



HAL
open science

Kolmogorov inequalities for the partial sum of independent bernoulli random variables

Petroula M. Mavrikiou

► **To cite this version:**

Petroula M. Mavrikiou. Kolmogorov inequalities for the partial sum of independent bernoulli random variables. *Statistics and Probability Letters*, 2009, 77 (11), pp.1117. 10.1016/j.spl.2007.02.001 . hal-00565446

HAL Id: hal-00565446

<https://hal.science/hal-00565446>

Submitted on 13 Feb 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Author's Accepted Manuscript

Kolmogorov inequalities for the partial sum of independent bernoulli random variables

Petroula M. Mavrikiou

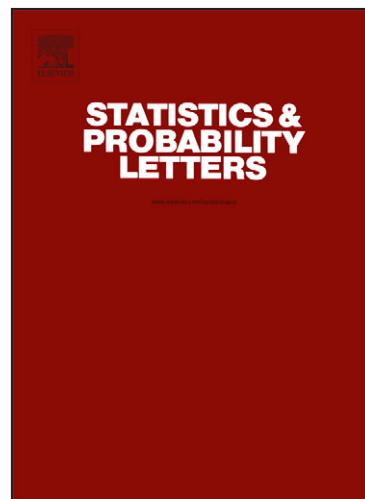
PII: S0167-7152(07)00063-6
DOI: doi:10.1016/j.spl.2007.02.001
Reference: STAPRO 4584

To appear in: *Statistics & Probability Letters*

Received date: 19 April 2006
Revised date: 21 December 2006
Accepted date: 21 February 2007

Cite this article as: Petroula M. Mavrikiou, Kolmogorov inequalities for the partial sum of independent bernoulli random variables, *Statistics & Probability Letters* (2007), doi:10.1016/j.spl.2007.02.001

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



www.elsevier.com/locate/stapro

Kolmogorov Inequalities for the Partial Sum of Independent Bernoulli Random Variables

Petroula M. Mavrikiou¹
Department of Business Administration
Frederick Institute of Technology
P.O. Box 24729
1303 Nicosia, Cyprus

Abstract

In this paper two Kolmogorov inequalities are presented for the sample average of independent (but not necessarily identically distributed) Bernoulli random variables.

Keywords and phrases: Kolmogorov inequalities, Bernoulli random variables.

¹Corresponding author: Tel: +357 22 345159 Ext 132; Fax: +357 22 438234
Email address: bus.mp@fit.ac.cy (P. M. Mavrikiou)

1. Introduction

For a sequence of independent and identically distributed Bernoulli random variables X_1, X_2, \dots , with $E(X_1) = p$, Kolmogorov (1963) provided the following inequality

$$P(\sup_{k \geq n} |\bar{X}_k - p| > \epsilon) \leq 2e^{-2n\epsilon^2(1-\epsilon)}$$

where

$$\bar{X}_k = \frac{1}{k} \sum_{i=1}^k X_i, \quad \epsilon > 0.$$

Improvements, extensions and related results can be found in Hoeffding (1963), Young, Seaman and Marco (1987), Turner, Young and Seaman (1992), Young, Turner and Seaman (1988), Christofides (1991), Christofides (1994), Kambo and Kotz (1966) and Banjevic (1985). In this paper, two Kolmogorov inequalities are provided for the case of independent but not necessarily identically distributed Bernoulli random variables.

2. Preliminaries

The following results will be used:

Lemma 2.1

Let p_1, \dots, p_n be positive real numbers and $\bar{p} = \frac{1}{n} \sum_{i=1}^n p_i$. Then for $t > 0$

$$\prod_{i=1}^n (p_i e^t + 1 - p_i) \leq (\bar{p} e^t + 1 - \bar{p})^n.$$

Proof: From the arithmetic-geometric mean inequality, we have

$$\begin{aligned} \prod_{i=1}^n (p_i e^t + 1 - p_i) &\leq \left[\sum_{i=1}^n \frac{1}{n} (p_i e^t + 1 - p_i) \right]^n \\ &= (\bar{p} e^t + 1 - \bar{p})^n. \end{aligned}$$

The following result is due to Christofides (1994).

