https://doi.org/10.26219/heal.aueb.6126

ATHENS UNIVERSITY
OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

POSTGRADUATE PROGRAM

ON DIFFUSION PROCESSES AND THEIR STATISTICAL INFERENCE: A REVIEW

By

Miltiadis D. Peponoulas

A THESIS
Submitted to the Department of Statistics
of the Athens University of Economics and Business
in partial fulfilment of the requirements for
the degree of Master of Science in Statistics

Athens, Greece
2004

https://doi.org/10.26219/heal.aueb.6126

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

POSTGRADUATE PROGRAM

ON DIFFUSION PROCESSES AND THEIR STATISTICAL INFERENCE: A REVIEW

By

Miltiadis D. Peponoulas

A THESIS
Submitted to the Department of Statistics
of the Athens University of Economics and Business
in partial fulfilment of the requirements for
the degree of Master of Science in Statistics

Athens, Greece
February 2004

https://doi.org/10.26219/heal.aueb.6126

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

ΣΥΝΟΨΗ ΣΤΙΣ ΔΙΑΔΙΚΑΣΙΕΣ ΔΙΑΧΥΣΗΣ ΚΑΙ ΤΗΝ ΣΧΕΤΙΚΗ ΣΤΑΤΙΣΤΙΚΗ ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ

Μιλτιάδης Δ. Πεπονούλας

ΔΙΑΤΡΙΒΗ

Που υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος Ειδίκευσης στη Στατιστική

Αθήνα
Φεβρουάριος 2004

https://doi.org/10.26219/heal.aueb.6126

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

A Thesis submitted in partial fulfillment of
the requirements for the degree of
Master of Science

ON DIFFUSION PROCESSES AND THEIR STATISTICAL INFERENCE: A REVIEW

Miltiadis D. Peponoulas

Approved by the Graduate Committee

H. Pavlopoulos	P. Dellaportas	V. Vasdekis
Assistant Professor	Associate Professor	Assistant Professor
Thesis Supervisor	Members of the Commitee	

Athens, May 2004

Michael Zazanis, Associate Professor
Director of the Graduate Program

https://doi.org/10.26219/heal.aueb.6126

DEDICATION

To my parents.

To Helen.

https://doi.org/10.26219/heal.aueb.6126

ACKNOWLEDGEMENTS

I would like to give many thanks to my supervisor Assistant Professor Harry Pavlopoulos for his valuable help and guidance towards the completion of this dissertation.

I would also like to thank my parents for their support.

VITA

I was born in Athens in 1973. After my graduation from High School in 1991, I enrolled the Mathematics Department of the University of Patras. I concluded my studies at February 1996 emphasizing on Statistics, Probabilities and Operational Research. In March 1997 I joined the Army and my service was completed by September 1998. I was admitted as a post-graduate student in the M.Sc. Program of the Department of Statistics of the Athens University of Economics and Business in September 1999.

https://doi.org/10.26219/heal.aueb.6126

https://doi.org/10.26219/heal.aueb.6126

ABSTRACT

Miltiadis Peponoulas

ON DIFFUSION PROCESSES AND THEIR STATISTICAL INFERENCE: A REVIEW

February 2004

This paper reviews basic stochastic theory applied in diffusions and provides current statistical inference for such processes. Apart from the definitions of diffusions and the stochastic differential equations that they fulfill, basic results of the Itô theory are presented. We establish the connection between stochastic differential equations and diffusions and present Kolmogorov's backward and forward equations. Statistical inference comprises the problem of identification and estimation of the infinitesimal drift and diffusion function along to the estimation of the diffusion's marginal density. Both nonparametric and parametric methods are discussed. Strict stationarity for the marginal density is assumed in all cases but one; in that case recurrence is the identifying feature of the density.

ΠΕΡΙΛΗΨΗ

Μιλτιάδης Πεπονούλας

ΣΥΝΟΨΗ ΣΤΙΣ ΔΙΑΔΙΚΑΣΙΕΣ ΔΙΑΧΥΣΗΣ ΚΑΙ ΤΗΝ ΣΧΕΤΙΚΗ ΣΤΑΤΙΣΤΙΚΗ ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ

Φεβρουάριος 2004

Η παρούσα διατριβή συνοψίζει βασική στοχαστική θεωρία εφαρμοσμένη σε διαχύσεις και παρέχει τρέχουσα στατιστική συμπερασματολογία για τέτοιες διαδικασίες. Εκτός των ορισμών των διαχύσεων και των στοχαστικών διαφορικών εξισώσεων που αυτές ικανοποιούν, παρουσιάζονται και βασικά αποτελέσματα της θεωρίας του Itô. Θεμελιώνουμε την σύνδεση μεταξύ στοχαστικών διαφορικών εξισώσεων και διαχύσεων και παρόντων ανάδρομων και πρόδρομων εξισώσεων του Kolmogorov. Η στατιστική συμπερασματολογία συναποτελείται από το πρόβλημα του προσδιορισμού και εκτίμησης των συναρτήσεων απειροστού συντελεστή μετατόπισης και συντελεστή διαχύσεως, παράλληλα με την εκτίμηση της οριακής πυκνότητας πιθανότητας της διάχυσης. Και οι δυο μέθοδοι, παραμετρικές και απαραμετρικές, συζητώνται. Η αυστηρά στασιμότητα έχει υποτεθεί για την οριακή πυκνότητα πιθανότητας για όλες τις περιπτώσεις πλην μιας˙ σε αυτήν την περίπτωση η επισκεψιμότητα είναι το χαρακτηριστικό γνώρισμα της πυκνότητας πιθανότητας.

https://doi.org/10.26219/heal.aueb.6126

https://doi.org/10.26219/heal.aueb.6126

TABLE OF CONTENTS

https://doi.org/10.26219/heal.aueb.6126

https://doi.org/10.26219/heal.aueb.6126

Chapter 1

Introduction

Stochastic processes in general are being increasingly used in scientific research in order to meet the growing demand for modelling the evolution of various random systems. Constricting ourselves into the class of continuous time and state space stochastic processes, where diffusion processes lie, it is foreseeable that it will burden the greater amount of such demand. Diverse fields of academic endeavor and/or more applied activity, such as optimal control theory, financial economics, and statistical thermodynamics are substantially benefitted by the theory of diffusions and the stochastic differential equation that describe them. Numerous examples that can be found in Physics, Biology, Economics or even Social Sciences assert the validity of the previous sentence and raise the expectations for possible applications of the theory in discuss. Such examples range from molecular motions of enumerable particles subject to interactions to security price fluctuations in a perfect market and from neurophysiological activity with disturbances to variations of population growth. The main equation describing diffusions(see (2.4)) involves two components: the infinitesimal drift $\mu(x)$, and the infinitesimal variation $\sigma^2(x)$ whose identification, along to the related problem of the estimation of the process's marginal density, are the composing parts of the diffusion identification problem. A decisive step for the analysis of diffusions and interrelated problems was the development of a fully operational "stochastic calculus" by Itô (1951). Itô managed to extend the standard tools of calculus to functions of a wide class of continuous-time random processes, now known as Itô processes. Stochastic integration was the key part for the derivation of a solution to the stochastic differential equation (2.4) describing Itô diffusions. Even more complex models driven by Itô processes can be readily accomodated by the Itô transformation formula. This formula actually describes the closure of the class of diffusions under quite general nonlinear transformations. This was undoubtly the result that introduced diffusion theory to a hole new variety of disciplines e.g. in finance's field of derivative pricing where derivative security prices when expressed as functions of the prices of underlying assets, easily identified as diffusions, are then subject to Itô's stochastic calculus.

However the purpose of this dissertation is not only to review the probabilistic framework of diffusion processes but also to provide current statistical inference for these processes. The problems of identification, estimation and investigation of the asymptotic sampling properties of the continuous-time diffusion process estimators have proved to be quite difficult and still remain a challenge to researchers. These problems arise of course by the highly dynamic behavior of diffusions along to unavailability of continuous sampling observations. With the, already considerable, relative literature being updated the moment we speak, we have no choice but to make a brief assessment to the most important papers that formed the existing diffusion theory and report the latest major developments especially in the domain of nonparametric estimation. Diffusion theory's historic advance began as expected with minimum considerations on sampling observations which were originally considered as continuous by all distinguished authors. For instance parametric estimators for the drift and/or diffusion function have been proposed by Brown and Hewitt (1975), Vasicek (1977), Lanska (1979), Brennan and Schwartz (1979, 1982), Kutoyants (1984) and Cox, Ingersoll and Ross (1985). Nonparametric counterparts have been proposed by Geman (1979), Pham Dinh (1981), and Banon and Nguyen (1981). Also Banon (1978) is usually accounted in the previous category by reviewers, but we prefer to identify his approach as semi-parametric. Discrete observations is a set that can be definitely considered as the starting point for drift and diffusion coefficient estimation; it is quite straightforward to assert the maximum likelihood method applicability to the problem at hand provided the exact transitional density function or marginal density function is known. But even so the complexity of the required calculations is discouraging except for special cases. The introducing paper to parametric estimation of the diffusion process based on a set of discrete observations was the one by Dacunha-Castelle and Florens-Zmirou (1986). They were followed by Dohnal (1987), Lo (1988), Chan, Karolyi, Longstaff and Sanders (1992), Duffie and Singleton (1993), Pedersen (1995), and Hansen and Scheinkman (1995). Alternating to nonparametric estimation still on the set of discrete sampling observations we study the pioneering paper of Florens-Zmirou (1993) for diffusion term estimation. Jiang and Knight (1997) refined some of her findings along to providing the variance of his consistent diffusion estimator. A series of approximations to the true drift and diffusion were constructed by Stanton (1996) while Chapman and Pearson (2000), Li and Tcacz (2002) and Fan and Zhang (2003) commented on his work and extended some of his findings. Bandi and Phillips (2002) and Moloche (2001) try to surpass the demand of stationarity of the process and build their estimators based on reccurence. Ait-Sahalia's (1996) approach is confronted as semi-parametric.

Diffusion's marginal density estimation problem can be viewed as a part of the coefficient estimation problem since density needs to be consistent with the drift and diffusion coefficients. Banon (1978) for example exploits this consistency in order to deduct the estimator of the drift. Nevertheless diffusion's density estimation is more often treated as a standalone issue and relative papers are provided by Prakasa (1979, 1983), Leblanc (1995) and Yamato (1971) among others.

The remainder of the dissertation is organized as follows. In chapter 2 we present the probabilistic and stochastic framework of diffusions. In section 2.1 we derive the stochastic differential equation that diffusions fulfill. Definitions of diffusion processes are discussed in section 2.2. Issues concerning stochastic integrals and the Itô equation are presented in 2.3. Theorems that connect diffusions with the stochastic differential equation built to describe them, are the issue of section 2.4. In section 2.5 we are concerned with three interesting problems of diffusion processes which are also introducing to important functionals. Kolmogorov's backward and forward equations are discussed in 2.6 along to the stationary distribution's connection with infinitesimal drift and diffusion. Behavior at the boundaries is the subject of section 2.7 and the chapter concludes with the Brownian motion paradigm in section 2.8. In chapter 3 we mainly provide statistical inference for diffusion processes. Preliminaries about kernel density estimators and kernel smoothing are provided in section 3.1. Various methods for diffusion marginal density estimation are presented in section 3.2. Nonparametric estimators for infinitesimal drift and diffusion are discussed in section 3.3 while their parametric analogues are confronted in section 3.4. Chapter 4 concludes.

https://doi.org/10.26219/heal.aueb.6126

Chapter 2

A review on probabilistic theory for one-dimensional diffusions

2.1 An illuminating physical example

In terms of being comprehensive about diffusion processes we shall study a physical example of a diffusion phenomenon and draw some useful conclusions.

Such an example that is very popular through the bibliography, is the motion of small particles suspended in a homogeneous liquid under the influence of collisions with the molecules of the liquid in chaotic thermal motion. The motion of each particle can be attributed to two principal forces. First an underlying fluid flow or some external force impressed on the system that engenders the determenistic (nonrandom) part of the motion. Second, collisions or other more general interaction relationships with other particles which cause generally random movements. So if by X_t we denote the coordinate of the particle at instant t, and by $\mu(t, x)$ the velocity of the motion of the liquid at point x and instant t, then the displacement of the particle for a small duration from time t to $t + \Delta t$ will be approximated by:

$$X_{t+\Delta t} - X_t = \mu(t, X_t)\Delta t + \xi_{t,X_t,\Delta t} \tag{2.1}$$

https://doi.org/10.26219/heal.aueb.6126

The fluctuational component of the displacement is indicated in the equation above by $\xi_{t,X_t,\Delta t}$, a random variable whose distribution depends on the position x of the particle, the time instant t of displacement observance and Δt the length of time interval. Under reasonable assumptions this is a process with independent increments for which stands that: $E\xi_{t,X_t,\Delta t} = 0$. Since the properties of the medium are naturally assumed to change only slightly for small changes in t and x, the process is also homogeneous. Therefore we may assume that:

$$\xi_{t,X_t,\Delta t} = \sigma(t, X_t)\xi_{t,\Delta t} \tag{2.2}$$

where $\sigma(t,x)$ characterizes the properties of the medium at the time-space point (t,x) and $\xi_{t,\Delta t}$ is the value of the increment that is obtained in the homogeneous case under the condition that $\sigma(t,x) = 1$.

