Available online at www.sciencedirect.com
SCIENCE DIRECT®

ELSEVIER

Applied Mathematical Modelling 30 (2006) 816–823

APPLIED MATHEMATICAL MODELLING

www.elsevier.com/locate/apm

On high-order discrete derivatives of stochastic variables

N. Moriya *

siOnet Ltd., Applied Modeling Research, Herzelia 46445, Israel

Received 1 March 2003; accepted 27 June 2005
Available online 19 August 2005

Abstract

We derive an explicit expression for the probability density function of the *m*th numerical derivative of a stochastic variable. It is shown that the proposed statistics can analytically be obtained based on the original probability characteristics of the observed signal in a simple manner. We argue that this allows estimating the statistical parameters of the original distribution and further, to simulate the noise contribution in the original stochastic process so that the noise component is statistically indistinguishable from the true contribution of the noise in the originally observed data signal.

1. Introduction

The importance of the fundamental properties of stochastic systems is recently acknowledged by a growing portion of the scientific community. This is mostly due to the intensive use of stochastic models considered to account for numerous applications related to Audio/Image signal processing [1], chaotic quantum processes [2], molecular dynamic calculations, atomic diffusion models [3], economy, statistical physics, communication systems and non-linear, optimal filtering methods, commonly used in different fields of engineering [4], to name just a few. Among other aspects, the careful treatment of noise-induced, non-equilibrium phenomena, represented by non-linear time-series signals and the prediction of the noise-level characteristics in most physical

* Tel.: +972 9 956 9893; fax: +972 9 957 5336.
E-mail address: netzer@si-o-net.com

0307-904X/$ - see front matter
doi:10.1016/j.apm.2005.06.008

stochastic systems, is of great relevance [5]. It is of particular importance in dynamical processes where non-linear behavior is expected and may seriously alter any estimation of the states of the system, if not to cause a total divergence of the model parameters. Such conditions are common in non-linear systems when modeled by recursive or adaptive methods such as Weiner or Kalman filtering [6].

Although the dynamical properties of various systems are extensively considered by their stochastic differential-equation representation, not much is known about the probability transfer propagation of the derivative operator of random variables. A fundamental property of stochastic processes, where a distributional derivative is involved, is the well-known definition of white noise, interpreted as the first derivative of Brownian motion (Weiner process). In the classical approach, which roots back to the pioneering work of Weiner [7], the derivative of a process identified as a Weiner process $W(\zeta_t)$, with ζ_t, the stochastic variable of the process, can be shown to have an expectation value given by $E\langle\zeta_t\rangle = 0$ and $\mathrm{cov}[\zeta_s, \zeta_t] = \delta(s, t)$ with a constant valued spectral density. An excellent review on stochastic analysis with a particular consideration for stochastic derivatives and their properties, may be found in [8]. Some other theoretical aspects of stochastic derivative in a standard L_2-space have recently been considered for the purpose of an alternative representation of the integrator for the general Levy processes [9] or the generalization of supersymmetric path integral representation to stochastic processes with high fermion interaction fields [10]. Aside from the general theoretical interest in stochastic derivatives of random processes, many experimental methods also use the numerical derivative of noise-containing data routinely, as for instance in synthetic aperture radar (SAR) interferometry [11] and other spectroscopic analysis methods [12]. Additional developments in the use of probability density function analysis relevant to signal processing applications may be found in [13].

The aim of this paper is to derive a general approach to obtain the probability density function (pdf) of the mth derivative of a stochastic variable. Given the explicit probability density function of the original distribution of an observation, we claim that a simple expression can be obtained to describe the probability density function of the derivative of order m and show that the result is a the sum of $m+1$ independent, non-identical, random variables each of which is a variant of the original density function.

2. Stochastic derivative matrix

We first define a stochastic process and let $(\Omega, \wp)$ be a measurable space with $\wp$, a collection of events n_i in Ω and (R, Γ) another measurable space (state-space) consisting of the real structure R and the σ-field of Borel sets Γ. Let the probability measure P be defined on $(\Omega, \wp)$ and assume $\xi(n_i)$ is a P-measurable function on σ-algebra of Γ for all P-measurable sets. The above definitions also ensure that the variance of the stochastic variables considered here are finite.

As remarked above, we differentiate the signal vector with respect to the index of the signal data points in their sequenced order (or equivalently, treating the signal as a time-series vector with a unit time step). By doing this, one may realize that a differentiation procedure, of the first order, is equivalent to subtracting the element n_i from the element n_{i+1} in the stochastic signal. Since in such random set of points each point is totally independent of all other points and controlled to any data point in the set only by the mutual statistics belongs to the sample space, denoted by Ω (i.e. all points (i,j) are uncorrelated where $i \neq j$), the equivalence to subtracting the element n_i from the element n_{i+1} in the noise signal would be the equivalent of the subtraction of two independent Random Variables with identical statistical distribution (IID).

