Occurrence Probabilities of Stochastic Paths

Dirk Helbing and Rolf Molini

II. Institute for Theoretical Physics, University of Stuttgart, 70550 Stuttgart, Germany

Abstract

An analytical formula for the occurrence probability of Markovian stochastic paths with repeatedly visited and/or equal departure rates is derived. This formula is essential for an efficient investigation of the trajectories belonging to random walk models and for a numerical evaluation of the ‘contracted path integral solution’ of the discrete master equation [Phys. Lett. A 195, 128 (1994)].
I. INTRODUCTION

Stochastic processes play an important role in all scientific fields dealing with systems that are subject to inherent or external random influences (‘fluctuations’). Therefore, numerous methods and models have been developed for the description of stochastically behaving systems \[1–4\]. The fields of application reach from physics \[2\] over chemistry \[2,5\] and biology \[6\] to the economic and social sciences \[4,7\].

In this paper we will focus on Markovian stochastic processes which do not essentially depend on so-called memory effects so that correlations of present transitions with past states of the considered system (except the last one) can be neglected. We can distinguish cross-section oriented methods describing the temporal evolution of the distribution of states and longitudinal-data oriented methods delineating single stochastic trajectories (time series). If one is confronted with a continuous state space the distribution of states is governed by the Fokker-Planck equation \[8,1,4\] and corresponding trajectories follow the Langevin equation (stochastic differential equation) \[8,1,4\]. Related methods exist for quantum mechanical systems \[9\]. In the following we will concentrate on systems with a discrete state space. Then, the distribution of states satisfies the master equation \[1,4\] whereas the corresponding time series are determined by random walk models \[3\] and generated by means of Monte-Carlo simulations \[10\].

In a recent paper \[11\], an important relation between the distribution of states and the occurrence probabilities of paths has been established which was called the ‘contracted path-integral solution’. However, during its numerical implementation, this relation turned out to be restricted to the infrequent case of pure birth processes (uni-directional transitions), since it did not provide an analytical formula for paths with repeatedly visited states. Therefore, this paper presents the non-trivial derivation of the missing occurrence-probability formula for ‘degenerate paths’ (Sec. III).

The numerical implementation of the ‘contracted path-integral solution’ is outlined in Section IV. Since the ‘breadth-first’ procedure \[12\] is very inefficient with respect to computer
time and memory (even for a few system states only), our algorithm bases on the ‘depth-first’ procedure \[12\]. The suitability and correctness of this new numerical method is illustrated by an example concerning Brownian motion.

Section V summarizes the results of the paper and discusses further fields of application.

II. OCCURRENCE PROBABILITIES OF PATHS

Let \( \mathcal{M} = \{1, 2, \ldots, N\} \) be the discrete set of possible states of the considered stochastically behaving system. Moreover, let \( w(j|i) \) represent the transition rate (i.e. the transition probability per unit time) for state changes from state \( i \in \mathcal{M} \) to state \( j \neq i \). In the following we will assume that \( w(j|i) \) is time-independent. Then, the master equation of the corresponding Markovian stochastic process reads

\[
\frac{d}{dt} P(i, t) = \sum_{j=1}^{N} \left[ w(i,j)P(j, t) - w(j|i)P(i, t) \right],
\]

where \( P(i, t) \) denotes the probability of the system to be in state \( i \) at the time \( t \).

Alternatively, we can consider the associated random walk. Let \( i_0 \) be the state of the system at the initial time \( t_0 \), and let \( t_l \) denote the transition times at which the system changes its state from \( i_{l-1} \) to \( i_l \). The corresponding stochastic time series up to time \( t \) with \( t_n \leq t < t_{n+1} \) is

\[
(i_0, t_0) \rightarrow (i_1, t_1) \rightarrow \ldots \rightarrow (i_n, t_n)
\]

and can be numerically generated by means of a Monte-Carlo simulation method \[10\].

