

Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be “used for any purpose other than private study, scholarship, or research.” If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of “fair use” that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation

Printing note: If you do not wish to print this page, then select “Pages from: first page # to: last page #” on the print dialog screen

The Van Houten library has removed some of the personal information and all signatures from the approval page and biographical sketches of theses and dissertations in order to protect the identity of NJIT graduates and faculty.

ABSTRACT

APPLICATIONS OF BINARY SEQUENCE OF ORDER k

by
Xulun Jiang

The cumulative distribution of the finite sum of the binary sequence of order k is studied and some of its applications discussed. Certain properties of this sequence are studied and uniformly superior bounds for the cumulative distribution under minimal information on the "success" probabilities are derived.

As an application, an optimal randomized response model to collect sensitive information with dependence in the sample is proposed. This dependence is caused by untruthful response to stigmatizing questions and has been ignored in the past procedures.

The proposed method is useful in collecting reliable information in situations where the response is difficult to get, e.g., gathering data regarding the incidence of AIDS.

APPLICATIONS OF
BINARY SEQUENCE OF ORDER k

by
Xulun Jiang

RECEIVED BY THE LIBRARY OF THE
NEW JERSEY INSTITUTE OF TECHNOLOGY
ON MAY 11, 1994

A Thesis
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of Requirements for the Degree of
Master of Science in Applied Mathematics

Department of Mathematics

May 1994

APPROVAL PAGE

APPLICATIONS OF
BINARY SEQUENCE OF ORDER k

Xulun Jiang

Dr. Sunil K. Dhar, Thesis Adviser
Assistant Professor of Applied Mathematics, NJIT

Date

Dr. Manish Bhattacharjee, Committee Member
Associate Professor of Applied Mathematics, NJIT

Date

Dr. Ponniah Elancheran, Committee Member
Associate Professor of Applied Mathematics, NJIT

Date

BIOGRAPHICAL SKETCH

Author: Xulun Jiang

Degree: Master of Science in Applied Mathematics

Date: May 1994

Undergraduate and Graduate Education:

- Master of Science in Applied Mathematics,
New Jersey Institute of Technology, Newark, NJ, 1994
- Bachelor of Science in Computational Mathematics,
Fudan University, Shanghai, P. R. China, 1982

Major: Applied Mathematics

ACKNOWLEDGMENT

The author wishes to express his sincere gratitude to his supervisor, Dr. Sunil Dhar, for his guidance, friendship, and moral support throughout this research.

Special thanks to Professors Manish Bhattacharjee and Ponniah Elancheran for serving as members of the committee.

The author is grateful to the Department of Mathematics of New Jersey Institute of Technology for the support throughout this research.

And finally, a thank you to the Department members for their timely and expert assistance.

This work was supported, in part, by the New Jersey Institute of Technology under Grant No. SBR 1993–94, #421740.

TABLE OF CONTENTS

Chapter	Page
1 INTRODUCTION	1
2 PROBABILITY BOUNDS ON THE FINITE SUM	3
2.1 Some Properties of the Sequence	3
2.2 Uniformly Tight Bounds	5
2.3 Numerical Examples	11
3 MODIFIED RANDOMIZED RESPONSE MODEL	13
3.1 Introduction and Summary	13
3.2 Properties of the Estimator	15
3.3 Central Limit Theorem	16
APPENDIX	21
REFERENCE	32

CHAPTER 1

INTRODUCTION

The sum of independent, identically or non identically distributed binomial random variables is one of the oldest random variable in probability and statistics. Estimation of their cumulative distribution functions has been studied extensively by Kolmogorov (1956), Hoeffding (1956), Anderson and Samuels (1956), Hodges and Le Cam (1960) and Gastwirth (1977). Percus and Percus (1985) obtained uniformly superior bounds for the sum of independent, non identically distributed random variables with minimal information on the underlying "success" probabilities $\{p_1, p_2, \dots\}$. In the present work, we study sum of a particular type of dependent, non identically distributed random variables. This sum is defined in terms of a binary distribution of order k , given by Aki (1985). We also obtain uniformly superior bounds for the cumulative distribution under minimal information on the "success" probabilities as in Percus and Percus (1985). However, the mathematical problem encountered in the present work, and hence the solution, turn out to be entirely different from theirs.

It is also noted that the optimal upper bound for the distribution of this sum is independent of k . Further, if the p 's are close to zero then the upper bound will be close to the true value.

Further study in this paper shows the asymptotic results of the binary sequence of order k and these results can be applied to the procedure for collecting sensitive information.

Definition 1.1 Let X_i , $i = 0, 1, 2, \dots$ be a sequence of $\{0, 1\}$ – valued random variables defined on a probability space (Ω, \mathcal{F}, P) . Then, this sequence $\{X_i\}$ is said to be a binary sequence of order k if there exist a positive integer k

and real numbers $0 \leq p_1, p_2, \dots, p_k \leq 1$ such that

$$(i) X_0 = 0,$$

$$(ii) P(X_n = 1 \mid X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}) = p_j$$

is satisfied for any positive integer n , where $j - 1 = (r-1) \bmod k$, i.e., the remainder in the division of $(r-1)$ by k and r is the smallest positive integer which satisfies $x_{n-r} = 0$. Here, the case $k = 1$ corresponds to the sequence of independent identically distributed Bernoulli random variables and $k = \infty$ to that of a renewal sequence. The latter follows from the fact that each time a failure occurs, the process starts again from the beginning and is independent of all the preceding events.

In addition to the notations in Definition 1.1, $q_i = 1 - p_i$, $i = 1, 2, \dots, k$, from here onwards. Also, we shall follow the convention that any product or sum over an empty set is one or zero, respectively.

To introduce the practical background of this distribution, we repeat the Example 2.2 from Aki (1985) here. In chapter 3 we will show another application of the sum of first n terms of the binary sequence in the field of non evasive sample survey.

Example An electric bulb is lighted and checked daily at a given time. Based on the result of the check on the i^{th} day X_i takes the value 0 if the bulb has burnt out, or 1 if it is working. A burnt out bulb is replaced by a new one immediately. A new bulb is replaced after k consecutive days, even if it is still working. Here p_i represents the probability of that the bulb will work on the i^{th} day, given it has not failed for the past $i - 1$ days. Then $\{X_i\}$ is a binary sequence of order k and the sum of its first n variables represents the total number of days, out of n , when the bulb was working.

CHAPTER 2

PROBABILITY BOUNDS ON THE FINITE SUM

2.1 Some Properties of the Sequence

In this section we present certain conditional and joint distributions of a binary sequence. We also forward a new approach to the binary sequence of order k , which avoids the use of r and j of Definition 1.1. We begin by restricting ourselves to binary sequence of order $k \geq n \geq 1$. Here, n is the sample size, i.e., the first n realizations of the binary sequence of order k . Let 0 represent the failure of a light bulb and 1 be its state of functioning, i.e., not failure. Since p^x vanishes as a factor in a product of terms when $x = 0$, and is p when $x = 1$, let x 's be the state of working of the light bulbs in the following discussion. Thus, $P\left(X_n = x_n \mid X_1 = x_1, X_2 = x_2, \dots, X_{n-1} = x_{n-1}\right)$ could be any one of the p_1, \dots, p_n or q_1, \dots, q_n depending on whether $x_n = 1$ or 0, respectively. Let j be as in Definition 1.1, which satisfies $1 \leq j \leq n$. In the case $n \leq k$, j equals r because $r - 1 < k$. This conditional probability, for a specific $i = r$, is equal to

$$p_i^{(1-x_{n-i}) \prod_{t=n-i+1}^n x_t} q_i^{(1-x_{n-i})(1-x_n) \prod_{t=n-i+1}^{n-1} x_t}.$$

However, i need not be r and could be any integer in $[1, n]$. Hence, this conditional probability is equal to

$$(2.1) \quad \prod_{i=1}^n p_i^{(1-x_{n-i}) \prod_{t=n-i+1}^n x_t} q_i^{(1-x_{n-i})(1-x_n) \prod_{t=n-i+1}^{n-1} x_t}.$$

The joint density of (X_1, X_2, \dots, X_n) in the case $n \leq k$ can now be obtained from the fact that $P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$

$$(2.2) \quad = P(X_n = x_n \mid X_1 = x_1, X_2 = x_2, \dots, X_{n-1} = x_{n-1}) \\ \times P(X_{n-1} = x_{n-1} \mid X_1 = x_1, X_2 = x_2, \dots, X_{n-2} = x_{n-2}) \dots$$

$$\times P(X_2 = x_2 | X_1 = x_1) P(X_1 = x_1).$$

This gives the joint density as

$$\prod_{i=1}^n p_i \sum_{m=1}^n (1-x_{m-i}) \prod_{t=m-i+1}^m x_t q_i \sum_{m=1}^n (1-x_{m-i})(1-x_m) \prod_{t=m-i+1}^{m-1} x_t.$$

The above results for the conditional probability and joint density can be derived without the restriction $k \geq n \geq 1$ by defining a function $S_k(t;n-1)$ as follows

$$(2.3) \quad \begin{cases} \sum_{i=0}^{\lfloor (n-t)/k \rfloor} (1-x_{n-ik-t}) \prod_{l=n-ik-t+1}^{n-1} x_l, & t \leq n, \\ 0, & t > n, \end{cases}$$

where $0 < t \leq k$ and S_k depends on the first $n-1$ random variables of the binary sequence. In general, the conditional and the joint distributions in (2.1) and (2.2) can now be described by the following lemma and theorem, respectively. For the proof of the lemma please see Appendix.

Lemma 2.1 Let $\{X_i\}$ be a binary sequence of order k with p_1, p_2, \dots, p_k , then

$$\begin{aligned} & P\left(X_n = x_n \mid X_1 = x_1, X_2 = x_2, \dots, X_{n-1} = x_{n-1}\right) \\ &= \prod_{t=1}^k \left\{ p_t^{S_k(t;n-1)x_n} q_t^{S_k(t;n-1)(1-x_n)} \right\}. \end{aligned}$$

The proof of the following theorem follows from Lemma 2.1 and (2.3).

Theorem 2.1 The density function of the binary sequence of order k with p_1, p_2, \dots, p_k is

$$f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) = \prod_{j=1}^k \left\{ p_j^{\sum_{m=1}^n S_k(j;m-1)x_m} q_j^{\sum_{m=1}^n S_k(j;m-1)(1-x_m)} \right\}.$$

In order to state and prove the following theorem we need the notation $(a)_k$

to be the remainder in the division of a by k , i.e., $(a)_k = a - [a/k]k$, $a, k \in$ integers.

