

Three papers in probability – some notes and recollections

Elja Arjas

Writing a brief commentary on three of Terry Speed's papers in probability brings to mind many memories from a time now almost forty years away. Two of these papers were written while Terry worked as a Lecturer in Sheffield, and during this period my encounters with Terry were very frequent. The third paper was written after Terry had already moved on to Perth.

These were times "when we were very young", and there was a great deal of excitement about new developments in probability. One of the main sources of inspiration was Volume 2 of *Introduction to Probability Theory and its Applications* by Feller [1966], which had come out sixteen years after the publication of Volume 1 [Feller, 1950], and was then followed five years later by an expanded Second Edition. Feller was a master in making probability theory look like it were a collection of challenging puzzles, for which one, if only sufficiently clever, could find an elegant solution by some ingenious trick that actually made the original problem look like it had been trivial. Feller's books offered also a large number of examples leading to potentially important applications. This idea of making probability a tool for practical mathematical modeling was gaining ground in other ways, too. An important initiative in this direction, in 1964, was founding, at the initiative of Joe Gani, of the *Applied Probability* journals. The Department of Probability and Statistics in Sheffield, also Gani's creation, was a hub of these developments and it attracted a number of young talents to its circles from around the world, Terry being of them.

Another source of inspiration at the time was 'the general theory of stochastic processes', which was represented, most importantly, by the French and the Russian schools of probability. The key figure behind this in France was Paul-André Meyer and his book *Probability and Potentials* [Meyer, 1966] was one of the favorites in Terry's impressive home library in Sheffield. (A sign of Terry's interest in the works

Elja Arjas
University of Helsinki and National Institute for Health and Welfare, e-mail: Elja.Arjas@helsinki.fi

coming from the French school is that he translated into English J. Neveu's book *Martingales à temps discret* [Neveu, 1972], which appeared in 1975 with the title *Discrete Parameter Martingales* [Neveu, 1975]. I remember Terry wondering why the French publishers did not seem to make any effort towards marketing their books outside France, or even making them available in the largest bookstores in UK.)

Chronologically the earliest of the three papers on probability in this collection is the one entitled *Symmetric Wiener-Hopf factorisations in Markov additive processes*, which Terry and I submitted to the prestigious Springer journal 'ZW' in November 1972 [Arjas and Speed, 1973b]. For me, the background story leading to this is as follows: Not finding anyone in Finland to suggest a topic to work on for a PhD in probability, let alone to act as a supervisor, I had in desperation written to Professor Gani, asking him whether he would let me come and spend some time in his Department in Sheffield. I was immediately welcomed, and I stayed there for the winter and spring 1970–71. Sheffield turned out to be an excellent choice, with lots of academically interesting things going on all the time. There were many visitors, good weekly seminars, and if this wasn't sufficient, the Department paid train trips for us to go to London and Manchester to listen to more. But above all, there were people roughly of my age some of whom were working towards a PhD just like I was, and others who were already much beyond, like Terry. There I learned what doing research in probability might involve in practice. My contact with Terry, which grew into a friendship, was particularly important in this respect. During the first and longest stay in Sheffield in the spring of 1971 I lived next door from Terry and Sally, and on my later visits I enjoyed their hospitality as a guest in their home.

This paper on Wiener-Hopf factorizations was inspired, in particular, by the ideas on Random Walks in \mathbb{R}^1 that were contained in Chapter XII of Feller's Volume 2, with that same title. On the introductory page of this chapter Feller writes: "The theory presented in the following pages is so elementary and simple that the newcomer would never suspect how difficult the problems used to be before their natural setting was understood." The key to such elementary understanding offered by Feller is the concept of 'ladder point', a pair of random variables consisting of a 'ladder epoch' and 'ladder height'. Consecutive ascending (descending) ladder points make up the sequence of new maximal (minimal) record values of the random walk. The sample path of the random walk arising from its first n steps can now be divided into random excursions, each ending with a new maximal (minimal) record value, and finally including an incomplete excursion from the last maximum to where the random walk is after n steps. Due to the assumed iid structure of the random walk, the differences between the successive ascending (descending) ladders are also iid, and therefore the distribution of the sum of any k of them can be handled by forming a k -fold convolution 'power' of the distribution of one. These convolution powers of the common distribution of the ascending ladder heights make up the 'positive part' of the Wiener-Hopf factorization. The 'negative part' stems from the incomplete excursion, by first noting that its distribution remains the same when the order of its steps is reversed and that, when considered in this manner 'backwards in time',

