STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **9**, December 2021, pp 984–989.
Published online in International Academic Press (www.IAPress.org)

A Berry-Esseen Bound for Nonlinear Statistics with Bounded Differences

Nguyen Tien Dung *, Hoang Thi Phuong Thao , Pham Trung Hieu

Department of Mathematics, VNU University of Science, Vietnam National University, Vietnam

Abstract In this paper, we obtain an explicit Berry-Esseen bound in the central limit theorem for nonlinear statistics with bounded differences. Some examples are provided as well.

Keywords Central limit theorem, Berry-Esseen bound, nonlinear statistic.

AMS 2010 subject classifications 60F05, 60B10.

DOI: 10.19139/soic-2310-5070-1305

1. Introduction

Let $X = (X_1, X_2, ..., X_n)$ be a vector of independent random variables (not necessarily identically distributed), defined on some probability space $(\Omega, \mathcal{F}, P)$ and taking values in a measurable space $\mathcal{X}$. We consider the problem of investigating the rate of convergence in the central limit theorem for nonlinear statistics of the form

$$W = f(X) = f(X_1, X_2, ..., X_n), \tag{1}$$

where $f : \mathcal{X}^n \to \mathbb{R}$ is a measurable function. This problem, of course, is one of the most fundamental topics in statistics and its study has a long history. A significant amount of general results for W and its special forms can be found in the literature, see e.g. [1, 2, 6, 8, 9, 10] and the references therein.

Let $X' = (X'_1, X'_2, ..., X'_n)$ be an independent copy of $X = (X_1, X_2, ..., X_n)$. For each $A \subseteq [n] := \{1, 2, ..., n\}$, define the random vector X^A as

$$X_i^A = \begin{cases} X'_i, & \text{if } i \in A, \\ X_i, & \text{if } i \notin A. \end{cases}$$

For each $j \in [n]$, we write X^j instead of $X^{\{j\}}$ and define the difference operator Δ_j by

$$\Delta_j f(X) := f(X) - f(X^j). \tag{2}$$

Definition 1.1. We say that the nonlinear statistic W has the bounded differences property if there exist nonnegative deterministic constants $c_1, ..., c_n$ such that, almost surely,

$$|\Delta_j f(X)| \leq c_j, \ 1 \leq j \leq n. \tag{3}$$

The motivation of the present paper comes from the observation that many existing statistics belong to the class of nonlinear statistics with bounded differences (the reader can consult Sections 3.2 and 6.1 in [3] for several specific examples provided there). Hence, it is necessary to study the central limit theorem and the rate of convergence for

*Correspondence to: Nguyen Tien Dung (Email: dung_nguyentien10@yahoo.com, dung@hus.edu.vn). Department of Mathematics, VNU University of Science, Vietnam National University, Hanoi, 334 Nguyen Trai, Thanh Xuan, Hanoi, Vietnam.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

this class. In fact, we can use the results obtained in [4, 13] to derive the following bound

$$d_W(\sigma^{-1}W, Z) \leq \frac{1}{\sigma^2}\sqrt{\mathrm{Var}(E[T|W])} + \frac{1}{2\sigma^3}\sum_{j=1}^n c_j^3, \tag{4}$$

$$d_K(\sigma^{-1}W, N) \leq \frac{1}{\sigma^2}\sqrt{\mathrm{Var}(E[T|W])} + \frac{1}{\sigma^2}\sqrt{\mathrm{Var}(E[T'|W])} + \frac{4+\sqrt{2\pi}}{16\sigma^3}\sum_{j=1}^n c_j^3, \tag{5}$$

where $T' := \frac{1}{2}\sum\limits_{A\subsetneq[n]}\frac{1}{\binom{n}{|A|}(n-|A|)}\sum\limits_{j\notin A}\Delta_j f(X)|\Delta_j f(X^A)|$ and

$$T := \frac{1}{2}\sum_{A\subsetneq[n]}\frac{1}{\binom{n}{|A|}(n-|A|)}\sum_{j\notin A}\Delta_j f(X)\Delta_j f(X^A). \tag{6}$$

