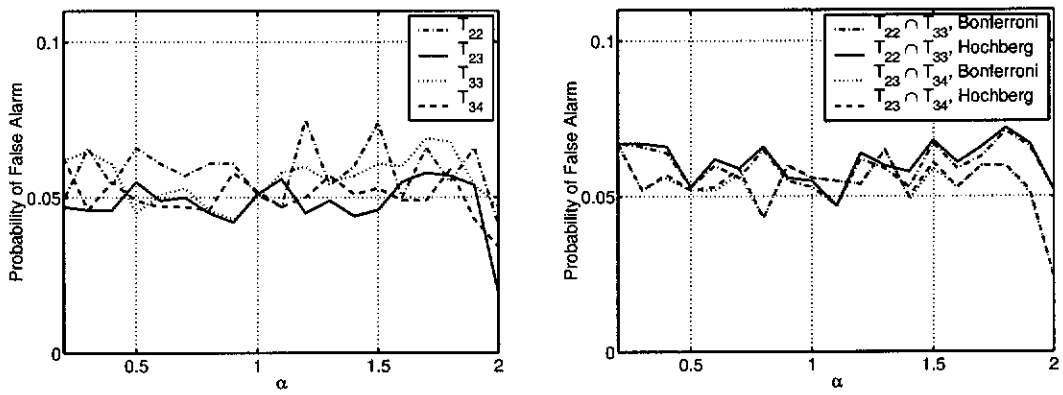


Figure 3.6: Probability of false alarm for the stability test using the bootstrap. No MHTs (left) and with MHTs (right).

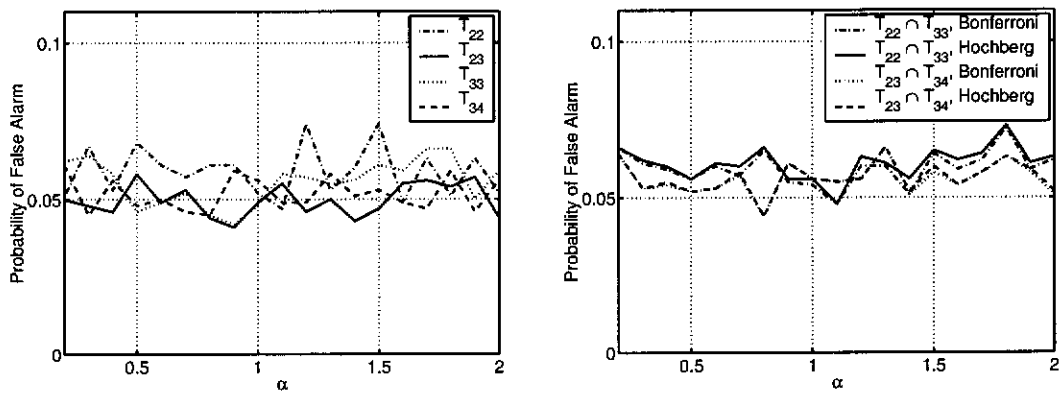


Figure 3.7: Probability of false alarm for the stability test using the bootstrap with pivoting. No MHTs (left) and with MHTs (right).

When non-pivotal statistics are used the $P_{FA}$ is close to the set level except for $\alpha$ near 2, where it can be as low as 2%. Note that this is for both overlapping and non-overlapping segments, showing that the bootstrap accounts for dependence between the characteristic exponents estimated from overlapping segments. The drop in $P_{FA}$ for $\alpha$ near 2 is again attributed to the rapid change in variance of $\hat{\alpha}$ in this region. The use of a pivotal statistic eliminates this drop in the $P_{FA}$ for $\alpha$ near 2, so that the set level is always closely maintained.

There does not appear to be a significant change in $P_{FA}$ when MHTs are used, though again Hochberg's procedure is slightly less conservative than Bonferroni's procedure.

#### 3.5.1.3 Empirical Distribution Function statistics

Empirical thresholds for the edf tests were calculated for $0.5 \leq \alpha \leq 2$ over $10,000$ Monte Carlo simulations. Calculation of the thresholds below $\alpha = 0.5$ becomes computationally expensive. The $P_{FA}$ for the statistics $D_{KS}$, $Q_{CM}$ and $Q_{AD}$ are shown in Figure 3.8. As the empirical thresholds were only found for $\alpha \geq 0.5$, results were limited to $\alpha \geq 0.6$ to ensure that $\hat{\alpha} \geq 0.5$. All three tests appear to approximately maintain the set level.

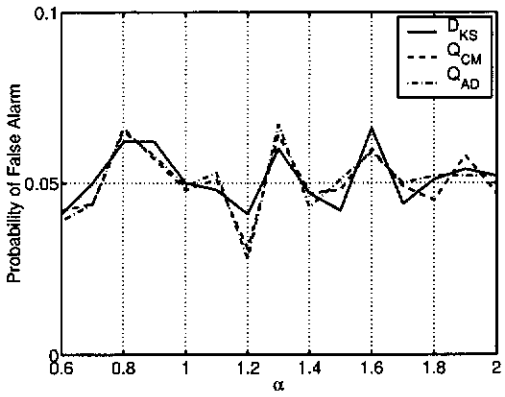


Figure 3.8: Probability of false alarm for several edf tests.

### 3.5.2 Power of the Tests

The tests were evaluated under a variety of alternative distributions:

- The chi-squared distribution with 1, 2 and 4 degrees of freedom, $\chi^2_1$, $\chi^2_2$, $\chi^2_4$.
- Student's $t$ distribution with 2, 3, 4 and 10 degrees of freedom, $t_2$, $t_3$, $t_4$, $t_{10}$.
- The Laplace distribution with zero mean and unit variance, $\mathrm{L}(0, 1)$.
- The uniform distribution on $[-1, 1]$, $\mathrm{U}(-1, 1)$.
- The beta distribution with both parameters equal to 4, $\beta(4, 4)$.
- The $\varepsilon-\mathrm{mix}(\varepsilon, \kappa)$ distribution for several values of $(\varepsilon, \kappa)$.
- The Gaussian mixture distribution with pdf

$$f_X(x; \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{w}) = \sum_{k=1}^{K} w_k \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(x-\mu_k)^2}{2\sigma_k^2}\right).$$

Three cases of the Gaussian mixture distribution, denoted by $GM_1$, $GM_2$ and $GM_3$, were used, the parameters chosen are shown in Table 3.2. All are heavier tailed than the Gaussian, slightly skewed, and except for $GM_3$, nonzero mean distributions.

Table 3.2: Parameters of the Gaussian mixture distributions used in assessing $P_D$.

| | $\mu$ | $\sigma$ | $w$ |
|---|---|---|---|
| $GM_1$ | $(-0.1, 0.2, 0.5, 1)^\top$ | $(0.25, 0.4, 0.9, 1.6)^\top$ | $(0.5, 0.1, 0.2, 0.2)^\top$ |
| $GM_2$ | $(-0.3, 0.7, 0.8)^\top$ | $(0.5, 0.25, 0.4)^\top$ | $(0.2, 0.5, 0.3)^\top$ |
| $GM_3$ | $(-0.2, 0, 0.2, 0.3)^\top$ | $(0.25, 0.4, 0.9, 1.6)^\top$ | $(0.5, 0.1, 0.2, 0.2)^\top$ |

Table 3.3 shows the power of the stability test using asymptotic distributions. Tables 3.4 and 3.5 show the same using bootstrap distributions for non-pivotal and pivotal statistics respectively. The power of the edf tests is shown in Table 3.6.

Table 3.3: Power of the stability test using asymptotic statistics.

| | | | | | $T_{22} \cap T_{33}$ | | $T_{23} \cap T_{34}$ | |
|---|---|---|---|---|---|---|---|---|
| | $T_{22}$ | $T_{23}$ | $T_{33}$ | $T_{34}$ | Bonf. | Hoch. | Bonf. | Hoch. |
| $\chi^2_1$ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| $\chi^2_2$ | 0.18 | 0.11 | 1.00 | 0.98 | 1.00 | 1.00 | 0.97 | 0.97 |
| $\chi^2_4$ | 1.00 | 1.00 | 0.99 | 0.89 | 1.00 | 1.00 | 1.00 | 1.00 |
| $t_2$ | 0.15 | 0.14 | 0.36 | 0.21 | 0.25 | 0.27 | 0.21 | 0.22 |
| $t_3$ | 0.25 | 0.21 | 0.54 | 0.28 | 0.44 | 0.45 | 0.30 | 0.31 |
| $t_4$ | 0.28 | 0.18 | 0.53 | 0.29 | 0.42 | 0.45 | 0.30 | 0.30 |
| $t_{10}$ | 0.12 | 0.08 | 0.31 | 0.09 | 0.27 | 0.27 | 0.09 | 0.09 |
| $L(0,1)$ | 0.97 | 0.75 | 1.00 | 0.86 | 1.00 | 1.00 | 0.92 | 0.93 |
| $U(-1,1)$ | 0.00 | 0.00 | 0.37 | 0.08 | 0.17 | 0.17 | 0.02 | 0.02 |
| $\beta(4,4)$ | 0.25 | 0.15 | 0.25 | 0.09 | 0.30 | 0.30 | 0.13 | 0.13 |
| $\varepsilon-\text{mix}(0.01, 10)$ | 0.00 | 0.03 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| $\varepsilon-\text{mix}(0.01, 100)$ | 0.01 | 0.12 | 0.04 | 0.16 | 0.02 | 0.02 | 0.18 | 0.19 |
| $\varepsilon-\text{mix}(0.1, 10)$ | 0.01 | 0.03 | 0.11 | 0.09 | 0.07 | 0.07 | 0.06 | 0.06 |
| $\varepsilon-\text{mix}(0.1, 100)$ | 0.87 | 0.47 | 0.98 | 0.65 | 0.97 | 0.98 | 0.66 | 0.68 |
| $GM_1$ | 1.00 | 0.98 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| $GM_2$ | 0.80 | 0.45 | 0.24 | 0.11 | 0.73 | 0.73 | 0.33 | 0.35 |
| $GM_3$ | 1.00 | 0.91 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 |

Table 3.4: Power of the stability test using the bootstrap.

| | | | | | $T_{22} \cap T_{33}$ | | $T_{23} \cap T_{34}$ | |
|---|---|---|---|---|---|---|---|---|
| | $T_{22}$ | $T_{23}$ | $T_{33}$ | $T_{34}$ | Bonf. | Hoch. | Bonf. | Hoch. |
| $\chi^2_1$ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| $\chi^2_2$ | 0.60 | 0.33 | 1.00 | 0.98 | 1.00 | 1.00 | 0.98 | 0.98 |
| $\chi^2_4$ | 1.00 | 1.00 | 1.00 | 0.97 | 1.00 | 1.00 | 1.00 | 1.00 |
| $t_2$ | 0.51 | 0.16 | 0.60 | 0.26 | 0.62 | 0.63 | 0.26 | 0.27 |
| $t_3$ | 0.71 | 0.17 | 0.76 | 0.28 | 0.81 | 0.82 | 0.27 | 0.28 |
| $t_4$ | 0.71 | 0.16 | 0.79 | 0.25 | 0.82 | 0.83 | 0.26 | 0.26 |
| $t_{10}$ | 0.40 | 0.06 | 0.35 | 0.06 | 0.37 | 0.40 | 0.07 | 0.07 |
| $\mathrm{L}(0,1)$ | 1.00 | 0.71 | 1.00 | 0.85 | 1.00 | 1.00 | 0.90 | 0.91 |
| $\mathrm{U}(-1,1)$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| $\beta(4,4)$ | 0.87 | 0.56 | 0.89 | 0.56 | 0.95 | 0.95 | 0.71 | 0.72 |
| $\varepsilon-\mathrm{mix}(0.01,10)$ | 0.04 | 0.03 | 0.03 | 0.03 | 0.04 | 0.04 | 0.02 | 0.03 |
| $\varepsilon-\mathrm{mix}(0.01,100)$ | 0.36 | 0.14 | 0.46 | 0.18 | 0.43 | 0.45 | 0.20 | 0.20 |
| $\varepsilon-\mathrm{mix}(0.1,10)$ | 0.13 | 0.03 | 0.31 | 0.07 | 0.26 | 0.26 | 0.06 | 0.06 |
| $\varepsilon-\mathrm{mix}(0.1,100)$ | 1.00 | 0.46 | 1.00 | 0.63 | 1.00 | 1.00 | 0.67 | 0.68 |
| $\mathrm{GM}_1$ | 1.00 | 0.98 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| $\mathrm{GM}_2$ | 0.93 | 0.68 | 0.52 | 0.24 | 0.92 | 0.93 | 0.63 | 0.64 |
| $\mathrm{GM}_3$ | 1.00 | 0.93 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |

#### 3.5.2.1 Discussion

The stability test based on overlapping segments and asymptotic distributions cannot be compared to the others fairly as it is conservative and does not maintain the set level. This particular case is included only for completeness and is not considered any further.

Some general conclusions can be drawn regarding the power of the statistics $T_{22}$, $T_{33}$, $T_{23}$, $T_{34}$:

- $T_{22}$ and $T_{33}$ are more powerful than $T_{23}$ and $T_{34}$.
- The power of $T_{34}$ exceeds or is comparable to $T_{23}$.
- The power of $T_{33}$ exceeds or is comparable to $T_{22}$.

That $T_{22}$ and $T_{33}$ have a higher power than $T_{23}$ and $T_{34}$ is not unexpected. The larger number of samples reduces the error in the estimates of the charac-

Table 3.5: Power of the stability test using the bootstrap with pivoting.

| | | | | | $T_{22} \cap T_{33}$ | | $T_{23} \cap T_{34}$ | |
|---|---|---|---|---|---|---|---|---|
| | $T_{22}$ | $T_{23}$ | $T_{33}$ | $T_{34}$ | Bonf. | Hoch. | Bonf. | Hoch. |
| $\chi^2_1$ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| $\chi^2_2$ | 0.60 | 0.33 | 1.00 | 0.98 | 1.00 | 1.00 | 0.98 | 0.98 |
| $\chi^2_4$ | 1.00 | 1.00 | 1.00 | 0.97 | 1.00 | 1.00 | 1.00 | 1.00 |
| $t_2$ | 0.50 | 0.16 | 0.60 | 0.26 | 0.62 | 0.63 | 0.26 | 0.27 |
| $t_3$ | 0.71 | 0.17 | 0.76 | 0.28 | 0.80 | 0.82 | 0.27 | 0.27 |
| $t_4$ | 0.71 | 0.16 | 0.79 | 0.25 | 0.82 | 0.83 | 0.25 | 0.26 |
| $t_{10}$ | 0.40 | 0.06 | 0.35 | 0.07 | 0.38 | 0.40 | 0.08 | 0.08 |
| $\mathrm{L}(0,1)$ | 1.00 | 0.71 | 1.00 | 0.85 | 1.00 | 1.00 | 0.89 | 0.91 |
| $\mathrm{U}(-1,1)$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| $\beta(4,4)$ | 0.87 | 0.56 | 0.89 | 0.56 | 0.95 | 0.95 | 0.72 | 0.72 |
| $\varepsilon-\mathrm{mix}(0.01,10)$ | 0.04 | 0.04 | 0.04 | 0.03 | 0.04 | 0.04 | 0.03 | 0.03 |
| $\varepsilon-\mathrm{mix}(0.01,100)$ | 0.35 | 0.13 | 0.46 | 0.17 | 0.42 | 0.44 | 0.20 | 0.21 |
| $\varepsilon-\mathrm{mix}(0.1,10)$ | 0.12 | 0.02 | 0.30 | 0.07 | 0.26 | 0.26 | 0.06 | 0.06 |
| $\varepsilon-\mathrm{mix}(0.1,100)$ | 1.00 | 0.47 | 1.00 | 0.63 | 1.00 | 1.00 | 0.67 | 0.68 |
| $\mathrm{GM}_1$ | 1.00 | 0.98 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| $\mathrm{GM}_2$ | 0.93 | 0.68 | 0.52 | 0.24 | 0.92 | 0.93 | 0.64 | 0.64 |
| $\mathrm{GM}_3$ | 1.00 | 0.93 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |

teristic exponents, reducing the variance of the test statistics under the null and alternative hypotheses and the error when estimating the threshold.

The latter two points have no simple explanation, but an intuitive reason is offered. Consider what happens as the number of segments grows very large, but the number of samples in each stays constant. Assuming that the sum of these segments converges to a limiting distribution, this limiting distribution is $\alpha S$ by the GCLT. As the number of segments increases, the difference in power between a test which uses $M$ and $M+1$ segments will diminish. Since in any practical implementation of the test the number of samples in each segment decreases as the number of segments increase, the power of the test must drop. This must be balanced against the use of too few segments, that is, the sum of three segments may have a distribution, and hence a characteristic exponent, farther from the original than for two segments.

To summarise, for finite samples there is a trade-off between using few segments to obtain better estimates of the characteristic exponents whose true values

Table 3.6: Power of several edf tests.

| | $D_{KS}$ | $Q_{CM}$ | $Q_{AD}$ |
|---|---|---|---|
| $\chi^2_1$ | 1.00 | 1.00 | 1.00 |
| $\chi^2_2$ | 1.00 | 1.00 | 1.00 |
| $\chi^2_4$ | 1.00 | 1.00 | 1.00 |
| $t_2$ | 0.05 | 0.05 | 0.06 |
| $t_3$ | 0.05 | 0.06 | 0.10 |
| $t_4$ | 0.07 | 0.06 | 0.09 |
| $t_{10}$ | 0.07 | 0.07 | 0.08 |
| $\mathrm{L}(0,1)$ | 0.42 | 0.27 | 1.00 |
| $\mathrm{U}(-1,1)$ | 1.00 | 1.00 | 1.00 |
| $\beta(4,4)$ | 1.00 | 1.00 | 1.00 |
| $\varepsilon-\mathrm{mix}(0.01,10)$ | 0.04 | 0.04 | 0.05 |
| $\varepsilon-\mathrm{mix}(0.01,100)$ | 0.07 | 0.06 | 0.06 |
| $\varepsilon-\mathrm{mix}(0.1,10)$ | 0.05 | 0.05 | 0.06 |
| $\varepsilon-\mathrm{mix}(0.1,100)$ | 0.24 | 0.12 | 1.00 |
| $\mathrm{GM}_1$ | 1.00 | 1.00 | 1.00 |
| $\mathrm{GM}_2$ | 1.00 | 1.00 | 1.00 |
| $\mathrm{GM}_3$ | 1.00 | 1.00 | 1.00 |

are close, and using many segments to obtain poorer estimates of the characteristic exponents whose true values are farther apart. The attained power shows that in the majority of cases, the trade-off is in favour of 3 segments as this results in a more powerful test.

Overall, the MHTs are more powerful compared to the test statistics taken separately. When using the bootstrap to calculate p-values, the power of the test based on $T_{22}, T_{33}$ exceeds that of $T_{23}, T_{34}$. Again, this is due to the larger number of samples available for estimation of the characteristic exponents. The power of tests that use Hochberg's procedure is slightly but consistently greater than when Bonferroni's procedure is used. This is to be expected as Hochberg's procedure is less conservative than Bonferroni's.

Pivoting does not appear to have a significant effect on the power of the tests, a favourable result since pivoting leads to a conservative test for $\alpha$ near 2.

As a final comparison, consider the power of the following tests

1. $T_{23} \cap T_{34}$ using asymptotic distributions and Hochberg's procedure.

2. $T_{22} \cap T_{33}$ using bootstrap distributions, a pivotal test statistic and Hochberg's procedure.

3. The Anderson-Darling edf goodness-of-fit test, $Q_{AD}$.

Each yielded the best performance in terms of maintenance of the set level and power for the stability test using asymptotic and bootstrap distributions and the edf tests, results are shown in Table 3.7.

Table 3.7: Comparison of the three types of test. The stability test using $T_{23} \cap T_{34}$ and Hochberg's procedure with asymptotically derived thresholds (Asymptotic); the stability test using $T_{22} \cap T_{33}$ and Hochberg's procedure with pivotal statistics and bootstrap derived thresholds (Bootstrap); the Anderson-Darling edf goodness-of-fit test (edf).

| | Asymptotic | Bootstrap | edf |
|---|---|---|---|
| $\chi^2_1$ | 1.00 | 1.00 | 1.00 |
| $\chi^2_2$ | 0.97 | 1.00 | 1.00 |
| $\chi^2_4$ | 1.00 | 1.00 | 1.00 |
| $t_2$ | 0.22 | 0.63 | 0.06 |
| $t_3$ | 0.31 | 0.82 | 0.10 |
| $t_4$ | 0.30 | 0.83 | 0.09 |
| $t_{10}$ | 0.09 | 0.40 | 0.08 |
| $\mathrm{L}(0,1)$ | 0.93 | 1.00 | 1.00 |
| $\mathrm{U}(-1,1)$ | 0.02 | 0.00 | 1.00 |
| $\beta(4,4)$ | 0.13 | 0.95 | 1.00 |
| $\varepsilon-\mathrm{mix}(0.01,10)$ | 0.01 | 0.04 | 0.05 |
| $\varepsilon-\mathrm{mix}(0.01,100)$ | 0.19 | 0.44 | 0.06 |
| $\varepsilon-\mathrm{mix}(0.1,10)$ | 0.06 | 0.26 | 0.06 |
| $\varepsilon-\mathrm{mix}(0.1,100)$ | 0.68 | 1.00 | 0.97 |
| $\mathrm{GM}_1$ | 1.00 | 1.00 | 1.00 |
| $\mathrm{GM}_2$ | 0.35 | 0.93 | 1.00 |
| $\mathrm{GM}_3$ | 1.00 | 1.00 | 1.00 |

The stability test using bootstrap thresholds is generally the most powerful. An exception occurs for the $\mathrm{U}(-1,1)$ distribution where the edf test always rejected the null hypothesis, but the stability test did not. If was found that the estimated characteristic exponent for this distribution was always near 2, even for sums of two and three segments. For $S\alpha S$ distributions the optimal number

of points at which to sample the ecf for $\alpha \approx 2$ is 2, which meant the ecf was not sampled sufficiently far out to detect a significant difference between the $U(-1,1)$ cf and the $S\alpha S$ cf for $\alpha \approx 2$. Increasing the number of points to 10, and hence sampling the ecf at larger values of $\omega$ where there is a significant difference between these cfs, increased the power to approximately 0.98. Of course, a uniform distribution is easily rejected by cursory examination.

Of more importance is the ability of the test to reject symmetric heavy tailed alternatives which are very difficult to distinguish from $S\alpha S$ distributions. These include the $t$, Laplace and $\varepsilon-$mix distributions. Here the stability test using bootstrap thresholds clearly outperforms the edf test.

## 3.6 Summary

In this Chapter a test for $S\alpha S$ distributions was developed based on the stability property of $\alpha S$ distributions. Thresholds for the test were found from asymptotic theory and estimated using bootstrap techniques. The bootstrap technique maintains the set level of the test while achieving high power in detecting alternative distributions which possess very similar tail behaviour to $S\alpha S$ distributions.

# Chapter 4

# Estimation for the Symmetric Alpha Stable Gaussian Sum Distribution

*Your theory is crazy, but it's not crazy enough to be true.*
*— Niels Bohr*

A sum of symmetric alpha stable and Gaussian noise has been proposed as a realistic model for impulsive phenomena which also includes thermal noise. Just as in the case of the symmetric alpha stable distribution, the probability density function of the symmetric alpha stable Gaussian sum distribution does not exist in closed form while its characteristic function is relatively simple. The difficulty of evaluating the probability density function has limited the use of the maximum likelihood estimator in favour of characteristic function domain estimators. An estimator is proposed for this distribution based on a nonlinear weighted least squares regression in the characteristic function domain.

## 4.1 Introduction

Impulsive noise generally enters a system from external sources, in communications systems the channel introduces impulsive noise. However, the receiver is an internal source of thermal Gaussian noise. The result is an additive combination of impulsive and Gaussian noise.

Modelling the impulsive component by a $S\alpha S$ distribution leads directly to the following model for the total noise $X$ which is said to follow a $S\alpha S$ Gaussian sum ($S\alpha SG$) distribution,

$$X = X_{S\alpha S} + X_G, \tag{4.1}$$

where $X_{S\alpha S}$ and $X_G$ are the $S\alpha S$ and Gaussian components respectively. The components are assumed to be mutually independent as they arise from not just physically distinct, but fundamentally different sources. When sampling from this distribution it will be assumed the samples are iid.

The $S\alpha SG$ distribution has not been investigated as extensively as $S\alpha S$ distributions or the other common impulsive noise models. Most of the results on detection can be found in [5, 6, 82, 83, 87, 86] while estimation is addressed in [85].

This Chapter begins by defining the $S\alpha SG$ distribution and describing some of its properties in Section 4.2. Existing estimators for the $S\alpha SG$ distribution are reviewed in Section 4.3 before the proposed NWLS estimator is developed in Section 4.4. Finally, the performance of these estimators is assessed in Section 4.5.

## 4.2 The Symmetric Alpha Stable Gaussian Sum Distribution

The $S\alpha SG$ pdf is obtained by convolving a $S\alpha S$ pdf and a Gaussian pdf. The same difficulties which arise with the $S\alpha S$ pdf exist here as well, making cf domain techniques attractive. The $S\alpha SG$ distribution is easily defined in the cf domain as

$$\phi_X(\omega) = \phi_{X_{S\alpha S}}(\omega)\phi_{X_G}(\omega) = \exp\left(-\gamma_{S\alpha S}|\omega|^\alpha - \gamma_G\omega^2\right). \tag{4.2}$$

$\gamma_{S\alpha S} > 0$ and $\gamma_G > 0$ are the dispersions of $X_{S\alpha S}$ and $X_G$ respectively. The dispersions are related to the scale, $c$, of $X_{S\alpha S}$ and the variance, $\sigma^2$, of $X_G$ as $\gamma_{S\alpha S} = c^\alpha$ and $\gamma_G = \sigma^2/2$ respectively. $0 < \alpha \leq 2$ is the characteristic exponent of $X_{S\alpha S}$ and controls the rate of decay of the pdf. For $\alpha = 2$, $X$ has a Gaussian distribution with variance $2(\gamma_{S\alpha S} + \gamma_G)$. Note that the $S\alpha SG$ distribution is not a mixture model in the usual sense as the pdf is the convolution of a $S\alpha S$ pdf and a Gaussian pdf, not the sum.

### 4.2.1 Properties

In general, the properties of the $S\alpha SG$ distribution are similar to those of the $S\alpha S$ distribution.

#### 4.2.1.1 The Symmetric Alpha Stable Gaussian Sum Pdf

The $S\alpha SG$ pdf is obtained from an inverse Fourier transform of the cf,

$$f_X(x) = \frac{1}{\pi}\int_0^\infty \phi_X(\omega)\cos(x\omega)\, d\omega. \tag{4.3}$$

This integral is not expressible in closed form for $\alpha < 2$.

#### 4.2.1.2 Fractional Lower Order Moments

The FLOMs, $\mathsf{E}[|X|^p]$, exist for $0 < p < \alpha$. Unlike $S\alpha S$ distributions, there is no simple expression for the FLOMs, but they can be found by numerical integration

$$\mathsf{E}[|X|^p] = \frac{p}{\Gamma(1-p)\cos\left(\frac{\pi}{2}p\right)} \int_0^\infty \frac{1-\phi_X(\omega)}{\omega^{p+1}}\, d\omega, \tag{4.4}$$

where $\Gamma(\cdot)$ is Euler's gamma function.

#### 4.2.1.3 Moments

From the existence of the FLOMs, all moments of order 2 and above do not exist for $\alpha < 2$. The mean only exists for $\alpha > 1$.

## 4.3 Existing Estimators

Existing estimators for $S\alpha SG$ distributions are the moment and integrated MSE estimators [85]. Maximum likelihood estimators (MLE) are not considered due to the large computational burden.

### 4.3.1 Integrated Mean Square Error Estimator

A general method proposed for estimation in the cf domain is to minimise the weighted integrated MSE (IMSE) between the ecf $\hat{\phi}_X(\omega)$ and cf $\phi_X(\omega;\boldsymbol{\theta})$ with respect to the parameters $\boldsymbol{\theta}$ [70],

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \int_{-\infty}^{\infty} \left|\hat{\phi}_X(\omega) - \phi_X(\omega;\boldsymbol{\theta})\right|^2 W(\omega)\, d\omega. \tag{4.5}$$

The IMSE estimator has been used with some success on $\alpha S$ distributions [145] and was implemented for the $S\alpha SG$ distribution in [85]. The weighting function $W(\omega)$ is used to counteract the poor statistical behaviour of the ecf for large $|\omega|$, namely $\lim_{|\omega|\to\infty} \mathsf{Var}[\hat{\phi}(\omega)] = 1/N$. A Gaussian weighting function, $W(\omega) = \exp(-\omega^2)$, is suggested so that Gauss-Hermite quadrature can be used for the numerical integration. When applied to the $S\alpha SG$ distribution this method suffers from ill convergence due to the presence of many local minima, correct convergence is only obtained if the initial values are close to the true parameters. Under some conditions IMSE estimates are consistent and asymptotically Gaussian [176, 53].

### 4.3.2 Moment Type Estimator

The moment type estimator is similar to one developed for $\alpha S$ distributions [153] and is based on the relation [85]

$$\frac{\omega_1^{\Delta_\alpha} - \omega_2^{\Delta_\alpha} + \omega_1^{-\Delta_\alpha} - \omega_2^{-\Delta_\alpha}}{\omega_3^{\Delta_\alpha} - \omega_4^{\Delta_\alpha} + \omega_3^{-\Delta_\alpha} - \omega_4^{-\Delta_\alpha}} = \frac{\log|\phi_X(\omega_1)|\log|\phi_X(\omega_1^{-1})| - \log|\phi_X(\omega_2)|\log|\phi_X(\omega_2^{-1})|}{\log|\phi_X(\omega_3)|\log|\phi_X(\omega_3^{-1})| - \log|\phi_X(\omega_4)|\log|\phi_X(\omega_4^{-1})|}, \quad (4.6)$$

where $\Delta_\alpha = 2 - \alpha$. Replacing $\phi_X(\omega)$ with $\hat{\phi}_X(\omega)$ and solving the nonlinear equation in $\Delta_\alpha$ gives $\hat{\alpha}$. Similarly, the dispersions are found from the relation

$$\begin{bmatrix} \log\phi_X(\omega_1) \\ \log\phi_X(\omega_2) \end{bmatrix} = -\begin{bmatrix} |\omega_1|^\alpha & \omega_1^2 \\ |\omega_2|^\alpha & \omega_2^2 \end{bmatrix} \begin{bmatrix} \gamma_{S\alpha S} \\ \gamma_G \end{bmatrix}. \quad (4.7)$$

As any suitable combination of $\omega_1, \omega_2, \omega_3, \omega_4$ can be used, it is suggested to average the results from all 6 possible combinations. Performance of the estimates is dependent on where the ecf is sampled. Since $\omega$ and $1/\omega$ are used it is recommended to cluster the points about 1 to avoid the increasing variance of the ecf for large $\omega$. The minimal separation between samples is based on the requirement that the matrix in (4.7) not be ill-conditioned. Final selection of the points was based on the parameters and Monte Carlo performance. When either of the dispersions are sufficiently far from unity, rescaling by $\mathrm{E}[|X|^p]^{1/p}$ where $p \leq \alpha/2$ is required.

An advantage of the moment type estimator over the IMSE estimator is its reduced computational complexity. Both require large data sets on the order of $10^4$.

## 4.4 Nonlinear Weighted Least Squares Estimation

Consider the following expression as the basis of an estimator for the $S\alpha SG$ distribution,

$$-\log|\phi_X(\omega)|^2 = 2\gamma_{S\alpha S}|\omega|^\alpha + 2\gamma_G\omega^2. \quad (4.8)$$

The relation is nonlinear in $\alpha$, but linear in both dispersions, motivating the use of a nonlinear regression to determine the parameters. This parallels the generalisation of the moment estimator for $\alpha S$ distributions [153] to the regression estimator of Koutrouvelis [101], which greatly improved the performance.

An improvement is anticipated not only because the ecf can be sampled at an arbitrary number of points, but since there is no longer any constraint to cluster the points about $\omega = 1$ and advantage can be taken of the superior performance of the ecf near $\omega = 0$.

Taking a regression over $K$ points and using the same notation as in Chapter 3 gives

$$\boldsymbol{u} = \boldsymbol{V}(\alpha)\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \tag{4.9}$$

where $\boldsymbol{u} = -\log|\hat{\phi}_X|^2$, $\boldsymbol{V}(\alpha) = 2(|\boldsymbol{\omega}|^\alpha, |\boldsymbol{\omega}|^2)$ and $\boldsymbol{\gamma} = (\gamma_{S\alpha S}, \gamma_G)^\mathsf{T}$.

The estimates are found by minimising the residual sum of squares (RSS) $||\hat{\boldsymbol{\varepsilon}}||^2 = ||\boldsymbol{u} - \boldsymbol{V}(\alpha)\boldsymbol{\gamma}||^2$. The dimension of this search is easily reduced to one. For a given $\alpha$, $\boldsymbol{\gamma}$ is a linear function of the *true* $\boldsymbol{u}$, so that $\boldsymbol{\gamma} = \boldsymbol{V}^\dagger(\alpha)\boldsymbol{u}$ where $\boldsymbol{V}^\dagger(\alpha) = \left(\boldsymbol{V}^\mathsf{T}(\alpha)\boldsymbol{V}(\alpha)\right)^{-1}\boldsymbol{V}^\mathsf{T}(\alpha)$ [121]. The RSS becomes $\boldsymbol{u}^\mathsf{T}\boldsymbol{V}^\perp(\alpha)\boldsymbol{u}$ where $\boldsymbol{V}^\perp(\alpha) = \boldsymbol{I} - \boldsymbol{V}(\alpha)\boldsymbol{V}^\dagger(\alpha)$ is the projection matrix onto the orthogonal subspace of $\boldsymbol{V}$. The estimator for $\alpha$ is then the one dimensional minimisation,

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}}\ \boldsymbol{u}^\mathsf{T}\boldsymbol{V}^\perp(\alpha)\boldsymbol{u}. \tag{4.10}$$

Given this estimate for $\alpha$, the dispersions are found by linear least squares (LLS),

$$\hat{\boldsymbol{\gamma}} = \boldsymbol{V}^\dagger(\hat{\alpha})\boldsymbol{u}. \tag{4.11}$$

The parameters of a $S\alpha SG$ distribution are unique if $\alpha < 2$ and since $\hat{\phi}_X(\omega)$ converges to $\phi_X(\omega)$ w.p.1, the estimator is strongly consistent for $\alpha < 2$.

Although the elements of the error term $\boldsymbol{\varepsilon}$ are correlated, the nonlinear least squares estimator is easily adapted to account for this to give the final nonlinear weighted least squares (NWLS) estimator. Next, the statistics of $\boldsymbol{\varepsilon}$ are found.

### 4.4.1 Covariance

The covariance of $\boldsymbol{u}$ and hence $\boldsymbol{\varepsilon}$ may be approximated by expanding $-\log|\hat{\phi}(\omega)|^2$ about $\mathsf{E}[|\hat{\phi}(\omega)|^2]$ in a Taylor series up to first order. From Appendices A and C,

$$\mathsf{Cov}\left[-\log\left|\hat{\phi}_X(\omega_i)\right|^2, -\log\left|\hat{\phi}_X(\omega_j)\right|^2\right] \approx \boldsymbol{C} = \frac{\mathsf{Cov}\left[\left|\hat{\phi}_X(\omega_i)\right|^2, \left|\hat{\phi}_X(\omega_j)\right|^2\right]}{\mathsf{E}\left[\left|\hat{\phi}_X(\omega_i)\right|^2\right]\mathsf{E}\left[\left|\hat{\phi}_X(\omega_j)\right|^2\right]}. \tag{4.12}$$

In Figure 4.1 the analytic correlation and covariance are shown for a $S\alpha SG$ distribution with parameters ($\alpha = 1.5, \gamma_{S\alpha S} = 1, \gamma_G = 1$) and a sample size of

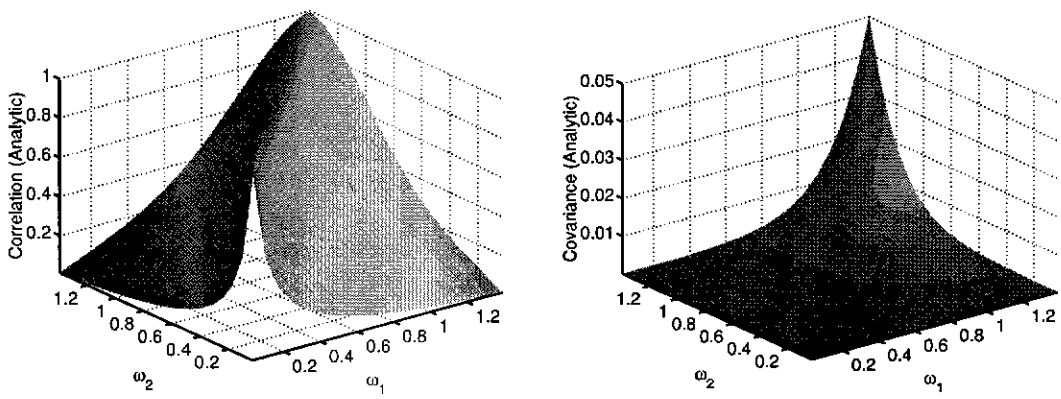


Figure 4.1: The first order Taylor series expansion for the correlation (left) and covariance (right) of $\epsilon$.

$N = 50000$. The correlation structure is clearly nonzero off the main diagonal, while the covariance is dominated by a rapid rise in variance as $\omega$ increases. Incorporating this into the estimator will reduce the weighting of points where the variance is large and lead to improved behaviour.

Figure 4.2 shows the error between the analytic and empirical correlation as estimated from 1000 independent Monte Carlo realisations. The first order Taylor series expansion was found to be accurate over a wide range of the parameters.

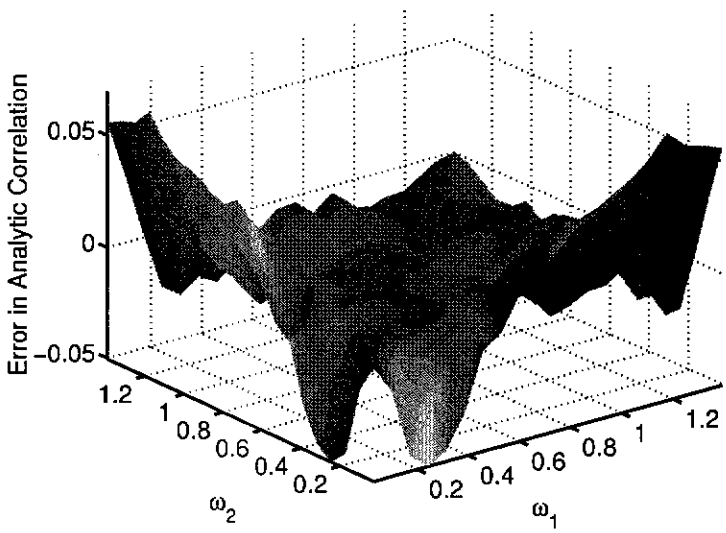


Figure 4.2: Error between the analytic and empirical correlation of $\epsilon$ estimated though Monte Carlo simulation.

The bias in $\mathsf{E}[-\log|\hat{\phi}(\omega)|^2]$ was found to be zero up to first order. A second

order Taylor series expansion gave the bias as $O(N^{-1})$, which is negligible for large $N$. Experiments have shown that accounting for bias does not lead to any improvement.

