O-COCOSDA2010

Nov 24-25, 2010
Kathmandu, Nepal

O-COCOSDA 2010 Committee

Conference Chair
Professor Jai Raj Awasthi, Tribhuban University, Kathmandu

Conference Co-Chair
Dr. Chiu-yu Tseng, Academia Sinica, Taipei

Conference Secretary
Mr. Bhim Narayan Regmi, CECODES, Kathmandu

Organizing Committee
Mr. Krishna Prasad Parajuli, CECODES, Kathmandu
Mr. Sagun Dhakhwa, CECODES, Kathmandu
Mr. Gelu Sherpa, CECODES, Kathmandu
Mr. Kamal Poudel, CECODES, Kathmandu
Mr. Bhim Lal Gautam, CECODES, Kathmandu
Mr. Sandeep Khatri, CECODES, Kathmandu

Mr. Asoke Datta, CDAC, Kolkata
Dr. Om Vikas, CDAC, Noida
Pushpak Bhattacharya, IIT, Mumbai
Mrs. Swarn Lata, DIT, MCIT

Advisory Committee

Honorable Nilkantha Upreti, Executive Chief Commissioner, Election Commission, Nepal
Honorable Tirtha Raj Khaniya, Member, National Planning Commission, Nepal
Honorable Manohar Kumar Bhattarai, Vice Chairman, HLCIT, Nepal
Mr. Bairagi Kainla (Til Bikram Nembang), Chancellor, Nepal Academy
Prof. Dr. Surendra Raj Kafle, Vice Chancellor, Nepal Academy of Science and Technology
Professor Chura Mani Bandhu, Nepal
Professor Tej Ratna Kansakar, Nepal
Professor Novel Kishore Rai, CNAS, Tribhuvan University,Nepal
Professor Govinda Raj Bhattarai, Tribhuvan University, Nepal

B.Yegnanarayana, IIIT, Hyderabad, India
Hema Murthy, IIT, Madras, India
Professor Hiroya Fujisaki, Japan

D.D. Majumdar, ISI, Kolkata, India
P.V.S. Rao,Former-Head-CSC & Speech Tech. Group
Rajeev Sanjal, IIIT, Hyderabad, India
Dr. RMK Sinha, IIT, Kanpur, India
Professor Shuichi Itahashi, NII/AIST, Japan
President IETE
ED-CDAC, Noida , India
ED CDAC, Kolkata, India
Director-IIIT, HYD, India

Technical Program Committee

Chairs
Dr. Shyam S. Agrawal (CDAC, Noida & KIIT, Gurgaon, India)
Professor Yoshinori Sagisaka (Waseda University, Japan)

Committee
Dr. Aijun Li (Chinese Academy of Sciences, Beijing, China)
Dr. B. Mallikarjun, CIIL, Mysore, India
Bimal Acharya, Nepal Telecom Company Limited, Nepal
Dr. Chai Wutiwiwatchai (NECTEC, Thailand)
Dr. Chiu-yu Tseng (Academia Sinica, Taipei)
Dr. Haizhou Li (Institute of Infocom Research, Singapore)
Dr. Hammam Riza (BTTP, Indonesia)
Professor Hideaki Kikuchi (Waseda University, Japan)
Professor Hsiao-Chuan Wang, (National Tsing Hua University, Taiwan)
K K Arora, CDAC, Noida
Dr. K. Samudravijaya (Tata Institute of Fundamental Research, India)
Professor Kanika Kaur, KIIT, Gurgaon, India
Dr. Luong Chi Mai (The Vietnamese Academy of Sciences, Vietnam)
Professor Nick Campbell (Trinity College, Ireland)
Professor Rachael Roxas, (De la Salle University-Manila, Philippines)
Dr. Satoshi Nakamura (NICT/ATR, Japan)
Shyamal Das Mandal, CDAC, Kolkata
Swarn lata, DIT, Govt.India
Professor Tan Lee (Chinese University of Hong Kong, Hong Kong)
Professor Thomas Fang Zheng (Tsing Hua University, Beijing)
Professor Yong Ju Lee (Wonkwang University, Korea)
Professor Zuraidah Mohd Don (University of Malaysia, Malaysia)

The Development of a Thai Speech Set for Telephonometry

Therdpong Daengsi[1], Apiruck Preechayasomboon[2], Saowanit Sukparungsee[3], Patcharika Chootrakool[4], Chai Wutiwiwatchai[5]