Here and in the sequel, Z is a standard normal random variable and d_W, d_K denote the Wasserstein and Kolmogorov distances, respectively. A bound on d_K is called the Berry-Esseen bound. For reader's convenience, we recall that

$$d_W(W,Z) := \sup_{|h(x)-h(y)|\leq|x-y|}|E[h(W)] - E[h(Z)]|,$$

$$d_K(W,Z) := \sup_{z\in\mathbb{R}}|P(W\leq z) - P(Z\leq z)|.$$

It should be noted that, from the statistical point of view, the Kolmogorov distance is more informative in practice. Indeed, for instance, the Berry-Esseen bound can be used for the construction of confidence intervals. However, by its definition, the additional term $\sqrt{\mathrm{Var}(E[T'|W])}$ is sometimes not easy to compute. In this paper, our aim is to point out that, for nonlinear statistics with bounded differences, the term $\sqrt{\mathrm{Var}(E[T'|W])}$ in (5) can be removed. In other words, our Berry-Esseen bound is the same as the Wasserstein bound (4) (up to constant). This allows us to employ the existing results in the literature and we are able to obtain the Berry-Esseen bound without doing any further computations.

The rest of this article is organized as follows. In the next Section, after stating and proving the main result in Theorem 2.1, we provide several examples to illustrate the theory.

2. The main results

Our proof is based on the techniques of Stein's method and the smoothed indicator functions. Let us recall here the following fundamental results from Lemma 2.3 in [4] and Lemma 2.5 in [7].

Lemma 2.1
For any $g, f : \mathcal{X}^n \to \mathbb{R}$ such that $E[g^2(X)]$ and $E[f^2(X)]$ are both finite, we have

$$\mathrm{Cov}(g(X), f(X)) = \frac{1}{2}\sum_{A\subsetneq[n]}\frac{1}{\binom{n}{|A|}(n-|A|)}\sum_{j\notin A}E[\Delta_j g(X)\Delta_j f(X^A)].$$

Lemma 2.2
Given $z\in\mathbb{R}$ and $\alpha > 0$, define the function $h_{\alpha,z}$ by

$$h_{\alpha,z}(w) = \begin{cases} 1 & w\leq z, \\ 1 + (z-w)/\alpha & z < w \leq z+\alpha, \\ 0 & w > z. \end{cases}$$

Let φ be the solution to the Stein equation

$$\varphi'(w) - w\varphi(w) = h_{\alpha,z}(w) - Eh_{\alpha,z}(Z).$$

Then, for all $w, v \in \mathbb{R}$, we have

$$0 \leq \varphi(w) \leq 1, \ \ |\varphi'(w)| \leq 1, \ \ |\varphi'(w) - \varphi'(v)| \leq 1 \tag{7}$$

and

$$|\varphi'(w+v) - \varphi'(w)| \leq |v| \left(1 + |w| + \frac{1}{\alpha} \int_0^1 \mathbf{1}_{[z,z+\alpha]}(w+rv)dr\right). \tag{8}$$

We now are in a position to state and prove the main result of the present paper.

Theorem 2.1
Let $W = f(X)$ be a nonlinear statistic with bounded differences. Suppose that $EW = 0$ and $\sigma^2 := \mathrm{Var}(W) < \infty$. Then, we have the following Berry-Esseen bound

$$d_K(\sigma^{-1}W, Z) \leq \frac{2}{\sigma^2}\sqrt{\mathrm{Var}(E[T|W])} + \frac{5}{\sigma^3}\sum_{j=1}^n c_j^3, \tag{9}$$

where T is defined by (6) and $c_j, j \in \{1, 2, ..., n\}$ are as in (3).

