Calcutta Statistical Association Bulletin
Vol. 47, March & June 1997, Nos. 185-186

ON EXTENSIONS OF AN INEQUALITY OF KOLMOGOROV

TAPAS K. CHANDRA AND SUBHASHIS GHOSAL

Indian Statistical Institute, Calcutta

ABSTRACT : The inequality of Kolmogorov (Sankhya, 1963) has been extended to a sequence of independent sub-Gaussian and other random variables. All the earlier results in the literature on this problem concerned only on the very special case of Bernoulli variables. We use martingale inequalities to establish a key result. A similar inequality is also proved for U-statistics based on exchangeable random variables.

Key words and phrases : Exchangeability; martingale; probability inequalities; U-statistics.

AMS (1991) Subject Classification : Primary 60E15, 60G42.

1. INTRODUCTION

Kolmogorov (1963) stated without proof that if $\{X_n\}$ are independent and identically distributed (iid) Bernoulli variables with probability of a success p $(0 < p < 1)$, then

$$P\left\{\sup_{k \geq n} |\overline{X_k} - p| \geq \varepsilon\right\} \leq 2\exp[-2n\varepsilon^2(1-\varepsilon)],\ \varepsilon \geq 0,\ n \geq 1. \qquad (1.1)$$

where $\overline{X}_k = k^{-1}\sum_{j=1}^{k} X_j$. This inequality has been sharpened by Banjevic (1984) and later by Young, Seaman and Marco (1987) as well; as a consequence, the bound on the right hand side (RHS) of (1.1) can, in fact, be improved to $2\exp(-2n\varepsilon^2)$ for all $0 \leq \varepsilon \leq 1$ and $n \geq 1$. Further improvements are obtained by Young, Turner and Seaman (1988) and Turner, Young and Seaman (1992,1995). In this paper, we shall establish a similar exponential bound for a class of independent random variables (rvs) *which need not be indentically distributed nor need be Bernoulli*. We shall use martingale techniques to establish a key result (Theorem 2.1) which is subsequently used to extend the Kolmogorov inequality (1.1). An analogous result for U-statistics based on exchangeable rvs is also provided with.

2. MAIN RESULTS

Let $\{X_n\}$, $n \geq 1$, be an independent sequence of rvs with finite moment generating functions (MGF) $M_n(t) := E(\exp(tX_n))$ for $t \geq 0$. Let $S_n = \sum_{j=1}^{n} X_j$ and let $\Re_n$ be the σ-field generated by $X_1,...,X_n$, $n \geq 1$.

THEOREM 2.1. *Under the above set up, we have*

$$\{P\left\{\sup_{k \geq n} \bar{X}_k > 0\right\} \leq \inf_{t \geq 0} \sup_{k \geq n} \prod_{j=1}^{k} M_j(t), \quad n \geq 1. \tag{2.1}$$

Proof. Fix $t > 0$ and let $Y_n = \exp(tS_n)/\prod_{j=1}^{n} M_j(t))$. Clearly, $\{Y_n, \Re_n\}$ is a positive martingale having means 1. Now for any $N>n$,

$$P\left\{\max_{n \leq k \leq N} \bar{X}_k \geq 0\right\} = P\left\{\max_{n \leq k \leq N} S_k \geq 0\right\}$$

$$= P\left\{\max_{n \leq k \leq N} \exp(t S_k) \geq 1\right\}$$

$$= P\left\{\max_{n \leq k \leq N} c_k Y_k \geq 1\right\} \tag{2.2}$$

where $c_k = \prod_{j=1}^{k} M_j(t) = E(\exp(tS_k))$. By Doob's maximal inequality (see, for example, Chow and Teicher (1988, p.247)),

$$P\left\{\max_{n \leq k \leq N} c_k Y_k \geq 1\right\} \leq P\left\{(\max_{n \leq k \leq N} c_k)(\max_{n \leq k \leq N} Y_k) \geq 1\right\}$$

$$\leq \left(\max_{n \leq k \leq N} c_k\right) EY_N = \max_{n \leq k \leq N} c_k \tag{2.3}$$

