Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2009, Article ID 373572, 9 pages
doi:10.1155/2009/373572

Research Article

Data Depth Trimming Counterpart of the Classical t (or T^2) Procedure

Yijun Zuo

Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA

Correspondence should be addressed to Yijun Zuo, yijun.zuo@gmail.com

Received 20 July 2009; Revised 11 October 2009; Accepted 14 October 2009

Recommended by Zhidong Bai

The classical t (or T^2 in high dimensions) inference procedure for unknown mean μ : $\overline{X} \pm t_\alpha(n-1)S_n/\sqrt{n}$ (or $\{\mu : n(\overline{x}-\mu)'S^{-1}(\overline{x}-\mu) \le \chi^2_{(1-\alpha)}(p)\}$) is so fundamental in statistics and so prevailing in practices; it is regarded as an optimal procedure in the mind of many practitioners. It this manuscript we present a new procedure based on data depth trimming and bootstrapping that can outperform the classical t (or T^2 in high dimensions) confidence interval (or region) procedure.

1. Introduction

Let $X^n := \{X_1, \ldots, X_n\}$ be a random sample from distribution F with an unknown mean parameter μ. The most prevailing procedure for estimating μ is the classical t-confidence interval. A $100(1-2\alpha)\%$ confidence interval (CI) for μ and large n is

$$\overline{X} \pm t_\alpha(n-1)\frac{s}{\sqrt{n}}, \tag{1.1}$$

where $\overline{X} = (1/n)\sum_{i=1}^n X_i$ is the standard sample mean, $s = \sqrt{(1/(n-1))\sum_{i=1}^n (X_i - \overline{X})^2}$ is the standard sample deviation, and $t_r(N)$ is the rth upper quantile of a t distribution with degrees of freedom N. The rule of thumb in most textbooks for the sample size n is: $n < 15$, do not use t procedure, $15 < n \le 40$, do not use it if outliers present, use it if $n > 40$. The procedure is based on the large sample property and the central limit theorem. So it is not exact but an approximation for large sample size n and arbitrary population distribution.

In higher dimensions, the counterpart to procedure (1.1) is the celebrated Hotelling's T^2 procedure: A $100(1-\alpha)\%$ confidence region for the unknown vector μ and large n is the region:

$$\left\{\mu : n(\overline{x}-\mu)' S^{-1}(\overline{x}-\mu) \le \chi^2_{(\alpha)}(p)\right\}, \tag{1.2}$$

where S is the sample covariance matrix and $\chi^2_{(\alpha)}(p)$ is the upper αth quantile of χ^2 distribution with degrees of freedom p.

Procedure (1.1) and (1.2) are so prevailing in practices that in many practitioners, mind they are regarded as optimal and unbeatable procedure. Are they really unbeatable? In this manuscript we introduce a new procedure that can outperform these seemingly optimal procedures.

The rest of the paper is organized as follows. Section 2 introduces the new procedure and Section 3 conducts some simulation studies. The paper ends in Section 4 with some concluding remarks.

2. A New Procedure for the Unknown μ

2.1. A Univariate Location Estimator

It is well known that the sample mean in the t procedure is extreme sensitive to outliers, heavy tailed distributions, or contamination. The procedure therefore is not *robust*. So naturally, one would replace the sample mean with a robust counterpart. We will utilize a special univariate location estimator $\widehat{\mu}$ to replace the sample mean $\overline{X}$ in the t procedure.

Now we consider a special univariate "projection depth-trimmed mean" (PTM_β) for X^n in $\mathbb{R}^1$, $\beta > 0$ (see Wu and Zuo [1] in $\mathbb{R}^1$, also see Zuo [2] for a multidimensional PTM_β)

$$\widehat{\mu}(X^n) := \mathrm{PTM}_\beta(X^n) = \frac{1}{k}\sum_{i=1}^{k} X_{j_i}, \tag{2.1}$$

where $j_1, \ldots, j_k$ are distinct numbers from $\{1, \ldots, n\}$ such that $\mathrm{PD}(X_{j_i}, X^n) \ge \beta$ for some $\beta > 0$, $i = 1, \ldots, k$, and $\mathrm{PD}(X_i, X^n) := 1/(1 + |X_i - \mathrm{Med}(X^n)|/\mathrm{MAD}(X^n))$; where $\mathrm{Med}(X^n)$ is standard sample median and $\mathrm{MAD} := \mathrm{med}\{|X_i - \mathrm{Med}(X^n)|, i = 1, \ldots, n\}$ is standard median of absolute deviations (see Zuo and Serfling [3] and Zuo [4] for the study of PD in high-dimensions).