In contrast with the case of the first derivative, where one could assume that all individual data points were uncorrelated, higher order derivatives involve correlated expressions that lead, in the general case, to non-trivial expressions for the resultant probability functions.

Considering the above definitions and referring to some arbitrary random variable function $V(n_i, \xi)$, referred here as the original data signal with ξ as the stochastic random variable, one can now derive the second-order derivative index series, $V^{(2)}(n_i^{(2)}, \xi^{(2)})$, with $\xi^{(2)}$ refers to the (yet) unknown stochastic random variable corresponding to the second-order derivative vector by realizing that $n_i^{(1)} = n_i - n_{i-1}$ and $n_{i+1}^{(1)} = n_{i+1} - n_i$ so that $n_i^{(2)} = n_{i+1}^{(1)} - n_i^{(1)} = n_{i+1} - 2n_i + n_{i-1}$. These expressions imply that the probability density function of the second order derivative is the equivalent pdf of the sum of three independent, however non-identical, random variables (InID), all with similar, however not identical, probability density functions. Referring now to the general result that given two independent random variables ξ_1 and ξ_2 on the space R^k, with μ and v their respective distribution functions and f and g denote their respective density functions, than the distribution of the sum $\xi_1 + \xi_2$ is the convolution $\mu * v$ and the analogue density function of the sum equals to the convolution integral denoted by $f * g$.

Using the notation $f_{\xi_1,\xi_2,\xi_3} = f_{\xi_1} * f_{\xi_2} * f_{\xi_3}$ this implies

$$f_{n_i^{(2)}} = \left(f_{n_i} * f_{(-2n_i)}\right) * f_{n_i}. \tag{1}$$

Following the above arguments, for higher derivatives, it can now easily be deduced that the mth derivative of a random variable derived from an arbitrary statistically defined variable can be obtained by noting that the correlation elements that dictate the derivative expressions are given by the matrix

$$\begin{array}{c}
1 \\
1 \quad -1 \\
1 \quad -2 \quad 1 \\
1 \quad -3 \quad 3 \quad -1 \\
1 \quad -4 \quad 6 \quad -4 \quad 1 \\
1 \quad -5 \quad 10 \quad -10 \quad 5 \quad -1 \\
1 \quad -6 \quad 15 \quad -20 \quad 15 \quad -6 \quad 1 \\
1 \quad -7 \quad 21 \quad -35 \quad 35 \quad -21 \quad 7 \quad -1 \\
1 \quad -8 \quad 28 \quad -56 \quad 70 \quad -56 \quad 28 \quad -8 \quad 1 \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots
\end{array}$$

and are governed by the following expression, denoted here as the Stochastic-Derivative Matrix S_k^m, and given by $S_j^{(m)} = (-1)^{j+1}\binom{m}{j}$, where $\binom{m}{j}$ the elements the binomial coefficients [14].

In terms of a summation of the individual elements needed to account for the probability density function of the mth order numerical derivative, one may write the summation as

$$f_m = \sum_{j=1}^{m} S_j^{(m)} f(z) \tag{2}$$

with $f(z)$ representing the probability density function of the original random variable.

3. The general case

Generalizing now the above, for a set of random variables $\xi_1, \xi_2, \ldots, \xi_m$ and a function $z = g(\xi_1, \xi_2, \ldots, \xi_m)$, one can form a new random variable

$$\xi_z = g(\xi_1, \xi_2, \ldots, \xi_m). \tag{3}$$

In particular, the density and distribution functions of ξ_z, in terms of the density and distribution functions of $\xi_1, \xi_2, \ldots, \xi_m$ can easily be obtained.

To do so one denotes $D_z = \{(\xi_1, \xi_2, \ldots, \xi_m) : g(\xi_1, \xi_2, \ldots, \xi_m) \leqslant z\}$ noting that $(\xi_z \leqslant z) = \{g(\xi_1, \xi_2, \ldots, \xi_m) \leqslant \ z\} = \{(\xi_1, \xi_2, \ldots, \xi_m) \in D_z\}$ so that

$$F_Z(z) = P(Z \leqslant z) = P((\xi_1, \xi_2, \ldots, \xi_m) \in D_z), \tag{4}$$

which gives

$$F_Z(z) = \int_{D_z} \int f_{\xi_1 \xi_2, \ldots, \xi_m}(\xi_1, \xi_2, \ldots, \xi_m) \mathrm{d}\xi_1 \mathrm{d}\xi_2, \ldots, \mathrm{d}\xi_m. \tag{5}$$

Thus, in order to find the distribution probability function of the new random variable ξ_z, given the distribution functions's of the random variables ξ_j's, one needs to define the range of the validity of the new variable z and to evaluate the integral using the mutual density function.

