AN ALGEBRAIC TREATMENT OF FLUCTUATIONS OF SUMS OF RANDOM VARIABLES

E. SPARRE ANDERSEN
AARHUS UNIVERSITY

1. Introduction

Let $X_1, \cdots, X_n$ be n symmetrically dependent random variables defined on a probability space $(\Omega, \mathfrak{F}, P)$. By symmetrically dependent it is meant that the joint distribution of $X_1, \cdots, X_n$ is invariant under permutations of the random variables.

The present paper is concerned with relations between the distributions of several functions of the variables $X_1, \cdots, X_n$.

The notion of symmetrically dependent random variables is closely connected to the concept of interchangeable random variables introduced by de Finetti [1]. De Finetti, however, assumes the existence of an infinite sequence of random variables, such that each finite subsequence is a set of symmetrically dependent random variables. It is easy to show by examples that there exists for each n a set of n random variables, which are symmetrically dependent, but such that it is not possible to extend the set to an infinite sequence of interchangeable random variables.

For interchangeable random variables it has been proved [1] that the distribution is a mixture with positive weights of distributions of independent, identically distributed random variables. This result does not hold in general for symmetrically dependent random variables as the following example shows.

Let X_1, X_2 be symmetrically dependent random variables such that $P(X_1 = 1, X_2 = 0) = P(X_1 = 0, X_2 = 1) = \frac{1}{2}$; then no mixture with positive weights of distributions of pairs of independent, identically distributed random variables yields the distribution of X_1 and X_2.

2. Symbolic convolutions and their relations

The symmetrically dependent random variables $X_1, \cdots, X_n$ will be called basic random variables. All other random variables will be defined in terms of these basic ones.

The research reported in this document has been sponsored by the Air Office of Scientific Research, OAR under Grant AF EOAR 62-70 with the European Office of Aerospace Research, United States Air Force.

The basic random variables map Ω into R^n. This mapping induces a measure P' defined on the Borel sets of R^n. Since all random variables, which we shall consider, are functions of the basic random variables, we may consider $(R^n, \mathcal{B}^n, P')$ as our basic probability space. For convenience, we shall from now on write P for P'.

Let $Y_1, \cdots, Y_k$ be random variables; we shall then denote the probability distribution of $Y_1, \cdots, Y_k$ by $P_{Y_1, \cdots, Y_k}$, such that $P_{Y_1, \cdots, Y_k}(A) = P((Y_1, \cdots, Y_k) \in A)$ for $A \in \mathcal{B}_k$.

When Y and Z are defined respectively as functions of $X_{i_1, \cdots, i_\mu}$ and $X_{j_1, \cdots, j_\nu}$ and the two sets of indices $i_1, \cdots, i_\mu$ and $j_1, \cdots, j_\nu$ are disjoint, then we shall define a symbolic convolution of P_Y and P_Z by $P_Y * P_Z = P_{Y+Z}$. If Y and Z are not defined on disjoint sets of basic random variables, but $\mu + \nu \leq n$, then we can extend the definition of the symbolic convolution. Let $Z = f(X_{j_1, \cdots, j_\nu})$ and define Z' as $f(X_{k_1, \cdots, k_\nu})$, where $k_1, \cdots, k_\nu$ and $i_1, \cdots, i_\mu$ are disjoint. Evidently, $P_Z = P_{Z'}$; we therefore define $P_Y * P_Z$ as $P_{Y+Z'}$. This is possible since the probability distribution of $Y + Z'$ is independent of the choice of $k_1, \cdots, k_\nu$. We shall also use the symbolic convolution for vector-valued random variables Y and Z.

In the case where the basic random variables $X_1, \cdots, X_n$ are independent, the symbolic convolution is the usual convolution of probability distributions.

We shall be interested in the following random variables defined in terms of the basic random variables:

(i) $S_m = X_1 + \cdots + X_m$, $m = 0, \cdots, n$,

(ii) $R_{m,k}$ = the k-th order statistics of $S_0, \cdots, S_m$, for which $R_{m,0} \leq R_{m,1} \leq \cdots \leq R_{m,m}$,

(iii) $R_{m,k}^{(j)}$ = the k-th order statistics of the random variables $S_{j+i} - S_j$, $i = 0, \cdots, m$.

