View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by KU ScholarWorks

IDENTIFICATION AND ADAPTIVE CONTROL METHODS FOR SOME STOCHASTIC SYSTEMS

By Ioannis Zachariou

Submitted to the graduate degree program in Mathematics and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Chairperson Bozenna Pasik-Duncan

Tyrone Duncan

Jack Porter

Zsolt Talata

Arvin Agah

Date Defended: February 1, 2011

The Dissertation Committee for Ioannis Zachariou certifies that this is the approved version of the following dissertation:

IDENTIFICATION AND ADAPTIVE CONTROL METHODS FOR SOME STOCHASTIC SYSTEMS

Chairperson Bozenna Pasik-Duncan

Date approved: February 1, 2011

Abstract

This dissertation is focused on the identification and adaptive control of some stochastic systems. Initially a survey of some adaptive control problems for both discrete and continuous time stochastic systems is provided. Discrete time branching processes are described and some results on parameter estimation and adaptive control for these processes are reviewed. Then continuous time branching processes are introduced and the main results in this dissertation concerning estimation and adaptive control are given. The family of estimators is shown to be strongly consistent and the optimal rate of convergence of this family of estimators is obtained. Furthermore some other asymptotic properties of these estimators are verified. An adaptive control is given that posses self-tuning property. It is shown that it does not achieve the optimal asymptotic cost for the known system. Finally some computational methods and simulations are given for a variety of stochastic differential equations driven by a Brownian motion or an arbitrary fractional Brownian motion and computational properties of the parameter estimates for the branching processes are given.

Acknowledgments

I would like to express my gratitude to all the people that supported me and helped me during my studies. I especially would like to thank my advisor, Professor Bozenna Pasik-Duncan for all of her guidance, support, feedback and her determination to complete tasks. Professor Pasik-Duncan is the reason I followed this particular field of study as she inspired me through her class during my senior year, both through her teachings and guests hosts that demonstrated to us the multiple applications of our field. Furthermore, I would like to thank Professor Tyrone Duncan for his inputs, his help and his willingness to give feedback on any problem I would throw on the whiteboard. Also, special thanks to Professor Duncan for assuming the role of a co-advisor. I would like to thank Prof. Bernard Bercu. His inputs and his past work were invaluable for the completion of my dissertation. Also, I would like to thank Professors Yaozhong Hu and David Nualart for serving in my Oral Exam. I did particularly appreciate the feedback, comments and suggestions on numerical approximations and their rates of convergence. Though in the final dissertation I did not address the rates of convergence, I appreciated and personally enjoyed the discussions on the subject. Special thanks go to my defense committee, Professors Jack Porter, Zsolt Talata and Arvin Agah. I appreciate your support and I especially appreciate your commitment to participate in my defense in such weather conditions. I would also like to thank the Department of Mathematics, present and past Heads of the Department, Professors Satya Mandal, Jack Porter and Charlie Himmelberg, current and past staff, Gloria Prothe, Debbie Garcia, Samantha Reinblatt and Kerrie Brecheisen. In addition I would like to thank the Directors of Graduate Studies during my time Professors Dan Katz, Albert Sheu, Judy Roitman and Jack Porter. Moreover, my gratitude goes to the Graduate School, particularly Lea Smith and Dean Becca Peterson for their support, and to the ISSS for the years of assistance and the original offer through Fulbright for my undergraduate

studies at the University of Kansas. I owe my deepest gratitude to my family. I would like to thank my brother Giorgos Zacharia, Ph.D. and sister Margarita Zachariou, PhD. for their continuous feedback and support. Also I would like to thank my parents Maria and Christakis Zachariou for their unconditional love, support and inspiration. Additionally, I would like to thank my wife whose can-do attitude, her determination, patience and love helped me significantly for the last five years. Finally I would like to thank my closest of friends, Antonis Stylianou, Ph.D. for his support and suggestions for Matlab coding, Brian Russell, Ph.D. for his advice, and Evgeniya Duzhak Ph.D. and Leila Ahvarhoush for their support and motivation.

Table of Contents:

Preface

The problem of adaptive control of stochastic systems has been studied extensively. While the majority of the work has been for discrete time stochastic systems, the problem that is studied here is for continuous time stochastic systems. The study of continuous time stochastic systems is necessary for some problems that evolve naturally in continuous time. The continuous time theory is in fact relevant for discrete time schemes using high sampling rates. In other words it is important in discrete time systems for the case of small sampling times and for the analysis for computational questions. The continuous time stochastic theory is one way of addressing the question of robustness in continuous time adaptive schemes designed for deterministic signal models. The stochastic systems are modeled by Stochastic Differential Equations where the unknown parameters appear in the Stochastic Differential Equations that describe the state. We shall investigate a problem of the adaptive control of continuous time stochastic systems and we shall obtain a solution of it. The problem of adaptive control is to identify the system and simultaneously to control it. A solution of the adaptive control problem means that the family of parameter estimates is strongly consistent and that the average costs for the control problem using the parameter estimates converge almost surely to the optimal average cost. Our major focus here is on identification and adaptive control of discrete time and continuous time Branching Processes. We provide also some computational methods for Stochastic Differential Equations. In Chapter 1 we provide a short survey of adaptive control of some discrete time and continuous time stochastic systems. In Chapter 2 we provide the history and some examples of Branching Processes with their properties. In Chapter 3 we review the results on estimation and adaptive control of discrete time Branching Processes that were established mostly by B. Bercu. We extend these results to continuous time Branching Processes in Chapter 4. In Chapter 5 we present some computational methods along with the simulation

results for stochastic systems governed by Stochastic Differential Equations modeled by a Brownian Motion and Stochastic Differential Equations governed by a fractional Brownian Motion. Finally, in this version of the dissertation, we took in account the suggestions, comments and corrections suggested after the presentation of the material to the defense committee.

1 Chapter 1: Stochastic Adaptive Control

"In everyday language, 'to adapt' means to change a behavior to conform to new circumstances. Intuitively, an adaptive controller is thus a controller that can modify its behavior in response to changes in the dynamics of the process and the character of the disturbances." - Astrom and Wittenmark, Adaptive Control, 1995

The general approach to adaptive control that is described here exhibits a splitting or separation of identification and adaptive control. The solution to the adaptive control consists of strong consistency of the family of estimates and self-optimizing property of the adaptive control that uses the family of estimates. For the identification purposes we would like to find estimators such as: the maximum likelihood, least squares or weighted least squares estimators. For some cases the weighted least squares estimator is strongly consistent while the least squares estimator is not. The following issues are important for identification problems: strong consistency, recursivity, rate of convergence and asymptotic behavior of estimates. The adaptive control that we consider here is constructed by the so-called 'certainty equivalence' principle, that is the optimal stationary control, is computed by replacing the unknown parameter values by the current estimators of these values. The following issues are important for adaptive control:

1. Self-tuning property which means that asymptotically the adaptive control using the estimate of the unknown parameter is as good as the optimal control if we knew the system.

2. Self-optimizing property which means that the family of average costs converges to the optimal average costs.

3. Numerical and simulation methods for adaptive control.

To obtain a complete solution to the adaptive control is a challenging task. In many cases we

obtain a partial solution only.

Let us focus on important questions related to the adaptive control problem mentioned above. The first group of questions is related to the identification of stochastic systems. For the identification of a stochastic systems our desire is to obtain a good estimator of the unknown parameters in the system. What we mean by a good estimator is that we would like an estimator to be consistent or strongly consistent which means that it converges in probability or with probability one called also almost sure convergence. We would also like to have an estimator that is recursive and have a good rate of the convergence. It is also desired for the controller to know how an estimator behaves asymptotically therefore it is desired to analyze whether the limit theorems of probability such as Central Limit Theorem or Iterated Law of Logarithm hold as well as other theorems such as strong quadratic law or large deviation law. For adaptive control part it is desired as mentioned above to know the self-tuning property which means the performance of adaptive control coincides with the performance that would be obtained if the system parameters were known exactly. We will point out that it is not part of the self-tuning property that the parameter estimates should converge to their true values, this may not be necessary. To desire an adaptive control, a very natural way is to adopt a 'certainty equivalence' approach, consisting of the following steps (M.H.A. Davis and R.B. Vinter, 1985): Supposing an estimate on Θ is available at time n, then we apply the desired control u_n which would be optimal if $\hat{\Theta}_n$ were the true parameter value. The output of the system is observed and the estimate is updated to one that is obtained at the next time step. Then the procedure is repeated. It is important to notice that the desired control should be recursive. Such a procedure may be optimal in the sense of long-run average cost per unit time or in some other asymptotic sense. Choosing a right cost for some problems is a true challenge as we will see it in chapter 4. Often the self-tuning property is very desirable for the controller. In designing a proposed self-tuning algorithm one has to choose

1. A Class of Models to represent the system.

2. An estimation procedure.

3. A Control system.

There are many excellent textbooks and monographs on Stochastic Adaptive Control with a rich list of references there. Let us mention those that are used as recommended books in the course Math 750 - Stochastic Adaptive Control, the course that was developed at the Mathematics Department of the University of Kansas, by Bozenna Pasik-Duncan. Those include (Kumar and Varaiya, 1986), (Davis and Vinter, 1985), (Chen and Guo, 1991), (Astrom and Wittenmark, 1934), (Pasik-Duncan, 1986) and [3]. Math 750 Lecture notes were developed [1]. There were several masters and Ph.D thesis written under direction of B. Pasik-Duncan focusing on stochastic adaptive control and system identification including (A, Gao, 1996), (Z. Chen, 1995), (O. Zane, 1995), (D. Matache, 2000), (J. Murakami, 1995). Many research projects on stochastic adaptive control and identification were prepared by the members of the KU Stochastic Adaptive Control Group and several lecture notes were written by Shane Haas, Peter Zimmer and Dora Matatche and have been used by the students in the class of Math 750.

Let us describe the concept of adaptive control procedure. At each time t, the process x_t is observed based upon its value, the control u_t is selected from a set U. The parameter α has the constant value α_0 which is not known in advance. It is known, however, that α_0 belongs to a fixed finite set I or a compact set. The following adaptive control is considered. At each time t the estimate $\hat{\alpha}_t$ of the unknown parameter α is made. Having the estimate $\hat{\alpha}_t$ the control action is selected to be $u_t = \Phi(\hat{\alpha}_t, x_t)$.

One of the most interesting objectives is to analyze the asymptotic behavior of $\{\hat{\alpha}_t\}_{t\to\infty}$ and $\{u_t\}_{t\to\infty}$.

Some questions are the following:

1. Does the sequence $\{\hat{\alpha}_t\}_{t=1}^{\infty}$ converge almost surely?

2. Does the sequence $\{\hat{\alpha}_t\}_{t=1}^{\infty}$ converge to the true value α_0 almost surely?

3. Does the sequence $\{u_t\}_{t=1}^{\infty}$, where $u_t = \Phi(\hat{\alpha}_t, \cdot)$ converge almost surely?

4. Does the sequence $\{u_t\}_{t=1}^{\infty}$, where $u_t = \Phi(\hat{\alpha}_t, \cdot)$ converge to $u_0 = \Phi(\alpha_0, \cdot)$ almost surely?

5. Does the average cost $C_t = \frac{1}{t}\sum_{k=0}^{t-1} C(x_k.u_k)$ converge almost surely?

6. Does the average cost $C_t = \frac{1}{t}\sum_{k=0}^{t-1} C(x_k.u_k)$ converge to $J(\alpha_0)$ almost surely? Here $J(\alpha_0)$ is the optimal cost achievable for the parameter α_0, and we assume that it does not depend on the initial state x_0.

7. At what rate do these quantities converge, if they do so?

Let us review some specific stochastic adaptive control problems. For identification of a discrete time stochastic system, let us consider a discrete linear system given by the equation

$$x_{k+1} = \varphi_k'\theta + w_{k+1}, \qquad k = 0, 1, 2, ..., n-1 \tag{1.1}$$

where x_k are scalars. Assume that $\varphi_0, ..., \varphi_{n-1}, x_1, ..., x_n$ are available and $\{w_k\}$ is a white noise process. We wish to estimate the parameter θ. The least squares LS method is to find θ so that the quantity

$$\sum_{k=0}^{n-1}(x_{k+1} - \varphi_k'\theta)^2 \tag{1.2}$$

in minimized. The estimate of θ is

$$\hat{\theta}_n = (\sum_{k=0}^{n-1}\varphi_k\varphi_k')^{-1}\sum_{k=0}^{n-1}\varphi_k x_{k+1} \tag{1.3}$$

Even when the previous inverse does not exist, one can still show that the least square estimate exists. In the hypothesis that invertibility holds, and the matrix is positive definite, it is not hard to see that the following recursive least squares RLS estimates hold (see Kumar and Varaiya)

$$\hat{\theta}_{n+1} = \hat{\theta}_n + P_n \varphi_n (x_{n+1} - \varphi_n' \hat{\theta}_n);$$
$$P_n = P_{n-1} - \frac{P_{n-1}\varphi_n \varphi_n' P_n}{1 + \varphi_n' P_{n-1} \varphi_n} \tag{1.4}$$

We refer now to the well-known ARMAX model, Auto Recursive Average Systems with Exogenous inputs

$$x_{k+1} = a_0 x_k + a_1 x_{k-1} + ... + a_p x_{k-p} + b_0 u_k + b_1 u_{k-1} + ... + b_p u_{k-p} +$$

$$+ c_0 w_k + c_1 w_{k-1} + ... + c_p w_{k-p} + w_{k+1}$$

which can be written as

$$x_{k+1} = \varphi_k' \theta + w_{k+1}$$

where $\varphi_k = [x_k, x_{k-1}, ..., x_{k-p}, u_k, u_{k-1}, .., u_{k-p}, w_k, w_{k-1}, ..., w_{k-p}]'$, $\theta = [a_0, a_1, .., a_p, b_0, b_1, ..., b_p, c_0, c_1, ..., c_p]$ and $\{w_k\}$ is a white noise process. In the case that $c_0 = c_1 = ... = c_p = 0$ (ARX model), the least squares estimate of θ can be written as $\hat{\theta}_n = \theta + P_{n-1} \sum_{k=0}^{n-1} \varphi_k w_{k+1}$. Let $F_k = \sigma(x_s, u_s, w_s, s \leq k)$. Then the following result holds (Kumar and Varaiya, 1986)

Theorem 1.1 If $E[w_{k+1}|F_k] = 0$ for $k = 0, 1, ...$ and $E[w_{k+1}^2|F_k] \leq \sigma^2$, then $\hat{\theta}_n \to \theta$ a.s. on the set $\{\lim_{n\to\infty} \lambda_{min}(P_n^{-1}) = \infty, \frac{1}{\mathrm{Tr} P_n^{-1}} P_n^{-1} \geq \epsilon I$ for some $\epsilon > 0$ and all large $n\}$. (I is the identity matrix).

The theorem states that the LSE is strongly consistent. As a corollary, we have the next result.

Corollary 1.2 Suppose that the previous assumptions hold and that there exist some positively defined matrices U, V, such that $U \leq \frac{1}{n} P_n^{-1} \leq V$, for all $n \in \mathbb{N}$ a.s.. Then $\hat{\theta}_n \to \theta$ a.s..

The first of the two inequalities is the persistency of excitation condition, while the second one represents a stability condition.

If in the previous ARX model we consider $a_0 = a_1 = ... = a_p = 0$, then the LSE has the same formula as before, and under the assumption that $E[w_{k+1}|\varphi_s, s < \infty] = 0$ for $k = 0, 1, ...$ we get $E[\hat{\theta}_n] = \theta E[E[\hat{\theta}_n|\varphi_s, s \leq n-1]] = \theta + E[P_{n-1} \sum_{k=0}^{n-1} \varphi_k E[w_{k+1}|\varphi_s, s \leq n-1]] = \theta$. We have shown the following result.

