Chapter 2
Probability

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 [8], which had come out sixteen years after the publication of Volume 1 [7], 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 move 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 one 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* [10] was one of the favorites in Terry's impressive home library in Sheffield. (A sign of Terry's interest in the works coming from the French school is that he translated into English J. Neveu's book *Martingales à temps discret* [11], which appeared in 1975 with the title *Discrete Pa-*

E. Arjas
University of Helsinki and National Institute for Health and Welfare, Finland
e-mail: elja.arjas@helsinki.fi

S. Dudoit (ed.), *Selected Works of Terry Speed*, Selected Works in Probability and Statistics,
DOI 10.1007/978-1-4614-1347-9_2, © Springer Science+Business Media, LLC 2012

rameter Martingales [12]. 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 [2]. 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 such a record value 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 [14] and Çinlar [4] in connection of semi-Markov and Markov renewal processes, and it was also an essential ingredient in Hidden Markov Models (HMMs) introduced by Baum and Petrie [3]. The general definitions and properties of conditional independence were expressed in measure theoretic terms in Meyer's book [10]. In statistics, it seems to have taken a few more years, to the well-known discussion paper of Dawid [6], 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 [5, 1]. 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* [13], 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 that 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 [9], which is famous in the time series and econometrics literature, as well as, for example, to the property of local independence introduced by Schweder [15].

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* [16]. 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 1970s, 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

[1] E. Arjas and T. P. Speed. Topics in Markov additive processes. *Math. Scand.*, 33:171–192, 1973.

[2] E. Arjas and T. P. Speed. Symmetric Wiener-Hopf factorisations in Markov additive processes. *Z. Wahrscheinlichkeitstheorie und verw. Geb.*, 26(2):105–118, 1973.

[3] L. E. Baum and T. Petrie. Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Statist.*, 37:1554–1563, 1966.

[4] E. Çinlar. Markov renewal theory. *Adv. Appl. Prob.*, 1:123–187, 1969.

[5] E. Çinlar. Markov additive processes. *Z. Wahrscheinlichkeitstheorie und verw. Geb.*, 24:85–93, 1972.

[6] A. P. Dawid. Conditional independence in statistical theory. *J. Roy. Stat. Soc. B*, 41:1–31, 1979.

[7] W. Feller. *Introduction to Probability Theory and its Applications*, volume 1. Wiley, New York, 1950.

[8] W. Feller. *Introduction to Probability Theory and its Applications*, volume 2. Wiley, New York, 1966.

[9] C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37:424–438, 1969.

[10] P. A. Meyer. *Probability and Potentials*. Blaisdell, Waltham, MA, 1966.

[11] J. Neveu. *Martingales à temps discret*. Masson, Paris, 1972.

[12] J. Neveu. *Discrete-Parameter Martingales*, volume 10 of *North-Holland Mathematical Library*. North-Holland, Amsterdam, 1975. Translated from French by T. P. Speed.

[13] J. W. Pitman and T. P. Speed. A note on random times. *Stoch. Proc. Appl.*, 1: 369–374, 1973.

[14] R. Pyke. Markov renewal processes: Definitions and preliminary properties. *Ann. Math. Statist.*, 32:1231–1242, 1961.

[15] T. Schweder. Composable Markov processes. *J. Appl. Probab.*, 7:400–410, 1970.

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

Z. Wahrscheinlichkeitstheorie verw. Geb. 26, 105–118 (1973)

Symmetric Wiener-Hopf Factorisations in Markov Additive Processes

E. Arjas and T. P. Speed

The classical Wiener-Hopf factorisation of a probability measure is extended to an operator factorisation associated with a semi-Markov transition function. Some consequences of this factorisation are indicated including a set of duality relations.

1. Introduction

The classical Wiener-Hopf factorisation of a probability measure F on $(\mathbb{R}^1, \mathscr{R}^1)$ has been put in a symmetric form by Spitzer [14] and Feller [7] and can be written as follows:

$$\delta_0 - F = (\delta_0 - H^-) * (\delta_0 - \zeta\,\delta_0) * (\delta_0 - H^+) \tag{1.1}$$

where δ_0 is the unit mass at zero, $0 \leqq \zeta < 1$ and H^+, H^- are possibly defective probability measures concentrated on $(0, \infty)$ and $(-\infty, 0)$ respectively. In fact H^+ (resp. H^-) is identified as the distribution of the strict ascending (resp. descending) ladder variable.

In his very interesting extension of (1.1) Dinges [6] considered a substochastic transition function P on a measurable space $(E, \mathscr{E})$ with a total order, and constructed a factorisation:

$$I - \tau P = \left(I - \sum_1^\infty \tau^k P_k^-\right) \circ \left(I - \sum_1^\infty \tau^k P_k^{\cdot}\right) \circ \left(I - \sum_1^\infty \tau^k P_k^+\right) \tag{1.2}$$

where P_k^-, $P_k^{\cdot}$, and $P_k^+, k=0,1,\ldots,$ are suitable operators or sub-stochastic transition functions, $0 \leqq \tau < 1$ and "$\circ$" denotes composition. Dinges' result gives (1.1) as a special case, but first a few rearrangements are required to do this. The reason is that although P_k^- and P_k^+ are notationally dual their constructions are not immediately seen to be so, and thus it is desirable to clarify this point. Further Presman [11, 12] has unsymmetric matrix factorisations which are similar to ones derived below, but these are obtained algebraically.

It is the purpose of this paper to obtain a symmetric factorisation which generalises (1.1) in two distinct ways: for we deal with Markov additive processes $\{(X_n, S_n): n \geqq 0\}$, which reduce to the classical random walk by specialising the first component to a single value, or by suppressing the second component and specialising the first to be a random walk. Thus we can also obtain a result like (1.2) with the difference that our factorisation is manifestly symmetric. We formulate our results in an abstract way and the different results referred to are special cases. One aspect we emphasise throughout is the duality obtained from, and implicit in the proof of, our symmetric factorisations. In this respect our method is quite analogous to that of Feller's [7] Fourier analytic derivation of (1.1) in Chapter XVIII.

We now describe the contents of this paper. After some preliminaries concerning Markov additive processes we consider briefly Markov additive processes in duality. Next we formulate our abstract Wiener-Hopf factorisation and give its simple proof. The following two sections give concrete applications of this result and give a selection of corollaries. We close with some purely probabilistic duality results which are of some interest in themselves, and which can also be used to give alternative (probabilistic) proofs of our factorisations.

2. Markov Additive Processes

Our approach and notation will be based as far as possible upon Çinlar [4, 5] which in turn, is modelled upon Blumenthal and Getoor [3]. We recall some terminology. If $(G, \mathscr{G})$ and $(H, \mathscr{H})$ are measurable spaces and if $f: G \to H$ is measurable with respect to $\mathscr{G}$ and $\mathscr{H}$ then we write $f \in \mathscr{G}/\mathscr{H}$. If $H = \bar{\mathbb{R}}^1 = [-\infty, \infty]$ and $\mathscr{H} = \bar{\mathscr{R}}^1$, the Borel subsets of $\bar{\mathbb{R}}^1$, then we write $f \in \mathscr{G}$ instead of $f \in \mathscr{G}/\mathscr{H}$. Further $b\mathscr{G} = \{f \in \mathscr{G}: f \text{ is bounded}\}$, $\mathscr{G}_+ = \{f \in \mathscr{G}: f \geqq 0\}$ and $b\mathscr{G}_+ = b\mathscr{G} \cap \mathscr{G}_+$.

A mapping $N: F \times \mathscr{G} \to [0, 1]$ is called a *transition function* from $(F, \mathscr{F})$ into $(G, \mathscr{G})$ if a) $A \to N(x, A)$ is a measure on $\mathscr{G}$ for all fixed $x \in F$, and b) $x \to N(x, A)$ is in $b\mathscr{F}$ for all fixed $A \in \mathscr{G}$. Analogously, we define a mapping $Q: E \times (\mathscr{E} \times \bar{\mathscr{R}}^m) \to [0, 1]$ to be a *semi-Markov transition function* (abbrev. SMTF) on $(E, \mathscr{E}, \bar{\mathscr{R}}^m)$ if a) $x \to Q(x, A \times B)$ is in $b\mathscr{E}$ for every $A \in \mathscr{E}$, $B \in \bar{\mathscr{R}}^m$, b) $A \times B \to Q(x, A \times B)$ is a measure on $\mathscr{E} \times \bar{\mathscr{R}}^m$ for every $x \in E$.

If Q, R are two SMTF's on $(E, \mathscr{E}, \bar{\mathscr{R}}^m)$ we may define the *convolution product* $Q \circ R$ as the function,

$$(2.1) \qquad (x, A \times B) \to (Q \circ R)(x, A \times B) = \int_E \int_{\bar{\mathbb{R}}^m} Q(x, dx' \times ds)\, R(x', A \times (B - s)).$$

$Q \circ R$ is easily checked to be an SMTF. For any SMTF Q we define $Q^0 \equiv I$ where $I(x, A \times B) = \delta_x(A)\, \delta_0(B)$, and for $n \geqq 1$ $Q^n = Q^{n-1} \circ Q$.

There are many different ways of viewing a SMTF Q, and at various times we will be doing this. Thus Q may be viewed as a positive contraction valued measure defined on $(\bar{\mathbb{R}}^m, \bar{\mathscr{R}}^m)$ by the map $B \to Q(B)$, where $(Q(B) I_A)(x) = Q(x, A \times B)$; as a transition function on $(E \times \bar{\mathbb{R}}^m, \mathscr{E} \times \bar{\mathscr{R}}^m)$ which is homogeneous in the second component by the map $((x, s), A \times B) \to Q(x, A \times (B - s))$; as a transition function from $(E, \mathscr{E})$ to $(E \times \bar{\mathbb{R}}^m, \mathscr{E} \times \bar{\mathscr{R}}^m)$ by $(x, A \times B) \to Q(x, A \times B)$ (cf. Çinlar [4] (1.2)); and finally as giving a sequence $\{Q^n: n \geqq 0\}$ satisfying Definition (1.1) of Çinlar [5].

Any SMTF Q induces a family $\{Q(\theta): \theta \in \mathbb{R}^m\}$ of contractions on the Banach space $b\mathscr{E}$ by writing $(Q(\theta) f)(x) = \iint Q(x, dx' \times dy) \cdot f(x')\, e^{i(\theta, y)}$, where $(\cdot, \cdot)$ denotes the usual inner product in $\mathbb{R}^m$. We call $\{Q(\theta)\}$ the *Fourier transform* of Q.