Given $\boldsymbol{C}$, a NWLS estimate is obtained by solving (4.10) and (4.11) with $\boldsymbol{u}$ and $\boldsymbol{V}$ pre-multiplied by $\boldsymbol{D}$, $\boldsymbol{D}$ being obtained from the Cholesky decomposition $\boldsymbol{D}^{\mathsf{T}}\boldsymbol{D} = \boldsymbol{C}^{-1}$. Since $\boldsymbol{C}$ depends on the unknown parameters $(\alpha, \gamma_{S\alpha SG}, \gamma_G)$, an iterative scheme is employed where at each iteration the parameters and $\boldsymbol{D}$ are updated once. The NWLS procedure is summarised in Table 4.1. Finally, there

Table 4.1: The NWLS Estimator

| |
|---|
| **Step 1.** Obtain initial estimates of the parameters $\hat{\alpha}_0$, $\hat{\gamma}_0$ by nonlinear least squares: Solve (4.10) and (4.11). |
| **Step 2.** Set $i \leftarrow i + 1$. |
| **Step 3.** Calculate $\boldsymbol{D}_i$ using $\hat{\alpha}_i$, $\hat{\gamma}_i$. |
| **Step 4.** Obtain updated estimates of the parameters by NWLS: Solve (4.10) and (4.11) with $\boldsymbol{u}$ replaced by $\boldsymbol{D}_i\boldsymbol{u}$ and $\boldsymbol{V}(\hat{\alpha}_i)$ by $\boldsymbol{D}_i\boldsymbol{V}(\hat{\alpha}_i)$. |
| **Step 5.** Check for convergence: If $\lvert\hat{\gamma}_i - \hat{\gamma}_{i-1}\rvert < 10^{-6}$ for both $\gamma_{S\alpha S}$ and $\gamma_G$, then stop, otherwise go to Step 1. |

is the question of how to sample the ecf, this is considered next.

### 4.4.2 Selection of Empirical Characteristic Function Sample Points

The selection of points at which to sample the ecf is not as straightforward for the $S\alpha SG$ distribution as it was for the $S\alpha S$ distribution. For the latter, rescaling standardised the observations to a $S\alpha S$ distribution with unit dispersion. Estimator performance was then optimised only over $\alpha$. The $S\alpha SG$ distribution has two dispersion parameters, $(\gamma_{S\alpha S}, \gamma_G)$, so that rescaling cannot standardise both components simultaneously.

For the moment estimator the observations were rescaled using FLOMs. The points selected were based on experimental performance and were dependent on all parameters. Here, the region over which the ecf is sampled will be based on the variance of $\boldsymbol{u}$. A similar idea, where the truncation region for Koutrouvelis

estimator was based on the variance of the ecf, has proved successful [30]. To motivate this approach consider Figure 4.3 which shows the analytic variance of $\boldsymbol{u}$ versus $\omega$. The variance of this statistic was always observed to increase with $\omega$

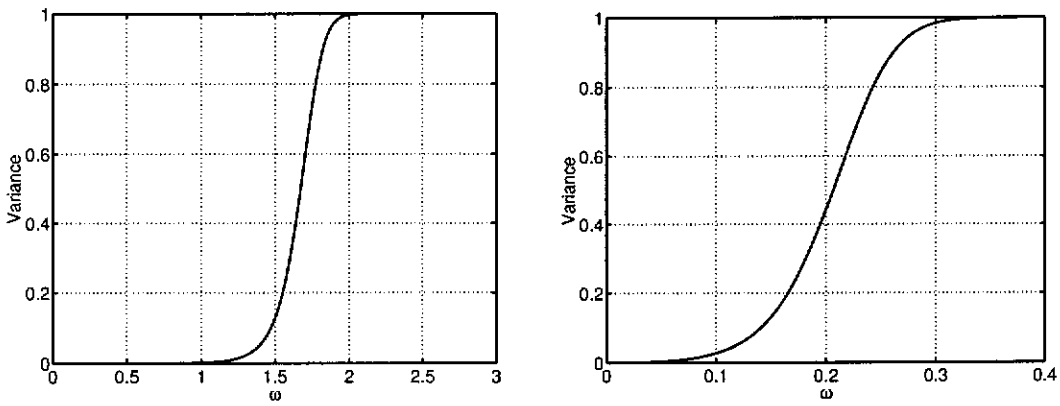


Figure 4.3: Analytic variance of $\epsilon$ for the $S\alpha SG$ distribution with parameters ($\alpha = 1.5, \gamma_{S\alpha S} = 1, \gamma_G = 1$) (left) and ($\alpha = 0.5, \gamma_{S\alpha S} = 10, \gamma_G = 10$) (right).

before reaching a plateau for sufficiently large $\omega$. From the Taylor series expansion this limiting value is $1 - 1/N$, which is approximated by 1 for sufficiently large samples. Including points in the region where the variance can be said to have attained its limiting value produces poor results.

Consider a standardised set of points $\omega'_{LB} \leq \omega'_k \leq 1$, $k = \mathbf{1}, \ldots, K$, which are scaled by the upper bound $\omega_{UB}$ so that $\omega_{LB} \leq \omega_k \leq \omega_{UB}$ where $\omega_k = \omega'_k \omega_{UB}$, $\omega_{LB} = \omega'_{LB} \omega_{UB}$. The bounds, $K$ and location of the points are determined as follows.

$\omega_{UB}$: Experience has shown that choosing a region such that at the upper bound $\mathrm{Var}[|\hat{\phi}_X(\omega)|^2]$ is at most 5% its limiting of 1 yields good performance.

$\omega_{LB}$: The lower bound must be set larger than but as close to zero as possible to avoid numerical problems which arise because $-\log|\hat{\phi}_X(\omega)|^2 \rightarrow \infty$ as $\omega \rightarrow 0$, $\omega_{LB} = 0.02$ was found to be a reasonable choice.

$K$: $K$ should be at least 3 to ensure a unique solution. Too large a $K$ increased computational complexity for little or no gain, $K = 10$ was found to be a good compromise.

Location: For $\alpha S$ distributions, Koutrouvelis showed that there was no gain by spacing the $\omega_k$ nonlinearly. Preliminary experiments have shown

the same to be true here and in the absence of any theoretical motivation for nonlinear spacing the $\omega_k$ are equi-spaced.

## 4.5 Experiments

The proposed NWLS method was compared with the distance and moment type estimators of [85] for a variety of parameter sets. Results are shown for $N = 50000$ over 50 independent Monte Carlo realisations in Tables 4.2 and 4.3.

To ensure the estimates are unique and $\boldsymbol{V}(\alpha)$ is not ill-conditioned, $\hat{\alpha}$ was bounded from above by 1.99. A lower bound of 0.1 was placed on $\hat{\alpha}$, this representing an extreme impulsiveness unlikely in practice. The dispersions were limited to $\hat{\gamma}_{S\alpha S}, \hat{\gamma}_G \in [10^{-8}, 10^3]$.

For the results in Table 4.2 both dispersions are near unity and rescaling is not critical. The NWLS estimates appear to be unbiased and possess a smaller variance than both the IMSE and moment estimators. The only exception occurs when estimating $\gamma_G$ for $(\alpha = 1.5, \gamma_{S\alpha S} = 1, \gamma_G = 1)$ where the variance of the NWLS estimator is 11e-2 compared to 7.7e-2 and 9.0e-2 for the IMSE and moment estimators respectively.

For Table 4.2 the dispersions are far from unity and rescaling becomes critical. Again the NWLS estimates are unbiased and yield the smallest variance when estimating $\alpha$. The variance of the dispersion estimates are comparable except when estimating $\gamma_G$ for $(\alpha = 0.5, \gamma_{S\alpha S} = 1, \gamma_G = 1)$ where the variance is notably larger. Figure 4.4 shows the $S\alpha SG$ cf compared to its $S\alpha S$ and Gaussian components for this case and $(\alpha = 1.5, \gamma_{S\alpha S} = 1, \gamma_G = 1)$. Note how for the former parameter set the $S\alpha S$ component is indistinguishable from the $S\alpha SG$ cf, while the Gaussian component is far from both. Clearly, the $S\alpha S$ component will accurately model observations for this specific $S\alpha SG$ distribution, effectively concealing the Gaussian component and resulting in poor estimates for $\gamma_G$. Only for very large $\omega$ does the $S\alpha S$ component become distinguishable from the $S\alpha SG$ cf, however, very large samples are necessary to ensure the variance of the ecf is sufficiently small for this difference to stand out.

In Figure 4.5 the MSE of the estimates versus $N$ over 500 independent Monte Carlo realisations is shown for $(\alpha = 1.5, \gamma_{S\alpha S} = 1, \gamma_G = 1)$. For $N \approx 2000$ the mean and standard deviation of $(\alpha, \gamma_{S\alpha S}, \gamma_G)$ was $(1.497, 1.117, 0.884)$ and $(0.14, 0.48, 0.52)$ respectively. The estimate of $\alpha$ appears unbiased, though there is evidence of a small bias in the dispersion estimates.

The ability of the NWLS estimator to cope with small data sets is an advan-

Table 4.2: Mean and standard deviation for the three estimators with $N = 50000$ samples.

| Method | $\alpha = 1.5$ | $\gamma_{S\alpha S} = 1$ | $\gamma_G = 1$ |
|---|---|---|---|
| IMSE | 1.501 (13e-2) | 1.010 (1.0e-1) | 0.990 (7.7e-2) |
| Moment | 1.550 (11e-2) | 1.100 (1.2e-1) | 0.960 (9.0e-2) |
| NWLS | 1.499 (3.5e-2) | 1.001 (1.0e-1) | 0.999 (11e-2)) |
| | $\alpha = 1.5$ | $\gamma_{S\alpha S} = 0.5$ | $\gamma_G = 1$ |
| IMSE | 1.490 (19e-2) | 0.480 (14e-2) | 0.980 (17e-2) |
| Moment | 1.520 (12e-2) | 0.480 (46e-2) | 0.990 (54e-2) |
| NWLS | 1.502 (4.7e-2) | 0.506 (7.1e-2) | 0.995 (7.8e-2) |
| | $\alpha = 0.5$ | $\gamma_{S\alpha S} = 1$ | $\gamma_G = 1$ |
| IMSE | 0.501 (48e-3) | 1.010 (7.6e-2) | 0.993 (8.9e-2) |
| Moment | 0.530 (110e-3) | 1.070 (4.2e-2) | 0.960 (4.5e-2) |
| NWLS | 0.500 (6.9e-3) | 1.000 (1.6e-2) | 1.004 (2.7e-2) |
| | $\alpha = 0.5$ | $\gamma_{S\alpha S} = 0.5$ | $\gamma_G = 1$ |
| IMSE | 0.493 (4.9e-2) | 0.498 (4.0e-2) | 0.990 (4.7e-2) |
| Moment | 0.488 (11e-2) | 0.505 (2.0e-2) | 0.995 (2.1e-2) |
| NWLS | 0.501 (0.92e-2) | 0.502 (1.1e-2) | 0.997 (1.7e-2) |

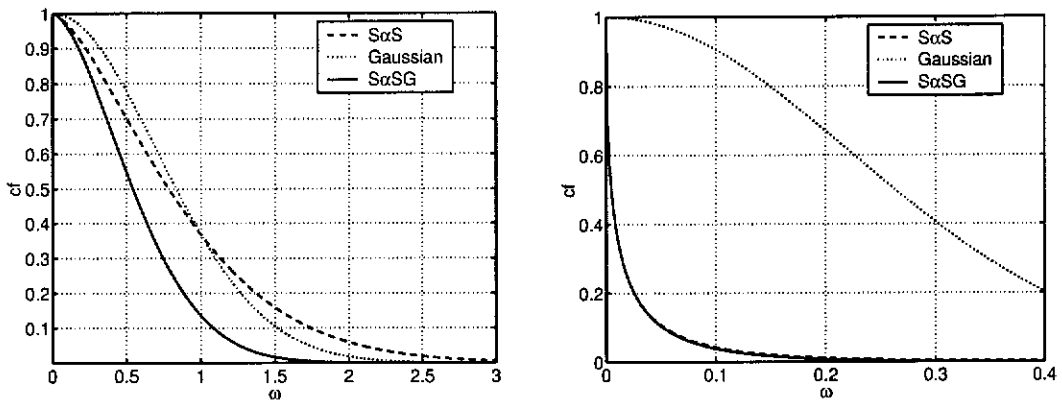


Figure 4.4: The $S\alpha SG$ cf for $(\alpha = 1.5, \gamma_{S\alpha S} = 1, \gamma_G = 1)$ (left) and $(\alpha = 0.5, \gamma_{S\alpha S} = 10, \gamma_G = 10)$ (right).

tage over previous estimators. Even for small sample sizes there were not any convergence problems and initial estimates are not required as for the IMSE estimator. The algorithm involves a one dimensional search coupled with weighted least squares, which may be performed inexpensively.

Table 4.3: Mean and standard deviation for the three estimators with $N = 50000$ samples.

| | $\alpha = 1.5$ | $\gamma_{S\alpha S} = 10$ | $\gamma_G = 10$ |
|---|---|---|---|
| IMSE | 1.551 (21e-2) | 11.01 (1.1) | 8.900 (8.0e-1) |
| Moment | 1.550 (19e-2) | 11.30 (1.2) | 9.860 (9.6e-1) |
| NWLS | 1.501 (2.7e-2) | 10.20 (1.2) | 9.710 (18e-1) |
| | $\alpha = 1.5$ | $\gamma_{S\alpha S} = 5$ | $\gamma_G = 10$ |
| IMSE | 1.530 (22e-2) | 4.800 (43e-2) | 10.80 (6.0e-1) |
| Moment | 1.520 (19e-2) | 4.780 (43e-2) | 10.60 (8.7e-1) |
| NWLS | 1.499 (3.4e-2) | 5.037 (71e-2) | 9.973 (9.9e-1) |
| | $\alpha = 0.5$ | $\gamma_{S\alpha S} = 10$ | $\gamma_G = 10$ |
| IMSE | 0.571 (8.8e-2) | 11.41 (1.6) | 8.790 (1.39) |
| Moment | 0.580 (11e-2) | 13.07 (2.22) | 9.560 (2.1) |
| NWLS | 0.500 (0.47e-2) | 10.00 (0.25) | 10.94 (7.2) |
| | $\alpha = 0.5$ | $\gamma_{S\alpha S} = 5$ | $\gamma_G = 10$ |
| IMSE | 0.593 (6.9e-2) | 5.398 (0.90e-1) | 8.490 (0.97) |
| Moment | 0.458 (11e-2) | 7.205 (14e-1) | 8.400 (1.6) |
| NWLS | 0.500 (0.60e-2) | 5.001 (1.2e-1) | 10.10 (1.2) |

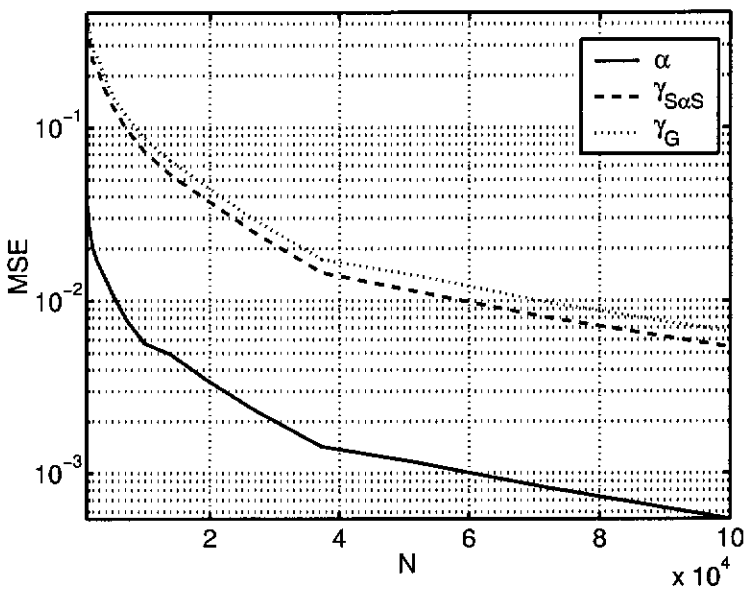


Figure 4.5: MSE of the NWLS parameter estimates versus sample size N for ($\alpha = 1.5, \gamma_{S\alpha S} = 1, \gamma_G = 1$).

## 4.6 Summary

The $S\alpha SG$ distribution is a model for impulsive noise which includes the effects of thermal Gaussian noise. An iterative NWLS estimator for the parameters of this distribution was developed in the cf domain. The NWLS estimator may be considered as a generalisation of the moment based estimator where the ecf is sampled at an arbitrary number of points, the selection of which was addressed. The proposed estimator performs well compared to existing IMSE and moment based estimators. Estimation of $\alpha$, the defining characteristic of the level of impulsive behaviour, was consistently superior. The NWLS method is computationally more efficient than the IMSE estimator and does not suffer from convergence problems nor does it require initial estimates. Furthermore, because the NWLS estimator does not suffer from convergence problems, it produces reasonable estimates for smaller sample sizes than is possible when using existing estimators.

# Chapter 5

# Detection of Sources in Array Processing

*We cannot learn without pain.*

*— Aristotle*

The number of sources impinging on an array is generally required knowledge for many array processing procedures such as high resolution direction of arrival estimation. If unknown, the number of sources may be found by determining the multiplicity of the smallest eigenvalues of the sample covariance. Here, a multiple test procedure which considers all pairwise comparisons between eigenvalues is developed. With a view to improving performance for small samples and when the sources or noise are heavy tailed, the bootstrap is used to estimate the null distributions of the test statistics.

## 5.1 Introduction

Detection of the number of sources is usually the first step in array processing algorithms. Essentially a model order selection problem, classical techniques for source detection tend to be based on information theoretic criteria such as Akaike's information criterion (AIC) and Rissanen's minimum description length criterion (MDL) [190]. The AIC has seen little use as it is not consistent and will over-estimate the number of sources, more popular is the MDL which is known to be consistent. Even so, consistency does not put any guarantee on the performance for small sample sizes or low SNR, which is often the case of interest. Since first being suggested, several modifications to the MDL have been developed to improve performance in these regimes [58].

Another approach to the problem is to use hypothesis tests. Considered first in the statistical literature, the sphericity test is a hypothesis test for equality of all eigenvalues [111]. For source detection it is necessary to test in a sequential manner for equality of subsets of the eigenvalues. Modifications to the original sphericity test have been developed to account for the presence of eigenvalues not included in the subset being tested. Unlike the original sphericity test these modified statistics converge to the correct asymptotic $\chi^2$ density giving higher detection rates [192].

All the aforementioned methods are based on the assumption of Gaussian observations, that is Gaussian sources and Gaussian noise. Under non-Gaussianity their behaviour is uncertain, the best that can be expected is that their performance degrades gracefully, though it is known that the distribution of the sample eigenvalues can be sensitive to departures from Gaussianity [189]. Reformulation of the AIC, MDL or sphericity tests to deal with non-Gaussian observations may be possible but the multitude of alternatives makes this approach problematic.

The problem of source detection for non-Gaussian observations is of interest here, particularly for heavy tailed observations. In the context of impulsive noise, direction of arrival (DOA) estimation has been addressed when the observations are modelled by $S\alpha S$ or Gaussian mixture distributions. The general approach for the $S\alpha S$ model is based on covariation, a FLOM based descriptor for multivariate dependence valid when $\alpha > 1$ [156]. The almost exclusive use of covariation may be attributed to the ease with which it is estimated and that it yields similar structures to those obtained from second order statistics. This allows the formulation of DOA estimators from existing subspace techniques by replacing sample covariance with sample covariation [177, 178, 179, 180, 183, 185]. The use of covariation is not optimal in any sense and so alternate estimators for dependence have been suggested.

In [168] it was shown that a suitably normalised sample correlation exists for all $S\alpha S$ distributions, as do normalised sample moments and cumulants, and that DOA estimates obtained in this way show less variability than covariation estimators [167, 169]. Preprocessing the observations by passing them through zero memory nonlinearities to reduce the effects of heavy tailed noise was investigated in [166, 169]. The application of standard second order DOA techniques to the nonlinearly transformed observations led to improved performance over the covariation and normalised correlation approaches. While source detection has not been addressed in any detail, it has been suggested to simply apply existing techniques such as the AIC or MDL to the eigenvalues of the sample covariation

matrix [177] or similar estimates of dependence based on the normalised moments or nonlinearly transformed observations.

For Gaussian mixtures, joint DOA estimation and source detection methods were developed for the deterministic signal model using Markov chain Monte Carlo techniques in [91] and for the general case using expectation-maximisation techniques in [108, 107]. Both take a fundamentally different approach to source detection. The Markov chain Monte Carlo technique jumps between models of different orders to determine the most likely number of sources, while the expectation-maximisation algorithm includes a robust estimate of the covariance to which the existing MDL criterion is applied.

Here the assumption of Gaussian signals and large sample sizes is removed. Fundamentally, what is proposed is a hypothesis test for equality of the smallest eigenvalues, as is the sphericity test. The procedure differs in that all pairwise comparisons, or differences, of the eigenvalues are considered and then combined using MHTs.

The null distributions of these statistics are estimated using the bootstrap, a technique valid for both Gaussian and a large class of non-Gaussian distributions [148, 198]. An additional advantage in using the bootstrap is that the finite sample, not the asymptotic, distributions are estimated, in contrast to the asymptotically correct distributions used in the sphericity test. The proposed procedure then makes minimal assumptions on the distribution of the signal and behaves well even for small sample sizes.

Bias in the sample eigenvalues can have a significant effect on the proposed method. This is especially true for small samples or when population eigenvalues are not well separated, such as at low SNR. Based on an expansion for the expectation of the sample eigenvalues a bias estimate is proposed which behaves well in these situations. Two resampling methods for bias estimation, using the jackknife and subsampling, are also suggested.

This Chapter is organised as follows. In Section 5.2 the signal model is described before discussing the detection procedure and the use of MHTs in Section 5.3. In Section 5.4 the bootstrap is applied to estimate the null distributions of the test statistics. Section 5.5 points out the need to correct for bias in sample eigenvalues and some methods for correction. Finally, Section 5.6 compares the proposed method against the sphericity test and the MDL.

## 5.2 Signal Model

The setting for the source detection problem is as follows, $N$ snapshots of iid zero mean complex observations are received from a $p$ element array,

$$\boldsymbol{x}_n = \boldsymbol{A}\boldsymbol{s}_n + \boldsymbol{v}_n, \qquad n = 1, \ldots, N, \tag{5.1}$$

where $\boldsymbol{A}$ is the $p \times q$ array steering matrix, $\boldsymbol{s}_n$ is a $q$ $(q < p)$ vector valued white source signal and $\boldsymbol{v}_n$ is noise with covariance $\sigma^2\boldsymbol{I}$, this is the narrowband stochastic signal model. The sources and noise are assumed to be independent, so that the array covariance is

$$\boldsymbol{R} = \mathsf{E}\left[\boldsymbol{x}_n\boldsymbol{x}_n^{\mathsf{H}}\right] = \boldsymbol{A}\boldsymbol{R}_s\boldsymbol{A}^{\mathsf{H}} + \sigma^2\boldsymbol{I}, \tag{5.2}$$

where $\boldsymbol{R}_s = \mathsf{E}\left[\boldsymbol{s}_n\boldsymbol{s}_n^{\mathsf{H}}\right]$ is the covariance of the sources and $(\cdot)^{\mathsf{H}}$ denotes the Hermitian transpose. The ordered eigenvalues of $\boldsymbol{R}$, the population eigenvalues, are

$$\lambda_1 \geq \cdots \geq \lambda_q > \lambda_{q+1} = \cdots = \lambda_p = \sigma^2, \tag{5.3}$$

so that the smallest $p - q$ population eigenvalues are equal. Hence the problem of detecting the number of sources is one of determining the multiplicity of the smallest eigenvalues. The ordered sample eigenvalues $l_i$, $i = 1, \ldots, p$, are estimated from the sample covariance

$$\hat{\boldsymbol{R}} = \frac{1}{N-1}\sum_{n=1}^{N}\boldsymbol{x}_n\boldsymbol{x}_n^{\mathsf{H}} \tag{5.4}$$

and are distinct w.p.1 for finite sample sizes [8],

$$l_1 > \cdots > l_q > l_{q+1} > \cdots > l_p > 0. \tag{5.5}$$

The sample eigenvalues are biased and mutually correlated. Their finite sample joint distribution is known in the Gaussian case and is represented as a series of zonal polynomials [89], a form too cumbersome for general use. A mathematically tractable form for their asymptotic joint distribution does exist in the Gaussian case [7], though it may be unreliable for the small sample sizes, on the order of 100, considered here. Also, this joint distribution is sensitive to departures from Gaussianity [189].

## 5.3 Detection of Sources by Determining Eigenvalue Multiplicity

Whether the source detection scheme is based on information theoretic criteria or hypothesis tests, the common assumption of Gaussian observations generally

leads to some variant of the statistic

$$\frac{(\prod_{i=k}^{p} l_i)^{\frac{1}{p-k+1}}}{\frac{1}{p-k+1}\sum_{i=k}^{p} l_i}, \qquad k = 1, \ldots, p-1, \tag{5.6}$$

which is the ratio of the geometric mean to the arithmetic mean of the smallest $p-k+1$ sample eigenvalues [7, 8, 111, 134, 190, 192, 191].

An intuitive understanding of why this statistic is appropriate follows by substituting the population eigenvalues into (5.6). Under the null hypothesis of equal eigenvalues, the geometric and arithmetic means are equal and (5.6) is 1. As the eigenvalues move apart under the alternative hypothesis (5.6) drops below unity.

This statistic can be regarded as a measure of the degree of sphericity, or ellipticity, of the subspace corresponding to the smallest $p-k+1$ eigenvalues. For equal eigenvalues the corresponding subspace is composed of iid Gaussian random variables, so that their multivariate distribution is spherical and (5.6) is 1. In contrast, if the eigenvalues are not all equal, their multivariate distribution becomes elliptic and (5.6) is less than 1.

This has provided the basis for hypothesis test based methods such as the sphericity test [192] where (5.6) comprises the test statistic, up to a multiplying factor. For information theoretic criteria such as the MDL (5.6) enters instead as the likelihood function of the observations [190, 191]. In a sense, the MDL can be considered a type of sphericity test where the set level is adaptive. For non-Gaussian observations this statistic still has relevance, but the performance of a test based on it is uncertain due to the assumption of Gaussian observations.

As mentioned, (5.3) suggests that $q$ should be estimated by determining the multiplicity of the smallest ordered sample eigenvalues. This can be accomplished by considering the following set of hypothesis tests,

$$\begin{array}{lcccccc}
\mathsf{H}_0 & : & \lambda_1 & = & \cdots & = & \lambda_p \\
\vdots & & & \vdots & & & \\
\mathsf{H}_k & : & \lambda_{k+1} & = & \cdots & = & \lambda_p \\
\vdots & & & \vdots & & & \\
\mathsf{H}_{p-2} & : & \lambda_{p-1} & = & \lambda_p & &
\end{array} \tag{5.7}$$

with corresponding alternative hypotheses $\mathsf{K}_k$ : not $\mathsf{H}_k$, $k = 0, \ldots, p-2$. Acceptance of $\mathsf{H}_k$ leads to the estimate $\hat{q} = k$. A practical procedure to estimate $q$ starts with testing $\mathsf{H}_0$ and proceeds to the next hypothesis test only on rejection of the hypothesis currently being tested. Upon acceptance the procedure stops, implying all remaining hypotheses are true. The procedure is outlined in Table 5.1.

Table 5.1: Hypothesis test procedure for determining the number of sources.

| |
|---|
| **Step 1.** Set $k = 0$. |
| **Step 2.** Test $\mathsf{H}_k$. |
| **Step 3.** If $\mathsf{H}_k$ is accepted then set $\hat{q} = k$ and stop. |
| **Step 4.** If $\mathsf{H}_k$ is rejected and $k < p-1$ then set $k \leftarrow k+1$ and return to step 2. Otherwise set $\hat{q} = p-1$ and stop. |

By taking the case of Gaussian signals and making simplifications to obtain an asymptotically correct closed form expression for the distribution of a test statistic based on (5.6), the sphericity test is obtained. Similarly, by following an information theoretic approach the MDL is arrived at.

At this point the sphericity test assumes Gaussianity, here this assumption is not made. Continuing with the hypothesis testing approach, consider all possible pairwise differences among the sample eigenvalues. This leads to the following test statistics

$$T_{ij} = l_i - l_j, \qquad i = k+1, \ldots, p-1, \quad j = i+1, \ldots, p. \tag{5.8}$$

These differences will be small when $l_i$ and $l_j$ are both noise eigenvalues, but relatively large if one or both of $l_i$ and $l_j$ are source eigenvalues. Representing the pairwise comparisons in a hypothesis testing framework gives

$$\begin{aligned} \mathsf{H}_{ij} &: \lambda_i = \lambda_j, \\ \mathsf{K}_{ij} &: \lambda_i \neq \lambda_j, \quad i = k+1, \ldots, p-1, \quad j = i+1, \ldots, p. \end{aligned} \tag{5.9}$$

The hypotheses $\mathsf{H}_k$ can then be formulated as intersections between the pairwise comparisons,

$$\begin{aligned} \mathsf{H}_k &= \cap_{i,j} \mathsf{H}_{ij}, \\ \mathsf{K}_k &= \cup_{i,j} \mathsf{K}_{ij}, \quad i = k+1, \ldots, p-1, \quad j = i+1, \ldots, p. \end{aligned} \tag{5.10}$$

The pairwise comparisons are carried out using a MHT procedure to maintain a global level of significance, as discussed next.

### 5.3.1 Multiple Hypothesis Testing

From Appendix E a critical property of MHT procedures is the degree to which the family wise error (FWE) rate is controlled. Weak control implies that the set

level, $\zeta$, is maintained under the global null hypothesis $\mathsf{H}_0$. In this problem the global null hypothesis is that all eigenvalues are equal, so that when no sources are present the probability of correctly choosing $\hat{q} = 0$ should be at least $1 - \zeta$. Strong control implies that the probability of rejecting any true hypothesis is at most $\zeta$ regardless of whether other hypotheses are true or false.

Bonferroni's MHT procedure strongly controls the level when the hypotheses, or equivalently, the p-values, are independent. Each hypothesis, $\mathsf{H}_{ij}$, is tested at a level $\zeta/h$ where $h = p(p-1)/2$ is the number of hypotheses comprising the global null hypothesis.

In this problem the p-values are not independent due to correlation between sample eigenvalues and logical implications between hypotheses. For instance, if $\mathsf{H}_{1p}$ were true, this would imply all the $\mathsf{H}_{ij}$ were true. In this case Holm's SRB MHT procedure better maintains the set level and is used exclusively. Although Hochberg's MHT procedure can also be used, the difference between these tests was found to be negligible.

Once the $\mathsf{H}_{ij}$ have been tested, all that remains is to step through Table 5.1. From (5.10) it is evident that $\mathsf{H}_k$ is rejected if any of the hypotheses $\mathsf{H}_{ij}$, $i = k+1, \ldots, p$, $j > i$, are rejected. P-values are determined from estimates of the null distributions which are obtained using the bootstrap, as discussed next. A similar approach was considered in [199] to find p-values for complicated test statistics.

## 5.4 Null Distribution Estimation

Evaluation of the p-values needed to carry out the hypothesis tests requires knowledge of the null distributions of the test statistics, the pairwise differences $T_{ij} = l_i - l_j$. The bootstrap is used as a nonparametric estimator of the null distributions [43]. It avoids making assumptions about the distribution of the signals, an important advantage when working with eigenvalues since their distribution is too complex for general use [89], while asymptotic expansions derived under Gaussianity [7] may not be valid for the small sample sizes considered. Asymptotic approximations developed for non-Gaussian cases require knowledge of the higher order moments of the observations, which are difficult to estimate for small sample sizes [189, 60]. Next, an application of the bootstrap method to this problem is presented.

### 5.4.1 Bootstrapping Eigenvalues

Here, the bootstrap is used to estimate the distribution of the test statistics from the sample. Details on the bootstrap principle and its application appear in Appendix D, however, a brief description of its application to this problem follows.

The sample is a set of vectors, the array snapshots, collected into the observation matrix

$$\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)\,. \tag{5.11}$$

The bootstrap resampling procedure for multivariate observations is a simple extension of the univariate case. First, let the empirical density be given by unit mass functions located at each snapshot and weighted by $1/N$. Resampling from this empirical density is equivalent to randomly resampling from the original observations with replacement, giving the bootstrap observations

$$\boldsymbol{X}^* = (\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_N^*)\,. \tag{5.12}$$

The sample eigenvalues of these observations, the bootstrapped eigenvalues, are

$$l_1^* > \cdots > l_p^*. \tag{5.13}$$

From the bootstrapped eigenvalues the test statistics are found. Repeating the procedure $B$ times gives an estimate of the distribution of this test statistic which can be used for inference. The bootstrap procedure for eigenvalues is shown in Table 5.2.

For linear statistics, such as the sample mean, the bootstrap is known to perform well and is a consistent estimator of the distribution. For complex nonlinear operations such as eigenvalue estimation these properties may not apply. In [17, 18] the statistical properties of bootstrapped eigenvalues are considered. It is shown that while the bootstrap converges to the correct asymptotic distributions for distinct eigenvalues, the same is not true for multiple eigenvalues. Bootstrapping with fewer resamples, $M$, where $M < N$, $\min(M, N) \to \infty$ and $M/N \to 0$, ensures the bootstrap converges weakly to the asymptotic distribution for multiple eigenvalues with sample size $M$. The sample sizes considered here are quite small, being on the order of 100. It was found that to fulfill the conditions which ensure weak convergence for multiple eigenvalues the decrease in resample size increased the error in eigenvalue estimation such that the overall error in the distribution increased.

For large sample sizes subsampling may be used to ensure weak convergence of the multiple eigenvalues to their asymptotic distributions. The conditions under

Table 5.2: Bootstrap procedure for resampling eigenvalues.

| |
|---|
| **Step 1.** Define the matrix of array snapshots $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$. |
| **Step 2.** Randomly select a snapshot from $\boldsymbol{X}$ with replacement. Repeat $N$ times to form the bootstrap observations $\boldsymbol{X}^* = (\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_N^*)$. |
| **Step 3.** Centre $\boldsymbol{X}^*$ by subtracting the sample mean from each row, $\boldsymbol{x}_n^* \leftarrow \boldsymbol{x}_n^* - \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_i^*$, $i = 1, \ldots, N$. |
| **Step 4.** Estimate the sample correlation matrix, $\hat{\boldsymbol{R}}^*$, of the centred $\boldsymbol{X}^*$. |
| **Step 5.** The bootstrapped eigenvalues, $l_1^*, \ldots, l_p^*$, are estimated from $\hat{\boldsymbol{R}}^*$. |
| **Step 6.** Repeat steps 2 to 5 $B$ times to obtain the bootstrap set of eigenvalues $l_1^*(b), \ldots, l_p^*(b)$, $b = 1, \ldots, B$. |

which subsampling is valid encompass a wider range of distributions and more complex statistics than with the bootstrap [149], a brief review of subsampling is given in Appendix D. Subsampling essentially means the observations are resampled $M < N$ times, either with or without replacement. To account for this decrease in sample size the rate of convergence to the asymptotic distribution must be estimated. Usually the rate is of the form $(M/N)^\tau$ where $\tau \approx 0.5$. The use of subsampling for null distribution estimation is considered to be outside the scope of this thesis.

### 5.4.2 Bootstrap Procedure

Given the bootstrap set of eigenvalues the test statistic of (5.8) is recalculated giving $T_{ij}^* = l_i^* - l_j^*$. Repeating this procedure $B$ times gives the set of bootstrapped test statistics, $T_{ij}^*(b)$, $b = 1, \ldots, B$. From these bootstrapped statistics the distribution of $T_{ij}$ under the null hypothesis is estimated as $\hat{T}_{ij}^{\mathsf{H}}(b) = T_{ij}^*(b) - T_{ij}$ [43]. Note that the test statistics are not studentised, preliminary investigations having shown the extra computational cost required to be unwarranted. P-values for the two-sided hypothesis tests are then evaluated in the usual fashion.

## 5.5 Bias Correction

Distinct sample eigenvalues, or those corresponding to sources, are asymptotically unbiased. Multiple sample eigenvalues, corresponding to the noise only, are

asymptotically biased. In the small sample case the bias becomes quite significant.

The hypothesis tests upon which the detection scheme rests is based upon detecting a statistically significant difference between the eigenvalues. Bias in the multiple sample eigenvalues falsifies the assumption that noise only sample eigenvalues have equal means. This necessitates some form of bias correction.

### 5.5.1 Lawley's Expansion

Lawley developed an expression for the expectation of the distinct sample eigenvalues by considering the propagation of error from the sample covariance to the eigenvalues for Gaussian observations [111],

$$\mathsf{E}[l_i] = \lambda_i + \frac{1}{N} \sum_{j=1,\ j\neq i}^{p} \frac{\lambda_i \lambda_j}{\lambda_i - \lambda_j} + O(N^{-2}), \qquad i = 1, \ldots, q, \tag{5.14}$$

remembering that the $\lambda_i$ are the unknown population eigenvalues. Hence the bias in the distinct sample eigenvalues is of order $O(N^{-1})$. A similar expression for the multiple eigenvalues does not exist, however extensive experiments have shown that the bias of multiple eigenvalues is of order $O(N^{-1/2})$, as discussed in Sections 5.5.4.2 and 5.5.4.3.

Replacing the population eigenvalues by their sample values gives Lawley's estimate for the bias in the $q$ distinct eigenvalues,

$$\widehat{Bias}_{\mathsf{LAW}}(l_i) = \frac{1}{N} \sum_{j=1,\ j\neq i}^{q} \frac{l_i l_j}{l_i - l_j} + \frac{p-q}{N} \frac{l_i \sigma^2}{l_i - \sigma^2}, \qquad i = 1, \ldots, q, \tag{5.15}$$

where $\sigma^2$, the population multiple eigenvalue, is replaced with its maximum likelihood estimate under Gaussianity,

$$\hat{\sigma}^2 = \frac{1}{p-q} \sum_{j=q+1}^{p} l_j. \tag{5.16}$$

After applying the bias estimates of (5.15, 5.16) the corrected distinct eigenvalues will have a bias of order $O(N^{-2})$, while the estimate of the multiple population eigenvalue is unbiased under Gaussianity. However, this correction cannot be used in the current problem for the following reasons,

1. The bias estimate is only valid when the difference between successive distinct eigenvalues is large compared to the sampling errors, which are of order $O(N^{-1/2})$. If this condition is not fulfilled and $\lambda_i \approx \lambda_j$ for $i \neq j$,

the variance of this estimator increases quickly [111]. This is easily understood by considering the effect of very close distinct eigenvalues on the denominator of the summation in (5.15).

2. If $q$ is unknown and multiple eigenvalues are present, application of (5.15) to all sample eigenvalues gives unpredictable results as the assumption of well separated population eigenvalues is invalid.

Essentially all problems related to applying the above correction without regard to the eigenvalue multiplicity is a result of the denominator term of (5.15) becoming very small. Next a simple solution to this problem is proposed.

### 5.5.2 A Robust Bias Estimate

Based on Lawley's expansion a bias estimate is proposed which overcomes the aforementioned problems by taking a binomial expansion in the denominator of the summand of (5.15) and truncating to a finite number of terms. For simplicity, assume that all the population eigenvalues are distinct. Then the bias estimate for $l_i$, $i = 1, \ldots, p$, becomes

$$\widehat{Bias}_{\text{LBE}}(l_i) = \frac{1}{N} \sum_{j=i+1}^{p} l_j \sum_{k=0}^{K} \left(\frac{l_j}{l_i}\right)^k - \frac{1}{N} \sum_{j=1}^{i-1} l_i \sum_{k=0}^{K} \left(\frac{l_i}{l_j}\right)^k \tag{5.17}$$

for some suitable $K$. If required, the upper limit on the outer summation can be changed from $p$ to $q$ and the term corresponding to multiple eigenvalues from (5.15) included. The choice of $K$ involves a compromise between the accuracy of the bias estimate and its variance. As $K$ increases the bias estimate improves, with an attendant rise in variance. Setting $K$ to a moderate value will retain the bias correction properties while guarding against large increases in variance when the population eigenvalues are not well separated or multiple eigenvalues are present. Experience has shown a value between 20 and 30 is acceptable, $K = 25$ was used here. $\widehat{Bias}_{\text{LBE}}$ can then be used without any knowledge of the multiple eigenvalues by assuming $q$ to be $p$, this bias estimate can be applied blindly to correct all eigenvalues irrespective of multiplicity.

