NAME OF AUTHOR/*NOM DE L'AUTEUR* TZER-LIN CHEN

TITLE OF THESIS/*TITRE DE LA THÈSE* NORMAL APPROXIMATIONS TO POSTERIOR DISTRIBUTIONS

UNIVERSITY/*UNIVERSITÉ* SIMON FRASER UNIVERSITY

DEGREE FOR WHICH THESIS WAS PRESENTED/ *GRADE POUR LEQUEL CETTE THÈSE FUT PRÉSENTÉE* MASTER OF SCIENCE

YEAR THIS DEGREE CONFERRED/*ANNÉE D'OBTENTION DE CE GRADE* 1978

NAME OF SUPERVISOR/*NOM DU DIRECTEUR DE THÈSE* PROFESSOR C. VILLEGAS

DATED/*DATÉ* APRIL 21, 1978 SIGNED/*SIGNÉ*

PERMANENT ADDRESS/*RÉSIDENCE FI*[redacted]

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

NORMAL APPROXIMATIONS TO POSTERIOR DISTRIBUTIONS

by

Tzer-Lin Chen

M. A., Villanova University, 1972

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in the Department

of

Mathematics

SIMON FRASER UNIVERSITY

February 1978

APPROVAL

Name: Tzer-Lin Chen

Degree: Master of Science

Title of Thesis: Normal Approximations to Posterior Distributions

Examining Committee:

Chairperson: Dr. A. H. Lachlan

Professor C. Villegas
Senior Supervisor

Dr. G. Bojadziev

Dr. D. M. Eaves

Dr. P. de Jong
External Examiner

Date Approved: April 20, 1978

PARTIAL COPYRIGHT LICENSE

Title of Thesis/Dissertation:

NORMAL APPROXIMATIONS TO POSTERIOR DISTRIBUTIONS

Author: ______________________
(signature)

TZER-LIN CHEN
(name)

April 21, 1978
(date)

ABSTRACT

Scheffé's Theorem and the Lebesgue Dominated Convergence Theorem are important tools in the derivation of normal approximations to posterior distributions. DeGroot first introduced the definition of supercontinuity and used the definition to provide normal approximations to posterior distribution in the case of one unknown parameter. The purpose of this thesis is to review an alternative approach, based on the lectures of Professor C. Villegas, which is applicable whenever the relative likelihood function has a special form.

The thesis is divided into five chapters, the first presenting the basic theory.

In Chapter 2 we give an alternative proof of Scheffé's Theorem and an alternative proof of the generalization of the Lebesgue Dominated Convergence Theorem. Also we offer a counterexample to the converse of Scheffé's Theorem.

In Chapter 3 we discuss a concept of supercontinuity that was introduced by DeGroot to derive normal approximations to posterior distributions. We offer four propositions which survey the basic properties of supercontinuity. We also illustrate DeGroot's method by five practical example.

In Chapter 4 we review an alternative approach based on the lectures of Professor C. Villegas. We also elucidate this alternative approach by considering five practical examples.

Finally in Chapter 5 we consider some problems which arise in the application of the above-mentioned methods in time series analysis. We illustrate the normal approximation to the distributions of the parameters of autoregressive processes by considering one practical example.

ACKNOWLEDGMENT

I would like to express my appreciation to Professor C. Villegas and Dr. P. de Jong for their suggestions and help during the writing of this thesis. I would also like to thank Mrs. Nancy Hood for typing the entire work.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

If a random variable x has a probability density $f(x \mid \alpha)$ which depends on a single unknown real valued parameter α, and $\Pi(\alpha)$ is the prior density for α, then the posterior density of α given x is proportional to $f(x \mid \alpha)\Pi(\alpha)$. In other words, the posterior density is proportional to the product of the likelihood function and the prior density.

Normal approximations to posterior distributions are available in many important cases. In the derivation of these approximations, DeGroot (1970) uses a concept of supercontinuity. In Chapter 3 we review DeGroot's approach and discuss the concept of supercontinuity.

In Chapter 4 we review an alternative approach based on the lectures of Professor C. Villegas, which is applicable whenever the likelihood function has a simple special form. In both Chapters 3 and 4 the methods are illustrated by several examples.

When the observed values constitute an autoregressive process, additional problems arise because of the lack of independence. Some of these problems are discussed in Chapter 5.

The following definitions are used throughout this thesis.

<u>Definition 1.1.1</u>. The function L_x defined by

$$L_x(\alpha) = f(x \mid \alpha) \tag{1.1}$$

is called the likelihood function, and its value for any α is called the likelihood of α.

Definition 1.1.2. Let α denote an arbitrary parameter and let the likelihood function be such that $L_x(\alpha) = f(x \mid \alpha)$. If $\tilde{\alpha}$ is a possible value of α and $\hat{\alpha}$ is the maximum likelihood estimator, then the likelihood ratio

$$(1.2) \qquad \frac{L(\tilde{\alpha})}{L(\hat{\alpha})}$$

is called the relative likelihood of $\tilde{\alpha}$.

CHAPTER II

WEAK CONVERGENCE

2.1 Introduction

The purpose of this chapter is to present a new proposition which gives an alternative proof of Scheffé's Theorem and an alternative proof of the generalization of the Lebesgue Dominated Convergence Theorem.

The chapter is divided into three sections, the first presenting the introduction.

The second section presents a new proposition on the General Lebesgue Integral and explores the relation of this to the generalization of the Lebesgue Dominated Convergence Theorem.

The last section gives an alternative proof of Scheffé's Theorem by means of the new proposition and offers an original counterexample to the converse of Scheffé's Theorem.

2.2 A Proposition on the General Lebesgue Integral

In order to give an alternative proof of Scheffé's Theorem, we find the following new proposition by using the generalization of the Lebesgue Dominated Convergence Theorem. This proposition also gives an alternative proof of the generalization of the Lebesgue Dominated Convergence Theorem.

Proposition 2.2.1. If

(i) $f_n \to f,\ g_n \to g,\ G_n \to G$ a.e.

(ii) $g_n \leqq f_n \leqq G_n$ for all n a.e.

(iii) $\int g_n \to \int g$ and $\int G_n \to \int G$ with $\int g$ and $\int G$ finite,

then $\int f_n \to \int f$ and $\int f$ is finite.