Theorem 2.2 Let $\{X_i, i \geq 1\}$ be a binary sequence of order k . Define $b = n - y$, $\Delta j_m = j_m - j_{m-1}$, $j_0 = 0$, $p_0 = 1$. Then the density of $\sum_{i=1}^n X_i$ is given by

$$P\left(\sum_{i=1}^n X_i = y\right) = \sum_{0 < j_1 < j_2 < \dots < j_b \leq n} \left(\prod_{m=1}^b \left((p_0 p_1 \dots p_k)^{[(\Delta j_m - 1)/k]} p_0 p_1 \dots \right. \right. \\ \left. \left. \times P_{(\Delta j_m - 1)_k} Q_{(\Delta j_m - 1)_k + 1} \right) \right) (p_0 p_1 \dots p_k)^{[(n - j_b)/k]} p_0 p_1 \dots p_{(n - j_b)_k}.$$

2.2. Uniformly Tight Bounds

Let $\{X_i, i \geq 1\}$ be a binary sequence of order k with parameters p_1, p_2, \dots, p_k .

Denote $F_n(a) = P(\sum_{i=1}^n X_i \leq a)$, where a is a nonnegative integer. In order to

present the results of this section need the following notation. When $p_{m+1} = p_{m+2} = \dots = p_k = 1$ or $p_{m+1} = p_{m+2} = \dots = p_k = 0$, $1 \leq m \leq k$, in $F_n(a)$, then denote it by $\tilde{F}_n(a)$ or $\tilde{\tilde{F}}_n(a)$. Even though intuitively $\tilde{F}_n(a)$ and $\tilde{\tilde{F}}_n(a)$ seem to be the lower and upper bounds of $F_n(a)$, the proof of this is not so obvious.

Theorem 2.3 For a binary sequence of order k with given p_1, p_2, \dots, p_m ($m < k$), $F_n(a)$ reaches its maximum or minimum when $p_{m+1} = p_{m+2} = \dots = p_k = 0$ or $p_{m+1} = p_{m+2} = \dots = p_k = 1$, respectively.

For the proof of this theorem please see Appendix.

Throughout the following work the definition of combinations $\binom{m}{n}$ is $m(m-1)\dots(m-n+1)/n!$.

Case (i) $F_n(0)$ is given.

Since, $F_n(0) = P(\sum_{i=1}^n X_i = 0) = q_1^n$, if $F_n(0)$ is known then so is $q_1 =$

$\sqrt[n]{F_n(0)}$ and therefore, $k > m = 1$, and $\tilde{F}_n(a)$, $\tilde{\tilde{F}}_n(a)$, as in Theorem 2.3., are the lower and upper bounds of $F_n(a)$, respectively, when $F_n(0)$ is known.

In order to compute $\tilde{F}_n(a)$, we first compute the $P(\sum_{i=1}^n X_i = y)$, where $p_i = 1$, $i = 2, 3, \dots, k$. In the event $\left\{ \sum_{i=1}^n X_i = y \right\}$ there are y number of random variables in the set $\{X_i, 1 \leq i \leq n\}$ with value 1. If for some $i \geq 0$, $X_i = 0$ and $X_{i+1} = 1$ then $X_{i+2} = \dots = X_{(i+k) \wedge n} = 1$, where \wedge represents minimum. This implies that these y ones must appear in groups of size k except for the last $(y)_k$ ones. The probability of a typical outcome of the type $[y/k]$ bunches of ones, each of size k , $n - y$ zeros and $(y)_k$ ones is

$$(2.4) \quad (p_0 p_1 \dots p_k)^{[y/k]} (p_0 p_1 \dots p_{(y)_k}) q_1^{n-y}.$$

If y is a multiple of k then $[y/k] = [(y-1)/k] + 1$, otherwise, $[y/k] = [(y-1)/k]$. Since $p_0 = p_2 = p_3 = \dots = p_k = 1$, therefore (2.4) becomes $p_1^{[(y-1)/k]+1} q_1^{n-y}$.

The number of all outcomes in the event $\sum_{i=1}^n X_i = y$, which are described above

and are equally probable, is $\binom{n-y+[y/k]}{[y/k]}$. Therefore, $P(\sum_{i=1}^n X_i = y) = \binom{n-y+[y/k]}{[y/k]} p_1^{[(y-1)/k]+1} q_1^{n-y}$. Thus, the lower bound of $F_n(a)$ is

$$\tilde{F}_n(a) = \sum_{y=0}^a \binom{n-y+[y/k]}{[y/k]} p_1^{[(y-1)/k]+1} q_1^{n-y}.$$

Interestingly, in the case when $k \rightarrow \infty$, $\tilde{F}_n(a)$ converges to the distribution

function evaluated at a of the density function $f(y) = \begin{cases} q_1^n & y = 0 \\ p_1 q_1^{n-y} & 1 \leq y \leq n \end{cases}$.

For computing $\tilde{\tilde{F}}_n(a)$, the condition $p_2 = 0$ means that in the binary sequence no consecutive random variables will take the value 1. The event

$\left\{ \sum_{i=1}^n X_i = y \right\}$ is equal to the disjoint union of $\left\{ \sum_{i=1}^n X_i = y, X_n = 1 \right\}$ and

$\left\{ \sum_{i=1}^n X_i = y, X_n = 0 \right\}$. However, in view of the preceding statement the events $\left\{ \sum_{i=1}^n X_i = y, X_n = 1 \right\}$ and $\left\{ \sum_{i=1}^n X_i = y, X_n = 0 \right\}$ satisfy $2y - 1 \leq n$ and $2y \leq n$, respectively.

We shall now compute the probability of $\left\{ \sum_{i=1}^n X_i = y, X_n = 1 \right\}$, where $X_{n-1} = 0$. In this joint event, $2(y-1)$ of X_1, \dots, X_n are such that a zero follows one, except for $X_n = 1$ and the remaining X 's take the value zero. Thus, $2(y-1) + 1$ must be at the most n , which puts the constraint $1 \leq y \leq [(n+1)/2]$ on y . Any outcome of this type must have its probability equal to

$$(2.5) \quad (p_1 q_2)^{y-1} q_1^{n-2y+1} p_1.$$

However, (2.5) becomes $p_1^y q_1^{n-2y+1}$ since $q_2 = 1$. Consider the one followed by a zero as a single piece of a special zero and ignore the last $X_n = 1$. Now there are a total of $y - 1$ special zeros and $(n-y) - (y-1)$ single zeros. Therefore, there are $\binom{n-y}{y-1}$ distinct equally probable outcomes in this case. Thus,

$$(2.6) \quad P\left(\sum_{i=1}^n X_i = y, X_n = 1\right) = \binom{n-y}{y-1} p_1^y q_1^{n-2y+1}, \quad 1 \leq y \leq [(n+1)/2].$$

Similarly,

$$(2.7) \quad P\left\{\sum_{i=1}^n X_i = y, X_n = 0\right\} = \binom{n-y}{y} p_1^y q_1^{n-2y}, \quad 0 \leq y \leq [n/2].$$

By adding up (2.6) and (2.7), we get the upper bound of $F_n(a)$ to be

$$\tilde{F}_n(a) = \sum_{y=1}^{a^*} \binom{n-y}{y-1} p_1^y q_1^{n-2y+1} + \sum_{y=0}^{a^{**}} \binom{n-y}{y} p_1^y q_1^{n-2y},$$

where $a^* = \min(a, [(n+1)/2])$ and $a^{**} = \min(a, [n/2])$.

Case (ii) Both $F_n(0)$ and $F_n(1)$ are given.

The bounds for $F_n(a)$ are computed using Theorem 2.3 with $m = 2 < k$.

Since,

$$P\left(\sum_{i=1}^n X_i = 1\right) = P(X_n = 1, X_i = 0 \text{ for all } i \neq n)$$

$$+ P(X_n = 0, \sum_{i=1}^n X_i = 1) = q_1^{n-1} p_1 + (n-1)q_1^{n-2} p_1 q_2, \text{ we get}$$

$$F_n(1) = F_n(0) + P\left(\sum_{i=1}^n X_i = 1\right) = q_1^{n-1} + (n-1)q_2(q_1^{n-2} - q_1^{n-1}).$$

Assuming $0 < q_1 < 1$, we get

$$(2.8) \quad q_2 = (F_n(1) - q_1^{n-1}) / [(n-1)(q_1^{n-2} - q_1^{n-1})], p_2 = 1 - q_2.$$

From Case (i) and (2.8), p_1 and p_2 are fixed because $F_n(0)$ and $F_n(1)$ are given. Thus, n must be at least 2 to compute the upper and lower bound for $F_n(a)$.

As in the previous section, for calculating the lower bound, we compute

$P\left(\sum_{i=1}^n X_i = y\right)$ first. For $p_i = 1$, $i = 3, 4, \dots, k$, if $X_i = 0$ and $X_{i+1} = X_{i+2} = 1$ for some $i \geq 1$, then $X_{i+3} = \dots = X_{(i+k) \wedge n} = 1$. Also, if $X_i = 0$ and $X_{i+1} = X_{i+2} = \dots = X_{i+jk+2} = 1$ for some positive integer i and j , then $X_{i+jk+3} = X_{i+jk+4} = \dots = X_{i+jk+k} = 1$. Therefore, each realization of $\{X_i, 0 \leq i \leq n\}$ in the event $\left\{\sum_{i=1}^n X_i = y\right\}$ can be written as the union of four disjoint

distinct groups. Group 1 consists of M pieces of consecutive ones of size k , $0 \leq M \leq [y/k]$. Fixing the elements of Group 1, Group 2 is the set of consecutive ones of size l [$\leq (k-1) \wedge (y-Mk)$], located near the n^{th} position and Group 3 consists of J pieces of a one followed by a zero, i.e., 10. Here, $\sum_{i=1}^n X_i = y$ gives

J in terms of M and l through $Mk + J + l = y$. Having fixed Groups 1, 2, and 3, Group 4 consists of all the remaining zeros. The number of zeros in Group 4 must therefore be equal to $n - y - J$. When M and l are fixed, the probability

of a typical outcome of this kind in the event $\left\{\sum_{i=1}^n X_i = y\right\}$ is

$$(p_1 \cdots p_k)^M (p_1 q_2)^J p_0 p_1 \cdots p_l q_1^{n-y-J}.$$

Substitute J by $y - Mk - l$, and take $p_0 = p_3 = p_4 = \cdots = p_k = 1$, this probability becomes

$$p_1^{y-Mk+M-l+1+[(l-1)/k]} p_2^{M+1+[(l-2)/k]} q_2^{y-Mk-l} q_1^{n-2y+Mk+l}.$$

The number of distinct, equally probable outcomes of this type, with M and l fixed, is counted by ignoring the l fixed ones located near the n^{th} position. Thus, the number of all possible arrangements of $n-y+M$ pieces with M of them alike, as described in Group 1, $y-Mk-l$ in Group 3 and $n-2y+Mk+l$ of the pieces are zeros of Group 4. This number is equal to

$$(n-y+M)! / \{M!(y-Mk-l)!(n-2y+Mk+l)!\}, \text{ i.e., } \binom{n-y+M}{M+y-Mk-l} \binom{M+y-Mk-l}{M}.$$

$$\begin{aligned} \text{Therefore, } P\left(\sum_{i=1}^n X_i = y\right) &= \sum_{M=0}^{\lfloor y/k \rfloor} \sum_{l=0}^{l^*} \binom{n-y+M}{M+y-Mk-l} \binom{M+y-Mk-l}{M} \\ &\times p_1^{y-Mk+M-l+1+[(l-1)/k]} p_2^{M+1+[(l-2)/k]} q_2^{y-Mk-l} q_1^{n-2y+Mk+l}, \end{aligned}$$

where $l^* = \min(k-1, y-Mk)$. Thus, the lower bound of $F_n(a)$ is

$$(2.9) \quad \sum_{y=0}^a \sum_{M=0}^{\lfloor y/k \rfloor} \sum_{l=0}^{l^*} \binom{n-y+M}{M+y-Mk-l} \binom{M+y-Mk-l}{M} p_1^{y-Mk+M-l+1+[(l-1)/k]} \\ \times p_2^{M+1+[(l-2)/k]} q_2^{y-Mk-l} q_1^{n-2y+Mk+l}.$$

In this case, as $k \rightarrow \infty$, the limit of (2.9) is given by

$$\sum_{y=0}^a \sum_{l=0}^{l^*} \binom{n-y}{y-l} p_1^{y-l+1} p_2 q_2^{y-l} q_1^{n-2y+l}.$$

This limit can be obtained by just taking $k > n$ and observing that $M = \lfloor y/k \rfloor = \lfloor (l-1)/k \rfloor = \lfloor (l-2)/k \rfloor = 0$. Also, it has been computationally seen that the above limit of (2.9) is again a distribution function.

