ON SOME STATISTICS FOR MEASUREMENT OF DEPENDENCE AND/OR INDEPENDENCE OF TWO RANDOM VARIABLES

Yalçın Tuncer
Ankara University, Faculty of Science, Department of Statistics
06100, Tandogan, Ankara, Turkey

Abstract

Two basic statistics seem to have been introduced in literature to measure and detect a possible quadrant dependence between two random variables. This work compares these two statistics and dwells on one of these discussing distributional aspects for the empirical case. The results presented are preliminary findings of an ongoing research on the subject.

Key Words: Probability ratio, bivariate empirical processes, simple random walk, negative binomial (Pascal) probability.

1. Introduction

Two arbitrary random variables $X$ and $Y$ are in consideration. These variables have the respective given marginal distributions $F_X(x)$ and $F_Y(y)$ and a symmetric joint distribution $F(x, y)$ at the point $(x, y) \in \mathbb{R}^2$, such that little information exist about the latter for reasons to be given later, the marginal and joint distributions are not assumed to converge concurrently to the same functional values as $x \to \pm \infty$ and (or) $y \to \pm \infty$.

When independence of $X$ and $Y$ to be emphasized, the joint distribution $F(x, y)$ will restrictively be denoted by $F^{(0)}(x, y)$, i.e.,

$$F(x, y) = F^{(0)}(x, y) = F_X(x)F_Y(y);$$

(1.1)

$F^{(1)}(x, y)$ will in general stand for quadrant dependence case (c.f., for the concept, Lehmann (1966)), i.e.,

$$F^{(1)}(x, y) \neq F_X(x)F_Y(y).$$

(1.2)

When neither $F^{(1)}(x, y)$ nor $F^{(0)}(x, y)$ are to be emphasized, the joint distribution will be denoted by $F(x, y)$. To measure and detect dependence of $X$ and $Y$, two statistics
Y. Tuncer

seem to be favored in literature which form the bases of many other alternatives used for the same purpose:

\[ i) \quad S_F(x, y) = \frac{F(x,y)}{F^{(0)}(x,y)}, \quad F^{(0)}(x,y) \neq 0, \quad (x,y) \in \mathbb{R}^2 \]

\[ ii) \quad H_F(x, y) = F(x, y) - F^{(0)}(x, y), \quad (x, y) \in \mathbb{R}^2 \]

such that \( S_F(x, y) = 1 \) and \( H_F(x, y) = 0 \) discloses in dependence of \( X \) and \( Y \) at the point \((x,y) \in \mathbb{R}^2\). To give some examples of other statistics based on \( S_F(x, y) \), the \textit{generalized likelihood (probability) ratio} is a typical one used in sequential analysis (c.f., Wald (1947), pp.37-61), i.e., \( S_F(x, y) = \frac{F^{(1)}(x,y)}{F^{(0)}(x,y)} \); also, the \textit{Kullback-Leibler separator} (c.f., Kullback-Leibler (1951)) is another example used in measurement of dependence, i.e.,

\[ s_F(x, y) = \ln S_F(x, y) = \ln \frac{F^{(1)}(x,y)}{F^{(0)}(x,y)}; \]

and we have also the \textit{concentration ratio (function)} of Cifarelli and Regazzini (1987) and of Scarsini (1991), i.e.,

\[ S_F^{(1)}(x, y) = \frac{F^{(1)}(x,y)}{F^{(0)}(x,y)}. \]

Similarly, it is well-known that

\[ \sup_{(x,y)\in \mathbb{R}^2} |H_F(x, y)| = \sup_{(x,y)\in \mathbb{R}^2} \left| F(x, y) - F^{(0)}(x, y) \right| \]

corresponds to the \textit{bivariate Kolmogorov-Smirnov statistic};

\[ [H_F(x,y)]^2 = \left[ F(x, y) - F^{(0)}(x, y) \right]^2 \]

yields the bivariate version of \textit{Cramer-von Mises statistic} (cf., Kendall and Stuart (1973), p.467); and

\[ [H_F(x,y)]^2 = \frac{\left[ F(x, y) - F^{(0)}(x, y) \right]^2}{F^{(0)}(x, y)} \]

forms the well-known \textit{chi-square statistic}.

