Institute for Mathematics and its Applications

IMA

The Institute for Mathematics and its Applications was established by a grant from the National Science Foundation to the University of Minnesota in 1982. The IMA seeks to encourage the development and study of fresh mathematical concepts and questions of concern to the other sciences by bringing together mathematicians and scientists from diverse fields in an atmosphere that will stimulate discussion and collaboration.

The IMA Volumes are intended to involve the broader scientific community in this process.

Avner Friedman, Director
Willard Miller, Jr., Associate Director

IMA ANNUAL PROGRAMS

1982–1983 Statistical and Continuum Approaches to Phase Transition
1985–1986 Stochastic Differential Equations and Their Applications
1986–1987 Scientific Computation
1987–1988 Applied Combinatorics
1988–1989 Nonlinear Waves
1989–1990 Dynamical Systems and Their Applications
1990–1991 Phase Transitions and Free Boundaries
1993–1994 Emerging Applications of Probability

IMA SUMMER PROGRAMS

1987 Robotics
1988 Signal Processing
1989 Robustness, Diagnostics, Computing and Graphics in Statistics
1990 Radar and Sonar
1990 Time Series
1991 Semiconductors
1992 Environmental Studies: Mathematical, Computational, and Statistical Analysis

SPRINGER LECTURE NOTES FROM THE IMA:

The Mathematics and Physics of Disordered Media

Editors: Barry Hughes and Barry Ninham
(Lecture Notes in Math., Volume 1035, 1983)

Orienting Polymers

Editor: J. L. Ericksen
(Lecture Notes in Math., Volume 1063, 1984)

New Perspectives in Thermodynamics

Editor: James Serrin
(Springer-Verlag, 1986)

Models of Economic Dynamics

Editor: Hugo Sonnenschein
(Lecture Notes in Econ., Volume 264, 1986)
IDENTIFICATION OF NONLINEARITIES AND NON-GAUSSIANITIES IN TIME SERIES

TOHRU OZAKI

Abstract. The identification of nonlinearities and non-Gaussianities of time series is discussed based on the observation that many nonlinear and/or non-Gaussian time series are generated from Gaussian white noise. The importance of nonlinear stochastic dynamical systems as their generation mechanisms is pointed out. A statistical identification method for nonlinear stochastic dynamical systems is presented. The identification method employs the maximum likelihood method which is based on Gaussian innovations obtained by applying a nonlinear filtering method to the observation data. The application of the method for the identification of nonlinear random vibration systems is presented with numerical results. Implications of the present method in nonlinear and non-Gaussian time series analysis are discussed.

Key words. Gaussian white noise, System identification, chaos, nonlinear stochastic dynamical systems, maximum likelihood method, local linearization, Markov diffusion process, nonlinear random vibrations, van der Pol oscillations, innovations, nonlinear state-space representation, nonlinear filtering, non-Gaussian distribution, Pearson system.

1. Introduction. In linear Gaussian time series analysis, Gaussian white noise has played a very important role. An ARMA (AutoRegressive Moving Average) time series is understood as the output of a system with rational transfer function driven by a Gaussian white noise. The model identification and diagnostic checking of the Box-Jenkins procedure for the ARMA time series modelling ([3]) are based entirely on this understanding, which suggests finding a linear dynamic mechanism to whiten the given time series data. In nonlinear and/or non-Gaussian time series analysis, however, we do not have any general standard model like the ARMA models in linear case. Instead we have many varieties of nonlinear time series models (see [23]). Some of them, such as Exponential AR models, Threshold AR models, Bilinear models, and Random Coefficient models are based on Gaussian white noise, but some are not. Some models are based on non-Gaussian white noise with fat tail distribution like the Cauchy distribution, and some models are based on Exponentially distributed driving noise. How are we supposed to understand these nonlinear and/or non-Gaussian time series models? Isn’t Gaussian white noise important in nonlinear and non-Gaussian time series analysis anymore?

In this paper we will see that Gaussian white noise will still play as important a role in nonlinear and/or non-Gaussian time series analysis as in the linear Gaussian case. We will see that very wide class of nonlinear and/or non-Gaussian stochastic processes are in the category of models driven by Gaussian white noise. What is different from the linear Gaussian case is that the dynamics of the system driven by Gaussian white noise are nonlinear. This means the idea of finding the whitening operation systematically used in linear Gaussian case in Box-Jenkins procedure is still valid in nonlinear and/or non-Gaussian time series model identification.

*Institute of Statistical Mathematics, 4-6-7 Minami Azabu Minato-ku, Tokyo Japan-106
What is required in nonlinear and/or non-Gaussian case is to find nonlinear
dynamics to whiten the time series data. This data has been brought in through
analysing real nonlinear and non-Gaussian time series data in ocean engineering,
where the time series data come from the dynamics of a moored ship which is
described by a continuous time stochastic dynamical system model. Typical non-
linearity and non-Gaussianity observed in real time series of moored vessel motion
is shown in section 2. Many dynamic phenomena in scientific fields are modeled by
a continuous time stochastic dynamical system driven by a Gaussian white noise
$\mathbf{w}(t)$ as in,

$$\dot{z} = f(z | \mathbf{z}) + \mathbf{w}(t).$$

This is because modelling dynamics in continuous time makes things easier to un-
derstand than in discrete time when the dynamics is nonlinear.

Data are mostly sampled at discrete time points. In some cases we can record
analog data of some phenomena, but that does not mean we have continuous time
observation data of the stochastic dynamical system. It is because the system is
driven by a continuous time Gaussian white noise and the continuous time record
of such system is supposed to contain infinite information even in a very short mi-
croscopic time interval. A machine which records such fast movement of continuous
time white noise does not exist. Even though it exists it is impossible to store
the exact continuous time data in a computer even before we apply any estimation
procedure to the data. To store the data we need to cut off the information of the
high frequency fluctuations of the process outside a certain frequency band.

In practice we need to use a discrete time dynamic model to whiten the time
series data measured in discrete time. This suggests us to introduce a nonlinear
discrete time dynamic model which is obtained from the continuous time model
by some time discretization method. A useful discretization method, called local
linearization method, is introduced in section 3. The discretization method leads
us to the identification method for a nonlinear stochastic dynamical system, which
generates the time series data, based on the approximate discrete time model. The
identification of nonlinearities in time series is realized by identifying the nonlinear
stochastic dynamical system generating the time series.

The dimension $r$ of the observation data $\mathbf{z}_t$ is sometimes less than the dimension
$k$ of the state $\mathbf{z}(t)$. In this situation the idea of finding a whitening operator for
$\mathbf{z}_t$ is still realized by the above method combined with a nonlinear filtering method
which transforms the time series data to Gaussian white innovation sequences. A
nonlinear filtering scheme based on the locally linearized model is given in section
4.

The identification method can be used to judge if a deterministic dynamical
system, i.e. a chaos model, is really more appropriate than a stochastic model by
checking the estimated variance of the system noise. Topics related to the chaos
model are discussed in section 5.

The process defined by the continuous time dynamical system driven by a Gauss-
ian white noise is a Markov diffusion process. A class of Markov diffusion processes
in general is a very wide class of stochastic processes indeed, including many inter-
esting non-Gaussian distributed processes. In section 6 it is pointed out that any Markov diffusion process has a stochastic dynamical system representation associated with an instantaneous variable transformation. It means that any diffusion process can be whitened into a Gaussian white noise process by a nonlinear dynamic model combined with an instantaneous variable transformation. It is also shown that there can be infinitely many different nonlinear dynamical system models which produce one and the same non-Gaussian marginal distribution. Thus the identification of a non-Gaussianities in time series is very much dependent on the identification of nonlinearities in a dynamical system associated with the process which generates the time series.

2. Time Series for Moored Vessel Motion. It is a common view of applied time series analysts that most non-linear and/or non-Gaussian time series analysis consists of a great deal of complicated mathematics without much contribution to real and general applications. To avoid confusion and uselessness in applications we consider nonlinearities and non-Gaussianities in time series based on some real time series data which have clear and significant nonlinearity and non-Gaussianity.

A typical example of nonlinear and non-Gaussian time series is the measurement of the motion of moored vessel in the sea. Fig. 2.1 shows time series data obtained in an experiment where a ship is moored in a experimental water tank and random waves are artificially generated to excite the moored ship. The data include time series of the height of the waves near the moored ship, the tension of the mooring rope and the dynamics of the moored ship, i.e. surging, pitching, heaving and yawing. The data was sampled every 0.03 seconds and the number of data points for each series is 2000.

Fig. 2.2 shows the histogram of these time series data. The tension data of Fig. 2.1 shows a clear non-Gaussian characteristic which can be noticed without even in looking at the asymmetric shape of its histogram in Fig. 2.2. The histogram of yawing does not contain reliable information since the data length is rather short compared with the length of the period of its main frequency (see Fig. 2.1). Interesting data is the surge data whose non-Gaussian characteristic is not obvious in Fig. 2.1, but is clearly seen in the asymmetric histogram in Fig. 2.2. The tail in the left hand side decays fast and cut off, while its right hand side tail decays more slowly. This happens because when the ship surges to the left direction the mooring rope starts pulling it back. This is confirmed by the diagram of the two dimensional plot of the tension data and other variable in Fig. 2.3 where the tension increases nonlinearly when the surge goes to the left hand side and decreases to zero when the surge goes to the right hand side.