Proof

$$0 \leqq f_n - g_n \Rightarrow 0 \leqq f - g \quad \text{and} \quad 0 \leqq G_n - f_n \Rightarrow 0 \leqq G - f \qquad \text{a.e.}$$

From Fatou's lemma and $\limsup(-X_n) = -\liminf X_n$

$$\liminf X_n + \liminf Y_n \leqq \liminf(X_n + Y_n) \leqq \liminf X_n + \limsup Y_n$$

$$\int G - \int f = \int \liminf(G_n - f_n) \leqq \liminf \int (G_n - f_n)$$

$$= \liminf(\int G_n + \int - f_n)$$

$$\leqq \limsup \int G_n + \liminf \int - f_n = \int G - \limsup \int f_n .$$

Hence

(2.1) $\limsup \int f_n \leqq \int f.$

Similarly,

(2.2) $\int f \leqq \liminf \int f_n .$

From (2.1) and (2.2)

$$\limsup \int f_n \leqq \int f \leqq \liminf \int f_n .$$

Thus

$$\int f_n \to \int f \quad \text{and} \quad \int f \quad \text{is finite.}$$

Theorem 2.2.1 (Generalization of the Lebesgue Dominated Convergence Theorem). Let $\{g_n'\}$ be a sequence of integrable functions which converge almost everywhere to an integrable function g'. Let $\{f_n\}$ be a sequence of measurable functions such that $|f_n| \leqq g_n'$ and $\{f_n\}$ converges to f almost everywhere. If

$$\int g' = \lim \int g_n'$$

then

$$\int f = \lim \int f_n .$$

For the standard proof see Royden (1968, p. 89). The proof that follows utilizes Proposition 2.2.1.

Suppose the condition of Proposition 2.2.1 holds and let $\{g_n'\}$ be a sequence of integrable functions which converge almost everywhere to an integrable function g'. Let $\{f_n\}$ be a sequence of measurable functions such that $|f_n| \leqq g_n'$ and $\{f_n\}$ converges to f almost everywhere.

If $\int g' = \lim \int g_n'$ then ~~$\int g_n' \to \int g'$ and $\int g'$ is finite.~~

We define G_n and g_n by $G_n = -g_n = g_n'$. Hence

$g_n \leqq f_n \leqq G_n$ for all n a.e.

$$\int g_n = \int -g_n' \to \int -g' \text{ and } \int -g' \text{ is finite.}$$

$$\int G_n = \int g_n' \to \int g' \text{ and } \int g' \text{ is finite.}$$

Now from Proposition 2.2.1

$$\int f_n \to \int f \text{ and } \int f \text{ is finite then } \int f = \lim \int f_n$$

which completes the proof.

2.3 Convergence in Distribution

We apply both the notion of convergence in distribution and Scheffe's Theorem in Chapter 3, Chapter 4 and Chapter 5. This section gives an alternative proof of Scheffé's Theorem by means of Proposition 2.2.1. This section also offers an original counterexample to the converse of Scheffé's Theorem.

Definition 2.3.1. A sequence $\{F_n\}$ of distribution functions is said to converge weakly to F if

$$F_n(x) \to F(x)$$

at all continuity points x of F.

Definition 2.3.2. We say a sequence $\{X_n\}$ of random variables converges in distribution to the random variable X, and we write

$$X_n \overset{\mathcal{D}}{\to} X$$

if the distributions F_n of the X_n converge weakly to the distribution F of X.

Theorem 2.3.1. A sequence $\{F_n\}$ of distribution functions converges weakly to F if and only if for every bounded continuous real function h

$$\int hdF_n \to \int hdF.$$

Proof. See Kingman (1966, pp. 315-317).

Theorem 2.3.2. If a sequence $\{X_n\}$ of random variables converges in distribution to the random variable X and if f is a continuous function, then $\{f(X_n)\}$ converges in distribution to $f(X)$.

Proof. See Billingsley (1968, p. 31).

Theorem 2.3.3 (Scheffé's Theorem). If $\{f_n\}$ is a sequence of probability densities in a Euclidean space and if $f_n \to f$ a.e., then the sequence $\{F_n\}$ of distribution functions converges weakly to F.

Proof.

$$f_n \geqq 0, \quad f \geqq 0$$

$$\int f_n = 1 \quad \text{and} \quad \int f = 1$$

$$|f_n - f| \to 0 \qquad \text{a.e.}$$

$$0 \leqq |f_n - f| \leqq |f_n| + |f| = f_n + f \to 2f \qquad \text{a.e.}$$

$$\int 0 \to \int 0 = 0 \quad \text{is finite}$$

$$\int (f_n + f) \to \int 2f = 2 \quad \text{is finite.}$$

It follows from Proposition 2.2.1 that

$$\int |f_n - f| \to \int 0 = 0.$$

Let h be any bounded continuous real-valued function on the Euclidean space on which the f_n and f are defined.

There exists M such that $h(x) \leqq M$ for every x.

$$\left|\int hf_n - \int hf\right| = \left|\int h(f_n - f)\right| \leqq \int |h(f_n - f)|$$

$$\leqq M \int |f_n - f| \to M \cdot 0 = 0.$$

Hence

$$\left|\int hf_n - \int hf\right| \to 0.$$

Thus

$$\int hf_n \to \int hf.$$

It follows from Theorem 2.3.1 that $\{F_n\}$ converges weakly to F.

<u>Remark 2.3.1.</u> The converse of Theorem 2.3.3 is not true, as is shown by the following counterexample

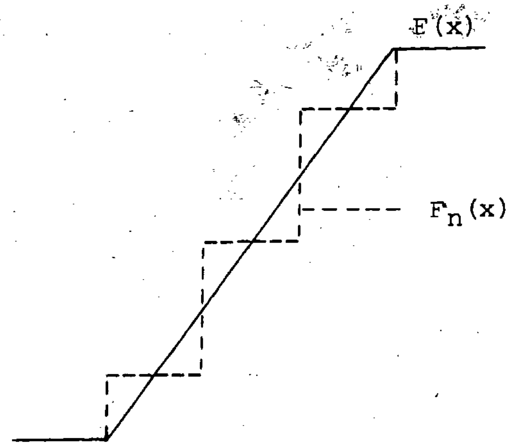

Figure 2.1

In Figure 2.1, $F_n(x) \to F(x)$, the derivatives of F_n at points except jump points are zero, but the derivatives of F_n at jump points do not exist. In words, a sequence of distribution functions without probability densities can converge weakly to a distribution function with a probability density.

CHAPTER III

DEGROOT'S NORMAL APPROXIMATIONS TO POSTERIOR DISTRIBUTIONS

3.1 Introduction

The definition of supercontinuity was first introduced by DeGroot (1970). If α_0 is an unknown parameter, observations $X_1, \ldots, X_n$ form a large random sample from the probability density $f(\cdot \mid \alpha_0)$, the prior density of α is positive and continuous in the neighbourhood, supercontinuity assumptions and regularity assumptions are satisfied, DeGroot (1970) has shown the validity of normal approximation to the posterior distribution of α when α is scalar.

The purpose of this chapter is to present four propositions which scrutinize the basic properties of supercontinuity and to elucidate DeGroot's normal approximations to the posterior distributions of one unknown parameter by studying five practical examples.