We will consider a Markov process with state space $(E, \mathscr{E})$ to be a sextuple $X = (\Omega, \mathscr{M}, \mathscr{M}_n, X_n, \theta_n, P^x)$ $(x \in E)$, and all such processes will be assumed non-terminating (see Blumenthal and Getoor [3]). Following Çinlar [5] we have:

(2.2) **Definition.** Let X be a Markov process with state space $(E, \mathscr{E})$, write $(F, \mathscr{F}) = (\bar{\mathbb{R}}^m, \bar{\mathscr{R}}^m)$, and let $S = \{S_n: n \geqq 0\}$ be a family of functions from $(\Omega, \mathscr{M})$ into $(F, \mathscr{F})$. Then $(X, S) = (\Omega, \mathscr{M}, \mathscr{M}_n, X_n, S_n, \theta_n, P^x)$ is called a *Markov additive process* (abbrev. MAP) provided the following hold:

a) $S_0=0$ a.s.;

b) for each $n\geqq 0$, $S_n\in\mathcal{M}_n/\mathcal{F}$;

c) for each $n\geqq 0$, $A\in\mathcal{E}$, $B\in\mathcal{F}$, the mapping $x\to P^x\{X_n\in A, S_n\in B\}$ of E into $[0,1]$ is in $\mathcal{E}_+$;

d) for each $k, l\geqq 0$, $S_{k+l}=S_k+S_l\circ\theta_k$ a.s.;

e) for each $k, l\geqq 0$, $x\in E$, $A\in\mathcal{E}$, $B\in\mathcal{F}$

$$P^x\{X_l\circ\theta_k\in A, S_l\circ\theta_k\in B|\mathcal{M}_k\}=P^{X_k}\{X_l\in A, S_l\in B\}.$$

We follow Çinlar [5] in our notation for objects associated with the definition,

$$Q(x,C)=P^x\{(X_1,S_1)\in C\},\qquad C\in\mathcal{E}\times\mathcal{F}; \tag{2.3}$$

$$P(x,A)=Q(x,A\times F),\qquad A\in\mathcal{E}. \tag{2.4}$$

The action of $Q(B)$ mentioned above is as follows: for $f\in\mathcal{E}_+$

$$(Q(B)f)(x)=E^x[f(X_1); S_1\in B]. \tag{2.5}$$

Let N be a stopping time on Ω relative to $\{\mathcal{M}_n\}$; we define the (operator) transforms associated with (X_N, S_N) and with the behaviour of (X_n, S_n) for $n<N$: for $f\in b\mathcal{E}_+$ $\theta\in\mathbb{R}^m$, $0\leqq\tau<1$:

$$(Gf)(x)=E^x\left[\sum_0^{N-1}\tau^n e^{i(\theta,S_n)}f(X_n)\right], \tag{2.6}$$

$$(Hf)(x)=E^x[\tau^N e^{i(\theta,S_N)}f(X_N); N<\infty]. \tag{2.7}$$

A fundamental passage-time identity relating the transforms $G=G_N(\tau,\theta)$, $H=H_N(\tau,\theta)$ and $Q(\theta)$ is the following proved in Arjas and Speed [2] (I is the identity operator):

(2.8) **Proposition.** $G_N(\tau,\theta)[I-\tau Q(\theta)]=I-H_N(\tau,\theta)$.

3. Markov Additive Processes in Duality

Let us suppose that we are given a σ-finite measure π over our fixed state space $(E,\mathcal{E})$. We shall say that the MAP's

$$(X,S)=(\Omega,\mathcal{M},\mathcal{M}_n,X_n,S_n,\theta_n,P^x)\quad\text{and}\quad(\hat{X},\hat{S})=(\hat{\Omega},\hat{\mathcal{M}},\hat{\mathcal{M}}_n,\hat{X}_n,\hat{S}_n,\hat{\theta}_n,\hat{P}^x)$$

with SMTF's Q, $\hat{Q}$ respectively, are in duality relative to π if

a) for every $x\in E$, $P(x,\cdot)\ll\pi$, $\hat{P}(x,\cdot)\ll\pi$;

b) for every $B\in\bar{\mathcal{R}}^m$, $f,g\in\mathcal{E}_+$

$$\langle f, Q(B)g\rangle=\langle f\hat{Q}(-B), g\rangle \tag{3.1}$$

where, for $f_1, g_1\in\mathcal{E}_+$, we have $\langle f_1, g_1\rangle=\int f_1(x)\,g_1(x)\,\pi(dx)$. In this case we say also that Q and $\hat{Q}$ are in duality relative to π.

It can be proved (cf. Blumenthal and Getoor [3]) that π is P-excessive where $P=Q(\bar{\mathbb{R}}^m)$ is the Markov transition function of X, and similar results hold for $\hat{P}$.

Thus (cf. Nelson [10]) the operators $Q(B)$ (resp. $\hat{Q}(B)$) defined by (2.5) act as linear contractions on $L^p(\pi)$ for $1 \leq p \leq \infty$. With this interpretation (3.1) expresses the fact that $\hat{Q}(-B)$, acting on $L^p(\pi)$, is the Banach space adjoint of $Q(B)$ acting on $L^q(\pi)$ where $p^{-1}+q^{-1}=1$. Slightly modifying this terminology we will speak of T and T^* being *adjoint* if $\langle f, T(B)g\rangle = \langle f T(-B)^*, g\rangle$ for every $B \in \bar{\mathscr{R}}^m$, $f, g \in \mathscr{E}_+$.

4. The Factorisation

In this section we present an axiomatic approach to symmetric Wiener-Hopf factorisations of SMTF's. A special case of our work is the unsymmetric matrix factorisation of Presman [12] whose derivation is abstract algebraic in nature. We would like to emphasise that while the discussion to follow is in a sense abstract, probabilistic considerations are used throughout and thus our arguments could hardly be termed algebraic.

Our formulation of the Wiener-Hopf factorisation will be in terms of the Fourier transforms of certain operator-valued measures. Explicitly, we will call a map $B \to T(B)$ from $\bar{\mathscr{R}}^m$ into the space of all bounded linear operators over $L^p(\pi)$ an operator-valued measure if for every $f \in L^p$, $g \in L^q$, the set function $B \to \langle f, T(B)g\rangle$ is countably additive. In this case the Fourier transform of the operator-valued measure is the operator-valued function $\theta \to T(\theta)$ from $\mathbb{R}^m$ into the space of all bounded linear operators over $L^p(\pi)$ where we write, for $f \in L^p$, $g \in L^q$, $\langle f, T(\theta)g\rangle = \int e^{i(\theta, y)}\langle f, T(dy)g\rangle$. It is easy to see that the functions $\theta \to G_N(\tau, \theta)$ and $\theta \to H_N(\tau, \theta)$ are Fourier transforms of suitable operator-valued measures. The space of all such Fourier transforms will be denoted $\mathscr{A}$, clearly an algebra over $\mathbb{C}$.

We make the following convention which shortens somewhat our statements: We say that a statement holds

(i) *symmetrically* (abbrev. s.) if it holds when all "+" symbols are replaced by "−" symbols and vice versa;

(ii) *dually* (abbrev. d.) if it holds when (X, S) and the possible other elements associated with it are replaced by $(\hat{X}, \hat{S})$ and the corresponding associated elements.

As we conceive them, symmetric Wiener-Hopf factorisations of transforms of SMTF's have three essential ingredients. We assume the following (I–III) throughout this section (almost surely):

I: A decomposition $\mathbf{A} = \mathbf{A}^- \oplus \mathbf{A}^{\cdot} \oplus \mathbf{A}^+$ of a subalgebra $\mathbf{A} \subset \mathscr{A}$ with

(i) $\mathbf{A}^-$, $\mathbf{A}^{\cdot}$, $\mathbf{A}^+$ all subalgebras of $\mathbf{A}$;

(ii) $\mathbf{A}^- \mathbf{A}^{\cdot} \subset \mathbf{A}^-$, $\mathbf{A}^{\cdot} \mathbf{A}^- \subset \mathbf{A}^-$, and s.;

(iii) $(\mathbf{A}^+)^* = \mathbf{A}^-$ and s., $(\mathbf{A}^{\cdot})^* = \mathbf{A}^{\cdot}$.

Here $\mathbf{A}^- \mathbf{A}^{\cdot} = \{ST\colon S \in \mathbf{A}^-, T \in \mathbf{A}^{\cdot}\}$ etc., and $(\mathbf{A}^+)^* = \{S^*\colon S \in \mathbf{A}^+\}$ and s.

We call a decomposition as in I a *symmetric W-decomposition*. The letter W is to stand for "Wendel" as there is a close relationship between the above and the so-called Wendel-projections of Kingman [9].

II: A system of stopping times N^+, $N^{\cdot +}$, N_+ relative to $\{\mathscr{M}_n\}$, and s. and d., such that almost surely

(i) $N_+ = N^{\cdot +} < N^+$ if $N^{\cdot +} < \infty$ and $N_+ = N^+$ if $N^{\cdot +} = \infty$, and s. and d.;

(ii) on $\{N^{\cdot +} < \infty\}$ $N^+ = N^{\cdot +} + N^+ \circ \theta_{N^{\cdot +}}$, and s. and d.

The stopping time N^+ will be sometimes described as a *strict ladder index* and N_+ as a *weak ladder index*, and s. and d.

We require that the above stopping times be adapted to the symmetric W-decomposition, by which we mean:

III: (i) $I \in \mathbf{A}^{\cdot}$;

(ii) $H_{N^+} \in \mathbf{A}^+$, $G_{N^+} \in \mathbf{A}^- \oplus \mathbf{A}^{\cdot}$, and s. and d.;

(iii) $H_{N_+} \in \mathbf{A}^{\cdot} \oplus \mathbf{A}^+$, $G_{N_+} - I \in \mathbf{A}^-$, and s. and d.;

where $\mathbf{A}^-$, $\mathbf{A}^{\cdot}$ and $\mathbf{A}^+$ stay fixed when statements are dualised.

We now prove two important preliminary lemmas, which give the desired factorisation as an almost immediate corollary. In the first lemma only II is used, whereas the second lemma is based on I and III.

(4.1) **Lemma** *(Relation between strict and weak ladder indices).*

$$I - H_{N_+} = (I - H_{N^{\cdot +}})(I - H_{N^+}), \textit{ and s. and d.}$$

Proof. We note first that for $x \in E$, $0 \leqq \tau \leqq 1$, $\theta \in \mathbb{R}^m$, $f \in L^p$

(4.2) $$E^x[\tau^{N^+} e^{i(\theta, S_{N^+})} f(X_{N^+}); \quad N^{\cdot +} < N^+ < \infty] = (H_{N^{\cdot +}} H_{N^+} f)(x).$$

To see this we write

$$\begin{aligned} &E^x[\tau^{N^+} e^{i(\theta, S_{N^+})} f(X_{N^+}); N^{\cdot +} < N^+ < \infty] \\ &\quad = E^x[\tau^{N^{\cdot +}} e^{i(\theta, S_{N^{\cdot +}})} E^x[\tau^{N^+ \circ \theta_{N^{\cdot +}}} e^{i(\theta, S_{N^+} \circ \theta_{N^{\cdot +}})} f(X_{N^+} \circ \theta_{N^{\cdot +}}); \\ &N^+ \circ \theta_{N^{\cdot +}} < \infty \,|\, \mathcal{M}_{N^{\cdot +}}]; N^{\cdot +} < \infty] \qquad \text{by II and the general properties of conditional expectations} \\ &\quad = E^x[\tau^{N^{\cdot +}} e^{i(\theta, S_{N^{\cdot +}})} (H_{N^+} f)(X_{N^{\cdot +}}); N^{\cdot +} < \infty] \\ &\qquad \text{by the (strong) Markov property} \\ &\quad = (H_{N^{\cdot +}} H_{N^+} f)(x). \end{aligned}$$

Then, using II(i) and (4.2), we observe that

$$\begin{aligned} (H_{N_+} f)(x) &= E^x[\tau^{N_+} e^{i(\theta, S_{N_+})} f(X_{N_+}); \; N_+ < \infty] \\ &= E^x[\tau^{N_+} e^{i(\theta, S_{N_+})} f(X_{N_+}); \; N_+ = N^{\cdot +} < \infty] \\ &\quad + E^x[\tau^{N_+} e^{i(\theta, S_{N_+})} f(X_{N_+}); \; N_+ = N^+ < \infty] \\ &= E^x[\tau^{N^{\cdot +}} e^{i(\theta, S_{N^{\cdot +}})} f(X_{N^{\cdot +}}); \; N^{\cdot +} < \infty] \\ &\quad + E^x[\tau^{N^+} e^{i(\theta, S_{N^-})} f(X_{N^+}); \; N^+ < \infty] \\ &\quad - E^x[\tau^{N^+} e^{i(\theta, S_{N^+})} f(X_{N^+}); \; N^{\cdot +} < N^+ < \infty] \\ &= (H_{N^{\cdot +}} f)(x) + (H_{N^+} f)(x) - (H_{N^{\cdot +}} H_{N^+} f)(x) \qquad \text{by (4.2)} \end{aligned}$$

which completes the proof. The symmetric and dual statements are proved similarly.

