

# UC Berkeley

## UC Berkeley Previously Published Works

### Title

The least variable phase type distribution is erlang

### Permalink

<https://escholarship.org/uc/item/7w51x4rb>

### Journal

Stochastic Models, 3(3)

### ISSN

1532-6349

### Authors

David, Aldous

Larry, Shepp

### Publication Date

1987

### DOI

10.1080/15326348708807067

Peer reviewed

**The least variable phase type distribution is Erlang**

*David Aldous*

U C Berkeley

*Larry Shepp*

AT&T Bell Laboratories

**Abstract**

Let  $X_t, t \geq 0$  be any continuous-time Markov process on states  $0, 1, \dots, n$  where  $X_0 = n$  and  $T_0$  is the time to reach 0 which is absorbing. We prove that  $T_0$  is most nearly constant in the sense of minimizing the coefficient of variation  $\text{var}(T_0) / (E T_0)^2$  over all transition matrices  $P_{ij}$  and exponential delay parameters  $\lambda_i$  in each state when  $P_{ii-1} = 1, i = n, n-1, \dots, 1$  and  $\lambda_i \equiv \text{constant}$ . The latter chain is Erlang's process on  $n$  fictitious states and has been used to show that an arbitrary semi-Markov process can be approximated by a Markov process. It has been a long-open problem since the work of Kendall, Cox, and others to try to improve on Erlang's scheme by generalizing the transition structure of  $X$ , i.e. adding loops, twists, and turns in order to make the overall waiting time have smaller coefficient of variation. We destroy this hope by showing at last that Erlang's original method is not improvable.

Our proof is simple and elegant and is a nice example of the power of martingales; it seems intractible without them.

**§1. The inequality**

R. P. Kurshan [4] has introduced the notion of tensor product of Markov processes  $X_1$  and  $X_2$  which (in a special case) is the ordinary product. The product of  $X_k$  on states  $i_k \in I_k, k = 1, 2$  has states  $(i_1, i_2) \in I_1 \times I_2$  and if  $\tau_k$  is the exit time from state  $i_k$  of  $X_k, k = 1, 2$  then a transition is made at time  $\tau = \min(\tau_1, \tau_2)$  from  $(i_1, i_2)$  to  $(j_1, i_2)$  if  $\tau_1 < \tau_2$  or to  $(i_1, j_2)$  if  $\tau_1 > \tau_2$ . Here  $j_k$  is the new state of  $X_k$ . This product construction results in a Markov process,

but if  $X_k$  are only semi-Markov processes (i.e. if the delays in each state  $i$  are not exponential but follow some distribution  $F_i$ ) then this product construction fails in general to be even semi-Markov. Thus to perform the calculations needed to analyze the behavior of first passage times for a product of semi-Markov processes unwieldy integral equations must be solved. To avoid these it is (apparently) required to first approximate each component semi-Markov process by a Markov process. Kurshan [4] needs to perform simulations and since his component state spaces are rather large, he has to do this Markov approximation very efficiently, i.e. he wants to use a minimum number of fictitious states. In [4] the non-exponential semi-Markov delay  $\theta$  in each state is well-modelled by a so-called delayed exponential, where  $\delta$  and  $\lambda$  depend on the state,

$$(1.1) \quad P(\theta \geq t) = e^{-(t-\delta)^+/\lambda}, \quad t \geq 0$$

so that the process waits a fixed time  $\delta$  and then an exponential time with mean  $\lambda$ . Kurshan's definition of tensor product allows more generally for the transition of  $X_1$  to also depend on the state of  $X_2$ . But the special case we have mentioned where it does not already illustrates the problem and the need to preserve Markovianness, i.e. to have exponential delay times. This is no doubt similar to the reasoning that prompted Erlang and others to invent Markov chains approximating semi-Markov ones.

There is an extensive literature on approximating semi-Markov processes by Markov processes including [1-5]. Neuts [5] gives a survey and extended discussion of algorithms for choosing a so-called "phase type" distribution, which is a distribution of the delay time of a Markov process on a fixed number  $n+1$  of

states until absorption in one of the states, to approximate an arbitrary delay distribution.