The chapter is divided into five sections, the first presenting the introduction. The second section offers four new propositions which survey the basic properties of supercontinuity. The third section deals with solutions of the likelihood equation when the observations $X_1, \ldots, X_n$ form a large random sample from the probability density $f(\cdot \mid \alpha_0)$. Section 4 illustrates DeGroot's normal approximations to the posterior distributions in the case of one unknown parameter. The last section studies five practical examples which explain DeGroot's normal approximations to the posterior distributions in the case of one unknown parameter.

3.2 Supercontinuity

DeGroot (1970) first introduced the definition of supercontinuity. Supercontinuity is equivalent to the ordinary continuity under a condition which will be illustrated in Remark 3.2.1. This section also presents four new propositions in supercontinuity. These propositions survey the basic properties of supercontinuity.

Definition 3.2.1. Following DeGroot (1970) a real-valued function g is *supercontinuous* at the value $\alpha_0 \in \Omega$ if

$$\lim_{\delta \to 0} E\left[\sup_{\alpha \in N(\alpha_0,\delta)} |g(X,\alpha) - g(X,\alpha_0)|\right] = 0$$

and $g(X,\alpha)$ is specified at every point (x,α) of the product space $S \times \Omega$, where S is the sample space of a single observation X, Ω is an open interval of the real line and $N(\alpha_0,\delta)$ is the interval around α_0 containing every point in Ω whose distance from α_0 is less than δ, and the expectation is computed assuming that the random variable X has the distribution indexed by α_0.

Remark 3.2.1. If $E[\sup_{\alpha \in N(\alpha_0,\delta)} |g(x,\alpha) - g(x,\alpha_0)|]$ exists for all sufficiently small values of δ, then supercontinuity of the function g at α_0 is equivalent to ordinary continuity of the function $g(x,\cdot)$ at α_0 for all values of $x \in S$ except on a subset T of S whose probability is 0.

The following four new propositions survey the basic properties of supercontinuity.

Proposition 3.2.1. If g and h are supercontinuous at the value $\alpha_0 \in \Omega$, then $g + h$ is supercontinuous at the value $\alpha_0 \in \Omega$.

Proof.

$$0 \leq \lim_{\delta \to 0} E[\sup_{\alpha \in N(\alpha_0,\delta)} |(g+h)(x,\alpha) - (g+h)(x,\alpha_0)|]$$

$$\leq \lim_{\delta \to 0} E[\sup_{\alpha \in N(\alpha_0,\delta)} [|g(x,\alpha) - g(x,\alpha_0)| + |h(x,\alpha) - h(x,\alpha_0)|]]$$

$$\leq \lim_{\delta \to 0} E[\sup_{\alpha \in N(\alpha_0,\delta)} |g(x,\alpha) - g(x,\alpha_0)|] + $$

$$+ \lim_{\delta \to 0} E[\sup_{\alpha \in N(\alpha_0,\delta)} |h(x,\alpha) - h(x,\alpha_0)|]$$

$$= 0 + 0 = 0.$$

Hence, $g + h$ is supercontinuous at the value $\alpha_0 \in \Omega$.

Proposition 3.2.2. If g and h are supercontinuous at the value $\alpha_0 \in \Omega$, and there exists $M > 0$ such that $\mathrm{Max}(|g(x,\alpha)|, |h(x,\alpha_0)|) \leq M$ for all values of $\alpha \in N(\alpha_0,\delta)$ and for all values of $x \in S$, then gh is supercontinuous at the value $\alpha_0 \in \Omega$.

Proof.

$$0 \leq \lim_{\delta \to 0} E[\sup_{\alpha \in N(\alpha_0,\delta)} |g(x,\alpha)h(x,\alpha) - g(x,\alpha_0)h(x,\alpha_0)|]$$

$$\leqq \lim_{\delta\to 0} E[\sup_{\alpha\in N(\alpha_0,\delta)} |g(x,\alpha)||h(x,\alpha) - h(x,\alpha_0)|$$

$$+ \sup_{\alpha\in N(\alpha_0,\delta)} |h(x,\alpha_0)||g(x,\alpha) - g(x,\alpha_0)|]$$

$$\leqq M\left\{\lim_{\delta\to 0} E[\sup_{\alpha\in N(\alpha_0,\delta)} |h(x,\alpha) - h(x,\alpha_0)|\right.$$

$$\left.+ \lim_{\delta\to 0} E[\sup_{\alpha\in N(\alpha_0,\delta)} |g(x,\alpha) - g(x,\alpha_0)|]\right\}$$

$$= M\{0 + 0\} = 0.$$

Hence, gh is supercontinuous at the value $\alpha_0 \in \Omega$.

Proposition 3.2.3. If h is supercontinuous at the value $\alpha_0 \in \Omega$, and there exists $K > 0$ such that $|h(x,\alpha)| \geqq K$ for all values of $\alpha \in N(\alpha_0,\delta)$ and for all values of $x \in S$, then $\frac{1}{h}$ is supercontinuous at the value $\alpha_0 \in \Omega$.

Proof.

$$0 \leqq \lim_{\delta\to 0} E[\sup_{\alpha\in N(\alpha_0,\delta)} \left|\frac{1}{h(x,\alpha)} - \frac{1}{h(x,\alpha_0)}\right|]$$

$$\leqq \frac{1}{K^2} \lim_{\delta\to 0} E[\sup_{\alpha\in N(\alpha_0,\delta)} |h(x,\alpha) - h(x,\alpha_0)|]$$

$$= \frac{\{0\}}{K^2} = 0.$$

Hence, $\frac{1}{h}$ is supercontinuous at the value $\alpha_0 \in \Omega$.

Remark 3.2.2. Cramer (1946, pp. 67-68) has stated the following result: If for almost every value of $x \in S$ and for a fixed value α_0 of α, the following conditions are satisfied:

(i) The partial derivative $\left\{\frac{\partial g(x,\alpha)}{\partial \alpha}\right\}_{\alpha=\alpha_0}$ exists,

(ii) we have $\left\{\left|\frac{g(x,\alpha + h) - g(x,\alpha)}{h}\right|\right\}_{\alpha=\alpha_0} < G(x)$ for

$0 < |h| < h_0$, where h_0 is independent of x, and G is an integrable function over S with respect to probability density $f(x \mid \alpha_0)$ and S is the sample space, then

$$\left\{\frac{\partial}{\partial \alpha} \int_S g(x,\alpha) f(x \mid \alpha_0) d\mu(x)\right\}_{\alpha=\alpha_0} = \left\{\int_S \frac{\partial g(x,\alpha)}{\partial \alpha} f(x \mid \alpha_0) d\mu(x)\right\}_{\alpha=\alpha_0}.$$

Proposition 3.2.4. If (ii) in Remark 3.2.2 is true, then g is supercontinuous at the value α_0.