If we are interested in the statistical properties of longitudinal (trajectory-related) quantities belonging to the considered stochastic process, we need to evaluate a large number of time series with respect to certain characteristics. This is connected with a considerable computational effort. However, if one is not interested in the respective times \( t_l \) at which the single transitions take place, but only in the path (sequence of states) which the system takes, this effort can be very much simplified.
In order to illustrate this, let
\[ C_n := i_0 \rightarrow i_1 \rightarrow \ldots \rightarrow i_n \tag{3} \]
be the path which the system takes up to the time \( t \). Note that \( w(i_l|i_{l-1})d\tau_{l-1} \) is the probability of changing from state \( i_{l-1} \) to \( i_l \) between time \( t_l \) and \( t_l + d\tau_{l-1} \) and that \( e^{-w_l\tau_l} \) with the departure rate (overall transition rate)
\[ w_l := \sum_{(i \neq i_l)}^N w(i|i_l) \tag{4} \]
is the (survival) probability with which the system stays in state \( i_l \) for a time interval (the survival time) \( \tau_l := t_{l+1} - t_l > 0 \) \( (\tau_n := t - t_n) \). Therefore, multiplying the probability \( P(i_0, t_0) \) of the initial state \( i_0 \) with the survival probabilities as well as the probabilities of the transitions involved and, afterwards, integrating with respect to the possible survival times \( \tau_l \), we obtain the occurrence probability \( P(C_n, \tau) \) with which the system takes the path \( C_n \) up to the time \( t := t_0 + \tau \) \[ P(C_n, \tau) = e^{-w_n\tau_n} \int_0^{\infty} d\tau_{n-1} w(i_n|i_{n-1})e^{-w_{n-1}\tau_{n-1}} \ldots \int_0^{\infty} d\tau_0 w(i_1|i_0)e^{-w_0\tau_0} \delta\left(\sum_{l=0}^n \tau_l - \tau\right) P(i_0, t_0). \tag{5} \]
Here, the \( \delta \)-function guarantees that the survival times sum up to the available time \( \tau \). Inserting its Laplace-representation
\[ \delta\left(\sum_{l=0}^n \tau_l - \tau\right) = \frac{1}{2\pi i} \int_{-c-i\infty}^{-c+i\infty} du \exp\left[u \left(\sum_{l=0}^n \tau_l - \tau\right)\right] \tag{6} \]
with a suitable constant \( c > 0 \) and carrying out the integrations with respect to the survival times \( \tau_l \), we finally obtain the formula \[ P(C_n, \tau) = \frac{1}{2\pi i} \int_{-c-i\infty}^{-c+i\infty} du \frac{e^{-u\tau}}{\prod_{l=0}^n (w_l - u)} w(C_n) P(i_0, t_0), \tag{7} \]
where we have introduced the abbreviation
\[ w(\mathcal{C}_n) := \begin{cases} 
\delta_{ii_0} & \text{if } n = 0 \\
\prod_{l=1}^{n} w(i_l|i_{l-1}) & \text{if } n \geq 1 
\end{cases} \] (8) 

with the Kronecker function \( \delta_{ij} \).

It is obviously much easier to calculate formula (7) than to count and evaluate a huge number of stochastic trajectories. However, for this calculation we must first evaluate the integral by means of the residue theorem. If all departure rates \( w_l \) are identical (i.e. \( w_l \equiv w \)), we get

\[ P(\mathcal{C}_n, \tau) = \frac{\tau^n}{n!} e^{-w\tau} w(\mathcal{C}_n) P(i_0, t_0), \] (9)

whereas we obtain

\[ P(\mathcal{C}_n, \tau) = \sum_{k=0}^{n} f_n(w_k, \tau) w(\mathcal{C}_n) P(i_0, t_0) \] (10)

with

\[ f_n(u, \tau) := \frac{e^{-u\tau}}{n! \prod_{l=0}^{n} (w_l - u)} \] (11)

if the departure rates \( w_l \) are pairwise different from each other. However, serious problems arise in deriving a general result for cases where different departure rates \( w_l \) occur, but some of them multiple times.

### III. DEGENERATE PATHS

In the following, we will call paths for which some departure rate \( w_l \) occurs twice or more often degenerate paths. Note that all paths with repeatedly visited states are degenerate!

