Ann. Inst. Statist. Math.
39 (1987), Part A, 85-102

ON SUM OF 0-1 RANDOM VARIABLES
I. UNIVARIATE CASE

KEI TAKEUCHI AND AKIMICHI TAKEMURA

(Received Nov. 28, 1985; revised Mar. 12, 1986)

Summary

Distribution of sum of 0-1 random variables is considered. No assumption is made on the independence of the 0-1 variables. Using the notion of "central binomial moments" we derive distributional properties and the conditions of convergence to standard distributions in a clear and unified manner.

1. Introduction

Let $X_1, \cdots, X_n$ be 0-1 random variables and let $S_n = X_1 + \cdots + X_n$ be the sum. In this article we discuss the distribution of S_n. The main point of this article is that we do not assume any condition on dependence among X_i's. In usual discussions on sum of random variables X_i's are assumed to be independent or close to being independent. Clearly some simplifying assumption is needed. Our simplifying assumption is only on the marginal distribution of X_i's.

When X_i's are 0-1 random variables, S_n only takes values $0, 1, \cdots, n$. In this case there is an explicit relationship between the probability distribution of S_n and its factorial moments as shown in (2.3) and (2.7). Therefore we can discuss the distribution of S_n in terms of its factorial moments. Actually we use a one-to-one function of factorial moments, which we call "central binomial moments". It will be shown that central binomial moments are especially useful when the distribution of S_n is approximated by standard distributions, e.g., binomial, Poisson, or normal distributions.

For discussing approximations by standard distributions we use expansions based on orthogonal polynomials associated with the standard distributions. These are Krawtchouk polynomials for binomial, Charlier polynomials for Poisson, and Hermite polynomials for normal distribu-

Key words and phrases: Binomial distribution, central binomial moments, finite exchangeability, orthogonal polynomials.

tions. The approximation theory using these polynomials are fully discussed in Takeuchi [14]. Other useful references include Kendall and Stuart [8], Ord [10], and Johnson and Kotz [6].

Our development here has a rather close connection with the literature on (finite) exchangeability. This is because we can assume the exchangeability among X_i's without loss of generality as far as the distribution of S_n is concerned. This point is discussed in Galambos [4] in detail. Although there is an extensive literature on infinite exchangeability, literature on finite exchangeability is rather scarce. Kendall [7] is notable in this respect. In fact our development in Section 4 partly overlaps with Kendall [7]. More recently Diaconis and Freedman [2] gave a clear discussion on finite exchangeable sequences which can be extended to longer (but finite) exchangeable sequences. In this article we are not concerned about infinite exchangeability or extending finite exchangeability. For our discussion therefore it would be more precise to consider triangular array of 0–1 random variables $X_{i,n}$ and write $S_n = X_{1,n} + \cdots + X_{n,n}$. Since this should be clear from the context, we do not repeat this point later. Actually, limits of finite exchangeable sequences and infinite exchangeable sequence can be quite different. See the discussion following Theorem 5.1 for example.

It is interesting to note that Watanabe [16] already gave a very detailed discussion of the case where the sample mean S_n/n has a limiting distribution. In this article we do not discuss this type of convergence in distribution.

In Section 2 we set up appropriate definitions and notations. In Section 3 we discuss approximations by binomial distribution. Convergence to Poisson distribution is discussed in Section 4 and convergence to normal distribution is discussed in Section 5. Generalization to multivariate case is subjects of a subsequent paper.

2. Notations and definitions

In this section we prepare definitions and appropriate notations for quantities used in this article. The key quantity for our discussion is "central binomial moment" defined in (2.4). It can be interpreted to indicate deviation from independence. We also discuss generating functions useful in the subsequent analysis. Finally we state a lemma which treats the convergence in distribution in terms of convergence of moments.

Let $X_1, \cdots, X_n$ be random variables taking either 0 or 1. No assumption is made on the dependence among X_i's. Let

$$\Pr(X_{i_1}=1, \cdots, X_{i_k}=1) = p_{i_1 \cdots i_k} . \tag{2.1}$$

Their "average" is denoted as

$$p_n(k)=\frac{1}{\binom{n}{k}}\sum_{i_1<\cdots<i_k} p_{i_1\cdots i_k}\,, \tag{2.2}$$

for $k\geqq 1$, and $p_n(0)=1$ for all n. Note that $p_n(k)=\Pr(X_1=1,\cdots,X_k=1)$ if X_i's are exchangeable. $\binom{n}{k}p_n(k)$ is called "binomial moment" by Galambos [4] in view of Lemma 2.1 below. Because $S_n=X_1+\cdots+X_n$ is invariant under the random permutation of the indices, we could assume the exchangeability among X's, as far as the distribution of S_n is concerned. This point is already discussed in Introduction. For a full discussion see Section 3.2 of Galambos [4]. For any nonnegative integer k let $x^{(k)}=x(x-1)\cdots(x-k+1)$. The k-th factorial moment of S_n is denoted as $\mu_{(k)}=E(S_n^{(k)})$. Then we have

LEMMA 2.1.

$$\mu_{(k)}=n^{(k)}p_n(k)\,. \tag{2.3}$$

PROOF. Lemma 1.4.1 of Galambos [4].

As the distribution of S_n is determined by its factorial moments, the lemma shows that it is completely specified by the binomial moments as well.

Now we define the following key quantity, which we call "central binomial moment".

DEFINITION 2.1.

$$q_n(0)=1\,,\qquad q_n(k)=\sum_{j=0}^{k}(-1)^j\binom{k}{j}p^j p_n(k-j)\,,\quad k\geqq 1\,, \tag{2.4}$$

where $p=p_n(1)$.