Let F_n be the empirical distribution based on X^n which places mass $1/n$ at points X_i, $i = 1, \ldots, n$. We sometimes write F_n (for X^n) for convenience. Let F be the distribution of X_i. Replacing X^n with F in the above definition, we obtain the population version. For example, the popular version of PTM_β for $F \in \mathbb{R}^1$ is

$$\mathrm{PTM}_\beta\ (F) = \frac{\int_{\mathrm{PD}(x,F)>\beta} x dF(x)}{\int_{\mathrm{PD}(x,F)>\beta} dF(x)}. \tag{2.2}$$

Table 1: Average coverage (length) of 95% CI's by t and PTM_β.

$n = 100$	method	$N(0,1)$		$t(3)$	
$m = 300$	PTM_β	.9390	(.3819)	.9515	(.5855)
	t	.9550	(.3967)	.9605	(.6607)
$m = 500$	PTM_β	.9470	(.3842)	.9505	(.5902)
	t	.9520	(.3956)	.9520	(.6618)
$m = 1000$	PTM_β	.9410	(.3864)	.9470	(.5916)
	t	.9490	(.3959)	.9540	(.6582)
$m = 2000$	PTM_β	.9435	(.3876)	.9550	(.5943)
	t	.9480	(.3960)	.9545	(.6589)

2.2. The New Procedure

Let $X_n^* = \{X_1^*, \ldots, X_n^*\}$ be a random sample from the empirical distribution F_n. It is often called a *bootstrap sample*. Let $Y^m := \{X_{n1}^*, \ldots, X_{nm}^*\}$ be m bootstrap samples from F_n.

We calculate $y_j := \text{PTM}_\beta(X_{nj}^*)$ for $j = 1, \ldots, m$. Now we calculate depth of y_j with respect to $y^m := \{y_1, \ldots, y_m\}$: $\text{PD}(y_j, y^m)$ and then order y_j with respect to their depth from the smallest to the largest: $y_{(1)}, \ldots, y_{(m)}$ where $\text{PD}(y_{(1)}, y^m), \leq, \ldots, \leq \text{PD}(y_{(m)}, y^m)$.

Finally, we simply delete first $100 \cdot 2\alpha \cdot m\%$ points from $y_{(1)}, \ldots, y_{(m)}$. Then the interval (or closed convex hull in high-dimensions) formed by $y_{(\lfloor 100 \cdot 2\alpha \cdot m\% \rfloor + 1)}, \ldots, y_{(m)}$ is our $100(1 - 2\alpha)\%$ confidence interval for μ, where $\lfloor \cdot \rfloor$ is the floor function.

3. Simulation Study

Now we conduct simulation study to examine the performance of the new and classical t (or T^2) procedure based on 2000 (replication) samples from various distribution F (including $N(0,1)$, $t(3)$, and others). Set $\alpha = 0.025$ and $\beta = 0.078$; we consider the combinations of $n = 100$ with the bootstrap number $m = 300$, 500, 1000, and 2000.

We will confine attention to the average length (or area) of the confidence interval (or region) from both procedures as well as their coverage frequency of true parameter μ (which is assumed to be the mean of the F), which ideally should be close to 95%. If both procedures can reach the nominal level 95%, then it is obviously better to have a shorter (or smaller) confidence interval (or region) or smaller average length (or area) of the intervals (or regions).

3.1. One Dimension

Table 1 lists the simulation results at the normal and $t(3)$ distributions.

Inspecting the table immediately reveals that the bootstrap number m affects the average coverage of the new procedure, with the increase of m it gets closer to the nominal level 95%, while the average length of intervals gets slightly larger. Of course, it does not affect the t procedure which has nothing to do with bootstrap. Overall, both procedures are indeed (roughly) 95% procedure and the new one produces an interval on the average about 2%-3% shorter than that of the classical t procedure even at $N(0,1)$ case, and it becomes 12%-13% shorter in the $t(3)$ case.

