Remarks on Non-Markov Processes*

N.G. van Kampen

Institute for Theoretical Physics, University Utrecht

Princetonplein 5, 3584 CC Utrecht,

The Netherlands

Received March 9, 1998

It is emphasized that non-Markovian processes, which occur for instance in the case of colored noise, cannot be considered merely as corrections to the class of Markov processes but require special treatment. Many familiar concepts, such as first-passage times, no longer apply and need be reconsidered. Several methods of dealing with non-Markov processes are discussed. As an example a recent application to the transport of ions through a membrane is briefly mentioned.

I. Definition of Markov processes.

The term 'non-Markov Process' covers all random processes with the exception of the very small minority that happens to have the Markov property.

FIRST REMARK. Non-Markov is the rule, Markov is the exception.

It is true that this minority has been extensively studied, but it is not proper to treat non-Markov processes merely as modifications or corrections of the Markov processes – as improper as for instance treating all nonlinear dynamical systems as corrections to the harmonic oscillator. I therefore have to start by reviewing some general facts [1, 2].

A *stochastic process* is a collection of random variables X_t, labeled by an index t which may be discrete but more often covers all real numbers in some interval. The stochastic properties of $\{X_t\}$ are expressed by the joint distribution functions

$$P_n(x_1,t_1;x_2,t_2;\ldots;x_n,t_n)dx_1dx_2\ldots dx_n = \text{probability that}$$

$$x_1 < X_{t_1} < x_1 + dx_1, \quad x_2 < X_{t_2} < x_2 + dx_2, \ldots, x_n < X_{t_n} < x_n + dx_n. \tag{1}$$

The process is uniquely defined by the entire set of these distribution functions for $n = 1, 2, \ldots$, which in general is infinite.

When the values $X_{t_1} = x_1, X_{t_2} = x_2, \ldots, X_{t_k} = x_k$ are given, the remaining variables obey the *conditional probability distribution function*

$$P(x_{k+1},t_{k+1};\ldots;x_n,t_n|x_1,t_1;\ldots;x_k,t_k) = \frac{P_n(x_1,t_1;\ldots;x_k,t_k;x_{k+1},t_{k+1};\ldots;x_n,t_n)}{P(x_1,t_1;\ldots;x_k,t_k)}.$$

*This text corresponds to an invited talk at the "Workshop on the Foundations of Statistical Mechanics and Thermodynamics" held in Natal, Brazil, in October, 1997.

This is a probability distribution of $X_{t_{k+1}}, \ldots, X_{t_n}$, in which $x_1, \ldots, x_k$ enter as parameters. Let us take the t_i in chronological order, then the process is *Markov* if this conditional probability depends on the latest value x_k at t_k alone and is independent of the earlier values $x_{i<k}$. This must hold for all n, for any choice of k, and for any $t_1, \ldots, t_k$ and $x_1, \ldots, x_k$. If this is true, all P_n can be constructed once P_1 and P_2 are given. For example,

$$\begin{aligned} P_3(x_1,t_1;x_2,t_2;x_3,t_3) &= P(x_3,t_3|x_1,t_1;x_2,t_2)P_2(x_1,t_1;x_2,t_2) \\ &= P(x_3,t_3|x_2,t_2)P(x_2,t_2|x_1,t_1)P_1(x_1,t_1). \end{aligned} \tag{2}$$

SECOND REMARK. The reason for the popularity of Markov processes is the fact that they are fully determined by these two functions alone. For non-Markov processes the distribution functions(1) must be determined by some other means, usually an entirely different mathematical construction. For M-processes it makes therefore sense to honor the function $P(x_2,t_2|x_1,t_1)$ with the name *transition probability.*