Hence (2.2) and (2.3) together yield

$$P\left\{\sup_{k \geq n} \bar{X}_k > 0\right\} \leq \lim_{N \to \infty} P\left\{\sup_{n \leq k \leq N} \bar{X}_k > 0\right\} \leq \sum_{k \geq n} \prod_{j=1}^{k} M_j(t) \tag{2.4}$$

Now (2.1) follows by taking infimum over $t \geq 0$. **Q.E.D**

COROLLARY 2.1. *In addition to the hypotheses of Theorem 2.1, suppose that $E(\exp(t_0 S_j)) \leq E(\exp(t_0 S_m))$ for all $j \geq n$ where $t_0 > 0$ and $m \geq n$, then $P\{\sup_{k \geq n} \bar{X}_k \geq 0\} \leq E(\exp(t_0 S_m))$.*

Definition 2.1. A rv X with finite expectation is said to be a sub-Gaussian variable with parameters $\alpha \geq 0$ and $\beta > 0$, X is SG(α,β) in notation, if

$$E(\exp[t(X-EX)]) \leq \exp(\alpha t^2) \text{ for all } t \in [0,\beta).$$

It is shown in Hoeffding (1963) that if $P(\alpha \leq X \leq \beta) = 1$ and $EX = \mu$, then $E(\exp[t(X-\mu)]) \leq \exp[t^2(\beta-\alpha)^2/8]$ for all $t \geq 0$; thus a bounded rv is sub-Gaussian with $\beta = \infty$. Also, any $N(\mu,\sigma^2)$ variable with $0 \leq \sigma^2 < \infty$ is trivially sub-Gaussian with $\beta = \infty$.

Definition 2.2. Let $\{X_n\}$ be a sequence of independent rvs with finite MGF on $[0,\beta)$. Then $\{X_n\}$ is said to satisfy the condition (*) if there exist functions $f(\alpha_n,t)$ such that

$$E(\exp[t(X_j-EX_j)]) \leq f(\alpha_j,t), \ t \in [0,\beta), \ j \geq 1$$

and

$$\prod_{i=1}^{j} f(\alpha_i, t) \leq (f(\bar{\alpha}_j, t))^j, \ t \in (0,\beta), \ j \geq 1$$

where $\bar{\alpha}_j = (\alpha_1 + \ldots + \alpha_j)/j$.

Thus a sequence of sub-Gaussian independent rvs satisfies the condition (*).

Definition 2.3. A rv X with finite expectation is said to be a *sub-Bernoulli variable with parameters* $p \in [0,1]$ and $\beta > 0$, X is SB(p,β) in notation, if

$$E(\exp(t(X-EX))) \leq e^{-pt}(1-p+pe^t), \ t \in [0, \beta).$$

From the elementary inequality $e^{tx} \leq 1 + x(e^t-1)$, $0 \leq x \leq 1$, $t \geq 0$, it follows that, for a rv X with $P(0 \leq X \leq 1) = 1$ and $EX = \mu$, $E(\exp(tX)) \leq 1-\mu + \mu e^t$, $t \geq 0$. Thus X is SB(μ,β) with $\beta = \infty$. A sequence of independent sub-Bernoulli rvs also satisfies the condition (*).

THEOREM 2.2. *Let $n \geq 1$ be a fixed integer and $\varepsilon > 0$. Let $\{X_i\}$ be a sequence of independent rvs such that X_i is SG(α_i,β_i). Put $\bar{\alpha}_n=(\alpha_1 + \ldots + \alpha_n)/n$. Assume that $\varepsilon/(2\bar{\alpha}_n) < \beta_j$ for all $j \geq n$ and $\bar{\alpha}_{nj} \leq 2\bar{\alpha}_n$ for all $j>n$, where $\bar{\alpha}_{nj} = (j-n)^{-1} \sum_{i=n+1}^{j} \alpha_i$.* Then

$$P\left\{\sup_{k \geq n} (\bar{X}_k - E\bar{X}_k) \geq \varepsilon\right\} \leq \exp[-n\varepsilon^2/(4\bar{\alpha}_n)]. \tag{2.5}$$

