72-11,524

CHEN, Louis Hsiao-Yun, 1940-
POISSON APPROXIMATION FOR SUMS OF
DEPENDENT BERNOULLI RANDOM VARIABLES.

Stanford University, Ph.D., 1971
Statistics

University Microfilms, A XEROX Company, Ann Arbor, Michigan

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED

POISSON APPROXIMATION FOR SUMS OF DEPENDENT BERNOULLI RANDOM VARIABLES

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF STATISTICS

AND THE COMMITTEE ON GRADUATE STUDIES

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

By

Louis Hsiao-Yun Chen

August 1971

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Charles M. Stein

(Principal Adviser)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Herman Chernoff

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Approved for the University Committee on Graduate Studies:

Dean of Graduate Studies

PLEASE NOTE:

Some Pages have indistinct print. Filmed as received.

UNIVERSITY MICROFILMS

ACKNOWLEDGMENT

I am deeply indebted to Professor Charles Stein for his suggestions, guidance, generosity with his time and constant encouragement throughout the preparation of this dissertation. I am particularly thankful to him for the knowledge and insight which he has imparted to me and I will always remember my experience of working with him.

I would like to thank Professor Herman Chernoff and Professor Bradley Efron who have kindly agreed to be on my Reading Committee.

Finally, I would also like to thank Rosemarie Stampfel for her excellent job in typing the manuscript.

TABLE OF CONTENTS

CHAPTER	TITLE	PAGE
0	INTRODUCTION AND SUMMARY	1
I	PRELIMINARIES	6
	§1. Stein's method	6
	§2. Conditional expectations	10
II	POISSON APPROXIMATION	13
	§1. The basic identity	13
	§2. A characterization of the Poisson distribution	15
	§3. Preliminary results	19
	§4. The sum of a weakly dependent sequence of Bernoulli random variables	24
	§5. Special cases of weak dependence	34
	§6. A randomly selected sum of Bernoulli random variables	48
	§7. Remarks	63
III	ITERATION OF POISSON APPROXIMATION	64
	§1. Preliminary results	64
	§2. Main theorems	69
IV	COMPOUND POISSON CONVERGENCE	79
	§1. Introduction	79
	§2. A characterization of the compound Poisson distribution	81
	§3. The solution of the equation $wf(w) - \lambda \int tf(w+t)\, d\mu(t) = v(w)$	85
	§4. The limit theorem	97
	REFERENCES	118

CHAPTER 0

INTRODUCTION AND SUMMARY

This dissertation is concerned with approximating the distributions of sums of dependent Bernoulli random variables by either the Poisson distribution or a compound Poisson distribution. Our main objective is to bound the errors in the approximation. In the compound Poisson case, our result is in the form of a limit theorem. However, the same proof can also be used to obtain a bound on the rate of convergence. The method we use is based on a simple perturbation argument. It is originally due to Stein (1970), who has successfully applied it to the normal approximation problem for sums of dependent random variables.

The Poisson approximation problem for sums of independent Bernoulli random variables has been studied by Prohorov (1953), Le Cam (1960), Hodges and Le Cam (1960), Kerstan (1964) and Vervaat (1969). Perhaps the most significant results are those of Le Cam who proved that if $X_1, X_2, \ldots, X_n$ are independent Bernoulli random variables with $P(X_i = 1) = 1 - P(X_i = 0) = p_i$, then for every real-valued function h defined on $\{0, 1, 2, \ldots\}$ such that $|h| \leq 1$,

$$|Eh(\sum_{i=1}^{n} X_i) - \mathcal{P}_\lambda h| \leq 2 \sum_{i=1}^{n} p_i^2$$

$$|Eh(\sum_{i=1}^{n} X_i) - \mathcal{P}_\lambda h| \leq 9 \max_{1 \leq i \leq n} p_i$$

$$|Eh(\sum_{i=1}^{n} X_i) - \mathcal{P}_\lambda h| \leq \frac{16}{\lambda} \sum_{i=1}^{n} p_i^2 \qquad (\max_{1 \leq i \leq n} p_i \leq \frac{1}{4})$$

where

$$\mathcal{P}_\lambda h = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} h(k)$$

and

$$\lambda = \sum_{i=1}^{n} p_i$$

Prohorov considered only the binomial distribution. Kerstan generalized Le Cam's results to sums of independent non-negative integer-valued random variables and also reduced the values of the absolute constants in the bounds.

In this work, emphasis is placed on obtaining the correct order of magnitude of the errors rather than minimizing the absolute constants in the bounds. The contents of this dissertation are divided into four chapters. Chapter I contains a brief discussion of Stein's method and a review of conditional expectations. In Chapter II, we prove two theorems in which bounds in fairly general forms are obtained for the error in the Poisson approximation to the distribution of the sum of an arbitrary sequence of dependent Bernoulli random variables. These two theorems are then used to obtain Le Cam's results as well as explicit bounds for three special cases of dependence. They are m-dependence, exponentially decreasing dependence and Markovian dependence. The results for the m-dependent case are as follows. Let $X_1, X_2, \ldots, X_n$ be m-dependent Bernoulli random variables

with $P(X_i = 1) = 1 - P(X_i = 0) = p_i$, then

$$|Eh(\sum_{i=1}^{n} X_i) - \mathcal{P}_\lambda h| \leq \frac{\lambda}{\lambda^{1/2} \vee 1} \left[A_1(m)\tilde{p} + A_2 \frac{|C_m|}{\lambda} \right]$$

$$|Eh(\sum_{i=1}^{n} X_i) - \mathcal{P}_\lambda h| \leq B_1(m)\tilde{p} + B_2(m) \frac{|C_m|}{\lambda}$$

where

$$A_1(m) = 32m + 8$$

$$A_2 = 16$$

$$B_1(m) = 168m^2 + 95m + 14$$

$$B_2(m) = 90m + 28$$

$$\tilde{p} = \max_{1 \leq i \leq n} p_i$$

$$C_m = \sum_{i<j} \operatorname{cov}(X_i, X_j)$$

and λ, h and $\mathcal{P}_\lambda h$ are defined as above.

Also contained in Chapter II is an approximation theorem for a randomly selected sum of Bernoulli random variables. The theorem can be stated as follows. Let X_{ij}, $i, j = 1, 2, \ldots, n$ be a square array of independent Bernoulli random variables with $P(X_{ij} = 1) = 1 - P(X_{ij} = 0) = p_{ij}$ and π be a random permutation of $(1, 2, \ldots, n)$, independent of the X_{ij}. Then

$$|Eh(\sum_{i=1}^{n} X_{i\pi(i)}) - \mathcal{P}_\lambda h| \leq \frac{24}{\lambda^{1/2} \vee 1} \left(\sum_{i=1}^{n} p_{i.}^2 + \sum_{j=1}^{n} p_{.j}^2 \right)$$

$$|Eh(\sum_{i=1}^{n} X_{i\pi(i)}) - \mathcal{P}_\lambda h| \leq \frac{84}{\lambda} \left(\sum_{i=1}^{n} p_{i.}^2 + \sum_{j=1}^{n} p_{.j}^2 \right)$$

where

$$p_{i.} = \frac{1}{n} \sum_{j=1}^{n} p_{ij}$$

$$p_{.j} = \frac{1}{n} \sum_{i=1}^{n} p_{ij}$$

$$\lambda = \sum_{i=1}^{n} p_{i.} = \sum_{j=1}^{n} p_{.j} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} p_{ij}$$

and $\pi(i)$ is the i^{th} component of π.

Chapter III is mainly computational, in which, by iterating the basic identity derived in Chapter II, we obtain asymptotic expansions up to the second order term. We also estimate the order of magnitude of the error of second order. Due to the complications involved, only the independent and the Markovian cases are considered.

Chapter IV is devoted to the study of the asymptotic behavior of the distribution of the row sum of an m-dependent triangular array of Bernoulli random variables and some related problems. Let X_{in}, $i = 1, 2, \ldots, n$, $n = 1, 2, \ldots$ be such an array with $P(X_{in} = 1) = 1 - P(X_{in} = 0) = p_{in}$ such that $\max_{1 \le i \le n} p_{in} \longrightarrow 0$ as $n \longrightarrow \infty$. In this chapter, we obtain a necessary and sufficient condition for the proper convergence of $\mathcal{L}(\sum_{i=1}^{n} X_{in})$ and also show that the limit distribution must be $\mathcal{L}(\sum_{i=1}^{m+1} iZ_i)$ where the Z_i are independent and Poisson distributed with parameters bearing a simple and explicit relation to the necessary and sufficient condition. Philipp (1969), showed that, under certain conditions, the limit distributions for

the row sums of certain mixing triangular arrays of random variables must be infinitely divisible and he also obtained a necessary and sufficient condition for convergence to any specififed distribution in the class of possible limit distributions. Although our problem is a special case of Philipp's, our proof is different and our necessary and sufficient condition is much simpler and more explicit than his. Moreover, our result also reveals a direct and explicit relation between $\mathcal{L}(\sum_{i=1}^{n} X_{in})$ and $\mathcal{L}(\sum_{i=1}^{m+1} iZ_i)$. The related problems contained in this chapter are the characterization theorems for the compound Poisson distribution and the study of the solution of a certain integral equation. These problems are interesting in their own rights as well as being preliminaries to the main limit theorem. Finally, it should be added that the material contained in Chapter IV provides the machinery not only for extending the limit theorem to mixing triangular arrays of Bernoulli random variables, but also for obtaining the rate of convergence.

CHAPTER I

PRELIMINARIES

§1. Stein's method

Let $(\Omega, \mathcal{B}, P)$ be a probability space. Suppose we wish to compute $\int \xi \, dP$ where ξ is an integrable random variable. We may think of P as a linear functional P^* on the linear space $\mathcal{Y}$ of all integrable random variables, defined by $P^*\xi = \int \xi \, dP$. Then to evaluate $P^*\xi$ is equivalent to finding a constant c such that $P^*(\xi - c) = 0$, i.e., finding in $\ker P^* = \{y : P^*y = 0\}$, an element differing from ξ by a constant c. It is possible to use conditional expectations to construct an operator $Q : \mathcal{X} \longrightarrow \mathcal{Y}$ such that $\operatorname{im} Q \subset \ker P^*$, where $\mathcal{X}$ is a subspace of $\mathcal{Y}$ and $\operatorname{im} Q = \{y : y = Qx$ for some $x \in \mathcal{X}\}$. Often one is able to see the possibility of $\xi - c \in \operatorname{im} Q$ without being able to evaluate c immediately. Under such circumstances, if we can approximate Q by another operator Q_0 which is easy to study we may be able to study Q by perturbation methods and then evaluate c approximately.

Thus we are led to the following situation.

(1)
$$\mathcal{X} \underset{Q_0}{\overset{Q}{\rightrightarrows}} \mathcal{Y} \underset{S_0}{\overset{S}{\rightrightarrows}} \mathcal{Y}$$

where $\mathcal{X}$ and $\mathcal{Y}$ are linear spaces, and Q, Q_0, S, S_0 are operators such that

(2) $$S^2 = S$$

(3) $$S_0^2 = S_0$$

(4) $$\text{im } Q = \ker S$$

(5) $$\text{im } Q_0 = \ker S_0$$

In the above context, S will be P*. Now, (2), (3), (4) and (5) imply

(6) $$\text{im } Q = \ker S = \text{im}(I_{\mathcal{Y}} - S)$$

(7) $$\text{im } Q_0 = \ker S_0 = \text{im}(I_{\mathcal{Y}} - S_0)$$

where $I_{\mathcal{Y}}$ is the identity mapping from $\mathcal{Y}$ into itself. Let Q_0^i be a generalized inverse of Q_0, i.e.,

(8) $$Q_0 Q_0^i Q_0 = Q_0$$

then by (7),

(9) $$Q_0 Q_0^i (I_{\mathcal{Y}} - S_0) = I_{\mathcal{Y}} - S_0$$

By (4), (5) and (9),

(10) $$0 = (SQ - S_0 Q_0)\, Q_0^i (I_{\mathcal{Y}} - S_0) = [(S-S_0)Q_0 + S(Q-Q_0)]\, Q_0^i (I_{\mathcal{Y}} - S_0)$$
$$= [S - S_0 + S(QQ_0^i - I_{\mathcal{Y}})]\, (I_{\mathcal{Y}} - S_0)$$

We shall be interested in the case where we also have

(11) $$SS_0 = S_0$$

Then (10) yields

(12) $$S = S_0 + S(I_{\mathcal{Y}} - QQ_0^i)(I_{\mathcal{Y}} - S_0)$$

It is clear that if Q is approximated by Q_0 in an appropriate sense, then the second term on the right hand side of (12) will be small in some sense. In this case, we can approximate S by S_0. Note that we did not actually use (4) and (6), but only $\text{im } Q \subset \ker S$. However, in order that S may be approximated by S_0 effectively, $\text{im } Q$ must be a large subset of $\ker S$ in some sense. Heuristically, if Q_0, S_0 are good approximations of Q, S respectively and $\text{im } Q_0 = \ker S_0$, then $\text{im } Q$ must necessarily be almost equal to $\ker S$. Thus it is not necessary that we must have $\xi - c \in \text{im } Q$. It will be good enough for the purpose of approximation if $\text{im } Q$ contains elements which are close to $\xi - c$ in some sense.

In applying this method to our problem, $\mathcal{Y}$ will be the linear space of all real-valued and bounded functions defined on $\{0, 1, 2, \ldots\}$, $\mathcal{X}$ will be a subspace of $\mathcal{Y}$, Q will be constructed using conditional expectations, Q_0 will be a difference operator, and S and S_0 will be integration operators with respect to μ and μ_0 respectively, where μ and μ_0 are two probability measures on $\{0, 1, 2, \ldots\}$. In order that we may carry out this idea with greater flexibility, we shall first derive an identity, containing an arbitrary function $f \in \mathcal{X}$, which is $SQf = 0$ and then choose f to be $Q_0^i(I_{\mathcal{Y}} - S_0)h$ where $h \in \mathcal{Y}$. Our aim is to approximate Sh by S_0h and estimate the error by bounding $S(I_{\mathcal{Y}} - QQ_0^i)(I_{\mathcal{Y}} - S_0)h$.

Perhaps it should be explained, at this point, how Q can be constructed using conditional expectations when $\mathcal{X}$ and $\mathcal{Y}$ are only linear spaces of functions defined on $\{0, 1, 2, \ldots\}$. Consider a

particular example in which $X_1, X_2, \ldots, X_n$ are m-dependent Bernoulli random variables with $P(X_i = 1) = 1 - P(X_i = 0) = p_i$. Let I be a random variable which is uniformly distributed on $\{1, 2, \ldots, n\}$ and is independent of the X_i's. Let $W = \sum_{i=1}^{n} X_i$, $W^* = \sum_{|i-I| > m} X_i$ and $\lambda = \sum_{i=1}^{n} p_i$. Then by using the m-dependence and the properties of conditional expectations, we have for every $f \in \mathcal{F}$,

$$(13)\quad E[\lambda f(W+1) - Wf(W) + nX_I[f(W) - f(W^*+1)] + np_I[f(W^*+1) - f(W+1)]] = 0$$

But this is also,

$$(14)\quad EE^W[\lambda f(W+1) - Wf(W) + nX_I[f(W) - f(W^*+1)] + np_I[f(W^*+1) - f(W+1)]] = 0$$

Thus (14) can be rewritten as

$$(15)\quad SQf = 0$$

where Q is defined by

$$(16)\quad Qf(w) = E^{W=w}[\lambda f(W+1) - Wf(W) + nX_I[f(W) - f(W^*+1)] + np_I[f(W^*+1) - f(W+1)]]$$

and S is defined by

(17) $SQf = \int Qf\, d\mu$, μ being the distribution of W.

For greater flexibility, we shall be using (13) rather than (15).

§2. Conditional expectations.

As has been mentioned in §1, we shall be using conditional expectations to construct the operator Q. In fact, in bounding the error $S(I_{\mathcal{Y}} - QQ_0^i)(I_{\mathcal{Y}} - S_0)h$, we shall be using their properties rather frequently. For this reason, we shall, in this section, briefly review the definition and some of the properties of conditional expectations.

Let $(\Omega, \mathcal{B}, P)$ be a probability space, $\mathcal{F}$ a sub-σ-algebra of $\mathcal{B}$ and X an integrable random variable. Recall that the conditional expectation of X given $\mathcal{F}$ is an $\mathcal{F}$-measurable random variable, which we shall denote by $E^{\mathcal{F}}X$, such that for every $A \in \mathcal{F}$, $\int_A E^{\mathcal{F}}X\, dP = \int_A X\, dP$. If $\mathcal{F} = \mathcal{B}(Z_\alpha, \alpha \in J)$ = the σ-algebra generated by the random variables Z_α, $\alpha \in J$, then there exists a function $g : \mathbb{R}^J \longrightarrow \mathbb{R}$ which is $\mathcal{B}(\mathbb{R})^J$-measurable such that $E^{\mathcal{F}}X = g(Z)$, where $\mathbb{R}$ is the real line, $\mathcal{B}(\mathbb{R})$ the σ-algebra of Borel subsets of $\mathbb{R}$, $\mathcal{B}(\mathbb{B})^J$ the σ-algebra generated by cylinder sets in $\mathbb{R}^J$ and $Z : \Omega \longrightarrow \mathbb{R}^J$ such that for every $\omega \in \Omega$, the α^{th} coordinate of $Z(\omega) = Z_\alpha(\omega)$. In this case, we shall sometimes write $E^{\mathcal{F}}X$ as $E^{Z}X$ or $E^{Z_\alpha,\ \alpha\in J}X$. The following are the properties of conditional expectations which we shall use throughout the subsequent chapters. They can also be found in Breiman [1]. All equalities and inequalities are to hold almost surely.

(1) $E^{\mathcal{F}}1 = 1$

(2) $E^{\mathcal{F}}(\alpha X_1 + \beta X_2) = \alpha E^{\mathcal{F}} X_1 + \beta E^{\mathcal{F}} X_2$ where $E|X_1| < \infty$ and $E|X_2| < \infty$

(3) $X \geq 0 \Longrightarrow E^{\mathcal{F}} X \geq 0$

(4) If $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$, then $E^{\mathcal{F}} E^{\mathcal{G}} X = E^{\mathcal{G}} E^{\mathcal{F}} X = E^{\mathcal{G}} X$

(5) If $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$ is a convex function, then $\varphi(E^{\mathcal{F}} X) \leq E^{\mathcal{F}} \varphi(X)$ (Jensen's inequality)

(6) Let Y be a random vector taking values in $(\mathbb{R}^J, \mathcal{B}(\mathbb{R})^J)$. Then for every $\varphi : \mathbb{R}^J \longrightarrow \mathbb{R}$ which is $\mathcal{B}(\mathbb{R})^J$-measurable such that $E|\varphi(Y)| < \infty$ and $E|X\varphi(Y)| < \infty$, $E^Y[\varphi(Y)X] = \varphi(Y)\, E^Y X$

(7) Let Y_1 and Y_2 be random vectors such that $\mathcal{B}(Y_1)$ is independent of $\mathcal{B}(X, Y_2)$ then $E^{Y_1, Y_2} X = E^{Y_2} X$.

In particular, $E^{Y_1} X = EX$

(8) Let Y_i be random vectors taking values in $(\mathbb{R}^{J_i}, \mathcal{B}(\mathbb{R})^{J_i})$, $i = 1, 2, 3$, respectively such that $\mathcal{B}(Y_1)$ is independent of $\mathcal{B}(Y_2, Y_3)$ and let J_2 be countable. Then for every $\varphi : \mathbb{R}^{J_1} \times \mathbb{R}^{J_2} \longrightarrow R$ which is $\mathcal{B}(\mathbb{R})^{J_1} \times \mathcal{B}(\mathbb{R})^{J_2}$-measurable such that $E|\varphi(Y_1, Y_2)| < \infty$, $E^{Y_1, Y_3} \varphi(Y_1, Y_2) = g(Y_1, Y_3)$ where $E^{Y_3} \varphi(y_1, Y_2) = g(y_1, Y_3)$. In particular, $E^{Y_1} \varphi(Y_1, Y_2) = h(Y_1)$ where $E\varphi(y_1, Y_2) = h(y_1)$.

Since the probability spaces considered in the subsequent chapters are finite, it is possible to prove the approximation

theorems in those chapters without using conditional expectations. However, the author has also elegance and abstraction in mind and it is for this reason that he has chosen to employ conditional expectations.

CHAPTER II

POISSON APPROXIMATION

§1. The basic identity

We begin this chapter with the derivation of the identity for the problem of Poisson approximation in its preliminary and abstract form. The identity will later be applied to two special cases, namely, sums of weakly dependent Bernoulli random variables and a randomly selected sum of Bernoulli random variables.

Let $(\Omega, \mathcal{B}, P)$ be a probability space and $\mathcal{C}$, $\mathcal{F}$, $\mathcal{G}$ be sub-σ-algebras of $\mathcal{B}$ such that $\mathcal{F}$ and $\mathcal{G}$ are independent. Suppose G is a $\mathcal{B}$-measurable random variable and W^* is a $\mathcal{C}$-measurable random variable, and that

$$W = E^{\mathcal{F}}G \tag{1}$$

$$EW = \lambda \quad (> 0) \tag{2}$$

Then for every bounded real-valued Borel function f on $\mathbb{R}$,

$$\begin{aligned} (3)\quad EWf(W) &= E(E^{\mathcal{F}}G)\, f(W) = EE^{\mathcal{F}}Gf(W) = EGf(W) \\ &= EG(f(W) - f(W^* + 1)) + E(G - E^{\mathcal{G}}G)\, f(W^* + 1) \\ &\qquad + E(E^{\mathcal{G}}G)(f(W^* + 1) - f(W + 1)) + E(E^{\mathcal{G}}G)\, f(W + 1) \end{aligned}$$

Using the independence of $\mathcal{F}$ and $\mathcal{G}$ and the fact that $EE^{\mathcal{G}}G = EE^{\mathcal{F}}G = EW = \lambda$, we have

$$(4) \qquad E(E^{\mathcal{G}}G)\, f(W + 1) = E\lambda f(W + 1)$$

Substituting (4) into (3) and rearranging the terms, we then obtain the following identity

$$(5) \quad E\{\lambda f(W + 1) - Wf(W) + G(f(W) - f(W^* + 1)) + f(W^* + 1)\, E^{\mathcal{C}}(G - E^{\mathcal{G}}G) + (E^{\mathcal{G}}G)(f(W^* + 1) - f(W + 1))\} = 0$$

To make the Poisson approximation more apparent, we choose f such that

$$(6) \qquad \lambda f(w + 1) - wf(w) = h(w) - \mathcal{P}_\lambda h$$

where h is another bounded real-valued Borel function on $\mathbb{R}$ and

$$(7) \qquad \mathcal{P}_\lambda h = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}\, h(k)$$

Since our aim is to approximate the distribution of W by the Poisson distribution with parameter λ, we may assume G to be non-negative and W and W^* to be non-negative integer-valued. In this case we may think of h as being defined on $\{0, 1, 2, \ldots\}$ and we see that the value of f at $w = 0$ does not enter into our

consideration at all. Therefore we may think of f as being defined on $\{1, 2, \ldots\}$.

The unique bounded solution of (6) is

$$(8)\quad f(w) = \frac{(w-1)!}{\lambda^w} \sum_{k=0}^{w-1} \frac{\lambda^k}{k!} \left(h(k) - \mathcal{P}_\lambda h\right)$$

$$= -\frac{(w-1)!}{\lambda^w} \sum_{k=w}^{\infty} \frac{\lambda^k}{k!} \left(h(k) - \mathcal{P}_\lambda h\right)$$

Substituting (8) into identity (5), we obtain

$$(9)\quad Eh(W) = \mathcal{P}_\lambda h - EG(f(W) - f(W^* + 1)) - Ef(W^* + 1)\, E^{\mathcal{L}}(G - E^{\mathcal{G}}G) + E(E^{\mathcal{G}}G)(f(W + 1) - f(W^* + 1))$$

where f is given by (8). Thus it becomes clear that the question of how well $Eh(W)$ can be approximated by $\mathcal{P}_\lambda h$ can be answered by bounding the error terms on the right hand side of (9).

§2. A characterization of the Poisson distribution

In this section T_λ is an operator on all real-valued bounded continuous functions on the real line $\mathbb{R}$ and is defined by

$$(1)\qquad T_\lambda f(w) = \lambda f(w + 1) - wf(w)\ ,$$

μ_λ is the Poisson distribution on $\mathbb{R}$ with parameter $\lambda > 0$, and all functions considered are real-valued, Borel and defined on $\mathbb{R}$.