Proposition 1.3 Under the assumption that $E[w_{k+1}|\varphi_s, s < \infty] = 0$ for $k = 0, 1, ...$, the LSE for the simple ARX model described above is unbiased.

Moreover,

Proposition 1.4 If $\lim_{n\to\infty} E[\mathrm{Tr}P_{n-1}] = 0$, the LSE $\hat{\theta}_n$ is consistent, i.e. $\hat{\theta}_n \to \theta$ in probability.

Let us come back to the general ARMAX model. The only available information is about inputs and outputs, so $w_k, w_{k-1}, ..., w_{k-p}$ are unknown. If we use the LS method to estimate only $a_0, a_1, ..., a_p, b_0, b_1, ..., b_p$ by redefining θ and φ_k in the convenient way, the LSE is biased since the noise process $\{c_0 w_k + c_1 w_{k-1} + ... + c_p w_{k-p} + w_{k+1}\}$ is not a white noise process, unless $c_0 = c_1 = ... = c_p = 0$. However, we could consider an extension of the LS procedure. We approximate the nonavailable information by $w_k = x_k - \varphi'_{k-1}\theta$. If at time $t, \hat{\theta}_k$ is known and if it is close to θ, then $\hat{w}_k = x_k - \varphi'_{k-1}\hat{\theta}_k$ is an approximation of w_k. We obtain the recursive procedure:

$$\hat{\theta}_{n+1} = \hat{\theta}_n + P_n \varphi_n (x_{n+1} - \varphi'_n \hat{\theta}_n); \tag{1.5}$$

$$P_{n+1}^{-1} = P_n^{-1} + \varphi_{n+1}\varphi'_{n+1}; \tag{1.6}$$

$$\varphi'_n = [x_n, x_{n-1}, ..., x_{n-p}, u_n, u_{n-1}, ..., u_{n-p}, \hat{w}_n, \hat{w}_{n-1}, ..., \hat{w}_{n-p}]; \tag{1.7}$$

$$\hat{w}_n = x_n - \varphi_n - \varphi'_{n-1}\hat{\theta}_n \tag{1.8}$$

This is called the extended least squares algorithm ELS or the pseudolinear regression algorithm PLR. The strong consistency holds under a few assumptions (Kumar and Varaiya, 1986)

Theorem 1.5 Under the assumption:

(i) $E[w_{k+1}|w_s, s \leq k] = 0, E[w_{k+1}^2|w_s, s \leq k] = \sigma^2, E[w_{k+1}^4|w_s, s \leq k] = \gamma$;

(ii) The polynomials $A(z) = 1 - a_0 z - a_1 z^2 - ... - a_p z^{p+1}$ and $C(z) = a - c_0 z - c_1 z^2 - ... - c_p z^{p+1}$, have all the roots outside the closed unit disc.;

(iii) The positive real condition holds, i.e. $\mathrm{Re}[\frac{1}{C(e^{i\omega})} - \frac{1}{2}] \geq 0$ for all ω;.

(iv) $\{u_k\}$ is F_k-adapted.

(v) $\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \varphi_k \varphi_k' = P > 0$;

then $\hat{\theta}_n \to \theta$ a.s.

Sometimes, the weighted least squares method (Bercu, 1995) can be used in the identification problem for linear systems. In this case one wants to minimize

$$\sum_{k=0}^{n-1} \gamma_k (x_{k+1} - \varphi_k' \theta)^2 \tag{1.9}$$

where $\{\gamma_k\}_n$ is called weighting sequence and $0 < \gamma_n \leq 1$. The weighting sequance is called admissible if $\sum_{n=0}^{\infty} \gamma_n^2 \varphi_n' P_{n+1} \varphi_n < \infty$. Here $P_n^{-1} = \sum_{k=0}^{n} \gamma_k \varphi_k \varphi_k'$. Recursive formulas are obtained and we have

$$\hat{\theta}_{n+1} = \hat{\theta}_n + \gamma_n \varphi_n (x_{n+1} - \varphi_n' \hat{\theta}_n); \tag{1.10}$$

$$P_n^{-1} = P_0^{-1} + \sum_{k=0}^{n} \gamma_k \varphi_k \varphi_k'. \tag{1.11}$$

When dealing with unknown controlled systems, one needs a way to identify the system and a method to control the identified system. As we mentioned in the beginning of this section, these two methods are integrated together in the stochastic adaptive control problem.

Example 1.6 Consider the discrete-time system

$$x_{k+1} = \theta u_k + w_{k+1}, \tag{1.12}$$

where $\{w_k\}$ is a white noise process and $\theta \in \mathbb{R}$ is unknown (non-Bayesian approach in which no prior distribution is attributed to θ). Assume that the control u_k is to keep x_k close to some fixed value $x \neq 0$ for all $k = 0, 1, 2, ...$ The cost criterion can be chosen as $\overline{\lim_{n\to\infty}} \frac{1}{n} \sum_{k=0}^{n-1} (x_{k+1} - x)^2$ (infinite-horizon cost criterion), or just $\sum_{k=0}^{n} (x_{k+1} - x)^2$ for any fixed n. If θ were known, then the optimal control would be $u_k = \frac{x}{\theta}$. If θ is unknown, the LSE is

$$\hat{\theta}_n = (\sum_{k=0}^{n-1} u_k^2)^{-1} \sum_{k=0}^{n-1} u_k x_{k+1}, \qquad n = 0, 1, 2, ...$$

Then $\hat{u}_n = \frac{x}{\hat{\theta}_n}, n = 1, 2, 3....$ minimized the previously mentioned cost criteria. In this control law the estimates are used as if they were the true value of the parameter. The control obtained this way is called certainty equivalence control. The following properties hold: (Kumar, Varaiya, 1986)

(i) $\hat{\theta}_n \to \theta$ a.s., that is the estimates are strongly consistent.

(ii) $\lim_{n\to\infty} \hat{u}_n = \frac{x}{\theta}$ a.s. Observe that this is exactly the control that would be used if θ were known from the beginning. We say that the adaptive control law is self-tunning with respect to the cost criteria described earlier.

(iii) $\overline{\lim_{n\to\infty}} \frac{1}{n} \sum_{0}^{n-1} (x_{k+1} - x)^2 = \overline{\lim_{n\to\infty}} \frac{1}{n} \sum_{0}^{n-1} w_{k+1}^2$. If θ were known, this would be the minimum cost. We say that with respect to this cost criterion, the adaptive control law is self-optimizing.

In continuous-time case the complete solution to the adaptive control of the linear quadratic Gaussian problem can be found in [5].

The problem is solved using only the natural assumptions of controllability and observability. The weighted least squares scheme is used to obtain the convergence of the family of estimates (self convergence). The scheme is modified by a random regularization to obtain the uniform controllability and observability of the family of estimates. A diminishing excitation white noise

is used to obtain strong consistency. The excitation is sufficient to include the identification of unknown deterministic linear systems. The approach eliminates some other assumptions that have previously been used that are unnecessary for the control problem for a known system and are often difficult to verify.

Let us describe weighted least squares identification in the continuous time case. Let $(X(t), t \geq 0)$ be the process that satisfies the stochastic differential equation

$$dX(t) = AX(t)dt + BU(t)dt + DdW(t) \tag{1.13}$$

or

$$dX(t) = \Theta^T \varphi(t)dt + DdW(t) \tag{1.14}$$

where

$$\Theta^T = [A, B], \varphi(t) = \begin{bmatrix} X(t) \\ U(t) \end{bmatrix}$$

$X(0) = X_0$, $X(t) \in \mathbb{R}^n$, $U(t) \in \mathbb{R}^m$, $(W(t), t \geq 0)$ is an $\mathbb{R}^p$ valued standard Wiener process, and $(U(t), t \geq 0)$ is a control from a family that is specified.

The random variables are defined on a fixed complete probability space (Ω, F, P) and there is a filtration $(F_t, t \geq 0)$ defined on this space. It is assumed that A and B are unknown.

A family of weighted least squares (WLS) estimates $(\hat{\Theta}(t), t \geq 0)$ is given by

$$d\hat{\Theta}(t) = a(t)P(t)\varphi(t)(dX^T(t) - \varphi(t)\hat{\Theta}(t)dt) \tag{1.15}$$

$$dP(t) = -a(t)P(t)\varphi(t)\varphi^T(t)P(t)dt \tag{1.16}$$

where $\Theta(0)$ and $P(0) > 0$ are arbitrary, and

$$a(t) = \frac{1}{f(\tau(t))} \tag{1.17}$$

with

$$\tau(t) = e + \int_0^1 |\varphi(s)|^2 ds \tag{1.18}$$

and

$$f \in F = \left\{ f | f : \mathbb{R}_t \to \mathbb{R}_t, f \text{ is slowly increasing}, \int_e^\infty \frac{dx}{xf(x)} < \infty \text{ and some } c \geq 0 \right\} \tag{1.19}$$

The following ergodic functional is used

$$J(U) = \limsup_{T \to \infty} \int_0^T [X^T(t) Q_1 X(t) + U^T(t) Q_2 U(t)] dt \tag{1.20}$$

where $(U(t), t \geq 0)$ is an admissible control, $Q_1 \geq 0$, $Q > 0$, and the following assumptions are made: (A, B) is controllable and $(A, Q_1^{1/2})$ is observable. For adaptive control, the diminishing excited lagged certainty equivalence control is used. For identification, to obtain the strong consistency for the family of estimates, a diminishing excitation is added to the adaptive control. The complete solution as defined at the beginning of this chapter to the adaptive control problem is obtained with the most natural assumptions. (Duncan, Guo, and Pasik-Duncan, 1999).

More on stochastic adaptive control can be also found in [4].

Survey on different types of convergence of random variables, limit theorems of probability as well as the introduction to martingale can be found in [11], [12], [13]. Survey on adaptive control tracking with all relevant references can be found in [14].

References

[1]. B. Pasik-Duncan, Lecture Notes, Math 750 - Stochastic Adaptive Control.

[2]. B. Pasik-Duncan, On Adaptive Contrl, SGH's - Publishers, Warsaw (1986).

[3]. Stochastic Theory and Control, LNCIS Vol. 2002 (2002). Springer - Verlag, Berlin. (B. Pasik-Duncan ed.).

[4]. T. Duncan, B. Pasik-Duncan, Stochastic Adaptive Control. The Control Handbook (W. Levine ed.). Taylor and Francis, to appear.

[5]. T. Duncan, L. Guo, B. Pasik-Duncan, Adaptive continuous time linear quadratic Gaussian control, IEEE Trans Autom. Control, 44 (1999).

[6]. P.R. Kumar, P. Varaiya, Stochastic Systems, Estimation, Identification and Adaptive Control, Prentice Hall, (1986).

[7]. B. Bercu, Weighted Estimation and Tracking for ARMAX models, SIAM J. Control and Optimization, Vol. 33 (1995).

[8] H.F. Chen, L. Guo, Identification and Stochastic Adaptive Control, Birkhauser, 1991.

[9] K.F Astrom, B. Wittenmark, Adaptive Control, Addison-Wesley Publishing Company, 1989.

[10] M.H.A. Davis, R.B. Vinter, Stochastic Modeling and Control, Chapman and Hall, 1985.

[11] R. Iranpour, P. Chacon, Basic Stochastic Processes, The Mark Kac Lectures, Macmillian Publishing Company, 1988

[12] S. Karlin, H.M. Taylor, A First Course in Stochastic Processes, Academic Press, Second Edition, 1975

[13] S. Ross, Stochastic Processes, John Wiley and Sons INC, Second Edition, 1996.

[14] Guo, Jinxin, Technical Report: Adaptive Control of Markov Chains, 1986.

[15] Yang, Zijiang, Technical Report: Introduction of adaptive control to queuing systems, 1991.

[16] Murakami, Junko, Technical Report: Parameter Identification for Stochastic Systems, 1995

[17] Haas, Shane, Technical Report: Ocular Artifact Removal through ARMAX systems, 1999

[18] Zane, Omar, Dissertation: Stochastic Adaptive control and its applications to the theory of finance, 1995

[19] Chen, Zhisheng, Dissertation: Some applications of adaptive control: Diffusion Approximations and heirarchial Approximation Schemes. 1995

[20] Gao, Ai-Jun, Dissertation: New approaches in continuous time stochastic adaptive control, 1996

[21] Matache, Dora, Dissertation: From reality to abstract in stochastic processes applied to telecommunications, 2000.

[22] Haas, Shane, Technical Report: ARMAX Model Identification

[23] Zimmer, Peter, Technical Report: Linear System Identifications.

[24] Matache, Dora, Oral Examination: on Stochastic Adaptive Control, technical report.

2 Chapter 2: Branching Process; History and Examples

2.1 History of Branching Process

Branching Processes have been traditionally studied as a part of Markov Processes and Renewal Theory. Traditionally they have been considered as a tool for applications in natural sciences and more particularly in the fields of Biology, Microbiology and Epidemiology (Jagers 1975). One can see the correlation as a Branching Process can describe populations according to the birth and/or death rates. In addition, a great interest exists in the continuous Branching Processes, particularly when they can be written in the form of Stochastic Differential equations. Another aspect of the Branching Processes that we are interested in is the estimation of unknown parameters that occur in the description of the Branching Processes.

A Branching Process (BP) is a stochastic process, and more specifically a Markov process, that models the size of the population at a given time t, based on assumptions on the length of the life of any individual and the resulting generation that is produced by that particular individual. It comes as no surprise that the process was first described, in its simplest form, in an attempt to study the survival of family names. The first recorded scholar to work on that problem was the French statistician Irenee-Jules Bienayme in 1845 (Jagers, 1975 and Heyde and Seneta, 1873). His success on the subject was that he provided a first formulation to represent mathematically the extinction of noble family names based on the mean number of males (offspring) for every male ancestor of the prior generation (Jagers, 1975). The process is more known as the Galton-Watson Process based on the work of English nobleman and science scholar Francis Galton and the mathematician Rev. Henry W. Watson. Galton was interested in investigating the survival/extinction of English aristocratic surnames. He proceeded to publish the question in the 1873 Educational Times. Watson answered his inquiry with a solution, which led to the publication of (Galton and

Watson, 1875), (Jagers, 1975). As such, the process is either referred to as Galton-Watson (GW) process or Bienayme-Galton-Watson (BGW) process. The process is commonly denoted as the family $\{Z(t,\omega), t \geq 0)\}$ with $Z(t,\omega)$ being the size of the population at time $t \geq 0$ and $\omega \in \Omega$, a sample space

Definition 2.1 Branching Process (Kimmel and Axelrod, 2002)

Let $\{Z(t,\omega), t \geq 0\}$ be a family of non-negative random variables defined on Ω with elements ω, where $Z(t,\omega)$ is the number of existing members of the colony at time t and ω the index of the generation. Assuming that the birth of the original ancestor occurs at $t = 0$, let $\tau(\omega)$ denote the life expectancy of the ancestor at generation ω and $X(\omega)$ be the count of the descendants that occurred at the death of the ancestor. Then:

$$Z(t,\omega) = \begin{cases} \sum_{i=1}^{X(\omega)} Z^i(t, \tau(\omega), \omega), & t \geq \tau(\omega) \\ 1 & t < \tau(\omega) \end{cases} \tag{2.1}$$

Since it has a self recurrent property we have

$$Z^i(t, \tau(\omega), \omega) = Z^i(t - \tau(\omega), \omega) \text{in distribution} \tag{2.2}$$

Replacing (2.2) in (2.1) we get:

$$Z(t,\omega) = \begin{cases} \sum_{i=1}^{X(\omega)} Z^i(t - \tau(\omega), \omega), & t \geq \tau(\omega) \\ 1 & t < \tau(\omega) \end{cases} \tag{2.3}$$

Next, it is important to discuss the probability generating function (p.g.f) of the process. Let us introduce the definition of p.g.f. of a random variable as it appears in (Kimmel and Axelrod, 2002):

Definition 2.2 The Probability Generating Function (pgf)

The pgf of a Z_+-valued random variable X is a function

$$f_X(s) = E(s^X) = \sum_{i=0}^{\infty} p_i s, \text{ where } s \in U \equiv [0,1] \text{ and } p_i = P(\{X = x_i\}) \tag{2.4}$$

The following theorem is a collection of the properties of the pgf (Kimmel and Axelrod, 2002), which are useful in the analysis of Branching Processes.

