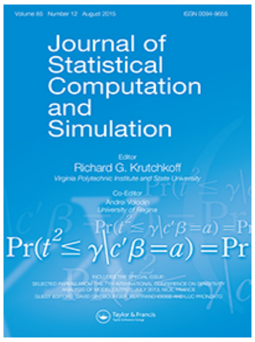


Journal of Statistical Computation and Simulation

ISSN: 0094-9655 (Print) 1563-5163 (Online) Journal homepage: http://www.tandfonline.com/loi/gscs20

# Confidence sets based on the positive part James–Stein estimator with the asymptotically constant coverage probability

**S. Ejaz Ahmed, Iskander Kareev, Sujitta Suraphee, Andrei Volodin & Igor Volodin**

**To cite this article:** S. Ejaz Ahmed, Iskander Kareev, Sujitta Suraphee, Andrei Volodin & Igor Volodin (2015) Confidence sets based on the positive part James–Stein estimator with the asymptotically constant coverage probability, Journal of Statistical Computation and Simulation, 85:12, 2506-2513, DOI: 10.1080/00949655.2014.933223

**To link to this article:** http://dx.doi.org/10.1080/00949655.2014.933223

Published online: 15 Jul 2014.

*Journal of Statistical Computation and Simulation*, 2015
Vol. 85, No. 12, 2506–2513, http://dx.doi.org/10.1080/00949655.2014.933223



# Confidence sets based on the positive part James–Stein estimator with the asymptotically constant coverage probability

S. Ejaz Ahmed[a], Iskander Kareev[b], Sujitta Suraphee[c], Andrei Volodin[d]* and Igor Volodin[b]

*[a]Department of Mathematics, Brock University, Niagara Region, 500 Glenridge Ave., St. Catharines, ON, Canada L2S 3A1; [b]Department of Mathematical Statistics, Kazan State University, Kazan-8, Kremlevskaya st. 18, 420008, Russia; [c]Department of Mathematics, Faculty of Science, Mahasarakham University, Mahasarakham 44150, Thailand; [d]Department of Mathematics and Statistics, University of Regina, Regina, SK, Canada S4S 0A2*

(*Received 1 December 2013; accepted 6 June 2014*)

The asymptotic expansions for the coverage probability of a confidence set centred at the James–Stein estimator presented in our previous publications show that this probability depends on the non-centrality parameter $\tau^2$ (the sum of the squares of the means of normal distributions). In this paper we establish how these expansions can be used for a construction of confidence region with constant confidence level, which is asymptotically (the same formula for both case $\tau \to 0$ and $\tau \to \infty$) equal to some fixed value $1-\alpha$. We establish the shrinkage rate for the confidence region according to the growth of the dimension $p$ and also the value of $\tau$ for which we observe quick decreasing of the coverage probability to the nominal level $1-\alpha$. When $p \to \infty$ this value of $\tau$ increases as $O(p^{1/4})$. The accuracy of the results obtained is shown by the Monte-Carlo statistical simulations.

**Keywords:** confidence sets; positive part James–Stein estimator; multivariate normal distribution; coverage probability; asymptotical expansions; second-order asymptotic

*AMS Subject Classification*: Primary: 62E20; Secondary: 62F10

## 1. Introduction

The shrinkage method plays an important role in studying estimation of the mean vector of multivariate normal distribution, estimation of common mean of several populations and estimation of regression parameters in a host of models. Since the inception of Stein effect by Stein [1,2] and James and Stein,[3] the shrinkage method has been studied in the field of multivariate statistical inference. Stein effect indicates that a suitable shrinkage estimator method may be better than a usual unbiased estimator. The shrinkage method for the mean parameter vector of multivariate normal population is recently addressed by, e.g. Tsukuma et al. [4] and Eldar and Chernoi.[5] Hartung et al. [6, Chapter 5] summarized finite sample and asymptotic results of estimating the common mean of two or more populations.

Research on the statistical implications of proposed and related estimators is ongoing and it is practically not possible to mention all important publications in this area. Because of that

*Corresponding author. Emails: andrei@uregina.ca; volodin@uregina.ca

we restricted ourselves to present in the References only publications that are crucial for the current article.

