

LA-UR-12-26980 (Accepted Manuscript)

A Stochastic Diffusion Process for the Dirichlet Distribution

Bakosi, Jozsef
Ristorcelli, Raymond

Provided by the author(s) and the Los Alamos National Laboratory (2016-08-31).

To be published in: International Journal of Stochastic Analysis

DOI to publisher's version: 10.1155/2013/842981

Permalink to record: <http://permalink.lanl.gov/object/view?what=info:lanl-repo/lareport/LA-UR-12-26980>

Disclaimer:

Approved for public release. Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the Los Alamos National Security, LLC for the National Nuclear Security Administration of the U.S. Department of Energy under contract DE-AC52-06NA25396. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

A stochastic diffusion process for the Dirichlet distribution

LA-UR-12-26980

February 15, 2013

J. Bakosi, J.R. Ristorcelli

`{jbakosi,jrrj}@lanl.gov`

Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Abstract

The method of potential solutions of Fokker-Planck equations is used to develop a transport equation for the joint probability of N coupled stochastic variables with the Dirichlet distribution as its asymptotic solution. To ensure a bounded sample space, a coupled nonlinear diffusion process is required: the Wiener-processes in the equivalent system of stochastic differential equations are multiplicative with coefficients dependent on all the stochastic variables. Individual samples of a discrete ensemble, obtained from the stochastic process, satisfy a unit-sum constraint at all times. The process may be used to represent realizations of a fluctuating ensemble of N variables subject to a conservation principle. Similar to the multivariate Wright-Fisher process, whose invariant is also Dirichlet, the univariate case yields a process whose invariant is the beta distribution. As a test of the results, Monte-Carlo simulations are used to evolve numerical ensembles toward the invariant Dirichlet distribution.

Keywords: Fokker-Planck equation; Stochastic diffusion; Dirichlet distribution; Monte-Carlo simulation

1 Objective

We develop a Fokker-Planck equation whose statistically stationary solution is the Dirichlet distribution [1, 2, 3]. The system of stochastic differential equations (SDE), equivalent to the Fokker-Planck equation, yields a Markov process that allows a Monte-Carlo method to numerically evolve an ensemble of fluctuating variables that satisfy a unit-sum requirement. A Monte Carlo solution is used to verify that the invariant distribution is Dirichlet.

The Dirichlet distribution is a statistical representation of non-negative variables subject to a unit-sum requirement. The properties of such variables have been of interest in a variety of fields, including evolutionary theory [4], Bayesian statistics [5], geology [6, 7], forensics [8], econometrics [9], turbulent combustion [10], and population biology [11].

2 Preview of results

The Dirichlet distribution [1, 2, 3] for a set of scalars $0 \leq Y_\alpha$, $\alpha = 1, \dots, N-1$, $\sum_{\alpha=1}^{N-1} Y_\alpha \leq 1$, is given by

$$\mathcal{D}(\mathbf{Y}, \boldsymbol{\omega}) = \frac{\Gamma\left(\sum_{\alpha=1}^N \omega_\alpha\right)}{\prod_{\alpha=1}^N \Gamma(\omega_\alpha)} \prod_{\alpha=1}^N Y_\alpha^{\omega_\alpha - 1}, \quad (1)$$

where $\omega_\alpha > 0$ are parameters, $Y_N = 1 - \sum_{\beta=1}^{N-1} Y_\beta$, and $\Gamma(\cdot)$ denotes the gamma function. We derive the stochastic diffusion process, governing the scalars, Y_α ,

$$dY_\alpha(t) = \frac{b_\alpha}{2} [S_\alpha Y_N - (1 - S_\alpha) Y_\alpha] dt + \sqrt{\kappa_\alpha Y_\alpha Y_N} dW_\alpha(t), \quad \alpha = 1, \dots, N-1, \quad (2)$$

where $dW_\alpha(t)$ is an isotropic vector-valued Wiener process [12], and $b_\alpha > 0$, $\kappa_\alpha > 0$, and $0 < S_\alpha < 1$ are coefficients. We show that the statistically stationary solution of Eq. (2) is the Dirichlet distribution, Eq. (1), provided the SDE coefficients satisfy

$$\frac{b_1}{\kappa_1}(1 - S_1) = \dots = \frac{b_{N-1}}{\kappa_{N-1}}(1 - S_{N-1}). \quad (3)$$

The restrictions imposed on the SDE coefficients, b_α , κ_α , and S_α , ensure reflection towards the interior of the sample space, which is a generalized triangle or tetrahedron (more precisely, a simplex) in $N-1$ dimensions. The restrictions together with the specification of the N^{th} scalar as $Y_N = 1 - \sum_{\beta=1}^{N-1} Y_\beta$ ensure

$$\sum_{\alpha=1}^N Y_\alpha = 1. \quad (4)$$

Indeed, inspection of Eq. (2) shows that for example when $Y_1 = 0$, the diffusion is zero and the drift is strictly positive, while if $Y_1 = 1$, the diffusion is zero ($Y_N = 0$) and the drift is strictly negative.