This kind of nonlinear dynamics of the moored vessel is modeled by equation (2.1),

\[
\ddot{x}(t) + \dot{x}(t) + bx = n(t) - T(x)
\]

where \(T(x)\) is some nonlinear function of the surge position \(x\) and represents the tension force on the surge motion caused by the mooring rope. A typical example
of the function $T(x)$ is

$$T(x) = c[(x - x_0)^3]^+$$

(2.2)

$$= c(x - x_0)^3 \quad \text{for } x > x_0$$

$$0 \quad \text{for } x \leq x_0$$
The model (2.1) can be rewritten as

\[
\ddot{x}(t) + a\dot{x}(t) + bx + T(x) = n(t).
\]

This is a random vibration with restoring force \( bx + T(x) \). It means the vibrating system of the moored ship is nonlinear when the restoring force \( T(x) \) is nonlinear.
Wave x Tension

Surge x Tension

Heave x Tension

Pitch x Tension

Fig. 2.3
even if the restoring force of the ship is linear. If we can estimate the parameters \( a \) and \( b \) and identify the function \( T(x) \) for this model, we can characterize the nonlinear dynamics and non-Gaussian feature of the surge motion of the moored vessel.

3. How to Identify the Dynamics? In this section we try to identify the model,

\[
\dddot{x}(t) + a\dot{x}(t) + bx + T(x) = n(t)
\]

from the observation data. If we employ the parametric form (2.2) for \( T(x) \), the model (3.1) can be rewritten in a two dimensional stochastic dynamical system representation as

\[
\dot{x} = f(x, t) + w(t)
\]

where

\[
\begin{align*}
\theta &= (a, b, c, x_0)' \\
x(t) &= (x(t), \dot{x}(t))' \\
w(t) &= (n(t), 0)' \\
f(x, t) &= (-a\dot{x} - bx - c[(x - x_0)^3]_+, \dot{x})'.
\end{align*}
\]

\( n(t) \) is a Gaussian white noise with variance \( \sigma^2 \). We first consider the identification of the model (3.2) when we have the observation data \( \tilde{z}_1 = (\tilde{x}_1, \dot{\tilde{x}}_1)' \), \( \tilde{z}_2 = (\tilde{x}_2, \dot{\tilde{x}}_2)' \), ..., \( \tilde{z}_N = (\tilde{x}_N, \dot{\tilde{x}}_N)' \). The identification problem for the model when we have only observations \( x_1, x_2, ..., x_N \) without the observation of the velocity of \( x \) is discussed in the next section.

Our basic idea for model identification is the same as in ARMA modelling, i.e. to find a whitening mechanism for the data \( \tilde{z}_1, \tilde{z}_2, ..., \tilde{z}_N \) into Gaussian white noise \( w_1, w_2, ..., w_N \). For this purpose we introduce a discrete time dynamical system model of the following type from model (3.2),

\[
\tilde{z}_{t+\Delta t} = A(\tilde{z}_t, \theta)\tilde{z}_t + B(\tilde{z}_t, \theta)w_{t+\Delta t}
\]

If this discrete time model is a good approximation for the original continuous time model (3.2), we can whiten the data \( \tilde{z}_1, \tilde{z}_2, ..., \tilde{z}_N \) into Gaussian white noise \( w_1, w_2, ..., w_N \), by

\[
w_{t+\Delta t} = B(\tilde{z}_t, \theta)^{-1}\{\tilde{z}_{t+\Delta t} - A(\tilde{z}_t, \theta)\tilde{z}_t\}
\]

The estimates of the parameter \( \theta \) of the model are given by the maximum likelihood estimates \( \hat{\theta} \) for the model (3.3).
3.1 Time Discretization. For the above purpose, we try to introduce a “good” discrete time approximation to the continuous time stochastic dynamical system model,

\[ \dot{z} = f(z | \theta) + w(t) \]

for a nonlinear and non-Gaussian process \( z(t) \), where \( w(t) \) is a Gaussian white noise and \( f(z | \theta) \) is a nonlinear vector function of \( z(t) \). Before we go into the definition of our discretization method, which we call the local linearization method, let us see why conventional and common methods of time discretization methods are not satisfactory and why our new “good” method is needed.

For simplicity of explanation we first confine ourselves to the time discretization of the scalar stochastic dynamical system model,

\[ \dot{z} = f(z | \theta) + w(t). \]  

The most simple and common discrete time model for the scalar stochastic dynamical system (3.5) is

\[ z_{t+\Delta t} = z_t + \Delta tf(z_t | \theta) + \sqrt{\Delta t} w_{t+\Delta t}. \]  

This discretization method is a stochastic analogue of Euler’s method for the discretization of deterministic dynamical system,

\[ \dot{z} = f(z | \theta). \]

The time discretization method (3.6) is consistent, i.e.

\[ \frac{z_{t+\Delta t} - z_t}{\Delta t} \rightarrow f(z_t | \theta) + w(t) \quad \text{for} \quad \Delta t \rightarrow 0. \]

Actually the Itô stochastic differential equation (and the Stratonovich stochastic differential equation as well, since the variance of \( w(t) \) is independent of \( z(t) \)) is defined as a limit of this discrete time model (see [13]). Therefore the distribution of \( z_t \) defined by the approximate discrete model is known to converge to the distribution of \( z(t) \) in probability. The only disadvantage of this model is that the discrete time process, i.e. the Markov chain in a continuous state space defined by (3.6) is not stationary for most nonlinear function \( f(z | \theta) \) if the time interval \( \Delta t \) is fixed, no matter how small it is (see [8] and [31]). We may have a non-explosive sample path if we simulate the model with a small \( \Delta t \) for a certain finite time interval. However the probability of computational explosion in a finite time step is always 1, i.e. if we carry on the simulation it always ends up with computational explosion in finite steps. For example, for the process defined by (3.5) with

\[ f(z | \theta) = -z^3, \]

the model (3.6) explodes if \( z_t \) starts from an initial value greater than \( \sqrt{2}/(\Delta t) \), even though the original continuous time process is stationary and never explodes no matter how large an initial value it starts from.
Some computationally stable time discretization schemes of the deterministic dynamical system,

$$\dot{z} = f(z | \theta)$$

have been used in the engineering field (see p264 of [28]). The discrete model derived from the scheme applied to the deterministic dynamical system (3.7) is

$$z_{t+\Delta t} = e^{J_t \Delta t} z_t$$

where $J_t$ is the Jacobian of $f(z | \theta)$ at time point $t$. The model (3.8) is known to be computationally stable whenever the original nonlinear deterministic dynamical system is stable. When $f(z | \theta)$ is linear the trajectory of (3.8) coincides with the exact trajectory of the original system (3.7) on the discrete time points $t = \Delta t, 2\Delta t, 3\Delta t, \ldots$. If we replace the deterministic part of the model (3.6) by (3.8) we have the following discrete time stochastic dynamical system model,

$$z_{t+\Delta t} = e^{J_t \Delta t} z_t + \sqrt{\Delta t} w_{t+\Delta t}.$$  

The discrete time stochastic process $z_t$ of (3.9), which is a Markov chain with a continuous state space, is known to be computationally stable and ergodic. Unfortunately the model is not consistent unless $f(z | \theta)$ is linear. If we let $\Delta t \to 0$, the process defined by the model (3.9) converges to

$$\dot{z}(s) = J_t z(s) + w(s).$$

Incidentally we note that if we assume the model (3.10) on the interval $[t, t + \Delta t)$, the autocovariance function $\gamma_t(s)$ of the process $z(s)$ of (3.10) is

$$\gamma_t(s) = \left( \frac{\sigma_w^2}{2 \Delta t} \right) e^{J_t s}.$$ 

Since

$$\frac{\gamma_t(\Delta t)}{\gamma_t(0)} = e^{J_t \Delta t},$$

it is reasonable to have $e^{J_t \Delta t}$ as a coefficient in its discretized model (3.9). This is used later when we introduce our local linearization scheme, which is consistent and computationally stable.

Some more sophisticated time discretization methods for the deterministic dynamical system (3.7) are known in numerical analysis, such as the Heun method and the Runge-Kutta method. It may be natural to substitute the deterministic part of (3.6) by some of the deterministic discrete time models obtained by these more sophisticated methods. We call these discrete time stochastic dynamical system models Heun scheme model and Runge-Kutta scheme model respectively. Actually these models are also used for simulation of stochastic differential equations (see [2] and [11]). Since the difference of these models and model (3.6) is of order $(\Delta t)^2$.
with \( r > 1 \), these models have the same consistency property as the Euler scheme model (3.6).

Unfortunately, however, the disadvantage of model (3.6), i.e. its non-stationarity, is not remedied by these more sophisticated discrete time models. For example the Runge-Kutta scheme, applied to

\[
\dot{z} = -z^3 + w(t),
\]
gives us a stochastic discrete time dynamic model,

\[
z_{t+\Delta t} = p_{81}(z_t) + \sqrt{\Delta t}w_{t+\Delta t}
\]

where \( p_{81}(z_t) \) is a 81-st order polynomial of \( z_t \). Therefore the Markov chain defined by (3.11) is, like the one in (3.6) introduced by the Euler scheme, transient (see [8] and [31]) and is not stationary.