For example

$$q_n(2)=p_n(2)-p^2\,,\qquad q_n(3)=p_n(3)-3pp_n(2)+2p^2\,.$$

The algebraic relation between $p_n(k)$ and $q_n(k)$ is the same as the relation between moments about the origin and moments about the mean, if $p_n(k)$ is regarded as k-th moment about the origin and $q_n(k)$ is regarded as k-th moment about the mean. Another justification of the term "central binomial moments" can be given as follows.

LEMMA 2.2. *Suppose that the distribution of $X_1,\cdots,X_n$ is a mixture of independent Bernoulli trials with random success probability* P. *Then*

$$p_n(k)=E(\mathrm{P}^k)\ ,\qquad q_n(k)=E(\mathrm{P}-E(\mathrm{P}))^k\ . \tag{2.5}$$

PROOF. The first relation follows from

$$\begin{aligned}p_n(k)&=\Pr(X_1=1,\cdots,X_k=1)\\&=E(\Pr(X_1=1,\cdots,X_k=1\mid \mathrm{P}))=E(\mathrm{P}^k)\ .\end{aligned}$$

The second equation is obvious.

In any case the relation (2.4) can be readily inverted to yield

$$\begin{aligned}p_n(k)&=\sum_{j=0}^{k}\binom{k}{j}p^j q_n(k-j)\\&=p^k+\binom{k}{2}p^{k-2}q_n(2)+\binom{k}{3}p^{k-3}q_n(3)+\cdots\ .\end{aligned} \tag{2.6}$$

This relation can be interpreted as follows: (i) the first term on the right hand side approximates the left hand side by the independent Bernoulli trials, (ii) the second term takes care of the dependence between $\binom{k}{2}$ pairs of X's, (iii) the third term takes care of the dependence among $\binom{k}{3}$ triples of X's, etc. Therefore $q_n(k)$ can be regarded as representing k-th order dependence among X's, when the lower order dependence has been taken into consideration. Clearly $q_n(k)=0$, $k\geqq 1$ if and only if S_n is binomial.

Now we discuss generating functions of factorial moments and the central binomial moments. Let the probability generating function of S_n be

$$G_n(\theta)=\sum_{k=0}^{n}\theta^k \Pr(S_n=k)\ .$$

Then $M_n(\theta)=G_n(1+\theta)$ is the factorial moment generating function:

$$M_n(\theta)=\sum_{k=0}^{n}(\theta^k/k!)\mu_{(k)}=\sum_{k=0}^{n}\binom{n}{k}p_n(k)\theta^k\ . \tag{2.7}$$

See Kendall and Stuart [8], Section 3.11. Note that in our case there is no question about the convergence because the series is finite. The central binomial moment generating function $Q_n(\theta)$ is defined as

$$Q_n(\theta)=\sum_{k=0}^{n}\binom{n}{k}q_n(k)\theta^k\ . \tag{2.8}$$

Then we have the following relation:

LEMMA 2.3.

$$M_n(\theta)=(1+p\theta)^n Q_n\left(\frac{\theta}{1+p\theta}\right). \tag{2.9}$$

PROOF. Using (2.6) we have

$$\begin{aligned}M_n(\theta)&=\sum_{k=0}^{n}\binom{n}{k}p_n(k)\theta^k\\&=\sum_{k=0}^{n}\binom{n}{k}\theta^k\sum_{j=0}^{k}\binom{k}{j}p^{k-j}q_n(j)\\&=\sum_{j=0}^{n}\binom{n}{j}\theta^j q_n(j)\sum_{k=j}^{n}\binom{n-j}{k-j}(p\theta)^{k-j}\\&=\sum_{j=0}^{n}\binom{n}{j}\theta^j(1+p\theta)^{n-j}q_n(j)\\&=(1+p\theta)^n Q_n\left(\frac{\theta}{1+p\theta}\right).\end{aligned}$$

Note that $(1+p\theta)^n$ is the factorial moment generating function of the binomial distribution. $Q_n(\theta/(1+p\theta))$ can be regarded as a correction factor. As mentioned above $Q_n\equiv 1$ if and only if S_n is binomial.

In Section 4 and Section 5 we discuss convergence of the distribution of S_n. Since we do not assume any condition on independence, we shall treat convergence in distribution in terms of convergence of moments. For our purposes we rely on the following fact.

LEMMA 2.4. *Let $\{F_n\}$ be a sequence of distribution functions such that moments of all orders $\mu_{k,n}=\int x^k dF_n$ are finite. Suppose that $\mu_{k,n}$ converges to μ_k (finite) for each k as $n\to\infty$. Then μ_k's are the moments of a distribution function F. If, moreover, μ_k's determine F uniquely, then F_n converges to F.*

Conversely if $F_n\to F$ and $\mu_{k,n}$ are bounded in n for each k, then $\mu_{k,n}$ converges to μ_k for each k.

For a proof see Loeve [9], Section 11.4. Note that this lemma can be also applied with other types of moments, e.g., factorial moments or cumulants.

3. Approximation by binomial distribution

Since S_n takes the values $0,\cdots,n$, it is natural to approximate its distribution by binomial distribution. Let

$$p_{BN}(x;n,p)=\binom{n}{x}p^x(1-p)^{n-x} \tag{3.1}$$

be the probability function of binomial distribution with parameters n

and p. Let

$$L_j^n(x;p)=\frac{d^j}{dp^j}p_{BN}(x;n,p)/p_{BN}(x;n,p)\ ,\quad j=0,\cdots,n\ . \tag{3.2}$$

Then L_j^n is a j-th degree polynomial in x and $\{L_j^n,\ j=0,\cdots,n\}$ forms a set of orthogonal polynomials with respect to $p_{BN}(x;n,p)$. These are often called Krawtchouk polynomials (see Appendix A of Ord [10], or Section 11.12 of Johnson and Kotz [6]). Approximations using these polynomials are discussed in Takeuchi [14] in detail. We note that L_j^n can be expressed also as

$$L_j^n=(-\Delta)^k n^{(k)}p_{BN}(x;n-k,p)/p_{BN}(x;n,p)\ , \tag{3.3}$$

where Δ is the difference operator: $\Delta f(x)=f(x)-f(x-1)$. See Chapter 3 of Takeuchi [14].