For the case of independent random variables the above expression simplifies with the integrand replaced by $\prod_{j=1}^{m} f_{\xi_j}$.

Finally we obtain

$$F_Z(z) = \int_{D_z} \int \prod_{j=1}^{m} f_{\xi_j} \mathrm{d}\xi_1 \mathrm{d}\xi_2, \ldots, \mathrm{d}\xi_N = \int_{D_z} \int \left[\prod_{j=1}^{m} S_j^m f_j(\xi) \right] \mathrm{d}\xi^{(\mathrm{m})}. \tag{6}$$

Since the density function is the same for all individual elements of the multiplication term under the integral, this can symbolically be written as

$$F_{n_i}^{(m)} = \int \int_{D_z} \left\{ \prod_{j=1}^{m} S_j^m f(\xi) \right\} \mathrm{d}\xi^{(\mathrm{m})}, \tag{7}$$

where $F_{n_i}^{(m)}$ represents the probability distribution function of $\frac{\partial^{\mathrm{m}} V(n_i, \xi)}{\partial i^{\mathrm{m}}}$ that can easily be evaluated to derive the respective density function, recalling that the term $\{\circ\}$ really represents a convolution of the original probability function weighted accordingly.

The above expressions constitute the main result of this contribution.

It is worthwhile to note that since the above expression inherently holds symmetry with respect to permutation, Eq. (9) in Ref. [15] may be applied here with only minor modifications [15].

It should be clear that the above derivation, although mathematically simple, is general and is not restricted to any particular probability density function of the observation.

However, it is of practical interest to focus the attention on stochastic processes characterized by some of the more important probability distribution functions such as the normal, Poisson and the exponential (Pearson Type X) distributed probability functions. As will be demonstrated in the following, the first two yields relatively straightforward analytical expression as belong to the few probability functions that convolve into similar functions while the third (as well as the uniform distribution) results in somewhat less intuitive expressions, yet simply derived results. We start with the normal distribution where ξ is referred to as the random variable, $N(0, \sigma_0^2)$, i.e. a Gaussian distribution with the first moment equals zero and variance given by σ_0^2 as an illustrative probability. It should be noted that the derivation of the following with mean values other than zero is straightforward. Also, the generalization of the following treatment to other Probability Density Function statistics (including non-analytical distributions) is almost trivial and can be easily performed.

For the above we argue that an expression of the form

$$\frac{\mathrm{d}^m N(0, \sigma_0^2)}{\mathrm{d}i^m} = \alpha(m) N(0, \beta(m)\sigma_0^2) \tag{8}$$

is explicitly describing the resultant statistics with $\beta(m)$ given by the sum of the squares of the elements of the $m + 1$'s row in the Stochastic-Derivative matrix and with $\alpha(m)$ given by the inverse of square-root of the sum of the squares of the elements of the $m + 1$'s row of the Stochastic-Derivative matrix. We leave the derivation of the above expression unproved as being straightforward. Note that for a normal distribution function, as used above, the condition $\alpha \propto 1/\sqrt{\beta}$ is required by the normalization condition.

As an illustrative example we use the above to derive the probability density function of a zero mean normal distribution for the case of the first, second and fifth derivatives.

Using Eq. (8) and the arguments above one can easily obtain

$$\begin{aligned}
\frac{\mathrm{d}N(a, \sigma_0^2)}{\mathrm{d}i} &= \frac{1}{\sqrt{2}} N\left(0, \left(\sqrt{2}\sigma_0\right)^2\right), \\
\frac{\mathrm{d}^2 N(0, \sigma_0^2)}{\mathrm{d}i^2} &= \frac{1}{\sqrt{2^2 + 1^2 + 1^2}} N\left(0, \left[\left(\sqrt{2^2 + 1^2 + 1^2}\right)\sigma_0\right]^2\right) = \frac{1}{\sqrt{6}} N\left(0, \left(\sqrt{6}\sigma_0\right)^2\right), \\
\frac{\mathrm{d}^5 N(0, \sigma_0^2)}{\mathrm{d}i^5} &= \frac{1}{\sqrt{252}} N\left(0, \left(\sqrt{252}\sigma_0\right)^2\right).
\end{aligned} \tag{9}$$

This was indeed verified by numerical simulations where a normal distributed random set of 200 K elements was generated (Figs. 1 and 2) and the difference vectors were obtained. The histograms of the resultant vectors were then taken and are shown to have Gaussian shapes with variance values compatible with the above results. Similar expression can be derived also for the Poisson density function having the convenient characteristics of convolving into a Poissonic distribution with the new decaying parameter, the sum of the original parameters.