In all of the above variations of \( S_F(x, y) \) and \( H_F(x, y) \), the statistics cannot distinguish between dependence and independence at the points \( x \rightarrow \pm \infty \) and/or \( y \rightarrow \pm \infty \), where \( F(x, y) \) on the one hand, \( F_X(x) \) and \( F_Y(y) \) on the other converge concurrently to the same limiting value. Hence, these two basic statistics have similar performances at the extremes. This means that the restriction \( F^{(0)}(x, y) \neq 0 \) for \( S_F(x, y) \) is not disadvantage for the statistic in question, because \( H_F(x, y) \) has also the same disadvantage. However, because of the major works of Kolmogorov (1933) and Smirnov
MEASUREMENT OF DEPENDENCE OF TWO RANDOM VARIABLES

(1935), and in view of time-proven applicability ease of chi-square, \( H_F(x, y) \) seem to favored and known more in applications.

In the following sections, empirical counterparts \( S_{F,n}(x, y) \) and \( H_{F,n}(x, y) \) of the dependence statistics \( S_F(x, y) \) and \( H_F(x, y) \) will be discussed; then, an attempt will made to tackle the distributional features of the former empirical statistic \( S_{F,n}(x, y) \). The results presented are preliminary findings of an ongoing research.

2. Empirical dependence statistics

When, on the other hand, all the distributions \( F_X(x), F_Y(y) \) and \( F(x, y) \) are unknown and when we have a finite-sized (i.e., \( n < \infty \)) random sample of observations \( (X_i, Y_i), i = 1, 2, \ldots, n \), on \( (X, Y) \), the following empirical counterparts of (1.3) are used:

\[
S_{F,n}(x, y) = \frac{F_n(x, y)}{F_n^{(0)}(x, y)}, \quad F_n^{(0)}(x, y) \neq 0, \quad (2.1)
\]

\[
H_{F,n}(x, y) = F_n(x, y) - F_n^{(0)}(x, y)
\]

where, by definition (cf., Gaenssler and Stute(1985) and Tuncer (1995), we have

\[
F_n(x, y) = \frac{1}{n} \sum_{i=1}^{n} I_{A \times B}(X_i, Y_i) = \frac{1}{n} \sum_{i=1}^{n} I_A(X_i) \cdot I_B(Y_i) \quad (2.2)
\]

\[
F_n^{(0)}(x, y) = F_{X,n}(x) \cdot F_{Y,n}(y) = \left( \frac{1}{n} \sum_{i=1}^{n} I_A(X_i) \right) \cdot \left( \frac{1}{n} \sum_{j=1}^{n} I_B(Y_j) \right)
\]

with the usual indicator function for an arbitrary set \( E \) being defined as

\[
I_E(\omega) = \begin{cases} 
1, & \text{if } \omega \in E, \\
0, & \text{if } \omega \notin E,
\end{cases}
\]

and \( A = (-\infty, x] \) and \( B = (-\infty, y] \).

For convenience of notation, we shall re-set

\[
\zeta_i = I_{A \times B}(X_i, Y_i) = I_A(X_i) \cdot I_B(Y_i) = \xi_i \cdot \eta_i, \\
\xi_i = I_A(X_i) \quad \text{and} \quad \eta_i = I_B(Y_i) \quad (2.3)
\]

which, as they will be easily noted, are independent Bernoulli trials yielding two-state Markovian processes

\[
Z_m = \sum_{i=1}^{m} \zeta_i, \quad T_m = \sum_{j=1}^{m} \xi_j \quad \text{and} \quad U_m = \sum_{k=1}^{m} \eta_k, \quad (2.4)
\]

\( m = 0, 1, \ldots, \) to be discussed in the next section.
Note however that for \( n > 1 \) and for any \( (x, y) \in \mathbb{R}_2 \), the events that \( S_{F,n}(x, y) = 1 \) and \( H_{F,n}(x, y) = 0 \) are analytically negligible, inasmuch as,

\[
\frac{1}{n} \sum_{k=1}^{n} \xi_k \cdot \eta_{*k} \neq \left( \frac{1}{n} \sum_{i=1}^{n} \xi_i \right) \cdot \left( \frac{1}{n} \sum_{j=1}^{n} \eta_j \right),
\]

(2.5)

unless of course all \( \xi \)'s (or all \( \eta \)'s or both) either vanish or are equal to unity — clearly, for \( n = 1 \), \( S_n(x, y) \equiv 1 \) — This point of negligibility is also supported by the fact that the set

\[ M = \{(x, y) \mid S_n(x, y) = 1, (x, y) \in \mathbb{R}_2 \} \]

is a line in \( \mathbb{R}_2 \) which has Lebesque measure zero, and therefore, for any distribution \( G(x, y) \) that is absolutely continuous with respect to Lebesque measure, the probability measure is zero:

\[
\mathcal{P}(M) = \int \int_M dG(x, y) = 0.
\]