Erlang [2] gave a simple class of (phase type) delay distributions by considering a Markov process  $X_t, t \geq 0$  on states  $0, 1, \dots, n$  where the exponential delay in state  $i > 0$  has fixed mean  $\lambda$  and the process moves deterministically with constant expected delay through the states,

$$(1.2) \quad X_0 = n, \quad P_{ii-1} = 1, \quad i = n, \dots, 1.$$

The time  $T_0$  until absorption at 0 then has Laplace transform

$$(1.3) \quad E e^{-sT_0} = \left( \int_0^\infty e^{-st} e^{-t/\lambda} \frac{dt}{\lambda} \right)^n = \left( \frac{1}{1 + \lambda s} \right)^n$$

since  $T_0 = \frac{\delta}{n} \sum_{i=1}^n e_i$  where  $e_i$  are i.i.d unit exponentials so the mean and variance of  $T_0$  are

$$(1.4) \quad E T_0 = n \lambda, \quad \text{var}(T_0) = n(n+1)\lambda^2 - (n\lambda)^2 = n\lambda^2$$

and the coefficient of variation is

$$(1.5) \quad \frac{\text{var}(T_0)}{(E T_0)^2} = \frac{n(n+1)\lambda^2 - (n\lambda)^2}{(n\lambda)^2} = \frac{1}{n}$$

Since  $(1 + \delta s/n)^{-n} \rightarrow e^{-\delta s}$  as  $n \rightarrow \infty$ , and  $e^{-\delta s}$  is the Laplace transform of the constant  $\delta$ , we see that the convergence as  $n \rightarrow \infty$  of  $T_0$  to a fixed delay  $\delta$  could be achieved with  $\lambda = \delta/n$  but the convergence of Erlang's scheme (to the  $\delta$ -distribution at  $T_0$ ) is rather slow, only  $O(1/n)$  since  $\text{var}(T_0) = \delta^2/n$ . Nevertheless we will see that this is best possible. Note that the question of choosing a fictitious semi-markov chain to minimize the coefficient of variation of the delay-time is implicit in [1,3,5] and is very natural.

To approximate the delayed exponential (1.1) with Erlang's method one might first generate an approximate fixed delay  $\delta$  with  $n$  fictitious states and then when the process reaches state 0 after time  $T_0$  one more state would be entered after an exponential delay with mean  $\lambda$ . Thus  $n+1$  fictitious states would be needed for each state of the original semi-Markov process for each component. It may be that there is no better way to approximate a delayed exponential.

Can the delayed exponential be approximated more efficiently by using a more involved  $P_{ij}$  matrix with variable  $\lambda_i$  as suggested by the extensive literature and discussion in [5]? We don't know the answer to this question. Certainly there are distributions which are efficiently approximated by using more general phase-type distributions than the simple Erlang ones (eg. the delay of any phase-type distribution is an exact realization of itself!). However, fixed delays are important in themselves for the general theory. It is only through the use of mixtures of fixed delays (and Erlang distributions) that it can be shown [1] that any distribution can be approximated by phase-type distributions. The fact that a fixed delay or a delayed exponential is difficult to approximate with a phase-type distribution has been observed [5, p. 79]. Kendall [3] and Cox [1] as well as Neuts [5] seem all to be unable to quantify this statement in terms of minimizing the coefficient of variation over phase type distributions. Although coefficient of variation is discussed by the above authors, apparently they never actually conjectured or explicitly discussed the following result.

*Theorem.* Let  $n$  be fixed and consider the class of all phase distributions, i.e. those of the delay times  $T_0$  to reach state 0 by a Markov process  $X_t, t \geq 0$ , on states

$0, 1, \dots, n$  starting in state  $n$  with arbitrary transition matrix  $P_{ij}$  and exponential delay with mean  $\lambda_i$  in state  $i$ . Then the coefficient of variation of  $T_0$ ,

$$(1.6) \quad \frac{\text{var}(T_0)}{(ET_0)^2} \geq \frac{1}{n}$$

with equality if and only if  $X$  is the Erlang process (1.2).

**§2. Proof of the theorem**

We assume that  $X_t, t \geq 0$  is a Markov process on states  $0, 1, \dots, n$ , absorbing at 0, starting at  $n$ . Thus set  $T_0 =$  time to hit state 0 and

$$(2.1) \quad Y_t = h(X_t) + \min(t, T_0) - h(X_0), t \geq 0$$

where  $h$  is the expected time to hit state 0 starting in state  $i$ ,

$$(2.2) \quad h(i) = E_i T_0 .$$

Since for  $i \neq 0, E[h(X_{t+\delta}) - h(X_t) | X_t = i] = -\delta + o(\delta)$  as  $\delta \rightarrow 0$ , it is easy to show that  $Y_t, t \geq 0$  is a martingale (with respect to the  $\sigma$ -fields generated by  $X$ ).