3 Development of the diffusion process

The diffusion process (2) is developed by the method of potential solutions.

We start from the Itô diffusion process [12] for the stochastic vector, Y_α ,

$$dY_\alpha(t) = a_\alpha(\mathbf{Y})dt + b_{\alpha\beta}(\mathbf{Y})dW_\beta(t), \quad \alpha, \beta = 1, \dots, N-1, \quad (5)$$

with drift, $a_\alpha(\mathbf{Y})$, diffusion, $b_{\alpha\beta}(\mathbf{Y})$, and the isotropic vector-valued Wiener process, $dW_\beta(t)$, where summation is implied for repeated indices. Using standard methods given in [12] the equivalent Fokker-Planck equation governing the joint probability, $\mathcal{F}(\mathbf{Y}, t)$, derived from Eq. (5), is

$$\frac{\partial \mathcal{F}}{\partial t} = -\frac{\partial}{\partial Y_\alpha} [a_\alpha(\mathbf{Y})\mathcal{F}] + \frac{1}{2} \frac{\partial^2}{\partial Y_\alpha \partial Y_\beta} [B_{\alpha\beta}(\mathbf{Y})\mathcal{F}], \quad (6)$$

with diffusion $B_{\alpha\beta} = b_{\alpha\gamma}b_{\gamma\beta}$. Since the drift and diffusion coefficients are time-homogeneous, $a_\alpha(\mathbf{Y}, t) = a_\alpha(\mathbf{Y})$ and $B_{\alpha\beta}(\mathbf{Y}, t) = B_{\alpha\beta}(\mathbf{Y})$, Eq. (5) is a statistically stationary process and the solution of Eq. (6) converges to a stationary distribution, [12] Sec. 6.2.2. Our task is to specify the functional forms of $a_\alpha(\mathbf{Y})$ and $b_{\alpha\beta}(\mathbf{Y})$ so that the stationary solution of Eq. (6) is $\mathcal{D}(\mathbf{Y})$, defined by Eq. (1).

A potential solution of Eq. (6) exists if

$$\frac{\partial \ln \mathcal{F}}{\partial Y_\beta} = B_{\alpha\beta}^{-1} \left(2a_\alpha - \frac{\partial B_{\alpha\gamma}}{\partial Y_\gamma} \right) \equiv -\frac{\partial \phi}{\partial Y_\beta}, \quad \alpha, \beta, \gamma = 1, \dots, N-1, \quad (7)$$

is satisfied, [12] Sec. 6.2.2. Since the left hand side of Eq. (7) is a gradient, the expression on the right must also be a gradient and can therefore be obtained from a scalar potential denoted by $\phi(\mathbf{Y})$. This puts a constraint on the possible choices of a_α and $B_{\alpha\beta}$ and on the potential, as $\phi_{,\alpha\beta} = \phi_{,\beta\alpha}$ must also be satisfied. The potential solution is

$$\mathcal{F}(\mathbf{Y}) = \exp[-\phi(\mathbf{Y})]. \quad (8)$$

Now functional forms of $a_\alpha(\mathbf{Y})$ and $B_{\alpha\beta}(\mathbf{Y})$ that satisfy Eq. (7) with $\mathcal{F}(\mathbf{Y}) \equiv \mathcal{D}(\mathbf{Y})$ are sought. The mathematical constraints on the specification of a_α and $B_{\alpha\beta}$ are as follows:

1. $B_{\alpha\beta}$ must be symmetric positive semi-definite. This is to ensure that

- the square-root of $B_{\alpha\beta}$ (e.g. the Cholesky-decomposition, $b_{\alpha\beta}$) exists, required by the correspondence of the SDE (5) and the Fokker-Planck equation (6),
- Eq. (5) represents a diffusion, and
- $\det(B_{\alpha\beta}) \neq 0$, required by the existence of the inverse in Eq. (7).

2. For a potential solution to exist Eq. (7) must be satisfied.

With $\mathcal{F}(\mathbf{Y}) \equiv \mathcal{D}(\mathbf{Y})$ Eq. (8) shows that the scalar potential must be

$$\phi(\mathbf{Y}) = - \sum_{\alpha=1}^N (\omega_{\alpha} - 1) \ln Y_{\alpha}. \quad (9)$$

It is straightforward to verify that the specifications

$$a_{\alpha}(\mathbf{Y}) = \frac{b_{\alpha}}{2} [S_{\alpha} Y_N - (1 - S_{\alpha}) Y_{\alpha}], \quad (10)$$

$$B_{\alpha\beta}(\mathbf{Y}) = \begin{cases} \kappa_{\alpha} Y_{\alpha} Y_N & \text{for } \alpha = \beta, \\ 0 & \text{for } \alpha \neq \beta, \end{cases} \quad (11)$$

satisfy the above mathematical constraints, 1. and 2. Here $b_{\alpha} > 0$, $\kappa_{\alpha} > 0$, $0 < S_{\alpha} < 1$, and $Y_N = 1 - \sum_{\beta=1}^{N-1} Y_{\beta}$. Summation is not implied for Eqs. (9–11).

