WHO INTRODUCED THE BIRTH- AND DEATH TECHNIQUE?

J.W. COHEN

MATHEMATICAL INSTITUTE, UNIVERSITY OF UTRECHT
UTRECHT, THE NETHERLANDS

ABSTRACT

The technique of Birth- and Death equations plays a vital role in the theoretical investigations concerning the performance of models encountered in teletraffic and computer systems. The Danish mathematician A.K. Erlang is usually credited for being the originator of this analytic technique. His ideas have been of paramount importance in the development of the analysis of the basic teletraffic models. However, he seems never to have used explicitly the Birth- and Death technique, and it may be questioned whether he applied this type of analysis.

In Erlang's papers no explicit indications are available that he has applied the Birth- and Death technique. So the question arises who developed this type of analytical approach. It seems difficult to give a definite answer. In the early twenties Fry and Molina undoubtedly had fully mastered this technique, but also in the researches of the biological statisticians McKendrick and Yule basic principles of this analytic technique can be found.

1. INTRODUCTION

The technique of Birth- and Death equations plays a vital role in the theoretical investigations concerning the performance of models encountered in teletraffic and computer systems. The Danish mathematician A.K. Erlang is usually credited for being the originator of this analytic technique. His ideas have been of paramount importance in the development of the analysis of the basic teletraffic models. However, he seems never to have used explicitly the Birth- and Death technique and it may be questioned whether he has applied this type of analysis.

In the present study we intend to throw some light on this question. Herefore we shall start with a derivation of the stationary distribution of the number of busy servers in an M/G/\infty model, a problem already considered by Erlang. Our derivation, which is based on geometrical considerations and which uses an idea stemming from D. van Dantzig, is extremely simple and provides a sharp insight in several phenomena such as insensitivity, reversibility, and the use of local and global properties of sample functions in the deduction of mathematical relations, cf. [2].

Comparison of the present derivation with the arguments as used by Erlang does strongly conjecture that Erlang mainly uses the global property of the sample functions, viz. "upcrossing intensity is equal to downcrossing intensity".

If as it seems very likely Erlang did not know the Birth- and Death technique, so the question arises who developed this type of analytical approach. A definite answer seems to be difficult to give. In the early twenties Fry and Molina undoubtedly had fully mastered this technique, but also in the researches of the biological statisticians McKendrick and Yule basic principles of this analytical technique can be found. It may be conjectured that Fry and Molina on the one side, McKendrick and Yule on the other side independently developed the principals of the Birth- and Death technique; they all used the technique of formulating equations between probabilities by means of recursive relations, a type of argumentation already applied by P.S. Laplace in his famous book: Théorie Analytique des Probabilités.

2. ON A GEOMETRIC MODEL OF M/G/\infty

We shall start with a Poisson point process with intensity \lambda on the horizontal line. Its points will be marked as "1" or "2" points, independently of each other with probabilities p_1 and p_2, p_1 + p_2 = 1. To a "1" point a line segment of length \tau_1 is assigned, similarly to a "2" point one with length \tau_2; these line segments are plotted vertically, see figure 1, and their end points constitute a point process in the plane.

It is readily verified that the number of points contained in any rectangle (a,b) x (\alpha,\beta) has a Poisson distribution with intensity depending on (a,\alpha) and on the position of a and b with respect to \tau_1 and \tau_2. Further it is evident that the number of points in disjoint rectangles are independent. By applying Carathéodory's extension theorem (cf. [1]) it follows that the distribution...
of the number of points in any measurable set in the plane is Poisson, i.e. the point process in the plane is a nonhomogeneous Poisson process. In particular it follows that the number of points in the infinite wedge of angle $\phi=45^\circ$ to the vertical axis and with its top at the point $t$ has a Poisson distribution with intensity $\lambda(p_1 T_1 + p_2 T_2)$, as a simple calculation shows.