Now from Lemma 2.1 applied to binomial distribution we see that the k-th factorial moment is $\mu_{(k)}=n^{(k)}p^k$. Differentiating this relation j times with respect to p we obtain:

LEMMA 3.1.

$$\sum_{x=0}^{n} x^{(k)}L_j^n(x;p)p_{BN}(x;n,p)=\begin{cases}0\ , & \textit{if } k<j\ ;\\ n^{(k)}k^{(j)}p^{k-j}\ , & \textit{if } k\geqq j\ .\end{cases} \tag{3.4}$$

Using this lemma we can prove the following theorem:

THEOREM 3.1.

$$\Pr(S_n=x)=p_{BN}(x;n,p)\left\{1+\sum_{j=2}^{n}\frac{q_n(j)}{j!}L_j^n(x;p)\right\}\ . \tag{3.5}$$

$$\Pr(S_n\leqq x)=\sum_{y=0}^{x}p_{BN}(y;n,p)-np_{BN}(x;n-1,p)\sum_{j=2}^{n}\frac{q_n(j)}{j!}L_{j-1}^{n-1}(x;p)\ . \tag{3.6}$$

PROOF. Because the factorial moments $\mu_{(0)},\cdots,\mu_{(n)}$ determine the distribution uniquely, it suffices to check that the right hand side of (3.5) has the same factorial moments as the left hand side. Now by Lemma 3.1 and Lemma 2.1

$$\begin{aligned}&\sum_{x=0}^{n}x^{(k)}p_{BN}(x;n,p)\left\{1+\sum_{j=2}^{n}\frac{q_n(j)}{j!}L_j^n(x;p)\right\}\\&\quad=n^{(k)}\sum_{j=0}^{k}\binom{k}{j}p^{k-j}q_n(j)=n^{(k)}p_n(k)=E(S_n^{(k)})\ .\end{aligned}$$

This proves the first equality. The second equality follows from (3.3).

Remark 3.1. Using inclusion-exclusion principle $\Pr(S_n=x)$ can also be expressed in terms of $p_n(k)$'s in a simple manner. This is called Jordan identity or Bonferroni identity. See Chapter 4 of Feller [3], Galambos [5], or Takacs [12] for example.

This theorem can be also proved using the following explicit expression of L_k^n

$$p^k(1-p)^k L_k^n = \sum_{j=0}^{k} \binom{k}{j} (-1)^j \frac{n^{(k)}}{n^{(k-j)}} p^j x^{(k-j)} . \tag{3.7}$$

See formula (121), Section 11.12 of Johnson and Kotz [6]. Then by Parseval's identity

$$q_n(k) = \frac{p^k(1-p)^k}{n^{(k)}} E(L_k^n) . \tag{3.8}$$

The following useful recurrence formula is used later in the proof of Proposition 3.1. Define $\tilde{L}_j^n = p^j(1-p)^j L_j^n$. Then

$$\tilde{L}_{j+1}^n = (x-np-j(1-2p))\tilde{L}_j^n + \{j(j-1)p(1-p) - jnp(1-p)\}\tilde{L}_{j-1}^n . \tag{3.9}$$

Initial conditions are $\tilde{L}_0^n = 1$, $\tilde{L}_1^n = x - np$. See A2 of Appendix.

As an application of Theorem 3.1 consider Hypergeometric distribution with parameters N, M, and n. Denote the probability function of Hypergeometric distribution as

$$p_{HG}(x; N, M, n) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} .$$

Then

PROPOSITION 3.1. *Let n be fixed. Let $p = M/N$ and assume that p is bounded away from 0 and 1. Then as $N \to \infty$,*

$$p_{HG}(x; N, M, n) = p_{BN}(x; n, p)\left\{1 + \sum_{j=2}^{k} \frac{q_{n,N}(j)}{j!} L_j^n(x; p)\right\} + O(N^{-[(k/2)+1]}) , \tag{3.10}$$

where [ ] denotes the integer part and $q_{n,N}(j)$ is the j-th central binomial moment of $p_{HG}(x; N, M, n)$.

To prove this theorem we investigate $q_{n,N}(j)$. An explicit expression can be given as follows.

LEMMA 3.2.

$$q_{n,N}(k) = \frac{p^k(1-p)^k}{N^{(k)}} L_k^N(M; p) , \tag{3.11}$$

where $p=M/N$. *Hence*

$$\tilde{q}_{n,N}(k+1)=-k(1-2p)\tilde{q}_{n,N}(k)+\{k(k-1)p(1-p)-kNp(1-p)\}\tilde{q}_{n,N}(k-1)\ , \tag{3.12}$$

where $\tilde{q}_{n,N}(h)=N^{(h)}q_{n,N}(h)$.

PROOF. For Hypergeometric distribution, k-th binomial moment is given as $p_{n,N}(k)=M^{(k)}/N^{(k)}$. Hence

$$q_{n,N}(k)=\sum_{j=0}^{k}\binom{k}{j}(-1)^{k-j}\frac{M^{(j)}}{N^{(j)}}\left(\frac{M}{N}\right)^{k-j}\ .$$

But this is equal to (3.7) divided by $N^{(k)}$ if x is replaced by M and p is replaced by M/N. This proves (3.11). Now (3.12) follows from (3.9) with the same substitutions.

PROOF OF PROPOSITION 3.1. Since the summation is finite, it suffices to show that $q_{n,N}=O(N^{-[(k+1)/2]})$. This follows from (3.12) since by recursion $N^{(k)}q_{n,N}(k)=\tilde{q}_{n,N}(k)=O(N^{[k/2]})$.