We know that every continuous homogeneous process with independent increments is a Gaussian process. Recalling the definition of a Brownian motion $B(t)$, $t \succeq 0$ as a stationary Gaussian process with independent increments of zero mean and variance $\sigma^2 t$, σ fixed, brings us to the conclusion that $\xi_{t,\Delta t}$ must be distributed like the increment of a process of Brownian motion. Thus we reach the formula:

$$X_{t+\Delta t} - X_t = \mu(t, X_t)\Delta t + \sigma(t, X_t)[B(t + \Delta t) - B(t)] \tag{2.3}$$

If we replace the increments by differentials, we obtain the differential equation:

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB(t) \tag{2.4}$$

which can be taken as a starting point for defining diffusion processes. The methodology of stochastic differential equations is the purely probabilistic approach to diffusions, in a range of approaches that varies from these to the purely analytical. It was suggested by P.Lèvy as an alternative to the existing analytical approach and was carried out in a masterly way by K.Itô(1942a,1946,1951). Since certain theoretical analysis regarding

https://doi.org/10.26219/heal.aueb.6126

stochastic integrals and stochastic differentials must be preceded to the solution of the equation (2.4), along with a comment on the precise meaning of the derivative of $B(t)$, we shall henceforward illustrate the traditional analytical approach to a diffusion.

2.2 The definition of diffusions and regarding issues

Diffusion processes are in fact special cases of Markov processes, therefore we must illuminate some theoretical issues regarding them.

A *Markov process* is a stochastic process $\{X_t, t_0 \leq t \leq T\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$ with state space R for which the following Markov property is satisfied: For $n \geq 1, t_0 \leq t_1 \leq t_2 \leq ... < t_n \leq t$ and $B \in \mathcal{B}$,

$$P(X_t \in B/X_{t_1}, ...X_{t_n}) = P(X_t \in B/X_{t_n}) \tag{2.5}$$

where $\mathcal{B}$ is the σ-algebra of Borel sets of $\mathbf{R}$.

Loosely speaking, Markov property implies that if the state of the system at a particular time s is known, additional information regarding the behaviour of the system at past times $t < s$ has no effect on our knowledge of the probable development of the system at future times t, $t > s$.

A function that is basic to the study of Markov processes is the *transition probability function* which is defined as follows:

$$P(s, x; t, A) = P(X_t \in A/X_s = x) \tag{2.6}$$

for $t > s$. When the transition probability $P(s, x; t, A)$ is stationary, that is, if the condition:

$$P(s + h, x; t + h, A) = P(s, x; t, A) \tag{2.7}$$

holds for $t_0 \leq s \leq t \leq T$ and $t_0 \leq s + h \leq t + h \leq T$,the Markov process is said to be *homogeneous* (with respect to time).

A final issue to be introduced before diffusions are to be defined, is the *strong Markov property.* This property restricts the probability distribution of

$$X(t_1+\sigma), X(t_2+\sigma), ..., X(t_\kappa+\sigma) \tag{2.8}$$

with $t_1 < t_2 < ... < t_\kappa$, given $X(s)$, $s \leq \sigma$, and $X(\sigma) = x$, to be identical with the probability distribution of

$$X(t_1), X(t_2), ..., X(t_\kappa) \tag{2.9}$$

given $X(0) = x$ for any Markov time σ (which is actually a random time). We remind that a random variable σ is considered to be a Markov time relative to a given process $\{X_t\}$, $0 \leq t < +\infty$, if for two sample functions of the process X_t and Y_t, such that $X_\tau = Y_\tau$ for $0 \preceq \tau \prec s$ and $\sigma(X_t) < s$, it holds that $\sigma(X_t) = \sigma(Y_t)$.

According to Karlin and Taylor (1981), a *diffusion process* is a continuous time parameter stochastic process which possesses the strong Markov property and for which the sample paths $X(t)$ are (almost always) continuous functions of the time parameter t.

A diffusion process $\{X(t), t > 0\}$ is said to be *regular* if starting from any point in the interior of its state space I any other point in the interior of I may be reached with positive probability. The state space of $X(t)$ is actually an interval with endpoints l, r with $l < r$ and necessarily of the form $(l,r), (l,r], [l,r), [l,r]$, where $l = -\infty$ and/or $r = +\infty$ is an option.

Although the definition above is more in order with the modern setup, we cannot ignore a more traditional approach found through a great part of the bibliography, which is also confronting to essential notions of diffusion theory. This definition is the following:

A Markov process $\{X_t, t_0 \leq t \leq T\}$ with state space R and continuous sample functions is called diffusion process if its transition probability $P(s, x; t, A)$ satisfies the following three conditions for every $s \in [t_0, T]$, $x \in R$ and $\varepsilon > 0$:

$$\text{(i)} \lim_{t \downarrow s} \int_{|y-x|>\varepsilon} P(s, x; t, dy) = 0 \tag{2.10}$$

(ii) there exists an R-valued function $\mu(s,x)$ such that

$$\lim_{t\downarrow s}(t-s)^{-1}\int_{|y-x|\leq\epsilon}(y-x)P(s,x;t,dy)=\mu(s,x) \tag{2.11}$$

(iii) there exists a real non-negative function $\sigma^2(s,x)$ such that

$$\lim_{t\downarrow s}(t-s)^{-1}\int_{|y-x|\leq\epsilon}(y-x)^2P(s,x;t,dy)=\sigma^2(s,x) \tag{2.12}$$

the functions μ and σ^2 are termed the *infinitesimal parameters* of the process, and, in particular, $\mu(t,x)$ the drift coefficient or *infinitesimal mean* and $\sigma^2(t,x)$ the diffusion coefficient or *infinitesimal variance* of the diffusion process X. The justification of the terminology used, becomes more visible when the truncated moments of the above relationships are replaced by the regular moments. The latest need not necessarily exist, but this is not the case when the following condition holds

$$E\{|X_t-X_s|^{2+\delta}/X_s=x\}=\int_R|y-x|^{2+\delta}P(s,x;t,dy)=o(t-s) \tag{2.13}$$

for some $\delta>0$. Supposing (2.13) the regions of integration in (ii) and (iii) can be chosen to be R and we are brought to

$$E(X_t-X_s/X_s=x)=\mu(s,x)(t-s)+o(t-s) \tag{2.14}$$

and

$$E\{(X_t-X_s)^2/X_s=x\}=\sigma^2(s,x)(t-s)+o(t-s) \tag{2.15}$$

which if seen under the equivalent form

$$E\{\frac{X_t-X_s}{t-s}/X_s=x\}=\mu(s,x)+\frac{o(t-s)}{t-s} \tag{2.16}$$

and

$$E\{\frac{(X_t-X_s)^2}{t-s}/X_s=x\}=\sigma^2(s,x)+\frac{o(t-s)}{t-s} \tag{2.17}$$

can have the following interpretation: $\mu(s,x)$ is a mean rate of change of the process while

https://doi.org/10.26219/heal.aueb.6126

visiting x at time s, while $\sigma^2(s,x)$ is a measure of the local magnitude of the fluctuation of $X_t - X_s$ about this mean value.

As far as condition (i) is concerned it develops to

$$P(|X_t - X_s| > \varepsilon / X_s = x) = o(t-s) \tag{2.18}$$

which actually means that large changes in X_t over a short period of time are improbable.

It is of value to have verifiable sufficient conditions under which a Markov process is a diffusion process. The concept of *standard process* is now introduced in order to facilitate this purpose. A strong Markov process $\{X(t), t \geq 0\}$ is called a standard process if the sample paths possess the following regularity properties:

(i) $X(t)$ is right continuous i.e. $\lim_{t \downarrow s} X(t) = X(s)$ for all $s \succeq 0$

(ii) left limits of $X(t)$ exist, i.e. $\lim_{t \uparrow s} X(t)$ exists for all $s > 0$ and

(iii) $X(t)$ is continuous from the left through Markov times, i.e. if $T_1 \leq T_2 \leq \ldots$ are Markov times converging to finite T then $\lim_{n \to \infty} X(T_n) = X(T)$

A sufficient condition that a standard process be a diffusion is then the fulfillment of the *Dynkin condition:*

$$\frac{1}{h} P\{|X(t+h) - X(t)| > \varepsilon / X(t) = x\} \to 0 \tag{2.19}$$

when $\varepsilon > 0$, as $h \downarrow 0$, where the convergence prevails uniformly for x restricted to any compact subinterval of the state space and t traversing any finite interval $[0, N]$.

https://doi.org/10.26219/heal.aueb.6126

2.3 Stochastic Integrals: solution of the stochastic differential equations

We can now focus to the solution of the stochastic differential equation (2.4), which by applying the usual integration notation can be expressed

$$X_t = X_0 + \int_0^t \mu(s, X_s)ds + \text{“}\int_0^t \sigma(s, X_s)dB_s\text{”} \tag{2.20}$$

In attaching meaning to the last term of (2.20) it is impossible to employ the standard calculus of integrals because almost every sample path of $B(t)$ is of unbounded variation. Therefore it is imperative to prove the existence of this term and study the sense of integration. Let us first consider an elementary function φ,which is a functions of the form

$$\phi(t,\omega) = \sum_{j\geq 0} e_j(\omega) I_{[t_j,t_{j+1})}(t) \tag{2.21}$$

where I is the indicator function and each e_j is an $\mathcal{F}_{t_j}$-measurable function. We remind at this point that $\mathcal{F}_t$ is the σ-algebra generated by the random variables B_s with $s \leq t$. For such functions it is reasonable to define

$$\int_S^T \varphi(t,\omega)dB_t(\omega) = \sum_{j\geq 0} e_j(\omega)[B_{t_{j+1}} - B_{t_j}](\omega) \tag{2.22}$$

The next step is to approximate a given function $f(t, \omega)$ by

$$\sum_j f(t_j^*, \omega) I_{[t_j,t_{j+1}]}(t) \tag{2.23}$$

where $t_j^* \in [t_j, t_{j+1}]$ and then define $\int_S^T f(t,\omega)dB_t(\omega)$ as an appropriate limit of the quantity (2.23) as $n \to \infty$. Unlike the Riemann-Stieltjes integral the choice of t_j^* influences the final result. The most useful choices have turned out to be the

https://doi.org/10.26219/heal.aueb.6126

1. $t_j^* = t_j$ (the left end point), which leads to the *Ito integral* from now on simply denoted as $\int_S^T f(t,\omega)dB_t(\omega)$

2. $t_j^* = \frac{1}{2}(t_j + t_{j+1})$ (the mid point), which leads to the *Stratonovich integral*, from now on denoted by $S - \int_S^T f(t,\omega)dB_t(\omega)$

It can be proven that for every $f \in \mathcal{V}$, there is a sequence of elementary functions $\phi_v \in \mathcal{V}$ such that

$$E\left\{\int_S^T |f - \phi_n|^2 \, dt\right\} \to 0 \tag{2.24}$$

for such a sequence we define

$$\int_S^T f(t,\omega)dB_t(\omega) = \lim_{n\to\infty} \int_S^T \phi_n(t,\omega)dB_t(\omega) \tag{2.25}$$

The S-integral generally differs from the Ito integral by a corrective term. If we name the solutions of stochastic differential equations after the type of the integrals been used, then for the most pertinent equation to our study

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB(t) \tag{2.26}$$

the S-solution, or coincidingly the Wong-Zakai solution, will be $X(t)$ satisfying

$$X_t = X_0 + \int_0^t \mu(s, X_s)ds + S - \int_0^t \sigma(s, X_s)dB_s \tag{2.27}$$

For the I-solution to agree with the solution given above, the stochastic differential equation (2.26) should be modified to

$$dX_t = [\mu(t, X_t) + \frac{1}{2}\sigma_x(t, X_t)\sigma(t, X_t)]dt + \sigma(t, X_t)dB(t) \tag{2.28}$$

involving the correction term $\frac{1}{2}\sigma_x\sigma$ contributing to the infinitesimal drift coefficient. This implies that X_t is the solution of the following modified *Ito equation*

$$X_t = X_0 + \int_0^t \mu(s, X_s)ds + \frac{1}{2}\int_0^t \sigma_x(s, X_s)\sigma(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s \tag{2.29}$$

For $\sigma(t, x)$ independent of x the S and the I-solution obviously coincide since $\sigma_x = 0$. Even though the Stratonovich integral presents several advantages such as compliance with the ordinary chain rule formula under a transformation, the Ito integral will be our tool of integration henceforward. This is mainly due to the exclusive martingale property of the Ito integral.

Calculations concerning Ito integrals are facilitaded by the *Ito stochastic transformation formula* which for $Y(t) = f(t, X(t))$, $f \in C^2([0, \infty) \times R)$ (i.e. f twice continiously differentiable on $[0, \infty) \times R$) and $X(t)$ the Ito solution of (2.26) gives that Y_t is again an Ito process, and

$$\begin{aligned} dY_t &= [f_x(t, X(t))\mu(t, X(t)) + f_t(t, X(t)) + \frac{1}{2}f_{xx}(t, X(t))\sigma^2(t, X(t))]dt \\ &\quad + f_x(t, X(t))\sigma(t, X(t))dB(t) \end{aligned} \tag{2.30}$$

or in integral form

$$\begin{aligned} Y(\tau) - Y(0) &= \int_0^\tau [f_x(t, X(t)\mu(t, X(t)) + f_t(t, X(t)) \\ &\quad + \frac{1}{2}f_{xx}(t, X(t))\sigma^2(t, X(t))]dt + \int_0^\tau f_x(t, X(t))\sigma(t, X(t))dB(t) \end{aligned} \tag{2.31}$$

The conditions that are necessary for the existence and uniqueness for solutions of the stochastic differential equation

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB(t) \tag{2.32}$$

equivalent to the integral notation

$$X_t = X_0 + \int_0^t \mu(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s \tag{2.33}$$

are the following

1. *Growth condition:* There exists a constant K independent of $0 \leq t \leq T$ and $-\infty < x < \infty$ such that

$$\mu^2(t,x)d + \sigma^2(t,x) \leq K(1+x^2), \qquad -\infty < x < \infty \tag{2.34}$$

   In the context of ordinary differential equations a growth restriction is essential in order to be assured that the solution can be continued for the total time horizon $0 \leq t \leq T$ without exploding to infinite at an intermediate time point.

2. *Lipschitz conditions:* There exists a constant L independent of t, $0 \leq t \leq T$, and of x, $-\infty < x, y < \infty$ such that

$$|\mu(t,x) - \mu(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leq L\,|x-y| \tag{2.35}$$

   The Lipschitz condition is guaranteeing uniqueness for solutions of ordinary differential systems.