II. Example.

Symmetric random walk in 1 dimension. Here $t = 0, 1, 2, \ldots$ and x takes integer values i. The process is Markovian with symmetric transition probability

$$P(i,t+1|i',t) = \tfrac{1}{2}\delta_{i,i'+1} + \tfrac{1}{2}\delta_{i,i'-1}.$$

But suppose the walker has a tendency to persist in his direction: probability p to step in the same direction, and q to return [3]. Then X_t is no longer Markovian since the probability of X_t depends not just on x_{t-1} but also on x_{t-2}. This may be remedied by introducing the two-component variable $\{X_t, X_{t-1}\}$. This joint variable is again Markovian, with transition probability

$$P(i_1,i_2,t+1|i_1',i_2',t) = \delta_{i_2,i_1'}[p\delta_{i_1-i_2,i_1'-i_2'} + q\delta_{i_1,i_2'}].$$

THIRD REMARK. A physical process (i.e. some physical phenomenon evolving in time) may or may not be Markovian, depending on the variables used to describe it.

If the memory of our random walk involves more preceding steps, more additional variables are needed. That does no longer work, however, if the memory extends over *all* previous steps. Example: polymers with excluded volume. This problem is often modelled as a random walk, but it is irremediably non-Markov and has not been solved [4].

III. The master equation

Take a Markov process in which t is the time while X_t takes discrete values $i = 0, 1, 2, \ldots$ In eq. (2) take $t_3 = t_2 + \Delta t$,

$$\frac{P_3(i_1,t_1;i_2,t_2;i_3,t_2+\Delta t)}{P_1(i_1,t_1)} = P(i_3,t_2+\Delta t|i_2,t_2)P(i_2,t_2|i_1,t_1).$$

Sum over i_2 and take the limit to obtain the *master equation*

$$\dot{P}(i,t|i_1,t_1) = \sum_{i'}\{W_{i,i'}P(i',t|i_1,t_1) - W_{i',i}P(i,t|i_1,t_1)\}. \tag{3}$$

The $W_{i,i'}$ are *transition probabilities per unit time* and are properties belonging to the physical system (such as squares of matrix elements), while P refers to the *state* of the system. The parameters i_1, t_1 are often not written, which may lead to the misconception that P in (3) is the same as P_1 in (1).

FOURTH REMARK. *The master equation is an equation for the transition probability of a Markov process*, valid for any initial i_1, t_1. If one knows in addition $P_1(i_1, t_1)$ the whole hierarchy (1) and thus the process is uniquely determined (for $\geq t_1$).

Warning. In the literature one occasionally encounters something called a "master equation with memory",

$$\dot{P}(i,t|i_i,t_1) = \int_{t_1}^{t} dt' \sum_{i'} \{W_{i,i'}(t-t')P(i',t'|i_1,t_1) - W_{i',i}(t-t')P(i,t'|i_1,t_1)\} \ ,$$

with the claim that it defines a non-Markov process. *Objections.* (i) A non-Markov process is not defined when merely $P(i,t|i_1,t_1)$ is known. (ii) The equation cannot be true for every x_1, t_1. (iii) The equation is no guarantee that the process is not Markovian [5].

IV. Diffusion.

Let x be continuous as well. Then $W_{i,i'}$ takes the form of an integral kernel $W(x|x')$. If the process is such that during an infinitely short Δt only infinitely small jumps are possible, then the kernel reduces to a differential operator [6]. The simplest example is the diffusion equation for the coordinate x of a Brownian particle,

$$\frac{\partial P(x,t)}{\partial t} = D\frac{\partial^2 P(x,t)}{\partial x^2}. \tag{4}$$

The solution of this equation specified by the initial condition $P(x,t_1) = \delta(x-x_1)$ is the transition probability $P(x,t|x_1,t_1)$.

Here the *coordinate* is treated as Markovian, although the particle has a *velocity* v as well. One ought therefore to consider the joint variable $\{x,v\}$ as Markovian, with master equation

$$\frac{\partial P(x,v,t)}{\partial t} = -v\frac{\partial P}{\partial x} + U'(x)\frac{\partial P}{\partial v} + \gamma\left(\frac{\partial}{\partial v}vP + T\frac{\partial^2 P}{\partial v^2}\right). \tag{5}$$

This is Kramers' equation [7]; γ is a friction coefficient and $U(x)$ an external potential.

