



HAL
open science

A Kolmogorov inequality for weighted U-statistics

Petroula M. Mavrikiou

► **To cite this version:**

Petroula M. Mavrikiou. A Kolmogorov inequality for weighted U-statistics. *Statistics and Probability Letters*, 2009, 78 (18), pp.3294. 10.1016/j.spl.2008.06.013 . hal-00508475

HAL Id: hal-00508475

<https://hal.science/hal-00508475>

Submitted on 4 Aug 2010

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.

Accepted Manuscript

A Kolmogorov inequality for weighted U-statistics

Petroula M. Mavrikiou

PII: S0167-7152(08)00317-9
DOI: 10.1016/j.spl.2008.06.013
Reference: STAPRO 5130

To appear in: *Statistics and Probability Letters*

Received date: 18 June 2007
Revised date: 28 April 2008
Accepted date: 2 June 2008

Please cite this article as: Mavrikiou, P.M., A Kolmogorov inequality for weighted U-statistics. *Statistics and Probability Letters* (2008), doi:10.1016/j.spl.2008.06.013

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 proof before it is published in its final 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.



A Kolmogorov inequality for weighted U-statistics

Petroula M. Mavrikiou¹
Department of Business Administration
School of Economic Sciences and Administration
Frederick University Cyprus
P.O. Box 24729
1303 Nicosia, Cyprus

Abstract

In this paper a Kolmogorov probability inequality for weighted U-statistics based on Bernoulli kernels is presented. This inequality which extends the results of Turner et al. (1995) is a Hoeffding type exponential inequality without any assumptions or restrictions.

Keywords and phrases: Bernoulli kernels, weighted U-statistics.

¹Corresponding author. Tel: +35722345159 Ext 132; Fax: +35722438234; Email address: bus.mp@fit.ac.cy

1. Introduction

Let X_1, X_2, \dots, X_n be independent random variables from a distribution F . Consider a parametric function $\theta = \theta(F)$ for which there is an unbiased estimator. That is,

$$\theta = \theta(F) = Eh(X_1, \dots, X_m), \quad m \leq n,$$

for a function $h = h(X_1, \dots, X_m)$ which is assumed to be symmetric without any loss of generality. Let w_{i_1, \dots, i_m} be non-negative numbers such that

$$\sum_{1 \leq i_1 < \dots < i_m \leq n} w_{i_1, \dots, i_m} = \binom{n}{m}.$$

We define the weighted U-statistic for estimation of the parameter $\theta = \theta(F)$ as

$$U_n = \binom{n}{m}^{-1} \sum_{1 \leq i_1 < \dots < i_m \leq n} w_{i_1, \dots, i_m} h(X_{i_1}, \dots, X_{i_m}),$$

where $\sum_{1 \leq i_1 < \dots < i_m \leq n}$ denotes summation over the $\binom{n}{m}$ combinations of (i_1, \dots, i_m) from $\{1, \dots, n\}$.

U_n is an unbiased estimator of θ . In the case where $w_{i_1, \dots, i_m} = 1$ for all $\binom{n}{m}$ combinations of (i_1, \dots, i_m) from $\{1, \dots, n\}$ we have the usual unweighted U-statistic whose theory and applications can be found in many references, e.g. Serfling (1980).

For many U-statistics of interest the kernel h is a Bernoulli random variable (e.g., indicator function). Improvements, extensions and results related to Kolmogorov inequalities can be found among others, in Young et. al (1987), Turner et. al (1995) and Mavrikiou (2007). Exponential and Kolmogorov inequalities have been constructed for U-statistics based on Bernoulli kernels in Christofides (1991, 1994).

In this paper a Kolmogorov probability inequality for weighted U-statistics based on Bernoulli kernels is presented. This result generalize the inequality found in Turner et al. (1995).

2. Preliminaries

For proving the result, a representation of a weighted U-statistic as an average of averages of independent random variables is used. The representation is a straightforward extension of the one introduced and utilized by Hoeffding (1963).

Lemma 2.1. *Let $k = [n/m]$ be the greatest integer less than or equal to n/m , and let U_n be the weighted U-statistic introduced in Section 1.*

Then,

$$U_n = \frac{1}{n!} \sum_p H_{i_1, \dots, i_n}$$

where $H_{i_1, \dots, i_n} = \frac{1}{k} [h(X_{i_1}, \dots, X_{i_m})w_{i_1, \dots, i_m} + h(X_{i_{m+1}}, \dots, X_{i_{2m}})w_{i_{m+1}, \dots, i_{2m}} + \dots$
 $\dots + h(X_{i_{k(m-1)+1}, \dots, X_{i_{km}}})w_{i_{k(m-1)+1}, \dots, i_{km}}]$

and \sum_p denotes summation over the $n!$ permutations of (i_1, \dots, i_n) from $\{1, \dots, n\}$.