[1]*Faculty of Information Technology, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand*
[2]*International Business Development Department, TOT Public Company Limited, Bangkok, Thailand*
[3]*Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology, North Bangkok, Thailand*
[4,5]*Human Language Technology Laboratory, National Electronics and computer Technology Center, Pathumthani Thailand*
therdpong1@yahoo.com, apiruck@tot.co.th, saowanits@hotmail.com, patcharika.chootrakool@nectec.or.th, chai.wutiwiwatchai@nectec.or.th

This paper presents the development for the Thai Speech Set for Telephonometry (TSST), which is mainly required for voice quality measurement for telecommunications. TSST was designed and developed by following the International Telecommunication Union – Telecommunication Standardization Sector (ITU-T) recommendation. Tasks were divided into three parts. The first part was about survey to find frequently used sentences (or phrases). The second part was to investigate for fifty frequently used sentences and to create the representative sentences of those, called Thai Text Set for Telephonometry (TTST). Finally, the last part was about speech recording in a high standard studio for TSST. The output from this work will be useful for telecommunication research and related research areas in Thai environments.

Speaker Recognition Based on Multilingual Speech Features using Neural Network Models

Sanjay Decate, Anupam Shukla, Sanjay Kumar Singh, Ritu Tiwari

ABV-IIITM, Gwalior, India
sanjaydekate@gmail.com, dranupamshukla@gmail.com, sksinghiiitm@gmail.com, tiwariritu2@gmail.com

In the present paper an attempt is made to develop multilingual speaker recognition system which is used to identify the identity of an unknown speaker among several speakers of known speech characteristics, from a sample of his or her utterances.

Every speaker has different individual characteristics embedded in his/her speech utterances. To evaluate speech characteristics from utterances they are stored in digitized form. Speech features namely LPC, RC, APSD, Number of zero crossing and Formant frequencies are extracted from speech signal and formed speech feature vectors. The database used for this system consists of 30 speakers including both male and female from different parts of India and languages are Hindi, English and Sanskrit containing total 14 words.

The average identification rate was 87.82% with BPA and improved by 2.87% with RBF and further by 3.57% with LVQ neural network models.

Orthography development for the standardization of Bhujel: Issues and approaches

Dr. Dan Raj Regmi

Central Department of Linguistics, Tribhuvan University, Nepal
danraj_regmi@hotmail.com

This paper attempts to examine the issues and approaches to orthography development for the standardization of Bhujel, a preliterate and endangered Tibeto-Burman language with different dialects. It is spoken in different villages of Gorkha, Tanahun, Nawalparasi and Chitwan districts of Nepal. Developing orthography is indispensable for literacy and multilingual education. It is predominantly central for the standardization of the language itself. Writing system is a linguistic as well as a social reality (Robinson, 2003). There are three scripts choices for Bhujel: Roman, Tibetan and Devanagari. However, it is hard to find common voice regarding script choice.

Recognition of Hindi Phoneme in Rhyming Words using Vector Quantization

Shweta Sinha[1], Shalini Goyal[2], Mona Gaur[2], S.S Agrawal[2]

[1]*KIIT College Of IT and Management, Gurgaon*
[2]*KIIT College Of Engineering, Gurgaon*
meshweta_7@rediffmail.com, shalinigoyal1980@gmail.com
monagaur1311@gmail.com, ss_agrawal@hotmail.com

This paper presents and discusses the recognition of Phoneme in Rhyming word environment using Vector Quantization. Recognition

The Development of a Thai Speech Set for Telephonometry

Therdpong Daengsi[1], Apiruck Preechayasomboon[2], Saowanit Sukparungsee[3], Patcharika Chootrakool[4], Chai Wutiwiwatchai[5]

[1]*Faculty of Information Technology, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand*

[2]*International Business Development Department, TOT Public Company Limited, Bangkok, Thailand*

[3]*Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North, Bangkok, Thailand*

[4,5]*Human Language Technology Laboratory, National Electronics and computer Technology Center, Pathumthani Thailand*

Email: therdpong1@yahoo.com, apiruck@tot.co.th, saowanits@hotmail.com, patcharika.chootrakool@nectec.or.th, chai.wutiwiwatchai@nectec.or.th

Abstract

This paper presents the development for the Thai Speech Set for Telephonometry (TSST), which is mainly required for voice quality measurement for telecommunications. TSST was designed and developed by following the International Telecommunication Union – Telecommunication Standardization Sector (ITU-T) recommendation. Tasks were divided into three parts. The first part was about survey to find frequently used sentences (or phrases). The second part was to investigate for fifty frequently used sentences and to create the representative sentences of those, called Thai Text Set for Telephonometry (TTST). Finally, the last part was about speech recording in a high standard studio for TSST. The output from this work will be useful for telecommunication research and related research areas in Thai environments.