Substituting Eqs. (9–11) into Eq. (7) yields a system with the same functions on both sides with different coefficients, yielding the correspondence between the N coefficients of the Dirichlet distribution, Eq. (1), and the Fokker-Planck equation (6) with Eqs. (10–11) as

$$\omega_{\alpha} = \frac{b_{\alpha}}{\kappa_{\alpha}} S_{\alpha}, \quad \alpha = 1, \dots, N-1, \quad (12)$$

$$\omega_N = \frac{b_1}{\kappa_1} (1 - S_1) = \dots = \frac{b_{N-1}}{\kappa_{N-1}} (1 - S_{N-1}). \quad (13)$$

For example, for $N=3$ one has $\mathbf{Y} = (Y_1, Y_2, Y_3 = 1 - Y_1 - Y_2)$ and from Eq. (9) the scalar potential is

$$-\phi(Y_1, Y_2) = (\omega_1 - 1) \ln Y_1 + (\omega_2 - 1) \ln Y_2 + (\omega_3 - 1) \ln(1 - Y_1 - Y_2). \quad (14)$$

Eq. (7) then becomes the system

$$\frac{\omega_1 - 1}{Y_1} - \frac{\omega_3 - 1}{Y_3} = \left(\frac{b_1}{\kappa_1} S_1 - 1 \right) \frac{1}{Y_1} - \left[\frac{b_1}{\kappa_1} (1 - S_1) - 1 \right] \frac{1}{Y_3}, \quad (15)$$

$$\frac{\omega_2 - 1}{Y_2} - \frac{\omega_3 - 1}{Y_3} = \left(\frac{b_2}{\kappa_2} S_2 - 1 \right) \frac{1}{Y_2} - \left[\frac{b_2}{\kappa_2} (1 - S_2) - 1 \right] \frac{1}{Y_3}, \quad (16)$$

which shows that by specifying the parameters, ω_{α} , of the Dirichlet distribution as

$$\omega_1 = \frac{b_1}{\kappa_1} S_1, \quad (17)$$

$$\omega_2 = \frac{b_2}{\kappa_2} S_2, \quad (18)$$

$$\omega_3 = \frac{b_1}{\kappa_1} (1 - S_1) = \frac{b_2}{\kappa_2} (1 - S_2), \quad (19)$$

the stationary solution of the Fokker-Planck equation (6) with drift (10) and diffusion (11) is $\mathcal{D}(\mathbf{Y}, \boldsymbol{\omega})$ for $N=3$. The above development generalizes to N variables, yielding Eqs. (12–13), and

reduces to the beta distribution, a univariate specialization of \mathcal{D} for $N = 2$, where $Y_1 = Y$ and $Y_2 = 1 - Y$, see [13].

If Eqs. (12–13) hold, the stationary solution of the Fokker-Planck equation (6) with drift (10) and diffusion (11) is the Dirichlet distribution, Eq. (1). Note that Eqs. (10–11) are one possible way of specifying a drift and a diffusion to arrive at a Dirichlet distribution; other functional forms may be possible. The specifications in Eqs. (10–11) are a generalization of the results for a univariate diffusion process, discussed in [13, 14], whose invariant distribution is beta.

The shape of the Dirichlet distribution, Eq. (1), is determined by the N coefficients, ω_α . Eqs. (12–13) show that in the stochastic system, different combinations of b_α , S_α , and κ_α may yield the same ω_α and that not all of b_α , S_α , and κ_α may be chosen independently to keep the invariant Dirichlet.

4 Corroborating that the invariant distribution is Dirichlet

For any multivariate Fokker-Planck equation there is an equivalent system of Itô diffusion processes, such as the pair of Eqs. (5–6) [12]. Therefore, a way of computing the (discrete) numerical solution of Eq. (6) is to integrate Eq. (5) in a Monte-Carlo fashion for an ensemble [15]. Using a Monte-Carlo simulation we show that the statistically stationary solution of the Fokker-Planck equation (6) with drift and diffusion (10–11) is a Dirichlet distribution, Eq. (1).

The time-evolution of an ensemble of particles, each with $N = 3$ variables (Y_1, Y_2, Y_3) , is numerically computed by integrating the system of equations (5), with drift and diffusion (10–11), for $N = 3$ as

$$dY_1^{(i)} = \frac{b_1}{2} \left[S_1 Y_3^{(i)} - (1 - S_1) Y_1^{(i)} \right] dt + \sqrt{\kappa_1 Y_1^{(i)} Y_3^{(i)}} dW_1^{(i)} \quad (20)$$

$$dY_2^{(i)} = \frac{b_2}{2} \left[S_2 Y_3^{(i)} - (1 - S_2) Y_2^{(i)} \right] dt + \sqrt{\kappa_2 Y_2^{(i)} Y_3^{(i)}} dW_2^{(i)} \quad (21)$$