For μ and σ satisfying conditions (2.34) and (2.35) and $E[\{X(0)\}^2] < \infty$ there exists a unique solution of (2.33) as a continuous process. This proposition is actually the first existence theorem for solutions of stochastic differential equations.

There are two kinds of solution for (2.32), the *weak* and *strong* solution. Sketching out their differences we should note that if we are only given the functions $\mu(t,x)$ and $\sigma(t,x)$ and ask for a pair of processes $((X_t, B_t), \mathcal{F}_t)$ on a probability space $(\Omega, \mathcal{F}, P)$ such that (2.32) holds, then the solution (X_t, B_t) is called a weak solution. If furthermore the solution X_t is constructed upon a given version of Brownian motion B_t and is

https://doi.org/10.26219/heal.aueb.6126

$\mathcal{F}_t^Z$–adapted, then is called strong solution.

2.4 The connection between stochastic differential equations and diffusions

We shall henceforward establish the connection between stochastic differential equations and diffusion processes assisted by the following theorems:

Theorem 1: Assume that $\eta(t)$ is a solution of

$$d\eta(t) = \mu(t, \eta(t))dt + \sigma(t, \eta(t))dB(t) \tag{2.36}$$

where coefficients satisfy the conditions for existence and uniqueness of solution. Then $\eta(t)$ will be a Markov process whose transition probability is defined by

$$P(t, x; s, A) = P(\eta_{x,t}(s) \in A) \tag{2.37}$$

where $\eta_{x,t}(s)$ is a solution of

$$\eta_{x,t}(s) = x + \int_t^s a(u, \eta_{x,t}(u))du + \int_t^s \sigma(u, \eta_{x,t}(u))dB(u) \tag{2.38}$$

on the interval $[t, T]$.

This theorem certifies that equation (2.36) is valid as a stochastic integral equation in the sense of Ito. The following theorem clarifies the conditions under the process $\eta(t)$ is actually a diffusion

Theorem 2: Let $\mu(t, x)$ and $\sigma(t, x)$ be continuous in both arguments and assume that for some K

$$|\mu(t, x)|^2 + |\sigma(t, x)|^2 \leq K(1 + |x|^2) \tag{2.39}$$

and that for each N, there exists L_N with $|x| \leq N$, $|y| \leq N$ for which

$$|\mu(t,x) - \mu(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leq L_N |x-y| \tag{2.40}$$

Then the process $\eta(t)$ as a solution of (2.36) will be a diffusion with diffusion coefficient $\sigma^2(t,x)$ and drift coefficient $\mu(t,x)$.

Inversely a diffusion process satisfies the stochastic differential equation (2.36) only if the requirements of the following theorem are met:

Theorem 3: Let $\xi(t)$ be a diffusion process on $[0,T]$ with coefficients $\mu(t,x)$ and $b(t,x)$ which satisfy the following properties

(i) $\mu(t,x)$ is continuous in both arguments and for some K satisfies

$$|a(t,x)| \leq K(1+|x|) \tag{2.41}$$

(ii) $b(t,x)$ is continuous in both arguments and has continuous bounded derivatives $\partial/\partial t\, b(t,x)$ and $\partial/\partial x\, b(t,x)$, also $1/b(t,x)$ is bounded

(iii) there exists a function $\psi(x)$, independent of t and Δ, for which

$$\psi(x) > 1+|x|, \quad \sup_{0\leq t\leq T} E[\psi(\xi(t))] < \infty \tag{2.42}$$

and

$$\left|\int (y-x)P(t,x;t+\Delta,y)\right| + \int (y-x)^2 P(t,x;t+\Delta,dy) \leq \psi(x)\Delta, \tag{2.43}$$

$$\int (|y|+y^2)P(t,x;t+\Delta,dy) \leq \psi(x) \tag{2.44}$$

Then there exists a Wiener process $B(t)$ for which $\xi(t)$ satisfies the stochastic differential equation

$$d\xi(t) = \mu(t,\xi(t))dt + [b(t,\xi(t))]^{\frac{1}{2}} dB(t) \tag{2.45}$$

In view of these theorems, the solutions of stochastic differential equations and dif-

https://doi.org/10.26219/heal.aueb.6126

fusion processes comprise one and the same class of processes, under the already stated conditions. Proofs of the theorems can be found in Gikhman and Skorokhod (1972), Arnold (1974) and Friedman (1975).

2.5 Functionals associated with differential equations

We shall now introduce three problems concerning diffusions. During the presentation of their solutions a number of important functionals will be discussed which apart from their physical importance and extensiveness in the way of anticipating diffusions, they are of great value in problem solving.

We assume for the pending problems a time homogeneous diffusion process $\{X(t), t \succeq 0\}$ which satisfies the following conditions:

1. The state space is an interval of the form $[l, r], (l, r], [l, r)$, or (l, r) where $-\infty \leq l < r \leq \infty$.

2. The process is regular in the interior of I.i.e.

$$\Pr\{T(y) < \infty / X(0) = x\} > 0, \qquad l < x, y < r \tag{2.46}$$

   where $T(y)$ is the hitting time of the value y

3. The process has infinitesimal parameters $\mu(x)$ and $\sigma^2(x)$ which are continuous functions of x, $l < x < r$

We also let a and b be fixed, with $l < a < b < r$, and note $T(y) = T_y$ and

$$T^* = T_{a,b} = \min\{T(a), T(b)\} \tag{2.47}$$

the first time the process reaches either a or b. The problems are now concentrated on finding the following quantities.

https://doi.org/10.26219/heal.aueb.6126

- **Problem A** The probability that the process reaches b before a or

$$u(x) = \Pr\{T(b) < T(a)/X(0) = x\}, \qquad a < x < b \tag{2.48}$$

- **Problem B** The mean time to reach either a or b

$$v(x) = E[T^*/X(0) = x], \qquad a < x < b \tag{2.49}$$

- **Problem C** The mean value of the integral stated below

$$w(x) = E\left[\int_0^{T^*} g(X(s))ds/X(0) = x\right], \qquad a < x < b \tag{2.50}$$

  where g bounded and continuous function.

Functions u, v and w, are correspondingly found to satisfy the following equations:

- **Equation A**

$$\mu(x)\frac{du}{dx} + \frac{1}{2}\sigma^2(x)\frac{d^2u}{dx^2} = 0 \tag{2.51}$$

  for $a < x < b$ and boundary conditions $u(a) = 0$, $u(b) = 1$

- **Equation B**

$$\mu(x)\frac{dv}{dx} + \frac{1}{2}\sigma^2(x)\frac{d^2v}{dx^2} = -1 \tag{2.52}$$

  for $a < x < b$ and boundary conditions $v(a) = v(b) = 0$

- **Equation C**

$$\mu(x)\frac{dw}{dx} + \frac{1}{2}\sigma^2(x)\frac{d^2w}{dx^2} = -g(x) \tag{2.53}$$

  for $a < x < b$ and boundary conditions $w(a) = w(b) = 0$

We observe that each equation **A-C** involve the differential operator L defined by

$$Lf(x) = \mu(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x) \tag{2.54}$$

https://doi.org/10.26219/heal.aueb.6126

for $f(x)$ a twice conditionally differentiable function on (a, b). For the purpose of solving equations **A-C** we introduce the following functions

$$s(x) = \exp\left\{-\int^{x}[2\mu(\xi)/\sigma^2(\xi)]d\xi\right\} \quad \text{for } l < x < r \tag{2.55}$$

and

$$S(x) = \int^{x} s(\eta)d\eta = \int^{x} \exp\left\{-\int^{x}[2\mu(\xi)/\sigma^2(\xi)]d\xi\right\} d\eta \tag{2.56}$$

Indefinite integrals are used since the results will prove to be independent of the lower limits of integration. Function $S(x)$ is called *scale function* of the process, while $m(x)$, which is given right away, is the *speed density*

$$m(x) = 1/[\sigma^2(x)s(x)] \quad \text{for } l < x < r \tag{2.57}$$

We also introduce the *speed measure* M which similarly to S is given by

$$M(x) = \int^{x} m(u)du \tag{2.58}$$

Of course we expect then that $dS = s(x)dx$ and $dM = m(x)dx$.

The L operator is proved by theory to be connected to the functions above through the expression

$$Lf(x) = \frac{1}{2}\frac{d}{dM}\left[\frac{df(x)}{dS}\right] \tag{2.59}$$

called *the canonical representation of the differential infinitesimal operator* associated with the diffusion process. Conclusively **equation A** will be brought to

$$\frac{1}{2}\frac{d}{dM}\left[\frac{du(x)}{dS}\right] = 0 \tag{2.60}$$

which by two succesive integrations will produce solution A

$$u(x) = \frac{S(x) - S(a)}{S(b) - S(a)} \quad \text{for } a \le x \le b \tag{2.61}$$

https://doi.org/10.26219/heal.aueb.6126

Proceding to **equation C** it will equivalently become

$$\frac{1}{2}\frac{d}{dM}\left[\frac{dw(x)}{dS}\right] = -g(x) \quad \text{for } a \leq x \leq b \tag{2.62}$$

which will similarly produce **solution C**

$$w(x) = 2\left\{u(x)\int_x^b [S(b) - S(\xi)]m(\xi)g(\xi)d\xi + [1 - u(x)]\int_a^x [S(\xi) - S(a)]m(\xi)g(\xi)d\xi\right\} \tag{2.63}$$

We can easily ascertain that **problem B** is only a special case of **problem C** for $g(x) = 1$ for all x, therefore by substituting $g(\xi)$ with 1 in solution C we obtain **solution B**

$$v(x) = 2\left\{u(x)\int_x^b [S(b) - S(\xi)]m(\xi)d\xi + [1 - u(x)]\int_a^x [S(\xi) - S(a)]m(\xi)d\xi\right\} \tag{2.64}$$

2.6 Backward and Forward Equations and Stationary Distribution

For what follows we shall assume that $\{X(t), t \succeq 0\}$ is a regular time homogeneous diffusion process on the open interval $I = (l, r)$ with transition distribution function $P(t, x, y) = \Pr\{X(t) \leq y / X(0) = x\}$. The initial distribution is assumed to be concentrating at point x i.e.

$$P(0, x, y) = \begin{cases} 1, & \text{if } x \leq y \\ 0, & \text{if } x > y \end{cases} \tag{2.65}$$

Furthermore it is assumed that the transition density function $p(t, x, y)$ given by

$$p(t, x, y) = \frac{dP(t, x, y)}{dy}, \qquad t > 0 \tag{2.66}$$

is continuous on (l, r). In the spirit of the previous section we shall produce a partial differential equation for the function

$$u(t, x) = E[g(X(t)) / X(0) = x] \tag{2.67}$$

https://doi.org/10.26219/heal.aueb.6126

with $g(x)$ bounded and piecewise continuous on I. Under mild conditions $u(t,x)$ satisfies the partial differential equation

$$\frac{\partial u}{\partial t} = \mu(x)\frac{\partial u}{\partial x} + \frac{1}{2}\sigma^2(x)\frac{\partial^2 u}{\partial x^2} \tag{2.68}$$

with the initial condition $u(0+,x) = g(x)$, where $u(0+,x) = \lim_{h\longrightarrow 0} u(h,x)$. If we specify

$$g(\eta) = \begin{cases} 1, & \text{if } \eta \leq y \\ 0, & \text{if } \eta > y \end{cases} \tag{2.69}$$

we acquire that $u(t,x) = P(t,x,y)$. In this case equation (2.68) is formulated

$$\frac{\partial P(t,x,y)}{\partial t} = \mu(x)\frac{\partial P(t,x,y)}{\partial x} + \frac{1}{2}\sigma^2(x)\frac{\partial^2 P(t,x,y)}{\partial x^2} \tag{2.70}$$

and refered to as the *Kolmogorov backward equation,* applicable for $t > 0$ and $l < x, y < r$. As expected the initial condition attendant to (2.70) is

$$P(0+,x,y) = \begin{cases} 1, & \text{if } x \leq y \\ 0, & \text{if } x > y \end{cases} \tag{2.71}$$

The Kolmogorov backward equation is also satisfied by the transition density $p(t,x,y)$ thus

$$\frac{\partial p(t,x,y)}{\partial t} = \mu(x)\frac{\partial p(t,x,y)}{\partial x} + \frac{1}{2}\sigma^2(x)\frac{\partial^2 p(t,x,y)}{\partial x^2} \tag{2.72}$$

for $t > 0$ and $l < x, y < r$. It should be noted that equations (2.70) and (2.72) do not always admit unique solutions. This problem is actually connected to the boundary conditions of diffusions studied later on in this chapter.

The transition density $p(t,x,y)$ also satisfies the *forward* or *evolution equation* which admits

$$\frac{\partial p(t,x,y)}{\partial t} = \frac{1}{2}\frac{\partial^2}{\partial y^2}[\sigma^2(y)p(t,x,y)] - \frac{\partial}{\partial y}[\mu(y)p(t,x,y)] \tag{2.73}$$

The derivation of the forward equation is considerably more complex than that of the

https://doi.org/10.26219/heal.aueb.6126

backward equation and exceeds the scope of the present review. It should however be noted that very stringent assumptions have to be made that result in some cases the non-fullfilment of equation (2.73) by $p(t,x,y)$.

In the case of a regular homogeneous diffusion process a stationary density $\psi(x)$ exists in the interior of the state space I. This stationarity measure is the limit to which the transition probability density eventually settles, regardless of the initial state of the homogeneous process i.e.

$$\lim_{t\to\infty} p(t,x,y) = \psi(y) \tag{2.74}$$

The stationary density is then given analytically by a formula involving only the scale function and the speed density of the diffusion process as shown in the equation below

$$\psi(x) = C_1 \frac{S(x)}{s(x)\sigma^2(x)} + C_2 \frac{1}{s(x)\sigma^2(x)} = m(x)[C_1 S(x) + C_2] \tag{2.75}$$

where C_1, C_2 constants. This is indeed an achievement since the actual density of the homogeneous diffusion, even for a protracted time perspective, is expressed through functions of the familiar drift $\mu(x)$ and diffusion $\sigma^2(x)$ coefficients.