Proof. We may write

$$k \sum_p H_{i_1, \dots, i_n} = k \cdot m!(n-m)! \sum_c h(X_{i_1}, \dots, X_{i_m})w_{i_1, \dots, i_m}$$

or

$$\sum_p H_{i_1, \dots, i_n} = \frac{n!}{\binom{n}{m}} \sum_c h(X_{i_1}, \dots, X_{i_m})w_{i_1, \dots, i_m}$$

thus,

$$U_n = \frac{1}{n!} \sum_p H_{i_1, \dots, i_n}.$$

In addition to the previous lemma we will use the following result which is due to Hoeffding (1963).

Lemma 2.2. *Let X be a random variable such that $c \leq X \leq b$ and $EX = \mu$. Then,*

$$Ee^{t(X-\mu)} \leq e^{\frac{1}{8}t^2(b-c)^2} \text{ for } t > 0.$$

3. Main Result

Throughout this section will be assumed that X_1, X_2, \dots, X_n is a sequence of independent random variables and $h(X_{i_1}, \dots, X_{i_m})$ is a Bernoulli kernel with

$$Eh(X_{i_1}, \dots, X_{i_m}) = p_{i_1, \dots, i_m}.$$

In addition \bar{p} will denote the expected value of the weighted U-statistic based on the kernel h , i.e.,

$$\bar{p} = EU_n = \binom{n}{m}^{-1} \sum_{1 \leq i_1 < \dots < i_m \leq n} w_{i_1, \dots, i_m} p_{i_1, \dots, i_m}.$$

The following result is a Hoeffding type exponential inequality for weighted U-statistics without any assumptions on w_{i_1, \dots, i_m} 's and the p_{i_1, \dots, i_m} 's.

Theorem 3.1. *Let U_n be a weighted U-statistic based on a Bernoulli kernel. For $\epsilon > 0$ and $k = \lceil n/m \rceil$*

$$P(U_n - \bar{p} \geq \epsilon) \leq \frac{1}{n!} \sum_p \frac{1}{k} \sum_{j=0}^{k-1} \exp\left(\frac{-2k\epsilon^2}{w_{i_{mj+1}, \dots, i_{mj+m}}^2}\right).$$

Proof. Let s be an arbitrary positive number. Then, using Markov's inequality and the representation of the weighted U-statistic as an average of averages of independent random variables we have

$$\begin{aligned} P(U_n - \bar{p} \geq \epsilon) &= P[sU_n \geq s(\epsilon + \bar{p})] \\ &\leq e^{-s(\epsilon + \bar{p})} Ee^{sU_n} \\ &= e^{-s(\epsilon + \bar{p})} E \exp\left(s \frac{1}{n!} \sum_p H_{i_1, \dots, i_n}\right) \\ &= e^{-s(\epsilon + \bar{p})} E \exp\left[s \frac{1}{n!} \sum_p \frac{1}{k} \sum_{j=0}^{k-1} h(X_{i_{mj+1}}, \dots, X_{i_{mj+m}}) w_{i_{mj+1}, \dots, i_{mj+m}}\right]. \end{aligned} \quad (1)$$

Clearly,

$$\binom{n}{m}^{-1} \sum_{1 \leq i_1 < \dots < i_m \leq n} h(X_{i_1}, \dots, X_{i_m}) w_{i_1, \dots, i_m} = \frac{1}{n!} \sum_p \frac{1}{k} \sum_{j=0}^{k-1} h(X_{i_{mj+1}}, \dots, X_{i_{mj+m}}) w_{i_{mj+1}, \dots, i_{mj+m}}$$

and therefore using expected value

$$\bar{p} = \binom{n}{m}^{-1} \sum_{1 \leq i_1 < \dots < i_m \leq n} p_{i_1, \dots, i_m} w_{i_1, \dots, i_m} = \frac{1}{n!} \sum_p \frac{1}{k} \sum_{j=0}^{k-1} p_{i_{mj+1}, \dots, i_{mj+m}} w_{i_{mj+1}, \dots, i_{mj+m}}.$$

Thus,

$$P(U_n - \bar{p} \geq \epsilon) \leq e^{-s\epsilon} \text{Exp} \left\{ s \frac{1}{n!} \sum_p \frac{1}{k} \sum_{j=0}^{k-1} [h(X_{i_{mj+1}}, \dots, X_{i_{mj+m}}) - p_{i_{mj+1}, \dots, i_{mj+m}}] w_{i_{mj+1}, \dots, i_{mj+m}} \right\}$$

$$\leq e^{-s\epsilon} \frac{1}{n!} \sum_p \text{Exp} \left\{ \frac{s}{k} \sum_{j=0}^{k-1} [h(X_{i_{mj+1}}, \dots, X_{i_{mj+m}}) - p_{i_{mj+1}, \dots, i_{mj+m}}] w_{i_{mj+1}, \dots, i_{mj+m}} \right\} \quad (2)$$

$$= e^{-s\epsilon} \frac{1}{n!} \sum_p \prod_{j=0}^{k-1} \text{Exp} \left\{ \frac{s}{k} [h(X_{i_{mj+1}}, \dots, X_{i_{mj+m}}) - p_{i_{mj+1}, \dots, i_{mj+m}}] w_{i_{mj+1}, \dots, i_{mj+m}} \right\} \quad (3)$$