Recall that in §1, in order to obtain (1.9) from identity (1.5), we chose f such that

$$T_\lambda f = h - \mathcal{P}_\lambda h \tag{2}$$

We shall here show that the Poisson distribution with parameter λ can be characterized by using T_λ.

Proposition 2.1. Let ν be a probability measure on $\mathbb{R}$. Then $\nu = \mu_\lambda$ if and only if for every bounded continuous function f,

$$\int T_\lambda f \, d\nu = 0 \tag{3}$$

Proof. Suppose $\nu = \mu_\lambda$. Then

$$\int T_\lambda f \, d\nu = \lambda \sum_{k=0}^{\infty} f(k+1) \, e^{-\lambda} \frac{\lambda^k}{k!} - \sum_{k=1}^{\infty} kf(k) \, e^{-\lambda} \frac{\lambda^k}{k!} = 0 \tag{4}$$

Note that for this to hold, f need only be defined on $\{1, 2, \ldots\}$. Conversely, by letting $f \equiv 1$, (3) implies

$$\int_{-\infty}^{\infty} |w| \, d\nu(w) < \infty \tag{5}$$

Next, by letting $f(w)$ be the real and imaginary parts of e^{itw}

separately, where $t \in \mathbb{R}$, (3) implies

$$\lambda \int_{-\infty}^{\infty} e^{it(w+1)}\, d\nu(w) - \int_{-\infty}^{\infty} we^{itw}\, d\nu(w) = 0 \tag{6}$$

Since this is true for every $t \in \mathbb{R}$ and (5) holds, it follows from the dominated convergence theorem that

$$\frac{1}{i}\frac{\partial}{\partial t} \int_{-\infty}^{\infty} e^{itw}\, d\nu(w) - \lambda e^{it} \int_{-\infty}^{\infty} e^{itw}\, d\nu(w) = 0 \tag{7}$$

Solving this differential equation and using the condition

$$\int_{-\infty}^{\infty} d\nu(w) = 1 \tag{8}$$

we obtain

$$\int_{-\infty}^{\infty} e^{itw}\, d\nu(w) = e^{\lambda(e^{it}-1)} \tag{9}$$

It follows that $\nu = \mu_\lambda$, since a distribution is characterized by its characteristic function. This completes the proof of Proposition 2.1.

In the next proposition we shall omit the proof of the necessity of (3), since it will be the same as that in Proposition 2.1.

Proposition 2.2. Let ν be a probability measure on $\mathbb{R}$. Then $\nu = \mu_\lambda$ if and only if for every continuous function f with compact support, (3) holds.

Proof. By Proposition 2.1, it suffices to show that (3) holds for every bounded continuous function f. Let h_n be a sequence of non-negative trapezoidal functions increasing monotonically to

$$h(w) = \begin{cases} 1 & \text{if } w > 0 \\ 0 & \text{if } w \leq 0 \end{cases} \tag{10}$$

Then for every n, (3) implies

$$\lambda \int_{-\infty}^{\infty} h_n(w + 1)\, d\nu(w) = \int_{-\infty}^{\infty} w h_n(w)\, d\nu(w) \tag{11}$$

By the monotone convergence theorem, the right hand side converges to $\int_0^\infty w\, d\nu(w)$ as $n \longrightarrow \infty$, while the left hand side remains uniformly bounded. Thus

$$\int_0^\infty w\, d\nu(w) < \infty \,. \tag{12}$$

Similarly,

$$\int_{-\infty}^{0} -w\, d\nu(w) < \infty \tag{13}$$

Now let f be a bounded continuous function and let f_n be a sequence of continuous functions with compact support such that

(14) $$|f_n| \leq |f| \quad \text{for every} \quad n,$$

and

(15) $$f_n \longrightarrow f \quad \text{pointwise as} \quad n \longrightarrow \infty .$$

Then

(16) $$T_\lambda f_n \longrightarrow T_\lambda f \text{ pointwise as } n \longrightarrow \infty$$

and

(17) $$\begin{aligned} |T_\lambda f(w)| &\leq \lambda |f_n(w+1)| + |w|\ |f_n(w)| \\ &\leq \lambda |f(w+1)| + |w|\ |f(w)| \end{aligned}$$

which, by the boundedness of f and (12) and (13), is integrable with respect to ν. Therefore, by the dominated convergence theorem,

(18) $$\int T_\lambda f_n d\nu \longrightarrow \int T_\lambda f d\nu \quad \text{as} \quad n \longrightarrow \infty .$$

But for every n,

(19) $$\int T_\lambda f_n d\nu = 0.$$

Hence (3) holds for bounded and continuous f and this proves the proposition.

§3. Preliminary results.

In order to obtain sharp bounds for the errors in our approximation problems, it is necessary to study adequately the solution of the difference equation (1.6). This section contains those results

concerning the solution of (1.6) that are relevant to our problems.

Lemma 3.1. For $0 < \lambda \leq w$, $1 \leq w$,

$$(1) \qquad \frac{(w-1)!}{\lambda^w} \sum_{k=w}^{\infty} \frac{\lambda^k}{k!} \leq \frac{3}{\sqrt{w}}$$

Proof. Let $\varphi(w)$ be an increasing function of w such that $1 \leq \varphi(w) \leq w$ and $\varphi(w) \longrightarrow \infty$ as $w \longrightarrow \infty$. Then

$$(2) \qquad \frac{(w-1)!}{\lambda^w} \sum_{k=w}^{\infty} \frac{\lambda^k}{k!} \leq \frac{1}{w} \sum_{k=0}^{\infty} \frac{w^k}{(w+1)\cdots(w+k)}$$

$$= \frac{1}{w} \left(\sum_{k=0}^{[\varphi(w)]-1} + \sum_{k=[\varphi(w)]}^{\infty} \right) \frac{w^k}{(w+1)\cdots(w+k)}$$

where $[\varphi(w)]$ is the greatest integer $\leq \varphi(w)$

$$\leq \frac{1}{w} \left\{ \varphi(w) + \sum_{k=0}^{\infty} \left(\frac{w}{w+\varphi(w)} \right)^k \right\}$$

$$= \frac{1}{w} \left(\varphi(w) + \frac{w}{\varphi(w)} + 1 \right)$$

Clearly, the optimal choice of $\varphi(w)$ is $\varphi(w) = \sqrt{w}$. Hence the lemma.

Lemma 3.2. For $1 \leq w \leq \lambda$,

$$(3) \qquad \frac{(w-1)!}{\lambda^w} \sum_{k=0}^{w-1} \frac{\lambda^k}{k!} \leq \frac{2}{\sqrt{\lambda}}$$

Proof. Let $\varphi(\lambda)$ be as in the proof of Lemma 1. Then by a similar argument,

$$(4) \qquad \frac{(w-1)!}{\lambda^w}\sum_{k=0}^{w-1}\frac{\lambda^k}{k!} \le \frac{1}{\lambda}\{\varphi(\lambda) + \sum_{k=0}^{\infty}\left(\frac{\lambda - \varphi(\lambda)}{\lambda}\right)^k\}$$

$$= \frac{1}{\lambda}\left(\varphi(\lambda) + \frac{1}{\varphi(\lambda)}\right)$$

$$= \frac{2}{\sqrt{\lambda}}$$

In the following, $\|\cdot\|$ will denote the sup norm μ_λ and T_λ will be the same as in §2. Thus T_λ^{-1} will be given by

$$(5) \qquad T_\lambda^{-1}g(w) = \frac{(w-1)!}{\lambda w}\sum_{k=0}^{w-1}\frac{\lambda^k}{k!}\,g(k) \qquad \text{for} \quad w = 1, 2, \ldots,$$

for every real-valued and bounded function g defined on $\{0,1,2,\ldots\}$. We also define two operators S_λ and U_λ as follows

$$(6) \qquad S_\lambda h = T_\lambda^{-1}(h - \mathcal{P}_\lambda h)$$

$$(7) \qquad U_\lambda h(w) = S_\lambda h(w+2) - S_\lambda h(w+1)$$

where h is any real-valued and bounded function defined on $\{0, 1, 2, \ldots\}$ and $\mathcal{P}_\lambda h$ is defined by (1.7). Note that while $T_\lambda^{-1}g$ and $S_\lambda h$ are defined on $\{1, 2, \ldots\}$, $U_\lambda h$ is defined on $\{0, 1, 2, \ldots\}$.

Proposition 3.1. $T_\lambda^{-1}g$ is bounded if and only if

$$\int g \, d\mu_\lambda = 0. \tag{8}$$

In the "if" part, we actually have

$$|T_\lambda^{-1}g| \le \frac{3}{(\lambda^{1/2} \vee 1)} \|g\|$$

Proof. We first prove the "if" part. (8) implies

$$T_\lambda^{-1}g(w) = \frac{(w-1)!}{\lambda^w} \sum_{k=0}^{w-1} \frac{\lambda^k}{k!} g(k) \tag{9}$$

$$= -\frac{(w-1)!}{\lambda^w} \sum_{k=w}^{\infty} \frac{\lambda^k}{k!} g(k)$$

If $0 < \lambda \le w$, then by Lemma 3.1

$$|T_\lambda^{-1}g(w)| \le \|g\| \frac{(w-1)!}{\lambda^w} \sum_{k=w}^{\infty} \frac{\lambda^k}{k!} \le \frac{3}{\lambda^{1/2} \vee 1} \|g\| \tag{10}$$

On the other hand, if $1 \le w \le \lambda$, then by Lemma 3.2,

$$|T_\lambda^{-1}g(w)| \le \|g\| \frac{(w-1)!}{\lambda^w} \sum_{k=0}^{w-1} \frac{\lambda^k}{k!} \le \frac{3}{\lambda^{1/2}} \|g\| \tag{11}$$

Next, the "only if" part. Since $T_\lambda^{-1}g$ is bounded, it follows from Proposition 2.1 that

(12) $$\int g \, d\mu_\lambda = \int T_\lambda(T_\lambda^{-1} g) \, d\mu_\lambda = 0.$$

This completes the proof.

Proposition 3.2.

(13) $$|S_\lambda h| \le \frac{6}{\lambda^{1/2} \vee 1} \|h\|$$

Proof. It follows immediately from Proposition 3.1 and the definition of S_λ.

Proposition 3.3.

(14) $$|U_\lambda h| \le \frac{8}{\lambda^{1/2} \vee 1} \|h\|$$

(15) $$|U_\lambda h(w)| \le \left(\frac{2}{\lambda} + \frac{6}{\lambda(\lambda^{1/2} \vee 1)} |w + 1 - \lambda|\right) \|h\|$$

Proof. It is easy to show that

(16) $$U_\lambda h(w) = \frac{1}{\lambda} [h(w + 1) - \mathcal{P}_\lambda h] + \frac{w + 1 - \lambda}{\lambda} S_\lambda h(w + 1)$$

$$= \frac{1}{w + 1} [h(w + 1) - \mathcal{P}_\lambda h] + \frac{w + 1 - \lambda}{w + 1} S_\lambda h(w + 2)$$

(14) and (15) then follow from (16) and Proposition 3.2.

§4. The sum of a weakly dependent sequence of Bernoulli random variables.

In this section, we prove two theorems in which bounds in fairly general forms are obtained for the error in approximating the distribution of a sum of weakly dependent Bernoulli random variables by the Poisson distribution. Theorem 4.1 provides a sharper bound than Theorem 4.2 when the parameter λ of the approximating Poisson distribution is small, and is more effective than the second theorem for proving limit theorems holding λ fixed. On the other hand, the bound obtained from Theorem 4.2 is sharper when λ is large and is in fact a theoretical improvement of Theorem 4.1. To illustrate the last point, let us consider the sum of n independent and identically distributed Bernoulli random variables, $X_1, X_2, \ldots, X_n$, with $P(X_i = 1) = 1 - P(X_i = 0) = p$. The bound obtained from Theorem 4.1 is $A(\frac{\lambda}{\lambda^{1/2} \vee 1})p$ with $\lambda = np$, whereas by Theorem 4.2, it is Bp, where A and B are absolute constants. Clearly, the bound obtained from Theorem 4.2 depends on the parameter of the approximating distribution only through p, but it is not so with the other bound.

Let $X_1, X_2, \ldots, X_n$ be a sequence of Bernoulli random variables with

$$P(X_i = 1) = 1 - P(X_i = 0) = p_i, \qquad i = 1, 2, \ldots, n \tag{1}$$

and

$$\operatorname{cov}(X_i, X_j) = C_{ij} \tag{2}$$

and let

$$\lambda = \sum_{i=1}^{n} p_i \tag{3}$$

Recall that in §1 we have derived a fairly abstract formula (1.9) for the Poisson approximation. We shall now use this to derive a formula expressing the error in approximating the distribution of $\sum_{i=1}^{n} X_i$ by the Poisson distribution with parameter λ. For this, we introduce a random variable I which is uniformly distributed on $\{1, 2, \ldots, n\}$ and is independent of $X_1, X_2, \ldots, X_n$ and let

$$
(4)\quad
\begin{cases}
\mathcal{B} = \mathcal{B}(I, X_1, \ldots, X_n) \\
\quad = \text{the } \sigma\text{-algebra generated by } I, X_1, \ldots, X_n \\
\mathcal{F} = \mathcal{B}(X_1, \ldots, X_n) \\
\mathcal{G} = \mathcal{B}(I) \\
\mathcal{C} = \mathcal{B}(I, X_i \text{ for } |i - I| > m), \text{ where } m \text{ is an integer} \geq 0 \\
G = nX_I \\
W^* = \sum_{|i-I| > m} X_i \\
W^{(I)} = \sum_{i \neq I} X_i \\
X_i = 0 \text{ for } i \leq 0 \text{ or } i \geq n + 1 \\
Y_i = W^* + \sum_{k=I-m}^{i} X_k \\
Y_i^{(I)} = W^* + \sum_{\substack{k=I-m \\ k \neq I}}^{i} X_k
\end{cases}
$$

Then

$$
(5)\qquad \begin{cases} W = \sum_{i=1}^{n} X_i \\ E^{\mathcal{G}} G = E^{I} n X_I = n p_I \\ Y_i = Y_{I+m} = W \quad \text{for} \quad i \geq n+1 \\ Y_i^{(I)} = Y_{I+m}^{(I)} = W^{(I)} \quad \text{for} \quad i \geq n+1 \\ Y_i = Y_{I-m-1} = W^* \quad \text{for} \quad i \leq 0 \\ Y_i^{(I)} = Y_{I-m-1}^{(I)} = W^* \quad \text{for} \quad i \leq 0 \end{cases}
$$

and from (1.9) we obtain

$$
(6)\qquad Eh(W) = \mathcal{P}_\lambda h - nEX_I(f(W) - f(W^* + 1)) - nEf(W^* + 1)\, E^{\mathcal{C}}(X_I - p_I)
$$
$$
+ nEp_I(f(W + 1) - f(W^* + 1))
$$

where as in §1, h is a bounded real-valued function defined on $\{0, 1, 2, \ldots\}$ and

$$
(7)\qquad \mathcal{P}_\lambda h = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}\, h(k)
$$

and f is the unique bounded solution of

$$
(8)\qquad \lambda f(w + 1) - wf(w) = h(w) - \mathcal{P}_\lambda h\ ,
$$

f being defined on $\{1, 2, \ldots\}$. Now

$$(9)\qquad X_I[f(W^* + 1) - f(W)] = X_I[f(W^* + 1) - f(W^{(I)} + 1)]$$

$$= \sum_{\substack{i=I-m \\ i\neq I}}^{I+m} X_I[f(Y_{i-1}^{(I)} + 1) - f(Y_i^{(I)} + 1)]$$

$$= \sum_{\substack{i=I-m \\ i\neq I}}^{I+m} X_I X_i[f(Y_{i-1}^{(I)} + 1) - f(Y_{i-1}^{(I)} + 2)]$$

$$= - \sum_{\substack{i=I-m \\ i\neq I}}^{I+m} X_I X_i U_\lambda h(Y_{i-1}^{(I)})$$

where U_λ is defined by (3.7). The first equality of (9) follows from the simple observation that

$$(10)\qquad X_I f(W) = X_I f(W^{(I)} + 1)$$

and the third equality follows from a further observation that

$$(11)\qquad f(Y_{i-1}^{(I)} + 1) - f(Y_i^{(I)} + 1) = f(Y_{i-1}^{(I)} + 1) - f(Y_{i-1}^{(I)} + X_i + 1)$$

$$= X_i[f(Y_{i-1}^{(I)} + 1) - f(Y_{i-1}^{(I)} + 2]$$

Of course both (10) and (11) are due to the simple fact that a Bernoulli random variable takes values 0 and 1.

Similarly,

$$(12)\qquad f(W^* + 1) - f(W + 1) = \sum_{i=I-m}^{I+m} [f(Y_{i-1} + 1) - f(Y_i + 1)]$$

$$= \sum_{i=I-m}^{I+m} X_i [f(Y_{i-1} + 1) - f(Y_{i-1} + 2)]$$

$$= - \sum_{i=I-m}^{I+m} X_i U_\lambda h(Y_{i-1})$$

Finally, writing $f = S_\lambda h$ where S_λ is defined by (3.6), we can rewrite (6) as

$$(13) \quad Eh(W) = \mathcal{P}_\lambda h - nE \sum_{\substack{i=I-m \\ i \neq I}}^{I+m} X_I X_i U_\lambda h(Y_{i-1}^{(I)}) - nE[E\,\mathcal{C}(X_I - p_I)] S_\lambda h(W^* + 1)$$

$$- + nE \sum_{i=I-m}^{I+m} p_I X_i U_\lambda h(Y_{i-1})$$

Theorem 4.1. For $|h| \leq 1$ and $m = 0, 1, 2, \ldots$,

$$(14) \quad |Eh(\sum_{i=1}^{n} X_i) - \mathcal{P}_\lambda h| \leq R$$

where

$$(15) \quad R = \frac{8}{\lambda^{1/2} \vee 1} \sum_{i=1}^{n} p_i^2 + \frac{32\lambda m\tilde{p}}{\lambda^{1/2} \vee 1}$$

$$+ \frac{16C_m}{\lambda^{1/2} \vee 1} + \frac{6n}{\lambda^{1/2} \vee 1} E|E\,\mathcal{C}(X_i - p_I)|$$

$$(16) \quad \tilde{p} = \max_{1 \leq i \leq n} p_i ,$$

and

$$(17) \quad C_m = \sum_{\substack{i < j \\ j-i \leq m}} C_{ij}$$

Proof. Applying Proposition 3.2 and (3.14) of Proposition 3.3 to (13), we obtain

$$(18)\quad |Eh(W) - \mathcal{P}_\lambda h| \le \frac{8n}{\lambda^{1/2} \vee 1} E \sum_{\substack{i=I-m \\ i\neq I}}^{I+m} X_I X_i + \frac{6n}{\lambda^{1/2}\vee 1} E|E^{I}(X_I - p_I)| + \frac{8n}{\lambda^{1/2}\vee 1} E \sum_{i=I-m}^{I+m} p_I X_i$$

The theorem then follows from

$$(19)\quad E \sum_{\substack{i=I-m \\ i\neq I}}^{I+m} X_I X_i = E \sum_{\substack{i=I-m \\ i\neq I}}^{I+m} E^{I} X_I X_i$$

$$= E \sum_{\substack{i=I-m \\ i\neq I}}^{I+m} (p_I p_i + C_{Ii}) \le 2m\tilde{p}Ep_I + \frac{2C_m}{n}$$

$$= \frac{2\lambda m\tilde{p}}{n} + \frac{2C_m}{n}$$

and

$$(20)\quad E \sum_{i=I-m}^{I+m} p_I X_i = E \sum_{i=I-m}^{I+m} E^{I} p_I X_i = E \sum_{i=I-m}^{I+m} p_I p_i$$

$$= Ep_I^2 + 2m\tilde{p}Ep_I = \frac{1}{n}\sum_{i=1}^{n} p_i^2 + \frac{2\lambda m\tilde{p}}{n}$$

Theorem 4.2. For $|h| \le 1$ and $m = 0, 1, 2, \ldots$

$$(21)\quad |Eh(\sum_{i=1}^{n} X_i) - \mathcal{P}_\lambda h| \le R$$

where

$$(22)\quad R = \frac{8(3m+1)}{\lambda}\sum_{i=1}^{n} p_i^2 + \frac{6\delta}{\lambda(\lambda^{1/2}\vee 1)}\sum_{i=1}^{n} p_i^2$$

$$+ 4m(27m+8)\tilde{p} + \frac{24m\delta\tilde{p}}{\lambda^{1/2}\vee 1} + \frac{4(15m+4)C_m}{\lambda} + \frac{12\delta C_m}{\lambda(\lambda^{1/2}\vee 1)}$$

$$+ \frac{6n}{\lambda(\lambda^{1/2}\vee 1)} E \sum_{\substack{i=I-m \\ i\neq I}}^{I+m} \min\left\{ nE^I\left|E^{\mathcal{C}_{I,i}}(X_I X_i - E^I X_I X_i)\right|, \right.$$

$$\left. \delta\sqrt{\mathrm{var}^I(E^{\mathcal{C}_{I,i}} X_I X_i)} \right\}$$

$$+ \frac{6n}{\lambda(\lambda^{1/2}\vee 1)} E \sum_{i=I-m}^{I+m} p_I \min\left\{ nE^I\left|E^{\mathcal{C}_i}(X_i - p_i)\right|, \right.$$

$$\left. \delta\sqrt{\mathrm{var}^I(E^{\mathcal{C}_i} X_i)} \right\}$$

$$+ \frac{6n}{\lambda^{1/2}\vee 1} E\left|E^{\mathcal{C}}(X_I - p_I)\right|$$

$$(23)\quad \mathcal{C}_{I,i} = \mathcal{B}(I,\ X_k \text{ for } |k-I| > m \text{ and } |k-i| > m)$$

$$(24)\quad \mathcal{C}_i = \mathcal{B}(X_k, \text{ for } |k-i| > m)$$

$$(25)\quad \delta = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} |c_{ij}|}$$

and $\tilde{p}$ and C_m are defined by (16) and (17) respectively.