To compute the upper bound $\tilde{F}_n(a)$, note that the condition $p_3 = 0$ imposes the restriction that no three consecutive random variables are each equal to 1. Therefore, each realization of $\{X_i, 0 \leq i \leq n\}$ in the event $\left\{ \sum_{i=1}^n X_i = y \right\}$

can be written as the union of three disjoint distinct groups. Fix j elements in a group of the type double ones followed by a zero, i.e., 110; the next group consists of M elements of the type single one followed by a zero, i.e., 10; and the final group contains all the remaining zeros. The number of these remaining zeros is $n - y - M - j$. Further, consider the set $\{X_i, 0 \leq i \leq n\}$ as the union of three mutually disjoint sets with restrictions, one with $X_{n-1} = X_n = 1$, the second with $X_{n-1} = 0, X_n = 1$, and the third with $X_n = 0$. In each of these three cases, M is equal to $y - 2j - 2$, $y - 2j - 1$, or $y - 2j$, and j satisfies the restriction $0 \leq j \leq (y-2) \wedge [(n-2)/3] = j_1$, $0 \leq j \leq (y-1) \wedge [(n-1)/3] = j_2$ or $0 \leq j \leq y \wedge [n/3] = j_3$, respectively. Also, y satisfies the restriction $0 \leq y \leq n - [n/3]$, $0 \leq y \leq n - [(n+1)/3]$ or $0 \leq y \leq n - [(n+2)/3]$, because the least number of zeros we must have under each case is $[n/3]$, $[(n+1)/3]$ or $[(n+2)/3]$, respectively. To achieve this, count backwards and fill in as many ones as possible and imagine for the second and third case the $n+1^{\text{th}}$ and/or $n+2^{\text{th}}$ positions are each 1. Again for each of these three cases, the total number of all possible arrangements of the $n - y$ zeros, which are of three types with sizes M, j and $n - y - M - j$, is $(n-y)!/\{M!j!(n-y-M-j)!\}$. The three types of zeros are 10, 110 and single zeros, as described above. Replacing M by $y - 2j - 2$, $y - 2j - 1$, or $y - 2j$, respectively, we get

$$(2.10) \quad \begin{aligned} \tilde{F}_n(a) = & \sum_{y=2}^{y_1} \sum_{j=0}^{j_1} \binom{n-y}{y-j-2} \binom{y-j-2}{j} p_1^{y-j-1} p_2^{j+1} q_2^{y-2j-2} q_1^{n-2y+j+2} \\ & + \sum_{y=1}^{y_2} \sum_{j=0}^{j_2} \binom{n-y}{y-j-1} \binom{y-j-1}{j} p_1^{y-j} p_2^j q_2^{y-2j-1} q_1^{n-2y+j+1} \\ & + \sum_{y=0}^{y_3} \sum_{j=0}^{j_3} \binom{n-y}{y-j} \binom{y-j}{j} p_1^{y-j} p_2^j q_2^{y-2j} q_1^{n-2y+j}, \end{aligned}$$

where $y_1 = \min(a, n - [n/3])$, $y_2 = \min(a, n - [(n+1)/3])$ and $y_3 = \min(a, n - [(n+2)/3])$.

In both the cases (i) and (ii) above $\tilde{F}_n(a)$ and $\tilde{F}_n(a)$ are particular values

of $F_n(a)$, implying that the bounds can not be improved.

2.3 Numerical Examples

Following are some examples comparing the upper and lower bounds with the true value of $F_n(a)$, for the case when both $F_n(0)$ and $F_n(1)$ are known. In reality, the remaining p 's would be unknown. Under this minimal information one cannot use Theorem 2.2 directly. Even though we have assumed in Examples 2 and 4 that all p_i 's are same, this information may not be a priori known. Such an assumption is made only to get a feel for the difference between the estimated bounds versus the true values of the cumulative distribution function.

FORTTRAN program was used to compute these values. The true values of $F_n(a)$ were obtained from Theorem 2.2 and that of $\tilde{F}_n(a)$ and $\tilde{\tilde{F}}_n(a)$ from (2.9) and (2.10), respectively. A cross check was performed and it was noted that the upper and lower bounds matched the true value of the cumulative distribution function when all p_3, p_4, \dots, p_k were 0 and 1, respectively.

Generally, the upper and lower bounds are close to each other when the given p_1 and p_2 are near zero. Besides, when $p_i, i \geq 3$, are closer to 1 (0), we notice, as expected, that the lower (upper) bound does better.

Example 1 $k = 4, p_i \equiv 0.05i,$
 $n = 12$

y	$F_n(y)$	Upper bound	Lower bound
0	.54036	.54036	.54036
1	.86517	.86517	.86517
2	.96959	.97432	.94280
3	.99404	.99664	.95547
4	.99905	.99971	.98722
5	.99988	.99998	.99807
6	.99999	1.00000	.99944

Example 2 $k = 5, p_i \equiv 0.3,$
 $n = 12$

y	$F_n(y)$	Upper bound	Lower bound
0	.01384	.01384	.01384
1	.08503	.08503	.08503
2	.25282	.26371	.22739
3	.49252	.53034	.37158
4	.72365	.78369	.46113
5	.88215	.93541	.56821
6	.96140	.98956	.74575

Example 3 $k = 5$, $p_i = 0.54, 0.19,$
 $0.04, 0.64, 0.42$, $n = 12$

y	$F_n(y)$	Upper bound	Lower bound
0	.00009	.00009	.00009
1	.00224	.00224	.00224
2	.02263	.02266	.02172
3	.12075	.12136	.10543
4	.37377	.37735	.27925
5	.71478	.72420	.44623
6	.93391	.94514	.56140

Example 4 $k = 4$, $p_i \equiv 0.5$,
 $n = 12$

y	$F_n(y)$	Upper bound	Lower bound
0	.00024	.00024	.00024
1	.00317	.00317	.00317
2	.01929	.02173	.01685
3	.07300	.09277	.04883
4	.19385	.27100	.09912
5	.38721	.55640	.19409
6	.61279	.83521	.35962

CHAPTER 3

MODIFIED RANDOMIZED RESPONSE MODEL

3.1 Introduction and Summary

A question like, "Do you have AIDS?" is offensive and does not guarantee a truthful answer. Collecting information of such sensitive and personal nature requires carefully thought out procedures. The techniques employed currently do address the sensitivity of the issue but do not take into account the basic instinct to hide the truth in such matters.

Consider a community being surveyed by a government agency for the incidence of AIDS. Due to the very stigmatizing nature of the disease, the community may want to hide the truth to present a positive image. To get an accurate estimate in a situation like this, the proposed work assumes that k subjects in the sample collaborate to distort the truth, where $k = 1$ gives rise to the existing procedures. Further, people giving truthful answers are doing so independently.

Warner (1965) proposed a randomized response procedure assuming that the yes and no reports on sensitive information are made independently and truthfully. Abul-Ela et al. (1967) generalized this idea to t disjoint categories of the population, of which at most $t-1$ categories are stigmatizing. Under the same assumptions of truthful reporting and independence among responses of different individuals as in Warner (1965), but with no direct replies needed from the respondents, Kuk (1990) designed a randomized response model with a more efficient estimator. To capture the bias due to the possibility of the truth being concealed in a specific manner, the binary sequence of order k , as defined in the Chapter 1, is introduced in the randomized response model of Kuk. This includes

dependence and changes the probability of an affirmative from person to person, due to negative implications.

As in Kuk (1990), imagine an enclosed booth with two packs of cards, each with red and green colors. The percentages of red cards are θ_1 and θ_2 , $\theta_1 \neq \theta_2$. Pack 1 relates to an affirmative for AIDS and pack 2 to its negation. Each respondent shuffles and draws a card from each pack and puts it back after noting its color. Depending on whether the person does or does not have AIDS he reports the color of the card from either pack 1 or 2, whichever ever relates to him. Let these responses be realizations from binary sequence of order k , X_1, \dots, X_n . Assuming that everyone tell the truth ($k = 1$) the probability of obtaining a red card is given by

$$\rho = P(\text{Red Card}) = \theta_1\pi + \theta_2(1 - \pi),$$

where π is the proportion of people in the population that have AIDS. Further,

when ρ is estimated by $\bar{X} = n^{-1} \sum_{i=1}^n X_i$, $k = 1$, the above equation gives

maximum likelihood estimator of π which is also a moment estimator. The effects of using the binary sequence of order k are seen through the following

facts. The probability of the first person saying "no" to having AIDS is p_1 .

Influenced by the previous number of "no's", the probability that each of the next $k-1$ individuals will give the same answer is p_2, \dots, p_k , respectively. After k negations have been noted, $k+1^{\text{th}}$ person saying "no" has the same probability as that of the first person with this answer. If a person says "yes" to the above question then the next set of answers will be independent of all the previous answers. These facts can be derived from Definition 1.1 in Section 2. Note that for a given problem there may be different k 's involved which need to be estimated.

In view of the Chapter 2 of this paper the sum in the estimate of ρ can more generally be replaced by the finite sums of first few random variables of several

independent binary sequences, which needs to be investigated.

Asymptotic results such as consistency and normality of the estimator taking the bias into account are determined. Reduction in bias strategies is investigated.

3.2 Properties of the Estimator

In this section, $\bar{X}_n = \sum_{i=1}^n X_i/n$ is shown to be a strongly consistent estimator of μ .

The μ turns out to be the asymptotic mean of \bar{X}_n , which is also computed in this section.