Proof.

$$|g(x,\alpha) - g(x,\alpha_0)| < |\alpha - \alpha_0| G(x)$$

$$\sup_{\alpha \in N(\alpha_0,h)} |g(x,\alpha) - g(x,\alpha_0)| < \sup_{\alpha \in N(\alpha_0,h)} |\alpha - \alpha_0| G(x) = hG(x)$$

$$E\Big[\sup_{\alpha\in N(\alpha_0,h)} |g(x,\alpha) - g(x,\alpha_0)|\Big] < hE[G(x)]$$

$$\lim_{h\to 0} E\Big[\sup_{\alpha\in N(\alpha_0,h)} |g(x,\alpha) - g(x,\alpha_0)|\Big] = 0.$$

Hence, g is supercontinuous at the value α_0 .

3.3 Solutions of the Likelihood Equation

We now proceed to find the solutions of the likelihood equation when the observations $X_1, \ldots, X_n$ form a large random sample from the probability density $f(\cdot \mid \alpha_0)$.

We first make the following assumptions (primes indicate differentiation with respect to α only):

Assumption L_1 . The second-order derivatives $f''(x \mid \alpha)$ exist for all values of $x \in S$ and all values of α in some neighbourhood $N(\alpha_0, \delta)$ of α_0 , where $\delta > 0$.

Assumption L_2 . $\int_S f'(x \mid \alpha_0) d\mu(x) = 0$ and $\int_S f''(x \mid \alpha_0) d\mu(x) = 0$.

Assumption L_3 . For any values of $x \in S$ and $\alpha \in N(\alpha_0, \delta)$, it is assumed that $f(x \mid \alpha) > 0$. Furthermore, if $\lambda(x \mid \alpha)$ is defined by

(3.1) $\lambda(x \mid \alpha) = \log f(x \mid \alpha)$

it is assumed that functions λ, λ' and λ'' are supercontinuous at

α_0, and for all values of $\alpha \in N(\alpha_0,\delta)$, the expectations $E[\lambda(x \mid \alpha)]$, $E[\lambda'(x \mid \alpha)]$ and $E[\lambda''(x \mid \alpha)]$ are finite, where

$$\lambda'(x \mid \alpha) = \frac{\partial\lambda(x \mid \alpha)}{\partial\alpha}, \quad \lambda''(x \mid \alpha) = \frac{\partial^2\lambda(x \mid \alpha)}{\partial\alpha^2}.$$

Assumption L_4. For any values $\alpha \in N(\alpha_0,\delta)$, let $I(\alpha)$ be defined as follows:

$$I(\alpha) = \int_S [\lambda'(x \mid \alpha)]^2 f(x \mid \alpha)d\mu(x). \tag{3.2}$$

The function I is positive and continuous throughout the neighbourhood $N(\alpha_0,\delta)$.

Theorem 3.3.1. If α_0 is an unknown parameter, observations $X_1, \ldots, X_n$ form a random sample from the probability density $f(\cdot \mid \alpha_0)$, and Assumptions L_1 to L_4 are satisfied, then, with probability 1, there will exist an integer n_0 such that for each value of $n \geqq n_0$, the likelihood equation

$$\sum_{i=1}^{n} \lambda'(x_i \mid \alpha) = 0 \tag{3.3}$$

where $\lambda(x \mid \alpha) = \log f(x \mid \alpha)$

has a solution $\alpha = \hat{\alpha}_n(X_1, \ldots, X_n)$ and $\lim_{n\to\infty} \hat{\alpha}_n(X_1, \ldots, X_n) = \alpha_0$.

Proof. See DeGroot (1970, pp. 209-210).

3.4 Normal Approximations to Posterior Distributions

If α_0 is an unknown parameter, observations $X_1, \ldots, X_n$ form a large random sample from the probability density $f(\cdot \mid \alpha_0)$, the prior density of α is positive and continuous in the neighbourhood $N(\alpha_0, \delta)$, and Assumptions L_1 to L_4 are satisfied, DeGroot (1970) has shown that the posterior distribution of α is approximately a normal distribution with mean $\hat{\alpha}_n$ and variance $\frac{1}{nI(\hat{\alpha}_n)}$, where $\hat{\alpha}_n(X_1, \ldots, X_n)$ denote solutions of the likelihood equation (3.3) and $\lim_{n\to\infty} \hat{\alpha}_n(X_1, \ldots, X_n) = \alpha_0$.

3.5 Examples

This section illustrates DeGroot's normal approximations to the posterior distributions by considering five practical examples.

Example 3.5.1. Suppose that $X_1, \ldots, X_n$ is a random sample from a Bernoulli distribution with an unknown value of the parameter α $(0 < \alpha < 1)$.

It can be checked that Assumptions L_1 to L_4 are satisfied and

$$\lambda'(x \mid \alpha) = \frac{x}{\alpha} - \frac{1-x}{1-\alpha}, \quad I(\alpha) = \frac{1}{\alpha(1-\alpha)}.$$

It also can be checked that the normal approximation to the posterior distribution of α is a normal distribution with mean $\bar{x}$

and variance $\frac{\bar{x}(1 - \bar{x})}{n}$.

Example 3.5.2. Suppose that $X_1, \ldots, X_n$ is a random sample from a binomial distribution with an unknown value of the parameter α $(0 < \alpha < 1)$

$$f(x \mid \alpha) = \binom{m}{x}\alpha^x (1 - \alpha)^{m-x} .$$

It can be checked that Assumptions L_1 to L_4 are satisfied and

$$\lambda'(x \mid \alpha) = \frac{x}{\alpha} - \frac{(m - x)}{1 - \alpha}, \quad I(\alpha) = \frac{m}{\alpha(1 - \alpha)} .$$

It also can be checked that the normal approximation to the posterior distribution of α is a normal distribution with mean $\frac{\bar{x}}{m}$ and variance $\frac{\bar{x}(m - \bar{x})}{nm^3}$.

Example 3.5.3. Suppose that $X_1, \ldots, X_n$ is a random sample from a Poisson distribution with an unknown value of the mean α $(\alpha > 0)$.

It can be checked that Assumptions L_1 to L_4 are satisfied and

$$\lambda'(x \mid \alpha) = \frac{x - \alpha}{\alpha}, \quad I(\alpha) = \frac{1}{\alpha} .$$

It also can be checked that the normal approximation to the posterior distribution of α is a normal distribution with mean $\bar{x}$ and variance $\frac{\bar{x}}{n}$.

Example 3.5.4. Suppose that $X_1, \ldots, X_n$ is a random sample from a normal distribution with an unknown value of the mean α $(-\infty < \alpha < \infty)$ and a specified value of the variance σ^2.

It can be checked that Assumptions L_1 to L_4 are satisfied and

$$\lambda'(x \mid \alpha) = \frac{(x - \alpha)}{\sigma^2}, \quad I(\alpha) = \frac{1}{\sigma^2}.$$

It also can be checked that the normal approximation to the posterior distribution of α is a normal distribution with mean $\bar{x}$ and variance $\frac{\sigma^2}{n}$. This is also the posterior distribution itself if the prior for α is uniform.