1. Introduction

Voice over Internet Protocol (VoIP) is one of the high potential technologies in this era but one of its major issues is voice quality. Voice quality measurement can be conducted using either the subjective and/or objective methods [1]. However, subjective measurement method requires some speech resources that must be in the standard of ITU-T. The ITU-T recommendation mainly provides guidelines for the English language. As Thai is a tonal language and tones might be an important factor affecting voice quality, this kind of resource should be designed specifically for Thai. This paper describes the development of the resource called Thai Speech Set for Telephonometry (TSST).

2. Background

2.1. Tones in Thai

Tonal languages use syllabic tones to distinguish lexical words. As in Figure 1, five tones in Thai show different fundamental frequency (F0) shapes.

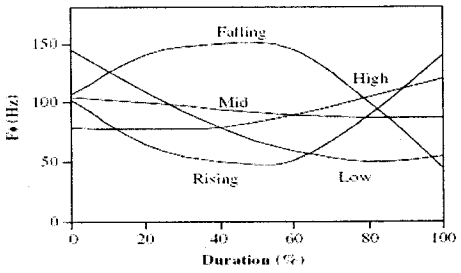

Figure 1. An example of characteristic of fundamental frequency (F0) contours of Thai tones [3]

There are middle tone, low tone, falling tone, high tone and rising tone. Different tones could result

The Pyrenees International Workshop on Statistics, Probability and Operations Research

SPO 2009

ABSTRACTS

Jaca, Spain, September 15-18, 2009

The Pyrenees International Workshop on Statistics, Probability and Operations Research, SPO 2009 Jaca, September 15–18th 2009

Combining martingale and integral equations approaches for finding optimal designs of EWMA procedure

Saowanit Sukparungsee[1], Yupaporn Areepong[1]

SUMMARY

This research is concerned with the use of Statistical Process Control (SPC) charts for detecting a discrete time change-point in parameter in quality control. The most popular characteristics of a control chart are Average Run Length (ARL) and Average Delay (AD) are used for comparing the performances of control charts. We used the analytical closed-form formulas based on martingale technique and numerical integral equations obtained by the Gauss-Legendre rule for evaluating the ARL and AD in the case of Gaussian observations. In particular, we have developed a numerical algorithm for finding optimal parameters of an Exponentially Weighted Moving Average (EWMA) procedure. Furthermore, this algorithm can be numerically used to approximate an overshoot which the distribution of the overshoot is unknown. The numerical results are compared with the algorithm which combined the results from the martingale-based and the simulations which is very time consuming. The results show that the accuracy from the new numerical algorithm is as good as that from the simulations. Moreover, the developed algorithm takes much less the computing time than the former technique[1].

Keywords: Martingale, Integral equation, Gauss-Legendre, Overshoot, Average run length, Average delay and Exponentially weighted moving average

AMS Classification: 45B05, 60G42, 93E20

References

[1] SUKPARUNGSEE, S. AND NOVIKOV, A.A. (2006). On EWMA procedure for detection of a change in observations via martingale approach. *KMITL Science Journal* **6**(a), 373–380.

[1]Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok, 10800, Thailand
saowanits@hotmail.com
Yupaporn.Areepong@hotmail.com

The International Conference in Mathematics and Applications

is sponsored by

CEM
PERDO

Centre of Excellence in Mathematics, PERDO,
Commission on Higher Education, Thailand

National Research Council of Thailand

Mahidol University, Thailand

COMMITTEES

International Scientific Committee

Chair:

Yongwimon Lenbury	Mahidol University

Executive Chair:

Nguyen Van Sanh	Mahidol University

Members:

Ravi P. Agarwal	Florida Institute of Technology
Jonathan Bell	University of Maryland, Baltimore County
Philip S. Crooke	Vanderbilt University
Andrea De Gaetano	CNR IASI Laboratorio di Biomatematica, Italy
Sompong Dhompongsa	Chiang Mai University
Dinh Van Huynh	Ohio University
Hideaki Kaneko	Old Dominion University
Gerasimos Ladas	University of Rhode Island
Wayne M. Lawton	National University of Singapore
Christian Licht	Université Montpellier II
Horst Martini	Chemnitz University
Charles A. Micchelli	State University of New York
Michael Ng	Hong Kong Baptist University
Vu Quoc Phong	Ohio University
Kar Ping Shum	The Chinese University of Hong Kong
Bikas K. Sinha	Indian Statistical Institute
Bimal K. Sinha	University of Maryland
Glenn F. Webb	Vanderbilt University
Bernd Wegner	Technische Universität Berlin
Yong Hong Wu	Curtin University of Technology
Yuesheng Xu	Syracuse University
Zhimin Zhang	Wayne State University

Some Analytical Results on CUSUM and EWMA Control Charts*

Gabriel Mititelu[1], Yupaporn Areepong[2], Saowanit Sukparungsee[2], Alexander A. Novikov[1]

[1]Department of Mathematical Sciences, University of Technology Sydney, Broadway, NSW 2007, Sydney, Australia.
Gabriel.Mititelu@uts.edu.au; Alex.Novikov@uts.edu.au
[2]Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand.
yupaporna@kmutnb.ac.th; swns@kmutnb.ac.th

Abstract

The control charts CUSUM (CUmulative SUM) and EWMA (Exponentially Weighted Moving Average) are widely used in a great variety of practical applications such as economics, finance, medicine, and engineering. The Average Run Length (ARL) is the most common characteristic used to design EWMA and CUSUM.

Here we use the Fredholm type integral equations to derive analytical closed form representations for the ARL for some special cases. In particular, we derive a closed form representation for the ARL of CUSUM chart assuming that the random observations have a hyperexponential distribution. For EWMA we solve the corresponding ARL integral equation when the observations have the Laplace distribution.

Keywords: analytical closed form solutions, control charts, hyperexponential distribution, integral equations, Laplace distribution.

*Corresponding Author:Gabriel.Mititelu@uts.edu.au

New Solutions in Planar Re-Entry Aerodynamics*

Gabriel Mititelu[1], Yupaporn Areepong[2]

[1]Department of Mathematical Sciences, University of Technology Sydney, Broadway, NSW 2007, Sydney, Australia.
Gabriel.Mititelu@uts.edu.au

[2]Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand.
yupaporna@kmutnb.ac.th

Abstract

In this paper we present integrable solutions for the atmospheric re-entry dynamical equations of motion of a space vehicle, under the assumptions of standard atmospheric model.

This is a very important and practical problem encountered during the atmospheric re-entry phase and our solution can be effectively applied to investigate and control the rocket flight characteristics.

By setting the initial conditions for the speed, re-entering flight-path angle, altitude, atmosphere density, lift and drag coefficients, the nonlinear differential equations of motion are linearized by a proper choice of the re-entry range angles. After integration, the solutions are expressed with the Exponential Integral, and Generalized Exponential Integral functions. Theoretical frameworks for proposed solutions as well as, several numerical examples, are presented.

Keywords: exponential integral function, generalized exponential integral function, integrable solutions, re-entry aerodynamics.

*Corresponding Author:Gabriel.Mititelu@uts.edu.au

The Pyrenees International Workshop on Statistics, Probability and Operations Research, SPO 2009 Jaca, September 15–18th 2009

The Performance of Forecasting Models for Non-Linear Financial data

Pathom Glannamtip[1], Kanyapat Pussara[1]

SUMMARY

The purpose of this paper is to analyze the performance of forecasting models when the financial data show a non-linear structure. The daily exchange rates between US dollars and Thai baht from 2006 to 2008 were analyzed. Two forecasting models have been considered. The first is the Autoregressive Conditional Heteroscedasticity (ARCH) and the second is the Smooth Transition Autoregressive (STAR). Non-linear financial data were tested by Tsays F test and Keenans test. Dot plot, Linear plot and Mean Square Error (MSE) were used to compare the performance of the forecasting models by R programming. The results showed that the STAR forecasting model has a better perfomance than the ARCH model when parameters α_0 and α_1 are low (0.1 to 0.5) but ARCH model is better than STAR model when both parameters are equal to 0.9 independently of selected sample size.