These results are obtained by showing that $E \bar{X}_n$ satisfies the following sequence property. A sequence $\{a_n\}$ is said to be a weighted mean sequence of order k if there exist weights $\{w_i, i = 1, 2, \dots, k\}$ such that $0 < w_i < 1, \sum_{i=1}^k w_i = 1$, and for any $n > k$, $a_n = \sum_{i=1}^k w_i a_{n-i}$. If not stated below, please see

Appendix for the proofs of lemmas and theorems.

Lemma 3.1 If $\{\mu_n\}$ is a weighted mean sequence of order k , there exist real numbers $\mu, 0 < q < 1$ and $M > 0$ such that $|\mu_n - \mu| \leq q^n M$ for all n . All the results obtained here onwards will inherently assume that $0 < p_i < 1, i = 1, 2, \dots, k$. In the subsequent results Lemma 2.2 is repeatedly used.

Lemma 3.2 The binary sequence of order k satisfies the property that the joint conditional distributions of $X_{\lambda+1}, X_{\lambda+2}, \dots, X_n | X_1 = x_1, X_2 = x_2, \dots, X_{\lambda-1} = x_{\lambda-1}, X_\lambda = 0$, and $X_{k+1}, X_{k+2}, \dots, X_n | X_1 = X_2 = \dots = X_k = 1$ are the same as those of $X_1, X_2, \dots, X_{n-\lambda}$ and X_1, X_2, \dots, X_{n-k} , respectively.

Proof These properties automatically follow from the definition of binary sequence of order k .

Here onwards, the condition $0 < p_i < 1, 1 \leq i \leq k$ is assumed.

Theorem 3.1 Suppose $\{X_n\}$ is a binary sequence of order k with parameters p_1, p_2, \dots, p_k . Then there exist real numbers μ, q and M as in

Lemma 2.1. such that

$$|E(X_n) - \mu| \leq q^n M \text{ for all } n.$$

Theorem 3.2 Let X_1, X_2, X_3, \dots be a binary sequence of order k . Then $\bar{X}_n = \sum_{i=1}^n X_i/n$ converges to μ with probability 1 as $n \rightarrow \infty$.

Since, from Theorem 3.1 EX_n converges as $n \rightarrow \infty$ and Kronecker's Lemma give $\lim_{n \rightarrow \infty} E \bar{X}_n = \mu$ exists. There is an asymptotic bias in estimating π when using binary sequence, $k > 1$. Aki [2, Proposition 2.1] gives μ satisfying

$$(2.1) \quad \mu = \sum_{i=1}^{\infty} p_1 \cdots p_{i-k[i/k]} (p_1 \cdots p_k)^{[i/k]} (1 - \mu).$$

In the special case when $k = 1$, $\mu = \rho = p_1$, i.e., no bias, where product over empty set is taken to be 1.

3.3 Central Limet Theorem

This section proves the limit distribution of $(\bar{X}_n - \mu)/(\sigma/\sqrt{n})$ to be $N(0, 1)$.

Where μ is given by (2.1) and σ^2 , the variance of the limit distribution, is equal to

$$(3.1) \quad \mu - \mu^2 + 2(1-\mu) \sum_{j=1}^{\infty} (\mu - EX_j).$$

This is achieved by first showing $n \text{Var}(\bar{X}_n) \rightarrow \sigma^2$ and then proving $(\bar{X}_n - \mu)/\sqrt{\text{Var}(\bar{X}_n)}$ converges in distribution to $N(0, 1)$. The later is shown by using the central limit theorem of Philipp (1969, Section 3, Theorem 3, p. 164) given here for the sake of completeness.

Let $\{\mathcal{X}_{Nn}\}$ be a double sequence of random variables centered at expectations and with finite variance $\sigma_{Nn}^2 = E \mathcal{X}_{Nn}^2$. Assume that $\sigma_N^2 \equiv \max_{1 \leq n \leq N} \sigma_{Nn}^2 \rightarrow 0$, $c(N) \equiv \max_{1 \leq n \leq N} \|\mathcal{X}_{Nn}\|_{\infty} \rightarrow 0$ ($N \rightarrow \infty$) and $\Sigma_N^2 \equiv E \left\{ \sum_{n=1}^N \mathcal{X}_{Nn} \right\}^2 \leq c$, where c is a constant not depending on N and that $\Sigma_N/\sigma_N \rightarrow \infty$ ($N \rightarrow \infty$).

Moreover suppose that the following condition holds.

Denote by $M_{ab}^{(N)}$ the σ -algebra generated by the events $\{\mathcal{X}_{Nn} < \alpha\}$, $1 \leq a \leq n \leq b \leq N$ and any real number α . The mixing condition that is satisfied by the process $\{\mathcal{X}_{Nn}\}$ is given by

$$(3.2) \quad \sup_t \sup_{A \in M_{1t}^{(N)}} \sup_{B \in M_{t+n,N}^{(N)}} |P(AB) - P(A)P(B)| \leq \alpha(n) \downarrow 0,$$

with $\sum_{n=1}^{\infty} \alpha^{1/2}(n) < \infty$.

To simplify the notation let n assume the values $1, 2, \dots, N$ and hence \sum_n^N stands for $\sum_{n=1}^N$. Further, omit the index N in the random variables $\mathcal{X}_{Nn}, y_{Nn}, z_{Nn}$ defined below. With this convention for fixed N write

$$\sum_n \mathcal{X}_n = \sum_{j=1}^l y_j + \sum_{j=1}^{l+1} z_j,$$

where

$$\begin{aligned} y_1 &= \mathcal{X}_1 + \dots + \mathcal{X}_{h_1}, & z_1 &= \mathcal{X}_{h_1+1} + \dots + \mathcal{X}_{h_1+k}, \\ &\dots\dots\dots & &\dots\dots\dots \\ y_l &= \mathcal{X}_{\rho_l+1} + \dots + \mathcal{X}_{\rho_l+h_l}, & z_l &= \mathcal{X}_{\rho_l+h_l+1} + \dots + \mathcal{X}_{\rho_{l+1}}, \\ & & z_{l+1} &= \mathcal{X}_{\rho_{l+1}+1} + \dots + \mathcal{X}_N. \end{aligned}$$

Here, put $\rho_i = \sum_{v < i} (h_v + k)$, the integers h_v and k being at our disposal.

Theorem 3.3 Let $\{\mathcal{X}_{Nn}\}$ be a stochastic process satisfying all the conditions described above and that $\sum_N \rightarrow 1$ ($N \rightarrow \infty$). Let (κ_n, S_n) be any admissible pair [Philipp 1969, Definition p. 164]. Then $\sum_{i=1}^n \mathcal{X}_{Nn}$ converges to $N(0,1)$ in distribution and $c_N \rightarrow 0$ if and only if, for any $\epsilon > 0$,

$$\sum_{j \leq l} \int_{|y| \geq \epsilon} y^2 dF_{N_j}(y) \rightarrow 0 \text{ as } N \rightarrow \infty.$$

Let $\mathcal{X}_{Nn} = (X_n - E X_n) / \sqrt{\text{Var}(\sum_n X_n)}$ in the above theorem. The following work shows all the conditions of the above theorem hold.

Theorem 3.4 The variance of X_n multiplied by n converges to σ^2 as in

(3.1).

Proof Rewrite $n \text{Var}(\bar{X}_n)$ as

$$\begin{aligned}
& E(n^{1/2}(\bar{X}_n - \mu) - n^{1/2}[E\bar{X}_n - \mu])^2 = n^{-1} \sum_{1 \leq i, j \leq n} E(X_i - \mu)(X_j - \mu) - n[E\bar{X}_n - \mu]^2 \\
(3.3) \quad & = n^{-1} \sum_{i=1}^n E(X_i - \mu)^2 + 2n^{-1} \sum_{1 \leq i < j \leq n} E(X_i - \mu)(X_j - \mu) - n[E\bar{X}_n - \mu]^2.
\end{aligned}$$

Since $X_i^2 = X_i$, $E(X_i - \mu)^2 = \mu^2 + (1-2\mu)EX_i$ and therefore $n^{-1} \sum_{i=1}^n E(X_i - \mu)^2 \rightarrow \mu - \mu^2$ as $n \rightarrow \infty$.

When $j > i$, $E(X_i X_j) = P(X_i = X_j = 1) = P(X_j = 1) - P(X_i = 0, X_j = 1) = EX_j - (1 - EX_i)EX_{j-i}$. The preceding equality gives

$$(3.4) \quad E(X_i - \mu)(X_j - \mu) = (1-\mu)(EX_{j-i} - \mu) - (1-\mu)(EX_{j-i} - \mu) + (EX_i - \mu)(EX_{j-i} - \mu).$$

From Theorem 3.1.

$$(3.5) \quad \left| n^{-1} \sum_{1 \leq i < j \leq n} (EX_{j-i} - \mu) \right| \leq n^{-1} \sum_{j=2}^n \sum_{i=1}^{j-1} q^j = n^{-1} \sum_{j=2}^n (j-1)q^j \rightarrow 0 \text{ as } n \rightarrow \infty,$$

because $\sum_{j=1}^{\infty} jq^j$ converges. Since $|(EX_i - \mu)(EX_{j-i} - \mu)| \leq q^i q^{j-i} = q^j$, the

calculation in (3.5) again gives

$$(3.6) \quad n^{-1} \sum_{1 \leq i < j \leq n} (EX_i - \mu)(EX_{j-i} - \mu) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Consider $EX_{j-i} - \mu$ as elements of an upper triangular matrix say $A = (a_{ij})$. In A , add all the elements along diagonals parallel to the major diagonal to give

$$\begin{aligned}
(3.7) \quad n^{-1} \sum_{1 \leq i < j \leq n} (EX_{j-i} - \mu) &= n^{-1} \sum_{m=1}^{n-1} (n-m)(EX_m - \mu) = \\
& \sum_{m=1}^{n-1} (EX_m - \mu) - n^{-1} \sum_{m=1}^{n-1} m(EX_m - \mu).
\end{aligned}$$

Thus (3.7) converges to $-\sum_{j=1}^{\infty} (\mu - EX_j)$. Substitute (3.4) in (3.3) and then apply

to it (3.5) through (3.7) to get the desired result.

Subsequent results need the following notation. Let A, B denote subsets of 0–1 valued vectors of the vector space with dimensions t and N , respectively.