It is worth mentioning that this is one of the two areas Bradley Efron predicted for the early twenty-first century (RSS News, January 1995). Shrinkage and likelihood-based methods continue to be extremely useful tools for efficient estimation.

To our knowledge, however, there are few papers to deal with a confidence set problem based on shrinkage estimator. The confidence set problem for shrinkage estimators are developed by Ahmed et al.[7,8] Alternatively, Kazimi and Brownstone [9] proposed confidence bands for shrinkage estimators using a simple percentile bootstrapping method. They find that '... simple percentile bootstrap confidence bands perform well enough to support empirical applications of shrinkage estimators',[9, p. 99] although there remain issues in using bootstrapping methods in this way, for the typical type of econometric model one encounters, where bootstrap sub-sampling is required to ensure consistency, and does not yield consistency in all cases.

In this paper we address the confidence estimation problem of the mean vector $\theta = (\theta_1, \ldots, \theta_p)$ of the $p$-variate normal distribution with independent components and equal unit variances $\sigma^2 = 1$. Let $\bar{X} = (\bar{X}_1, \ldots, \bar{X}_p)$ be the sample mean vector that is calculated from samples of equal size $n$ of marginal distributions. The confidence set

$$D_{\bar{X}} = \{\theta : n\|\theta - \bar{X}\|^2 \leq c^2\}$$

has the given confidence coefficient $1 - \alpha$, if $c^2$ is defined as the quantile of the central chi-square distribution with $p$ degrees of freedom according to the formula $K_p(c^2) = 1 - \alpha$, where $K_p(\cdot)$ is the chi-square distribution function.

The confidence set $D_{\bar{X}}$ possesses the minimax property, but there exist other minimax sets which may provide improved coverage probability. In this paper we are mostly interested in the coverage properties of one of such sets, namely

$$D_{\delta^+} = \{\theta : n\|\theta - \delta^+(\bar{X})\|^2 \leq c^2\},$$

which is centred at the positive modification of the James–Stein estimator, cf.[3]

$$\delta^+(\bar{X}) = \left(1 - \frac{p-2}{n\|\bar{X}\|^2}\right) \bar{X}\, I\left(n\|\bar{X}\|^2 > p-2\right) = \left(1 - \frac{p-2}{n\|\bar{X}\|^2}\right)^+ \bar{X}.$$

We restrict $p$ to be greater than 2, the notation $a^+ = \max(0, a)$, and we consider euclidian norm $\|\theta\| = \sqrt{\sum_{i=1}^p \theta_i^2}$.

There are many publications on shrinkage estimation topic by various authors. We are mostly interested in the results connected with asymptotic investigations of the coverage probability of the true value of vector $\theta$, and we refer to the bibliography presented in Ahmed et al.[7]

In Ahmed et al.,[7] a novel approach to the approximation of the coverage probability was developed, which is based on a combination of geometrical and analytical methods. It was established that $Q_p^+(\tau) = P(D_{\delta^+})$ depends on the values of the vector $\theta$ via the parameter $\tau^2 = n\|\theta\|^2$; in fact it is a decreasing function of $\tau^2$, with $Q_p^+(\tau) = K_p(w(c^2, \tau)) + R_p(\tau)$, where

$$w(c^2, \tau) = p - 2 + \frac{c^2 - \tau^2}{2} + \sqrt{\frac{(c^2 - \tau^2)^2}{4} + c^2\tau^2 - (p-2)(\tau^2 - c^2)}. \quad (1)$$

The second term $R_p(\tau)$ is represented as a double integral and it is established that $R_p(\tau) = O(\tau^2)$ for $\tau \to 0$, and $R_p(\tau) = O(\tau^{-2})$ for $\tau \to \infty$. It is important to note that we have the general formula that deals with both cases of the asymptotic behaviour of $\tau$ for the coverage probability by the confidence set centred at the positive-part James–Stein estimator.