the position at which the original random walk had its maximum now becomes a minimal record value. Therefore the distribution of this incomplete excursion gets a similar representation as the original sample path up to the maximal value, but now in terms convolution powers arising from the descending ladder points.

A second ingredient leading to our 'ZW' paper was the emergence, in varying formulations and uses, of the concept of conditional independence. Conditional independence had been previously considered, for example, by Pyke [1961] and Çinlar [1969] in connection of semi-Markov and Markov renewal processes, and it was also an essential ingredient in Hidden Markov Models (HMM's) introduced by Baum and Petrie [1966]. The general definitions and properties of conditional independence were expressed in measure theoretic terms in Meyer's book [Meyer, 1966]. In statistics, it seems to have taken a few more years, to the well-known discussion paper of Dawid [1979], until the fundamentally important ideas relating to conditional independence were fully appreciated and elaborated on. Presently, as is well known, conditional independence plays a major role particularly in Bayesian statistical modeling.

By replacing 'time' in Markov renewal processes by an additive real valued variable led us to consider, in a straightforward manner, a stochastic process called 'random walk defined on a Markov chain', or somewhat more generally, to Markov additive processes [Çinlar, 1972, Arjas and Speed, 1973a]. It was relatively easy to see that the key ideas of Feller's treatment of random walks could be retained if the model was extended to include an underlying Markov chain, then assuming that the increments of the additive variable were conditionally independent given the states of this chain. In the case where the state space of the chain is finite, ordinary univariate convolutions used in the original random walk would be replaced by the corresponding matrix convolutions. Our paper in ZW adds a further level of generality to these results, by stating them in terms of transition kernels defined on a measurable state space. The technically most demanding aspect here was the construction of the dual or adjoint operators, corresponding to the time reversal in the original process. For the record, I should say that it was Terry who was primarily responsible for correctly adding all necessary mathematical bells and whistles to these general formulations.

The second paper, entitled *A note on random times* [Pitman and Speed, 1973], provides the natural definition of, as it is called there, randomized stopping time in the case of processes of a discrete time parameter. In this brief note Jim and Terry, not only define this concept, but actually exhaust the topic completely by listing all its relevant properties and by linking it to different variants of essentially the same concept that existed in the literature at the time. Here, too, the key concept is conditional independence: Definition 1 says that a random time is a randomized stopping time relative to a family of histories if its occurrence, given the past, has no predictive value concerning the future. Of the properties derived, of most interest would seem to be the equivalence of (i) and (ii) of Proposition 2.5, and the intuitive

explanation which is provided afterwards. To put it simply, a randomized stopping time is an 'ordinary' stopping time if it is considered relative to a family of bigger histories. What is required of these larger histories is that, at any given time point and given the past of the 'original' history, events in the past of this larger history do not help in predicting the future of the original. When expressed in this way, one can see how close it is to the concept of 'non-causality' of Granger [1969], which is famous in the time series and econometrics literature, as well as, for example, to the property of local independence introduced by Schweder [1970].

Looking at a result like this one gets the feeling that the message it conveys should have been read, and understood, by generations of statisticians working in the area of survival analysis, in need of a natural definition of the concept of non-informative right censoring. They should have been thinking in terms of randomized stopping times! Instead, the common assumption stated in nearly all of the survival analysis literature is that of the 'random censoring model', which postulates for each considered individual the existence of two independent random variables, of which only the smaller is actually observed in the data. This model leads to strange events such as 'censoring of a person who is already dead'.