For this reason let \( m_l \equiv m(w_l) \) be the multiplicity of the departure rate \( w_l \) in path \( \mathcal{C}_n \) \((l \in \{0,1,\ldots,n\})\). Then we have the relation \( \sum_{w_l} m_l = n + 1 \), and formula (7) can be rewritten in the form
\[ P(C_n, \tau) = -\frac{1}{2\pi i} \int_{-c-i\infty}^{-c+i\infty} \frac{e^{-u\tau} w(C_n)}{\prod_{w_l} (w_l - u)^{m_l}} P(i_0, t_0), \quad (12) \]

where we have changed the integration direction. Applying the residue theorem, we obtain

\[ P(C_n, \tau) = -\sum_{w_k} \frac{1}{(m_k - 1)!} \frac{\partial^{m_k - 1}}{\partial u^{m_k - 1}} \left( \frac{e^{-u\tau} (-1)^{m_k}}{\prod_{w_l} (w_l - u)^{m_l}} \right) \bigg|_{u = w_k} w(C_n) P(i_0, t_0) \]

\[ = \sum_{k=0}^{n} \frac{(-1)^{m_k + 1}}{m_k} f_n^{(m_k - 1)}(w_k, \tau) w(C_n) P(i_0, t_0), \quad (13) \]

where \( f_n^{(m)}(u, \tau) := \frac{\partial^m f_n(u, \tau)}{\partial u^m} \). Introducing the functions

\[ p_m(u, \tau) := \frac{(-1)^{m+1}}{f_n(u, \tau)m!} f_n^{(m-1)}(u, \tau) \]

we can represent (13) in the form

\[ P(C_n, \tau) = \sum_{k=0}^{n} p_m(w_k, \tau) f_n(w_k, \tau) w(C_n) P(i_0, t_0). \quad (15) \]

For \( p_m(u, \tau) \) we can derive the recursive relation

\[ p_m(u, \tau) = -\frac{1}{m f_n(u, \tau)} \frac{\partial}{\partial u} \left[ f_n(u, \tau) \frac{(-1)^m}{f_n(u, \tau)(m-1)!} f_n^{(m-2)}(u, \tau) \right] \]

\[ = -\frac{1}{m} \left[ \left( \sum_{l=0}^{n} \frac{1}{w_l - u - \tau} \right) p_{m-1}(u, \tau) + \frac{\partial}{\partial u} p_{m-1}(u, \tau) \right]. \quad (16) \]

This shows that \( p_m(u, \tau) \) is a polynomial in \( \tau \) of order \( m - 1 \) since a comparison with (10) provides us with \( p_1(u, \tau) = 1 \). However, formula (16) is not very suitable for a numerical implementation or a derivation of the explicit form of \( p_m(u, \tau) \). In order to evaluate (14), we try to find the \((m-1)\)st derivative \( f_n^{(m-1)} \) of \( f_n \). Since we would not succeed with the usual procedure, we need a trick and take the ‘detour’ over the first derivative

\[ f_n^{(1)}(u, \tau) = \left( \sum_{l=0}^{n} \frac{1}{w_l - u - \tau} \right) f_n(u, \tau). \quad (17) \]

The \((n_0 - 1)\)st derivative of this relation can now be determined as usual:

\[ f_n^{(n_0)}(u, \tau) = \sum_{l=0}^{n_0-1} \binom{n_0 - 1}{l} \frac{\partial^l}{\partial u^l} \left( \sum_{l=0}^{n} \frac{1}{w_l - u - \tau} \right) \frac{\partial^{n_0-1-l}}{\partial u^{n_0-1-l}} f_n(u, \tau) \]

\[ = \sum_{n_1=0}^{n_0-1} \frac{(n_0 - 1)!}{n_1!} g^{(n_0-n_1)}(u, \tau) f_n^{(n_1)}(u, \tau). \quad (18) \]
Here, we have used the definitions \( n_1 := n_0 - 1 - l \) and

\[
g^{l+1}(u, \tau) := \frac{1}{l!} \frac{\partial^l}{\partial u^l} \left( \sum_{\substack{i=0 \atop (w_i \neq u)}}^n \frac{1}{w_i - u} - \tau \right) = \sum_{i=0 \atop (w_i \neq u)}^n \frac{1}{(w_i - u)^{l+1}} - \tau \delta_{l0}. \tag{19}\]