This undoubtedly is true for continuous distributions. Corresponding to the case where the quantities \( \frac{1}{n} \sum_{k=1}^{n} \xi_k \cdot \eta_{*k} \) and \( \left( \frac{1}{n} \sum_{i=1}^{n} \xi_i \right) \cdot \left( \frac{1}{n} \sum_{j=1}^{n} \eta_j \right) \) are each obtained from two distinct samples from an identical population, Karlin and Taylor(1981, pp. 113-116) find on the other hand that the probability \( \mathcal{P}(M) \) in general is equal to \( \frac{1}{2n-1} \). This point of negligibility warrants that there is almost no information lost when such points are discarded from the analysis.

In fact, as it is the case with \( S_F(x, y) \) and \( H_F(x, y) \) at \( x \to \pm \infty \) and/or \( y \to \pm \infty \), the empirical measures \( S_{F,n}(x, y) \) and \( H_{F,n}(x, y) \) will not discern dependence from independence at points \( x_{n:n} = \max \{x_{1:n}, x_{2:n}, \ldots, x_{n:n}\} \) and/or \( y_{n:n} = \max \{y_{1:n}, y_{2:n}, \ldots, y_{n:n}\} \).

These observations yield thus negligible results.

Note that the distributions that govern \( H_{F,n}(x, y) \) are well-worked-out in literature, but this is hardly the case with \( S_{F,n}(x, y) \). The latter is a ratio of two positive integers, which may be statistically dependent on each other. As a matter of fact, when the numerator and denominator of \( S_{F,n}(x, y) \) are both estimated from the same sample, then, as will be noted in (2.3), both will be based on identical observations, so that dependence will be inevitable. The form of such a dependence will be taken up in the next section. Nonetheless, when the numerator and denominator are each computed from two distinct (independent) samples, such a dependence is eliminated, so that it may be possible to enquire into some discrete distributions, say, (cf. Johnson and Kotz (1969, p.31))

\[
\mathcal{P}[X = r/s] = (e - 1)^2 \left(e^{r+s} - 1\right)^{-2},
\]

(2.6)

which has positive finite moments and where \( r \) and \( s \) positive integers. However, when based on the same distribution, the numerator and the denominator are dependent, so that the distributions that govern the numerator and the denominator of \( S_{F,n}(x, y) \) must be studied separately, and ways must be sought to eliminate the dependence. This line of reasoning will be followed in the following pages:
MEASUREMENT OF DEPENDENCE OF TWO RANDOM VARIABLES

Letting \( N = \{1, 2, \ldots, n\} \), the set of first positive integers, then

\[
S_{F,n}(x, y) = \frac{\sum Z_{n}^*}{V_{n}^*},
\]

\[
Z_{n}^* = n \cdot Z_{n} \text{ and } V_{n} = T_{n} U_{n}
\]

where, for a given \( n \), \( Z_{n} \), \( T_{n} \) and \( U_{n} \) are as in (2.4). Accordingly, \( V_{n} \in N^2 = \{k \cdot l/ k \in N \text{ and } l \in N\} \) and \( Z_{n} \in N \), the latter holding under specific conditions to be discusses in the next section. Thus, \( S_{F,n}(x, y) \) becomes a quotient of two interdependent positive integers, and in the next section, we shall discuss the distributions of \( Z_{n} \) as eliminated from its dependence on \( V_{n} \) and the subsquent section will shortly dwell on the distribution of \( V_{n} \). Under the circumstance, it will be possible to enquire into potential uses of such distributions as in (2.6). The last point will be resumed in another article.

3. Bivariate empirical distribution \( n \cdot F_{n}(x, y) \)

Since the variable \( \zeta_{i} \) in (2.3) is a Bernoulli trial which takes the value \( \zeta_{i} = 1 \) with \( \mathcal{P}( (X_{i}, Y_{i}) \in A \times B ) = p \) and the value \( \zeta_{i} = 0 \) with \( \mathcal{P}( (X_{i}, Y_{i}) \notin A \times B ) = q \), then their sums \( Z_{n} = \sum_{i=1}^{n} \zeta_{i} \), \( (m = 0, 1, 2, \ldots) \), will obviously be a simple random walk governed by the probability \( p \). The sample size \( n \) is fixed, so that the process cannot continue indefinitely. Our concern will thus be the probability governing the sum \( Z_{n} = n \cdot F_{n}^{(1)}(x, y) = (\sum_{i=1}^{n} \zeta_{i}) \).