Theorem 2.1 Properties (Kimmel and Axelrod, 2002)

Suppose X is a Z_+ valued random variable with pgf $f_X(s)$ which may not be proper. Also assume the non-triviality condition: $p_0 + p_1 < 1$ (2.5), then:

1. f_X is a non-negative and continuous function with all derivatives on [0,1). Under (2.5) f_X is an increasing and convex function.

2. If X is proper, $f_X(1) = 1$; otherwise $f_X(1) = P[X < \infty]$.

3. $d^k f_X(0)/(ds^k) = k! p_k$.

4. If X is proper, the k^{th} factorial moment of X given by $\mu_k = E[X(X-1)(X-2)(X-k+1)]$ is finite if and only if $f_X^k(1-) = \lim_{s\to 1} f_X^k(s)$ is finite. In such case $\mu_k = f_X^k(1-)$.

5. If X and Y are independent Z_+ valued random variables then $f_{X+Y}(s) = f_X(s) f_Y(s)$

6. If Y is a Z_+ valued random variable and $\{X^i, i \geq 1\}$ is a sequence of independent identically distributed Z_+ valued random variables independent of Y, then $V = \sum_{i=1}^{Y} X^{(i)}$ has the pgf $f_V(s) = f_Y[f_{X^{(1)}}(s)]$.

7. Suppose that $\{X_i, i \geq 1\}$ is a sequence of Z_+ valued random variables and $\lim_{i\to\infty} f_{X_i}(s) = f_X(s)$ exists for each $s \in [0,1)$ if and only if the sequence $\{X_i, i \geq 1\}$ converges in distribution to a random variable X. Then $f_X(s)$ is the pgf of the limit of X.

By combining theorem 2.1 and the result from (2.3) one can deduce that:

$$f_t(s) = f(f_{t-\tau}(s)), t \geq \tau \tag{2.6}$$

As we have discussed the general properties of Branching processes, we would like to proceed to a classification that depends on the lifespan of the members of the colony.

Definition 2.3 The Branching Galton Watson Process (BGW Process)

Let $\{Z_t, t \geq 0)\}$ be Branching Process and assume the ancestor generates offsprings $\{X_i\}_{i=0}^{\infty}$ at time of death, where $X_n \in \mathbb{N}$. Furthermore, assume that the lifespan is identical for each member of the colony and it is equal to 1. Then the Branching Process is a BGW Process.

Definition 2.4 Markov Branching Process

Let $\{Z_t, t \geq 0)\}$ be a Branching process whose individuals have continuous lifetimes that are exponentially distributed. Then the process $\{Z_t, t \geq 0)\}$ is a Markov Branching Process.

Definition 2.5 Bellman-Harris Branching Process

Let $\{Z_t, t \geq 0)\}$ be a Branching process where the lifetime of the members is a non-negative random variable. Then the process is a Bellman-Harris Branching Process.

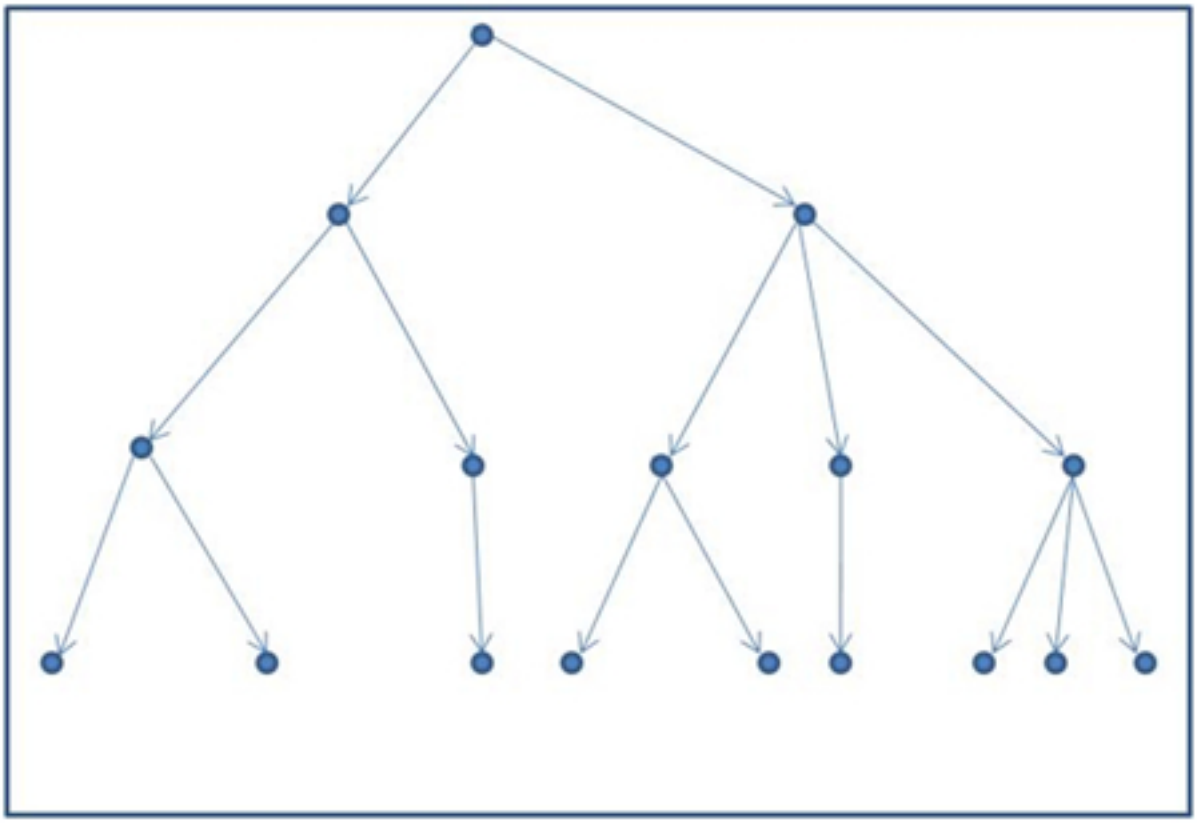


Figure 1. Example of a Galton-Watson Branching Process

The figure above gives an example of a single ancestor at time t=0 that produces two new members of the colony. Therefore at the original ancestor's death, the size of the colony is equal to the size of the progeny of that original member. Then the one member produces two new members and the second member produces three members. At the death of the two ancestors, the size of the colony is equal to the sum of their progeny. Therefore at $t = 2$ the colony has five members etc.

Next we want to discuss the mean for the offspring of every ancestor for the Branching Process, i.e. $m = E[X]$. The criticality of the Branching Process will be stated for the BGW process, though it holds as well for all cases of the Branching Process. Based on the properties from Theorem 2.1 and Definition 2.3 we have that

$$m = E[X] = f'(1-), \text{ and therefore} \tag{2.7}$$

$$E[Z_t] = f_t'(1-) \tag{2.8}$$

In addition, we have that

$$f_t(s) = f(f...f(s)...), \text{ composed t times, and we can therefore get that} \tag{2.9}$$

$$E[Z_t] = f_t'(1-) = f'(1-)^t = m^t \tag{2.10}$$

As the expression above gives as the expected value of the Branching Process based on the value of the mean of the offspring generated by each ancestor m, the trichotomy below can be stated:

i For $m > 1$, $\lim_{t\to\infty} E[Z_t] = \infty$, therefore supercritical

ii For $m = 1, \lim_{t\to\infty} E[Z_t] = 1$, therefore critical

iii For $m < 1, \lim_{t\to\infty} E[Z_t] = 0$, therefore subcritical

In the last two cases one can see that the process does not have enough energy, i.e. the members of the colony are not generating enough offspring to substitute the original members as they generate on average at most one offspring each. In fact the population will extinct almost surely.

Next we would like to show the probability generating function for the continuous Markov Branching Process (Kimmel and Axelrod, 2002). By the definition, the lifetime of an individual follows an exponential distribution with parameter λ. Then the cumulative distribution for Y, the lifetime of the individual, takes the form

$$G(t) = P(y \leq t) = 1 - e^{-\lambda t} \text{ for } t \geq 0, \tag{2.11}$$

which gives the p.d.f.

$$g(t) = G'(t) = \begin{cases} \lambda e^{-\lambda t} & \text{for } t > 0 \\ \\ 0 & \text{elsewhere} \end{cases} \tag{2.12}$$

Any ancestor, during its lifetime, will produce progeny according to the pgf $f(s) = \sum_{k=0}^{\infty} p_k s^k$. Defining $Z(t)$ as the total population size at time t, the process $\{Z(t), t \geq 0\}$ is a continuous time Markov Process with initial condition $Z(0) = 1$. Our purpose here is to derive a differential equation for the pgf $Q(s,t)$ for the population size $Z(t)$. We first use the fact that

$$Q(s, t + \Delta t) = Q(Q(s,t), \Delta t) \tag{2.13}$$

Taking $Q(z, 0) = z$ and small enough Δt we get

$$Q(s, \Delta t) = sP(y > \Delta t) + f(s)P(y \leq \Delta) + o(\Delta t) = se^{-\lambda t} + f(s)(1 - e^{-\lambda t} + o(\Delta t) \tag{2.14}$$

As such we can write

$$\begin{aligned} Q(s, t + \Delta t) - Q(s,t) &= Q(Q(s,t), \Delta t) - Q(s,t) \\ &= Q(s,t)e^{-\lambda \Delta t} + f(Q(s,t))(1 - e^{-\lambda \Delta t}) + o(\Delta t) - Q(s,t) \\ &= [-Q(s,t) + f(Q(s,t))](1 - e^{-\lambda \Delta t}) + o(\Delta t) \end{aligned} \tag{2.15}$$

Now, we divide both sides by Δt and take the limit as Δt goes to zero so that

$$lim_{\Delta t \to 0} \frac{Q(s, t + \Delta t) - Q(s,t)}{\Delta t} = [-Q(s,t) + f(Q(s,t))] \lim_{\Delta t \to 0} \frac{1 - e^{-\lambda \Delta t}}{\Delta t} + 0$$

which by the L'Hospital rule applied to the ratio $\frac{1 - e^{-\lambda \Delta t}}{\Delta t}$ we get:

$$\frac{dQ(s,t)}{dt} = -\lambda [Q(s,t) - f(Q(s,t)] \tag{2.16}$$

with initial condition $Q(s,0) = s$ and unique solution if $\lim_{s\to 1^-} Q(s,t) = 1$.

Example 2.1 Drug Resistance in Cancer Cells (Kimmel and Axelrod, 2002) and (Allen 2003) As mentioned earlier, Branching Processes can be applied in various fields, including but not limited to Biology. What we found particularly attractive was its application to cancer and cancer therapy. An excellent example is in chemotherapy and drug resistance found in various books (Kimmel and Axelrod, 2002) and (Allen 2003). In the scenario presented here the existence of two different types of cancer cells is assumed: Type A being the drug sensitive cells in a tumor and Type B being the drug resistant cells. Define y as the time it takes for a cell to divide (split), which y is going to be exponentially distributed with parameter $\lambda > 0$. Also lets define p the probability that out of the two generated cells from a Type A cell one of them will be a Type B cell. Another reasonable assumption is that every Type B cell will generated always two type B cells. In such a case, where members produce offspring in more than one way, a multi-type Branching Process is necessary. Therefore the probability generating functions satisfy:

$$f_1(z_1, z_2) = (1-p)z_1^2 + pz_1z_2 \text{ and} \tag{2.17}$$

$$f_2(z_1, z_2) = z_2^2 \tag{2.18}$$

Also, $Q(t) = \begin{bmatrix} Q_1(t) \\ \\ Q_2(t) \end{bmatrix}$ satisfies the differential equation (2.16) therefore by separation of variables we have the set of differential equations

$$\frac{dQ_1}{dt} = -\lambda Q_1 + \lambda[(1-p)Q_1^2 + pQ_1Q_2] \text{ and} \tag{2.19}$$

$$\frac{dQ_2}{dt} = -\lambda Q_2 + \lambda Q_2^2, \tag{2.20}$$

with initial condition $Q(z,0) = z$ Substituting the solution of the differential equation (2.20) in (2.19) we get

$$Q_1(z_1, z_2, t) = \frac{z_1 e^{-\lambda t}[z_2 e^{-\lambda t} + 1 - z_2]^{-p}}{1 + z_1[e^{-\lambda t} z_2 + 1 - z_2]^{-p} z_2^{-1}} \tag{2.21}$$

$$Q_2(z_1, z_2, t) = \frac{z_2}{z_2 + (1 - z_2)e^{\lambda t}} \tag{2.22}$$

The most interesting result for this example comes when the probability of no resistant cells at time t is evaluated, i.e.:

$$p(t) = \lim_{z_1 \to 1} \lim_{z_2 \to 0} Q_1 = \frac{1}{(1 - p + pe^{\lambda t})} \tag{2.23}$$

It is easy to see that the limit of (2.23) as $t \to \infty$, is equal to zero. This is a very interesting result as it indicates that without any outside influence, the probability of having no resistant cells is zero, except when $p = 0$. If p is actually equal to zero, it would imply that the probability of a drug sensitive cell to generate a drug resistant cell is itself zero (Allen 2003) and (Kimmel and Axelrod, 2002). More examples of Branching Processes and the proofs of most of the results in this chapter, as well as generalization of the theory of more than one type of Branching Processes (i.e. multi-type Branching Process mentioned earlier) can be found in (Karlin and Taylor, 1975).

References

1. H. W. Watson and F. Galton, On the Probability of the Extinction of Families, Journal of the Anthropological Institute of Great Britain, volume 4, pages 138-144, 1875.

2. M. Kimmel and D.E. Axelrod, Branching Processes in Biology, Springer, 2002.

3. L.J.S. Allen, An Application to Stochastic Processes with Applications to Biology, Prentice Hall, 2003

4. P. Jagers, Branching Processes with Biological Applications, Wiley, 1975

5. C.C. Heyde and E. Seneta, The simple branching process, a turning point test and a fundamental identity : a historical note on I.J. Bienayme, Biometrika, 59, 680-683, 1972

6. S. Karlin and H.M Taylor, A First Course in Stochastic Processes, Second Edition, Academic Press, Chapter 8, 1975

3 Chapter 3: Identification and Adaptive Control of Discrete Time Branching Processes

3.1 Introduction

As we mentioned in the prior chapter, we were attracted to branching processes due to the biological applications and more particularly the applications in cancer therapy and chemotherapy. What attracted us to investigate Branching Process's estimators for its parameters is the work of Bernard Bercu. The field is not necessarily new, and many authors have provided estimators (Wei and Winnicki, 1989 and 1990), (Winnicki, 1991) and (Heyde, 1974). The innovation in Bercu's (Bercu, 1999 and 2001) approach though is that he introduces the process in its ARMAX form and by choosing appropriate form of the control function he can manipulate the process with fewer restrictions on the mean and variance. As such, we present his approach as well as some of his theorems that we were motivated to investigate the continuous time case. In 3.2 we present a first scenario with no Immigration and in 3.3 we present the scenario were Immigration is nonzero.