In Ahmed et al.,[8] the second-order asymptotic expansions of coverage probabilities for $\tau \to 0$ and $\tau \to \infty$ were established. Numerical illustrations for the same values of $p$ and $\tau$ as in Ahmed et al. [7] show that for $\tau$ much bigger than $c$ and $\tau$ very small, the third term of the asymptotic has the magnitude of order $10^{-4}$, but the general picture is such that the third term makes the approximation even worse, especially for small values of $p$. Therefore, the initial approximations presented above might be interpreted from the practical point of view as approximations of the order $\tau^{\pm 3}$.

Note that James–Stein estimators are applied in the case when it is known that the true values of the means $\theta_1, \dots, \theta_p$ are concentrated close to some common value $\mu$. This is the reason why we use the term 'shrinkage' for these estimates. Usually the value of $\mu$ is known from the essence of a statistical problem that we consider. For example, $\mu$ may be the value of a characteristic of some standard object, while $\theta_1, \dots, \theta_p$ are the values of the same characteristic for some other objects, which should be close to $\mu$. On the other hand, according to data of measures of the characteristics for each of these objects, or according to data of measures of the characteristic for some similar objects, we can estimate the value of the centre of shrinkage $\mu$ of our observations. Note that we need also an estimate of the variance $\sigma^2$ in order to construct asymptotically confidence sets that are suggested in this paper.

Without loss of generality in this paper it is assumed that $\mu = 0$ and $\sigma = 1$. If the concentration point and the variance is different from these values, in all formulae we should substitute the vector $\bar{X}$ by $(\bar{X} - \mu)/\sigma$. All properties of the shrinkage estimations with a large amount of examples can be found in Chapter 5 of Lehmann and Casella.[10]

In this paper we show how using the approximation above, it is possible to construct a confidence region with constant, asymptotically equal ($\tau \to 0$ and $\tau \to \infty$) to the fixed confidence level $1 - \alpha$. The method that we use to construct these confidence intervals borrows some ideas from Hwang and Casella.[11] We establish the rate of shrinkage of the confidence region as the dimension $p$ increases. Also we find the value of $\tau$ for which happens the quick drop of the coverage probability to the nominal level $1 - \alpha$. Note that the coverage probability before this value of $\tau$ is close to one and for large $p$ even equals one. When $p \to \infty$ this value of $\tau$ increases as $O(p^{1/4})$. The accuracy of the asymptotic obtained is illustrated by the statistical simulations.

## 2. Confidence region with asymptotically constant coverage probability

The basis of the confidence region $D_{\delta^+}$ consists of the random function

$$T^2 = T^2(\theta, \delta^+) = n \sum_{i=1}^{p} (\theta_i - \delta_i^+)^2$$

of the vector parameter $\theta$ and vector statistic $\delta^+(\bar{X})$ with components

$$\delta_i^+ = \left(1 - \frac{p-2}{n \sum_{j=1}^{p} \bar{X}_j^2}\right)^+ \bar{X}_i, \quad i = 1, \dots, p.$$

The transformation of $D_{\delta^+}$ to a region with asymptotically constant coverage probability is based on the fact that $w(c^2, \tau)$ is a strictly increasing function of the variable $c$, if the expression under the root in Equation (1) is positive. This is simple to establish by rewriting the function in the form

$$w(c^2, \tau) = q + \lambda + \sqrt{\lambda^2 + 2\lambda\tau^2 + \tau^4 + 2q\lambda},$$

where $q = p - 2$ and $\lambda = (c^2 - \tau^2)/2$.

The region of real values for the function $w(c^2, \tau)$ is defined by the inequality

$$c^2 \geq c_0 = 2\sqrt{2\tau^2 q + q^2} - 2q - \tau^2.$$

Note that $c_0 < \tau^2$ and $c_0 \sim \tau^2$ for $\tau \to 0$, that is, the complex values of the function $w(c, \tau)$ belong to the regions of small values of $\tau$ and $c$.

Introduce the subregions

$$D = \{\theta : w(T^2, \tau) \leq c^2, T^2 \geq c_0\} \cup \{T^2 < c_0\} \quad \text{and} \quad D_0 = \left\{\theta : T^2 \leq \left(c - \frac{q}{c}\right)^2\right\}$$

of the parametric space $\mathrm{R}^p$.