Adopting the vectorial notation \( \xi^{t} = (\xi_{1}, \xi_{2}, \ldots, \xi_{n}) \) and \( \eta^{t} = (\eta_{1}, \eta_{2}, \ldots, \eta_{n}) \), where the components \( \xi_{i}'s \) and \( \eta_{i}'s \), \( (i = 1, 2, \ldots, n) \), are the indicator functions defined in (2.3), it is possible to define three distinct sums which are related to the three distinct empidistributions mentioned earlier (cf., 2.4):

\[
Z_{n} = nF_{n}(x, y) = \xi^{t} \eta, \quad T_{n} = nF_{X,n}(x) = \xi^{t} \xi \quad \text{and} \quad U_{n} = nF_{Y,n} = \eta^{t} \eta,
\]

where the sum \( Z_{n} \) corresponds to the bivariate Markovian process \( \zeta_{i}'s \); the sums \( T_{n} \) and \( U_{n} \) result from the respective univariate Markovian processes \( \xi_{i}'s \) and \( \eta_{i}'s \). On the other hand, the Cauchy-Schwarz inequality as applied to the case, i.e.,

\[
(\xi^{t} \eta)^2 \leq (\xi^{t} \xi) \cdot (\eta^{t} \eta)
\]

yields the following relations between these three sums:

\[
(\xi^{t} \eta) \leq \min \left\{ (\xi^{t} \xi), (\eta^{t} \eta) \right\}, \quad (3.1)
\]

and all the more

\[
(\xi^{t} \eta) \leq (\xi^{t} \xi) \cdot (\eta^{t} \eta). \quad (3.2)
\]
The sum $Z_n$ is thus bounded from above by the random quantity $r = \min\left\{\left(\xi^t\xi\right), \left(\eta^t\eta\right)\right\}$ as in (3.1) above. As such, the sum $Z_n$ can only take values $0, 1, 2, \ldots, r$; whereas the sums $T_n$ and $U_n$ can freely take values $0, 1, 2, \ldots, n$. Since these latter sums obey Binomial probability laws with respective parameters $(n, F_X(x))$ and $(n, F_Y(y))$ (cf., Karlin and Taylor (1981), pp. 112-123; Gaensler and Stute (1985), pp.1-9), one is tempted to think that $Z_n$ is also governed by the Binomial law with parameters $(n, F(x, y))$. This, however, is in strict conflict with the random bound given by (3.1).

In fact, corresponding to the bivariate case, the random sums

$$Z_m = \sum_{i=1}^{m} \xi_i, \quad m = 0, 1, 2, \ldots, n, \quad \text{with} \quad Z_0 = \xi_0 = 0,$$

(3.3)

display a two-state simple Markov process property, such that the process $Z_m$ starts at $Z_0$ and is either absorbed at $Z_m = r, m < n$, or stopped whenever $0 < Z_n < r$. The fixed sample size $n$ and the bound $r$ play thus some important role on the stochastic behavior of the process.

From an intuitive standpoint, since $n$ is fixed and since the bound $r$ must be obeyed, Negative Binomial probability law seems to be appropriate for the issue in hand. This intuitive result is also supported by analytical methods often utilized in practice, i.e., by probability generating function technique (cf., Cox and Miller (1965) pp.22-75). In fact, the probability generating function of such a process (for analytical derivations, cf., Appendix) is

$$\Phi(t) = \left(\frac{tp}{1-tq}\right)^r = \sum_{j=0}^{\infty} t^j \left(\frac{j-1}{r-1}\right)p^r q^{j-r},$$

(3.4)

where $t \in \mathbb{R}_1$, such that $t < q^{-1}$, and, as before,

$$r = \min\left\{\sum_{i=1}^{n} \xi_i, \sum_{i=1}^{n} \eta_i\right\} > 0,$$

$$p = P(X \leq x, Y \leq y),$$

$$q = 1 - p.$$

(3.5)

