بسم الله الرحمن الرحيم

شبكة المعلومات الجامعية

التوثيق الالكتروني والميكروفيلم

شبكة المعلومات الجامعية

جامعة عين شمس

التوثيق الالكتروني والميكروفيلم

قسم

نقسم بالله العظيم أن المادة التي تم توثيقها وتسجيلها

علي هذه الأفلام قد أعدت دون أية تغيرات

يجب أن

تحفظ هذه الأفلام بعيدا عن الغبار

في درجة حرارة من ١٥-٢٥ مئوية ورطوبة نسبية من ٢٠-٤٠%

To be Kept away from Dust in Dry Cool place of 15-25- c and relative humidity 20-40%

بعض الوثائـــــق الاصليـــة تالفـــة

بالرسالة صفحات

لم ترد بالاصل

Cairo University
Institute of Statistical Studies and Research
Department of Mathematical Statistics

PARAMETER ESTIMATION AND TEST OF FIT FOR SKEW NORMAL DISTRIBUTION

By

Mahmoud Abd El-Hamed Mostafa Abd El-Hamed

Diploma in Statistics, I.S.S.R., Cairo University, 2006

Under the Supervision of

Prof. Samir Kamel Ashour

Professor of Mathematical Statistics

**A Thesis Submitted toThe
Institute of Statistic Studies and Research, Cairo University in
Partial Fulfillment of The Requirement for The Degree of
Master of Science
in the Department of Mathematical Statistics**

2009

APPROVAL SHEET

PARAMETER ESTIMATION AND TEST OF FIT FOR SKEW NORMAL DISTRIBUTION

By:

Mahmoud Abd-El Hamed Mostafa

This thesis for Master Degree in Mathematical Statistics, Department of Mathematical Statistics, Institute of Statistical Studies and Research, Cairo University, has bee approved by:

Name	**Signature**
S. K. Ashour	
A. M. Nigm	
Ahmed Fouad	

ACKNOWLEDGMENT

I wish to express my deep gratitude and respect to my supervisor **Prof. Samir Kamel Ashour**, for his valuable guidance and the honest support. Indeed, his helpful and remarkable suggestions were the main reason for a compliment this work. Special thanks also dedicated to him for his close following up, continuous advice and all throughout the review of this work.

Contents

Chapter I
Introduction

The celebrated Gaussian (Normal) distribution has been known for centuries. Its popularity has been driven by its analytical simplicity and the associated Central Limit Theorem. The multivariate extension is straightforward because the marginals and conditionals are both normal, a property rarely found in most of the other multivariate distributions. Yet there have been doubts, reservations, and criticisms about the unqualified use of normality. There are numerous situations when the assumption of normality is not validated by the data. In fact Geary (1947) remarked, Normality is a myth; there never was and never will be a normal distribution." As an alternative, many near normal distributions have been proposed. Some families of such near normal distributions, which include the normal distribution and to some extent share its desirable properties, have played a crucial role in data analysis.

It has been observed in various practical applications that data do not conform to the normal distribution, which is symmetric with no skewness. The skew normal distribution proposed by Azzalini (1985) is appropriate for the analysis of data which is unimodal but exhibits some skewness. The skew normal distribution includes the normal distribution as a special case where the skewness parameter is zero. The skew normal family of probability distributions is a fairly recent family of distribution that attracted wide attention in the literature due to its strict inclusion of the normal distribution, its mathematical tractability and because it reproduces some properties of the normal distribution. Since the normal distribution is still the most commonly used distribution both in statistical theory and applications, a family of distributions that possesses the same properties have a great potential impact in theoretical and applied probability and statistics. However, this potential impact, there are still relatively few statisticians who use this family in their theoretical and applied works. The main reason of this potential impact that research on characterization and statistical inference for this family is still in its early stage.

In this thesis we will review the statistical properties, goodness of fit, and the estimation problem for unknown parameter of the skew normal distribution. We will introduce a new approximate form to the skew normal distribution and its cumulative distribution function. This approximate is easier for calculating, convenient, mathematically tractable and will be in a closed form and particularly useful when the probability density function occurs in an expression to be used for further mathematical derivation or in programs for them. Also, we will introduce mean and the variance of the new approximation. Finally, we will provide a comparison between the exact and the approximate form of the probability density function of the skew normal distribution. A numerical comparison between the exact and the approximate to cumulative distribution function of the skew normal distribution will be carried.

This thesis organized as follows. Chapter 2, contains some important definitions and notation. Chapter 3, contains properties of the two parameters generalized skew normal distribution and the univriate skew normal distribution with the estimation of the unknown parameters the goodness of fit test for the skew normal distribution will be reviewed. Chapter 4, contains a new approximate form for the skew normal distribution and its cumulative distribution function. Also, we will introduce a numerical comparison between the exact and the approximation to the cumulative distribution function of the skew normal distribution. Finally, all numerical results are carried out using the Mathcad(2001) package. The programs, tables and figures are listed in appendix.

Chapter II

Definitions and Notation

This chapter is concerned with some definitions and notation which will be needed in the next chapters.