An example with multivariate Gaussian observations is shown in Figures 5.1 and 5.2 where the largest sample eigenvalue is considered and both corrections are applied. The observations have a diagonal covariance matrix with population eigenvalues $(1.15, 1.1, 1.05, 1)^{\mathsf{T}}$. While Figure 5.1 shows the mean value of the corrected eigenvalues are very similar there is a notable decrease in the variance, as seen in Figure 5.2, when using $\widehat{Bias}_{\text{LBE}}$. For small sample sizes the reduction in

variance is significant as the separation of population eigenvalues is of the same order as the standard error.

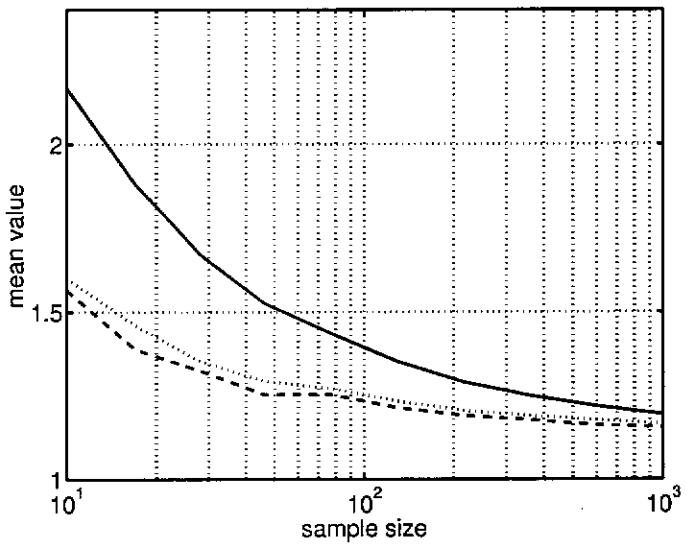


Figure 5.1: Mean of the largest sample eigenvalue with no bias estimation $(-)$, $\widehat{Bias}_{\text{LAW}}$ $(--)$ and $\widehat{Bias}_{\text{LBE}}$ $(\cdots)$ versus sample size for multivariate Gaussian observations, $p = 4$, with diagonal covariance matrix and population eigenvalues $(1.15, 1.1, 1.05, 1)^{\mathsf{T}}$.

For non-Gaussian signals Lawley's expansion includes an additive term in the numerator of (5.14), consisting of the second order cross-cumulant between the eigenvalues $\lambda_i$ and $\lambda_j$ [189]. As long as this additive term does not dominate, $\widehat{Bias}_{\text{LBE}}$ is still quite effective for non-Gaussian signals.

### 5.5.3 Jackknife Bias Estimation

Several alternative techniques for bias correction based on resampling methods were considered. The advantage of resampling techniques to bias correction in this case is that they can be applied blindly, with no knowledge of the eigenvalue multiplicity. The jackknife [43] was found to be the most effective scheme, it reduced the bias at least as much as (5.15) and did not suffer from any large increases in variance, even for multiple eigenvalues.

The effects of non-Gaussianity on the bias are also important. Here the jackknife has an advantage over (5.15) which was derived under the assumption of Gaussianity, for non-Gaussian observations the bias also depends on the cumulants of the underlying distribution [189]. In the non-Gaussian case then, the

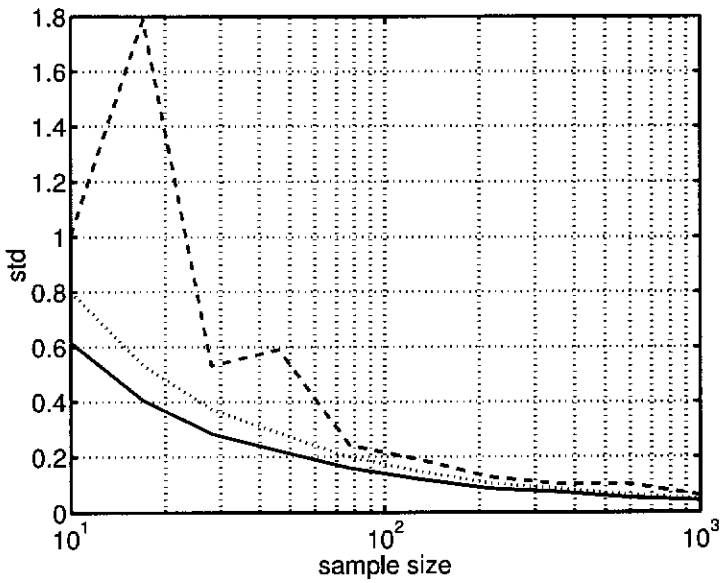


Figure 5.2: Standard Deviation of the largest sample eigenvalue with no bias estimation (−), $\widehat{Bias}_{\text{LAW}}$ (−−) and $\widehat{Bias}_{\text{LBE}}$ (···) versus sample size for multivariate Gaussian observations, $p = 4$, with diagonal covariance matrix and population eigenvalues $(1.15, 1.1, 1.05, 1)^{\mathsf{T}}$.

jackknife is still valid as it is a distribution free, though not distribution insensitive, method.

Jackknife resampling is reviewed in Appendix D, given the resampled eigenvalues $l_i^*(b)$, $b = 1, \ldots, N$, the jackknife estimate of bias for $l_i$, $i = 1, \ldots, p$, is

$$\widehat{Bias}_{\text{JCK}}(l_i) = (N-1)\left(\frac{1}{N}\sum_{b=1}^{N} l_i^*(b) - l_i\right), \tag{5.18}$$

where $l_i$ was estimated from the entire sample.

As will be shown in Section 5.5.4.3, the behaviour of $\widehat{Bias}_{\text{LBE}}$ and $\widehat{Bias}_{\text{JCK}}$ were found to be very similar, even for non-Gaussian distributions. Even though the former estimate is based on Gaussian observations it proves to be quite robust with respect to changes in distribution. Computationally, $\widehat{Bias}_{\text{LBE}}$ is also significantly more efficient than $\widehat{Bias}_{\text{JCK}}$, which increases by approximately $n$ times the computations needed to estimate a set of eigenvalues.

### 5.5.4 Subsampling Bias Estimation

In Section 5.4 subsampling was mentioned for distribution estimation as it is known to be applicable in the case of complicated nonlinear statistics where

the bootstrap may not converge to the correct asymptotic distribution. Here subsampling is applied to the problem of bias estimation.

Subsampling generalises the jackknife, removing $d$ samples at a time giving a subsample of size $M = N - d$. The number of unique subsamples, $N!/(M!(N - M)!)$, may be very large, in which case a smaller number of $B$ subsamples are chosen at random. As subsampling estimates of a statistic are based on a sample size $M < N$, any statistic obtained through subsampling must be adjusted for its dependence on sample size. For example, most statistics have a variance proportional to $1/\tau_N^2$ where $\tau_N$ is a function of the sample size $N$, while the variance of a statistic obtained from a subsample of size $M < N$ is proportional to $1/\tau_M^2$, hence the need for rescaling. The function $\tau_N$ is usually of the form $N^{\beta_\tau}$ with $\beta_\tau \in (0, 1)$ [148].

The subsampling bias estimate for $l_i$, $i = 1, \ldots, p$, is

$$\widehat{Bias}_{\text{SUB}}(l_i) \approx \frac{\tau_r}{\tau_N} \left( \frac{1}{B} \sum_{b=1}^{B} l_i^*(b) - l_i \right), \tag{5.19}$$

where $l_i^*(b)$ is the $b^{\text{th}}$ subsample estimate of the statistic, $r = MN/(N - M)$, and the approximation sign ($\approx$) is used to denote the fact that not all possible subsamples are used. The reason for using $\tau_r$ as opposed to $\tau_M$ is due to the resampling being done without replacement from a finite sample as opposed to with replacement from an infinite population, which is assumed with $\tau_M$. The ratio $\tau_r^2/\tau_M^2$ is known as the finite population correction [148]. When $M$ is much smaller than $N$ the finite population correction is close to unity and $r$ can be replaced by $M$.

Each subsample should also be a unique one from the $N!/(M!(N - M)!)$ possibilities, though in practice it is not necessary to check for this unless $N$ is very small or $M \approx N$. Application of subsampling to this problem is not as straight-forward as with the jackknife, two parameters need to be set. First, the subsample size, $M$, must be decided on and second, the rate, $\beta_\tau$, requires estimation. Next the question of how to appropriately choose the subsample size is addressed.

#### 5.5.4.1 Choice of subsample size

Guidelines for the appropriate subsample size are generally based on the complexity of the estimator and the true rate. In [148] the following conditions are given which ensure subsampling is accurate given a large enough sample size,

1. The statistic is linear and Gaussian: The jackknife, or subsampling with $M = N - 1$ is valid.

2. The statistic is nonlinear and asymptotically normal with $\tau_N = \sqrt{N}$: Both $M$ and $N - M$ must be large, a suitable choice being $M = kN$, $k \in (0, 1)$.

3. The statistic is complex, nonlinear, not asymptotically normal and $\tau_N$ is not necessarily $\sqrt{N}$: $M$ must be large while $M/N$ is small. A suitable choice is $M = N^k$, $k \in (0, 1)$.

In essence the more complex the estimator, the smaller the subsample size should be. For rate estimation it was found that the rate of convergence of the bias depends on whether the eigenvalues are distinct or multiple. Due to this difference in rate, distinct eigenvalues fall into the second category while multiple eigenvalues fall into the last. This means a small value of $M$ should be used.

The difference in rate for distinct and multiple eigenvalues suggests that $\beta_\tau$ cannot be set just once, but should be estimated for each sample eigenvalue. Next the rate of convergence for sample eigenvalue bias and its estimation is considered.

#### 5.5.4.2 Rate Estimation For Eigenvalue Bias

Subsampling estimators for the bias of a statistic are dependent on the rate of convergence $\tau_N$. Herein, $\tau_N$ is assumed to be of the form $N^{\beta_\tau}$ where $\beta_\tau$ is referred to as the rate.

From (5.14) the bias of distinct eigenvalues is $O(N^{-1})$, so a first order approximation to the rate for the bias of distinct eigenvalues is 1. A similar expression for the multiple eigenvalues does not exist, though empirical results suggest the rate is 0.5. Figure 5.3 shows the behaviour of eigenvalue bias with respect to sample size for uncorrelated Gaussian observations. The two scenarios shown are for distinct and multiple eigenvalues. The population eigenvalues chosen in the distinct case were $(4, 3, 2, 1)^\mathsf{T}$, this ensures that they behave as distinct eigenvalues over the range of sample sizes shown. The multiple population eigenvalues chosen were $(1, 1, 1, 1)^\mathsf{T}$.

A model for the bias of the eigenvalue $l_i$ is $Bias(l_i, N) = c/N^{\beta_\tau}$, where $c$ is a constant. In the figure several curves are fitted to the empirically determined rate: $\beta_\tau = 1$, $\beta_\tau = 1/2$ and $\beta_\tau$ being estimated.

The estimated rate for the largest distinct eigenvalue was 0.98 and for the largest multiple eigenvalue 0.53. Similar values are obtained for the other eigenvalues. Overall, $\beta_\tau = 1$ for distinct eigenvalues and $\beta_\tau = 0.5$ for multiple eigenvalues is a good approximation.

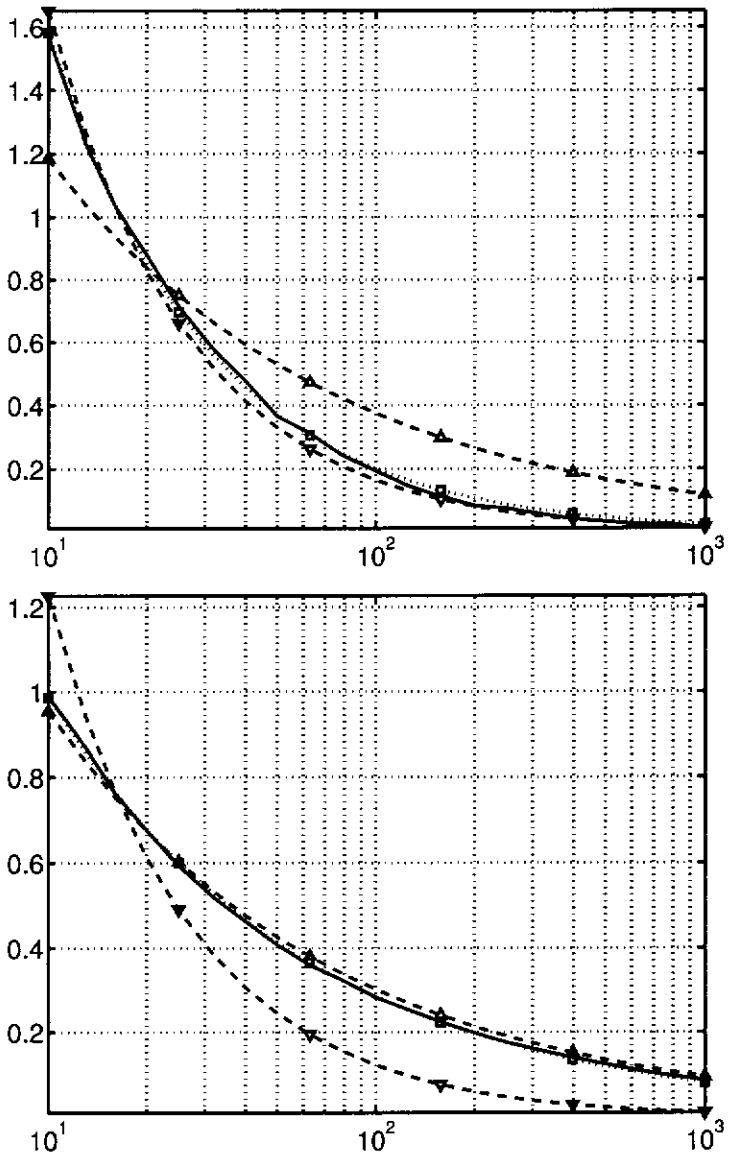


Figure 5.3: Bias of the largest sample eigenvalue versus sample size for distinct eigenvalues (top) and multiple eigenvalues (bottom). Shown are Monte Carlo ($-$) and the fitted models: $\beta_\tau = 1$ ($\bigtriangledown$), $\beta_\tau = 1/2$ ($\bigtriangleup$) and $\beta_\tau$ estimated ($\square$). The observations are multivariate Gaussian, $p = 4$, with diagonal covariance matrix. The distinct eigenvalues have population values $(4, 3, 2, 1)^\mathsf{T}$ while the multiple eigenvalues have population values $(1, 1, 1, 1)^\mathsf{T}$.

The most important point to note is that the convergence rate is clearly different for distinct and multiple eigenvalues. Hence any attempt to use the subsampling estimate for the bias of the eigenvalues should include estimation of the rate.

It is proposed to estimate $\beta_\tau$ as follows. First, express (5.19) so as to clarify its dependence on the subsample size $M < N$,

$$\widehat{Bias}_{\text{SUB}}(l_i) \approx \frac{\tau_r}{\tau_N} \widehat{Bias}_{\text{SUB}}(l_i; M), \tag{5.20}$$

where

$$\widehat{Bias}_{\text{SUB}}(l_i; M) \approx \left( \frac{1}{B} \sum_{b=1}^{B} l_i^*(b) - l_i \right) \tag{5.21}$$

is found from $B$ subsamples of size $M$. Clearly, (5.19) explicitly depends on $M$ and $\beta_\tau$,

$$Bias_{\text{SUB}}(l_i; M) \approx \left( \frac{N}{r} \right)^{\beta_\tau} Bias_{\text{SUB}}(l_i). \tag{5.22}$$

Remembering that $r$ is a simple function of $N$ and $M$, it can be seen that by varying $M$ and taking the logarithm of the above equation, $\beta_\tau$ can be determined by least squares.

The result can be understood intuitively by considering subsampling with size $M$ as an approximation to taking $M$ samples from the entire population, i.e. subsampling approximates Monte Carlo sampling. Obviously this becomes more accurate as $M$ decreases and $N$ increases.

To estimate the rate several estimates of the bias for various subsample sizes are required. Following the guidelines above, smaller values of $M$ are suggested.

The entire process of rate estimation can in fact be avoided by choosing $M = N/2$, in which case the finite population correction becomes unity and the subsampling bias estimate is not dependent on the rate.

#### 5.5.4.3 Examples of Bias Correction

In Table 5.3 some results on subsampling bias estimators are shown. The observations are uncorrelated and Gaussian with distinct population eigenvalues $(4, 3, 2, 1)^\mathsf{T}$ and a sample size of $N = 100$. Two subsample sizes are used, $M = 30$ and $M = N/2 = 50$, in both cases $B = 100$. For $M = 30$ three corrections are used, first assuming $\beta_\tau = 1$, then assuming $\beta_\tau = 0.5$ and finally where $\beta_\tau$ is estimated. The subsample sizes used for rate estimation were $M = kN$ with $k = (0.10, 0.20, 0.30, 0.40, 0.50)$. Corrections using the jackknife are also shown. The results obtained are for 1000 independent Monte Carlo realisations.

Whether $\beta_\tau$ is estimated or not makes very little difference. Even if an incorrect $\beta_\tau$ of 0.5 is used there is a reduction in bias. The subsampling corrections are also quite similar to the jackknife corrections. The same observations hold

Table 5.3: Bias corrected eigenvalues (mean ± standard error) for uncorrelated Gaussian observations with distinct population eigenvalues $(4, 3, 2, 1)^{\mathsf{T}}$, $N = 100$. Results shown are for uncorrected Monte Carlo estimates, the four subsampling corrected estimates and the jackknife corrected estimate.

| Correction | $l_1$ | $l_2$ | $l_3$ | $l_4$ |
|---|---|---|---|---|
| Monte Carlo | $4.19 \pm 0.54$ | $2.94 \pm 0.37$ | $1.91 \pm 0.26$ | $0.95 \pm 0.14$ |
| $M = 30$, $\beta_\tau = 0.5$ | $3.95 \pm 0.60$ | $3.01 \pm 0.45$ | $2.04 \pm 0.31$ | $1.04 \pm 0.17$ |
| $M = 30$, $\beta_\tau = 1.0$ | $4.04 \pm 0.58$ | $3.00 \pm 0.42$ | $2.00 \pm 0.29$ | $1.01 \pm 0.16$ |
| $\beta_\tau$ estimated | $4.03 \pm 0.60$ | $3.00 \pm 0.46$ | $2.00 \pm 0.30$ | $1.01 \pm 0.16$ |
| $M = N/2$ | $4.02 \pm 0.57$ | $2.99 \pm 0.47$ | $2.00 \pm 0.30$ | $1.00 \pm 0.15$ |
| Jackknife | $3.96 \pm 0.59$ | $3.03 \pm 0.49$ | $2.01 \pm 0.32$ | $1.00 \pm 0.15$ |

for multiple eigenvalues. Since subsampling produces bias estimates very close to the jackknife [24], only the latter will be used.

Finally note that whichever method of bias correction is used, it must be applied to both the sample eigenvalues and the bootstrapped eigenvalues. A summary of the entire detection procedure is given in Table 5.4.

Table 5.4: Bootstrap Detection Procedure.

**Step 1.** Estimate the eigenvalues, $l_1, \ldots, l_p$, from the matrix of array snapshots and apply one of the bias corrections $\widehat{Bias}_{\mathsf{LBE}}$ or $\widehat{Bias}_{\mathsf{JCK}}$.

**Step 2.** Obtain the $B$ bootstrapped eigenvalue sets as in Table 5.2 and bias correct each set using the same correction procedure as above.

**Step 3.** Calculate the test statistics $T_{ij}$ and the bootstrap estimate of their distributions under the null hypothesis, $\hat{T}_{ij}^{\mathsf{H}}$, as in Section 5.4.2.

**Step 4.** Given the level, carry out the multiple hypothesis test as in Table 5.1.

To demonstrate the behaviour of the bias estimators for the array signal model, their performance is shown for a specific scenario. The model was a $p = 4$ element array with a single ($q = 1$) Gaussian source at 20 degrees (with respect to

broadside), an SNR of $-7$dB and heavy tailed Laplacian noise. The associated population eigenvalues were $(1.8, 1, 1, 1)^{\mathsf{T}}$, so that the largest eigenvalue is distinct from the remaining eigenvalues, which are multiple. Figure 5.4 shows the mean of the third sample eigenvalue, one of the multiple eigenvalues, versus sample size over 1000 independent Monte Carlo realisations with no bias correction and after applying $\widehat{Bias}_{\text{LBE}}$, $\widehat{Bias}_{\text{JCK}}$ and $\widehat{Bias}_{\text{SUB}}$. As expected, the bias corrections behave similarly, further experiments have confirmed this. Between the three corrections, the jackknife appeared to yield slightly better bias estimates coupled with a greater increase in variance. Subsampling gave bias estimates with the smallest variance, while the performance of the proposed correction was somewhere between the two.

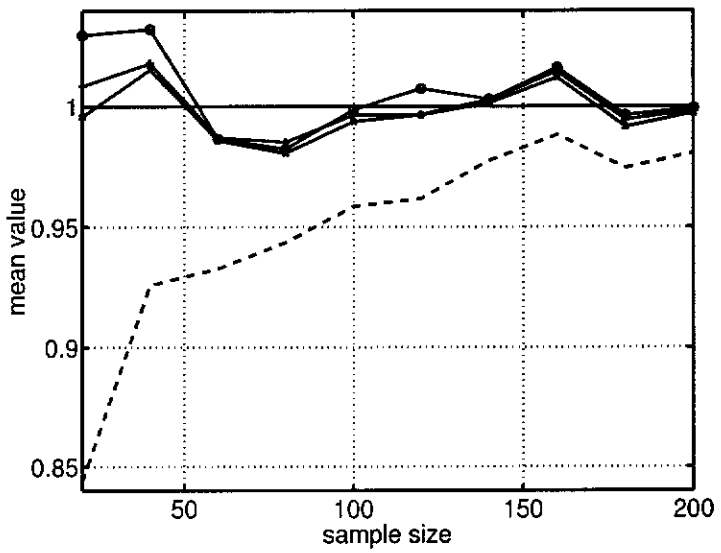


Figure 5.4: Mean of the third sample eigenvalue with no bias estimation ($--$), $\widehat{Bias}_{\text{LBE}}$ ($\times$), $\widehat{Bias}_{\text{JCK}}$ ($\circ$), $\widehat{Bias}_{\text{SUB}}$ ($+$), versus sample size for a Gaussian source in Laplacian noise. The population eigenvalues were $(1.8, 1, 1, 1)^{\mathsf{T}}$.

## 5.6 Experiments

In the following experiments the proposed method is evaluated by comparing it to the MDL [190] and sphericity test [192] in a variety of scenarios. Some parameters which remain unchanged throughout the tests are: the number of resamples, $B = 200$, the global level of significance, $\zeta = 2\%$ and the element spacing which was one half the wavelength. The signals are also Gaussian, unless

otherwise stated. All results were averaged over 100 Monte Carlo realisations. Results using $\widehat{Bias}_{\text{JCK}}$ are denoted by 'JCK' those using $\widehat{Bias}_{\text{LBE}}$ by 'LBE', the MDL and sphericity tests are denoted by 'MDL' and 'SPH' respectively.

The angular resolution, SNR threshold and effect of source correlation are important indicators of performance for source detection. In the following three experiments the relative performance of the bootstrap based method against existing ones is evaluated with respect to these criteria for Gaussian observations.

*Angular Separation* For Figure 5.5 $q = 2$ sources impinge on an array with $p = 4$ elements. The first source is fixed at 20 degrees (with respect to broadside) while the other is allowed to vary between 20 and 36 degrees. Both sources were at 0dB SNR and $N = 100$ snapshots were taken.

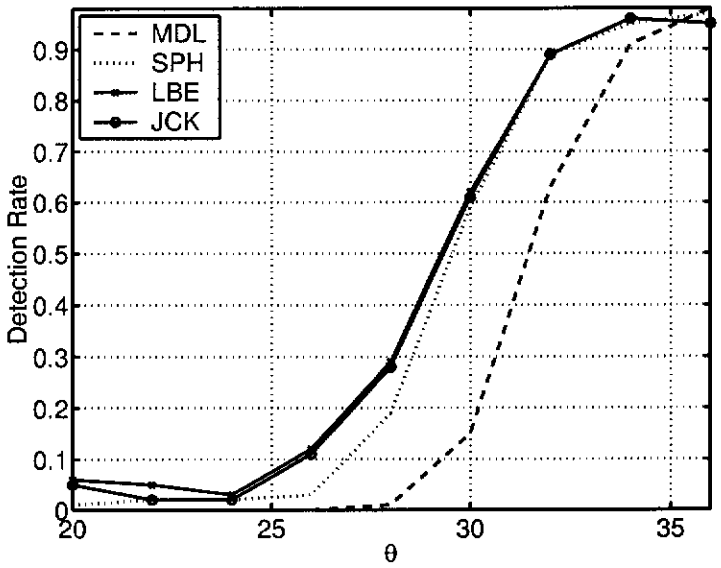


Figure 5.5: Empirical probability of correctly detecting two narrowly separated sources as the direction of one is varied.

*Effect of* SNR For Figure 5.6, $q = 3$ sources at 10, 30 and 50 degrees impinge on an array with $p = 4$ elements. The SNR of the second source is varied from 0 to 20dB while the SNRs of the first and third sources are $-2$ and 6dB respectively. $N = 50$ snapshots were taken.

*Correlated Sources* For Figure 5.7 $q = 2$ correlated sources at 20 and 40 degrees and SNRs of $-3$ and 0dB respectively impinge on a $p = 6$ element array, $N = 100$ snapshots were taken. The correlation coefficient between the two sources is varied from 0.69 to 0.99.

These three examples show the bootstrap method performing similarly to the sphericity test when the assumption of Gaussian signals is fulfilled. It should be

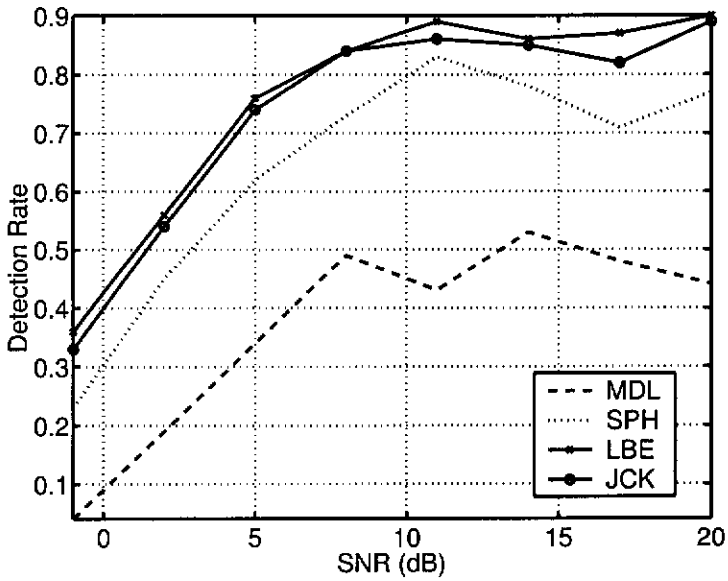


Figure 5.6: Empirical probability of correct detection as the source SNR is varied.

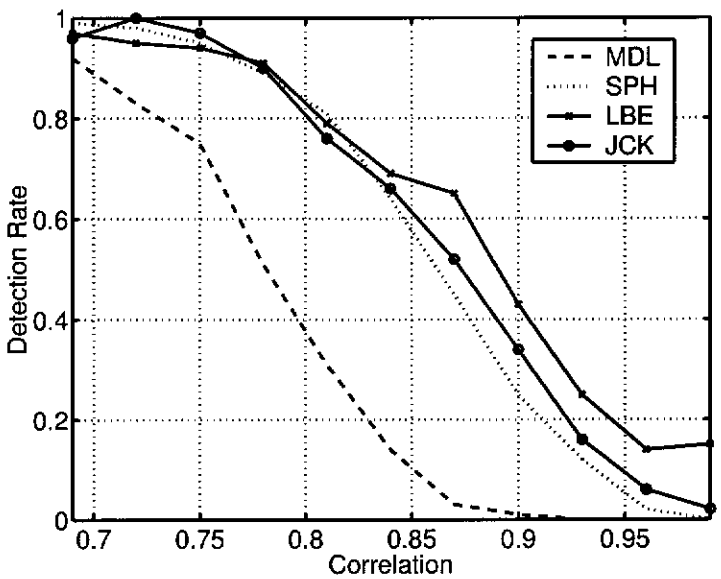


Figure 5.7: Empirical probability of correctly detecting two correlated sources as the correlation coefficient is varied.

remembered that the bootstrap method is based on minimal assumptions about the signal distribution while the sphericity test assumes Gaussian signals. Any gain attained under Gaussianity are due to the bootstrap estimating the finite sample distributions as opposed to the asymptotic approximations made in the sphericity test. Gains in detection rate may then be achieved by removing the

assumption of Gaussianity and using the bootstrap.

Due to its consistency properties, the MDL approaches a 100% detection rate as the sample size increases. In this sense it is fundamentally different to the hypothesis testing approach proposed here, or that of the sphericity test, where for large sample sizes the detection rate stays below 100%. By reducing $\zeta$ the detection rate of the bootstrap test can be increased for large sample sizes or when ideal conditions, such as high SNR or weakly correlated sources prevail. The reason for this is that under such conditions the distinct eigenvalues are clearly separated from multiple eigenvalues. Reducing $\zeta$ has the effect of lowering the false alarm rate while not significantly affecting the correct rejection of the null hypothesis when the eigenvalues are distinct. This decreases the probability of an erroneous decision, increasing the detection rate. Contrary-wise, increasing $\zeta$ improves the detection rates when conditions are not ideal. In this sense $\zeta$ can be regarded as a tuning parameter, though is should be keep small ($< 5\%$) to reduce false detections when no sources are present. However, as the level decreases more resamples are required to properly estimate critical points and the computational load increases [43].

*Sample Size* Maintenance of a high detection rate for small sample sizes is another favourable property of source detection procedures. In Figure 5.8 the effects of sample size are demonstrated for Gaussian observations. There were $q = 3$ sources at 10, 30 and 50 degrees and at SNRs of $-2$, 2 and 6dB respectively impinging on a $p = 4$ element array. The sample size was varied over $20 \leq N \leq 160$.

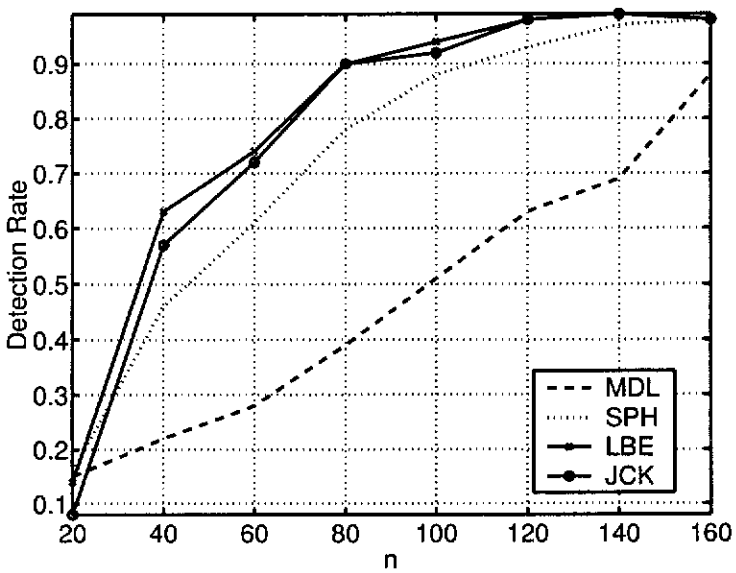


Figure 5.8: Empirical probability of correct detection as the sample size is varied.

For small sample sizes of 100 snapshots or less the bootstrap method achieves a consistently higher detection rate than the sphericity test. Further experiments have shown that while the gain is not always so marked the bootstrap still performs comparably. As already discussed, this is a result of the bootstrap estimating the finite sample distributions in contrast to the asymptotic correctness of the sphericity test.

*Noise Only* In Figure 5.9 the probability of correctly accepting the global null hypothesis is shown when there are $q = 0$ sources in Gaussian noise. In this case all the eigenvalues are equal and so this may be interpreted as the probability of correctly detecting $q = 0$ sources. The array has $p = 4$ elements and the sample size was varied over $20 \leq N \leq 160$.

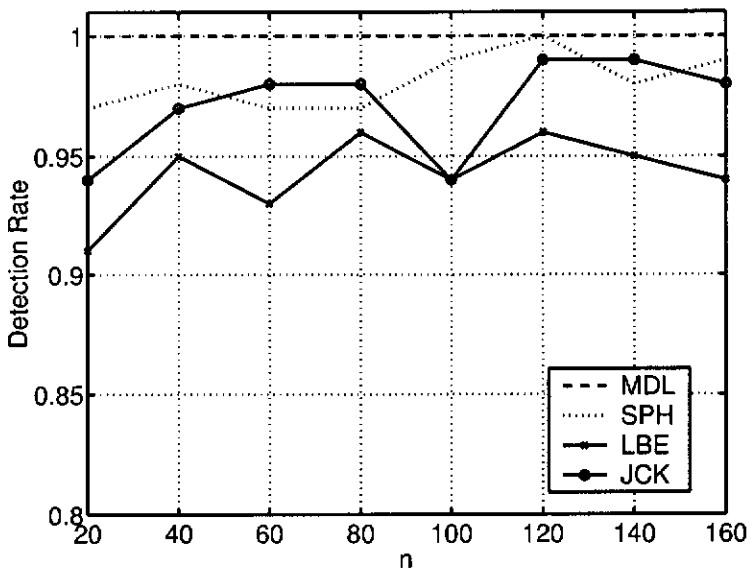


Figure 5.9: Empirical probability of correctly detecting $q = 0$ sources in a source free environment as the sample size varies.

In a source free environment the global null hypothesis is in force and the global level of significance should be maintained close to the set level of $\zeta = 2\%$, that is, the probability of correctly detecting $q = 0$ sources should be near $1 - \zeta = 98\%$.

Given that the finite number of Monte Carlo realisations induces some variability in the results, the level of the test does not appear to be dependent on sample size. This is a desirable and expected property as the level of the test should be independent of the array size and the sample size. Note that the MDL does not have the ability to adjust $\zeta$, instead it is adjusted implicitly so that the

detection rate approaches 100% under ideal conditions such as large samples.

On average, the attained level for the sphericity test is approximately 2%, while for the bootstrap with jackknife bias correction it is 3% and 6% when Lawley's modified bias estimator is used. Several sources of error are present which may account for this, the most obvious being the actual estimates of the null distribution. Taking a larger number of resamples decreases this, but at the expense of computation. The logical implications between hypotheses also affect the SRB procedure. While more powerful tests which are specifically tailored to the implications among the hypotheses can be developed, they are often complicated [72].

The larger error in the attained level when using $\widehat{Bias}_{\text{LBE}}$ may be attributed to small sample effects, truncation error and applying the estimator to resampled eigenvalues. Lawley's estimator was developed for *continuous* Gaussian observations, whereas the resampled eigenvalues can be considered to be generated from a *discrete* distribution formed by the array snapshots, thereby introducing an error. The jackknife estimator overcomes this problem as it is a distribution free method.

*Non-Gaussianity* For the next two examples the observations are heavy tailed. In Figure 5.10 the conditions are the same as in Figure 5.6, except that the sources are Laplacian and the noise Gaussian, while for Figure 5.11 the sources are Gaussian and the noise Laplacian.

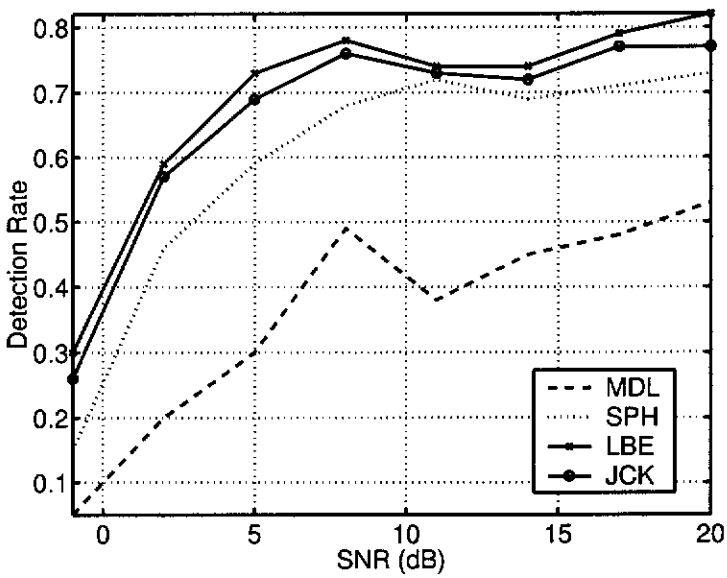


Figure 5.10: Empirical probability of correctly detecting Laplacian sources in Gaussian noise as the SNR varies.

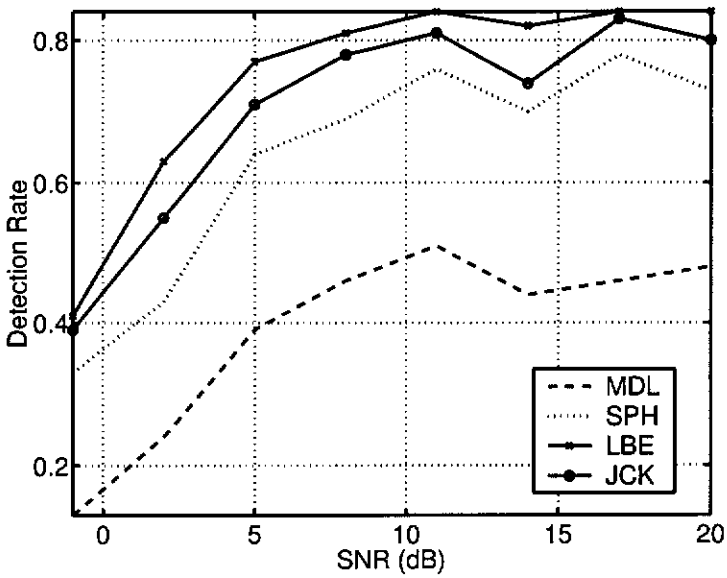


Figure 5.11: Empirical probability of correctly detecting Gaussian sources in Laplacian noise as the SNR varies.

In both cases the bootstrap is an improvement over the sphericity test and MDL. For low SNR the improvement in detection rate is approximately 10%, or up to a 5dB gain in SNR threshold. Similar results were obtained for other heavy tailed distributions such as Gaussian mixtures.

Considering computational load, the bootstrap detector using $\widehat{Bias}_{\text{JCK}}$ takes approximately $BN$ times more computations than the MDL or sphericity test, using $\widehat{Bias}_{\text{LBE}}$ reduces this to $B$ times. The trade-off in using $\widehat{Bias}_{\text{LBE}}$ to decrease the load is that the set level is not exactly maintained, but is still acceptable. For small sample sizes of $N < 100$ and the $B = 200$ resamples necessary to maintain a level of a few percent, the increase in computational complexity is easily met.