3.2 BGW (No Immigration)

As we discussed earlier, the value of $E[X] = m$ is of great significance. Estimating m, as well as estimating σ^2, is a challenging task that requires strict assumptions. As the Least Squares Estimation (LSE) method does not necessarily provide strong consistency for the family of estimators, (Bercu, 2001) offers an approach by applying Weighted Least Squares (WLS) method with the choice of appropriate weights. Though his ultimate task is to estimate the parameters in the Bienayme-Galton-Watson with Immigration (BGWI) Process, it is only reasonable to initially investigate the case where Immigration is zero (null) i.e.

$$X_{n+1} = \sum_{i=1}^{X_n+U_n} Y_{n,i} \tag{3.1}$$

3.3 BGW Results

As we discussed earlier, the value of $E[X] = m$ is of great significance. Estimating m, as well as estimating σ^2, is a challenging task that requires strict assumptions. As the LSE does not necessarily provide strong convergence for the family of estimators, in [7], B. Bercu offers an approach by applying Weighted Least Squares (WLS) and appropriate weights. Though his ultimate task is to estimate the parameters in the BGWI (Bienayme-Galton-Watson with Immigration) scenario, it is only reasonable to initially investigate the case where Immigration is zero (null) i.e.:

$$X_{n+1} = \sum_{i=1}^{X_n+U_n} Y_{n,i} \tag{3.1}$$

where X_n is the size of the colony for generation $n, Y_(n, i)$ is the progeny count of each member of the nth generation and (U_n) is an adaptive control. The purpose of the control is to control X, i.e. boost it when it drastically declines to zero or force it lower when it drastically grows. Also, another important element of the approach in [7] is the rewriting of (3.1) in ARMAX form i.e. by setting

$$X_n = ma_n + \epsilon_n, \text{ and } a_n = X_{n-1} + U_{n-1} \tag{3.2}$$

where $\epsilon_n = X_n - am_n$.

In order to formulate the estimators the quadratic criterion is

$$\Delta_n(m) = \frac{1}{2} \sum_{k=1}^{n} a_k^{-1} (X_k - ma_k)^2 \tag{3.3}$$

with weights a_k^{-1}. The above is minimized by

$$\hat{m}_n = A_n^{-1} \sum_{k=1}^{n} X_k, \text{ where } A_n = \sum_{k=1}^{n} a_k \tag{3.4}$$

Also, the variance can be estimated by

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{k=1}^{n} a_k^{-1} (X_k - \hat{m}_n a_k)^2 \tag{3.5}$$

and the choice of the contrl for U_n is

$$U)n = \begin{cases} 1 - X_n & \text{if } P(\hat{m}_n^{-1} x_{n+1}) = 0 \\ P(\hat{m}_n^{-1} x_{n+1} - X_n & \text{otherwise} \end{cases} \tag{3.6}$$

where (x_n) is a sequence of non-negative integer valued random variables and P the projection operator on $\mathbb{N}$.

This particular choice of control satisfies the purposes we stated earlier. By employing the projection operator to the Natural numbers, we are assured that the summation makes sense. In addition, for the case where $P(\hat{m}_n^{-1} x_{n+1}) = 0$, i.e. the process is running out of energy and therefore would die out, it forces $X_n + U_n$ is at least equal to 1. Furthermore, due to the setup, the following theorems and lemmas were obtained (Bercu 2001):

Theorem 1. Assume that $(Y_{n,i})$ has a finite moment of order > 2 and that (x_n) converges a.s. to an integer $x \geq 0$. If we use the adaptive control given by (3.6), then $\hat{m}_n$ is a strongly consistent estimator of m.

$$(\hat{m}_n - m)^2 = O\left(\frac{\log n}{n}\right) a.s. \tag{3.7}$$

In addition, if $\alpha = \max\left(1, P(m^{-1}x)\right)$, we have the central limit theorem

$$\sqrt{n}\,(\hat{m} - m) \rightarrow^{\mathcal{L}} N\left(o, \frac{\sigma^2}{\alpha}\right) \tag{3.8}$$

i.e. it converges in distribution to a normal distribution.

Also, by the law of iterated logarithm we get that

$$\limsup_{n\to\infty}\left(\frac{n}{2\log\log n}\right)(\hat{m}_n - m)^2 = \frac{\sigma^2}{\alpha} \qquad a.s. \tag{3.9}$$

and the quadratic strong law we have that

$$\lim_{n\to\infty}\frac{1}{\log n}\sum_{k=1}^{n}(\hat{m}_n - m)^2 = \frac{\sigma^2}{\alpha} \qquad a.s. \tag{3.10}$$

In fact, the above Theorem (Bercu, 2001) is in essence the base for what we want to prove in the next chapter for continuous case as it covers the main properties for estimation and most importantly it provides the strongest rate of convergence possible for a family of estimators.

In addition, the same properties were shown (Bercu, 2001) for the family of estimators for the variance:

Theorem 2. Assume that $(Y_{n,i})$ has a finite moment of order > 2 and that (x_n) converges a.s. to an integer $x \geq 0$. If we use the adaptive control given by (3.6), then $\hat{\sigma}_n^2$ is a strongly consistent estimator of σ^2, i.e.

$$\left|\hat{\sigma}_n^2 - \sigma_n^2\right| = O\left(\frac{\log n}{n}\right) \qquad a.s. \tag{3.11}$$

In addition let τ^4 the fourth order centered moment of $(Y_{n,i})$ and set $\rho = \alpha^{-1}\tau^4 + (2 - 3\alpha^{-1})\sigma^4$, then, we have the central limit theorem

$$\sqrt{n}\left(\hat{\sigma}_n^2 - \sigma^2\right) \to^{\mathcal{L}} N\left(O, \rho\right) \tag{3.12}$$

Also by the law of iterated logarithm we get that

$$\limsup_{n\to\infty}\left(\frac{n}{2\log\log n}\right)\left(\hat{\sigma}_n^2 - \sigma^2\right)^2 = \rho \qquad a.s. \tag{3.13}$$

and by the quadratic strong law

$$\lim_{n\to\infty} \frac{1}{\log n} \sum_{k=1}^{n} \left(\hat{\sigma}_n^2 - \sigma^2\right)^2 = \rho \qquad a.s. \tag{3.14}$$

3.4 BGWI Results (with Immigration)

Another interesting result (Bercu, 2001) is when the same approach is used in the case where the BGW process has non-zero Immigration involved (therefore the BGWI case) given by

$$X_{n+1} = \sum_{i=1}^{X_n+U_n} Y_{n,i} + I_{n+1} \tag{3.15}$$

in which case we are also interested in the parameter of (I_n), namely λ and b^2. So, in similar fashion he sets (Bercu, 2001)

$$X_n = Ma_n + \lambda + \epsilon_n, \quad a_n = X_{n-1} + U_{n-1} \text{ where } \epsilon_n = X_n - ma_n - \lambda \tag{3.16}$$

therefore, attaining the stochastic regression equation in its traditional form used in the literature on identification and adaptive control:

$$X_n = \theta^T \Phi_n + \epsilon_n \text{ where } \theta^T = (m, \lambda) \text{ and } \Phi_n^t = (a_n, 1) \tag{3.17}$$

Then the quadratic criterion

$$\Delta_n(\theta) = \frac{1}{2} \sum_{k=1}^{n} \zeta_k (X_k - \theta^t \Phi_k)^2 \tag{3.18}$$

with weights $\alpha_n^{-1} = a_n (\log A_n)^\gamma$, with $A_n = \sum_{k=1}^{n} a_k + \gamma$ where $\gamma > 0$.

This was solved in (Bercu, 1999) to get

$$\hat{\theta}_n = S_n^{-1} \sum \alpha_k \Phi_k X_k \text{ where } S_n = \sum_{k=1}^{n} \alpha_k \Phi_k \Phi_k^T S \tag{3.19}$$

where S is a deterministic, symmetric and positive definite matrix.

For the variances of the two parameters

$$\eta^T = (\sigma^2, b^2) \tag{3.20}$$

the suggested family of estimators is:

$$\hat{\eta}_n = Q_n^{-1} \sum_{k=1}^{n} \beta_k \Phi_k \hat{\epsilon}_k^2 \tag{3.21}$$

where $\beta_n = a_n^{-1}\alpha_n$ and $Q_n = \sum \beta_k \Phi_k \Phi_k^t + Q$ and where Q is a deterministic, symmetric and positive definite matrix. Similarly to the case of the BGW Process in the earlier section , the adaptive control would be expected to be:

$$U_n = \begin{cases} 1 - X_n & \text{if } P\left(\hat{m}_n^{-1}\left(x_{n+1} - \lambda_n\right)\right) = 0 \\ P\left(\hat{m}_n^{-1}\left(x_{n+1} - \lambda_n\right)\right) - X_n & \text{otherwise} \end{cases} \tag{3.22}$$

Using the above control though, the estimator for Θ is not strongly consistent. Therefore an excitation (V_n) is considered, i.e.

$$U_n = P\left(\hat{m}_n^{-1}\left(x_{n+1} + \hat{\lambda}_n\right)\right) - X_n + V_n \tag{3.23}$$

where (V_n) is taken as an exogenous bounded sequence of identically distributed positive, integer valued random variables, V the nondegenerate distribution of (V_n). (Bercu, 2001).

Lemma 1. Assume that (x_n) converges to an integer $x \geq 0$. By using the adaptive control in (3.23), then we get a.s.

$$(\log n)^{1+\gamma}\frac{S_n}{n} \to H = \begin{bmatrix} E[h+V] & 1 \\ 1 & E[(h+v)^{-1}] \end{bmatrix} \text{ where } h = P(m^{-1}(x-\lambda)) \text{ and } S_n = \sum_{k=1}^{n} \alpha_k \Phi_k \Phi_k^T S \tag{3.24}$$

Theorem 3 Assume that (x_n) converges to an integer $x \geq 0$. By using the adaptive control in (3.9), then $\hat{\theta}_n$ is a strongly consistent estimator of θ, i.e.

$$\left\|\hat{\theta}_n - \theta\right\|^2 = O\left(\frac{(\log n)^{1+\gamma}}{n}\right) \qquad a.s. \tag{3.25}$$

In addition, assume that both $(Y_{n,i})$ and (I_n) have finite moments of order > 2. Then the central limit theorem states that

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \to^{\mathcal{L}} N(0, H^{-1}WH^{-1}) \tag{3.26}$$

where L denotes the convergence in probability and H is defined as in Lemma 1.

By the law of iterated logarithm we have that

$$\limsup_{n\to\infty}\left(\frac{n}{2\log\log n}\right)\left\|\hat{\theta}_n - \theta\right\|^2 \leq \lambda_{\max}(H^{-1}W^{-1}H^{-1}) \qquad a.s. \tag{3.27}$$

where

$$W = \begin{bmatrix} E[\sigma^2(h+V)+b^2] & E\left[\frac{\sigma^2(h+V)+b^2}{(h+V)}\right] \\ E\left[\frac{\sigma^2(h+V)+b^2}{(h+V)}\right] & E\left[\frac{\sigma^2(h+V)+b^2}{(h+V)}\right] \end{bmatrix}$$

Lemma 2. Assume that (x_n) converges to an integer $x \geq 0$ and that both $(Y_{n,i})$ and (I_n) have finite moments of order 4. If the adaptive control from (3.9) is used then a.s. we have

$$(\log n)^{1+\gamma}\frac{Q_n}{n} \to K = \begin{bmatrix} 1 & E[(h+V)^{-1}] \\ E[(h+V)^{-1}] & E[(h+V)^{-2}] \end{bmatrix} \tag{3.28}$$

where $Q_n = \sum_{k=1}^{n} \beta_k \Phi_k \Phi_k^T + Q$.

Finally, the last theorem (Bercu, 2001) we would like to mention is

Theorem 4. Assume that (x_n) converges to an integer $x \geq 0$ and that both $(Y_{n,i})$ and (I_n) have finite moments of order 4. By using the adaptive control in (3.23), then $\hat{\eta}_n$ is a strongly consistent estimator of η, i.e.

$$\|\hat{\eta}_n - \eta\|^2 = O\left(\frac{(\log n)^{1+\gamma}}{n}\right) a.s. \tag{3.29}$$

Also by the central limit theorem we have that

$$\sqrt{n}(\hat{\eta}_n - \eta) \rightarrow^{\mathcal{L}} N(0, K^{-1}PK^{-1}) \tag{3.30}$$

Finally, by the law of iterated logarithm we have that

$$\limsup_{n\to\infty} \left(\frac{n}{2\log\log n}\right) \|\hat{\eta}_n - \eta\|^2 \leq \lambda_{\max}(K^{-1}PK^{-1}) a.s. \tag{3.31}$$

where

$$P = \begin{bmatrix} E\left[\frac{R(h+V)}{(h+V)^2}\right] & E\left[\frac{R(h+V)}{(h+V)^3}\right] \\ E\left[\frac{R(h+V)}{(h+V)^3}\right] & E\left[\frac{R(h+V)}{(h+V)^4}\right] \end{bmatrix}$$

and $R(z) = 2z^2\sigma^4 + z(\tau^4 - 3\sigma^4 + 4b^2\sigma^2) + v^4 - b^4$.

The proofs for the theorems above, along with some other related theorems can be found in (Bercu, 1999 and 2001).

References

1. J. Winnicki, Estimation of the variances in the branching process with immigration, Th. Rel. Fields 88, 77 - 106 (1991).

2. Wei, C.Z. Wei and J. Winnicki, Some asymptotic results for the branching process with immigration, Stochastic Processes and their Applications, vol. 31m pp. 261 - 282 (1989).

3. C.Z. Wei and J. Winnicki, Estimation of the means in the branching process with immigration, Ann. Stat. 18, pp.1757 - 1773 (1990).

4. C.C Heyde, On estimating the variance of the o?spring distribution in a simple branching process, Adv. Appl. Probab., vol 6, pp. 421 - 433 (1974).

5. B. Bercu, Weighted estimation and tracking for branching processes with immigration, IEEE Transactions on Automatic Control, 46, pp. 43-50, 2001.

6. B. Bercu, Bernard, Weighted estimation and tracking for Bienayme Galton Watson processes with adaptive control, Statistics and Probability Letters 42, 415 - 421 (1999).

4 Chapter 4: Identification and Adaptive Control of Continuous Time Branching Processes

Abstract

In this chapter a controlled branching process is modeled in continuous time with a Brownian motion. An estimation problem is solved for an unknown intensity parameter in the stochastic equation that describes the branching process and an adaptive control is given that is self-tuning. Furthermore, similar results are obtained for a branching process with immigration. These results provide the initial work on parameter identification and adaptive control for continuous time branching processes.