Lemma 2.2

Let $\epsilon < \frac{1}{2}$ and

$$g(\bar{p}, \epsilon) = (1 - \bar{p} - \epsilon) \ln\left(\frac{1 - \bar{p} - \epsilon}{1 - \bar{p}}\right) + (\bar{p} + \epsilon) \ln\left(\frac{\bar{p} + \epsilon}{\bar{p}}\right).$$

Then for $\bar{p} + \epsilon < \frac{1}{2}$ or $\frac{1}{2} + \frac{1}{3}\epsilon \leq \bar{p} \leq 1$

$$g(\bar{p}, \epsilon) \geq -\frac{1}{2} \ln(1 - 4\epsilon^2).$$

Lemma 2.3

Let $x = 2(\bar{p} + \epsilon) - 1$ and $y = 1 - 2\bar{p}$. Then,

$$\sum_{r=1}^{\infty} \frac{1}{2r(2r-1)} [x^{2r} + (2r-1)y^{2r} + 2rxy^{2r-1}] = g(\bar{p}, \epsilon)$$

where $g(\bar{p}, \epsilon)$ is the function defined in Lemma 2.2.

Proof: We have $\bar{p} + \epsilon = \frac{x+1}{2}$ and $\bar{p} = \frac{1-y}{2}$. Then, $g(\bar{p}, \epsilon)$ from Lemma 2.2 is

$$g(\bar{p}, \epsilon) = \left(\frac{1-x}{2}\right) \ln\left(\frac{1-x}{1+y}\right) + \left(\frac{1+x}{2}\right) \ln\left(\frac{1+x}{1-y}\right).$$

Using the Taylor series expansions

$$\ln(1+x) = \sum_{r=1}^{\infty} (-1)^{r-1} \frac{x^r}{r} \quad \text{and} \quad \ln(1-x) = -\sum_{r=1}^{\infty} \frac{x^r}{r}$$

we have that

$$g(\bar{p}, \epsilon) = \frac{(1-x)}{2} \left[-\sum_{r=1}^{\infty} \frac{x^r}{r} - \sum_{r=1}^{\infty} (-1)^{r-1} \frac{y^r}{r} \right] + \frac{(1+x)}{2} \left[\sum_{r=1}^{\infty} (-1)^{r-1} \frac{x^r}{r} + \sum_{r=1}^{\infty} \frac{y^r}{r} \right]$$

and after algebraic manipulations we arrive at

$$g(\bar{p}, \epsilon) = \sum_{r=1}^{\infty} \frac{1}{2r(2r-1)} [x^{2r} + (2r-1)y^{2r} + 2rxy^{2r-1}].$$

Lemma 2.4

Let ν be a positive integer and $x > 1$. Then

$$2^{2\nu-2}(x^{2\nu-1} + 1) - (x+1)^{2\nu-1} > 0.$$

Proof: Let $F(x) = 2^{2\nu-2}(x^{2\nu-1} + 1) - (x+1)^{2\nu-1}$. Then

$$\begin{aligned} F'(x) &= 2^{2\nu-2}(2\nu-1)x^{2\nu-2} - (2\nu-1)(x+1)^{2\nu-2} \\ &= (2\nu-1)[2^{2\nu-2}x^{2\nu-2} - (x+1)^{2\nu-2}] \\ &= (2\nu-1)[(2x)^{2\nu-2} - (x+1)^{2\nu-2}] > 0 \text{ since } x > 1. \end{aligned}$$

Thus, $F(x)$ is an increasing function and $F(x) > F(1) = 0$.