Example 3.5.5. Suppose that $X_1, \ldots, X_n$ is a random sample from a normal distribution with a specified value of the mean μ and an unknown value of the standard deviation α $(\alpha > 0)$.

It can be checked that Assumptions L_1 to L_4 are satisfied and

$$\lambda'(x \mid \alpha) = \frac{(x - \mu)^2}{\alpha^3} - \frac{1}{\alpha}, \quad I(\alpha) = \frac{2}{\alpha^2}.$$

It also can be checked that the normal approximation to the posterior distribution of α is a normal distribution with mean s and variance $\frac{s^2}{2n}$, where s is the sample standard deviation.

3.6 Conclusions

In this chapter we have presented four propositions which scrutinize the basic properties of supercontinuity and we elucidated DeGroot's normal approximations to the posterior distribution of one unknown parameter by studying five practical examples.

CHAPTER IV

AN ALTERNATIVE APPROACH

4.1 Introduction

In the present chapter we review an alternative approach to the problem of finding normal approximations to posterior distributions that was developed in the lectures of Professor C. Villegas. We assume that the unknown parameter $\underline{\alpha}$ is a column vector belonging to a known open set Ω in a p-dimensional space. The basic assumption is that, for sufficiently large n, the relative likelihood function has the special form $\exp[-n\varphi(\underline{\tilde{\alpha}},\underline{u}_n(X_1, \ldots, X_n))]$, where φ is a known function of two arguments, a parameter $\underline{\tilde{\alpha}} \in \Omega$ and a vector $\underline{u}_n$ which is a known function of the observations. We assume also that, with probability 1, there is, for a sufficiently large sample size n, a uniquely defined maximum likelihood estimate $\underline{\hat{\alpha}}$ which converges to the true parameter $\underline{\alpha}$ when n increases indefinitely. Finally, we assume that φ is a twice-differentiable function of $\underline{\tilde{\alpha}}$ when $X_1, \ldots, X_n$ are fixed, and that, with probability 1, $\underline{u}_n(X_1, \ldots, X_n)$ converges to an unknown vector $\underline{u}$.

The purpose of this chapter is to review an alternative approach, under Professor C. Villegas's assumptions. These normal approximations to posterior distributions are obtained by applying the Taylor expansion technique, the strong law of large numbers and

Scheffé's Theorem respectively to a special form of relative likelihood. In the alternative approach we do not apply DeGroot's supercontinuity notion.

This chapter is divided into three sections, the first presenting the introduction. The second section reviews an alternative approach of normal approximations to posterior distributions under Professor C. Villegas's assumptions. The last section elucidates these normal approximations to posterior distributions by considering five practical examples.

4.2 Normal Approximations to Posterior Distributions

Suppose these assumptions in 4.1 hold. It follows from Definition 1.1.2 that

$$\exp[-\, n\varphi(\underline{\hat{\alpha}},\underline{u_n}(X_1, \ldots, X_n))] = 1.$$

Hence

$$\varphi(\underline{\hat{\alpha}},\underline{u_n}(X_1, \ldots, X_n)) = 0.$$

Maximizing $\exp[-\, n\varphi(\underline{\tilde{\alpha}},\underline{u_n}(X_1, \ldots, X_n))]$ is equivalent to minimizing $\varphi(\underline{\tilde{\alpha}},\underline{u_n}(X_1, \ldots, X_n))$, it follows that

$$\varphi'(\underline{\hat{\alpha}},\underline{u_n}(X_1, \ldots, X_n)) = \underline{0},$$

where $\varphi'(\underline{\tilde{\alpha}},\underline{u_n}(X_1, \ldots, X_n)) = \dfrac{\partial\varphi(\underline{\tilde{\alpha}},\underline{u_n}(X_1, \ldots, X_n))}{\partial\underline{\tilde{\alpha}}}$

A second-order Taylor expansion gives, assuming that $\underline{\tilde{\alpha}}$ and $\underline{\hat{\alpha}}$ are column vectors,

$$\varphi(\underline{\tilde{\alpha}},\underline{u_n}(X_1, \ldots ,X_n)) = (\underline{\tilde{\alpha}} - \underline{\hat{\alpha}})'\phi_n(\underline{\tilde{\alpha}} - \underline{\hat{\alpha}})/2.$$

The prime denotes transposition and ϕ_n is the matrix whose i,j-th entry is $\varphi_{ij}(\underline{\hat{\alpha}} + \theta(\underline{\tilde{\alpha}} - \underline{\hat{\alpha}}), \underline{u_n}(X_1, \ldots ,X_n)) = \dfrac{\partial^2\varphi(\underline{\hat{\alpha}} + \theta(\underline{\tilde{\alpha}} - \underline{\hat{\alpha}}),\underline{u_n}(X_1, \ldots ,X_n))}{\partial\alpha_i\partial\alpha_j}$ for some θ in the interval $0 < \theta < 1$.

We define Λ_n to be the upper triangular matrix with positive diagonal elements which is uniquely defined by

$$\Lambda_n'\Lambda_n = \{\varphi_{ij}(\underline{\hat{\alpha}},\underline{u_n}(X_1, \ldots ,X_n))\} = \hat{\phi}.$$

For fixed vector $\underline{\tau}$, suppose that $\underline{\tilde{\alpha}}$ is specified by the equation

$$\sqrt{n}\Lambda_n(\underline{\tilde{\alpha}} - \underline{\hat{\alpha}}) = \underline{\tau}.$$

Hence

$$\sqrt{n}(\underline{\tilde{\alpha}} - \underline{\hat{\alpha}}) = \Lambda_n^{-1}\underline{\tau}\ , \quad \sqrt{n}(\underline{\tilde{\alpha}} - \underline{\hat{\alpha}})' = \underline{\tau}'\Lambda_n'^{-1}.$$

Let $\Pi(\underline{\alpha})$ be the prior for the parameter $\underline{\alpha}$. Since the Jacobian $J = \left|\{\frac{\partial\underline{\tilde{\alpha}}}{\partial\underline{\tau}}\}\right|$ is constant, the prior for $\underline{\tau}$ will be proportional to $\Pi(\underline{\hat{\alpha}} + \Lambda_n^{-1}\frac{\underline{\tau}}{\sqrt{n}})$. Assuming that with probability 1, $\underline{\hat{\alpha}} \to \underline{\alpha}$, it follows that the prior density for $\underline{\tau}$ converges to a constant. We also have, with probability 1,

$$\phi_n \rightarrow \phi = \{\varphi_{ij}(\underline{\alpha},\underline{u})\}$$

and

$$\Lambda_n \rightarrow \Lambda \quad \text{where} \quad \Lambda'\Lambda = \phi .$$

Therefore, for any fixed $\underline{\tau}$, $\Lambda_n'^{-1}\phi_n\Lambda_n^{-1}$ converges to the identity matrix with probability 1. Hence

$$\exp[-\, n\varphi(\underline{\tilde{\alpha}},\underline{u_n}(X_1, \ . \ . \ . \ ,X_n))] = \exp\left[-\ \frac{\{n(\underline{\tilde{\alpha}} - \underline{\hat{\alpha}})'\phi_n(\underline{\tilde{\alpha}} - \underline{\hat{\alpha}})\}}{2}\right]$$

$$= \exp\left[-\ \frac{(\underline{\tau}'\Lambda_n'^{-1}\phi_n\Lambda_n^{-1}\underline{\tau})}{2}\right]$$

converges, with probability 1, to

$$\exp\left[-\frac{(\underline{\tau}'I\underline{\tau})}{2}\right] = \exp\left[-\ \frac{(\underline{\tau}'\underline{\tau})}{2}\right]$$

This provides an approximation to the relative likelihood for large values of n.