The second of the preliminary lemmas is

(4.3) **Lemma** *(Duality).*

(i) $G_{N^+} = (I - \hat{H}^*_{N_+})^{-1}$, *and s. and d.*;

(ii) $G_{N_+} = (I - \hat{H}^*_{N^+})^{-1}$, *and s. and d.*

Proof. By Proposition (2.8) applied to N_+, and its dual form applied to $\hat{N}^+$, for $0 \leqq \tau < 1$,

$$(I-\tau Q)^{-1} = (I - H_{N_+})^{-1} G_{N_+}$$

and

$$(I-\tau \hat{Q})^{-1} = (I - \hat{H}_{\hat{N}^+})^{-1} \hat{G}_{\hat{N}^+}.$$

These equations are mutually adjoint because $\hat{Q} = Q^*$, and so comparing the right hand sides we get

$$(I - H_{N_+})^{-1} G_{N_+} = \hat{G}^*_{\hat{N}^+} (I - \hat{H}^*_{\hat{N}^+})^{-1},$$

and further

$$G_{N_+} (I - \hat{H}^*_{\hat{N}^+}) = (I - H_{N_+}) \hat{G}^*_{\hat{N}^+}.$$

From I and III follows that the left hand side is of the form $I + K$ where $K \in \mathbf{A}^-$, and the right hand side is in $\mathbf{A}^{\cdot} \oplus \mathbf{A}^+$. Hence both sides must be I, giving (4.3)(ii) and the dual statement of (4.3)(i). Other symmetric and dual statements are proved similarly.

(4.4) **Corollary.** (i) $H_{N\cdot+} = \hat{H}^*_{\hat{N}\cdot+}$ *and s.;*

(ii) $H_{N\cdot+} \in \mathbf{A}^{\cdot}$ *and s. and d.*

Proof. (i)

$$\begin{aligned} I - H_{N\cdot+} &= (I - H_{N_+})(I - H_{N^+})^{-1} && \text{by (4.1)} \\ &= G_{N_+}(I - \tau Q)(I - \tau Q)^{-1} G_{N^+}^{-1} && \text{by (2.8)} \\ &= G_{N_+} G_{N^+}^{-1} && \text{cancelling} \\ &= (I - \hat{H}^*_{\hat{N}^+})^{-1} (I - \hat{H}^*_{\hat{N}_+}) && \text{by (4.3)} \\ &= [(I - \hat{H}_{\hat{N}_+})(I - \hat{H}_{\hat{N}^+})^{-1}]^* = I - \hat{H}^*_{\hat{N}\cdot+} && \text{by (4.1)}. \end{aligned}$$

(ii) $H_{N\cdot+} \in \mathbf{A}^{\cdot} \oplus \mathbf{A}^+$ follows from the first line of the above proof when using III, and $\hat{H}^*_{\hat{N}\cdot+} \in \mathbf{A}^{\cdot} \oplus \mathbf{A}^-$ can be proved similarly. The assertion then follows from (4.4)(i).

(4.5) **Theorem** *(Wiener-Hopf factorisation). Let (X, S) and $(\hat{X}, \hat{S})$ be in duality relative to π, and assume* I–III *to be valid. Then, for $0 \leqq \tau < 1$, $\theta \in \mathbb{R}^m$:*

(4.6) $$I - \tau Q(\theta) = [I - \hat{H}^*_{\hat{N}^+}(\tau, \theta)]\,[I - H_{N\cdot+}(\tau, \theta)]\,[I - H_{N^+}(\tau, \theta)], \quad \textit{and s. and d.},$$

where the middle term is interchangeable with $I - \hat{H}^*_{\hat{N}\cdot+}(\tau, \theta)$, and s. and d. Further, the factorisation (4.6) *is unique in the sense that for a given W-decomposition there are no other factorisations with the non-unit term of the first (resp. second, third) factor in $\mathbf{A}^-$ (resp. $\mathbf{A}^{\cdot}$, $\mathbf{A}^+$), and s., and d.*

Proof.

$$\begin{aligned} I - \tau Q(\theta) &= G_{N^+}^{-1}(\tau, \theta)\,[I - H_{N_+}(\tau, \theta)] && \text{by (2.8)} \\ &= [I - \hat{H}^*_{\hat{N}^+}(\tau, \theta)]\,[I - H_{N_+}(\tau, \theta)] && \text{by (4.3)(ii)} \\ &= [I - \hat{H}^*_{\hat{N}^+}(\tau, \theta)]\,[I - H_{N\cdot+}(\tau, \theta)]\,[I - H_{N^+}(\tau, \theta)] && \text{by (4.1)}, \end{aligned}$$

which is the required factorisation. The interchangeability of $I - H_{N\cdot+}(\tau, \theta)$ with $I - \hat{H}^*_{\hat{N}\cdot+}(\tau, \theta)$ follows from (4.4)(i).

We now prove uniqueness. To do this let us abbreviate the notation and assume that

$$I - \tau Q = K^- K^{\cdot} K^+ = L^- L^{\cdot} L^+$$

are two factorisations with factors invertible such that $I-K^{-}$, $I-L^{-}\in\mathbf{A}^{-}$; $I-K^{\cdot}$, $I-L^{\cdot}\in\mathbf{A}^{\cdot}$ and $I-K^{+}$, $I-L^{+}\in\mathbf{A}^{+}$. Then

$$K^{\cdot}K^{+}(L^{+})^{-1}(L^{\cdot})^{-1}=(K^{-})^{-1}L^{-},$$

and arguing as in the proof of (4.4)(ii) we see that both sides must be equal I, giving

$$K^{-}=L^{-} \quad\text{and}\quad K^{\cdot}K^{+}=L^{\cdot}L^{+}.$$

A similar argument on the latter equation shows that $K^{+}=L^{+}$ and $K^{\cdot}=L^{\cdot}$. (This proof followed a familiar pattern, cf. Dinges [6].)

We also state the factorisation in a measure form, allowing a direct comparison to the factorisation (1.2) of Dinges. Without going through the lengthy preliminaries (regarding the decomposition of the convolution algebra of operator-valued measures etc.) or making qualifications regarding uniqueness we simply describe the form of the factorisation and briefly explain some details of its components.

(4.7) **Theorem** *(Wiener-Hopf factorisation, measure form). For suitable operator-valued measures $H_n^{+}, H_n^{\cdot +}$ $\hat{H}_n^{+}$, $n\geqq 1$, we have*

$$(4.8)\qquad [I-\tau Q](B)=\left[I-\sum_1^\infty \tau^n(\hat{H}_n^{+})^*\right]\circ\left[I-\sum_1^\infty \tau^n H_n^{\cdot +}\right]\circ\left[I-\sum_1^\infty \tau^n H_n^{+}\right](B),$$

and s. and d.

Interpretation. (i) "$\circ$" denotes the convolution product (see (2.1)) and "$*$" the adjoint as in § 3;

(ii) for $x\in E$, $B\in\bar{\mathscr{R}}^m$, $f\in L^p$ and $n\geqq 1$:

$$(H_n^{+}(B)f)(x)=E^x[f(X_n);\ N^{+}=n,\ S_n\in B],$$
$$(H_n^{\cdot +}(B)f)(x)=E^x[f(X_n);\ N^{\cdot +}=n,\ S_n\in B],$$
$$(\hat{H}_n^{+}(B)f)(x)=\hat{E}^x[f(\hat{X}_n);\ \hat{N}^{+}=n,\ \hat{S}_n\in B].$$

5. A Factorisation for Markov Chains with Totally Ordered State Space

We now specialise the results of the previous section to give a symmetrised factorisation for a transition function P, analogous to Dinges' [6] result. Recall however that we have assumed our process to be non-terminating, whereas in Dinges' case no extra assumptions of this kind are made save the necessary ones regarding order. These are that E has a reflexive, transitive binary relation, denoted $\leqq$, such that for any $x, x'\in E$ either $x\leqq x'$ or $x'\leqq x$. Further, if we write $x\sim x$ iff $x\leqq x'$ and $x'\leqq x$, and $x<x'$ if $x\leqq x'$ and $x\sim x'$ is false, then we require that $\{(x,x'): x'<x\}$ belong to the product σ-field $\mathscr{E}\times\mathscr{E}$.

For our algebra $\mathbf{A}$ (subalgebra of $\mathscr{A}$) we choose the real algebra generated by the set of all positive contractions on $L^p(\pi)$; this arises by putting $\theta=0$ in each element of $\mathscr{A}$. Using the well-known equivalence between positive contractions and transition functions on $(E,\mathscr{E})$ we define the appropriate symmetric W-decomposition as follows: for $T\in\mathbf{A}$, $x\in E$, $A\in\mathscr{E}$ put

$$(5.1)\qquad \begin{aligned} T^{+}(x,A)&=T(x,\{x': x<x'\}\cap A);\\ T^{\cdot}(x,A)&=T(x,\{x': x'\sim x\}\cap A);\\ T^{-}(x,A)&=T(x,\{x': x'<x\}\cap A);\end{aligned}$$

clearly $T = T^- + T^{\cdot} + T^+$ and this is easily seen to define a direct sum decomposition of **A** satisfying I (i), (ii) of § 4. To see how the decomposition can be defined directly in terms of its action on functions, we refer to Dinges [6].

The system of stopping times is the familiar one – the usual ladder indices:

(5.2)
$$\begin{aligned} &N^+ = \inf\{n>0: X_0 < X_n\};\\ &N_+ = \inf\{n>0: X_0 \leqq X_n\};\\ &N^{\cdot +} = N_+ \quad \text{if } N_+ < N^+, \text{ and } N^{\cdot +} = \infty \text{ otherwise;}\\ &\text{and s. and d.}\end{aligned}$$

We omit the verification of the fact that (5.2) satisfies II and III of § 4; II (ii) follows because on $\{N^{\cdot +} < \infty\}$ $X_{N^{\cdot +}} \sim X_0$ and $N^{\cdot +} < N^+$ so that $N^+ = \inf\{n > N^{\cdot +}: X_{N^{\cdot +}} < X_n\}$, and other requirements are satisfied quite obviously. Thus we can read off the following theorem, where we write $H_{N^+}(\tau) = H_{N^+}(\tau, 0)$ etc.:

(5.3) **Theorem.** *Let P and $\hat{P}$ be in duality relative to π, and consider the stopping times* (5.2). *Then as a relation between contractions on $L^p(\pi)$ for $0 \leqq \tau < 1$*

(5.4) $$I - \tau P = [I - \hat{H}^*_{\hat{N}^+}(\tau)]\,[I - H_{N^{\cdot +}}(\tau)]\,[I - H_{N^+}(\tau)], \qquad \textit{and s. and d.,}$$

where the middle term is interchangeable with $I - \hat{H}^*_{\hat{N}^{\cdot +}}(\pi)$, and s. and d. The uniqueness is as in Theorem (4.5).

(5.5) *Application 1.* The one-dimensional random walk. Suppose that $X_n = \sum_1^n Z_k$ where the $\{Z_k\}$ are i.i.d. random variables with law μ. Let λ denote Lebesgue measure on $(\mathbb{R}^1, \mathscr{R}^1)$; then it is easy to see that λ is P-excessive with $\hat{P}(x, A) = \hat{\mu}(A - x)$ where $\hat{\mu}$ is the measure μ reflected in the origin i.e. for $B \in \mathscr{R}^1$ $\hat{\mu}(B) = \mu(-B)$.