$$Y_3^{(i)} = 1 - Y_1^{(i)} - Y_2^{(i)} \quad (22)$$

for each particle i . In Eqs. (20–21) dW_1 and dW_2 are independent Wiener processes, sampled from Gaussian streams of random numbers with mean $\langle dW_\alpha \rangle = 0$ and covariance $\langle dW_\alpha dW_\beta \rangle = \delta_{\alpha\beta} dt$. 400,000 particle-triplets, (Y_1, Y_2, Y_3) , are generated with two different initial distributions, displayed in the upper-left of Figures 1 and 2, a triple-delta and a box, respectively. Each member of both initial ensembles satisfy $\sum_{\alpha=1}^3 Y_\alpha = 1$. Eqs. (20–22) are advanced in time with the Euler-Maruyama scheme [16] with time step $\Delta t = 0.05$. Table 1 shows the coefficients of the stochastic system (20–22), the corresponding parameters of the final Dirichlet distribution, and the first two moments at the initial times for the triple-delta initial condition case. The final state of the ensembles are determined by the SDE coefficients, constant for these exercises, also given in Table 1, the same for both simulations, satisfying Eq. (19).

The time-evolutions of the joint probabilities are extracted from both calculations and displayed at different times in Figures 1 and 2. At the end of the simulations two distributions are plotted at the bottom-right of both figures: the one extracted from the numerical ensemble and the Dirichlet distribution determined analytically using the SDE coefficients – in excellent agreement in both figures. The statistically stationary solution of the developed stochastic system is the Dirichlet distribution.

For a more quantitative evaluation, the time evolution of the first two moments, $\mu_\alpha = \langle Y_\alpha \rangle = \int_0^1 \int_0^1 Y_\alpha \mathcal{F}(Y_1, Y_2) dY_1 dY_2$, and $\langle y_\alpha y_\beta \rangle = \langle (Y_\alpha - \langle Y_\alpha \rangle)(Y_\beta - \langle Y_\beta \rangle) \rangle$, are also extracted from the numerical simulation with the triple-delta-peak initial condition as ensemble averages and displayed in Figures 3 and 4. The figures show that the statistics converge to the precise state given by the Dirichlet distribution that is prescribed by the SDE coefficients, see Table 1.

The solution approaches a Dirichlet distribution, with non-positive covariances [2], in the statistically stationary limit, Figure 4(b). Note that during the evolution of the process, $0 < t \lesssim 80$, the solution is not necessarily Dirichlet, but the stochastic variables sum to one at all times. The point (Y_1, Y_2) , governed by Eqs. (20–21), can never leave the $(N-1)$ -dimensional (here $N=3$) convex polytope and by definition $Y_3 = 1 - Y_1 - Y_2$. The rate at which the numerical solution converges to a Dirichlet distribution is determined by the vectors b_α and κ_α .

The above numerical results confirm that starting from arbitrary realizable ensembles the solution of the stochastic system converges to a Dirichlet distribution in the statistically stationary state, specified by the SDE coefficients.

5 Relation to other diffusion processes

It is useful to relate the Dirichlet diffusion process, Eq. (2), to other multivariate stochastic diffusion processes with linear drift and quadratic diffusion.

A close relative of Eq. (2) is the multivariate Wright-Fisher (WF) process [11], used extensively in population and genetic biology,

$$dY_\alpha(t) = \frac{1}{2}(\omega_\alpha - \omega Y_\alpha)dt + \sum_{\beta=1}^{N-1} \sqrt{Y_\alpha(\delta_{\alpha\beta} - Y_\beta)} dW_{\alpha\beta}(t), \quad \alpha = 1, \dots, N-1, \quad (23)$$

where $\delta_{\alpha\beta}$ is Kronecker's delta, $\omega = \sum_{\beta=1}^N \omega_\beta$ with ω_α defined in Eq. (1) and, $Y_N = 1 - \sum_{\beta=1}^{N-1} Y_\beta$. Similarly to Eq. (2), the statistically stationary solution of Eq. (23) is the Dirichlet distribution [17]. It is straightforward to verify that its drift and diffusion also satisfy Eq. (7) with $\mathcal{F} \equiv \mathcal{D}$, i.e. WF is a process whose invariant is Dirichlet and this solution is potential. A notable difference between Eqs. (2) and (23), other than the coefficients, is that the diffusion matrix of the Dirichlet diffusion process is diagonal, while that of the WF process it is full.

Another process similar to Eqs. (2) and (23) is the multivariate Jacobi process, used in econometrics,

$$dY_\alpha(t) = a(Y_\alpha - \pi_\alpha)dt + \sqrt{cY_\alpha} dW_\alpha(t) - \sum_{\beta=1}^{N-1} Y_\alpha \sqrt{cY_\beta} dW_\beta(t), \quad \alpha = 1, \dots, N \quad (24)$$

of Gourieroux & Jasiak [9] with $a < 0$, $c > 0$, $\pi_\alpha > 0$, and $\sum_{\beta=1}^N \pi_\beta = 1$.