Formula (18) obviously allows to express the derivatives \( f_n^{(n_0)}(u, \tau) \) in terms of derivatives \( f_n^{(n_1)}(u, \tau) \) of lower order \( n_1 < n_0 \). Therefore, we insert for \( f_n^{(n_1)}(u, \tau) \) again relation (18), etc. After a couple of iterations this procedure leads, for \( n_0 \geq 1 \), to

\[
f_n^{(n_0)} = (n_0 - 1)! \left( g_n^{(n_0)} f_n^{(0)} + \sum_{n_1=1}^{n_0-1} g_n^{(n_0-n_1)} \left( g_n^{(n_1)} f_n^{(0)} + \sum_{n_2=1}^{n_1-1} g_n^{(n_1-n_2)} \right) \right) \times \left( g_n^{(n_2)} f_n^{(0)} + \cdots \right), \tag{20}\]

where we have separated the terms containing \( f_n^{(0)}(u, \tau) = f_n(u, \tau) \) from the rest and omitted the arguments \((u, \tau)\) of the functions. In equation (20) we have to apply the convention \( \sum_{n=l}^{l_2} (...) := 0 \) if \( l_2 < l_1 \) (as will be shown by Fig. 1). Consequently, the term with the \( n_{n_0} \)th derivative \( f_n^{(n_{n_0})} \) does not give a contribution to \( f_n^{(n_0)} \), since \( \text{max}(n_1) = n_0 - 1 \), \( \text{max}(n_2) = n_1 - 1 = n_0 - 2 \), \ldots, \( \text{max}(n_{n_0}) = n_0 - n_0 = 0 \). Therefore, we obtain the final result

\[
p_m = \frac{(-1)^{m+1}}{m(m-1)} \left( g^{(m-1)} + \sum_{n_1=1}^{m-2} g^{(m-1-n_1)} \left( g^{(n_1)} + \sum_{n_2=1}^{n_1-1} g^{(n_1-n_2)} \right) \right) \times \left( \cdots + \sum_{n_{n_0-1}=1}^{n_{n_0-2} \cdots - 1} g^{(n_{n_0-2} \cdots - n_{n_0-1})} \left( g^{(n_{n_0-1})} \right) \right) \right), \tag{21}\]

for \( m \geq 2 \). The correctness of this formula is illustrated by Fig. 1.

**IV. NUMERICAL EVALUATION OF THE ‘CONTRACTED PATH-INTEGRAL SOLUTION’**

In Refs. [11,4] it has been proved that the solution \( P(i, t) \) of master equation (1) can be represented in the form

\[
P(i, t_0 + \tau) = \sum_{n=0}^{\infty} \sum_{C_n} P(C_n, \tau) := \sum_{n=0}^{\infty} \sum_{i_{n-1}=1}^{N} \sum_{i_{n-2}=1}^{N} \cdots \sum_{i_0=1}^{N} P(i_0 \rightarrow i_1 \rightarrow \ldots \rightarrow i_n, \tau) \tag{22}\]
with $i_n := i$. According to (22), the probability $P(i, t_0 + \tau)$ to find state $i$ at the time $t = t_0 + \tau$ is given as the sum over the occurrence probabilities $P(C_n, \tau)$ of all paths $C_n$ which have an arbitrary length $n$ but have led to state $i_n = i$ within the time interval $\tau$.