Now the operator P on $L^\infty(\lambda)$ is

(5.6) $$(Pf)(x) = E^x[f(X_1)] = \int f(x + x')\,\mu(dx').$$

Following Dinges [6] we call this operator T_μ; note that if $e(x) = e^{i\theta x}$ for $\theta \in \mathbb{R}^1$ then $(T_\mu e)(x) = \phi(\theta)\, e(x)$ i.e. scalar multiplication by the characteristic function $\phi(\theta)$ of μ. The following expressions are readily checked: with notation as in Feller [7], Chapter XVIII (3.5)

(5.7)
$$\begin{aligned} &(H_{N^+} e)(0) = \chi(\tau, \theta),\\ &(H_{N^{\cdot +}} e)(0) = f(\tau),\\ &(\hat{H}^*_{\hat{N}^+} e)(0) = \chi^-(\tau, \theta).\end{aligned}$$

Note that in the last case the adjoint simply means complex-conjugation; the Eq. (3.5) of Feller is now seen to be an immediate consequence of (5.4) above acting on $e(x)$ and evaluating at $x = 0$.

(5.8) *Application 2.* The m-dimensional random walk.

Here $X_n = \sum_1^n Z_k$ where $\{Z_k\}$ is a sequence of i.i.d. random variables with law μ. The dual process $\hat{X}$ is constructed as in the previous example, with respect to λ_m, m-dimensional Lebesgue measure. We order the state space $(\mathbb{R}^m, \mathscr{R}^m)$ by selecting a basis for $\mathbb{R}^m$ so that each Z_k can be written $Z_k=(Z_k^{(1)}, \ldots, Z_k^{(m)})$ and we then write:

$$(x'^{(1)}, \ldots, x'^{(m)}) \underset{[\sim]}{>} (x^{(1)}, \ldots, x^{(m)}) \quad \text{iff } x'^{(m)} \underset{[=]}{>} x^{(m)}.$$

In terms of this order the ladder indices N^+ etc. relate to the hyperplane $x^{(m)}=0$. Exactly as we found in the preceding example a factorisation arises by operating on $e(x)=e^{i(\theta,x)}$ for $\theta\in\mathbb{R}^m$.

(5.9) *Application 3.* A duality principle.

We now briefly describe a duality principle which is implicitly contained in Lemma (4.3). We express it as adjointness of two transition functions or rather, their associated contractions. For $x\in E$, $A\in\mathscr{E}$, $n\geqq 1$, define:

(5.10) (i) $D_n(x, A)=P^x\{X_1\leqq x, \ldots, X_n\leqq x,\ X_n\in A\}$,
(ii) $\hat{D}_n(x, A)=\hat{P}^x\{\hat{X}_1\leqq\hat{X}_n, \ldots, \hat{X}_{n-1}\leqq\hat{X}_n, \hat{X}_n\in A\}$.

Clearly these transition functions induce contractions D_n and $\hat{D}_n$ on $L^p(\pi)$ and $L^q(\pi)$ respectively, and the duality result is:

(5.11) **Propostion.** *$D_n^*=\hat{D}_n$ for all $n>0$.*

(5.12) *Remark.* The symmetric statements, where $\leqq$ in (5.10) is replaced systematically by $<$, $\geqq$ or $>$, and the dual statements hold also.

Proof. With the stopping times N^+ and $\hat{N}_+$ and the duality being used in this section we see that with definition (5.10)(i)

$$G_{N^+}(\tau)=\sum_0^\infty \tau^n D_n \qquad \textit{where } D_0=I.$$

Further, observing that

$$\hat{D}_n(x, A)=\hat{P}^x\{n \text{ is a weak ascending ladder index, } \hat{X}_n\in A\}$$

we readily find that

$$(I-\hat{H}_{\hat{N}_+}(\tau))^{-1}=\sum_0^\infty \tau^n \hat{D}_n \qquad \text{where } \hat{D}_0=I,$$

and the proof is an immediate consequence of Lemma (4.3)(ii).

(5.13) *Remark.* We can express Proposition (5.11) as follows: for $f\in L^p(\pi)$, $g\in L^q(\pi)$, $n>0$:

$$\begin{aligned}\langle f, D_n g\rangle &= \iint f(x)\, P^x\{X_1\leqq x, \ldots, X_n\leqq x, X_n\in(dx')\}\, g(x')\,\pi(dx)\\ &= \iint f(x)\, \hat{P}^{x'}\{\hat{X}_1\leqq\hat{X}_n, \ldots, \hat{X}_{n-1}\leqq\hat{X}_n, \hat{X}_n\in(dx)\}\, g(x')\,\pi(dx')\\ &= \langle f\hat{D}_n, g\rangle.\end{aligned}$$

In this form it is easy to give a direct probabilistic proof, and with this proof of Lemma (4.3), combined with a direct probabilistic proof of Lemma (4.1), we have an alternative method of obtaining Theorem (5.3).

6. A Factorisation Associated with the Second Component of a MAP

As a second specialisation we derive a factorisation using the ladder indices associated with the S-component of a MAP (X, S). This was our original aim and amongst many possible applications, it gives an alternative way of deriving the result (1.1). Throughout we suppose the dimension $m=1$, see Remark (6.6).

The algebra which we decompose is the full algebra $\mathscr{A}$ of all Fourier transforms $T(\theta)$. For any such transform we have $T(\theta)=\int e^{i\theta y}\,T(dy)$, and we define

$$\begin{aligned} T(\theta)^- &= \int_{-\infty}^{0-} e^{i\theta y}\,T(dy), \\ T(\theta)^{\cdot} &= T(\{0\}), \\ T(\theta)^+ &= \int_{0+}^{\infty} e^{i\theta y}\,T(dy), \end{aligned} \tag{6.1}$$

where the right sides can be interpreted formally or precisely, as operator integrals. For example, if $f\in L^p$, $g\in L^q$, $p^{-1}+q^{-1}=1$, then we define such integrals by

$$\langle f, T(\theta)^- g\rangle = \int_{-\infty}^{0-} e^{i\theta y}\langle f, T(dy)\,g\rangle$$

and similarly for $T(\theta)^+$. Clearly $T(\theta)=T(\theta)^-+T(\theta)^{\cdot}+T(\theta)^+$ and this decomposition induces a decomposition of $\mathscr{A}$ satisfying I(i), (ii) of § 4. The system of stopping times is the family of ladder indices for S:

$$\begin{aligned} &N^+=\inf\{n>0\colon S_n>0\}; \\ &N_+=\inf\{n>0\colon S_n\geqq 0\}; \\ &N^{\cdot+}=N_+ \quad \text{if } N_+<N^+, \text{ and } N^{\cdot+}=\infty \text{ otherwise;} \\ &\text{and s. and d.} \end{aligned} \tag{6.2}$$

We again omit the verification of the fact that (6.2) satisfies II and III of § 4; II(ii) now follows because $S_{N^{\cdot+}}=0$ on $\{N^{\cdot+}<\infty\}$. We have the following theorem, where $H_{N^{\cdot+}}(\tau)=H_{N^{\cdot+}}(\tau,0)$:

(6.3) **Theorem.** *Let Q and $\hat{Q}$ be in duality relative to π, and consider the stopping times* (6.2) *and s. and d. Then as a relation between contractions on $L^p(\pi)$, for $0\leqq\tau<1$, $\theta\in\mathbb{R}^1$:*

$$\begin{aligned} &I-\tau Q(\theta)=[I-\hat{H}^*_{\hat{N}^+}(\tau,\theta)]\,[I-H_{N^{\cdot+}}(\tau)]\,[I-H_{N^+}(\tau,\theta)], \\ &\text{and s. and d.,} \end{aligned} \tag{6.4}$$

where the middle term is interchangeable with $I-\hat{H}^*_{\hat{N}^{\cdot+}}(\tau)$, and s. and d. The uniqueness is as in Theorem (4.5).

We now suppose that the state space $E=\{1,2,\ldots,s\}$ and for a given SMTF Q the underlying chain P is ergodic. Thus there is a unique invariant measure π such that $\pi(i)>0$, $i\in E$. Put $\Delta=(\delta_{ij}\,\pi(i))$.

(6.5) **Corollary.** *In the finite-state case just described, if t denotes matrix transpose:*

$$I-\tau Q(\theta)=\Delta^{-1}[I-\hat{H}_{\hat{N}^+}(\tau,\theta)]^t\Delta\,[I-H_{N^{\cdot+}}(\tau)]\,[I-H_{N^+}(\tau,\theta)]$$

and s. and d.

This result is a symmetrised form of Theorem (2.1) of Presman [12], and if the last two factors are combined it becomes exactly his result.

(6.6) *Remark.* Before going on to give applications of Theorem (6.3) we will observe that the restriction to $m=1$ in this section is purely for simplicity. At least one interesting situation in $m>1$ dimensions is when N is the hitting time to a half-space through 0, as described in § 5. This topic can be treated exactly as the 1-dimensional case has been, giving rise to a generalised form of (6.3).

(6.7) *Application 1.* A duality principle.

The following discussion is a generalisation of the result Feller [7], p. 609, as indeed was the result (5.9). In a manner similar to our previous discussion we define SMTF's $D_n, \hat{D}_n$: for $x\in E$, $A\in\mathscr{E}$, $B\in\bar{\mathscr{R}}^1$ and $n\geqq 1$

(6.8) (i) $D_n(x, A\times B)=P^x\{X_n\in A, S_1\leqq 0, \ldots, S_n\leqq 0, S_n\in B\}$;

(ii) $\hat{D}_n(x, A\times B)=\hat{P}^x\{\hat{X}_n\in A, \hat{S}_1\leqq \hat{S}_n, \ldots, \hat{S}_{n-1}\leqq \hat{S}_n, \hat{S}_n\in B\}$.

It is easy to see that these induce contractions on $L^p(\pi)$ and $L^q(\pi)$ respectively, and the duality result here is:

(6.9) **Proposition.** *$D_n^*(B)=\hat{D}_n(B)$ for all $B\in\bar{\mathscr{R}}^1$, $n>0$.*

Proof. The proof is almost identical to that given for Proposition (5.11).

Remark (5.12) applies here as well. Also as in § 5 we can give a direct proof of this result, but we refer to the final section for a fuller discussion.

We now discuss briefly the above duality in the context of the bivariate processes $(X,W)=\{(X_n, W_n): n\geqq 0\}$ and $(X,M)=\{(X_n, M_n): n\geqq 0\}$ where we define

(6.10)
$$(X_0, W_0)=(X_0, 0)$$
$$(X_n, W_n)=(X_n, (W_{n-1}+S_n-S_{n-1})^+), \qquad n>0;$$

and

(6.11)
$$(X_n, M_n)=(X_n, \min(0, S_1, \ldots, S_n)), \qquad n\geqq 0.$$

We now formulate this duality explicitly as:

(6.12) **Theorem.** *For (X,S) and $(\hat{X},\hat{S})$ in duality the bivariate processes (X,W) and $(\hat{X},\hat{M})$ are adjoint.*

Proof. As shown in Arjas and Speed [2] the resolvent of (X, W) is

$$\Lambda(\tau,\theta)=[I-H_{N_-}(\tau,0)]^{-1}\, G_{N_-}(\tau,\theta)$$

and that of $(\hat{X},\hat{M})$ is

$$\hat{\Phi}(\tau,\theta)=[I-\hat{H}_{\hat{N}^-}(\tau,\theta)]^{-1}\,\hat{G}_{\hat{N}^-}(\tau,0),$$

where the stopping times are the ladder indices (6.2). Now if we take the adjoint of $\Lambda(\tau,\theta)$ we find

$$\begin{aligned}\Lambda^*(\tau,\theta)&=G^*_{N_-}(\tau,\theta)\,[I-H^*_{N_-}(\tau,0)]^{-1}\\&=[I-\hat{H}_{\hat{N}^-}(\tau,\theta)]^{-1}\,\hat{G}_{\hat{N}^-}(\tau,0) \quad \text{by Lemma (4.3)}\\&=\hat{\Phi}(\tau,\theta) \qquad \text{as stated.}\end{aligned}$$

(6.13) *Application 2.* A moment identity.