Proof. Applying Proposition 3.2 and (3.15) of Proposition 3.3 to (13), we obtain

$$(26)\quad |Eh(W) - \mathcal{P}_\lambda h|$$

$$\le nE \sum_{\substack{i=I-m \\ i\neq I}}^{I+m} E^I X_I X_i \left\{ \frac{2}{\lambda} + \frac{6}{\lambda(\lambda^{1/2}\vee 1)} |Y_{i-1}^{(I)} + 1 - \lambda| \right\}$$

$$+ \frac{6n}{\lambda^{1/2}\vee 1} E|E^{\mathcal{L}}(X_i - p_I)$$

$$+ nE \sum_{i=I-m}^{I+m} E^I p_I X_i \left\{ \frac{2}{\lambda} + \frac{6}{\lambda(\lambda^{1/2}\vee 1)} |Y_{i-1} + 1 - \lambda| \right\}$$

Now

$$(27)\quad E^I X_I X_i \left\{ \frac{2}{\lambda} + \frac{6}{\lambda(\lambda^{1/2}\vee 1)} |Y_{i-1}^{(I)} + 1 - \lambda| \right\}$$

$$\le E^I X_I X_i \left\{ \frac{30m + 8}{\lambda} + \frac{6}{\lambda(\lambda^{1/2}\vee 1)} |Y_{i-1}^{(I)} - Z_i^{(I)} - EY_{i-1}^{(I)} + EZ_i^{(I)}| \right\}$$

$$= \frac{(30m + 8)(p_I p_i + C_{Ii})}{\lambda} + \frac{6(p_I p_i + C_{Ii})}{\lambda(\lambda^{1/2}\vee 1)} E^I |Y_{i-1}^{(I)} - Z_i^{(I)} - EY_{i-1}^{(I)} + EZ_i^{(I)}|$$

$$+ \frac{6}{\lambda(\lambda^{1/2}\vee 1)} E^I \left[E^{\mathcal{L}_{I,i}} (X_I X_i - E^I X_I X_i) \right] |Y_{i-1}^{(I)} - Z_i^{(I)} - EY_{i-1}^{(I)} + EZ_i^{(I)}|$$

where

$$(28)\quad Z_i^{(I)} = \sum_{\substack{k=(i-m)\wedge(I-m) \\ k\neq I}}^{i-1} X_k + \sum_{k=I+m+1}^{i+m} X_k$$

By Jensen's inequality,

$$(29)\quad E^I|Y_{i-1}^{(I)}-Z_i^{(I)}-EY_{i-1}^{(I)}+EZ_i^{(I)}|$$

$$\le \sqrt{E^I(Y_{i-1}^{(I)}-Z_i^{(I)}-EY_{i-1}^{(I)}+EZ_i^{(I)})^2} \le \delta$$

and by Hölder's inequality

$$(30)\quad \left|E^I[E^{\mathcal{C}_{I,i}}(X_IX_i-E^IX_IX_i)]|Y_{i-1}^{(I)}-Z_i^{(I)}-EY_{i-1}^{(I)}+EZ_i^{(I)}|\right|$$

$$\le \sqrt{E^I(E^{\mathcal{C}_{I,i}}(X_IX_i-E^IX_IX_i))^2\ E^I(Y_{i-1}^{(I)}-Z_i^{(I)}-EY_{i-1}^{(I)}+EZ_i^{(I)})^2}$$

$$\le \delta\sqrt{\mathrm{var}^I(E^{\mathcal{C}_{I,i}}X_IX_i)}$$

Also,

$$(31)\quad \text{the left hand side of (30)} \le nE^I|E^{\mathcal{C}_{I,i}}(X_IX_i-E^IX_IX_i)|$$

Thus we have from (27), (29), (30) and (31),

$$(32)\quad nE\sum_{\substack{i=I-m\\ i\neq I}}^{I+m} E^IX_IX_i\left\{\frac{2}{\lambda}+\frac{6}{\lambda(\lambda^{1/2}\vee 1)}\ |Y_{i-1}^{(I)}+1-\lambda|\right\}$$

$$\le 4m(15m+4)\tilde{p}+\frac{4(15m+4)C_m}{\lambda}+\frac{12m\delta\tilde{p}}{\lambda^{1/2}\vee 1}+\frac{12\delta C_m}{\lambda(\lambda^{1/2}\vee 1)}$$

$$+\frac{6n}{\lambda(\lambda^{1/2}\vee 1)}E\sum_{\substack{i=I-m\\ i\neq I}}^{I+m}\min\left\{nE^I|E^{\mathcal{C}_{I,i}}(X_IX_i-E^IX_IX_i)|,\ \delta\sqrt{\mathrm{var}^I(E^{\mathcal{C}_{I,i}}X_IX_i)}\right\}$$

Similarly,

$$(33)\quad E^I p_I X_i \left\{ \frac{2}{\lambda} + \frac{6}{\lambda(\lambda^{1/2} \vee 1)} |Y_{i-1} + 1 - \lambda| \right\}$$

$$\leq E^I p_I X_i \left\{ \frac{24m+8}{\lambda} + \frac{6}{\lambda(\lambda^{1/2} \vee 1)} |Y_{i-1} - Z_i - EY_{i-1} + EZ_i| \right\}$$

$$= \frac{(24m+8)p_I p_i}{\lambda} + \frac{6p_I p_i}{\lambda(\lambda^{1/2} \vee 1)} E^I |Y_{i-1} - Z_i - EY_{i-1} + EZ_i|$$

$$+ \frac{6p_I}{\lambda(\lambda^{1/2} \vee 1)} E^I [E^{\mathcal{L}_i}(X_i - p_i)] |Y_{i-1} - Z_i - EY_{i-1} + EZ_i|$$

where

$$(34)\quad Z_i = \sum_{k=i-m}^{i-1} X_k + \sum_{k=I+m+1}^{i+m} X_k$$

and

$$(35)\quad nE \sum_{i=I-m}^{I+m} E^I p_I X_i \left\{ \frac{2}{\lambda} + \frac{6}{\lambda(\lambda^{1/2} \vee 1)} |Y_{i-1} + 1 - \lambda| \right\}$$

$$\leq \frac{8(3m+1)}{\lambda} \sum_{i=1}^{n} p_i^2 + \frac{6\delta}{\lambda(\lambda^{1/2} \vee 1)} \sum_{i=1}^{n} p_i^2 + 16m(3m+1)\tilde{p} + \frac{12m\delta\tilde{p}}{\lambda^{1/2} \vee 1}$$

$$+ \frac{6n}{\lambda(\lambda^{1/2} \vee 1)} E \sum_{i=I-m}^{I+m} p_I \min \left\{ nE^I |E^{\mathcal{L}_i}(X_i - p_i)|, \delta \sqrt{\mathrm{var}^I(E^{\mathcal{L}_i} X_i)} \right\}$$

(21) and (22) then follow from (26), (32) and (35).

§5. Special cases of weak dependence.

In this section, we shall use Theorems 4.1 and 4.2 to obtain explicit bounds for three special cases of weak dependence. They are m-dependence, exponentially decreasing dependence and Markovian dependence, all with small covariances or correlations. Note that m-dependence is a special case of mixing conditions and the exponentially decreasing dependence described in Theorem 5.3 is a consequence of a certain mixing condition (see Philipp (1969)). We shall also prove Le Cam's result (1960). All notation will be the same as in §4 unless otherwise defined.

Theorem 5.1. (Le Cam) If $X_1, X_2, \ldots, X_n$ are independent, then for $|h| \leq 1$

$$(1) \qquad |Eh(\sum_{i=1}^{n} X_i) - \wp_\lambda h| \leq \begin{cases} \frac{A}{\lambda^{1/2}\vee 1} \sum_{i=1}^{n} p_i^2 \\ \frac{B}{\lambda} \sum_{i=1}^{n} p_i^2 \end{cases}$$

where A and B are absolute constants, which can be taken to be 8 and 14 respectively.

Proof. Since $X_1, X_2, \ldots, X_n$ are independent, we take $m = 0$. Then many terms in (4.15) and (4.22) vanish. Also

$$(2)\quad E|E^{\mathcal{C}}(X_I - p_I)| \le \sqrt{E(E^{\mathcal{C}}(X_I - p_I))^2}$$

$$= \sqrt{E(X_I - p_I)\, E^{\mathcal{C}}(X_I - p_I)} = \sqrt{EE^I[(X_I - p_I)\, E^{\mathcal{C}}(X_I - p_I)]}$$

$$= \sqrt{E[E^I(X_I - p_I)\, E^I E^{\mathcal{C}}(X_I - p_I)]} = 0$$

where the second last equality follows from the conditional independence of X_I and $E^{\mathcal{C}}X_I$ given I.

Applying the same argument to two other terms in (4.22) and noting

$$(3)\qquad \delta \le \lambda^{1/2}\ ,$$

we obtain (1) from (4.15) and (4.22) of Theorems (4.1) and (4.2) respectively.

Definition 5.1. A sequence (finite or infinite) of random variables $Y_1, Y_2, \ldots$ is m-dependent if for every two disjoint subsets A and B of natural numbers such that $\inf_{\substack{i\in A\\ j\in B}} |i-j| > m$, $(Y_i)_{i\in A}$ and $(Y_j)_{j\in A}$ are independent.

Lemma 5.1.

$$(4)\qquad -p_i p_j \le C_{ij} \le \min(p_i, p_j)$$

Proof. It follows from

$$EX_iX_j = p_ip_j + C_{ij} \tag{5}$$

and

$$0 \le EX_iX_j \le \min(p_i,p_j) \tag{6}$$

Lemma 5.2. Suppose $Y_1, Y_2, \ldots, Y_n$ is any m-dependent sequence of random variables with finite variances. Then

$$\Big|\sum_{i<j} \operatorname{cov}(Y_i,Y_j)\Big| \le \frac{m}{2} \sum_{i=1}^{n} \operatorname{var} Y_i \tag{7}$$

Proof. By the positive semi-definiteness of the covariance matrix of any two random variables Y and Z with finite variances, we have

$$|\operatorname{cov}(Y,Z)| \le \sqrt{\operatorname{var} Y \operatorname{var} Z} \le \frac{1}{2}(\operatorname{var} Y + \operatorname{var} Z) \tag{8}$$

Now let

$$Z_i = Y_i + Y_{i+m+1} + Y_{i+2m+2} + \cdots, \qquad i = 1, 2, \ldots, m+1 \tag{9}$$

Then for every pair (i,j), $i \neq j$,

$$|\operatorname{cov}(Z_i,Z_j)| \le \frac{1}{2}(\operatorname{var} Z_i + \operatorname{var} Z_j) \tag{10}$$

Summing all possible pairs (i,j), $i \neq j$, and using the mutual independence of the Y's in each Z_i, $i = 1, 2, \ldots, m+1$, we obtain (7).

Lemma 5.3. If $X_1, X_2, \ldots, X_n$ are m-dependent, then

$$\delta \leq \lambda^{1/2}\left(\frac{5m}{2} + 1\right) \tag{11}$$

Proof.

$$\begin{aligned} \delta^2 &= \sum_{i=1}^{n} \sum_{j=1}^{n} |c_{ij}| = \sum_{|i-j| \leq m} |c_{ij}| \\ &= \sum_{i=1}^{n} c_{ii} + 2 \sum_{\substack{i<j \\ j-i \leq m}} |c_{ij}| \\ &= \sum_{i=1}^{n} c_{ii} + 2 \sum_{i<j} c_{ij} - 4 \sum_{\substack{i<j \\ j-i \leq m \\ c_{ij} \leq 0}} c_{ij} \end{aligned} \tag{12}$$

which by Lemmas 5.1 and 5.2

$$\leq \lambda + \lambda m + 4 \sum_{\substack{i<j \\ j-i \leq m}} p_i p_j$$

$$\leq \lambda + \lambda m + 4\lambda m \tilde{p} \leq \lambda(5m + 1)$$

This immediately implies (11).

Theorem 5.2. If $X_1, X_2, \ldots, X_n$ are m-dependent, then for $|h| \le 1$,

$$(13) \qquad |Eh(\sum_{i=1}^{n} X_i) - \mathcal{P}_\lambda h| \le \begin{cases} \frac{\lambda}{\lambda^{1/2} \vee 1}\left[A_1(m)\tilde{p} + A_2 \frac{|C_m|}{\lambda}\right] \\ B_1(m)\tilde{p} + B_2(m)\frac{|C_m|}{\lambda} \end{cases}$$

where

$$(14) \qquad \begin{cases} A_1(m) = 32m + 8 \\ A_2 = 16 \\ B_1(m) = 168m^2 + 95m + 14 \\ B_2(m) = 90m + 28 \end{cases}$$

Proof. We first choose the m in (4.15) and (4.22) of Theorems 4.1 and 4.2 respectively to be the same m as the m-dependence in the hypothesis. Then by arguments similar to (2), the last term of (4.15) and the last three terms of (4.22) vanish. Finally, using Lemma 5.3 for three of the remaining terms in (4.22), we obtain (13).

Corollary 5.1. Suppose $X_1, X_2, \ldots, X_n$ are stationary and m-dependent with

$$(15) \qquad P(X_i = 1) = 1 - P(X_i = 0) = p$$

and

$$(16) \qquad \operatorname{corr}(X_i, X_{i+k}) = \rho_k$$

Then for $|h| \le 1$,

$$(17)\qquad |\mathrm{E}h(\sum_{i=1}^{n} X_i) - \mathcal{P}_\lambda h| \le \begin{cases} \dfrac{\lambda}{\lambda^{1/2} \vee 1}\,[A_1(m)p + A_2\,|\sum_{i=1}^{m} \rho_i|] \\ B_1(m)p + B_2(m)\,|\sum_{i=1}^{m} \rho_i| \end{cases}$$

where $A_1(m)$, A_2, $B_1(m)$ and $B_2(m)$ are given by (14).

Theorem 5.3. Let $X_1, X_2, \ldots, X_n$ be stationary with

$$(18)\qquad P(X_i = 1) = 1 - P(X_i = 0) = p$$

$$(19)\qquad \mathrm{corr}(X_i, X_{i+k}) = \rho_k$$

Suppose that for $m = 0, 1, 2, \ldots, n-2$ and for every pair of bounded random variables Y and Z which depend on $\{X_i\}_{i\in A}$ and $\{X_j\}_{j\in B}$ respectively, where A and B are two disjoint subsets of $\{1, 2, \ldots, n\}$ such that $\inf_{\substack{i\in A \\ j\in B}} |i-j| > m$, we have

$$(20)\qquad |\mathrm{corr}(Y,Z)| \le e^{-\alpha(m+1)}, \quad \alpha > 0$$

Then for $n \ge e^{(2\alpha)/3}$ and $|h| \le 1$,

$$(21)\qquad |\mathrm{E}h(\sum_{i=1}^{n} X_i) - \mathcal{P}_\lambda h|$$

$$\leq \begin{cases} \dfrac{\lambda}{\lambda^{1/2} \vee 1}\left[A_1(\alpha)(\log p^{-1})p + A_2\left|\sum_{i=1}^{[\frac{3}{2\alpha}\log p^{-1}]} \rho_i\right|\right], & p \leq e^{-(2\alpha)/3} \\ B_1(\alpha)(\log \lambda p^{-1})^2 p + B_2(\alpha)(\log \lambda p^{-1})\left|\sum_{i=1}^{[\frac{3}{2\alpha}\log \lambda p^{-1}]} \rho_i\right|, & \lambda \geq 1 \end{cases}$$

where

$$(22) \qquad \begin{cases} A_1(\alpha) = \dfrac{69}{\alpha} \\ A_2 = 16 \\ B_1(\alpha) = \dfrac{441}{\alpha^2} + \dfrac{189}{2\alpha^2}\left(\dfrac{e^{\alpha}+1}{e^{\alpha}-1}\right)^{1/2} \\ B_2(\alpha) = \dfrac{114}{\alpha} + \dfrac{18}{\alpha}\left(\dfrac{e^{\alpha}+1}{e^{2}-1}\right)^{1/2} \end{cases}$$

Proof. It consists of the following lemmas.

Lemma 5.4.

$$(23) \qquad \sqrt{\mathrm{var}^I(E^{\mathcal{E}_{I,i}} X_I X_i)} \leq p^{1/2}\, e^{-\alpha(m+1)}$$

$$(24) \qquad E^I\left|E^{\mathcal{E}_I}(X_I - p)\right| \leq p^{1/2}\, e^{-\alpha(m+1)}$$

$$(25) \qquad E^I\left|E^{\mathcal{E}_i}(X_i - p)\right| \leq p^{1/2}\, e^{-\alpha(m+1)}$$

Proof of Lemma 5.4.

(23) follows from

$$(26)\quad \mathrm{var}^I(E^{\mathcal{C}_{I,i}X_IX_i}) = E^I[E^{\mathcal{C}_{I,i}(X_IX_i - E^IX_IX_i)}]^2$$

$$= E^I[(X_IX_i - E^IX_IX_i)\, E^{\mathcal{C}_{I,i}(X_IX_i - E^IX_IX_i)}]$$

$$\le e^{-\alpha(m+1)} \sqrt{\mathrm{var}^I X_IX_i \ \mathrm{var}^I(E^{\mathcal{C}_{I,i}X_IX_i})}$$

$$\le p^{1/2}\, e^{-\alpha(m+1)} \sqrt{\mathrm{var}^I(E^{\mathcal{C}_{I,i}X_IX_i})}$$

For (24), first obtain

$$(27)\qquad E^I|E^{\mathcal{C}}(X_I - p)| \le \sqrt{E^I[E^{\mathcal{C}}(X_I - p)]^2}$$

and then apply the same argument as (26). The proof for (25) is the same as that for (24).

Lemma 5.5.

$$(28)\qquad \delta \le \lambda^{1/2}\left(\frac{e^{\alpha}+1}{e^{\alpha}-1}\right)^{1/2}$$

Proof of Lemma 5.5.

$$(29)\qquad \delta^2 = \sum_{i,j} |C_{ij}| \le npq + 2pq \sum_{k=1}^{n-1} k e^{-\alpha(n-k)} \le \lambda\left(\frac{e^{\alpha}+1}{e^{\alpha}-1}\right)$$

Finally, in order to obtain (21), we apply Lemma 5.4 to (4.15) of Theorem 4.1 with $m = [\frac{3}{2\alpha} \log p^{-1}]$ and Lemmas 5.4 and 5.5 to (4.22) of Theorem 4.2 with $m = [\frac{3}{2\alpha} \log \lambda p^{-1}]$.

Theorem 5.4. Let $X_1, X_2, \ldots, X_n$ be stationary and Markovian with

$$P(X_i = 1) = 1 - P(X_i = 0) = p \tag{30}$$

and

$$\operatorname{corr}(X_i, X_{i+1}) = \rho \tag{31}$$

Then for $|h| \leq 1$,

$$|Eh(\sum_{i=1}^{n} X_i) - \Phi_\lambda h| \leq \begin{cases} \frac{\lambda}{\lambda^{1/2} \vee 1} [A_1 p + A_2 |\rho|] \\ B_1 p + B_2 \lambda^{1/2} |\rho| \end{cases} \tag{32}$$

where A_1, A_2, B_1 and B_2 are absolute constants, which can be taken to be 8, 48, 20 and 108 respectively.

Proof. It consists of the following lemmas.

Lemma 5.6.

$$\rho_k = \rho^k \tag{33}$$

where ρ_k is the correlation between X_i and X_{i+k}

Proof of Lemma 5.6. If $pq = 0$, (33) is trivially true. Thus we consider the case $pq > 0$. By definition $\rho_1 = \rho$. For $k \geq 2$,

$$(34)\quad \rho_k pq = \operatorname{cov}(X_i, X_{i+k}) = EX_iX_{i+k} - p^2 = P(X_i = 1,\ X_{i+k} = 1) - p^2$$

$$= \sum_{x=0}^{1} P(X_{i+k} = 1 | X_{i+1} = x)\ P(X_{i+1} = x | X_i = 1)\ P(X_i = 1) - p^2$$

$$= \sum_{x=0}^{1} \frac{P(X_{i+k} = 1,\ X_{i+1} = x)\ P(X_{i+1} = x,\ X_i = 1)}{P(X_{i+1} = x)} - p^2$$

$$= \frac{(\rho_{k-1}pq + p^2)(\rho pq + p^2)}{p} + \frac{(p - \rho_{k-1}pq - p^2)(p - \rho pq - p^2)}{q} - p^2$$

$$= \rho\rho_{k-1}pq$$

Hence

$$(35)\qquad \rho_k = \rho\rho_{k-1} = \rho^k \qquad \text{by induction.}$$

Lemma 5.7. For $m = 0$,

$$(36)\qquad E^I |E^{\mathcal{L}}(X_I - p)| \leq 8|\rho|pq$$

Proof of Lemma 5.7. We first show that $E^{\mathcal{L}}(X_I - p)$ is a function of I, X_{I-1} and X_{I+1} using the Markov property. In the case $I = 1$ or n, we shall take X_0 and X_{n+1} to be identically zero.

$$(37)\qquad E^{\mathcal{C}}(X_I - p) = P(X_I = 1 | I,\ X_k,\ k = 1, 2, \ldots, n,\ k \neq I) - p$$

$$= \frac{P(I,\ X_I = 1,\ X_k,\ k \neq I)}{P(I,\ X_k,\ k \neq I)} - p$$

$$= \frac{P(X_{I+1} | I,\ X_I = 1)\ P(X_I = 1 | I,\ X_{I-1})}{P(X_{I+1} | I,\ X_{I-1})} - p$$

$$= \varphi(I,\ X_{I-1},\ X_{I+1}) \quad \text{say,}$$

where we define for any two random vectors Y and Z $P(Y)$, $P(Y|Z)$, $P(Y = y|Z)$ and $P(Y|Z = z)$ to be functions $g_1(Y)$, $g_2(Y,Z)$, $g_3(Z)$ and $g_4(Y)$ such that

$$(38)\qquad \begin{cases} g_1(y) = P(Y = y) \\ g_2(y,z) = P(Y = y | Z = z) \\ g_3(z) = P(Y = y | Z = z) \\ g_4(y) = P(Y = y | Z = z) \end{cases}$$

Now we evaluate $\varphi(I,\ X_{I-1},\ X_{I+1})$

We consider three cases

Case (i). $I = 1$.

$$(39)\qquad \varphi(1,0,X_2) = \frac{P(X_2 | X_1 = 1)\ P(X_1 = 1)}{P(X_2)} - p = \frac{P(X_2,\ X_1 = 1)}{P(X_2)} - p$$

Thus

(40)
$$\begin{cases} \varphi(1,0,0) = -\rho p \\ \varphi(1,0,1) = \rho p \end{cases}$$

Case (ii). $I = n$

(41)
$$\varphi(n,X_{n-1},0) = P(X_n = 1|X_{n-1}) - p$$

Thus

(42)
$$\begin{aligned} \varphi(n,0,0) &= -\rho p \\ \varphi(n,1,0) &= \rho q \end{aligned}$$

Case (iii). $2 \le I \le n-1$

(43)
$$\begin{cases} \varphi(I,0,0) = \dfrac{-2\rho pq + \rho^2 p(q-p)}{q + \rho^2 p} \\ \varphi(I,1,1) = \dfrac{2\rho pq + \rho^2 q(q-p)}{p + \rho^2 q} \end{cases}$$

By Lemma 5.6, the correlation between X_{I+1} and X_{I-1} gives I for $2 \le I \le n-1$ is ρ^2. Therefore if $\rho^2 = 1$,

(44)
$$\varphi(I,0,1) = \varphi(I,1,0) = 0$$

and if $\rho^2 < 1$,

(45)
$$\varphi(I,0,1) = \varphi(I,1,0) = \frac{\rho(q-p)}{1+\rho}$$

where

$$q = 1 - p \tag{46}$$

Clearly, for $I = 1$ or n

$$E^I|\varphi(I, X_{I-1}, X_{I+1})| \le 2|\rho|pq \tag{47}$$

For $2 \le I \le n-1$, we first note that

$$\begin{cases} P(X_{I-1} = 0, X_{I+1} = 0|I) = q^2 + \rho^2 pq \\ P(X_{I-1} = 1, X_{I+1} = 1|I) = p^2 + \rho^2 pq \\ P(X_{I-1} = 0, X_{I+1} = 1|I) = P(X_{I-1} = 1, X_{I+1} = 0|I) \\ \qquad\qquad = (1 - \rho^2)pq \end{cases} \tag{48}$$

Then

$$\begin{aligned} &E^I|\varphi(I, X_{I-1}, X_{I+1})| \\ &= pq|-2\rho q + \rho^2(q-p)| + pq|2\rho p + \rho^2(q-p)| + 2|\varphi(I,0,1)|(1-\rho^2)pq \\ &\le 6|\rho|pq + 2|\rho|(1-\rho)\, pq|q-p| \le 8|\rho|pq \end{aligned} \tag{49}$$

where the last inequality follows from the natural constraint

$$-\min\left(\frac{p}{q}, \frac{q}{p}\right) \le \rho \le 1 \tag{50}$$

This completes the proof of the lemma.

Lemma 5.8.

$$\delta \leq (3\lambda)^{1/2} + 2^{1/2}\,\lambda p^{-1/2}|\rho| \tag{51}$$

Proof.

$$\begin{aligned} \delta^2 &= npq + 2pq \sum_{k=1}^{n-1} k\rho^{n-k} \\ &\leq 3\lambda + 2\lambda^2 p^{-1}\rho^2 \end{aligned} \tag{52}$$

Finally, in order to obtain (32), we apply Lemmas 5.7 and 5.8 to Theorems 4.1 and 4.2 with $m = 0$.

The most satisfactory results obtained in this section seem to be those of the m-dependent case. For example, in Corollary 5.1, the second bound depends on λ only through the probability of success and the correlations. Unfortunately, this has not been so with the case with exponentially decreasing dependence and the Markovian case. The author has not been able to improve the results of these two cases obtained so far. Although it is possible to obtain a bound in the Markovian case similar to those in Theorem 5.3 so that it will have a weaker overall dependence on λ than the second bound in Theorem 5.4, this bound will not reduce to Cp when $\rho = 0$.

§6. A randomly selected sum of Bernoulli random variables.