In this picture x by itself is not Markov. How is that compatible with (4)? The answer is that for large γ the collisions are so frequent that the velocity distribution rapidly becomes locally Maxwellian,

$$P(x,v,t) \longrightarrow P(x,t)\frac{\exp[-v^2/2T]}{\sqrt{2\pi T}}. \tag{6}$$

This makes it possible to eliminate v so that there remains an equation for $P(x,t)$ by itself [7,1], namely

$$\frac{\partial P(x,t)}{\partial t} = \frac{1}{\gamma}\left\{\frac{\partial^2 P}{\partial x^2} + \frac{\partial}{\partial x}\left(\frac{U'(x)}{T}P\right)\right\} + \mathcal{O}(\gamma^{-2}). \tag{7}$$

V. First-passage problems.

Consider one-dimensional diffusion in a potential field as given by (7). Let it take place in a finite medium $x_a < x < x_c$ (Figure 1). When the diffusing particle starts at an interior point x_b, what are the chances that it will exit through x_a or x_c, respectively?

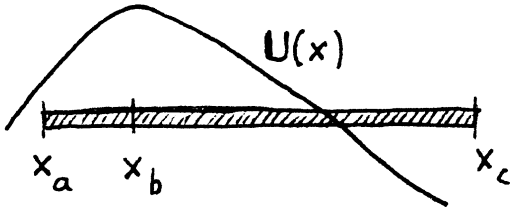

Figure 1

The answer is obtained by solving (7) with absorbing boundary conditions: $P(x_a,t) = 0$ and $P(x_c,t) = 0$. (The solution can be obtained explicitly thanks to the fact that the time does not occur in the equation for the probabilities.) It is clear that when x_b is at the top of a high maximum of $U(x)$ the exit probabilities will be fifty-fifty. It is also possible to find the mean time for either exit [2].

In three dimensions the question is: When I surround the potential maximum by a closed surface, what is the probability distribution of the exit points on that surface and how long does it take? This problem can be formulated in the same way, but cannot usually be solved analytically.

For a particle described by (5), however, the coordinate is not Markovian and it does not suffice therefore to know that $x(t_1) = x_b$: one also has to know its preceding history. In the present case, that history is represented by the value of v at t_1. Of course it is possible to simply pick an initial $P_1(x,v,t_1)$, but the correct choice depends on the problem one wants to solve. For instance I want to compute the autocorrelation function of x,

$$\begin{aligned} < x(t_1)x(t_2) > &= \int x_1 x_2 P(x_1,v_1,t_1;x_2,v_2,t_2)dx_1dx_2dv_1dv_2 \\ &= \int x_1 x_2 P_1(x_1,v_1,t_1)P(x_2,v_2,t_2|x_1,v_1,t_1)dx_1dx_2dv_1dv_2 \ . \end{aligned}$$

Evidently one needs to know the correct initial distribution $P_1(x,v,t)$. Only in the limit of large γ may it be replaced with $P_1(x,t_1)$ with the aid of (6).

FIFTH REMARK. For a non-Markov process the initial value problem is not well-defined unless further information about the problem is supplied.

Of more interest is the question of *escape from a potential minimum* such as x_a in Fig. 2. How long does it take to get across the barrier? Here the ambiguity of the initial velocity is harmless because the particle moves around in the potential valley long enough for the Maxwellian (6) to prevail.

