Applicable Analysis and Discrete Mathematics, **2** (2008), 38–50.
Available electronically at **http://pefmath.etf.bg.ac.yu**

A MARKOV - BINOMIAL DISTRIBUTION

E. Omey, J. Santos, S. Van Gulck

Let $\{X_i, i \geq 1\}$ denote a sequence of $\{0,1\}$-variables and suppose that the sequence forms a Markov Chain. In the paper we study the number of successes $S_n = X_1 + X_2 + \cdots + X_n$ and we study the number of experiments $Y(r)$ up to the *r-th* success. In the i.i.d. case S_n has a binomial distribution and $Y(r)$ has a negative binomial distribution and the asymptotic behaviour is well known. In the more general Markov chain case, we prove a central limit theorem for S_n and provide conditions under which the distribution of S_n can be approximated by a Poisson-type of distribution. We also completely characterize $Y(r)$ and show that $Y(r)$ can be interpreted as the sum of r independent r.v. related to a geometric distribution.

1. INTRODUCTION

Many papers are devoted to sequences of Bernoulli trials and they form the basis of many (known) distributions. Applications are numerous. To mention only a few:

– the one-sample runs test can be used to test the hypothesis that the order in a sample is random;

– the number of successes can be used for testing for trends in the weather or in the stock market;

– Bernoulli trials are important in matching DNA-sequences;

– the number of (consecutive) failures can be used in quality control.

For further use we suppose that each X_i takes values in the set $\{0,1\}$ and for $n \geq 1$, let $S_n = \sum_{i=1}^{n} X_i$ denote the number of successes in the sequence $(X_1, X_2, \ldots, X_n)$. If the X_i are i.i.d. with $P(X_i = 1) = p$ and $P(X_i = 0) = q = 1 - p$, it is well known that S_n has a binomial distribution $S_n \sim \mathrm{BIN}(n,p)$. In the classical theory one either calculates probabilities concerning S_n by using the binomial distribution or by using a normal- or a Poisson-approximation. A related variable

2000 Mathematics Subject Classification. 60J10, 60F05, 60K99.
Keywords and Phrases. Markov chain, generalized binomial distribution, central limit theorem.

of interest is $Y(r)$ where for $r \geq 1$ the variable $Y(r)$ counts the number of experiments until the r-th success. In the i.i.d. case it is well known that $Y(r)$ has a negative binomial distribution and that $Y(r)$ can be interpreted as the sum of i.i.d. geometrically distributed r.v.

In the section 2 below we first list the MARKOV chain properties that we need and then study S_n (section 2.1) and $Y(r)$ (section 2.2). We finish the paper with some concluding remarks.

2. MARKOV CHAINS

Let the initial distribution be given by $P(X_1 = 1) = p$ and $P(X_1 = 0) = q = 1 - p$ and for $i, j = 0, 1$, let $p_{i,j} = P(X_2 = j \mid X_1 = i)$ denote the transition probabilities. To avoid trivialities we suppose that $0 < p_{i,j} < 1$. The one-step transition matrix of the MARKOV chain is given by

$$P = \begin{pmatrix} p_{0,0} & p_{0,1} \\ p_{1,0} & p_{1,1} \end{pmatrix}.$$

We list some elementary properties of this MARKOV chain, cf. [**3**, Chapter XVI.3]. First note that the MARKOV chain has a unique stationary vector given by $(x, y) = (p_{1,0},\ p_{0,1})/(p_{0,1} + p_{1,0})$. The eigenvalues of P are $\lambda_1 = 1$ and $\lambda_2 = 1 - p_{0,1} - p_{1,0} = p_{0,0} + p_{1,1} - 1$. Note that $|\lambda_2| < 1$. By induction it is easy to show that the n-step transition matrix is given by $P^n = A + \lambda_2^n B$ where

$$A = \begin{pmatrix} x & y \\ x & y \end{pmatrix}, \quad B = \begin{pmatrix} y & -y \\ -x & x \end{pmatrix}.$$

It follows that $p_{0,0}^{(n)} = x + \lambda_2^n y$ and $p_{1,0}^{(n)} = x - \lambda_2^n x$. Using these relations for $n \geq 1$ we have

$$\begin{aligned} P(X_n = 1) &= y + \lambda_2^{n-1}(px - qy) = y - \lambda_2^{n-1}(y - p), \\ P(X_n = 0) &= x + \lambda_2^{n-1}(y - p). \end{aligned}$$

Information about moments is given in the following result.

Lemma 1. *For $n \geq 1$ we have*

(i) $E(X_n) = P(X_n = 1) = y - \lambda_2^{n-1}(y - p)$.

(ii) $\mathrm{Var}\,(X_n) = \big(y - \lambda_2^{n-1}(y - p)\big)\big(x + \lambda_2^{n-1}(y - p)\big)$.

(iii) *For $n \geq m$ we have* $\mathrm{Cov}\,(X_n, X_m) = \lambda_2^{n-m}\mathrm{Var}\,(X_m)$.