Proof. Put $t_0 = \varepsilon/(2\bar{\alpha}_n)$. Note that $\forall j > n$

$$E(exp\,[t_0(S_j - ES_j - j\,\varepsilon)])/E(exp\,[t_0(S_n - ES_n - n\varepsilon])$$

$$= \prod_{i=n+1}^{j} E(exp\,[t_0(X_i - EX_i - \varepsilon)])$$

$$\leq \prod_{i=n+1}^{j} \exp\,[-t_0\varepsilon + \alpha_i\; t_0^2]$$

$$= (exp\,[-(j-n)\,t_0\,\varepsilon + (j-n)\,\overline{\alpha}_{nj}\,t_0^2] \tag{2.6}$$

$$\leq (exp\,[-t_0\varepsilon + 2t_0^2\,\overline{\alpha}_n])^{j-n} = 1$$

Applying Corollary 2.1 with m = n and $X_i - EX_i - \varepsilon$ in place of X_i, we have

$$P\left\{\sup_{k \geq n}(\overline{X}_k - E\overline{X}_k) > \varepsilon\right\}$$

$$\leq E(\exp\,[t_0(S_n - ES_n - n\varepsilon)])$$

$$\leq \exp\,[-n\,t_0\,\varepsilon + n\,\overline{\alpha}_n\,t_0^2]$$

$$= \exp\,[-n\varepsilon^2/(4\overline{\alpha}_n)]. \tag{2.7}$$

The RHS is continuous in ε. Replacing ε by ε - η and letting $\eta \downarrow 0$ in (2.7), the desired inequality (2.5) follows. **Q. E. D.**

COROLLARY 2.2. If in Theorem 2.2 $\alpha_n \leq \alpha$ for all $n \geq 1$, then for each $n \geq 1$,

$$P\left\{\sup_{k \geq n}(\overline{X}_k - E\,\overline{X}_k) \geq \varepsilon\right\} \leq \exp\,[-n\,\varepsilon^2/(4\alpha)]. \tag{2.8}$$

In particular, if for all $n \geq 1$, $P\{a_n \leq X_n \leq b_n\}=1$ with $(b_n - a_n) \leq c$, $n \geq 1$, then

$$P\left\{\sup_{k \leq n}(\overline{X}_k - E\,\overline{X}_k) \geq \varepsilon\right\} \leq \exp\,(-2n\,\varepsilon^2/c^2), \quad \varepsilon > 0,\, n \geq 1.$$

Remark 2.1. The results of Young, Seaman and Marco (1987) and Turner, Young amd Seaman (1995) are special cases of Corollary 2.2. Also, Theorem 2 of Hoeffding (1963) is a special case of Theorem 2.2 (by letting $0 = X_{n+1} = X_{n+2} = ...$). Finally, it is clear that a version of Theorem 2.2 can be proved if the condition "the X_i's are sub-Gaussian" is replaced by the condition "the X_i's satisfy the condition (*)".

Remark 2.2. Consider the set up of Inequality (1.1) due to Kolmogorov (1963). It is known that

$$P\left\{\sup_{k \geq n} |\bar{X}_k - p| \geq \varepsilon\right\} \leq 2\exp\left[-n\left(2\varepsilon^2 + (4/3)\,\varepsilon^4 \exp(\varepsilon^2)\right)\right] \qquad (2.10)$$

provided either $p \geq 1/2$ or $p + \varepsilon \leq 1/2$ (see, for example, Young, Turner and Seaman (1988) and Christofides (1971)). As $\bar{X}_n \to p$ a.s., it is a question of great importance to find an integer $N(\varepsilon,n)$, as small as possible at least for moderately small values of ε and η, satisfying the condition that

$$P\{|\bar{X}_n - p| < \varepsilon \text{ for all } n \geq N(\varepsilon,\eta)\} > 1-\eta. \qquad (2.11)$$

This is an old problem of the probability theory which had been investigated, for example, by Bernoulli and Cantelli (see Uspensky (1937, pp. 101-103)). Using the above inequality, one can see that if $\varepsilon = 0.01$ and $\eta = 0.001$, then $N(\varepsilon,\eta) \geq 38002$. The best possible value of $N(\varepsilon,n)$ appears to be not yet known.