Lemma 2.5

Let $y \geq 1$ and $\nu = 1, 2, \dots$. Then

$$H(y) = y^{2\nu} + 2\nu - 1 + 2\nu y - \frac{4}{2^{2\nu}}(y+1)^{2\nu} \geq 0.$$

Proof:

$$\begin{aligned} H'(y) &= 2\nu \cdot y^{2\nu-1} + 2\nu - \frac{4}{2^{2\nu}} 2\nu (y+1)^{2\nu-1} \\ &= 2\nu \frac{4}{2^{2\nu}} \left[\frac{2^{2\nu}}{4} y^{2\nu-1} + \frac{2^{2\nu}}{4} - (y+1)^{2\nu-1} \right] \end{aligned}$$

$$\begin{aligned}
&= 2^{3-2\nu} \cdot \nu [2^{2\nu-2} y^{2\nu-1} + 2^{2\nu-2} - (y+1)^{2\nu-1}] \\
&= 2^{3-2\nu} \cdot \nu [2^{2\nu-2} (y^{2\nu-1} + 1) - (y+1)^{2\nu-1}].
\end{aligned}$$

By Lemma 2.4 $H'(y) > 0$ implying that H is increasing and therefore $H(x) \geq H(1) = 0$.

Lemma 2.6

Let x, y be as in Lemma 2.3 and $r = 1, 2, \dots$. Then

$$x^{2r} + (2r-1)y^{2r} + 2rxy^{2r-1} \geq 4\left(\frac{x+y}{2}\right)^{2r}.$$

Proof: By lemma 2.5 for $c \geq 1$

$$c^{2r} + 2r - 1 + 2rc - \frac{4}{2^{2r}}(c+1)^{2r} \geq 0.$$

Take $c = \frac{x}{y}$. Then

$$\begin{aligned}
\frac{x^{2r}}{y^{2r}} + 2r - 1 + 2r\frac{x}{y} &\geq \frac{4}{2^{2r}}\left(\frac{x}{y} + 1\right)^{2r} \\
x^{2r} + (2r-1)y^{2r} + 2rxy^{2r-1} &\geq 4\left(\frac{x+y}{2}\right)^{2r}.
\end{aligned}$$

3. Main Results

Theorem 3.1

Let Y_1, Y_2, \dots, Y_n be a sequence of independent Bernoulli random variables with $E(Y_i) = p_i$, $i = 1, \dots, n$ and $\epsilon < \frac{1}{2}$. Then, for $\bar{p} + \epsilon < \frac{1}{2}$ or $\bar{p} \geq \frac{1}{2} + \frac{1}{3}\epsilon$,

$$P(\bar{Y} - \bar{p} > \epsilon) \leq (1 - 4\epsilon^2)^{\frac{n}{2}}$$

where $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$ and $\bar{p} = \frac{1}{n} \sum_{i=1}^n p_i$.

Proof: Let $s > 0$. Then,

$$\begin{aligned}
P(\bar{Y} - \bar{p} > \epsilon) &= P[s(\bar{Y} - \bar{p} - \epsilon) > 0] \\
&\leq E[e^{s(\bar{Y} - \bar{p} - \epsilon)}] \\
&= e^{-s(\bar{p} + \epsilon)} E[e^{s\bar{Y}}] \\
&= e^{-s(\bar{p} + \epsilon)} E\left(e^{s \frac{1}{n} \sum_{i=1}^n Y_i}\right) \\
&= e^{-s(\bar{p} + \epsilon)} \prod_{i=1}^n E\left(e^{\frac{s}{n} Y_i}\right) \\
&= e^{-s(\bar{p} + \epsilon)} \prod_{i=1}^n (p_i e^{\frac{s}{n}} + 1 - p_i) \\
&\leq e^{-s(\bar{p} + \epsilon)} (\bar{p} e^{\frac{s}{n}} + 1 - \bar{p})^n \tag{1} \\
&= e^{-f(s)}
\end{aligned}$$

where $f(s) = s(\bar{p} + \epsilon) - n \ln(\bar{p} e^{\frac{s}{n}} + 1 - \bar{p})$ and the last inequality follows from Lemma 2.1.