THEOREM 1. *For $m = 0, 1, \cdots, n$ and $k = 0, 1, \cdots, m$ we have*

$$P_{R_{m,k},S_m} = P_{R_{k,k},S_k} * P_{R_{m-k,0},S_{m-k}}. \tag{1}$$

PROOF. Equation (1) is trivial for $k = 0$ and $k = m$. We use induction with respect to m and assume that (1) is true if m is replaced by $1, 2, \cdots$, or $m - 1$. Using the notations Y^+ and Y^- for max $(0, Y)$ and min $(0, Y)$, we have

$$P_{Y^+,Z} + P_{Y^-,Z} = P_{Y,Z} + P_{0,Z}. \tag{2}$$

If, furthermore, we adopt the conventions that $P_{Y,Z}^+ = P_{Y^+,Z}$, $P_{Y,Z}^- = P_{Y^-,Z}$, then (2) may be written as

$$P_{Y,Z}^+ + P_{Y,Z}^- = P_{Y,Z} + P_{0,Z}. \tag{3}$$

We note for later use that these operators working on P have the property that if $Y_2 \leq 0$, then

$$(P_{Y_1,Z_1} * P_{Y_2,Z_2})^+ = (P_{Y_1,Z_1}^+ * P_{Y_2,Z_2})^+, \tag{4}$$

and if $Y_2 \geq 0$, then

$$(P_{Y_1,Z_1} * P_{Y_2,Z_2})^- = (P_{Y_1,Z_1}^- * P_{Y_2,Z_2})^-. \tag{5}$$

For the proof of (4) we remark that the operator + throws the probability mass of the points where Y_1 or the "sum" of Y_1 and Y_2 is negative into 0; when Y_2 is nonpositive, it therefore makes no change if in the left-hand side of (4) we replace Y_1 by Y_1^+. The proof of (5) is analogous.

Since the random variables $X_1 + R^{(1)}_{m-1,k}$, $k = 0, \cdots, m - 1$ are the order statistics of the sums $S_1, \cdots, S_m$, we obtain, using $S_0 = 0$,

(6) $$R^+_{m,k} = (X_1 + R^{(1)}_{m-1,k-1})^+$$

and

(7) $$R^-_{m,k} = (X_1 + R^{(1)}_{m-1,k})^-.$$

From these equations we obtain, using $P_{R^{(1)}_{m-1,k}} = P_{R_{m-1,k}}$, the equations

(8) $$P^+_{R_{m,k},S_m} = (P_{X_1,X_1} * P_{R_{m-1,k-1},S_{m-1}})^+$$

and

(9) $$P^-_{R_{m,k},S_m} = (P_{X_1,X_1} * P_{R_{m-1,k},S_{m-1}})^-.$$

From the induction assumptions follow

(10) $$P_{R_{m-1,k-1},S_{m-1}} = P_{R_{k-1,k-1},S_{k-1}} * P_{R_{m-k,0},S_{m-k}}$$

and

(11) $$P_{R_{m-1,k},S_{m-1}} = P_{R_{m-k-1,0},S_{m-k-1}} * P_{R_{k,k},S_k}.$$

These expressions may be introduced in (8) and (9). We shall only treat (8), since the treatment of (9) is analogous.

Using $R_{m-k,0} \leq S_0 = 0$, we obtain by (4)

(12) $$\begin{aligned} P^+_{R_{m,k},S_m} &= (P_{X_1,X_1} * P_{R_{k-1,k-1},S_{k-1}} * P_{R_{m-k,0},S_{m-k}})^+ \\ &= ((P_{X_1,X_1} * P_{R_{k-1,k-1},S_{k-1}})^+ * P_{R_{m-k,0},S_{m-k}})^+. \end{aligned}$$

If in (8) we let $m = k$, then we obtain

(13) $$P^+_{R_{k,k},S_k} = (P_{X_1,X_1} * P_{R_{k-1,k-1},S_{k-1}})^+,$$

and using $R_{k,k} \geq 0$, we have

(14) $$P_{R_{k,k},S_k} = P^+_{R_{k,k},S_k}.$$

We therefore obtain from (12) the equation

(15) $$P^+_{R_{m,k},S_m} = (P_{R_{k,k},S_k} * P_{R_{m-k,0},S_{m-k}})^+.$$

Similarly, (9) yields

(16) $$P^-_{R_{m,k},S_m} = (P_{R_{k,k},S_k} * P_{R_{m-k,0},S_{m-k}})^-.$$

The addition of (15) and (16) and the subtraction of P_{0,S_m} give, by using (3), equation (1). The proof by induction is completed.

We shall use theorem 1 to derive the following theorem.

