Sankhyā : The Indian Journal of Statistics
1984, Volume 46, Series A, Pt. 1, pp. 110-116.

ON LARGE DEVIATION PROBABILITIES OF U-STATISTICS IN NON I.I.D CASE

By RATAN DASGUPTA

Indian Statistical Institute

SUMMARY. Under a rather stringent moment assumption on the kernel we compute large deviation probabilities of U-statistics.

1. Introduction

After the work of Hoeffding on limiting distribution of U-statistics, various authors studied the rates of convergence of U-statistics to normality see e.g. Grams and Serfling (1973), Bickel (1974), Chan and Wierman (1977), Callert and Janseen (1978) and Ghosh and Dasgupta (1980). However the problem of computing the large deviation probabilities for U-statistics has not received any attention. In this note using a simple technique we approximate the said probability in terms of that for independent summands. Our main result is stated below.

Let U_n be a U-statistic with kernel ϕ and degree r, based on independent observations $X_1, \ldots, X_n$. In the i.i.d case Hoeffding showed if $E\phi(x_{i_1} \ldots x_{i_r}) = 0$, $\forall\, i_1 \ldots i_r$ then,

$$nU_n/(rs_n) = s_n^{-1} \sum_{i=1}^{n} \bar{\psi}_n^{(1)}(X_i)(1+o_p(1))$$

where

$$\psi^{i_1}_{i_2 \ldots i_r}(x_{i_1}) = E[\phi(X_{i_1} \ldots X_{i_r})/X_{i_1} = x_{i_1}]$$

$$\bar{\psi}_n^{(1)}(X_i) = \binom{n-1}{r-1}^{-1} \sum_{1 < i_2 \ldots < i_r < n} \psi^{i}_{i_2 \ldots i_r}(X_i);\ i_2 \neq i, \ldots i_r \neq i$$

and

$$s_n^2 = \sum_{i=1}^{n} E\bar{\psi}^{(1)2}(X_i).$$

Suppose the kernel satisfies a certain moment condition (vide 3.6);

$$\inf_n n^{-1} s_n^2 > 0$$

AMS (1980) *subject classification* : 60F10.

Key words and phrases : Large deviation, U-statistic, Hoeffding decomposition.

and

$$\lim n^{-1}\log P\left(s_n^{-1}\sum_{i=1}^{n}\tilde{\psi}_n^{(1)}(X_i) > a\sqrt{n}\right) = \rho$$

is finite; then our main result is

$P(U_n/(rs_n) > a/\sqrt{n}) = P\left(s_n^{-1}\sum_{i=1}^{n}\psi_n^{(1)}(X_i) > a\sqrt{n}\right)(1+o(1))$ which implies that $\lim n^{-1}\log P(U_n/(rs_n) > a/\sqrt{n}) = \rho$.

2. Estimate of remainder in U-statistics

Let $\{X_n, n \geqslant 1\}$ be a sequence of independent but not necessarily identically distributed random variables. A U-statistics with kernel ϕ and degree r, based on $X_1, X_2, \ldots, X_n$ $(n \geqslant r)$, is defined by

$$U_n = \binom{n}{r}^{-1} \sum_{1 \leqslant i_1 < \ldots < i_r < n} \phi(X_{i_1}, \ldots, X_{i_r}) \qquad \ldots \quad (2.1)$$

where the kernel ϕ is symmetric in its arguments.

For the sake of completeness and further reference below (vide 2.9), we give a Hoeffding (1961) decomposition for U-statistics in the non i.i.d case (see also Ghosh and Dasgupta, 1980). We give the decomposition for the case $r = 3$; generalisation for other values of r follows easily. Without loss of generality let $E\phi(X_{i_1}, X_{i_2}, X_{i_3}) = 0 \;\forall\; 1 \leqslant i_1 < i_2 < i_3 \leqslant n$. Let

$$\psi^{i_1}_{i_2,i_3}(x_{i_1}) = \varphi^{i_1}_{i_2,i_3}(x_{i_1}) = E[\phi(X_{i_1}, X_{i_2}, X_{i_3}) \mid X_{i_1} = x_{i_1}]$$