The function f is maximized at $s^{max} = n \ln\left[\frac{(1-\bar{p})(\bar{p}+\epsilon)}{\bar{p}(1-\bar{p}-\epsilon)}\right]$ and

$$\begin{aligned}
f(s^{max}) &= n(\bar{p} + \epsilon) \ln\left[\frac{(\bar{p} + \epsilon)(1 - \bar{p})}{\bar{p}(1 - \bar{p} - \epsilon)}\right] - n \ln\left[\frac{1 - \bar{p}}{1 - \bar{p} - \epsilon}\right] \\
&= n[(1 - \bar{p} - \epsilon) \ln\left(\frac{1 - \bar{p} - \epsilon}{1 - \bar{p}}\right) + (\bar{p} + \epsilon) \ln\left(\frac{\bar{p} + \epsilon}{\bar{p}}\right)] \\
&= n.g(\bar{p}, \epsilon).
\end{aligned}$$

Thus, by Lemma 2.2

$$f(s^{max}) = n.g(\bar{p}, \epsilon) \geq -\frac{n}{2} \ln(1 - 4\epsilon^2)$$

and therefore

$$P(\bar{Y} - \bar{p} > \epsilon) \leq (1 - 4\epsilon^2)^{\frac{n}{2}}.$$

Notice that the restrictions on \bar{p} and ϵ arise from Lemma 2.2 which is due to Christofides (1994).

The following theorem gives an exponential bound under different conditions on \bar{p} and ϵ .

Theorem 3.2

Let Y_1, Y_2, \dots, Y_n be a sequence of independent Bernoulli random variables, with $E(Y_i) = p_i$, $i = 1, \dots, n$. Then for $\bar{p} + \epsilon > \frac{1}{2}$ or $\bar{p} < \frac{1}{2}$ and $\forall \epsilon < 1$,

$$P(\bar{Y} - \bar{p} > \epsilon) \leq e^{-n(2\epsilon^2 + \frac{1}{3}\epsilon^4 e^{\frac{\epsilon^2}{4}})}$$

where $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$ and $\bar{p} = \frac{1}{n} \sum_{i=1}^n p_i$.

Proof: From the proof of Theorem 3.1

$$P(\bar{Y} - \bar{p} > \epsilon) \leq e^{-ng(\bar{p}, \epsilon)}$$

where

$$g(\bar{p}, \epsilon) = (1 - \bar{p} - \epsilon) \ln\left(\frac{1 - \bar{p} - \epsilon}{1 - \bar{p}}\right) + (\bar{p} + \epsilon) \ln\left(\frac{\bar{p} + \epsilon}{\bar{p}}\right).$$

By Lemma 2.3

$$g(\bar{p}, \epsilon) = \sum_{r=1}^{\infty} \frac{1}{2r(2r-1)} [x^{2r} + (2r-1)y^{2r} + 2rxy^{2r-1}].$$

x and y are positive values, imposing the restrictions on \bar{p} and ϵ .

Now, since $x + y = 2\epsilon$ and using Lemma 2.6 we have

$$\begin{aligned}
g(\bar{p}, \epsilon) &\geq \sum_{r=1}^{\infty} \frac{1}{2r(2r-1)} \cdot 4 \left(\frac{2\epsilon}{2}\right)^{2r} \\
&= \sum_{r=1}^{\infty} \frac{2\epsilon^{2r}}{r(2r-1)} \\
&= 2\epsilon^2 + \epsilon^4 \sum_{r=2}^{\infty} \frac{2(\epsilon)^{2r-4}}{r(2r-1)} \\
&\geq 2\epsilon^2 + 2\epsilon^4 \sum_{r=2}^{\infty} \frac{(\epsilon^2)^{r-2}}{6.4^{r-2}(r-2)!}
\end{aligned}$$

since $r(2r-1) \leq 6.4^{r-2}(r-2)!$ for $r \geq 2$. The last inequality is clearly true for $r = 2$, and can be established (e.g., by induction) for $r > 2$. Then,

$$\begin{aligned}
g(\bar{p}, \epsilon) &\geq 2\epsilon^2 + 2\epsilon^4 \frac{1}{6} \sum_{r=2}^{\infty} \left(\frac{\epsilon^2}{4}\right)^{r-2} \frac{1}{(r-2)!} \\
&= 2\epsilon^2 + \frac{1}{3}\epsilon^4 \cdot e^{\frac{\epsilon^2}{4}}.
\end{aligned}$$

Thus

$$P(\bar{Y} - \bar{p} > \epsilon) \leq e^{-n(2\epsilon^2 + \frac{1}{3}\epsilon^4 \cdot e^{\frac{\epsilon^2}{4}})}$$

and the proof is complete.