For example up to the order N^{-2}

$$\begin{aligned} q_{n,N}(2)&=-\frac{p(1-p)}{N-1}\ ,\\ q_{n,N}(3)&=\frac{2p(1-p)(1-2p)}{(N-1)(N-2)}\ ,\\ q_{n,N}(4)&=\frac{3Np^2(1-p)^2-6p(1-p)(1-3p+3p^2)}{(N-1)(N-2)(N-3)}\ . \end{aligned} \tag{3.13}$$

Therefore using these quantities we obtain an approximation of Hypergeometric distribution up to the order N^{-2}.

4. Convergence to Poisson distribution

In this section we discuss the case, where $n\to\infty$ and the distribution of S_n approaches Poisson. We also discuss (i) asymptotic expansion around Poisson, (ii) convergence to a distribution which can be expanded around Poisson distribution, (iii) nonregular case, where normalized moments are not bounded as $n\to\infty$.

Applying Lemma 2.4 with factorial moments, convergence to Poisson distribution can be easily discussed. The k-th factorial moment of Poisson distribution with parameter λ is given as $\mu_{(k)}=\lambda^k$. Therefore noting $n^{(k)}/n^k\to 1$ as $n\to\infty$ we obtain from Lemma 2.4:

THEOREM 4.1. *The distribution of* S_n *converges to Poisson with pa-*

rameter λ *if* $n^k p_n(k) \to \lambda^k$ *(as* $n \to \infty$*) for each* k. *Converse is true if* $n^k p_n(k)$ *is bounded in* n *for each* k.

The first part of this theorem was already proved by Kendall [7] (see his Theorem II). In terms of $q_n(k)$ we have

COROLLARY 4.1. *The distribution of* S_n *converges to Poisson with parameter* λ *if* $np_n(1) \to \lambda$ *and* $n^k q_n(k) \to 0$ *(as* $n \to \infty$*) for each* $k \geqq 2$. *Converse is true if* $np_n(1)$, *and* $n^k q_n(k)$ *are bounded in* n *for each* k.

If $q_n(k)$ is interpreted as indicating deviation from independence, this corollary shows that if S_n approaches Poisson and $n^k q_n(k)$ are bounded in n then the 0–1 variables X_i's must approach the independence in a certain manner.

A more detailed treatment of this convergence can be given by an asymptotic Charlier Type B expansion around the Poisson distribution (see Chap. 1 of Takeuchi [14] or Kendall and Stuart [8]). Let $p(x;\lambda) = (\lambda^x/x!)e^{-\lambda}$ be the probability function of Poisson distribution with parameter λ. Then the j-th degree Charlier polynomial $L_j(x;\lambda)$ is defined as

$$L_j(x;\lambda) = \frac{d^j}{d\lambda^j} p(x;\lambda)/p(x;\lambda) . \tag{4.1}$$

$\{L_j(x;\lambda),\ j=0,1,\cdots\}$ forms a complete set of orthogonal polynomials with respect to $p(x;\lambda)$. See Section 11.12 of Johnson and Kotz [6]. Now we consider the factorial moment generating function (2.9). Let $\lambda = np_n(1)$. Taking the logarithm of (2.9) we have

$$\log G_n(1+\theta) = \log M_n(\theta) = n \log\left(1+\frac{\lambda}{n}\theta\right) + \log Q_n\left(\frac{\theta}{1+\lambda\theta/n}\right) . \tag{4.2}$$

Now assume that up to the order n^{-2}, $\log Q_n(\tau)$ can be expressed as

$$\log Q_n(\tau) = \left(\frac{1}{n}b_2 + \frac{1}{n^2}\tilde{b}_2\right)\tau^2 + \frac{1}{n^2}b_3\tau^3 + O(n^{-3}) , \tag{4.3}$$

where as a regularity condition we assume that the remainder term is of order n^{-3} uniformly for $|\tau| \leqq 1+\varepsilon$. Substituting (4.3) into (4.2) we obtain

$$\log M_n(\theta) = \lambda\theta + \theta^2\left(-\frac{\lambda^2}{2n} + \frac{b_2}{n} + \frac{\tilde{b}_2}{n^2}\right) + \theta^3\left(\frac{\lambda^3}{3n^2} - \frac{2\lambda b_2}{n^2} + \frac{b_3}{n^2}\right) + O(n^{-3}) . \tag{4.4}$$

Therefore up to the order n^{-2}

$$M_n(\theta) = e^{\lambda\theta}[1 + c_2\theta^2 + c_3\theta^3 + c_4\theta^4] + O(n^{-3}) , \tag{4.5}$$

where

$$
(4.6)\qquad
\begin{aligned}
c_2 &= -\frac{\lambda^2}{2n}+\frac{b_2}{n}+\frac{\tilde{b}_2}{n^2}\,, \qquad c_3=\frac{\lambda^3}{3n^2}-\frac{2\lambda b_2}{n^2}+\frac{b_3}{n^2}\,,\\
c_4 &= \frac{\lambda^4}{8n^2}+\frac{b_2^2}{2n^2}-\frac{b_2\lambda^2}{2n^2}\,.
\end{aligned}
$$

From this we obtain

THEOREM 4.2. *Assume that* $\log Q_n(\tau)$ *can be expanded as in* (4.3). *Then*

$$
(4.7)\qquad \Pr(S_n=x)=\frac{\lambda^x}{x!}e^{-\lambda}[1+c_2L_2(x;\lambda)+c_3L_3(x;\lambda)+c_4L_4(x;\lambda)]+O(n^{-3})\,,
$$

where c_2, c_3, c_4 *are given as* (4.6).

Since derivation of (4.7) from (4.6) is given in Chapter 1 of Takeuchi [14] we omit the proof. Higher order expansions can be derived in a similar manner.