THEOREM 2. *For $m = 0, \cdots, n$ we have*

(17) $$P_{R_{m,m},S_m} = \sum_{(\alpha_\mu)}^{(m)} \prod_{\nu=1}^{m*} \frac{1}{\nu^{\alpha_\nu}\alpha_\nu!} (P_{S_\nu^+,S_\nu})^{\overset{*}{\alpha_\nu}},$$

and

$$(18)\qquad P_{R_{m,0},S_m} = \sum_{(\alpha_\mu)}^{(m)} \prod_{\nu=1}^{m*} \frac{1}{\nu^{\alpha_\nu}\alpha_\nu!} (P_{S_\nu^-,S_\nu})^{\alpha_\nu^*},$$

where $\sum_{(\alpha_\mu)}^{(m)}$ *denotes the summation over those* $\alpha_1, \alpha_2, \cdots$ *for which* $\alpha_\mu \geq 0$ *and* $1\alpha_1 + 2\alpha_2 + \cdots = m$, *where* Π^* *denotes the symbolic convolution of measures, and the exponent* α_ν^* *denotes the symbolic convolution of* α_ν *measures.*

PROOF. In the case where the finite sequence of symmetrically dependent random variables $X_1, \cdots, X_n$ can be extended to an infinite sequence of interchangeable random variables, the use of generating functions leads to a rather simple proof of theorem 2. We shall, however, give a proof which does not need this extra assumption.

We first remark that it follows from theorem 1 that

$$(19)\qquad \sum_{k=0}^{m} P_{R_{k,k},S_k} * P_{R_{m-k,0},S_{m-k}} = \sum_{k=0}^{m} P_{R_{m,k},S_m}, \qquad m = 0, \cdots, n.$$

Since the order statistics $R_{m,k}$, $k = 0, \cdots, m$ are locally in the sample space a rearrangement of the sums S_k, $k = 0, \cdots, m$, we have

$$(20)\qquad \sum_{k=0}^{m} P_{R_{m,k},S_m} = \sum_{k=0}^{m} P_{S_k,S_m} = \sum_{k=0}^{m} P_{S_k,S_k} * P_{0,S_{m-k}}, \qquad m = 0, \cdots, n.$$

For the proof we need the following lemma.

LEMMA. *For* $m = 0, \cdots, n$ *we have*

$$(21)\qquad \sum_{k=0}^{m} \sum_{(\alpha_\mu)}^{(k)} \prod_{\mu=1}^{k*} \frac{1}{\mu^{\alpha_\mu}\alpha_\mu!} (P_{S_\mu^+,S_\mu})^{\alpha_\mu^*} * \sum_{(\beta_\nu)}^{(m-k)} \prod_{\nu=1}^{m-k} \frac{1}{\nu^{\beta_\nu}\beta_\nu!} (P_{S_\nu^-,S_\nu})^{\beta_\nu^*} = \sum_{k=0}^{m} P_{S_k,S_k} * P_{0,S_{m-k}}.$$

Before we prove this lemma, we shall use it to prove theorem 2 by induction. Equations (17) and (18) are trivially true for $m = 0$. Assume that they are true also for $1, 2, \cdots, m-1$. We may then in all terms of the left-hand side of (21), except the first and the last one, replace the $\sum^{(k)}\cdots$ and $\sum^{(m-k)}\cdots$ expressions by $P_{R_{k,k},S_k}$ and $P_{R_{m-k,0},S_{m-k}}$. Using (20) we may replace the right-hand side of (21) by the left-hand side of (19). After these changes in (21), corresponding terms on the left and the right—except the first and last terms—cancel out, and we obtain

$$(22)\qquad \sum_{(\alpha_\mu)}^{(m)} \prod_{\mu=1}^{m*} \frac{1}{\mu^{\alpha_\mu}\alpha_\mu!} (P_{S_\mu^+,S_\mu})^{\alpha_\mu^*} + \sum_{(\alpha_\mu)}^{(m)} \prod_{\mu=1}^{m*} \frac{1}{\mu^{\alpha_\mu}\alpha_\mu!} (P_{S_\mu^-,S_\mu})^{\alpha_\mu^*} = P_{R_{m,m},S_m} + P_{R_{m,0},S_m}.$$

In (22) the first term on both sides is a probability distribution on the plane which places all the probability mass in the part of the plane where the first coordinate is nonnegative. The last term on both sides places all the probability mass on the part of the plane where the first coordinate is nonpositive. It follows that the first terms are equal and that the last terms are equal. We have thus obtained (17) and (18), and the proof of theorem 2 follows by induction.