Proof. The first and the second part are easy to prove. To prove (iii), note that $E(X_n X_m) = P(X_n = 1, X_m = 1) = p_{1,1}^{(n-m)} P(X_m = 1)$. It follows that $\mathrm{Cov}\,(X_n, X_m) = \big(p_{1,1}^{(n-m)} - P(X_n = 1)\big)P(X_m = 1)$. Using the expressions obtained before we obtain

$$\begin{aligned} \mathrm{Cov}\,(X_n, X_m) &= \big(y + \lambda_2^{n-m}x - y + \lambda_2^{n-1}(y - p)\big)P(X_m = 1) \\ &= \lambda_2^{n-m}\big(x + \lambda_2^{m-1}(y - p)\big)P(X_m = 1), \end{aligned}$$

and the result follows. □

As a special case we consider the type of correlated Bernoulli trials studied in [**2**] and [**9**]. In this model we assume that $P(X_n = 1) = p$, $P(X_n = 0) = q = 1-p$ and $\rho = \rho(X_n, X_{n+1}) \neq 0$ for all $n \geq 1$. From this it follows that $\mathrm{Cov}(X_n, X_{n+1}) = \rho pq$ and that $P(X_n = 1, X_{n+1} = 1) = p(p + \rho q)$. Since $P(X_n = 1) = p$ we also find that

$$P(X_n = 0, X_{n+1} = 1) = P(X_n = 1, X_{n+1} = 0) = pq(1 - \rho).$$

It turns out that $P(X_{n+1} = j \mid X_n = i) = p_{i,j}$ $(i, j = 0, 1)$, where the $p_{i,j}$ are given by

$$P(p, \rho) = \begin{pmatrix} q + \rho p & p(1 - \rho) \\ q(1 - \rho) & p + \rho q \end{pmatrix}.$$

In this case we have $(x, y) = (q, p)$ and $\lambda_2 = \rho$. For $n \geq m$ it follows from Lemma 1 that $\rho(X_n, X_m) = \rho^{n-m}$.

2.1. THE NUMBER OF SUCCESSES S_n

2.1.1. Moments

In this section we study the number of successes $S_n = \sum_{i=1}^{n} X_i$. In our first result we study moments of S_n and extend the known i.i.d. results.

Proposition 2. (i) *We have* $E(S_n) = ny - (y - p)(1 - \lambda_2^n)/(1 - \lambda_2)$.

(ii) *We have*

$$\mathrm{Var}(S_n) = nxy(1 + \lambda_2)/(1 - \lambda_2) + \sum_{k=0}^{n-1} (A\lambda_2^k + B\lambda_2^{2k} + Ck\lambda_2^k),$$

where A, B, C *will be determined in the proof of the result.*

(iii) *If* $P = P(p, \rho)$ *we have* $E(S_n) = np$ *and*

$$\mathrm{Var}(S_n) = pq\big(n(1 + \rho) - 2\rho(1 - \rho^n)/(1 - \rho)\big)/(1 - \rho).$$

Proof. Part (i) follows from Lemma 1(i). To prove (ii) we start from $\mathrm{Var}(S_{k+1}) = Var(S_k + X_{k+1})$. Using Lemma 1, we see that

$$\mathrm{Var}(S_{k+1}) - \mathrm{Var}(S_k) = \mathrm{Var}(X_{k+1}) + 2\sum_{i=1}^{k} \lambda_2^{k+1-i}\mathrm{Var}(X_i).$$

Again from Lemma 1 we see that

$$\mathrm{Var}(X_i) = xy + a\lambda_2^{i-1} - b\lambda_2^{2i-2}$$

where $a = (y - p)(y - x)$ and $b = (y - p)^2$. Straightforward calculations show that

$$\mathrm{Var}(S_{k+1}) - \mathrm{Var}(S_k) = xy(1 + \lambda_2)/(1 - \lambda_2) + A\lambda_2^k + B\lambda_2^{2k} + Ck\lambda_2^k$$

where $C = 2a$ and

$$A = \big(a(1-\lambda_2) - 2xy\lambda_2 - 2b\big)/(1-\lambda_2),$$
$$B = b(1+\lambda_2)/(1-\lambda_2).$$

If we define $S_0 = 0$ this result holds for all $k \geq 0$. Using $\mathrm{Var}\,(S_n) = \sum_{k=0}^{n-1} \big(\mathrm{Var}\,(S_{k+1}) - \mathrm{Var}\,(S_k)\big)$, result (ii) follows. Result (iii) is easier to prove. Using Lemma 1 we have

$$\mathrm{Var}\,(S_{k+1}) - \mathrm{Var}\,(S_k) = xy\Big(1 + 2\sum_{i=1}^{k} \rho^{k+1-i}\Big) = xy(1+\rho-2\rho^{k+1})/(1-\rho).$$

The result follows by taking sums as before. □

The expression for the variance can be simplified asymptotically. We use the notation $u(n) \sim cv(n)$ to indicate that $u(n)/v(n) \to c$.

Corollary 3. *As $n \to \infty$ we have*

(i) $E(S_n) \sim ny$ *and* $\mathrm{Var}\,(S_n) \sim nxy(1+\lambda_2)/(1-\lambda_2)$.