Charlier Type B series expansion can be also applied to the first order convergence to a limiting distribution which can be expanded around Poisson distribution in L^2. Consider the formal limit of (3.5) as $n\to\infty$. As $n\to\infty$ and $np\to\lambda$, $n^{-j}L_j^n(x;p)\to L_j(x;\lambda)$ (see A1 of Appendix for a proof). Therefore if $n^jq_n(j)\to q^*(j)$ as $n\to\infty$, as a formal limit we obtain $\lim_{n\to\infty}\Pr(S_n=x)=p(x;\lambda)\left(1+\sum_{j=2}^{\infty}(1/j!)q^*(j)L_j(x;\lambda)\right)$. In fact this can be made rigorous as follows.

THEOREM 4.3. *Suppose that* $np_n(1)\to\lambda$ *and* $n^jq_n(j)\to q^*(j)$ *as* $n\to\infty$ *such that* $\sum_{j=2}^{\infty}q^*(j)^2/j!\lambda^j<\infty$. *Then*

$$
(4.8)\qquad \lim_{n\to\infty}\Pr(S_n=x)=p(x;\lambda)\left\{1+\sum_{j=2}^{\infty}\frac{q^*(j)}{j!}L_j(x;\lambda)\right\}\,.
$$

Remark 4.1. The regularity condition on the square summability of $q^*(j)$ guarantees that the series within the parentheses on the right hand side converges in L^2 with respect to $p(x;\lambda)$. However since the sample space is discrete convergence in L^2 implies pointwise convergence. Actually this regularity condition can be weakened substantially. See Takemura [13].

PROOF. The right hand side of (4.8) defines a signed measure P^* over nonnegative integers. From the assumption the factorial moments of S_n converge to those of P^*. Now from Lemma 2.4 it follows that

P^* is a distribution and the distribution of S_n converges to P^* if P^* is uniquely determined by its moments. This is the case because as shown in Lemma 4.1 below P^* has moment generating function. This proves the theorem.

The following lemma completes the above proof.

LEMMA 4.1. *Let P^* be a distribution over nonnegative integers such that $\sum_{x=0}^{\infty} (P^*(x)/p(x;\lambda))^2 p(x;\lambda)<\infty$. Then P^* has moment generating function defined everywhere.*

PROOF. Let

$$M=\sum_{x=0}^{\infty} P^*(x)^2/p(x;\lambda)=e^{\lambda}\sum P^*(x)^2 x!/\lambda^x<\infty .$$

Then by Schwarz for any positive a,

$$\sum a^x P^*(x)\leqq\left[\sum a^{2x}\frac{\lambda^x}{x!}\sum P^*(x)^2 x!/\lambda^x\right]^{1/2}=e^{(a^2-1)\lambda/2}M^{1/2}<\infty .$$

Finally we consider nonregular case, where the distribution converges but the moments do not converge. Applying the continuity theorem for the probability generating function we obtain the following theorem.

THEOREM 4.4. *Suppose that $np_n(1)\to\lambda$. Then S_n converges to the Poisson distribution with parameter λ, if and only if $Q_n(\tau)\to 1$ for every τ in the interval $-1\leqq\tau<0$.*

PROOF. Consider the factorial moment generating function of S_n given by (2.9). As $n\to\infty$ and $np\to\lambda$ the first term on the right hand side converges to the factorial moment generating function of the Poisson distribution. Therefore the distribution of S_n converges to the Poisson, if and only if Q_n converges to 1. Note that from the relation between probability generating function and the factorial moment generating function it suffices to consider θ in the range $-1\leqq\theta<0$. Because for every ε the convergence is uniform for $-1\leqq\theta\leqq-\varepsilon<0$, the difference between θ and $\tau=\theta/(1+p\theta)$ does not matter asymptotically. This proves the theorem.

To illustrate this theorem, consider the following mixture of binomial distributions. Let the success probability P be random with $\Pr(\mathrm{P}=1/n)=1-1/n$ and $\Pr(\mathrm{P}=1/n+1/2)=1/n$. In this case clearly the limiting distribution of S_n is Poisson with parameter 1. Now from Lemma 2.2, $q_n(k)=E(\mathrm{P}-1/n)^k=1/2^k n$. Hence $n^k q_n(k)$ diverges. On the other hand $Q_n(\tau)=\sum_{k=0}^{n}\binom{n}{k}\tau^k/2^k n+1-1/n=(1+\tau/2)^n/n+1-1/n$ which converges to 1 for $-1\leqq\tau<0$. Note that the coefficients in the expansion of $(1+\tau/2)^n$ diverge.

5. Convergence to normal distribution

In this section we discuss central limit theorem. We obtain conditions on moments such that $Z_n=(S_n-np)/\sqrt{n}$ converges to normal in distribution. The development of this section closely follows that of the previous section. We discuss (i) ordinary convergence to a normal distribution, (ii) asymptotic Edgeworth expansion, (iii) convergence to a distribution which can be expanded around the normal distribution, and finally, (iv) nonregular case.

Corresponding to Theorem 4.1 we have the following result.

THEOREM 5.1. *Assume that* $\lim_{n\to\infty} p_n(1)=p$ *exists and let* $Z_n=(S_n-np)/\sqrt{n}$. *If for some real* $c\geqq -p(1-p)$

$$(5.1)\qquad \begin{aligned}&\lim_{n\to\infty} n^{k/2}q_n(k)=0\,, && k:\ odd\,,\\ &\lim_{n\to\infty} n^{k/2}q_n(k)=c^{k/2}1\cdot 3\cdots(k-1)\,, && k:\ even\,,\end{aligned}$$

then Z_n *converges to* $N(0,\sigma^2)$ *in distribution, where* $\sigma^2=p(1-p)+c$. *Converse of this holds if* $n^{k/2}q_n(k)$ *is bounded in* n *for each* k.