## 5.7 Summary

The source detection problem in array processing was approached from a hypothesis testing viewpoint. The proposed test was based on the difference between eigenvalues, the assumption being that under the null hypothesis of equality this difference should be small relative to under the alternative hypothesis.

The assumption of Gaussian signals and large sample sizes to enable calculation of the test statistics' null distributions was removed. Instead the bootstrap was used to estimate the null distributions nonparametrically under minimal as-

sumptions on the signal distribution.

It was shown that for the cases of interest, such as small sample size, bias in the sample eigenvalues is non-negligible and should be corrected when carrying out inference on the eigenvalues. A modified improved bias estimate was proposed based on Lawley's expansion which overcame the need to know the multiplicity of the eigenvalues, performing well in spite of their presence. The jackknife was also presented as a distribution free method for bias correction. Both these estimators perform similarly, though the jackknife, like the bootstrap, makes minimal assumptions on the signal distribution. The method of bias correction using subsampling was covered, with consideration given to the problems of choosing the subsample size and the rate, for which an estimator was proposed.

Experiments have shown that the bootstrap based method outperforms the MDL for small sample sizes, when the observations are heavy tailed or when the conditions are not ideal, such as for strongly correlated sources or low SNR. Performance was also comparable to the more powerful sphericity test, with a demonstrated increase in detection rate for small sample sizes and heavy tailed observations.

When using the bootstrap, the number of computations is increased by approximately a factor of $B$, the number of resamples. This includes the increase associated with bias correction if the proposed bias estimator is used, which itself represents an $N$-fold reduction in computational complexity over the jackknife bias estimator. While the overall increase in computational complexity may seem large, it should be remembered that it is entirely predictable and that the nature of the bootstrap makes the algorithm amenable to parallel processing. Hence, the use of the bootstrap is not outside the capabilities of modern processors.

# Chapter 6

# Robust Estimation and Detection

*This isn't right, this isn't even wrong.*
*— Wolfgang Pauli, upon reading*
*a young physicist's paper.*

The problem of estimation in heavy tailed noise is considered using the M-estimation concept of robust statistics. An adaptive robust estimator is presented which while following the M-estimator structure, simultaneously estimates the score function of the noise directly from the observations. Two schemes are used to estimate the score function. The first employs a nonparametric kernel density estimator, an application to multiuser detection in impulsive noise demonstrates superior performance in very impulsive noise. The second models the score function as a linear combination of basis functions. Theoretical results show clear improvements in performance over static schemes such as the minimax estimator for proper choice of the bases.

## 6.1 Introduction

The presence of impulsive noise in communications channels has driven the development of detectors and associated estimators which perform well in both Gaussian and impulsive noise, not suffering the rapid degradation in performance which plagues those based on Gaussian noise or linear statistics, such as the classical linear correlator detector [1].

A typical approach is to assume a parametric form for the noise distribution, chosen from among one of the many impulsive noise models and then to develop an optimal or sub-optimal procedure based on it. Such an approach assumes the chosen distribution accurately models the noise, of course it is only an approximation to physical reality. How far the model deviates from reality and what

effect these deviations have becomes problematic, it can only be hoped that the procedure is insensitive to changes in the model.

To decrease the sensitivity of an estimator to the underlying distribution one may turn to the theory of robust estimation and use M-estimators to implement a sub-optimal nonlinear detector which is robust to changes in distribution [74, 188]. While M-estimators can be chosen to give a minimal level of performance over a large set of impulsive models, their performance for any one may be significantly far from optimal.

In this Chapter the design of adaptive robust procedures is addressed. By adaptive it is meant that the estimator or detector alters itself to follow changes in the noise distribution [73]. Two distinct methods based on the M-estimation concept are considered, both of which extract information from the observations about the noise distribution and incorporate this into an M-estimator. The first explicitly estimates the noise density through a nonparametric kernel density estimator modified for improved performance in impulsive noise. This enables the detector to adapt to various noise models with minimal a priori information, at most the density is assumed unimodal and symmetric. The second estimates the noise density indirectly, a linear combination of bases functions being used to model the score function of the noise distribution. Originally developed in [170], improvements are proposed which greatly improve small sample performance. The asymptotic covariance of the estimator is derived and the small and large sample performance is analysed. Results show that near optimal performance is attainable when the bases are suitably chosen.

The organisation of this Chapter is as follows. In Section 6.2 the signal model is presented. In Section 6.3 the theory of M-estimation is reviewed and the theoretical performance under the signal model is derived. The concept of adaptive robust estimation is introduced in Section 6.4. The nonparametric adaptive robust estimator is developed in Section 6.5 and results from an application to the multiuser detection (MUD) problem is shown. The parametric adaptive robust estimator and associated theoretical results appear in Section 6.6.

## 6.2 Signal Model

Consider the general signal in additive noise model,

$$y_n = s_n(\boldsymbol{\theta}) + x_n, \qquad n = 1, \ldots, N, \tag{6.1}$$

where $x_n$ is iid noise and the signal, $s_n$, is parameterised by $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)^{\mathsf{T}}$. The aim is to estimate the signal parameters $\boldsymbol{\theta}$ from the $N$ observations $y_n$, as a priori knowledge required for further processing, such as detection.

Given that the noise density, $f_X(x)$, is known, the ML solution is obtained as

$$\hat{\boldsymbol{\theta}}_{\mathrm{ML}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{n=1}^{N} -\log f_X\left(y_n - s_n(\boldsymbol{\theta})\right) \tag{6.2}$$

or equivalently, is found by solving the $K$ coupled equations,

$$\sum_{n=1}^{N} \psi\left(y_n - s_n(\boldsymbol{\theta})\right) \frac{\partial s_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}, \tag{6.3}$$

where $\psi(x) = -f_X'(x)/f_X(x)$ is the location score function of $f_X(x)$.

As ML estimators require that $f_X(x)$ be known, any associated distributional parameters must also be known or estimated. It is clear that without a priori knowledge of $f_X(x)$, estimation of $\boldsymbol{\theta}$ cannot be optimal in the ML sense and performance is uncertain with respect to deviations from the assumed model. Furthermore, the computational cost of the MLE can be too high, $S\alpha S$ distributions being a prime example due to the lack of a closed form expression for their pdf.

## 6.3 M-Estimation

M-estimators may be viewed as a generalisation of ML in the face of uncertainty about the noise distribution [46, 81]. Here, the uncertainty stems from the presence of outliers injected by impulsive noise. In an M-estimator the log-likelihood function $-\log f_X(x)$ of the MLE is replaced with a similarly behaved penalty function, $\rho(x)$. The penalty function is chosen to confer robustness on the estimator under deviations from the assumed density. $\boldsymbol{\theta}$ is estimated from

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{n=1}^{N} \rho\left(y_n - s_n(\boldsymbol{\theta})\right), \tag{6.4}$$

or alternatively from the $K$ coupled equations,

$$\sum_{n=1}^{N} \varphi\left(y_n - s_n(\boldsymbol{\theta})\right) \frac{\partial s_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}, \tag{6.5}$$

where $\varphi(x) = \rho'(x)$. When $f_X(x)$ is unknown it is unsure how close $\varphi(x)$ is to $\psi(x)$. Selection of the penalty function is then of prime importance in ensuring the performance of the estimator is not highly sensitive to $f_X(x)$ but is robust over a wide class of noise models.

### 6.3.1 Theoretical Performance for Arbitrary $\varphi$

Under some mild regularity conditions such as $\mathsf{E}[\varphi(x)] = 0$, M-estimates possess desirable properties such as consistency and asymptotic normality [46, 81]. Herein, $f_X(x)$ is assumed to be symmetric so that an antisymmetric $\varphi(x)$ is used to ensure this condition is met.

Let $F_X$ be the distribution (cdf) of the noise and $F_N$ be its empirical counterpart from a sample of size $N$. Then an estimate of $\boldsymbol{\theta}$ can be defined in terms of a functional $T$ operating on $F_N$, $T(F_N)$, while the true parameters are obtained as $T(F_X)$.

#### 6.3.1.1 Consistency

Let $F_X$ belong to a family of distributions $\mathcal{F}$, then $T(F_N)$ converges in probability to $T(F_X)$ as $N \to \infty$,

$$\mathsf{Pr}[|T(F_N) - T(F_X)| > \epsilon] \to 0 \text{ as } N \to \infty, \qquad F_X \in \mathcal{F}, \tag{6.6}$$

for every $\epsilon > 0$.

#### 6.3.1.2 Asymptotic Normality

Define $V(T, F_X)$ to be the asymptotic variance, then

$$N^{1/2}(T(F_N) - T(F_X)) \xrightarrow{d} \mathrm{N}(0, V(T, F_X)), \tag{6.7}$$

where $\mathrm{N}(0, 1)$ is the standard normal distribution.

#### 6.3.1.3 Influence Function

The influence function measures the effect of a deviation from the assumed distribution on a descriptive statistic, $T$, in other words, robustness. It is defined as the change induced in $T$ by an infinitesimal contaminant at $x$,

$$\mathrm{IF}(x; T, F_X) = \lim_{\varepsilon \to 0} \frac{T((1-\varepsilon)F_X + \varepsilon \Delta_x) - T(F_X)}{\varepsilon}, \tag{6.8}$$

where $\Delta_x$ denotes a unit step in the distribution.

The utility of the influence function is that it allows calculation of the asymptotic covariance of the M-estimates,

$$V(T, F_X) = \int \mathrm{IF}(x; T, F_X)\mathrm{IF}(x; T, F_X)^{\mathsf{T}} \, dF_X. \tag{6.9}$$

$\mathrm{IF}(x; T, F_X)$ and $V(T, F_X)$ will now be found for the signal model (6.1).

#### 6.3.1.4 Asymptotic Covariance of M-estimates

The functional corresponding to (6.5) is

$$\int \sum_{n=1}^{N} \left\{ \varphi(y_n - s_n(\boldsymbol{\theta}(\bar{F}))) \frac{\partial s_n(\boldsymbol{\theta}(\bar{F}))}{\partial \boldsymbol{\theta}} \right\} d\bar{F} = \mathbf{0}, \tag{6.10}$$

where $\bar{F} = \Pi_{n=1}^{N} F_n$ is defined as, $F_n = F_X(y_n - s_n(\boldsymbol{\theta}(\bar{F})))$, $\bar{H} = \Pi_{n=1}^{N} H_n$, $H_n = (1-\varepsilon)F_n + \varepsilon \Delta_{y_n}$. Following the approach in [81], replace $\bar{F}$ in (6.10) by $\bar{H}$, differentiate with respect to $\varepsilon$ and evaluate at $\varepsilon = 0$,

$$\frac{\partial}{\partial \varepsilon} \int \sum_{n=1}^{N} \left\{ \varphi(y_n - s_n(\boldsymbol{\theta}(\bar{H}))) \frac{\partial s_n(\boldsymbol{\theta}(\bar{H}))}{\partial \boldsymbol{\theta}} \right\} d\bar{H} \Bigg|_{\varepsilon=0} = \mathbf{0}. \tag{6.11}$$

Applying the product rule first to the term in braces and taking the partial differential inside the summation gives

$$\int \sum_{n=1}^{N} \frac{\partial}{\partial \varepsilon} \left\{ \varphi(y_n - s_n(\boldsymbol{\theta}(\bar{H}))) \frac{\partial s_n(\boldsymbol{\theta}(\bar{H}))}{\partial \boldsymbol{\theta}} \right\} \Bigg|_{\varepsilon=0} d\bar{F}. \tag{6.12}$$

The operations of differentiation and integration can be interchanged subject to some conditions on the continuity of the integrand [92]. These regularity conditions are assumed to be fulfilled, see [81] for more details. Since

$$\begin{aligned} \frac{\partial s_n(\boldsymbol{\theta}(\bar{H}))}{\partial \varepsilon} \Bigg|_{\varepsilon=0} &= \frac{\partial s_n(\boldsymbol{\theta}(\bar{F}))}{\partial \boldsymbol{\theta}^{\mathsf{T}}} \cdot \frac{\partial \boldsymbol{\theta}(\bar{H})}{\partial \varepsilon} \Bigg|_{\varepsilon=0} \\ &= \frac{\partial s_n(\boldsymbol{\theta}(\bar{F}))}{\partial \boldsymbol{\theta}^{\mathsf{T}}} \cdot \mathrm{IF}(\boldsymbol{y}; T, F_X), \end{aligned} \tag{6.13}$$

where $\boldsymbol{y} = (y_1, \ldots, y_N)^{\mathsf{T}}$, it follows that

$$\begin{aligned} &\frac{\partial \varphi(y_n - s_n(\boldsymbol{\theta}(\bar{H})))}{\partial \varepsilon} \Bigg|_{\varepsilon=0} = \\ &\quad -\varphi'(y_n - s_n(\boldsymbol{\theta}(\bar{F}))) \cdot \frac{\partial s_n(\boldsymbol{\theta}(\bar{F}))}{\partial \boldsymbol{\theta}^{\mathsf{T}}} \cdot \mathrm{IF}(\boldsymbol{y}; T, F_X). \end{aligned} \tag{6.14}$$

Also,

$$\frac{\partial^2 s_n(\boldsymbol{\theta}(\bar{H}))}{\partial \varepsilon \partial \boldsymbol{\theta}} \Bigg|_{\varepsilon=0} = \frac{\partial^2 s_n(\boldsymbol{\theta}(\bar{F}))}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\mathsf{T}}} \cdot \mathrm{IF}(\boldsymbol{y}; T, F_X). \tag{6.15}$$

Since the estimator is consistent, $\boldsymbol{\theta}(\bar{F}) \to \boldsymbol{\theta}$, as $N \to \infty$ and the redundant reference to $\bar{F}$ is dropped. Hence (6.12) becomes

$$\begin{aligned} \sum_{n=1}^{N} \Bigg\{ &\int -\varphi'(y_n - s_n(\boldsymbol{\theta}))\, d\bar{F} \cdot \frac{\partial s_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \cdot \frac{\partial s_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathsf{T}}} + \\ &\int \varphi(y_n - s_n(\boldsymbol{\theta}))\, d\bar{F} \cdot \frac{\partial^2 s_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\mathsf{T}}} \Bigg\} \cdot \mathrm{IF}(\boldsymbol{y}; T, F_X). \end{aligned} \tag{6.16}$$

Since $\varphi(x)$ is antisymmetric and $f_X(x)$ is symmetric, the second integral is zero. Furthermore, the first integral is simply $-\mathsf{E}[\varphi'(x)]$, hence (6.12) reduces to

$$-\mathsf{E}[\varphi'(x)] \sum_{n=1}^{N} J_n J_n^{\mathsf{T}} \cdot \mathrm{IF}(\boldsymbol{y}; T, F_X), \tag{6.17}$$

where $J_n$ is the gradient of $s_n(\boldsymbol{\theta})$. Finally, evaluate the remaining term from (6.11),

$$\int \sum_{n=1}^{N} \varphi(y_n - s_n(\boldsymbol{\theta})) \frac{\partial s_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \, d \left\{ \frac{\partial \bar{H}}{\partial \varepsilon} \right\} \Big|_{\varepsilon=0}, \tag{6.18}$$

$$\begin{aligned} \frac{\partial \bar{H}}{\partial \varepsilon}\Big|_{\varepsilon=0} &= \frac{\partial \Pi_{n=1}^{N} (1-\varepsilon) F_X(y_n - s_n(\boldsymbol{\theta})) + \varepsilon \Delta_{y_n}}{\partial \varepsilon}\Big|_{\varepsilon=0} \qquad (6.19) \\ &= \sum_{n=1}^{N} \{\Delta_{y_n} - F_X(y_n - s_n(\theta))\} \, \Pi_{m \neq n}^{N} F_X(y_m - s_m(\boldsymbol{\theta})) \\ &= \sum_{n=1}^{N} \Delta_{y_n} \Pi_{m \neq n}^{N} F_X(y_m - s_m(\boldsymbol{\theta})) - N\bar{F}. \qquad (6.20) \end{aligned}$$

The second term of (6.20), when substituted back into (6.18), yields zero by application of (6.5). The first term gives

$$\sum_{n=1}^{N} \varphi(y_n - s_n(\boldsymbol{\theta})) \frac{\partial s_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{n=1}^{N} \varphi(y_n - s_n(\boldsymbol{\theta})) J_n. \tag{6.21}$$

Combining (6.17) and (6.21) gives $\mathrm{IF}(\boldsymbol{y}; T, F_X)$ as

$$\mathrm{IF}(\boldsymbol{y}; T, F_X) = \frac{(\sum_{n=1}^{N} J_n J_n^{\mathsf{T}})^{-1} \sum_{n=1}^{N} \varphi(y_n - s_n(\boldsymbol{\theta})) J_n}{\mathsf{E}[\varphi'(x)]}. \tag{6.22}$$

It is then straightforward to obtain the asymptotic covariance from (6.9) and (6.22) as

$$V(T, F_X) = \frac{\mathsf{E}[\varphi^2(x)]}{\mathsf{E}[\varphi'(x)]^2} \cdot \left( \sum_{n=1}^{N} J_n J_n^{\mathsf{T}} \right)^{-1}. \tag{6.23}$$

A similar result was obtained in [59], where the asymptotic behaviour of $T(F_N)$ was analysed by taking a first order Taylor series expansion of (6.5). The only degree of freedom available for minimising the asymptotic variance is the appropriate choice of $\varphi(x)$. It is then critical to choose $\varphi(x)$ such that the asymptotic relative efficiency (ARE) with respect to Fisher information,

$$\mathrm{ARE} = \frac{\mathsf{E}[\varphi'(x)]^2}{\mathsf{E}[\varphi^2(x)] \, \mathsf{E}[\psi^2(x)]}. \tag{6.24}$$

is maximised over a wide class of $\psi(x)$.

### 6.3.2 Selection of the Penalty Function

The penalty and score functions are equivalent representations from the view of M-estimation, although the later is the more useful for two reasons. First, estimation is typically performed using $\varphi(x)$ by solving equation (6.5) while the performance and operation of M-estimators are more easily described in terms of the score function.

Many penalty functions are suggested in the literature based on experience, intuitive reasoning and the optimisation of performance measures such as the asymptotic variance or ARE.

#### 6.3.2.1 Score Functions Based on Heavy Tailed Distributions

The score functions may be of heavy tailed distributions such as the Cauchy and Laplace. The reason for such a choice is clear, near optimal performance will be attained at or near the chosen heavy tailed distribution. As shown in the literature this generally results in robust behaviour, avoiding the catastrophic failure of estimates based on Gaussianity. Furthermore, score functions chosen for their simple form reduce the computational complexity. The Laplace score function is particularly simple, being a hard limiter. Figure 6.1 shows the Gaussian, Laplace and Cauchy score functions. Note their differing limiting behaviour,

Gaussian score : $\lim_{|x|\to\infty} |\varphi(x)| = \infty$,

Laplace score : $\lim_{|x|\to\infty} |\varphi(x)| = 1$,

Cauchy score : $\lim_{|x|\to\infty} |\varphi(x)| = 0$.

The limiting behaviour of a score function is essentially what separates those which yield robust behaviour from those that do not. A general guide is as follows

Not robust : $|\varphi(x)| \propto |x|^p$ as $|x| \to \infty$ with $p \geq 1$. The score function constantly increases, placing a heavier weight on larger outliers. An example is any pdf whose tails decay at a rate equal to or greater than $\exp(-x^2)$, i.e. a light-tailed distribution.

Mildly robust : $|\varphi(x)| \propto |x|^p$ as $|x| \to \infty$ with $0 < p < 1$. The score function constantly increases, but at a decreasing rate. An example is the generalised Gaussian score function for $1 < \nu < 2$.

Robust : $|\varphi(x)| \propto |x|^p$ as $|x| \to \infty$ with $p = 0$. The score function has a nonzero limiting value for large $x$. Equal weight is given to all outliers. An example is any distribution with Laplacian tail behaviour.

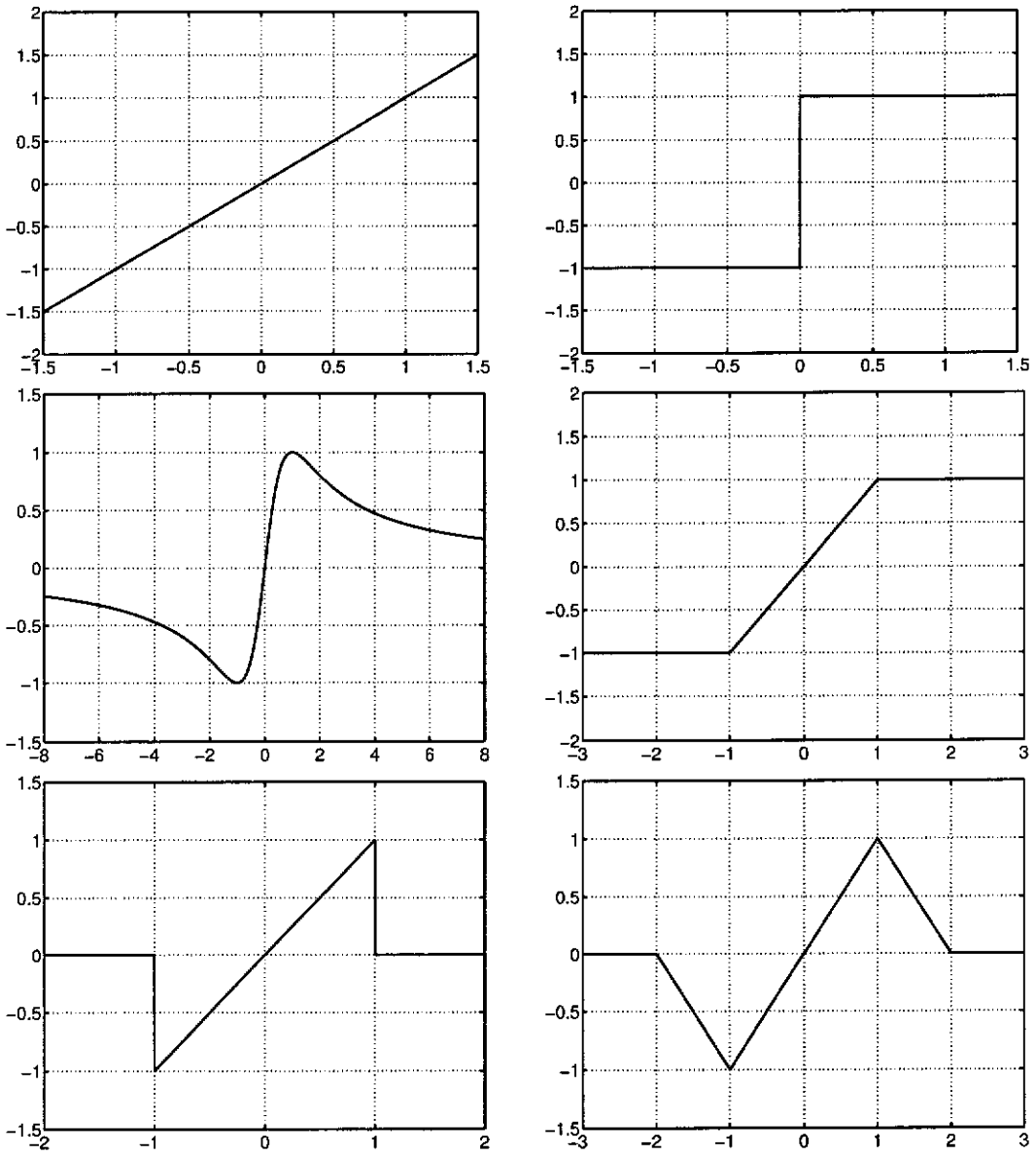

Figure 6.1: Some typical score functions used in robust estimation, the Gaussian score (top left), Laplace score or hard limiter (top right), Cauchy score (middle left), soft limiter (middle right), hole puncher (bottom left) and triangular score (bottom right)

Very robust : $|\varphi(x)| \propto |x|^p$ as $|x| \to \infty$ with $p < 0$. The score function redescends to 0, either in the limit or for a finite $x$ greater than some threshold. An example is any distribution with algebraic tail behaviour.

Redescending score functions are more suited to very heavy tailed distributions as they completely remove observations greater than a set value or in the limit.

#### 6.3.2.2 Generic Score Functions

As noted above, a common trait linking the score functions of heavy tailed distributions is their limiting behaviour, which is generally $|\varphi(x)| \propto |x|^p$ as $|x| \to \infty$ with $p \leq 0$. The effect of this is to reduce the influence of observations with very large magnitudes on the estimate, clipping, downweighting or even removing likely outliers. As $|x| \to 0$ the behaviour tends to be linear, so that observations small in magnitude are treated as they would be under the Gaussian model. Based on these general properties many score functions have been suggested for robust estimation. Some of the more common appear in Figure 6.1. Score functions derived from heavy tailed distributions, or generic ones such as those just shown, are used extensively in non-Gaussian detection theory where they appear as the zero memory nonlinearities of locally optimal, locally suboptimal and generalised nonlinear correlator detectors [95, 150, 135, 182, 31].

#### 6.3.2.3 Score Functions Based on Optimising Asymptotic Performance

Huber considered estimation of location in a two component mixture model, one component representing a nominal distribution, the other an unknown contaminating distribution [81]. The contaminant, being arbitrary and constrained only in that it is assumed to be symmetric, may be a heavy or light tailed distribution. However, its function here is the introduction of outliers and as such is assumed to possess heavier tails than the nominal.

A minimax criterion is used to obtain an explicit solution for the penalty function. The criterion being to minimise the maximum asymptotic variance of the M-estimate over the entire class of contaminating distributions. Another interpretation is to first find the least favourable distribution, that which minimises Fisher information and then to use the score function of this distribution in the M-estimator.

For a nominal Gaussian distribution the mixture model is

$$f_X(x) = (1-\varepsilon) f_G(x; \sigma_G^2) + \varepsilon f_C(x), \tag{6.25}$$

where $f_G(x; \sigma_G^2)$ is the nominal zero mean Gaussian pdf with variance $\sigma_G^2$ and $f_C(x)$ is the unknown symmetric contaminant. The minimax solution for $\varphi(x)$ is the soft limiter,

$$\varphi(x) = \begin{cases} \frac{x}{\sigma_G^2}, & \text{for } |x| \leq k\sigma_G^2 \\ k\text{sgn}(x), & \text{for } |x| > k\sigma_G^2 \end{cases}, \tag{6.26}$$

where $k$ is determined from $2f_G(k\sigma_G)/(k\sigma_G) - 2F_G(-k\sigma_G) = \varepsilon/(1-\varepsilon)$ and $f_G(x)$ and $F_G(x)$ are the standard normal pdf and cdf respectively. As explained previously, the principal of the soft limiter is to clip outliers, reducing their deleterious effect on the estimator. Given $\varphi(x)$, solutions to (6.5) can be obtained, for example, by iteratively reweighted least squares or the iterative modified residual algorithm [81].

Next, a variation on M-estimation is suggested where the penalty function is not static but is estimated from the observations.

## 6.4 Adaptive Robust Estimation

Although the soft limiter score function is well motivated, like any minimax solution it may be far from optimal for many distributions within the class for which it was designed. Similarly, near optimal performance cannot be maintained over a wide class of distributions for specific choices of the score function such as the Cauchy.

If the penalty function could adapt itself to the observations the estimates may still exhibit robust behaviour over a wide class of distributions, but with improved performance. This is what is proposed, instead of using a static penalty function in the M-estimator, it is estimated from the observations. Score function estimation is incorporated into any of the standard iterative parameter estimation procedures used to obtain the M-estimates. At each step the parameters are updated and the residuals found, these residuals are used to estimate the score function which in the next step are utilised to update the parameters and so forth. The algorithm is summarised in Table 6.1.

The most important component of the iterative algorithm is estimation of the score function. An obvious approach is to find an estimate of the noise density, followed by analytic or numerical differentiation to find the associated score function. Another is to estimate the score function directly, as in [170] where a linear combination of basis functions were used to approximate $\psi(x)$.

The former approach is considered next, based on a nonparametric density estimator which provides a nonparametric estimate for the score function.

## 6.5 Nonparametric Adaptive Robust Estimation

The use of nonparametric density estimates in the context of detection was considered in [194, 195]. There, a training sequence was used to estimate the locally

Table 6.1: Iterative algorithm for the adaptive robust estimator.

**Step 1.** Initialisation:

Set $i = 0$. Obtain an initial estimate of $\boldsymbol{\theta}$ based on the Gaussian assumption or using a simple robust estimator such as an M-estimate with the soft limiter.

**Step 2.** Determine the residuals:

$$\hat{x}_n = y_n - s_n(\hat{\boldsymbol{\theta}}_i).$$

**Step 3.** Estimate the score function:

From $\hat{x}_n$, estimate the score function $\varphi(x)$.

**Step 4.** Update the parameter estimates:

Using $\varphi(x)$, update the parameters from $\hat{\boldsymbol{\theta}}_i$ to $\hat{\boldsymbol{\theta}}_{i+1}$ by advancing one step in the iterative solution for a static M-estimator.

**Step 5.** Check for convergence:

If $|(\hat{\boldsymbol{\theta}}_{i+1})_k - (\hat{\boldsymbol{\theta}}_i)_k| < \epsilon|(\hat{\boldsymbol{\theta}}_i)_k|$ for $k = 1, \ldots, K$, stop, otherwise set $i \rightarrow i + 1$ and go to step 2.

optimum nonlinearity for detection of a signal in impulsive noise. A training sequence gives a priori information about $f_X(x)$, necessary to estimate the distribution of the observations under the null hypothesis that no signal is present. Without such knowledge the Neyman-Pearson detector cannot be implemented.

The proposed method as outlined in Table 6.1 requires no training sequence as it relies on an iterative scheme where the score function and parameters are estimated jointly. Next the problem of density estimation is addressed.

### 6.5.1 Density Estimation

The problem of density estimation is as follows, given $N$ iid observations, $x_n$, $n = 1, \ldots, N$, estimate the density $f_X(x)$. Density estimators may be separated into two broad classes, parametric and nonparametric. The former are based on parametric density models and imply estimation of any distributional parameters. Hybrids between parametric and nonparametric density estimators also exist, a case in point being the Gaussian-mixture distribution with a sufficient number

of components. Using the expectation-maximisation algorithm to estimate the distributional parameters, it was shown in [107, 106, 110] that a 4 component Gaussian-mixture distribution could model a wide variety of distributions including the $S\alpha S$. This approach is not pursued any further here, rather, the focus is on nonparametric methods as they are not constrained by any model, but allow the most general structure for a density.

A survey reveals many nonparametric density estimators including kernel and nearest neighbour methods, series estimators and maximum penalised likelihood estimators. Kernel methods are popular because of their wide applicability and the properties of the estimates [41, 163, 175]. Various disadvantages are associated with each of the other estimators. Nearest neighbour methods result in densities with discontinuous derivatives, which is to be avoided as the derivative is required for estimation of the score function. Series estimators may yield negative densities and possible oscillatory behaviour in the tails, accurate estimation of the tails is of prime importance for robust behaviour in heavy tailed distributions. Penalised likelihood estimators again specify parametric models for the density and require the distributional parameters to be estimated.

Here an adaptive kernel estimator is used, which is briefly reviewed before considering several extensions to cater for specific properties of $f_X(x)$ such as heavy tails, symmetry, unimodality and multimodality.

Fundamentally, kernel methods smooth the empirical density by placing a kernel at each observation, their sum giving an estimate of $f_X(x)$. The adaptive kernel estimator (AKE) is defined as

$$\hat{f}_X(x) = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{hb_n}K_f\left(\frac{x-x_n}{hb_n}\right), \tag{6.27}$$

where $K_f(x)$ is the kernel function, $h$ is the global bandwidth and the $b_n$ are the local bandwidths. As kernel methods inherit the properties of $K_f(x)$ it is recommended that $K_f(x)$ be differentiable, non-negative and integrate to one to guarantee a valid density. A Gaussian kernel was used since it fulfills all these requirements and gives good results over a wide range of distributions. The local bandwidths are used to adjust for local structure in $f_X(x)$ such as heavy tails, while the global bandwidth affects $\hat{f}_X(x)$ globally and may be used to control smoothness.

#### 6.5.1.1 Bandwidth Selection

The local and global bandwidths are found as follows. First, a pilot estimate for the density, $\hat{f}_0(x)$, is found by setting $b_n = 1$ and choosing to a suitable $h$, as will be discussed.

Second, estimate the local bandwidths as $b_n = (\hat{f}_0(x_n)/\Pi_{n=1}^N \hat{f}_0(x_n)^{1/N})^{-\vartheta}$ where $0 \leq \vartheta \leq 1$ controls the sensitivity to $\hat{f}_0(x)$. A recommended choice is $\vartheta = 1/2$ since asymptotically the resulting estimator will have a bias of lower order than when $\vartheta = 0$. Division by the geometric mean of the density estimate at the observations conserves the scale of the density. Finally, evaluate $\hat{f}_X(x)$ at the required values.

For fixed $b_n$, an optimal $h$ can be obtained by minimising a measure of distance between $\hat{f}_X(x)$ and $f_X(x)$ with respect to $h$. An often used measure is the mean integrated squared error, MISE $= \mathsf{E}[\int_{-\infty}^{\infty}(\hat{f}_X(x) - f_X(x))^2\,dx]$. For the special case of a Gaussian distribution and kernel $h_{\text{opt}} = 0.79\text{IQR}N^{-1/5}$ is obtained, where IQR is the interquartile range. In general $h_{\text{opt}}$ is a reasonable choice even for non-Gaussian $K_f(x)$ and $f_X(x)$, however, should $f_X(x)$ be multimodal this choice gives large errors. In this case, or merely to refine $h$, more intelligent methods of bandwidth selection such as cross-validation or resampling schemes such as the bootstrap are required [51, 163].

#### 6.5.1.2 Bandwidth Corrections for Heavy Tailed Distributions

Even with local bandwidth adjustments the AKE may undersmooth and produce spurious peaks in the tails for heavy tailed distributions. Modes in the tails are undesirable as they produce a score function which oscillates in the extremities. This may lead to inaccurate estimates as convergence of the iterative parameter estimation algorithms depend on the properties of $|\varphi'(x)|$, on which more will be said later.

Possibly the simplest way to remove these peaks is to selectively increase the local bandwidths. The tails are then corrected while not distorting the estimate near the mode. First define the tail regions as the lower and upper $100\xi\%$ of the observations, $\xi = 0.25$ was found to be adequate. For each tail move from the extremal observations inwards, adjusting local bandwidths for successive pairs of observations. If $\hat{f}_X(x)$ at the midpoint between the observations is less than that at both observations a local minima is deemed present and the two local bandwidths are increased. The correction is based on the assumption of a Gaussian kernel. If two Gaussian kernels are separated by a distance of $2\Delta$ their sum will

be unimodal if their local bandwidths are at least $\Delta/h$. The component with the smaller local bandwidth is assigned this critical value, while the other is assigned the critical value scaled by the ratio of the original bandwidths to preserve their original proportionality. It can occur that one of the modified local bandwidths is less than the original, in which case the greater one is kept. This simple scheme was found to produce good results, removing spurious peaks when the tails were heavy.

#### 6.5.1.3 Symmetry

Recall that for (6.5) to give consistent estimates $f_X(x)$ should be symmetric and $\psi(x)$ antisymmetric. $\psi(x)$ is ensured to be antisymmetric by imposing symmetry on $\hat{f}_X(x)$. Let $\hat{f}_X(x)$ be the raw asymmetric estimate, then a symmetric estimate, $\hat{f}_s(x)$, is obtained by taking the even part of $\hat{f}_X(x)$, $\hat{f}_s(x) = (\hat{f}_X(x) + \hat{f}_X(-x))/2$.

#### 6.5.1.4 Unimodality

If $f_X(x)$ is known to be unimodal, then incorporating this information into the estimator will reduce error. To impose unimodality on kernel estimators it is suggested to increase $h$, this follows from a result for Gaussian kernels which states that the number of modes is a decreasing function of $h$ [163]. To check for unimodality determine the number of peaks in $\hat{f}_X(x)$, should there be more than a single peak $h$ is increased by a factor $\eta$. The process is then repeated until unimodality is attained, with $\eta = 1.05$ a few iterations was usually enough.

#### 6.5.1.5 Consistency of the Estimates

Pointwise and uniform consistency using the kernel estimators can be achieved under some mild regularity conditions on $f_X(x)$ and $K_f(x)$, and some further conditions on $h$, both of which are usually fulfilled. Uniform consistency implies that $\Pr[\sup_x |\hat{f}_X(x) - f_X(x)| \to 0] = 1$ as $N \to \infty$, more details and other consistency results can be found in [41, 163, 175].

### 6.5.2 Application to Multiuser Detection

Present and emerging wireless communications standards such as CDMA2000 (code division multiple access) and WCDMA (wideband code division multiple access) implement direct sequence spread spectrum CDMA. The primary advantages of CDMA are that it allows multiple users access to the channel, gives a

soft degradation in performance as the number of users increase and can offer secure transmission [77, 142].

MUD techniques are being intensively investigated for CDMA communications because they have the ability to combat the multiple access interference caused by the presence of more than one user in the channel. At best MUD can achieve single user performance by removing multiple access interference altogether. Most MUD schemes are designed under the assumption of Gaussian noise and tend to be linear in nature, the decorrelating detector which removes multiple access interference for high SNR is simply a linear transformation [187].

Impulsive noise can severely degrade the performance of MUD detectors based on the Gaussian assumption, warranting the design of MUD schemes for impulsive noise [1, 2, 3].

#### 6.5.2.1 Multiuser Detection Signal Model

Consider a direct sequence spread spectrum multiple access communications channel where $K$ users transmit synchronously. The received signal is

$$y_n = \sum_{k=1}^{K} (\boldsymbol{S})_{nk} A_k b_k + x_n \qquad n = 1, \ldots, N. \tag{6.28}$$

$(\boldsymbol{S})_{nk}$ denotes element $(n, k)$ of the matrix $\boldsymbol{S}$ whose columns consist of the normalised spreading codes of the $K$ users, $\boldsymbol{S} = (\boldsymbol{s}_1, \ldots, \boldsymbol{s}_K)$, the codes being of length $N$. $A_k > 0$ is the amplitude of user $k$ and $b_k \in \{-1, 1\}$ is the bit sent by user $k$. $x_n$ is additive iid zero mean noise. The vector form of the above model, $\boldsymbol{y} = \boldsymbol{S}\boldsymbol{\theta} + \boldsymbol{x}$, will be used, where $\boldsymbol{\theta} = \boldsymbol{A}\boldsymbol{b}$, $\boldsymbol{A} = \mathrm{diag}(A_1, \ldots, A_K)$, $\boldsymbol{b} = (b_1, \ldots, b_K)^{\mathsf{T}}$ and $\theta_k = A_k b_k$ denotes element $k$ of $\boldsymbol{\theta}$, so that all the unknowns are collected into $\boldsymbol{\theta}$.

Classical least squares (LS) gives $\hat{\boldsymbol{\theta}}_{\mathrm{LS}}$ as

$$\hat{\boldsymbol{\theta}}_{\mathrm{LS}} = \underset{\boldsymbol{\theta}}{\mathrm{argmin}} \sum_{n=1}^{N} \left( y_n - \sum_{k=1}^{K} (\boldsymbol{S})_{nk} \theta_k \right)^2, \tag{6.29}$$

for which the solution is

$$\hat{\boldsymbol{\theta}}_{\mathrm{LS}} = (\boldsymbol{S}^{\mathsf{T}}\boldsymbol{S})^{-1}\boldsymbol{S}^{\mathsf{T}}\boldsymbol{y}. \tag{6.30}$$

Existence of $(\boldsymbol{S}^{\mathsf{T}}\boldsymbol{S})^{-1}$ is ensured as the columns of the code matrix are linearly independent spreading codes. The users' amplitudes and data bits are easily recovered as

$$\hat{A}_k = \left| \hat{\theta}_k \right|, \qquad \hat{b}_k = \mathrm{sgn}\left( \hat{\theta}_k \right). \tag{6.31}$$

This gives the so called decorrelating detector, aptly named since it decorrelates the users, removing the multiple access interference [187]. For Gaussian $x_n$ the LS solution for $\boldsymbol{\theta}$ is equivalent to the ML solution, obtained by solving

$$\sum_{n=1}^{N}(\boldsymbol{S})_{nk}\ \psi\left(x_n - \sum_{m=1}^{K}(\boldsymbol{S})_{nm}\theta_m\right) = 0, \qquad k = 1,\ldots,K, \tag{6.32}$$

where in general $x_n$ has a density $f_X(x)$. In impulsive noise of unknown distribution both these methods have severe disadvantages. First, the decorrelating detector is sensitive to impulsive noise, its performance rapidly degrades as the level of impulsive behaviour increases. Second, even for known $f_X(x)$, the computational cost of ML is often too high.