4.1 Introduction

In this chapter the problems of identification and adaptive control are considered for continuous time branching processes. These processes are diffusion limits of discrete state space branching processes where each particle produces a random number of offsprings at death. The mean number of offsprings is $a+1$ where a is a parameter (the intensity) in the model. The diffusion coefficient in the stochastic differential equation for the branching process arises from the fact that the number of particles in a new generation $X(t+\Delta t)$ given that $X(t)=n$ is a sum of n independent, identically distributed random variables that are the offsprings of each particle at $X(t)$. Thus the total variance is proportional to $X(t)$. These processes have been considered for many years, e.g. Feller [8]. An extensive description of discrete branching processes is provided by Athreya and Ney [1]. For the intensity $a > 0$ and without a control the process explodes and is called supercritical. For the intensity $a < 0$ and without a control the process dies almost surely and is called subcritical. A maximum likelihood estimation approach to continuous branching processes without control is given by Overbeck [10]. The work in this chapter has been motivated by the work of Bercu [3, 4]

and Duncan, Guo and Pasik-Duncan [7].

4.2 Branching without Immigration

Initially consider a branching process without immigration. It is described by the following stochastic differential equation

$$dX(t) = aA(t)dt + c\sqrt{A(t)}dW(t)$$
$$X(0) = X_0 \tag{4.1}$$

where $(W(t), t \geq 0)$ is a standard Brownian motion on the complete probability space $(\Omega, \mathcal{F}, P)$, c is a constant, a is a constant, $(A(t), t \geq 0)$ is a stochastic process adapted to the filtration of W. It is assumed that $A(t) > 0$ a.s. for all $t \geq 0$. Furthermore it is assumed that the intensity a is an unknown nonzero parameter. To estimate the branching intensity a the following least squares family of estimators, $(\hat{a}(t), t \geq 0)$, is used

$$\hat{a}(t) = \frac{\int_0^t dX(s)}{\int_0^t A(s)ds} \tag{4.2}$$

so that

$$\hat{a}(t) - a = c\frac{V(t)}{C(t)} \tag{4.3}$$

where

$$V(t) = \int_0^t \sqrt{A(s)}dW(s) \tag{4.5}$$
$$C(t) = \int_0^t A(s)ds \tag{4.5}$$

It is verified subsequently that the family of estimators $(\hat{a}(t), t \geq 0)$ is strongly consistent. The parameter c^2 can be estimated from the quadratic variation of the Brownian motion or it can be done asymptotically using the quadratic strong law for Brownian motion. This quadratic strong

law is given now. For a standard Brownian motion it is verified by Brosamler [6] and more generally for the family of fractional Brownian motions by Berkes and Horvath [5].

Theorem 4.1

Let $(B(t), t \geq 0)$ be a standard Brownian motion on a complete probability space. The following equality is satisfied

$$lim_{t\to\infty} \frac{1}{log(t)} \int_1^t \frac{B^2(s)}{s^2} ds = 1 \qquad a.s. \tag{4.6}$$

It is noted subsequently how to use this result to obtain a strongly consistent family of estimators of c^2 which is the square of c in (4.1).

The control $(U(t), t \geq 0)$ in the drift term of (4.1) is required to contribute to the consistent estimation of the unknown intensity a and also follow a given curve or independent process $(x(t), t \geq 0)$ that is absolutely continuous such that

$$x(t) = \int_0^t z(s) ds \tag{4.7}$$

Such a problem is called a tracking problem because the branching process is required to follow a known curve or process. This tracking requirement is often very natural to obtain a desired performance for a branching process. Since

$$dX(t) - dx(t) = aA(t)dt - z(t)dt + c\sqrt{A(t)}dW(t) \tag{4.8}$$

it follows that if the intensity a were known then it is natural to determine $(U(t), t \geq 0)$ by requiring the drift in (4.8) to be zero, that is,

$$z(t) = aA(t) \tag{4.9}$$

because the noise process (martingale) cannot be directly removed or predicted. Since the intensity a is unknown a natural choice of an adaptive control $(U(t), t \geq 0)$ is to satisfy the equation

$$A(t) = max\{\hat{a}^{-1}(t-\delta)z(t), d\} \tag{4.10}$$

for $t \geq \delta$ and $A(t) = d$ for $t \in [0, \delta)$ where $d > 0$ is a sufficiently small constant and $\delta > 0$ is a constant. The use of the lagged estimators of a in (4.10) is natural for the implementation of the adaptive control and it has been used in other cases for adaptive controls, e.g. [7]. The performance of the tracking problem can be quantified by the following quadratic cost functional

$$J(t) = \frac{1}{log(t)} \int_1^t \frac{(X(\tau_C(s)) - x(\tau_C(s)))^2}{C^2(\tau_C(s))} ds \tag{4.11}$$

where $C(\tau_C(s)) = s.$ and the associated asymptotic cost is defined as

$$limsup_{t \to \infty} J(t) \tag{4.12}$$

Since the cost is of asymptotic-type it suffices to let $X(0) = 0$ so that

$$X(t) = aC(t) + cV(t) \tag{4.13}$$

$$X(t) - x(t) = \pi(t) + cV(t) \tag{4.14}$$

where

$$\pi(t) = aC(t) - x(t) \tag{4.15}$$

Now let $c^2(t)$ satisfy

$$c^2(t) = \frac{1}{log(t)} \int_1^t \frac{c^2 V^2(\tau_C(s))}{C^2(\tau_C(s))} ds \tag{4.16}$$

where τ_C satisfies $C(\tau_C(s)) = s$ and the parameter c^2 is the square of c in (4.1). The family of estimators $(c^2(t), t \geq 0)$ is strongly consistent for c^2 by the quadratic strong law for a standard Brownian motion because $(V(\tau_C(s), s \geq 0)$ is a Brownian motion by a well known result of P. Levy. From the definitions of $J(t)$ and $c^2(t)$ it follows that

$$|J(t) - c^2(t)| = O(\frac{1}{log(t)} \int_1^t \frac{\pi^2(\tau_C(s))}{C^2(\tau_C(s))} ds) \tag{4.17}$$

and

$$liminf_{t \to \infty} J(t) \geq c^2 \qquad \text{a.s.} \tag{4.18}$$

There is equality in the above inequality if and only if

$$lim_{t\to\infty}\frac{1}{log(t)}\int_1^t \frac{\pi^2(\tau_C(s))}{C^2(\tau_C(s))}ds = 0 \qquad \text{a.s.} \tag{4.19}$$

Thus the tracking problem has an optimal asymptotic-type control if $(J(t), t \geq 0)$ converges a.s. to c^2.

The following result describes some important properties of the family of estimators $(\hat{a}(t), t \geq 0)$.

Theorem 4.2

Assume that a in (4.1) is nonzero and that $(z(t), t \geq 0)$ in (4.7) converges a.s. to a positive real number z. Let $\delta > 0$ be fixed. If $(\hat{U}(t), t \geq 0)$ is the adaptive control from the equality

$$A(t) = max\{\hat{a}^{-1}(t-\delta)z(t), d\} \tag{4.20}$$

for $t \geq \delta$ and $A(t) = d$ for $t \in [0, \delta)$, then $(\hat{a}(t), t \geq 0)$ given by (4.2) is a strongly consistent family of estimators for a. Furthermore if $\alpha = max\{a^{-1}z, d\}$, then there is the following law of the iterated logarithm for the estimation error

$$\begin{aligned} &limsup_{t\to\infty}(\frac{t}{2loglog(t)})^{\frac{1}{2}}(\hat{a}(t)-a) \\ &= -liminf_{t\to\infty}(\frac{t}{2loglog(t)})^{\frac{1}{2}}(\hat{a}(t)-a) \\ &= \frac{c}{\sqrt{\alpha}} \qquad \text{a.s.} \end{aligned} \tag{4.21}$$

so that

$$limsup_{t\to\infty}(\frac{t}{2loglog(t)})(\hat{a}(t)-a)^2 = \frac{c^2}{\alpha} \qquad \text{a.s.} \tag{4.22}$$

There are also the following other asymptotic properties for the estimation error of the family of estimators $(\hat{a}(t), t \geq 0)$.

- A central limit theorem

$$\sqrt{t}(\hat{a}(t)-a) \to N(0, \frac{c^2}{\alpha}) \tag{4.23}$$

where the convergence is in distribution and $N(\cdot,\cdot)$ is a normal distribution.

- A quadratic strong law

$$\frac{1}{log(t)}\int_1^t(\hat{a}(\tau_C(s))-a)^2ds=1 \qquad \text{a.s.} \tag{4.24}$$

where τ_C satisfies $C(\tau_C(s))=s$.

- A large deviations principle

For each $\gamma>0$

$$lim_{t\to\infty}logP(|\hat{a}(\tau_C(t))-a|\geq\gamma)=-\frac{\gamma^2}{2} \tag{4.25}$$

Proof. Recall the equality (4.3)

$$\hat{a}(t)-a=c\frac{\int_0^t\sqrt{A(s)}dW(s)}{\int_0^t A(s)ds} \tag{4.26}$$

The strong consistency of $(\hat{a}(t),t\geq 0)$ can be obtained from the Strong Law of Large Numbers for a Brownian motion by a random time change or directly from the Strong Law of Large Numbers for a continuous martingale. Recall that $lim_{t\to\infty}C(t)=\infty \qquad a.s.$ so that

$$lim_{t\to\infty}(\hat{a}(t)-a)=0 \qquad \text{a.s.} \tag{4.27}$$

Since it is assumed that $z(t)\to z$ a.s. and from the previous equality $\hat{a}(t)\to a$ a.s. as $t\to\infty$, it follows that

$$lim_{t\to\infty}A(t)=A(\infty)=max\{a^{-1}z,d\}=\alpha \tag{4.28}$$

and α is strictly positive.

The Law of the Iterated Logarithm for a Brownian motion provides the following equality

$$limsup_{t\to\infty}(\frac{C(t)}{2loglog(C(t))})^{\frac{1}{2}}(\hat{a}(t)-a) \tag{4.29}$$

$$= -liminf_{t\to\infty}(\frac{C(t)}{2loglog(C(t))})^{\frac{1}{2}}(\hat{a}(t)-a) = c \qquad \text{a.s.}$$

and

$$limsup_{t\to\infty}(\frac{C(t)}{2loglog(C(t))})(\hat{a}(t)-a)^2 = \frac{c^2}{\alpha} \qquad \text{a.s.} \tag{4.30}$$

Since $A(\infty)=\alpha$ it follows that

$$lim_{t\to\infty}(\frac{C(t)}{t})^{\frac{1}{2}} = \sqrt{\alpha} \qquad a.s. \tag{4.31}$$

If $f(x) = loglog(x)$ then $Df(x) = \frac{1}{xlog(x)}$. Applying Taylor's formula up to the first derivative or a Mean Value Theorem it follows that

$$lim_{t\to\infty}\frac{loglog(t)}{loglog(C(t))} = 1 \qquad a.s. \tag{4.32}$$

Thus it follows that

$$limsup_{t\to\infty}(\frac{t}{2loglog(t)})(\hat{a}(t)-a)^2 = \frac{c^2}{\alpha} \qquad \text{a.s.} \tag{4.33}$$

By the Central Limit Theorem for a Brownian motion and Slutsky's Theorem, it follows directly that

$$\sqrt{t}(\hat{a}(t)-a) \to N(0,\frac{c^2}{\alpha}) \tag{4.34}$$

where the convergence is in distribution.

The quadratic strong law for a standard Brownian motion given in the previous theorem verifies (4.24).

The result for the large deviations principle follows from the result for Brownian motion, that is, for each $\gamma > 0$

$$lim_{t\to\infty}\frac{1}{t}logP(|\frac{B(t)}{t}| \geq \gamma) = -\frac{\gamma^2}{2} \tag{4.35}$$

■

The asymptotic behavior of the estimation error in (4.30) that uses the Law of the Iterated Logarithm provides the optimal convergence rate for the family of estimators $(\hat{a}(t), t \geq 0)$.

Remark If $z > 0$, then the theorem is valid for a control $(U(t), t \geq 0)$ that for large t is determined from the equality $A(t) = \hat{a}(t-\delta)z(t)$

The adaptive control defined by (4.10) is self-tuning, that is, it converges to the optimal control given by

$$A^*(t) = max\{a^{-1}z(t)\} \tag{4.36}$$

where it is assumed that $a^{-1}z > d > 0$. From the definition of the A(t) in (4.10) it is clear that

$$lim_{t\to\infty}A(t) = limA^*(t) = a^{-1}z \qquad a.s. \tag{4.37}$$

by the strong consistency of $(\hat{a}(t), t \geq 0)$. This self-tuning is an important and desirable property of an adaptive control.

In the determination of the properties of the adaptive control (4.10) it is natural to consider the question of self-optimality, that is, can the optimal control A^* be replaced by the adaptive control (4.10) and still achieve the optimal cost? It is shown in the next result that the answer to this question is negative, that is, the adaptive control is not self-optimizing for the cost (4.12).

Theorem 4.3

The adaptive control given by (4.10) is not self-optimizing for the cost (4.12), that is, this control does not achieve the optimal control for the control with the intensity a known.

Proof.

$$\begin{aligned} J(t) &= \frac{1}{log(t)}\int_1^t \frac{(X(\tau_C(s)) - x(\tau_C(s)))^2}{C^2(\tau_C(s))}ds \\ &= \frac{1}{log(t)}\int_1^t \frac{(\int_1^{\tau_C(s)}(a\hat{a}^{-1}(r-\delta)z(r) - z(r))dr + cV(\tau_C(s)))^2}{C^2(\tau_C(s))}ds \end{aligned} \tag{4.38}$$

To verify self-optimality it is necessary to show that

$$lim_{t\to\infty}\frac{1}{log(t)}\int_1^t \frac{(\int_0^{\tau_C(s)}(a\hat{a}^{-1}(r-\delta)z(r)-z(r))dr)^2}{C^2(s)}ds=0 \qquad a.s. \tag{4.39}$$

The interchange of limit and expectation can be justified by bounding any sequence of integrals from (4.39). It suffices to consider only this term because the other term cannot cancel this term e.g. V and $-V$ have the same probability law. If the equality (4.39) is satisfied then the following equality is satisfied.

$$lim_{t\to\infty}E\frac{1}{log(t)}\int_0^t \frac{(\int_0^{s}(a\hat{a}^{-1}(r-\delta)z(r)-z(r))dr)^2}{C^2(\tau_C(s))}ds=0 \tag{4.40}$$

It follows directly that

$$\begin{aligned}&(\int_1^s a\hat{a}^{-1}(r-\delta)z(r)-z(r)dr)^2\\ &=(\int_1^s \frac{(a-\hat{a}(r-\delta)z(r)}{\hat{a}(r-\delta)a}dr)^2\end{aligned} \tag{4.41}$$

As $s\to\infty$ a lower bound on (4.41) is

$$\begin{aligned}&K\int_1^s\frac{B(p)}{p}dp\int_1^s\frac{B(q)}{q}dqdp\\ &=2K\int_1^s\int_1^p\frac{B(p)B(q)}{pq}dqdp\end{aligned} \tag{4.42}$$

where $(B(t),t\geq 0)$ is a standard Brownian motion. Then

$$\begin{aligned}&E\int_1^s\int_1^p\frac{B(p)B(q)}{pq}dqdp\\ &=\int_1^s\frac{p-1}{p}dp=s-1-logs\end{aligned} \tag{4.43}$$

Since $1+log(s)=o(s)$ as $s\to\infty$ and

$$lim_{t\to\infty}\frac{1}{log(t)}\int_1^t\frac{1+log(s)}{s^2}ds=0 \tag{4.44}$$

so it suffices to consider the first term from the integration

$$\frac{1}{log(t)}\int_1^t \frac{1}{s}ds = 1 \tag{4.45}$$

Thus the expectation of (4.39) is strictly positive so there is measurable set Λ such that $P(\Lambda) > \gamma > 0$ and

$$limsup_{t\to\infty} 1_\Lambda \frac{1}{log(t)}\int_0^t \frac{(\int_0^{\tau_C(s)}(a\hat{a}^{-1}(r-\delta)z(r)-z(r))dr)^2}{C^2(\tau_C(s))}ds > \gamma \tag{4.46}$$

Thus the adaptive control (4.10) is not self-optimizing.