Proof
Assume without loss of generality that $\sigma = 1$. For each $A \subseteq [n]$, we denote $W^A = f(X^A)$. When $A = \{j\}$, we write W^j instead of $W^{\{j\}}$. Let $h_{\alpha,z}$ be as in Lemma 2.2. It follows from Lemma 2.1 that

$$\begin{aligned} Eh_{\alpha,z}(W) - Eh_{\alpha,z}(Z) &= E[\varphi'(W)] - E[W\varphi(W)] \\ &= E[\varphi'(W)] - \frac{1}{2}\sum_{A \subsetneq [n]} \frac{1}{\binom{n}{|A|}(n-|A|)} \sum_{j \notin A} E[\Delta_j \varphi(W) \Delta_j W^A]. \end{aligned} \tag{10}$$

For each $j \in \{1, 2, ..., n\}$, we have

$$\begin{aligned} \Delta_j \varphi(W) &= \varphi(W) - \varphi(W^j) \\ &= \int_0^{W-W^j} \varphi'(W-t)dt = \int_0^{\Delta_j W} \varphi'(W-t)dt \\ &= \varphi'(W)\Delta_j W + \int_0^{\Delta_j W} [\varphi'(W-t) - \varphi'(W)]dt. \end{aligned}$$

We now put $R_j := \int_0^{\Delta_j W} [\varphi'(W-t) - \varphi'(W)]dt$ and use the notations $x_- = \min(x, 0), x_+ = \max(x, 0)$. Then, we have

$$\Delta_j \varphi(W) = \varphi'(W)\Delta_j W + R_j, \ \ 1 \leq j \leq n, \tag{11}$$

and, by (8), the remainder term R_j satisfies the following estimate

$$\begin{aligned} |R_j| &\leq \int_{(\Delta_j W)_-}^{(\Delta_j W)_+} |\varphi'(W-t) - \varphi'(W)|dt \\ &\leq \int_{(\Delta_j W)_-}^{(\Delta_j W)_+} |t| \left(1 + |W| + \frac{1}{\alpha}\int_0^1 \mathbf{1}_{[z,z+\alpha]}(W - rt)dr\right) dt \\ &= \frac{1}{2}(1+|W|)(\Delta_j W)^2 + \frac{1}{\alpha}\int_{(\Delta_j W)_-}^{(\Delta_j W)_+} \left(\int_{t_-}^{t_+} \mathbf{1}_{[z,z+\alpha]}(W+u)du\right) dt. \end{aligned}$$

Furthermore, by bounded differences property, we can get

$$|R_j| \leq \frac{c_j^2}{2}(1+|W|) + \frac{1}{\alpha}\int_{-c_j}^{c_j} \left(\int_{t_-}^{t_+} \mathbf{1}_{[z,z+\alpha]}(W+u)du\right) dt, \ \ 1 \leq j \leq n. \tag{12}$$

Inserting (11) into (10) yields

$$Eh_{\alpha,z}(W) - Eh_{\alpha,z}(Z) = E[\varphi'(W)(1-T)] - \frac{1}{2}\sum_{A\subsetneq[n]} \frac{1}{\binom{n}{|A|}(n-|A|)} \sum_{j\notin A} E[R_j\Delta_j W^A]. \tag{13}$$

Note that we also have $|\Delta_j W^A| \le c_j$, and hence, the estimates (12) give us

$$\begin{aligned}
|E[R_j\Delta_j W^A]| &\le \frac{c_j^3}{2}(1+E|W|) + \frac{c_j}{\alpha}\int_{-c_j}^{c_j}\left(\int_{t_-}^{t_+} E[\mathbf{1}_{[z,z+\alpha]}(W+u)]du\right)dt \\
&= \frac{c_j^3}{2}(1+E|W|) + \frac{c_j}{\alpha}\int_{-c_j}^{c_j}\left(\int_{t_-}^{t_+} P(z\le W+u\le z+\alpha)du\right)dt,\ 1\le j\le n.
\end{aligned}$$