Theorem 3.3

Let Y_1, Y_2, \dots , be a sequence of independent Bernoulli random variables with $E(Y_i) = p_i$. Then,

(1) $\forall \epsilon < \frac{1}{2}$ and for $\bar{p}_k + \epsilon < \frac{1}{2}$ or $\bar{p}_k \geq \frac{1}{2} + \frac{1}{3}\epsilon$,

$$P[\sup_{k \geq n} (\bar{Y}_k - \bar{p}_k) > \epsilon] \leq (1 - 4\epsilon^2)^{\frac{n}{2}} \quad n = 0, 1, 2, \dots$$

while

(2) $\forall \epsilon < 1$ and for $\bar{p}_k + \epsilon > \frac{1}{2}$ or $\bar{p}_k < \frac{1}{2}$,

$$P[\sup_{k \geq n} (\bar{Y}_k - \bar{p}_k) > \epsilon] \leq e^{-n(2\epsilon^2 + \frac{1}{3}\epsilon^4 e^{\frac{\epsilon^2}{4}})} \quad n = 0, 1, 2, \dots$$

where $\bar{Y}_k = \frac{1}{k} \sum_{i=1}^k Y_i$ and $\bar{p}_k = \frac{1}{k} \sum_{i=1}^k p_i$.

Proof: Using Lemma 1 of Turner, Young and Seaman (1995) we can arrive at inequality (1) of Theorem 3.1 having the quantity $P[\sup_{k \geq n} (\bar{Y}_k - \bar{p}_k) > \epsilon]$ as our left-hand side. Then, one can follow the same steps of theorems 3.1 and 3.2 to reach the right-hand side, and the proof is complete.

Remarks

(1) Theorem 3.3 provides sharper bounds than that of the main result of Turner, Young and Seaman (1995), under of course restrictions on \bar{p} and ϵ .

(2) Theorem 3.1 is an extension of Corollary 3.2 of Christofides (1994).

References

- Banjevic, D. (1985), On a Kolmogorov inequality, *Theory Probab. Appl.* **29**, 391-394.
- Christofides T.C. (1991), Probability inequalities with exponential bounds for U-statistics, *Statist. Probab. Lett.* **12**, 257-261.
- Christofides T.C. (1994), A Kolmogorov inequality for U-statistics based on Bernoulli kernels, *Statist. Probab. Lett.*, **21**, 357-362.
- Hoeffding, W. (1963), Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.* **58**, 13-30.
- Kambo, N.S. and S. Kotz (1966), On exponential bounds for binomial probabilities, *Ann. Inst. Statist. Math.* **18**, 277-287.
- Kolmogorov, A.N. (1963), On tables of random numbers, *Sankhyā*, Ser. A **25**, 369-376.
- Turner, D.W., D.M. Young and J.W. Seaman (1992), Improved Kolmogorov inequalities for the binomial distribution, *Statist. Probab. Lett.* **13**, 223-227.
- Turner, D.W., D.M. Young and J.W. Seaman (1995), A Kolmogorov inequality for the sum of independent Bernoulli random variables with unequal means, *Statist. Probab. Lett.* **23**, 243-245.
- Young, D.M., J.W. Seaman and V.R. Marco (1987), A note on a Kolmogorov inequality, *Statist. Probab. Lett.* **5**, 217-218.
- Young D.M., D.W. Turner and J.W. Seaman (1988), A note on a Kolmogorov inequality for the Binomial distribution, *Theory Probab. Appl.* **33**, 747-749.