3. The Uniformly Bounded Case

In view of the results of Turner, Young and Seaman (1992), Hoeffding (1963) and Kraft (1969), we discuss now the case of uniformly bounded independent rvs. Without loss of generality, we can assume that $0 \leq X_n \leq 1$ for all $n \geq 1$. Actually, we shall work with the sub-Bernoulli variables $\{X_n\}$ (in place of uniformly bounded ones). So let the X_i be $SB(p_i, \beta_i)$ where $p_i = EX_i$. Let $n \geq 1$ be a fixed integer. Put $\bar{p}_n = n^{-1} \sum_{i=1}^{n} p_i$. Let $0 < \varepsilon < \bar{q}_n$ where $\bar{q}_n = 1-\bar{p}_n$. Assume that $0 < \bar{p}_n < 1$. Further set $\bar{p}_{nj} = (n-j)^{-1} \cdot \sum_{i=n+1}^{j} p_i,\ j > n.$

By using independence, sub-Bernoulli property and the inequality that the geometric mean is less than or equal to the arithmetic mean respectively, we obtain for $j > n$, and any $0 \leq t < \inf_{n<k\leq j} \beta_k$,

$$E(\exp[t(S_j - ES_j - j\varepsilon)]) / E(\exp[t(S_n - ES_n - n\varepsilon)])$$

$$= \prod_{i=n+1}^{j} \exp[-t\varepsilon]\, E(\exp(t(X_i - EX_i)))$$

$$\leq \exp[-(j-n)\,t(\bar{p}_{nj} + \varepsilon)] \prod_{i=n+1}^{j} (1 - p_i + p_i e^t)$$

$$\leq (\exp [-t(\bar{p}_{nj}+\varepsilon)]\,(1-\bar{p}_{nj}+\bar{p}_{nj}e^{t}))^{j-n} \qquad (3.1)$$

We shall apply Corollary 2.1 on the sequence X_i-EX_i-ε. Now for a fixed $0 < t_0 < \inf_{k>n} \beta_k$, the hypothesis of Corollary 2.1 is satisfied for m = n if with $t=t_0$, the last term in (3.1) is less than or equal to one. We shall need this assumption for a special choice of t_0 to be described shortly. Under such a condition, Corollary 2.1 yields the bound

$$P\left\{\sup_{k\geq n}(\bar{X}_k-E\bar{X}_k)>\varepsilon\right\} \leq \exp[-nt_0\varepsilon]\prod_{i=1}^{n}E(\exp[t_0(X_i-EX_i)])$$

$$\leq \exp[-nt_0(\bar{p}_n+\varepsilon)]\prod_{i=1}^{n}(1-p_i+p_ie^{t_0})$$

$$\leq \exp[-n\,t_0(\bar{p}_n+\varepsilon)]\,(1-\bar{p}_n+\bar{p}_ne^{t_0})^n \qquad (3.2)$$

The RHS of (3.2) is minimized for the choice

$$t_0 = \log\left(\frac{1+\varepsilon\bar{p}_n^{-1}}{1-\varepsilon\bar{q}_n^{-1}}\right) \qquad (3.3)$$

provided $t_0 < \inf_{k>n}, \beta_k$; we assume that $p_i+\varepsilon < 1$ for all i in order to make the denominator positive. The corresponding minimum value is given by $\exp[-nL(\bar{p}_n,\varepsilon)]$, where

$$L(p,\varepsilon) = (p+\varepsilon)\log(1+\varepsilon p^{-1}) + (1-p-\varepsilon)\log(1-\varepsilon(1-p)^{-1}). \qquad (3.4)$$