2.7 Boundary Behaviour of Regular Diffusion Processes

In order to attempt a classification of regular diffusion processes with reference to their behavior near the boundaries l and r of the state space $I = (l,r)$ we can simplify the necessary statements by introducing

1. The *scale measure* i.e. a function $S[J]$ of closed intervals $J = [c,d] \subset (l,r)$ defined by

$$S[J] = S[c,d] = S(d) - S(c) \tag{2.76}$$

https://doi.org/10.26219/heal.aueb.6126

where $S(x)$ the scale function already met. The similar notation is not coincidental, since considering the definition of $S(x)$, becomes obvious that

$$S[c,d] = \int_c^d s(\xi)d\xi \tag{2.77}$$

2. The *speed measure* M where

$$M[J] = M[c,d] = \int_c^d m(x)dx \tag{2.78}$$

where $m(x)$ the speed density.

We shall henceforward be concentrated on one of the boundaries, permit l, keeping in mind that definitions concerning the other will be entirely symmetrical.

First of all if $S(l,x] < \infty$ and this criterion applies independently of x in (l,r) then the boundary l is *attracting*. We clarify that $S(l,b] = \lim_{a \downarrow l} S[a,b]$. Having as a starting point an attracting boundary we move to the question: when this boundary is attainable, meaning in finite expected time. This question is answered by the theorem that suggests that a boundary is *attainable* if and only if $\lim_{a \downarrow l} \int_a^x S[a,\xi]dM(\xi) < \infty$. Otherwise the boundary is *unattainable*. Obviously such a boundary cannot be reached by our diffusion process. A rough justification of the theorem stems by the fact that the $\lim_{a \downarrow l} \int_a^x S[a,\xi]dM(\xi)$ measures the time it takes our process to reach the boundary l or an alternative interior state b starting from an interior point $x < b$. If we also define $M(l,x] = \lim_{a \downarrow l} M[a,x]$, then the requirements $S(l,x] < \infty$ and $M(l,x] < \infty$ establish the *regular* boundary behavior. This represents the ability of the diffusion process to enter and leave this boundary. The behavior on a regular boundary can range from *reflection* (for $M(l,x] = 0$), to the non-regular case of *absorption* ($M(l,x] = \infty$). For cases $0 < M(l,x] < \infty$ in between the behavior on the boundary is *sticky*, a term which corresponds to the strictly positive duration spent at the boundary.

Back to the absorbing case and speaking only for diffusions, i.e. processes which exclude discontinuous trajectories, the boundary is called *exit* boundary. This term is used to characterize state l as a trap state from which the process can never escape thus we claim to exit the process. An *entrance* boundary contrary is a boundary that cannot be reached from the interior of the state space but we can consider processes that start there. Such processes quickly move to the interior of the state space and never return to the entrance boundary. The sufficient condition for a boundary to be considered as entrance boundary is $S(l,x] = \infty$ and $\int_l^x S[\xi, x]dM(\xi) < \infty$.

Our discussion on boundary behavior cannot be complete without the description of the so called *Natural* Boundary (in the *Feller sense).* Such a boundary can neither be reached nor be a start point for a diffusion process. It is comprehensible that such a boundary is omitted from the state space. A boundary is natural when $\int_l^x S(l,\xi]dM(\xi) = \infty$ and $\int_l^x S[\xi, x]dM(\xi) = \infty$.

A very helpful notion for the purpose of constructing processes with interesting boundary behavior is the concept of *local time.* Examining the local time process at the boundary provides us an additional tool of boundary behavior classification. Before mentioning this classification let us first examine what is local time. If we define *occupation time* $L_A(t)$ of the set A up to time t as follows

$$L_A(t) = \int_0^t I_A(X(\tau))d\tau \tag{2.79}$$

where I_A the indicator function of the set A, then the limit random variable

$$\phi(t,a) = \lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon} L_{(a-\varepsilon, a+\varepsilon)}(t), \qquad t > 0 \tag{2.80}$$

defines a family of random variables $\phi(t,a)$ called the local time process. We should emphasize that local time $\phi(t,a)$ is a density and is not the same quantity as the occupation of a point a,

$$L_{\{a\}}(t) = \int_0^t I_{(a)}(X(\tau))d\tau \tag{2.81}$$

https://doi.org/10.26219/heal.aueb.6126

which is usually identical to zero.

According to the local time approach the boundary behaviors are combinations of five basic types

1. absorbing barrier phenomenon,
2. reflecting barrier action,
3. elastic boundary structure,
4. sticky boundary complex,
5. jump boundary behavior and instaneous return processes

Our discussion on boundary behavior concludes with an example in which local time $\phi(t, a)$ is used for constructing a process with one of the above boundary behaviors. Specifically we shall briefly discuss the sticky boundary example for the endpoint l of a diffusion process $\{X(t), t \succeq 0\}$ on the state space $[l, r)$. Using the notion of occupation time, a regular boundary l is said to be sticky when $L_{\{l\}}(t) > 0$, for all $t > 0$, where $X(0) = l$. If we take $Y(t) = |B(t)|$, with $B(t)$ a standard Brownian motion, then $Y(t)$ is a reflecting Brownian motion. If also $\phi(t) = \phi(t, 0)$ the local time at the origin, we form the additive increasing process $U(t)$ with respect to $Y(t)$

$$U(t) = t + \kappa\phi(t) \tag{2.82}$$

where κ is a fixed positive constant. We are able now to construct the process

$$\Gamma(s) = Y(U^{-1}(s)) \tag{2.83}$$

wich inherits the state space $[0, +\infty)$ of reflecting Brownian motion and the regularity of the boundary point 0. As shown in Karlin and Taylor (1981, pg.257-258) the occupation time functional $L_{\{0\}}(t; \Gamma)$ is strictly positive for all $t > 0$. This of course implies that the barrier $\{0\}$ of the $\Gamma(s)$ process is a sticky boundary.

https://doi.org/10.26219/heal.aueb.6126

As a conclusive remark to this section we should point out that the knowledge of the exact boundary behavior is crucial into determining the overall diffusion process evolution, which clearly cannot be solidly attained only by the infinitesimal parameters awareness. This issue along with others is elucidated in the following section of diffusion examples.

2.8 Certain Diffusion Examples

We have already encountered the case of Brownian motion $B(t)$ as a stationary Gaussian process with independent increments. It is easy now to view Brownian motion as a regular diffusion process on $(-\infty, +\infty)$ with infinitesimal parameters $\mu(x) = 0$ and $\sigma^2(x) = \sigma^2$ a constant, for all x. Dividing $B(t)$ by σ we built a Brownian motion process of variance equal to one, called *standard Brownian motion.* If instead we add the trend μt to a Brownian motion $B(t)$ we will produce the *Brownian motion with drift* $Y(t) = B(t) + \mu t$. The process $Y(t)$ will have drift parameter μ and variance parameter σ^2 and will satisfy the equation

$$dY(t) = \mu dt + \sigma dW(t) \tag{2.84}$$

μ, σ constants, $\sigma > 0$, $W(t)$ standard Brownian motion.

A couple of interesting cases of Brownian motion are the *Absorbed* and *Reflected Brownian motion.* If $X(t)$ a Brownian motion with initial value $X(0) = x$, $x > 0$ and τ the first time it reaches zero then Brownian motion absorbed at the origin is called the process $Z(t)$ with

$$Z(t) = \left\langle \begin{array}{l} X(t), \text{ for } t \leq \tau \\ \quad 0, \text{ for } t > \tau \end{array} \right. \tag{2.85}$$

Correspondingly we define the reflected Brownian motion as a stochastic process having the distribution of $Y(t)$ with

$$Y(t) = |X(t)|, \quad t \succeq 0 \tag{2.86}$$

where the characteristic property of reflection is also met at the origin. These two processes are defined on the state space $[0, +\infty)$ and both act like Brownian motion, thus determined by the same μ and σ of the typical Brownian motion, until the level zero is first reached. There we have an alternation to their behavior according to their specific boundary conditions. We must here note that oftenly physical considerations dictate the choice of the boundary conditions in a natural way...

https://doi.org/10.26219/heal.aueb.6126

https://doi.org/10.26219/heal.aueb.6126

Chapter 3

Estimation of parameters

3.1 An introduction to kernel estimators

We feel that accounting for certain density and regression estimation methods in general is a necessity in order to cope with the issues of nonparametric estimation of diffusion processes's coefficients and marginal density. One of the oldest and natural approaches in the estimation problem is the naive estimator. For n observations $X_1, X_2, ..., X_n$ of a real variable X the naive estimator is bound to be

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}w\left(\frac{x-X_i}{h}\right) \tag{3.1}$$

with weight function w defined by

$$w(x) = \begin{cases} \frac{1}{2} & \text{if } |x| < 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

It would help our perception if we could think of the naive estimator as a construction of "boxes" of width $2h$ and height $(2nh)^{-1}$, placed on each observation, that sum up to the estimator.

The problems that arise by the stepwise nature of the former estimator can be overcome with the use of the kernel estimator. In such an estimator we replace the weight function w by a *kernel function* K which satisfies the condition

$$\int_{-\infty}^{\infty} K(x)dx = 1 \tag{3.3}$$

Usually, but not always, K is considered to be a symmetric probability function. In what follows K will be a positive function that satisfies (3.3) along with

$$\int tK(t)dt = 0 \qquad \text{and} \qquad \int t^2K(t)dt = k_2 \neq 0. \tag{3.4}$$

The kernel estimator will unsurprisingly be

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{3.5}$$

where h is the window width, also called the smoothing parameter or bandwidth by some authors. Similarly to the naive estimator, the kernel estimator can be considered as a sum of "bumps" placed at the observations. The advantages of such an estimator are obvious. Given that K is a probability density function it follows from the definition that $\hat{f}$ will itself be a probability density. Even not so, $\hat{f}$ inherits all the continuity and differentiability properties of the kernel K.

We should also note that the actual choice of the kernel K and the bandwidth h is conducted on criteria of discrepancy measures's minimazation. Such measures are the *mean square error* (MSE) defined by

$$MSE_x(\hat{f}) = E\{\hat{f}(x) - f(x)\} \tag{3.6}$$

and the global measure of *mean integrated square error* (MISE) defined by

$$MISE(\hat{f}) = E\int\{\hat{f}\ (x) - f(x)\}^2 dx \tag{3.7}$$

https://doi.org/10.26219/heal.aueb.6126

Elementary manipulations of the definitions above bring us to

$$MSE_x(\hat{f}) = \{E\hat{f}(x) - f(x)\}^2 + var\,\hat{f}(x) \tag{3.8}$$

and

$$MISE(\hat{f}) = \int \{E\hat{f}(x) - f(x)\}^2 dx + \int var\,\hat{f}(x)dx \tag{3.9}$$

Clearly MSE and MISE involve the familiar quantities of bias and variance into a minimazation problem in which a "trade-off" between the reduction of these two quantities is necessary.

In view of the nonparametric estimators $\hat{\mu}$ and $\hat{\sigma}$ that we provide in the following sections, a small discussion on nonparametric regression is essential. The goal of regression curve fitting is to find a relationship between variables $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ where X is considered the explanatory variable of the Y_i values. For n independent observations $\{(X_i, Y_i)\}_{i=1}^n$the regression relationship is modeled as follows

$$Y_i = m(X_i) + \varepsilon_i \qquad i = 1, ..., n \tag{3.10}$$

where ε is a random variable denoting the variation of Y around $m(X)$, the mean regression curve $E[Y/X = x]$. So the approximation of mean response function m, is actually a problem of estimation of the conditional mean curve

$$m(x) = E(Y/X = x) = \frac{\int yf(x,y)dy}{f(x)} \tag{3.11}$$

where $f(x, y)$ denotes the joint density of (X, Y) and $f(x)$ the marginal density of X. Keeping in mind that it is now more of our interest to weight the response variable Y in a certain neighborhood of x, a nonparametric regression smoother of the following general form arises

$$\hat{m}_h(x) = \frac{1}{n}\sum_{i=1}^{n} W_{hi}(x)Y_i \tag{3.12}$$

If kernels are involved in the weights, the most popular estimator is the Nadaraya-Watson proposed in 1964

$$\hat{m}_h(x) = n^{-1}\frac{\sum_{i=1}^{n} K_h(x - X_i)Y_i}{n^{-1}\sum_{j=1}^{n} K_h(x - X_j)} \tag{3.13}$$

For a more detailed discussion see Hardle (1990) among others.

3.2 Density estimation

Let us consider once more the case of a regular homogeneous diffusion process $\{X_t, t \in [0, +\infty)\}$. We already know that the limit of the transition probability density, met as $\psi(y)$ in (2.74), is actually the limiting density of the process. Supposing that the initial density of X, $p(x)$, is chosen so as to $p(x) = \psi(x)$ for all admissible x, it results that each X_t has the same density $p(x)$. The problem of interest is the estimation of $p(x)$ when the process is observed up to time t. For this purpose a kernel type estimator will originally be examined.

Let $K(x)$ be a bounded probability density on R and h be a bounded strictly positive function on R_+ such that

1. $h_t \downarrow 0$ as $t \to \infty$
2. $\gamma_t = \int_0^t h_s ds < \infty$, and
3. $\gamma_t \to \infty$ as $t \to \infty$.