#### 6.5.2.2 Robust Multiuser Detection

An M-estimator for the MUD problem is obtained by replacing the true score function $\psi(x)$ in (6.32) by $\varphi(x)$,

$$\sum_{n=1}^{N}(\boldsymbol{S})_{nk}\ \varphi\left(x_n - \sum_{m=1}^{K}(\boldsymbol{S})_{nm}\theta_m\right) = 0, \qquad k = 1,\ldots,K. \tag{6.33}$$

In [188] a robust MUD was developed using an M-estimator with Huber's penalty function and the modified residual method, herein this is referred to as the minimax detector. The same approach was used in [151] to develop a robust MUD for flat fading channels. An alternative was proposed in [14, 15, 13] where adaptive chip based nonlinearities were used to mitigate impulsive noise, followed by adaptive filtering which is needed to combat the multiple access interference.

The adaptive equivalent of the M-estimator based robust MUD uses a nonparametric density estimator to estimate the score function and is summarised in Table 6.2, it will be referred to as the nonparametric detector.

The iterative parameter estimation algorithms for both the minimax and nonparametric detectors will only converge to a unique minimum if the step size, $\mu_{ss}$, satisfies $\mu_{ss} \leq 1/|\psi'(x)|$ [188]. For the minimax detector the step size is set a priori to $1/\sigma^2$, the inverse noise variance, where it exists or to the noise scale where it does not. For the nonparametric detector the step size is set adaptively to

$$\mu_{ss} = \frac{1}{\varrho \max\limits_{n} |\psi'(x_n)|}, \tag{6.34}$$

where $\varrho \geq 1$ allows for a margin of error when estimating $\psi'(x_n)$, $\varrho = 1.25$ was used.

Table 6.2: Iterative algorithm for the nonparametric adaptive robust estimator as applied to MUD.

**Step 1.** Initialisation:

Set $i = 0$. Obtain an initial estimate of $\boldsymbol{\theta}$ from the decorrelator, $\hat{\boldsymbol{\theta}}_0 = \hat{\boldsymbol{\theta}}_{\text{LS}}$.

**Step 2.** Determine the residuals:

$$\hat{\boldsymbol{x}} = \boldsymbol{y} - \boldsymbol{S}\hat{\boldsymbol{\theta}}_i.$$

**Step 3.** Estimate the score function:

From $\hat{\boldsymbol{x}}$, estimate the density, $\hat{f}_X(x)$ and evaluate the score function estimate as

$$\varphi(x) = -\frac{\hat{f}'_X(x)}{\hat{f}_X(x)}.$$

**Step 4.** Update the parameter estimates:

Evaluate $\boldsymbol{z} = \varphi(\hat{\boldsymbol{x}})$ and update the parameters,

$$\hat{\boldsymbol{\theta}}_{i+1} = \hat{\boldsymbol{\theta}}_i + \mu_{ss}(\boldsymbol{S}^{\mathsf{T}}\boldsymbol{S})^{-1}\boldsymbol{S}^{\mathsf{T}}\boldsymbol{z},$$

where $\mu_{ss}$ is the step size.

**Step 5.** Check for convergence:

If $|(\hat{\boldsymbol{\theta}}_{i+1})_k - (\hat{\boldsymbol{\theta}}_i)_k| < \epsilon|(\hat{\boldsymbol{\theta}}_i)_k|$ for $k = 1, \ldots, K$, stop, otherwise set $i \rightarrow i + 1$ and go to step 2.

#### 6.5.2.3 Experiments

Three detectors are compared here, the classical decorrelating detector, the minimax detector and the proposed nonparametric detector, all over a variety of noise models. Unless otherwise stated all noise models are symmetric and unimodal, this being incorporated into the density estimator. Bit error rates (BER) versus SNR are shown for $K = 6$ users. The spreading codes are shifted maximal length sequences of length $N = 31$, meaning density estimation is performed with a sample size of 31. The amplitude of the first user is at -10dB relative to the others, BERs are shown for this user. The SNR is defined as $A^2/\sigma^2$ where $A$ is the amplitude of the user of interest, the first user, and $\sigma^2$ is the noise variance. For $S\alpha S$ noise, $\sigma$ is replaced by $c$, the scale parameter.

For Gaussian noise in Figure 6.2 the decorrelator and minimax detectors perform similarly. This is expected as for Gaussian noise the influence function for the minimax detector clips only a minority of the incoming observations, but appears linear, the optimum shape in Gaussian noise, for the majority. Since the nonparametric detector estimates the score function it is never perfectly linear for Gaussian noise, leading to a decrease in performance of approximately 1dB.

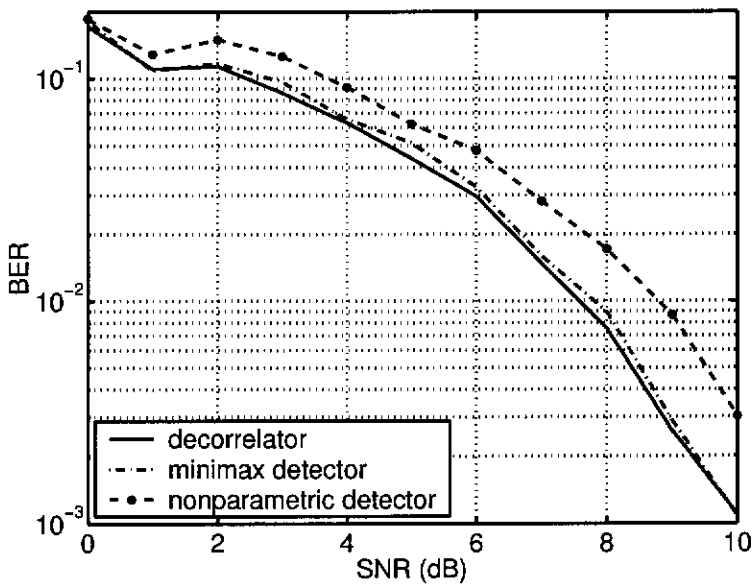


Figure 6.2: Bit error rate versus SNR in Gaussian noise.

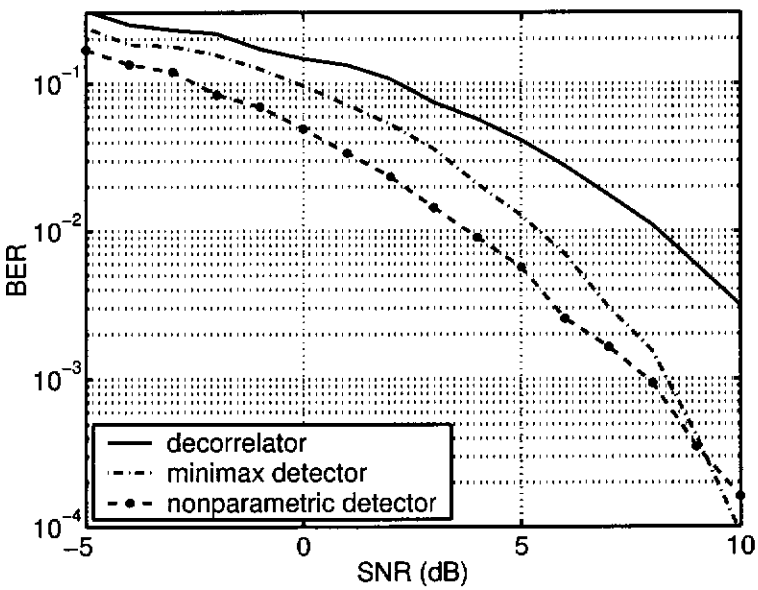


Figure 6.3: Bit error rate versus SNR in generalised Gaussian noise with $\nu = 0.5$.

Figure 6.3 shows results for a generalised Gaussian distribution with $\nu = 0.5$. The nonparametric detector has the lowest BER up to an SNR of 9dB. There appears to be a range of $\nu$ and SNR which favor use of the nonparametric detector, namely for low $\nu$ and SNR, a region where accurate detection is difficult. This suggests that the nonparametric detector is more robust in highly impulsive noise.

Next, results are shown for the $\varepsilon-$mix model with $\varepsilon = 0.01$, $\kappa = 100$ and $\varepsilon = 0.1$, $\kappa = 100$ in Figures 6.4 and 6.5 respectively. Both the minimax and nonparametric detectors outperform the decorrelator, while the minimax detector outperforms the nonparametric detector in the first case but not the second. Which detector is best depends on how the model parameters affect the accuracy of the density estimation. In the first case there are a small proportion of very large noise values, making tail estimation difficult as there are so few observations in the tail regions. In the second case large noise values occur with a greater probability, aiding in tail estimation and hence estimation of the score function, leading to smaller BERs. The nonparametric detector will then perform well given $N\varepsilon$ is sufficiently large to allow accurate estimation of the tails.

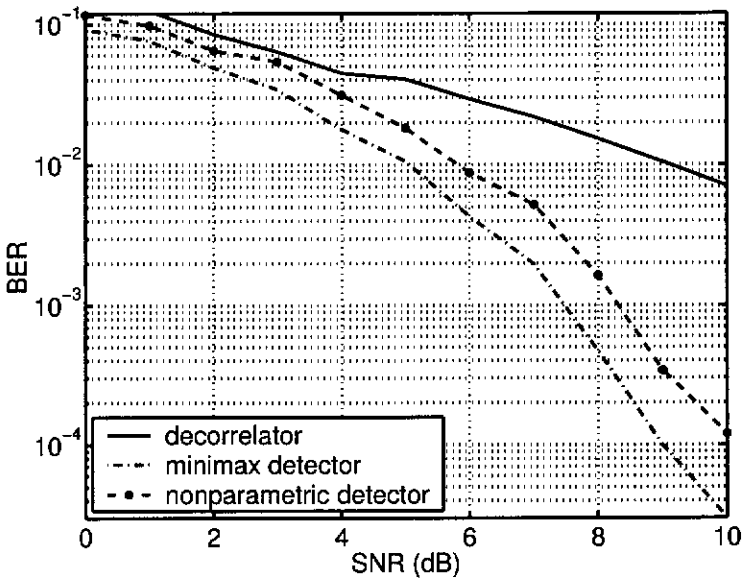


Figure 6.4: Bit error rate versus SNR in $\varepsilon$ mixture noise with $\varepsilon = 0.01$, $\kappa = 100$.

Next, consider two cases where the noise is $S\alpha S$. Theoretical motivation for $S\alpha S$ distributions as a model for impulsive noise in multiuser environments can be found in [84]. Results for $\alpha = 1.5$ and $\alpha = 1.0$ (Cauchy) are shown in Figures 6.6 and 6.7 respectively. In these cases the nonparametric detector outperforms both the minimax and decorrelating detectors. Further experiments have shown this

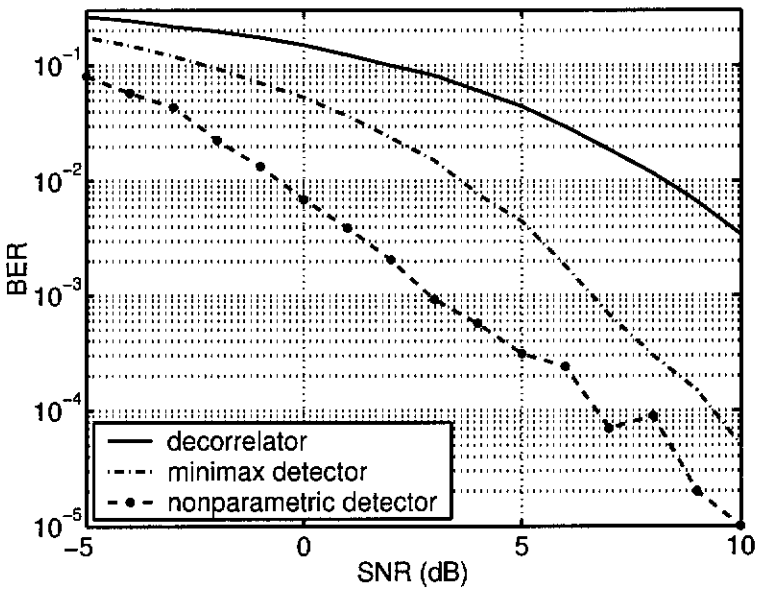


Figure 6.5: Bit error rate versus SNR in $\varepsilon$ mixture noise with $\varepsilon = 0.1$, $\kappa = 100$.

to be true for smaller values of $\alpha$. This gives more evidence to the suggestion that the nonparametric detector is more robust to highly impulsive noise.

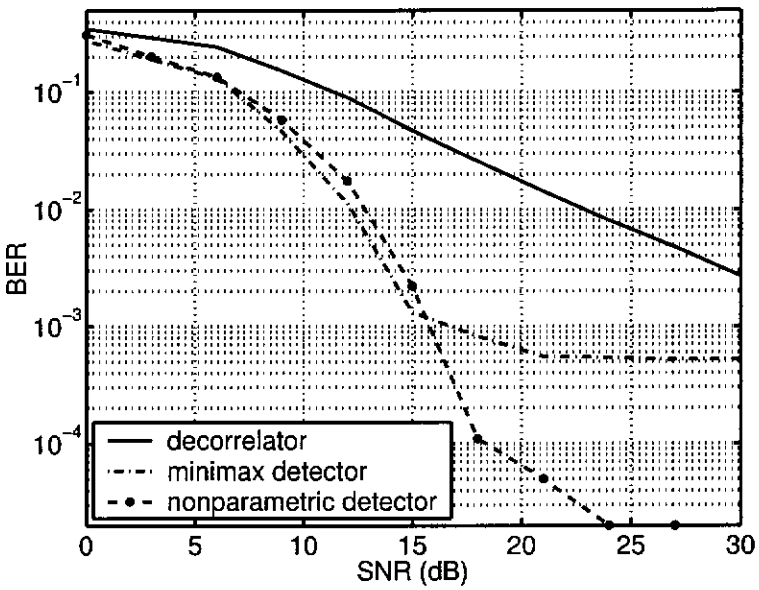


Figure 6.6: Bit error rate versus SNR in $S\alpha S$ noise for $\alpha = 1.5$.

For the final example the noise distribution is symmetric but bimodal, this is incorporated into the density estimation procedure as outlined in Section 6.5.1. The bimodal density is composed of two Gaussian densities with means of $\pm 1$ and a variance of $\sigma^2 = 0.01$.

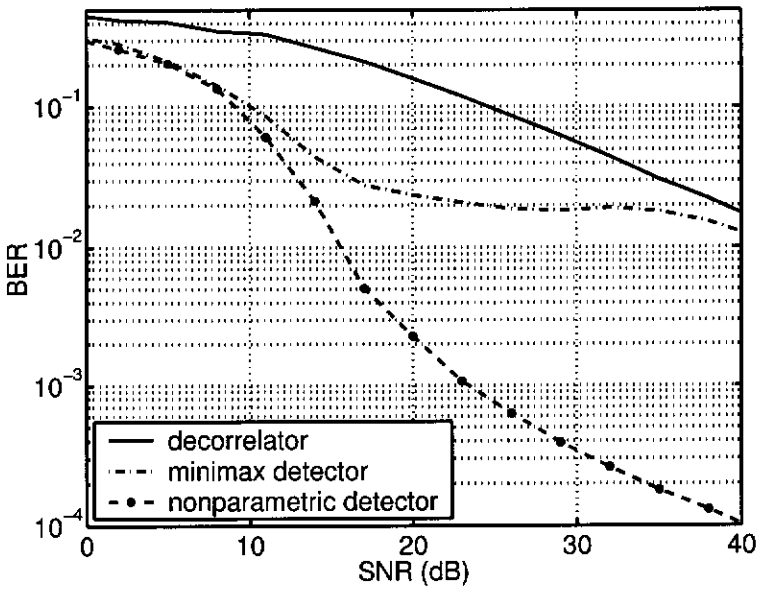


Figure 6.7: Bit error rate versus SNR in $S\alpha S$ noise for $\alpha = 1.0$.

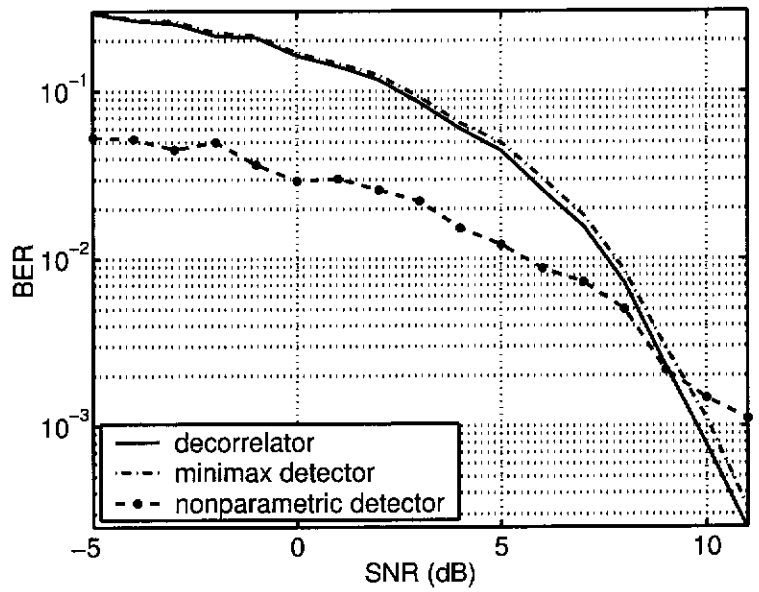


Figure 6.8: Bit error rate versus SNR in symmetric bimodal noise generated from a two component Gaussian mixture model where the components have a mean of $\pm 1$ and a common variance of $0.1^2$.

Figure 6.8 shows results for this scenario. The nonparametric detector is able to adapt itself to the bimodal nature of the noise and outperforms the other methods at low SNR.

To briefly comment on computational complexity, it was found the nonpara-

metric detector took on average 4 iterations to converge while the minimax detector took 2.

## 6.6 Parametric Adaptive Robust Estimation

The nonparametric adaptive robust estimator makes minimal assumptions about the noise and as such can adapt to a fairly broad range of models. The flexibility of nonparametric methods is countered by a drop in performance relative to parametric ones, suggesting the use of a parametric model for the score function. In [170] an adaptive robust estimator was proposed which modelled the score function parametrically as a linear combination of basis functions,

$$\varphi(x) = \sum_{q=1}^{Q} a_q g_q(x) = \boldsymbol{a}^{\mathsf{T}} \boldsymbol{g}(x), \tag{6.35}$$

where $\boldsymbol{a} = (a_1, \ldots, a_Q)^{\mathsf{T}}$ are the weights and $\boldsymbol{g}(x) = (g_1(x), \ldots, g_Q(x))^{\mathsf{T}}$ are the bases.

The bases are chosen for their ability to approximate $\psi(x)$. For instance, the $g_q(x)$ can simply be a set of score functions obtained from distributions known to be close, in some sense, to $f_X(x)$. The weights can then be chosen to minimise some measure of distance between $\varphi(x)$ and $\psi(x)$ or to maximise the performance of the estimator.

A sensible measure of distance between $\varphi(x)$ and $\psi(x)$ is the MSE, from which the weights are determined as

$$\boldsymbol{a} = \underset{\boldsymbol{a}}{\operatorname{argmin}}\ \mathsf{E}\big[(\varphi(x) - \psi(x))^2\big]. \tag{6.36}$$

Given that the following condition holds

$$\lim_{x \to \pm\infty} g_q(x) f_X(x) = 0, \tag{6.37}$$

$\boldsymbol{a}$ is obtained as the solution to the normal equations,

$$\mathsf{E}\big[\boldsymbol{g}(x)\boldsymbol{g}^{\mathsf{T}}(x)\big]\, \boldsymbol{a} = \mathsf{E}[\boldsymbol{g}'(x)]\,. \tag{6.38}$$

Estimates of the weights obtained in this fashion are referred to as the MMSE (minimum MSE) solution,

$$\boldsymbol{a} = \mathsf{E}\big[\boldsymbol{g}(x)\boldsymbol{g}^{\mathsf{T}}(x)\big]^{-1}\, \mathsf{E}[\boldsymbol{g}'(x)]\,. \tag{6.39}$$

In practice expectation is replaced with an empirical mean. An alternate interpretation for the solution (6.39) was developed in [170]. Consider the expression for the ARE with $\varphi(x) = \boldsymbol{a}^{\mathsf{T}}\boldsymbol{g}(x)$,

$$\text{ARE} = \frac{\boldsymbol{a}^{\mathsf{T}}\mathsf{E}[\boldsymbol{g}'(x)]\,\mathsf{E}[\boldsymbol{g}'(x)]^{\mathsf{T}}\boldsymbol{a}}{\boldsymbol{a}^{\mathsf{T}}\mathsf{E}\left[\boldsymbol{g}(x)\boldsymbol{g}(x)^{\mathsf{T}}\right]\boldsymbol{a}} \cdot \frac{1}{\mathsf{E}[\psi(x)^2]}, \tag{6.40}$$

then estimating the weights according to (6.39) also maximises the above expression for the ARE, or equivalently, minimises the asymptotic variance. Since this approach attempts to minimise the asymptotic variance it can be expected to improve upon the minimax solution when the set of basis functions model the true score function satisfactorily.

This estimate of the score function is then incorporated into the adaptive robust estimator show in Table 6.1, the algorithm is summarised in Table 6.3.

Table 6.3: Iterative algorithm for the parametric adaptive robust estimator.

**Step 1.** Initialisation:

Set $i = 0$. Obtain an initial estimate of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}_0$.

**Step 2.** Determine the residuals:

$$\hat{x}_n = y_n - s_n(\hat{\boldsymbol{\theta}}_i).$$

**Step 3.** Estimate the score function:

From $\hat{x}_n$, estimate the weights,

$$\hat{\boldsymbol{a}} = \left(\sum_{n=1}^{N} \boldsymbol{g}(\hat{x}_n)\boldsymbol{g}^{\mathsf{T}}(\hat{x}_n)\right)^{-1} \sum_{n=1}^{N} \boldsymbol{g}'(\hat{x}_n),$$

the score function estimate is then $\varphi(x) = \hat{\boldsymbol{a}}^{\mathsf{T}}\boldsymbol{g}(x)$.

**Step 4.** Update the parameter estimates:

Using $\varphi(x)$, update the parameters from $\hat{\boldsymbol{\theta}}_i$ to $\hat{\boldsymbol{\theta}}_{i+1}$ by advancing one step in the iterative solution for a static M-estimator.

**Step 5.** Check for convergence:

If $|(\hat{\boldsymbol{\theta}}_{i+1})_k - (\hat{\boldsymbol{\theta}}_i)_k| < \epsilon|(\hat{\boldsymbol{\theta}}_i)_k|$ for $k = 1, \ldots, K$, stop, otherwise set $i \rightarrow i + 1$ and go to step 2.

#### 6.6.0.4 Existence of the Minimum Mean Square Error Solution for $a$

To estimate the weights using (6.39) the two expectations, $\mathsf{E}\big[\boldsymbol{g}(x)\boldsymbol{g}^{\mathsf{T}}(x)\big]$ and $\mathsf{E}[\boldsymbol{g}'(x)]$ must exist. Their existence depends on the limiting behaviour of both the noise distribution and the bases. To gain some insight into this, consider distributions with algebraic tails such as the $S\alpha S$ or Pareto distributions which have the limiting behaviour

$$f_X(x) \propto |x|^{-\alpha-1}, \quad |x| \to \infty, \tag{6.41}$$

where $\alpha > 0$. These are the heaviest tails which can be expected in practice. The limiting behaviour of the bases will be modelled as

$$g_q(x) \propto \operatorname{sgn}(x)\,|x|^{-p_q}, \quad |x| \to \infty, \tag{6.42}$$

where $p_q$ is the limiting rate of decay of the basis function $g_q(x)$. There are three regimes for the limiting behaviour of $g_q(x)$ depending on the sign of the decay rate,

$$\begin{aligned} p_q < 0 &: \; g_q(x) \to \infty, \\ p_q = 0 &: \; g_q(x) \to c, \text{ a constant}, \\ p_q > 0 &: \; g_q(x) \to 0. \end{aligned} \tag{6.43}$$

Element $(i,j)$ of $\mathsf{E}\big[\boldsymbol{g}(x)\boldsymbol{g}^{\mathsf{T}}(x)\big]$ will exist if

$$\int_{-\infty}^{\infty} g_i(x) g_j(x) f_X(x)\,dx < \infty. \tag{6.44}$$

As the bases are antisymmetric and the noise pdf symmetric, the expression inside the integral is even and so becomes

$$2\int_{0}^{\infty} g_i(x) g_j(x) f_X(x)\,dx, \tag{6.45}$$

which will exist if

$$\lim_{x\to\infty} \int g_i(x) g_j(x) f_X(x)\,dx < \infty. \tag{6.46}$$

For sufficiently large $x$, the approximations (6.41) and (6.42) may be substituted into the above integral, yielding

$$\begin{aligned} \lim_{x\to\infty} \int x^{-p_i} x^{-p_j} x^{-\alpha-1}\,dx &< \infty \\ \lim_{x\to\infty} x^{-p_i-p_j-\alpha} &< \infty \\ p_i + p_j &> -\alpha. \end{aligned} \tag{6.47}$$

Since $\alpha > 0$, the final condition for existence of the integral and hence the expectation is $p_i + p_j \geq 0$. If $p$ is the smallest limiting rate of decay across all bases, $\mathsf{E}\left[\boldsymbol{g}(x)\boldsymbol{g}^{\mathsf{T}}(x)\right]$ will exist if $p \geq 0$.

The $q^{\text{th}}$ element of $\mathsf{E}[\boldsymbol{g}'(x)]$ will exist if,

$$\int_{-\infty}^{\infty} g_q'(x) f_X(x)\, dx < \infty. \tag{6.48}$$

Following a similar argument as above leads to $p_q > -\alpha - 1$ as the condition for existence of the integral. It follows that $\mathsf{E}[\boldsymbol{g}'(x)]$ will exist when $p \geq -1$.

Combining these two requirements gives $p \geq 0$, so that provided every basis has a non-increasing limiting behaviour, a solution will exist. The soft limiter is an example for which $p = 0$.

By the strong law of large numbers the empirical expectations are consistent estimates of the statistical expectations if

$$\mathsf{E}[g_i(x)g_j(x)] < \infty, \tag{6.49}$$

$$\mathsf{E}\left[g_q'(x)\right] < \infty, \tag{6.50}$$

that is, $p \geq 0$.

#### 6.6.0.5 Inclusion of the True Score Function in the Basis Set

Let the set of bases contain the true score function $\psi(x)$. Then the estimated score function $\varphi(x)$ is asymptotically equal to $\psi(x)$.

Proof: Without loss of generality define $\boldsymbol{g}(x)^{\mathsf{T}} = \left(\psi(x),\, \boldsymbol{g}_0(x)^{\mathsf{T}}\right)^{\mathsf{T}}$ where $\boldsymbol{g}_0(x)$ does not contain $\psi(x)$. The weights are obtained by solving the normal equations,

$$\begin{aligned}
\boldsymbol{a} &= \mathsf{E}\left[\boldsymbol{g}(x)\boldsymbol{g}(x)^{\mathsf{T}}\right]^{-1} \mathsf{E}[\boldsymbol{g}'(x)] \\
&= \mathsf{E}\left[\boldsymbol{g}(x)\boldsymbol{g}(x)^{\mathsf{T}}\right]^{-1} \mathsf{E}[\psi(x)\boldsymbol{g}(x)] \\
&= \begin{pmatrix} \mathsf{E}[\psi^2(x)] & \mathsf{E}\left[\psi(x)\boldsymbol{g}_0(x)^{\mathsf{T}}\right] \\ \mathsf{E}[\psi(x)\boldsymbol{g}_0(x)] & \mathsf{E}\left[\boldsymbol{g}_0(x)\boldsymbol{g}_0(x)^{\mathsf{T}}\right] \end{pmatrix}^{-1} \begin{pmatrix} \mathsf{E}[\psi^2(x)] \\ \mathsf{E}[\psi(x)\boldsymbol{g}_0(x)] \end{pmatrix} \\
&= \begin{pmatrix} G_{\psi\psi} & \boldsymbol{G}_{\psi 0}{}^{\mathsf{T}} \\ \boldsymbol{G}_{\psi 0} & \boldsymbol{G}_{00} \end{pmatrix}^{-1} \begin{pmatrix} G_{\psi\psi} \\ \boldsymbol{G}_{\psi 0} \end{pmatrix}.
\end{aligned} \tag{6.51}$$

By the partitioned matrix inversion formula this becomes

$$\begin{pmatrix} \boldsymbol{G}_{11} & \boldsymbol{G}_{12} \\ \boldsymbol{G}_{21} & \boldsymbol{G}_{22} \end{pmatrix} \begin{pmatrix} G_{\psi\psi} \\ \boldsymbol{G}_{\psi 0} \end{pmatrix}, \tag{6.52}$$

where

$$\begin{aligned} \boldsymbol{G}_{11} &= \left(G_{\psi\psi} - \boldsymbol{G}_{\psi 0}{}^{\mathsf{T}}\boldsymbol{G}_{00}^{-1}\boldsymbol{G}_{\psi 0}\right)^{-1} \\ \boldsymbol{G}_{12} &= -\left(G_{\psi\psi} - \boldsymbol{G}_{\psi 0}{}^{\mathsf{T}}\boldsymbol{G}_{00}^{-1}\boldsymbol{G}_{\psi 0}\right)^{-1}\boldsymbol{G}_{\psi 0}{}^{\mathsf{T}}\boldsymbol{G}_{00}^{-1} \\ \boldsymbol{G}_{21} &= -\left(\boldsymbol{G}_{00} - \boldsymbol{G}_{\psi 0}G_{\psi\psi}^{-1}\boldsymbol{G}_{\psi 0}{}^{\mathsf{T}}\right)^{-1}\boldsymbol{G}_{\psi 0}G_{\psi\psi}^{-1} \\ \boldsymbol{G}_{22} &= \left(\boldsymbol{G}_{00} - \boldsymbol{G}_{\psi 0}G_{\psi\psi}^{-1}\boldsymbol{G}_{\psi 0}{}^{\mathsf{T}}\right)^{-1}, \end{aligned} \tag{6.53}$$

so that

$$\boldsymbol{a} = \begin{pmatrix} \boldsymbol{G}_{11}\boldsymbol{G}_{11}^{-1} \\ \boldsymbol{G}_{22}\left(\boldsymbol{G}_{\psi 0} - \boldsymbol{G}_{\psi 0}G_{\psi\psi}^{-1}G_{\psi\psi}\right) \end{pmatrix} = \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix} \tag{6.54}$$

and finally,

$$\varphi(x) = \boldsymbol{a}^{\mathsf{T}}\boldsymbol{g}(x) = \begin{pmatrix} 1 & \mathbf{0}^{\mathsf{T}} \end{pmatrix}\begin{pmatrix} \psi(x) \\ \boldsymbol{g}_0(x) \end{pmatrix} = \psi(x), \tag{6.55}$$

given that the inverses exist.

### 6.6.1 Small Sample Performance

To gain some insight into the small sample performance a simulation study was carried out using the expression for the ARE (6.40) as a performance criterion. $\boldsymbol{a}$ was estimated from a finite sample and substituted into the expression for the ARE, results were averaged over 1000 realisations. The benefit of such a criterion is that it is independent of any signal parameters, so the results can be regarded as representative.

The $\varepsilon-$mix distribution was used with parameters $\varepsilon = 0.01$, $\kappa = 10$ and $\sigma = 1$. The bases were $\varepsilon-$mix score functions with parameters $(\varepsilon_1 = 0.1, \kappa_1 = 100)$, $(\varepsilon_2 = 0.1, \kappa_2 = 10)$, $(\varepsilon_3 = 0.01, \kappa_3 = 100)$, $(\varepsilon_4 = 0.01, \kappa_4 = 10)$.

Results are shown in Figure 6.9 for the MMSE estimator where the performance measure is referred to as the 'efficiency'. Even though this measure cannot strictly be interpreted as information loss, it gives a good approximation and clearly should be as near unity as possible. For large sample sizes performance does appear optimal, though the MMSE solution performs quite poorly for small sample sizes.

Several methods were assessed in an attempt to reduce this small sample effect. A weighting function introduced into the MSE criterion of (6.36) in an attempt to arrive at a weighted least squares solution gave unpredictable results. Consistently improved performance was attained by constraining the weights, that is, by solving the normal equations of (6.38) subject to linear constraints on $\boldsymbol{a}$. First the constraint $\boldsymbol{a} > \mathbf{0}$, was imposed. The reasoning being that for

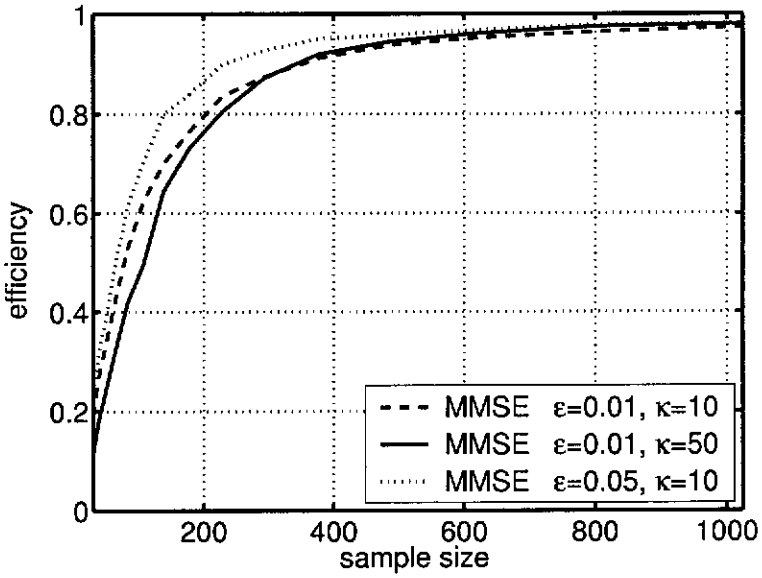


Figure 6.9: Small sample performance of the MMSE estimator for the weights.

small sample sizes it is possible for negative values of $a_q$ to allow the estimated score function to become negative for $x > 0$ or visa-versa, an undesirable solution corresponding to a multimodal distribution. Second, the constraint $\sum_{q=1}^{Q} a_q = 1$ was imposed since the score function only needs to be known to within a constant of proportionality. Combining these two constraints gives the constrained MMSE (CMMSE) estimator. In Figure 6.10 the performance of the CMMSE estimator is shown under the same conditions as in Figure 6.9 for the MMSE estimator. Clearly, the small sample performance has been greatly improved by using the CMMSE estimator. As is shown next, the attendant loss of the CMMSE for large sample sizes is very small.

### 6.6.2 Large Sample Performance

Consider estimation when the noise follows an $\varepsilon-$mix distribution and there is uncertainty about the model parameters $\varepsilon$ and $\kappa$. Figure 6.11 shows the ARE of the proposed CMMSE estimator relative to Huber's minimax estimator as the parameters of the $\varepsilon-$mix distribution are varied over the region $0.01 \leq \varepsilon \leq 0.1$, $10 \leq \kappa \leq 100$. This region covers the parameter space most likely to be encountered. The 4 bases chosen for estimation of the score function were $(\varepsilon_1 = 0.01, \kappa_1 = 10)$, $(\varepsilon_2 = 0.02, \kappa_2 = 100)$, $(\varepsilon_3 = 0.1, \kappa_3 = 50)$, $(\varepsilon_4 = 0.1, \kappa_4 = 100)$. The CMMSE estimator is superior to the minimax estimator over this range of parameters, its efficiency is up to 15% higher. The

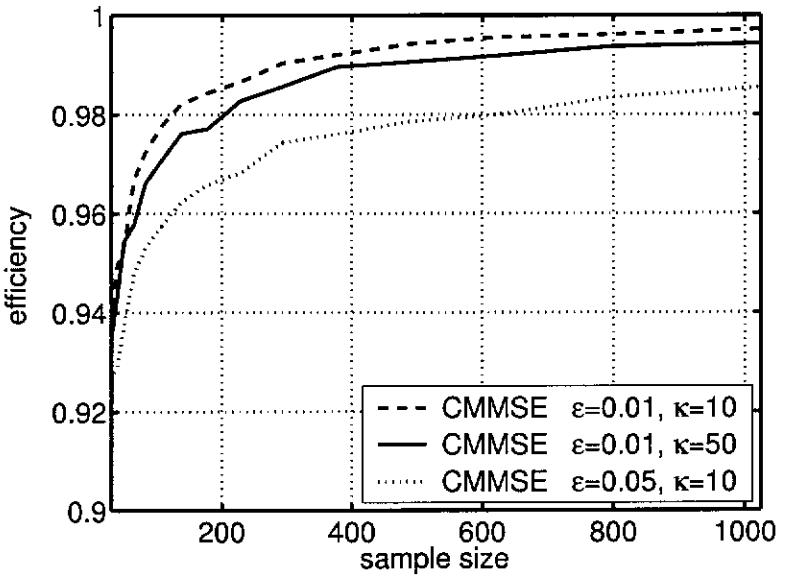


Figure 6.10: Small sample performance of the CMMSE estimator for the weights.

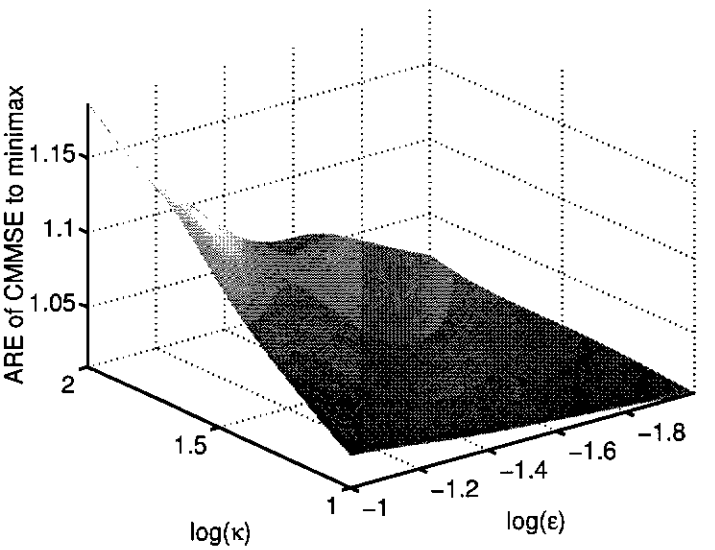


Figure 6.11: ARE of CMMSE estimation relative to minimax estimation.

ARE of the the CMMSE, MMSE and minimax estimators with respect to ML are shown in Figures 6.12, 6.13 and 6.14 respectively over the same region of the $\varepsilon-$mix parameter space and for the same bases as in Figure 6.11. The MMSE estimator looses least of all, at most 0.8%, next comes the CMMSE estimator with a 2% loss relative to ML, finally, the minimax estimator looses up to 15% relative to ML. It follows that for large samples the CMMSE and MMSE estima-

tors yield near optimal performance over a large range of the parameter space.