PROOF. We apply Lemma 2.4 with cumulants. Letting $\theta=e^{it}-1$ in (2.9) and taking logarithm, the cumulant generating function of Z_n is written as

$$(5.2)\qquad \begin{aligned}\log E(\exp(itZ_n))&=-it\sqrt{n}\,p+n\log(1+p(\exp(it/\sqrt{n})-1))\\ &\quad+\log Q_n\left(\frac{\exp(it/\sqrt{n})-1}{1+p(\exp(it/\sqrt{n})-1)}\right).\end{aligned}$$

The first two terms constitute the binomial part and k-th order cumulant coming from these terms converges to 0 if $k\geqq 3$ and to $p(1-p)$ if $k=2$. Therefore if the k-th order cumulant from the third term converges to 0 for $k\geqq 3$ and converges to c for $k=2$ the conclusion of the theorem holds. Let $a_{n,k}$ denote the k-th cumulant from the third term, i.e., let

$$\log Q_n\left(\frac{\exp(it/\sqrt{n})-1}{1+p(\exp(it/\sqrt{n})-1)}\right)=a_{n,2}\frac{(it)^2}{2}+a_{n,3}\frac{(it)^3}{3!}+\cdots.$$

Note that since Q_n is a finite series for a finite n, $\log Q_n$ can be always expanded around the origin. Now let

(5.3) $$\log Q_n\left(\frac{\tau}{\sqrt{n}}\right)=\tilde{a}_{n,2}\frac{\tau^2}{2}+\tilde{a}_{n,3}\frac{\tau^3}{3!}+\cdots .$$

Since $\tau=it+O(n^{-1/2})$ we have $a_{n,k}=\tilde{a}_{n,k}+O(n^{-1/2})$ for each k. Therefore if $\tilde{a}_{n,k}$ converges to 0 for $k\geqq 3$ and to c for $k=2$ the conclusion of the theorem holds. But exponentiating (5.3) we see that this is equivalent to

$$n^{k/2}q_n(k)\to 0\ , \qquad k:\text{ odd}\ ,$$

$$n^{k/2}q_n(k)\to c^{k/2}1\cdot 3\cdots(k-1)\ , \qquad k:\text{ even}\ .$$

This proves the first part of the theorem. The converse part is a direct consequence of the converse part of Lemma 2.4.

The condition of the theorem requires that the normalized central binomial moments approach the central moments of the normal distribution with variance c. An interesting fact here is that c can be negative. When c is negative the asymptotic variance σ^2 is smaller than the asymptotic variance $p(1-p)$ of the binomial distribution and therefore the asymptotic distribution is more concentrated toward origin than the binomial case. This contrasts with the case where S_n is a mixture of binomial distributions and hence c is nonnegative. Here are examples of this phenemenon.

Example 5.1. Consider the Hypergeometric distribution of Proposition 3.1. This time let n approach ∞ with the same order as N approaches ∞. Letting $n=\alpha N$, $M=pN$, it is easy to see that $\mathrm{Var}(Z_n)\to p(1-p)(1-\alpha)=p(1-p)-\alpha p(1-p)$. Therefore $c=-\alpha p(1-p)$ in this case.

Example 5.2. Let the 0-1 variables X_i be independent with $\Pr(X_i=1)=p_i$. Then $\mathrm{Var}(Z_n)=\bar{p}(1-\bar{p})-\sum(p_i-\bar{p})^2/n$, where $\bar{p}=\sum p_i/n$. Therefore $c=-\lim\sum(p_i-\bar{p})^2/n$ if this limit exists.

For more detailed discussion of these examples see Chapter 2 of Takeuchi [14].

Closer approximation can be given by asymptotic Edgeworth expansion. Edgeworth expansion for discrete random variable is fully discussed in Chapter 2 of Takeuchi [14]. Here we derive the expansion up to the order n^{-1}. Suppose that

(5.4) $$\log Q_n\left(\frac{i\tau}{\sqrt{n}}\right)=\frac{(i\tau)^2}{2}\left(c_2+\frac{1}{n}\tilde{c}_2\right)+c_3\frac{(i\tau)^3}{6\sqrt{n}}+c_4\frac{(i\tau)^4}{24n}+R_n(\tau)\ .$$

A possible regularity condition on the remainder term $R_n(\tau)$ is given in A3 of Appendix. Now

$$i\tau=\sqrt{n}\left(\frac{\exp(it/\sqrt{n})-1}{1+p(\exp(it/\sqrt{n})-1)}\right) \tag{5.5}$$
$$=it+\frac{1-2p}{2\sqrt{n}}(it)^2+\frac{1-6p+6p^2}{6n}(it)^3+o(n^{-1}) .$$

From (5.4) and (5.5) we obtain

$$\log Q_n\left(\frac{i\tau}{\sqrt{n}}\right)=\frac{(it)^2}{2}c_2+\frac{(it)^3}{6\sqrt{n}}\{c_3+(3-6p)c_2\}+\frac{(it)^2}{2n}\tilde{c}_2 \tag{5.6}$$
$$+\frac{(it)^4}{24n}\{c_4+(6-12p)c_3+(7-36p+36p^2)c_2\}+o(n^{-1}) .$$

Similar expansion for the binomial part is given in Chapter 2 of Takeuchi [14]. Combining these two parts we can obtain the following result.