It is interesting to note that in contrast with the Gaussian distribution density function (and the Poissonic distribution as discussed above), the general case of arbitrary distribution need not necessarily result a density function similar to the original function after the differentiation of

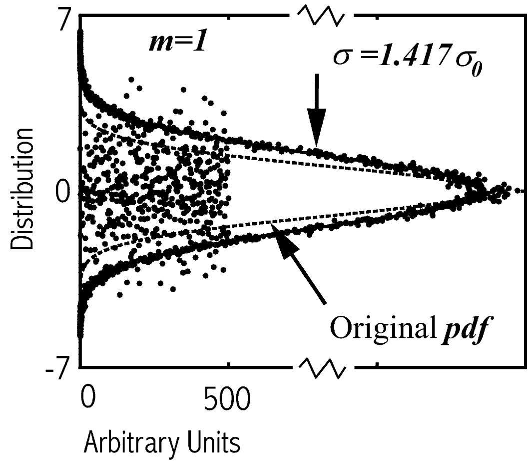

Fig. 1. A simulation of the first derivative ($m = 1$) of a 200 K normal distributed, $N(0, \sigma_0^2 = 1)$, noise signal (only partially shown) with the corresponding deduced histogram (fitted to a Gaussian and arbitrarily scaled). Also shown is the original density function used to generate the noise signal. As expected $\sigma = \sqrt{2}\sigma_0$.

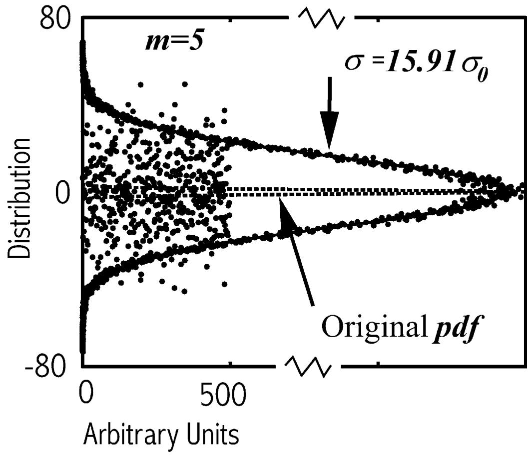

Fig. 2. A simulation of the fifth derivative ($m = 5$) of a normal distributed, noise signal (see caption of Fig. 1). As expected $\sigma = \sqrt{252}\sigma_0$.

order m. This may simply be concluded if one experiments with a uniform probability density function which convolves to obtain a triangular shaped probability function even for $m = 1$. However, recalling that the mth order derivative is in fact a series of convolving terms, each weighted with correlation to the elements of the Stochastic-Derivative matrix, the derivation of the respective expression may be straightforward. For instance, the probability density function of the first derivative of the Exponential pdf (i.e. Pearson Type X) $f(x) = \frac{1}{\tau}e^{-x/\tau}$ for $x \geqslant 0$ is given by $f'(z) = \frac{1}{\tau^2}ze^{-z/\tau}$ for $z \geqslant 0$, respectively, i.e. a non-exponential pdf (a variant of the Pearson Type III density function).

To demonstrate one of the proposed motivations for the use of high-order numerical derivative of a stochastic signal, we mention the derivation of the noise-level of experimental output where noise, either due to experimental set-up or to the process itself (or from both sources), convolves with the signal and is screening the, usually required, pure data signal. A relevant study, related to the above may be pointed out [5]. However, in the referred work, the authors derive an estimate approach to separate the measurement noise from the model uncertainty, most suitable to be implemented in adaptive or recursive filtering-like procedures assuming known (normal) probability density. In this paper we suggest, that the analytical expressions presented here may be used to test the validity of the normality assumption as well as to set upper (lower) limits to the Noise-Level in such procedures such as, for instance, Kalman or Weiner filtering.

For simplicity we thus assume that the arbitrary noisy signal $P = S + N$, with N being the noise that is added to the pure signal S, can be represented by an arbitrary smooth and continuous signal contaminated by noise. Let us further assume that on the interval of validity of S, one can approximate S (for instance, in the least mean square sense) by an m-degree polynomial function. This can be proved to be possible for any bounded, smoothed and continuous function S [16], but may be of practical use only when the interval is not too long, as compared to the structure of the signal, and for a relatively low polynomial degree.