For $t > 0$, let

$$p_t(x_0) = \gamma_t^{-1} \int_0^t K\left[(x_0 - X_s)/h_s\right] ds \tag{3.14}$$

be an estimator of $p(x_0)$. The asymptotic properties of the estimator $p_t(x_0)$ are examined by Prakasa Rao (1983). In theorem 6.3.1 page 328, P.Rao asserts that "if $\{X_t, t \in [0, +\infty)\}$ is a stationary Markov process satisfying the condition $G_2(s', a)$ for some $s' > 0$

https://doi.org/10.26219/heal.aueb.6126

and the initial density $p(x)$ is bounded and continuous on R, then

$$E\left[p_t(x_0) - p(x_0)\right]^2 \rightarrow 0 \qquad \text{as} \quad t \rightarrow \infty." \tag{3.15}$$

$K(x)$ and $h(x)$ are assumed to satisfy conditions (i)-(iii) stated above. In what follows we describe the $G_2(s,a)$ condition and introduce relating notions. For each $t \in [0,+\infty)$, the transition probability operator H_t of X is defined as follows

$$H_t f(x) = E[f(X_t) \ / \ X_0 = x], \qquad x \in R \tag{3.16}$$

with f any bounded Borel measurable function on R. We then define $|H_t|_2$ to be

$$|H_t|_2 = \sup_{\{f : E[f(X)]=0\}} \frac{E^{1/2}(H_t f)^2}{E^{1/2}(f^2)} \tag{3.17}$$

The transition operator H_t is said to satisfy the condition $G_2(s,a)$ if there exists $s > 0$ such that $|H_s|_2 \leq a$ with $0 < a < 1$.

A wider family of estimators than the one introduced by (3.14) is also examined in Prakasa (1983) whose exact form is

$$p_t(x) = \left\{\int_0^t h(s)H[h(s)]ds\right\}^{-1} \int_0^t H[h(s)]K\left[\frac{X_s - x}{h(s)}\right] ds \tag{3.18}$$

with h, H functions from R_+ to R_+, $h(s) \rightarrow 0$ as $s \rightarrow \infty$. For $H(s) \equiv 1$ the estimator above developes to (3.14). Strong consistency of the estimator is also proven under explicit conditions stated in theorem 6.3.2 page 331 of Prakasa (1983)

An alternative estimator using delta sequences is proposed in Prakasa Rao (1979a). This estimator generalizes the recursive estimator of a distribution's density f that for a random sample $X_1, X_2, ..., X_n$ is provided by the formula

$$f_n(x) = \gamma_n^{-1} \sum_{i=1}^{n} h_i \delta_{h_i}(x - X_i) \tag{3.19}$$

with $0 < h_n \to 0$ as $n \to \infty$,so that $\gamma_n = \sum_{i=1}^{n} h_i$ diverges (e.g. $h_n = n^{-s}, 0 < s \leq 1$). For the better comprehension of the present estimator we recall the definition of a delta family (as found in Prakasa (1983) pg305). A family $\{\delta_t, t > 0\}$ of nonnegative bounded functions is called a *delta family of positive type* $a > 0$, if there exist $A > 0, B > 0$ such that

1. $\left|1 - \int_{-A}^{B} \delta_t(x)dx\right| = O(t^a)$
2. $\sup\{|\delta_t(x)| : |x| \geq t^a\} = O(t^a)$
3. $\|\delta_t\|_\infty \approx \infty$

The exact form of the estimator of $p(x)$ as proposed in Prakasa (1979a) is

$$p_t^*(x) = \gamma(t)^{-1} \int_0^t h(s)\delta_{h(s)}(x - X_s)ds \tag{3.20}$$

where $h(t)$ a nonnegative real-valued function such that $h(t) \downarrow 0$ as $t \to \infty$ and $h(t)$ is locally integrable, i.e.,

$$\gamma(t) = \int_0^t h(s)ds < \infty. \ \gamma(t) \to \infty \qquad \text{as} \qquad t \to \infty \tag{3.21}$$

It can be shown that $p_t^*(x)$ converges to the real function according to the

$$\sup_x E[p_t^*(x) - p(x)]^2 = O\left([1/\gamma(t)] + \left\{[1/\gamma(t)] \int_0^t h(s)^{1+a\lambda}ds\right\}^2\right) \tag{3.22}$$

under conditions identified specifically in Prakasa (1979a) or problem 5, Chapter 6 of Prakasa (1983)

Density estimation for continuous time processes with application to inference for diffusion processes is also conducted by Leblanc (1995) using wavelets. A wavelet is an orthogonal system that can be expressed as an infinite collection of translated and scaled versions of functions ϕ and ψ called the scaling function and the primary wavelet function

https://doi.org/10.26219/heal.aueb.6126

respectively. The function $\phi(x)$ is a solution of the difference equation

$$\phi(x) = \sum_{k=-\infty}^{\infty} c_k \phi(2x-k) \tag{3.23}$$

with normalization

$$\int_{-\infty}^{\infty} \phi(x)dx = 1 \tag{3.24}$$

while function $\psi(x)$ is defined by

$$\psi(x) = \sum_{k=-\infty}^{\infty} (-1)^k c_{-k+1} \phi(2x-k) \tag{3.25}$$

The coefficients c_k are called the filter coefficients and their choice determines the wavelet system. The wavelet system expansion is actually a tool used to decompose a function into a set of weighted basis functions that are localized in time and frequency. It is analogous to the Fourier series expansion which represents a signal by a summation of complex sinusoids weighted by a set of coefficients. Therefore engaging wavelets to the problem of estimating the marginal density f of X_t over an interval $[-K,K]$ of the observed sample path $\{X_t, 0 \leq t \leq T\}$, produces the following expansion

$$fI([-K,K]) = \sum_{k\in K_{j_0}} a_{j_0,k}\phi_{j_0,k} + \sum_{j\geq j_0}\sum_{k\in K_{j_0}} \beta_{j,k}\psi_{j,k} \tag{3.26}$$

where

$$\phi_{j_0 k}(x) = 2^{j_0/2}\phi(2^{j_0}x-k) \tag{3.27}$$

and

$$\psi_{j,k} = 2^{j/2}\psi(2^j x - k) \tag{3.28}$$

By orthogonality of the wavelet basis, it follows that

$$a_{j_0,k} = \int f(x)\phi_{j_0,k}(dt) \quad \text{and} \quad \beta_{j,k} = \int f(x)\psi_{j,x}(dx) \tag{3.29}$$

https://doi.org/10.26219/heal.aueb.6126

Leblanc (1995) considers the estimator

$$\hat{f}_T(x) = \sum_{k \in K_{j_0}} \hat{a}_{k_0,k}\, \phi_{j_0,k}(x) \tag{3.30}$$

with

$$\hat{a}_{j_0,k} = \frac{1}{T}\int_0^T \phi_{j_0,k}(X_s)ds \tag{3.31}$$

and obtains the upper bound for $E\left\|(f - \hat{f}_T)I(-K,K)\right\|_p^2$.

Advantages and disantvantages of density estimation with wavelets were discussed by Walter and Ghorai (1992). They indicated that application of wavelet bases give better asymptotic properties of the estimators but, for small samples, they have little advantage over the kernel methods and do not give as close an approximation to the true density.

3.3 Drift and diffusion parameters estimation: Non-parametric approach

Following the historic progression of research concerning the nonparametric estimation of diffusion processes's coefficients, we shall first review the procedure followed when the sampling observations are continuous. So beginning with the usual considerations of a diffusion process X_t, that satisfies the stochastic differential equation

$$dX_t = \mu(X_t)dt + \sigma(X_t)dB_t \tag{3.32}$$

with initial condition $X_0 = X$, we suppose that X_t can be observed continuously throughout the time interval $[0,T]$. Observations of this kind enable the true diffusion function to be determined, at least for states x visited by X_t during $[0,T]$, through

$$\lim_{n\to\infty} \sum_{j=1}^{2^n} (X_{jt2^{-n}} - X_{(j-1)t2^{-n}})^2 = \int_0^t \sigma^2(X_s)ds \tag{3.33}$$

https://doi.org/10.26219/heal.aueb.6126

which holds almost surely for all $t \in [0, T]$, where $\{t_j = jt2^{-n}, j = 0, 1, 2, ..., 2^n\}$ is a sequence of divisions of the interval $[0, T]$ such that $\max\{(t_j - t_{j-1}), 1 \leq j \leq 2^n\} \to 0$ as $n \to \infty$ (see, e.g. Brown and Hewitt, 1975). Based on the fact that $\mu(X_t)$ is the following conditional expectation limit

$$\mu(X_t) = \lim_{h \to 0} E\left\{\frac{X_{t+h} - X_t}{h} / X_t(\omega) = X_t\right\}, \tag{3.34}$$

Geman (1979) considered the commonsense nonparametric estimator of the drift function of the following form:

$$\mu_n(x) = n^{-1} \sum_{i=1}^{n} \tau_i^{-1} (X_{t_i + \tau_i} - x) \tag{3.35}$$

where $0 < \tau_i \to 0$ as $i \to \infty$ and $\{t_i\}$ is a sequence of random times defined by $t_1 = \inf\{t \geq 0 : X_t = x\}$ and $t_{i+1} = \{t \geq t_i + \tau_i : X_t = x\}$. Geman (1979) proved the consistency and asymptotic normality of this estimator.

Banon (1978) proposed an alternative nonparametric procedure to estimate point by point the function $\mu(x)$ when the function $\sigma(x)$ is known or when $\sigma(x)$ is unknown but takes a constant value. Banon (1978) considered the relation found in Wong (1971)

$$\frac{1}{2}(\sigma^2 p)' = \mu p \tag{3.36}$$

that explicitly relates the pair $(\mu(x), \sigma(x))$ to the limiting density p of the X_t process, when it exists. Therefore estimation of $\mu(x)$ is related to estimation of $p(x)$ and $p'(x)$ discussed above. Banon (1978) first assumes for reasons of simplicity that the process $\{X_t\}$ is stationary with probability density function p and satisfies the condition $G_2(s, a)$ defined above. He then estimates σ^2 when $\sigma(x)$ is constant. Considering that

$$\sigma^2(x) = \lim_{t \to 0} \frac{1}{t} E[(X_{t+s} - X_t)^2 / X_s = x], \qquad x \epsilon \mathcal{R}, \ s \geq 0 \tag{3.37}$$

Banon (1978) proposed the estimator

$$\sigma_n^2 = n^{-1} \sum_{i=1}^{n} \frac{1}{\tau_i} (X_{t_i+\tau_i} - X_{t_i})^2 \tag{3.38}$$

with $\{\tau_i\}$ a bounded sequence of positive numbers tending to zero and $\{t_i\}$ be such that $t_1 \succeq 0$ and $t_i + \tau_i \leq t_{i+1}$, $i = 1, 2, \ldots$. Under the initial condition $EX^4 < \infty$, σ_n^2 is proven to be a quadratic mean convergent estimate of σ^2, that is

$$\sigma_n^2 \xrightarrow{q.m.} \sigma^2, \qquad \text{as } n \to \infty \tag{3.39}$$

The described estimate is recursive, i.e., is the solution to:

$$\sigma_n^2 = \frac{n-1}{n} \sigma_{n-1}^2 + \frac{1}{n\tau_n} (X_{t_n+\tau_n} - X_{t_n})^2 \tag{3.40}$$

The starting point for Banon to estimate the drift $\mu(x)$ is equation (3.62), the origin of which is explicated in pages 41 and 42 of the present dissertation. Banon sets $q(x) = p'(x)/p(x)$ and proposes for q the following estimate

$$q_t(x_0) = \frac{\left[\int_0^t \frac{1}{h_s} K_2' \left(\frac{x_0 - X_s}{h_s}\right) ds\right]}{\left[\int_0^t \frac{1}{h_s} K_1 \left(\frac{x_0 - X_s}{h_s}\right) ds + \varepsilon\right]} \tag{3.41}$$

with $K_1(x)$ a bounded probability density function, $K_2(x)$ a continuous probability density function of bounded variation such that K_2' is bounded, $\varepsilon > 0$ fixed and h_t bounded positive function such that

1. $h_t \downarrow 0$ as $t \to \infty$
2. $\gamma_t = \int_0^t h_s ds < \infty$
3. $\gamma_t \to \infty$ as $t \to \infty$ and
4. $h_t^2 \gamma_t \to \infty$ as $t \to \infty$

If $p'(x)$ is continuous and bounded then the estimate is proven to converge in probability to the true quantity, thus

$$q_t(x_0) \xrightarrow{p} q(x_0) \qquad \text{as } t \to \infty \tag{3.42}$$

(Proof can be found in Banon (1978) Theorem 5.1 page 393). It is easily anticipated at this point that the suggested estimate for the drift $\mu(x)$ is

$$\mu_t(x_0) = \frac{1}{2}[\sigma^{2'}(x_0) + \sigma^2(x_0)q_t(x_0)] \tag{3.43}$$

for $\sigma^2(x)$ known. Under the conditions assuring the convergence of q_t, it is proven that

$$\mu_t(x_0) \xrightarrow{p} \mu(x_0), \qquad \text{as } t \to \infty \tag{3.44}$$

For $\sigma(x)$ unknown but constant replacing in (3.43) the estimate given in (3.38) provides the estimate $\mu_{t,n}(x)$ with

$$\mu_{t,n}(x_0) = \frac{1}{2}\sigma_n^2 q_t(x_0) \tag{3.45}$$

Under the conditions stated for $q_t(x)$ it is proven that

$$\mu_{t,n}(x_0) \xrightarrow{p} \mu(x_0), \qquad \text{as } t, n \to \infty \tag{3.46}$$

It is obvious though that in most practical situations continuous sampling of the stochastic process is impossible. This comes as a consequence of slow sampling rates comparing to rapid characteristic dynamics of these systems. So the frequently arising question in practical applications is how to estimate the parameters of a stochastic differential equation from discrete time observations.