■

It is natural to try some other scaling for the integrand in (4.39). However it is seen by choosing some of these scalings or considering the ratios that occur in the integrand that the term (4.39) is comparable to the square of the martingale so the term corresponding to (4.39) will not converge to zero. This lack of self-optimality is in contrast to the adaptive control problems for discrete or continuous time linear systems or discrete time branching processes ([2, 7, 9]). To obtain a self-optimizing adaptive control it is necessary to determine a strongly consistent family of estimators that has a better rate of convergence than the least squares family. It is natural to consider some type of weighted least squares (WLS) estimators. However it seems that the weights that are used for WLS estimators for parameter estimation for linear systems ([2] for discrete time and [9] for continuous time) are not sufficient for the problem here.

4.3 Branching with Immigration

Now it is assumed that the branching process also has immigration. Such a situation can be described by the following stochastic differential equation

$$dX(t) = aA(t)dt + bdt + c\sqrt{A(t)}dW(t) \tag{4.47}$$

This equation is similar to (4.1) with an additional constant drift b which describes the immigration for the branching process. If it is assumed that a is unknown as in (4.47) and that b is known then a family of strongly consistent estimators is obtained by removing the immigration drift b, that is,

$$\hat{a}(t) = \frac{\int_0^t (dX(s) - bds)}{\int_0^t A(s)ds} \tag{4.48}$$

It follows directly from the previous section that $(\hat{a}(t), t \geq 0)$ is a strongly consistent family of estimators of a.

On the other hand if b is an unknown parameter but a is known then the family of estimators of b given by

$$\hat{b}(t) = \frac{\int_0^t (dX(s) - aA(s)ds)}{t} \tag{4.49}$$

is strongly consistent where it is assumed that $C(t)$ and t have comparable growth as is the case for the application here.

If the two parameters a and b are both unknown then because they both appear linearly in the drift, they cannot be determined separately from an estimation procedure on the same equation (4.47). Thus it is assumed that only one of the parameters is unknown.

In the former case of unknown a it is natural to choose the control $(U(t), t \geq 0)$ such that

$$A(t) = max\{\hat{a}^{-1}(t-\delta)(x(t) - b), d\} \tag{4.50}$$

for $t \geq \delta$ and $A(t) = d$ for $t \in [0, \delta)$ and in the latter case of unknown b it is natural to choose the control $(U(t), t \geq 0)$ such that

$$A(t) = max\{a^{-1}(x(t) - \hat{b}(t-\delta)), d\} \tag{4.51}$$

for $t \geq \delta$ and $A(t) = d$ for $t \in [0, \delta)$. In both cases the adaptive control is self-tuning. These results are collected in the following theorem.

Theorem 4.4

- Assume that a is nonzero and unknown and that b is known and that $(z(t), t \geq 0)$ converges a.s. to a positive real number z. If $(\hat{U}(t), t \geq 0)$ is the adaptive control from the equation (4.50) then $(\hat{a}(t), t \geq 0)$ is a strongly consistent family of estimators for a and the asymptotic properties of this family is the same as in the corresponding theorem in the previous section. Furthermore, the adaptive control is self-tuning.

- If a is nonzero and known and b is unknown and $(z(t), t \geq 0)$ converges a.s. to a positive real number z and the adaptive control is given by (4.51), then the family of estimators of b, $(\hat{b}(t), t \geq 0)$, is strongly consistent for b. The asymptotic properties of this family of estimators is the same as $(\hat{a}(t), t \geq 0)$ in the corresponding theorem in the previous section. Furthermore, the adaptive control is self-tuning.

References

[1] K. B. Athreya and P. E. Ney, *Branching Processes*, Springer, Berlin, 1972.

[2] B. Bercu, Weighted estimation and tracking for ARMAX models, *SIAM J. Control Optim.* 33 (1995), 89-106.

[3] B. Bercu, Weighted estimation and tracking for Bienayme Galton Watson processes with adaptive control, *Stat. Prob. Lett.* 42 (1999), 415-421.

[4] B. Bercu, Weighted estimation and tracking for branching processes with immigration, *IEEE Trans. Autom. Control* 46 (2001), 43-50.

[5] I. Berkes and L. Horvath, Limit theorems for logarithmic averages of fractional Brownian motions, *J. Theor. Probab.* 12 (1999), 985-1009.

[6] G. Brosamler, An almost everywhere central limit theorem, *Math. Proc. Cambridge Phil. Soc.* 104 (1988), 561-574.

[7] T. E. Duncan, L. Guo, and B. Pasik-Duncan, Adaptive continuous-time linear quadratic Gaussian control *IEEE Trans. Autom. Control* 44 (1999), 1653-1662.

[8] W. Feller, Two singular diffusions, *Ann. Math.* 54 (1951), 173-182.

[9] A. J. Gao and B. Pasik-Duncan, Stochastic linear quadratic adaptive contol for continuous-time first-order systems, *Systems Control Lett.* 31 (1997), 149-154.

[10] L. Overbeck, Estimation for continuous branching processes, Scand. J. Statist. 23 (1998), 111-126.

5 Chapter 5: Computational Methods for Stochastic Differential Equations Involving Standard Brownian Motion and Fractional Brownian Motion

Abstract

As more applied science researchers are attempting to use Stochastic Differential Equations (SDEs) in their modeling, especially when involving Fractional Brownian Motion (fBM), one common issue appears: an exact solution cannot always be found. Therefore, in this paper, we test various Numerical methods in solving SDEs with standard BM that have non-linear coefficients. In addition we test our method for SDEs with fBM

5.1 Introduction

Stochastic Differential Equations (SDEs) involving both Brownian Motion (BM) or fractional Brownian Motion (fBM) have been becoming more prevalent in applied mathematics and modeling of various systems. Some examples of these areas, and not limited to them, are finance (i.e Black-Scholes formula), networks (i.e. data transfer in wireless communications), biology (i.e. arrhythmia, brain signaling after a stroke) etc. In many of those cases, years of research and collection of empirical data is performed in order to build an appropriate model. More often than not though, the SDE that best fits the data is an SDE that does not have a simple analytical solution. Therefore the need appears for numerical methods.

In section 5.2 we cover some brief preliminaries about BM, fBM and SDEs that are essential for the numerical approximations we intent to use. In section 5.3 we will state the three different methods tested for numerical solutions of SDEs involving BM, present the results of

the three methods and identify the best. Once we derive the best method, we extend it to SDEs involving fBM. In section 5.4, we state our conclusions.

5.2 Preliminaries

What is a Brownian Motion (BM)? The honor for the discovery of the BM belongs to the Scottish botanist Robert Brown that originally described it in 1828 (Brown, 1828) as he observed it in the movement of pollen particles floating in liquid. The first one to actually construct the process was the Missourian mathematician Norbert Wiener in 1923. Ergo the process itself is also referred to as Wiener Process.

Definition 1 The process $(B(t), t \geq 0)$ is a Brownian Motion (BM) if it is a process of independent Gaussian increments with zero first moment, i.e. a standard Brownian Motion over $[0, T]$ is a random variable $B(t)$ that depends continuously on $t \in [0, T]$ and satisfies (Higham, 2001):

$B(0) = 0$ with probability 1. (5.1)

For $0 \leq s < t \leq T$, the random variable given by the increment $B(t) - B(s)$ is $N(0, t - s)$. (5.2)

For $0 \leq s < t < u < v \leq T$, the increments $B(v) - B(u)$ and $B(t) - B(s)$ are independent. (5.3)

Some basic properties that are easily attained by the definition above are:

$E[B(t)] = 0$, from (5.2) (5.4)

$E[B(t)^2] = t$, from (5.1) and (5.2) (5.5)

Also, for $0 \leq s < t \leq T$ we can write:

$$E[B(t).B(s)] = E[B(s)^2 + B(s)(B(t) - B(s)] = s + 0.(E[B(t - s)]) = s,$$

that is for any $s, t \in [0, T]$ we have that:

$$E[B(t).B(s)] = \min(s, t) = s \wedge t \tag{5.6}$$

Furthermore, let $\alpha > 0$ and define $C(t) = B(\alpha t)$. Then $E[C(t)] = E[B(\alpha t)] = \alpha E[B(t)] = 0$ and $E[C(t)]^2 = E[B(\alpha t)^2] = \alpha^2 E[B(t)^2] = \alpha^2 t$

As we are planning to discuss Stochastic Differential Equations with Brownian Motion, we feel the need to also discuss the continuity of the process. To prove continuity we refer to the Kolmogorov theorem as in (Oksendal, 1998):

Theorem 1 (Kolmogorovs Continuity theorem)

Let $X = \{X_t\}_(t \geq 0)$ a process that for all $T > 0$ there exist $\alpha, \beta, D > 0$ such that

$$E[|X_t - X_s|^\alpha] \leq D|t - s|^{(1+\beta)}, \text{ for } 0 \leq s, t \leq T,$$

then there exists a continuous version of X.

A proof of the theorem can be found in (Strook and Varadhan, 1979).

For the Brownian Motion, it can be shown (Onsendal, 1998) that $E[|B_t - B_s|^4 = n(n+2)|t-s|^2$, which by Theorem 1 we have that B_t has a continuous version. In fact, from now we will be referring to that continuous version of B_t.

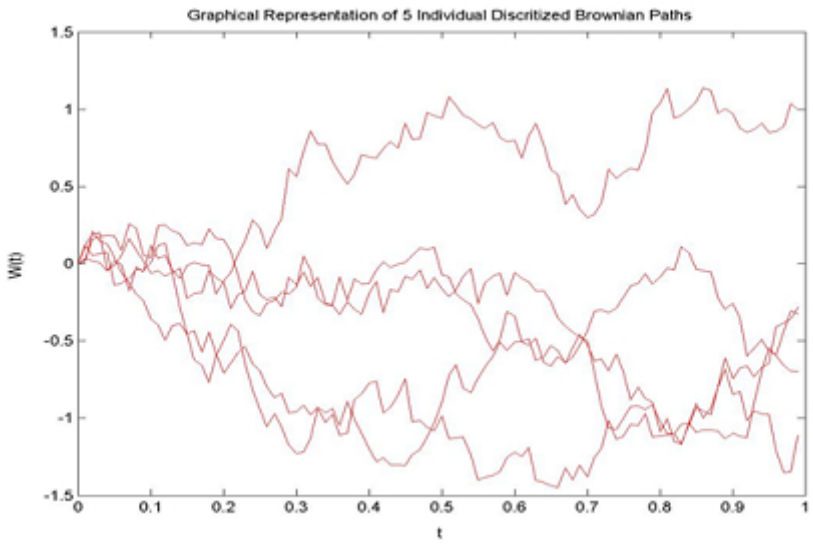


Figure 1. Standard Brownian Motion Paths

As one of the intentions is to investigate numerical approximations of Stochastic Differential Equations, the next natural step is to briefly discuss integration in terms of dB_t. Though there are multiple approaches in various research papers, we are interested in the one shown by (Higham, 2001) as in it is more lined up with numerical approximations. Another side benefit of the approach above is that it provides an interesting connection to Classical Riemann calculus. As such, recall the left end-point Riemann sum representation of the Riemann integral given by

$$\int_0^T f(t)dt = \lim_{\Delta t_0} \sum_0^{N-1} f(t_j)\Delta t, \text{ where } \Delta t = t_{(j+1)} - t_j, \tag{5.7}$$

or by using the midpoint of the partitions we can get

$$\int_0^T f(t)dt = \lim_{\Delta t_0} \sum_{j=0}^{N-1} f((t_{(j+1)} - t_j)/2)\Delta t \tag{5.8}$$

Next we set $f(t) \equiv B(t)$. Therefore we write from (5.7) that

$$\sum_{j=0}^{N-1} B(t_j)[B(t_{(j+1)}) - B(t_j)] = 1/2 \sum_{j=0}^{N-1} \{B(t_{(j+1)})^2 - B(t_j)^2 - [B(t_{(j+1)}) - B(t_j)]^2\} =$$

$$1/2\{B(T)^2 - B(0)^2 - \sum_{j=0}^{N-1}[B(t_{(j+1)}) - B(t_j)]^2\}, \tag{5.9}$$

by using telescoping series. The second term drops off as it is equal to zero. For the third term, we have that:

$$E[\sum_{j=0}^{N-1}(B(t_{(j+1)}) - B(t_j))^2 = \sum_{j=0}^{N-1} E[(B(t_{(j+1)}) - B(t_j))]^2 = \sum_{j=0}^{N-1}(t_{(j+1)} - t_j) = T \tag{5.10}$$

Also, the variance of the third term is of $O(\delta t)$. Therefore by applying limits on both sides of (5.9) we get

$$\int_0^T B(t)sB(t) = (1/2)B(T)^2 - (1/2)T, \tag{5.11}$$

which is the Ito Integral.

By following a similar logic on (5.8) we get

$$\int_0^T B(t)odB(t) = (1/2)B(T)^2, \tag{5.12}$$

which is the Stratonovich Integral.

As explained by Oksendal in (Oksendal, 1998), even though the two integrals look to be different, the choice of which one to be used is really a matter of which properties the user is interested in. The more general and usual choice is usually looking into the Ito Integral is due to the fact that it is not looking into the future, which is a property we care for in Biology. Also Stratonovich is handled better under transformations, especially for SDEs on manifolds. On the other hand, the Ito integrals are martingales, therefore gaining a computational advantage.

As with classical calculus, we could not possibly apply the above approach every time we need to calculate a stochastic integral. The biggest breakthrough in Stochastic Calculus could possibly be due to Kiyoshi Ito.

Lemma 1 (Itos Lemma, (Oksendal, 1998)

Let X_t be an Ito process given by

$$dX_t = udt + vdB_t \tag{5.13}$$

Let $g(t,x) \in C^2([0,\infty) \times \mathbb{R})$. Then $Y_t = g(t, X_t)$ is again an Ito process and

$$dY_t = \frac{\partial g}{\partial t}(t, X_t)dt + \frac{\partial g}{\partial x}(t, X_t)dX_t + \frac{1}{2}\frac{(\partial^2 g)}{\partial x^2}(t, X_t)(dX_t)^2, \tag{5.14}$$

where $(dX_t)^2 = (dX_t)(dX_t)$ is computed according to the rules

$$dt \cdot dt = dt \cdot dB_t = dB_t \cdot dt = 0 \text{ and } dB_t \cdot dB_t = t \tag{5.15}$$

The Ito Lemma, or otherwise known as the Ito formula, is the equivalent of a change of variable formula. One could fairly easily notice from the structure of the formula that it stems from a Taylor series expansion to the second partial derivative in terms of the stochastic process.

As an example, we would like to confirm the result (5.11), i.e. evaluate $\int_0^t B_s dB_s$. Therefore we set $X_t = B_t$ and $g(t,x) = 1/2x^2$. Then $g(t, B_t) = 1/2B_t^2$ and by Itos formula we get

$$d((1/2)B_t^2) = dY_t = \frac{\partial g}{\partial t}dt + \frac{\partial g}{\partial x}dB_t + (1/2)\frac{(\partial^2 g)}{\partial x^2}(dB_t)^2 = B_t dB_t + (1/2)(dB_t)^2 \tag{5.16}$$

$$= B_t dB_t + 1/2dt,$$

which leads to the same answer as (5.11), namely

$$\int_0^t B_s dB_s = (1/2)B_t^2 - (1/2)t \tag{5.17}$$

Preliminaries for fractional Brownian Motion (fBM)

Our investigation will not be limited to the Brownian Motion and to SDEs with BM. We are interested in testing numerical methods to the fractional Brownian motion as well to SDEs with fBM. According to (Lewis, 2005), the process has been defined in 1940 by Kolmogorov in (Kolmogorou, 1940) and its properties, i.e. self similarity and long term dependence, were developed by (Mandelbrot and Van Ness, 1968). Another important contributor was the British hydrologist Harold Edwin Hurst (Hurst, 1951). In his studies on the Nile River, he

observed through 800 years worth of empirical data, that the water levels had a long term dependency and self similarity. To describe that dependency, he estimated a parameter, now known as the Hurst Parameter and denoted by H, based on his data.