$$= \frac{1}{n!} \sum_p \prod_{j=0}^{k-1} e^{-\frac{s\epsilon}{k}} \text{Exp} \left\{ \frac{s}{k} [h(X_{i_{mj+1}}, \dots, X_{i_{mj+m}}) - p_{i_{mj+1}, \dots, i_{mj+m}}] w_{i_{mj+1}, \dots, i_{mj+m}} \right\}$$

$$\leq \frac{1}{n!} \sum_p \prod_{j=0}^{k-1} e^{-\frac{s\epsilon}{k}} \cdot e^{\frac{s^2}{8k^2} w_{i_{mj+1}, \dots, i_{mj+m}}^2} \quad (4)$$

$$= \frac{1}{n!} \sum_p \prod_{j=0}^{k-1} \exp \left(-\frac{s\epsilon}{k} + \frac{s^2}{8k^2} w_{i_{mj+1}, \dots, i_{mj+m}}^2 \right)$$

$$= \frac{1}{n!} \sum_p \prod_{j=0}^{k-1} \exp[-g(s)]. \quad (5)$$

Inequality (2) is due to the convexity of the exponential function, equality (3) follows from independence and finally, inequality (4) is due to Lemma 2.2.

For notational simplicity let $w_{i_{mj+1}, \dots, i_{mj+m}} = w$. Therefore, the function $g(s)$ in equation (5) equals to $g(s) = \frac{s\epsilon}{k} - \frac{1}{8} \frac{s^2 w^2}{k^2}$.

Maximizing $g(s)$ we get $s^{max} = \frac{4\epsilon k}{w^2}$ and $g(s^{max}) = \frac{2\epsilon^2}{w^2}$, implying that

$$P(U_n - \bar{p} \geq \epsilon) \leq \frac{1}{n!} \sum_p \prod_{j=0}^{k-1} \exp \left(\frac{-2\epsilon^2}{w_{i_{mj+1}, \dots, i_{mj+m}}^2} \right)$$

$$\begin{aligned}
&\leq \frac{1}{n!} \sum_p \left[\frac{1}{k} \sum_{j=0}^{k-1} \exp\left(\frac{-2\epsilon^2}{w_{i_{mj+1}, \dots, i_{mj+m}}^2}\right) \right]^k \quad (6) \\
&\leq \frac{1}{n!} \sum_p \frac{1}{k^k} k^{k-1} \sum_{j=0}^{k-1} \exp\left(\frac{-2k\epsilon^2}{w_{i_{mj+1}, \dots, i_{mj+m}}^2}\right) \\
&= \frac{1}{n!} \sum_p \frac{1}{k} \sum_{j=0}^{k-1} \exp\left(\frac{-2k\epsilon^2}{w_{i_{mj+1}, \dots, i_{mj+m}}^2}\right).
\end{aligned}$$

Notice that inequality (6) is due to the arithmetic-geometric mean while the following one is due to the elementary inequality $(\sum_{i=1}^n x_i)^k \leq n^{k-1} \sum_{i=1}^n x_i^k$ for $k > 1$. Thus, the proof of the theorem is now complete.

Remarks

- (i) Although in Section 1 the weighted U-statistic is defined based on independent and identically distributed random variables, the result proved in this paper does not require that the observations are identically distributed.
- (ii) For the special case of the weighted U-statistic being the sample weighted average of independent Bernoulli random variables Theorem 3.1, can be stated as:

Corollary 3.2. *Let Y_1, \dots, Y_n be independent Bernoulli random variables with $E(Y_i) = p_i$ for $i = 1, \dots, n$. Let $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i w_i$, $\bar{p} = \frac{1}{n} \sum_{i=1}^n p_i w_i$, $\epsilon > 0$ and $w_i, i = 1, \dots, n$ nonnegative numbers with $\frac{1}{n} \sum_{i=1}^n w_i = 1$. Then,*

$$P(\bar{Y} - \bar{p} \geq \epsilon) \leq e^{-2n\epsilon^2} \quad \epsilon > 0 \quad (7)$$

- (iii) Inequality (7) extends the result of Turner et al. (1995) to the case of weighted Bernoulli random variables.

Acknowledgments

The author would like to thank the referee for helpful comments.

References

- 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.
- Mavrikiou, P.M., 2007. Kolmogorov inequalities for the partial sum of independent Bernoulli random variables, *Statist. Probab. Lett.*, **77**, 1117-1122 doi:10.1016/j.spl.2007.02.001.
- Serfling, R.J., 1980. *Approximation Theorems of Mathematical Statistics* (Wiley, New York).
- Young, D.M., Seaman, W.J., and Marco, R.V., 1987. A note on a Kolmogorov inequality, *Statist. Probab. Lett.*, **5**, 217-218
- Turner, D.W., Young, D.M. and Seaman, J.W., 1995. A Kolmogorov inequality for the sum of independent Bernoulli random variables with unequal means, *Statist. Probab. Lett.* **23**, 243-245.