Florens-Zmirou's (1993) paper was a pioneering one in nonparametric estimation with discrete sampling observations. She is concentraded in non-parametrically estimating the diffusion coefficient of a diffusion process using discrete observations X_{t_i} at times

$(t_1, t_2, ..., t_n)$ of the finite time interval $[0, T]$. It can be assumed that $T = 1$ and $t_i = i/n$ without loss of generality. Primarily the local time $\phi(t, x)$ of X in x during $[0, t]$ is estimated by developing the discrete approximation

$$\phi^{(n)}(t, x) = \frac{1}{2nh_n} \sum_{i=1}^{[nt]} I_{(x-h_n, x+h_n)}(X_{i/n}) \tag{3.47}$$

with $h_n \to 0$ and $nh_n \to \infty$. Let be noted that regarding local time we use terminology introduced in Section 2.7 of the present paper instead of the one found in Florens-Zmirou for reasons of continuity. The $\phi^{(n)}(t, x)$ approximation is proved to converge to the local time $\phi(t, x)$ in the $\mathbf{L}^2$ sense for nh_n^4 tending to zero as n tends to infinity. Under the same conditions it is also proven that

$$V_t^{(n)} \xrightarrow{\mathbf{L}^2} \sigma^2(x)\phi(t, x) \tag{3.48}$$

with $V_t^{(n)}(x)$ defined as follows

$$V_t^{(n)}(x) = \frac{1}{2h_n} \sum_{i=1}^{[nt]-1} I_{\{|X_{i/n}-x|<h_n\}}[X_{(i+1)/n} - X_{i/n}]^2 \tag{3.49}$$

$$V_1^{(n)}(x) = V^{(n)}(x) \quad \text{and} \quad V^{(n)}(0) = V^{(n)} \tag{3.50}$$

These propositions allowed Florens-Zmirou to conclude that for (h_n) such that nh_n^4 tends to zero the $S_n(x)$ with

$$S_n(x) = \frac{\sum_{i=1}^{n-1} I_{\{|X_{i/n}-x|<h_n\}}[X_{(i+1)/n} - X_{i/n}]^2}{\sum_{i=1}^{n} I_{\{|X_{i/n}-x|<h_n\}}} \tag{3.51}$$

is a consistent estimator of $\sigma^2(x)$. An equivalent approach to the estimation of $\phi(t, x)$, also mentioned in Florens-Zmirou (1993), is the familiar kernel estimator

$$\phi^{(n)}(t, x) = \frac{1}{nh_n} \sum_{i=1}^{[nt]} K\left(\frac{X_i - x}{h_n}\right) \tag{3.52}$$

As we shall see forward Jiang and Knight (1997) advanced the idea of kernel estimator in accordance to Florens-Zmirou (1993) suggestions and preceding context. Still in Florens-Zmirou (1993) a finding of greater significance is also formulated and that is the derivation of asymptotic distribution for $S_n(x)$ epitomized in the following theorem.

Theorem 1 *If nh_n^3 tends to zero, then*

$$\sqrt{nh_n}\left(\frac{S_n(x)}{\sigma^2(x)}-1\right) \tag{3.53}$$

converges in distribution to $\phi(x)^{-1/2}Z$, where Z is a standard normal variable independent of $\phi(x)$. (By $\phi(x)$ we note $\phi(1,x)$ (with $1=T$))

Stanton (1997) illustrated a procedure for nonparametrically estimating both drift and diffusion functions. Beginning with the usual considerations of a diffusion process X_t, that satisfies the stochastic differential equation

$$dX_t = \mu(X_t)dt + \sigma(X_t)dB_t \tag{3.54}$$

with initial condition $X_0 = X$, Stanton wrote the conditional expectation $E_t[f(X_{t+\Delta_n})]$ in the form of a Taylor series expansion

$$E_t[f(X_{t+\Delta_n})] = f(X_t) + Lf(X_t)\Delta_n + \frac{1}{2}L^2f(X_t)\Delta_n^2 + \dots + \frac{1}{N!}L^Nf(X_t)\Delta_n^N + O(\Delta_n^{N+1}) \tag{3.55}$$

Δ_n is the sampling interval between the n observations of X_t over the time period $[0,T]$ i.e. $\Delta_n = \frac{T}{n}$ while L is the already known infinitesimal generator of the process X_t

$$Lf(x) = \frac{\partial}{\partial x}f(x)\mu(x) + \frac{1}{2}\frac{\partial^2}{\partial x^2}f(x)\sigma^2(x) \tag{3.56}$$

Assuming $f(x) = x$, Stanton (1997) used the N-W kernel method to provide the following

https://doi.org/10.26219/heal.aueb.6126

approximation for $\mu(x)$:

$$\hat{\mu_1}(x) = \frac{\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta_n}-x}{h_n}\right)\left(X_{(i+1)\Delta_n} - X_{i\Delta_n}\right)}{\Delta_n \sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta_n}-x}{h_n}\right)} \tag{3.57}$$

where h_n is the smoothing parameter of the nonparametric estimator. Actually $\hat{\mu_1}(x)$ is the kernel regression estimator of the first order approximation of $\mu(X_t)$ provided by Stanton

$$\mu(X_t) = \frac{1}{\Delta_n} E_t[X_{t+\Delta} - X_t] + O(\Delta_n) \tag{3.58}$$

This approximation had already been used by previous authors, such as Chan et al. (1992), but Stanton's contribution lies at the provided sequence of approximations for $\mu(x)$ with (3.58) being only the first.

In order to construct the approximations for the diffusion parameter σ, Stanton used the $f(x,t) = (x - X_t)^2$ and concluded with a series of approximations first of which is

$$\sigma(X_t) = \sqrt{\frac{1}{\Delta_n} E_t[(X_{t+\Delta} - X_t)^2]} + O(\Delta_n) \tag{3.59}$$

Therefore the corresponding estimator is

$$\hat{\sigma_1}(X_t) = \sqrt{\frac{\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta_n}-x}{h_n}\right)\left(X_{(i+1)\Delta_n} - X_{i\Delta_n}\right)^2}{\Delta_n \sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta_n}-x}{h_n}\right)}} \tag{3.60}$$

The family of approximations estimated by Stanton (1997) converge to the true functions at a rate Δ^k, where Δ is the time between successive observations, and k is an arbitrary positive integer. Stanton's (1997) approach is appealing due to the simplicity of the separate estimation of σ and μ but still problems concerning these estimators were identified. Chapman and Pearson (2000) applied Stanton's procedure to simulated sample paths of a diffusion with linear drift, namely the Cox-Ingersoll-Ross Squared root Diffusion, and

https://doi.org/10.26219/heal.aueb.6126

detected nonlinearity of the Stanton estimator. This spurious nonlinearity was concluded to be caused by "mean reversion" and small sample at the boundary which as expected worsened the diversion from the true drift near the boundary regions. Fan and Zhang (2003) continues the discussion launched in the two previous papers trying to answer the following essential questions:

- Do higher order approximations outperform their lower order counterparts?
- Can reasonable and formal procedures be found in order to help determine whether the observed nonlinearity in the drift is real or due to chance variation?

In order to facilitate the first question Fan and Zhang (2003) built explicit formulas for the variance and bias of the drift and diffusion estimators with regards to the order k and notice that though asymptotic bias is indeed reduced, asymptotic variance escalates nearly exponentially as k grows. The variance inflation phenomenon applies to parametric modeling as well and thus is not only an artifact of nonparametric fitting. As far as the second question is concerned it can be brought to a hypothesis testing problem with a parametric model null hypothesis against a nonparametric alternative; the statistic used for the particular problem by Fan and Zhang was the Generalized Likelihood Ratio (GLR) developed by Fan et al. (2001)

As illustrated in (2.73) μ and σ are participants of the forward equation which by proper manipulations can be brought to

$$\frac{d^2}{dx^2}(\sigma^2(x)\psi(x)) = 2\frac{d}{dx}(\mu(x)\psi(x)) \tag{3.61}$$

with $\psi(x)$ the marginal density. Equation (3.61) shows a relationship between the drift, the diffusion and the marginal density function of the diffusion process. Integrating (3.61) with the boundary conditions $\lim_{x\to\infty} \sigma^2(x)\psi'(x) = 0$, $\lim_{x\to\infty} \sigma^{2\prime}(x)\psi(x) = 0$ and

$\lim_{x\to\infty} \mu(x)\psi(x) = 0$, we obtain

$$\mu(x) = \frac{1}{2}\left[\frac{d\sigma^2(x)}{dx} + \sigma^2(x)\frac{\psi'(x)}{\psi(x)}\right] \tag{3.62}$$

as Banon (1978) illustrated. This is the formula that permitted Jiang and Knight (1997) to estimate the drift coefficient function with discrete observations and avoid the extreme sensitivity to the sampling intervals and the length of the total sampling period that emerges by the use of formula (3.35).

More explicitly Jiang and Knight (1997) considered the local time estimators (3.52) proposed by Florens-Zmirou (1993) and thus constructed the estimator $S_n(x)$ of $\sigma^2(x)$ that is given by the following formula

$$S_n(x) = \frac{\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta_n}-x}{h_n}\right)\left(X_{(i+1)\Delta_n} - X_{i\Delta_n}\right)^2}{\Delta_n \sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta_n}-x}{h_n}\right)} \tag{3.63}$$

They proceeded to proving pointwise consistency and asymptotic normality for their diffusion estimator in a theorem very much alike to Florens-Zmirou's theorem 1 of the present section. The variance of $S_n(x)$ is also provided in the same theorem by the consistent estimator

$$V\left[S_n(x)\right] = \frac{S_n^2(x)}{\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta_n}-x}{h_n}\right)} \tag{3.64}$$

By utilizing the kernel method instead of the naive method, Jiang and Knight (1997) tried to achieve better performance for their estimators on the basis of integrated mean squared error (IMSE), following the conclusions of Kumar and Markman (1975) derived from their Monte Carlo studies; furthermore they avoided the discontinuity of the naive estimator which is a serious drawback when constructing the nonparametric drift function estimator since derivatives of the diffusion estimator are involved.

The next step for Jiang and Knight (1997) was to suggest a non parametric estimate of $\mu(x)$ based on equation (3.62) and a proposition of Banon and Nguyen (1981) concerning a strongly consistent estimator of $Q(x) = \psi'(x)/\psi(x)$ with continuous sampling observations. So, providing, as an intermediate step, the following consistent estimator of $Q(x)$

$$q_n(x) = \frac{\sum_{i=0}^{n-1} \frac{1}{h_n} K'\left(\frac{X_{i\Delta_n} - x}{h_n}\right)}{\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta_n} - x}{h_n}\right)} \tag{3.65}$$

from discrete sampling observations, Jiang and Knight (1997) suggest the following estimate for drift

$$\hat{\mu}_2(x) = \frac{1}{2}\left[\frac{dS_n(x)}{dx} + S_n(x)\frac{\sum_{i=0}^{n-1} \frac{1}{h_n} K'\left(\frac{X_{i\Delta_n} - x}{h_n}\right)}{\sum_{i=0}^{n-1} K\left(\frac{X_{i\Delta_n} - x}{h_n}\right)}\right] \tag{3.66}$$

which is proven to be a pointwise consistent estimator (this conclusion is valid under the explicit conditions A1-A8, see Jiang and Knight (1997))

Li and Tkacz (2002) proved that the nonparametric estimator $\hat{\mu}_1(x)$ of the drift function proposed by Stanton (1997) is not consistent and provided the convergence rate of Jiang and Knight (1997) estimator $\hat{\mu}_2(x)$. These essential conclusions lie in a single theorem of Li and Tcacz (2002) that assumes growth and Lipschitz conditions, stationarity of the process and restrictiveness to the amount of the allowed dependence in the observable sequence. The basic results of this theorem are

1. $\sqrt{h_n}[\hat{\mu}_1(x) - \mu(x)] \to N\left[0, \frac{\sigma^2(x)}{\psi(x)} \int K^2(u)du\right]$ in distribution
2. if $nh_n^3 \to \infty$ and $nh_n^5 \to 0$, then $\sqrt{nh_n^3}[\hat{\mu}_2(x) - \mu(x)] = O_p(1)$
3. if $nh_n^3 \to 0$ then $\sqrt{nh_n}(S_n(x) - \sigma^2(x)) \to N\left(0, \frac{\sigma^4(x)}{\psi(x)} \int K^2(u)du\right)$ in distribution

Bandi and Phillips (2002) in their recent paper find it more useful for empirical applications that a nonparametric estimation method for diffusion processes with nonstationary behavior is constructed. Instead the substantially milder assumption of reccurence is the new identifying condition. *Reccurence* requires that the continuous trajectory of the process must visit any level in the permittable range of values an infinite number of times over time.