THEOREM *Let $c^2 = K_p^{-1}(1-\alpha)$, $\alpha \ll 0,5$. Then for $Q_p$, the coverage probability of the true value of the parameter $\theta$ by subregion $D$ the following asymptotic equalities are true $Q_p = 1 - \alpha + O(\tau^2)$ if $\tau \to 0$, and $Q_p = 1 - \alpha + O(\tau^{-2})$ if $\tau \to \infty$.*

*The region $D$ is a part of the region $D_{\delta^+}$ and covers the region $D_0$, that is,*

$$D_0 \subseteq D \subseteq D_{\delta^+}.$$

*Proof* The first statement of the theorem concerning the confidence region $D$ follows directly from the results of Ahmed et al.,[7] see formula (1). Note that

$$\begin{aligned}
Q_p &= P(T^2 \leq w_\tau^{-1}(c^2), T^2 \geq c_0) + P(T^2 < c_0) \\
&= K_p(w(w_\tau^{-1}(c^2), \tau)) - K_p(c_0) + K_p(c_0) + O(\tau^2) \\
&= K_p(c^2) + O(\tau^2) = 1 - \alpha + O(\tau^2).
\end{aligned}$$

The region

$$\begin{aligned}
D_0 &= \{\theta : w(T^2, 0) \leq c^2\} \\
&= \left\{\theta : \frac{T^2}{2} + q + \left(\frac{T^4}{4} + qT^2\right)^{1/2} \leq c^2\right\} \\
&= \left\{\theta : T^2 \leq \left(c - \frac{q}{c}\right)^2\right\}
\end{aligned}$$

corresponds to the maximum value $w(c^2, \tau)$ as a function of $\tau$ for each fixed value of $c$ (recall that $w(c^2\alpha \ll 0, 5., \tau)$ is strictly decreasing function of $\tau$). The region $D_{\delta^+}$ corresponds to the minimal value of $w$ for $\tau \to \infty$. These two remarks prove the required inclusion relations. The theorem is proved. ■

The theorem provides the smallest (by inclusion) set $D_0$ which is not a confidence set for $\tau \neq 0$, but is the 'limit' of confidence sets $D$ when $\tau \to 0$. Region $D_0$ corresponds to the maximum gain in the size of confidence regions that were constructed on the principle of constant (at least asymptotically) coverage probability. If the true value $\tau = 0$, then the confidence coefficient of this region equals exactly to the given level $1 - \alpha$. We also note that when $\tau = 0$ there are no negative values under the square root of the function $w(\tau, c)$.

At the same time, set $D$ may be much smaller by size than confidence set $D_{\delta^+}$, which is usually used. Region $D_0$ is a ball of the radius $c - (p-2)/c$. With respect to the region $D_{\delta^+}$, the radius is smaller in $(1 - (p-2)/c^2)^{-1}$ times. Since for $p \to \infty$ the quantile $c^2 = K^{-1}(1-\alpha) \sim p$, we

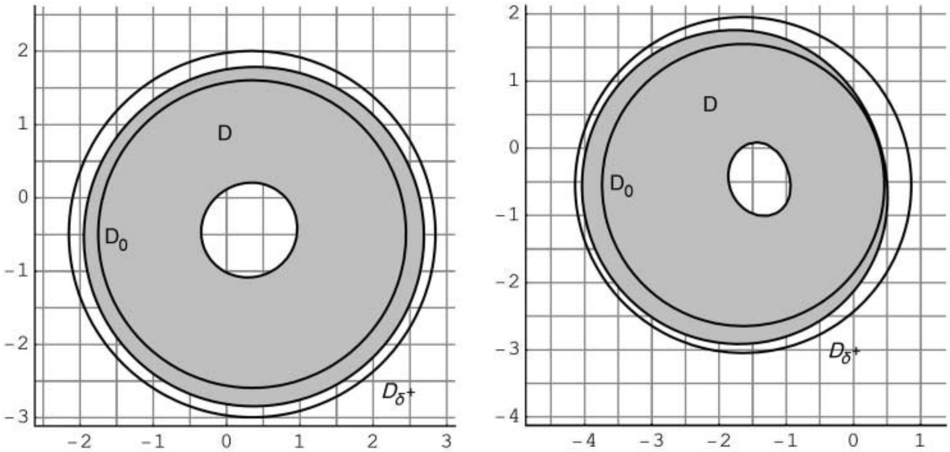


Figure 1. Relations between the regions $D_0$, $D$, and $D_{\delta^+}$ for $p = 3$.