(ii) $E(S_n) - ny \to (p-y)/(1-\lambda_2)$ *and* $\mathrm{Var}\,(S_n) - nxy(1+\lambda_2)/(1-\lambda_2) \to c$, *where*

$$c = \frac{A}{1-\lambda_2} + \frac{B}{1-\lambda_2^2} + \frac{C\lambda_2}{(1-\lambda_2)^2}\,.$$

2.1.2. Distribution of S_n

In this section we determine $p_n(k) = P(S_n = k)$. It is convenient to condition on X_n and to this end we define

$$p_n^i(k) = P(S_n = k, X_n = i),\ i = 0, 1.$$

Obviously $p_n(k) = p_n^0(k) + p_n^1(k)$. Note that $p_1^1(1) = p$, $p_1^0(0) = q$ and that $p_1^1(0) = p_1^0(1) = 0$. In the next result we show how to calculate $p_n^i(k)$ recursively.

Lemma 4. *For $n \geq 1$ we have*

(i) $p_{n+1}^0(k) = p_{0,0}p_n^0(k) + p_{1,0}p_n^1(k)$;

(ii) $p_{n+1}^1(k) = p_{0,1}p_n^0(k-1) + p_{1,1}p_n^1(k-1)$.

Proof. We have $p_{n+1}^0(k) = I + II$ where

$$I = P(S_{n+1} = k, X_{n+1} = 0, X_n = 0),$$
$$II = P(S_{n+1} = k, X_{n+1} = 0, X_n = 1).$$

Clearly $I = P(S_n = k, X_{n+1} = 0, X_n = 0)$. Now note that

$$\begin{aligned} I &= P(X_{n+1} = 0 \mid S_n = k, X_n = 0)p_n^0(k) \\ &= P(X_{n+1} = 0 \mid S_{n-1} = k, X_n = 0)p_n^0(k) \\ &= p_{0,0}p_n^0(k). \end{aligned}$$

In a similar way we find that $II = p_{1,0}p_n^1(k)$ and the first result follows. The second result can be proved in a similar way. □

For small values of n we can use Lemma 4 to obtain the p.d. of S_n. It does not seem to be easy to obtain an explicit expression for $p_n(k)$. For an alternative approach we refer to the end of section 2.2 below.

2.1.3. Central limit theorem

For fixed P and (q,p) and large values of n we can approximate the p.d. of S_n by a normal distribution. We prove the following central limit theorem.

Theorem 5. *As $n \to \infty$ we have*

$$\frac{S_n - E(S_n)}{\sqrt{\mathrm{Var}\,(S_n)}} \overset{d}{\Rightarrow} Z \sim N(0,1).$$

Proof. We prove the theorem using generating functions. For $|z| \le 1$ and $i = 0,1$ let $\Psi_n^i(z) = \sum_{k=0}^{n} p_n^i(k)z^k$ and let $\Psi_n(z) = \Psi_n^0(z) + \Psi_n^1(z)$. Furthermore, let $\Lambda_n(z) = (\Psi_n^0(z), \Psi_n^1(z))$ and note that $\Lambda_1(z) = (q, pz)$. Using Lemma 4 we have

$$\Psi_{n+1}^0(z) = p_{0,0}\Psi_n^0(z) + p_{1,0}\Psi_n^1(z),$$
$$\Psi_{n+1}^1(z) = p_{0,1}z\Psi_n^0(z) + p_{1,1}z\Psi_n^1(z).$$

Switching to matrix notation, we find that $\Lambda_{n+1}(z) = \Lambda_n(z)A(z)$, where

$$A(z) = \begin{pmatrix} p_{0,0} & p_{0,1}z \\ p_{1,0} & p_{1,1}z \end{pmatrix}.$$

It follows that $\Lambda_{n+1}(z) = \Lambda_1(z)A^n(z)$. We can find $A^n(z)$ by using the eigenvalues of $A(z)$ and to this end we use the characteristic equation

$$|A(z) - \lambda I| = \lambda^2 - \lambda(p_{0,0} + p_{1,1}z) + (p_{0,0}p_{1,1} - p_{1,0}p_{0,1})z = 0.$$

Solving this equation gives

$$\lambda_1(z) = (p_{0,0} + p_{1,1}z + \sqrt{D})/2,$$
$$\lambda_2(z) = (p_{0,0} + p_{1,1}z - \sqrt{D})/2,$$

where $D = (p_{0,0} - p_{1,1}z)^2 + 4p_{0,1}p_{1,0}z$. Now define $U = U(z)$ and $W = W(z)$ by the equations $I = A^0(z) = U + W$ and $A(z) = \lambda_1(z)U + \lambda_2(z)W$. We find that $U(z) = (A(z) - \lambda_2(z)I)/\sqrt{D}$ and $W(z) = -(A(z) - \lambda_1(z)I)/\sqrt{D}$. Using the theorem of Cayley and Hamilton we obtain for $n \ge 0$ that $A^n(z) = \lambda_1^n(z)U + \lambda_2^n(z)W$. Note that as $z \to 1$ we have $\lambda_1(z) \to 1$ and $\lambda_2(z) \to 1 - p_{0,1} - p_{1,0} = \theta$, where by assumption we have $|\theta| < 1$. Since $|\lambda_2(z)/\lambda_1(z)| \to |\theta|$ as $z \to 1$, for all

ε sufficiently small we can find $\delta = \delta(\varepsilon)$ such that $0 < \delta < 1$ and such that $|\lambda_2(z)/\lambda_1(z)| \le |\theta| + \varepsilon < 1$, for all z such that $1-\delta \le z \le 1$. We conclude that for any sequence $z_n \to 1$, we have $|\lambda_2(z_n)/\lambda_1(z_n)|^n \to 0$. From this it follows that $A^n(z_n)/\lambda_1^n(z_n) \to U(1)$ where both rows of $U(1)$ are equal to (x,y). Using $\Psi_{n+1}(z) = \Lambda_1(z)A^n(z)\,(1,1)^t$ it follows that $\Psi_{n+1}(z_n)/\lambda_1^n(z_n) \to 1$.