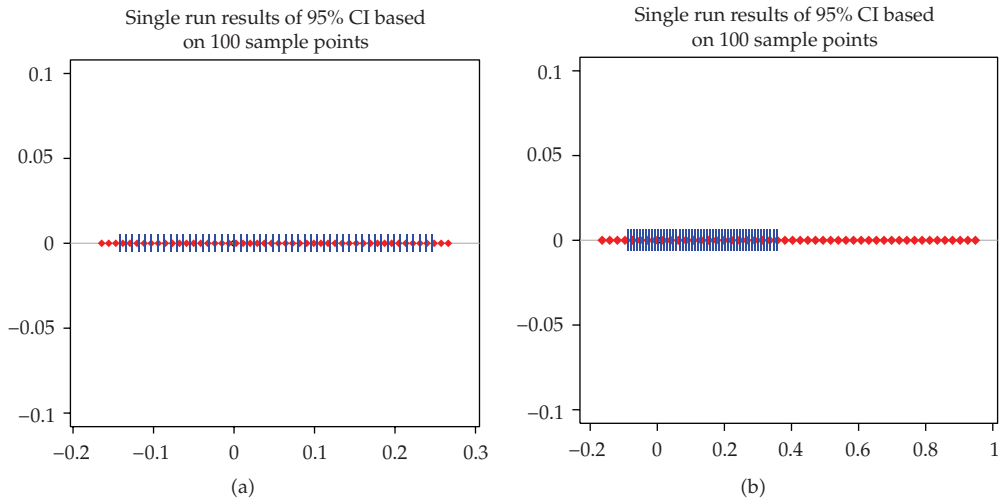

Figure 1: 95% confidence intervals by t (red one) and new procedures (blue one) for the mean of F based on 100 sample points from F: (a) $F = N(0,1)$ and (b) $F = t(3)$.

Figure 1 displays the typical single run results from two procedures based on 100 sample points from $N(0,1)$ (a) and $t(3)$ (b). We see even at $N(0,1)$ case, the new procedure outperforms the classical t procedure with a 95% confidence interval $[-0.1422550, 0.2463009]$ 10% shorter than that of t $[-0.1644447, 0.2661397]$, both cover the target parameter $\mu = 0$. At $t(3)$ case, new procedure produces an interval $[-0.0874566, 0.357108]$ 60% shorter than that of t : $[-0.1644486, 0.9482213]$. Both cover the target parameter $\mu = 0$.

In our simulation studies, we also compare our new procedure with the existing *bootstrap percentile confidence procedure* (i.e., it orders means of m bootstrap samples and then just to trim the upper and lower $\lceil 100 \cdot m \cdot \alpha\% \rceil$ points, the left points form an interval which is called bootstrap percentile conference interval, where $\lceil x \rceil$ is the ceiling function of x), our new procedure also outperforms this one. But the later performs better than the classical t procedure in term of the average length of intervals at the same confidence level.

Our experiments with n also reveal that small n (the real situation in practice) is in favor of our new procedure. Note that this is the exact case where it is difficult to determine if the data are *close to normal* and hence to decide if one is able to use the classical CI. This is what we expected since the classical CI is based on normal F (or on the large sample property for large n). But this does not mean that the classical CI has an edge over the new procedure at really large sample size n (say, 10,000) even for the perfect $N(0,1)$ case.

In addition to the distributions we considered in Table 1, we also conduct simulation studies to compare the performance of the new and classical t procedure at contaminated normal model: $(1-\epsilon)N(0,1) + \epsilon N(\mu, \sigma^2)$ with different choices of ϵ and (μ, σ^2) since we know in practice, there is never a pure (exact) $N(0,1)$; we may have just a slight departure from the pure normal or some contamination. Our results reveal that the new procedure is overwhelmingly more robust than the classical t; this is what we would expect since the t procedure depends on the sample mean which is notorious for its extreme sensitivity to outliers or contaminations. We also compare the performance of the two procedures at Cauchy distribution since we know that sample mean $\overline{x}$ performs extremely well at