In numerical analysis it is known that the error of the Euler method applied to the deterministic dynamical system (3.7) is of order \( O(\Delta t) \), whereas the Heun method is of order \( O(\Delta t^2) \), and the Runge-Kutta method is of order \( O(\Delta t)^4 \). In the stochastic situation one way of evaluating the goodness of the approximations may be by the mean square one step error

\[
E_z[z(t + \Delta t) - z_{t+\Delta t}]^2,
\]

where \( E_z \) denotes the conditional expectation with respect to \( z_t \). The Euler scheme model (3.6) is known to have expected mean square error of order \( O(\Delta t)^3 \) ([12]). Higher order expected mean square errors are expected for the Heun scheme model and the Runge-Kutta scheme model. However it is shown that the order of the expected mean square error of these models are also \( O(\Delta t)^3 \). Moreover it was proved that the maximum possible speed of convergence attained by a discrete time approximate model is \( O(\Delta t)^3 \) ([26]). This result is, in a sense, natural because even though we approximate the deterministic part of the model (3.5) more accurately by Runge-Kutta method etc., the approximation of the white noise part in these models stays as poor as in the Euler scheme model, dominating the overall performance, and is impossible to improve. Thus other discrete time models with more sophisticated time discretization schemes are no better than the Euler method (3.6).

From the above discussion we now know that we should not expect to have a new time discretization scheme with a higher order convergence speed of the approximation. What we need is a new discretization method which is consistent as the Euler scheme method and at the same time yields a computationally stable and stationary Markov chain as (3.9) for a fixed \( \Delta t \) whenever the original continuous time Markov diffusion process is stationary.

3.2. Local Linearization. A key for the introduction of a new time discretization scheme lies in the scheme (3.9) which is the only one giving a computationally stable scheme out of the conventional methods. The scheme is also the only one whose deterministic part gives a trajectory which exactly coincides with the true
trajectory of the continuous deterministic system (3.7) at the discrete time points when \( f(z \mid \theta) \) is linear. The scheme (3.9) can be introduced by approximating the original process by the linear process (3.10) on each short interval \([t, t + \Delta t]\), where we assume that the coefficient of the linear function of the dynamical system on the interval is given by the Jacobian \( J_t \) of \( f(z \mid \theta) \) at time point \( t \). A disadvantage of this scheme is that it is not consistent. This is because \( J_t \) is constant on the interval \([t, t + \Delta t]\) and the function \( J_t z \) does not converge to \( f(z(t) \mid \theta) \) for \( \Delta t \to 0 \) unless \( f(z(t) \mid \theta) \) is linear. This consideration suggests that we may be able to get a computationally stable and consistent scheme if we use some different function \( L_t \) to approximate the original continuous model on each short time interval \([t, t + \Delta t]\) by using, instead of (3.9), a linear stochastic dynamical system,

\[
(3.12) \quad \dot{z}(s) = L_t z(s) + w(s),
\]

where \( L_t \) is some function of \( z_t \) and

\[
L_t z_t \to f(z(t)) \quad \text{for} \quad \Delta t \to 0.
\]

A simple scheme which satisfies this requirement is obtained by a simple assumption, i.e. "The Jacobian of the linear dynamical system,

\[
(3.13) \quad \dot{z}(s) = L_t z(s)
\]

is equivalent to the Jacobian,

\[
J_t = \left( \frac{\partial f}{\partial z} \right)_{z=z_t}
\]

of the original dynamical system on each short time interval \([t, t + \Delta t]\)." We call it the Local Linearization (L.L.) assumption. From this assumption we have

\[
\ddot{z}(s) = J_t \dot{z}(s),
\]

on \([t, t + \Delta t]\). If we integrate this on the interval \([t, t + \tau]\), where \( 0 \leq \tau < \Delta t \), we have

\[
\dot{z}(t + \tau) = e^{J_t \tau} \dot{z}(t) = e^{J_t \tau} f(z(t)).
\]

By integrating this again with respect to \( \tau \) on the interval \([0, \Delta t]\) we have,

\[
(3.14) \quad z(t + \Delta t) = z(t) + J_t^{-1}(e^{J_t \Delta t} - 1)f(z(t))
\]

Since \( J_t \) is given as a function of \( z(t) \), \( z(t + \Delta t) \) is explicitly given as a function of \( z(t) \) at time point \( t \). Thus the trajectory of the system (3.13) takes the value (3.14) at the discrete time point, \( t + \Delta t \). On the other hand the solution \( z(t + \Delta t) \) of the
linear system (3.13) is given explicitly as a function of $L_t$ and $z(t)$ by integrating $z(s)$ of (3.13) from $t$ to $t + \Delta t$ thus,

\begin{equation}
(3.15) \quad z(t + \Delta t) = e^{L_t \Delta t} z(t).
\end{equation}

From (3.14) and (3.15) we can write down $L_t$ explicitly as a function of $z(t)$, as

\begin{equation}
(3.16) \quad L_t = \frac{1}{\Delta t} \log \{1 + J_t^{-1} (e^{J_t \Delta t} - 1) F_t \}
\end{equation}

where

\[ F_t = \frac{f(z(t))}{z(t)}. \]

$L_t$ of (3.16) is defined on the region of $z(t)$ where $J_t z_t \neq 0$. When $J_t z_t \neq 0$, it holds, for sufficiently small $\Delta t$, that

\[ 1 + J_t^{-1} (e^{J_t \Delta t} - 1) F_t > 0. \]

Therefore $L_t$ of (3.16) is well defined. When $z(t)$ is on the region where $J_t z_t = 0$ (which is of measure zero), the $z(t + \Delta t)$ is defined separately in a way consistent with the rest of the region (see [20]).

Next we try to use $L_t$ of (3.16) to introduce a stochastic version of the discrete time approximation to the continuous time system. One simple way is, like those conventional discretization schemes for stochastic dynamical systems, to add $\sqrt{\Delta t} w_{t+\Delta t}$ to the deterministic part, as

\begin{equation}
(3.17) \quad z_{t+\Delta t} = e^{L_t \Delta t} z_t + \sqrt{\Delta t} w_{t+\Delta t}.
\end{equation}

However if we stick to the L.I. assumption (3.16) and extend it to stochastic dynamical systems, we have more consistent model than (3.17). In the stochastic situation it will be natural to start from the assumption, "The stochastic dynamical system is locally linear and so it is Gaussian on each short time interval $[t, t + \Delta t]$, i.e. we employ the model (3.12) on the interval to approximate the original process, where $L_t$ is given by (3.16)". Then $z(s)$ of (3.12) can be integrated from $t$ to $t + \Delta t$ on the interval giving

\[ z_{t+\Delta t} = e^{L_t \Delta t} z_t + a_{t+\Delta t} \]

where

\[ a_{t+\Delta t} = \int_t^{t+\Delta t} e^{L_s (t+\Delta t - s)} w(s) ds. \]

The variance of $a_{t+\Delta t}$ is

\[ \sigma_a^2 = \int_t^{t+\Delta t} e^{2L_s (t+\Delta t - s)} \sigma_w^2 ds \]

\[ = \frac{(e^{2L_t \Delta t} - 1)}{2L_t} \sigma_w^2. \]
where $\sigma^2_w$ is the variance of the continuous time white noise $w(t)$. Therefore it will be natural and more consistent if we use, instead of model (3.17), the following discrete time model,

\begin{align}
  z_{t+\Delta t} &= A_t z_t + B_t w_{t+\Delta t} \\
  A_t &= e^{L_t \Delta t} \\
  B_t &= \sqrt{\frac{(e^{2L_t \Delta t} - 1)}{2L_t}}
\end{align}

where $w_{t+\Delta t}$ is a discrete time Gaussian white noise of variance $\sigma^2_w$. Since $L_t$ of (3.16) satisfies

$$L_t z(t) \rightarrow f(z(t)) \text{ for } \Delta t \rightarrow 0,$$

model (3.18) is consistent. The deterministic part of the model (3.18) inherits the zero points and Jacobian of the original continuous time deterministic dynamical system (3.18). Therefore the stability property of the original continuous time stochastic dynamical system is also preserved in scheme (3.18). Ergodicity of the discrete time model is proved by direct application of Tweedie's theorem (see [31]) when a proper $f(\mathbf{z} | \mathbf{\theta})$ of the original continuous time system is given ([16], [20]). Thus the Markov chain process defined by (3.18) gives us a stationary Markov chain if the original continuous time diffusion process is stationary.

Since

$$\sqrt{\frac{(e^{2L_t \Delta t} - 1)}{2L_t}} \approx \sqrt{\Delta t}$$

we can replace $B_t$ of (3.18) by $\sqrt{\Delta t}$. Then we have the following nonlinear autoregressive model driven by a Gaussian white noise,

\begin{align}
  z_{t+\Delta t} &= A_t z_t + \sqrt{\Delta t} w_{t+\Delta t} 
\end{align}

which is equivalent to (3.17). When $f(\mathbf{z} | \mathbf{\theta})$ of (3.5) is nonlinear the autoregressive coefficient $A_t$ of the model (3.19) is a nonlinear function of $z_t$. Figures of the functions $A_t$ for the following three examples are illustrated in Fig. 3.1 where we set $\Delta t = 0.1$.

Example 3.1. $\dot{z} = -z^3 + w(t)$

Example 3.2. $\dot{z} = -6z + 5.5z^3 - z^5 + w(t)$

Example 3.3. $\dot{z} = -\tanh(\sqrt{2}z) + w(t)$

An interesting common feature seen in the figures of the functions $A_t$ in Fig. 3.1 is that the $A_t$'s are all smooth functions of $z_t$ and tend to a constant when $z_t$ tends to $\pm \infty$.