By the fundamental property of the standard normal random variable $P(Z\le b) - P(Z\le a) \le |b-a|/2$ for all $a,b\in\mathbb{R}$, we obtain

$$\begin{aligned}
P(z\le W+u\le z+\alpha) &= P(W\le z+\alpha-u) - P(Z\le z+\alpha-u) \\
&\quad + P(Z\le z+\alpha-u) - P(Z\le z-u) + P(Z\le z-u) - P(W\le z-u) \\
&\le 2d_K(W,Z) + \frac{\alpha}{2}.
\end{aligned}$$

This, together with the fact that $E|W| \le \mathrm{Var}(W) = 1$, implies

$$\begin{aligned}
|E[R_j\Delta_j W^A]| &\le \frac{c_j^3}{2}(1+E|W|) + \frac{c_j}{\alpha}\int_{-c_j}^{c_j}\left(\int_{t_-}^{t_+}(2d_K(W,Z)+\alpha/2)du\right)dt \\
&\le 2c_j^3\left(\frac{3}{2} + \frac{d_K(W,Z)}{\alpha}\right),\ 1\le j\le n.
\end{aligned}\tag{14}$$

Recalling (7), we combine (13) and (14) to get the following

$$\begin{aligned}
|Eh_{\alpha,z}(W) - Eh_{\alpha,z}(Z)| &\le E|E[T|W]-1| + \left(\frac{3}{2}+\frac{d_K(W,Z)}{\alpha}\right)\sum_{A\subsetneq[n]}\frac{1}{\binom{n}{|A|}(n-|A|)}\sum_{j\notin A} c_j^3 \\
&= E|E[T|W]-1| + \left(\frac{3}{2}+\frac{d_K(W,Z)}{\alpha}\right)\sum_{j=1}^n c_j^3\ \ \forall\, z\in\mathbb{R}, \alpha>0,
\end{aligned}$$

or equivalently

$$\sup_{z\in\mathbb{R}}|Eh_{\alpha,z}(W) - Eh_{\alpha,z}(Z)| \le E|E[T|W]-1| + \left(\frac{3}{2}+\frac{d_K(W,Z)}{\alpha}\right)\sum_{j=1}^n c_j^3\ \ \forall\,\alpha>0.$$

To finish the proof, we observe that

$$\begin{aligned}
P(W\le z) - P(Z\le z) &\le Eh_{\alpha,z}(W) - P(Z\le z) \\
&= Eh_{\alpha,z}(W) - Eh_{\alpha,z}(Z) + Eh_{\alpha,z}(Z) - P(Z\le z) \\
&\le Eh_{\alpha,z}(W) - Eh_{\alpha,z}(Z) + \frac{\alpha}{2}\ \ \forall\, z\in\mathbb{R}.
\end{aligned}$$

The same argument will gives us a corresponding lower bound. That is to say

$$d_K(W,Z) \le \sup_{z\in\mathbb{R}}|Eh_{\alpha,z}(W) - Eh_{\alpha,z}(Z)| + \frac{\alpha}{2}\ \ \forall\,\alpha>0,$$

and hence,

$$d_K(W,Z) \leq E|E[T|W]-1| + \left(\frac{3}{2} + \frac{d_K(W,Z)}{\alpha}\right)\sum_{j=1}^n c_j^3 + \frac{\alpha}{2} \;\; \forall\, \alpha > 0.$$

Choosing $\alpha = 2\sum\limits_{j=1}^n c_j^3$, we obtain

$$d_K(W,Z) \leq 2E|E[T|W]-1| + 5\sum_{j=1}^n c_j^3 \leq 2\sqrt{\mathrm{Var}(E[T|W])} + 5\sum_{j=1}^n c_j^3,$$

which leads us to the desired conclusion (9). Notice that the second inequality follows from the Hölder inequality and the fact that $E[T] = \sigma^2 = 1$. □

Once again, we would like to emphasize that our Berry-Esseen bound (9) is the same as the Wasserstein bound (4). Hence, we can use the results obtained previously by other authors to derive the Berry-Esseen bound without doing any further computations. Let us provide some examples to illustrate this advantage.