can state that with the growth of the number $p$ of components of the vector $\overline{X}$ the region $D_0$ shrinks to a point. Hence for small values of $\|\theta\|$ (as we already mentioned, the true value lies close by the shrinkage point) and large $p$ the radius $D$ is close to the radius of the region $D_0$ and the confidence set $D$ has much smaller size than the confidence region $D_{\delta^+}$.

Relations between the regions $D_0$, $D$, and $D_{\delta^+}$ for $p = 3$ are presented in Figure 1. Note that the region $D$ is the shaded area with the white spot in the middle. This white spot corresponds to the negative values under square root of the function $w(\tau, c)$ and from the figures we see that it is a significant part of the confidence set $D$. The left picture corresponds to $\tau^2 = 0.03, \theta_1 = \theta_2 = \theta_3 = 0.1$, the right picture to $\tau = 1, \theta_1 = \theta_2 = \theta_3 = 1/\sqrt{3}$. For the given values of $\theta$, realizations of $(x_1, x_2, x_3)$ and $(y_1, y_2, y_3)$ as independent normal $(\theta_i, 1)$ random variables were obtained and their corresponding regions $D_0$, $D$ and $D_{\delta^+}$ are calculated. The figures provide projections on the plane $(\theta_1, \theta_2)$ of levels of these regions when $\theta_3 = x_3$ (left figure) and when $\theta_3 = y_3$ (right figure).

## 3. Graphical illustrations of the coverage probabilities

As we already mentioned, the coverage probability by the region $D$ of the true value of the parametric vector $\theta$ depends on $\theta$ only through the values of the parametric function $\tau$. Therefore, $Q_p = Q_p(\tau)$ and the plots of these functions for the values $p = 3, 5, 10, 15$ and $n = 1$ are presented in Figure 2 ($\alpha = 0.01$), Figure 3 ($\alpha = 0.10$), and Figure 5 ($\alpha = 0.05$). In Figure 4, we present similar plots of $Q_p(\tau)$ when $\alpha = 0.05$ and the sample size $n = 10$. For a comparison in Figure 6 we present ($\alpha = 0.05$) the coverage probabilities of $\theta$ by the James–Stein confidence region $D_{\delta^+}$ (thick lines) and their approximation $K_p(w(c^2, \tau))$ (thin lines). Calculations were conducted by the Monte-Carlo method with $10^6$ replications for each fixed values of $p$ and $\tau = 0.0(0.1)10.0(0.5)30.0$. The values of $\theta_i,\ i = 1, \ldots, p$ were chosen identical and equal to $\sqrt{\tau^2/p}$ because the coverage probabilities depend on the coordinates of vector $\theta$ only through the symmetric functions $T^2$ and $\tau^2$. With high probability guaranteed, the accuracy of calculations of the coverage probability values is of the order 0.002.

The graphical illustrations provide the following conclusions. The coverage probability constancy corresponding to the nominal $1 - \alpha$ can be guaranteed only in the neighbourhood $\tau < 1$ of the shrinkage point $\theta = 0$. Approximately up to the values $2 \le \tau \le 3$, the coverage probability is insignificantly higher than the nominal. ‘A catastrophe’ happens after these values of $\tau$, and as