This suggests that among several well-known conventional nonlinear autoregressive time series models, the exponential autoregressive (ExpAR) model, Ozaki [14], and nonlinear threshold model, Ozaki [15], are more appropriate than others such as the linear threshold autoregressive (TAR) model of Tong and Lim [30] whose autoregressive coefficient is a discontinuous step function.
The local linearization can be applied to a $k$-dimensional stochastic dynamical system,

(3.20) \[ \dot{x} = f(x | \theta) + \psi(t) \]

without any essential change, where $\psi(t)$ is a $k$-dimensional Gaussian white noise.
with variance-covariance matrix \( \sigma^2 I \). Since the solution of the linear \( k \)-dimensional differential equation,

\[
\dot{z} = L\dot{z}(t)
\]

is

\[
z(t) = \text{Exp}(Lt)z(0),
\]

the discrete time model for the multi-dimensional stochastic dynamical system (3.20) is given by

\[
\begin{align*}
\tilde{z}_{t+\Delta t} & = A_t \tilde{z}_t + \tilde{a}_{t+\Delta t} \\
\tilde{a}_{t+\Delta t} & = \int_{t-\Delta t}^{t} \text{Exp} \{ L(t + \Delta t - s) \} \mu(s) ds \\
A_t & = \text{Exp}(L_t \Delta t) \\
L_t & = \frac{1}{\Delta t} \text{Log} \{ 1 + J_t^{-1} (e^{J_t \Delta t} - 1) F_t \} \\
J_t & = \{ \frac{\partial f(z)}{\partial z} \}_{z=\tilde{z}_t}
\end{align*}
\]

\( F_t \) is such that \( F_t \tilde{z}_t = f(\tilde{z}_t) \). The matrix functions \( \text{Exp}(\cdot) \) and \( \text{Log}(\cdot) \) are defined by

\[
\text{Exp}(L) = \sum_{k=0}^{\infty} \frac{1}{k!} L^k
\]

and

\[
\text{Log}(L) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} (L - I)^k
\]

respectively. The variance-covariance matrix of \( \tilde{a}_{t+\Delta t} \) is \( \Sigma_t \) whose elements are given as functions of the eigenvalues of \( L_t \) ([18], [22]). Using the unitary matrix \( U_t \) and the eigenvalues \( \lambda_1, \ldots, \lambda_k \) of the diagonalized representation of \( \Sigma_t \),

\[
\Sigma_t = \sigma^2 U_t \begin{bmatrix}
\lambda_1, 0 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & 0, \lambda_k
\end{bmatrix} U_t'
\]

we have, from (3.21), the following model,

\[
\tilde{z}_{t+\Delta t} = A_t \tilde{z}_t + B_t \tilde{a}_{t+\Delta t}
\]

where

\[
B_t = U_t \begin{bmatrix}
\sqrt{\lambda_1}, 0 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & 0, \sqrt{\lambda_k}
\end{bmatrix},
\]

and \( \tilde{a}_{t+\Delta t} \) is a discrete time \( k \)-variate Gaussian white noise with variance-covariance matrix \( \sigma^2 I \). Since the elements of the variance-covariance matrix \( \Sigma_t \) of \( \tilde{a}_{t+\Delta t} \) are explicitly given as a function of the eigenvalues of the matrix \( L_t \) the elements of
the unitary matrix $U_t$ and the eigenvalues $\lambda_1, \ldots, \lambda_k$, and so of $B_t$, are all given functions of $z_t$.

An example of a two dimensional stochastic dynamical system is given by the following model for moored vessel motion.

\begin{equation}
\ddot{z}(t) + 0.25\dot{z}(t) + 14.8z = n(t) - 4[z^3]_+.
\end{equation}

From this example we have, through the two dimensional stochastic dynamical system representation (3.2), a bi-variate time series model representation,

\begin{equation}
\ddot{z}_{t+\Delta t} = A_t z_t + B_t \varepsilon_{t+\Delta t}.
\end{equation}

The eigenvalues of $A_t$, which are given by $\exp(\lambda_1 \Delta t)$ and $\exp(\lambda_2 \Delta t)$ using the eigenvalues $\lambda_1$ and $\lambda_2$ of $L$, are functions of $z_t$. If $z$ is negative, the mooring force $4[z^3]_+$ is zero and the system is linear. When the system is linear, the eigenvalues of $A_t$, i.e. $\exp(\lambda_1 \Delta t)$ and $\exp(\lambda_2 \Delta t)$, are constant and the model (3.24) is known to be equivalent to a linear ARMA(2,1) model. The autoregressive coefficients $\phi_1$ and $\phi_2$ of the ARMA(2,1) model are given by

\begin{align*}
\phi_1 &= \exp(\lambda_1) + \exp(\lambda_2) \\
\phi_2 &= -\exp(\lambda_1) \exp(\lambda_2).
\end{align*}

The first autoregressive coefficient of ARMA(2,1) model is known to characterize the proper frequency of the vibrating system and the second coefficient to characterize the damping property of the vibrating system. However for $z > 0$ the system becomes nonlinear because of the presence of $4[z^3]_+$. We can see how $\phi_1$ and $\phi_2$ change for positive $z$ in Fig. 3.2, where the figures of the function $\phi_1$ and $\phi_2$ are calculated from (3.24) where $\Delta t = 0.1$ (see Fig. 3.2).

In the figure we notice that $\phi_1$ decreases when $z$ increases. This means that the proper frequency of the vibrating system shifts to higher frequency when $z$ is positive and increasing. This frequency shift phenomenon is a common nonlinearity seen in hard spring type random vibrations described by a stochastic Duffing equation,

\begin{equation}
\ddot{z}(t) + a\dot{z}(t) + bx + cz^3 = n(t).
\end{equation}

### 3.3. Maximum Likelihood Method

The log-likelihood function of the model (3.24) is given by

\begin{equation}
\log p(z_1, z_2, \ldots, z_N | \theta, \sigma^2) = \log p(z_2, z_3, \ldots, z_N | z_1, \theta, \sigma^2) + \log p(z_1 | \theta, \sigma^2) \\
= \log p(z_2, z_3, \ldots, z_N | z_1, \theta, \sigma^2) + \log |J(z, \varepsilon)| + \log p(z_1 | \theta, \sigma^2) \\
= \sum_{t=1}^{N-1} \frac{\|B_t^{-1}(z_{t+1} - A_t z_t)\|^2}{2\sigma^2} - \frac{N-1}{2} \log |\sigma^2 I| - \frac{N-1}{2} \log 2\pi \\
+ \sum_{t=1}^{N-1} \log \det(B) + \log p(z_1 | \theta, \sigma^2)
\end{equation}
where \( J(\mathbf{x}, \mathbf{u}) \) is the Jacobian of the transformation from \( \mathbf{z} = (\mathbf{z}_2, \mathbf{z}_3, \ldots, \mathbf{z}_N) \) to \( \mathbf{w} = (\mathbf{w}_2, \mathbf{w}_3, \ldots, \mathbf{w}_N) \) which is equivalent to
\[
N-1 \prod_{i=1}^{N-1} \det(B_i)
\]

When \( N \), the data length, is large the last term of the log-likelihood (3.25) is negligibly small compared with the rest and we can drop the term from the log-likelihood representation. Since the log-likelihood function satisfies the relation,
\[
\left[ \frac{\partial \log p(\mathbf{z}_2, \mathbf{z}_3, \ldots, \mathbf{z}_N | \mathbf{z}_1, \theta, \sigma^2) + \log p(\mathbf{z}_1 | \theta, \sigma^2)}{\partial \sigma^2} \right]_{\sigma^2 = \hat{\sigma}^2} = 0
\]
at the maximum likelihood estimate \( \hat{\sigma}^2 \) of \( \sigma^2 \), \( \sigma^2 \) takes the following form
\[
(3.26) \quad \sigma^2 = \frac{1}{N-1} \sum_{i=1}^{N-1} \| B_i^{-1}(\mathbf{z}_{i+1} - A_i \mathbf{z}_i) \|^2.
\]

Then in the log-likelihood representation (3.25) the first term of the right-hand side becomes constant at \( \sigma^2 = \hat{\sigma}^2 \) and the maximum log-likelihood is obtained by maximizing
\[
(3.27) \quad -\frac{N-1}{2} \log \sigma^2 + \sum_{i=1}^{N-1} \log \det(B_i)
\]
with respect to \( \theta \), where \( \sigma^2 \) in (3.27) is given by (3.26). To maximize the log-likelihood we need to use a numerical nonlinear optimization method such as the Davidon-Fletcher-Powell method or the Newton-Raphson method.

We can check how the above maximum likelihood method works in the following two examples of typical nonlinear random vibration time series data. One example is the data \( \xi_1 = (\xi_1, x_1)', \xi_2 = (\xi_2, x_2)', \ldots, \xi_N = (\xi_N, x_N)' \) \((N = 1000)\) generated from the van der Pol type nonlinear vibration model where the damping coefficient is nonlinear as in,

\[
(3.28) \quad \ddot{x}(t) + (x^2 - 1)\dot{x}(t) + 14.8x = n(t)
\]

The variance of the white noise is 1. The time series \( x_1, x_2, \ldots, x_N \) of the data \( \xi_1, \xi_2, \ldots, \xi_N \) is shown in Fig. 3.3.a. For this case we assumed the following parameteric model,

\[
(3.29) \quad \ddot{x}(t) + g_1(x)\dot{x}(t) + g_2(x)x = n(t)
\]

with

\[
(3.30) \quad g_1(x) = a_1 + a_2x + a_3x^2 \\
g_2 = b
\]

The estimated parameters obtained by applying the maximum likelihood method are shown in Table 3.1. The second example is the data \( \xi_1, \xi_2, \ldots, \xi_N \) \((N = 1000)\) generated from the Duffing type nonlinear vibration model where the restoring force \( g_2(x)x \) is nonlinear and so the coefficient \( g_2(x) \) is non-constant. The variance of the white noise is 1. The time series \( x_1, x_2, \ldots, x_N \) of the data \( \xi_1, \xi_2, \ldots, \xi_N \) is shown in Fig. 3.3.b. For this case we assumed the model (3.29) with

\[
(3.31) \quad g_1(x) = a, \\
g_2(x) = b_1 + b_2x + b_3x^2.
\]

The estimation results are in Table 3.1. In both examples the results are satisfactory and convincing. Since we are using a numerical optimization method, standard errors of the estimates are easily obtained numerically from the inverse of the Hessian, the second derivative of the log-likelihood function.