In the univariate case the Dirichlet, WF, and Jacobi diffusions reduce to

$$dY(t) = \frac{b}{2}(S - Y)dt + \sqrt{\kappa Y(1 - Y)} dW(t), \quad (25)$$

see also [13], whose invariant is the beta distribution, which belongs to the family of Pearson diffusions, discussed in detail by Forman & Sorensen [14].

6 Summary

The method of potential solutions of Fokker-Planck equations has been used to derive a transport equation for the joint distribution of N fluctuating variables. The equivalent stochastic process, governing the set of random variables, $0 \leq Y_\alpha$, $\alpha = 1, \dots, N-1$, $\sum_{\alpha=1}^{N-1} Y_\alpha \leq 1$, reads

$$dY_\alpha(t) = \frac{b_\alpha}{2}[S_\alpha Y_N - (1 - S_\alpha)Y_\alpha]dt + \sqrt{\kappa_\alpha Y_\alpha Y_N} dW_\alpha(t), \quad \alpha = 1, \dots, N-1, \quad (26)$$

where $Y_N = 1 - \sum_{\beta=1}^{N-1} Y_\beta$, and b_α , κ_α and S_α are parameters, while $dW_\alpha(t)$ is an isotropic Wiener process with independent increments. Restricting the coefficients to $b_\alpha > 0$, $\kappa_\alpha > 0$ and $0 < S_\alpha < 1$,

and defining Y_N as above ensure $\sum_{\alpha=1}^N Y_\alpha = 1$ and that individual realizations of (Y_1, Y_2, \dots, Y_N) are confined to the $(N-1)$ -dimensional convex polytope of the sample space. Eq. (26) can therefore be used to numerically evolve the joint distribution of N fluctuating variables required to satisfy a conservation principle. Eq. (26) is a coupled system of nonlinear stochastic differential equations whose statistically stationary solution is the Dirichlet distribution, Eq. (1), provided the coefficients satisfy

$$\frac{b_1}{\kappa_1}(1 - S_1) = \dots = \frac{b_{N-1}}{\kappa_{N-1}}(1 - S_{N-1}). \quad (27)$$

In stochastic modeling, one typically begins with a physical problem, perhaps discrete, *then* derives the stochastic differential equations whose solution yields a distribution. In this paper we reversed the process: we assumed a desired stationary distribution and derived the stochastic differential equations that converge to the assumed distribution. A potential solution form of the Fokker-Planck equation was posited, from which we obtained the stochastic differential equations for the diffusion process whose statistically stationary solution is the Dirichlet distribution. We have also made connections to other stochastic processes, such as the Wright-Fisher diffusions of population biology and the Jacobi diffusions in econometrics, whose invariant distributions possess similar properties but whose stochastic differential equations are different.

Acknowledgements

It is a pleasure to acknowledge a series of informative discussions with J. Waltz. This work was performed under the auspices of the U.S. Department of Energy under the Advanced Simulation and Computing Program.

References

- [1] N. L. Johnson. An approximation to the multinomial distribution some properties and applications. *Biometrika*, 47(1-2):93–102, 1960.
- [2] J. E. Mosimann. On the compound multinomial distribution, the multivariate-distribution, and correlations among proportions. *Biometrika*, 49(1-2):65–82, 1962.
- [3] S. Kotz, N.L. Johnson, and N. Balakrishnan. *Continuous Multivariate Distributions: Models and applications*. Wiley series in probability and statistics: Applied probability and statistics. Wiley, 2000.
- [4] K. Pearson. Mathematical Contributions to the Theory of Evolution. On a Form of Spurious Correlation Which May Arise When Indices Are Used in the Measurement of Organs. *Royal Society of London Proceedings Series I*, 60:489–498, 1896.
- [5] C.D.M. Paulino and de Braganca Pereira C.A. Bayesian methods for categorical data under informative general censoring. *Biometrika*, 82(2):439–446, 1995.
- [6] F. Chayes. Numerical correlation and petrographic variation. *J. Geol.*, 70:440–452, 1962.
- [7] P. S. Martin and J. E. Mosimann. Geochronology of pluvial lake cochise, southern arizona; [part] 3, pollen statistics and pleistocene metastability. *Am. J. Sci.*, 263:313–358, 1965.
- [8] K. Lange. Applications of the Dirichlet distribution to forensic match probabilities. *Genetica*, 96:107–117, 1995. 10.1007/BF01441156.
- [9] C. Gourieroux and J. Jasiak. Multivariate Jacobi process with application to smooth transitions. *Journal of Econometrics*, 131:475–505, 2006.
- [10] S. S. Girimaji. Assumed β -pdf model for turbulent mixing: validation and extension to multiple scalar mixing. *Combust. Sci. Technol.*, 78(4):177 – 196, 1991.
- [11] M. Steinrucken, Y.X. Rachel Wang, and Y.S. Song. An explicit transition density expansion for a multi-allelic Wright–Fisher diffusion with general diploid selection. *Theoretical Population Biology*, 83(0):1–14, 2013.
- [12] C. W. Gardiner. *Stochastic methods, A Handbook for the Natural and Social Sciences*. Springer-Verlag, Berlin Heidelberg, 4 edition, 2009.
- [13] J. Bakosi and J.R. Ristorcelli. Exploring the beta distribution in variable-density turbulent mixing. *J. Turbul.*, 11(37):1–31, 2010.
- [14] J. L. Forman and M. Sorensen. The Pearson Diffusions: A Class of Statistically Tractable Diffusion Processes. *Scandinavian Journal of Statistics*, 35:438–465, 2008.
- [15] S. B. Pope. PDF methods for turbulent reactive flows. *Prog. Energ. Combust.*, 11:119–192, 1985.
- [16] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, Berlin, 1999.
- [17] S. Karlin and H.M. Taylor. *A Second Course in Stochastic Processes*. Number v. 2. Acad. Press, 1981.