Let us next rotate every point of the point process in the plane clockwise by $90^\circ$ around its projection on the horizontal line. The point process on the line so obtained is the sum of two independent Poisson processes with intensities $\lambda p_1$ and $\lambda p_2$, and as such is again a Poisson process with intensity $\lambda$. Obviously, the latter process may be regarded as the departure process on the line so obtained is the sum of two independent Poisson processes with intensities $\lambda p_1$ and $\lambda p_2$, and as such is again a Poisson process with intensity $\lambda$. Hence their number has a Poisson distribution. Hence we reach the conclusion that the wedge at level $T_2$ is equal to $M/G/N$ model, which includes the $M/D/N$ and $M/G/\infty$ systems. In the infinite wedge of angle $45^\circ$, the sloping strip between the sides of the wedges contains just one point is of order $\lambda t + o(\lambda t)$. To calculate the probability that the sloping strip between the sides of the wedges contains just one point, note that the probability of having $K$ points in the wedge at $t$ with $n$ points at level $T_2$ is equal to

$$\frac{(\lambda p_2 T_2)^n}{n!} \frac{(\lambda p_1 T_1)^{k-n}}{(k-n)!} e^{-\lambda(p_1 T_1 + p_2 T_2)}.$$

Given that at level $T_2$ there are $n$ points, then they are all uniformly distributed on $[0, T_2]$, and similarly for level $T_1$. Hence the probability of having $K$ points in the wedge at $t$ and no points in the sloping strip is equal to

$$\lambda(p_1 T_1 + p_2 T_2)^k \frac{(\lambda p_1 T_1)^{k-n}}{k!} e^{-\lambda(p_1 T_1 + p_2 T_2)}.$$

Another evident consequence of the argument above is the reversibility of the process, i.e. the stochastic structure of the process formed by the number of busy servers at time $t$ is identical for $t$ increasing or decreasing (if the process is stationary).

3. ERLANG'S PRINCIPLE OF STATISTICAL EQUILIBRIUM.

Next consider two wedges (as in figure 2) a distance $\Delta t$ apart, again with two levels $T_1$ and $T_2$.

Obviously, the probability that the vertical strip formed by the vertical boundaries of the wedges contains just one point is $\lambda \Delta t + o(\lambda \Delta t)$. To calculate the probability that the sloping strip between the sides of the wedges contains just one point, note that the probability of having $K$ points in the wedge at $t$ with $n$ points at level $T_2$ is equal to

$$\frac{(\lambda p_2 T_2)^n}{n!} \frac{(\lambda p_1 T_1)^{k-n}}{(k-n)!} e^{-\lambda(p_1 T_1 + p_2 T_2)}.$$

Given that at level $T_2$ there are $n$ points, then they are all uniformly distributed on $[0, T_2]$, and similarly for level $T_1$. Hence the probability of having $K$ points in the wedge at $t$ and no points in the sloping strip is equal to

$$\lambda(p_1 T_1 + p_2 T_2)^k \frac{(\lambda p_1 T_1)^{k-n}}{k!} e^{-\lambda(p_1 T_1 + p_2 T_2)}.$$

Another evident consequence of the argument above is the reversibility of the process, i.e. the stochastic structure of the process formed by the number of busy servers at time $t$ is identical for $t$ increasing or decreasing (if the process is stationary).

Another evident consequence of the argument above is the reversibility of the process, i.e. the stochastic structure of the process formed by the number of busy servers at time $t$ is identical for $t$ increasing or decreasing (if the process is stationary).

Using the probabilities derived above, the equality of these intensities is expressed by

$$(1) \lambda \Pr(x_k = k) = \frac{k+1}{p_1 T_1 + p_2 T_2} \Pr(x_k = k+1), k=0,1,2,\ldots$$

From this relation it is readily seen that $x_k$ has a Poisson distribution with parameter $\lambda(p_1 T_1 + p_2 T_2)$, the same result as derived in the preceding section.