In Feller [7] one of the more immediate consequences of the factorisation (1.1) is a relation between the expectations of the hitting times to half-lines (assuming both exist) which reads

$$-\tfrac{1}{2}\sigma^2 = E[S_{N^-}]\,[1-\zeta]\; E[S_{N^+}]. \tag{6.14}$$

We now derive an analogue of (6.14) for the stopping times under discussion in this section. Let $E^\pi[f]$ be an abbreviation for $\langle 1, f\rangle = \int f(x)\,\pi(dx)$ and let us consider (when possible) the limited expansions:

$$\begin{aligned} Q(\theta) &= P + i\theta Q_1 - \tfrac{1}{2}\theta^2 Q_2 + o(\theta^2);\\ H_{N^+}(1,\theta) &= H^+ + i\theta M^+ + o(\theta);\\ H_{N^{\cdot+}}(1) &= H^{\cdot+};\\ &\text{and d.} \end{aligned} \tag{6.15}$$

(6.16) **Theorem.** *Let Q and $\hat{Q}$ be in duality relative to π, and consider the stopping times* (6.2). *Then, if S_{N^+}* (resp. $\hat{S}_{\hat{N}^+}$) *is proper and has a finite expectation irrespective of the starting point X_0 of X* (resp. $\hat{X}_0$ *of* $\hat{X}$),

$$Q_1 = 0, \qquad Q_2 < \infty$$

and

$$-\tfrac{1}{2}E^\pi[S_1^2] = \iint E^x[\hat{S}_{\hat{N}^+}]\,[I - H^{\cdot+}](x, dx')\, E^{x'}[S_{N^+}]\,\pi(dx).$$

Proof. We use the factorisation (6.4) at $\tau = 1$, giving

$$\begin{aligned} &\langle 1, [I - Q(\theta)]\,1\rangle\\ &\quad= \langle 1, [I - \hat{H}^*_{\hat{N}^+}(1,\theta)]\,[I - H_{N^{\cdot+}}(1)]\,[I - H_{N^+}(1,\theta)]\,1\rangle\\ &\quad= \langle [I - \hat{H}^+ - i\theta\hat{M}^+ + o(\theta)]\,1, [I - H^{\cdot+}]\,[I - H^+ - i\theta M^+ + o(\theta)]\,1\rangle\\ &\quad= -\theta^2\langle \hat{M}^+ 1, [I - H^{\cdot+}]\,M^+ 1\rangle + o(\theta^2), \end{aligned}$$

since, by the assumption of properness, $\hat{H}^+ 1 = 1$ and $H^+ 1 = 1$. On the other hand we can use the expansion

$$\begin{aligned} &\langle 1, [I - Q(\theta)]\,1\rangle\\ &\quad= \langle 1, [I - P - i\theta Q_1 + \tfrac{1}{2}\theta^2 Q_2 + o(\theta^2)]\,1\rangle\\ &\quad= -i\theta\langle 1, Q_1 1\rangle + \tfrac{1}{2}\theta^2\langle 1, Q_2 1\rangle + o(\theta^2), \end{aligned}$$

and the assertion follows by comparing the coefficients of θ and θ^2.

7. Two-Barrier Duality Relations in MAP's

In this final section we show that some general duality relations obtained recently by one of us in the case of one-dimensional random walks carry over to the present situation. In particular we can use them to give a direct probabilistic proof of (6.3).

Let (X, S) be as before, $m = 1$, and define the "reflected" process (X', S') with SMTF Q' by $Q'(B) = Q(-B)$, $B \in \bar{\mathscr{R}}^1$. Further, let (X, V) (resp. (X', V')) be the process obtained from (X, S) (resp. (X', S')) by placing two absorbing barries for the second component at specified positions, and (X, W) (resp. (X', W')) be the process obtained from (X, S) (resp. (X', S')) by placing two impenetrable barries for the second component at 0 and $a>0$. In the latter case we have inductively

$$W_0 = S_0; \quad W_n = \min\big(a, \max(W_{n-1} + S_n - S_{n-1}, 0)\big), \quad n>0.$$

The dual processes $(\hat{X}, \hat{S})$, $(\hat{X}', \hat{S}')$, $(\hat{X}, \hat{V})$, $(\hat{X}', \hat{V}')$, $(\hat{X}, \hat{W})$ and $(\hat{X}', \hat{W}')$ have their obvious meanings. We remark that the definition of an MAP can easily be extended to allow S to have a non-zero starting position.

Our duality relations are expressed in terms of the equality and adjointness of certain operators on $L^p(\pi)$. We define the following transition functions, where absorbing barriers are placed in braces following the expressions: for $x \in E$, $A \in \mathscr{E}$, an interval $I \in \tilde{\mathscr{R}}^1$, $y, z \in \mathbb{R}^1$, $n \geqq 0$, $a>0$:

$$\text{(7.1)} \quad \begin{aligned} &D_n(x, A, I, y, z) = P^x\{X_n \in A, W_n \leqq z, S_n \in I + y \mid S_0 = y\}; \\ &\hat{D}_n(x, A, I, y, z) = \hat{P}^x\{\hat{X}_n \in A, \hat{V}_n \leqq a - y, \hat{S}_n \in I + a - z \mid \hat{S}_0 = a - z\}, \quad \{0, a+\}; \\ &D'_n(x, A, I, y, z) = P^x\{X'_n \in A, W'_n \geqq a - z, S'_n \in -I + a - y \mid S'_0 = a - y\}; \\ &\hat{D}'_n(x, A, I, y, z) = \hat{P}^x\{\hat{X}'_n \in A, \hat{V}'_n \geqq y, \hat{S}'_n \in -I + z \mid \hat{S}'_0 = z\}, \quad \{0-, a\}. \end{aligned}$$

The associated operators are denoted by dropping the first two arguments e.g. $D_n(I, y, z)$ arises from $D_n(x, A, I, y, z)$.

(7.2) **Proposition.** *The following operators coincide*:

(1) $D_n(I, y, z)$,
(2) $\hat{D}_n^*(I, y, z)$,
(3) $D'_n(I, y, z)$,
(4) $\hat{D}'^*_n(I, y, z)$.

Further, if the inequalities on the right side of (7.1) are made strict and the barriers changed to $\{0-, a\}$ and $\{0, a+\}$ respectively, the above result is still true.

Proof. The result (1)=(2) follows from the corresponding result of Speed [13] by proving that for $f \in L^p$, $g \in L^q$:

$$\iint f(x)\, P^x\{X_n \in (dx'), W_n \leqq z, S_n \in I + y \mid S_0 = y\}\, g(x')\, \pi(dx)$$
$$= \iint f(x)\, \hat{P}^{x'}\{\hat{X}_n \in (dx), \hat{V}_n \leqq a - y, \hat{S}_n \in I + a - z \mid \hat{S}_0 = a - z\}\, g(x')\, \pi(dx').$$

All the other assertions are proved similarly.

Finally we remark that the case $a = \infty$ (one impenetrable or absorbing barrier only) can be formulated as (7.2) above using the analogous results in the i.i.d. case.

References

1. Arjas, E.: On a fundamental identity in the theory of semi-Markov processes. Advances Appl. Probab. **4**, 258–270 (1972).
2. Arjas, E., Speed, T.P.: Topics in Markov additive processes. To appear in Math. Scand.
3. Blumenthal, R.M., Getoor, R.K.: Markov Processes and Potential Theory. New York: Academic Press 1968.
4. Çinlar, E.: Markov additive processes. I. Z. Wahrscheinlichkeitstheorie verw. Gebiete **24**, 85–93 (1972).

5. Çinlar, E.: Markov additive processes. II. Z. Wahrscheinlichkeitstheorie verw. Gebiete **24**, 95–121 (1972).
6. Dinges, H.: Wiener-Hopf-Faktorisierung für substochastische Übergangsfunktionen in angeordneten Räumen. Z. Wahrscheinlichkeitstheorie verw. Gebiete **11**, 152–164 (1969).
7. Feller, W.: An Introduction to Probability Theory and its Applications, Volume 2, Second Edition. New York: Wiley 1971.
8. Kemeny, J.G., Snell, J.L., Knapp, A.: Denumerable Markov Chains. Princeton: Van Nostrand 1966.
9. Kingman, J.F.C.: On the algebra of queues. J. Appl. Probab. **3**, 285–326 (1966).
10. Nelson, E.: The adjoint Markoff process. Duke Math. J. **25**, 671–690 (1958).
11. Presman, E.L.: A boundary value problem for the sum of lattice random variables given on a finite regular Markov chain. Theor. Probab. Appl. **12**, 323–328 (1967). (English translation.)
12. Presman, E.L.: Factorization methods and boundary problems for sums of random variables given on Markov chains. Math. USSR Izvestija **3**, 815–852 (1969). (English translation.)
13. Speed, T.P.: A note on random walks. II. To appear in J. Appl. Probab. **10**, 218–222 (1973).
14. Spitzer, F.: Principles of Random Walk. Princeton: Van Nostrand 1964.

E. Arjas
C.O.R.E.
de Croylaan 54
B-3030 Heverlee
Belgium

T.P. Speed
Department of Probability and Statistics
The University of Sheffield
Sheffield S3 7RH
England

(Received November 15, 1972)

Stochastic Processes and their Applications 1 (1973) 369–374. © North-Holland Publ. Co.

A NOTE ON RANDOM TIMES

J.W. PITMAN and T.P. SPEED

Department of Probability and Statistics, University of Sheffield, Sheffield, UK

Received 26 June 1973

Abstract. A generalisation of the notion of stopping time is stated, and related to similar generalisations introduced by Bahadur, Kemperman, Siegmund and others with a view to permitting auxiliary experimentation to enter into the definition of stopping rule. The main aim of this note is to draw attention to the conditional independence implicit in the definitions of these writers, and briefly indicate some consequences of this.

random time
stopping time
conditional independence

1. Introduction and description of results

Suppose that $(X_n, n = 1, 2, ...)$ is a sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$, and let $\mathcal{F}_n$ denote the σ-field generated by the random variables $X_1, X_2, ..., X_n$. In the theory of optional stopping of such a process (X_n), the random times considered are commonly assumed to be *stopping times of* $(\mathcal{F}_n)$, that is to say, extended positive integer-valued random variables t such that for each $n = 1, 2, ...$, the event $\{t > n\}$ lies in the σ-field $\mathcal{F}_n$ determined by the evolution of the process up to time n. Several authors have also considered stopping procedures involving the outcomes of random experiments auxiliary to the basic process (X_n); in this connection we mention Bahadur [2], Kemperman [6], Singh [9], Siegmund [8], Chow, Robbins and Siegmund [4], and Arjas and Speed [1].

In order to provide a unified approach to the work of these authors we make the definition which follows. Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space and let us refer to an extended positive integer-valued random variable defined on Ω as a *random time*. Suppose that $(\mathcal{F}_n, n = 1, 2, ...)$ is an increasing sequence of sub-σ-fields of $\mathcal{F}$, and let $\mathcal{F}_\infty$ denote the smallest σ-field containing every $\mathcal{F}_n$, $n = 1, 2, ...$.

Definition 1.1. A random time t is a *randomised stopping time* of $(\mathcal{F}_n)$ if for each $n = 1, 2, ...,$ the event $\{t > n\}$ and the σ-field $\mathcal{F}_\infty$ are conditionally independent given $\mathcal{F}_n$.

The point of this note is to state Propositions 2.4 and 2.5 below which utilise some elementary properties of conditional independence to give various equivalent formulations of the above definition. These formulations show that the kinds of random times considered by Bahadur and Siegmund are essentially the same and just randomised stopping times according to the above definition, and it is clear that the random times considered by Kemperman and Singh are also included. We do not discuss here any applications of randomised stopping times, but refer the reader to the papers and books mentioned above.