Now we discuss the asymptotic behaviour of $\lambda_1(z)$ as $z \to 1$. For convenience we write $\lambda(z) = \lambda_1(z)$. Note that $\lambda(z)$ satisfies $\lambda(1) = 1$ and the characteristic equation

$$\lambda^2(z) - \lambda(z)(p_{0,0} + p_{1,1}z) + (p_{0,0}p_{1,1} - p_{1,0}p_{0,1})z = 0.$$

Taking derivatives with respect z we find that

$$2\lambda(z)\lambda'(z) - \lambda'(z)(p_{0,0} + p_{1,1}z) - \lambda(z)p_{1,1} + p_{0,0}p_{1,1} - p_{1,0}p_{0,1} = 0$$

and

$$2\lambda(z)\lambda''(z) + 2\big(\lambda'(z)\big)^2 - 2\lambda'(z)p_{1,1} - \lambda''(z)(p_{0,0} + p_{1,1}z) = 0.$$

Replacing z by $z = 1$ we find that

$$2\lambda(1)\lambda'(1) - \lambda'(1)(p_{0,0} + p_{1,1}) - \lambda(1)p_{1,1} + p_{0,0}p_{1,1} - p_{1,0}p_{0,1} = 0,$$
$$2\lambda(1)\lambda''(1) + 2\big(\lambda'(1)\big)^2 - 2\lambda'(1)p_{1,1} - \lambda''(1)(p_{0,0} + p_{1,1}) = 0.$$

Since $\lambda(1) = 1$, straightforward calculations show that

$$\lambda'(1) = p_{0,1}/(p_{0,1} + p_{1,0}) = y,$$
$$\lambda''(1) = 2xy(p_{1,1} - p_{0,1})/(p_{0,1} + p_{1,0}).$$

Using the first terms of a Taylor expansion, it follows that

$$\lambda(z) = 1 + y(z-1) + \frac{1}{2}\,\lambda''(1)(z-1)^2\big(1 + o(1)\big).$$

Using twice the expansion $\log(x) = -(1-x) - (1-x)^2\big(1+o(1)\big)/2$, we obtain

$$\frac{\log(\lambda(z)) - y\log(z)}{(1-z)^2} \to \frac{1}{2}\,\big(\lambda''(1) + y - y^2\big).$$

It follows that

$$\frac{\log\big(\lambda(z)z^{-y}\big)}{(1-z)^2} \to xy\,\frac{p_{1,1} - p_{0,1}}{p_{0,1} + p_{1,0}} + xy\,\frac{1}{2} = \frac{1}{2}\,\beta,$$

where, after simplifying,

$$\beta = xy\,\frac{p_{1,1} + p_{0,0}}{p_{0,1} + p_{1,0}} = xy\,\frac{1 + \lambda_2}{1 - \lambda_2}\,.$$

To complete the proof of Theorem 5 we replace z by $z_n = z^{1/\sqrt{n}}$. In this case we find that

$$\frac{\log\big(\lambda(z_n)z_n^{-y}\big)}{n} \to \frac{1}{2}\,\beta\,\big(\log(z)\big)^2$$

and then that

$$z_n^{-yn}\,\Psi_{n+1}(z_n) \to \exp\Big(\beta\big(\log(z)\big)^2/2\Big).$$

It follows that $(S_{n+1} - ny)/\sqrt{n} \overset{d}{\Rightarrow} W$ where $W \sim N(0,\beta)$. Using Corollary 3, the proof of the theorem is finished. □

From Theorem 5 we obtain the following result.

Corollary 6. *Let $Z \sim N(0,1)$. As $n \to \infty$ we have the following results:*

(i) $(S_n - ny)/\sqrt{n\beta} \overset{d}{\Rightarrow} Z$, *where* $\beta = xy(1+\lambda_2)/(1-\lambda_2)$;

(ii) *If* $p_{0,1} = p_{1,1} = p$, *we have* $(S_n - pn)/\sqrt{npq} \overset{d}{\Rightarrow} Z$;

(iii) *If* $P = P(p,\rho)$, *we have* $(S_n - pn)/\sqrt{n\beta} \overset{d}{\Rightarrow} Z$, *where*

$$\beta = pq(1+\rho)/(1-\rho).$$

2.1.4. Poisson approximation

In the i.i.d. case it is well known how we can approximate a binomial distribution by a suitable Poisson distribution. The same can be done in the more general Markov setting.