PROOF OF LEMMA. The left-hand side of (21) can be rearranged as

$$(23)\quad \sum_{k=0}^{m} \sum_{(\alpha_\mu)}^{(k)} \sum_{(\beta_\nu)}^{(m-k)} \prod_{\mu=1}^{k*} \prod_{\nu=1}^{m-k*} \frac{1}{\mu^{\alpha_\mu}\alpha_\mu!\nu^{\beta_\nu}\beta_\nu!} (P_{S_\mu^+,S_\mu})^{\alpha_\mu^*} * (P_{S_\nu^-,S_\nu})^{\beta_\nu^*}$$

$$= \sum_{k=0}^{m} \sum_{(\alpha_\mu)}^{(k)} \sum_{(\beta_\mu)}^{(m-k)} \prod_{\mu=1}^{m*} \frac{1}{\mu^{\alpha_\mu+\beta_\mu}\alpha_\mu!\beta_\mu!} (P_{S_\mu^+,S_\mu})^{\alpha_\mu^*} * (P_{S_\mu^-,S_\mu})^{\beta_\mu^*}.$$

Introducing $\gamma_\mu = \alpha_\mu + \beta_\mu$, $\mu = 0, \cdots, m$, we see that $\gamma_\mu \geq 0$ and $1\gamma_1 + 2\gamma_2 + \cdots + m\gamma_m = m$. We may therefore rearrange (23) as

$$(24)\quad \sum_{(\gamma\mu)}^{(m)} \sum_{\alpha_1=0}^{\gamma_1} \cdots \sum_{\alpha_m=0}^{\gamma_m} \prod_{\mu=1}^{m*} \frac{1}{\mu^{\gamma_\mu}\alpha_\mu!(\gamma\mu - \alpha\mu)!} (P_{S_\mu^+,S_\mu})^{\alpha_\mu^*} * (P_{S_\mu^-,S_\mu})^{(\gamma_\mu-\alpha_\mu)*}$$

$$= \sum_{(\gamma\mu)}^{(m)} \frac{1}{\mu^{\gamma_\mu}\gamma\mu!} \prod_{\mu=1}^{m*} \sum_{\alpha_\mu=0}^{\gamma_\mu} \binom{\gamma_\mu}{\alpha_\mu} (P_{S_\mu^+,S_\mu})^{\alpha_\mu^*} * (P_{S_\mu^-,S_\mu})^{(\gamma_\mu-\alpha_\mu)*}$$

$$= \sum_{(\gamma\mu)}^{(m)} \frac{1}{\mu^{\gamma_\mu}\gamma\mu!} \prod_{\mu=1}^{m*} (P_{S_\mu^+,S_\mu} + P_{S_\mu^-,S_\mu})^{\gamma_\mu^*}$$

$$= \sum_{(\gamma\mu)}^{(m)} \frac{1}{\mu^{\gamma_\mu}\gamma\mu!} \prod_{\mu=1}^{m*} (P_{S_\mu,S_\mu} + P_{0,S_\mu})^{\gamma_\mu^*},$$

since $P_{S_\mu^+,S_\mu} + P_{S_\mu^-,S_\mu} = P_{S_\mu,S_\mu} + P_{0,S_\mu}$. After this change we may perform the rearrangements in inverse order. We then obtain for the left side of (21) the expression

$$(25)\quad \sum_{k=0}^{m} \sum_{(\alpha_\mu)}^{(k)} \prod_{\mu=1}^{k*} \frac{1}{\mu^{\alpha_\mu}\alpha_\mu!} (P_{S_\mu,S_\mu})^{\alpha_\mu^*} * \sum_{(\beta_\nu)}^{(m-k)} \sum_{\nu=1}^{m-k*} \frac{1}{\nu^{\beta_\nu}\beta_\nu!} (P_{0,S_\nu})^{\beta_\nu^*}.$$

We now consider

$$(26)\quad \sum_{(\alpha_\mu)}^{(k)} \prod_{\mu=1}^{k*} \frac{1}{\mu^{\alpha_\mu}\alpha_\mu!} (P_{S_\mu,S_\mu})^{\alpha_\mu^*}.$$

Since the S_μ's are sums of the basic random variables, we get

$$(27)\quad \prod_{\mu=1}^{k*} (P_{S_\mu,S_\mu})^{\alpha_\mu^*} = \prod_{\mu=1}^{k*} P_{S_{\mu\alpha_\mu},S_{\mu\alpha_\mu}} = P_{S_k,S_k}.$$