Assuming the above, it turns that $\frac{\mathrm{d}^{m+1}P}{\mathrm{d}k^{m+1}} = \frac{\mathrm{d}^{m+1}N}{\mathrm{d}k^{m+1}}$, as the mth derivative of S, under the above assumptions, is constant. For most experimental data m would not exceed 5. However the present approach holds for arbitrarily higher order [17].

Now, if the characteristics of the statistical properties of the high-order derivative of the original noise $(\partial^{m+1}/\partial i^{m+1})N$ is known, i.e. the probability density function that statistically describes the initial noise, subject to high-order numerical derivative, in terms of the parameters (assumed to be unknown) of the statistical nature of the noise (assumed to be known), one can obtain the specific parameters of the original noise and thus to deduce the noise-level of the original signal P. Additionally, when any prior knowledge, with respect to the original probability function of the noise is not at hand, one may, based on the above approach, conduct a simple test, to deduce the specific probability density function of the original stochastic process in the observed data signal by reproducing only the noise component from the noisy signal and further process it so that no unintentional influences, related to the pure signal rather than to the noise, are affecting the probability testing procedure.

4. Conclusions

In conclusion, we have derived an explicit expression for the probability density function of the mth numerical derivative of a stochastic variable. It was shown that the proposed statistics can analytically be obtained based on the original probability characteristics of the observed signal for a normal distribution or easily be derived for the general case as a sum of independent, though non-identical, random variables, based on a simple weighting procedure of the original probability density function. We suggest that this allows estimating the statistical parameters of the original distribution parameters and further, to simulated the noise contribution in the stochastic system so that the noise component reconstructed is statistically indistinguishable from the true contribution of the noise in the originally observed data signal.

References

[1] B. Aiazzi et al., IEEE Signal Processing Lett. 6 (1999) 138; R. Chandramouli et al., IEEE Signal Processing Lett. 6 (1999) 129.
[2] Z.P. Karkuszewski, C. Jarzynski, W.H. Zurek, Phys. Rev. Lett. 89 (2002) 170405.
[3] N. Moriya, L.C. Feldman, H.S. Luftman, C.A. King, J. Bevk, B. Freer, Phys. Rev. Lett. 71 (1993) 883; N. Moriya, L.C. Feldman, S.W. Downey, C.A. King, A.B. Emerson, Phys. Rev. Lett. 75 (1995) 1981.
[4] A.F. Faruqi, K.J. Turner, Appl. Math. Comput. 115 (2000) 213.
[5] J.P.M. Heald, J. Stark, Phys. Rev. Lett. 84 (2000) 2366.
[6] See for instance in M. Grewal, A. Andrewa, Kalman Filtering, Prentice-Hall, 1993 (Chapter 6).
[7] N. Wiener, J. Math. Phys. Sci. 2 (1923) 132.
[8] A.A. Dorogovtsev, Stochastic Analysis and Random Maps in Hilbert Space, VSP Publishing, The Netherlands, 1994, In particular see the consideration for high-order stochastic derivative in Chapter 1.
[9] G. Di Nunno, Pure Math. 12 (2001) 1.
[10] H. Kleinert, S.V. Shabanov, Phys. Lett. A 235 (1997) 105.
[11] C. Elachi, Science 209 (1980) 1073–1082.
[12] (a) L. Davis et al., Handbook of Auger Electron Spectroscopy, Physical Electronics, Eden Prairie, MN, 1976; (b) A. Joshi, L.E. Davis, P.W. Palmberg, in: A.W. Czanderna (Ed.), Methods of Surface Analysis, Elsevier Scientific Publishing Co., Amsterdam, 1975.
[13] V. Kontorovich et al., IEEE Signal Processing Lett. 3 (1996) 19; S. Kay, IEEE Signal Processing Lett. 5 (1998) 318.
[14] The Stochastic-Derivative matrix S_k^m, as defined above is in fact a variant of Pascal Triangle.
[15] M.I. Tribelsky, Phys. Rev. Lett. 89 (2002) 070201, Note that for the general case of Eq. (9) in this reference, the variance parameter should be indexed and be moved inside the summation term.
[16] See for example the classical proof by K. Weierstrass, Mathematische Werke, Bd. III, Berlin 1903, pp. 1–17. Can also be found in most textbooks on Functional Analysis.
[17] This approach can easily be generalized also for orthogonal, complete functional systems other than polynomials.