Terry is sole author of the third paper discussed here, entitled *Geometric and probabilistic aspects of some combinatorial identities* [Speed, 1976]. It is rather difficult to describe its contents in an understandable way in only a few sentences. In geometrical terms, it is concerned with certain hyperplanes in the positive orthant of the $(k + 1)$ -dimensional integer lattice. The main focus is on a particular combinatorial expression, which is shown to correspond to the number of minimal lattice paths from the origin to the considered hyperplane and such that the paths do not touch that plane until at the last point. This geometric interpretation then leads to concise derivations of some convolution type identities between the combinatorial expressions. Later on, the paper provides probabilistic interpretations, and corresponding proofs, for these results by considering the first passage time of a random walk from the origin to the hyperplane. There are also results on the associated moment generating functions, which have interesting analogues in the theory of branching processes. Although these combinatorial identities were not included in Feller's two books, one could say that Terry's approach to deal with them is very much Feller-like: when going through the mathematical derivations, at some point there is a phase transition from mysterious to intuitive and obvious. Another thing about this paper which I liked is its careful citing of the work of all authors who had earlier contributed, in various versions, to this same topic. But it looks like Terry just about exhausted this topic since, according to Google Scholar, to date this paper has been cited only once, and it isn't even listed in the ISI Web of Knowledge database.

Epilogue

When looking at the list of contents of this volume, which covers fifteen topics starting from algebra and ending with analysis of microarray data, one soon concludes that it would be hopeless to try to compete with Terry in terms of scientific output. In fact, competing with him in anything turned out to be a futile attempt. I once tried, in the late 1970'ies, when Terry visited me in Oulu and we went jogging. As we came back, I believe Terry was a bit more out of breath than I. Later on, however, Terry started practicing regularly by running up and down the steep hills surrounding Berkeley, and at some point I was told that he had run the marathon in less than three hours. My first marathon is still due. But luckily, there may be a sport where I have a chance of beating him: cross-country skiing. This is an open invitation to Terry to try.

References

- E. Arjas and T. P. Speed. Topics in Markov additive processes. *Math. Scand.*, 33: 171–192, 1973a.
- E. Arjas and T. P. Speed. Symmetric Wiener-Hopf factorisations in Markov additive processes. *Z. Wahrscheinlichkeitstheorie und verw. Geb.*, 26:105–118, 1973b.
- L. E. Baum and T. Petrie. Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Stat.*, 37:1554–1563, 1966.
- E. Çinlar. Markov renewal theory. *Advances in Applied Probability*, 1:123–187, 1969.
- E. Çinlar. Markov additive processes. *Z. Wahrscheinlichkeitstheorie und verw. Geb.*, 1972.
- A. P. Dawid. Conditional independence in statistical theory. *Journal of the Royal Statistical Society. Series B*, 41:1–31, 1979.
- W. Feller. *Introduction to Probability Theory and its Applications*, volume 1. Wiley, New York, 1950.
- W. Feller. *Introduction to Probability Theory and its Applications*, volume 2. Wiley, New York, 1966.
- C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37:424–438, 1969.
- P. A. Meyer. *Probability and Potentials*. Blaisdell, Waltham, MA, 1966.
- J. Neveu. *Martingales à temps discret*. Masson, Paris, 1972.
- J. Neveu. *Discrete Parameter Martingales*. North-Holland, Amsterdam, 1975.
- J. W. Pitman and T. P. Speed. A note on random times. *Stoch. Proc. Appl.*, 1: 369–374, 1973.
- R. Pyke. Markov renewal processes: Definitions and preliminary properties. *Ann. Math. Statist.*, 32:1231–1242, 1961.
- T. Schweder. Composable Markov processes. *Journal of Applied Probability*, 7: 400–410, 1970.

- T. P. Speed. Geometric and probabilistic aspects of some combinatorial identities.
J. Aust. Math. Soc., 22:462–468, 1976.