In this section we prove an approximation theorem for the distribution of a sum of random variables randomly selected from a square array of Bernoulli random variables such that no two of those selected came from the same row or column. If we let X_{ij}, $i, j = 1, 2, \ldots, n$ be the array with $P(X_{ij} = 1) = 1 - P(X_{ij} = 0) = p_{ij}$ and let π be a uniformly random permutation of $(1, 2, \ldots, n)$, independent of the X_{ij}, we can write the sum as $\sum_{i=1}^{n} X_{i\pi(i)}$, where $\pi(i)$ is the i^{th} coordinate of π. Intuitively, because the number of the selected random variables is small compared to those in the array, the distribution of $\sum_{i=1}^{n} X_{i\pi(i)}$ will still be approximately Poisson even when some of the p_{ij} are large, provided the number of large p_{ij} is small compared to n^2. In particular, the approximation is still good if the p_{ij} are either 0 or 1 and the number of those p_{ij} equal to 1 is small compared to n^2. In order to obtain a bound for the error in the approximation, which is consistent with this property, we have to operate with π and the X_{ij} at the same time. As a result of this and also due to the fact that the nature of the dependence in this problem is different from that in sums of weakly dependent Bernoulli random variables, the technique used in bounding the error here will be very different from that used in Sections 4 and 5. We now state and prove

Theorem 6.1. Let X_{ij}, $i, j = 1, 2, \ldots, n$ be mutually independent Bernoulli random variables with $P(X_{ij} = 1) = 1 - P(X_{ij} = 0) = p_{ij}$. Suppose π is a uniformly random permutation of $(1, 2, \ldots, n)$ and is independent of the X_{ij}'s. Then for $n \geq 2$, and every real-valued function h on $\{0, 1, 2, \ldots\}$ such that $|h| \leq 1$,

$$(1) \qquad |Eh(\sum_{i=1}^{n} X_{i\pi(i)}) - \mathcal{P}_\lambda h| \leq \begin{cases} \frac{A}{\lambda^{1/2} \vee 1} \left(\sum_{i=1}^{n} p_{i.}^2 + \sum_{j=1}^{n} p_{.j}^2 \right) \\ \frac{B}{\lambda} \left(\sum_{i=1}^{n} p_{i.}^2 + \sum_{j=1}^{n} p_{.j}^2 \right) \end{cases}$$

where A and B are absolute constants which can be taken to be 24 and 84 respectively, and

$$(2) \qquad \mathcal{P}_\lambda h = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} h(k)$$

$$(3) \qquad p_{i.} = \frac{1}{n} \sum_{j=1}^{n} p_{ij}, \qquad p_{.j} = \frac{1}{n} \sum_{i=1}^{n} p_{ij}$$

$$(4) \qquad \lambda = \sum_{i=1}^{n} p_{i.} = \sum_{j=1}^{n} p_{.j} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} p_{.j}$$

Proof. Let I, J, K, L, M be uniformly distributed on $\{1, 2, \ldots, n\}$ and $\pi, \tilde{\pi}, \tilde{\tilde{\pi}}$ be uniformly random permutations of $(1, 2, \ldots, n)$ such that

(5) $\{I, J, K, L, M, \pi, \tilde{\pi}, \tilde{\tilde{\pi}}\}$ is independent of $\{X_{ij}\}_{i,j=1}^{n}$

(6) $$P\{(I,K) = (i,k),\ i \neq k\} = \frac{1}{n(n-1)}$$

(7) $$P\{(L,M) = (\ell,m),\ \ell \neq m\} = \frac{1}{n(n-1)}$$

(8) J, (I,K), (L,M), $\tilde{\tilde{\pi}}$ are mutually independent

(9) J, (I,K), $\tilde{\pi}$ are mutually independent

(10) I, π are independent

(11)
$$\tilde{\pi}(\alpha) = \begin{cases} \tilde{\tilde{\pi}}(\alpha), & \alpha \neq I, K, \tilde{\tilde{\pi}}^{-1}(L), \tilde{\tilde{\pi}}^{-1}(M) \\ L, & \alpha = I \\ M, & \alpha = K \\ \tilde{\tilde{\pi}}(I), & \alpha = \tilde{\tilde{\pi}}^{-1}(L) \\ \tilde{\tilde{\pi}}(K), & \alpha = \tilde{\tilde{\pi}}^{-1}(M) \end{cases}$$

and

(12)
$$\pi(\alpha) = \begin{cases} \tilde{\pi}(\alpha), & \alpha \neq I, \tilde{\pi}^{-1}(J) \\ J, & \alpha = I \\ \tilde{\pi}(I), & \alpha = \tilde{\pi}^{-1}(J) \end{cases}$$

where $\tilde{\pi}(\alpha)$ etc. denote the α^{th} component of $\tilde{\pi}$ etc. In order to show that the definitions of I, J, K, L, M, π, $\tilde{\pi}$, $\tilde{\tilde{\pi}}$ and conditions (5)-(12) are consistent, it suffices to verify that $\tilde{\pi}$ defined by

(11) and π defined by (12) are uniformly random permutations of $(1, 2, \ldots, n)$ and that (9) and (10) hold. We shall first verify that $\tilde{\pi}$ is uniform. For every permutation $(w_1, w_2, \ldots, w_n)$ of $(1, 2, \ldots, n)$,

$$
\begin{aligned}
(13)\quad & P\{\tilde{\pi} = (w_1, w_2, \ldots, w_n)\} \\
&= \sum_{i=1}^{n} \sum_{\substack{k=1 \\ k\neq i}}^{n} P\{(I,K) = (i,k),\ \tilde{\pi} = (w_1, \ldots, w_n)\} \\
&= \sum_{i=1}^{n} \sum_{\substack{k=1 \\ k\neq i}}^{n} P\{(I,K) = (i,k),\ (L,M) = (w_i, w_k),\ \tilde{\pi} = (w_1, \ldots, w_n)\} \\
&= \sum_{i=1}^{n} \sum_{\substack{k=1 \\ k\neq i}}^{n} \sum_{\substack{r=1 \\ s\neq r}}^{n} \sum_{s=1}^{n} P\left\{\begin{array}{l}(I,K) = (i,k),\ (L,M) = (w_i, w_k), \\ \tilde{\pi} = (w_1, \ldots, w_n),\ \tilde{\pi}^{-1}(w_i) = r,\ \tilde{\pi}^{-1}(w_k) = s\end{array}\right\} \\
&= \sum_{i=1}^{n} \sum_{\substack{k=1 \\ k\neq i}}^{n} \sum_{\substack{r=1 \\ s\neq r}}^{n} \sum_{s=1}^{n} P\left\{\begin{array}{l}(I,K) = (i,k),\ (L,M) = (w_i, w_k), \\ \approx{\pi} = (\approx{w}_1, \ldots, \approx{w}_n),\quad \text{where}\end{array}\right\}
\end{aligned}
$$

$$
\approx{w}_\alpha = \left\{\begin{array}{ll} w_\alpha, & \alpha \neq i, k, r, s \\ w_r, & \alpha = i \\ w_s, & \alpha = k \\ w_i, & \alpha = r \\ w_k, & \alpha = s \end{array}\right\}
$$

which by (6), (7), (8) and that $\approx{\pi}$ is uniform,

$$= \sum_{i=1}^{n} \sum_{\substack{k=1 \\ k\neq i}}^{n} \sum_{r=1}^{n} \sum_{\substack{s=1 \\ s\neq r}}^{n} \frac{1}{n^2(n-1)^2} \frac{1}{n!} = \frac{1}{n!}$$

Next, (9) follows from arguments similar to the intermediate steps of (13). Similarly, π is uniform and is independent of I.

Now we let $(\Omega, \mathcal{B}, P)$ be the probability space on which all the random variables and the random vectors concerned are defined and let

(14)
$$\begin{cases}
\mathcal{F} = \mathcal{B}(\pi, X_{ij}, i, j = 1, 2, \ldots, n) \\
\mathcal{G} = \mathcal{Q}(I) \\
G = nX_{I\pi(I)} \\
\tilde{W} = \sum_{i=1}^{n} X_{i\tilde{\pi}(i)} \\
\tilde{\tilde{W}} = \sum_{i=1}^{n} X_{i\tilde{\tilde{\pi}}(i)} \\
W^* = \sum_{i=1, i\neq I}^{n} X_{i\pi(i)} \\
\tilde{W}^* = \sum_{i=1, i\neq I}^{n} X_{i\tilde{\pi}(i)} \\
\tilde{W}^{**} = \sum_{i=1, i\neq I, K}^{n} X_{i\tilde{\pi}(i)} \\
\tilde{\tilde{W}}^{**} = \sum_{i=1, i\neq I, K}^{n} X_{i\tilde{\tilde{\pi}}(i)} \\
\lambda^* = \sum_{i=1, i\neq I}^{n} p_i . \\
\lambda^{**} = \sum_{i=1, i\neq I, K}^{n} p_i .
\end{cases}$$

Then

$$W = \sum_{i=1}^{n} X_{i\pi(i)} \tag{15}$$

$$E^{\mathcal{P}}G = nE^{I}X_{I\pi(I)} = np_{I.} \tag{16}$$

and from (1.9) we obtain

$$\begin{aligned} (17)\quad & Eh(W) \\ &= \mathcal{P}_\lambda h - nEX_{I\pi(I)}(f(W) - f(W^*+1)) - nE(X_{I\pi(I)} - p_{I.})\, f(W^*+1) \\ &\qquad + nEp_{I.}(f(W+1) - f(W^*+1)) \\ &= \mathcal{P}_\lambda h - nE(X_{I\pi(I)} - p_{I.})\, f(W^*+1) + nEp_{I.}X_{I\pi(I)}(f(W^*+2) - f(W^*+1)) \\ &= \mathcal{P}_\lambda h - nE(X_{I\pi(I)} - p_{I.})\, S_\lambda h(W^*+1) + nEp_{I.}X_{I\pi(I)}\, U_\lambda h(W^*) \end{aligned}$$

where the second equality follows from the same arguments as in §4 and S_λ and U_λ are defined by (3.6) and (3.7) respectively.

Before proceeding to bound the error terms in (17), we shall first prove three lemmas.

Lemma 6.1. Let $\hat{\pi}$ be a uniformly random permutation of $(1, 2, \ldots, n)$ independent of the X_{ij}'s. Then

$$\delta \leq \sqrt{2\lambda} \tag{18}$$

where

$$(19) \qquad \delta = \Big(\sum_{i=1}^{n} \sum_{j=1}^{n} |E(X_{i\hat{\pi}(i)} - p_{i.})(X_{j\hat{\pi}(j)} - p_{j.})| \Big)^{1/2}$$

Proof of Lemma 6.1.

$$(20) \qquad \delta^2 \leq \lambda + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} |E(X_{i\hat{\pi}(i)} - p_{i.})(X_{j\hat{\pi}(j)} - p_{j.})|$$

Now by the conditional independence of $X_{i\hat{\pi}(i)}$ and $X_{j\hat{\pi}(j)}$ given $\hat{\pi}$, where $j \neq i$,

$$(21) \qquad E(X_{i\hat{\pi}(i)} - p_{i.})(X_{j\hat{\pi}(j)} - p_{j.})$$

$$= EE^{\hat{\pi}}(X_{i\hat{\pi}(i)} - p_{i.})(X_{j\hat{\pi}(j)} - p_{j.})$$

$$= E[E^{\hat{\pi}}(X_{i\hat{\pi}(i)} - p_{i.})][E^{\hat{\pi}}(X_{j\hat{\pi}(j)} - p_{j.})]$$

$$= E(p_{i\hat{\pi}(i)} - p_{i.})(p_{j\hat{\pi}(j)} - p_{j.})$$

$$= E(p_{i\hat{\pi}(i)} - p_{i.})\, E^{\hat{\pi}(i)}(p_{j\hat{\pi}(j)} - p_{j.})$$

$$= E(p_{i\hat{\pi}(i)} - p_{i.}) \left(\frac{np_{j.} - p_{j\hat{\pi}(i)}}{n - 1} - p_{j.} \right)$$

$$= \frac{1}{n-1} (p_{i.}p_{j.} - Ep_{i\hat{\pi}(i)}\, p_{j\hat{\pi}(i)})$$

Thus, for $j \neq i$,

$$(22) \qquad |E(X_{i\hat{\pi}(i)} - p_{i.})(X_{j\hat{\pi}(j)} - p_{j.})|$$

$$\leq \frac{1}{n-1} \max(p_{i.}, Ep_{i\hat{\pi}(i)})$$

$$= \frac{1}{n-1} p_{i.}$$

and (18) follows from (20) and (22).

Lemma 6.2.

$$(23) \qquad E|X_{IJ} - p_{I.}| X_{KL} 1_{\{J=M\}} \leq \frac{3\lambda^2}{n^2(n-1)}$$

Proof of Lemma 6.2.

$$(24) \qquad E|X_{IJ} - p_{I.}| X_{KL} 1_{\{J=M\}}$$

$$\leq EX_{IJ}X_{KL} 1_{\{J=M\}} + Ep_{I.}X_{KL} 1_{\{J=M\}}$$

$$= Ep_{IJ}p_{KL} 1_{\{J=M\}} + Ep_{I.}p_{KL} 1_{\{J=M\}}$$

$$= \frac{1}{n} Ep_{IM}p_{KL} + \frac{1}{n} Ep_{I.}p_{KL}$$

$$\leq \frac{3\lambda^2}{n^2(n-1)}$$

Lemma 6.3.

$$E|X_{IJ} - p_{I.}|\, X_{KJ}\, 1_{\{J=M\}} \le \frac{1}{n(n-1)} \sum_{j=1}^{n} p_{.j}^2 + \frac{\lambda^2}{n^2(n-1)} \tag{25}$$

Proof of Lemma 6.3.

$$\begin{aligned} & E|X_{IJ} - p_{I.}|\, X_{KJ}\, 1_{\{J=M\}} \qquad (26) \\ & \le EX_{IJ}X_{KJ}\, 1_{\{J=M\}} + Ep_{I.}X_{KJ}\, 1_{\{J=M\}} \\ & = Ep_{IJ}p_{KJ}\, 1_{\{J=M\}} + Ep_{I.}p_{KJ}\, 1_{\{J=M\}} \\ & = \frac{1}{n} Ep_{IM}p_{KM} + \frac{1}{n} Ep_{I.}p_{KM} \\ & \le \frac{1}{n(n-1)} \sum_{j=1}^{n} p_{.j}^2 + \frac{\lambda^2}{n^2(n-1)} \end{aligned}$$

We now bound the error terms in (17). First the second error term. By (3.14) of Proposition 3.3,

$$|nEp_{I.}X_{I\pi(I)}U_\lambda h(W^*)| \le \frac{8}{\lambda^{1/2} \vee 1} \sum_{i=1}^{n} p_{i.}^2 , \tag{27}$$

and by (3.15) of Proposition 3.3,

$$
\begin{aligned}
(28)\qquad & |nEp_{I.}X_{I\pi(I)}U_\lambda h(W^*)| \\
& \le \frac{8}{\lambda}\sum_{i=1}^{n} p_{i.}^2 + \frac{6n}{\lambda(\lambda^{1/2}\vee 1)}\, Ep_{I.}X_{IJ}|W^*-\lambda^*| \\
& \le \frac{8}{\lambda}\sum_{i=1}^{n} p_{i.}^2 + \frac{6n}{\lambda(\lambda^{1/2}\vee 1)}\, Ep_{I.}X_{IJ}\{|\tilde{W}^*-\lambda^*| + \big||W^*-\lambda^*|-|\tilde{W}^*-\lambda^*|\big|\} \\
& \le \frac{14}{\lambda}\sum_{i=1}^{n} p_{i.}^2 + \frac{6n}{\lambda(\lambda^{1/2}\vee 1)}\, E(E^I p_{I.}X_{IJ})(E^I|\tilde{W}^*-\lambda^*|) \\
& \le \frac{14}{\lambda}\sum_{i=1}^{n} p_{i.}^2 + \frac{6n}{\lambda(\lambda^{1/2}\vee 1)}\, Ep_{I.}^2\sqrt{E^I(\tilde{W}^*-\lambda^*)^2} \\
& \le \frac{14}{\lambda}\sum_{i=1}^{n} p_{i.}^2 + \frac{6\delta}{\lambda(\lambda^{1/2}\vee 1)}\sum_{i=1}^{n} p_{i.}^2 \\
& \le \frac{23}{\lambda}\sum_{i=1}^{n} p_{i.}^2
\end{aligned}
$$

where, in the third inequality, the conditional independence of X_{IJ} and $\tilde{W}^*$ given I is used and the last inequality follows from Lemma 6.1. Next, the first error term

$$
\begin{aligned}
(29)\qquad & nE(X_{I\pi(I)} - p_{I.})\, S_\lambda h(W^*+1) \\
& = nE(X_{IJ} - p_{I.})\, S_\lambda h(\tilde{W}^*+1) + nE(X_{IJ}-p_{I.})(S_\lambda h(W^*+1)-S_\lambda h(\tilde{W}^*+1)) \\
& = nE(X_{IJ} - p_{I.})(S_\lambda h(W^*+1) - S_\lambda h(\tilde{W}^*+1))
\end{aligned}
$$

where again the last equality follows from the conditional independence of X_{IJ} and $\tilde{W}^*$ given I. Now

(30) $$W^* - \tilde{W}^* = X_{\tilde{\pi}^{-1}(J),\tilde{\pi}(I)} - X_{\tilde{\pi}^{-1}(J),J}$$

Thus

(31) $$nE(X_{IJ} - p_{I.})(S_\lambda h(W^* + 1) - S_\lambda h(\tilde{W}^* + 1))$$

$$= nE(X_{IJ} - p_{I.})\, X_{\tilde{\pi}^{-1}(J),\tilde{\pi}(I)}(1 - X_{\tilde{\pi}^{-1}(J),J})\, U_\lambda h\Big(\sum_{\substack{\alpha\neq I \\ \alpha\neq\tilde{\pi}^{-1}(J)}} X_{\alpha\tilde{\pi}(\alpha)}\Big)$$

$$-nE(X_{IJ} - p_{I.})\, X_{\tilde{\pi}^{-1}(J),J}(1 - X_{\tilde{\pi}^{-1}(J),\tilde{\pi}(I)})\, U_\lambda h\Big(\sum_{\substack{\alpha\neq I \\ \alpha\neq\tilde{\pi}^{-1}(J)}} X_{\alpha\tilde{\pi}(\alpha)}\Big)$$

$$= R_1 + R_2 \quad \text{say}$$

Now, by (3.15) of Proposition 3.3, and noting that $0 \le p_{I.} + p_{\tilde{\pi}^{-1}(J).} \le 2$ implies $|1 - p_{I.} - p_{\tilde{\pi}^{-1}(J).}| \le 1$, we have

(32) $$|R_1| \le nE|X_{IJ} - p_{I.}|\, X_{\tilde{\pi}^{-1}(J),\tilde{\pi}(I)}(1 - X_{\tilde{\pi}^{-1}(J),J})$$

$$\cdot \left\{\frac{8}{\lambda} + \frac{6}{\lambda(\lambda^{1/2} \vee 1)} \left|\sum_{\alpha\neq I,\tilde{\pi}^{-1}(J)} (X_{\alpha\tilde{\pi}(\alpha)} - p_{\alpha.})\right|\right\}$$

$$= \sum_{i=1}^{n} \sum_{\substack{k=1 \\ k\neq i}}^{n} E|X_{iJ} - p_{i.}|\, X_{k,\tilde{\pi}(i)}(1 - X_{kJ})\, 1_{\{\tilde{\pi}(k)=J\}}$$

$$\cdot \left\{\frac{8}{\lambda} + \frac{6}{\lambda(\lambda^{1/2} \vee 1)} \left|\sum_{\alpha\neq i,k} (X_{\alpha\tilde{\pi}(\alpha)} - p_{\alpha.})\right|\right\}$$

$$= n(n-1)\, E|X_{IJ} - p_{I.}|\, X_{K,\tilde{\pi}(I)}\,(1 - X_{KJ})\, 1_{\{\tilde{\pi}(K)=J\}}$$

$$\cdot \left\{ \frac{8}{\lambda} + \frac{6}{\lambda(\lambda^{1/2} \vee 1)} |\tilde{W}^{**} - \lambda^{**}| \right\}$$

$$\leq n(n-1)\, E|X_{IJ} - p_{I.}|\, X_{KL}\, 1_{\{J=M\}}$$

$$\cdot \left\{ \frac{20}{\lambda} + \frac{6}{\lambda(\lambda^{1/2} \vee 1)} |\tilde{\tilde{W}}^{**} - \lambda^{**}| \right\}$$

which by Lemma 6.2 and the conditional independence of $|X_{IJ} - p_{I.}|\, X_{KL}\, 1_{\{J=M\}}$ and $\tilde{\tilde{W}}^{**}$ given (I,K)

$$\leq \frac{60\lambda}{n} + \frac{6n(n-1)}{\lambda(\lambda^{1/2} \vee 1)} E\Big\{ [E^{I,K}|X_{IJ} - p_{I.}|\, X_{KL}\, 1_{\{J=M\}}]$$

$$\cdot [E^{I,K}|\tilde{\tilde{W}}^{**} - \lambda^{**}|] \Big\}$$

$$\leq \frac{60\lambda}{n} + \frac{6n(n-1)\delta}{\lambda(\lambda^{1/2} \vee 1)} E|X_{IJ} - p_{I.}|\, X_{KL}\, 1_{\{J=M\}}$$

which by Lemmas 6.1 and 6.2

$$\leq \frac{87\lambda}{n}$$

and by (3.14) of Proposition 3.3,

$$|R_1| \leq \frac{8n(n-1)}{\lambda^{1/2} \vee 1} E|X_{IJ} - p_{I.}|\, X_{KL}\, 1_{\{J=M\}} \tag{33}$$

which by Lemma 6.2

$$\leq \frac{24\lambda^2}{(\lambda^{1/2} \vee 1)n}$$

Similarly, by (3.15) of Proposition 3.3, Lemmas 6.1 and 6.3,

$$(34) \quad |R_2| \leq nE|X_{IJ} - p_{I.}| \, X_{\tilde{\pi}^{-1}(J),J} \left(1 - X_{\tilde{\pi}^{-1}(J),\tilde{\pi}(I)}\right) \cdot \left\{\frac{8}{\lambda} + \frac{6}{\lambda(\lambda^{1/2} \vee 1)} \left| \sum_{\alpha \neq I, \tilde{\pi}^{-1}(J)} (X_{\alpha\tilde{\pi}(\alpha)} - p_{\alpha.}) \right| \right\}$$

$$= n(n-1) \, E|X_{IJ} - p_{I.}| \, X_{KJ}(1 - X_{KL}) \, 1_{\{J=M\}} \cdot \left\{\frac{8}{\lambda} + \frac{6}{\lambda(\lambda^{1/2} \vee 1)} |\tilde{W}^{**} - \lambda^{**}|\right\}$$

$$\leq n(n-1) \, E|X_{IJ} - p_{I.}| \, X_{KJ} \, 1_{\{J=M\}} \cdot \left\{ \frac{20}{\lambda} + \frac{6}{\lambda(\lambda^{1/2} \vee 1)} |\tilde{\tilde{W}}^{**} - \lambda^{**}|\right\}$$

$$\leq n(n-1) \, E|X_{IJ} - p_{I.}| \, X_{KJ} \, 1_{\{J=M\}} \cdot \left\{\frac{20}{\lambda} + \frac{68}{\lambda(\lambda^{1/2} \vee 1)} \right\}$$

$$\leq \frac{29}{\lambda} \sum_{j=1}^{n} p_{.j}^2 + \frac{29\lambda}{n} ,$$

and by (3.14) of Propositon 3.3 and Lemma 6.3,

$$(35)\qquad |R_2| \le \frac{8n(n-1)}{\lambda^{1/2} \vee 1} E|X_{IJ} - p_{I\cdot}|\, X_{KJ}\, 1_{\{J=M\}}$$

$$\le \frac{8}{\lambda^{1/2} \vee 1} \sum_{j=1}^{n} p_{\cdot j}^2 + \frac{8\lambda^2}{(\lambda^{1/2} \vee 1)n}$$

Finally, using

$$(36)\qquad \lambda^2 \le n \sum_{i=1}^{n} p_{i\cdot}^2, \qquad \lambda^2 \le n \sum_{j=1}^{n} p_{\cdot j}^2$$

we obtain (1) from (27), (28), (29), (31), (32), (33), (34) and (35). Hence the theorem.