Before providing their estimators Bandi and Phillips (2002) discuss of course the requirements of their model but also introduce a standardized version of the conventional local time $\phi(t,a)$ that is defined in terms of pure time units. A natural way to define this local time, called chronological local time for the first time by Phillips and Park (1998), is the following

$$\bar{\phi}(t,a) = \frac{1}{\sigma^2(a)}\phi(t,a) \tag{3.67}$$

Assume a process X_t in the time interval $[0,T]$ for which n equispaced observations $\{X_{\Delta_{n,T}}, X_{2\Delta_{n,T}}, ..., X_{n\Delta_{n,T}}\}$ are available ($\Delta_{n,T} = T\ /n$). The proposed estimators now are

$$\begin{aligned}\hat{\mu}_{(n,T)} &= \frac{\sum_{i=1}^{n} K\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\left(\frac{1}{m_{n,T}(i\Delta_{n,T})\Delta_{n,T}}\sum_{j=o}^{m_{n,T}(i\Delta_{n,T})-1}[X_{t(i\Delta_{n,T})_j+\Delta_{n,T}} - X_{t(i\Delta_{n,T})_j}]\right)}{\sum_{i=1}^{n} K\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)} \\ &= \frac{\sum_{i=1}^{n} K\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\bar{\mu}_{n,T}\left(X_{i\Delta_{n,T}}\right)}{\sum_{i=1}^{n} K\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)}\end{aligned} \tag{3.68}$$

for the drift and

$$\begin{aligned}\hat{\sigma}^2_{(n,T)} &= \frac{\sum_{i=1}^{n} K\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\left(\frac{1}{m_{n,T}(i\Delta_{n,T})\Delta_{n,T}}\sum_{j=0}^{m_{n,T}(i\Delta_{n,T})-1}[X_{t(i\Delta_{n,T})_j+\Delta_{n,T}} - X_{t(i\Delta_{n,T})_j}]^2\right)}{\sum_{i=1}^{n} K\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)} \\ &= \frac{\sum_{i=1}^{n} K\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)\bar{\sigma}^2_{n,T}\left(X_{i\Delta_{n,T}}\right)}{\sum_{i=1}^{n} K\left(\frac{X_{i\Delta_{n,T}}-x}{h_{n,T}}\right)}\end{aligned} \tag{3.69}$$

for the diffusion function. $K(x)$ is a kernel function with properties already discussed at section 3.1, while $\{t(i\Delta_{n,T})_j\}$ is a sequence of random times defined in the following manner:

$$t(i\Delta_{n,T})_0 = \inf\{t \geq 0 : \left|X_t - X_{i\Delta_{n,T}}\right| \leq \varepsilon_{n,T}\} \tag{3.70}$$

and

$$t(i\Delta_{n,T})_{j+1} = \inf\{t \geq t(i\Delta_{n,T})_j + \Delta_{n,T} : \left|X_t - X_{i\Delta_{n,T}}\right| \leq \varepsilon_{n,T}\} \tag{3.71}$$

The number $m_{n,T}(i\Delta_{n,T}) \leq n$ counts the stopping times associated with the value $X_{i\Delta_{n,T}}$

and is defined as

$$m_{n,T}(i\Delta_{n,T}) = \sum_{j=1}^{n} I_{(X_{j\Delta_{n,T}}-\varepsilon_{n,T}, X_{j\Delta_{n,T}}+\varepsilon_{n,T})}(X_{i\Delta_{n,T}}) \qquad \forall i \leq n. \tag{3.72}$$

The quantity $\varepsilon_{n,T}$ is a bandwidth like parameter that is taken to depend on the time span and on the sample size. The first conclusions that are derived concern the case of simultaneous increase in the sampling frequency and the observation period, mathematicaly speaking $n \to \infty$, $T \to \infty$ and $\Delta_{n,T} = T\ /n \to 0$. By using such observations the estimators $\hat{\mu}_{(n,T)}$and $\hat{\sigma}^2{}_{(n,T)}$aim to reconstruct as well as possible the path of the process in terms of their true analogs. As far as the $\hat{\mu}_{(n,T)}$is concerned it is proven to converge to the true function with probability one and that its asymptotic distribution is of the form

$$\sqrt{\varepsilon_{n,T}\ \bar{\hat{\phi}}\ (T,x)}\left\{\hat{\mu}_{(n,T)}\ (x) - \mu(x)\right\} \xrightarrow{d} \mathbf{N}\left(0, K_2^{ind}\sigma^2(x)\right) \tag{3.73}$$

where $K_2^{ind} = \frac{1}{4}\int_{-\infty}^{\infty} I^2_{\{|a|\leq 1\}}da = \frac{1}{2}$ if $h_{n,T} = o(\varepsilon_{n,T})$. The explicit conditions under which the convergence holds are presented in Theorems 2 and 3 of Bandi and Phillips (2002) along to the asymptotic bias of the drift estimator. Promptly we can notice that these conditions are concerning the speed that n, T diverge to ∞ and $\Delta_{n,T}$, $h_{n,T}$ and $\varepsilon_{n,T}$ converge to 0 in relation to $\bar{\hat{\phi}}\ (T, x)$. Moving forward to the asymptotic theory for the diffusion estimator, Bandi and Phillips prove its convergence to the true function with probability one and find its asymptotic distribution to be of the form

$$\sqrt{\frac{\varepsilon_{n,T}\ \bar{\hat{\phi}}\ (T,x)}{\Delta_{n,T}}}\left\{\hat{\sigma}^2{}_{(n,T)}\ (x) - \sigma^2(x)\right\} \xrightarrow{d} \mathbf{N}\left(0, 4K_2^{ind}\sigma^4(x)\right) \tag{3.74}$$

Again the explicit conditions are described in Theorems 4 and 5 of Bandi and Phillips along to the bias term of the diffusion estimator.

Analysis is also conducted for the fixed T case which consequenses the non identification of the drift function. Contrary the features of the diffusion function can be

https://doi.org/10.26219/heal.aueb.6126

better reflected by the local dynamics of the underlying continuous process. This results a meaningful definition for the diffusion function estimator even over a fixed time span of observations, a result accordant to other authors as Geman (1979). The limiting distribution of the estimator $\hat{\sigma}^2{}_{(n,T)}$is then established as a Mixed Normal with principal terms actually determined by the relation the observation rate $\Delta_{n,T}$ to the spatial bandwidth $\varepsilon_{n,T}$. If $\Delta_{n,T}$ is small relative to $\varepsilon_{n,T}$, so that $n\varepsilon^4_{n,T} \to \infty$, then the bias effect dominates the asymptotics. If the spatial bandwidth $\varepsilon_{n,T}$ is small relative to the observation interval and $n\varepsilon^4_{n,T} = o(1)$, then the bias effect is eliminated asymptotically and the martingale effect governs the limit theory. Rates of convergence for $\hat{\mu}_{(n,T)}$and $\hat{\sigma}^2{}_{(n,T)}$are discussed as well as the case of single smoothers $\bar{\mu}_{(n,T)}$and $\bar{\sigma}^2{}_{(n,T)}$already met in our discussion of (3.57) and (3.63). Relation to Florens-Zmirou (1993) is also commented by Bandi and Phillips, drawing attention to the resemblance between the limiting distribution of their diffusion estimator and the one provided by Florens-Zmirou (see Theorem 1 of present section) for choices of $\varepsilon_{n,T}$ and $h_{n,T}$ that make the bias term negligible.

Moloche (2001) advances the methodology introduced by Bandi and Phillips (2002) using the local polynomial kernel approach to estimate the drift and diffusion function of recurrent scalar diffusion processes. Bandi and Phillips's theory is accordingly extended in three directions: first the decrease of the small sample bias due to the implementation of the local polynomial approach is illustrated; second, the convergence rates are further analyzed and their dependence on infinitesimal coefficients is explicated; third, the positive recurrent case is comprehensively studied.

3.4 Drift and diffusion parameters estimation: parametric approach

We consider once more the diffusion process X_t satisfying the time-homogeneous stochastic differential equation

$$dX_t = \mu(X_t)dt + \sigma(X_t)dB_t \tag{3.75}$$

where B_t is a standard Brownian motion. In estimating the functions μ and σ in equation (3.75) the usual approach remains the specification of parametric forms for the functions μ and σ and subsequently the estimation of the values of the parameters existing in them. Relating examples can be located in Econometrics by Cox, Ingersoll, and Ross (CIR) (1985), Vasicek (1977), Brennan and Schwartz (1979, 1982), and Chan , Karolyi, Longstaff and Sanders (CKLS) (1992).

The first paper to deal with parametric estimation of the coefficients of a stationary diffusion process from discrete sampling observations is the one by Dacunha-Castelle and Florens-Zmirou (1986). The measure of the amount of information lost due to discretization is also provided in the particular paper. Dohnal (1987) also considered the parametric estimation of the diffusion term and proved the local asymptotic mixed normality property of its likelihood function. He then used the demonstrated property to achieve better results than those obtained by the use of formulas (3.33) and (3.35). The method used in both papers is the expansion of the transitional density of the underlying Markov process for small changes in time.

Given functions μ and σ the transition density from value x to value y in period t, $p(t,x,y)$, must satisfy the Kolmogorov backward equation (2.72) and the forward equation (2.73). In principle, for a given parametrization of μ and σ, we can solve equation (2.72) for the conditional density p as a function of the parameters, then use maximum likelihood to estimate the model's parameters. Lo (1988) for example derived the ML estimation method of the parameters based on the Markovian properties of a diffusion process with jumps. Another example can be found in Pearson and Sun (1994)) Pedersen (1995) suggested an approximate maximum likelihood (AML) parameter estimator for multidimensional diffusion processes, but his framework was purely theoretical.

Unfortunately, except in a few cases such as the one confronted in Pearson and Sun (1994), equation (2.72) can only be solved numerically, making implementation of maximul likelihood extremely inconvenient. Hansen's (1982) Generalized Method of Moments (GMM) can often be used instead of full Maximum Likelihood, either when the full like-

https://doi.org/10.26219/heal.aueb.6126

lihood function is too complicated or time-consuming to calculate, or where we wish to specify only certain properties of the distribution, rather than the full likelihood function.

Duffie and Singleton (1993) use an "indirect inference" approach to estimating non-linear stochastic differential equations called simulated GMM. They use simulation to calculate arbitrary population moments as functions of the parameters of the process being estimated. The simulated moments are subsequently compared with the sample moments estimated from the data and the minimization of this difference provides the parameter estimates.(see also Gourieroux, Monfort, and Renault (1993)). Monfort (1996) reviewed methods of indirect inference and focused on misspecified models.Gallant and Tauchen (1994) also use simulation, generating moment conditions from the score function of an auxiliary (quasi) maximum likelihood estimation. Hansen and Scheinkman (1995) show how to derive analytic moment restrictions from equation (3.75) using the infinitesimal generator $\mathcal{L}$ of X_t. For example, their first class of moment conditions (C1) can be obtained by noting that, if X_t is stationary, $E[\phi(X_t)]$ must be independent of calendar time for any function ϕ. This implies that its unconditional expected rate of change must be zero, i.e.

$$C1: \qquad E[\mathcal{L}\phi(X_t)] = E[\phi'(X_t)\mu(X_t) + \frac{1}{2}\phi''(X_t)\sigma^2(X_t)] = 0 \tag{3.76}$$

While these moment conditions are less computationally intensive than those of Duffie and Singleton (1993) or Gallant and Tauchen (1994), they do not take advantage of all of the information contained in the discretely observed data. An alternative approach is to use GMM with approximate moment conditions. A well-known example is Chan *et al.* (1992). In estimating their continuous-time interest rate model,

$$dr_t = (a + \beta r_t)dt + \sigma r^\gamma dB_t \tag{3.77}$$

https://doi.org/10.26219/heal.aueb.6126

they use approximate conditional moments of the form

$$E_t(\epsilon_{t+\Delta}) = 0 \tag{3.78}$$

$$E_t(\epsilon_{t+\Delta}^2) = \sigma^2 r_t^{2\gamma} \Delta \tag{3.79}$$

where $\epsilon_{t+\Delta} = r_{t+\Delta} - r_t - (a+\beta r_t)\Delta$. Despite the approximate correctness of these requirements, this approach is, according to Stanton (1997), the simplest of all to implement and likely to introduce small approximation errors for reasonably frequent data available. Still Ait-Sahalia stands with criticism towards discretization methods as the one in Chan et al.. He points that although this method is commonly used, the discretization of the model involved rises considerations about the method's practical value; discretization-based methods implicitly assume that more data means more frequent data on a fixed period of observation, an assumption hardly matching real data deliverance.

It is certain that parametric estimators of drift and diffusion coefficients of a diffusion process are desirable for comprehensing the mechanism that generates the process. The true weakness of parametric estimators is misspecification a problem deteriorating when no apparent reasons can safely permit the selection of a model over another. The answer of recent research to the misspecification problem is the adoption of nonparametric estimation techniques. The nonparametric estimator is purely data driven and therefore avoids the specification of arbitrary functional forms for μ and σ. We have already revised a number of nonparametric estimators for both the unknown coefficients μ and σ.

Still distinguished researchers of the field created estimators that combined the parametric and non-parametric approach into a new semi-parametric approach. Banon's (1978) approach, analytically described in the previous section, can be considered as such. Briefly resuming, Banon (1978) integrated equation (3.61) and obtained

$$\mu(x) = \frac{1}{2\psi(x)} \frac{d}{dx} [\sigma^2(x)\psi(x)] \tag{3.80}$$

which allowed him to estimate the drift nonparametrically, given a nonparametric estimate of the stationary density ψ, while σ was known or constant.

Symmetricaly to Banon, Ait-Sahalia (1996a) assumed a linear drift,

$$\mu(x, \theta) = \beta[a - x] \tag{3.81}$$

with $\theta = (\alpha, \beta)'$ which by use of the backward Kolmogorov equation derives the conclusion

$$E[X_{t+\Delta}/X_t] = \alpha + e^{-\beta\Delta}(X_t - a) \tag{3.82}$$

Since ordinary least squares (OLS) clearly identifies the parameters γ and δ in

$$E[X_{t+\Delta} - X_t/X_t] = \gamma + \delta X_t \tag{3.83}$$

it is comprehensible that α and β of (3.82) are indirectly identified. The thus identified θ permits the non-parametric estimation of the diffusion function σ^2 since a marginal distribution ψ of the diffusion process is set. This is feasible due to the formula

$$\sigma^2(x) = \frac{2}{\psi(x)} \int_0^x \mu(u, \theta)\psi(u)du \tag{3.84}$$

which can be obtained by integrating (3.80) once more. More explicitly in order to obtain an estimate of $\sigma^2(x)$ we replace θ and $\psi(x)$ in (3.84) by their consistent estimators. Starting with $\psi(x)$ the smooth density estimator

$$\hat{\psi}(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) \tag{3.85}$$

can be formed by our data $\{X_i, i = 1, ..., n\}$ based on a kernel function $K(x)$ and bandwidth h_n while we provide an OLS estimator for $\theta = (\alpha, \beta)'$ by transformating the OLS

https://doi.org/10.26219/heal.aueb.6126

estimate of (γ, δ) in (3.83) as follows:

$$\alpha = -\gamma/\delta \qquad \text{and} \qquad \beta = -\ln(1+\delta)/\Delta \tag{3.86}$$

As far as the asymptotic distribution of the described estimator $\hat{\sigma}^2$ is concerned, Ait-Sahalia (1996) presents his findings in the following theorem (see also Ait-Sahalia (1996) page 535 and Appendix 1 page 550)

Theorem 2 *Under assumptions A1-A5 (see Appendix 1 in Ait Sahalia (1996)):*

- *The estimator $\hat{\sigma}^2$ is pointwise consistent and asymptotically normal, i.e.*

$$\sqrt{nh_n}\{\hat{\sigma}^2(x) - \sigma^2(x)\} \xrightarrow{d} N(0, V_{\sigma^2}(x)),$$

with asymptotic variance

$$V_{\sigma^2}(x) = \left\{\int_{-\infty}^{\infty} K(u)^2 du\right\} \sigma^4(x)/\psi(x)$$

- *The asymptotic variance can be consistently estimated by*

$$\hat{V}_{\sigma^2}(x) = \left\{\int_{-\infty}^{\infty} K(u)^2 du\right\} \hat{\sigma}^4(x)/\hat{\psi}(x)$$

- *At different points x and x' in $(0, \infty)$, $\hat{\sigma}^2(x)$ and $\hat{\sigma}^2(x')$ are asymptotically independent.*

Part (iii) of the theorem is typical of pointwise kernel estimators (see Robinson (1983)) and is useful to know for inference purposes. The consistent estimator of the pointwise asymptotic variance makes it possible to construct pointwise confidence intervals for the diffusion coefficient's estimate.