$$1 \leqslant i_1 \neq i_2 \neq i_3 \leqslant n. \qquad \ldots \quad (2.2)$$

$$\phi^{i_1,i_2}_{i_3}(x_{i_1}, x_{i_2}) = E[\phi(X_{i_1}, X_{i_2}, X_{i_3}) \mid X_{i_1} = x_{i_1}, X_{i_2} = x_{i_2}]$$

$$1 \leqslant i_1 \neq i_2 \neq i_3 \leqslant n; \qquad \ldots \quad (2.3)$$

$$\psi^{i_1,i_2}_{i_3}(x_{i_1}, x_{i_2}) = \phi^{i_1,i_2}_{i_3}(x_{i_1}, x_{i_2}) - \psi^{i_1}_{i_2,i_3}(x_{i_1}) - \psi^{i_2}_{i_1,i_3}(x_{i_2})$$

$$1 \leqslant i_1 \neq i_2 \neq i_3 \leqslant n \qquad \ldots \quad (2.4)$$

$$\psi^{i_1,i_2,i_3}(x_{i_1}, x_{i_2}, x_{i_3}) = \phi(x_{i_1}, x_{i_2}, x_{i_3}) - \psi^{i_1}_{i_2,i_3}(x_{i_1}) - \psi^{i_2}_{i_1,i_3}(x_{i_2})$$

$$-\psi^{i_3}_{i_1,i_2}(x_{i_3}) - \psi^{i_1,i_2}_{i_3}(x_{i_1}, x_{i_2}) - \psi^{i_1,i_3}_{i_2}(x_{i_1}, x_{i_3}) - \psi^{i_2,i_3}_{i_1}(x_{i_2}, x_{i_3})$$

$$1 \leqslant i_1 \neq i_2 \neq i_3 \leqslant n. \qquad \ldots \quad (2.5)$$

Writing

$$\bar{\psi}_n^{(1)}(X_i) = \binom{n-1}{2}^{-1} \sum_{\substack{1 \leqslant j < l \leqslant n \\ j \neq i \\ l \neq i}} \psi^i_{j,l}(X_i),$$

$$\bar{\psi}_n^{(2)}(X_i, X_j) = (n-2)^{-1} \sum_{k=1(\neq i \neq j)}^{n} \psi_k^{i,j}(X_i, X_j) \qquad \dots \quad (2.6)$$

$$V_n^{(1)} = n^{-1} \sum_{i=1}^{n} \bar{\psi}_n^{(1)}(X_i),$$

$$V_n^{(2)} = \binom{n}{2}^{-1} \sum\sum_{1 \leqslant i < j \leqslant n} \bar{\psi}_n^{(2)}(X_i, X_j) \qquad \dots \quad (2.7)$$

and

$$V_n^{(3)} = \binom{n}{3}^{-1} \sum\sum\sum_{1 < i_1 < i_2 < i_3 < n} \psi^{i_1,i_2,i_3}(X_{i_1}, X_{i_2}, X_{i_3}), \qquad \dots \quad (2.8)$$

one has the representation

$$U_n = 3V_n^{(1)} + 3V_n^{(2)} + V_n^{(3)}. \qquad \dots \quad (2.9)$$

The following facts can be easily verified

$$E\psi^{i_1}_{i_2,i_3}(X_{i_1}) = 0, \; E[\psi^{i_1,i_2}_{i_3}(X_{i_1}, X_{i_2}) \mid X_{i_1} = x_{i_1}] = 0 \quad \text{a.e.}, \qquad \dots \quad (2.10)$$

$$E[\psi^{i_1,i_2,i_3}(X_{i_1}, X_{i_2}, X_{i_3}) \mid X_{i_1} = x_{i_1}, X_{i_2} = x_{i_2}] = 0 \quad \text{a.e.}, \qquad \dots \quad (2.11)$$

$$E[\psi^{i_1,i_2,i_3}(X_{i_1}, X_{i_2}, X_{i_3}) \mid X_{i_1} = x_{i_1}] = 0 \quad \text{a.e.} \qquad \dots \quad (2.12)$$