Table 3.1 Maximum likelihood estimates of the model (3.30) and the model (3.31).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model (3.30)</th>
<th>Model (3.31)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a_1 )</td>
<td>-1.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>( a_2 )</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>( a_3 )</td>
<td>1.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>( b )</td>
<td>1.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>( \sigma^2_n )</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
instead of introducing an approximate continuous time filtering scheme as in most nonlinear filtering methods ([10]). In (4.1), \( w(t) \) is a continuous time \( k \)-dimensional Gaussian white noise. We assume that \( \Sigma_w \), the variance-covariance matrix of \( w(t) \) is \( \sigma^2 I \) for simplicity. The variance-covariance matrix of discrete time \( r \)-dimensional Gaussian white noise \( \nu_t \) is \( \Sigma_v \) and \( w(t) \) and \( \nu_t \) are independent. Thus from (4.1), using the L.L. method in the previous section, we have,

\[
\begin{align*}
\dot{z}_{t+\Delta t} &= A(z_t)z_t + B(z_t)w_{t+\Delta t} \\
z_t &= Cz_t + \nu_t
\end{align*}
\]

where \( A(z_t) \) and \( B(z_t) \) are given by \( A_t \) of (3.20) and \( B_t \) of (3.21). \( w_{t+\Delta t} \) is a \( k \)-dimensional discrete time Gaussian white noise with variance-covariance matrix \( \Sigma_w \). Also between \( i \)-th component \( w^{(i)} \) of \( w_{t+\Delta t} \) and \( j \)-th component \( v^{(j)} \) of \( \nu_t \) it follows that \( E[w^{(i)}v^{(j)}] = 0 \) for \( 1 \leq i \leq k \) and \( 1 \leq j \leq r \). Based on the minimum variance principle ([28]), which is to minimize

\[
E[(\dot{z}_t - \dot{z}_{t|t})'(\dot{z}_t - \dot{z}_{t|t})],
\]

the filtering equation for (4.2) is obtained as follows,

\[
\begin{align*}
\dot{z}_{t+\Delta t|t} &= A(z_t)\dot{z}_{t|t} \\
\dot{z}_{t|t} &= z_{t|t-\Delta t} + K_t\nu_t \\
\nu_t &= Cz_{t|t-\Delta t} \\
K_t &= P_tC(P_tC^* + \Sigma_v)^{-1} \\
P_t &= E[(\dot{z}(t) - \dot{z}_{t|t-\Delta t})(\dot{z}(t) - \dot{z}_{t|t-\Delta t})'] \\
V_t &= P_t - K_tC
\end{align*}
\]

where \( \nu_t \) is the innovation of the filtering model, \( K_t \) is the filter gain and \( P_t \) is the variance-covariance matrix of the one step ahead prediction error of the state \( z_t \). The evolution of \( P_t \) is obtained from

\[
P_{t+\Delta t} = E[(\dot{z}_{t+\Delta t|t} - \dot{z}_{t+\Delta t|t})(\dot{z}_{t+\Delta t|t} - \dot{z}_{t+\Delta t|t})']
\]

\[
= E[(A(z_t)(\dot{z}_t - \dot{z}_{t|t}) + B(z_t)w_{t+\Delta t})(A(z_t)(\dot{z}_t - \dot{z}_{t|t}) + B(z_t)w_{t+\Delta t})'].
\]

Since we are assuming \( A(z_t) \) and \( B(z_t) \) are constant on \([t, t + \Delta t]\), we have

\[
P_{t+\Delta t} = A(z_t)V_tA(z_t)' + B(z_t)\Sigma_wB(z_t)'.
\]

As is mentioned in the previous section, the function \( B(\cdot) \) is almost constant for small \( \Delta t \), while \( A(\cdot) \) is not. Since we assume the system (4.2) is linear on \([t, t + \Delta t]\) and the system transition is characterized by \( A(z_t) \) as in (3.21) on the interval, it will be reasonable to replace \( A(z_t) \) and \( B(z_t) \) in (4.3) and (4.4) by \( A(z_{t|t}) \) and \( B(z_{t|t}) \) respectively. Then we have

\[
\begin{align*}
\dot{z}_{t+\Delta t|t} &= A(z_{t|t})\dot{z}_{t|t} \\
\dot{z}_{t|t} &= z_{t|t-\Delta t} + K_t\nu_t \\
\nu_t &= Cz_{t|t-\Delta t} \\
K_t &= P_tC(P_tC^* + \Sigma_v)^{-1}CP_t
\end{align*}
\]
We note that the maximum likelihood method which we have used is not for
the continuous time model (3.30) but for the discrete time model (3.21). If \( \hat{\theta} \) is the
estimated parameter the model which we obtained is

\[
\hat{z}_{t+\Delta t} = A(\hat{z}_t | \hat{\theta})\hat{z}_t + B(\hat{z}_t | \hat{\theta})\nu_{t+\Delta t}
\]

Then the following continuous time model (3.32) will be a natural estimate for the
continuous time model (3.20),

\[
\hat{z} = f(\hat{z} | \hat{\theta}) + \nu(t)
\]

since the local linearization of (3.33) is equivalent to (3.32). However since the
local linearization is an approximation this correspondence is not unique and we
have some other function \( f_{\Delta t}(z | \hat{\theta}) \) which produces (3.32). A numerical method
to obtain \( f_{\Delta t}(z | \hat{\theta}) \) from \( A(z_t | \hat{\theta}) \) is given in Ozaki ([17], [19]). The numerical
method provides us with a useful method to guess the functional form of \( f(z | \theta) \)
from an estimated nonlinear autoregressive coefficient of some general nonlinear
autoregressive model with fairly general parameterization such as the ExpAR model
(see [17] and [20] for numerical examples).

4. Nonlinear Filtering as a Whitening Operator. The above maximum
likelihood method is not valid when we observe, instead of a \( k \)-dimensional state
vector \( z_t \), a scalar time series data \( x_t \), which is a linear transformation of the state
de vector \( z_t \). In many practical identification problems we have only observations,
\( x_1, x_2, \ldots, x_N \), whose dimension \( r \) is smaller than the system dimension \( k \). Also the
observation error is not negligibly small. In these situations what we need is an
identification method for the following nonlinear state space representation model,

\[
\hat{z} = f(z | \theta) + \nu(t)
\]

\[
z_t = Cz(t) + w_t
\]

where \( C \) is an \( r \times k \) rectangular observation matrix, and \( w_t \) is an observation error
vector which is an \( r \)-dimensional Gaussian white noise. In this situation we can still
identify the model (4.1) based on the same idea in the previous section, i.e. to try to
find a whitening operator which transforms the \( r \)-dimensional data, \( z_1, z_2, \ldots, z_N \)
into \( r \)-dimensional Gaussian white innovations, \( \nu_1, \nu_2, \ldots, \nu_N \). This is realized by
using a nonlinear filtering method for (4.1).

4.1. Local Linearization Filter. The filtering problem for (4.1) is sometimes
called continuous-discrete nonlinear filtering ([7]) since the first equation for the
state dynamics is in continuous time and the other observation equation is in discrete
time. It is well known that nonlinear filtering is a difficult problem even in the
scalar case. Its main difficulty is the computational instability which comes from
the nonlinearity of the state dynamics. Our idea of obtaining a computationally
stable filtering scheme for (4.1) is to introduce a stable approximate discrete time
L.I. model from (4.1) and give a discrete time filtering scheme for the L.I. model,
and

\[ P_{t+\Delta t} = A(\mathbf{x}_{t|t})V_tA(\mathbf{x}_{t|t})' + B(\mathbf{x}_{t|t})\Sigma_u B(\mathbf{x}_{t|t})' \]

The local linearization filter works very well compared with conventional nonlinear filtering method such as the Extended Kalman filter or the Minimum Variance filter. The error reduces to one tenth or less compared with the conventional methods (see [21] for numerical results).