The effects of the location of the bases on the ARE curves of Figures 6.12

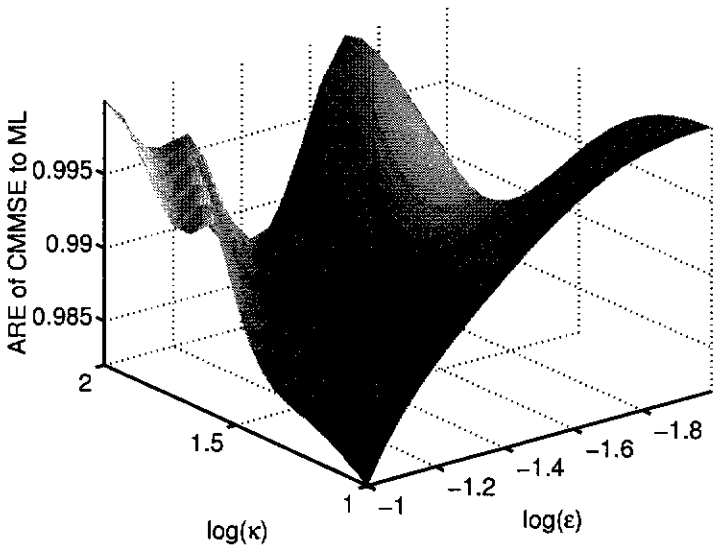


Figure 6.12: ARE of CMMSE estimation relative to ML estimation.

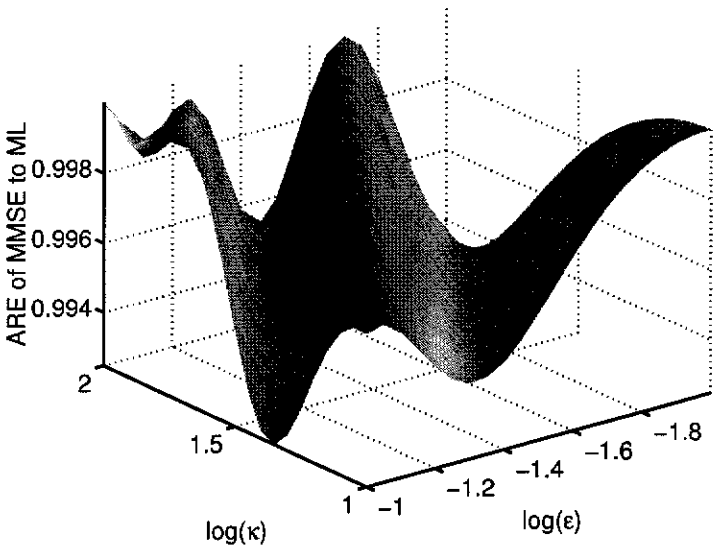


Figure 6.13: ARE of MMSE estimation relative to ML estimation.

and 6.13 is, in general, complicated and cannot be predicted easily. What can be predicted is the location of local maxima where the ARE is 1. Recall the result of Section 6.6.0.5 where it was shown that including the true score function in the basis set led to the estimated score function asymptotically equaling the true

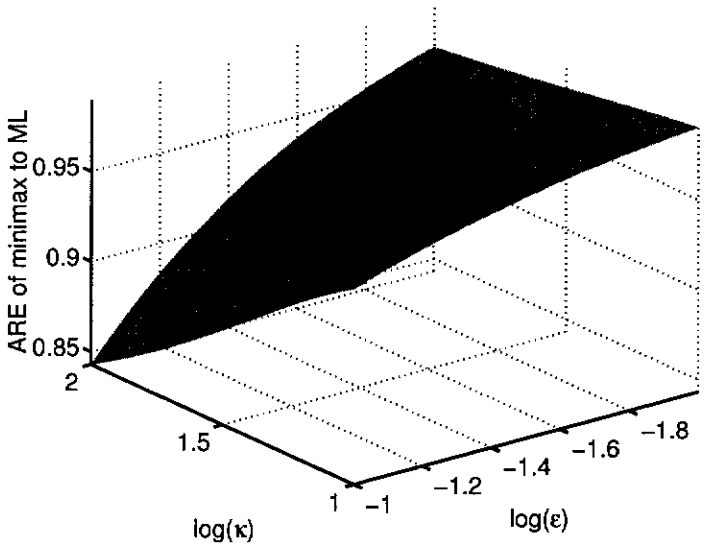


Figure 6.14: ARE of minimax estimation relative to ML estimation.

score function. From this result it follows that the ARE of both the CMMSE and MMSE estimators takes the maximum value of 1 at points in the parameter space where an $\varepsilon-$mix basis function is placed, that is, a local maximum exists. These points can also be regarded as global maxima, their multiplicity being determined by the number of $\varepsilon-$mix bases used. Note though, that if none of the bases is proportional to the true score function over the parameter space of interest of the noise distribution, the ARE will not reach its maximum value of 1 over the chosen parameter space. Global or local maxima may still exist, but their location is not easily found.

Asymptotically, the efficiency of the CMMSE estimator with respect to Fisher information can be forced arbitrarily close to unity by increasing the number of bases in the parameter space of interest. Likewise, including the score functions of other impulsive noise models in the set of bases yields near optimal performance for those impulsive noise distributions.

## 6.7 Summary

A generic adaptive robust estimator was proposed for the general signal in additive noise model. The influence function of the M-estimator corresponding to the signal model was derived, from which the asymptotic covariance of the M-estimates was found. Based on the M-estimation concept, the estimator adap-

tively estimates the score function of the noise distribution directly from the observations. Both a nonparametric and parametric approach are taken to estimate the score function.

The first approach uses a nonparametric adaptive kernel density estimator modified for use with heavy tailed distributions. Assumptions on the noise distribution are minimal, including symmetry and unimodality, which are likely in practice. Experiments in the context of MUD in impulsive noise show the nonparametric estimator can outperform the minimax M-estimate for very heavy tailed distributions. When the noise is slightly impulsive performance is improved relative to the decorrelating detector, the least squares solution. For complicated noise models, such as for multimodal densities, the nonparametric detector is able to adapt to the noise distribution.

In the second approach, the score function was approximated in a parametric fashion by a linear combination of bases, the weights being found by minimising the MSE between the estimated and true score functions. For small sample sizes the MMSE estimate of the weights was found to perform poorly. Several constraints on the weights were proposed which improved performance for small samples with a negligible loss in efficiency for large samples. Using the $\varepsilon-$mix distribution as an example it was shown that careful selection of the bases gives near optimal performance over the parameter space of practical interest, surpassing the minimax solution.

# Chapter 7

# Conclusion

*A witty saying proves nothing.*

*— Voltaire*

This thesis dealt with some problems associated with signal processing in heavy tailed noise. The first part concentrated on both validation and parameter estimation for the noise model. The second part considered two specific signal processing problems, source detection and signal estimation for heavy tailed models.

The $S\alpha S$ distribution is a popular heavy tailed model motivated by theoretical and empirical results. Goodness-of-fit tests for $S\alpha S$ distributions tend to be qualitative or exhibit low power against heavy tailed alternatives. In Chapter 3 a goodness-of-fit test for $S\alpha S$ distributions was developed based on their unique stability property. Null distributions were derived using asymptotic theory and estimated with the parametric bootstrap. Results show the bootstrap procedure has a higher power when detecting heavy tailed alternatives compared to asymptotic theory or edf tests, while accurately maintaining the set level. In conjunction with this work optimal sampling of the ecf for $S\alpha S$ parameter estimation using Koutrouvelis' cf domain procedure was addressed.

From a practical standpoint, an inadequacy of the $S\alpha S$ model is the lack of a Gaussian component to describe thermal noise. The $S\alpha SG$ distribution, which results from the sum of independent $S\alpha S$ and Gaussian random variables, has appeared in the literature in reply to this. In Chapter 4 an estimator the $S\alpha SG$ distribution was developed based on NWLS in the cf domain. The NWLS estimator performs well compared to existing methods. In particular, estimation of the characteristic exponent, which describes the level of impulsive behaviour, was much improved.

Source enumeration is an important precursor to advanced array processing techniques such as high resolution DOA estimation. Chapter 5 presented a source

enumeration procedure which determines the multiplicity of the smallest sample eigenvalues of the array covariance matrix. A hypothesis testing framework was developed to compare all pairs of eigenvalues for equality, which are combined using MHTs. The bootstrap was used to estimate the null distributions of the test statistics. Improvements over traditional techniques such as the MDL and sphericity tests were demonstrated for small sample sizes and heavy tailed models for the sources and noise.

Robust estimators are designed to be insensitive to changes in the statistical model and have been widely used when outliers or impulsive noise are present. A generic adaptive robust estimator was proposed in Chapter 6 based on the concept of M-estimation, but with the score function of the noise distribution being estimated directly from the observations. This allows a greater flexibility over static M-estimators and leads to improved performance. Two estimates for the score function was considered. First, a nonparametric density estimator was proposed which performed well in very heavy tailed noise compared to robust procedures based on the minimax M-estimator. The second modelled the score function as a linear combination of basis functions, constraints were proposed which greatly improved small sample performance and still gave significant improvements over minimax M-estimation for large samples.

## 7.1 Future Directions

### 7.1.1 Testing for Symmetric Alpha Stable Distributions

*Optimising the number of segments*: Though the minimal number of segments necessary to determine stability were used, the power of the test may be increased by using more. As discussed in Chapter 3 the power of the single tests can be expected to decrease when the number of segments is increased beyond a certain number, this number being determined by the rate of convergence of sums of the random variable to its limiting distribution and the sample size. For instance, the increase in power when using three as opposed to two segments was relatively large for the $\varepsilon-\text{mix}(0.01, 100)$ and $\varepsilon-\text{mix}(0.1, 10)$ distributions.

*Testing for univariate and multivariate $\alpha S$ distributions*: Subject to some modifications, the stability property is a definition for both univariate and multivariate $\alpha S$ distributions, hence the stability test can be generalised to these cases.

The stability property for univariate $\alpha S$ distributions states that sums of iid $\alpha S$ random variables are also $\alpha S$ with the same characteristic exponent and

skewness parameter. The test is then

$$\begin{aligned} \mathsf{H} &: \alpha_X = \alpha_{Z_2} \cap \alpha_X = \alpha_{Z_3}, \qquad \beta_X = \beta_{Z_2} \cap \beta_X = \beta_{Z_3} \\ \mathsf{K} &: \alpha_X = \alpha_{Z_2} \cup \alpha_X = \alpha_{Z_3}, \qquad \beta_X = \beta_{Z_2} \cup \beta_X = \beta_{Z_3} \end{aligned} \tag{7.1}$$

and is a simple extension of the stability test for $S\alpha S$ distributions. Estimates of $\alpha$ and $\beta$ are obtained using Koutrouvelis' procedure and the bootstrap is used to estimate the null distributions of the test statistics which are combined using a MHT. Alternatively, the asymptotic distributions of the test statistics can be found. Edf tests for $\alpha S$ distributions are difficult to implement because of the associated computational complexity involved in evaluation of the $\alpha S$ cdf and estimation of the thresholds for the edf statistics over a suitable range of $\alpha$ and $\beta$.

Multivariate $\alpha S$ distributions have a more complex structure, the skewness and scale parameters being combined into a finite measure on the unit sphere. Perhaps the most straightforward approach to testing for multivariate $\alpha S$ distributions was suggested in [137, 139] and is based on the following theorem (Theorem 2.1.5 [156]).

Let $\boldsymbol{X}$ be a $d$-dimensional random vector in $\mathbb{R}^d$. $\boldsymbol{X}$ is a 1) strictly $\alpha S$, 2) $S\alpha S$, 3) $\alpha S$ random vector in $\mathbb{R}^d$ if all linear combinations of the components of $\boldsymbol{X}$ are 1) strictly $\alpha S$, 2) $S\alpha S$, 3) $\alpha S$ with $\alpha \geq 1$.

If follows that to test for either of these three cases, every linear combination must be assessed using a univariate test. In practice only a finite number of directions along which to project the random vector are chosen.

The advantage of this approach is that only univariate tests and hence estimators are required. A general multivariate approach would require estimators for the characteristic exponent of a multivariate $\alpha S$ distribution, for which only a few limited techniques exist [153].

*Effect of estimator on stability test*: It is emphasised that the performance of the stability test is largely dependent on the performance of the estimator for $\alpha$, hence ML estimates could be used to improve the power of the test at the cost of increased computational complexity.

### 7.1.2 Estimation for the Symmetric Alpha Stable Gaussian Sum Distribution

*Optimal sampling of the ecf*: Just as the asymptotic distribution of Koutrouvelis' estimator for the $S\alpha S$ distribution was found, so too can the asymptotic

distributions for the parameters of the $S\alpha SG$ distribution based on the estimator presented in Chapter 4. Again, this would allow optimal selection of the points at which to sample the ecf, based on minimising the asymptotic MSE of the estimates.

*Testing for $S\alpha SG$ distributions*: Specific tests for $S\alpha SG$ distributions have yet to be considered. For a $S\alpha SG$ null hypothesis versus an open alternative hypothesis, that is, a goodness-of-fit test for $S\alpha SG$ distributions, ecf methods similar to those of [200] can be developed.

Another useful test determines whether impulsive noise is present in addition to thermal noise. In this case the test is for a Gaussian null hypothesis versus a $S\alpha SG$ alternative hypothesis. The classical likelihood or generalised likelihood ratio test is disadvantaged by problems associated with numerical evaluation of the $S\alpha S$ and $S\alpha SG$ densities. Simpler cf domain techniques are envisioned motivated by the k-L procedure [53]. The concept rests on the ecf being asymptotically MVN, with a mean and covariance easily determined from the parametric cf under each hypothesis. Consider the above problem, a test for the Gaussian versus $S\alpha SG$ distributions. For several points in the cf domain, let $\hat{\boldsymbol{\phi}}_G$ be an estimate for the parametric cf under the null hypothesis and $\hat{\boldsymbol{\phi}}$ be the ecf. Then under the null hypothesis, $\boldsymbol{C}(\hat{\boldsymbol{\phi}} - \hat{\boldsymbol{\phi}}_G) \overset{a}{\sim} \text{MVN}(\boldsymbol{0}, \boldsymbol{I})$, where $\boldsymbol{C}$ is the Cholesky decomposition of the inverse covariance matrix of the ecf. A similar result holds under the alternative hypothesis from which it follows that a generalised likelihood ratio type test can be implemented based on the asymptotic normality of the ecf.

### 7.1.3 Detection of Sources in Array Processing

*Alternate measures of dependence*: Throughout Chapter 5 the sample covariance was used in the source detection algorithm. Improvements in performance were observed for non-Gaussian observations where the sample covariance is not an optimal estimate because of the flexibility of the bootstrap in estimating the null distributions of the test statistics.

A $S\alpha S$ model for either the sources or noise poses a problem as the analytic covariance no longer exists. Even though the sample covariance will always be finite, it is a very poor estimate of the true dependence structure, itself possessing infinite variance.

Alternate measures of dependence with acceptable performance such as the sample covariation should then be used. While the sample covariation is specific to the $S\alpha S$ distribution for $\alpha > 1$, it is still a useful measure for distributions which are not $\alpha S$. If generic measures are sought, a multitude of robust estimators

for dependence also exist and may be applied under any heavy tailed model including $S\alpha S$ observations with $\alpha \leq 1$.

Use of these estimators in conjunction with existing techniques such as the MDL and sphericity test should improve performance when the observations are heavy tailed. Though well motivated, it should be remembered that both are based on the asymptotic statistics of the covariance matrix for Gaussian observations and that in dropping these assumptions unknown losses are incurred.

It is here that bootstrap source detection has an advantage, still estimating the finite sample distributions of the test statistics regardless of the measure of dependence used or the distribution of the observations. Under the same observation model and using identical estimators for dependence, the bootstrap method should offer an improvement over the MDL and sphericity test.

Further extensions include application of bootstrap source detection to the wideband and deterministic signals model.

### 7.1.4 Robust Estimation and Detection

*Kernel selection*: It has been suggested that the kernel should match the noise model, so that a Cauchy kernel should be used when the noise is Cauchy. However, the Cauchy kernel results in heavy tail estimates even for light tailed distributions, hence this is not recommended. Instead, an automatic kernel selector is needed. Given a repertoire of possible kernels a scheme can be envisaged where a goodness-of-fit criterion is minimised with respect to the kernel, using for instance, cross-validation or resampling methods. The set of kernels need not be very large but must capture the essential characteristics of the distributions expected, so that kernels with heavy tails should be included to cater for heavy tailed distributions. *Bandwidth selection*: Existing automatic bandwidth selection methods based on cross-validation or the bootstrap assign equal weight across the entire density. In goodness-of-fit tests a weighting function is used so that different regions of the density contribute more or less to the final decision. When testing for heavy tailed distributions more weight is placed on the tail regions by using a weighting function concentrated away from the mode. This idea can be incorporated into cross-validation and bootstrap IMSE bandwidth selection methods, producing bandwidths favoring good estimation of the tails, possibly leading to better estimates for heavy tailed distributions. Alternatively, adaptive kernel density estimators can be developed which use the nearest neighbour distance to determine local bandwidths.

*Alternate pdf estimators*: Nonparametric kernel density estimators were considered here due to their favourable properties, nevertheless, other estimators exist. A promising alternative mentioned in Chapter 5 is to use the expectation-maximisation algorithm to estimate the parameters of a Gaussian-mixture distribution. As shown in [107, 106, 110] a Gaussian-mixture distribution with only 4 components can accurately model a wide variety of distributions including the $S\alpha S$.

*Rank based score functions*: The nonparametric estimator explicitly includes estimation of the noise scale, while the parametric estimator does not. Many techniques exist for joint estimation of the scale and signal parameters. An alternative approach is the use of rank based score functions, analogous to the rank based zero memory nonlinearities used in nonparametric detection theory. As rank based functions are invariant to changes of scale the problem of scale estimation is neatly avoided.

*Alternate nominal distributions for minimax estimation*: Robust minimax estimators can be developed for any nominal symmetric distribution. By using one of the heavy tailed distributions as the nominal performance may be improved for very heavy tailed distributions.

*Basis selection*: Although a large number of bases enable accurate modelling of a wide range of noise distributions, two problems may arise. First, too many bases will result in the signal being modelled in addition to the noise, second, computational complexity is increased. Given a repertoire of basis functions, a parsimonious model selection procedure is required. When subsets of the same cardinality are chosen, a criterion based on the MSE between the estimated and true score functions is feasible.

The estimated score function can also be related back to a distribution, so that likelihood procedures can be developed both for estimation of the weights and model selection. In this case the Kullback-Leibler information criterion can be used to choose between subsets with the same cardinality [158], otherwise the more general AIC and MDL criterion are required.

# Appendix A

# Characteristic Functions

*There are three kinds of lies: lies,*
*damn lies, and statistics.*
*— Benjamin Disraeli*

The characteristic function (cf) of a random variable $X$ with pdf $f_X(x)$ is defined as the Fourier transform of $f_X(x)$,

$$\phi_X(\omega) = \int_{-\infty}^{\infty} f_X(x) e^{\jmath x\omega}\, dx = \mathsf{E}\left[e^{\jmath X\omega}\right], \quad -\infty < \omega < \infty. \tag{A.1}$$

The inverse transform is

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_X(\omega) e^{-\jmath x\omega}\, d\omega. \tag{A.2}$$

There is a one to one correspondence between the cf and pdf, so the cf retains all the information present in the latter description. Furthermore, the cf always exists. The cf is a bounded function,

$$0 \leq |\phi_X(\omega)| \leq 1 \tag{A.3}$$

and so all of its moments exist regardless of whether the same is true of $X$. This is one advantage of using cf domain techniques.

A comprehensive coverage of cf theory may be found in [114], for a concise overview see [44] and the references therein.

## A.1 The Empirical Characteristic Function and its Statistical Properties

Let $X_n$, $n = 1, \ldots, N$, be iid observations of the random variable $X$. An estimate of $\phi_X(\omega)$ is given by the empirical characteristic function (ecf)

$$\hat{\phi}_X(\omega) = \frac{1}{N} \sum_{n=1}^{N} e^{\jmath X_n \omega}. \tag{A.4}$$

### A.1.1 Pointwise Convergence

For a fixed $\omega$, $\hat{\phi}_X(\omega)$ is an average of bounded iid random variables with means $\phi_X(\omega)$. By the strong law of large numbers $\hat{\phi}_X(\omega)$ converges to $\phi_X(\omega)$ w.p.1 [54]

$$\Pr\left[\lim_{N\to\infty} \left|\hat{\phi}_X(\omega) - \phi_X(\omega)\right| = 0\right] = 1. \tag{A.5}$$

### A.1.2 Uniform Convergence

The ecf does not converge over the entire real line, but only over a finite region. Take $T < \infty$, then

$$\Pr\left[\lim_{N\to\infty} \sup_{\omega \le T} \left|\hat{\phi}_X(\omega) - \phi_X(\omega)\right| = 0\right] = 1. \tag{A.6}$$

Exceptions where convergence is uniform over the entire real line are when $X$ is a purely discrete random variable or when a kernel cf estimator, subject to some conditions, is used in place of the ecf [54].

### A.1.3 Asymptotic Distribution

Consider $Y_N(\omega) = N^{1/2}(\hat{\phi}_X(\omega) - \phi_X(\omega))$ which has zero mean and covariance structure $\mathrm{Cov}[Y_N(\omega_i), Y_N(\omega_j)] = \phi_X(\omega_i - \omega_j) - \phi_X(\omega_i)\phi_X^*(\omega_j)$. Let $Y(\omega)$ be a zero mean complex Gaussian process with the same mean and covariance structure as $Y_N(\omega)$ and satisfying $Y^*(\omega) = Y(-\omega)$. Then subject to some conditions [37]

**Theorem 2** *$Y_N(\omega)$ converges in distribution to $Y(\omega)$ over every finite interval of the real line.*

This result originally appears as Theorem 3.1 in [54], but is incorrect in its stated generalisation. However, for finite dimensional distributions the result does generally hold. Define $\hat{\boldsymbol{\phi}}_X = (\hat{\phi}_X(\omega_1), \ldots, \hat{\phi}_X(\omega_K))^{\mathsf{T}}$ for $K < \infty$. $\hat{\boldsymbol{\phi}}_X$ has mean $\boldsymbol{\mu}_{\boldsymbol{\phi}} = (\phi_X(\omega_1), \ldots, \phi_X(\omega_K))^{\mathsf{T}}$ and covariance $(\boldsymbol{R}_{\boldsymbol{\phi}})_{ij} = 1/N(\phi_X(\omega_i - \omega_j) - \phi_X(\omega_i)\phi_X{}^*(\omega_j))$.

**Theorem 3** *By the multidimensional CLT*

$$N^{1/2}(\hat{\boldsymbol{\phi}}_X - \boldsymbol{\phi}_X) \xrightarrow{d} \text{CMVN}(\mathbf{0}, N\boldsymbol{R}_{\boldsymbol{\phi}}) \tag{A.7}$$

*and hence*

$$\hat{\boldsymbol{\phi}}_X \overset{a}{\sim} \text{CMVN}(\boldsymbol{\phi}_X, \boldsymbol{R}_{\boldsymbol{\phi}}) . \tag{A.8}$$

Only the real part of the ecf is used here, for which the asymptotic distribution is MVN with mean $(\boldsymbol{\mu}_{\text{Re}(\boldsymbol{\phi})})_i = \mathsf{E}[\text{Re}(\hat{\boldsymbol{\phi}}_X(\omega_i))]$ and covariance $(\boldsymbol{R}_{\text{Re}(\boldsymbol{\phi})})_{ij} = \mathsf{Cov}[\text{Re}(\hat{\phi}_X(\omega_i)), \text{Re}(\hat{\phi}_X(\omega_j))]$,

$$\text{Re}\left(\hat{\boldsymbol{\phi}}_X\right) \overset{a}{\sim} \text{MVN}\left(\boldsymbol{\mu}_{\text{Re}(\boldsymbol{\phi})}, \boldsymbol{R}_{\text{Re}(\boldsymbol{\phi})}\right) . \tag{A.9}$$

### A.1.4 Statistics of $\text{Re}(\hat{\phi}_X(\omega))$

The following results, (A.10, A.11, A.13, A.14), are stated without proof in [101]. The proofs are shown here for completeness.

#### A.1.4.1 Mean

Since $\hat{\phi}_X(\omega)$ is an unbiased estimator of $\phi_X(\omega)$,

$$\mathsf{E}\left[\text{Re}\left(\hat{\phi}_X(\omega)\right)\right] = \text{Re}(\phi_X(\omega)) . \tag{A.10}$$

#### A.1.4.2 Covariance

$$\begin{aligned}
&\mathsf{Cov}\left[\text{Re}\left(\hat{\phi}_X(\omega_i)\right), \text{Re}\left(\hat{\phi}_X(\omega_j)\right)\right] \\
&= \mathsf{E}\left[\text{Re}\left(\hat{\phi}_X(\omega_i)\right) \text{Re}\left(\hat{\phi}_X(\omega_j)\right)\right] - \mathsf{E}\left[\text{Re}\left(\hat{\phi}_X(\omega_i)\right)\right] \mathsf{E}\left[\text{Re}\left(\hat{\phi}_X(\omega_j)\right)\right] \\
&= \frac{1}{N^2} \sum_{n=1}^{N} \sum_{m=1}^{N} \mathsf{E}[\cos(X_n\omega_i) \cos(X_m\omega_j)] - \text{Re}(\phi_X(\omega_i)) \text{Re}(\phi_X(\omega_j)) \\
&= \frac{1}{2N} \mathsf{E}[\cos(X(\omega_i - \omega_j)) + \cos(X(\omega_i + \omega_j))] \\
&\quad + \frac{N-1}{N} \mathsf{E}[\cos(X\omega_i)] \mathsf{E}[\cos(X\omega_j)] \\
&\quad - \text{Re}(\phi_X(\omega_i)) \text{Re}(\phi_X(\omega_j)) \\
&= \frac{1}{2N} \left(\text{Re}(\phi_X(\omega_i - \omega_j)) + \text{Re}(\phi_X(\omega_i + \omega_j)) - 2\text{Re}(\phi_X(\omega_i)) \text{Re}(\phi_X(\omega_j))\right)
\end{aligned} \tag{A.11}$$

It follows that the variance of $\text{Re}(\hat{\phi}_X(\omega))$ is

$$\mathsf{Var}\left[\text{Re}\left(\hat{\phi}_X(\omega)\right)\right] = \frac{1}{2N} \left(1 + \text{Re}(\phi_X(2\omega)) - 2\text{Re}(\phi_X(\omega))^2\right) \tag{A.12}$$

### A.1.5 Statistics of $|\hat{\phi}_X(\omega)|^2$

#### A.1.5.1 Mean

$$\begin{aligned}
\mathsf{E}\left[\left|\hat{\phi}_X(\omega)\right|^2\right] &= \mathsf{E}\left[\frac{1}{N}\sum_{n=1}^{N} e^{\jmath X_n\omega}\frac{1}{N}\sum_{m=1}^{N} e^{-\jmath X_m\omega}\right] \\
&= \frac{1}{N^2}\sum_{\substack{n=1\\m=n}}^{N}\sum_{m=1}^{N}\mathsf{E}\left[e^{\jmath(X_n-X_m)\omega}\right] + \frac{1}{N^2}\sum_{\substack{n=1\\m\neq n}}^{N}\sum_{m=1}^{N}\mathsf{E}\left[e^{\jmath X_n\omega}e^{-\jmath X_m\omega}\right] \\
&= \frac{1}{N} + \frac{N-1}{N}\phi(\omega)\phi^*(\omega) \\
&= |\phi_X(\omega)|^2 + 1/N(1-|\phi_X(\omega)|^2). \qquad \text{(A.13)}
\end{aligned}$$

#### A.1.5.2 Covariance

The covariance is tedious but straightforward,

$$\mathsf{Cov}\left[\left|\hat{\phi}_X(\omega_i)\right|^2, \left|\hat{\phi}_X(\omega_j)\right|^2\right] = \mathsf{E}\left[\left|\hat{\phi}_X(\omega_i)\right|^2\left|\hat{\phi}_X(\omega_j)\right|^2\right] - \mathsf{E}\left[\left|\hat{\phi}_X(\omega_i)\right|^2\right]\mathsf{E}\left[\left|\hat{\phi}_X(\omega_j)\right|^2\right].$$

Taking the first term on the RHS gives

$$\begin{aligned}
&\frac{1}{N^4}\mathsf{E}\left[\sum_{n=1}^{N} e^{\jmath X_n\omega_i}\sum_{m=1}^{N} e^{-\jmath X_m\omega_i}\sum_{l=1}^{N} e^{\jmath X_l\omega_j}\sum_{k=1}^{N} e^{-\jmath X_k\omega_j}\right] \\
= \frac{1}{N^4}&\sum_{n=1}^{N}\sum_{m=1}^{N}\sum_{l=1}^{N}\sum_{k=1}^{N}\mathsf{E}\left[e^{\jmath X_n\omega_i}e^{-\jmath X_m\omega_i}e^{\jmath X_l\omega_j}e^{-\jmath X_k\omega_j}\right].
\end{aligned}$$

There are several possible relationships between the indices of the sums which result in distinct expressions giving the following,

$$n = m = l = k \qquad \frac{1}{N^3}$$

$$\begin{aligned}
n = m = l &\neq k \\
n = m = k &\neq l
\end{aligned} \qquad 2\frac{N-1}{N^3}|\phi(\omega_j)|^2$$

$$\begin{aligned}
n = k = l &\neq m \\
m = k = l &\neq n
\end{aligned} \qquad 2\frac{N-1}{N^3}|\phi(\omega_i)|^2$$

$$n = m \neq l = k \qquad \frac{N-1}{N^3}$$

$$n = k \neq m = l \qquad \frac{N-1}{N^3}|\phi(\omega_i-\omega_j)|^2$$

$$n = l \neq m = k \qquad \frac{N-1}{N^3}|\phi(\omega_i+\omega_j)|^2$$

$$
\begin{array}{llll}
n \neq m & n \neq l & m \neq l & n = k \\
n \neq m & n \neq l & m \neq l & m = k \\
n \neq m & n \neq l & m \neq l & l = k \\
n \neq m & n \neq k & m \neq k & n = l \\
n \neq m & n \neq k & m \neq k & m = l \\
m \neq l & m \neq k & l \neq k & n = m
\end{array}
\quad
\begin{array}{l}
\frac{(N-1)(N-2)}{N^3}\phi(\omega_i - \omega_j)\phi^*(\omega_i)\phi(\omega_j) \\
\frac{(N-1)(N-2)}{N^3}\phi^*(\omega_i + \omega_j)\phi(\omega_i)\phi(\omega_j) \\
\frac{(N-1)(N-2)}{N^3}\left|\phi(\omega_i)\right|^2 \\
\frac{(N-1)(N-2)}{N^3}\phi(\omega_i + \omega_j)\phi^*(\omega_i)\phi^*(\omega_j) \\
\frac{(N-1)(N-2)}{N^3}\phi^*(\omega_i - \omega_j)\phi(\omega_i)\phi^*(\omega_j) \\
\frac{(N-1)(N-2)}{N^3}\left|\phi(\omega_j)\right|^2
\end{array}
$$

$$
n \neq m \quad n \neq l \quad n \neq k \quad m \neq l \quad m \neq k \quad l \neq k
\qquad \frac{(N-1)(N-2)(N-3)}{N^3}\left|\phi(\omega_i)\right|^2\left|\phi(\omega_j)\right|^2 .
$$

After some simplifications,

$$
\begin{aligned}
\mathrm{Cov}\left[\left|\hat{\phi}_X(\omega_i)\right|^2, \left|\hat{\phi}_X(\omega_j)\right|^2\right] &= \frac{N-1}{N^3}\left(\left|\phi_X(\omega_i+\omega_j)\right|^2 + \left|\phi_X(\omega_i-\omega_j)\right|^2\right. \\
&\left. -2\left|\phi_X(\omega_i)\right|^2\left|\phi_X(\omega_j)\right|^2\right) + \frac{2(N-1)(N-2)}{N^3}\left(\mathrm{Re}(\phi_X^*(\omega_i+\omega_j)\phi_X(\omega_i)\phi_X(\omega_j)) + \right. \\
&\left. \mathrm{Re}(\phi_X(\omega_i-\omega_j)\phi_X^*(\omega_i)\phi_X(\omega_j)) - 2\left|\phi_X(\omega_i)\right|^2\left|\phi_X(\omega_j)\right|^2\right). \quad \text{(A.14)}
\end{aligned}
$$

It follows that the variance of $|\hat{\phi}_X(\omega)|^2$ is

$$
\begin{aligned}
\mathrm{Var}\left[\left|\hat{\phi}_X(\omega)\right|^2\right] &= \frac{N-1}{N^3}\left(\left|\phi_X(2\omega)\right|^2 + 1 - 2\left|\phi_X(\omega)\right|^4\right) + \\
&\frac{2(N-1)(N-2)}{N^3}\left(\mathrm{Re}\left(\phi_X^*(2\omega)\phi_X^2(\omega)\right) + \left|\phi_X(\omega)\right|^2\left(1 - 2\left|\phi_X(\omega)\right|^2\right)\right). \quad \text{(A.15)}
\end{aligned}
$$

If $\phi_X(\omega)$ is real valued, such as when $X$ is a symmetric random variable with zero mean or location, then

$$
\begin{aligned}
\mathrm{Cov}\left[\left|\hat{\phi}_X(\omega_i)\right|^2, \left|\hat{\phi}_X(\omega_j)\right|^2\right] &= \frac{N-1}{N^3}\left(\phi_X^2(\omega_i+\omega_j) + \phi_X^2(\omega_i-\omega_j)\right. \\
&\left. -2\phi_X^2(\omega_i)\phi_X^2(\omega_j)\right) + \frac{2(N-1)(N-2)}{N^3}\phi_X(\omega_i)\phi_X(\omega_j)\left(\phi_X(\omega_i+\omega_j) + \right. \\
&\left. \phi_X(\omega_i-\omega_j) - 2\phi_X(\omega_i)\phi_X(\omega_j)\right) \quad \text{(A.16)}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{Var}\left[\left|\hat{\phi}_X(\omega)\right|^2\right] &= \frac{N-1}{N^3}\left(\phi_X^2(2\omega) + 1 - 2\phi_X^4(\omega)\right) + \\
&\frac{2(N-1)(N-2)}{N^3}\phi_X^2(\omega)\left(\phi_X(2\omega) + 1 - 2\phi_X^2(\omega)\right). \quad \text{(A.17)}
\end{aligned}
$$

# Appendix B

# Optimal Sampling of the Empirical Characteristic Function for Symmetric Alpha Stable Parameter Estimation

*No, no, you're not thinking;*
*you're just being logical.*
*— Niels Bohr*

The points at which the empirical characteristic function is sampled determine the performance of Koutrouvelis' estimator for the parameters of the alpha stable distribution. An optimal solution in the mean square error sense samples the empirical characteristic function at points which minimise the mean square error of the characteristic exponent and scale estimates. For $K$ points in the characteristic function domain, $\omega_k$, $k = 1, \ldots, K$, this is a $K$-dimensional optimisation problem. The optimal set also depends on the characteristic exponent $\alpha$, the scale $c$ and the sample size $N$. Koutrouvelis takes a simplified approach to minimising the mean square error and selects the points through a Monte Carlo study, here the theoretical asymptotic mean square error is used.

## B.1 Koutrouvelis' Estimator

Koutrouvelis' approach was to determine the optimal set of points given $\alpha$, $c$ and $N$. The estimator was structured as follows:

1. Find initial estimates $\hat{\alpha}_0, \hat{c}_0$.

2. Select the optimal set of points based on $\hat{\alpha}_0, \hat{c}_0$.

3. Estimates of the characteristic exponent and scale parameter, $\hat{\alpha}$ and $\hat{c}$, are found using this set of points.

Since $c$ is a scaling constant, the optimal set is found only for $c = 1$. The data is then standardised using $\hat{c}_0$, or uses a standardised set of points $\omega_k/\hat{c}_0$. An interpercentile range estimate, $(Q_X(0.72) - Q_X(0.28))/1.654$, where $Q_X(q)$ is an estimate of the $q^{\text{th}}$ percentile, was suggested for $\hat{c}_0$ [49, 101]. An initial value for $\alpha$ was then obtained using (3.33), after standardisation, with the predetermined set of points $\omega_k = \pi k/25$, $K = 11$.

The difficult $K$-dimensional optimisation problem was avoided by choosing $K$ equi-distant samples separated by $\pi/25$, $\omega_k = \pi k/25$. The reasoning being:

- A lower limit of $\pi/25$ was recommended since $\log(-\log|\phi_X(\omega)|^2) \to -\infty$ and $\log|\omega| \to -\infty$ as $\omega \to 0$, causing numerical problems.

- In general the variance of the ecf increases with $|\omega|$, while for a given $\omega$ and $c$, it increases with $\alpha$. It follows that as $\alpha$ decreases, the ecf can be sampled farther from 0 with comparable variance at $\omega_K$.

- A nonlinear relation between $k$ and $\omega$, i.e. non-equal spacing, was not found to significantly affect the variance of the estimator. From the previous point, the width of the region, $\omega_K - \omega_1$ and hence $K$, was found to be the dominant factor in estimator performance.

A limited Monte Carlo study was performed for a standard $S\alpha S$ distribution using several values of $\alpha$ and $N$, from which the number of samples $K$ that minimised the MSE of the estimates was found. $K(\alpha, N)$ is then found by interpolating a lookup table.

Though an iterative scheme can be envisioned where the parameters and $K$ are updated in an alternating fashion, estimator performance was found to not improve significantly while convergence problems may arise. It is therefore suggested to use only one iteration [101].

## B.2 Selection Based on Asymptotic Theory

A similar approach to the one above will be followed, except that the optimal number of points will be based on the asymptotic MSE of $\hat{\alpha}$ and $\hat{c}$ as found in Chapter 3. Figure B.1 shows the optimal $K$ for estimation of $\alpha$, $K_\alpha$. Figure B.2

shows the same for estimation of $c$, $K_c$. The scenario chosen was for $N = 1000$ observations following a standard $S\alpha S$ distribution, though the curves were not found to differ significantly for $N > 100$. There are several interesting points to note regarding the shape of these curves.

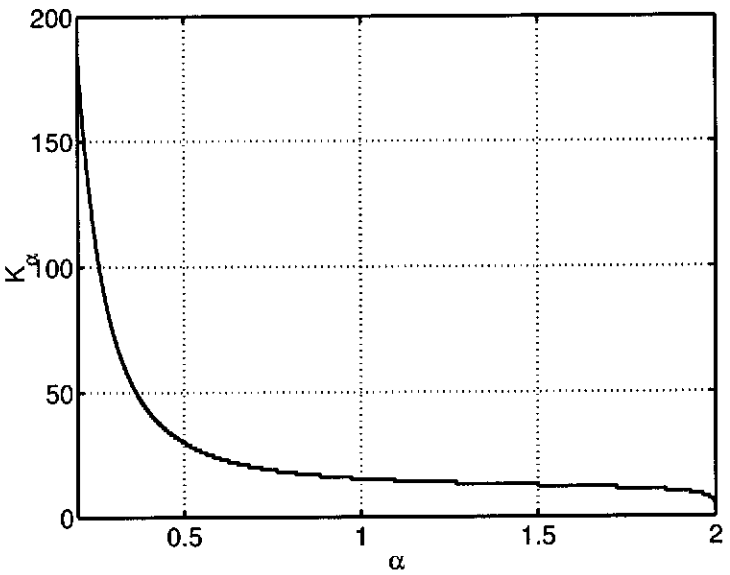


Figure B.1: Optimal $K$ for estimation of $\alpha$ for $N = 1000$ observations with a standard $S\alpha S$ distribution.