Definition 2 We define a Gaussian process $(B^H(t), t \geq 0)$ with continuous sample paths as a standard fractional Brownian Motion (fBM) with Hurst parameter $H \in (0, 1)$ if it satisfies:

$$E[B^H(t)] = 0 \tag{5.18}$$

$$E[B^H(t)B^H(s)] = 1/2(t^{2H} + s^{2H} - |t - s|^{2H}), \tag{5.19}$$

for all $s, t \in \mathbb{R}_+$.

Just by simply looking at expression (5.19), it is obvious that we should consider a trichotomy on the value of the power in the right hand side, more particularly at the value $H = 1/2$:

$$\text{For} H = (1/2), E[B^H(t)B^H(s)] = \min(t, s), \text{ therefore } B^{1/2}(t) \text{ is the standard B.M.} \tag{5.20}$$

$$\text{For } H > 1/2 \text{ the increments are positively correlated} \tag{5.21}$$

$$\text{For } H < 1/2 \text{ the increments are negatively correlated} \tag{5.22}$$

As we mentioned above, two very important properties of fBM are self similarity and long term dependence.

Definition 3 A process $X(t), t \geq 0)$ is said to be self similar with parameter H if for each $a > 0$

$$(X(at), t \geq 0) \rightarrow^L (a^H X(t), t \geq 0)) \tag{5.23}$$

So, for the process $(B^H(at), t \geq 0)$ we can write $E[(B^H(at)B^H(as)] = 1/2\{(at)^{2H} + (as)^{2H} - |at - as|^{2H}\} = a^{2H}(1/2)\{(t^{2H} + s^{2H} - |t - s|^{2H}\}]$ (5.24)

Therefore fBM is a self similar process with parameter H and

$$B^H(t) \sim |a|^H B^H(t) \tag{5.25}$$

Also, regarding the long range dependence property, let $r(n) = E[B^H(1)(B^H(n+1) - B^H(n)]$.

Then for $H \in (1/2, 1)$

$$\sum_{n=1}^{\infty} r(n) = +\infty \tag{5.26}$$

and therefore the process is long range dependent.

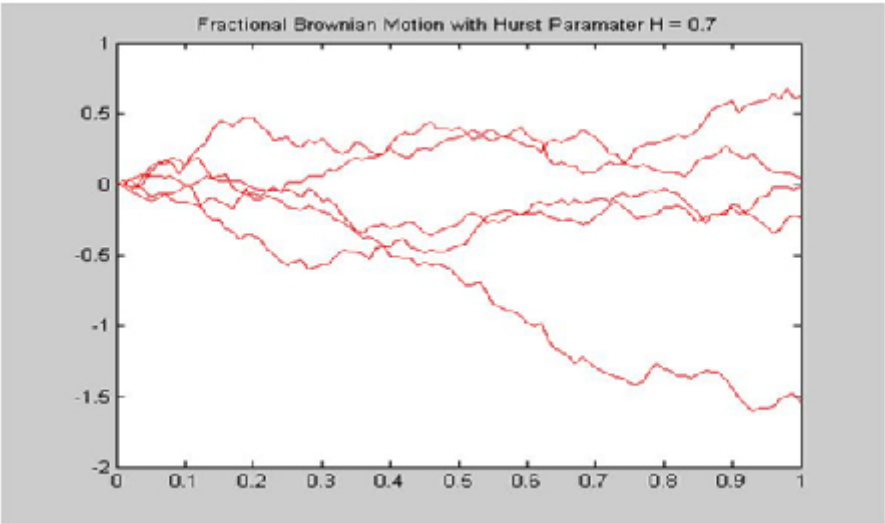


Figure 2. Fractional Brownian Motion Paths with $H = 0.7$

Also, we are interested in the following theorem as a tool for solving SDEs involving fBM:

Theorem 2 if $f : \mathbb{R} \to \mathbb{R}$ is C^2 with derivatives to order two, then

$$f(B^H(T)) - f(B^H(0)) = \int_0^T f'(B^H(s))\delta B^H(s) + H \int_0^T s^{2H-1} f''(B^H(s))ds \text{ a.s.} \tag{5.27}$$

If we let $H = 1/2$ then we have the usual Ito formula.

5.3 Numerical Approximation and Simulations

The main scope of our work Is to develop tolls and methods that can be used to numerically represent Brownian Motion paths, fractional Brownian Motion paths and SDEs with either BM or fBM. The purpose of simulating the first two is so that we can use them as inputs in the SDEs in both cases of actual explicit solutions and numerical approximations. The purpose to simulate SDEs comes as we can approximate numerically their solutions in cases where an explicit solution cannot be found. The programs used for this paper can be found in Appendix A. We will start by defining our error measurement formulas.

Definition 4 (Error formulas)

Let $(X_1, X_2, , X_n)$ be the actual values of X and $(X_1, X_2, ...X_3)$ the numerical approximated values of X at time points $t_1 < t_2 < ... < t_n$. Then

$E_{abs} = \sum_{J=1}^{N} |X_j - \hat{X}_j|$ is the absolute error, (5.28)

$E_{rel} = \sum_{J=1}^{N} |(X_j - \hat{X}_j)/X_j|$ is the relative error, and (5.29)

$E_{ave} = \sum_{J=1}^{N} |(X_j - \hat{X}_j)/N|$ is the average error (5.30)

We use different forms of error measurements so that we are not susceptible to misleading results.

Next we deal with our approach to simulate the different processes. The basic and common principle is to discretize the process as we are using Matlab. Starting with the standard Brownian Motion, we use its properties, i.e. the fact that it is a Gaussian process whose increments follow a normal distribution with mean 0 and variance equal to the time-step. Therefore we use a build-in random number generator that provides us with a $N(0, 1)$ and we scale it by $\sqrt{\Delta t}$, where Δt is the time-step. For our work we considered equidistant partitions, i.e. $\Delta t = T/N$, where T is the stopping time and N is the number of time-steps desired. Also, we usually investigate our processes on $t \in [0, 1]$ in order to reduce as much complexity and cost on the program. As expected, we produce different paths of the Brownian Motion even if we preserve all the constants (Figure 1). Though our original code was successful, the code suggested in (Higham, 2001) by Higham is slimmer and very efficient.

We also employ the properties of the fractional Brownian motion in order to simulate its paths. The following steps are needed (Beran, 1994):

1. Form an $N \times N$ matrix A whose entries are given by (5.19), i.e the covariance of the process.

2. Evaluate the square root of A using the Cholesky decomposition method.

3. Generate a $1 \times N$ vector v whose entries are from a standard Gaussian distribution

4. Apply $\sqrt{A}$ to v.

A sample of five fBM paths with parameter $H = 0.7$ can be seen in Figure 2.

As we now have the tools to simulate both BM and fBM, we proceed to discuss the approximations of SDEs. We start by investigating three methods for Stochastic Differential Equations involving standard Brownian Motion as defined in (Kloeden and Platen, 1992). The best performing method will be applied to Stochastic Differential Equations with fractional Brownian Motion. So, the task is to approximate the stochastic process $X = \{X_t, t_0 \leq t \leq T\}$ satisfying the SDE:

$$dX_t = a(t, X_t)d_t + b(t, X_t)dW_t \text{ on } t_0 \leq t \leq T \text{ and initial value } X(t_0) = X_0 \tag{5.31}$$

For simplicity purposes we set $a(t, X_t) = \alpha X_t$ and $b(t, X_t) = \beta X_t$. So we get

$$dX_t = \alpha X_t d_t + \beta X_t dW_t. \tag{5.32}$$

Applying the Ito formula to (5.32) we have that

$$X_t = X_0 exp\{(\alpha - (1/2)\beta^2)t + \beta W_t\} \tag{5.33}$$

We now introduce the three methods:

Definition 4 Euler Method (Kloeden amd Platen, 1992)

For $t_0 < t_1 < t_2 < ... < t_n = T$ on the interval $[t_0, T]$, the Euler approximation is a continuous time stochastic process $Y = Y_t, t_0 \leq t \leq T$ satisfying the iterative scheme:

$$Y_{n+1} = Y_n + a(Y_n)\Delta t + b(Y_n)\Delta B_t \tag{5.34}$$

More specifically in our case that we wish to apply the method to (5.32), we get:

$$Y_{n+1} = Y_n + \alpha Y_n \Delta t + \beta Y_n \Delta B_t \tag{5.35}$$

Definition 5 Heun Method (Kloeden and Platen, 1992)

For $t_0 < t_1 < t_2 < ... < t_n = T$ on the interval $[t_0, T]$, the Heun method is satisfying the iterative scheme:

$$Y_{n+1} = Y_n + (1/2)a(\hat{Y}_n + Y_n)\Delta t + (1/2b(\hat{Y}_n + Y_n)\Delta B_t \tag{5.36}$$

where

$$\hat{Y}_n = Y_n + a(Y_n)\Delta t + b(Y_n)\Delta B_t \tag{5.37}$$

More specifically in our case that we wish to apply the method to (3.6), we get:

$$Y_{n+1} = Y_n + 1/2\alpha(\hat{Y}_n + Y_n)\Delta t + 1/2\beta(\hat{Y}_n + Y_n)\Delta B_t \tag{5.38}$$

where $\hat{Y}_n = Y_n + \alpha Y_n \Delta t + \beta Y_n \Delta B_t$.

The principle behind the Heun method is very much alike to the Euler one, with the difference that instead of the process being evaluated at the endpoints, the trapezoid rule is being used.

Definition 6 (Milstein Method)

For $t_0 < t_1 < t_2 < ... < t_n = T$ on the interval $[t_0, T]$, the Milstein approximation is a continuous time stochastic process $Y = \{Y_t, t_0 \leq t \leq T\}$ satisfying the iterative scheme:

$$Y_{n+1} = Y_n + a(Y_n)\Delta t + b(Y_n)\Delta B + (1/2)b(Y_n)b'(Y_n)[(\Delta B)^2 - \Delta t] \tag{5.39}$$

More specifically in our case that we wish to apply the method to (5.32), we get:

$$Y_{n+1} = Y_n + \alpha Y_n \Delta t + \beta Y_n \Delta B + (1/2)\beta(Y_n)\beta'(Y_n)[(\Delta B)^2 - \Delta t] \tag{5.40}$$

The Milstein method is in a sense an evolutionary form of the Euler method. The basic difference is that one extra term is included in the method. Another important remark is that the Ito-Taylor expansion is used in order to derive this method, therefore providing

an order 1.0 strong Taylor scheme (Kloeden and Plates, 1998). Next we compare the three methods with the actual solution graphically.

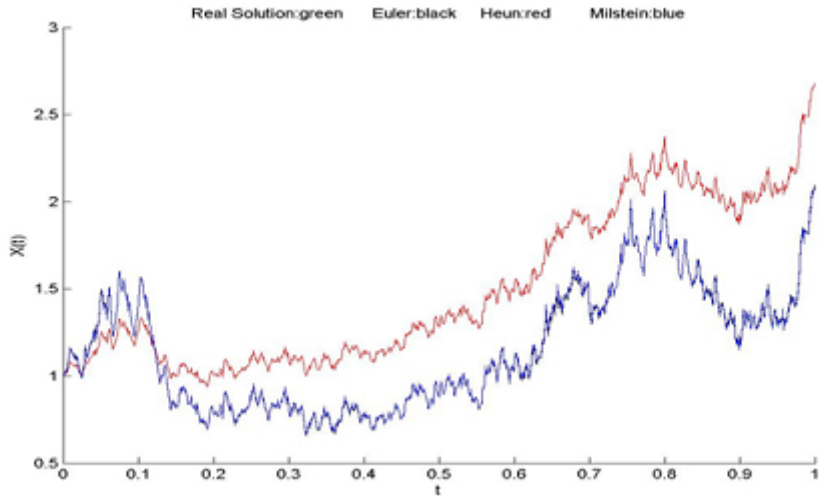


Figure 3 Simulations for $N = 1000$ and $\alpha > (1/2)\beta^2$

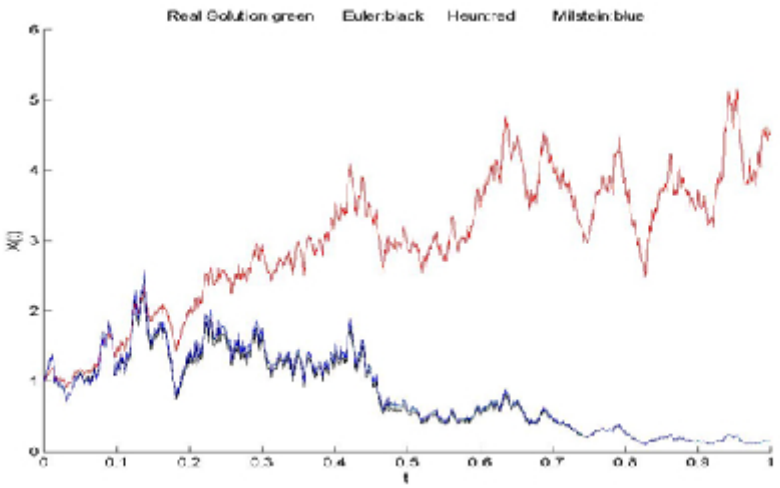


Figure 4. Simulations for $\alpha < (1/2)\beta^2$

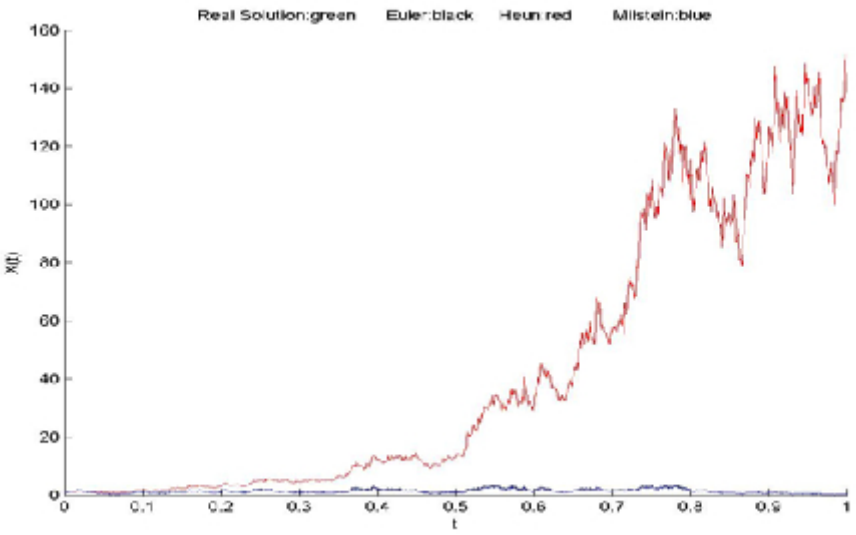


Figure 5. Simulations for $\alpha = (1/2)\beta^2$

Euler	Heun	Milstein
4.8317	372.3806	0.6604
12.0207	278.0391	0.9094
26.0531	679.4561	1.8465
10.3954	179.0899	0.5445
13.6615	321.4044	1.0420
15.8134	321.8179	1.2279
9.2588	404.3833	1.0448
22.5554	577.9528	3.8917
40.3136	798.9531	3.4242
13.9642	378.7190	0.7851