The coefficient of $t^n$ in (3.4) is instrumental for and is central to the subsequent discussions on obtainement of the probability governing $Z_n$ which we aim to obtain. The coefficient corresponds to

$$P(N = n; r, p) = \binom{n-1}{r-1} p^r q^{n-r}, \quad n \geq r \geq 1.$$  

(3.6)

and $P(N = n; r, p)$ in (3.6) is the probability mass function of a random variable $N$, which represents varying sample size and yields the relative frequency of the number of observations required to obtain $Z_n = r$ is $n$. The quantity $r$ is the parameter of this mass function, which, as it will be noted in (3.1) above, itself is random. Another well-known version of (3.6) is
\[ \mathcal{P}(W = w; r, p) = \binom{w + r - 1}{w} p^r q^{w}, \ w \geq 0 \text{ and } r \geq 1. \quad (3.7) \]

which stands for the relative frequency of the waiting time \( w \) (in terms of failures) to \( Z_n = r \).

As it is noted earlier, the positive integer \( r \) is random, and therefore, the probabilities in (3.6) and (3.7) must be compounded with the distribution of this random positive integer. The distribution of \( R \) is known to be Binomial with parameters \( n \) and \( \pi \), where \( \pi \) is either \( \mathcal{P}(X \leq x) = F_X(x) \) or \( \mathcal{P}(Y \leq y) = F_Y(y) \), depending on \( \min \{F_X(x), F_Y(y)\} \). However, since \( r > 0 \), the relevant Binomial distribution is the so-called modified Binomial distribution, in which only positive Binomial observations are considered (cf., for the concept, Johnson and Kotz (1969), pp.204-209):

\[ \mathcal{P}(R^* = r; n, \pi) = \frac{\mathcal{P}(R=r; n, \pi)}{\mathcal{P}(R=0; n, \pi)} = \frac{1}{1 - (1 - \pi)^n} \binom{n}{r} \pi^r (1 - \pi)^{n-r}, \quad r = 1, 2, \ldots, n \quad (3.8) \]

where \( R^* \) is the modified Binomial variate and \( R \) is the standard Binomial variate.

The corresponding probability function is:

\[ \Phi_{R^*}(t) = \alpha + (1 - \alpha) \cdot \Phi_R(t), \]

with \( \alpha = -\frac{(1-\pi)^n}{1-(1-\pi)^n} \) and \( \Phi_R(t) = ((1 - \pi) + \pi t)^n \), so that

\[ \Phi_{R^*}(t) = \frac{(1 - \pi)^n}{1 - (1 - \pi)^n} \sum_{k=1}^{n} \frac{n!}{k!} \left( \frac{\pi}{1 - \pi} \right)^k t^k. \quad (3.9) \]

Accordingly, the compound distribution has repective probability and moment generating functions

\[ \Phi_{W^*_n}(t) = \alpha + (1 - \alpha) \cdot \Phi_{R^*}(\Phi_{W_n}(t)) \]

\[ = \frac{(1-\pi)^n}{1-(1-\pi)^n} \sum_{r=1}^{n} \binom{n}{r} \left( \frac{\pi}{1 - \pi} \right)^r (\Phi_{W_n}(t))^r \]

\[ = \frac{(1-\pi)^n}{1-(1-\pi)^n} \sum_{r=1}^{n} \binom{n}{r} \left( \frac{\pi}{1 - \pi} \right)^r \left( \frac{p}{1 - qe^t} \right)^r \quad (3.10) \]

\[ M_{W^*_n}(t) = \frac{(1 - \pi)^n}{1 - (1 - \pi)^n} \sum_{k=1}^{n} \frac{n!}{k!} \left( \frac{\pi}{1 - \pi} \right)^k \left( \frac{p}{1 - qe^t} \right)^k. \quad (3.11) \]

Hence, the mean can be calculated to be equal to

\[ \mathcal{E}(W^*_n) = \frac{nq}{p} \cdot \frac{(1 - \pi)^n}{1 - (1 - \pi)^n} \sum_{r=1}^{n} \frac{n-1}{r-1} \left( \frac{\pi}{1 - \pi} \right)^r, \quad (3.12) \]

and the variance is given by

\[ \mathcal{V}(W^*_n) = \frac{nq^2}{p^2} \cdot \frac{(1 - \pi)^n}{1 - (1 - \pi)^n} \sum_{r=1}^{n} (r+1) \left( \frac{\pi}{1 - \pi} \right)^r + \mathcal{E}(W^*_n)(1 - \mathcal{E}(W^*_n)). \quad (3.13) \]
Y. TUNCER

It is only obviously a routine matter to calculate \( E(Z_n) \) and \( V(Z_n) \) through the relation \( Z_n + W_n^* = n \). Also it should be noted that, in (3.10), instead of the standard characteristic function of \( w \), we can used the characteristic function of the modified Negative Binomial variate \( w \). where \( w = 0 \) is eliminated.