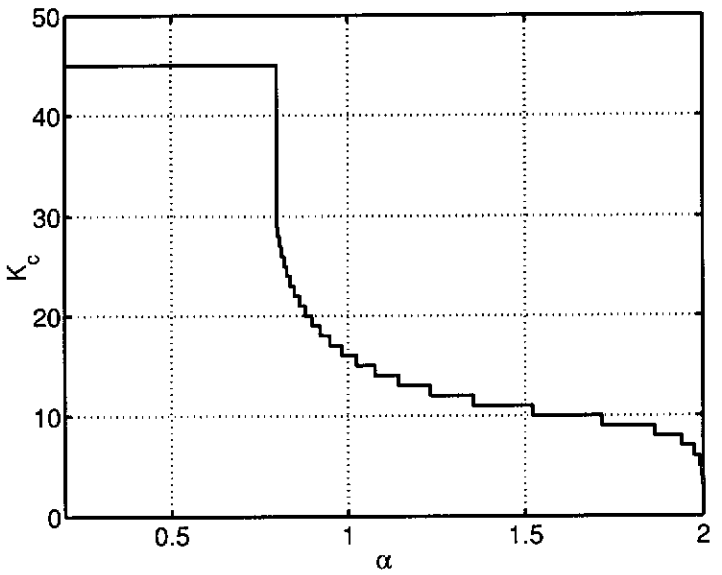


Figure B.2: Optimal $K$ for estimation of $c$ for $N = 1000$ observations with a standard $S\alpha S$ distribution.

First, $K_\alpha$ and $K_c$ differ, suggesting that the ecf should be sampled at different points depending on whether $\alpha$ or $c$ is being estimated. It was found that estimating $\alpha$ using $K_c$ increased its asymptotic MSE by up to 80% above that obtained using $K_\alpha$ for $0.2 < \alpha \leq 2$. Estimating $c$ using $K_\alpha$ gave a corresponding increase of up to 5% only. Since $\alpha$ is the parameter of interest here, $K_\alpha$ will be used exclusively.

Second, Figure B.2 shows that $K_c$ is constant for values of $\alpha$ below approximately 0.8. This effect is caused by the presence of local minima in the function of the MSE of $\hat{c}$ versus $K$ and $\alpha$. For values of $\alpha$ greater than approximately 0.8 two minima exist, the global minimum occurring at lower values of $K$ than the local minimum which occurs near $K = 45$ regardless of $\alpha$. As $\alpha$ decreases from 2 to 0.8, the global minimum occurs at increasing values of $K$, until, at approximately $\alpha = 0.8$, the global and local minima meet. For $\alpha$ below approximately 0.8, a single minimum exists at $K = 45$.

Third, although not clearly visible in Figure B.1, as $\alpha \to 2$, $K_\alpha, K_c \to 2$, while Koutrouvelis uses $K = 9$ for $\alpha > 1.9$. The effect on the MSE of $\hat{\alpha}$ is quite significant, there being a large reduction near $\alpha = 2$, the effect on $c$ is negligible. Near $\alpha = 2$ the modified procedure yields estimates of $\alpha$ with a much smaller variance compared to Koutrouvelis' procedure.

Finally, the stepped nature of these curves is caused by $K$ being a discrete parameter.

## B.2.1 Performance for Finite Samples

### B.2.1.1 Example 1

Figures B.3 and B.4 show the variance of $\hat{\alpha}$ and $\hat{c}$ respectively for $0.2 \leq \alpha \leq 2$ compared to asymptotic theory. The modified estimator was used with $K$ and $c_0$ set a priori to their optimum values, $K = K_\alpha$ and $c_0 = c = 1$. Results were obtained over 10000 Monte Carlo realisations for a sample size of $N = 1000$. Finite sample performance generally agrees with asymptotic theory, regardless of how the $\omega_k$ are selected.

Figures B.5 and B.6 show the CRB of $\alpha$ and $c$ respectively for $0.5 \leq \alpha \leq 2$ compared to the variance obtained using asymptotic theory for the modified estimator with $K_\alpha$, $c = 1$ and a sample size of $N = 1000$. A smaller range for $\alpha$ was chosen due to the difficulty in evaluating the $S\alpha S$ pdf for small values of $\alpha$, all other parameters are the same as in Figures B.3 and B.4. For both $\alpha$ and $c$, it can be seen that the estimator variance as found from asymptotic theory follows

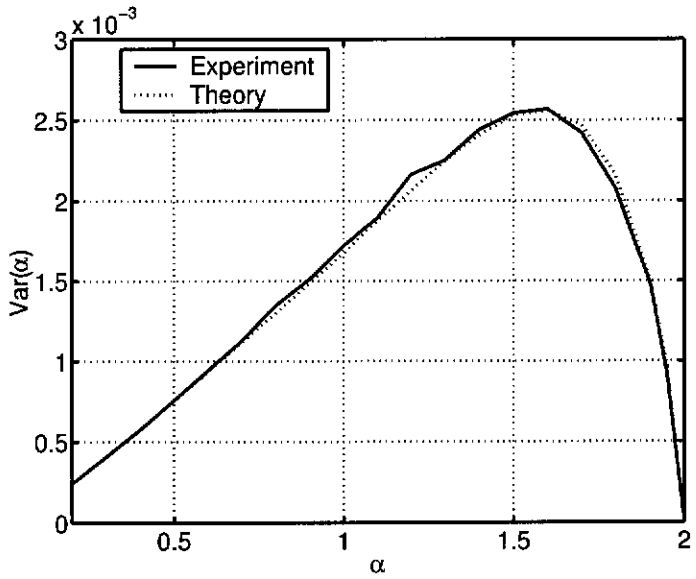


Figure B.3: Var$[\hat{\alpha}]$ using the modified estimator with $K_\alpha$ for $c = 1$. The experimental curve was taken over 10000 independent Monte Carlo realisations for a sample size of $N = 1000$.

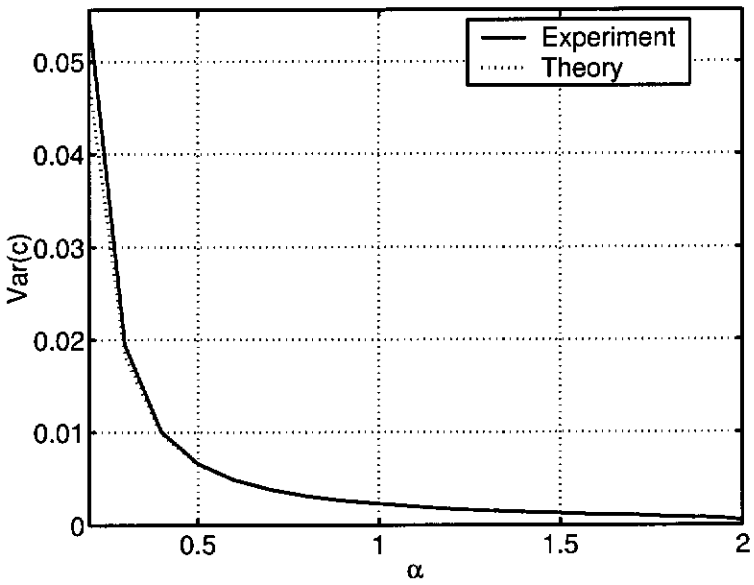


Figure B.4: Var$[\hat{c}]$ using the modified estimator with $K_\alpha$ for $c = 1$. The experimental curve was taken over 10000 independent Monte Carlo realisations for a sample size of $N = 1000$.

the same shape as the CRB.

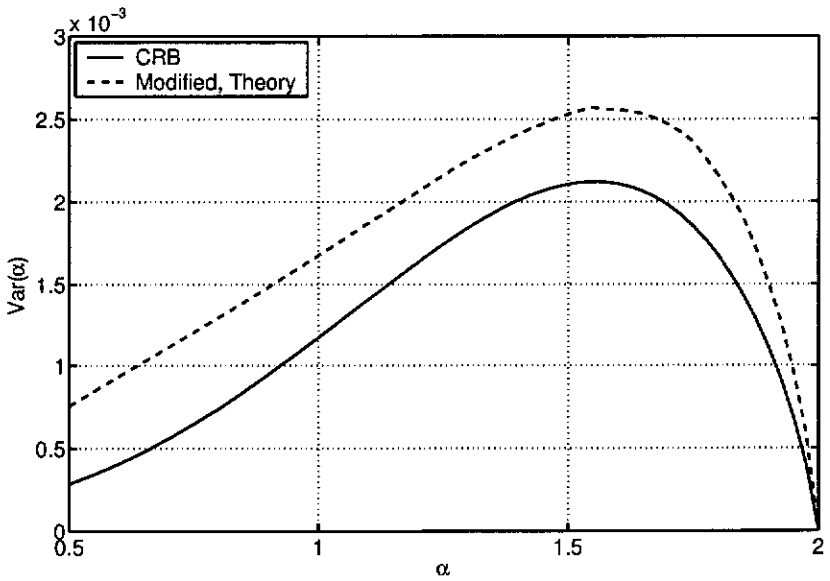


Figure B.5: CRB of $\alpha$ compared to Var$[\hat{\alpha}]$ obtained using asymptotic theory for the modified estimator with $K_\alpha$, $c = 1$ and a sample size of $N = 1000$.

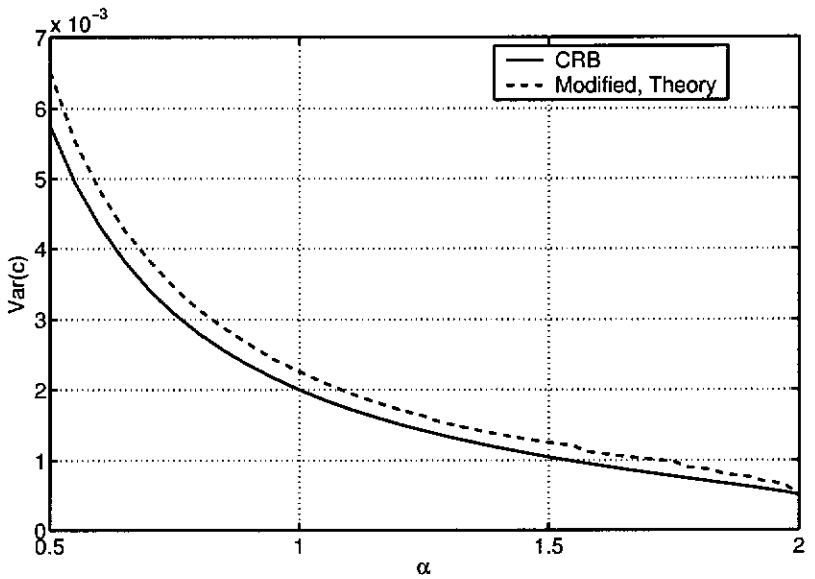


Figure B.6: CRB of $c$ compared to Var$[\hat{c}]$ obtained using asymptotic theory for the modified estimator with $K_\alpha$, $c = 1$ and a sample size of $N = 1000$.

### B.2.1.2 Example 2

Figures B.7 and B.8 show the variance of $\hat{\alpha}$ and $\hat{c}$ respectively for $50 \leq N \leq 2000$ compared to asymptotic theory. Results are shown for $\alpha = 1.5$, all other conditions being the same as in Example 1. There is close agreement with asymptotic

theory down to sample sizes of $N = 100$, though bias does become non-negligible for small sample sizes.

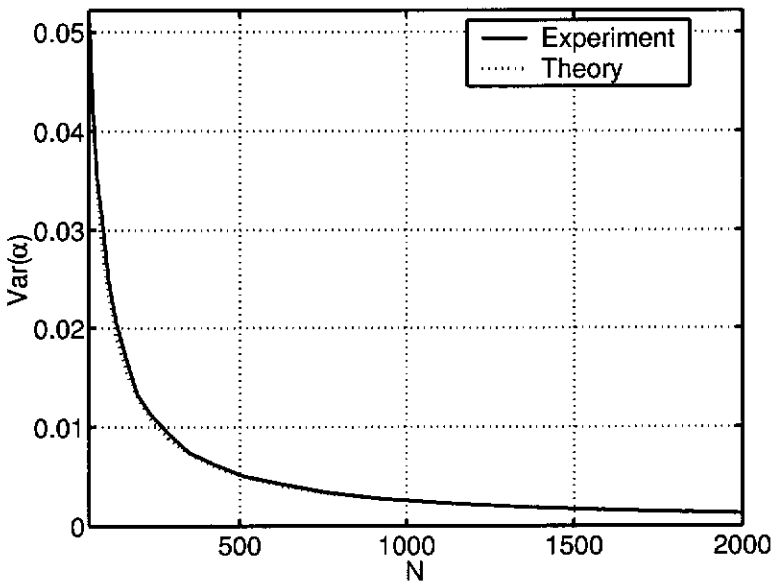


Figure B.7: Var$[\hat{\alpha}]$ using the modified estimator with $K_\alpha$ for $c = 1$. The experimental curve was taken over 10000 independent Monte Carlo realisations for $\alpha = 1.5$.

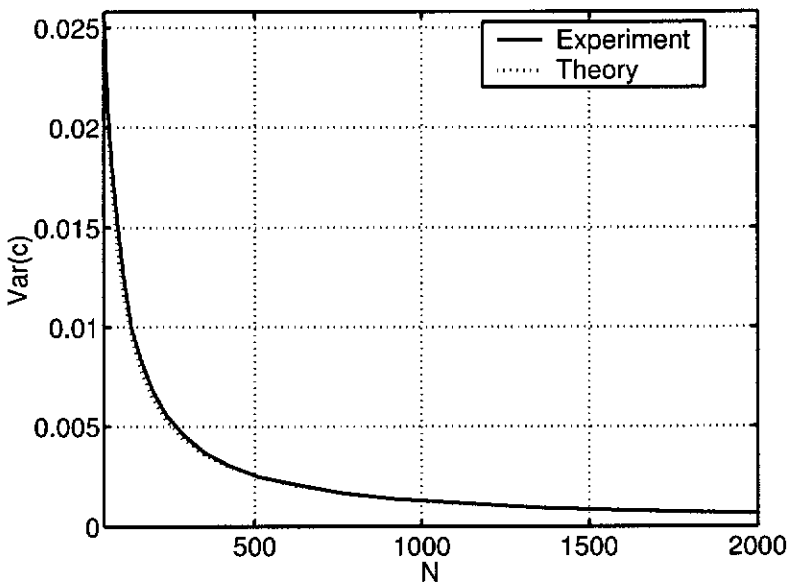


Figure B.8: Var$[\hat{c}]$ using the modified estimator with $K_\alpha$ for $c = 1$. The experimental curve was taken over 10000 independent Monte Carlo realisations for $\alpha = 1.5$.

It was found that if $K_k$, Koutrouvelis' choice for $K$, was used with the modified estimator, similar results to Koutrouvelis' estimator are obtained. This suggests

that selection of the $\omega_k$ is more critical than whether $\mathrm{Re}(\phi_X(\omega))^2$ or $|\phi_X(\omega)|^2$ is used. Likewise, asymptotic theory for the modified estimator compares well with the experimental performance of Koutrouvelis' estimator.

### B.2.1.3 Example 3

To gauge the improvement in performance, Figures B.9 and B.10 show the efficiency of Koutrouvelis' estimator with respect to the modified estimator of $\hat{\alpha}$ and $\hat{c}$ respectively. The scenario is the same as in Example 1.

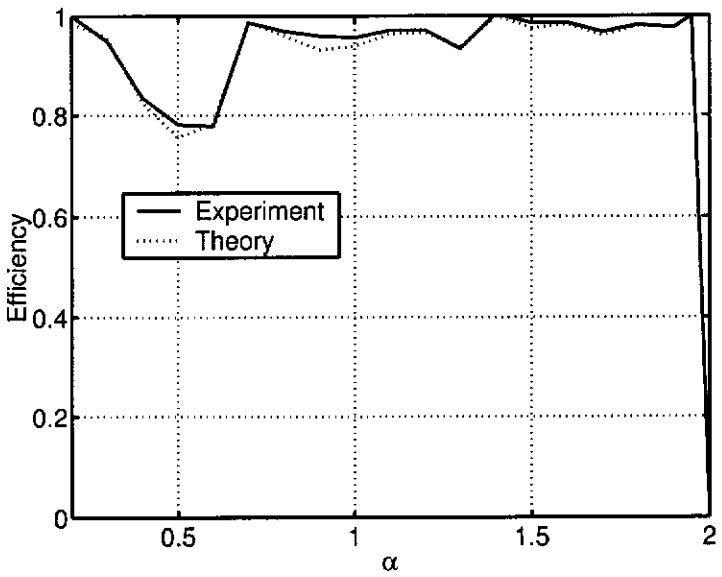


Figure B.9: Efficiency of Koutrouvelis' estimator using $K_k$ with respect to the modified estimator using $K_\alpha$ for $\hat{\alpha}$. The experimental curve was taken over 10000 independent Monte Carlo realisations for a sample size of $N = 1000$.

The asymptotic theory curve is based on the modified estimator only and therefore cannot be interpreted as the asymptotic theoretical efficiency for this scenario, even so, it gives a very close approximation to the experimental result. There is a uniform improvement in the estimation of $\alpha$, particularly near $\alpha = 2$. As the modified estimator used $K_\alpha$ to estimate $c$, there is not a uniform improvement over Koutrouvelis' estimator, though over $0.2 \leq \alpha \leq 2$ the performance of the two are comparable.

The computational complexity of Koutrouvelis' estimator and the modified one is approximately proportional to the value of $K$ used in both. For instance, when $\alpha = 1.5$ and $N = 1000$, Koutrouvelis' estimator uses $K = 11$, while the modified one uses $K = 13$. Hence, the modified estimator may be up to 20% less

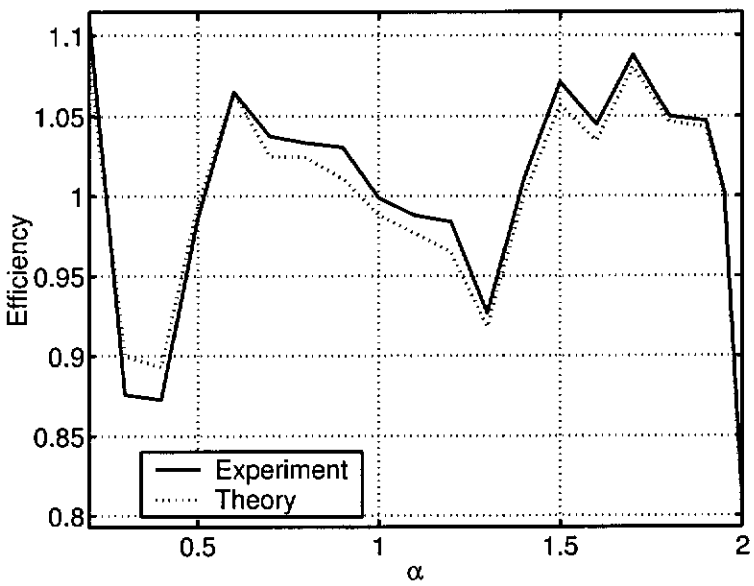


Figure B.10: Efficiency of Koutrouvelis' estimator using $K_k$ with respect to the modified estimator using $K_\alpha$ for $\hat{c}$. The experimental curve was taken over 10000 independent Monte Carlo realisations for a sample size of $N = 1000$.

computationally efficient. Likewise, when $\alpha = 0.5$ the modified estimator may be up to 55% more computationally efficient.

#### B.2.1.4 Example 4

Figures B.11 and B.12 show the distributions of $\hat{\alpha}$ and $\hat{c}$ respectively using the modified estimator with $K_\alpha$. Results are shown over 10000 independent Monte Carlo realisations for a sample size of $N = 1000$, the true parameters were $\alpha = 1.5, c = 1$.

The finite sample distributions of $\hat{\alpha}$ and $\hat{c}$ agree well with asymptotic theory down to sample size of approximately 500. Apart from some skewness in $\hat{\alpha}$ near $\alpha = 2$ the agreement appears uniform over $\alpha$.

#### B.2.1.5 Example 5

Finally, both estimators are compared when $c_0$ and $K$ are not provided, but are determined from initial estimates. Figure B.13 shows the variance of $\hat{\alpha}$ for $0.2 \leq \alpha \leq 2$ and $c = 1$. Results were obtained over 10000 independent Monte Carlo realisations for a sample size of $N = 1000$. At worst, the modified estimator does as well as Koutrouvelis' estimator.

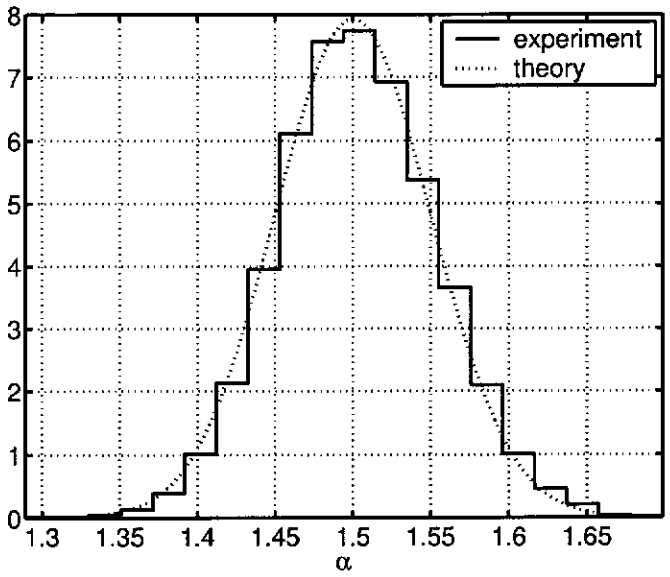


Figure B.11: Distribution of $\hat{\alpha}$ for the modified estimator using $K_\alpha$, $\alpha = 1.5, c = 1$. The experimental curve was taken over 10000 independent Monte Carlo realisations for a sample size of $N = 1000$.

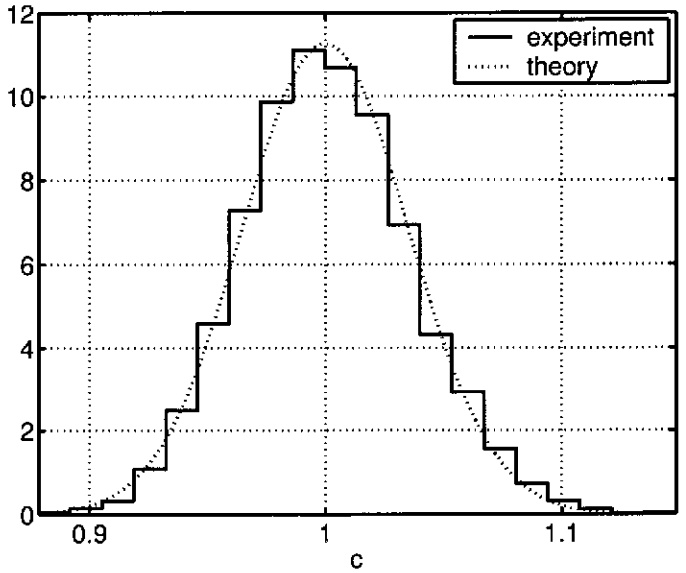


Figure B.12: Distribution of $\hat{c}$ for the modified estimator using $K_\alpha$, $\alpha = 1.5, c = 1$. The experimental curve was taken over 10000 independent Monte Carlo realisations for a sample size of $N = 1000$.

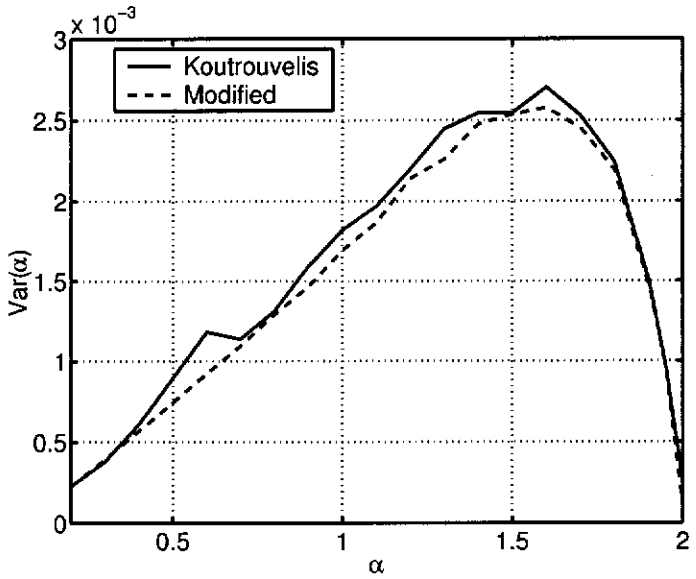


Figure B.13: Var$[\hat{\alpha}]$ for both estimators, $c = 1$. Results are over 10000 independent Monte Carlo realisations for a sample size of $N = 1000$.

# Appendix C

# Taylor Series Expansions of Random Variables

*Computer, compute to the last digit*
*the value of pi.*
*— Spock (Wolf in the Fold)*

Let $X$ and $Y$ be random variables with means $\mu_X$ and $\mu_Y$ respectively for which all moments exist. Also, let $g(x)$ and $h(y)$ be continuously differentiable functions with nonzero first order derivatives at $\mu_X$ and $\mu_Y$ respectively. Expand $g(X)$ and $h(Y)$ in a Taylor series up to first order about their respective means [159],

$$g(X) \approx g(\mu_X) + g'(\mu_X)(X - \mu_X), \tag{C.1}$$
$$h(Y) \approx h(\mu_Y) + h'(\mu_Y)(Y - \mu_Y). \tag{C.2}$$

A first order approximation for the expectations of $g(X)$ and $h(Y)$ are

$$\mathsf{E}[g(X)] \approx g(\mu_X), \tag{C.3}$$
$$\mathsf{E}[h(Y)] \approx h(\mu_Y). \tag{C.4}$$

A first order approximation for the covariance of $g(X)$ and $h(Y)$ is

$$\begin{aligned} \mathsf{Cov}[g(X), h(Y)] &= \mathsf{E}[(g(X) - \mathsf{E}[g(X)])(h(Y) - \mathsf{E}[h(Y)])] \\ &\approx \mathsf{E}[g'(\mu_X)(X - \mu_X)h'(\mu_Y)(Y - \mu_Y)] \\ &= g'(\mu_X)h'(\mu_Y)\mathsf{Cov}[X, Y]\,. \end{aligned} \tag{C.5}$$

# Appendix D

# The Bootstrap

*Insanity: doing the same thing over and over again and expecting different results.*
*— Albert Einstein*

A very general statistical problem is that of finding the sampling distribution of a statistic $T(\boldsymbol{X})$ formed from a sample $\boldsymbol{X} = (X_1, \ldots, X_N)^\mathsf{T}$. The sample is drawn from a distribution $F_X$ which may be multivariate. Knowledge of the sampling distribution allows one to carry out inference, typically involving hypothesis tests or confidence intervals.

Sampling distributions can be exceedingly difficult to find, which necessitates the use of approximations, usually of an asymptotic nature, to obtain workable expressions. The accuracy of inference based on such approximations in the finite and especially small sample cases is understandably reduced. A lack of knowledge concerning $F_X$ merely compounds the problem further.

The bootstrap solves this problem by replacing complex theoretical analysis with computational power. In an ideal world samples would be continually drawn from $F_X$ until the distribution of $T$ is known to sufficient accuracy. Repeating the experiment in such a fashion is untenable in practice. The bootstrap paradigm is to instead continue drawing from the empirical distribution $\hat{F}_X$, creating bootstrap samples $\boldsymbol{X}^*$. The bootstrap test statistics, $T^*(\boldsymbol{X}^*)$, obtained by recalculation of the test statistic for each of the bootstrap samples, then estimate the sampling distribution of $T(\boldsymbol{X})$. More succinctly, the 'real' and 'bootstrap' worlds can be compared as follows,

$$\begin{array}{lcccccc} \text{Real World} & : & F_X & \rightarrow & \boldsymbol{X} & \rightarrow & T(\boldsymbol{X}) \\ \text{Bootstrap World} & : & \hat{F}_X & \rightarrow & \boldsymbol{X}^* & \rightarrow & T^*(\boldsymbol{X}^*). \end{array}$$

A concise overview of the bootstrap and its application to real problems can be

found in [148, 198], for a more thorough treatment see [43].

The bootstrap is usually applicable whenever $\hat{F}_X$ is a consistent estimate for $F_X$. The general procedure of drawing from $\hat{F}_X$ to estimate a sampling distribution is called resampling. There are many types of resampling procedure depending on how $\hat{F}_X$ is obtained and how resamples are drawn.

## D.1 Bootstrap Resampling

Bootstrap resampling takes for $\hat{F}_X$ the edf, from which $N$ samples are drawn. Each observation has an equal probability of being drawn and may be drawn more than once in a single bootstrap resample. $B$ bootstrap resamples of size $N$ are drawn in this way, where $B$ is chosen sufficiently large to ensure the bootstrap resamples adequately model the sampling distribution.

Instead of the edf, a kernel edf estimator may be used, which leads to the smoothed bootstrap. Using a smoothed bootstrap means drawing from a continuous distribution so it is possible to obtain values not present in the original sample. The improvement offered by smoothing is usually negligible given the added complexity and the smoothed bootstrap is not considered any further here.

If $F_X$ is known to within a family parameterised by $\boldsymbol{\theta}$, the parametric bootstrap is used. Given that consistent estimates, $\hat{\boldsymbol{\theta}}$, are available, the parametric bootstrap draws $N$ samples from $F_X(\hat{\boldsymbol{\theta}})$. In this sense, the former two resampling procedures comprise the nonparametric bootstrap.

### D.1.1 Subsampling

The bootstrap may fail in certain situations, for example when $T$ is the sample mean and $F_X$ is within the domain of attraction of an $\alpha S$ distribution. The solution is to draw from the edf only $M < N$ times, creating $B$ resamples.

The jackknife is a special case, occurring when $M = N - 1$ and resamples are drawn without replacement giving subsamples of the original sample. Drawing without replacement means that once a sample has been drawn, it cannot be drawn again for the same subsample. This gives $B = N$ unique subsamples of size $M$. This idea is generalised in the delete-$d$ jackknife where $d = N - M$, $M < N$, giving $B = N!/(M!(N - M)!)$ possible unique subsamples.

For $M << N$ the number of unique subsamples becomes unwieldy and a subset $B < N!/(M!(N - M)!)$ is chosen instead. Additionally, it makes little difference whether the draws are made with or without replacement, so that bootstrap

resampling with a resample size $M << N$ and subsampling are essentially the same. An extensive treatment of subsampling theory appears in [149].

## D.2 Hypothesis Testing

Consider a test for the null hypothesis $\mathsf{H} : T(F_X) = T_H$ against the alternative hypothesis $\mathsf{K} : T(F_X) \neq T_H$, where $T(F_X)$ is the true value of some descriptive statistic as obtained from $F_X$. Assuming a symmetric distribution for the sample statistic $T(X)$, the null hypothesis is rejected if for the test statistic $|\hat{T}| = |T(X) - T_H|$, $|\hat{T}| \geq F_{\hat{T}}^{-1}(1-\zeta/2)$, where $F_{\hat{T}}$ is the distribution of $\hat{T}$ under the null hypothesis and $\zeta$ is the probability of false alarm.

$F_{\hat{T}}$ can be estimated using the bootstrap by resampling the test statistic under the null hypothesis. Denote the empirical distribution of the bootstrap test statistics $\hat{T}^* = T^*(X) - T(X)$, by $\hat{F}_{\hat{T}}$, $\hat{F}_{\hat{T}}$ is then an approximation to $F_{\hat{T}}$. The test is carried out by inverting $\hat{F}_{\hat{T}}$, so that the null hypothesis is rejected if $|\hat{T}| \geq \hat{F}_{\hat{T}}^{-1}(1-\zeta/2)$. Alternatively, an estimate for the p-value is $\frac{1}{B}\sum_{b=1}^{B} \mathrm{I}\Big(|\hat{T}^*| > |\hat{T}|\Big)$, the null hypothesis being rejected if $\mathcal{P} \leq \zeta$. $B \geq 10/\zeta$ is generally a sufficient number of resamples for accurate estimates, this placing at least 10 of the $\hat{T}^*$ in the critical region, though it has been suggested that only $B \geq 1/\zeta$ is necessary. Further guidelines can be found in [68].

### D.2.1 Pivotality

The accuracy of the bootstrap can be improved by using statistics which are asymptotically pivotal. Asymptotic pivotality implies the asymptotic distribution of the statistic is independent of any unknown parameters. Accuracy is improved in the sense that the error between the attained and set levels is $O(N^{-1})$, as compared to $O(N^{-1/2})$ without pivotal statistics or using asymptotic methods [66, 67].

Asymptotic pivotality may be achieved through variance stabilising transforms or studentisation. Studentisation scales the statistic by the square root of its variance or an available estimate $\hat{\sigma}$, giving a test statistic $\hat{T} = (T(X) - T_H)/\hat{\sigma}$ and bootstrap statistics $\hat{T}^* = (T^*(X) - T(X))/\hat{\sigma}^*$. Note that for each of the bootstrap statistics, $\hat{\sigma}^*$ must be found anew. In the absence of an estimate for the variance another round of resampling is required. The sample variance of the bootstrap statistics obtained in the second round is then used as an estimate for the variance and is known as the iterated or nested bootstrap. More details can

be found in [43, 148, 198].

# Appendix E

# Multiple Hypothesis Tests

*I think I speak for everyone*
*when I say 'huh?'*
*— Buffy*

Multiple hypothesis tests (MHT) are employed when several hypotheses are being tested and a global level of significance must be maintained [72]. Given a set of hypotheses $\mathsf{H} = \{\mathsf{H}_i : i = 1, \ldots, n\}$ consider a test for the global null hypothesis $\mathsf{H}_0$ formed by the intersection of every element in $\mathsf{H}$,

$$\mathsf{H}_0 = \bigcap\{\mathsf{H}_i\}. \tag{E.1}$$

$\mathsf{H}_0$ is rejected if at least one of the $\{\mathsf{H}_i\}$ are rejected. The probability of Type I error, that of incorrectly rejecting the global null hypothesis, should be maintained at or below the global level of significance $\zeta$.

When testing multiple hypotheses there is usually interest in testing not only for the global null hypothesis, but also for non-empty subsets $\mathsf{H}^s = \{\mathsf{H}^s_I : I \subseteq \{1, \ldots, n\}\}$ of $\mathsf{H}$ where the null hypothesis is formed by the intersection of every element in $\mathsf{H}^s$,

$$\mathsf{H}^s_0 = \bigcap\{\mathsf{H}^s_I\}. \tag{E.2}$$

In this case the Type I error is generalised to be the probability of incorrectly rejecting $\mathsf{H}^s_0$ irrespective of which $\{\mathsf{H}^s_I\}$ are true.

A MHT procedure is said to weakly control the family wise error (FWE) rate if

$$\mathsf{Pr}[\text{Reject } \mathsf{H}_0 \,|\, \mathsf{H}_0] = \mathsf{Pr}[\text{Reject any } \mathsf{H}_i \,|\, \mathsf{H}_0] \leq \zeta, \tag{E.3}$$

while strong control implies

$$\mathsf{Pr}[\text{Reject } \mathsf{H}^s_0 \,|\, \mathsf{H}^s_0] = \mathsf{Pr}[\text{Reject any element of } \mathsf{H}^s_I \,|\, \mathsf{H}^s_0] \leq \zeta. \tag{E.4}$$

Strong control is the more powerful condition and tests which strongly control the FWE rate are favoured over those which do not.

The most basic MHT procedure follows from Bonferroni's inequality on the upper and lower bounds for the probability of a union of events $E_i : i = 1, \ldots, n$,

$$\sum_{i=1}^{n} \mathsf{Pr}[E_i] - \sum\sum_{i<j} \mathsf{Pr}[E_i \cap E_j] \leq \mathsf{Pr}[\cup_{i=1}^{n} E_i] \leq \sum_{i=1}^{n} \mathsf{Pr}[E_i]\,. \tag{E.5}$$

Letting $E_i$ be the event that $\mathsf{H}_i$ is rejected given $\mathsf{H}_0$ is true, the centre term in the above equation is then the FWE rate under the global null hypothesis. Define $\mathcal{P}_i$ as the p-value corresponding to $\mathsf{H}_i$, then reject $\mathsf{H}_i$ if $\mathcal{P}_i \leq \zeta_i$. From the Bonferroni inequality [193],

$$\begin{aligned} \mathsf{Pr}[\text{Reject any } \mathsf{H}_i \,|\, \mathsf{H}_0] &= \mathsf{Pr}[\cup_{i=1}^{n} \text{Reject } \mathsf{H}_i \,|\, \mathsf{H}_0] \\ &= \mathsf{Pr}[\cup_{i=1}^{n} \mathcal{P}_i \leq \zeta_i \,|\, \mathsf{H}_0] \\ &\leq \sum_{i=1}^{n} \mathsf{Pr}[\mathcal{P}_i \leq \zeta_i \,|\, \mathsf{H}_0]\,. \end{aligned} \tag{E.6}$$

Given that the p-values are $\mathrm{U}(0,1)$ under $\mathsf{H}_0$ the last line becomes $\sum_{i=1}^{n} \zeta_i$ so that the FWE will be weakly controlled if $\sum_{i=1}^{n} \zeta_i \leq \zeta$. The simplest choice is to set $\zeta_i = \zeta/n$ so that all the $\mathsf{H}_i$ are tested at the same level, this is the original Bonferroni MHT procedure. Clearly, the Bonferroni MHT procedure strongly controls the FWE rate. If it is further assumed that the p-values are independent, the exact FWE rate under $\mathsf{H}_0$ can be found

$$\begin{aligned} \mathsf{Pr}[\cup_{i=1}^{n} \mathcal{P}_i \leq \zeta_i \,|\, \mathsf{H}_0] &= 1 - \mathsf{Pr}[\cap_{i=1}^{n} \mathcal{P}_i > \zeta_i \,|\, \mathsf{H}_0] \\ &= 1 - \Pi_{i=1}^{n} \mathsf{Pr}[\mathcal{P}_i > \zeta_i \,|\, \mathsf{H}_0] \\ &= 1 - \Pi_{i=1}^{n} (1 - \zeta_i). \end{aligned} \tag{E.7}$$

Setting this equal to $\zeta$ and allowing the $\zeta_i$ to be the same gives $\zeta_i = 1 - (1-\zeta)^{1/n}$. Since $\zeta_i = \zeta/n + O((\zeta/n)^2)$ and $\zeta << 1$, the change in FWE rate and power by testing the $\mathsf{H}_i$ at this level as opposed to $\zeta/n$ is usually negligible.

Let the ordered p-values be $\mathcal{P}_{(1)} \leq \cdots \leq \mathcal{P}_{(n)}$, with corresponding hypotheses, $\mathsf{H}_{(1)}, \ldots, \mathsf{H}_{(n)}$. When testing for $\mathsf{H}_0$ using Bonferroni's MHT procedure it is sufficient only to compare the smallest p-value, $\mathcal{P}_{(1)}$, with $\zeta/n$. $\mathcal{P}_{(1)}$ always contains the strongest evidence for rejecting $\mathsf{H}_0$ and since the $\zeta_i$ are equal, no other p-values need to be compared should $\mathcal{P}_{(1)} \leq \zeta/n$. Alternatively if $\mathcal{P}_{(1)} > \zeta/n$ $\mathsf{H}_0$ is accepted since this implies all other p-values are greater than $\zeta/n$. Bonferroni's MHT procedure is summarised in Figure E.1.

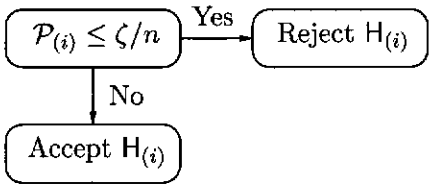


Figure E.1: Bonferroni's MHT procedure.