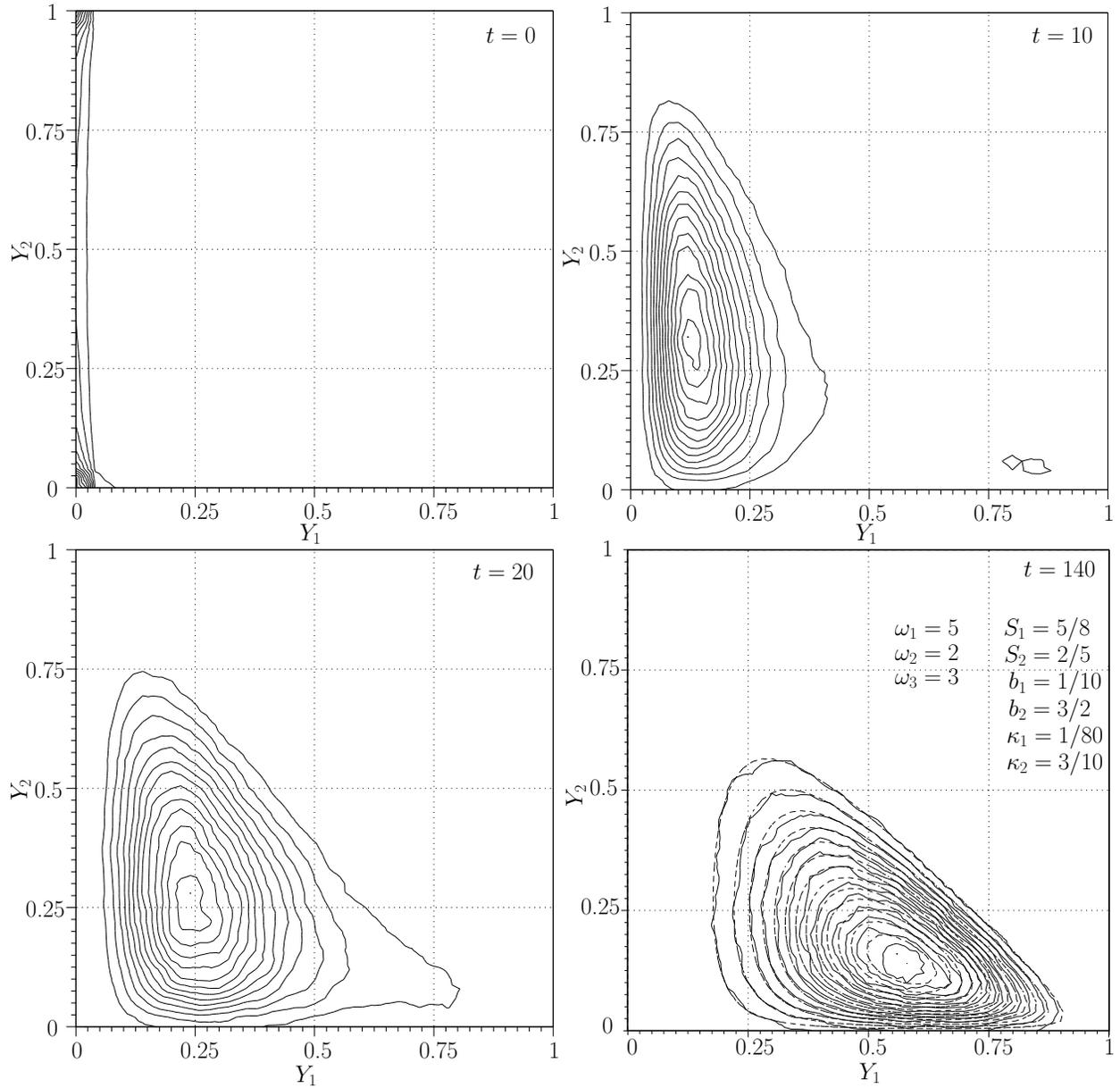


Fig. 1: Time evolution of the joint probability, $\mathcal{F}(Y_1, Y_2)$, extracted from the numerical solution of Eqs. (20–22). The initial condition is a triple-delta distribution, with unequal peaks at the three corners of the sample space. At the end of the simulation, $t = 140$, the solid lines are that of the distribution extracted from the numerical ensemble, dashed lines are that of a Dirichlet distribution to which the solution converges in the statistically stationary state, implied by the constant SDE coefficients, sampled at the same heights.