When (22) is evaluated numerically, one must restrict the summation to a finite number of relevant paths. Here, we will define a path $C_n$ to be relevant if it fulfills the condition

$$|\tau - \langle \tau \rangle_{C_n}| \leq a \sqrt{\theta_{C_n}},$$

(23)

where $\langle \tau \rangle_{C_n} := \sum_{k=0}^{n} 1/w_k$ is the mean value of the occurrence times of path $C_n$ and $\theta_{C_n} := \sum_{k=0}^{n} 1/(w_k)^2$ their variance (cf. [11,4]). The parameter $a$ is a measure for the accuracy of the above outlined approximation. According to our experience, $a \lesssim 3$ usually guarantees a reconstruction of 99% of the probability distribution $P(i, t)$ which can be checked by means of the normalization condition $\sum_{i=1}^{N} P(i, t) = 1$.

An efficient implementation of the approximate ‘contracted path-integral solution’ on a serial computer bases on the path-search procedure ‘depth-first’ [12]. After each step of this standard procedure which extends the previous path $C_n$ by an additional state $i_{n+1}$ (leading to the path $C_{n+1} := C_n \rightarrow i_{n+1}$) the following quantities are calculated and stored in a so-called ‘linked list’ [12]:

$$\langle \tau \rangle_{C_{n+1}} := \langle \tau \rangle_{C_n} + \frac{1}{w_{n+1}}, \quad \theta_{C_{n+1}} := \theta_{C_n} + \frac{1}{(w_{n+1})^2}, \quad w(C_{n+1}) := w(C_n) w(i_{n+1}|i_n),$$

(24)

and

$$f_{n+1}(w_k, \tau) := \frac{f_n(w_k, \tau)}{w_{n+1} - w_k} \quad \text{for} \quad k = 0, \ldots, n \quad \text{with} \quad w_k \neq w_{n+1}.$$  

(25)

If condition (23) is fulfilled, the occurrence probability $P(C_n, \tau)$ is calculated according to formula (15) and added to $P(i_{n+1}, \tau)$. Afterwards, the next step of the ‘depth-first’ procedure is carried out.

The path $C_{n+1}$ is not further extended if $\langle \tau \rangle_{C_{n+1}} - a \sqrt{\theta_{C_{n+1}}} > \tau$. This implies that the path $C_{n+1}$ and all longer paths $C_{n+1} \rightarrow i_{n+2} \rightarrow \ldots$ which include $C_{n+1}$ as subpath are
irrelevant. Then, the procedure traces back its path $C_{n+1}$ one step before it tries to extend the previous path $C_n$ by another state $i'_{n+1} = i_{n+1} + 1$.

Figure 2 illustrates the approximate ‘contracted path-integral solution’ for the example of Brownian motion which is characterized by nearest-neighbor transitions with the following transition rates:

$$w(j|i) = \begin{cases} D & \text{if } |j - i| = 1 \\ 0 & \text{otherwise.} \end{cases}$$  \hspace{1cm} (26)

V. SUMMARY AND FIELDS OF APPLICATION

In this paper we were able to derive an analytical formula for the occurrence probability of arbitrary paths including the complex case of repeatedly visited states and/or equal departure rates. This allows a numerical implementation of the ‘contracted path-integral solution’ of the discrete master equation. Other kinds of path-integral formalisms have been developed for chaotic mappings [14], the Schrödinger equation [15], the Fokker-Planck equation [8,16], and also the master equation [16].

The importance of the occurrence-probability formula goes far beyond the new solution method for the master equation. It considerably simplifies the evaluation of the trajectories related with random walk models. Therefore, a simulation program for the calculation of the occurrence probabilities of paths, the path-integral solution, most probable paths, and first-passage times has recently been developed at the University of Stuttgart. It is expected to be a useful tool for investigations in a number of current research fields concerning different types of random walks [17], noise-induced transitions [18], first-passage time problems [19], percolation [20], critical behavior [21], and diffusion in disordered or fractal media [22].
REFERENCES

[1] C. W. Gardiner, Handbook of Stochastic Methods (Springer, Berlin, 1985).

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

[3] J. W. Haus and K. W. Kehr, Physics Reports 150, 263 (1987).

[4] D. Helbing, Quantitative Sociodynamics. Stochastic Methods and Models of Social Interaction Processes (Kluwer Academic, Dordrecht, 1995).

[5] I. Oppenheim, K. E. Schuler, and G. H. Weiss, editors, Stochastic Processes in Chemical Physics: The Master Equation (MIT Press, Cambridge, MA, 1977).