Furthermore, let

$$\begin{aligned}
A_{x_1, \dots, x_j} &= \{\underline{a} = (x_1, \dots, x_j, a_{j+1}, a_t) : \underline{a} \in A\}, \quad j < t, \\
A_{x_1, \dots, x_t} &= \{(x_1, \dots, x_t)\} \cap A, \\
B_{x_1, \dots, x_j} &= \{\underline{b} = (x_1, \dots, x_j, b_{j+1}, b_N) : \underline{b} \in B\}, \quad j < N, \\
B_{x_1, \dots, x_t} &= \{(x_1, \dots, x_N)\} \cap B,
\end{aligned}$$

$$\begin{aligned}
\tilde{A}_{x_1, \dots, x_j} &= \{(a_{j+1}, \dots, a_t) : \underline{a} = (x_1, \dots, x_j, a_{j+1}, a_t) \in A_{x_1, \dots, x_j}\}, \\
\tilde{B}_{x_1, \dots, x_j} &= \{(b_{j+1}, \dots, b_N) : \underline{b} = (x_1, \dots, x_j, b_{j+1}, b_N) \in B_{x_1, \dots, x_j}\},
\end{aligned}$$

and $\tilde{X}_{i,j}, \tilde{Y}_{i,j}$, the random vectors $(X_i, X_{i+1}, \dots, X_j)$ and $(Y_i, Y_{i+1}, \dots, Y_j)$, respectively. In particular, denote $\tilde{X}_{i,i+t-1}$ and $\tilde{Y}_{i,i+N-1}$ as \tilde{X}_i and \tilde{Y}_i . In this section, the constants q and M are as in Theorem 3.1.

Lemma 3.3 The inequality $|\mathbb{P}(\tilde{Y}_{n+1} \in B) - \mathbb{P}(\tilde{Y}_n \in B)| \leq C_1 q^n$ holds for $C_1 = M/(1-q)$, any B , and all integers $n, N > 0$.

Lemma 3.4 The inequality $|\mathbb{P}(\tilde{Y}_{n+l} \in B) - \mathbb{P}(\tilde{Y}_n \in B)| \leq C_2 q^n$ holds for $C_2 = C_1/(1-q)$, any B , and all positive integers n and l .

Lemma 3.5 The inequality

$$|\mathbb{P}(\tilde{X}_m \in A, \tilde{Y}_{n+m} \in B) - \mathbb{P}(\tilde{X}_m \in A)\mathbb{P}(\tilde{Y}_{n+m+l} \in B)| \leq 2C_2 q^n$$

holds for any B and all positive integers m, n, l , where A is the subset of \mathbb{R}

Lemma 3.6 There exists a constant $C(t)$ depending only on t such that

$$(3.8) \quad |\mathbb{P}(\tilde{X}_m \in A, \tilde{Y}_{n+m+t-1} \in B) - \mathbb{P}(\tilde{X}_m \in A)\mathbb{P}(\tilde{Y}_{n+m+l+t-1} \in B)| \leq C(t)q^n$$

holds for all positive integers m, n, l, t and N .

Lemma 3.7 Let $C(k)$ be as in Lemma 3.4, where k is the order of the binary sequence. Then $|\mathbb{P}(\tilde{X}_1 \in A, \tilde{Y}_{n+t} \in B) - \mathbb{P}(\tilde{X}_1 \in A)\mathbb{P}(\tilde{Y}_{n+t} \in B)| \leq C(k)q^n$ holds for any positive integer n, t , and N .

Corollary 3.1 Let $C(k)$ be as in Lemma 3.5, then for any integers $1 \leq i_1 < i_2 < \dots < i_m < j_1 < j_2 < \dots < j_l$ and $j_1 - i_m \geq N$, the following inequality holds:

$$|\mathbb{E}(X_{i_1} X_{i_2} \dots X_{i_m} X_{j_1} X_{j_2} \dots X_{j_l}) - \mathbb{E}(X_{i_1} X_{i_2} \dots X_{i_m})\mathbb{E}(X_{j_1} X_{j_2} \dots X_{j_l})| \leq C(k)q^N.$$

Theorem 3.5 (Central Limit Theorem) The limit distribution of $(\tilde{X}_n - \mu)/(\sigma/\sqrt{n})$ is $N(0, 1)$. Where μ is given by (2.1) and σ by (3.1).

Proof Start with $(\mathcal{X}_{N1}, \dots, \mathcal{X}_{Nt}) \in A^*$ and $(\mathcal{X}_{N,t+n}, \dots, \mathcal{X}_{NN}) \in B^*$ in $M_{1t}^{(N)}$ and $M_{t+n,N}^{(N)}$. Notice that these events get translated into $\tilde{X}_1 \in E \tilde{X}_1 + A^* \sqrt{\text{Var}(\sum_n X_n)}$ and $\tilde{Y}_{n+t} \in E \tilde{Y}_{n+t} + B^* \sqrt{\text{Var}(\sum_n X_n)}$. Since, Lemma 3.5 holds

for any Borel measurable set A and B , (3.2) holds with $\sum_{n=1}^{\infty} \alpha^{1/2}(n) < \infty$.

Further, it is a fact that $\Sigma_N \equiv 1$. From Theorem 3.4 σ_N^2 and $c(N)$ are of the orders of $O(1)/N$ and $O(1)/N^{1/2}$ because $\text{Var}(X_n) \leq 1$ and $\|X_n - E X_n\|_{\infty} \leq 1$.

In Philipp (1969, Definition, p. 159), let $S_N = \sigma_N$, $\kappa_N = \sigma_N^{1/2}$. Then $\zeta_N = \sigma^{3/2}/c^2(N) \rightarrow \infty$, $\Sigma_N^2/S_N = \sigma_N^{-1} \rightarrow \infty$ and $\alpha(\zeta_N)\Sigma_N^2/S_N = \alpha[\sigma^{3/2}/c^2(N)]\sigma_N^{-1} \rightarrow 0$, by the same reasoning as in the preceding statement. Finally, the set $\{|y_{N_j}| > \epsilon\}$

$\equiv \{|X_{\rho_{j+1}} - E X_{\rho_{j+1}} + \dots + X_{\rho_j+h_j} - E X_{\rho_j+h_j}| > \epsilon \sqrt{\text{Var}(\sum_n X_n)}\}$, the right hand side of this inequality goes to ∞ as $N \rightarrow \infty$, whereas the left hand side is bounded by h_j . Thus for a suitable choice of h_j the Lindeberg type condition of Theorem 3.3 is satisfied and hence the result.

APPENDIX

We start with the proof of Lemma 2.1. In order to do so we will first prove Lemma A.1.

Lemma A.1 Let $\{X_i\}$ be a binary sequence of order k . Suppose $x_i, i = 1, 2, \dots, n-1$, are given and j is defined by Definition 1.1, then for $1 \leq t \leq k$,

$$(A.1) \quad S_k(t; n-1) = \begin{cases} 0, & t \neq j \\ 1, & t = j \end{cases}$$

Proof This result holds true for the following two cases, (i) $t > n$ and (ii) $n \geq t$. Under (i), the fact that $n \geq r \geq j$, with r and j as in Definition 1.1, gives $t > j$ and from (2.3) both sides of (A.1) are equal to zero. In case (ii), let $i^* = [(r-1)/k]$, then $r = i^*k + j$. Again, from Definition 1.1, we get

$$(A.2) \quad x_{n-1} = x_{n-2} = \dots = x_{n-i^*k-j+1} = 1, \quad x_{n-i^*k-j} = 0.$$

If $t = j$ then the i^{th} term in the summation (2.3) is the product of ones from (A.2). All terms $i > i^*$ and $i < i^*$ in (2.3) are zero because they include in their factors x_{n-i^*k-j} and $1 - x_{n-ik-j}$, respectively. Therefore, $S_k(t; n-1) = 1$.

If $t \neq j$, then each term of the summation in (2.3) has at least one zero factor. Let us see this when (a) $k \geq t > j \geq 1$ and (b) $1 \leq t < j \leq k$. In Case (a) the terms $i < i^*$ in (2.3) contain the factor $1 - x_{n-ik-t}$, which is zero because $n - ik - t \geq n - i^*k + k - t > n - i^*k - j$. Also, the terms with subscript $i \geq i^*$ contain the factor x_{n-i^*k-j} because $n - i^*k - j \geq n - ik - t + 1$. Hence, they are zero from (A.2). In Case (b) the terms $i \leq i^*$ in (2.3) contain the factor $1 - x_{n-ik-t}$, which is zero because $n - ik - t > n - i^*k - j$. Also, the terms with subscript $i > i^*$ contain the factor $x_{n-i^*k-j} (= 0)$ because $n - i^*k - j \geq n - ik + k - j \geq n - ik - t + 1$. Hence the result.

Proof of Lemma 2.1 From Definition 1.1

$$P\left(X_n = x_n \mid X_1 = x_1, X_2 = x_2, \dots, X_{n-1} = x_{n-1}\right) = p_j^{x_n} q_j^{1-x_n}.$$

From Lemma A.1, $S_k(t; n-1) = \begin{cases} 0, & t \neq j \\ 1, & t = j \end{cases}$. Therefore,

$$\prod_{t=1}^k \left\{ p_t^{S_k(t; n-1)x_n} q_t^{S_k(t; n-1)(1-x_n)} \right\} = p_j^{x_n} q_j^{1-x_n}. \text{ This completes the proof.}$$

Proof of Theorem 2.2 When $\sum_{i=1}^n X_i = y$, there are y number of X_i 's satisfying $X_i = 1$, and b number of X_i 's satisfying $X_i = 0$. Fix the positions of those X_i satisfying $X_i = 0$ at $j_1, j_2, \dots, j_b, j_1 < j_2 < \dots < j_b$, in the first n terms of the binary sequence. The probability of $[X_{j_1} = \dots = X_{j_b} = 0, X_i = 1, i \neq j_m, 1 \leq i \leq n, 1 \leq m \leq b]$ is

$$\begin{aligned} &= P[(X_1, X_2, \dots, X_{j_1-1}, X_{j_1}) = (1, \dots, 1, 0)] \\ (A.3) \times &P[X_{j_1+1} = \dots = X_{j_2-1} = 1, X_{j_2} = 0 \mid (X_1 = \dots = X_{j_1-1} = 1, X_{j_1} = 0)] \times \dots \\ &\times P[X_{j_{b-1}+1} = \dots = X_{j_b-1} = 1, X_{j_b} = 0 \mid X_i = 1, i < j_{b-1}, i \neq j_m, X_{j_m} = 0, 1 \leq m < b] \\ &\times P[X_{j_b+1} = \dots = X_n = 1 \mid X_i = 1, i < j_b, i \neq j_m, X_{j_m} = 0, 1 \leq m \leq b]. \end{aligned}$$

According to Definition 1.1, the m^{th} factor on the right hand side of (A.3) is

$$\left(p_0 p_1 \dots p_k \right)^{[(\Delta j_m - 1)/k]} p_0 p_1 \dots p_{(\Delta j_m - 1)_k} q_{(\Delta j_m - 1)_k}^{+1},$$

and the last factor is

$$\left(p_0 p_1 \dots p_k \right)^{[(n-j_b)/k]} p_0 p_1 \dots p_{(n-j_b)_k}.$$

Adding all probabilities corresponding to the various positions of $\{j_m\}$'s, we get the desired result.

To prove Theorem 2.3 we shall first prove some lemmas.

Lemma A.2 Let $1 < \lambda < n$, $P\left(\sum_{i=\lambda+1}^n X_i \leq a \mid X_1 = x_1, \dots, X_{\lambda-1} = x_{\lambda-1}, X_\lambda = 0\right) = F_{n-\lambda}(a)$ and $P\left(\sum_{i=k+1}^n X_i \leq a \mid X_1 = \dots = X_k = 1\right) = F_{n-k}(a)$.