In what follows and also in Section 2.2 we shall use the following notations. We use $U(a)$ to denote a Bernoulli r.v. with $P(U(a)=1) = a$ $(0 < a < 1)$, $V(a)$ is a Poisson(a)-variable ($a > 0$) and $G(a)$ is a geometric r.v. with $P(G(a)=k) = (1-a)a^{k-1}(k \geq 1,\ 0 < a < 1)$. Note that

$$\begin{aligned} E(z^{U(a)}) &= (1-a) + az, \\ E(z^{V(a)}) &= \exp(-a + az), \\ E(z^{G(a)}) &= (1-a)z/(1-az). \end{aligned}$$

A compound Poisson distribution with generator $G(a)$ is defined as follows. Let $G_0(a) = 0$ and for $i \geq 1$, let $G_i(a)$ denote i.i.d. copies of $G(a)$. Independent of the $G_i(a)$ let $V(b)$ denote a Poisson(b)-variable. Now consider the new r.v. $B(V(b)) = \sum_{i=0}^{V(b)} G_i(a)$. Clearly we have

$$E(z^{B(V(b))}) = \exp\big(-b + bE(z^{G(a)})\big)$$

and we say that $B\big(V(b)\big)$ has a compound Poisson distribution with generator $G(a)$.

In the next result, in the limiting distribution all r.v. involved are independent and we use the notations introduced above. In each case we take limits as $n \to \infty$.

Theorem 7. *Suppose that $np_{0,1} \to a > 0$ and $p_{1,1} \to c$ $(0 \leq c < 1)$.*

(i) *If* $c = 0$, *then* $S_n \overset{d}{\Rightarrow} U(p) + V(a)$.

(ii) *If* $0 < c < 1$, *then* $S_n \overset{d}{\Rightarrow} U(p)G(c) + B\big(V(a)\big)$.

Proof. Using the notations as in the proof of Theorem 5 we have $\Psi_{n+1}(z) = \Lambda_1(z)A^n(z)\,(1,1)^t$ where $\Lambda_1(z) = (q, pz)$ and $A^n(z) = \lambda_1^n(z)U(z) + \lambda_2^n(z)W(z)$. Recall that $U(z) = (A(z) - \lambda_2(z)I)/\sqrt{D}$ and $W(z) = -(A(z) - \lambda_1(z)I)/\sqrt{D}$ and that $\lambda_1(z) = (p_{0,0} + p_{1,1}z + \sqrt{D})/2$ and $\lambda_2(z) = (p_{0,0} + p_{1,1}z - \sqrt{D})/2$, where $D = (p_{0,0} - p_{1,1}z)^2 + 4p_{0,1}p_{1,0}z$.

Some straightforward calculations show that

$$1 - \lambda_1(z) = 2p_{0,1}(1-z)/\big(p_{0,1} + p_{1,0} + p_{1,1}(1-z) + \sqrt{D}\,\big).$$

Since by assumption $np_{0,1} \to a$ we have $p_{0,1} \to 0$ and $p_{0,0} \to 1$. By assumption we have $p_{1,1} \to c$ and hence also $p_{1,0} \to 1 - c$. It follows that $D \to (1-cz)^2$. Using $1 - cz > 0$ we readily see that $n\big(1 - \lambda_1(z)\big) \to a(1-z)/(1-cz)$. From this it follows that

$$\lambda_1^n(z) \to \theta(z) := \exp\big(-a(1-z)/(1-cz)\big).$$

Next we consider $\lambda_2^n(z)$. Clearly we have $\lambda_2(z) = \lambda_2(1)z/\lambda_1(z)$. Our assumptions imply that $\lambda_2(1) = p_{0,0} + p_{1,1} - 1 \to 0$. It follows that $\big(\lambda_2(1)z\big)^n \to 0$ and hence also that $\lambda_2^n(z) \to 0$. Using $\lambda_2(z) \to cz$, we obtain that

$$U(z) \to U^* = \begin{pmatrix} 1 & 0 \\ (1-c)/(1-cz) & 0 \end{pmatrix}$$

and hence also that that $A^n(z) \to \theta(z)U^*$. It follows that

$$\Psi_{n+1}(z) \to (q, pz)\theta(z)U^*\,(1,1)^t\,,$$

so that

$$\Psi_{n+1}(z) \to L(z) = \theta(z)\,\frac{q + z(p-c)}{1 - cz}\,.$$

It remains to identify the limit $L(z)$.

If $c = 0$, we have $L(z) = (q + pz)\exp\big(-a(1-z)\big)$. Now the interpretation is clear. Using the notations introduced before, let $U(p)$ and $V(a)$ denote independent r.v.. Clearly $L(z) = E(z^{V(a)}z^{U(p)})$. If $c = 0$, we conclude that $S_n \overset{d}{\Rightarrow} V(a) + U(p)$.

If $0 < c < 1$, we find $L(z) = \big(q + pK(z)\big)\exp\big(-a(1 - K(z)\big)$, where $K(z) = (1-c)z/(1-cz)$. Using the notations introduced before, we see that $K(z) = E(z^{G(c)})$. Let $G_0(c) = 0$ and let $G_i(c)$ denote independent copies of $G(c)$ and, independent of the other variables, let $V(a)$ denote a Poisson-variable with parameter a. The random sum $B\big(V(a)\big) = \sum_{i=0}^{V(a)} G_i(c)$ is well-defined and has generating function $E(z^{B(V(a))}) = \exp\big(-a(1 - K(z)\big)$. Finally, let $U(p)$ denote a Bernoulli variable, independent of all other variables. We conclude that $L(z) = E(z^{U(p)G(c)+B(V(a))})$ and as a consequence we find that $S_n \overset{d}{\Rightarrow} U(p)G(c) + B\big(V(a)\big)$.
□

Remark. If $c = 1$, then $L(z) = q\exp(-a)$ which is the generating function of a degenerate variable.