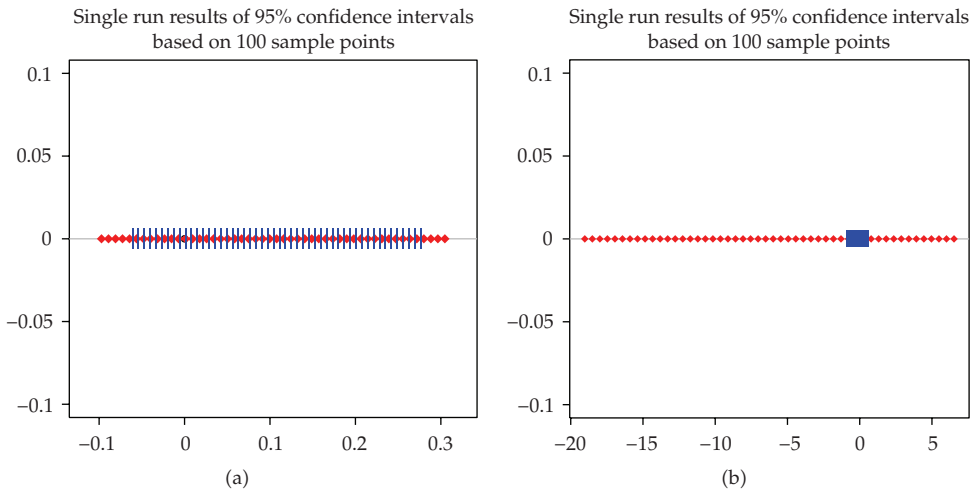

Figure 2: 95% CIs by t (red one) and new procedures (blue one) for the mean of F based on 100 sample points from F: (a) $F = 0.95N(0,1) + 0.5N(1.5, 0.1^2)$ and (b) $F = t(1)$.

Table 2: Average coverage (length) of 95% CIs by t and PTM_β.

$n = 100$	method	$.95N(0,1) + .05N(1.5, 0.1^2)$		Cauchy	
$m = 300$	PTM_β	.9455	(.3922)	.9585	(1.135)
	t	.9595	(.4070)	.9760	(57.76)
$m = 500$	PTM_β	.9525	(.3953)	.9765	(1.115)
	t	.9590	(.4081)	.9825	(20.10)
$m = 1000$	PTM_β	.9530	(.3972)	.9715	(1.164)
	t	.9600	(.4073)	.9775	(31.87)
$m = 2000$	PTM_β	.9485	(.3994)	.9725	(1.168)
	t	.9525	(.4077)	.9830	(25.61)

symmetric light tailed distributions like $N(0,1)$ but not so at heavy tailer ones like cauchy distribution.

We first display the typical single run results of 95% confidence intervals in Figure 2 to demonstrate the difference between the two procedures.

Here in Figure 2, on the left-hand side are 95% CI's by t (red one) and by our new procedure (blue one) at the model $0.95N(0,1) + 0.05N(1.5, 0.1^2)$ with an interval from t: $[-0.09723137, 0.3047054]$ and from new procedure $[-0.06083767, 0.2763990]$ which is 16% longer than that of t. These intervals are supposed to estimating the mean parameter μ in this case is $\epsilon \cdot u = 0.075$. So both intervals cover the unknown parameter μ.

On the right-hand side are 95% CIs by t (red one) and by new procedure (blue one) at the Cauchy distribution with an interval from t: $[-19.02593, 6.527279]$ and from new procedure: $[-0.8909354, 0.5884936]$ which is 94% shorter.

Of course, the single run results may not represent the overall performance of the two procedures. So we conduct a simulation over 2000 replications. The results are listed in Table 2.

Table 3: Average coverage (length) of 95% confidence regions by T^2 and PTM_β.