[6] L. Arnold and R. Lefever, editors, Stochastic Nonlinear Systems in Physics, Chemistry and Biology (Springer, Berlin, 1981).

[7] W. Weidlich and G. Haag, Concepts and Models of a Quantitative Sociology (Springer, Berlin, 1983). W. Weidlich, Physics Reports 204, 1 (1991). D. Helbing, Physica A 193, 241 (1993).

[8] H. Risken, The Fokker-Planck Equation (Springer, Berlin, 1989). R. L. Stratonovich, Topics in the Theory of Random Noise, Vols. 1, 2 (Gordon and Breach, New York, 1963, 1967).

[9] C. W. Gardiner, Quantum Noise (Springer, Berlin, 1991). P. Goetsch and R. Graham, Annalen der Physik 2, 706 (1993); P. Goetsch and R. Graham, in: Quantum Optics VI, ed. by D. F. Walls et al. (Springer, Berlin, 1994).

[10] K. Binder, editor, Monte Carlo Methods in Statistical Physics (Springer, Berlin, 1979).

[11] D. Helbing, Physics Letters A 195, 128 (1994).

[12] H. Schildt, C: The Complete Reference (McGraw Hill, Berkeley, 1990).
[13] N. Empacher, Die Wegintegrallösung der Mastergleichung (PhD thesis, University of Stuttgart, 1992).

[14] C. Beck and T. Tél, J. Phys. A 28, 1889 (1995).

[15] B. Felsager, Geometry, Particles and Fields (Odense University Press, 1981). R. P. Feynman, Reviews of Modern Physics 20, 367 (1948).

[16] H. Haken, Zeitschrift für Physik B 24, 321 (1976).

[17] M. Müller and W. Paul, Europhysics Lett. 25, 79 (1994). F. D. A. A. Reis, J. Phys. A 28, 3851 (1995).

[18] S. V. Lawande and Q. V. Lavande, Modern Phys. Lett. B 9, 87 (1995). G. Zumofen and J. Klafter, Chemical Physics Letters 219, 303 (1994). A. Hamm, T. Tél, and R. Graham, Phys. Lett. A 185, 313 (1994). Y.-C. Lai, T. Tél, and C. Grebogi, Phys. Rev. E 48, 709 (1993).

[19] P. Pechukas and P. Hänggi, Phys. Rev. Lett. 73, 2772 (1994). J. Masoliver and J. M. Porrà, Phys. Rev. Lett. 75, 189 (1995). M. O. Vlad and M. C. Mackey, Phys. Lett. A 203, 292 (1995).

[20] N. Inui, Phys. Lett. A 184, 79 (1993). S. Fujiwara and F. Yonezawa, Phys. Rev. Lett. 74, 4229 (1995).

[21] S. Zang, Q. Fan, and E. Ding, Phys. Lett. A 203, 83 (1995). B. Yu and D. A. Browne, Phys. Rev. E 49, 3496 (1993). I. Campos and A. Taracón, Phys. Rev. E 50, 91 (1993).

[22] R. Dasgupta, T. K. Ballabh, and S. Tarafdar, Phys. Lett. A 187, 71 (1994). S. Mukherjee, H. Nakanishi, and N. H. Fuchs, Phys. Rev. E 49, 5032 (1994).

[23] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, Numerical Recipes in C (Cambridge University Press, Cambridge, 1992).
FIGURES

FIG. 1. Comparison of the occurrence probability $P(C_n, \tau)$ calculated via formulas (15), (11) and (21) for the case $w_l \equiv w$, $m_l = n + 1 = 10$ (○) with the occurrence probability of the same path $C_n$ according to (9) (—).

FIG. 2. Comparison of the approximate path-integral solution for Brownian motion (bars) with the numerical result of Runge-Kutta integration [23] for the corresponding master equation (grid). The diffusion coefficient was assumed to be $D = 1$, and the accuracy parameter was set to $a = 3$ so that the path-integral method reconstructs about 99% of the exact probability distribution.