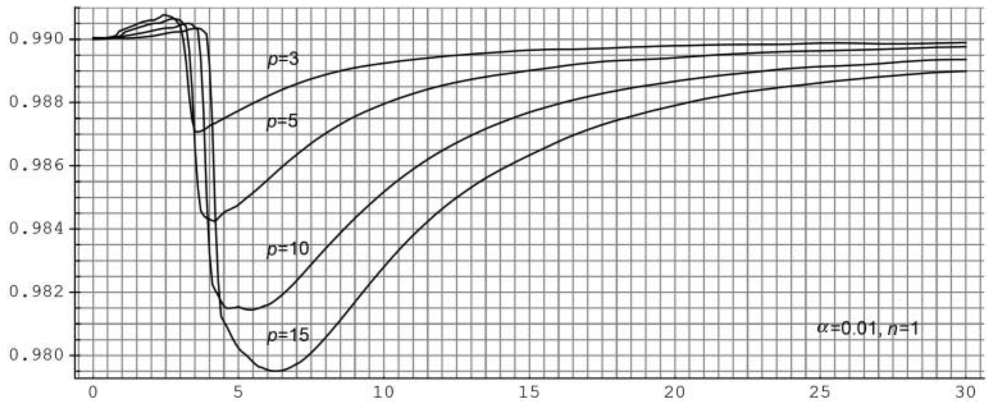


Figure 2. Plots of $Q_p$ for the values $p = 3, 5, 10, 15$ for $\alpha = 0.01$ and $n = 1$.

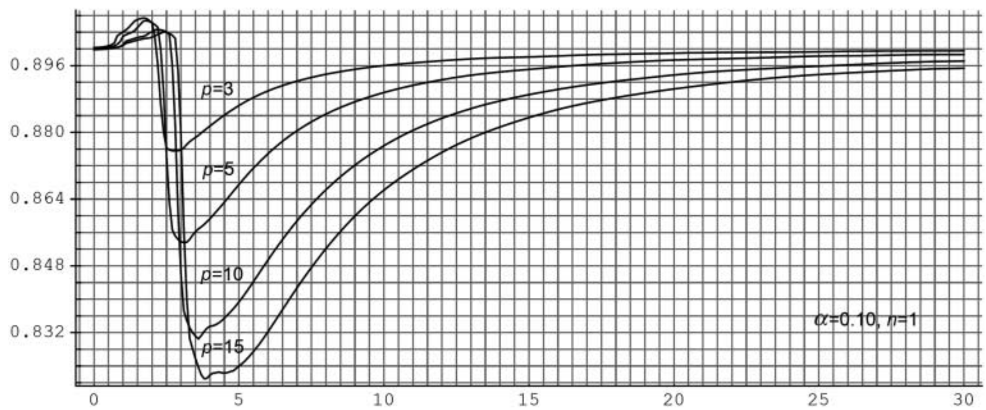


Figure 3. Plots of $Q_p$ for the values $p = 3, 5, 10, 15$ for $\alpha = 0.1$ and $n = 1$.

plots in Figures 5 and 6 show, the point of the breakdown corresponds to the point of the sharp slump of coverage probability by the James–Stein confidence region. Moreover, this is exactly the point where our approximation of the coverage probability is equal to its exact value.

The maximum value of the difference $(1 - \alpha) - Q_p(\tau)$ strongly depends on the values of $p$ and increases with a growth of $p$. Thus, for $p = 5$, the actual coverage probability is close to $1 - 1.5\alpha$, while for $p = 15$ it reduces to $1 - 2\alpha$.

Therefore, we could recommend using the confidence region $D$ only in the case when the statistician has information that the true values of the parametric vector $\theta$ are close to the shrinkage point, otherwise it is better to use confidence sets centred at the positive part James–Stein estimator $D_{\delta^+}$ or at the sample mean $D_{\bar{X}}$.

In connection with this, we could state a problem of the detection of point $\tau_0$ starting from which a quick decreasing of coverage probability of true parametric values by the region $D$ appears. Obviously this point is a function of $c$ and dimension $p$, as we can observe from the plots of $Q_p^+(\tau) = P_\tau(T^2 \leq c^2)$. If this point is passed, then we observe a rapid decreasing to the nominal level $1 - \alpha$ of the coverage probability by the confidence region $D_{\delta^+}$. Recall that at this point the coverage probability is $Q_p^+(\tau) = K_p(w(c^2, \tau))$, that is, coincides with the approximation suggested in Ahmed et al.[7, Figure 4] Hence there exists point $\tau(c)$ for $Q_p(\tau)$ for the function $Q_p^+(\tau)$ in which happens the quick decrease of coverage probability by region $D$.