$n = 100$	method	$N_2(0,1)$		$t_2(3)$	
$m = 300$	PTM_β	.9501	(.1492)	.9515	(.3045)
	T^2	.9524	(.1935)	.9495	(.5137)
$m = 500$	PTM_β	.9395	(.1582)	.9577	(.3247)
	T^2	.9515	(.1947)	.9507	(.5204)
$m = 1000$	PTM_β	.9436	(.1672)	.9475	(.3438)
	T^2	.9547	(.1949)	.9353	(.5111)
$m = 2000$	PTM_β	.9403	(.1736)	.9586	(.3554)
	T^2	.9470	(.1949)	.9488	(.5202)

Inspecting the table immediately reveals that the classical t procedure becomes useless in the heavy tailed Cauchy distribution case: exceeding the nominal level 95% and reaching 98% with an extremely wide confidence interval, no informative any more. At the same time, the new procedure can roughly reach the nominal level 95% (it is about 97%) and provide a meaningful estimation about the underlying unknown parameter. We list the results from the contaminated model with just 5% contamination to a pure $N(0,1)$ model with the contamination also come from a normal distribution centered at 1.5 and with a small variance 0.01. Under such a potential real situation, the classical t 95% procedure becomes again useless since it can never reach the nominal level 95%, it is a roughly 96% procedure with an interval slightly longer than that of the new procedure, while the new procedure still is a reasonable 95% procedure with an interval on the average 2%–4% shorter than that of t one.

3.2. Higher Dimensions

In higher dimensions, with the multivariate version of PTM and PD (see Zuo [4], Zuo [2]) it is straightforward to extend our new procedure described in Section 2. That is, with the m bootstrap sample: $Y^m = \{X^*_{n1}, \ldots, X^*_{nm}\}$ we calculate $y_j := \text{PTM}_\beta(X^*_{nj})$ for $j = 1, \ldots, m$. Then we calculate the projection depth of y_j with respect to $y^m := \{y_1, \ldots, y_m\}$: $\text{PD}(y_j, y^m)$ and then we order y_j's with respect to their depth from smallest to largest: $y_{(1)}, \ldots, y_{(m)}$ where $\text{PD}(y_{(1)}, y^m), \leq, \ldots, \leq \text{PD}(y_{(m)}, y^m)$. The final step is the same as before: trimming first $100 \cdot 2\alpha \cdot m\%$ points from y_j's the left formed a convex hull, that is our $100(1 - 2\alpha)\%$ confidence region for μ. We will examine the performance of this one and the classical Hotelling's T^2 given in (1.2) in term of their average area of confidence regions as well as their coverage frequency of true parameter μ (which is assumed to be the mean of the F). The latter ideally should be close to 95%. If both procedures can reach the nominal level 95%, then it is obviously better to have a smaller confidence region or smaller average area of confidence regions.

We first display single run results of two procedures at bivariate standard normal distribution $N_2(0,1)$ and bivariate t distribution with 3 degrees of freedom $t_2(3)$ in Figure 3.

Of course, single run result may not represent the overall performance of the two procedures. To see if the single run results are repeatable now we list the average of coverage and the area of the confidence regions based on two procedures in 2000 replications in Table 3. Here we set $\beta = 0.1$, $n = 100$ and $\alpha = 0.025$.

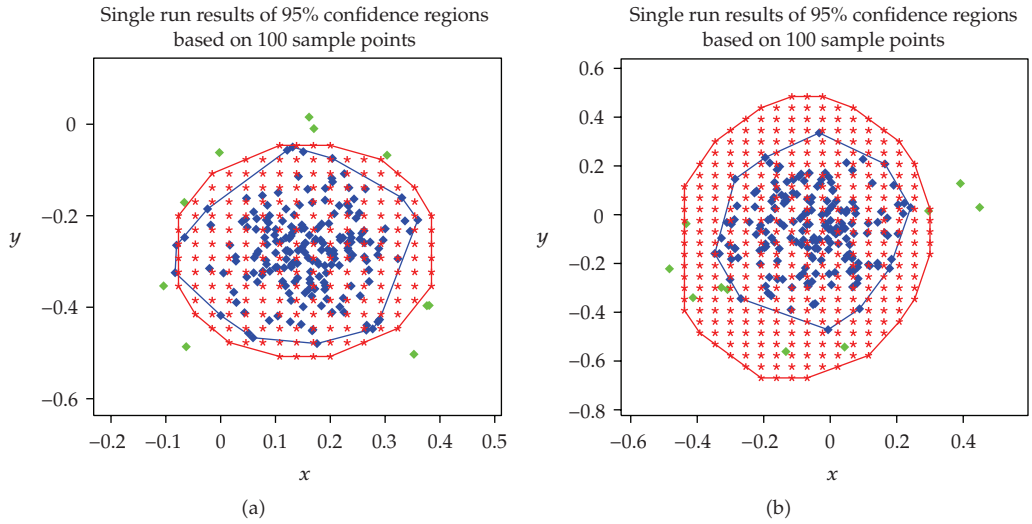

Figure 3: 95% confidence regions by T^2 (red one) and new procedures (blue one) for the mean of F based on 100 sample points from F: (a) $F = N_2(0,1)$, (b) $F = t_2(3)$.