In the case where p_{ij}'s are either 0 or 1, we have the following

Corollary 6.1. Let $\{a_{ij}\}_{i,j=1}^{n}$ be a square array of 0's and 1's and π be a uniformly random permutation of $(1, 2, \ldots, n)$. Then for $n \ge 2$ and $|h| \le 1$,

$$(37)\qquad |Eh(\sum_{i=1}^{n} a_{i\pi(i)}) - \mathcal{P}_\lambda h| \le \begin{cases} \frac{A}{\lambda^{1/2} \vee 1} (\sum_{i=1}^{n} p_i^2 + \sum_{j=1}^{n} q_j^2) \\ \frac{B}{\lambda} (\sum_{i=1}^{n} p_i^2 + \sum_{j=1}^{n} q_j^2) \end{cases}$$

where A and B are absolute constants, which can be taken to be 24 and 84 respectively, and

(38) p_i = proportion of 1's in i^{th} row

(39) q_j = proportion of 1's in j^{th} row

(40) $$\lambda = \sum_{i=1}^{n} p_i = \sum_{j=1}^{n} q_i = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}$$

It is well-known that, under certain conditions, the hypergeometric distribution is approximately Poisson. By choosing the a_{ij} in Corollary 6.1 appropriately, a bound can be obtained for the error in the approximation. To accomplish this, we let ξ_i, $i = 1, 2, \ldots, n$ and η_j, $j = 1, 2, \ldots,$ be 0's and 1's and let $a_{ij} = \xi_i \eta_j$. Suppose there are a 1's among the ξ_i and b 1's among the η_j with $a \geq b$. Then for $k = 0, 1, 2, \ldots, b$,

(41) $$P\left(\sum_{i=1}^{n} \xi_i \eta_{\pi(i)} = k \right) = \frac{\binom{a}{k} \binom{n-a}{b-k}}{\binom{n}{b}}$$

(42) $$p_{i.} = \frac{1}{n} \sum_{j=1}^{n} \xi_i \eta_j = \begin{cases} 0 & \text{if } \xi_i = 0 \\ \frac{b}{n} & \text{if } \xi_i = 1 \end{cases}$$

where A and B are absolute constants, which can be taken to be 24 and 84 respectively, and

(38) p_i = proportion of 1's in i^{th} row

(39) q_j = proportion of 1's in j^{th} row

(40) $$\lambda = \sum_{i=1}^{n} p_i = \sum_{j=1}^{n} q_i = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}$$

It is well-known that, under certain conditions, the hypergeometric distribution is approximately Poisson. By choosing the a_{ij} in Corollary 6.1 appropriately, a bound can be obtained for the error in the approximation. To accomplish this, we let ξ_i, $i = 1, 2, \ldots, n$ and η_j, $j = 1, 2, \ldots,$ be 0's and 1's and let $a_{ij} = \xi_i \eta_j$. Suppose there are a 1's among the ξ_i and b 1's among the η_j with $a \geq b$. Then for $k = 0, 1, 2, \ldots, b$,

(41) $$P\left(\sum_{i=1}^{n} \xi_i \eta_{\pi(i)} = k\right) = \frac{\binom{a}{k}\binom{n-a}{b-k}}{\binom{n}{b}}$$

(42) $$p_{i\cdot} = \frac{1}{n} \sum_{j=1}^{n} \xi_i \eta_j = \begin{cases} 0 & \text{if } \xi_i = 0 \\ \frac{b}{n} & \text{if } \xi_i = 1 \end{cases}$$

$$(43)\qquad p_{\cdot j} = \frac{1}{n}\sum_{i=1}^{n}\xi_i\eta_j = \begin{cases} 0 & \text{if } \eta_j = 0 \\ \frac{a}{n} & \text{if } \eta_j = 1 \end{cases}$$

$$(44)\qquad \lambda = \frac{ab}{n}$$

and we have

Corollary 6.2.

$$(45)\qquad \sum_{k=0}^{b}\left|\frac{\binom{a}{k}\binom{n-a}{b-k}}{\binom{n}{b}} - e^{-\lambda}\frac{\lambda^k}{k!}\right| \le \frac{C(a+b)}{n}$$

where C can be taken to be 84.

§7. Remarks.

It should be mentioned that the author has made no special attempt to minimize the absolute constants in the bounds obtained so far, and that the constants can be reduced by making some specific choices of h. Observe that $\sup_{|h|\le 1}|Eh(W) - \mathcal{P}_\lambda h| = \sum_{k=0}^{\infty}|P(W=k) - e^{-\lambda}\frac{\lambda^k}{k!}|$ Thus the reduction of the constants can be achieved by choosing h to be h_k for $k = 0, 1, 2, \ldots$ where

$$h_k(w) = \begin{cases} 1 & \text{if } w = k \\ 0 & \text{if } w \neq k\,. \end{cases}$$

By applying Cauchy-Schwartz's inequality to $E\sum_{i=I-m}^{I+m} p_I p_i$ in the proofs of Theorems 4.1 and 4.2 in §4, it is possible to improve slightly the bounds obtained in the theorems. Consequently, we can replace the term $\max_{1\le i\le n} p_i$ in the bounds in the m-dependent case studied in §5 by $\frac{1}{\lambda}\sum_{i=1}^{n} p_i^2$, which is indeed an improvement.

CHAPTER III

ITERATION OF POISSON APPROXIMATION

§1. Preliminary results

By iterating the basic identity, one can obtain the error of second order of magnitude in the approximation. In this section, we shall prove a few lemmas and propositions which will be used in bounding the error in two special cases in the next section. All notation will be the same as in §3 of Chapter II. We also define a new operator L by

$$Lf(w) = f(w+1) \tag{1}$$

Lemma 1.1. For $\lambda \geq \nu > 0$ and $k = 0,1,2,\ldots$

$$|e^{-\lambda}\lambda^k - e^{-\nu}\nu^k| \leq (\lambda-\nu)(e^{-\lambda}k\lambda^{k-1} + e^{-\nu}\nu^k) \tag{2}$$

Proof. We first obtain

$$\begin{aligned} |e^{-\lambda}\lambda^k - e^{-\nu}\nu^k| &\leq |e^{-\lambda}\lambda^k - e^{-\lambda}\nu^k| + |e^{-\lambda}\nu^k - e^{-\nu}\nu^k| \\ &= e^{-\lambda}\lambda^k|1 - (1 - \tfrac{\lambda-\nu}{\lambda})^k| + e^{-\nu}\nu^k|1 - e^{-\lambda+\nu}| \end{aligned} \tag{3}$$

Then (2) follows from (3) using

$$1 - kx \leq (1-x)^k \qquad \text{for } 0 \leq x < 1,\ k \geq 1 \tag{4}$$

and

$$1-e^{-x} \leq x \tag{5}$$

Lemma 1.1. immediately implies

Proposition 1.1. For $\lambda \geq \mu > 0$,

$$|(\mathcal{P}_\lambda - \mathcal{P}_\nu)f| \leq (\lambda-\mu)[\mathcal{P}_\lambda|Lf| + \mathcal{P}_\nu|f|] \tag{6}$$

Proposition 1.2. For $m = 0,1,2,\ldots$

$$\mathcal{P}_\nu|L^m U_\lambda h| \leq \frac{8}{\lambda^{1/2} \vee 1}\,\|h\| \tag{7}$$

$$\mathcal{P}_\nu|L^m U_\lambda h| \leq \left[\frac{2 + 6|m+1+\nu-\lambda|}{\lambda} + \frac{6\nu^{1/2}}{\lambda(\lambda^{1/2} \vee 1)}\right]\|h\| \tag{8}$$

Proof. (7) follows immediately from (3.14) of Proposition II 3.3. Let Y be Poisson distributed with parameter ν. Then by (3.15) of the same proposition,

$$\mathcal{P}_\nu|L^m U_\lambda h| = E|U_\lambda h(Y+m)| \leq E\left[\frac{2}{\lambda} + \frac{6}{\lambda(\lambda^{1/2} \vee 1)}\,|Y+m+1-\lambda|\right]\|h\| \tag{9}$$

This together with the following imply (8)

$$E|Y+m+1-\lambda| \leq |m+1+\nu-\lambda| + E|Y-\nu| \leq |m+1+\nu-\lambda| + \sqrt{E(Y-\nu)^2} \tag{10}$$

$$= |m+1+\nu-\lambda| + \nu^{1/2}$$

For brevity, we shall from now on omit proofs which are trivial.

Proposition 1.3. For $m = 0,1,2,\ldots$

$$|S_\nu L^m S_\lambda h| \le \frac{36}{(\lambda^{1/2} \vee 1)(\nu^{1/2} \vee 1)} \|h\| \tag{11}$$

Proposition 1.4. For $m = 0,1,2,\ldots$

$$|U_\nu L^m S_\lambda h| \le \frac{48}{(\lambda^{1/2} \vee 1)(\nu^{1/2} \vee 1)} \|h\| \tag{12}$$

$$|U_\nu L^m S_\lambda h(w)| \le \left[\frac{12}{\nu(\lambda^{1/2} \vee 1)} + \frac{36}{\nu(\nu^{1/2} \vee 1)(\lambda^{1/2} \vee 1)} |w+1-\nu|\right] \|h\| \tag{13}$$

Proof. We have

$$U_\nu L^m S_\lambda h(w) = \frac{1}{\nu} [L^m S_\lambda h(w+1) - \varphi_\nu L^m S_\lambda h] + \frac{1}{\nu} (w+1-\nu) - S_\nu L^m S_\lambda h(w+1) \tag{14}$$

This together with Proposition 1.3 and Proposition II 3.2 imply (13).

Lemma 1.2. For $0 < \nu \le w$, $w \ge 1$,

$$\frac{(w-1)!}{\nu^w} \sum_{k=w}^{\infty} \frac{\nu^k}{k!} |k-\nu| = 1 \tag{15}$$

Lemma 1.3. For $1 \le w \le \nu$

$$\frac{(w-1)!}{\nu^w} \sum_{k=0}^{w-1} \frac{\nu^k}{k!} |k-\nu| = 1 \tag{16}$$

Proposition 1.5.

$$|S_\nu U_\lambda h| \le \frac{48}{(\lambda^{1/2} \vee 1)(\nu^{1/2} \vee 1)} \|h\| \tag{17}$$

$$|S_\nu U_\lambda h(w)| \le \left[\frac{12 + 36|1+\nu-\lambda|}{\lambda(\nu^{1/2} \vee 1)} + \frac{24}{\lambda(\lambda^{1/2} \vee 1)} \right] \|h\| \tag{18}$$

Proof. By the definition of S_ν, we have

$$\begin{aligned} S_\nu U_\lambda h(w) &= \frac{(w-1)!}{\nu^w} \sum_{k=0}^{w-1} \frac{\nu^k}{k!} [U_\lambda h(k) - \mathcal{P}_\nu U_\lambda h] \\ &= -\frac{(w-1)!}{\nu^w} \sum_{k=w}^{\infty} \frac{\nu^k}{k!} [U_\lambda h(k) - \mathcal{P}_\nu U_\lambda h] \end{aligned} \tag{19}$$

for $w = 1,2,\ldots$

If $0 < \nu \le w$, the second equality in (19) yields

$$|S_\nu U_\lambda h(w)| \le \frac{(w-1)!}{\nu^w} \sum_{k=w}^{\infty} \frac{\nu^k}{k!} |U_\lambda h(k)| + \frac{(w-1)!}{\nu^w} \sum_{k=w}^{\infty} \frac{\nu^k}{k!} |\mathcal{P}_\nu U_\lambda h| \tag{20}$$

which together with Lemma II 3.1, Proposition II 3.3, Proposition 1.2 and Lemma 1.2,

$$\leq \frac{(w-1)!}{\nu^w} \sum_{k=w}^{\infty} \frac{\nu^k}{k!} \left[\frac{2}{\lambda} + \frac{6}{\lambda(\lambda^{1/2} \vee 1)} |k+1-\lambda|\right] \|h\| + \frac{3}{(\nu^{1/2} \vee 1)} |\mathcal{P}_\nu U_\lambda h|$$

$$\leq \frac{6 + 18|1+\nu-\lambda|}{\lambda(\nu^{1/2} \vee 1)} \|h\| + \frac{6\|h\|}{\lambda(\lambda^{1/2} \vee 1)} \frac{(w-1)!}{\nu^w} \sum_{k=w}^{\infty} \frac{\nu^k}{k!} |k-\nu|$$

$$+ \frac{6 + 18|1+\nu-\lambda|}{\lambda(\nu^{1/2} \vee 1)} \|h\| + \frac{18}{\lambda(\lambda^{1/2} \vee 1)} \|h\|$$

which implies (18).

On the other hand, if $1 \leq w \leq \nu$, the first equality in (19) yields

$$(21) \quad |S_\nu U_\lambda h(w)| \leq \frac{(w-1)!}{\nu^w} \sum_{k=0}^{w-1} \frac{\nu^k}{k!} |U_\lambda h(k)| + \frac{(w-1)!}{\nu^w} \sum_{k=0}^{w-1} \frac{\nu^k}{k!} |\mathcal{P}_\nu U_\lambda h|$$

which together with Lemma II 3.2, Proposition II 3.3, Proposition 1.2 and Lemma 1.3 imply (18).

Proposition 1.6.

$$(22) \quad |U_\nu U_\lambda h| \leq \frac{64}{(\lambda^{1/2} \vee 1)(\nu^{1/2} \vee 1)} \|h\|$$

$$(23) \quad |U_\nu U_\lambda h(w)| \leq \left[\frac{4 + 6|1-\nu-\lambda|}{\lambda\nu} + \frac{6}{\lambda(\lambda^{1/2} \vee 1)(\nu^{1/2} \vee 1)}\right] \|h\|$$

$$+ \frac{6}{\lambda\nu(\lambda^{1/2} \vee 1)} |w+2-\lambda| \|h\| + \left[\frac{12+36|1+\nu-\lambda|}{\lambda\nu(\nu^{1/2} \vee 1)} + \frac{24}{\lambda\nu(\lambda^{1/2} \vee 1)}\right]$$

$$\times |w+1-\mu| \|h\|$$

Proof. First we have

$$(24)\qquad U_\nu U_\lambda h(w) = \frac{1}{\nu}\,[U_\lambda h(w+1) - \mathcal{P}_\nu U_\lambda h] + \frac{1}{\nu}\,(w+1-\nu)\,S_\nu U_\lambda h(w+1)$$

This together with Proposition II 3.3, Propositions 1.2 and 1.5 imply (23).

§2. Main Theorems

While it is interesting to see how the idea of iteration can be carried out for a sum of dependent Bernoulli random variables, it is also interesting to see what bound for the error of second order of magnitude one can obtain in the independent but non-identical case. Thus, in this section, we shall consider both the independent but non-identical case and the Markovian case.

Theorem 2.1. Let $X_1, X_2, \ldots, X_n$ be mutually independent Bernoulli random variables with $P(X_i = 1) = 1 - P(X_i = 0) = p_i$. Then for every h defined on $\{0,1,2,\ldots\}$ such that $|h| \le 1$, we have for $n \ge 2$,

$$(1)\qquad \Big|Eh\Big(\sum_{i=1}^{n} X_i\Big) - \mathcal{P}_\lambda h - \Big(\sum_{i=1}^{n} p_i^2\Big)\mathcal{P}_\lambda U_\lambda h\Big| \le \frac{A}{\lambda}\sum_{i=1}^{n} p_i^3 + \frac{B}{\lambda}\sum_{i=1}^{n}\frac{p_i^2}{\lambda^{(i)}}\sum_{\substack{j=1\\ j\neq i}}^{n} p_j^2$$

where A and B are absolute constants which are not greater than 34 and 178 respectively, and

$$\lambda = \sum_{i=1}^{n} p_i \tag{2}$$

$$\lambda^{(i)} = \lambda - p_i \tag{3}$$

$\mathcal{P}_\lambda$ is defined by II (1.7) and U_λ by II (3.7). Note that when all p_i's are equal, say to p, the bound in (1) reduces to Cp^2 where $C = 212$.

Proof. By putting $m = 0$ and using independence, II (4.13) yields

$$Eh(W) = \mathcal{P}_\lambda h + nEp_I^2 U_\lambda h(W^*) = \mathcal{P}_\lambda h + nEp_I^2 E^I U_\lambda h(W^*) \tag{4}$$

where, as before, I is uniform on $\{1,2,\ldots,n\}$ and is independent of the X_i's, and

$$W = \sum_{i=1}^{n} X_i \tag{5}$$

$$W^* = \sum_{\substack{i=1 \\ i \neq I}}^{n} X_i \tag{6}$$

Now applying (4) again to $E^I U_\lambda h(W^*)$, we obtain

$$(7)\quad Eh(W) = \mathcal{P}_\lambda h + nEp_I^2\{\mathcal{P}_{\lambda*}U_\lambda h + (n-1)\,E^I p_J^2 U_{\lambda*}U_\lambda h(W^{**})\}$$

$$= \mathcal{P}_\lambda h + \Big(\sum_{i=1}^{n} p_i^2\Big)\,\mathcal{P}_\lambda U_\lambda h + nEp_I^2(\mathcal{P}_{\lambda*}U_\lambda h - \mathcal{P}_\lambda U_\lambda h)$$

$$+ n(n-1)\,Ep_I^2 p_J^2 U_{\lambda*}U_\lambda h(W^{**})$$

where (I,J) is independent of the X_i's,

$$(8)\quad P\{(I,J) = (i,j),\ i \neq j,\ i,\ j = 1,2,\ldots,n\} = \frac{1}{n(n-1)}$$

$$(9)\quad \lambda^* = \sum_{\substack{i=1\\ i\neq I}}^{n} p_i$$

and

$$(10)\quad W^{**} = \sum_{\substack{i=1\\ i\neq I,J}}^{n} X_i$$

By Proposition 1.1 and 1.2,

$$(11)\quad |\mathcal{P}_{\lambda*}U_\lambda h - \mathcal{P}_\lambda U_\lambda h| \le p_I[\mathcal{P}_\lambda |LU_\lambda h| + \mathcal{P}_{\lambda*}|U_\lambda h|] \le \frac{34}{\lambda}\,p_I$$

By Proposition (1.6),

$$(12)\quad U_{\lambda*}U_\lambda h(W^{**}) \le \frac{100}{\lambda\lambda^*} + \frac{78}{\lambda\lambda^{*3/2}}\,|W^{**} - \lambda^{**}|$$

where

$$(13)\quad \lambda^{**} = \sum_{\substack{i=1\\ i\neq I,J}}^{n} p_i$$

Thus

$$(14)\quad |n(n-1)\ Ep_I^2 p_J^2 U_{\lambda^*} U_\lambda h(W^{**})|$$

$$= |n(n-1)\ Ep_I^2 p_J^2 E^{I,J} U_{\lambda^*} U_\lambda h(W^{**})|$$

$$\le n(n-1)\ Ep_I^2 p_J^2 E^{I,J} \left[\frac{100}{\lambda\lambda^*} + \frac{78}{\lambda\lambda^{*3/2}} |W^{**} - \lambda^{**}|\right]$$

$$\le n(n-1)\ Ep_I^2 p_J^2 \left[\frac{100}{\lambda\lambda^*} + \frac{78 \sqrt{E^{I,J}(W^{**} - \lambda^{**})^2}}{\lambda\lambda^{*3/2}}\right]$$

$$\le 178n(n-1)\ E \frac{1}{\lambda\lambda^*} p_I^2 p_J^2$$

This together with (7) and (11) imply (1).

Theorem 2.2. Let $X_1, X_2, \ldots, X_n$ be a stationary and Markovian sequence of Bernoulli random variables with $P(X_i = 1) = 1 - p(X_i = 0) = p$ and $\mathrm{corr}(X_i, X_{i+1}) = \rho$ such that

$$(15)\qquad (n-3)p \ge 1$$

$$(16)\qquad |\rho| \le \frac{1}{2}$$

Then for every h such that $|h| \le 1$, we have for $n \ge 4$,

$$(17)\quad |Eh(\sum_{i=1}^{n} X_i) - \mathcal{P}_\lambda h - \lambda(p-2\rho)\mathcal{P}_\lambda U_\lambda h| \le C(p^2 + \lambda^{1/2}|\rho|p + \lambda\rho^2)$$

where C is an absolute constant,

(18) $$\lambda = np$$

and $\mathcal{P}_\lambda$ is defined by II (1.7) and U_λ by II (3.7).

Proof. Throughout the proof, C will denote an absolute constant but may have different values at different places. All notation unless otherwise defined will be the same as in II §4 and, as in Theorem II 5.4. With $m = 0$, II (4.13) yields

(19) $$Eh(W) = \mathcal{P}_\lambda h - nE[E^{\mathcal{C}}(X_I - p)]S_\lambda h(W^{*+1}) + \lambda EX_I U_\lambda h(W^*)$$

In view of Theorem II 5.4, this identity can be written as

(20) $$\begin{aligned} Eh(W) &= \mathcal{P}_\lambda h - nE\varphi(I, X_{I-1}, X_{I+1})\, LS_\lambda h(W^*) + \lambda EX_I U_\lambda h(W^*) \\ &= \mathcal{P}_\lambda h - nE\varphi(I,1,1)\, X_{I-1}X_{I+1}LS_\lambda h(W^{O+2}) \\ &\quad - nE\varphi(I,0,1)(1 - X_{I-1})\, X_{I+1}LS_\lambda h(W^{O+1}) \\ &\quad - nE\varphi(I,1,0)\, X_{I-1}(1 - X_{I+1})\, LS_\lambda h(W^{O+1}) \\ &\quad - nE\varphi(I,0,0)(1 - X_{I-1})(1 - X_{I+1})\, LS_\lambda h(W^O) + \lambda EX_I U_\lambda h(W^*) \\ &= \mathcal{P}_\lambda h - nE\varphi(I,1,1)\, X_{I-1}X_{X+1}E^{I,X_{I-1},X_{I+1}}L^3 S_\lambda h(W^O) \\ &\quad - nE\varphi(I,0,1)(1 - X_{I-1})\, X_{I+1}E^{I,X_{I-1},X_{I+1}}L^2 S_\lambda h(W^O) \\ &\quad - nE\varphi(I,1,0)\, X_{I-1}(1 - X_{I+1})\, E^{I,X_{I-1},X_{I+1}}L^2 S_\lambda h(W^O) \\ &\quad - nE\varphi(I,0,0)(1 - X_{I-1})(1 - X_{I+1})\, E^{I,X_{I-1},X_{I+1}}LS_\lambda h(W^O) \\ &\quad + \lambda EX_I E^{I,X_I}U_\lambda h(W^*) \\ &= \mathcal{P}_\lambda h + R_1 + R_2 + R_3 + R_4 + R_5, \text{ say,} \end{aligned}$$

where L is defined by (1.1)

$$W^{\circ} = \sum_{\substack{k=1 \\ k \neq I-1, I, I+1}}^{n} X_k \tag{21}$$

and, as before, $X_k \equiv 0$ if $k \leq 0$ or $\geq n+1$. Now applying (19) to $E^{I,X_I} U_\lambda h(W^*)$, we obtain

$$E^{I,X_I} U_\lambda h(W^*) = \mathcal{P}_{\lambda^*} U_\lambda h - (n-1)\, E^{I,X_I}[E^{\mathcal{C}'}(X_J - p)]\, LS_{\lambda^*} U_\lambda h(W^{**}) + \lambda^* E^{I,X_I} X_J U_{\lambda^*} U_\lambda h(W^{**}) \tag{22}$$

where

$$\lambda^* = (n-1)p \tag{23}$$

$$p\{(I,J) = (i,j),\ i \neq j,\ i, j = 1, 2, \ldots, n\} = \frac{1}{n(n-1)} \tag{24}$$

and (I,J) is independent of the X_i's,

$$\mathcal{C}' = \mathcal{B}(I, J, X_k \text{ for } k = 1, 2, \ldots, n, k \neq J\} \tag{25}$$

and

$$W^{**} = \sum_{\substack{i=1 \\ i \neq I, J}}^{n} X_k \tag{26}$$

By arguments similar to those in Lemma II 5.7,

$$E^{\mathcal{C}'}(X_J - p) = \varphi(J, X_{J-1}, X_{J+1}) \tag{27}$$

Thus we have

$$(28)\quad R_5 = \lambda p \mathcal{P}_\lambda U_\lambda h + \lambda EX_I(\mathcal{P}_{\lambda^*} U_\lambda h - \mathcal{P}_\lambda U_\lambda h)$$
$$- \lambda(n-1)\, EX_I \varphi(J, X_{J-1}, X_{J+1})\, LS_{\lambda^*} U_\lambda h(W^{**})$$
$$+ \lambda\lambda^* EX_I X_J U_{\lambda^*} U_\lambda h(W^{**})$$

By (11),

$$(29)\quad |\lambda EX_I(\mathcal{P}_{\lambda^*} U_\lambda h - \mathcal{P}_\lambda U_\lambda h)| \le Cp^2$$

By Proposition 1.5 and tedious computation

$$(30)\quad |\lambda(n-1)\, EX_I \varphi(J, X_{J-1}, X_{J+1})\, LS_{\lambda^*} U_\lambda h(W^{**})|$$
$$\le C(\lambda^{1/2}|\rho|p + \lambda\rho^2)$$