Table 1. Table of Absolute Errors

Euler	Heun	Milstein
3.4319	20.5666	0.6359
6.1103	23.7858	1.1628
1.3683	28.6385	0.8478
1.4295	74.8322	1.5371
2.8765	11.2171	0.4262
1.4639	14.8793	0.6616
3.2527	10.0544	0.4610
5.2890	23.4771	0.5525
4.9824	29.2045	1.6498
5.3197	16.8437	0.6747

Table 2. Table of Relative Errors

As shown by graphs 3-5 we get the idea that the Heun method is not appropriate for SDEs whatsoever. In fact, the scheme seems to diverge once BM is involved. Therefore it is abandoned for our purposes. In comparing the two remaining methods, even though both seem to follow the actual solution, the Milstein scheme seems to have a much smaller deviation from the actual solution (Tables 1 and 2). The result is not surprising as both Euler and Milstein can be derived by applying the Taylor polynomial expansion to the SDE, with the difference that the Milstein methods scheme is of higher order. The one main concern usually with higher order schemes, is the how computationally expensive it can be. Truth is though, that even a standard home computer can easily run the programs in matter of seconds. As such, we further test the Milstein scheme against the actual solutions of two more non-linear SDEs, namely:

$$dX_t = \alpha^2 X_t(1 + X_t^2)dt - \alpha(1 + X_t^2)dW_t, \tag{5.41}$$

that has as an explicit solution

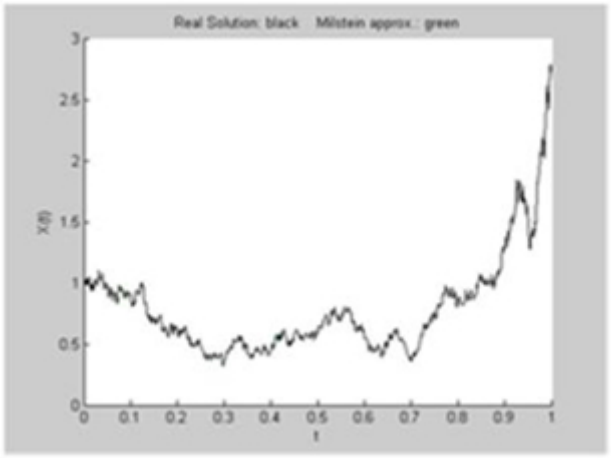
$$X_t = \cot(\alpha W_t + \operatorname{arccot}(X_0)). \tag{5.42}$$

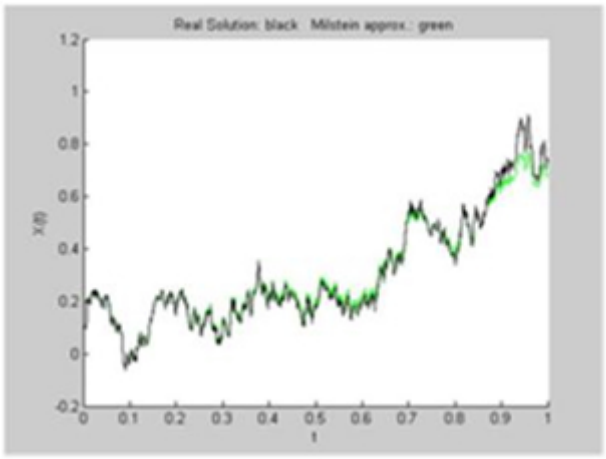
lso we test the SDE

$$dX_t = -1/2\alpha^2 X_t dt - \alpha\sqrt{1 - X_t^2}dW_t, \tag{5.43}$$

whose solution is

$$X_t = \cos(\alpha W_t + \arccos(X_0)) \tag{5.44}$$





Figures 6 and 7 Simulations of SDEs using the Milstein Method

Trial #	Average Error
1	0.0009
2	0.0019
3	0.0013
4	0.0002
5	0.0001
6	0.0006
7	0.0004

Trial #	Average Error
1	0.0238
2	0.0090
3	0.0098
4	0.0201
5	0.0316
6	0.0055
7	0.0060

Table 3 and 4 Errors of SDEs using the Milstein Method

Our next step is to extend our results to provide a method that works in SDEs with fBM. As with the Milstein method for SDEs involving Brownian Motion, we apply the Taylor polynomial to the general form of SDE with fBM. Our result and suggested method is given by:

$$Y_{n+1} = Y_n + a(Y_n)\Delta t + b(Y_n)(B(t_{(n+1)})^H - B(t_n)^H) + (1/2)b(Y_n)b'(Y_n)[(B(t_{(n+1)})^H - B(t_n)^H)^2 - (t_{(n+1)}^{2H} - t_n^{2H})] \tag{5.45}$$

One remark for our method is that if we set $H = 1/2$ we get expression (3.13) which is the Milstein method for SDEs involving standard Brownian Motion.

Proof:

The Milstein Scheme for standard Brownian motion can be produced by adding the term $bb'I((1,1))$ to the Euler method. In similar approach we have

$$Y_{n+1} = Y_n + a\Delta t + b(Y_n) + b(Y_n)b'(Y_n)I((1,1)) \tag{5.46}$$

Evaluating the last term we have:

$$I_{(1,1)} = \int_{t_n}^{t_{n+1}} \int_0^{s2} dB_{s1}^H dB_{s2}^H = \int_{t_n}^{t_{n+1}} B_{s2}^H dB_{s2}^H = [(B_{t_{n+1}}^H - B_{t_n}^H)^2 - (t_{n+1}^{2H} - t_n^{2H})] \tag{5.47}$$

Substituting back in (3.20) we get

$$Y_{n+1} = Y_n + a(Y_n)\Delta t + b(Y_n)(B_{t_{n+1}}^H - B_{t_n}^H) + 1/2b(Y_n)b'(Y_n)[(B_{t_{n+1}}^H - B_{t_n}^H)^2 - (t_{n+1}^{2H} - t_n^{2H})] \tag{5.48}$$

For the numerical simulation, we consider the SDE

$$dX_t = AX_t dt + CX_t dB_t^H \text{ with } X_0 = 1. \tag{5.49}$$

Its solution is given by:

$$X_t = exp\{At - (1/2)C^2 t^{2H} + CB_t^H\} \tag{5.50}$$

Next we run a comparison of the extended Milstein scheme to the actual solution of the SDE with $H = 0.7$. The outcome is very encouraging.

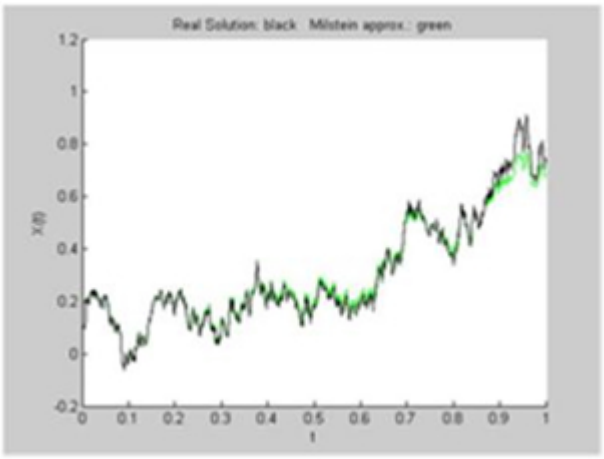


Figure 7. SDE with fBM using the extended Milstein Method - Milstein: Green, Actual: Black

Trial #	Average error
1	0.0010
2	0.0037
3	0.0074
4	0.0031
5	0.0014
6	0.0046
7	0.0041

Table 5. Average Errors

In a head to head comparison with the method suggested in [6], we resulted in an absolute error of zero. After further investigation it seems that the two schemes are in fact the same scheme. The main difference is that the suggested method in this paper is a much simpler expression and not dependent on summations of triple integrals.

5.4 Conclusions

We believe that our methods for simulating Brownian Motion and fractional Brownian Motion are adequate due to the fact that they are derived directly from the properties of the processes. Regarding SDEs with Brownian Motion, we reject the Heun method and choose to either use either Euler or Milstein method. The Milstein method is somewhat closer to the exact solution, but the Euler method might be more appropriate for finer partitions on t. Finally we suggest that for SDEs involving fBM, the extended Milstein method should be used.

References

[1] R. Brown, A brief account of microscopical observations made in the months of June, July and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies." Phil. Mag. 4, 161173, 1828.

[2] D.J. HIGHAM, An algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations, SIAM Review, September 2001, Vol. 43, Number 3, pp. 525-546

[3] B.Oksendal, Stochastic Differential Equations, An Introduction with Applications, Fifth Edition, Springer, 1998.

[4] D.W. Strook & S.R.S Varadhan, Multidimensional Diffusion Processes, Springer-Verlag, 1979, p51.

[5] P. Kloeden AND E. Platen, Numerical Solution of Stochastic Differential Equations, Springer, 1992

[6] I. Lewis, One Stage Approximation of SDEs with Noise Modeled as fractional Brownian Motion, Honors Undergraduate Thesis supervised by B. Pasik-Duncan, 2005

[7] A.N. Kolmogorov, Wienersche Spiralan and einige andere interessante Kurven im Hilbertschen Raum, C.R. (doklady) Acad. Sci. Urss (N.S.), 26, 1940, pp. 115-118

[8] B.B. Mandelbrot and J.W. Van Ness, Fractional Brownian motion, fractional noises and applications, SIAm Rev., 10, 1968, pp. 422-437

[9] H.E. Hurst, Long Term Storage Capacity of Reservoirs, Transactions of the American Society of Civil Engineers, 1951, 116, 770-799

[10] J. Beran (1994) Statistics for Long-Memory Processes, Chapman & Hall

6 Appendix - MATLAB Programs

```
Appendix A (MATLAB Programs)
% Investigation of Milstein Method as a Numerical approximation
% for SPDEs with fBM.
% Author: Yiannis Zachariou
% Date: 7/04/2009
% W is the standard BM
% B is the fBM

% Fractional Brownian Motion generator
% Method: Using the Cholesky decomposition

clear;
H = input('Hurst parameter?');

% Generating Brownian Motion and Independent Increments
m=5
k=100
N = m*k;
dt = 1/k;
randn('state',sum(100*clock));
s = randn(N,1);
dW = zeros(1,N);
W = zeros(1,N);
T = zeros(1,N);
dB = zeros(1,N);

dW(1) = sqrt(dt)*randn;
W(1) = 0;
T(1) = dt;

for j = 1:N-1
   dW(j+1) = sqrt(dt)*randn;
   W(j+1) = W(j) + dW(j+1);
    T(j+1) = (j+1)*dt;
end

R = zeros(N,N);
for i=1:N
    for j=1:N
        R(i,j) = (T(i)^(2*H) + T(j)^(2*H) - (abs(T(i)-T(j)))^(2*H))/2;
    end
end

G = chol(R);
A = G'*s;
```

```
B = [0 A'];
t = [0 T];

% Part 2
% SPDEs

Y = zeros(1,N); % Analytical Solution
X = zeros(1,N); % Milstein Approximation
Z = zeros(1,N); % Ian Lewis Approximation
    L = zeros(1,N);
    I = zeros(1,N);
% Set initial Conditions
Y(1) = 1;
X(1) = 1;
L(1) = 0;
I(1) = 0;
Z(1) = 1;

E(1) = 0;
dB(1) = B(1);
% Set coefficients of SPDE with fBM
a=2;
c=3;

for j = 2:N
   dW(j) = sqrt(dt)*randn;
   T(j) = (j-1)*dt;
   W(j) = W(j-1) + dW(j-1);
   dB(j) = B(j)-B(j-1);

 Y(j) = exp(a*T(j)-0.5*c^2*T(j)^(2*H)+c*B(j-1));
 X(j) = X(j-1)+X(j-1)*(a)*dt+c*X(j-1)*dB(j-1)+(1/2)*c^2*X(j-1)*(dB(j-1)^2-(T(j)^(2*H)-T(j-
    L(j) = a*Z(j-1)*dt + c*Z(j-1)*dB(j-1)+0.5*c^2*Z(j-1)*(dB(j-1)^2-(T(j)^(2*H)-T(j-1)^(2*
    I = cumsum(L);
 Z(j) = X(1)+ I(j);

   %error evaluation
   E(j) = ((Y(j)-X(j))^2)/N;
   F(j) = ((X(j)-Z(j))^2)/N;
end

hold on

p = num2str(H);
q = num2str(dt);

plot(T,Y,'r',T,X,'g')
```

```
%plot(T,X,'g')
xlabel('t')
ylabel('X(t)')
legend('Actual','Milstein')
%plot(T,Z,'k')

title(['Actual Solution vs. Approximation with H=',p, ' and N=',q])

hold off

M = sum(E)
K = sum(F)
hold off

% Numerical Approximation of intensity a
clear;
randn('state',sum(100*clock))          % set the state of randn

j=100000;
s = 10; dt = 1/s;
N=s*j
%N = 10000;
%dt = 1/N;

X = zeros(1,N+1); % Milstein Approximation
A = zeros(1,N+1);
K = zeros(1,N+1);
L = zeros(1,N+1);
M = zeros(1,N+1);
z= zeros (1,N+1);
Q = zeros(1,N+1);
P = zeros(1,N+1);
V = zeros(1,N+1);
U = zeros(1,N+1);
% Set the parameter for control
e=10;


d=2;
% Set the parameter for BP
f=3
c=1;

X(1) = 0;

dW = zeros(1,N+1);
W = zeros(1,N+1);
```

```
dB = zeros(1,N+1);
B = zeros(1,N+1);
T = zeros(1,N+1);
T(1)=0;

B(1) = 0;
dB(1) = sqrt(dt)*randn;

z(1) =1;

dB(1) = sqrt(dt)*randn;

K(1) = 1;  %  a hat

A(1) = z(1)/K(1);
%K(2)=1;
L(1) = A(1);
M(1)=L(1);
U(1)=f;

Q(1) = ((X(1)-f*M(1))^2)*dt/M(1);
P(1) = Q(1);
V(1)= 0;
for i = 2:N+1
   dB(i) = sqrt(dt)*randn;
   T(i) = (i-1)*dt;
   B(i) = B(i-1) + dB(i-1);
   %z(i) = -1/(i*dt);
   z(i) = exp(-i*dt)+1;
   X(i) = X(i-1) + (f*A(i-1)-(c^2)/4)*dt+c*sqrt(A(i-1))*dB(i-1)+(c^2)*((dB(i-1))^2)/4;
 A(i) = z(i)/K(i-1);
 if A(i)<e
      A(i)=e;
  else A(i)=z(i)/K(i-1);
  end

  L(i)= A(i)*dt;
 M(i) =L(i)+M(i-1);  % C(s)
  K(i) = (X(i)-X(1))/M(i);  % a hat
 L(i) = K(i)-f;
U(i)=f;
end

p=K(N);
%V(N)
q = num2str(f);
s = num2str(K(N));
```

```
d = num2str(N);
hold on
plot(T,K,'g',T,U,'--')
xlabel('t')
ylabel('a(t)')
legend('Estimator','Actual')
title(['Estimator for a=',q, '  a hat= ',s,' and N=',d])
%xlabel('t')
%ylabel('U(t)')
 hold off
%title(['Real Solution:green      Euler:black     Heun:red        Milstein:blue'])
```