Example 2.1 (An occupancy problem). Suppose n balls are dropped into αn boxes such that all $(\alpha n)^n$ possibilities are equally likely. Let $\mathcal{X}$ be the set of labels of the αn boxes and let X_i denote the label of the box into which ball i is dropped. Let $W = f(X_1,\cdots,X_n)$ be the number of empty boxes. It is known from Section 3.2 in [4] that $|\Delta_j W| \leq 1$ for all $1 \leq j \leq n$. Thus W is a nonlinear statistics with bounded differences and $c_j = 1$ for all $1 \leq j \leq n$. Moreover, also from Section 3.2 in [4], we have $\sigma^2 = \mathrm{Var}(W) \sim (\alpha e^{-1/\alpha} - (1+\alpha)e^{-2/\alpha})n$ as $n \to \infty$ and $\sqrt{\mathrm{Var}(E[T|W])} \leq C\sqrt{n}$ for some constant C that does not depend on n. So, by Theorem 2.1, we obtain the Berry-Esseen bound

$$d_K\left(\frac{W-E[W]}{\sigma}, Z\right) \leq \frac{C}{\sqrt{n}}.$$

Example 2.2 (A permutation statistic). Let S_n be the group of all $n!$ permutations of $\{1,\cdots,n\}$. Let the number of descents be defined as

$$D(\pi) = |\{i : 1 \leq i \leq n-1, \pi(i+1) < \pi(i)\}|,\;\; \pi \in S_n.$$

In seeking to make a metric on permutations using descents we were led to study

$$T(\pi) = D(\pi) + D(\pi^{-1}).$$

Note that $T(\pi)$ is a new permutation statistic introduced recently in [5]. In Theorem 1.1 of [5], Chatterjee and Diaconis obtained the following results

$$E[T(\pi)] = n-1\,,\;\; \sigma^2 := \mathrm{Var}(T(\pi)) = \frac{n+7}{6} - \frac{1}{n}\,,$$

and the Wasserstein bound, for some C not depending on n,

$$d_W\left(\frac{T(\pi)-E[T(\pi)]}{\sigma}, Z\right) \leq \frac{Cn^{1/2}}{\sigma^2} + \frac{Cn}{\sigma^3}.$$

Moreover, $T(\pi)$ can be considered as a measurable function of independent random variables: $T(\pi) = f(X_1,\cdots,X_n)$, where $X_1,\cdots,X_n$ are independent uniformly distributed points on $\mathcal{X} = [0,1]^2$. We have, in addition, that $|\Delta_j f(X)| \leq 4$ for every j. Thus we can apply Theorem 2.1 and we obtain

$$d_K\left(\frac{T(\pi)-E[T(\pi)]}{\sigma}, Z\right) \leq \frac{Cn^{1/2}}{\sigma^2} + \frac{Cn}{\sigma^3} \leq \frac{C}{\sqrt{n}},$$

which, to the best of our knowledge, is new.

Example 2.3 (Longest common subsequence). Let $(X_i)_{i\geq 1}$ and $(Y_i)_{i\geq 1}$ be two infinite sequences whose coordinates take their values in $\mathcal{A}_m = \{\alpha_1, \cdots, \alpha_m\}$, a finite alphabet of size m. Next, let LC_n be the length of the longest common subsequences (LCSs) of the random words $X_1 \cdots X_n$ and $Y_1 \cdots Y_n$, i.e. LC_n is the maximal integer $k \in \{1, \cdots, n\}$, such that there exist $1 \leq i_1 < \cdots < i_k \leq n$ and $1 \leq j_1 < \cdots < j_k \leq n$, such that $X_{i_s} = Y_{j_s}$ for all $1 \leq s \leq k$.