Apart from the use of a nonparametric estimate, it is also obvious that Banon (1978) and Ait-Sahalia's (1996) approach is also alternative to the common parametric approach as met in Vasicek (1977) or Cox, Ingersoll, and Ross (CIR) (1985) in more than one way. The joint parameterizations of (μ, σ^2) adopted in the literature of fully parametric estimation, imply specific forms for the marginal and transitional densities of the process that have to be accepted by the researchers with no further discussion. Instead the semi-parametric approach discussed advances towards the opposite direction; relying on the equivalence between (μ, σ^2) and densities it starts with non-parametric estimates of these densities and reconstructs the drift and diffusion of the continuous time process by matching these densities.

Finally we should not overlook the Bayesian analysis that was implemented by researchers to discretely observed diffusion processes. Roberts and Stramer (2001) as well as Elerian et al. (2001) use Markov chain Monte Carlo algorithms in order to sample from a properly transformed diffusion and subsequently draw conclusions about the likelihood function, the marginal likelihood and the parameters of the original diffusion process. The transformation of the diffusion's values aims at breaking down the dependency between the diffusion's volatility and the missing paths connecting two data points.

Chapter 4

Conclusion

This dissertation studies essential probabilistic and stochastic theory of diffusions and reviews the identification and estimation problem as confronted by various authors. It seems unavoidable that the extent of such an undertaking leaves plenty of room for overlooked issues. Such an issue can be considered the derivation of the exact confidence intervals that the drift, diffusion and marginal density estimators lie within, although in most cases the exact limiting distribution of these estimators is provided. Diffusion's invariant distribution is also estimated only through nonparametric methods with any parametric approach left aside.

A natural continuance of the present effort rests with the multidimensional case of the stochastic process X_t. More than a few researchers were concentrated in the multivariate case and important bibliography has been formatted, even not so extensive as the one concerning the univariate case. Density estimators for the case are revised for instance by Prakasa (1983) and Yamato (1971) while Genon-Catalot and Jacod (1993) estimate the diffusion coefficient matrix. Basawa and Prakasa (1980) include a broad discussion on multidimensional diffusion processes in their book. Bandi and Moloche (2002) recently discussed kernel methods for multivariate diffusion processes assuming only Harris recurrence instead of stationarity as a distributional property.

Separate from the issue of multidimensionality confronted in the particular paper, the recurrence consideration seems to be a promising area for research. Recurrence is an assumption allowing nonstationary behavior to the process and thus substantially milder than the extensively used stationarity assumption for the marginal process. Since in many empirical applications cross-restrictions delivered from the existence of a time-invariant marginal data density are unnatural to be imposed, work on reccurence seems to produce developments towards the right direction.

Bibliography

[1] Aït-Sahalia, Y. **(1996)**. Nonparametric Pricing of Interest Rate Derivative Securities, *Econometrica*, 64, 527-560

[2] **Bandi, F.M. and Moloche, G. (2002)**. On the Functional Estimation of Multivariate Diffusion Processes, *unpublished manuscript*, University of Chicago GSB and MIT

[3] **Bandi, F.M.** and Phillips, P.C.B. (2002). Fully Nonparametric Estimation of Scalar Diffusion Models, *unpublished manuscript*, University of Chicago GSB and Yale University

[4] **Banon, G.** (1978). Nonparametric Identification for Diffusion Processes, *SIAM Journal of Control and Optimization,* 16, 3, 380-395

[5] **Banon, G. and Nguyen, H.T. (1981)**. Recursive Estimation in Diffusion Models, *SIAM Journal of Control and Optimization,* 19, 676-685

[6] **Basawa, I.V. and Prakasa Rao,** B.L.S. (1980). *Statistical Inference for Stochastic Processes.* Academic Press, London

[7] **Bhattacharya, R.N. and Waymire, E.C. (1990)**. *Stochastic Processes with Applications.* Wiley, New York

[8] **Bowman, A.W. and Azzalini A. (1997)**. *Applied Smoothing Techniques for Data Analysis.* Clarendon Press, Oxford

[9] **Brennan, M.J. and Schwartz, E.S. (1979).** A Continuous Time Approach to the Pricing of Bonds, *Journal of Banking and Finance*, 3, 133-155

[10] **Brennan, M.J. and Schwartz, E.S. (1982).** An Equilibrium Model of Bond Pricing and a Test of Market Efficiency, *Journal of Financial and Quantitative Analysis*, 17, 210-239

[11] **Brown, B.M. and Hewitt, J.I. (1975).** Asymptotic Likelihood Theory for Diffusion Process, *Journal of Applied Probability*, 12, 228-238

[12] **Chan, K.C., Karolyi A.G., Longstaff F.A., and Sanders A.B. (1992).** An Empirical Comparison of Alternative Models of the Short-term Interest Rate, *Journal of Finance*, 47, 1209-1227

[13] **Chapman, D.A. and Pearson, N.D. (2000).** Is the Short Rate Drift Actually Nonlinear?, *Journal of Finance*, 55, 355-388

[14] **Cox, J.C., Ingersoll, J.E.Jr., and Ross, S.A. (1985).** A Theory of the Term Structure of Interest Rates, *Econometrica*, 53, 385-467

[15] **Dacunha-Castelle, D. and Florens-Zmirou, D. (1986).** Estimation of the Coefficient of a Diffusion from discrete observations, *Stochastics*, 19, 263-284

[16] **Dohnal, G. (1987).** On Estimating the Diffusion Coefficient, *Journal of Applied Probability*, 24, 105-114

[17] **Duffie, D. and Singleton, K.J. (1993).** Simulated Moments Estimation of Markov Models of Asset Prices, *Econometrica*, 61, 929-952

[18] **Elerian, O.S., Chib, S. and Shephard, N. (2001).** Likelihood Inference for Discretely Observed Nonlinear Diffusions, *Econometrica*, 69, 959-993

[19] **Fan, J. and Zhang, C. (2003).** A Reexamination of Diffusion Estimators With Applications to Financial Model Validation, *Journal of the American Statistical Association*, 98, 118-134

https://doi.org/10.26219/heal.aueb.6126

[20] **Fan, J., Zhang, C.M.,and Zhang, J. (2001).** Generalized Likelihood Ratio Statistics and Wilks Phenomenon, *The Annals of Statistics*, 29, 153-193

[21] **Florens-Zmirou, D. (1993).** On Estimating the Diffusion Coefficient from Discrete Observations, *Journal of Applied Probability*, 30, 790-804

[22] **Friedman, A. (1975).** *Stochastic Differential Equations and Applications, Volume 1.* Academic Press, New York

[23] **Gallant, R.A. and Tauchen, G.E. (1994).** Which moments to match?, *Econometric Theory*, Forthcoming

[24] **Geman, S.A. (1979).** On a Common Sense Estimator for the Drift of a Diffusion. Pattern Analysis 79, Division of Applied Mathematics, Brown University, Providence, Rhode Island.

[25] **Genon-Catalot, V. and Jacod, J. (1993).** On the Estimation of the Diffusion Coefficient for Multi-dimensional Diffusion Processes, *Ann. Inst. Henri Poincaré*, 29, 119-151

[26] **Gihman, I.I. and Skorohod, A.V. (1972).** *Stochastic Differential Equations.* Springer-Verlag, Berlin

[27] **Gourieroux, C., Monfort, A. and Renault, E. (1993).** Indirect Inference, *Journal of Applied Econometrics*, 8, S85-S118

[28] **Hansen, L.P. (1982).** Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica*, 50, 1029-1054

[29] **Hansen, L.P. and Scheinkman, J.A. (1995).** Back to the Future: Generating Moment Implications for Continuous-time Markov Processes, *Econometrica*, 63, 767-804

[30] **Härdle, W. (1990).** *Applied Nonparametric Regression.* Cambridge University Press, Cambridge

[31] **Härdle, W. (1991).** *Smoothing Techniques With Implemantation in S.* Springer-Verlag, New York

[32] **Itô, K. (1942).** Differential Equations Determining Markov Processes (in Japanese), *Zenkoku Shijō Sūgaku Danwakai*, 1077, 1352-1400

[33] **Itô, K. (1946).** On a Stochastic Integral Equation, *Proc. Imp. Acad. Tokyo*, 22, 32-35

[34] **Itô, K. (1951).** On Stochastic Differential Equations, *Mem. Amer. Math. Soc.*, 4, 1-51

[35] **Jiang, G.J. and Knight, J.L. (1997).** A Nonparametric Approach to the Estimation of Diffusion Processes, with an Application to a Short-term Interest Rate Model, *Econometric Theory*, 13, 615-645

[36] **Karatzas, I. and Shreve S.E. (1991).** *Brownian Motion and Stochastic Calculus. Second edition.* Springer-Verlag, New York

[37] **Karlin, S. and Taylor H.M. (1975).** *A First Course in Stochastic Processes. Second edition.* Academic Press, New York

[38] **Karlin, S. and Taylor H.M. (1981).** *A Second Course in Stochastic Processes.* Academic Press, New York

[39] **Kumar, T.K. and Markman, J.M. (1975).** Estimation of Probability Density Function: A Monte Carlo Comparison of Parametric and Nonparametric Methods. Mimeo.

[40] **Leblanc, F. (1995).** Wavelet Density Estimation of a Continuous Time Process and Application to Diffusion Process, *C. R. Acad. Sci. Paris*, 321, 345-350

[41] **Li, F. and Tcacz, G. (2002).** Nonparametric Estimation of Diffusion Processes with Discrete Observations, *Internet source:* http://math.ucsd.edu/~politis/CONTR/fuchun.pdf

[42] **Lo, A.W. (1988).** Maximum Likelihood Estimation of Generalized Ito Processes with Discretely Sampled Data, *Econometric Theory*, 4, 231-247

[43] **Moloche, G. (2001).** Local Nonparametric Estimation of Scalar Diffusions, *unpublished manuscript*, MIT

[44] **Monfort, A. (1996).** A Reappraisal of Misspecified Econometric Models, *Econometric Theory*, 12, 597-619

[45] **Oksendal, B.(1998).** *Stochastic Differential Equations, Fifth edition*, Springer-Verlag, Berlin

[46] **Pavlopoulos, H. and Kedem, B. (1992).** Stochastic modeling of rain rate processes: A diffusion model, *Commun. Statist.-Stochastic Models*, 8(3), 397-420

[47] **Pearson, N.D. and Sun, T.S. (1994).** Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, and Ross Model, *Journal of Finance*, 49, 1279-1304

[48] **Phillips, P.C.B. and Park, J. (1998).** Nonstationary Density Estimation and Kernel Autoregression, *Cowles Foundation Discussion Paper*, No. 1181, Yale University

[49] **Pinotsi, D.K. (1997).** *Non-parametric Diagnosis of Parametric Coefficients in a Diffusion Model for Processes of Spatially Averaged Rain Rate.* M.Sc. Dissertation, Department of Statistics, Athens University of Economics and Business, Athens, Greece. ISBN: 960-7929-23-3

[50] **Prakasa Rao, B.L.S. (1979).** Nonparametric Estimation for Continuous Time Markov Processes via Delta Families, *Publ. Inst. Statist. Univ. Paris*, 24, 79-97

[51] **Prakasa Rao, B.L.S. (1983).** *Nonparametric Functional Estimation.* Academic Press, Orlando

[52] **Prakasa Rao, B.L.S. (1999).** Nonparametric Functional Estimation: An overview. In *"Asymptotics, Nonparametrics and Time Series"*, 461-509, edited by S. Ghosh, Marcel Dekker

[53] **Roberts, G.O. and Stramer, O. (2001).** On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm, *Biometrika,* 88, 3, 603-621

[54] **Robinson, P.M. (1983).** Nonparametric Estimators for Time Series, *Journal of Time Series Analysis*, 4, 185-207

[55] **Silverman, B.W. (1986).** *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London

[56] **Stanton, R. (1997).** A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk, *The Journal of Finance*, 52, 1973-2002

[57] **Vasicek, O.A. (1977).** An Equilibrium Characterization of the Term Structure, *Journal of Financial Economics*, 5, 177-188

[58] **Walter, G. and Ghorai, J. (1992).** Advantages and Disadvantages of density estimation with wavelets. In *"Proceedings of the 24th Symp. on the Interface"*, edited by H. Joseph Newton, Interface FNA, VA 24: 234-243

[59] **Wong, E. (1971).** *Stochastic Processes in Information and Dynamical Systems.* McGraw-Hill, New York

[60] **Yamato, H. (1971).** Sequential Estimation of a Continuous Probability Density Function and Mode, *The Bulletin of Mathematical Statistics*, Vol. 14, No 3-4, pp.1-12

https://doi.org/10.26219/heal.aueb.6126

Δωρεά