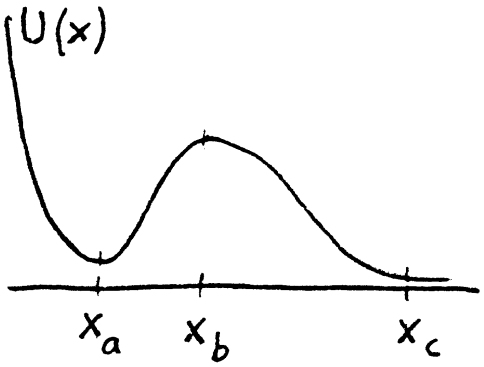

Figure 2

In the case of *diffusion* described by (7) one may take the mean time of first arrival at x_b – and multiply

by 2 because once in x_b there is equal probability to escape or go back. The mean first-passage time can again be computed analytically. In more dimensions however, the question is: How long does it take to escape from a minimum x_a surrounded by a potential ridge? This mean time is determined by the lowest mountain pass on the ridge and an elegant approximation is available [8].

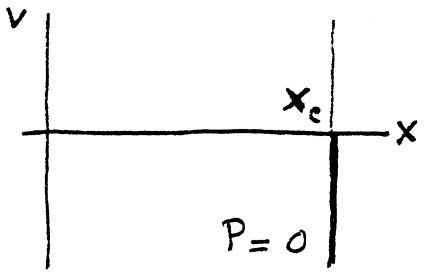

Figure 3

In Kramers' equation (5), however, it is not enough to compute the time for $x(t)$ to reach x_b because $x(t)$ is not Markovian and one cannot tell the probability of subsequent escape without taking into account v. It is necessary to reformulate the problem by picking a large x_c and solving (5) with boundary condition: $P(x_c, v, t) = 0$ for $v < 0$ (Fig. 3). This is the much discussed Kramers problem [7,9].

SIXTH REMARK. For non-Markov processes the first-passage problem may be formulated but its physical relevance is questionable.

VI. Langevin equation and colored noise.

We start from Kramers' equation (5) and consider the spatially homogeneous case $U' = 0$. Then it is possible to integrate over x and obtain an equation for the distribution $P(v,t)$ of v alone,

$$\frac{\partial P(v,t)}{\partial t} = \gamma \frac{\partial}{\partial v} vP + \gamma T \frac{\partial^2 P}{\partial v^2}. \tag{8}$$

The Markov process $v(t)$ described by this master equation is the Ornstein-Uhlenbeck process. A mathematically equivalent way of describing this process is the Langevin equation,

$$\dot{v} = -\gamma v + \xi(t), \tag{9}$$

where $\xi(t)$ is a stochastic function, called *Gaussian white noise*, whose stochastic properties are determined by

$$< \xi(t) >= 0, \qquad < \xi(t)\xi(t') >= C\delta(t - t'). \tag{10}$$

As a result, the solution $v(t)$ (with given initial value $v(t_1) = v_1$) is also a stochastic function and Markovian, its master equation being (8) if one takes $C = 2\gamma T$. The Langevin equation (9) with (10) is popular, because it is more intuitive than (8), but it is not better.

Numerous authors have generalized (9) by taking $\xi(t)$ non-white or colored noise, i.e., not delta-correlated [10]

$$< \xi(t)\xi(t') >= \phi(t - t'). \tag{11}$$

with some even function ϕ.

SEVENTH REMARK. When the noise in the L-equation is not white the variable $v(t)$ is not Markov. Hence for the Langevin equation with colored noise one cannot formulate a meaningful initial condition or a first-passage time.

A different generalization of the Langevin equation (9) is the nonlinear version [11],

$$\dot{x} = f(x) + g(x)\xi(t). \tag{12}$$

When $\xi(t)$ is *white*, then $x(t)$ is Markov. Unfortunately the equation as it stands has no well-defined meaning and has to be supplied by an "interpretation rule", either Itô or Stratonovich. To avoid the vagaries of the Itô calculus we choose the latter. It corresponds to the M-equation,

$$\frac{\partial P(x,t)}{\partial t} = -\frac{\partial}{\partial x} f(x)P + \frac{C}{2}\frac{\partial}{\partial x} g(x) \frac{\partial}{\partial x} g(x)P.$$

When $\xi(t)$ is *colored*, then $x(t)$ is not Markov and no M-equation exists. Attempts at constructing analogous partial differential equations are doomed to fail. Eq. (12) belongs to the general class of stochastic differential equations. They may be treated by approximation methods developed for the case that $\xi(t)$ is off-white, i.e. short correlation and sharply peaked $\phi(t-t')$ [12]. The M-equation appears as the first approximation.