Proof According to the definition of binary sequence of order k , the joint distributions of $X_{\lambda+1}, X_{\lambda+2}, \dots, X_n \mid X_1 = x_1, \dots, X_{\lambda-1} = x_{\lambda-1}, X_\lambda = 0$, and $X_{k+1}, X_{k+2}, \dots, X_n \mid X_1 = \dots = X_k = 1$ are the same as the distributions of $X_1, X_2, \dots, X_{n-\lambda}$ and X_1, X_2, \dots, X_{n-k} , respectively. Therefore,

$$\sum_{i=\lambda+1}^n X_i \leq a \mid X_a = 0, \quad \sum_{i=k+1}^n X_i \leq a \mid X_1 = \dots = X_k = 1 \quad \text{and} \quad \sum_{i=1}^{n-\lambda} X_i, \quad \sum_{i=1}^{n-k} X_i$$

have the same cumulative distribution functions, respectively.

Since $X_{n+j} \leq 1$, the proof of Lemma A.3 below follows from the fact that $\sum_{i=1}^{n+j} X_i \leq a+j$ implies that $\sum_{i=1}^{n+j+1} X_i \leq a+j+1$.

Lemma A.3 For any integer m ,

$$F_n(a) \leq F_{n+1}(a+1) \leq \dots \leq F_{n+m}(a+m).$$

Proof of Theorem 2.3 We shall use induction on n to give this proof. The hypothesis is true for $n \leq m$, because $F_n(a)$ does not depend on $p_{m+1}, p_{m+2}, \dots, p_k$ and is a known constant as a function of the remaining p_i 's. Suppose the statement is true for $n-1$ or less, then: If $m < n \leq k$,

$$\begin{aligned} F_n(a) &= P\left(\sum_{i=2}^n X_i \leq a, X_1 = 0\right) + P\left(\sum_{i=3}^n X_i \leq a-1, X_1 = 1, X_2 = 0\right) \\ &\quad + P\left(\sum_{i=4}^n X_i \leq a-2, X_1 = X_2 = 1, X_3 = 0\right) \\ &\quad + \dots + P\left(\sum_{i=m+1}^n X_i \leq a-m+1, X_1 = \dots = X_{m-1} = 1, X_m = 0\right) \\ &\quad + P\left(\sum_{i=m+2}^n X_i \leq a-m, X_1 = \dots = X_m = 1, X_{m+1} = 0\right) \\ &\quad + \dots + P(X_n \leq a-n+2, X_1 = \dots = X_{n-2} = 1, X_{n-1} = 0) \\ &\quad + P(X_n \leq a-n+1, X_1 = \dots = X_{n-1} = 1). \end{aligned}$$

The previous equality follows from the fact that the sample space is the union of mutually disjoint sets

$$(A.4) \quad \{X_1 = 0\}, \{X_1 = 1, X_2 = 0\}, \dots, \{X_1 = \dots = X_{n-2} = 1, X_{n-1} = 0\}, \\ \{X_1 = \dots = X_{n-1} = 1\}.$$

Applying Lemma A.2, we have

$$(A.5) \quad F_n(a) = F_{n-1}(a)q_1 + F_{n-2}(a-1)p_1q_2 + F_{n-2}(a-2)p_1q_2q_3 \\ + \dots + F_{n-m}(a-m+1)p_1 \dots p_{m-1}q_m + F_{n-m-1}(a-m)p_1 \dots p_m q_{m+1} \\ + \dots + F_1(a-n+2)p_1 \dots p_{n-2}q_{n-1} + F_1(a-n+1)p_1 \dots p_{n-1}.$$

In this expression, take $p_{m+1} = p_{m+2} = \dots = p_k = 1$, i.e., also $q_{m+1} = q_{m+2} = \dots = q_k = 0$, which gives all the last $n - m$ terms, except the last, to be zero. The last term simplifies to $\tilde{F}_1(a-n+1)p_1p_2 \dots p_m$ giving

$$(A.6) \quad \tilde{F}_n(a) = \tilde{F}_{n-1}(a)q_1 + \tilde{F}_{n-2}(a-1)p_1q_2 + \tilde{F}_{n-3}(a)p_1p_2q_3 \\ + \dots + \tilde{F}_{n-m}(a-m+1)p_1p_2 \dots p_{m-1}q_m + \tilde{F}_1(a-n+1)p_1p_2 \dots p_m.$$

Since $F_n(\cdot)$ is an increasing function, using Lemma A.3 yields

$$(A.7) \quad F_1(a-n+1) \leq F_1(a-n+2) \leq F_2(a-n+1) \leq \dots \leq F_{n-m-1}(a-m).$$

Applying these inequalities to the last $n - m$ terms of (A.5) yields

$$(A.8) \quad F_n(a) \geq F_{n-1}(a)q_1 + F_{n-2}(a-1)p_1q_2 + F_{n-3}(a)p_1p_2q_3 \\ + \dots + F_{n-m}(a-m+1)p_1p_2 \dots p_{m-1}q_m \\ + F_1(a-n+1)p_1p_2 \dots p_m [q_{m+1} + p_{m+1}q_{m+2} + \dots + p_{m+1}p_{m+2} \dots p_{n-2}q_{n-1} \\ + p_{m+1}p_{m+2} \dots p_{n-2}p_{n-1}].$$

Since $p_i + q_i = 1$, the factor in the last term on the right hand side of the above inequality

$$q_{m+1} + p_{m+1}q_{m+2} + \dots + p_{m+1}p_{m+2} \dots p_{n-2}q_{n-1} + p_{m+1}p_{m+2} \dots p_{n-2}p_{n-1} = 1.$$

By induction hypothesis, $\tilde{F}_j \leq F_j$, $1 \leq j \leq n - 1$. Applying these inequalities term by term to the right hand side of (A.8) gives $\tilde{F}_n(a) \leq F_n(a)$.

Similarly, in (A.5), take $p_{m+1} = p_{m+2} = \cdots = p_k = 0$, i.e., also $q_{m+1} = q_{m+2} = \cdots = q_k = 1$. We have:

$$\begin{aligned} \tilde{F}_n(a) &= \tilde{F}_{n-1}(a)q_1 + \tilde{F}_{n-2}(a-1)p_1q_2 + \tilde{F}_{n-3}(a)p_1p_2q_3 \\ &+ \cdots + \tilde{F}_{n-m}(a-m+1)p_1p_2 \cdots p_{m-1}q_m + \tilde{F}_{n-m-1}(a-m)p_1p_2 \cdots p_m. \end{aligned}$$

From (A.5) and (A.7) we get, as above,

$$\begin{aligned} F_n(a) &\leq F_{n-1}(a)q_1 + F_{n-2}(a-1)p_1q_2 + F_{n-3}(a)p_1p_2q_3 \\ &+ \cdots + F_{n-m}(a-m+1)p_1p_2 \cdots p_{m-1}q_m + F_{n-m-1}(a-m)p_1p_2 \cdots p_m. \end{aligned}$$

Therefore, once again $F_n(a) \leq \tilde{F}_n(a)$, $n \leq k$, by induction hypothesis.

Consider the case $n > k$. Use (A.4), with n replaced by $k+1$, in the proof of (A.5) to give

$$\begin{aligned} (A.9) \quad F_n(a) &= F_{n-1}(a)q_1 + F_{n-2}(a-1)p_1q_2 + F_{n-2}(a-2)p_1q_2q_3 \\ &+ \cdots + F_{n-m}(a-m+1)p_1 \cdots p_{m-1}q_m + F_{n-m-1}(a-m)p_1 \cdots p_m q_{m+1} \\ &+ \cdots + F_{n-k}(a-k+1)p_1 \cdots p_{k-1}q_k + F_{n-k}(a-k)p_1 \cdots p_k. \end{aligned}$$

Hence, applying Lemma 5.3 to last $k-m+1$, excluding the last one, of the terms on the right hand side of (A.9) and induction hypothesis give

$$\begin{aligned} F_n(a) &\geq F_{n-1}(a)q_1 + F_{n-2}(a-1)p_1q_2 + F_{n-3}(a-2)p_1p_2q_3 \\ &+ \cdots + F_{n-m}(a-m+1)p_1p_2 \cdots p_{m-1}q_m + F_{n-k}(a-k)p_1p_2 \cdots p_m \\ &\geq \tilde{F}_{n-1}(a)q_1 + \tilde{F}_{n-2}(a-1)p_1q_2 + \tilde{F}_{n-3}(a-2)p_1p_2q_3 \\ &+ \cdots + \tilde{F}_{n-m}(a-m+1)p_1p_2 \cdots p_{m-1}q_m + \tilde{F}_{n-k}(a-k)p_1p_2 \cdots p_m = \tilde{F}_n(a). \end{aligned}$$

Also, as seen earlier,

$$\begin{aligned} F_n(a) &\leq F_{n-1}(a)q_1 + F_{n-2}(a-1)p_1q_2 + F_{n-3}(a-2)p_1p_2q_3 \\ &+ \cdots + F_{n-m}(a-m+1)p_1p_2 \cdots p_{m-1}q_m + F_{n-m-1}(a-m)p_1p_2 \cdots p_m \\ &\leq \tilde{F}_{n-1}(a)q_1 + \tilde{F}_{n-2}(a-1)p_1q_2 + \tilde{F}_{n-3}(a-2)p_1p_2q_3 \end{aligned}$$

$$+ \cdots + \overset{\approx}{F}_{n-m}(a-m+1)p_1 p_2 \cdots p_{m-1} q_m + \overset{\approx}{F}_{n-m-1}(a-m)p_1 p_2 \cdots p_m = \overset{\approx}{F}_n(a).$$

Therefore the statement is also true for n bigger than k . This completes the proof.

Proof of Lemma 3.1 Let $M_n = \max(\mu_n, \mu_{n+1}, \dots, \mu_{n+k-1})$ and $m_n = \min(\mu_n, \mu_{n+1}, \dots, \mu_{n+k-1})$. Since $\mu_{n+k} = \sum_{i=1}^k w_i \mu_{n+k-i} \leq \sum_{i=1}^k w_i M_n = M_n$, then $M_{n+1} = \max(\mu_{n+1}, \mu_{n+2}, \dots, \mu_{n+k}) \leq M_n$, i.e., $\{M_n\}$ is an decreasing sequence. Similarly, we can show that $\{m_n\}$ is an increasing sequence. Both $\{M_n\}$ and $\{m_n\}$ are convergent because $\{M_n\}$ and $\{m_n\}$ are bounded according to $m_1 \leq m_n \leq M_n \leq M_1$. We shall now prove $\lim_{n \rightarrow \infty} M_n = \lim_{n \rightarrow \infty} m_n$ by proving

$$\lim_{n \rightarrow \infty} (M_n - m_n) = 0.$$

Let $w^* = \min(w_1, w_2, \dots, w_k)$, and i^* be such μ_{n+i^*} is the value M_n , $0 \leq i^* \leq k-1$. Then

$$\mu_{n+k} = w_1 \mu_{n+k-1} + \cdots + w_{k-i^*} \mu_{n+i^*} + \cdots + w_k \mu_n.$$

Replacing $w_{k-i^*} \mu_{n+i^*}$ by $w_{k-i^*} m_n + w_{k-i^*} (M_n - m_n)$, and using $\mu_{n+i} \geq m_n$ for $0 \leq i \leq k-1, i \neq i^*$, we get

$$\begin{aligned} \mu_{n+k} &\geq (w_1 + \cdots + w_{k-i^*} + \cdots + w_k) m_n + w_{k-i^*} (M_n - m_n) \\ &\geq m_n + w^* (M_n - m_n) \geq m_n + w^* (M_{n+k} - m_{n+k}), n \geq 1. \end{aligned}$$

The last step follows from the fact that $\{M_n - m_n\}$ is a decreasing sequence.