4.2. Innovation Likelihood. Before we go into the maximum likelihood method for the nonlinear state space models, let us see how the linear state space model parameters are estimated by innovation maximum likelihood method. The parameters of a discrete time linear state space model,

\[ \mathbf{\tilde{x}}_{t+\Delta t} = A(\varphi)\mathbf{\tilde{x}}_t + B(\varphi)\mathbf{w}_{t+\Delta t} \]

\[ \mathbf{\tilde{x}}_t = C\mathbf{z}(t) + \mathbf{\nu}_t \]

\[ C = \begin{pmatrix} 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 1 \end{pmatrix} \]

can be estimated by maximizing the likelihood,

\[ p(\mathbf{\tilde{x}}_1, \mathbf{\tilde{x}}_2, \ldots, \mathbf{\tilde{x}}_N | \varphi) \]

of the model for the given data \( \mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N \). If we use the Kalman filter, the data \( \mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N \) are transformed to innovations \( \mathbf{\nu}_1, \mathbf{\nu}_2, \ldots, \mathbf{\nu}_N \) as

\[ \mathbf{\nu}_i = \mathbf{z}_i - \hat{\mathbf{z}}_{i|i-1} \]

where \( \hat{\mathbf{z}}_{i|i-1} \) is the one-step ahead prediction value of \( \mathbf{z}_i \) at time point \( i-1 \) and is a function of \( \mathbf{\tilde{x}}_{i-1}, \mathbf{V}_{i-1} \) and \( \varphi \). If the initial values \( \mathbf{z}_0 \) and \( \mathbf{V}_0 \) are known, then the determinant of the Jacobian of the transformation from \( \mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N \) to the innovations \( \mathbf{\nu}_1, \mathbf{\nu}_2, \ldots, \mathbf{\nu}_N \) is 1. Usually they are unknown and may be specified by a probability density function \( p(\mathbf{z}_0, \mathbf{V}_0 | \varphi) \). Then the likelihood is written down in terms of the innovations, obtained by applying the Kalman filter, thus

\[ p(\mathbf{\tilde{x}}_1, \mathbf{\tilde{x}}_2, \ldots, \mathbf{\tilde{x}}_N | \varphi) = \int \int p(\mathbf{\tilde{x}}_1, \mathbf{\tilde{x}}_2, \ldots, \mathbf{\tilde{x}}_N | \mathbf{z}_0, \mathbf{V}_0, \varphi) p(\mathbf{z}_0, \mathbf{V}_0 | \varphi) d\mathbf{z}_0 d\mathbf{V}_0 \]

\[ = \int \int p(\mathbf{\nu}_1, \mathbf{\nu}_2, \ldots, \mathbf{\nu}_N | \mathbf{z}_0, \mathbf{V}_0, \varphi) p(\mathbf{z}_0, \mathbf{V}_0 | \varphi) d\mathbf{z}_0 d\mathbf{V}_0 \]

There are many ways of specifying \( p(\mathbf{z}_0, \mathbf{V}_0 | \varphi) \). One simple and practical way of specification is to use the delta function. If we assume a delta function for \( \mathbf{z}_0 \) and \( \mathbf{V}_0 \), the likelihood becomes

\[ p(\mathbf{\tilde{x}}_1, \mathbf{\tilde{x}}_2, \ldots, \mathbf{\tilde{x}}_N | \mathbf{z}_0, \mathbf{V}_0, \varphi) = p(\mathbf{\nu}_1, \mathbf{\nu}_2, \ldots, \mathbf{\nu}_N | \mathbf{z}_0, \mathbf{V}_0, \varphi), \]
where we have two more parameters \( \xi_0 \) and \( V_0 \). Since innovations are Gaussian white noise, we have

\[ p(\xi_1, \xi_2, \ldots, \xi_N | \xi_0, V_0, \theta) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi |\Sigma_{\xi_i}|}} \exp\left(-\frac{1}{2} \xi_i^T \Sigma^{-1}_{\xi_i} \xi_i\right). \]

Then we have

\[ (-2) \log p(\xi_1, \xi_2, \ldots, \xi_N | \xi_0, V_0, \theta) = \sum_{i=1}^{N} \left( \log |\Sigma_{\xi_i}| + \xi_i^T \Sigma^{-1}_{\xi_i} \xi_i\right) + N \log 2\pi. \]

Then the maximum likelihood estimates of the parameters \( \xi_0, V_0 \) and \( \theta \) are obtained by minimizing the \((-2)\log\)-likelihood (4.10). The maximum likelihood method based on this innovation likelihood has been used for the estimation of parameters of linear state space models ([27], [6] and [9]).

In the nonlinear case,

\[ \dot{x} = f(x | \theta) + \nu(t) \]
\[ \dot{z}_i = C \dot{x}(t) + \nu_i \]
\[ C = \begin{pmatrix} 0 \ldots 0 & 1 \ldots 0 \\ \ldots & \ldots \\ 0 \ldots 0 & 0 \ldots 1 \end{pmatrix} \]

we expect to use the L.L. filter to transform the data \( \xi_1, \xi_2, \ldots, \xi_N \) into Gaussian white noise innovations \( \xi_1, \xi_2, \ldots, \xi_N \) as we did in the linear case. When \( f(x | \theta) \) is nonlinear the state \( x(t) \) is not Gaussian. However if the sampling interval \( \Delta t \) is sufficiently small, we can assume that the data \( \xi_1, \xi_2, \ldots, \xi_N \) are generated from the discrete time state space model,

\[ \xi_{i+\Delta t} = A(\xi_i) \xi_t + B(\xi_i) \nu_{i+\Delta t} \]
\[ \dot{x}_i = C \dot{x}(t) + \xi_i \]

whose discrete time state space process \( \xi_t \) is driven by \( \nu_t \). Then we can get discrete time Gaussian white noise innovations, from observations \( \xi_t \), by applying the L.L. filter, since \( \nu_t \) is a discrete time Gaussian white noise. It is also proved by Frost and Kailath [5] that the continuous time innovation process for a non-Gaussian diffusion process is a Gaussian white noise. Since the continuous time innovations are defined as a limit of discrete time innovations as

\[ \xi(t) = \lim_{\Delta t \to 0} \xi_t \]
\[ = \lim_{\Delta t \to 0} (\dot{\xi}_t - C \xi_{t-[1-\Delta t]}) \]

their result implies that the innovation \( \xi_t \) obtained by the L.L. filter is very close to a Gaussian white noise. This consideration justifies and guarantees the use of the following representation for the approximation of the \((-2)\log\)-likelihood,

\[ (-2) \log p(\xi_1, \xi_2, \ldots, \xi_N | \xi_0, V_0, \theta) = \sum_{i=1}^{N} \left( \log |\Sigma_{\xi_i}| + \xi_i^T \Sigma^{-1}_{\xi_i} \xi_i\right) + N \log 2\pi \]
where $\Sigma_{e_i}$ is the variance-covariance matrix of the innovation $e_i$ and is given by the L.I. filter (4.3) as

\begin{equation}
\Sigma_{e_i} = CP_iC_i' + \Sigma_{e_i}.
\end{equation}

Thus the maximum likelihood estimates of the parameters $\hat{z}_0$, $\hat{V}_e$ and $\hat{\theta}$ of (4.7) are obtained by minimizing (4.13).

### 4.3. Numerical Results Applied to Surge Data

We applied the innovation maximum likelihood method to the surge data in section 2 (see Fig. 2.1). To estimate the nonlinear restoring function $kx + T(x)$ of the surge motion model (3.1) we used the following parametric model,

\begin{equation}
\ddot{x}(t) + \alpha \dot{x}(t) + b_1 x + b_2 x^2 + b_3 x^3 = n(t).
\end{equation}

With this model we expect to obtain an asymmetric estimated restoring function. Before we apply the maximum likelihood method we calibrate the data $x_i$ to the range between $-0.64$ and $1.2$ so that $-0.64 < x_i < 1.2$. The calibration is often useful to avoid computational difficulties, such as overflows, in likelihood calculation for nonlinear models. We assumed $\Delta t = 0.03$. We also assumed the observation error is zero because the measurement is quite accurate and the sampled data show quite smooth movements of the surge motion. The estimated parameters are as follows: $a = 51.4807$, $b_1 = 15.7963$, $b_2 = -14.3694$, $b_3 = 10.8584$ and $\sigma_N^2 = 221.9206$.

![Figure 4.1](image-url)
Fig. 4.1 shows the estimated restoring force function $b(x) = b_1 x + b_2 x^2 + b_3 x^3$. The function shows that when $x$ is on the left hand side it sharply decreases and pulls the ship back strongly, while if the surge motion goes to the right hand side the restoring force does not increase so sharply. It is known (see [4] and [18]) in Markov diffusion theory that the equilibrium density distribution $p(x)$ of $x$ of (4.15) is

$$p(x) = C \frac{1}{\sqrt{2\pi \sigma^2}} \exp\left(- \frac{2a(\frac{b_1}{3} x^2 + \frac{b_2}{3} x^3 + \frac{b_3}{4} x^4)}{\sigma^2}\right)$$

where $C$ is a normalizing constant. The density function (4.16) explains why the asymmetric non-Gaussian like histogram of surge data is seen in Fig. 2.2 through the nonlinear function $b(x)$ of Fig. 4.5.

5. Deterministic or Stochastic Models? The identification method introduced in the previous section can be used not only for stochastic dynamical system but for deterministic dynamical system models. Recently it has been shown that some non-deterministic process can be generated by a deterministic dynamical system model. There have been introduced many interesting nonlinear deterministic dynamical system models which show stochastic behaviour, called chaos ([25]). Some people seem to be led, from these results, to the belief in determinism, which claims that all the stochastic models should be replaced by deterministic models referring to Albert Einstein’s famous phrase (see [29] for example) which sounds misleading for those who know Einstein’s innovative contribution to Brownian motion theory in statistical mechanics.

With so much evidence of random fluctuations ([24]) at a microscopic level of quantum mechanics it does not seem to be so promising to revive the determinism of Newton and Laplace. Recently there has been some interesting work reformulating quantum mechanics based on Brownian motion theory. Actually it was Norbert Wiener who initiated the white noise approach to quantum mechanics as an alternative to the conventional quantum mechanics formulation of Niels Bohr (see [32] and [33]). Incidentally he started it after being stimulated by Albert Einstein’s hidden variable argument in the famous debate between Niels Bohr over the wave function formulation of quantum mechanics.