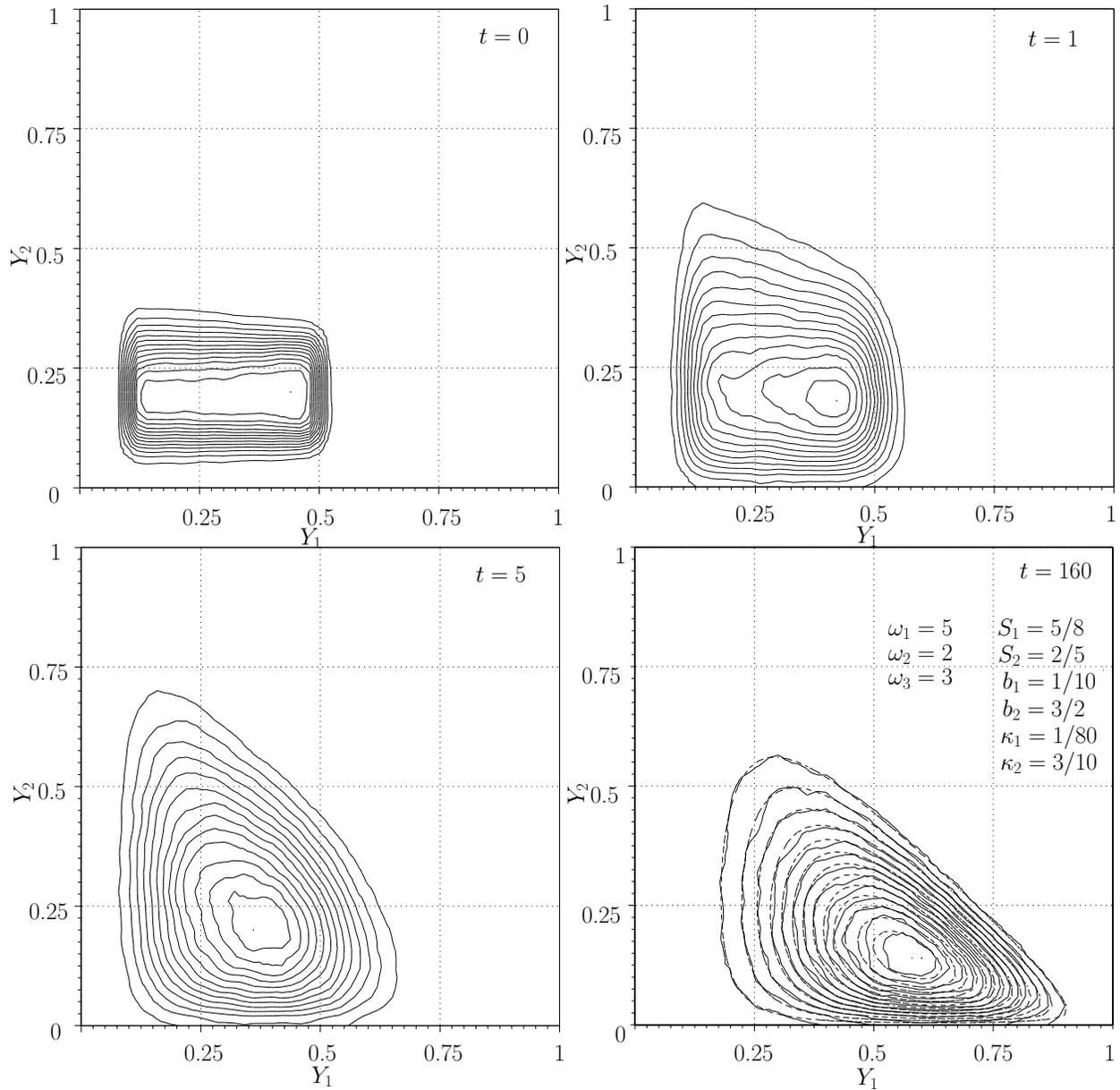


Fig. 2: Time evolution of the joint probability, $\mathcal{F}(Y_1, Y_2)$, extracted from the numerical solution of Eqs. (20–22). The top-left panel shows the initial condition: a box with diffused sides. By $t = 160$, bottom-right panel, the distribution converges to the same Dirichlet distribution as in Figure 1.