Keywords: Forecasting model, Non linear models, Financial data

AMS Classification: 62J99

[1]Department of Applied Statistics, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand
pathom@kmutnb.ac.th
Mangpor_bin@hotmail.com

The Pyrenees International Workshop on Statistics, Probability and Operations Research, SPO 2009 Jaca, September 15–18th 2009

Confidence intervals for the difference between two log-normal means

Thongkam Maiklad[1]

SUMMARY

This paper proposes two methods for obtaining confidence intervals for the difference between two log-normal means. The first method is based on the generalized pivotal quantity. The other method is the bootstrap percentile method. Monte Carlo simulations were used to compare the results of estimation from these two methods with results when using Krishnamoorthy and Mathew generalized pivotal quantity method [1]. The results found that the coverage probabilities obtained from both the proposed and Krishnamoorthy pivotal quantities were greater than the nominal level in all cases of study, but the proposed generalized pivotal quantity yielded shorter expected lengths. If sample sizes from the two log-normal distributions were large and the difference between variances from the two populations was small, the best method for estimating the difference between two log-normal means was the bootstrap percentile method.

Keywords: Two Log-normal Means, Generalized Pivotal Quantity, Bootstrap Percentile

AMS Classification: 62F25

References

[1] KRISHNAMOORTHY, K. AND MATHEW, T. (2003). Inferences on the means of log-normal distributions using generalized p-values and generalized confidence intervals. *Journal of Statistical planning and inference* **115**, 103–121.

[1]Department of Applied Statistics, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand
tks@kmutnb.ac.th

The Pyrenees International Workshop on Statistics, Probability and Operations Research, SPO 2009 Jaca, September 15–18th 2009

Effect of Unit Root Test on Predictors for a Gaussian AR(1) Process with an Unknown Drift and Additive Outliers

Sa-aat Niwitpong[1], Wararit Panichkitkosolkul[1]

SUMMARY

It is known from Diebold and Kilian [1] and Niwitpong [2] that the Dickey-Fuller [3] unit root test is useful for a one-step-ahead forecast of the trend of an AR(1) process and a Gaussian AR(1) process with an unknown drift respectively. In this paper, we present the standard predictor and the predictor following the Dickey-Fuller unit root test of an AR(1) process with additive outliers. The relative efficiencies of predictors using the scaled prediction mean square error are compared through Monte Carlo simulation studies. Simulation results have shown that the unit root test can improve the accuracy of the predictor when there is an uncertainty of an AR(1) process, i.e. ρ is close to one and there are additive outliers in time series.

Keywords: Additive Outlier, AR(1), Predictor, Unit Toot Test

AMS Classification: 62M10, 63F10, 63F30

References

[1] F.X. DIEBOLD, L.KILIAN (2000). , Unit root tests are useful for selecting forecasting Models. *Journal of Business and Economic Statistics* **18**, 265–273.

[2] S. NIWITPONG (2007). , Predictive inference for times series.,PhD thesis, Department of Statistical Science, La Trobe University.

[3] D.A. DICKEY (1979). , Distribution of estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* **74**, 427–431.

[1]Department of Applied Statistics, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand
snw@kmutnb.ac.th
wararit@mathstat.sci.tu.ac.th

The Pyrenees International Workshop on Statistics, Probability and Operations Research, SPO 2009 Jaca, September 15–18th 2009

Upper Bounds of the Generalized p-Values for the Behrens-Fisher Problem with a Known Ratio of Variances

Suparat Niwitpong[1], Wichai Surachedkiati[1], Gareth Clayton[1]

SUMMARY

Motivated by the recent work of Schechtman and Sherman [1]. Schechtman and Sherman described the situation where we know the ratio of variances, but the variance are unknown as follows: "when the two instruments of equal precision average a different number of replicates in arriving at a response. In this case the ratio of the variances is known, and it is the ratio between the number of readings going into the two (averaged) responses." In this paper, we provide analytic expressions for the generalized p-values, proposed by Tsui and Weerahandi [2], for the Behrens-Fisher problem when a ratio of variances is known, Schechtman and Sherman [1]. The upper bounds of the generalized p-values are derived mathematically using Jensen' inequality. By means of the Monte Carlo simulation, we compare our upper bounds of the generalized p-values of the Behrens-Fisher problem with the existing upper bounds of the p-values.