By (12),

$$(31)\quad |\lambda\lambda^* E\, X_I X_J U_{\lambda^*} U_\lambda h(W^{**})|$$
$$\le CEX_I X_J + \frac{C}{\lambda^{*1/2}}\, EX_I X_J |W^{**} - (n-2)p|$$
$$= CEX_I X_J + \frac{C}{\lambda^{*1/2}}\, \{EX_I[E\mathcal{C}'(X_J - p)]|W^{**} - (n-2)p|$$
$$+ pE[E\mathcal{C}(X_I - p)]|W^{**} - (n-2)p| + p^2 E|W^{**} - (n-2)p|\}$$
$$\le CEX_I X_J + \frac{Cn}{\lambda^{*1/2}}\, EX_I|\varphi(J, X_{J-1}, X_{J+1})| + \frac{Cnp}{\lambda^{*1/2}}\, E|\varphi(I, X_{I-1}, X_{I+1})|$$
$$+ \frac{Cp^2}{\lambda^{*1/2}} \sqrt{E(W^{**} - (n-2)p)^2}$$

which by Lemmas II 5.7 and II 5.8 and (30)

$$\le C(p^2 + \lambda^{1/2}|\rho|p + \lambda\rho^2)$$

Thus the combination of (28), (29), (30) and (31) yields

(32) $$R_5 = \lambda p \mathcal{P}_\lambda U_\lambda h + R_5'$$

with

(33) $$|R_5'| \le C(p^2 + \lambda^{1/2}|\rho|p + \lambda\rho^2)$$

Next, applying (19) to $E^{I,X_{I-1},X_{I+1}} LS_\lambda h(W^o)$, we obtain

(34) $$E^{I,X_{I-1},X_{I+1}} LS_\lambda h(W^o)$$

$$= \mathcal{P}_{\lambda^o} LS_\lambda h - (n-3)_I E^{I,X_{I-1},X_{I+1}}[E\mathcal{C}''(X_K - p)]S_{\lambda^o} LS_\lambda h(W^{**+1})$$

$$+ \lambda^o E^{I,X_{I-1},X_{I+1}} X_K U_{\lambda^o} LS_\lambda h(W^{o*})$$

$$= \mathcal{P}_{\lambda^o} LS_\lambda h - (n-3)_I E^{I,X_{I-1},X_{I+1}} \varphi(K,X_{K-1},X_{K+1}) S_{\lambda^o} LS_\lambda h(W^{o*} + 1)$$

$$+ \lambda^o E^{I,X_{I-1},X_{I+1}} X_K U_{\lambda^o} LS_\lambda h(W^{o*})$$

where

(35) $$\lambda^o = EW^o$$

(36) (I,K) is uniform on $\{(i,k) : i, k = 1,2,\ldots,n \text{ and } |i-k| \ge 2\}$

and is independent of the X_i's,

(37) $$\mathcal{C}'' = \mathcal{B}(I, K, X_k \text{ for } k = 1,2, \ldots, n, k \ne K)$$

(38) $$W^{o*} = W^o - X_K$$

and

(39) $$(n-3)_I = \begin{cases} n-3 & \text{if } 2 \le I \le n-1 \\ n-2 & \text{if } I = 1 \text{ or } n \end{cases}$$

then we have

(40) $$\begin{aligned} R_4 = {} & -nE\varphi(I,0,0)(1 - X_{I-1})(1 - X_{I+1})\, \mathcal{P}_\lambda LS_\lambda h \\ & -nE\varphi(I,0,0)(1 - X_{I-1})(1 - X_{I+1})(\mathcal{P}_\lambda {}^o LS_\lambda h - \mathcal{P}_\lambda LS_\lambda h) \\ & + nE\varphi(I,0,0)(1-X_{I-1})(1-X_{I+1})(n-3)_I \varphi(k, X_{K-1}, X_{K+1}) LS_\lambda {}^o LS_\lambda h(W^{o*}) \\ & - nE\varphi(I,0,0)(1-X_{I-1})(1-X_{I+1})\, \lambda^o X_K U_\lambda {}^o LS_\lambda h(W^{o*}) \end{aligned}$$

By very tedious computation, (40) yields

(41) $$R_4 = 2\lambda\rho \mathcal{P}_\lambda LS_\lambda h + R_4'$$

with

(42) $$|R_4'| \le C(\lambda^{1/2}|\rho|p + \lambda\rho^2)$$

Similarly, we have

(43) $$|R_1| \le C(\lambda^{1/2}|\rho|p + \lambda\rho^2)$$

(44) $$R_2 = -\lambda\rho\, \mathcal{P}_\lambda L^2 S_\lambda h + R_2'$$

with

(45) $$|R_2'| \le C(\lambda^{1/2}|\rho|p + \lambda\rho^2)$$

and

(46) $$R_3 = -\lambda\rho\,\mathcal{P}_\lambda L^2 S_\lambda h + R_3'$$

with

(47) $$|R_3'| \le C(\lambda^{1/2}|\rho|p + \lambda\rho^2)$$

Finally by noting that

(48) $$\mathcal{P}_\lambda L^2 S_\lambda h - \mathcal{P}_\lambda L S_\lambda h = \mathcal{P}_\lambda U_\lambda h \ .$$

The combination of (20), (32), (33), (41), (42), (43), (44), (45), (46), (47) yields (17). Hence the theorem.

CHAPTER IV

COMPOUND POISSON CONVERGENCE

§1. Introduction

The aim of this chapter is to answer a question which arose naturally from the Poisson approximation problem studied in Chapter II. Recall that in Section 5 of Chapter II, we have obtained the bound for the departure of the distribution of the row sum of an m-dependent triangular array of Bernoulli random variables $X_{1n}, X_{2n}, \ldots, X_{nn}$ from the Poisson distribution $\mathcal{P}(\lambda_n)$ to be of the order of $\max_{1 \le i \le n} p_{in} + |C_n|$ where

$$p_{in} = P(X_{in} = 1) = 1 - P(X_{in} = 0) \tag{1}$$

$$C_n = \sum_{i<j} \operatorname{cov}(X_{in}, X_{jn}) \tag{2}$$

and

$$\lambda_n = \sum_{i=1}^{n} p_{in} \tag{3}$$

It follows from this result that a sufficient condition for $\mathcal{L}(\sum_{i=1}^{n} X_{in}) \longrightarrow \mathcal{P}(\lambda)$ is

$$\max_{1 \le i \le n} p_{in} \longrightarrow 0 \tag{4}$$

(5) $$\lambda_n \longrightarrow \lambda$$

and

(6) $$C_n \longrightarrow 0$$

Conditions (5) and (6) are equivalent to

(7) $$E \sum_{i=1}^{n} X_{in} \longrightarrow \lambda$$

and

(8) $$\mathrm{var}\left(\sum_{i=1}^{n} X_{in} \right) \longrightarrow \lambda$$

Philipp (1969) has proved that under conditions (4) and (8), (7) is necessary and sufficient for $\mathcal{L}\left(\sum_{i=1}^{n} X_{in} \right) \longrightarrow \mathcal{P}(\lambda)$.

In Section 4, under the assumption (4), we obtain a necessary and sufficient condition, which is analogous to (7) and (8), for the proper convergence of $\mathcal{L}\left(\sum_{i=1}^{n} X_{in} \right)$ and show that the class of possible limit laws is $\mathcal{L}\left(\sum_{i=1}^{m+1} iZ_i \right)$ where the Z_i are independent and Poisson distributed with parameters bearing a simple and explicit relation to the necessary and sufficient condition. This result reveals a direct relation between $\mathcal{L}\left(\sum_{i=1}^{n} X_{in} \right)$ and $\mathcal{L}\left(\sum_{i=1}^{m+1} iZ_i \right)$ and also yields (7) and (8) as a necessary and sufficient condition for $\mathcal{L}\left(\sum_{i=1}^{n} X_{in} \right) \longrightarrow \mathcal{P}(\lambda)$. It should be mentioned that Philipp (1969) has shown that, under certain conditions, the class of limit distributions for the row sums of certain mixing triangular arrays of random variables must be infinitely divisible and has also obtained a necessary and sufficient condition for convergence to any specified distribution in the class of possible limit distributions.

However, our necessary and sufficient condition and the explicitness of our limit distributions do not seem to follow immediately from his results.

The study of this problem leads to some related problems such as a characterization of the compound Poisson distribution and the study of the solution of an integral equation. These problems are interesting in their own rights as well as being preliminaries to Section 4. They are contained in Sections 2 and 3 of this chapter.

§2. A characterization of the compound Poisson distribution

The results in this section are generalizations of those in Section 2 of Chapter II. The proofs of Theorems 2.1 and 2.2 will be very similar to those of Propositions II 2.1 and II 2.2. But we shall be brief in the proofs in order to avoid repetition. We begin with the definition and some basic properties of the compound Poisson distribution.

A compund Poisson distribution is one whose characteristic function has the form $e^{\lambda(\varphi(t)-1)}$ where $\lambda > 0$ and $\varphi(t)$ is a characteristic function. It is the distribution of $X_1 + X_2 + \cdots + X_N$ where $X_1, X_2, \ldots$ are i.i.d. with $\varphi(t)$ as their common characteristic function and N is Poisson distributed with parameter λ and independent of the X_i. If μ is the common distribution of the X_i, then the compound Poisson distribution can also be defined as $e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \mu^{k*}$ where μ^{k*} is the k-fold convolution of μ. Note that each pair (λ,μ) uniquely determines a compound Poisson distribution but not

conversely. However, a one-to-one correspondence can be achieved if we restrict ourselves to only those μ which do not have an atom at zero. The compound Poisson distribution corresponding to (λ,μ) is said to be generated by μ and have parameter λ.

In the following, all functions considered are defined on the real line $\mathbb{R}$, real-valued and Borel measurable, unless otherwise defined.

Theorem 2.1. Z has a compound Poisson distribution with parameter $\lambda > 0$ and generated by μ where $\int |t| \, d\mu(t) < \infty$ if and only if for every bounded and continuous function f

$$EZf(Z) = \lambda E \int tf(Z+t) \, d\mu(t) \tag{1}$$

Proof. We first prove the "only if" part. For every such f,

$$EZf(Z) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} E(X_1 + \cdots + X_k) f(X_1 + \cdots + X_k) \tag{2}$$

where $X_1, X_2, \ldots$ are i.i.d. with distribution μ

$$= e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} EX_k f(X_1 + \cdots + X_k)$$

$$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k+1}}{k!} E \int tf(X_1 + \cdots + X_k + t) \, d\mu(t)$$

$$= \lambda \int tE \left[\sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} f(X_1 + \cdots + X_k + t)\right] d\mu(t)$$

$$= \lambda \int tEf(Z+t) \, d\mu(t) = \lambda E \int tf(Z+t) \, d\mu(t)$$

In (2), the second equality follows from the invariance of the joint distribution of $X_1, X_2, \ldots, X_k$ under permutation and the last three equalities from Fubini theorem. For the "if" part, by choosing $f = 1$, (1) implies $E|Z| < \infty$. Next by choosing $f(x)$ to be the real and imaginary parts of e^{iux} separately where $u \in \mathbb{R}$, (1) further implies

$$EZe^{iuZ} = \lambda E \int te^{iu(Z+t)}\, d\mu(t) \tag{3}$$

Since this is true for every $u \in \mathbb{R}$ and both $E|Z| < \infty$, $\int |t|\, d\mu(t) < \infty$, it follows from the dominated convergence theorem that

$$\frac{\partial}{\partial u} Ee^{iuZ} = \lambda \left[\frac{\partial}{\partial u} \int e^{iut}\, d\mu(t)\right] Ee^{iuZ} \tag{4}$$

Solving this differential equation and using the condition that $Ee^{iuZ} = 1$ at $u = 0$, we obtain

$$Ee^{iuZ} = e^{\lambda(\int e^{iut}\, d\mu(t) - 1)} \tag{5}$$

and this proves the theorem.

By approximating bounded and continuous functions by continuous functions with compact support in the same way as in Proposition II 2.2, we also obtain

Theorem 2.2. Z has a compound Poisson distribution with parameter $\lambda > 0$ and generated by μ where $\int |t|\, d\mu(t) < \infty$ if and only if (1) holds for every continuous function f with compact support.

If μ is concentrated on $(0,\infty)$, then we can remove the condition $\int |t|\, d\mu(t) < \infty$ in Theorem 2.2. Thus we have

Theorem 2.3. Z has a compound Poisson distribution with parameter $\lambda > 0$ and generated by μ which is concentrated on $(0,\infty)$ if and only if (1) holds for every continuous function f defined on $[0,\infty)$ and with compact support. Actually, if Z is compound Poisson distributed, then (1) holds for every f such that $E|Zf(Z)| < \infty$.

Proof. The proof of the "only if" part is the same as in Theorem 2.1. For the "if" part we shall need Theorems 3.1 and 3.2 in the next section. First we note that (1) holding with every continuous f on $[0,\infty)$ and with compact support implies (1) holding with every function f defined on $[0,\infty)$ such that $wf(w)$ is bounded and continuous on $(0,\infty)$. Now choose f as defined by (3.4) of the next section with equality holding for every $w \in (0,\infty)$ and $v(w) = g(w) - \int g(x)\, dQ_{\lambda,\mu}(x)$ where g is any bounded and continuous function defined on $[0,\infty)$ and

$$Q_{\lambda,\mu} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \mu^{k*} . \tag{6}$$

Then by Theorem 3.1, we obtain from (1)

$$Eg(Z) = \int g(x)\, dQ_{\lambda,\mu}(x) \tag{7}$$

and this proves the theorem.

§3. The solution of the equation $wf(w) - \lambda \int tf(w+t)\, d\mu(t) = v(w)$

An appropriately adequate study of the solution of the integral equation

$$wf(w) - \lambda \int tf(w+t)\, d\mu(t) = v(w) \tag{1}$$

is necessary for obtaining the best possible rate of convergence for the limit theorem we are interested in. However, since we are only interested in proving the limit theorem as it is, we shall need only the boundedness of its solution. In this section, we shall give a necessary and sufficient condition for (1) to have a bounded solution. It turns out that a bounded solution of (1) is also unique in some sense.

We shall assume μ to be concentrated on $(0,\infty)$ and consider only complex-valued functions which are defined a.e. on the real line $\mathbb{R}$ w.r.t. the measure $Q_{\lambda,\mu} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \mu^{k*}$ and are measurable w.r.t. $N^c \cap \mathcal{B}(\mathbb{R}) = \{A : A = N^c \cap B,\ B \in \mathcal{B}(\mathbb{R})\}$, where $\mathcal{B}(\mathbb{R})$ is the σ-algebra of Borel subsets of $\mathbb{R}$ and N^c is the set on which each individual function is defined (and $Q_{\lambda,\mu}(N) = 0$). We shall call the σ-algebra $A \cap \mathcal{F}$ the trace of the σ-algebra $\mathcal{F}$ on A. By a solution of (1), we mean a function f such that the integral on the left hand side of (1) is finite a.e. w.r.t. $Q_{\lambda,\mu}$ and also (1) is satisfied a.e. w.r.t. $Q_{\lambda,\mu}$.

Throughout the remainder of this section, "a.e." will mean "a.e. w.r.t. $Q_{\lambda,\mu}$" unless otherwise stated. We begin by proving a combinatorial identity which seems essential in the proofs of the following theorems.

Lemma 3.1. Let $a_1, a_2, \ldots, a_k$ be any k complex numbers. Then

$$(2) \qquad \sum_{\substack{\text{over all} \\ \text{permutations} \\ \text{of}(a_1,a_2,\ldots,a_k)}} \frac{a_1 a_2 \cdots a_k}{a_1(a_1+a_2)\cdots(a_1+\cdots+a_k)} = 1$$

provided no denominator in (2) is zero.

Proof. We prove (2) by induction. Obviously (2) is true for $k = 1$. Assume it to be true for $k \geq 1$, then

$$(3) \qquad \sum_{\substack{\text{over all} \\ \text{permutation} \\ \text{of } (a_1,\ldots,a_{k+1})}} \frac{a_1 a_2 \cdots a_{k+1}}{a_1(a_1+a_2)\cdots(a_1+\cdots+a_{k+1})}$$

$$= \frac{1}{a_1+a_2+\cdots+a_{k+1}} \left\{ a_{k+1} \sum_{\substack{\text{over all} \\ \text{permutations} \\ \text{of } (a_1,\ldots,a_k)}} \frac{a_1 a_2 \cdots a_k}{a_1(a_1+a_2)\cdots(a_1+\cdots+a_k)} \right.$$

$$\left. + \sum_{i=1}^{k} a_i \sum_{\substack{\text{over all} \\ \text{permutations} \\ \text{of } (a_1,\ldots,a_{k+1},a_k) \\ \text{with } a_{k+1} \text{ staying} \\ \text{in } i^{\text{th}} \text{ position}}} \frac{a_1 \cdots a_{k+1} \cdots a_k}{a_1(a_1+a_2)\cdots(a_1+\cdots+a_{k+1}+\cdots+a_k)} \right\}$$

(a_{k+1} in i^{th} position)

$$= \frac{1}{a_1+\cdots+a_{k+1}} \left(a_{k+1} + \sum_{i=1}^{k} a_i\right)$$

$$= 1$$

Theorem 3.1. Suppose $v(w)$ is bounded a.e. Then

$$f(w) = \sum_{k=0}^{\infty} \lambda^k E \frac{X_1 \cdots X_k \, v(w + X_1 + \cdots + X_k)}{w(w+X_1) \cdots (w+X_1 + \cdots + X_k)} \tag{4}$$

for almost all $w \in (0,\infty) \cap \mathcal{S}(Q_{\lambda,\mu})$ and $f(0)$ arbitrary and finite,

is a solution of (1) if and only if

$$\int v(w) \, dQ_{\lambda,\mu}(w) = 0 \tag{5}$$

where $\mathcal{S}(Q_{\lambda,\mu})$ is the support of $Q_{\lambda,\mu}$, and $X_1, X_2, \ldots$ are i.i.d. with distribution μ. Moreover, $|wf(w)| \leq e^{\lambda}\|v\|$ a.e. where $\|v\|$ is the essential supremum of v. If v is bounded and continuous on $[0,\infty)$ then f defined by (4) with equality holding for every $w \in (0,\infty)$ is such that $wf(w)$ is bounded and continuous on $(0,\infty)$.

Proof. Define two operators $\tilde{w}$ and S^t by

$$\tilde{w}f(w) = wf(w) \tag{6}$$

$$S^t f(w) = f(w+t) \tag{7}$$

Then (1) can be written as

$$[\tilde{w} - \lambda \int t S^t \, d\mu(t)] \, f(w) = v(w) \tag{8}$$

Let

$$\mathcal{D} = \{[f]: f \text{ is restricted to } (0,\infty) \cap \mathcal{S}(Q_{\lambda,\mu}) \text{ and } wf(w) \text{ is bounded a.e.}\}$$

and

$$\mathcal{R} = \{[v]: v \text{ is restricted to } (0,\infty) \cap \mathcal{S}(Q_{\lambda,\mu}) \text{ and } v \text{ is bounded a.e.}\},$$

where $[f]$ denotes the equivalence class of functions which equal f a.e. except at $w = 0$, and similarly for $[v]$.

Then by noting that

$$(9) \quad \left|\int tf(w+t)\, d\mu(t)\right| \leq \int \frac{t}{w+t} \left|(w+t)\, f(w+t)\right| d\mu(t) \leq \operatorname{ess\,sup}|wf(w)| ,$$

$$(10) \quad [\tilde{w} - \lambda \int tS^t\, d\mu(t)] : \mathcal{D} \longrightarrow \mathcal{R},$$

where

$$(11) \quad [\tilde{w} - \lambda \int tS^t\, d\mu(t)][f] = [\tilde{w} - \lambda \int tS^t\, d\mu(t)]f \quad \text{a.e.}$$

In the following we shall actually mean $[f]$ or $[v]$ when we talk about f or v and an operator operating on $[f]$ or $[v]$ will be defined as in (11).

Next define $[\tilde{w} - \lambda \int tS^t\, d\mu(t)]^{-1}$ by

$$(12) \quad [\tilde{w} - \lambda \int tS^t\, d\mu(t)]^{-1} v(w)$$

$$= \sum_{k=0}^{\infty} \lambda^k [\tilde{w}^{-1} \circ \int tS^t\, d\mu(t)]^k \circ \tilde{w}^{-1} v(w)$$

$$= \sum_{k=0}^{\infty} \lambda^k \int\cdots\int \frac{t_1 \cdots t_k\, v(w+t_1+\cdots+t_k)\, d\mu(t_1) \cdots d\mu(t_k)}{w(w+t_k)(w+t_k+t_{k-1}) \cdots (w+t_k+\ldots+t_1)} .$$

$$= \sum_{k=0} \lambda^k E \frac{X_1 \cdots X_k \, v(w+X_1+ \cdots + X_k)}{w(w+X_1) \cdots (w+X_1+ \cdots + X_k)}$$

Note that Fubini Theorem ensures that for almost all $w \in (0,\infty) \cap \mathcal{S}(Q_{\lambda,\mu})$, $\frac{v(w+t_1+ \cdots + t_k)}{(w+t_k) \cdots (w+t_k+ \cdots + t_1)}$ is defined on $\mathbb{R}^k$ a.e. w.r.t. the k-fold product measure of μ and is measurable w.r.t. the trace of $\mathcal{B}(\mathbb{R})^k$ on its domain of definition. By Lemma 3.1 and the invariance of the joint distribution of $X_1, X_2, \ldots$ under permutation, we have for almost all $w \in (0,\infty) \cap \mathcal{S}(Q_{\lambda,\mu})$

(13) $$|w[\tilde{w} - \lambda \int tS^t \, d\mu(t)]^{-1} v(w)|$$

$$\leq \|v\| \sum_{k=0}^{\infty} \lambda^k E \frac{X_1 \cdots X_k}{X_1(X_1 + X_2) \cdots (X_1 + \cdots + X_k)}$$

$$= \|v\| \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}\|v\| < \infty$$

where $\|v\|$ is the essential supremum of v w.r.t. $Q_{\lambda,\mu}$. Thus

(14) $$[\tilde{w} - \lambda \int tS^t \, d\mu(t)]^{-1} : \mathcal{R} \to \mathcal{D}$$

and it follows from a simple verification that $[\tilde{w} - \lambda \int tS^t \, d\mu(t)]^{-1}$ defined by (12) is the inverse of $[\tilde{w} - \lambda \int tS^t \, d\mu(t)]$. Hence it follows from (9) and (12) that (4) is a solution of (1) if we can also verify that (8) holds also for $w = 0$ (which is a point of increase of $Q_{\lambda,\mu}$). Substituting (4) into (8), we obtain

$$(15) \qquad v(0) = -\lambda \int tf(t)\, d\mu(t)$$

$$= -\lambda \sum_{u=0}^{\infty} \lambda^k \int E \frac{X_1 \cdots X_k\, v(t+X_1+\cdots+X_k)}{(t+X_1)\cdots(t+X_1+\cdots+X_k)}\, d\mu(t)$$

$$= -\sum_{k=0}^{\infty} \lambda^{k+1} E \frac{X_1 \cdots X_{k+1}\, v(X_1+\cdots+X_{k+1})}{X_1(X_1+X_2)\cdots(X_1+\cdots+X_{k+1})}$$

which again by Lemma 3.1 and the invariance of the joint distribution of $X_1, X_2, \ldots$ under permutation

$$= -\sum_{k=1}^{\infty} \frac{\lambda^k}{k!} Ev(X_1+\cdots+X_k)$$

But (15) is equivalent to (5). Hence the necessity and sufficiency of (5). The rest of the theorem being clear, this proves the theorem.

We shall be interested in the case where μ is concentrated on $\{1,2,\ldots\}$. In this case, $Q_{\lambda,\mu}$ is concentrated on $\{0,1,2,\ldots\}$ and it is obvious that f given by (4) is bounded a.e. For then

$$(16) \qquad |f(w)| \le \|v\| \sum_{k=0}^{\infty} \lambda^k E \frac{X_1 \cdots X_k}{X_1(X_1+X_2)\cdots(X_1+\cdots+X_k)}$$

$$= \|v\|\, e^{\lambda} \quad \text{for} \quad w = 1, 2, \ldots$$

In general, f may not be bounded a.e. unless an additional condition is given. The next theorem gives two sufficient conditions under which f is bounded a.e.