In the preceding inequality, replace n by $n+1, \dots, n+k-1$, gives

$$\mu_{n+k+1} \geq m_{n+1} + w^* (M_{n+1} - m_{n+1}) \geq m_n + w^* (M_{n+k} - m_{n+k}).$$

...

$$\mu_{n+2k-1} \geq m_{n+k-1} + w^* (M_{n+k-1} - m_{n+k-1}) \geq m_n + w^* (M_{n+k} - m_{n+k}).$$

Therefore

$$(A.10) \quad m_{n+k} = \min(\mu_{n+k}, \mu_{n+k+1}, \dots, \mu_{n+2k-1}) \geq m_n + w^* (M_{n+k} - m_{n+k}),$$

$n \geq 1$.

Similarly, in the preceding paragraph, replace maximum by minimum and " \geq " by " \leq " to give

$$(A.11) \quad M_{n+k} \leq M_n - w^*(M_{n+k} - m_{n+k}).$$

Subtracting (A.10) from (A.11) yields

$$(A.12) \quad \begin{aligned} M_{n+k} - m_{n+k} &\leq (M_n - m_n) - 2w^*(M_{n+k} - m_{n+k}), \text{ i.e.,} \\ M_{n+k} - m_{n+k} &\leq (1+2w^*)^{-1}(M_n - m_n) \text{ and} \\ M_{n+jk} - m_{n+jk} &\leq (1+2w^*)^{-j}(M_n - m_n). \end{aligned}$$

Hence, $M_{n+jk} - m_{n+jk} \rightarrow 0$ as $j \rightarrow \infty$. Also, $\lim_{n \rightarrow \infty} M_n = \lim_{n \rightarrow \infty} m_n$ because

$\lim_{n \rightarrow \infty} (M_n - m_n)$ exists. Finally, the relation $M_n \geq \mu_n \geq m_n$ gives the

convergence of the sequence $\{\mu_n\}$.

Let $q = (1+2w^*)^{-1/k}$ and $M = \max_{1 \leq i \leq k} \{(M_i - m_i)/q^{k+1}\}$, from (A.12) $M_n - m_n \leq (1+2w^*)^{-[(n-1)/k]}(Mq^{k+1}) \leq q^n M$. It yields $|\mu_n - \mu| \leq M_n - m_n \leq q^n M$. This completes the proof.

Proof of Theorem 3.1 It suffices to consider the case when $n > k$. Thus,

$$\begin{aligned} E(X_n) &= P(X_n = 1) = P(X_n = 1; X_1 = 0) + P(X_n = 1; X_1 = 1, X_2 = 0) + \\ &P(X_n = 1; X_1 = X_2 = 1, X_3 = 0) + \cdots + P(X_n = 1; X_1 = \cdots = X_{k-1} = 1, X_k = 0) \\ &\quad + P(X_n = 1; X_1 = X_2 = \cdots = X_{k-1} = X_k = 1). \end{aligned}$$

Using Lemma 2.2 and Definition 1.1, $E(X_n)$ further reduces to

$$\begin{aligned} &q_1 E(X_{n-1}) + p_1 q_2 E(X_{n-2}) + p_1 p_2 q_3 E(X_{n-3}) + \cdots + p_1 p_2 \cdots p_{k-1} q_k E(X_{n-k}) + \\ &\quad p_1 p_2 \cdots p_k E(X_{n-k}) \\ &= w_1 E(X_{n-1}) + w_2 E(X_{n-2}) + \cdots + w_1 E(X_{n-1}) + w_k E(X_{n-k}), \end{aligned}$$

where $w_i = p_0 p_1 \cdots p_{i-1} q_i$, $1 \leq i \leq k-1$, $w_k = (p_1 \cdots p_{k-1} q_k) + (p_1 \cdots p_k)$.

The fact that $p_i + q_i = 1$ yields $\sum_{i=1}^k w_i = 1$. Also, $p_i \in (0,1)$ implies $w_i \in (0,1)$.

Thus, $\{E(X_n)\}$ is a weighted mean sequence of order k . Hence the result follows from Lemma 2.1.

Proof of Theorem 3.2 First, let us prove that there exists a constant D and

an integer $n_0 > 0$ such that

$$(A.13) \quad \left| \sum_{1 \leq i, j, l, m \leq n} E(X_i - \mu)(X_j - \mu)(X_l - \mu)(X_m - \mu) \right| \leq Dn^{8/3} \quad \text{for } n > n_0.$$

In order to achieve this, consider the term inside the expectation of (A.13) and change the positions of its factors so as to satisfy $i \leq j \leq l \leq m$. Then, identify each term inside the summation of (A.13) with one of the following mutually exclusive and exhaustive sets G_i , $i = 1, 2, 3, 4$ with their subscript satisfying

$$\begin{aligned} [1] \quad m - l \geq N & & [2] \quad m - l < N, j - i \geq N, i > N, \\ [3] \quad m - l < N, j - i \geq N, i \leq N, & & [4] \quad m - l < N, j - i < N, \end{aligned}$$

respectively, where N is any real number.

To determine the order of each term in group G_1 or G_2 consider a general term $E\{(X_i - \mu)(X_j - \mu)(X_l - \mu)(X_m - \mu)\}$ and expanded it as

$$\begin{aligned} & \left\{ E(X_i X_j X_l X_m) - E(X_i X_j X_l)E(X_m) \right\} + E(X_i X_j X_l)[E(X_m) - \mu] + \\ & \mu \left\{ \left([E(X_i X_j) + E(X_j X_l) + E(X_i X_l)] E(X_m) - E(X_i X_j X_m) + E(X_j X_l X_m) + \right. \right. \\ & E(X_i X_l X_m) \left. \right) + [E(X_i X_j) + E(X_j X_l) + E(X_i X_l)] [\mu - E(X_m)] \left. \right\} + \mu^2 \left\{ \left(E(X_i X_m) \right. \right. \\ & + E(X_j X_m) + E(X_l X_m) - [E(X_i) + E(X_j) + E(X_l)] E(X_m) \left. \right) + [E(X_i) + E(X_j) + \\ & E(X_l)] (E(X_m) - \mu) \left. \right\} + \mu^3 [\mu - E(X_m)]. \end{aligned}$$

Simplify the preceding equation by using the fact that $X_i^n = X_i$, for any integer n , when one or more of the subscripts are equal to each other. Then by Corollary 3.1, and the fact $m - l \geq N$, the expanded term above is observed to be of the order q^N . Due to the symmetry of the subscripts in the above expansion interchange the role of m and i to note the same order holds for each term in G_2 .

The inequality (A.13) can now be proved by showing that there is only an order of $N^2 n^2$ terms in G_3 and G_4 both combined and an appropriate choice of N . The number of different terms with subscripts $i \leq j \leq l \leq m$, is n^2 for the different pairs of (m, j) and for each fixed pair of these there are N^2 pair of (i, l) because $1 \leq i \leq N$, $m - N < l \leq m$ or $j - N < i \leq j$, $m - N < l \leq m$ for

G_3 or G_4 , respectively. To incorporate all the terms in the sum of (A.13) the condition $i \leq j \leq l \leq m$ can be removed by taking the number of terms in each of groups G_3 and G_4 to be no more than $4! N^2 n^2$ terms. Thus, choosing $N = n^{1/3}$, gives (A.13).

Therefore,

$$E(\bar{X}_n - \mu)^4 = n^{-4} \left| \sum_{1 \leq i, j, l, m \leq n} E\{(X_i - \mu)(X_j - \mu)(X_l - \mu)(X_m - \mu)\} \right| \leq Dn^{-4/3},$$

for all $n > n_0$. For any $\epsilon > 0$, and $n > n_0$, $P\{|\bar{X}_n - \mu| \geq \epsilon\} = P\{|\bar{X}_n - \mu|^4 \geq \epsilon^4\} \leq \epsilon^{-4} E|\bar{X}_n - \mu|^4 \leq D\epsilon^{-4} n^{-4/3}$ by the Markov inequality. The result now follows from the Borel–Cantelli Theorem.

Proof of Lemma 3.3 From the definition of the B 's and Lemma 2.2

$$\begin{aligned} P(\tilde{Y}_{n+1} \in B) - P(\tilde{Y}_n \in B) &= P(\tilde{Y}_{n+1} \in B_0) + P(\tilde{Y}_{n+1} \in B_1) - P(\tilde{Y}_n \in B_0) - \\ P(\tilde{Y}_n \in B_1) &= P(Y_{n+1} = 0)P(\tilde{Y}_{1, N-1} \in \tilde{B}_0) - P(Y_n = 0)P(\tilde{Y}_{1, N-1} \in \tilde{B}_0) + \\ P(\tilde{Y}_{n+2, N+n} \in \tilde{B}_1) - P(Y_{n+1} = 0)P(\tilde{Y}_{1, N-1} \in \tilde{B}_1) - P(\tilde{Y}_{n+1, N+n-1} \in \tilde{B}_1) + \\ P(Y_n = 0)P(\tilde{Y}_{1, N-1} \in \tilde{B}_1). \end{aligned}$$

The absolute value of the preceding expression is dominated by

$$|P(Y_{n+1} = 0) - P(Y_n = 0)| |P(\tilde{Y}_{1, N-1} \in \tilde{B}_0) - P(\tilde{Y}_{1, N-1} \in \tilde{B}_1)| + |P(\tilde{Y}_{n+2, N+n} \in \tilde{B}_1) - P(\tilde{Y}_{n+1, N+n-1} \in \tilde{B}_1)| \leq Mq^n + |P(\tilde{Y}_{n+2, N+n} \in \tilde{B}_1) - P(\tilde{Y}_{n+1, N+n-1} \in \tilde{B}_1)|,$$

which is obtained by applying Theorem 3.1. Using the inequality obtained so far

$$\begin{aligned} \text{iteratively yields } |P(\tilde{Y}_{n+1} \in B) - P(\tilde{Y}_n \in B)| \\ \leq Mq^n + Mq^{n+1} + |P(\tilde{Y}_{n+3, N+n} \in \tilde{B}_{1,1}) - P(\tilde{Y}_{n+2, N+n-1} \in \tilde{B}_{1,1})| \\ \leq Mq^n + Mq^{n+1} + \dots + Mq^{n+N-1} \leq C_1 q^n. \end{aligned}$$

Proof of Lemma 3.4 Note that $|P(\tilde{Y}_{n+l} \in B) - P(\tilde{Y}_n \in B)| \leq \sum_{j=1}^l |P(\tilde{Y}_{n+j} \in B) - P(\tilde{Y}_{n+j-1} \in B)|$. To each term in this sum apply Lemma 3.1, with n replaced by $n+j-1$, which gives a bound of $C_1 \sum_{j=1}^l q^{n+j-1} \leq q^n C_1 / (1-q) = C_2 q^n$.