From the computational standpoint, both versions of Negative Binomial noted in (3.6) and (3.7) above are known to have some affinity with a quotient of Incomplete Beta functions, and hence Binomial distribution. In fact, the variables \( Z, W \) and \( N \) are related through \( N = Z + W, \) so that the event \( \{ Z_n < r \} \) in the Binomial case is related to the event \( \{ n < N \} \) in the Negative Binomial case, etc., (cf., for details, Kendall and Stuart (1969), p.130; Johnson and Kotz (1969), p.127). As such, computation of Negative Binomial probabilities boils down to calculation of the corresponding Beta and hence Binomial percentage points. Nonetheless, separate tables of percentage points are available (cf., Williamson and Bretherton (1963)). Furthermore, avenues of Poisson approximation (cf., Rahman (1968), p.218), of Gamma approximation (cf., Woodroofe (1975), p.109) and of Normal approximation (cf., Wilks (1962), p.274) as well as other computational possibilities (cf., Johnson and Kotz (1969), pp.127-131) also exist.

4. Distributions of \( V_n \)

Since \( V_n \) is the product of \( T_n \) and \( U_n \), both being positive integers in \( N \), and since the distribution of these latter variables are modified Binomial distributions in which \( T_n = 0 \) and \( U_n = 0 \) are eliminated, then \( V_n \) will be a positive integer in \( N^2 \) which contains some of the positive integers ranging from one to \( n^2 \) and leaves out some like \( n^2 - n + 1, n^2 - n + 2, \ldots, n^2 - 1 \). Accordingly, the distribution of \( V_n \) will be a discrete distribution in which the unit mass is concentrated at points of \( N^2 \). The distributions of \( T_n \) and \( U_n \) in the through the distributions of \( T_n \) and \( U_n \) in the usual way:

\[
P(V = v) = \sum_{t = \frac{n}{t}}^{n} \mathcal{P}(t, \frac{v}{t}) \]
\[
= \sum_{t = \frac{n}{t}}^{n} \mathcal{P}(T = t) \cdot \mathcal{P}(U = \frac{v}{t}) \]
\[
= AB \sum_{t = \frac{n}{t}}^{n} \left( \binom{n}{t} \binom{n}{v} \right) \psi^t \phi^v \tag{4.1}.
\]

where \( AB = \frac{(1-\delta)^n(1-\theta)^n}{(1-(1-\delta)^n)(1-(1-\theta)^n)} \), \( \psi = \frac{\delta}{1-\delta} \) and \( \phi = \frac{\theta}{1-\theta} \)

with \( \delta = F_X(x) \) and \( \theta = F_Y(y) \).

54
Acknowledgement

I would like to thanks Drs. Ismihan Bairamov, Yilmaz Akdi and Bilgehan Güven for helpful suggestions on the subject.

References

Appendix

Let $Z_0$ stand for the initial stage of the process, so that we may have either $Z_0 = 0$ or $Z_0 = r > 0$. In the latter case, with probability one, the process starts and ends automatically, and this will be denoted with

$$p^{(0)}_{(r,r)} = P(\text{process starts and ends at initial stage}) = 1,$$

where the superscript $(0)$ indicates the stage 0 and the subscript $(r,r)$ shows that the process starts at $r$ and ends at $r$. In this appendix, we shall be concerned with probabilities like $p^{(n)}_{(a,r)}$ and shall seek the generating function which produces them. When the process starts at $Z_0 = 0$, $Z_m$ for any $m = 1, 2, \ldots, n$ can take the values (can reach the states) $0, 1, \ldots, r$ with probabilities $p^{(m)}_{(0,0)}, p^{(m)}_{(0,1)}, \ldots, p^{(m)}_{(0,r)}$. Likewise, for an intermediary state $c$, $0 < c < r$, we have

$$p^{(m)}_{(c,r)} = P(c \leq X_1, X_2, \ldots, X_{m-1} < r, X_m = r \mid X_0 = c).$$