The chaotists (as we might call scientists who believe in determinism) claim that $\xi(t)$ and $\nu_t$ of (4.11) should be zero. However, as one cannot deny the uncertainty principle in quantum mechanics, we cannot assume $\nu_t = 0$. If the system noise $\xi(t)$ really was zero, then the maximum log-likelihood of the model should occur when the estimated variance-covariance matrix $\Sigma_w$ is close to zero. This never happens in numerical studies except for artificially generated time series data. Real data does not come from a simple mathematical model but comes from nature, which is too complicated to be totally explained by a simple deterministic model. Actually what Akaike’s Entropy Maximization Principle ([1]) and AIC aim to do is to find a model with $\Sigma_w$ as small as possible and at the same time with as small a number of parameters as possible. The present identification method will provide chaotists with a clue to understanding the big gap between reality and fantasy in real data analysis.
5. Non-Gaussianities in Time Series. In this section we see how non-Gaussianities of time series are specified by parametric models. In the past decades high order spectra have been studied to characterize non-Gaussianities of time series. However there is not any guarantee that major non-Gaussianity is specified by the bispectrum. The fact is that not only the bispectrum, but also all the higher order spectra from fourth order to infinite order are needed to characterize a single non-Gaussian process. Estimating so many quantities from a finite set of time series data does not sound a sensible thing to do for statisticians. To obtain some suggestions for the direction of the study of non-Gaussian time series, again, a real example of non-Gaussian phenomena in the real world is very useful.

6.1. Non-Gaussianity of Moored Vessel Motion. We have seen that the dynamics of moored vessel motion is characterized by a nonlinear random vibration model,

\[ \ddot{z}(t) + a\dot{z}(t) + b\dot{z} + T(z) = n(t) \]

where

\[ T(z) = c((z - x_0)^3)_+ \]
\[ = c(z - x_0)^3 \quad \text{for} \quad z > x_0 \]
\[ = 0 \quad \text{for} \quad z \leq x_0. \]

From (6.1) we have the following two dimensional stochastic dynamical system representation,

\[ \dot{z} = f(z, \theta) + \omega(t) \]

where

\[ \theta = (a, b, c, x_0)' \]
\[ z(t) = (\dot{z}(t), z(t))' \]
\[ \omega(t) = (n(t), 0)' \]
\[ f(z, \theta) = (-a\dot{z} - b\dot{z} - c((z - x_0)^3)_+, \dot{z})' \]

The process \( z(t) \) defined by (6.2) is a bivariate Markov diffusion process. The evolution of the transition function \( p(\tilde{z}, t) \) of the Markov diffusion process is characterized by the Fokker-Planck equation,

\[ \frac{\partial p}{\partial t} = \sum_{i=1}^{2} \frac{\partial}{\partial z_i} [f_i(z)p] + \frac{1}{2} \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{\partial^2}{\partial z_i \partial z_j} [\sigma_{ij} p] \]

where \( \sigma_{ij} \) is \((i, j)\)-th element of the variance-covariance matrix of \( \omega(t) \). The marginal distribution of the process \( z(t) \) is given by \( p(\tilde{z}, t) \) with \( t \to \infty \), which is obtained as

\[ p(\tilde{z}, \infty) = p_0 \exp\left(-\frac{a\tilde{z}^2}{\sigma^2}\right) \exp\left(-\frac{ab\tilde{z}^2}{\sigma^2} - \frac{ac(z - x_0)^3}{2\sigma^2}\right) \]
by solving the Fokker-Planck equation (6.3) with $\partial p/\partial t = 0$ ([4]). $p_0$ is a normalizing constant. We see from (6.4) that the tail of the distribution of $x$ decreases faster than the Gaussian case for $x > x_0$ because of the presence of the nonlinear term $c(x - x_0)^2$ in the original dynamical system of (6.2). This explains the asymmetric shape of the histogram of the surge data of Fig. 2.2 where the left hand side of the tail decreases faster and is cut off compared with the right hand side tail. It is interesting to see that the distribution of the velocity is not affected by the nonlinear term and is Gaussian distributed.

The above consideration suggests that the nonlinear dynamics specified by $c(x - x_0)^2$ also specifies the non-Gaussianity of the distribution of $x(t)$. Then a natural question may be “Does the non-Gaussian distribution of $x(t)$ specify the nonlinear dynamics of the process?” The answer is “No” which comes from further analysis of the relationship between distribution systems and diffusion processes.

6.2. Pearson System and Markov Diffusion Process. Wong [34] pointed out an interesting relationship between the Pearson system and Markov diffusion process. He showed that for any distribution $W(x)$ which belongs to the Pearson system,

$$\frac{dW(x)}{dx} = \frac{c_0 + c_1 x}{d_0 + d_1 x + d_2 x^2} W(x)$$

it is possible to give a diffusion process whose marginal distribution is $W(x)$. Ozaki [16] showed that it can be extended to any distribution $W(x)$ defined by proper analytic functions $c(x)$ and $d(x)$ in a distribution system given by,

$$\frac{dW(x)}{dx} = \frac{c(x)}{d(x)} W(x)$$

and also showed that the corresponding Markov diffusion process is given, with $c(x)$ and $d(x)$, by the following Fokker-Planck equation,

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x} [c(x) + d'(x)p] + \frac{1}{2} \frac{\partial^2}{\partial x^2} [2d(x)p].$$

The class of distributions defined by the system (6.6) is very wide indeed, including the Pearson system and the exponential family of distributions. The distribution system (6.6) is sometimes called the Extended Pearson system.

From Markov diffusion theory ([13]) we know that the diffusion process $x(t)$ defined by (6.7) has a stochastic differential equation representation,

$$\dot{x} = a(x) + \sqrt{b(x)} n(t)$$

where

$$a(x) = c(x) + d'(x),$$

$$b(x) = 2d(x),$$
in the Ito form of stochastic calculus. In Stratonovich form they are,

\[ a(x) = c(x) - \frac{1}{2} d'(x) \]
\[ b(x) = 2d(x) \]

From (6.8) we have,

\[ \frac{\dot{x}}{\sqrt{b(x)}} = \frac{a(x)}{\sqrt{b(x)}} - \frac{1}{4} \frac{b'(x)}{\sqrt{b(x)}} + n(t) \]

With the variable transformation,

\[ y = \int^x \frac{1}{\sqrt{b(\zeta)}} d\zeta \]

we can replace (6.8) by a dynamical system driven by a Gaussian white noise \( n(t) \) as follows.

\[ x = h(y) \]
\[ \dot{y} = f(y) + n(t) \]

where

\[ f(y) = \frac{a(x)}{\sqrt{b(x)}} - \frac{1}{4} \frac{b'(x)}{\sqrt{b(x)}} \]

Since the white noise in (6.10) is free from \( x \), the representation (6.10) is uniquely given and coincides for both Ito calculus and Stratonovich calculus.

The above consideration suggests that we can generate, with the locally linearized time series model,

\[ x_t = h(y_t) \]
\[ y_{t+\Delta t} = A_t y_t + B_t n_{t+\Delta t} \]

a time series with a marginal distribution which is close to any density distribution belonging to the Extended Pearson system defined by (6.6). We can confirm this in simulated time series of these diffusion processes and their histograms shown in Fig. 6.1.

The simulated time series in Fig. 6.1 are obtained using the local linearization method in section 3 applied to examples shown below.

**Example 6.1.**

\[ W(x) = W_0 \exp\left\{ -6x^2 + \frac{11}{4} x^4 + \frac{1}{3} x^6 \right\} \]

\[ \dot{x} = -6x + 5.5x^3 + x^5 + n(t) \]

**Example 6.2.**

\[ W(x) = \frac{\Gamma(\alpha + 1)}{\Gamma(\frac{1}{4})\Gamma(\alpha)} \left(1 + x^2\right)^{-(\alpha+1)/2} \]

\[ x = \sinh \sqrt{2}y \]

\[ \dot{y} = -\sqrt{2}\alpha \tanh \sqrt{2}y + n(t) \]

**Example 6.3.**

\[ W(x) = \frac{\Gamma(\alpha + \gamma + 2)}{\Gamma(\alpha + 1)\Gamma(\gamma + 1)} \left(1 + x\right)\alpha \left(1 - x\right)^{\gamma} \]

\[ x = \sin \sqrt{2}y \]

\[ \dot{y} = \frac{\alpha - \gamma}{\sqrt{2} \cos \sqrt{2}y} - \frac{\alpha + \gamma + 1}{\sqrt{2}} \tan \sqrt{2}y + n(t) \]
Example 6.1 is a process with a trimodal marginal distribution. A part of simulated time series (N=800) is shown in Fig. 6.1.a. We can see the trimodal histogram from the simulated series (N=80,000) in Fig. 6.1.a. Example 6.2 is a process with the stable marginal distribution. The stable distribution is equivalent to the Cauchy distribution when $\alpha = 1/2$. A part of simulated time series (N=800) is shown in Fig. 6.1.b, where we used $\alpha = 1/2$. The histogram of the simulated series (N=80,000) shows long tails which are typical for the Cauchy distribution. A part of simulated time series (N=800) is in Fig. 6.1.c, where we used $\alpha = 1$ and $\gamma = 1$. The bell shaped histogram is shown in Fig. 6.1.c.
The representation (6.10) suggests that any Markov diffusion process has a stochastic dynamical system representation associated with an instantaneous variable transformation. It means that any non-Gaussian time series data which comes from a diffusion process can be whitened into a Gaussian white noise process using (6.11), i.e. by an instantaneous variable transformation, \( y_t = h^{-1}(x_t) \), and a nonlinear dynamic model for \( y_t \).