As a special case we have the following corollary.

Corollary 8. (Wang [**9**]) *Suppose that* $P = P(p, \rho)$.

(i) *If* $np \to u > 0$ *and* $\rho \to 0$, *then* $S_n \overset{d}{\Rightarrow} V(u)$.

(ii) *If* $np \to u > 0$ *and* $\rho \to v$ $(0 < v < 1)$, *then* $S_n \overset{d}{\Rightarrow} B(V(u(1-v)))$.

2.2. GENERALIZED NEGATIVE BINOMIAL DISTRIBUTION

In this section we invert S_n and for each $r \geq 1$ we define $Y(r)$ as the number of experiments until the r-*th* success. Since $Y(r) = \min\{n : S_n = r\}$ we have $P(S_n \leq r) = P(Y(r+1) \geq n+1)$ and $P(Y(r) = n) = P(S_n = r, X_n = 1)$. The last quantity has been studied before. Adapting the notations, for $n \geq r$ and $i = 0, 1$ we set $p_n^i(r) = P(S_n = r, X_n = i)$ and $p_n(r) = P(S_n = r)$. The corresponding generating functions will be denoted by $\Psi_r^i(z)$ and $\Psi_r(z)$. Using similar methods as before, for $n \geq r$ we obtain that

$$\begin{aligned} p_n^1(r) &= p_{1,1}p_{n-1}^1(r-1) + p_{0,1}p_{n-1}^0(r-1), \\ p_n^0(r) &= p_{0,0}p_{n-1}^0(r) + p_{1,0}p_{n-1}^1(r). \end{aligned}$$

Using generating functions, this leads to

$$\begin{aligned} \Psi_r^1(z) &= zp_{1,1}\Psi_{r-1}^1(z) + zp_{0,1}\Psi_{r-1}^0(z), \\ \Psi_r^0(z) &= zp_{0,0}\Psi_r^0(z) + zp_{1,0}\Psi_r^1(z). \end{aligned}$$

We find that

$$\begin{aligned} \Psi_r^0(z) &= \frac{zp_{1,0}}{1 - p_{0,0}z}\,\Psi_r^1(z), \\ \Psi_r^1(z) &= zk(z)\Psi_{r-1}^1(z), \end{aligned}$$

where $k(z) = p_{1,1} + zp_{0,1}p_{1,0}/(1 - p_{0,0}z)$. It follows that

$$\Psi_r^1(z) = \big(zk(z)\big)^{r-1}\Psi_1^1(z). \tag{1}$$

Using $\Psi_r(z) = \Psi_r^0(z) + \Psi_r^1(z)$ we also find that

$$\Psi_r(z) = \left(1 + \frac{zp_{1,0}}{1 - p_{0,0}z}\right)\Psi_r^1(z) = u(z)\big(zk(z)\big)^{r-1}\Psi_1^1(z), \tag{2}$$

where $u(z) = 1 + zp_{1,0}/(1 - p_{0,0}z)$.

It remains to determine $\Psi_1^1(z) = \sum\limits_{n=1}^{\infty} P(S_n = 1, X_n = 1)z^n$. Clearly we have

$$\begin{aligned} &P(S_n = 1, X_n = 1) = p, \text{ if } n = 1, \\ &P(S_n = 1, X_n = 1) = qp_{0,0}^{n-2}p_{0,1}, \text{ if } n \geq 2. \end{aligned}$$

It follows that

$$\Psi_1^1(z) = z\left(p + q\,\frac{p_{0,1}z}{1 - p_{0,0}z}\right).$$

This result can be interpreted in the following way. We use the notations as in the beginning of Section 2.1.4. Assuming that $U(s)$ and $G(t)$ are independent r.v., we have $E(z^{G(t)}) = z(1-t)/(1-tz)$ and

$$E(z^{U(s)G(t)}) = (1-s) + sz(1-t)/(1-tz). \tag{3}$$

Using these notations we can identify $k(z)$ and $\Psi_1^1(z)$. Using (3) we obtain that

$$k(z) = E(z^{U(p_{1,0})G(p_{0,0})}), \tag{4}$$

$$\Psi_1^1(z) = zE(z^{U(q)G(p_{0,0})}). \tag{5}$$

Using (1), (4) and (5) we obtain the following result.

Theorem 9. (i) *We have* $Y(1) \stackrel{d}{=} 1 + U(q)G(p_{0,0})$.