The normal approximation to the posterior distribution of $\underline{\tilde{\alpha}}$ is thus a normal distribution with mean vector $\underline{\hat{\alpha}}$ and covariance matrix $\frac{\hat{\phi}^{-1}}{n}$.

4.3 Examples

We illustrate the normal approximations to posterior distributions by considering five practical examples.

Example 4.3.1. Suppose that $X_1, \ldots, X_n$ is a random sample from a normal probability distribution.

$$f(x \mid \alpha_1, \alpha_2) = \frac{\left\{\exp\left[-\dfrac{(x-\alpha_1)^2}{2\alpha_2^2}\right]\right\}}{\sqrt{2\pi}\alpha_2}.$$

The maximum likelihood estimates are

$$\hat{\alpha}_1 = \bar{x}, \quad \hat{\alpha}_2 = s.$$

The relative likelihood is

$$R(\alpha_1, \alpha_2 \mid X_1, \ldots, X_n) = \{\exp[\{-n[s^2 + (\bar{x} - \alpha_1)^2]/2\alpha_2^2\} + n/2]\}s^n/\alpha_2^n$$

$$= \exp[-n\varphi(\alpha_1, \alpha_2, \bar{x}, s)],$$

where

$$\varphi(\alpha_1, \alpha_2, \bar{x}, s) = \{[s^2 + (\bar{x} - \alpha_1)^2]/2\alpha_2^2\} - 1/2 - \log s + \log \alpha_2 .$$

It is well known that, with probability 1,

$$\bar{x} \to E[x_i] = \alpha_1$$

$$s \to \alpha_2 .$$

It can be checked that the matrix $\hat{\phi}$ is

$$\hat{\phi} = \begin{bmatrix} \frac{1}{s^2} & 0 \\ 0 & \frac{2}{s^2} \end{bmatrix} .$$

It is obvious that

$$[n\hat{\phi}]^{-1} = \begin{bmatrix} \frac{s^2}{n} & 0 \\ 0 & \frac{s^2}{2n} \end{bmatrix} .$$

It is clear that the relative likelihood is approximately the density of a normal distribution with mean zero and covariance matrix $\frac{\hat{\phi}^{-1}}{n}$. In other words, the normal approximation to the posterior distribution of (α_1,α_2) is a normal distribution with mean vector $(\bar{x},s)$ and covariance matrix $[n\hat{\phi}]^{-1}$.

<u>Example 4.3.2</u>. Suppose that $X_1, \ldots ,X_n$ is a random sample from a Rayleigh distribution

$$f(x \mid \alpha) = \begin{cases} \{\exp[- x^2/2\alpha^2]\}x/\alpha^2 & \text{for } x \geqq 0 \\ 0 & \text{otherwise.} \end{cases}$$

The maximum likelihood estimate is

$$\hat{\alpha} = \left[\frac{\sum_{i=1}^{n} X_i^2}{2n}\right]^{\frac{1}{2}}.$$

It can be checked that

$$\varphi(\alpha, u_n(X_1, \ldots, X_n)) = \left[\frac{u_n(X_1, \ldots, X_n)}{\alpha^2}\right] - 1$$

$$- \log u_n(X_1, \ldots, X_n) + 2\log \alpha$$

where $u_n(X_1, \ldots, X_n) = \left[\frac{\sum_{i=1}^{n} X_i^2}{2n}\right]$.

It is well known that

$$u_n(X_1, \ldots, X_n) = \left[\frac{\sum_{i=1}^{n} X_i^2}{2n}\right] \to \left(\frac{\mathbf{E}[X_i^2]}{2}\right) = \alpha^2$$

with probability 1. It also can be checked that

$$\varphi''(\hat{\alpha}, u_n(X_1, \ldots, X_n)) = 4\hat{\alpha}^{-2}$$

where $\varphi''(\alpha, u_n(X_1, \ldots, X_n)) = \dfrac{\partial^2 \varphi(\alpha, u_n(X_1, \ldots, X_n))}{\partial \alpha^2}$.

It is clear that the relative likelihood is approximately a normal distribution with mean 0 and variance $\dfrac{\hat{\alpha}^2}{4}$. In other words, the normal approximation to the posterior distribution of α is a normal distribution with mean $\hat{\alpha}$ and variance $\dfrac{\hat{\alpha}^2}{4n}$.

Example 4.3.3. Suppose that $X_1, \ldots, X_n$ is a random sample from a Poisson distribution

$$f(x \mid \alpha) = \frac{(\alpha^x e^{-\alpha})}{X!} \; ; \quad X = 0,1,2, \ldots$$

The maximum likelihood estimate is

$$\hat{\alpha} = \bar{x}.$$

It can be checked that

$$\begin{aligned}\varphi(\alpha, u_n(X_1, \ldots, X_n)) = &- u_n(X_1, \ldots, X_n)\log \alpha + \alpha \\ &- u_n(X_1, \ldots, X_n) \\ &+ u_n(X_1, \ldots, X_n)\log u_n(X_1, \ldots, X_n)\end{aligned}$$

where $u_n(X_1, \ . \ . \ . \ , X_n) = \frac{[\sum_{i=1}^{n} X_i]}{n}$.

It is well known that

$$u_n(X_1, \ . \ . \ . \ , X_n) = \frac{[\sum_{i=1}^{n} X_i]}{n} \rightarrow E[X_i] = \alpha \text{ with probability 1.}$$

It also can be checked that

$$\varphi''(\hat{\alpha}, u_n(X_1, \ . \ . \ . \ , X_n)) = \frac{1}{\bar{x}}$$

where $\varphi''(\alpha, u_n(X_1, \ . \ . \ . \ , X_n)) = \frac{\partial^2 \varphi(\alpha, u_n(X_1, \ . \ . \ . \ , X_n)}{\partial \alpha^2}$

It is clear that the relative likelihood is approximately a normal distribution with mean 0 and variance $\bar{x}$. In other words, the normal approximation to the posterior distribution of α is a normal distribution with mean $\bar{x}$ and variance $\frac{\bar{x}}{n}$.