When $\mathcal{F}_n$ is the σ-field generated by random variables $X_1, ..., X_n$, we refer to a (randomised) stopping time of $(\mathcal{F}_n)$ as a *(randomised) stopping time of* (X_n). It will be seen that a randomised stopping time of (X_n) can be thought of as being generated in the following way: an observer watches the evolution of the process (X_n) as time n increases, until a random time t when he stops observing the process; if at time k he has not yet stopped observing the process, the observer notes the value of X_k and then decides according to the outcome of some random experiment whether to stop at time k or to continue to observe the process. The random time t is a randomised stopping time of (X_n) if for each k, the outcome of the random experiment at time k and the as yet unobserved future $(X_n, k < n < \infty)$ are conditionally independent given the observed past $(X_n, 1 \leq n \leq k)$. The random time t is a stopping time of (X_n) if for each k, the decision at time k is made deterministically (and measurably) according to the past $(X_n, 1 \leq n \leq k)$.

A consequence of Proposition 2.5 is that properties of randomised stopping times associated with Markov processes or martingales can be immediately deduced from the well-known properties of stopping times of these processes. Indeed, let $(X_n, n = 1, 2, ...)$ be a sequence of random variables adapted to an increasing sequence of σ-fields $(\mathcal{F}_n, n = 1, 2, ...)$, and suppose that (X_n) is a Markov process (respectively, martingale) with respect to $(\mathcal{F}_n)$. If t is a randomised stopping time of $(\mathcal{F}_n)$ and $\mathcal{F}_n^t$ denotes the σ-field generated by $\mathcal{F}_n$ and the events $\{t = 1\}, ...$ $..., \{t = n\}$, then it follows from the equivalence of (i) and (iv) in Proposition 2.5 that (X_n) is also a Markov process (respectively, martingale) with respect to $(\mathcal{F}_n^t)$, and since t is a stopping time of $(\mathcal{F}_n^t)$, all the standard results for stopping times of Markov processes and martingales can be applied at once to randomised stopping times of these processes.

For the sake of simplicity, we have only considered here random times associated with processes whose time set is the positive integers, but most of the discussion is easily adapted to the other usual time sets.

2. Details

Let $\mathbf{N}$ denote the set of natural numbers $\{1, 2, \ldots, n, \ldots\}$. Suppose throughout that $(\Omega, \mathcal{F}, \mathbf{P})$ is a probability space and that $(\mathcal{F}_n, n \in \mathbf{N})$ is an increasing sequence of sub-σ-fields of $\mathcal{F}$, with $\mathcal{F}_\infty$ the smallest sub-σ-field of $\mathcal{F}$ containing each $\mathcal{F}_n$. The reader is referred to [7] for a treatment of conditional independence.

Remark 2.1. Recalling from the introduction the definition of a randomised stopping time of $(\mathcal{F}_n)$, we observe that alternative but equivalent definitions are obtained by replacing the set $\{t > n\}$ apearing in the definition by any one of the sets $\{t \leq n\}$, $\{t = n\}$ and $\{t \neq n\}$.

Examples 2.2. Any random time independent of $\mathcal{F}_\infty$ is a randomised stopping time of $(\mathcal{F}_n)$, and so too is any stopping time of $(\mathcal{F}_n)$. For a less trivial example, consider a real-valued process (X_n) defined on $(\Omega, \mathcal{F}, \mathbf{P})$ and suppose that Y is a real-valued random variable independent of the process (X_n). Let $t = \inf\{n : X_n \geq Y\}$. Then it is easily seen that t is a randomised stopping time of (X_n). A similarly defined randomised stopping time of a continuous time process finds an application in [3, p. 276].

For another example, suppose that (X_n) is a Markov chain and let T_n be the time of the n^{th} visit to state i. Let T be any stopping time of (X_n) and define a random time t by $t = \inf\{n : T_n \geq T\}$, so that $t - 1$ is the number of visits to state i before time T. Then t is a randomised stopping time of (T_n), as may be seen from the fact that $\{t > n\} = \{T_n < T\}$, [7, IV T 41] and the strong Markov property (cf. [5, p. 27, proof of Theorem (76)]).

For sub-σ-fields $\mathcal{A}$ and $\mathcal{B}$ of $\mathcal{F}$, let us denote by $\mathcal{A} \vee \mathcal{B}$ the smallest sub-σ-field of $\mathcal{F}$ which contains both $\mathcal{A}$ and $\mathcal{B}$. Suppose we are given sub-σ-fields $\mathcal{E}_1$, $\mathcal{E}_2$ and $\mathcal{E}$ of $\mathcal{F}$.

Lemma 2.3. *The following statements are equivalent:*

(i) $\mathcal{E}_1$ *and* $\mathcal{E}_2$ *are conditionally independent given* $\mathcal{E}$;
(ii) $P[A \mid \mathcal{E} \vee \mathcal{E}_2] = P[A \mid \mathcal{E}]$ *a.s. for every set* A *in* $\mathcal{E}_1$;
(iii) $\mathcal{E} \vee \mathcal{E}_1$ *and* $\mathcal{E} \vee \mathcal{E}_2$ *are conditionally independent given* $\mathcal{E}$;
(iv) $E\{Y \mid \mathcal{E} \vee \mathcal{E}_2\} = E\{Y \mid \mathcal{E}\}$ *a.s. for every integrable* $\mathcal{E} \vee \mathcal{E}_1$*-measurable random variable* Y.

In (ii) and (iii) the subscripts 1 and 2 can be interchanged to give further statements equivalent to (i).

Proof. The equivalence of (i) and (ii) is proved in [7, II T 51]. The further equivalence of (iii) and (iv) follows by repeated application of this result.

With the aid of Lemma 2.3, the conditional independence condition in the definition of randomised stopping time can now be rephrased in a multitude of ways. Proposition 2.4 below displays some minimal conditions for a random time to be a randomised stopping time of $(\mathcal{F}_n)$, while in Proposition 2.5 the conditional independence is exploited to the full to give some strong properties of randomised stopping times.

Suppose that t is a random time on $(\Omega, \mathcal{F}, P)$.

Proposition 2.4. *The following statements are equivalent:*
(i) t *is a randomised stopping time of* $(\mathcal{F}_n)$;
(ii) *for all* $n \in \mathbf{N}$, $P[t > n \mid \mathcal{F}_\infty] = P[t > n \mid \mathcal{F}_n]$ *a.s.*;
(iii) *for all* $n \in \mathbf{N}$, $A \in \mathcal{F}_\infty$, $P[t > n, A] = \int_{\{t>n\}} P[A \mid \mathcal{F}_n]\, dP$.
Further statements equivalent to (i) *are obtained by replacing the set* $\{t > n\}$ *appearing in* (ii) *and* (iii) *by any of* $\{t \leq n\}$, $\{t = n\}$ *and* $\{t \neq n\}$.

Proof. Let $\mathcal{S}_n$ denote the sub-σ-field of $\mathcal{F}$ generated by the event $\{t > n\}$. By definition, t is a randomised stopping time of $(\mathcal{F}_n)$ if and only if for each n in $\mathbf{N}$, the σ-fields $\mathcal{S}_n$ and $\mathcal{F}_\infty$ are conditionally independent given $\mathcal{F}_n$. The equivalence of (i), (ii) and (iii) now follows from the equivalence of (i) and (ii) in Lemma 2.3 since $\mathcal{F}_\infty \vee \mathcal{F}_n = \mathcal{F}_\infty$ and $\mathcal{S}_n \vee \mathcal{F}_n$ has an obvious simple structure. Using Remark 2.1 the remaining assertions can be proved in an identical manner.

Continuing to suppose that t is a random time on $(\Omega, \mathcal{F}, P)$, let $\mathcal{F}'_n$ denote the smallest sub-σ-field of $\mathcal{F}$ containing $\mathcal{F}_n$ and the events $\{t = 1\}, \ldots, \{t = n\}$.

§ 2. *Details*

Proposition 2.5. *The following statements are equivalent:*

(i) *t is a randomised stopping time of $(\mathcal{F}_n)$;*

(ii) *for each $n \in \mathbf{N}$ the σ-fields $\mathcal{F}^t_n$ and $\mathcal{F}_\infty$ are conditionally independent given $\mathcal{F}_n$;*

(iii) *for each $n \in \mathbf{N}$, $\mathbf{E}\{Y \mid \mathcal{F}_\infty\} = \mathbf{E}\{Y \mid \mathcal{F}_n\}$ a.s. for each integrable $\mathcal{F}^t_n$-measurable random variable Y;*

(iv) *for each $n \in \mathbf{N}$, $\mathbf{E}\{Z \mid \mathcal{F}^t_n\} = \mathbf{E}\{Z \mid \mathcal{F}_n\}$ a.s. for each integrable $\mathcal{F}_\infty$-measurable random variable Z.*

Proof. Let $\mathcal{F}_n$ denote the sub-σ-field of $\mathcal{F}$ generated by the events $\{t = 1\}, \ldots, \{t = n\}$, so that $\mathcal{F}^t_n = \mathcal{F}_n \vee \mathcal{F}_n$. The fact that $(\mathcal{F}_n)$ is an increasing sequence of σ-fields ensures that t is a randomised stopping time of $(\mathcal{F}_n)$ if and only if for each n in $\mathbf{N}$, the σ-fields $\mathcal{F}_n$ and $\mathcal{F}_\infty$ are conditionally independent given $\mathcal{F}_n$, and the proposition now follows by applying Lemma 2.3.

Put another way, the equivalence of (i) and (ii) in Proposition 2.5 means that t is a randomised stopping time of $(\mathcal{F}_n)$ if and only if there exists an increasing sequence of σ-fields $(\mathcal{G}_n)$ within $\mathcal{F}$ such that $\mathcal{F}_n \subseteq \mathcal{G}_n$, t is a stopping time of $(\mathcal{G}_n)$, and $\mathcal{G}_n$ and $\mathcal{F}_\infty$ are conditionally independent given $\mathcal{F}_n$. With this conditional independence criterion written in the form

$$\mathbf{P}[A \mid \mathcal{G}_n] = \mathbf{P}[A \mid \mathcal{F}_n] \qquad \text{a.s.}$$

for all $n \in \mathbf{N}$ and $A \in \mathcal{F}_\infty$, this is just the property required by Siegmund of his 'randomised stopping variables for $(\mathcal{F}_n)$''.

Given an increasing sequence $(\mathcal{F}_n)$ of σ-fields in a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, it may be that the probability space as it stands is not large enough to support many randomised stopping times of $(\mathcal{F}_n)$. As an extreme case, if $\mathcal{F} = \mathcal{F}_\infty$, then the only randomised stopping times of $(\mathcal{F}_n)$ are stopping times of $(\mathcal{F}_n)$. For this reason, it seems reasonable to consider the possibility of enlarging the original probability space in some way to allow room for experimentation auxiliary to $\mathcal{F}_\infty$. Consider, for example, the following procedure used by Bahadur [2]. Suppose that there is given for each n, an $\mathcal{F}_n$-measurable function a_n with $0 \leq a_n \leq 1$.