It follows from (2.10)–(2.12) that

$$E\left[\psi^{i_1}_{i_k,i_l}(X_{i_1})\psi^{i_1,i_2}_{i_m}(X_{i_1}, X_{i_2})\right] = 0; \qquad \dots \quad (2.13)$$

for any $1 \leqslant i_k \neq i_l \neq i_1 \leqslant n, \;\; 1 \leqslant i_m \neq i_1 \neq i_2 \leqslant n$;

$$E\left[\psi^{i_1,i_2}_{i_r}(X_{i_1}, X_{i_2})\psi^{i_1,i_3}_{i_s}(X_{i_1}, X_{i_3})\right] = 0, \qquad \dots \quad (2.14)$$

for any $1 \leqslant i_r \neq i_1 \neq i_2 \leqslant n$, $1 \leqslant i_s \neq i_1 \neq i_3 \leqslant n$; $i_2 \neq i_3$;

$$E\left[\psi^{i_1}_{i_k,i_l}(X_{i_1})\psi^{i_1,i_2,i_3}(X_{i_1}, X_{i_2}, X_{i_3})\right] = 0, \qquad \dots \quad (2.15)$$

for any $1 \leqslant i_k \neq i_l \neq i_1 \leqslant n$, $1 \leqslant i_1 \neq i_2 \neq i_3 \leqslant n$;

$$E\left[\psi_{i_r}^{i_1, i_2}(X_{i_1}, X_{i_2})\psi^{i'_1, i'_2, i'_3}(X_{i'_1}, X_{i'_2}, X_{i'_3})\right] = 0, \qquad \dots \quad (2.16)$$

for any $1 \leqslant i_r \neq i_1 \neq i_2 \leqslant n$, $1 \leqslant i'_1 \neq i'_2 \neq i'_3 \leqslant n$, whenever $\{i_1, i_2\}$ and $\{i'_1, i'_2, i'_3\}$ are disjoint;

$$E\left[\psi^{i_1, i_2, i_3}(X_{i_1}, X_{i_2}, X_{i_3})\psi^{i'_1, i'_2, i'_3}(X_{i'_1}, X_{i'_2}, X_{i'_3})\right] = 0 \qquad \dots \quad (2.17)$$

whenever $\{i_1, i_2, i_3\} \neq \{i'_1, i'_2, i'_3\}$.

Henceforth, unless otherwise mentioned, we work with U-statistics with kernel ϕ and degree 3. The generalisation to an arbitrary $r(> 3)$ is immediate. It is assumed without loss of generality that

$$E\phi(X_{i_1}, X_{i_2}, \dots, X_{i_r}) = 0, \; 1 \leqslant i_1 \neq \dots \neq i_r \leqslant n. \qquad \dots \quad (2.18)$$

We now prove a lemma which gives moment bounds for $V_n^{(2)}$ and $V_n^{(3)}$ when ϕ has $(2m)$-th moment.

Lemma 2.1 : *If* (2.18) *holds and*

$$\delta_m = \sup_{n \geqslant 3} \binom{n}{3}^{-1} \sum_{1 \leqslant i_1 < i_2 < i_3 \leqslant n} E|\phi(X_{i_1}, X_{i_2}, X_{i_3})|^{2m} < \infty \qquad \dots \quad (2.19)$$

then

$$E(V_n^{(2)})^{2m} \leqslant n^{-2m} L^m m \,! \; \delta_m \qquad \dots \quad (2.20)$$

and

$$E(V_n^{(3)})^{2m} \leqslant n^{-3m} L^m m \,! \; \delta_m \qquad \dots \quad (2.21)$$

where in the above and in what follows $L(> 0)$ is a generic constant independent of m and n.