We now assume that $(X_i)_{i\geq 1}$ and $(Y_i)_{i\geq 1}$ are two independent sequences of independent identically distributed random variables having the same law. Thus, by its definition, LC_n is a measurable function of $2n$ independent random variables:

$$LC_n = LC_n(X_1, \cdots, X_n, Y_1, \cdots, Y_n).$$

Although the study of LC_n has a long history, the first limit theorem result for LC_n was recently obtained by Houdré and Işlak. In Theorem 1.1 of [11], they provide the following bound on the Wasserstein distance

$$d_W\left(\frac{LC_n - E[LC_n]}{\sqrt{\mathrm{Var}(LC_n)}}, Z\right) \leq C\frac{(\log n)^{3/4}}{n^{3/14}},$$

where $C > 0$ is a constant independent of n, provided that $\mathrm{Var}(LC_n) \geq Kn$ for some $K > 0$ not depending on n.

From the proof of Theorem 1.1 in [11] (also see Example 3.4 in [3]), we have $|\Delta_j LC_n| \leq 1$ for all $1 \leq j \leq 2n$. This means that LC_n is a nonlinear statistics with bounded differences and $c_j = 1$ for all $1 \leq j \leq 2n$. Hence, by using the estimates established in [11], our Theorem 2.1 provides the following Berry-Esseen bound

$$d_K\left(\frac{LC_n - E[LC_n]}{\sqrt{\mathrm{Var}(LC_n)}}, Z\right) \leq C\frac{(\log n)^{3/4}}{n^{3/14}}.$$

For further examples, we note that the nonlinear statistics, including U-statistics, V-statistics and Lipschitz L-statistics, discussed in [12] also satisfy the bounded difference property. The Wasserstein bound obtained there also holds true for the Berry-Esseen bound.

Acknowledgments.

This research was funded by the Vietnam National University, Hanoi under grant number QG.20.21.

REFERENCES

1. V. Bentkus, F. Götze, A. Tikhomirov, Berry-Esseen bounds for statistics of weakly dependent samples. *Bernoulli* 3 (1997), no. 3, 329–349.
2. I. Berkes, W. Philipp, Approximation theorems for independent and weakly dependent random vectors. *Ann. Probab.* 7 (1979), no. 1, 29–54.
3. S. Boucheron, G. Lugosi, P. Massart, Concentration inequalities. A nonasymptotic theory of independence. With a foreword by Michel Ledoux. Oxford University Press, Oxford, 2013.
4. S. Chatterjee, A new method of normal approximation. *Ann. Probab.* 36 (2008), no. 4, 1584–1610.
5. S. Chatterjee, P. Diaconis, A central limit theorem for a new statistic on permutations. *Indian J. Pure Appl. Math.* 48 (2017), no. 4, 561–573.
6. L. H. Y. Chen, Q.-M. Shao, Normal approximation for nonlinear statistics using a concentration inequality approach. *Bernoulli* 13 (2007), no. 2, 581–599.
7. L. H. Y. Chen, L. Goldstein, Q.-M. Shao, Normal approximation by Stein's method. Probability and its Applications (New York). Springer, Heidelberg, 2011.
8. N. T. Dung, Explicit rates of convergence in the multivariate CLT for nonlinear statistics. *Acta Math. Hungar.* 158 (2019), no. 1, 173–201.
9. K. O. Friedrich, A Berry-Esseen bound for functions of independent random variables. *Ann. Statist.* 17 (1989), no. 1, 170–183.
10. F. Götze, On the rate of convergence in the multivariate CLT. *Ann. Probab.* 19 (1991), no. 2, 724–739.
11. C. Houdré, Ü. Işlak, A Central limit theorem for the length of the longest common subsequences in random words. arXiv:1408.1559v4, 2017.
12. A. Maurer, M. Pontil, Empirical bounds for functions with weak interactions. *Proceedings of Machine Learning Research* vol 75:1–24, 2018
13. R. L-Rey, G. Peccati, New Berry-Esseen bounds for functionals of binomial point processes. *Ann. Appl. Probab.* 27 (2017), no. 4, 1992–2031.