An other device, applicable when this time is not short, was used by Kubo [13] and has been rediscovered many times [14]. Suppose $\xi(t)$ is itself a Markov process governed by an M-equation, for example the O.-U. process governed by (8). Then *the joint variable* (x,ξ) *is Markov*, with M-equation

$$\frac{\partial P(x,\xi,t)}{\partial t} = -\frac{\partial}{\partial x}\{f(x)+\xi g(x)\}P + \gamma\left(\frac{\partial}{\partial \xi}\xi P + T\frac{\partial^2 P}{\partial \xi^2}\right). \tag{13}$$

Again, the non-Markov x has been tamed by introducing an additional variable.

In practice this is of not much help unless one chooses for $\xi(t)$ an even simpler non-white process, viz., the "dichotomic Markov process". That is, ξ has two possible values +1, -1, and jumps with a constant probability γ per unit time, as described by the M-equation

$$\begin{aligned}\dot{P}_+ &= \quad - \quad \gamma P_+ + \gamma P_- \\ \dot{P}_- &= \qquad\quad \gamma P_+ - \gamma P_- .\end{aligned}$$

With this choice equation (13) reduces to

$$\begin{aligned}\frac{\partial}{\partial t}P_+(x,t) &= -\frac{\partial}{\partial x}(f+g)P_+ - \gamma P_+ + \gamma P_- \\ \frac{\partial}{\partial t}P_-(x,t) &= -\frac{\partial}{\partial x}(f-g)P_- + \gamma P_+ - \gamma P_-\end{aligned}$$

which is less forbidding, but still the subject of many articles [15]. Other choices for $\xi(t)$, two-valued but not Markovian, have also been considered [16].

VII. A class of non-Markov processes.

In various connections the following type of process occurs [17,18]. Let X_t have two or more possible states j, and in each it has a probability per unit time γ_{ij} to jump to i. If the γ_{ij} are constants, $\{X_t\}$ is a Markov process with M-equation (3). However, suppose they are functions $\gamma(\tau)$ of the time τ elapsed since arrival in j. *Examples:* Molecules in solution may get stuck to the wall temporarily; a bacterium produces offspring after reaching a certain age. Other examples in [18,1].

The probability that X is still in j at a time t after entering it, is given by

$$u_j(t) = \exp\left[-\int_o^t \Sigma_i \gamma_{ij}(\tau')d\tau'\right]. \tag{14}$$

When starting at $t=0$ in j_0, it may have arrived in the state j at time t through a sequence of s transitions, at times $t_1, t_2, \ldots, t_s$, taking it through the states $j_1, j_2, \ldots, j_s = j$. The probability for this particular history to happen is

$$u_{j_o}(t_1)\gamma_{j_1 j_o}(t_1)u_{j_1}(t_2-t_1)\gamma_{j_2 j_1}(t_2-t_1)\ldots\gamma_{j_s j_{s-1}}(t_s-t_{s-1})u_{j_s}(t-t_s).$$

The probability $P_{jj_o}(t)$ to be at time t in state j is obtained by summing over all histories, that is: summing over all s and all $j_s, \ldots, j_{s-1}$ and integrating over all intermediate times $t_1, \ldots, t_s$. The result, written in Laplace transforms, is

$$\hat{P}_{jj_o}(\lambda) = \hat{u}_j(\lambda)\left[\frac{1}{1-\hat{V}(\lambda)}\right]_{jj_o}, \tag{15}$$

where $\hat{u}_j(\lambda)$ is the transform of (14), and the matrix $\hat{V}$

is

$$\hat{V}_{ij}(\lambda) = \int_0^\infty \gamma_{ij}(\tau) u_j(\tau) e^{-\lambda\tau} d\tau.$$

In some cases the result (15) can be evaluated explicitly.