We now show that

$$\exp[-t_0(\bar{p}_{nj}+\varepsilon)]\,(1-\bar{p}_{nj}+\bar{p}_{nj}e^{t_0}) \leq 1 \qquad (3.5)$$

is satisfied (with t_0 given by (3.3)) if all the p_i's are equal. Note that for $p_i \equiv p$, the expression in (3.5) is the same as $\exp[-L(p,\varepsilon)]$, which, in fact, is less than or equal to $\exp[-2\varepsilon^2]$ by Theorem1 of Hoeffding (1963). Since $L(p,\varepsilon)$ is continuous in ε, we obtain .

Theorem 3.1. *Let $\{X_n\}$ be a sequence of independent sub-Bernoulli variables with parameters $\{p_n\}$ and $\{\beta_n\}$. Let $\varepsilon > 0$ satisfy $p_n + \varepsilon < 1$ for all n. Assume that $t_0 < \inf_{k>n} \beta_k$ and (3.5) holds where t_0 is given by (3.3). Then*

$$P\left\{\sup_{k\leq n}(\bar{X}_k-E\bar{X}_k)\geq\varepsilon\right\} \leq \exp[-n\,L(\bar{p}_n,\varepsilon)], \quad \varepsilon>0 \qquad (3.6)$$

where the function $L(p,\varepsilon)$ is defined by (3.4).

Theorem 3.1 is an extension of Theorem 1 of Hoeffding (1963). One can now use the arguments of the proof of Theorem 1 of Hoeffding (1963) and the results of Kraft (1969) to get the extensions of the results of Turner, Young and Seaman (1992); moreover, we have

$$P\left\{\sup_{k\geq n}(\bar{X}_k - E\bar{X}_k)) \geq \varepsilon\right\} \leq \exp\left[-n\varepsilon^2 \, g(\bar{p}_n)\right], \tag{3.7}$$

where $g(x) = (1-2x)^{-1}\log((1-x)/x)$ or $g(x) = (2x(1-x))^{-1}$ according as $0 < x < 1/2$ or $1/2 \leq x < 1$.

4. The Case of U-Statistics

Let $\{X_n, \mathfrak{R}_n\}$, $n \geq 1$, be a reverse submartingale; then

$$P\left\{\sup_{k\geq n} X_k \geq 0\right\} \leq \inf_{t\geq 0} E\left(exp\left[t X_n\right]\right) \tag{4.1}$$

Inequality (4.1) follows since for $t \geq 0$ and $N > n$,

$$P\left\{\max_{n\leq k\leq N} X_k \geq 0\right\} = P\left\{\max_{n\leq k\leq N} \exp(t X_k) \geq 1\right\} \leq E\left(exp\left[t X_n\right]\right) \tag{4.2}$$

by Doob's inequality applied to the reverse submartingale $\{\exp(tX_n), \mathfrak{R}_n\}$ (see, e.g., Shorack and Wellner, 1986, p.875) and the continuity argument used earlier (see the last part of the proof of Theorem 2.2).

Let $\{Y_n\}$ be an exchangeable sequence of rvs. Let

$$U_n = \binom{n}{m}^{-1} \sum_{1\leq i_1 < i_2 < \ldots < i_m \leq n} h(Y_{i_1}, \ldots, Y_{i_m})$$

be the U-statistics based on $Y_1,\ldots,Y_n$ and a kernel $h(.)$. Set $\theta = E_p h(y_1,\ldots,y_m)$. It is well known (e.g., Serfling, 1980, p.180) that $\{U_n, F_n\}$ is a reverse martingale where $\mathfrak{R}_n = \sigma(Y_{1:n},\ldots,Y_{n:n}, Y_{n+1}, Y_{n+2},\ldots)$ and $Y_{1:n} \leq \ldots \leq Y_{n:n}$ are the order statistics based on $Y_1,\ldots,Y_n$. Therefore, by (4.1), we obtain

$$P\left\{\sup_{k\geq n}(U_k - \theta) \geq \varepsilon\right\} \leq \inf_{t\geq 0} \exp(-t\varepsilon) E\left(exp\left[t(U_n - \theta)\right]\right). \tag{4.3}$$