Proof of Lemma 3.5 Consider all possible choices of A of dimension 1. If $A = \phi$, the statement is obviously true. If $A = \{0\}$, by Lemma 3.2,

$$\begin{aligned} & |P(X_m = 0, \tilde{Y}_{n+m} \in B) - P(X_m = 0)P(\tilde{Y}_{n+m+l} \in B)| = \\ & |P(X_m = 0)\{P(\tilde{Y}_n \in B) - P(\tilde{Y}_{n+l} \in B)\}| \leq C_2 q^n \leq C_3 q^n. \end{aligned}$$

$$\begin{aligned} \text{If } A = \{1\}, & P(X_m = 1, \tilde{Y}_{n+m} \in B) - P(X_m = 1)P(\tilde{Y}_{n+m+l} \in B) = \\ & (P(\tilde{Y}_{n+m} \in B) - P(X_m = 0)P(\tilde{Y}_n \in B) - [1 - P(X_m = 0)]P(\tilde{Y}_{n+m+l} \in B)), \end{aligned}$$

by Lemma 2.2. The preceding expression is bounded by

$$\begin{aligned} & |P(\tilde{Y}_{n+m} \in B) - P(\tilde{Y}_{n+m+l} \in B)| + |P(X_m = 0)[P(\tilde{Y}_n \in B) - P(\tilde{Y}_{n+m+l} \in B)]| \leq \\ & 2C_2 q^n, \text{ by Lemma 3.2. If } A = \{0, 1\}, \text{ the hypothesis reduces to that of Lemma} \\ & 3.2. \end{aligned}$$

Proof of Lemma 3.6 Apply induction to this hypothesis as a function of t and note that it is true for $t = 1$ by Lemma 3.3. Assume that the statement is true for any integer $t \leq t_0$, to show that it holds for $t = t_0 + 1$. Consider the expression on the left hand side of (3.8). It can be rewritten as

$$\begin{aligned} & P(X_m = 0, \tilde{X}_{m+1, m+t-1} \in \tilde{A}_0, \tilde{Y}_{n+m+t-1} \in B) + \\ & P(X_m = 1, \tilde{X}_{m+1, m+t-1} \in \tilde{A}_1, \tilde{Y}_{n+m+t-1} \in B) - \\ & P(X_m = 0, \tilde{X}_{m+1, m+t-1} \in \tilde{A}_0)P(\tilde{Y}_{n+m+l+t-1} \in B) - \\ & P(X_m = 1, \tilde{X}_{m+1, m+t-1} \in \tilde{A}_1)P(\tilde{Y}_{n+m+l+t-1} \in B), \end{aligned}$$

because A is the disjoint union of the sets A_0 and A_1 . This expression using Lemma 2.2 can be rearranged as

$$\begin{aligned} & P(X_m = 0) \left\{ P(\tilde{X}_{1, t-1} \in \tilde{A}_0, \tilde{Y}_{n+t-1} \in B) - P(\tilde{X}_{1, t-1} \in \tilde{A}_0)P(\tilde{Y}_{n+m+l+t-1} \in B) \right\} + \\ & \left\{ P(\tilde{X}_{m+1, m+t-1} \in \tilde{A}_1, \tilde{Y}_{n+m+t-1} \in B) - P(\tilde{X}_{m+1, m+t-1} \in \tilde{A}_1)P(\tilde{Y}_{n+m+l+t} \in B) \right\} - \\ & P(X_m = 0) \left\{ P(\tilde{X}_{1, t-1} \in \tilde{A}_1, \tilde{Y}_{n+t-1} \in B) - P(\tilde{X}_{1, t-1} \in \tilde{A}_1)P(\tilde{Y}_{n+m+l+t} \in B) \right\} \end{aligned}$$

Notice that \tilde{A}_0 and \tilde{A}_1 are subset of the space with dimension $t-1$. By induction hypothesis each of the terms in the curly brackets is less than $C(t-1)q^n$. Hence, take $C(t) = 3C(t-1)$, i.e., $C(t) = 2C_2 3^{t-1}$, where C_2 is as in Lemma 3.2. The statement being true for $t = t_0 + 1$ the proof is complete.

Proof of Lemma 3.7 This result is derived by applying induction on t . The hypothesis is true for $1 \leq t \leq k_0$, which follows from Lemma 3.4 and the fact that

$C(k) = \text{Max}_{1 \leq j \leq k} C(j)$. Assume that the statement is true for $t \leq t_0$, to show that

it is true for $t = t_0 + 1$, $t > k$. Since A can be written as the union of mutually exclusive disjoint sets $A_0, A_{1,0}, \dots, A_{\underbrace{1,1, \dots, 1,0}_{k-1}}$ and $A_{\underbrace{1,1, \dots, 1}_k}$, it gives

$$\begin{aligned} & |P(X_{\tilde{1}} \in A, Y_{\tilde{n+t}} \in B) - P(X_{\tilde{1}} \in A)P(Y_{\tilde{n+t}} \in B)| \leq |P(X_{\tilde{1}} \in A_0, Y_{\tilde{n+t}} \in B) - \\ & P(X_{\tilde{1}} \in A_0)P(Y_{\tilde{n+t}} \in B)| + |P(X_{\tilde{1}} \in A_{1,0}, Y_{\tilde{n+t}} \in B) - P(X_{\tilde{1}} \in A_{1,0})P(Y_{\tilde{n+t}} \in B)| + \\ & \dots + |P(X_{\tilde{1}} \in A_{\underbrace{1,1, \dots, 1,0}_{k-1}}, Y_{\tilde{n+t}} \in B) - P(X_{\tilde{1}} \in A_{\underbrace{1,1, \dots, 1,0}_{k-1}})P(Y_{\tilde{n+t}} \in B)| + \\ & |P(X_{\tilde{1}} \in A_{\underbrace{1,1, \dots, 1}_k}, Y_{\tilde{n+t}} \in B) - P(X_{\tilde{1}} \in A_{\underbrace{1,1, \dots, 1}_k})P(Y_{\tilde{n+t}} \in B)|. \end{aligned}$$

This in turn by Lemma 2.2 is equal to

$$\begin{aligned} & P(X_1 = 0) |P(X_{\tilde{1,t-1}} \in \tilde{A}_0, Y_{\tilde{n+t-1}} \in B) - P(X_{\tilde{1,t-1}} \in \tilde{A}_0)P(Y_{\tilde{n+t}} \in B)| + \\ & P(X_1 = 1, X_2 = 0) |P(X_{\tilde{1,t-2}} \in \tilde{A}_{1,0}, Y_{\tilde{n+t-2}} \in B) - P(X_{\tilde{1,t-2}} \in \tilde{A}_{1,0})P(Y_{\tilde{n+t}} \in B)| + \\ & \dots + P(X_1 = X_2 = \dots = X_{k-1} = 1, X_k = 0) |P(X_{\tilde{1,t-k}} \in \tilde{A}_{\underbrace{1,1, \dots, 1,0}_{k-1}}, \\ & Y_{\tilde{n+t-k}} \in B) - P(X_{\tilde{1,t-k}} \in \tilde{A}_{\underbrace{1,1, \dots, 1,0}_{k-1}})P(Y_{\tilde{n+t}} \in B)| + \\ & P(X_1 = X_2 = \dots = X_k = 1) |P(X_{\tilde{1,t-k}} \in \tilde{A}_{\underbrace{1,1, \dots, 1}_k}, Y_{\tilde{n+t-k}} \in B) - \\ & P(X_{\tilde{1,t-k}} \in \tilde{A}_{\underbrace{1,1, \dots, 1}_k})P(Y_{\tilde{n+t}} \in B)| \\ & \leq \left\{ P(X_1 = 0) + P(X_1 = 1, X_2 = 0) + \dots + P(X_1 = X_2 = \dots = X_{k-1} = 1, X_k = 0) \right. \\ & \quad \left. + P(X_1 = X_2 = \dots = X_k = 1) \right\} Cq^n = Cq^n. \end{aligned}$$

The last inequality follows from Lemma 3.5.

Proof of Corollary 3.1 Since $\{X_i\}$ is a binary sequence,

$$E(X_{i_1} X_{i_2} \dots X_{i_m} X_{j_1} X_{j_2} \dots X_{j_l}), E(X_{i_1} X_{i_2} \dots X_{i_l}) \text{ and } E(X_{j_1} X_{j_2} \dots X_{j_l})$$

are equal to their corresponding joint "success" probabilities and therefore it is a special case of Lemma 3.5.

REFERENCES

- Abul-Ela, A. A., Greenberg, B. G. and Horvitz, D. G. 1967. "A Multi Proportions Randomized Response Model." *ASA J.* Sept.: 990–1008.
- Aki S. 1985. "Discrete Distributions of Order k on a Binary Sequence." *Ann. Inst. Stat. Math.* 37: 205–224.
- Anderson, T. W. and Samvels, S. M. 1956. "Some Inequalities Among Binomial and Poisson Probabilities." *Fifth Berkeley Symposium* 1: 1–12.
- Dhar, S. K. and Jiang, X. "Probability Bounds on the Finite Sum of the Binary Sequence of Order k ." *Appl. Probability* (accepted).
- Gastwirth, J. L. 1977. "A Probability Model of a Pyramid Scheme." *Amer. Statistician* 31: 79–82.
- Hodges, J. L. and Le Cam, L. 1960. "The Poisson Approximation to the Poisson Binomial Distribution." *Ann. Math. Stat.* 3: 737–740.
- Hoefding, W. 1956. "On the Distribution of the Number of Successes in Independent Trials." *Ann. Math. Stat.* 27: 713–721.
- Kolmogorov, A. N. 1956. "Two Uniform Limit Theorems for Sums of Independent Random Variables." *Theory Prob. Appl.* 1: 384–394.
- Kuk, A. Y. C. 1990. "Asking Sensitive Questions Indirectly." *Biometrics* 77: 436–438.
- Percus O. E. and Percus, J. K. 1985. "Probability Bounds on the Sum of Independent Nonidentically Distributed Binomial Random Variables." *SIAM J. Appl. Math.*, 45: 621–640.
- Philipp, W. 1969. "The Central Limit Problem for Mixing Sequences of Random Variables." *Wahrscheinlichkeitstheorie verw. Geb.* 12: 155–171.
- Warner, S. L. 1965. "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias." *ASA J.* March: 63–69.