Example: transport of ions through a membrane. The following model has been suggested [19]. A membrane separates two reservoirs and ions may enter from the left or from the right reservoir. The rates at which they enter (i.e., the probabilities per unit time) are determined by the two outside liquids, with the restriction that there can be at no time more than one ion inside the membrane. Once inside, the ion may exit after a while on either side. Thus the interior of the membrane is a system having 5 states:

(0) empty;

(1) one ion that entered at left and is destined to exit at right;

(2) one ion from the left destined to exit at left;

(3, 4) two similar states with ions from the right.

The exits occur with probabilities γ_{oj} per unit time ($j = 1, 2, 3, 4$), which depend on time τ elapsed since the ion entered. This is a non-Markov process of the type described by (15). The equation can be solved and makes it possible to compute such physically relevant quantities as the average of the transported current, and the spectrum of fluctuations [20].

References

[1] N.G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam 1981, 1992).

[2] C.W. Gardiner, *Handbook of Stochastic Methods* (Springer, Berlin 1983).

[3] F. Zernike, in *Handbuch der Physik* III (Geiger and Scheel eds., Springer, Berlin 1926).

[4] M. Doi and S.F. Edwards, *The Theory of Polymer Dynamics* (Oxford Univ. Press 1986).

[5] For a counterexample see ref. 1, ch IV, eq. (2.9).

[6] W. Feller, *An Introduction to Probability Theory and its Applications* I (2^{nd} ed., Wiley, New York 1966) p. 323.

[7] H.A. Kramers, Physica **7**, 284 (1940).

[8] Z. Schuss and B. Matkowski, SIAM J. Appl. Math. **35**, 604 (1970).

[9] P. Hänggi, P. Talkner, M. Berkovich, Rev. Mod. Phys. **62**, 251 (1990).

[10] E.g. S. Faetti and P. Grigolini, Phys. Rev. A **36**, 441 (1987); M.J. Dykman, Phys. Rev. A **42**, 2020 (1990).

[11] P. Hänggi and P. Riseborough, Phys. Rev. A **27**, 3379 (1983); C. Van den Broeck and P. Hänggi, Phys. Rev. A **30**, 2730 (1984). For an application to lasers see S. Zhu, A.W. Yu, and R. Roy, Phys.. Rev. A **34**, 4333 (1986).

[12] N.G. van Kampen, Physica **74**, 215 and 239 (1974); Physics Reports **24**, 171 (1976); J. Stat. Phys. **54**, 1289 (1989); R.H. Terwiel, Physica **74**, 248 (1974).

[13] R. Kubo, in *Stochastic Processes in Chemical Physics* (K.L. Shuler ed., Interscience, New York 1969).

[14] E.g. A.J. Bray and A.J. McKane, Phys. Rev. Letters **62**, 493 (1989); D.T. Gillespie, Am. J. Phys. **64**, 1246 (1996).

[15] Much literature is given in A. Fuliński, Phys. Rev. E **50**, 2668 (1994).

[16] R.F. Pawula, J.M. Porrà, and J. Masoliver, Phys. Rev. E **47**, 189 (1993); J.M. Pontà, J. Masoliver, and K. Lindenberg, Phys. Rev. E **48**, 951 (1993).

[17] G.H. Weiss, J. Stat. Phys. **8**, 221 (1973); K.J. Lindenberg and R. Cukier, J. Chem. Phys. **67**, 568 (1977).

[18] N.G. van Kampen, Physica A **96**, 435 (1979).

[19] E. Barkai, R.S. Eisenberg, and Z. Schuss, Phys. Rev. E **54**, 1 (1996).

[20] N.G. van Kampen, Physica A **244**, 414 (1997).