The principle of statistical equilibrium, which in present day terminology is called "partial balance" obviously applies to the sample functions of the number of busy servers in the $M/G/N$ loss model, which includes the $M/D/N$ and
the $\text{M/M/N}$ case. Here it is readily seen that the upward jump intensity at $x = k$ is equal to

\[
\lambda \Pr(x = k),
\]

but it is rather hard to prove that the downward jump intensity at $x = k$ is equal to $(k+1)/(p_{01} + \rho_{31}) \Pr(x = k+1).$ The finiteness of $N$, the number of servers, complicates the stochastic structure of the process, and the argument used above for $N = \infty$ does not apply. Only if the service times are negative exponentially distributed (the case M/M/N) is a simple deduction possible, because of the memory less properties of this distribution. Erlang’s principle of statistical equilibrium is not questioned, but its application fails according to the difficulty of transforming it into a mathematical equation. It took nearly sixty years before the basic idea of transforming global sample function properties into quantitative relations between probabilities was applied again in queueing analysis, cf. [2], [3], [10].

4. ON THE BIRTH- AND DEATH EQUATION

A stochastic process $\{x(t), t \in (0, \infty)\}$ with a denumerable state space is a Birth- and Death process if the process is Markovian and if in every time interval $t + \Delta t$ only transitions to a neighbouring state are possible (i.e. with probability $1 - \rho(s, \Delta t))$. In this stochastic structure which motivates the formulation of the so called Birth- and Death equations for the transition probabilities $p(s, t, s') = \Pr(x(t) = s' | x(t) = s)$, $s > t$. Obviously, these Birth- and Death equations are based on a local property of the sample functions of the process $x$. For the stationary distribution $\Pr(x = x_\infty, x = 0, 1, 2, \ldots)$ of the M/M/\infty queueing model with arrival rate $\lambda$ and average service time $\beta$ the Birth- and Death equations read

\[
(2) \quad \lambda \Pr(x = 0) + \frac{1}{\beta} \Pr(x = 1) = 0,
\]

\[
(3) \quad \lambda \Pr(x = k) + \frac{k}{\beta} \Pr(x = k+1) = 0, \quad k = 0, 1, 2, \ldots
\]

It is readily seen that (2) is a first "integral" of the system (2) of second order "recurrence" equations.

Actually, the relation (3) is the same as (1), see the preceding section, which has been derived from the global property: "the rate of upcrossings is equal to the rate of downcrossings". In the present day literature Erlang is usually regarded as the first researcher who introduced the technique of Birth- and Death equations in the analysis of stochastic processes. The correctness of this viewpoint is questionable. A careful reading of Erlang’s work (cf. [4]), shows that Erlang has never formulated equations which have the typical structure of the Birth- and Death equations. Erlang is rather sparing in providing explicit mathematical arguments. His most explicit argument occurs on page 141 of [4], and here his arguments are much closer to the use of the global property of up- and downcrossings than to the local property on which the birth- and death equations are based.

The question arises: where do we find in literature the first explicit introduction of the Birth- and Death technique in the analysis of stochastic processes.

In Fry’s important book, cf. [5], published in 1928 the Birth- and Death technique is the standard method in the analysis of telephone traffic models. A similar explicit use of this technique is to be found in Molina’s paper [6] of 1927. Fry’s discussion of kinetic gas theory in his book points to the conjecture that he was strongly influenced in his analysis of stochastic processes by the ideas of kinetic gas theory. Explicit references concerning the origin of the Birth- and Death technique, however, do not occur in his book. There is no doubt, however, that both Molina and Fry completely understood this type of analysis.

The name Birth- and Death process presumably has been introduced by Feller. It raises the conjecture that such processes have been first encountered in medical and biological statistics. For an interesting historical account on this subject see the study [7]. Here a study of McKendrick [8] is mentioned in which the author is concerned with the analysis of a probability distribution related to the phenomenon of phagocytosis. In this study we encounter arguments and equations which have a structure that resembles that of the Birth- and Death analysis. Like Fry’s McKendrick’s reasoning seems to be influenced by the Methods of kinetic gas theory. Next to McKendrick’s work that of Yule needs to be mentioned. The influence of their studies (around 1914-1930) on the development of the theory of stochastic processes in biology and medicine is discussed in [9], and it turns out that both McKendrick and Yule apply a type of analysis which comes very close to that which at present is called the Birth- and Death technique. Whether Molina and Fry, on the one hand, and McKendrick and Yule on the other hand knew of each others researches is not clear. My personal feeling is that Fry and Molina should be credited for the first shaping of the Birth- and Death type of analysis.

REFERENCES