Now let  $S_t$  be the sum of the squares of the jumps of  $Y$  up to  $t$ ,

$$(2.3) \quad S_t = \sum_{s < t} (Y_s - Y_{s-})^2 .$$

Then

$$(2.4) \quad EY_t^2 = ES_t$$

since

$$(2.5) \quad E(Y_{t+u} - Y_t)^2 = EY_{t+u}^2 - 2EY_{t+u}Y_t + EY_t^2 = EY_{t+u}^2 - EY_t^2$$

and so if  $0 < t_1 < \dots < t_r < t, EY_t^2 = E \sum_{j=1}^r (Y_{t_j} - Y_{t_{j-1}})^2 \rightarrow ES_t$  as the partition

refines since

$$(2.6) \quad (Y_{t+u} - Y_t)^2 = O(u^2) \text{ as } u \rightarrow 0 \text{ unless there is a jump in } (t, t+u].$$

Since

$$(2.7) \quad Y_{T_0} = T_0 - E_n T_0$$

we have from (2.7), (2.4), and (2.3)

$$(2.8) \quad \text{var}(T_0) = EY_{T_0}^2 = ES_{T_0} = E\sum_s (h(X_s) - h(X_{s-}))^2.$$

Now consider a permutation  $\sigma$  of  $\{0, 1, \dots, n\}$  such that  $h(\sigma(i))$  is non-decreasing. Let  $n^* = \sigma^{-1}(n)$ . We assert that, for any path  $n = i_0, i_1, \dots, i_L = 0$ ,

$$(2.9) \quad \sum_{u=1}^L (h(i_u) - h(i_{u-1}))^2 \geq \sum_{i=1}^{n^*} (h(\sigma(i)) - h(\sigma(i-1)))^2.$$

For let  $I_u = \{j: \min(\sigma^{-1}(i_u), \sigma^{-1}(i_{u-1})) < j \leq \max(\sigma^{-1}(i_u), \sigma^{-1}(i_{u-1}))\}$ . Then

$$(h(i_u) - h(i_{u-1}))^2 \geq \sum_{j \in I_u} (h(\sigma(j)) - h(\sigma(j-1)))^2$$

by monotonicity. A moment's thought shows  $UI_u$  contains  $\{1, \dots, n^*\}$ , and (2.9) follows.

Now

$$\begin{aligned} \text{var}(T_0) &\geq \sum_{i=1}^{n^*} (h(\sigma(i)) - h(\sigma(i-1)))^2 \text{ by (2.8) and (2.9)} \\ &\geq \left(\sum_{i=1}^{n^*} h(\sigma(i)) - h(\sigma(i-1))\right)^2 \frac{1}{n^*} \text{ by Schwarz's inequality} \\ &\geq h^2(n) \frac{1}{n} = (E_n T_0)^2 \frac{1}{n} \end{aligned}$$

which proves the theorem. Equality holds only if no branching occurs and if  $h(i) - h(i-1)$  is a constant, i.e.  $\lambda_i \equiv \lambda$ .

#### Acknowledgement

We are grateful to R. P. Kurshan for mentioning the problem which led to his conjecture (1.6), and to J. E. Mazo and J. A. Reeds for discussions.

#### References

1. Cox, D. R. (1955) The use of complex probabilities in the theory of stochastic processes, Proc. Canad. Phil Soc, 51, 313-319.
2. Erlang, A. K. (1917-18) Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges, The Post Office Electrical Engineer's Journal, 10, 189-197.
3. Kendall, D. G. (1953) Stochastic processes occurring in the theory of queues and their analysis by the method of the embedded Markov chain, Ann Math Stat 24, 338-354.
4. Kurshan, R. P. and I. Gertner (1986) Stochastic analysis of coordinating systems, unpublished manuscript, to appear.
5. Neuts, M. F. (1981) Matrix-Geometric Solutions in Stochastic Models, an Algorithmic Approach, Johns Hopkins U. Press, Baltimore.

Received: 12/04/1986

Revised: 6/15/1987

Recommended by M. F. Neuts, Editor





4  
1  
3  
2