THEOREM 5.2. *Let x be an integer and let $z=(x-np+1/2)/\sqrt{n}\,\sigma$ where $p=p_n(1)$ is fixed and $\sigma^2=p(1-p)+c_2$. Suppose that $\log Q_2(i\tau/\sqrt{n})$ can be expanded as* (5.4). *Then under suitable regularity conditions (such as those given in* A3 *of Appendix)*

$$\Pr(S_n\leqq x)=\Phi(z)-\phi(z)\left\{\frac{\alpha_3}{6\sqrt{n}\,\sigma^3}h_2(z)+\frac{\alpha_4}{24n\sigma^4}h_3(z)\right. \tag{5.7}$$
$$\left.+\frac{\alpha_3^2}{72n\sigma^6}h_5(z)-\frac{1-12\tilde{c}_2}{24n\sigma^2}h_1(z)\right\}+o(n^{-1}) ,$$

where $\phi=\Phi'$ is the standard normal density, h_k is the k-th Hermite polynomial, and

$$\alpha_3=p(1-p)(1-2p)+(3-6p)c_2+c_3 ,$$

$$\alpha_4=p(1-p)(1-6p+6p^2)+c_4+(6-12p)c_3+(7-36p+36p^2)c_2 .$$

We omit detail of the proof.

As in the Poisson case S_n can converge to a distribution which can be expanded around normal distribution in L^2. As $n\to\infty$

$$n^{-k/2}L_k^n(np+x\sqrt{np(1-p)};p)\to[p(1-p)]^{-k/2}h_k(x) . \tag{5.8}$$

See A4 of Appendix. Now suppose that

$$n^{k/2}q_n(k)/[p(1-p)]^{k/2}\to q^*(k) .$$

Then as in the Poisson case we can take the formal limit in (3.6). More precisely we can prove

THEOREM 5.3. *Let $p=p_n(1)$ be fixed. Suppose that $q^*(k)=\lim_{n\to\infty} n^{k/2}\cdot q_n(k)/[p(1-p)]^{k/2}$ exists for each k with $\sum q^*(k)^2/k!<\infty$. Then*

$$(5.9)\qquad \lim_{n\to\infty} \Pr\,(S_n \leqq np + x\sqrt{np(1-p)}) = \Phi(x) - \phi(x)\left\{\sum_{k=2}^{\infty} \frac{q^*(k)}{k!} h_{k-1}(x)\right\} .$$

PROOF. As in the Poisson case it suffices to show that the distribution defined by the right hand side of (5.8) is uniquely determined by its moments. This follows from Lemma 5.1 below because the distribution on the right hand side of (5.9) has the density

$$f(x) = \phi(x)\left\{1 + \sum_{k=2}^{\infty} q^*(k) h_k(x)/k!\right\}$$

which satisfies the condition of Lemma 5.1 and hence has the moment generating function.

LEMMA 5.1. *Let F be a distribution function over the real line with a density function f such that $\int (f/\phi)^2 \phi dx < \infty$. Then F has moment generating function defined everywhere.*

PROOF. By Schwarz

$$\int f(x) e^{\theta x} dx \leqq \left[\int e^{2\theta x - x^2/2} dx \int e^{x^2/2} f(x)^2 dx\right]^{1/2} < \infty .$$

Remark 5.1. Theorem 5.3 is stated in terms of the distribution function. If we take the formal limit of (3.5) instead of (3.6) we may obtain the local central limit theorem. However it seems somewhat difficult to state the condition for the convergence in a precise manner. For example let S_n be a binomial random variable and consider the conditional distribution of S_n given that S_n be even: $\Pr\,(S_n = 2k \mid S_n$ is even). Clearly central limit theorem holds for this conditional distribution, but local central limit theorem does not.

Finally we briefly look at the nonregular case, where the distribution of Z_n approaches the normal but the moments of Z_n diverge. Consider the cumulant generating function (5.2) of Z_n again. We only need to consider real t. The sum of the first two terms of (5.2) converges to the cumulant generating function of the normal distribution. Therefore Z_n converges to $N(0, p(1-p))$ in distribution if and only if the third term converge to 0 for every real t. For the particular case of $p=1/2$, it is easy to show that if t is real then τ of (5.5) is also real and we have

$$(5.10)\qquad t = 2\sqrt{n}\,\tan^{-1}(\tau/2\sqrt{n})\,.$$

Therefore Z_n converges to $N(0, 1/4)$ in distribution if and only if $Q_n(i\tau/\sqrt{n}) \to 0$ for every real τ.

The following example illustrates this. Let the distribution of S_n

be the following mixture: with probability $1-1/n$ S_n is distributed according to the binomial distribution with parameters n and $p=1/2$ and with probability $1/2n$ S_n takes 0 or n. If the parameter P of the binomial distribution is considered as random, this corresponds to

$$\Pr(P=0)=\Pr(P=1)=\frac{1}{2n}, \qquad \Pr\left(P=\frac{1}{2}\right)=1-\frac{1}{n}.$$

Therefore

$$(5.11) \qquad q_n(k)=E(P-1/2)^k=\begin{cases} 1, & k=0 \\ 0, & k \text{ is odd} \\ \dfrac{1}{2^k n}, & k \text{ is even and } k\geqq 2. \end{cases}$$

Hence $n^{k/2}q_n(k)\to\infty$. However obviously S_n converges to $N(0, p(1-p))$ in distribution. Now from (5.11) we obtain

$$(5.12) \qquad Q_n\left(\frac{i\tau}{\sqrt{n}}\right)=1+\sum_{k=1}^{[n/2]}\binom{n}{2k}(-1)^k\frac{\tau^{2k}}{2^{2k}}\frac{1}{n^{k+1}}.$$

It is not easy to see from this whether this Q_n converges to 1. Another expression of Q_n can be given as follows. Because Z_n is a mixture of binomial and two-point distributions

$$E(\exp(\theta Z_n))=\left(1-\frac{1}{n}\right)\exp(-\sqrt{n}\,p\theta)(1+p(\exp(\theta/\sqrt{n})-1))^n$$
$$+\frac{1}{2n}(\exp(\theta\sqrt{n}/2)+\exp(-\theta\sqrt{n}/2)).$$

Using (2.9) and (5.10) it is now straightforward to show that $Q_n(i\tau/\sqrt{n})$ can be written as

$$Q_n\left(\frac{i\tau}{\sqrt{n}}\right)=1-\frac{1}{n}+\frac{1}{n}\left(1-\frac{\tau^2}{4n}\right)^{n/2}\cos\left(2\sqrt{n}\arctan\frac{\tau}{2\sqrt{n}}\right).$$

Therefore $Q_n(i\tau/\sqrt{n})$ converges to 1 for every real τ.