Example 4.3.4. Suppose that $X_1, \ . \ . \ . \ , X_n$ is a random sample from a Bernoulli distribution

$$f(x \mid \alpha) = \alpha^x (1 - \alpha)^{1-x} \quad \text{for } x = 0,1.$$

The maximum likelihood estimate is

$$\hat{\alpha} = \overline{x}.$$

It can be checked that

$$\begin{aligned}\varphi(\alpha, u_n(X_1, \ldots, X_n)) = &- u_n(X_1, \ldots, X_n)\log \alpha \\ &+ [u_n(X_1, \ldots, X_n) - 1]\log(1-\alpha) \\ &+ u_n(X_1, \ldots, X_n)\log u_n(X_1, \ldots, X_n) \\ &+ [1 - u_n(X_1, \ldots, X_n)] \\ &\quad \log[1 - u_n(X_1, \ldots, X_n)]\end{aligned}$$

where $u_n(X_1, \ldots, X_n) = \overline{x}$.

It is well known that

$$u_n(X_1, \ldots, X_n) = \overline{x} \rightarrow E[X_i] = \alpha \quad \text{with probability 1.}$$

It also can be checked that

$$\varphi''(\hat{\alpha}, u_n(X_1, \ldots, X_n)) = \frac{1}{\overline{x}(1-\overline{x})}$$

where $\varphi''(\alpha, u_n(X_1, \ldots, X_n)) = \dfrac{\partial^2 \varphi(\alpha, u_n(X_1, \ldots, X_n))}{\partial \alpha^2}$.

It is clear that the relative likelihood is approximately a normal distribution with mean 0 and variance $\bar{x}(1 - \bar{x})$. In other words, the normal approximation to the posterior distribution of α is a normal distribution with mean $\bar{x}$ and variance $\frac{\bar{x}(1 - \bar{x})}{n}$.

Example 4.3.5. Suppose that $X_1, \ldots, X_n$ is a random sample from an exponential density

$$f(x \mid \alpha) = \begin{cases} \dfrac{\{\exp[-\frac{x}{\alpha}]\}}{\alpha} & \text{for } x \geqq 0 \\ 0 & \text{otherwise.} \end{cases}$$

The maximum likelihood estimate is

$$\hat{\alpha} = \bar{x}.$$

It can be checked that

$$\varphi(\alpha, u_n(X_1, \ldots, X_n)) = \left[\frac{u_n(X_1, \ldots, X_n)}{\alpha}\right] - 1 - \log u_n(X_1, \ldots, X_n) + \log \alpha.$$

where $u_n(X_1, \ldots, X_n) = \bar{x}$.

It is well known that

$$u_n(X_1, \ldots, X_n) = \bar{x} \rightarrow E[X_i] = \alpha \text{ with probability 1.}$$

It also can be checked that

$$\varphi''(\hat{\alpha},u_n(X_1, \ldots ,X_n)) = \frac{1}{\bar{x}^2}$$

where $\varphi''(\alpha,u_n(X_1, \ldots ,X_n)) = \dfrac{\partial^2\varphi(\alpha,u_n(X_1, \ldots ,X_n))}{\partial\alpha^2}$.

It is clear that the relative likelihood is approximately a normal distribution with mean 0 and variance $(\bar{x})^2$. In other words, the normal approximation to the posterior distribution of α is a normal distribution with mean $\bar{x}$ and variance $\dfrac{(\bar{x})^2}{n}$.

4.4 Conclusions

In this chapter we review an alternative approach under Professor C. Villegas's assumptions. In the alternative approach we did not apply DeGroot's supercontinuity notion.

CHAPTER V

NORMAL APPROXIMATIONS TO THE DISTRIBUTIONS OF THE PARAMETERS OF AUTOREGRESSIVE PROCESSES

5.1 Introduction

In this chapter we consider the problem of finding asymptotic approximations to the posterior distributions of the parameters of autoregressive processes. In the examples of Chapter 3 and 4 we applied the ordinary strong law of large numbers, but the ordinary strong law of large numbers does not apply to autoregressive processes, and a version of the strong law of large numbers valid for autoregressive processes should be used instead.

The purpose of this chapter is to elucidate the normal approximations to the distributions of the parameters of autoregressive processes by studying one practical example. The normal approximations to the distributions of the parameters of autoregressive processes are obtained by applying strong law of large numbers to the conditional relative likelihood.

This chapter is divided into three sections, the first introducing the topic. The second section states two strong laws of large numbers. The last section considers one practical example which describes the normal approximations to the distributions of the parameters of autoregressive processes.

5.2 Two Strong Laws of Large Numbers in Autoregressive Processes

Now we shall state two strong laws of large numbers. We shall use the following definition.

Definition 5.2.1. We call the following equation the first-order p-component vector autoregressive process.

$$(5.1) \qquad \underline{X_t} + \underline{\alpha}\, \underline{X_{t-1}} = \underline{V_t}$$

where

$$\underline{X_t} = \begin{bmatrix} X_{1t} \\ X_{2t} \\ \cdot \\ \cdot \\ \cdot \\ X_{pt} \end{bmatrix}, \qquad \underline{V_t} = \begin{bmatrix} V_{1t} \\ V_{2t} \\ \cdot \\ \cdot \\ \cdot \\ V_{pt} \end{bmatrix}$$

$\underline{\alpha}$ is a $p \times p$ matrix of coefficients, $E[\underline{V_t}] = 0$, $E[\underline{V_t V_t'}] = \Sigma$.

Theorem 5.2.1. If $\underline{X_t}$ is defined by (5.1), $t = 1, 2, \ldots,$ with $-\underline{\alpha}$ having eigenvalues less than 1 in absolute value, and if $\underline{V_t}$'s are independently and identically distributed with $E[\underline{V_t}] = 0$ and $E[\underline{V_t V_t'}] = \Sigma$, then with probability 1

$$(5.2) \qquad \lim_{n\to\infty} \frac{[\sum_{t=1}^{n} \underline{X_t}\,\underline{X_t'}]}{n} = \lim_{n\to\infty} \frac{[\sum_{t=1}^{n} \underline{X_{t-1}}\,\underline{X_{t-1}'}]}{n} = E[\underline{X_t}\,\underline{X_t'}] = D$$

Proof. See Anderson (1971, p. 195).

Theorem 5.2.2. Under the conditions of Theorem 5.2.1 and if D is positive definite, then with probability 1

$$(5.3) \qquad \lim_{n\to\infty} \underline{\hat{\alpha}} = \underline{\alpha}$$

$$(5.4) \qquad \lim_{n\to\infty} \hat{\Sigma} = \Sigma$$

where

$$\underline{\hat{\alpha}} = - \frac{[\sum_{t=1}^{n} \underline{X_{t-1}}\underline{X_t'}]}{\sum_{t=1}^{n} \underline{X_{t-1}}\underline{X_{t-1}'}}$$

$$\hat{\Sigma} = \frac{[\sum_{t=1}^{n} (\underline{X_t} + \underline{\hat{\alpha}}\, \underline{X_{t-1}})(\underline{X_t} + \underline{\hat{\alpha}}\, \underline{X_{t-1}})']}{n} .$$

Proof. See Anderson (1971, p. 196).