Observe the sequence of σ-fields $(\mathcal{F}_n)$ in succession, and given that the first m σ-fields have been observed, conduct an auxiliary random experiment with probability of success equal to the observed value of a_m, stopping at the time of the first success. This procedure will define a random time t on a probability space $(\Omega', \mathcal{F}', \mathbf{P}')$ constructed from $(\Omega, \mathcal{F}, \mathbf{P})$ and all necessary auxiliary experiments. This probability space will contain an isomorphic image of $\mathcal{F}$ in $\mathcal{F}'$ on which $\mathbf{P}'$ agrees with $\mathbf{P}$, and after identifying σ-fields and random variables defined on $(\Omega, \mathcal{F}, \mathbf{P})$ with their isomorphic copies in $(\Omega', \mathcal{F}', \mathbf{P}')$ it is being assumed that for each $n \in \mathbf{N}$, the event $\{t > n\}$ and the σ-field $\mathcal{F}_\infty$ are conditionally independent given $\mathcal{F}_n$ and $\{t \geq n\}$, and that

$$\mathbf{P}'[\{t = n\} \mid \mathcal{F}_n, \{t \geq n\}] = a_n \qquad \text{on } \{t \geq n\}.$$

The construction of the probability space $(\Omega', \mathcal{F}', \mathbf{P}')$ and random time t is easily formalised, and it may be shown that t is a randomised stopping time of $(\mathcal{F}_n)$ in $(\Omega', \mathcal{F}', \mathbf{P}')$, with

$$\mathbf{P}'[\{t > n\} \mid \mathcal{F}_\infty] = (1 - a_1) \dots (1 - a_n), \qquad n \in \mathbf{N}.$$

Moreover, if t^* is any randomised stopping time of $(\mathcal{F}_n)$ defined (in the obvious way) on an enlargement $(\Omega^*, \mathcal{F}^*, \mathbf{P}^*)$ of $(\Omega, \mathcal{F}, \mathbf{P})$, then a randomised stopping time of $(\mathcal{F}_n)$ having the same joint distribution with $\mathcal{F}_\infty$ as t^* can be constructed in the manner described above by taking

$$a_n = \mathbf{P}^*[t = n \mid \mathcal{F}_n] / \mathbf{P}^*[t \geq n \mid \mathcal{F}_n].$$

References

[1] E. Arjas and T.P. Speed, Markov additive processes (1972), manuscript.
[2] R.R. Bahadur, Sufficiency and statistical decision functions, Ann. Math. Statist. 25 (1954) 423–462.
[3] L. Breiman, Probability (Addison–Wesley, Reading Mass., 1968).
[4] Y.S. Chow, H. Robbins and D. Siegmund, Great Expectations: The Theory of Optimal Stopping (Houghton Mifflin, Boston, Mass., 1971).
[5] D. Freedman, Markov Chains (Holden–Day, San Francisco, Calif., 1971).
[6] J.H.B. Kemperman, The Passage Problem for a Stationary Markov Chain (Univ of Chicago Press, Chicago, Ill., 1961).
[7] P.A. Meyer, Probability and Potentials (Blaisdell, Waltham, Mass., 1966).
[8] D. Siegmund, Some problems in the theory of optimal stopping, Ann. Math. Statist. 38 (1967) 1627–1640.
[9] R. Singh, Existence of unbiased estimates, Sankhya Ser. A 26 (1964) 93–96.

J. Austral. Math. Soc. 22 (Series A) (1976), 462–468.

GEOMETRIC AND PROBABILISTIC ASPECTS OF SOME COMBINATORIAL IDENTITIES

T. P. SPEED

(Received June 9, 1975; revised October 20, 1975)

Communicated by W. D. Wallis

Abstract

For positive integers a, b and n define the combinational expression

$$A_n(a,b) = \frac{a}{a+bn}\binom{a+bn}{n}.$$

We give geometric and probabilistic interpretations of these expressions (and their multidimensional extensions) and find new, simple proofs of the convolution identities known to hold for such expressions.

1. Introduction

For non-negative integers a, b with $a + bn > 0$ let us define the combinatorial expression

$$A_n(a,b) = \frac{a}{a+bn}\binom{a+bn}{n}. \tag{1}$$

In two papers written some twenty years ago Gould (1956, 1957) discussed the above (and related) expressions. He obtained, amongst other results, the following convolution identity: for positive integral c

$$\sum_{m=0}^{n} A_m(a,b)A_{n-m}(c,b) = A_n(a+c,b). \tag{2}$$

Gould's two papers contain different approaches to this identity whilst in his recent article Gould (1974) gives yet another. We also note that Riordan (1958) presents an inductive proof of (2), as do Gould and Kaucky (1966) where further comments and extensions can be found. The proof of Blackwell and Dubins (1966) is perhaps closest in spirit to the one given below. Mohanty (1966a) extended the argument of Gould's first paper, stating and proving multinomial analogues of (2) and the related identities. He also gave a probabilistic interpretation of these facts. To formulate these results we use bold letters to denote k-tuples of non-negative integers, $\boldsymbol{b} = (b_1, b_2, \cdots, b_k)$ and $\boldsymbol{n} = (n_1, n_2, \cdots, n_k)$. Further we use the usual dot-product notation $\boldsymbol{b} \cdot \boldsymbol{n} = b_1 n_1 + b_2 n_2 + \cdots + b_k n_k$ and write $\mathbf{1} = (1, 1, \cdots, 1)$. With these preliminaries we can extend the notation above when $a + \boldsymbol{b} \cdot \boldsymbol{n} > 0$ writing

$$(1') \qquad A_{\boldsymbol{n}}(a, \boldsymbol{b}) = \frac{a}{a + \boldsymbol{b} \cdot \boldsymbol{n}} \binom{a + \boldsymbol{b} \cdot \boldsymbol{n}}{\boldsymbol{n}}$$

where $\binom{N}{\boldsymbol{n}} = N(N-1) \cdots (N - \mathbf{1} \cdot \boldsymbol{n} + 1) / \prod_1^k n_i!$ denotes the usual multinomial coefficient. In this notation one of Mohanty's results (1966a, equation (9) p. 502), the generalisation of (2) above, can be written

$$(2') \qquad \sum_{\boldsymbol{m}=0}^{\boldsymbol{n}} A_{\boldsymbol{m}}(a, \boldsymbol{b}) A_{\boldsymbol{n}-\boldsymbol{m}}(c, \boldsymbol{b}) = A_{\boldsymbol{n}}(a + c, \boldsymbol{b}).$$

Here is the summation from $m_1 = 0$ to $m_1 = n_1, \cdots, m_k = 0$ to $m_k = n_k$ as the notation suggests, and $\boldsymbol{n} - \boldsymbol{m} = (n_1 - m_1, n_2 - m_2, \cdots, n_k - m_k)$.

It is the purpose of this note to provide new proofs of these identities, the first, it is believed, that involve the geometrical interpretation of the expression (1′). After doing this we reconsider the probabilistic aspects of (2′), being somewhat more concrete than Mohanty in obtaining a random walk whose first passage probabilities to a certain hyperplane provide yet another interpretation and proof of (2′).

2. Geometric interpretation of $A_{\boldsymbol{m}}(a, \boldsymbol{b})$

It is hoped that the notation will enable us to deal with the general case (arbitrary k) almost as as easily as one would the case $k = 1$, but this will involve some slightly unusual temporary usages. We will be working in the positive orthant of the integer lattice in $k + 1$ dimensions, the coordinate variables being denoted by $X_0, X_1, \cdots, X_k$ and an arbitrary element will be denoted by $(x_0, \boldsymbol{x})$ where $\boldsymbol{x} = (x_1, x_2, \cdots, x_k)$. The first coordinate will be treated differently, and all all bold letters will be *k-tuples* of non-negative integers.

Given any k-tuple $\boldsymbol{b}$ and non-negative integer a we can define a hyperplane by the equation

$$(P) \qquad X_0 = a + (\boldsymbol{b} - \mathbf{1}) \cdot \boldsymbol{X}.$$

Clearly the point $(a + (\boldsymbol{b} - \mathbf{1}) \cdot \boldsymbol{n}, \boldsymbol{n})$ lies on (P) for any k-tuple $\boldsymbol{n}$, and we may now state the desired interpretation as follows:

PROPOSITION 1 (Mohanty). *The number of minimal lattice paths from* $(0, \mathbf{0})$ *to* $(a + (\mathbf{b} - \mathbf{1}) \cdot \mathbf{n}, \mathbf{n})$ *which do not touch the plane* (P) *until the last point, is* $A_{\mathbf{n}}(a, \mathbf{b})$.

For the case $k = 1$ this result is in implicit in Mohanty and Narayana (1961) (following by duality from their Corollary on p. 256), and appears in the present generality in Mohanty (1972). The following proof is essentially Mohanty's but we include it for completeness.

PROOF. The minimal lattice paths from $(0, \mathbf{0})$ to $(a + (\mathbf{b} - \mathbf{1}) \cdot \mathbf{n}, \mathbf{n})$ can be put into one-one correspondence with N-tuples $L = (\lambda_1, \lambda_2, \cdots, \lambda_N)$, where $N = a + \mathbf{b} \cdot \mathbf{n}$; for each $i, 1 \leqq i \leqq N$, λ_i is one of the symbols $S_0, S_1, \cdots, S_k$; for each j, $1 \leqq j \leqq k$ there are precisely n_j symbols S_j, and there are $a + (\mathbf{b} - \mathbf{1}) \cdot \mathbf{n}$ symbols S_0.

Given such an N-tuple L we can build up a minimal lattice path, starting at either end, by interpreting a symbol S_j to mean 'move one unit along the X_j-axis towards the other end'. Conversely any minimal lattice path defines a unique such N-tuple in the obvious way.

It is also clear that there are precisely $\binom{N}{\mathbf{n}}$ such N-tuples and so this is the total number of minimal lattice paths connecting $(0, \mathbf{0})$ with $(a + (\mathbf{b} - \mathbf{1}) \cdot \mathbf{n}, \mathbf{n})$. But we want the number of these which do not touch the plane (P) other than at the last point. To express this requirement as a property of the N-tuple L we need a little more notation. For each h, $1 \leqq h \leqq N$ and j, $0 \leqq j \leqq k$ define

$$\sigma_{hj} = \begin{cases} 1 & \text{if } \lambda_h = S_j \\ 0 & \text{otherwise.} \end{cases}$$

Clearly $\sum_{j=0}^{k} \sigma_{hj} = 1$. Also put $\xi_{ij} = \sum_{h=1}^{i} \sigma_{hj}$, this being the number of times the symbol S_j appears in the first i positions of L, and finally write $\boldsymbol{\xi}_i = (\xi_{i1}, \xi_{i2}, \cdots, \xi_{ik})$.

We will build up the lattice path by working backwards from the endpoint using L. After i steps have been incorporated, the X_j coordinate has reduced by $\xi_{ij} (0 \leqq j \leqq k)$ and so we are at the point $(a + (\mathbf{b} - \mathbf{1}) \cdot \mathbf{n} - \xi_{i0}, \mathbf{n} - \boldsymbol{\xi}_i)$. This point lies in the half-space defined by (P) which contains the origin for all i, $1 \leqq i \leqq N$, if, and only if,

$$a + (\mathbf{b} - \mathbf{1}) \cdot \mathbf{n} - \xi_{i0} < a + (\mathbf{b} - \mathbf{1}) \cdot (\mathbf{n} - \boldsymbol{\xi}_i) \qquad (1 \leqq i \leqq N),$$

equivalently, upon expanding and using the fact that $\xi_{i0} + \mathbf{1} \cdot \boldsymbol{\xi}_i = i$, if and only if

$$\text{(C)} \qquad \mathbf{b} \cdot \boldsymbol{\xi}_i < i \qquad (1 \leqq i \leqq N).$$

Now $\mathbf{b} \cdot \boldsymbol{\xi}_i = \sum_{j=1}^{k} b_j \xi_{ij}$ is simply a partial sum along L of numerical terms if we replace S_j by the integer b_j, $1 \leqq j \leqq k$, and S_0 by 0. With this interpretation we can immediately recognise the condition (C) and use a well-known result to deduce that of the N cyclic permutations of the N-tuple L (with the numerical components just indicated), precisely $a = N - \boldsymbol{b} \cdot \boldsymbol{n}$ have the property (C); that is, satisfy the condition that for all i ($1 \leqq i \leqq N$) the partial sums of the first i terms are less than i.

We refer to Takacs (1967) p. 4 for this result; for a geometric proof more in the spirit of the present paper, see Mohanty (1966b).