6.3. Different Dynamics with Same Distribution. We note that \( c(x) \) and \( d(x) \) of the Extended Pearson system (6.6) need not to be mutually irreducible. We can think of a diffusion process which corresponds to

\[
\frac{dW(x)}{dx} = \frac{c(x)x}{d(x)x} W(x),
\]

from which we have a diffusion process defined by the following Fokker-Planck equation,

\[
\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x} \left[ c(x)x + d(x) + d'(x)x \right] p + \frac{1}{2} \frac{\partial^2}{\partial x^2} [2d(x)x] p.
\]

The diffusion process has the same marginal distribution as the diffusion process of (6.7). For example the density \( W(x) \) of Gamma distribution is given by

\[
W(x) = \frac{x^{\alpha-1} \exp(-\frac{x}{\beta})}{\Gamma(\alpha) \Gamma(\beta)}.
\]

Its Pearson system form is

\[
\frac{dW(x)}{dx} = \frac{(\alpha - 1)\beta - x}{\beta x} W(x).
\]

A diffusion process for the Gamma distribution with this Pearson system is given by Wong [34] by

\[
\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x} \left[ \alpha \beta - x \right] p + \frac{1}{2} \frac{\partial^2}{\partial x^2} [2\beta x] p.
\]

If we multiply the numerator and denominator of (6.14) by \( x \) we obtain

\[
\frac{dW(x)}{dx} = \frac{(\alpha - 1)\beta x - x^2}{\beta x^2} W(x)
\]

and we have another diffusion process defined by

\[
\frac{\partial p}{\partial t} = \frac{\partial}{\partial x} \left[ (\alpha + 1)\beta x - x^2 \right] p + \frac{1}{2} \frac{\partial^2}{\partial x^2} [2\beta x^2] p.
\]

We call the process of (6.15) a type-I Gamma distributed process and the process of (6.17) a type-II Gamma distributed process. From (6.15) we have the following
stochastic dynamical system representation with an instantaneous variable transformation,

\begin{align}
\dot{x} & = \frac{\beta y^2}{2} \\
\dot{y} & = -\frac{y}{2} + n(t).
\end{align}

From (6.17) we have the following model,

\begin{align}
\dot{x} & = \exp(\sqrt{2\beta} y) \\
\dot{y} & = \frac{\alpha \beta}{\sqrt{2\beta}} - \exp(\sqrt{2\beta} y) + n(t).
\end{align}

Figs. 6.2 and Figs. 6.3 show time series data simulated by our locally linearized models for Type-I Gamma processes and for Type-II processes for several different shape parameters. Their histograms are also in the figures. We can confirm from these examples that actually many different dynamics can produce the same distribution.

In time series data analysis the model (6.18) suggests taking the square root transformation of the positive valued time series data before fitting a nonlinear time series model, while the model (6.19) suggests taking the log-transformation before fitting a nonlinear time series model. It is interesting to see that the use of these variable transformations, commonly seen in empirical non-Gaussian time series data analysis, is suggested from our theoretical analysis. In some case (Type-I Gamma process with \( \alpha = 1/2 \)) the process is transformed to a Gaussian process by a memoryless transformation, where a linear modelling is sufficient. This means that memoryless variable transformations are important for nonlinear time series models in non-Gaussian time series data analysis.

The above discussion shows that the introduction of common factors in the extended Pearson system (6.6) could lead us to infinitely many different representations of (6.11), a memoryless variable transformation with a nonlinear time series model, which produce one and the same non-Gaussian marginal distribution. A variety of different nonlinear dynamics for the same marginal distribution shows how differently a \( k \)-step ahead prediction distribution of each model approaches the marginal distribution for \( k \to \infty \). Thus the identification of a non-Gaussian distribution character in time series is very much dependent on the identification of nonlinearities of a dynamical system and a variable transformation associated with the process which generates the time series.

6.4. Second Order Dynamics with a Given Distribution. The two examples of Gamma distributed processes in the previous section were generated by combinations of a memoryless variable transformation and a nonlinear dynamical system. In both cases the dynamical systems used are one dimensional, and cannot have a vibrating mechanism contributing to the existence of a peak in the spectrum of the process. A stochastic process having both vibration character
Fig. 6.2.
gamma 2

gamma 22

gamma 23

gamma 24

Fig. 6.3.
and non-Gaussian marginal distribution character is easily obtained by combining a stochastic differential equation model of (6.1) type and one of these models of (6.10) type. If a nonlinear random vibration model has a general nonlinear restoring function \( b(x) \), instead of \( b x + T(x) \) in (6.1) the model is given by

\[(6.20) \quad \ddot{x}(t) + \alpha \dot{x}(t) + b(x) = n(t),\]

where the variance of \( n(t) \) is \( \sigma^2 \). The marginal distribution of \( x \) of (6.20) is given, in the same way as for (6.1), by

\[(6.21) \quad p(x, t) = p_0 \exp\left(\frac{-\alpha x^2}{\sigma^2}\right) \exp\left(-\frac{2a \int_{\xi}^{x} b(\xi) d\xi}{\sigma^2}\right).\]

On the other hand we know that the marginal distribution of \( y \) of (6.10) is

\[p(y) = p_0 \exp\left(\frac{2 \int_{y}^{\infty} f(\eta) d\eta}{\sigma^2}\right)\]

Then if we replace \( b(x) \) in model (6.20) by \( f(.) \) of the dynamical system of the Type-II Gamma distributed process (6.19) in model (6.20), i.e. if we take

\[(6.22) \quad \ddot{y} + \alpha \dot{y} - \frac{\alpha \beta}{\sqrt{2 \beta}} + \exp(\sqrt{2 \beta} y) = n(t)\]

we have the same distribution \( p(y) \) as the associated variable \( y \) of the Type-II Gamma process. By combining (6.22) and the variable transformation,

\[(6.23) \quad x(t) = \exp(\sqrt{2 \beta} y(t))\]

we can have another Gamma-distributed process (Type-III Gamma-distributed process) \( x(t) \) which has a vibration mechanism. Some of the Type-III Gamma distributed processes simulated by the locally linearized time series models with several shape parameter \( \alpha \)'s and their histograms are shown in Fig.6.4.

From (6.22) and (6.23) we have

\[(6.24) \quad \ddot{x} + (\alpha - \frac{\alpha \beta}{\sqrt{2 \beta}}) \dot{x} + (\sqrt{2 \beta} \varepsilon - \alpha \beta) x = \sqrt{2 \beta} x n(t),\]

which is equivalent to

\[(6.25) \quad \ddot{x} + (\alpha - \frac{\alpha \beta}{\sqrt{2 \beta}}) \dot{x} + (\sqrt{2 \beta} - \alpha \beta - \sqrt{2 \beta} n(t)) x = 0.\]

The representation (6.25) implies not only a dynamical system driven by a Gaussian white noise but also a deterministic dynamical system whose coefficients are disturbed by a Gaussian white noise produce a random process. It implies that with chaos models not only an inexact initial value but also an inexact coefficient of a deterministic dynamical system model could produce to large future uncertainty.
7. Conclusions. We have seen that Gaussian white noise still plays a very important role in nonlinear and/or non-Gaussian time series analysis as it does in
the linear Gaussian case. Any diffusion process is generated from a dynamical system model driven by Gaussian white noise associated with a memoryless variable transformation. We have also seen that any time series generated from (possibly high dimensional) stochastic dynamical system can be transformed into Gaussian white innovation sequences with an identified model by maximizing the likelihood of innovations generated by the local linearization filter. It means any diffusion process is transformed into Gaussian white noise by a memoryless variable transformation and a nonlinear dynamic model. The idea of finding a whitening operation systematically used in the linear Gaussian case in the Box-Jenkins procedure is still valid in many cases of nonlinear and/or non-Gaussian time series analysis.

It has been also shown that there can be infinitely many different nonlinear dynamical system models which produce one and the same non-Gaussian marginal distribution. Each different dynamic produces different multi-step ahead prediction distributions although the infinite-step ahead prediction distributions have the same marginal distribution. Thus the identification of non-Gaussianity in time series reduces to choosing one out of many possible combinations of variable transformations and nonlinear time series models.

The fact that any non-Gaussian distributed diffusion processes are all generated from Gaussian white noise leads to a natural question “Is non-Gaussian white noise unnecessary in non-Gaussian time series analysis?”. The answer is “Yes!” if the time series is generated from a process which is Markov and continuous in type, i.e. the transition probability from \( x \) to \( R - (x - \epsilon, x + \epsilon) \) in \( \Delta t \) time goes to zero for \( \Delta t \to 0 \) where \( \epsilon \) is a small value. This is because if the process is Markov and of continuous type then the process is a Markov diffusion process and has the Gaussian white noise representation (6.10). Non-Gaussian distributed time series does not necessarily mean the presence of non-Gaussian white noise. A nonlinear time series model driven by non-Gaussian white noise may look more general than a nonlinear time series model driven by a Gaussian white noise, but actually in many cases it is an unnecessary generalization. However, in real applications, we sometime have time series data which come from a continuous type Markov diffusion process contaminated by a Markov jump process, where we need to consider some non-Gaussian white noise with a fat tail distribution. Thus the non-Gaussian generalization of white noise in nonlinear time series models is justified in the sense of robustification of the whitening method for the identification of the associated nonlinear dynamics and variable transformation.

Our final conclusion is “Nonlinear and non-Gaussian time series analysis can be less confusing and simple if we go back to Wiener’s idea, Explain time series with Gaussian white noise”.

Acknowledgement. The author would like to thank Prof. Y. Yamanouchi and D. H. Oda for useful discussions and for providing him with the data of Fig. 2.1.

REFERENCES