(ii) *For* $r \geq 2$, *we have*

$$Y(r) \stackrel{d}{=} \sum_{i=1}^{r-1}\big(1 + U_i(p_{1,0})G_i(p_{0,0})\big) + 1 + U(q)G(p_{0,0}),$$

where all r.v. $U(q)$, $U_i(p_{1,0})$, $G(p_{0,0})$ *and* $G_i(p_{0,0})$ *involved are independent.*

Using $\Psi_r^1(z) = zk(z)\Psi_{r-1}^1(z)$ it also follows for $r \geq 2$ that

$$Y(r) \stackrel{d}{=} Y(r-1) + 1 + U(p_{1,0})G(p_{0,0})$$

and that

$$Y(r) - Y(r-1) \stackrel{d}{=} 1 + U(p_{1,0})G(p_{0,0})$$

where $Y(r-1)$, $U(p_{1,0})$, and $G(p_{0,0})$ are independent r.v.. Together with $Y(1) \stackrel{d}{=} 1 + U(q)G(p_{0,0})$ we obtain the following probabilistic interpretation of the formulas. At the start, the first value is either a succes or a failure. If we start with a failure (which happens with probability q) we have to wait geometrically long until we have a first succes. Given another succes position in the sequence, either the next result is a succes or the next result is a failure (which happens with probability $p_{1,0}$) and then we have to wait geometrically long until we have a new success. Although we start from a sequence of (Markov-)dependent variables, it turns out that the times between consecutive successes are independent variables!

Now we take a closer look at $p_n(r) = P(S_n = r)$ and use (2). Note that $\Psi_r(1) = u(1) = 1/y$ and observe that

$$yu(z) = y + \frac{yzp_{1,0}}{1 - p_{0,0}z} = y + \frac{xzp_{0,1}}{1 - p_{0,0}z} = E(z^{U(x)G(p_{0,0})})$$

It follows that $y\Psi_r(z) = yu(z)\big(zk(z)\big)^{r-1}\Psi_1^1(z)$ and hence that

$$yP(S_n = r) = P\big(U(x)G(p_{0,0}) + Y(r) = n\big)$$

where $U(x)$, $G(p_{0,0})$ and $Y(r)$ are independent. In the next result we formulate the central limit theorem for $Y(r)$ and a Poisson-type of approximation. The results easily follow from the representation obtained in Theorem 9.

Corollary 10. (i) *As $r \to \infty$, we have*

$$\frac{Y(r) - r/y}{\sqrt{rp_{1,0}p_{1,1}/p_{0,1}^2}} \overset{d}{\Rightarrow} Z \sim N(0,1).$$

(ii) *If $rp_{1,0} \to a$ and $p_{0,0} \to c$, $0 \le c < 1$, then $Y(r) - r \overset{d}{\Rightarrow} B\big(V(a)\big)$.*

Remark. For $r \ge 0$, let $M(r) = \max\{n : S_n \le r\}$ and $K(r) = \max\{n : Y(n) \le r\}$. Clearly $M(r) = n$ if and only if $Y(r+1) = n+1$ so that $Y(r+1) = M(r)+1$. Using $\{M(r) \ge n\} = \{S_n \le r\} = \{Y(r+1) \ge n+1\}$ and $\{K(r) \ge n\} = \{Y(n) \le r\}$, it follows that $\{K(n) \le r\} = \{Y(r+1) \ge n+1\}$. As a consequence, we find that $K(n) \overset{d}{=} S_n$. Now note that $K(r)$ corresponds to the renewal counting process associated with the sequence $A, A_1, A_2, ...$ where $A \overset{d}{=} 1+U(q)G(p_{0,0})$ and where the A_i are i.i.d. random variables with $A_i \overset{d}{=} 1 + U(p_{1,0})G(p_{0,0})$. Standard (delayed) renewal theory could now be used to prove the central limit theorem for S_n.

3. CONCLUDING REMARKS

1. Earlier we have observed that for $n \ge 1$, $P(X_n = 1) = y - \lambda_2^{n-1}(y-p)$ and $P(X_n = 0) = x + \lambda_2^{n-1}(y-p)$. From this it is easy to see that for $n, m \ge 1$ we have

$$P(X_{n+m} = 1) = y(1-\lambda_2^n) + \lambda_2^n P(X_m = 1),$$
$$P(X_{n+m} = 0) = x(1-\lambda_2^n) + \lambda_2^n P(X_m = 0).$$

Recall that (x, y) was the stationary vector. In what follows we assume that $\lambda_2 > 0$. Now let $B_n \sim U(\lambda_2^n)$ and $B^\circ \sim U(y)$ denote Bernoulli r.v.. The formulas obtained above show that $X_{n+m} \overset{d}{=} (1-B_n)B^\circ + B_n X_n$ where B_n is independent of B° and X_n. In particular it follows that $\{X_n\}$ satisfies the stochastic difference equation: $X_{n+1} \overset{d}{=} (1-B_1)B^\circ + B_1 X_n$.