Proof :

$$E(V_n^{(2)})^{2m} = \binom{n}{2}^{-2m} \sum_{1 \leqslant i_1 < j_1 \leqslant n} \dots \sum_{1 \leqslant i_{2m} < j_{2m} \leqslant n} E[\bar{\psi}_n^{(2)}(X_{i_1}, X_{j_1}) \dots \bar{\psi}_n^{(2)}(X_{i_{2m}}, X_{j_{2m}})]. \qquad \dots \quad (2.22)$$

Note that if a pair of suffixes (i_k, j_k) occurs exactly once in $(\{i_1, j_1\}, \dots, \{i_{2m}, j_{2m}\})$, then in view of (2.14)

$$E[\bar{\psi}_n^{(2)}(X_{i_1}, X_{j_1}) \dots \bar{\psi}_n^{(2)}(X_{i_{2m}}, X_{j_{2m}})] = 0. \qquad \dots \quad (2.23)$$

Subject to the condition that each pair of suffixes (i_k, j_k) occurs at least twice, the maximum number of suffixes that can occur is $2m$, corresponding to m distinct pairs of suffixes.

If there are k distinct pairs, by repeated application of Hölders inequality, the product term inside the sum in (2.22) is

$$|E(\bar{\psi}_n^{(2)}(X_{i_1}, X_{j_1}))^{l_1} \dots (\bar{\psi}_n^{(2)}(X_{i_k}, X_{j_k}))^{l_k}|;\ l_1+\dots+l_k = 2m$$

$$\leqslant E^{l_1/2m}(\bar{\psi}_n^{(2)}(X_{i_1}, X_{j_1}))^{2m} \dots E^{l_k/2m}(\bar{\psi}_n^{(2)}(X_{i_k}, X_{j_k}))^{2m}$$

$$\leqslant \frac{l_1}{2m} E(\bar{\psi}_n^{(2)}(X_{i_1}, X_{j_1}))^{2m} + \dots + \frac{l_k}{2m} E(\bar{\psi}_n^{(2)}(X_{i_k}, X_{j_k}))^{2m}$$

$$\leqslant \sum_{p=1}^{k} E(\bar{\psi}^{(2)}(X_{i_p}, X_{j_p}))^{2m};\ \text{since } l_i \leqslant 2m,\ \forall\, i. \qquad \dots \quad (2.24)$$

Now note that k pairs can be taken out of $2m$ pairs in $2m_{p_k}$ ways, $k = 1, \dots, m$, m being the maximum number of distinct pairs. Therefore from (2.22)

$$E(V_n^{(2)})^{2m} \leqslant \binom{n}{2}^{-2m} \sum_{1 \leqslant i_1 < j_1 \leqslant n} \dots \sum_{1 \leqslant i_m < j_m \leqslant n}$$

$$\times \left[\sum_{k=1}^{m} \sum_{p=1}^{k} E|\bar{\psi}_n^{(2)}(X_{i_p}, X_{j_p})|^{2m} 2m_{p_k} \right]. \qquad \dots \quad (2.25)$$

Since $2m_{p_k} \leqslant L^m m\,!$ for $k = 1, \dots, m$ and since the maximum variation of the indexes $1 \leqslant i_1 < j_1 \leqslant n, \dots, 1 \leqslant i_m < j_m \leqslant n$ are in a total number of ways $\leqslant n^{2m}$, we have with an application of Jensen and C_{2m} inequalities

$$E(V_n^{(2)})^{2m} \leqslant \binom{n}{2}^{-2m} n^{2m} L^m m\,!\ \ m^2 \delta_m;\ \text{from (2.19)}$$

and since $p \leqslant k \leqslant m$,

$$\leqslant n^{-2m} L^m\, m\,!\, \delta_m.$$

Hence (2.20). Similarly (2.21) can be shown and the lemma follows.