Theorem 3.2. Let v be bounded a.e. the solution of (1) given by (4) is bounded a.e. if either one of the following condition is satisfied.

(17) μ is concentrated on $[c,\infty)$ for some $c > 0$.

(18) There exists a constant K such that $|v(x)-v(y)| \leq K|x-y|$.

Actually, if v is defined and bounded everywhere on $[0,\infty)$, then under condition (18), f given by (4) with equality holding for every $w \in (0,\infty)$ is bounded and continuous on $(0,\infty)$. If in addition, v has a bounded derivative on $[0,\infty)$, then f is bounded and uniformly continuous on $(0,\infty)$.

Proof. The sufficiency of (17) is obvious. We shall only prove that of (18). First, consider v to be bounded a.e. For this, it suffices to show that for almost all $w \in (0,\infty) \cap \mathcal{S}(Q_{\lambda,\mu})$,

$$(19)\quad \sum_{k=0}^{\infty} \lambda^k E \frac{X_1 \cdots X_k \, v(w + X_1 + \cdots + X_k)}{w(w+X_1) \cdots (w+X_1+\cdots+X_k)}$$

$$= e^{2\lambda} E \frac{(-1)^L}{w + \sum_{i=1}^{L} X_i} [v(w + \sum_{i=1}^{L} X_i + \sum_{j=1}^{M} Y_j) - v(\sum_{j=1}^{M} Y_j)]$$

where L and M are independent and Poisson distributed with parameter λ and the X_i and Y_j are i.i.d. with distribution μ and are independent of L and M. First, we prove by induction that

$$(20)\quad k!\, E\,\frac{X_1 \cdots X_k\, v(w+X_1+\cdots+X_k)}{w(w+X_1)\cdots(w+X_1+\cdots+X_k)}$$

$$= E\sum_{\ell=0}^{k}(-1)^{\ell}\binom{k}{\ell}\frac{v(w+X_1+\cdots+X_k)}{w+\sum_{i=1}^{\ell}X_i}$$

Clearly (20) is true for $k = 1$. Assume (20) to be true for some $k \geq 1$. Then

$$(21)\quad E\sum_{\ell=0}^{k+1}(-1)^{\ell}\binom{k+1}{\ell}\frac{v(w+X_1+\cdots+X_{k+1})}{w+\sum_{i=1}^{\ell}X_i}$$

$$= E\sum_{\ell=0}^{k}(-1)^{\ell}\binom{k}{\ell}\frac{v(w+X_1+\cdots+X_{k+1})}{w+\sum_{i=1}^{\ell}X_i}$$

$$- E\sum_{\ell=0}^{k}(-1)^{\ell}\binom{k}{\ell}\frac{v(w+X_1+\cdots+X_{k+1})}{w+\sum_{i=1}^{\ell+1}X_i}$$

$$= k!\, E\,\frac{X_1 \cdots X_u\, v(w+X_1+\cdots+X_{k+1})}{w(w+X_1)\cdots(w+X_1+\cdots+X_k)}$$

$$- k!\, E\,\frac{X_2 \cdots X_{k+1}\, v(w+X_1+\cdots+X_{k+1})}{(w+X_1)(w+X_1+X_2)\cdots(w+X_1+\cdots+X_{k+1})}$$

$$= k!\, E\,\frac{X_1 \cdots X_k\, v(w+X_1+\cdots+X_{k+1})}{w(w+X_1)\cdots(w+X_1+\cdots+X_k)}$$

$$- k!\, E\,\frac{X_1 \cdots X_k\, v(w+X_1+\cdots+X_{k+1})}{(w+X_{k+1})(w+X_{k+1}+X_1)\cdots(w+X_{k+1}+X_1+\cdots+X_k)}$$

Define for $i = 0,1,\ldots,k$,

$$(22)\quad g_i = \prod_{j=0}^{i}(w+X_{k+1}+X_1+\cdots+X_j)\prod_{j=i+1}^{k}(w+X_1+\cdots+X_j)$$

and

$$g_{-1} = \prod_{j=0}^{k} (w + X_1 + \cdots + X_j) \tag{23}$$

Then (21) yields

$$E \sum_{\ell=0}^{k+1} (-1)^{\ell} \binom{k+1}{\ell} \frac{v(w + X_1 + \cdots + X_{k+1})}{w + \sum_{i=1}^{\ell} X_i} \tag{24}$$

$$= k!\, E\, X_1 \cdots X_k \sum_{i=0}^{k} \left(\frac{g_i - g_{i-1}}{g_{-1} g_k} \right) v(w + X_1 + \cdots + X_{k+1})$$

$$= (k+1)!\, E \frac{X_1 \cdots X_{k+1}\, v(w+X_1+ \cdots + X_{k+1})}{w(w+X_1) \cdots (w+X_1+ \cdots +X_{k+1})}$$

and thus (20) is also true for $k+1$, where the last equality follows from the invariance of the joint distribution of $X_1, X_2, \ldots$ under permutation. Next by observing that $\mathcal{L}(L|L + M = k)$ is binomial $(k, 1/2)$, we have

$$E^{L+M=k}\, v(w + \sum_{i=1}^{L} X_i + \sum_{j=1}^{M} Y_j)\, E^{L+M=k, X_1, \ldots, X_L, Y_1, \ldots, Y_M} \frac{(-1)^L}{w + \sum_{i=1}^{L} X_i} \tag{25}$$

$$= \frac{1}{2^k} E \sum_{\ell=0}^{k} (-1)^{\ell} \binom{k}{\ell} \frac{v(w + X_1 + \cdots + X_k)}{w + \sum_{i=1}^{\ell} X_i}$$

Finally, the combination of (20), (25) and (5) yields (19) and this proves the sufficiency of (18). If v is defined and bounded everywhere on $[0,\infty)$, then (19) holds for every $w \in (0,\infty)$ and the continuity and boundedness of f follows from (18), (19) and the

continuity of $wf(w)$ (see Theorem 3.1). In addition, by applying the bounded convergence theorem to the right hand side of (19) as $w \longrightarrow 0$, (18) and the differentiability of v with bounded derivatives imply $\lim_{w \to 0^+} f(w)$ exists. This together with the continuity and boundedness of $wf(w)$ on $(0,\infty)$ imply the uniform continuity of f and this completes the proof of the theorem.

Condition (18) is not as stringent as it may seem. For $v(w) = e^{itw} - \int e^{itx} dQ_{\lambda,\mu}(x)$ for every $t \in \mathbb{R}$, satisfies (18). In fact it is defined, bounded and has bounded derivatives everywhere on $[0,\infty)$. Choosing such a v will enable us to approximate the characteristic function of the distribution under consideration by that of the compound Poisson distribution $Q_{\lambda,\mu}$. The result is not only good enough for proving a limit theorem but can also be used to obtain the rate of convergence. All that is needed is to apply the standard procedures to the approximating characteristic function. Of course, as has been mentioned at the beginning of this section, in order to obtain a good rate, one has to study the solution of (1) beyond its boundedness.

We have so far established conditions for the existence of a bounded solution of equation (1). It is natural to ask if such a solution is unique. The next theorem furnishes the answer to this question.

Theorem 3.3. If equation (1) has a solution f such that

$$\int |wf(w)| \, dQ_{\lambda,\mu}(w) < \infty , \tag{26}$$

Then, except at $w = 0$, it must be unique up to a set of $Q_{\lambda,\mu}$-measure zero among those solutions satisfying (26).

In particular, if (1) has a bounded solution, it must be unique in this sense. Note that v need not be bounded a.e.

Proof. It suffices to show that (26) and

$$wf(w) - \lambda \int tf(w+t) \, d\mu(t) = 0 \quad \text{a.e.} \tag{27}$$

imply $f = 0$ a.e.

By Theorem 2.3, (36) is equivalent to

$$\iint t|f(w+t)| \, d\mu(t) \, dQ_{\lambda,\mu}(w) < \infty \tag{28}$$

But the left hand side of (28) is

$$e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} E \int t|f(t + \sum_{i=1}^{k} X_i)| \, d\mu(t)$$

where $X_1, X_2, \ldots$ are i.i.d. with distribution μ. Therefore,

$$\sum_{k=n}^{\infty} \frac{\lambda^k}{k!} E \int t|f(t + \sum_{i=1}^{k} X_i)| \, d\mu(t) \longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty \tag{29}$$

Now (27) implies that for $n = 1,2, \ldots$

$$(30)\qquad wf(w) = \lambda^{n+1} E \frac{X_1 \cdots X_n \int t|f(w + t + \sum_{i=1}^{n} X_i)|\, d\mu(t)}{(w + X_1) \cdots (w + \sum_{i=1}^{n} X_i)} \quad \text{a.e.}$$

thus

$$(31)\qquad \int |wf(w)|\, dQ_{\lambda,\mu}(w)$$

$$\leq \lambda^{n+1} \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{k!} E \frac{X_1 \cdots X_n \int t|f(t + \sum_{j=1}^{k} Y_j + \sum_{i=1}^{n} X_i)|\, d\mu(t)}{(\sum_{j=1}^{k} Y_j + X_1) \cdots (\sum_{j=1}^{k} Y_j + \sum_{i=1}^{n} X_i)}$$

where $X_1, X_2, \ldots, Y_1, Y_2, \ldots$ are i.i.d. with distribution μ.

By Lemma 3.1 and the invariance-under-permutation argument applied only to $Y_1, \ldots, Y_k$ first and then to $Y_1, \ldots, Y_k, X_1, \ldots, X_n$,

$$(32)\qquad \text{the right hand side of (31)}$$

$$= e^{-\lambda} \lambda^{n+1} \sum_{k=0}^{\infty} \lambda^k E \frac{Y_1 \cdots Y_k X_1 \cdots X_n \int t|f(t + \sum_{j=1}^{k} Y_j + \sum_{i=1}^{n} X_i)|\, d\mu(t)}{Y_1(Y_1+Y_2)\cdots(\sum_{j=1}^{k} Y_j)(\sum_{j=1}^{k} Y_j + X_1)\cdots(\sum_{j=1}^{k} Y_j + \sum_{i=1}^{n} X_i)}$$

$$= \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k+n}}{(k+n)!} E \int t|f(t + \sum_{j=1}^{k} Y_j + \sum_{i=1}^{n} X_i)|\, d\mu(t)$$

$$= \lambda e^{-\lambda} \sum_{k=n}^{\infty} \frac{\lambda^k}{k!} E \int t|f(t + \sum_{i=1}^{k} X_i)|\, d\mu(t)$$

which by (29) $\longrightarrow 0$ as $n \longrightarrow \infty$. Hence $\int |wf(w)|\, dQ_{\lambda,\mu}(w) = 0$ and this proves the theorem.

Note that under the boundedness of v, the uniqueness of the solution is implicit in Theorem 3.1 if $wf(w)$ is bounded a.e. Theorem 3.3 shows that the uniqueness is still true under the weaker condition (26) and without assuming the boundedness of v.

Finally if we combine Theorem 3.1, 3.2 and 3.3, we have

Theorem 3.4. Let v be bounded a.e. If either one of conditions (17) and (18) is satisfied, then (1) has a bounded solution if and only if $\int v(w)\, dQ_{\lambda,\mu}(w) = 0$. The solution is given by (4) and is unique (among bounded solutions) up to a set of $Q_{\lambda,\mu}$-measure zero except at $w = 0$.

§4. The limit theorem

This section contains the main theorem in this chapter, which we first state as follows.

Theorem 4.1. Let $X_{1n}, X_{2n}, \ldots, X_{nn}$, $n = 1, 2, \ldots$ be a triangular array of m-dependent Bernoulli random variables with $P(X_{in} = 1) = 1 - P(X_{in} = 0) = p_{in}$ such that

$$\max_{1 \le i \le n} p_{in} \longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty \tag{1}$$

Then $\mathcal{L}(\sum_{i=1}^{n} X_{in})$ converges properly if and only if for $k = 1, 2, \ldots, m+1$,

$$(2)\qquad \sum_{\substack{i_1 < i_2 < \cdots < i_k \\ i_k - i_1 \le m}} EX_{i_1 n} X_{i_2 n} \cdots X_{i_k n} \longrightarrow \theta_k \quad \text{for some } \theta_k \quad \text{as } n \longrightarrow \infty$$

The limit distribution must be $\mathcal{L}(\sum_{i=1}^{m+1} iZ_i)$ where the Z_i are independent and Poisson distributed, the parameter of each Z_i being

$$(3)\qquad \gamma_i = \sum_{k=i}^{m+1} (-1)^{k-i} \binom{k}{i} \theta_k$$

and the γ_i are necessarily non-negative.

Before embarking on proving this theorem, we shall first prove three lemmas.

Lemma 4.1. Let X_{in}, $i = 1,2,\ldots,r_n$, $n = 1,2,\ldots$ be a triangular array of m-dependent Bernoulli random variables with $P(X_{in} = 1) = 1 - P(X_{in} = 0) = p_{in}$ such that

$$(4)\qquad \tilde{p} = \max_{1 \le i \le r_n} p_{in} \longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty$$

$$(5)\qquad \sum_{i=1}^{r_n} p_{in} \le c < \infty$$

where c is a constant independent of n. If $\mathcal{L}(\sum_{i=1}^{r_n} X_{in})$ converges properly to $\mathcal{L}(Z)$, then $\mathcal{L}(Z)$ must be infinitely divisible.

Proof. Define ρ_1 to be the smallest integer such that $E\sum_{i=1}^{\rho_1} X_{in} > \tilde{p}^{1/2}$ and ρ_k, $k = 2,3,\ldots$ be the smallest integer such that

$$E\sum_{i=\rho_1+\cdots+\rho_{k-1}+(k-1)m+1}^{\rho_1+\cdots+\rho_k+(k-1)m} X_{in} > \tilde{p}^{1/2}$$

Let

$$Y_{kn} = \sum_{i=\rho_1+\cdots+\rho_{k-1}+(k-1)m+1}^{\rho_1+\cdots+\rho_k+(k-1)m} X_{in} \tag{6}$$

and

$$Z_{kn} = \sum_{i=\rho_1+\cdots+\rho_k+(k-1)m+1}^{\rho_1+\cdots+\rho_k+km} X_{in} \tag{7}$$

where $k = 1,2, \ldots$ and $\rho_1 + \cdots + \rho_{k-1}$ is taken to be zero when $k = 1$.

Then we can write

$$\sum_{i=1}^{r_n} X_{in} = \sum_{k=1}^{\ell} Y_{kn} + \sum_{k=1}^{\ell} Z_{kn} + T \tag{8}$$

with

$$ET = O(\tilde{p}^{1/2}) \tag{9}$$

By the definitions of Y_{kn} and Z_{kn},

$$EY_{kn} = \tilde{p}^{1/2} + O(\tilde{p}) \tag{10}$$

and

$$EZ_{kn} = O(\tilde{p}) \tag{11}$$

Thus

$$(12) \qquad \sum_{i=1}^{r_n} p_{in} = E \sum_{i=1}^{r_n} X_{in} = \tilde{\ell p}^{1/2} + o(\tilde{\ell p})$$

This together with the boundedness of $\sum_{i=1}^{r_n} p_{in}$ imply

$$(13) \qquad \tilde{\ell p} \longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty$$

and hence

$$(14) \qquad E \sum_{k=1}^{\ell} Z_{kn} \longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty$$

Consequently,

$$(15) \qquad \mathcal{L}\left(\sum_{i=1}^{\ell} Y_{in}\right) \longrightarrow \mathcal{L}(Z)$$

But, since the X_{in} are m-dependent, it follows that the Y_{kn} are independent. Also by (4) and (10), $Y_{kn} \xrightarrow{P} 0$ uniformly in i as $n \longrightarrow \infty$. Therefore, by the classical limit theorem for independent triangular arrays of random variables (see Feller [2], p. 550), $\mathcal{L}(Z)$ must be infinitely divisible.

The next lemma follows from decomposition of characteristic functions (see Feller [2], p. 539). However, we shall here give a different proof.

Lemma 4.2. If the distribution Z is infinitely divisible and of lattice type with a minimum point of increase, then

$$(16) \qquad Z \overset{d}{=} \text{(equal in distribution to) } a + bZ'$$

where a is the minimum point of increase, b the span of the lattice distribution and Z' has a compound Poisson distribution concentrated on $\{0,1,2,\ldots\}$.

Proof. Without loss of generality, we may assume Z to be concentrated on $\{0,1,2,\ldots\}$ with $P(Z = 0) > 0$. Since Z is infinitely divisible, for every $n = 1,2,\ldots$

$$Z \overset{d}{=} Z_{1n} + Z_{2n} + \cdots + Z_{nn} \tag{17}$$

where the Z_{in} are i.i.d.

Therefore for every continuous function f defined on $[0,\infty)$ and with compact support,

$$\begin{aligned} EZf(Z) &= nEZ_{nn}f(Z^{(n)} + Z_{nn}) \\ &= nE\int tf(Z^{(n)}+t)\, d\mu_n(t) \end{aligned} \tag{18}$$

where μ_n is the common distribution of the Z_{in}, and

$$Z^{(n)} = Z_{1n} + \cdots + Z_{n-1,n} \tag{19}$$

In (18), the first equality follows from the invariance of the joint distribution of $Z_{1n}, \ldots, Z_{nn}$ under permutation and the second equality from the independence of the Z_{in}.

Now since Z is concentrated on $\{0,1,2,\ldots\}$ and $P(Z = 0) > 0$, it follows that each Z_{in} is concentrated on $\{0,1,2,\ldots\}$ and $P(Z_{in} = 0) > 0$. Let

(20) $$P(Z_{in} = k) = p_{kn}, \qquad k = 0,1,2, \ldots$$

and

(21) $$\lambda_n = \sum_{k=1}^{\infty} p_{kn}$$

Then (18) yields

(22) $$EZf(Z) = n\lambda_n \sum_{k=1}^{\infty} \frac{kp_{kn}}{\lambda_n} Ef(Z^{(n)} + k)$$

Now

(23) $$P(W=0) = [P(Z_{1n}=0)]^n = (1-\lambda_n)^n \qquad \text{for } n = 1,2, \ldots$$

So

(24) $$n\lambda_n \longrightarrow \lambda \text{ for some } 0 \leq \lambda < \infty$$

This in turn implies

(25) $$Z_{nn} \xrightarrow{P} 0 \text{ and hence } \mathcal{L}(Z^{(n)}) \longrightarrow \mathcal{L}(Z)$$

Choose f such that

$$f(x) = \begin{cases} 1 & \text{if } x = 1 \\ 0 & \text{if } x = \text{any other integer} \end{cases}$$

Then (22) yields

(26) $$P(Z=1) = n\lambda_n \frac{p_{1n}}{\lambda_n} P(Z^{(n)} = 0)$$

This together with (24), (25) and $P(Z=0) > 0$ imply

(27) $$\begin{cases} P(Z=1) = 0 & \text{if } \lambda = 0 \\ \dfrac{p_{1n}}{\lambda_n} \longrightarrow \alpha_1 & \text{for some } \alpha_1 \text{ if } \lambda > 0 \end{cases}$$

Next choose

$$f(x) = \begin{cases} 1 & \text{if } x = 2 \\ 0 & \text{if } x = \text{any other integer} \end{cases}$$

and by a similar argument

$$(28) \qquad \begin{cases} P(Z=2) = 0 & \text{if } \lambda = 0 \\ \dfrac{p_{2n}}{\lambda_n} \longrightarrow \alpha_2 \text{ for some } \alpha_2, & \text{if } \lambda > 0 \end{cases}$$

Proceeding inductively, we obtain for $k = 1,2,3, \ldots$

$$(29) \qquad \begin{cases} P(Z=k) = 0 & \text{if } \lambda = 0 \\ \dfrac{p_{kn}}{\lambda_n} \longrightarrow \alpha_k \text{ for some } k, & \text{if } \lambda > 0 \end{cases}$$

Thus if $\lambda = 0$, $\mathcal{L}(Z) = 0$ (which is compound Poisson). If $\lambda > 0$, combining (22), (24), (25) and (29), we obtain

$$(30) \qquad EZf(Z) = \lambda E \sum_{k=1}^{\infty} k\alpha_k f(W+k) = \lambda\left(\sum_{i=1}^{\infty} \alpha_i\right) E \int tf(W+t)\, d\mu(t)$$

where f is any continuous function defined on $[0,\infty)$ and with compact support and μ attributes probability mass $\alpha_k / \sum_{i=1}^{\infty} \alpha_k$ to the integer k, $k = 1,2, \ldots$. (Note that $0 < \sum_{i=1}^{\infty} \alpha_i \leq 1$.) Hence by Theorem 2.3, the distribution of Z is compound Poisson with parameter $\lambda \sum_{i=1}^{\infty} \alpha_i$ and generated by μ, and this proves the lemma.

The next lemma can easily be proved using characteristic function, but we shall again give a different proof.

Lemma 4.3. Let $Z_1, Z_2, \ldots$ be independent and Poisson distributed, the parameter of each Z_i being γ_i. Then $\mathcal{L}(\sum_{i=1}^{n} iZ_i)$ converges properly if and only if

$$\lambda = \sum_{i=1}^{\infty} \gamma_i < \infty \tag{31}$$

The limit distribution is compound Poisson which has parameter λ and is generated by the distribution μ which attributes probability mass γ_k/λ to the integer k, where $k = 1,2, \ldots$.

Proof. Clearly for every n, $\sum_{i=1}^{n} iZ_i$ is stochastically larger than $\sum_{i=1}^{n} Z_i$ which is Poisson distributed with parameter $\sum_{i=1}^{n} \gamma_i$. Therefore, if $\sum_{i=1}^{\infty} \gamma_i = \infty$, the probability mass of $\sum_{i=1}^{n} Z_i$ escapes to infinity as $n \longrightarrow \infty$ and so does the probability mass of $\sum_{i=1}^{n} iZ_i$. Conversely, for every real-valued function f defined on $\{0,1,2, \ldots\}$ such that $wf(w)$ is bounded, we have

$$E(\sum_{i=1}^{n} iZ_i)\, f(\sum_{i=1}^{n} iZ_i) = \sum_{i=1}^{n} iEZ_i\, f(\sum_{i=1}^{n} iZ_i) \tag{32}$$

$$= \sum_{i=1}^{n} i\gamma_i\, Ef(\sum_{i=1}^{n} iZ_i + i)$$

Now let f be given by (3.4) with

$$v(w) = g(w) - \int g(x)dQ_{\lambda,\mu}(x) \tag{33}$$

where g is any bounded real-valued function defined on $\{0,1,2, \ldots\}$ and $Q_{\lambda,\mu} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \mu^{k*}$. Then by Theorem 3.1, (32) yields

$$(34) \qquad \left|Eg\left(\sum_{i=1}^{n} iZ_i\right) - \int g(x)\, dQ_{\lambda,\mu}(x)\right| = \left|\sum_{i=n+1}^{\infty} i\gamma_i Ef\left(\sum_{i=1}^{n} iZ_i + i\right)\right|$$

$$\leq 2e^{\lambda} \|g\| \sum_{i=n+1}^{\infty} \gamma_i$$

which by (31) $\longrightarrow 0$ as $n \longrightarrow \infty$ and this proves the lemma.

We are now in the position to prove Theorem 4.1.

Proof of Theorem 4.1. The crux of the whole proof is the derivation of the identity that for every bounded real-valued function f defined on $\{0,1,2, \ldots\}$

$$(35) \qquad EW_n f(W_n) - \sum_{i=1}^{m+1} i\gamma_i^{(n)} Ef(W_n + i) = R$$

where

$$(36) \qquad W_n = \sum_{i=1}^{n} X_{in}$$

$$(37) \qquad \theta_k^{(n)} = \sum_{\substack{i_1 < i_2 < \cdots < i_k \\ i_k - i_1 \leq m}} EX_{i_1 n} X_{i_2 n} \cdots X_{i_k n}$$

$$(38) \qquad \gamma_i^{(n)} = \sum_{k=i}^{m+1} (-1)^{k-i} \binom{k}{i} \theta_k^{(n)}$$

(39) $$|R| \leq C(\theta_1^{(n)}, m, \|f\|) \max_{1 \leq i \leq n} p_{in}$$

and

(40) $C(\theta_1^{(n)}, m, \|f\|)$ depends on $\theta_1^{(n)}$, m and $\|f\| = \sup_x |f(x)|$, and remains bounded as long as $\theta_1^{(n)}$, m and $\|f\|$ are bounded.