Inspecting the Table reveals that the two procedures are indeed (roughly) 95% confidence procedures. Therefore it make sense to compare their average area of confidence regions. The table entries show that the new procedure produce a confidence region on the average 11%–22% smaller than that of the classical Hotelling's T^2 procedure in term of area even at $N_2(0,1)$. This becomes 32%–40% in $t_2(3)$ case.

4. Concluding Remarks

From the last section we see that the new procedure has some advantages over the classical (seemingly optimal) procedures. But we know that we cannot get all the advantages of the new procedure for free. What kind of price we have to pay here? For all the advantage of the new procedures possess over the classical ones, the price it has to pay is the intensive computing in the implement of the procedure. In our simulation study, there are 4 million basic operations (the case $n = 100$, replication $R = 2000$, and bootstrap number $m = 2000$). Computing the data depth in two or higher dimensions is very challenging. Fortunately, there is a R package (called ExPD2D) for the exact computation of projection depth of bivariate data already developed by Zuo and Ye [5] and is part of CRAN now. For high-dimensional computation, see Zuo [6]. In one dimension it is straightforward. One can compute the sample median in linear time (i.e., the worst case time complexity is $O(n)$) by employing special technic (see any computer science Algorithm textbook), for further discussion about the property of related remedian, see H. Chen and Z. Chen [7]. Fortunately, in practices, only one replication is needed. Also with the everlasting advance in computing power, the computation burden should not be an excuse for not using a better procedure.

A natural question is Why the new procedure has advantage over the classical one? The procedure clearly depends on bootstrap and data depth. Is it due to bootstrap or data depth? Who is the main contributor? If one just uses bootstrap, can one have some advantages? The answer for the latter is positive, Indeed, in our simulation we compare

the classical one with the bootstrap percentile procedure, it reveals that the bootstrap percentile one does have some mild advantage over the classical one but still is inferior to our new procedure. So both bootstrap and data depth make contributions to the advantages of the new procedure. But remember, it is data depth that allow the bootstrap percentile procedure (which originally was defined only in one dimension) implementable in high-dimensions: to order sample bootstrap mean vectors. Without the data depth, it is impossible to implement the procedure in high-dimensions. So overall, it is data depth that makes the major contribution towards the advantages of the new procedure.

We also like to point out at this point that there is different new procedure introduced and studied in Zuo [8], where depth-weighted mean used in the procedure instead of the depth-trimmed mean used in our current procedure. However, our simulation studies indicate that our current new procedure is superior to the one in Zuo [8] which confines attention mainly to one dimension.

Our empirical evidence for the new procedure in one and higher dimensions is very promising, but we still need some theoretical developments and justifications, which is beyond scope of this paper and will be pursued elsewhere. A heuristic argument is because the bootstrap percentile confidence interval has advantage over the classic confidence interval procedure in term of at the same nominal level it can produce an asymptotically shorter interval (see Hall [9], and Falk and Kaufmann [10]). But the classical bootstrap percentile interval procedure is limited to one dimension, here we use data depth to ordering high-dimensional estimators so that we can extend the procedure to high-dimensions. The advantage of bootstrap percentile confidence interval carries on to high-dimensions.

One question left about our new procedure in practices is how does one choose the β value? Well, there are at least two ways to deal with this β value problem. First, one can chose a fixed value, our empirical experience indicates a value between 0.01-0.1 will serve most of our purposes. Or (second), dynamically choose β value by minimizing some objective function which could be your interval length in our simulation case or variance in the efficiency evaluation case. With such a data dependant β, one natural question raised is: Is the theory in Zuo [2] established based on the fixed constant β still holds? Fortunately, all still hold if we employ a more powerful tool (empirical process theory) from Pollard [11] or van der Vaart and Wellner [12] to handle this situation with a data dependent β.