3. Results on probabilities of deviations for U-statistics

Now writing

$$n^{1/2}U_n/(3\sigma_n) = s_n^{-1} \sum_{i=1}^{n} \bar{\psi}_n^{(1)}(X_i)+R_n \qquad \ldots \quad (3.1)$$

where

$$s_n^2 = \sum_{i=1}^{n} E\bar{\psi}^{(1)^2}(X_i), \quad \sigma_n^2 = n^{-1}s_n^2 \qquad .. \quad (3.2)$$

and

$$R_n = n^{1/2}\sigma_n^{-1}\left(V_n^{(2)}+\frac{1}{3}\ V_n^{(3)}\right)$$

we have under

$$\inf_n \sigma_n^2 > 0, \qquad \ldots \quad (3.3)$$

the following

$$ER_n^{2m} \leqslant n^{-m}L^m m\,!\,\delta_m. \qquad \ldots \quad (3.4)$$

Also note that from (2.2), (2.6) using C_{rm} and Jensens inequality for conditioral expectation

$$\sup_n n^{-1} \sum_{i=1}^{n} E\,|\,\bar{\psi}_n^{(1)}(X_i)\,|^{2m} \leqslant L^m\,\delta_m. \qquad \ldots \quad (3.5)$$

The representation in (3.1) permits us to have large deviation results for U-statistics even for non i.i.d case. For i.i.d case under the assumption of existence of m.g.f. Chernoff (1952) has results on large deviation on sample mean which were generalised for independent random variables by Sievers (1969), Plachky and Stienebach (1975) etc. We may have similar results for U-statistics provided we show that the remainder has negligible contribution compared to the mean part in (3.1). Let

$$\delta_m \leqslant k^{-m}.\,(m)\,!\ \text{ for all sufficiently large } m \qquad \ldots \quad (3.6)$$

where k is fixed but may be taken arbitrary large. In i.i.d case this condition is equivalent to $E\,\exp(t\phi^2) < \infty$ for large t. Then from (3.4) it is clear that m.g.f. of $\sqrt{n}R_n$ exists for all values of t i.e.

$$f(t) = E\,\exp(t\sqrt{n}R_n) < \infty\ \forall\ t \text{ real}. \qquad \ldots \quad (3.7)$$

Hence by Marcov inequality we have the following

Theorem 1 : *Let δ_m satisfies* (3.6) *and define*

$$P_n^*(a) = P\left\{s_n^{-1} \sum_{i=1}^{n} \bar{\psi}_n^{(1)}(X_i) > a\, n^{1/2}\right\}.$$

If

$$-\lim_n n^{-1} \log P_n^*(a) < \infty \text{ for some } a > 0. \quad \ldots \quad (3.8)$$

Then under (3.3), *the following holds*

$$P_n(a) = P_n^*(a) + o(P_n^*(a)) \quad \ldots \quad (3.9)$$

where $P_n(a) = P\{U/(3\sigma_n) > a\}$.

Comment : (3.8) is nothing but an assumption about existence of large deviation for independent random variables under quite a strong assumption 2m-th moment is of order m !

It would be interesting to know if Theorem 1 holds under more natural condition $\delta_m = O(2m)$! Sethuraman has a result similar to Theorem 1 independently. Since crude moment bounds are used, sharper results may not be possible by this technique.

References

Bickel, P. J. (1974): Edgeworth expansions in non-parametric statistics. *Ann. Statist.*, **2**, 1–20.

Callert, H. and Janssen, P. (1978): The Berry-Esseen theorem for U-statistics. *Ann. Statist.*, **6**, 417–421.

Chan, Y. K., and Wierman, J. (1977): On the Berry-Esseen theorem for U-statistics. *Ann. Prob.*, **4**, 136–139.

Ghosh, M. and Dasgupta, R. (1980): Berry-Esseen theorems for U-statistics in the non-iid case. *Proc. on the Non Parametric Inference*, Hungary.

Grams, W. F. and Serfling, R. J. (1973): Convergence rates for U-statistics and related statistics. *Ann. Statist.*, **1**, 153–160.

Plachky, D. and Steinebach, J. (1975): A theorem about probabilities of large deviation with application of queing theory. *Period. Math. Hungar.* **6**, 343–346.

Sievers, G. L. (1969): On the probabilities of large deviation and exact slopes. *Ann. Math. Stat.*, **40**, 1908–1921.

Paper received : October, 1982.
Revised : March, 1983.