Clearly, since a process cannot start at $c$ and end at $r$ at the initial stage whenever $c \neq r$, we shall have

$$p^{(0)}_{(c,r)} = 0.$$

Together with $p^{(0)}_{(r,r)} = 1, p^{(0)}_{(c,r)} = 0$ forms initial conditions. Note that a markovian process has the property that the realization of $Z_s, s < t$, does not depend on the realization of $X_u, t \leq u$. Accordingly, if we start at the intermediary point $c$, i.e., $Z_0 = c$, and then the next will be

$$\zeta_1 = Z_1 = \begin{cases} c + 1, \text{with } p = \mathcal{P}((X_1, Y_1) \in A \times B) \\ c, \text{with } q = \mathcal{P}((X_1, Y_1) \in (A \times B)^c) \end{cases}, \quad (A.1)$$

where $A$ and $B$ will respectively be, say, $A = (-\infty, x]$ and $B = (-\infty, y]$, and this will be independent of the realizations of all $(m-1)$ subsequent steps. Therefore, when $Z_1 = c+1$, the $(m-1)^{th}$ step will be independent, so that we can multiply the relevant probabilities to have

$$p^{(1)}_{(c,c+1)} \cdot p^{(m-1)}_{(c+1,r)} = p^{(m)}_{(c,r)}. \quad (A.2)$$

Similarly, when $Z_1 = c$, the next steps will be independent, so that we have

$$q^{(1)}_{(c,c)} \cdot p^{(m-1)}_{(c,r)} = p^{(m)}_{(c,r)}. \quad (A.3)$$

Obviously, these two events are mutually exclusive, and these imply that. $Z_m = b$. Thus, adding (A.1) and (A.2), we obtain $p^{(m)}_{(c,r)}$

$$p^{(m)}_{(c,r)} = p^{(1)}_{(c,c+1)} \cdot p^{(m-1)}_{(c+1,r)} + q^{(1)}_{(c,c)} \cdot p^{(m-1)}_{(c,r)}$$
MEASUREMENT OF DEPENDENCE OF TWO RANDOM VARIABLES

which denotes the probability that the process starts at \( c \) and is absorbed at \( r \) at the \( m \)th stage. Clearly, \( p^{(m)}_{(c,r)} \) is an unknown function of two discrete variables \( m \) and \( c \).

By definition, the probability generating function which will produce \( p^{(m)}_{(c,r)} \) is

\[
\varphi_{(c,r)}(t) = \sum_{m=0}^{\infty} t^m p^{(m)}_{(c,r)} = t \left[ p \cdot \sum_{m=1}^{\infty} t^{m-1} p^{(m-1)}_{(c+1,r)} + q \cdot \sum_{m=1}^{\infty} t^{m-1} p^{(m)}_{(c,r)} \right]
\]

because, as stated earlier in connection with initial conditions, \( p^{(0)}_{(c,r)} = p^{(0)}_{(c+1,r)} = 0 \). As such,

\[
\varphi_{(c,r)}(t) = t \cdot p \left[ \varphi_{(c+1,r)}(t) \right] + t \cdot q \left[ \varphi_{(c,r)}(t) \right],
\]

which is a linear homogeneous difference equation in \( \varphi(t) \) with the boundary condition

\[
\varphi_{(r,r)}(t) = 1, \quad (A.4)
\]

due to the facts that \( p^{(0)}_{(r,r)} = 1 \) set initially and that \( p^{(m)}_{(r,r)} = 0 \), i.e., the process is not recurrent. A solution of this difference equation is given by

\[
\varphi_{(c+1,r)}(t) = \left( \frac{1 - t \cdot q}{t \cdot p} \right)^c \cdot \varphi_{(c,r)}(t).
\]

The equation now is free from \( m \) and is a function of \( t \) only. Thus,

\[
\varphi_{(c,r)}(t) = \left( \frac{1 - t \cdot q}{t \cdot p} \right)^c \cdot \varphi_{(0,r)}(t),
\]

i.e.,

\[
\varphi_{(r,r)}(t) = \left( \frac{1 - t \cdot q}{t \cdot p} \right)^r \cdot \varphi_{(0,r)}(t).
\]

However, by the boundary condition (A.4) above \( \varphi_{(r,r)}(t) = 1 \), so that

\[
\varphi_{(0,r)}(t) = \left( \frac{t \cdot p}{1 - t \cdot q} \right)^r,
\]

as stated in the text.

ÖZET