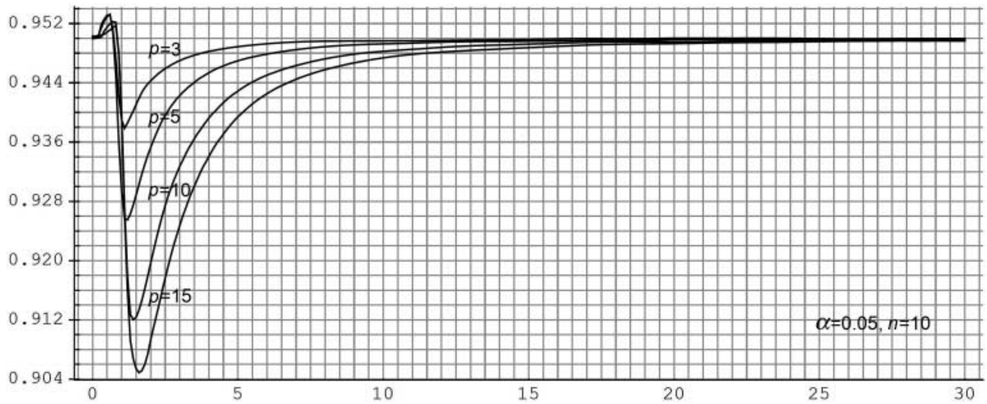


Figure 4. Plots of $Q_p$ for the values $p = 3, 5, 10, 15$ for $\alpha = 0.05$ and $n = 10$.

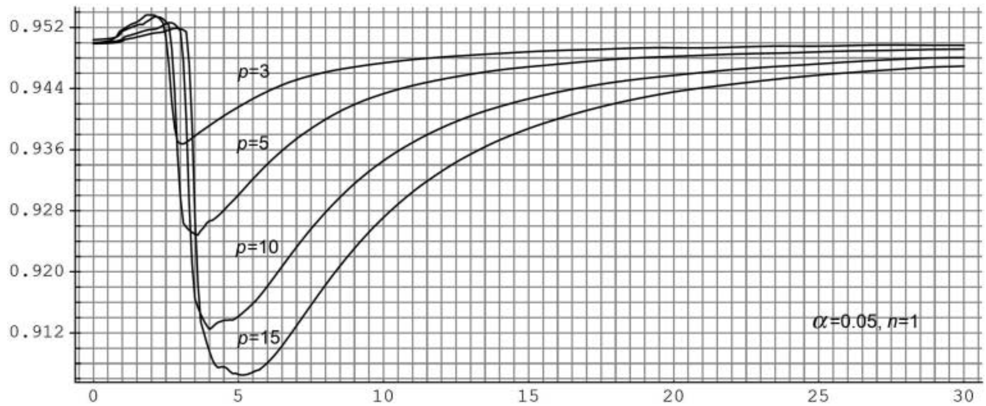


Figure 5. Plots of $Q_p$ for the values $p = 3, 5, 10, 15$ for $\alpha = 0.05$ and $n = 1$.