Keywords: Behrens-Fisher Problem, Generalized p-values, Known Ratio of Variances

AMS Classification: 62F05, 62F30

References

[1] E. SCHECHTMAN AND M. SHERMAN (2007). , The two sample t-test with one variance unknown. *Statistical Methodology* **4**, 508–514.

[2] K.W. TSUI AND S. WEERAHANDI (1989). , Generalized p-values in significance testing of hypothesis in the presence of nuisance parameters. *Journal of the American Statistical Association* **84**, 602–607.

[1]Department of Applied Statistics, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand 10800
suparatn@kmutnb.ac.th

The Pyrenees International Workshop on Statistics, Probability and Operations Research, SPO 2009 Jaca, September 15–18th 2009

Effect of Unit Root Tests on Prediction Intervals for an Unknown Mean Gaussian AR(1) Process with Additive Outliers

Wararit Panichkitkosolkul[1], Sa-aat Niwitpong[2]

SUMMARY

This paper presents the effect of unit root tests on prediction intervals for an unknown mean Gaussian AR(1) process with additive outliers. We present the standard prediction interval, PI_S, the prediction interval following the Dickey-Fuller unit root test [1], PI_{DF}, and the prediction interval following the Shin-Sarkar-Lee unit root test [2], PI_{SSL}. The performance of the proposed prediction intervals is assessed through Monte Carlo simulation studies. Simulation results have shown that the coverage probabilities of PI_{DF} are higher than $1-\alpha$ for all situations. When an autoregressive parameter value approaches one, the coverage probabilities of PI_{SSL} have minimum coverage probabilities $1-\alpha$ for moderate and large sample sizes. Furthermore, the expected length of PI_{SSL} is shortest prediction intervals compared to other prediction intervals.

Keywords: Additive Outlier, AR(1), Prediction Interval, Unit Toot Test

AMS Classification: 62M10, 62F30, 62F25

References

[1] D.A. DICKEY (1979). , Distribution of estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* **74**, 427–431.

[2] D.W. SHIN, S. SARKAR, J.H. LEE (1996). , Unit root tests for time series with outliers. *Statistics & Probability Letters* **30**, 189–197.

[1]Department of Applied Statistics, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand
snw@kmutnb.ac.th
wararit@mathstat.sci.tu.ac.th

The Pyrenees International Workshop on Statistics, Probability and Operations Research, SPO 2009 Jaca, September 15–18th 2009

Parameter Estimation Methods for Three-Parameter Generalized Pareto Distribution with Outliers

Wichitra Phonyiem[1], Prateep Subkongdee[1], Pathom Glannamtip[1]

SUMMARY

The generalized Pareto distribution, is a two-parameter distribution, the shape parameter a and the scale parameter b. It contains uniform, exponential, and Pareto distributions. It has three special cases $a = 0$ and $a = 1$ yield respectively, the exponential with mean b and the uniform distribution on $[0, b]$, and Pareto distributions are obtained when $a < 0$. If c is the lower bound of a random variable X having generalized Pareto distribution, the distribution of X is the three-parameter generalized Pareto distribution where c is a location parameter, b is a scale parameter, and a is a shape parameter. In this paper, the parameter estimation methods of the moments, probability weighted moments, and least squares to estimate the parameters in the three-parameter generalized Pareto distribution with outliers contaminated data are compared. The investigation was done using Monte-Carlo simulation runs with various sample sizes. Results show that the best method of parameter estimation varies with the value of the shape parameter a.

Keywords: Generalized Pareto distribution, Outliers

AMS Classification: 62F25

[1]Department of Applied Statistics, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand
wpy@kmutnb.ac.th
go_ma09@hotmail.com
pathom@kmutnb.ac.th

The Pyrenees International Workshop on Statistics, Probability and Operations Research, SPO 2009 Jaca, September 15–18th 2009

Bias Reduction of the Process Capability Indices

Jeerapa Sappakitkamjorn[1], Sa-aat Niwitpong[1]

SUMMARY

In parameter estimation, a fundamental problem is to reduce or remove the bias in an estimator in order to obtain an accurate estimator. Kotz and Johnson [1] have proved that the estimators of the commonly used process capability indices (PCIs) are biased estimators. Since the bias makes the estimates of these indices inaccurate, it may seriously affect judgment on the performance of the process. Therefore, the focus of this study is to examine methods to be used for bias reduction so that the estimates of these indices will be more accurate. The methods considered in the study are i) the direct bias correction method which is a common method for bias correction where the bias is directly estimated and then a biased corrected estimate is formed, ii) the jackknife method and iii) the method of stochastic approximation proposed by Leung and Wang [2]. To evaluate the efficiency of these methods, a Monte Carlo simulation of 10,000 replications was conducted. Although an increase in sample size can reduce the bias, there are some situations where large sample sizes cannot be applied. Thus, it is of interest to evaluate the efficiency of these methods in bias reduction particularly when the estimates are used to assess the process with small sample sizes. Graphical results are presented.