2.1 Some Methods of Estimation

The theory of estimation consists of these methods by which are make inference or generalization about a population parameter. The trend of today is to distinguish between the non Bayesian and Bayesian method of estimation which utilize prior subjective knowledge about the probability distribution of the unknown parameters in conjunction with information provided by the sample data.

Let Ω denoted the set of all possible values that the unknown parameter θ could assume which called the parameter space. If θ is a vector, then Ω will be a subset of a Euclidean space of the same distribution and the dimension of Ω will correspond to the number of unknown real parameters. Suppose that we want to estimate the population parameter θ there are two distinct forms. First, to obtain a single value that should be close to the unknown population parameter θ this type of estimation is called point estimation. Second, it might say that θ will fall between two numbers which are intended to enclose the parameter of interest this second type of estimation is called interval estimation.

A point estimation procedure utilizes the information in the sample to arrive at a single value or point that estimates the target parameter. A function of random sample, $T(\mathrm{x})=t(x_1,x_2,\ldots,x_n)$, that does not depend on any unknown parameters is called a statistic. A statistic, $T(\mathrm{x})=t(x_1,x_2,\ldots,x_n)$, that is used to estimate the value of θ is called an estimator of θ, and an observed value $t(x_1,x_2,\ldots,x_n)$ is called an estimate of θ. An approximate value of a population parameter on the basic sample statistic is called an estimator of the population parameter and it depends on the random sample. Frequently the distribution of observation x depends not only upon a set of k parameters $\Theta_1=(\theta_1,\theta_2,\ldots,\theta_k)$ of interest, but also on a set of, say (m-k) further nuisance parameters $\Theta_2 = (\theta_{k+1}, \theta_{k+2}, ..., \theta_{k+m})$. Thus to make inference about the parameters $\Theta = \Theta_1$ with unknown parameter Θ_2. Here the parameter of interest, or inference parameter, is $\Theta = \Theta_1$, while the nuisance of identical parameter is Θ_2. Let θ be an unknown real parameter taking values in a parameter space Ω. A real valued statistic T that is used to estimate θ is called an estimator of θ. Thus, the estimator is a random variable and hence has distribution, a mean, a variance, and so on. The bias is the expected value of the difference between the estimator and the parameter.

$$\text{Bias}(T) = E(T - \theta) = E(T) - \theta .$$

(i) Unbiased Estimator

Definition (2.1): The estimator T is said to be unbiased if the bias is zero for value of θ, equivalently if the expected value of the estimator is the parameter being estimated

$$\text{i.e.} \qquad E(T) = \theta \qquad \text{for all } \theta \in \Omega$$

The goodness of the estimator is usually measured by computing the mean square error as follows

$$\text{MSE}(T) = \text{Var}(T) + (\text{bias}(T))^2 .$$

In particular if the estimator is unbiased, then the mean square error of T is simply the variance of T. Ideally it would like to have unbiased estimate with small mean square error. However, if there exist two unbiased estimator of θ, denoted T_1 and T_2, we naturally prefer the one with the smallest variance. Comparison involving the variances of estimators are after used to decide which method makes more efficient use of the data.

(ii) Uniformly Minimum Variance Unbiased Estimator

Definition (2.2): An estimator T^* of θ is called UMVUE if

1. T^* is unbiased estimator of θ and
2. For any other unbiased estimator T of θ,

$$\text{Var}(T^*) \leq \text{Var}(T) \qquad \text{for all } \theta \in \Omega .$$

(iii) Efficiency

Definition (2.3): The relative efficiency of an unbiased estimator T_1 of θ to another unbiased estimator T_2 of θ is given by

$$re(T_1, T_2) = \frac{\text{var}(T_2)}{\text{var}(T_1)}.$$

An unbiased estimator T_2 of θ is said to be efficient if $re(T_1,T_2) \leq 1$ for all estimator T_1 of θ, and $\theta \in \Omega$. Notice that an efficient estimator is just a uniformly variance unbiased estimator UVUE.

In some cases it is possible to show that, in a certain sense, a particular statistic or set of statistics contain all the information in the sample concerning the parameters. It would then be reasonable to restrict attention to some definitions about this idea such as sufficient and completeness.

(iv) Sufficient Statistic

Definition (2.4): A statistic T is sufficient for θ if T contains all of the information about θ that is available in the entire data variable X, formally, T is sufficient for θ if the conditional distribution of X given T does not depend on θ.

(v) Complete Statistic

Definition (2.5): A statistic T is complete if:

$$E[K(T) \mid \theta] = 0 \qquad \text{for all } \theta \in \Omega$$

$$P[K(T) = 0 \mid \theta] = 1 \qquad \text{for all } \theta \in \Omega$$

where $K(T)$ is a function of statistic T.

(vi) Ancillary Statistics

Definition (2.6): A statistics whose distribution does not depend on the unknown parameter is called ancillary statistics.

(vii) Equivariant Estimators

Definition (2.7): If (μ^*, σ^*) are two estimators based on independent and identically