The bound on the RHS of (4.3) has been used for $P\{U_n - \theta \geq \varepsilon\}$ in Theorem 2.3 of Christofides (1991). Christofides (1991) further obtained a computable upper bound for the expression on the RHS of (4.3). It thus follows that Christofides' upper bound holds for the probability $P\{\sup_{k\geq n}(U_k - \theta) \geq \varepsilon\}$ as well.

Remark 4.1. In the above, if Y_n's are iid and U_n is the sample mean, then (2.1) and (4.3) yield, in two different ways, the same upper bound for $P\{\sup_{k\geq n}(\bar{X}_k - \theta) \geq \varepsilon\}$.

Remark 4.2. Specializing (4.3) to the case of sample means of iid rvs and using the proof of Chernoff's Theorem given in Bahadur (1971, p.7), we obtain the following variation of Chernoff's Theorem: If $\{X_n\}$ is an iid sequence, then

$$\lim_{n \to \infty} n^{-1} \log P\left\{\sup_{k \leq n} \bar{X}_k \geq 0\right\} = \log \rho$$

where $\rho = \inf\left\{E(\exp[t X_1]) : t \geq 0\right\}$

EXAMPLE 4.1. Let $X_1, X_2, \ldots$ be a sequence of N(0,1) rvs with $\mathrm{cov}(X_i, X_j) = \rho > 0$ if $i \neq j, i, j \geq 1$. An easy calculation shows that

$$\inf_{t \geq 0} \exp(-t\varepsilon) E(\exp[t \bar{X}_n]) = \exp[-n\varepsilon^2/(2(1+(n-1)\rho))] = \alpha_n \ (say).$$

Since $\{X_n\}$ is exchangeable, we have $P\{\sup_{k\geq n} \bar{X}_k \geq \varepsilon\} \leq \alpha_n$.

ACKNOWLEDGEMENT

The authors thank the referee for pointing out an error in an earlier version of the paper and also for several useful remarks.

REFERENCES

Bahadur, R.R. (1971) : *Some Limit Theorems in Statistics*. SIAM, Pennsylvania.

Banjevic, D. (1984) : 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.

Chow, Y.S. and Teicher, H. (1988) : *Probability Theory : Independence, Interchangeability, Martingales; Second edition.* Springer-Verlag, New York.

Hoeffding, W. (1963) : Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13-30.

Kolmogorov. A.N. (1963) : On tables of random numbers. *Sankhya, Ser.* A **25**, 369-376.

Kraft, O. (1969) : A note on exponential bounds for binominal probabilities. *Ann. Inst. Statist. Math.* **21**, 219-220.

Serfling, R.J. (1980) : *Approximation Theorems of Mathematical Statistics.* John Wiley, New York.

Shorack, G.R. and Wellner, J.A. (1986) : *Empirical Processes with Applications to Statistics.* John Wiley, New York.

Turner, D.W., Young, D.M. and Seaman, J.W. Jr. (1992) : Improved Kolmogorov inequalities for the binomial distribution. *Statist. Probab. Lett.* **13**, 223-227.

Turner, D.W., Young, D.M. and Seaman, J.W. Jr. (1995) : A Kolmogorov inequality for the sum of independent Bernoulli random variables with unequal means. *Statist. Probab. Lett.* **23**, 243-245.

Uspensky, J.V. (1937) : *Introduction to Mathematical Probability.* McGraw-Hill Book Company, Inc. New York.

Young, D.M., Seaman, J.W. and Marco, V.R. (1987) : A note on a Kolmogorov inequality. *Statist. Probab. Lett.* **5**, 217-218.

Young, D.M., Turner, D.W. and Seaman, J.W. (1988) : A note on a Kolmogorov inequality for the binomial distribution. *Theory Probab. Appl.* **33**, 747-749.