We shall leave the derivation of (35) until the end of the proof. For the time being, we shall assume that we have already derived (35) and proceed to prove the theorem. First, the sufficiency of (2). We observe that $\theta_k^{(n)} \longrightarrow \theta_k$ for $k = 1,2, \ldots, m+1$ is equivalent to $\gamma_i^{(n)} \longrightarrow \gamma_i$ for $i = 1,2, \ldots, m+1$. $\theta_1^{(n)} \longrightarrow \theta_1$ and Chebyshev inequality imply that the sequence of distributions $\mathcal{L}(W_n)$ is tight. Therefore, by Helly-Bray theorem, there exists a subsequence W_{n_k} such that $\mathcal{L}(W_{n_k})$ converges properly, say, to $\mathcal{L}(Z)$. By Lemma 4.1, $\mathcal{L}(Z)$ must be infinitely divisible. Since $\mathcal{L}(W_{n_k})$ are all concentrated on $\{0,1,2, \ldots\}$, $\mathcal{L}(Z)$ must also be concentrated on $\{0,1,2, \ldots\}$. It also follows that $P(Z=0) > 0$. Otherwise, by choosing

$$f(x) = \begin{cases} 1 & x = \ell \\ 0 & x = \text{any other integer} \end{cases}$$

where $\ell = 1,2, \ldots$, (35) implies that $\mathcal{L}(Z)$ has zero mass, which contradicts the proper convergence of $\mathcal{L}(W_{n_k})$. Therefore, by Lemma 4.2, $\mathcal{L}(Z)$ must be compound Poisson and by Theorem 2.3, we have for every real-valued bounded function f defined on $\{0,1,2, \ldots\}$ which vanishes outside a finite subset of $\{0,1,2, \ldots\}$,

$$EZ\ f(Z) = \sum_{i=1}^{\infty} i\beta_i\ Ef(Z+i) \tag{41}$$

where $\beta_i \geq 0$, $i = 1,2, \ldots$. But by (35) and the proper convergence of $\mathcal{L}(W_{n_k})$ to $\mathcal{L}(Z)$, we also have for every such f

$$EZ\ f(Z) = \sum_{i=1}^{m+1} i\gamma_i\ Ef(Z+i) \tag{42}$$

By comparing (41) with (42), $\gamma_i = \beta_i$ for $i = 1, 2, \ldots, m+1$ and $\beta_i = 0$ for $i = m+2, m+3, \ldots$. Thus γ_i must necessarily be non-negative. Now write (35) as

$$EW_n\ f(W_n) - \sum_{i=1}^{m+1} i\gamma_i\ Ef(W_n+i) = R+R' \tag{43}$$

where

$$|R'| \leq (m+1)\ \|f\| \max_{1 \leq i \leq m+1} i|\gamma_i^{(n)} - \gamma_i| \tag{44}$$

Choose f to be the solution of equation (3.1) given by (3.4) of Theorem 3.1 with $v(w) = g(w) - \int g(x)\, dQ_{\lambda,\mu}(x)$ where g is any real-valued bounded function defined on $\{0,1,2, \ldots\}$,

$$\lambda = \sum_{i=1}^{m+1} \gamma_i\ , \tag{45}$$

μ attributes probability mass γ_k/λ to the integer k, $k = 1,2,\ldots,m+1$, and $Q_{\lambda,\mu} = e^{-\lambda} \sum_{u=0}^{\infty} \frac{\lambda^k}{k!} \mu^{k*}$. By Theorem 3.2, this f is bounded and

its substitution into (43) yields

$$Eg(W_n) = \int g(x)\, dQ_{\lambda,\mu}(x) + R + R' \tag{46}$$

As $n \longrightarrow \infty$ we obtain

$$\lim_{n \to \infty} Eg(W_n) = \int g(x)\, dQ_{\lambda,\mu}(x) \tag{47}$$

By Lemma 4.3, $Q_{\lambda,\mu} = \mathcal{L}(\sum_{i=1}^{m+1} iZ_i)$ and thus we have

$$\mathcal{L}(W_n) \longrightarrow \mathcal{L}(\sum_{i=1}^{m+1} iZ_i) \tag{48}$$

Conversely, suppose $\mathcal{L}(W_n)$ converges properly, say, to $\mathcal{L}(Z)$. Let

$$W_{in} = X_{in} + X_{i+m+1,n} + X_{i+2m+2,n} + \cdots \tag{49}$$

where $i = 1,2, \ldots, m+1$. Then

$$W_n = \sum_{i=1}^{m+1} W_{in}\,, \tag{50}$$

and W_n is stochastically larger than W_{in} for any i. Suppose $\sum_{i=1}^{n} p_{in}$ does not remain bounded. Then there exists a subsequence n_k of n and another sequence $i(n_k)$, $1 \leq i(n_k) \leq m+1$ such that

$$(51) \qquad EW_{i(n_k),n_k} = p_{i(n_k),n} + p_{i(n_k)+m+1,n_k} + \cdots \longrightarrow \infty$$

$$\text{as} \quad n_k \longrightarrow \infty$$

As the X's in each $W_{i(n_k),n_k}$ are independent, it is easy to show that the probability mass of $\mathcal{L}(W_{i(n_k),n_k})$ escapes to infinity and consequently so does the probability mass of W_{n_k}. This contradicts the proper convergence of W_n. Thus $\sum_{i=1}^{n} p_{in}$ must remain bounded. Next, by observing that

$$(52) \qquad \theta_k^{(n)} \leq \binom{m}{k-1} \theta_1^{(n)} = \binom{m}{k-1} \sum_{i=1}^{n} p_{in}$$

it follows that the $\gamma_i^{(n)}$ are also bounded. Hence by the same argument as above $P(Z=0) > 0$ and it also follows that $\gamma_i^{(n)}$ converges for $i = 1,2, \ldots, m+1$. Consequently, $\theta_k^{(n)}$ converges for $k = 1,2, \ldots, m+1$ and this proves the theorem.

We shall now derive (35). The derivation is based on the repeated application of a simple but effective device with which we have derived (4.13) of Chapter II. The device can be described as follows. Let $X_1, X_2, \ldots, X_\ell$ be Bernoulli random variables. Then

$$(53) \quad f(Y + \sum_{i=1}^{\ell} X_i) - f(Y) = \sum_{k=1}^{\ell} [f(Y + \sum_{i=1}^{k} X_i) - f(Y + \sum_{i=1}^{k-1} X_i)]$$

$$= \sum_{k=1}^{\ell} X_k [f(Y + \sum_{i=1}^{k-1} X_i + 1) - f(Y + \sum_{i=1}^{k-1} X_i)]$$

We shall from now on omit the subscript n for brevity and write X_i for X_{in} and p_i for p_{in}. As before, let I be uniformly distributed on $\{1,2, \ldots , n\}$ and be independent of $X_1, X_2, \ldots , X_n$ and let

$$
(54) \qquad \begin{cases} W = \sum_{i=1}^{n} X_i \\ W^* = \sum_{|i-I| > m} X_i \\ W^{(I)} = \sum_{i \neq I} X_i \\ X_i = 0 \quad \text{for} \quad i \leq 0 \quad \text{or} \quad i \geq n+1 \\ Y_i = W^* + \sum_{k=I-m}^{i} X_k \\ Y_i^{(I)} = W^* + \sum_{\substack{k=I-m \\ k \neq I}}^{i} X_k \end{cases}
$$

Then

$$
(55) \qquad \begin{cases} Y_i = Y_{I+m} = W \quad \text{for} \quad i \geq n+1 \\ Y_i^{(I)} = Y_{I+m}^{(I)} = W^{(I)} \quad \text{for} \quad i \geq n+1 \\ Y_i = Y_{I-m-1} = W^* \quad \text{for} \quad i \leq 0 \\ Y_i^{(I)} = Y_{I-m-1}^{(I)} = W^* \quad \text{for} \quad i \leq 0 \end{cases}
$$

Also define an operator Δ by

$$
(56) \qquad \Delta f(x) = f(x+1) - f(x)
$$

and let

$$
(57)\quad \begin{cases} \xi(i_1,i_2,\ldots,i_k) = EX_{i_1}X_{i_2}\cdots X_{i_k} \\ \xi_1 = n(E\xi(I))\,Ef(W+1) \\ \xi_{k+1} = n\Big(E \sum_{i_1=I+1}^{I+m} \sum_{\substack{i_2=i_1-m+k-2 \\ i_2\neq I}}^{i_1-1} \sum_{\substack{i_3=i_1-m+k-3 \\ i_3\neq I}}^{i_2-1} \cdots \sum_{\substack{i_k=i_1-m \\ i_k\neq I}}^{i_{k-1}-1} \xi(I,i_1,\ldots,i_k)\Big) \\ \qquad\qquad \times (E\Delta^k f(W+1)) \\ \qquad +n\Big(E \sum_{i_1=I-m+k-1}^{I-1} \sum_{i_2=I-m+k-2}^{i_1-1} \sum_{i_3=I-m+k-3}^{i_2-1} \cdots \sum_{i_k=I-m}^{i_{k-1}-I} \xi(I,i_1,\ldots,i_k)\Big) \\ \qquad\qquad \times (E\Delta^k f(W+1)) \qquad \text{for } k = 1,2,\ldots . \end{cases}
$$

In deriving (35), we shall use R to denote any finite collection (independent of n) of terms which are bounded by $C(\theta_1^{(n)}, m, \|f\|) \max_{1\le i\le n} p_{in}$. The contributors to R will be terms of the forms

$$
(58)\qquad nE \sum_{(i_1,i_2,\ldots,i_k)\in A_I} X_I X_{i_1}\cdots X_{i_k}\,\Delta^{\ell} f(Y+1)
$$

and

$$
(59)\qquad nE \sum_{(i_1,i_2,\ldots,i_k)\in A_I} \xi(I,i_1,\ldots,i_k) \sum_{k\in B_I} X_k \Delta^{\ell} f(Y+1)
$$

where Y is the sum of a subset of $X_1, X_2, \ldots, X_n$, A_I and B_I are

finite sets (independent of n) given each I and at least two of $X_I, X_{i_1}, \ldots, X_{i_k}$ in (58) are independent given each I. In particular, the last condition is satisfied if $I \neq i_1 \neq \cdots \neq i_k$ and $k \geq m+1$. Thus we have for every real-valued bounded function f defined on $\{0,1,2,\ldots\}$,

$$
\begin{aligned}
(60) \qquad Ewf(W) &= nE[(E^{X_1,\ldots,X_n}{}_{X_I})\, f(W)] = nE[E^{X_1,\ldots,X_n}(X_I f(W))] = nEX_I f(W) \\
&= nEX_I[f(W^{(I)}+1) - f(W^*+1)] + nEp_I[f(W^*+1) - f(W+1)] \\
&\quad + \theta_1^{(n)} Ef(W+1) \\
&= nE \sum_{\substack{i=I-m \\ i\neq I}}^{I+m} X_I X_i \Delta f(Y_{i-1}^{(I)}+1) + nE \sum_{i=I-m}^{I+m} p_I X_i \Delta f(Y_{i-1}+1) \\
&\quad + \theta_1^{(n)} Ef(W+1) \\
&= nE \sum_{\substack{i=I-m \\ i\neq I}}^{I+m} X_I X_i \Delta f(Y_{i-1}^{(I)}+1) + \xi_1 + R
\end{aligned}
$$

As before the first three equalities of (60) follow from the properties of conditional expectations, the fourth equality from the independence of I and W and the conditional independence of X_I and W^* given I, the fifth equality from (53) and the last equality from (57) and (59). We sahll from now on continue with the derivation without making such references. Accordingly,

$$(61)\quad nE \sum_{\substack{i=I-m \\ i\neq I}}^{I+m} X_I X_i \Delta f(Y_{i-1}^{(I)}+1)$$

$$= nE \sum_{i=I+1}^{I+m} X_I X_i \Delta f(Y_{i-1}^{(I)}+1) + nE \sum_{i=I-m}^{I-1} X_I X_i \Delta f(Y_{i-1}^{(I)}+1)$$

$$= nE \sum_{i=I+1}^{I+m} X_I X_i \left\{\Delta f(W^*+1 + \sum_{\substack{k=I-m \\ k\neq I}}^{i-1} X_k) - \Delta f(W^*+1)\right\}$$

$$+ nE \sum_{i=I+1}^{I+m} X_I X_i \left\{\Delta f(W^*+1) - \Delta f(W^*+1 - \sum_{k=I+m+1}^{i+m} X_k)\right\}$$

$$+ nE \sum_{i=I-m}^{I-1} X_I X_i \left\{\Delta f(Y_{i-1}^{(I)}+1) - \Delta f(Y_{i-1}^{(I)}+1 - \sum_{k=i-m}^{i-1} X_k)\right\}$$

$$+ nE \sum_{i=I+1}^{I+m} \xi(I,i) \left\{\Delta f(W^*+1- \sum_{k=I+m+1}^{i+m} X_k) - \Delta f(W+1)\right\}$$

$$+ nE \sum_{i=I-m}^{I-1} \xi(I,i) \left\{\Delta f(Y_{i-1}^{(I)}+1- \sum_{k=i-m}^{i-1} X_k) - \Delta f(W+1)\right\}$$

$$+ n[E \sum_{i=I+1}^{I+m} \xi(I,i)][E\Delta f(W+1)] + n[E \sum_{i=I-m}^{I-1} \xi(I,i)][E\Delta f(W+1)]$$

$$= nE \sum_{i_1=I+1}^{I+m} \sum_{\substack{i_2=I-m \\ i_2\neq I}}^{i_1-1} X_I X_{i_1} X_{i_2} \Delta^2 f(W^*+1+ \sum_{\substack{k=I-m \\ k\neq I}}^{i_2-1} X_k)$$

$$+ nE \sum_{i_1=I-m}^{I-1} \sum_{i_2=i_1-m}^{i_1-1} X_I X_{i_1} X_{i_2} \Delta^2 f(Y_{i_1-1}^{(I)}+1- \sum_{k=i_2}^{i_1-1} X_k) + \xi_2 + R$$

$$= nE \sum_{i_1=I+1}^{I+m} \sum_{\substack{i_2=i_1-m \\ i_2\neq I}}^{i_1-1} X_I X_{i_1} X_{i_2} \Delta^2 f(Y_{i_2-1}^{(I)}+1)$$

$$+ nE \sum_{i_1=I-m+1}^{I-1} \sum_{i_2=I-m}^{i_1-1} X_I X_{i_1} X_{i_2} \Delta^2 f(Y_{i_2-1}^{(I)}+1) + \xi_2 + R$$

$$= nE \sum_{i_1=I+1}^{I+m} \sum_{\substack{i_2=i_1-m \\ i_2\neq I}}^{i_1-1} X_I X_{i_1} X_{i_2} \left\{\Delta^2 f(W^*+1+ \sum_{\substack{k=I-m \\ K\neq I}}^{i_2-1} X_k) - \Delta^2 f(W^*+1)\right\}$$

$$+ nE \sum_{i_1=I+1}^{I+m} \sum_{\substack{i_2=i_1-m \\ i_2\neq I}}^{i_1-1} X_I X_{i_1} X_{i_2} \left\{\Delta^2 f(W^*+1) - \Delta^2 f(W^*+1- \sum_{k=I+m+1}^{i_1+m} X_k - \sum_{k=i_2-m}^{I-m-1} X_k)\right\}$$

$$+ nE \sum_{i_1=I-m+1}^{I-1} \sum_{i_2=I-m}^{i_1-1} X_I X_{i_1} X_{i_2} \left\{\Delta^2 f(Y_{i_2-1}^{(I)}+1) - \Delta^2 f(Y_{i_2-1}^{(I)}+1- \sum_{k=i_2-m}^{i_2-1} X_k)\right\}$$

$$+ \xi_2 + \xi_3 + R$$

$$= nE \sum_{i_1=I+1}^{I+m} \sum_{\substack{i_2=i_1-m+1 \\ i_2\neq I}}^{i_1-1} \sum_{\substack{i_3=i_1-m \\ i_3\neq I}}^{i_2-1} X_I X_{i_1} X_{i_2} X_{i_3} \Delta^3 f(Y_{i_3-1}^{(I)}+1)$$

$$+ nE \sum_{i_1=I-m+2}^{I-1} \sum_{i_2=I-m+1}^{i_1-1} \sum_{i_3=I-m}^{i_2-1} X_I X_{i_1} X_{i_2} X_{i_3} \Delta^3 f(Y_{i_3-1}^{(I)}+1)$$

$$+ \xi_2 + \xi_3 + R$$

Continuing in this manner, we obtain

$$(62) \quad nE \sum_{\substack{i_1=I-m \\ i\neq I}}^{I+m} X_I X_i \Delta f(Y_{i-1}^{(I)}+1) = \xi_2 + \xi_3 + \cdots + \xi_{m+1} + R$$

This together with (60) yield

$$(63) \quad EWf(W) = \xi_1 + \xi_2 + \cdots + \xi_{m+1} + R$$

Now it is not difficult to check that

$$(64)\quad \theta_{k+1}^{(n)} = nE \sum_{i_1=I-m+k-1}^{I-1} \sum_{i_2=I-m+k-2}^{i_1-1} \sum_{i_3=I-m+k-3}^{i_2-1} \cdots \sum_{i_k=I-m}^{i_{k-1}-1} \xi(I,i_1,\ldots,i_k)$$

and that for every $\ell = 1,2,\ldots,k\ (\leq m)$

$$(65)\quad \theta_{k+1}^{(n)}$$

$$= nE \sum_{i_1=I+\ell}^{I+m-k+\ell} \sum_{i_2=I+\ell-1}^{i_1-1} \cdots \sum_{i_\ell=I+1}^{i_{\ell-1}-1} \sum_{i_{\ell+1}=i_1-m+k-\ell-1}^{I-1} \sum_{i_{\ell+2}=i_1-m+k-\ell-2}^{i_{\ell+1}-1} \cdots \sum_{i_k=i_1-m}^{i_{k-1}-1} \xi(I,i_1,\ldots,i_k)$$

Thus

$$(66)\quad nE \sum_{i_1=I+1}^{I+m} \sum_{\substack{i_2=i_1-m+k-2 \\ i_2\neq I}}^{i_1-1} \sum_{\substack{i_3=i_1-m+k-3 \\ i_3\neq I}}^{i_2-1} \cdots \sum_{\substack{i_k=i_1-m \\ i_k\neq I}}^{i_{k-1}-1} \xi(I,i_1,\ldots,i_k)$$

$$= nE \sum_{i_1=I+1}^{I+m-k+1} \sum_{i_2=i_1-m+k-2}^{I-1} \sum_{i_3=i_1-m+k-3}^{i_2-1} \cdots \sum_{i_k=i_1-m}^{i_{k-1}-1} \xi(I,i_1,\ldots,i_k)$$

$$+nE \sum_{i_1=I+m-k+2}^{I+m} \sum_{\substack{i_2=i_1-m+k-2 \\ i_2\neq I}}^{i_1-1} \sum_{\substack{i_3=i_1-m+k-3 \\ i_3\neq I}}^{i_2-1} \cdots \sum_{\substack{i_k=i_1-m \\ i_k\neq I}}^{i_{k-1}-1} \xi(I,i_1,\ldots,i_k)$$

$$+nE \sum_{i_1=I+2}^{I+m-k+1} \sum_{i_2=I+1}^{i_1-1} \sum_{\substack{i_3=i_1-m+k-3 \\ i_3\neq I}}^{i_2-1} \cdots \sum_{\substack{i_k=i_1-m \\ i_k\neq I}}^{i_{k-1}-1} \xi(I,i_1,\ldots,i_k)$$

$$=\theta_{k+1}^{(n)}$$

$$+nE \sum_{i_1=I+2}^{I+m} \sum_{i_2=I+1}^{i_1-1} \sum_{\substack{i_3=i_1-m+k-3 \\ i_3\neq I}}^{i_2-1} \cdots \sum_{\substack{i_k=i_1-m \\ i_k\neq I}}^{i_{k-1}-1} \xi(I,i_1,\ldots,i_k)$$

$$=\theta_{k+1}^{(n)}$$

$$+nE \sum_{i_1=I+2}^{I+m-k+2} \sum_{i_2=I+1}^{i_1-1} \sum_{i_3=i_1-m+k-3}^{I-1} \sum_{i_4=i_1-m+k-4}^{i_3-1} \cdots \sum_{i_k=i_1-m}^{i_{k-1}-1} \xi(I,i_1,\ldots,i_k)$$

$$+nE \sum_{i_1=I+m-k+3}^{I+m} \sum_{i_2=I+1}^{i_1-1} \sum_{\substack{i_3=i_1-m+k-3 \\ i_3\neq I}}^{i_2-1} \sum_{\substack{i_4=i_1-m+k-4 \\ i_4\neq I}}^{i_3-1} \cdots \sum_{\substack{i_k=i_1-m \\ i_k\neq I}}^{i_{k-1}-1} \xi(I,i_1,\ldots,i_k)$$

$$+nE \sum_{i_1=I+2}^{I+m-k+2} \sum_{i_2=I+1}^{i_1-1} \sum_{i_3=I+1}^{i_2-1} \sum_{\substack{i_4=i_1-m+k-4 \\ i_4\neq I}}^{i_3-1} \cdots \sum_{\substack{i_k=i_1-m \\ i_k\neq I}}^{i_{k-1}-1} \xi(I,i_1,\ldots,i_k)$$

$$=2\theta_{k+1}^{(n)}$$

$$+nE \sum_{i_1=I+3}^{I+m} \sum_{i_2=I+2}^{i_1-1} \sum_{i_3=I+1}^{i_2-1} \sum_{\substack{i_4=i_1-m+k-4 \\ i_4\neq I}}^{i_3-1} \cdots \sum_{\substack{i_k=i_1-m \\ i_k\neq I}}^{i_{k-1}-1} \xi(I,i_1,\ldots,i_k)$$

$$= k\theta_{k+1}^{(n)} \quad \text{by induction}$$

Finally, the substitution of (64) and (66) into (63) yields

$$(67)\quad EWf(W) = \sum_{k=1}^{m+1} k\theta_k^{(n)} E\Delta^{k-1} f(W+1) + R$$

$$= \sum_{k=1}^{m+1} \sum_{i=1}^{k} (-1)^{k-i} \binom{k-1}{i-1} k\theta_k^{(n)} Ef(W+i) + R$$

$$= \sum_{i=1}^{m+1} \sum_{k=i}^{m+1} (-1)^{k-1} \binom{k}{i} i\theta_k^{(n)} Ef(W+i) + R^R$$

$$= \sum_{i=1}^{m+1} i\gamma_i^{(n)} Ef(W+i) + R$$

and this completes the proof of Theorem 4.1.

It is possible to extend without much effort Theorem 4.1 to the case where m increases with n but at a much slower rate. It also seems very likely that an analogous result should hold under certain mixing conditions. A self-contained proof of Theorem 4.1, one without using the Classical Limit Theorem, can be achieved by proving

$$(68)\qquad \gamma_i^{(n)} \geq -C(\theta_1^{(n)}, m) \max_{1 \leq i \leq n} p_{in}$$

where $C(\theta_1^{(n)}, m)$ is a positive constant depending on $\theta_1^{(n)}$ and m, and remaining bounded as long as $\theta_1^{(n)}$ and m are bounded. The author believes that (68) is true but has not been able to prove it. However, at the moment, the author is satisfied with the present result and its proof.

REFERENCES

1. L. Breiman, Probability. Addison-Wesley, Reading (1968).

2. W. Feller, An introduction to Probability theory and its applications. Vol. II (1st edition), Wiley, New York (1966).

3. J. L. Hodges and Le Cam, The Poisson approximation to the Poisson binomial distribution. Ann. Math. Stat. 31 (1960), 737-740.

4. J. Kerstan, Verallgemeinerung eines Satzes von Prochorow und Le Cam. Z. Wahrscheinlichkeitstheorie verw. Geb. 2 (1964), 173-179.

5. L. Le Cam, An approximation theorem for the Poisson binomial distribution. Pacific J. Math. 10 (1960), 1181-1197.

6. W. Philipp, The central limit problem for mixing sequences of random variables. Z. Wahrscheinlichkeitstheorie verw. Geb. 12 (1969), 155-171.

7. Ju. V. Prohorov, Asymptotic behavior of the binomial distribution (Russian). Uspehi Matematiceskih Nauk 8 (1953), 3 : 135-142.

8. C. Stein, A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. Proceedings of Sixth Berkeley Symposium on Math. Stat. and Probability (1970) (to appear).

9. W. Vervaat, Upper bounds for the distance in total variation between the binomial or negative binomial and the Poisson distribution. Statistica Nurlandica 23 (1969), 79-86.