This completes the proof that the number of minimal lattice paths from $(0, \boldsymbol{0})$ to $(a + (\boldsymbol{b} - \boldsymbol{1}) \cdot \boldsymbol{n}, \boldsymbol{n})$ not touching (P) before their endpoint is $\frac{a}{\boldsymbol{N}}\binom{\boldsymbol{N}}{\boldsymbol{n}}$ where $N = a + \boldsymbol{b} \cdot \boldsymbol{n}$.

3. Derivation of identities

Let us consider the hyperplane (P) defined above, and the parallel hyperplane (c being another non-negative integer)

$$(\mathrm{P}') \qquad X_0 = a + c + (\boldsymbol{b} - \boldsymbol{1}) \cdot \boldsymbol{X}.$$

Clearly any minimal lattice path from $(0, \boldsymbol{0})$ to $(a + c + (\boldsymbol{b} - \boldsymbol{1}) \cdot \boldsymbol{n}, \boldsymbol{n})$ on (P') must hit (P) *for the first time* at $\boldsymbol{X} = \boldsymbol{m}$ for some $\boldsymbol{m}$, $\boldsymbol{0} \leqq \boldsymbol{m} \leqq \boldsymbol{n}$. Indeed there are precisely $A_{\boldsymbol{m}}(a, \boldsymbol{b})$ such paths. Each can be completed in $A_{\boldsymbol{n}-\boldsymbol{m}}(c, \boldsymbol{b})$ ways, as can be seen by viewing (P') relative to the coordinate system $(X_0', \boldsymbol{X}')$ where $X_0' = X_0 - a$ and $\boldsymbol{X}' = \boldsymbol{X} - \boldsymbol{m}$. This, plus an obvious counting argument, completes the proof of (2').

Another identity which can be derived in a similar way is:

$$(3) \qquad \sum_{\boldsymbol{m}=\boldsymbol{0}}^{\boldsymbol{n}} A_{\boldsymbol{m}}(a, \boldsymbol{b}) A_{\boldsymbol{n}-\boldsymbol{m}}(\boldsymbol{d} \cdot \boldsymbol{m}, \boldsymbol{b} + \boldsymbol{d}) = A_{\boldsymbol{n}}(a, \boldsymbol{b} + \boldsymbol{d}).$$

To get this one we consider the hyperplane (P) and the 'steeper' plane (P'') having the same X_0-intercept viz:

$$(\mathrm{P}'') \qquad X_0 = a + (\boldsymbol{b} + \boldsymbol{d} - \boldsymbol{1}) \cdot \boldsymbol{X}.$$

Any minimal lattice path from $(0, \boldsymbol{0})$ to $(a + (\boldsymbol{b} + \boldsymbol{d} - \boldsymbol{1}) \cdot \boldsymbol{n}, \boldsymbol{n})$ on (P'') must hit the hyperplane (P) *for the first time* at $\boldsymbol{X} = \boldsymbol{m}$ for some $\boldsymbol{m}$, $\boldsymbol{0} \leqq \boldsymbol{m} \leqq \boldsymbol{n}$. Again there are $A_{\boldsymbol{m}}(a, \boldsymbol{b})$ such, and each can be completed in $A_{\boldsymbol{n}-\boldsymbol{m}}(\boldsymbol{d} \cdot \boldsymbol{m}, \boldsymbol{b} + \boldsymbol{d})$ ways, as we can see by viewing (P'') relative to the coordinate system $(X_0'', \boldsymbol{X}'')$ where $X_0'' = X_0 - a - (\boldsymbol{b} - \boldsymbol{1}) \cdot \boldsymbol{m}$, $\boldsymbol{X}'' = \boldsymbol{X} - \boldsymbol{m}$. Thus (3) follows in the same way as (2').

The general identity in Mohanty (1966a) is seen to be a combination of (2') and (3). Another identity derived in Gould's papers involves the expressions

$$(4) \qquad A_n(a, b) = \frac{a}{a + bn} \frac{(a + bn)^n}{n!}.$$

Gould (1957 equation 6) shows that (2) holds with this definition of $A_n(a, b)$ and one might wonder whether a geometric interpretation exists for the entities

(4) similar to that derived for (2). I have been unable to find such an interpretation although a probabilistic one exists, and Raney (1964) gives the combinatorial interpretations of closely related expressions which lead to the proof of (2) in this case. The definition (4) also suggests a generalisation not discussed by Mohanty, namely

$$(4') \qquad A_{\boldsymbol{n}}(a, \boldsymbol{b}) = \frac{a}{a + \boldsymbol{b} \cdot \boldsymbol{n}} \frac{(a + \boldsymbol{b} \cdot \boldsymbol{n})^{\mathbf{1} \cdot \boldsymbol{n}}}{\boldsymbol{n}!}$$

where $\boldsymbol{n}! = n_1! \, n_2! \cdots n_k!$ and $\mathbf{1} \cdot \boldsymbol{n} = n_1 + n_2 + \cdots + n_k$. The coefficients $A_{\boldsymbol{n}}(a, \boldsymbol{b})$ defined by (1′) approximate those defined in (4′) when a and $\boldsymbol{b}$ are large so it is reasonable to suppose that the convolution identity (2′) also holds in this case. This is indeed true, the result being deducible (with a little effort) from Raney (1964).

4. Associated probability distributions

Let $(p_0, \boldsymbol{p})$ be a $(k+1)$-tuple with $p_0 > 0$, $p_1 > 0, \cdots, p_k > 0$ and $p_0 + p_1 + \cdots + p_k = 1$. Then if $\boldsymbol{b} \cdot \boldsymbol{p} \leqq 1$, Mohanty (1966a) proved that

$$(5) \qquad \sum_{\boldsymbol{n}=0}^{\infty} A_{\boldsymbol{n}}(a, \boldsymbol{b}) p_0^{a + (\boldsymbol{b} - \mathbf{1}) \cdot \boldsymbol{n}} \boldsymbol{p}^{\boldsymbol{n}} = 1$$

where $\boldsymbol{p}^{\boldsymbol{n}} = p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$. We will offer an alternative derivation of (5) based upon a random walk interpretation. To do this we consider the random walk on the lattice points in the positive orthant in $(k+1)$-space which begins at $(0, \mathbf{0})$ and at each step moves along the X_j axis one unit in the positive direction with probability p_j $(0 \leqq j \leqq k)$, steps being mutually independent and identically constructed.

PROPOSITION 2. *The probability that the above random walk ever hits the hyperplane (P) is π^a, where π is the smallest positive root of*

$$(6) \qquad \sum_{1}^{k} p_j x^{b_j} - x + p_0 = 0.$$

PROOF. Let us define the function, in fact a probability generating function:

$$(7) \qquad f(x) = p_0 + \sum_{1}^{k} p_j x^{b_j}$$

We will see that the probability that the walk ever hits (P) is π^a where π is a probability that the walk ever hits (P) when $a = 1$, and that π is the smallest positive root of the equation $f(x) = x$. The first assertion is an immediate consequence of the assumed independence of the steps in the walk, as the passage from $(0, \mathbf{0})$ to (P) can be viewed as a succession of independent and probabilistically identical passages from $(0, \mathbf{0})$ to $X_0 = 1 + (\boldsymbol{b} - \mathbf{1}) \cdot \boldsymbol{X}$, from this plane to $X_0 = 2 + (\boldsymbol{b} - \mathbf{1}) \cdot \boldsymbol{X}$, and so on up to (P).

Let $\pi_a(m)$ denote the probability that the walk hits the hyperplane (P) in less than m steps. Clearly $\pi_1(m) \uparrow \pi$ as $m \to \infty$. If $m > 1$ we may condition upon the outcome of the first (random) step and find that

$$\pi_1(m) = p_0 + \sum_1^k p_j \pi_{b_j}(m-1). \tag{8}$$

Now $\pi_{b_j}(m-1) \leqq \pi_{b_j}(m) \leqq [\pi_1(m)]^{b_j}$ and so we find that $\pi_1(m)$ satisfies the inequality

$$0 \leqq \pi_1(m) \leqq f(\pi_1(m)). \tag{9}$$

Letting $m \to \infty$ we see from (8) and the remarks opening this proof that $\pi = f(\pi)$ and it follows from (9) that π is the smallest such positive root.

COROLLARY. *$\pi = 1$ if and only if $\boldsymbol{b} \cdot \boldsymbol{p} \leqq 1$.*

PROOF. This is easily derived using methods well known in the theory of branching processes. See for example Harris (1963).

Let us define T to be random time, possibly infinite, which the walk takes to hit the hyperplane (P). Then we have the distribution of T involving our coefficients.

PROPOSITION 2. (i) *$P(T = a + \boldsymbol{b} \cdot \boldsymbol{n}) = A_{\boldsymbol{n}}(a, \boldsymbol{b}) p_0^{a+(\boldsymbol{b}-\mathbf{1}) \cdot \boldsymbol{n}} \boldsymbol{p}^{\boldsymbol{n}}$.*
(ii) *$P(T < \infty) = \pi^a$ where π is defined above.*

PROOF. Result (i) follows from Proposition 1 and the definition of the walk, whereas (ii) follows from the previous proposition.

COROLLARY 2. *Identity (5) holds if $\boldsymbol{b} \cdot \boldsymbol{p} \leqq 1$.*

If we denote by T_a the above random variable, then it is probabilistically obvious that the first passage time T_{a+c} should be distributed as the sum of a r.v. T_a and another, independent, r.v. T_c. This convolution property is equivalent to (2′) as is easily checked. Thus an alternative, probabilistic, proof of (2′) could be constructed. The details are left to the reader.

Finally we note that $E\{T_a\} = a/(1 - \boldsymbol{b} \cdot \boldsymbol{p})$ can be proved in a manner analogous to that used to obtain the equation for π. That is, by first deriving the equation $E\{T_a\} = aE\{T_1\}$, and then conditioning upon the outcome of the first step obtaining

$$E\{T_1\} = p_0 + \sum_1^k p_j (1 + E\{T_{b_j}\}).$$

The variance formula for T_a can be derived in a similar way.

5. Acknowledgements

I am indebted to H. W. Gould and G. N. Raney for providing references to some relevant work.

References

David Blackwell and Lester Dubins (1966), 'An elementary proof of an identity of Gould's', *Bol. Soc. Mat. Mexicana Ser.* 2 **11,** 108–110.

H. W. Gould (1956), 'Some generalisations of Vandermonde's convolution', *Amer. Math. Monthly* **63,** 84–91.

H. W. Gould (1957), 'Final analysis of Vandermonde's convolution', *Amer. Math. Monthly* **64,** 409–415.

H. W. Gould (1974), 'Coefficient identities for powers of Taylor and Dirichlet series', *Amer. Math. Monthly* **81,** 3–13.

H. W. Gould and J. Kaucky (1966), 'Evaluation of a class of binomial coefficient summations', *J. Combinatorial Theory* **1,** 233–247.

Theodore E. Harris (1963), *The theory of branching processes.* Springer-Verlag Berlin, Göttingen, Heidelberg.

S. G. Mohanty (1966a), 'Some convolutions with multinomial coefficients and related probability distributions', *SIAM Review* **8,** 501–509.

S. G. Mohanty (1966b), 'An urn problem related to the ballot problem', *Amer. Math. Monthly* **73,** 526–528.

S. G. Mohanty (1972), 'On queues involving batches', *J. Applied Probability* **9,** 430–435.

S. G. Mohanty and T. V. Narayana (1961), 'Some properties of compositions and their application to probability and statistics I', *Biom. Zeit.* **3,** 252–258.

George N. Raney (1964), 'A formal solution of $\Sigma_{i=1}^{\infty} A_i e^{B_i X} = X$', *Canad. J. Math.* **16,** 755–762.

John Riordan (1958), *An introduction to combinatorial analysis.* John Wiley and Sons, New York, London, Sydney.

Lajos Takács (1967) *Combinatorial methods in the theory of stochastic processes.* John Wiley and Sons, Inc., New York, London, Sydney.

Department of Mathematics,
University of Western Australia,
Nedlands, W.A. 6009.