UNIVERSITY OF TOKYO

REFERENCES

[1] Chung, K. L. (1974). *A Course in Probability Theory*, 2nd ed., Academic Press, New York.
[2] Diaconis, P. and Freedman, D. (1980). Finite exchangeable sequences, *Ann. Prob.*, 8, 745-764.
[3] Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, 1, 3rd

ed., Wiley, New York.

[4] Galambos, J. (1978). *The Asymptotic Theory of Extreme Order Statistics*, Wiley, New York.

[5] Galambos, J. (1982). The role of exchangeability on the theory of order statistics, in *Exchangeability in Probability and Statistics*, (eds. Koch, G. and Spizzichino), North-Holland, Amsterdam.

[6] Johnson, N. L. and Kotz, S. (1969). *Discrete Distributions*, Wiley, New York.

[7] Kendall, D. G. (1967). On finite and infinite sequences of exchangeable events, *Studia Sci. Math. Hungar.*, **2**, 319-327.

[8] Kendall, M. G. and Stuart, A. (1969). The advanced theory of statistics, **1**, *Distribution Theory*, 3rd ed., Griffin, London.

[9] Loeve, M. (1977). *Probability Theory I*, Springer, New York.

[10] Ord, J. K. (1972). *Families of Frequency Distributions*, Griffin, London.

[11] Shimizu, R. (1976). *Central Limit Theorems* (in Japanese), Kyoiku Shuppan, Tokyo.

[12] Takacs, L. (1967). On the method of inclusion and exclusion, *J. Amer. Statist. Ass.*, **62**, 102-113.

[13] Takemura, A. (1985). On convergence of Gram-Charlier expansion based on discrete distributions, in preparation.

[14] Takeuchi, K. (1975). *Approximations of Probability Distributions* (in Japanese), Kyoiku Shuppan, Tokyo.

[15] Takeuchi, K. and Takemura, A. (1987). On sum of 0-1 random variables, II, Multivariate case, to appear in *Ann. Inst. Statist. Math.*

[16] Watanabe, M. (1919). On a determinate system of non-independent trials, *Tohoku Math. J.*, **15**, 1-134.

Appendix

A1.

LEMMA A1. *As $n\to\infty$ and $np\to\lambda$*

$$n^{-j}L_j^n(x;p)\to L_j(x;\lambda)\ .$$

PROOF. From the definition of the Krawtchouk polynomials

$$p(x;n,p+t)=p(x;n,p)\left(1+\sum_{j=1}^{n}t^jL_j^n(x;p)/j!\right)\ .$$

Therefore the generating function of the Krawtchouk polynomials is given as

$$\text{(A1)}\qquad G_{BN}(x;t)=\frac{p(x;n,p+t)}{p(x;n,p)}=\left(1+\frac{t}{p}\right)^x\left(1-\frac{t}{1-p}\right)^{n-x}\ .$$

Hence the generating function of $n^{-j}L_j^n(x;n,p)$ is given as

$$G^*(x,t)=\left(1+\frac{t}{np}\right)^x\left(1-\frac{t}{n(1-p)}\right)^{n-x}\ .$$

Similarly for Charlier polynomials the generating function is given by

$$G_P(x, t)=\frac{p(x\,;\,\lambda+t)}{p(x\,;\,\lambda)}=e^{-t}\left(1+\frac{t}{\lambda}\right)^x .$$

Now as $n\to\infty$ and $np\to\lambda$, G^* converges to G_P. Since G^* and G_P are analytic, this proves the lemma.

A2.

LEMMA A2. *Let* $\tilde{L}_j^n(x\,;p)=p^j(1-p)^jL_j^n(x\,;p)$. *Then*

(A2) $\tilde{L}_{j+1}^n=(x-np-j(1-2p))\tilde{L}_j^n+\{j(j-1)p(1-p)-jnp(1-p)\}\tilde{L}_{j-1}^n$.

PROOF. Let $t=p(1-p)\omega$ in (A1) and differentiate G with respect to ω. Then

$$(1+(1-2p)\omega-p(1-p)\omega^2)\frac{\partial G}{\partial\omega}=(x-np-np(1-p)\omega)G .$$

Equating the coefficient of ω^k we obtain (A2).

A3.

A possible regularity condition on the remainder term in (5.4) can be given as follows:
(C1): There exist constants $0<c<\pi$ and M such that

$$|R_n(\tau)|\leqq M|\tau|^5/n^{3/2}$$

for all $|\tau|\leqq\sqrt{n}\,c$.

The following additional condition covers the region $\sqrt{n}\,c<|\tau|\leqq\sqrt{n}\,\pi$.
(C2): There exists a nonnegative continuous function $b(t)$ such that $b(t)<1$ and

$$|E(e^{itZ_n})|\leqq b(t/\sqrt{n}\,)^n$$

for $0<|t|\leqq\sqrt{n}\,\pi$.

For discussion on these regularity conditions, see Chapter 2 of Takeuchi [14] and Chapter 6 of Shimizu [11].

A4.

LEMMA A4. *As* $n\to\infty$

(A3) $n^{-k/2}L_k^n(np+x\sqrt{np(1-p)}\,;p)\to[p(1-p)]^{-k/2}h_k(x)$.

PROOF. For Hermite polynomials the generating function is given by

$$G_N(x, t)=e^{tx-t^2/2} .$$

The rest of the proof is entirely analogous to Lemma A1 and we omit the detail.