5.3 Example

We illustrate the normal approximations to the distributions of the parameters of autoregressive processes by considering the following simple example:

Let X_i denote the autoregressive process satisfying the following first-order stochastic difference equations:

$$X_i = \alpha X_{i-1} + \sigma V_i , \quad i = 0,1,2, \ldots ,$$

where σ is unknown, V_i is normal (0,1).

The conditional density of X_1 given X_0 is

$$\frac{\left\{\exp\left[-\frac{(X_1 - \alpha X_0)^2}{2\sigma^2}\right]\right\}}{\sqrt{2\pi}\,\sigma}$$

and similarly for the conditional density of X_j given $X_{j-1}, X_{j-2}, \ldots, X_1$ such that, for example the conditional density of X_n given $X_0, X_1, \ldots, X_{n-1}$ is

$$\frac{\left\{\exp\left[-\frac{(X_n - \alpha X_{n-1})^2}{2\sigma^2}\right]\right\}}{\sqrt{2\pi}\,\sigma} .$$

It can be checked that the joint conditional density of $X_1, \ldots, X_n$ given X_0 is

$$\frac{\left\{\exp\left[-\frac{\sum_{i=1}^{n}(X_i - \alpha X_{i-1})^2}{2\sigma^2}\right]\right\}}{(\sqrt{2\pi})^n\sigma^n} .$$

Upon substituting the actual values of $X_0, \ldots, X_n$ we obtain a function of the parameters α, σ which will be called the conditional likelihood function. The values

$$\hat{\alpha} = \frac{\sum_{i=1}^{n} X_i X_{i-1}}{\sum_{i=1}^{n} X_{i-1}^2} , \qquad \hat{\sigma} = \left[\frac{\sum_{i=1}^{n}(X_i - \hat{\alpha}X_{i-1})^2}{n}\right]^{\frac{1}{2}}$$

maximize the conditional likelihood function and will be called the (conditional) maximum likelihood estimates.

$$\sum_{i=1}^{n}(X_i - \alpha X_{i-1})^2 = \sum_{i=1}^{n}(X_i - \hat{\alpha}X_{i-1})^2 + (\alpha - \hat{\alpha})^2\left(\sum_{i=1}^{n} X_{i-1}^2\right)$$

$$- 2\sum_{i=1}^{n}(X_i - \hat{\alpha}X_{i-1})(\alpha - \hat{\alpha})X_{i-1}$$

$$= \sum_{i=1}^{n}(X_i - \hat{\alpha}X_{i-1})^2 + (\alpha - \hat{\alpha})^2\left(\sum_{i=1}^{n} X_{i-1}^2\right)$$

$$- 2(\alpha - \hat{\alpha})\sum_{i=1}^{n}(X_i - \hat{\alpha}X_{i-1})X_{i-1}$$

$$= \sum_{i=1}^{n} (X_i - \hat{\alpha}X_{i-1})^2 + (\alpha - \hat{\alpha})^2 (\sum_{i=1}^{n} X_{i-1}^2)$$

$$- 2(\alpha - \hat{\alpha}) \sum_{i=1}^{n} (X_i X_{i-1} - \hat{\alpha}X_{i-1}^2)$$

$$= \sum_{i=1}^{n} (X_i - \hat{\alpha}X_{i-1})^2 + (\alpha - \hat{\alpha})^2 (\sum_{i=1}^{n} X_{i-1}^2)$$

$$= n\hat{\sigma}^2 + (\alpha - \hat{\alpha})^2 (\sum_{i=1}^{n} X_{i-1}^2).$$

Hence the conditional likelihood is proportional to

$$\frac{\left\{\exp\left[-\frac{\left\{n\hat{\sigma}^2 + (\tilde{\alpha} - \hat{\alpha})^2 (\sum_{i=1}^{n} X_{i-1}^2)\right\}}{2\tilde{\sigma}^2}\right]\right\}}{(\sqrt{2\pi})^n \tilde{\sigma}^n}.$$

It follows that the conditional relative likelihood is equal to $\exp[- n\varphi(\tilde{\alpha},\tilde{\sigma},u_n(X_1, . . . ,X_n))]$, where

$$\varphi(\tilde{\alpha},\tilde{\sigma},u_n(X_1, . . . ,X_n)) = - \log(\frac{\hat{\sigma}}{\tilde{\sigma}}) +$$

$$+ \frac{[(\hat{\sigma}^2 - \tilde{\sigma}^2) + (\tilde{\alpha} - \hat{\alpha})^2 u_n(X_1, . . . ,X_n)]}{2\tilde{\sigma}^2}.$$

Here we have set

$$u_n(X_1, \ldots, X_n) = \frac{(\sum_{i=1}^{n} X_{i-1}^2)}{n}.$$

It follows from Theorem 5.2.1 that

$$u_n(X_1, \ldots, X_n) = \frac{(\sum_{i=1}^{n} X_{i-1}^2)}{n} \to E[X_{i-1}^2] \quad \text{with probability } 1.$$

It follows from Theorem 5.2.2 that

$\hat{\alpha} \to \alpha$ with probability 1

$\hat{\sigma} \to \sigma$ with probability 1.

A posterior distribution of the parameters α and σ may be derived by multiplying the conditional likelihood function by a prior density, and we can find an approximation to this posterior distribution using the approach developed in the previous chapter. The results are formally the same as in simple regression. In this approximation to the posterior distribution of α and σ, these parameters are independent, α has a normal distribution with mean $\hat{\alpha}$ and variance $\frac{\hat{\sigma}^2}{\sum_{i=1}^{n} X_{i-1}^2}$, and σ has a normal distribution with mean $\hat{\sigma}$ and variance $\frac{\hat{\sigma}^2}{2n}$.

5.4 Conclusions

The normal approximations to the distributions of the parameters of autoregressive processes are obtained by applying the strong law of large numbers to the conditional relative likelihood.

BIBLIOGRAPHY

1. Anderson, T. W., The Statistical Analysis of Time Series, John Wiley and Sons, Inc., New York, 1971.

2. Billingsley, Patrick, Convergence of Probability Measures, John Wiley and Sons, Inc., New York, 1968.

3. Cramer, Harald, Mathematical Methods of Statistics, Princeton University Press, 1946.

4. DeGroot, Morris H., Optimal Statistical Decisions, McGraw-Hill Book Company, New York, 1970.

5. Kingman, J. F. and Taylor, S. J., Introduction to Measure and Probability, Cambridge University Press, 1966.

6. Royden, H. L., Real Analysis, The Macmillan Company, 1968.