There are a number of depth functions and related depth estimators (see Tukey [13], Liu [14], Zuo and Serfling [3], and Bai and He [15]), but among them projection depth function used here is the most favorite one (see Zuo [4, 16]). Furthermore, the computation of depth functions all are very challenging but we have some algorithm at hand for the projection depth function, this is yet another motivation for us to pick the projection depth function in this paper.

Finally, we comment that findings in this paper are consistent with the results obtained in Bai and Saranadasa (BS) [17] which shows the *Effect* of high-dimension, that is, there are better procedures than the classical inference procedures like Hotrlling's T^2 one which is inferior compared to other procedures like Dempster's nonexact test (Dempster [18]) and BS proposed test even for moderately large dimension and sample sizes.

Acknowledgment

This research was partially supported by NSF Grants DMS-0234078 and DMS-0501174.

References

[1] M. Wu and Y. Zuo, "Trimmed and Winsorized means based on a scaled deviation," *Journal of Statistical Planning and Inference*, vol. 139, no. 2, pp. 350–365, 2009.

[2] Y. Zuo, "Multi-dimensional trimming based on projection depth," *The Annals of Statistics*, vol. 34, no. 5, pp. 2211–2251, 2006.

[3] Y. Zuo and R. Serfling, "General notions of statistical depth function," *The Annals of Statistics*, vol. 28, no. 2, pp. 461–482, 2000.

[4] Y. Zuo, "Projection-based depth functions and associated medians," *The Annals of Statistics*, vol. 31, no. 5, pp. 1460–1490, 2003.

[5] Y. Zuo and X. Ye, "ExPD2D: Exact Computation of Bivariate Projection Depth Based on Fortran Code. R package version 1.0.1," 2009, http://CRAN.R-project.org/package=ExPD2D.

[6] Y. Zuo, "Exact computation of the bivariate projection depth and Stahel-Donoho estimator," accepted to *Computational Statistics & Data Analysis*.

[7] H. Chen and Z. Chen, "Asymptotic properties of the remedian," *Journal of Nonparametric Statistics*, vol. 17, no. 2, pp. 155–165, 2005.

[8] Y. Zuo, "Is t procedure: $\overline{x} \pm t_{1-\alpha/2}(n-1)s/\sqrt{n}$ optimal?" accepted to *The American Statistician*.

[9] P. Hall, "Theoretical comparison of bootstrap confidence intervals," *The Annals of Statistics*, vol. 16, no. 3, pp. 927–985, 1988.

[10] M. Falk and E. Kaufmann, "Coverage probabilities of bootstrap-confidence intervals for quantiles," *The Annals of Statistics*, vol. 19, no. 1, pp. 485–495, 1991.

[11] D. Pollard, *Convergence of Stochastic Processes*, Springer Series in Statistics, Springer, New York, NY, USA, 1984.

[12] A. W. van der Vaart and J. A. Wellner, *Weak Convergence and Empirical Processes with Applications to Statistics*, Springer Series in Statistics, Springe, New York, NY, USA, 1996.

[13] J. W. Tukey, "Mathematics and the picturing of data," in *Proceedings of the International Congress of Mathematicians*, vol. 2, pp. 523–531, Vancouver, Canada, August 1974.

[14] R. Y. Liu, "On a notion of data depth based on random simplices," *The Annals of Statistics*, vol. 18, no. 1, pp. 405–414, 1990.

[15] Z.-D. Bai and X. He, "Asymptotic distributions of the maximal depth estimators for regression and multivariate location," *The Annals of Statistics*, vol. 27, no. 5, pp. 1616–1637, 1999.

[16] Y. Zuo, "Robustness of weighted L_p-depth and L_p-median," *Allgemeines Statistisches Archiv*, vol. 88, no. 2, pp. 215–234, 2004.

[17] Z. Bai and H. Saranadasa, "Effect of high dimension: by an example of a two sample problem," *Statistica Sinica*, vol. 6, no. 2, pp. 311–329, 1996.

[18] A. P. Dempster, "A high dimensional two sample significance test," *Annals of Mathematical Statistics*, vol. 29, pp. 995–1010, 1958.