2. A correlated binomial distribution has been introduced and studied in [**1**], [**4**], [**6**], [**7**]. Examples and applications can be found e.g. in quality control, cf. [**5**]. One of the models can be described as follows. Let $X, X_1, X_2, \dots, X_n$ denote i.i.d. Bernoulli variables with $X \sim U(p)$ and let $U(\alpha)$ and $U(\beta)$ denote independent Bernoulli variables, independent of the X_i. Now define $Y_i = U(\alpha)X_i + (1 - U(\alpha))U(\beta)$ and the corresponding sum

$$T_n = \sum_{i=1}^n Y_i = U(\alpha)S_n + n\big(1-U(\alpha)\big)U(\beta).$$

The following interpretation can be given. In quality control we can decide to check all produced units. The alternative is to check just one unit and then accept or reject all produced units. In the first scenario S_n counts the number of defects or successes. In the second scenario we conclude that we have either 0 or n defects or successes. Clearly $S_n \sim BIN(p,n)$ and for T_n we have $P(T_n = k) = \alpha P(S_n = k) + (1-\alpha)P(nU(\beta) = k)$. It follows that

$$P(T_n = k) = \alpha P(S_n = k) \text{ for } 1 \le k \le n-1,$$
$$P(T_n = 0) = \alpha P(S_n = 0) + (1-\alpha)(1-\beta),$$
$$P(T_n = n) = \alpha P(S_n = n) + (1-\alpha)\beta.$$

As to Y_i we have $Y_i \stackrel{d}{=} U(\lambda)$ where $\lambda = \alpha p + (1-\alpha)\beta$. For the variance $Var(Y_i) = \sigma^2$ we find that

$$\sigma^2 = \alpha(1-\alpha)(p-\beta)^2 + \alpha p(1-p) + (1-\alpha)\beta(1-\beta).$$

For $i \neq j$ we obtain that $\mathrm{Cov}(Y_i, Y_j) = \alpha(1-\alpha)(p-\beta)^2 + (1-\alpha)\beta(1-\beta)$. It follows that $\rho(Y_i, Y_j) = \rho = 1-\alpha$ if $\beta = p$. If $\beta \neq p$ we find that $\rho(Y_i, Y_j) = \rho = 1/(c+1)$ where

$$c = \alpha p(1-p)/\big(\alpha(1-\alpha)(p-\beta)^2 + (1-\alpha)\beta(1-\beta)\big).$$

We see that $\rho(Y_i, Y_j)$ is the same for all $i \neq j$. Clearly this implies that $\mathrm{Var}(T_n) = n\big(1 + (n-1)\rho/2\big)\sigma^2$. In the next result we formulate two asymptotic results.

Proposition 11. (i) *As* $n \to \infty$,

$$\Big(T_n - \big(1 - U(\alpha)\big)\, nU(\beta) - npU(\alpha)\Big)\Big/\sqrt{np(1-p)} \stackrel{d}{\Rightarrow} B(\alpha)Z,$$

where $Z \sim N(0,1)$.

(ii) *If* $np \to a > 0$ *and* $\beta \to 0$, *then* $T_n \stackrel{d}{\Rightarrow} Y$ *where* $Y \stackrel{d}{=} U(\alpha)V(a)$ *and* $U(\alpha)$ *and* $V(a)$ *are independent.*

3. From the physical point of view it seems reasonable to study random vectors $(X_1, X_2, \ldots, X_n)$ with a joint probability distribution of the following form: for $x_i = 0, 1$ we have

$$P(X_i = x_i, i = 1, 2, \ldots, n) = C \exp\Big(\alpha \sum_{i=1}^{n} a_i x_i + \beta \sum_{i=1}^{n} \sum_{j=1}^{n} a_{i,j} x_i x_j\Big).$$

The second term represents the interaction between particles. If $\beta = 0$, the X_i are independent and no interaction appears.

4. In our next paper [**8**] we study runs, singles and the waiting time until the first run of r consecutive successes for Bernoulli-sequences that are generated by a Markov chain.

Acknowledgement. The authors are grateful to the referees for their detailed comments and useful suggestions to improve the paper.

REFERENCES

1. P. Altham: *Two generalizations of the binomial distribution.* Applied Statistics, **27** (1978), 162–167.
2. A. W. F. Edwards: *The meaning of binomial distribution.* Nature, London, **186** (1960), 1074.
3. W. Feller: *An Introduction to Probability Theory and its Applications*, Vol. I (3rd edition). John Wiley and Sons, New York, 1968.
4. L. L. Kupper, J. K. Haseman: *The use of the correlated binomial model for the analysis of certain toxicological experiments.* Biometrics, **34** (1978), 69–76.
5. C. D. Lai, C. D., K. Govindaraju, M. Xie: *Effects of correlation on fraction non-conforming statistical process control procedures.* Journal of Applied Statistics, **25** (**4**) (1998), 535–543.
6. R. W. Madsen: *Generalized binomial distributions.* Comm. in Statistics: Theory and Methods, **22** (**11**)(1993), 3065–3086.
7. S. A. Mingoti: *A note on sample size required in sequential tests for the generalized binomial distribution.* Journal of Applied Statistics, **30** (**8**) (2003), 873–879.
8. E. Omey, S. Van Gulck: *Runs, singles and waiting times for Bernoulli trials governed by a Markov chain.* Forthcoming (2008).
9. Y. H. Wang: *On the limit of the Markov binomial distribution.* J. Appl. Prob., **18** (1981) 937–942.

(Received July 2, 2007)
(Revised October 5, 2007)

E. Omey and S. Van Gulck,
EHSAL,
Stormstraat 2, 1000 Brussels,
Belgium
E–mails: edward.omey@ehsal.be
stefan.vangulck@ehsal.be

J. Santos,
Public University of Navarre,
31006 Pamplona,
Spain
E–mail: santos@si.unavarra.es