Using

$$(28)\quad \sum_{(\alpha_\mu)}^{(k)} \prod_{\mu=1}^{k} \frac{1}{\mu^{\alpha_\mu}\alpha_\mu!} = 1,$$

we obtain for the expression (26) the value P_{S_k,S_k}. Equation (28) may be proved by a combinatorial argument or by comparing the coefficients of x^k in the left- and the right-hand terms of

$$(29)\quad \prod_1^\infty \exp\left(\frac{1}{n} x^n\right) = \exp\left(\sum_1^\infty \frac{1}{n} x^n\right) = \exp\,(-\ln(1 - x)) = \frac{1}{1 - x}.$$

An analogous argument shows that

$$(30)\quad \sum_{(\beta_\nu)}^{(m-k)} \prod_{\nu=1}^{m-k*} \frac{1}{\nu^{\beta_\nu}\beta_\nu!} (P_{0,S_\nu})^{\beta_\nu^*} = P_{0,S_{m-k}}.$$

Using these results we obtain from (25) the right-hand side of (21). This completes the proof of the lemma.

From theorem 1 and theorem 2 follows a corollary.

COROLLARY. *For $m = 0, \cdots, n$ and $k = 0, \cdots, m$ we have*

$$P_{R_{m,k},S_m} = \sum_{(\alpha_\mu)}^{(k)} \prod_{\mu=1}^{k*} \frac{1}{\mu^{\alpha_\mu}\alpha_\mu!} (P_{S_\mu^+,S_\mu})^{*}_{\alpha_\mu} * \sum_{\beta_\nu}^{(m-k)} \prod_{\nu=1}^{m-k*} \frac{1}{\nu^{\beta_\nu}\beta_\nu!} (P_{S_\nu^-,S_\nu})^{*}_{\beta_\nu}. \tag{31}$$

In order to show the connection between theorem 1 and theorem 2 and results obtained by Wendel [3] and Spitzer [2], we shall now assume that there is given an infinite sequence $X_1, X_2, \cdots$ of interchangeable random variables. Defining for $|s| < 1$, $|t| < 1$ the generating functions

$$p(s) = \sum_{n=0}^{\infty} P_{R_{n,n},S_n} s^n, \tag{32}$$

$$q(s) = \sum_{n=0}^{\infty} P_{R_{n,0},S_n} s^n, \tag{33}$$

$$r(s, t) = \sum_{n=0}^{\infty} \sum_{m=0}^{n} P_{R_{n,m},S_n} s^n t^m, \tag{34}$$

$$a(s) = \sum_{n=1}^{\infty} \frac{1}{n} P_{S_n^+,S_n} s^n, \tag{35}$$

$$b(s) = \sum_{n=1}^{\infty} \frac{1}{n} P_{S_n^-,S_n} s^n, \tag{36}$$

it follows from theorem 1 that

$$r(s, t) = p(st) * q(s), \tag{37}$$

and from theorem 2 that

$$p(s) = e^{a(s)}, \tag{38}$$

$$q(s) = e^{b(s)}. \tag{39}$$

Equation (37) follows using the multiplication rule for power series. In order to prove (38) (the proof of (39) is analogous), we write

$$\begin{aligned} e^{a(s)} &= \exp\left(\sum_{n=1}^{\infty} \frac{1}{n} P_{S_n^+,S_n} s^n\right) = \prod_{n=1}^{\infty *} \exp\left(\frac{1}{n} P_{S_n^+,S_n} s^n\right) \\ &= \prod_{n=1}^{\infty *} \sum_{\alpha_n=0}^{\infty} \frac{1}{\alpha_n!} \frac{1}{n^{\alpha_n}} (P_{S_n^+,S_n} s^n)^{*}_{\alpha_n} \\ &= \sum_{k=0}^{\infty} \sum_{(\alpha_n)}^{(k)} \prod_{n=1}^{k} \frac{1}{n^{\alpha_n}\alpha_n!} (P_{S_n^+,S_n})^{*}_{\alpha_n} s^k. \end{aligned} \tag{40}$$

If the random variables $X_1, X_2, \cdots$ are independent, and we change from probability distributions to characteristic functions, then (37) goes into Wendel's formula [3] and (38) goes into Spitzer's formula [2].

REFERENCES

[1] B. de Finetti, "La prévision: ses lois logiques, ses sources subjectives," *Ann. Inst. H. Poincaré*, Vol. 7 (1937), pp. 1–68.
[2] F. Spitzer, "A combinatorial lemma and its application to probability theory," *Trans. Amer. Math. Soc.*, Vol. 82 (1956), pp. 323–339.
[3] J. G. Wendel, "Order statistics of partial sums," *Ann. Math. Statist.*, Vol. 31 (1960), pp. 1034–1044.