Dependence between the hypotheses has a more serious effect on Bonferroni's MHT procedure, leading to a conservative test with reduced power. MHT procedures exist which improve on Bonferroni's by moving the FWE closer to the global level for dependent hypotheses.

Two tests for the global null hypothesis are the simultaneous procedures of Hommel and Simes. Hommel's MHT procedure rejects $\mathsf{H}_0$ if any $\mathcal{P}_{(i)} \leq i\zeta/(nC_n)$, where $C_n = \sum_{i=1}^{n} 1/i$, and weakly controls the level [78]. Simes' MHT procedure is the same but with $C_n = 1$, making it less conservative. In general this procedure does not weakly control the level, but does so for a wide range of test statistics or for independent hypotheses [164]. Hommel's MHT procedure exhibits weak control and has been extended for tests on individual hypotheses in which case the control is strong [79]. The procedures are shown in Figures E.2 and E.3 respectively.

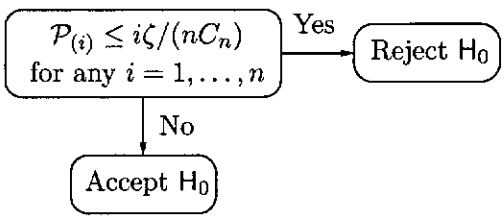


Figure E.2: Hommel's MHT procedure.

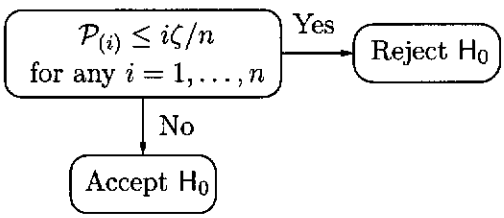


Figure E.3: Simes' MHT procedure.

Holm's sequentially rejective Bonferroni (SRB) MHT procedure is a stepwise procedure which rejects $\mathsf{H}_i$ if $\mathcal{P}_j \leq \zeta/(n-j+1)$ for $j = 1, \ldots, i$, as shown

in Figure E.4. Though less conservative than Bonferroni's MHT procedure, it strongly controls the level. A slightly less conservative test is obtained by replacing $\zeta_{(i)} = \zeta/i$ with $1-(1-\zeta)^{1/i}$. The generalised SRB MHT procedure replaces $\mathcal{P}_{(i)}$ by $\mathcal{P}_{(i)}/c_{(i)}$ and $\zeta_i$ by $\zeta/\sum_{j=i}^{n} c_{(j)}$, it also strong controls the level. If a particular hypothesis is considered more important it is given a larger weight $c_i$, this increases the power of that test at the cost of the others [76].

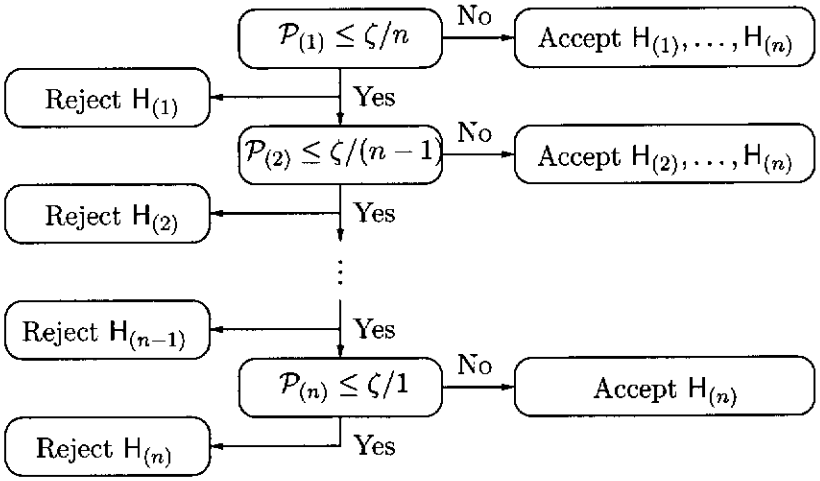


Figure E.4: Holm's MHT procedure.

A stepwise MHT procedure which is less conservative than Holm's SRB MHT procedure was proposed by Hochberg, it rejects $\mathsf{H}_j$ for $j = 1, \ldots, i$ if $\mathcal{P}_i \leq \zeta/(n-i+1)$. Based on an extended version of Simes' MHT procedure, it strongly controls the level [71] and is shown in Figure E.5.

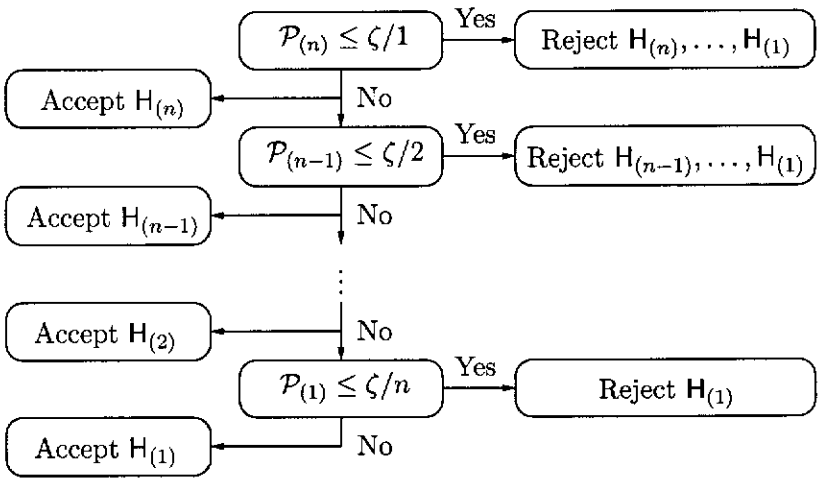


Figure E.5: Hochberg's MHT procedure.

Improvements in power over Holm's SRB MHT procedure can be obtained by taking into account any logical implications between the hypotheses. Shaffer used Holm's generalised MHT procedure, modifying the weights according to the maximum number of hypothesis that may still be true at each stage. Strong control is implicit as it is a version of Holm's generalised MHT procedure [160]. Further improvements over Shaffer's MHT procedure subject to positive orthant dependence of the test statistics appears in [75].

# Bibliography

*The secret to creativity is knowing*
*how to hide your sources.*
*— Albert Einstein*


[1] B. Aazhang and H. Poor. Performance of DS/SSMA communications in impulsive channels-part I: Linear correlation receivers. *IEEE Transactions on Communications*, COM-35(11):1179–88, November 1987.

[2] B. Aazhang and H. Poor. Performance of DS/SSMA communications in impulsive channels-part II: Hard-limiting correlation receivers. *IEEE Transactions on Communications*, 36(1):88–97, January 1988.

[3] B. Aazhang and H. Poor. An analysis of nonlinear direct-sequence correlators. *IEEE Transactions on Communications*, 37(7):723–31, July 1989.

[4] R. Adler, R. Feldman, and M. Taqqu, editors. *A Practical Guide to Heavy Tails: Statistical Techniques and Applications.* Birkhäuser, 1998.

[5] S. Ambike and D. Hatzinakos. Three receiver structures and their performance analysis for binary signaling in a mixture of Gaussian and $\alpha$ stable impulsive noises. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 4, pages 173–6, Adelaide, Australia, April 1994.

[6] S. Ambike, J. Ilow, and D. Hatzinakos. Detection for binary transmission in a mixture of Gaussian noise and impulsive noise modeled as an alpha-stable process. *IEEE Signal Processing Letters*, 1(3):55–57, March 1994.

[7] T. Anderson. Asymptotic theory for principal component analysis. *The Annals of Statistics*, 34:122–48, 1963.

[8] T. Anderson. *An Introduction to Multivariate Statistical Analysis.* John Wiley, 2$^{nd}$ edition, 1984.

[9] D. Andrews and C. Mallows. Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Series B*, B-36:99–102, 1974.

[10] R. Arad. Parameter estimation for symmetric stable distribution. *International Economic Review*, 21(1):209–20, February 1980.

[11] M. Arnold, D. Iskander, and A. Zoubir. Testing Gaussianity with the characteristic function. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 3, pages 2012–2015, Detroit, USA, May 1995.

[12] A. Banerjee, P. Burlina, and R. Chellappa. Adaptive target detection in foliage-penetrating SAR images using alpha-stable models. *IEEE Transactions on Image Processing*, 8(12):1823–31, December 1999.

[13] S. Batalama, M. Medley, and D. Pados. Robust adaptive recovery of spread-spectrum signals with short data records. In *Proceedings of the Defence Applications of Signal Processing (DASP) Workshop*, pages 19–24, LaSalle, USA, August 1999.

[14] S. Batalama, M. Medley, and D. Pados. Robust adaptive recovery of spread-spectrum signals with short data records. *IEEE Transactions on Communications*, 48(10):1725–31, October 2000.

[15] S. Batalama, M. Medley, and I. Psaromiligkos. Adaptive robust spread-spectrum receivers. *IEEE Transactions on Communications*, 47(6):905–17, June 1999.

[16] S. Bates and S. McLaughlin. The estimation of stable distribution parameters. In *Proceedings of the IEEE Signal Processing Workshop on Higher Order Statistics*, pages 390–4, Banff, Canada, July 1997.

[17] R. Beran and M. Srivastava. Bootstrap tests and confidence regions for functions of a covariance matrix. *The Annals of Statistics*, 13(1):95–115, 1985.

[18] R. Beran and M. Srivastava. Correction: Bootstrap tests and confidence regions for functions of a covariance matrix. *The Annals of Statistics*, 15(1):470–471, 1987.

[19] L. Berry. Understanding Middleton's canonical formula for class A noise. *IEEE Transactions on Electromagnetic Compatibility*, EMC-23(4):337–44, November 1981.

[20] K. Blackard and T. Rappaport. Measurements and models of radio frequency impulsive noise for indoor wireless communications. *IEEE Journal on Selected Areas in Communications*, 11(7):991–1001, September 1993.

[21] T. Blankenship, D. Krizman, and T. Rappaport. Measurements and simulation of radio frequency impulsive noise in hospitals and clinics. In *Proceedings of the IEEE 47$^{th}$ Vehicular Technology Conference. Technology in Motion.*, volume 3, pages 1942–6, Phoenix, USA, May 1997.

[22] R. Blum, R. Kozick, and B. Sadler. An adaptive spatial diversity receiver for non-gaussian interference and noise. *IEEE Transactions on Signal Processing*, 47(8):2100–11, August 1999.

[23] J. Bodenschatz and C. Nikias. Maximum-likelihood symmetric $\alpha$-stable parameter estimation. *IEEE Transactions on Signal Processing*, 47(5):1382–4, May 1999.

[24] R. Brcich and A. Zoubir. Resampling based techniques for source detection in array processing. In *Proceedings of the 11$^{th}$ IEEE Workshop on Statistical Signal Processing*, pages 26–9, Singapore, August 2001.

[25] L. Breiman. *Probability.* Society for Industrial and Applied Mathematics, 1992.

[26] P. Brockwell and B. Brown. High-efficiency estimation for the positive stable laws. *Journal of the American Statistical Association*, 76(375):626–31, September 1981.

[27] B. Brorsen and S. Yang. Maximum likelihood estimates of symmetric stable distribution parameters. *Communications in Statistics: Simulation and Computation*, 19(4):1459–64, 1990.

[28] C. Brown. *Goodness-of-Fit and Detection Problems in Impulsive Interference.* PhD thesis, Australian Telecommunications Research Institute and School of Electrical and Computer Engineering, Curtin University of Technology, 2000.

[29] C. Brown and S. Saliu. Testing of alpha-stable distributions with the characteristic function. In *Proceedings of the IEEE Signal Processing Workshop on Higher Order Statistics*, pages 224–7, Caesarea, Israel, June 1999.

[30] C. Brown and A. Zoubir. On the estimation of the parameters of $\alpha$-stable distributions using linear regression in the characteristic function domain. In *Proceedings of the 9$^{th}$ IEEE Workshop on Statistical Signal and Array Processing*, pages 423–6, Portland, USA, September 1998.

[31] C. Brown and A. Zoubir. Locally suboptimal and rank-based known signal detection in correlated alpha-stable interference. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 1, pages 53–6, Istanbul, Turkey, June 2000.

[32] D. Buckle. Bayesian inference for stable distributions. *Journal of the American Statistical Association*, 90(430):605–13, June 1995.

[33] S. Cambanis and G. Miller. Linear problems in $p^{\text{th}}$ order and stable processes. *SIAM Journal of Applied Mathematics*, 41(1):43–69, August 1981.

[34] J. Chambers, C. Mallows, and B. Stuck. A method for simulating stable random variables. *Journal of the American Statistical Association*, 71(354):340–4, June 1976.

[35] S. Chandrasekhar. Stochastic problems in physics and astronomy. *Reviews of Modern Physics*, 15(1):1–89, January 1943.

[36] E. Conte, M. Di Bisceglie, M. Longo, and M. Lops. Canonical detection in spherically invariant noise. *IEEE Transactions on Communications*, 43(2/3/4):347–353, February/March/April 1995.

[37] S. Csörgő. Limit behaviour of the empirical characteristic function. *The Annals of Probability*, 9(1):130–44, 1981.

[38] R. D'Agostino and M. Stephens, editors. *Goodness-of-Fit Techniques*. Marcel Dekker, 1986.

[39] X. Dan, Z. Huimin, and Y. Yang. Fast and efficient estimation of the symmetric alpha-stable impulsive signal or noise parameters. In *Proceedings of the 3$^{rd}$ International Conference on Signal Processing*, volume 1, pages 213–6, Beijing, China, October 1996.

[40] C. Dance and E. Kuruoglu. Estimation of the parameters of skewed $\alpha$-stable distributions. In *Proceedings of Heavy Tails '99, Applications of Heavy Tailed Distributions in Economics, Engineering and Statistics*, pages TAILS–12, Washington, DC, June 1999.

[41] L. Devroye. *A Course In Density Estimation.* Birkhäuser, 1987.

[42] W. DuMouchel. On the asymptotic normality of the maximum-likelihood estimate when sampling from a stable distribution. *The Annals of Statistics*, 1(5):948–57, 1973.

[43] B. Efron and R. Tibshirani. *An Introduction to the Bootstrap.* Chapman and Hall, 1993.

[44] T. Epps. Characteristic functions and their empirical counterparts: Geometrical interpretations and applications to statistical inference. *The American Statistician*, 47(1):33–8, February 1992.

[45] R. Kapoor et. al. UWB radar detection of targets in foliage using alpha-stable clutter models. *IEEE Transactions on Aerospace and Electronic Systems*, 35(3):819–34, July 1999.

[46] P. Rousseeuw F. Hampel, E. Ronchetti and W. Stahel. *Robust Statistics, The Approach Based on Influence Functions.* John Wiley, 1986.

[47] E. Fama. Mandelbrot and the stable Paretian hypothesis. *Journal of Business*, 36(4):420–9, October 1963.

[48] E. Fama. The behaviour of stock-market prices. *The Journal of Business*, 38:34–105, January 1965.

[49] E. Fama and R. Roll. Some properties of symmetric stable distributions. *Journal of the American Statistical Association*, 63:817–36, September 1968.

[50] E. Fama and R. Roll. Parameter estimates for symmetric stable distributions. *Journal of the American Statistical Association*, 66(334):331–8, June 1971.

[51] J. Faraway and M. Jhun. Bootstrap choice of bandwidth selection for density estimation. *Journal of the American Statistical Association*, 85(412):1119–1122, December 1990.

[52] W. Feller. *An Introduction to Probability Theory and its Applications*, volume II. John Wiley, 2nd edition, 1971.

[53] A. Feuerverger and P. McDunnough. On the efficiency of empirical characteristic function procedures. *Journal of the Royal Statistical Society, Series B*, 43(1):20–7, 1981.

[54] A. Feuerverger and R. Mureika. The empirical characteristic function and its applications. *The Annals of Statistics*, 5(1):88–97, 1977.

[55] Richard P. Feynman. *QED, The Strange Theory of Light and Matter.* Princeton University Press, 1988.

[56] Richard P. Feynman. *"Surely You're Joking Mr. Feynman!": Adventures of a Curious Character.* W. W. Norton & Company, 1997.

[57] B. Fielitz and E. Smith. Asymmetric stable distributions of stock price changes. *Journal of the American Statistical Association*, 67(340):813–4, December 1972.

[58] E. Fishler and H. Messer. Order statistics approach for determining the number of sources using an array of sensors. *IEEE Signal Processing Letters*, 6(7):179–82, July 1999.

[59] J. Friedmann, H. Messer, and J-F. Cardoso. Robust parameter estimation of a deterministic signal in impulsive noise. *IEEE Transactions on Signal Processing*, 48(4):935–42, April 2000.

[60] Y. Fujikoshi. Asymptotic expansions for the distributions of the sample roots under nonnormality. *Biometrika*, 67(1):45–51, 1980.

[61] Galileo Galilei. *Dialogue Concerning the Two Chief World Systems.* Modern Library, 2001 (originally published 1632). Translated by Stillman Drake.

[62] J. Gallardo, D. Makrakis, and L. Orozco-Barbosa. Use of $\alpha$-stable self-similar stochastic processes for modeling traffic in broadband networks. *Performance Evaluation*, 40(1/3):71–98, 2000.

[63] B. Gnedenko and A. Kolmogorov. *Limit Distributions for Sums of Independent Random Variables.* Addison Wesley, 1968.

[64] C. Granger and D. Orr. "Infinite variance" and research strategy in time series analysis. *Journal of the American Statistical Association*, 67(338):275–85, June 1972.

[65] P. Hall. A comedy of errors: The canonical form for a stable characteristic function. *Bulletin of the London Mathematical Society*, 13:23–7, May 1980.

[66] P. Hall. On the number of bootstrap simulations required to construct a confidence interval. *The Annals of Statistics*, 14(4), 1986.

[67] P. Hall and M. Martin. On the bootstrap and two-sample problems. *Australian Journal of Statistics*, 30A:179–92, 1988.

[68] P. Hall and D. Titterington. The effect of simulation order on level accuracy and power on Monte Carlo tests. *Journal of the Royal Statistical Society, Series B*, 51(3):459–67, 1989.

[69] C. Heathcote. A test of goodness of fit for symmetric random variables. *Australian Journal of Statistics*, 14(2):172–81, 1972.

[70] C. Heathcote. The integrated squared error estimation of parameters. *Biometrika*, 64(2):255–64, 1977.

[71] Y. Hochberg. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, 75(4):800–2, 1988.

[72] Y. Hochberg and A. Tamhane. *Multiple Comparison Procedures.* John Wiley, 1987.

[73] R. Hogg. Adaptive robust procedures: A partial review and some suggestions for future applications and theory. *Journal of the American Statistical Association*, 69(348):909–27, December 1974.

[74] R. Hogg. Statistical robustness: One view of its use in applications today. *The American Statistician*, 33(3):108–15, August 1979.

[75] B. Holland and M. Copenhaver. An improved sequentially rejective Bonferroni test procedure. *Biometrics*, 43:417–23, June 1987.

[76] S. Holm. A simple sequentially rejective multiple test procedure. *Scandanavian Journal of Statistics*, 6:65–70, 1979.

[77] H. Holma and A. Toskala, editors. *WCDMA for UTMS: Radio Access for Third Generation Mobile Communications.* John Wiley, 2000.

[78] G. Hommel. A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika*, 75(2):383–6, 1988.

[79] G. Hommel. A comparison of two modified Bonferroni procedures. *Biometrika*, 76(3):624–5, 1989.

[80] H. Hsu, R. Storwick, D. Schlick, and G. Maxam. Measured amplitude distribution of automotive ignition noise. *IEEE Transactions on Electromagnetic Compatibility*, EMC-16(2):57–63, May 1974.

[81] P. Huber. *Robust Statistics.* Wiley, 1981.

[82] J. Ilow. *Signal Processing in Alpha-Stable Environments: Noise Modeling, Detection and Estimation.* PhD thesis, Electrical and Computer Engineering, University of Toronto, December 1995.

[83] J. Ilow and D. Hatzinakos. Detection in alpha-stable noise environments based on prediction. *International Journal of Adaptive Control and Signal Processing*, 11:555–68, 1997.

[84] J. Ilow and D. Hatzinakos. Analytic alpha-stable noise modeling in a Poisson field of interferers or scatterers. *IEEE Transactions on Signal Processing*, 46(6):1601–11, June 1998.

[85] J. Ilow and D. Hatzinakos. Applications of the empirical characteristic function to estimation and detection problems. *Signal Processing*, 65:199–219, March 1998.

[86] J. Ilow, D. Hatzinakos, and A. Venetsanopoulos. Detection for binary transmission based on the empirical characteristic function. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 5, pages 2487–90, Atlanta, USA, May 1996.

[87] J. Ilow, D. Hatzinakos, and A. Venetsanopoulos. Performance of FH SS radio networks with interference modelled as a mixture of Gaussian and alpha-stable noise. *IEEE Transactions on Communications*, 46(4):509–20, April 1998.

[88] J. Ilow and H. Leung. No evidence of stable distributions in radar clutter. In *Proceedings of the IEEE Signal Processing Workshop on Higher Order Statistics*, pages 264–267, Banff, Alberta, July 1997.

[89] A. James. The distribution of the latent roots of the covariance matrix. *The Annals of Mathematical Statistics*, 31:151–8, 1960.

[90] A. Janicki and A. Weron. Can one see $\alpha$ stable variables and processes? *Statistical Science*, 9(1):109–26, 1994.

[91] B. Kannan, W. Fitzgerald, and E. Kuruoglu. Joint DOA, frequency and model order estimation in additive $\alpha$-stable noise. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 6, pages 3798–3801, Istanbul, Turkey, June 2000.

[92] W. Kaplan. *Advanced Calculus.* Addison Wesley, 1969.

[93] A. Karasaridis and D. Hatzinakos. On the modeling of network traffic and fast simulation of rare events using $\alpha$-stable self-similar processes. In *Proceedings of the IEEE Signal Processing Workshop on Higher Order Statistics*, pages 268–72, Banff, Canada, July 1997.

[94] A. Karasaridis and D. Hatzinakos. Bandwidth allocation bounds for alpha-stable self-similar internet traffic models. In *Proceedings of the IEEE Signal Processing Workshop on Higher Order Statistics*, pages 214–218, Caesarea, Israel, June 1999.

[95] S. Kassam. *Signal Detection in Non-Gaussian Noise.* Springer-Verlag, 1988.

[96] S. Kogon and D. Manolakis. Signal modeling with self-similar $\alpha$-stable processes: The fractional Lévy stable motion model. *IEEE Transactions on Signal Processing*, 44(4):1006–10, April 1996.

[97] S. Kogon and D. Williams. *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, chapter Characteristic Function Based Estimation of Stable Distribution Parameters. Birkhäuser, 1998.

[98] A. Kolmogorov. On the approximation of distributions of sums of independent summands by infinitely divisible distributions. *Sankhya Series A*, 25:159–74, 1963.

[99] K. Kolodziejski and J. Betz. Detection of weak random signals in IID non-Gaussian noise. *IEEE Transactions on Communications*, 48(2):222–230, February 2000.

[100] I. Koutrouvelis. A goodness-of-fit test of simple hypotheses based on the empirical characteristic function. *Biometrika*, 67(1):238–40, 1980.

[101] I. Koutrouvelis. Regression-type estimation of the parameters of stable laws. *Journal of the American Statistical Association*, 75(372), 1980.

[102] I. Koutrouvelis. An iterative procedure for the estimation of the parameters of stable laws. *Communications in Statistics: Simulation and Computation*, 10(1):17–28, 1981.

[103] I. Koutrouvelis and D. Bauer. Asymptotic distribution of regression-type estimators of parameters of stable laws. *Communications in Statistics: Theory and Methods*, 11(23):2715–30, 1982.

[104] I. Koutrouvelis and J. Kellermeier. A goodness-of-fit test based on the empirical characteristic function when parameters must be estimated. *Journal of the Royal Statistical Society, Series B*, 43(2):173–6, 1981.

[105] I. Koutrouvelis and S. Meintanis. Testing for stability based on the empirical characteristic function with applications to financial data. *Journal of Statistical Computation and Simulation*, 64:275–300, 1999.

[106] R. Kozick, R. Blum, and B. Sadler. Signal processing in non-gaussian noise using mixture distributions and the em algorithm. In *Conference Record of the 31st Asilomar Conference on Signals, Systems and Computing*, volume 1, pages 438–42, Pacific Grove, USA, November 1998.

[107] R. Kozick and B. Sadler. Maximum likelihood array processing in non-Gaussian noise with Gaussian mixtures. *IEEE Transactions on Signal Processing*, 48(12):3520–35, December 2000.

[108] R. Kozick and B. Sadler. Robust subspace estimation in non-Gaussian noise. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 6, pages 3818–3821, Istanbul, Turkey, June 2000.

[109] E. Kuruoglu. *Signal Processing In $\alpha$-Stable Noise Environments: A Least $l_p$-Norm Approach.* PhD thesis, Department of Engineering, University of Cambridge, 1998.

[110] E. Kuruoglu, W. Fitzgerald, and P. Rayner. Near optimal detection of signals in impulsive noise modeled with a symmetric $\alpha$-stable distribution. *IEEE Communications Letters*, 2(10):282–4, October 1998.

[111] D. Lawley. Tests of significance for the latent roots of covariance and correlation matrices. *Biometrika*, 43:128–136, 1956.

[112] E. Lehmann. *Testing Statistical Hypotheses.* Springer-Verlag, 2nd edition, 1997.

[113] E. Lehmann and G. Casella. *Theory of Point Estimation.* Springer-Verlag, 2nd edition, 1998.

[114] E. Lukacs. *Characteristic Functions.* Griffin, 2nd edition, 1970.

[115] X. Ma and C. Nikias. On blind channel identification for impulsive signal environments. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 3, pages 1992–5, Detroit, USA, May 1995.

[116] X. Ma and C. Nikias. Parameter estimation and blind channel identification in impulsive signal environments. *IEEE Transactions on Signal Processing*, 43(12):2884–97, December 1995.

[117] B. Mandelbrot. The variation of certain speculative prices. *Journal of Business*, 36(4):394–419, October 1963.

[118] B. Mandelbrot. The variation of some other speculative prices. *The Journal of Business*, 40:393–413, October 1967.

[119] S. Maymon, J. Friedmann, E. Fishler, and H. Messer. Estimation of the parameters of a stable distribution based on order statistics. In *Proceedings of Heavy Tails '99, Applications of Heavy Tailed Distributions in Economics, Engineering and Statistics*, pages TAILS–37, Washington, DC, June 1999.

[120] S. Maymon, J. Friedmann, and H. Messer. A new method for estimating parameters of a skewed alpha-stable distribution. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 6, pages 3822–5, Istanbul, Turkey, June 2000.

[121] P. McCullagh and J. Nelder. *Generalized Linear Models*. Chapman and Hall, 2nd edition, 1989.

[122] J. McCulloch. Simple consistent estimators of stable distribution parameters. *Communications in Statistics: Simulation and Computation*, 15(4):1109–36, 1986.

[123] J. McCulloch. *Statistical Methods in Finance*, volume 14 of *Handbook of Statistics*, chapter Financial Applications of Stable Distributions. Elsevier, 1996.

[124] J. McCulloch. *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, chapter Numerical Approximation of the Symmetric Stable Distribution and Density, pages 489–500. Birkhäuser, 1998.

[125] K. McDonald and R. Blum. A statistical and physical mechanisms-based interference and noise model for array observations. *IEEE Transactions on Signal Processing*, 48(7):2044–2056, July 2000.

[126] P. Mertz. Model of impulsive noise for data transmission. *IRE Transactions on Communications Systems*, CS-9(2):130–7, June 1961.

[127] James A. Michener. *The Source.* Random House, 2002.

[128] D. Middleton. Statistical-physical models of urban radio-noise environments - part 1: Foundations. *IEEE Transactions on Electromagnetic Compatibility*, EMC-14(2):38–56, May 1972.

[129] D. Middleton. Statistical-physical models of electromagnetic interference. *IEEE Transactions on Electromagnetic Compatibility*, EMC-19(3):106–27, August 1977.

[130] D. Middleton. Canonical and quasi-canonical probability models of class A interference. *IEEE Transactions on Electromagnetic Compatibility*, EMC-25(2):76–106, May 1983.

[131] D. Middleton. *An Introduction to Statistical Communication Theory.* IEEE Press, 1996.

[132] D. Middleton. Non-Gaussian noise models in signal processing for telecommunications: New methods and results for class A and class B noise models. *IEEE Transactions on Information Theory*, 45(4):1129–49, May 1999.

[133] J. Miller and J. Thomas. Detectors for discrete-time signals in non-Gaussian noise. *IEEE Transactions on Information Theory*, IT-18(2):241–50, March 1972.

[134] R. Muirhead. *Aspects of Multivariate Statistical Theory.* John Wiley, 1982.

[135] C. Nikias and M. Shao. *Signal Processing with Alpha-Stable Distributions and Applications.* John Wiley, 1995.

[136] J. Nolan. Numerical calculation of stable densities and distribution functions. *Communications in Statistics: Stochastic Models*, 13(4):759–74, 1997.

[137] J. Nolan. Parameter estimation and data analysis for stable distributions. In *Conference Record of the 31st Asilomar Conference on Signals, Systems and Computing*, volume 1, pages 443–7, Pacific Grove, USA, November 1998.

[138] J. Nolan. *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, chapter Univariate Stable Distributions: Parameterizations and Software, pages 527–33. Birkhäuser, 1998.

[139] J. Nolan. Fitting data and assessing goodness of fit with stable distributions. In *Proceedings of Heavy Tails '99, Applications of Heavy Tailed Distributions in Economics, Engineering and Statistics*, pages TAILS–42, Washington, DC, June 1999.

[140] J. Nolan. *Lévy processes. Theory and applications.*, chapter Maximum Likelihood Estimation and Diagnostics for Stable Distributions, pages 379–400. Birkhäuser, 2001.

[141] R. Officer. The distribution of stock returns. *Journal of the American Statistical Association*, 67(340):807–12, December 1972.

[142] T. Ojanperä and R. Prasad, editors. *WCDMA: Towards IP Mobility and Mobile Internet.* Artech House, 2001.

[143] M. Paolella. Testing the stable paretian assumption. *Mathematical and Computer Modelling*, 34:1095–112, 2001.

[144] A. Paulson and T. Delehanty. Modified weighted squared error estimation procedures with special emphasis on the stable laws. *Communications in Statistics: Simulation and Computation*, 14(4):927–72, 1985.

[145] A. Paulson, E. Holcomb, and R. Leitch. The estimation of the parameters of the stable laws. *Biometrika*, 62(1):163–70, 1975.

[146] A. Petropulu, J. Pesquet, X. Yang, and J. Yin. Power-law shot noise and its relationship to long-memory $\alpha$-stable processes. *IEEE Transactions on Signal Processing*, 48(7):1883–92, July 2000.

[147] R. Pierce. Application of the positive alpha-stable distribution. In *Proceedings of the IEEE Signal Processing Workshop on Higher Order Statistics*, pages 420–4, Banff, Canada, July 1997.

[148] D. Politis. Computer-intensive methods in statistical analysis. *IEEE Signal Processing Magazine*, pages 39–55, January 1998.

[149] D. Politis, J. Romano, and M. Wolf. *Subsampling.* Springer-Verlag, 1999.

[150] H. Poor. *An Introduction to Signal Detection and Estimation.* Springer-Verlag, 1988.

[151] H. Poor and M. Tanda. Multiuser detection in flat fading non-gaussian channels. *IEEE Transactions on Communications*, 50(11):1769–77, November 2002.

[152] S. Press. *Applied Multivariate Analysis.* Holt, Rinehart and Winston, 1972.

[153] S. Press. Estimation in univariate and multivariate stable distributions. *Journal of the American Statistical Association*, 67(340):842–6, December 1972.

[154] S. Press. Multivariate stable distributions. *Journal of Multivariate Analysis*, pages 444–62, 1972.

[155] Carl Sagan. *Cosmos.* Random House, 2002.

[156] G. Samorodnitsky and M. Taqqu. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance.* Chapman and Hall, 1994.

[157] L. Scharf. *Statistical Signal Processing : Detection, Estimation, and Time Series Analysis.* Addison Wesley, 1991.

[158] M. Sekine and Y. Mao. *Weibull Radar Clutter.* Peter Peregrinus, 1990.

[159] R. Serfling. *Approximation Theorems of Mathematical Statistics.* John Wiley, 1980.

[160] J. Shaffer. Modified sequentially rejective multiple test procedures. *Journal of the American Statistical Association*, 81(395):826–31, September 1986.

[161] R. Shepherd. Measurements of amplitude probability distributions and power of automobile ignition noise at HF. *IEEE Transactions on Vehicular Technology*, VT-23(3):72–83, August 1974.

[162] R. Shepherd, J. Gaddie, and D. Nielson. New techniques for suppression of automobile ignition noise. *IEEE Transactions on Vehicular Technology*, VT-25(1):2–12, February 1976.

[163] B. Silverman. *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, 1986.

[164] R. Simes. An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3):751–4, 1986.

[165] B. Stuck and B. Kleiner. A statistical analysis of telephone noise. *The Bell System Technical Journal*, 53(7):1263–1320, September 1974.

[166] A. Swami. Non-Gaussian mixture models for detection and estimation in heavy-tailed noise. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 6, pages 3802–3805, Istanbul, Turkey, June 2000.

[167] A. Swami and B. Sadler. TDE, DOA and related parameter estimation problems in impulsive noise. In *Proceedings of the IEEE Signal Processing Workshop on Higher Order Statistics*, pages 273–7, Banff, Canada, July 1997.

[168] A. Swami and B. Sadler. Parameter estimation for linear alpha-stable processes. *IEEE Signal Processing Letters*, 5(2):48–50, February 1998.

[169] A. Swami and B. Sadler. On some detection and estimation problems in heavy-tailed noise. *Signal Processing*, 82:1471–88, 2002.

[170] A. Taleb, R. Brcich, and M. Green. Suboptimal robust estimation for signal plus noise models. In *Conference Record of the 34$^{rd}$ Asilomar Conference on Signals, Systems and Computing*, volume 2, pages 837–41, Pacific Grove, USA, October 2000.

[171] J. Teichmoeller. A note on the distribution of stock price changes. *Journal of the American Statistical Association*, 66(334):282–4, June 1971.

[172] A. Teschioni, C. Sacchi, and C. Regazzoni. Non-Gaussian characterization of DS/CDMA noise in few-user systems with complex signature sequences. *IEEE Transactions on Signal Processing*, 47(1):234–7, January 1999.

[173] A. Tesei, R. Bozzano, and C. Regazzoni. Comparison between asymmetric generalized Gaussian (AGG) and symmetric-$\alpha$-stable (S$\alpha$S) noise models for signal estimation in non-Gaussian environments. In *Proceedings of the IEEE Signal Processing Workshop on Higher Order Statistics*, pages 259–63, Banff, Canada, July 1997.

[174] A. Tesei and C. Regazzoni. The asymmetric generalized Gaussian function: a new HOS-based model for generic noise pdfs. In *Proceedings of the IEEE Signal Processing Workshop on Higher Order Statistics*, pages 210–3, Corfu, Greece, June 1996.

[175] J. Thompson and R. Tapia. *Nonparametric Function Estimation, Modeling, and Simulation.* Society for Industrial and Applied Mathematics, 1990.

[176] J. Thornton and A. Paulson. Asymptotic distribution of characteristic function-based estimators for the stable laws. *Sankhya: The Indian Journal of Statistics*, 39(Series A, Pt. 4):341–54, 1977.

[177] P. Tsakalides. *Array Signal Processing with Alpha-Stable Distributions.* PhD thesis, Electrical Engineering, University of Southern California, 1995.

[178] P. Tsakalides and C. Nikias. Maximum likelihood localization of sources in noise modeled as a stable process. *IEEE Transactions on Signal Processing*, 43(11):2700–13, November 1995.

[179] P. Tsakalides and C. Nikias. The robust covariation-based MUSIC (ROC-MUSIC) algorithm for bearing estimation in impulsive environments. *IEEE Transactions on Signal Processing*, 44(7):1623–33, July 1996.

[180] P. Tsakalides and C. Nikias. A new model for non-Rayleigh clutter: Space-time adaptive processing in stable impulsive interference. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 5, pages 3517–20, Munich, Germany, April 1997.

[181] G. Tsihrintzis and C. Nikias. Fast estimation of the parameters of alpha-stable impulsive interference using asymptotic extreme value theory. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 3, pages 1840–3, Detroit, USA, May 1995.

[182] G. Tsihrintzis and C. Nikias. Performance of optimum and suboptimum receivers in the presence of impulsive noise modelled as an alpha stable process. *IEEE Transactions on Communications*, 43(2/3/4):904–14, Feb/March/April 1995.

[183] G. Tsihrintzis and C. Nikias. Data-adaptive algorithms for signal detection in impulsive noise modelled as a sub-Gaussian, alpha-stable process. In *Proceedings of the $8^{th}$ IEEE Workshop on Statistical Signal and Array Processing*, pages 238–41, Corfu, Greece, June 1996.

[184] G. Tsihrintzis and C. Nikias. Fast estimation of the parameters of alpha stable impulsive interference. *IEEE Transactions on Signal Processing*, 44(6):1492–1503, June 1996.

[185] G. Tsihrintzis and C. Nikias. Data-adaptive algorithms for signal detection in sub-Gaussian impulsive interference. *IEEE Transactions on Signal Processing*, 45(7):1873–8, July 1997.

[186] K. Vastola. Threshold detection in narrow-band non-Gaussian noise. *IEEE Transactions on Communications*, 32(2):134–9, February 1984.

[187] S. Verdu. *Multiuser Detection.* Cambridge University Press, 1998.

[188] X. Wang and H. Poor. Robust multiuser detection in non-Gaussian channels. *IEEE Transactions on Signal Processing*, 47(2):289–304, February 1999.

[189] C. Waternaux. Asymptotic distribution of the sample roots for a nonnormal population. *Biometrika*, 63(3):639–45, 1976.

[190] M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-33(2):387–92, April 1985.

[191] M. Wax and I. Ziskind. Detection of the number of coherent signals by the MDL principle. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(8):1190–6, August 1989.

[192] D. Williams and D. Johnson. Using the sphericity test for source detection with narrow-band passive arrays. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 38(11):2008–14, November 1990.

[193] K. Worsley. An improved Bonferroni inequality and applications. *Biometrika*, 69(2):297–302, 1982.

[194] S. Zabin and G. Wright. Nonparametric density estimation and detection in impulsive interference channels-part I: Estimators. *IEEE Transactions on Communications*, 42(2/3/4):1684–1697, February/March/April 1994.

[195] S. Zabin and G. Wright. Nonparametric density estimation and detection in impulsive interference channels-part II: Detectors. *IEEE Transactions on Communications*, 42(2/3/4):1698–1711, February/March/April 1994.

[196] V. Zolotarev. *One-dimensional Stable Distributions*, volume 65. American Mathematical Society, 1986.

[197] A. Zoubir and M. Arnold. Testing Gaussianity with the characteristic function: The i.i.d case. *Signal Processing*, 53:245–55, 1996.

[198] A. Zoubir and B. Boashash. The bootstrap and its application in signal processing. *IEEE Signal Processing Magazine*, pages 55–76, January 1998.

[199] A. Zoubir and J. Böhme. Bootstrap multiple tests applied to sensor location. *IEEE Transactions on Signal Processing*, 43(6):1386–96, June 1995.

[200] A. Zoubir and C. Brown. Testing for impulsive behavior: A bootstrap approach. *Digital Signal Processing*, 11(2):120–32, April 2001.

*Vision without action is dreaming.*
*Action without vision is passing the time.*
*Action with vision can change the world.*