Keywords: Bias reduction, Jackknife, Process capability indices, Stochastic approximation

AMS Classification: 62F10, 62F40

References

[1] KOTZ, S. AND JOHNSON, N. L. (1992). *Process Capability Indices.* Chapman & Hall, London.

[2] LEUNG, D. H. AND WANG, Y. (1998). Bias Reduction using Stochastic Approximation. *Austral. & New Zealand J. Statist.* **40**(1), 43–52.

[1]Department of Applied Statistics, King Mongkut's University of Technology North Bangkok
1518 Pibulsongkram Road, Bangsue, Bangkok, 10800 THAILAND
jsj@kmutnb.ac.th, snw@kmutnb.ac.th

The Pyrenees International Workshop on Statistics, Probability and Operations Research, SPO 2009 Jaca, September 15–18th 2009

Improvement of Randomized Response Model by Twice Sampling Method

Wichai Suracherdkiati[1], Pasakorn Phanrod[2]

SUMMARY

Conducting research by collecting data with a sensitive question usually obtained the responses that are not correct. This affects the research result. This research aims to make a random response technique to encourage the respondents to answer the questions correctly. This research have developed a mixed model for the random response technique. The mixed model uses a direct question to categorize the respondents into groups. Each respondent from the sample is instructed to answer the direct question. If a respondent answers "Yes" to the direct question, then she or he is instructed to go to randomization device R_1 consisting of the (i) Sensitive Question and (ii) Direct Question, each respondent has to answer one time using simple random sampling with replacement in the random device R_1. If a respondent answers "No" to the direct question, then the respondent is instructed to use a randomization device R_2 consisting of the (i) Sensitive Question and (ii) Say "Yes", each respondent has to answer two times using simple random sampling with replacement in the random device R_2. The results show that the proposed model is an unbiased estimator and more efficient than other models in terms of variance. This can be seen from the percent relative efficiency which is greater than or equal to 100.00 in all cases.

Keywords: Sensitive question, Random responses technique, Random device, Variances

AMS Classification: 62D05

[1]Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Thailand
wsc@kmutnb.ac.th

[2]Dean Office, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Thailand
pasakornp@kmutnb.ac.th

The Pyrenees International Workshop on Statistics, Probability and Operations Research, SPO 2009 Jaca, September 15–18th 2009

Confidence Interval for the Mean Difference of Paired Data Following a Pretest

Sthaporn Thepsumritporn[1], Sa-aat Niwitpong[1]

SUMMARY

This paper investigates the effect of the pretest, there is no correlation between the two sample data, on confidence interval for the difference between means of independent data, CI_i and confidence interval for the difference of means of paired data, CI_p. Typically, for two independent samples, confidence interval for the difference between the two normal means is based on the well-known Welch-Satterthwaite confidence interval and for paired data, the confidence interval for the mean difference is based on the paired t-test. Devore and Berk[1] described that there is no precise answer to the question "What confidence interval will use for paired data with an uncertainty of the degree of correlation of paired sample?". To answer this question, the t-pretest for $H_0 : \rho=0$ is conducted, then the suitable confidence interval between CI_i and CI_p is chosen. We call this new confidence interval for the difference of means following the pretest, CI_f. The performance of the proposed confidence interval, CI_f, CI_i and CI_p is assessed through Monte Carlo simulation studies for a variety of sample sizes and a values of ρ.

Keywords: Confidence Interval, Paired Data, Pretest

AMS Classification: 62F25

References

[1] JAY L. DEVORE AND KENNETH N. BERK (2007). *Modern Mathematical Statistics with Applications.* Thomson Learning, Belmont, CA.

[2]Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand
sthapornt@kmutnb.ac.th
snw@kmutnb.ac.th