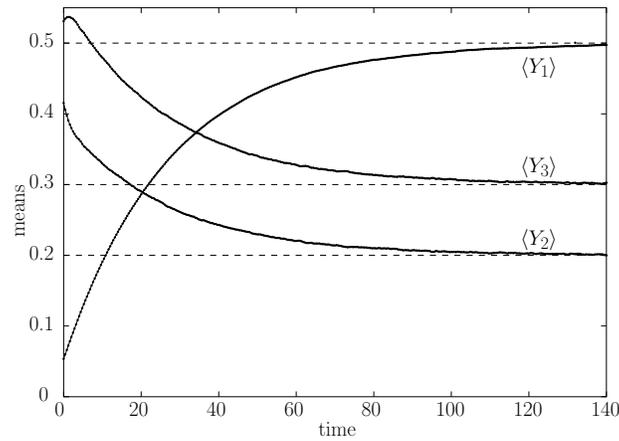


Fig. 3: Time evolution of the means extracted from the numerically integrated system of Eqs. (20–22) starting from the triple-delta initial condition. Dotted-solid lines – numerical solution, dashed lines – statistics of the Dirichlet distribution determined analytically using the constant coefficients of the SDE, see Table 1.

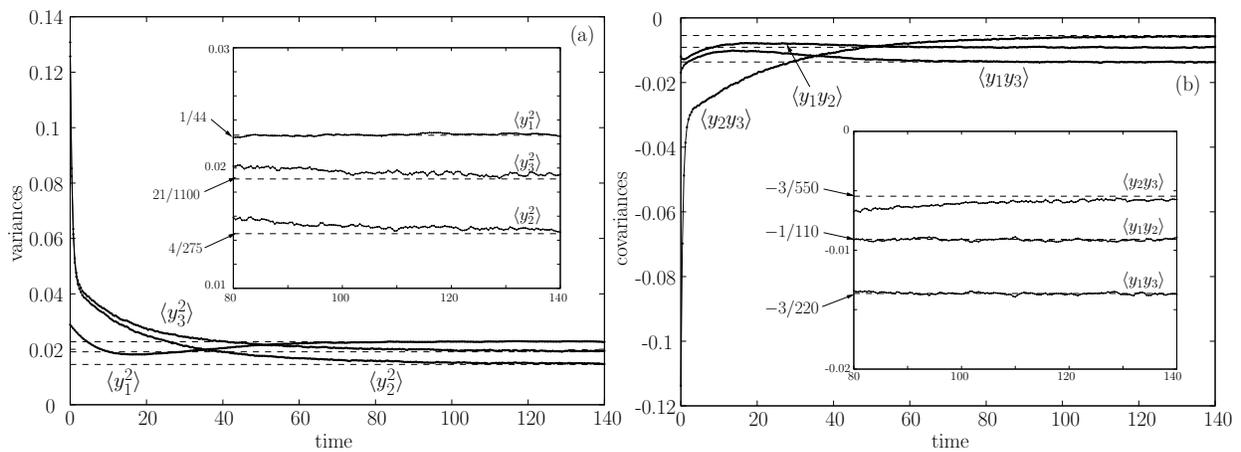


Fig. 4: Time evolution of the second central moments extracted from the numerically integrated system of Eqs. (20–22) starting from the triple-delta initial condition. The legend is the same as in Figure 3.

Tab. 1: Initial and final states of the Monte-Carlo simulation starting from a triple-delta. The coefficients, $b_1, b_2, S_1, S_2, \kappa_1, \kappa_2$, of the system of SDEs (20–22) determine the distribution to which the system converges. The Dirichlet parameters, implied by the SDE coefficients via Eqs. (17–19), are in brackets. The corresponding statistics are determined by the well-known formulae of Dirichlet distributions [2].

Initial state: triple-delta, see Figure 1	SDE coefficients and the statistics of their implied Dirichlet distribution in the stationary state
	$b_1 = 1/10 \quad b_2 = 3/2 \quad (\omega_1 = 5)$
	$S_1 = 5/8 \quad S_2 = 2/5 \quad (\omega_2 = 2)$
	$\kappa_1 = 1/80 \quad \kappa_2 = 3/10 \quad (\omega_3 = 3)$
$\langle Y_1 \rangle_0 \approx 0.05$	$\langle Y_1 \rangle_s = 1/2$
$\langle Y_2 \rangle_0 \approx 0.42$	$\langle Y_2 \rangle_s = 1/5$
$\langle Y_3 \rangle_0 \approx 0.53$	$\langle Y_3 \rangle_s = 3/10$
$\langle y_1^2 \rangle_0 \approx 0.03$	$\langle y_1^2 \rangle_s = 1/44$
$\langle y_2^2 \rangle_0 \approx 0.125$	$\langle y_2^2 \rangle_s = 4/275$
$\langle y_3^2 \rangle_0 \approx 0.13$	$\langle y_3^2 \rangle_s = 21/1100$
$\langle y_1 y_2 \rangle_0 \approx -0.012$	$\langle y_1 y_2 \rangle_s = -1/110$
$\langle y_1 y_3 \rangle_0 \approx -0.017$	$\langle y_1 y_3 \rangle_s = -3/220$
$\langle y_2 y_3 \rangle_0 \approx -0.114$	$\langle y_2 y_3 \rangle_s = -3/550$