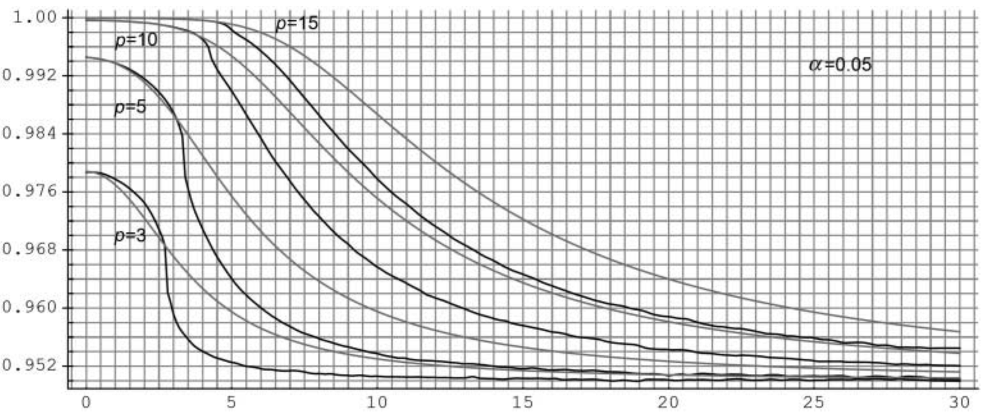


Figure 6. The coverage probabilities of $\theta$ by the James-Stein confidence region $D_{\delta^+}$ (thick lines) and their approximation $K_p(w(c^2, \tau))$ (thin lines) with $\alpha = 0.05$.

We can show that $\tau(c) = \sqrt{c^2 - (p-2)}$. Note that $Q_p(\tau) = P_\tau(T^2 \le w_\tau^{-1}(c^2)) = Q_p^+(w_\tau^{-1}(c^2))$ and hence $\tau(c)$ is the root of the equation $\tau = w_\tau^{-1}(c^2)$ or, which is equivalent, $w(\tau^2, \tau) = c^2$. The last equality gives $c^2 = p - 2 + \tau^2$.

Calculations of $\tau(c)$ for the values of $p$ and $\alpha$ that we used in the graphical illustrations provided above, show that this is correct and the obtained value of $\tau(c)$ is the change point. Note that since $c^2 \sim p + \Phi^{-1}(1-\alpha)\sqrt{2p}$ when $p \to \infty$, we have that $\tau^2(c) \sim \Phi^{-1}(1-\alpha)\sqrt{2p} + 2$. This gives us a conclusion that the interval $[0, \tau(c)]$ of preferable usage of the confidence region $D$ is increasing when the dimension $p$ of the observed vector increases. For $p = 8$ the length of this interval is approximately 3.

## 4. Concluding remarks

We constructed a confidence region $D$ with confidence level, which is asymptotically ($\tau \to 0$ or $\tau \to \infty$) equal to some fixed value $1 - \alpha$. The confidence region $D$ may be much smaller by size than usually used confidence set $D_{\delta^+}$. Moreover a point $\tau$ is detected, starting from which a quick decreasing of coverage probability of true parametric values by the region $D$ appears. This provides us an interval of values of $\tau$ when it is preferable to use of the confidence region $D$, than the usual confidence set $D_{\delta^+}$.

The point of the probability 'breakdown' is the most important result of this paper. The asymptotic obtained in our previous papers work poorly exactly in the neighbourhood of this point. It is interesting to investigate the asymptotic of the remainder term $R_p(\tau)$ from the asymptotic representation of the coverage probability as $\tau = \sqrt{c^2 - (p-2)} = O(p^{1/4})$ and $p \to \infty$.

## Acknowledgement

The authors would like to thank reviewers for useful remarks that strongly improved the presentation.

## References

[1] Stein C. Inadmissibility of the usual estimator for the mean of a multivariate distribution. Proc Third Berkeley Symp Math Statist Prob. 1956;1:197–206.
[2] Stein C. Confidence sets for the mean of a multivariate normal distribution. J R Stat Soc Ser B. 1962;24:265–296.
[3] James W, Stein C. Estimation with quadratic loss. Proc Fourth Berkeley Symp Math Statist Probab. 1961;1:361–380.
[4] Tsukuma H, Kubokawa T. Methods for improved in estimation of a normal mean matrix. J Multivariate Anal. 2007;98:1592–1610.
[5] Eldar YC, Chernoi JS. A pre-test like estimator dominating the least-squares method. J Statist Plann Inference. 2008;138:3069–3085.
[6] Hartung J, Knapp G, Sinha BK. Statistical meta-analysis with applications. Hoboken, NJ: John Wiley & Sons; 2011.
[7] Ahmed SE, Saleh AK, Volodin AI, Volodin IN. Asymptotic expansion of the coverage probability of James-Stein estimators. Theory Probab Appl. 2007;51:683–695.
[8] Ahmed SE, Volodin AI, Volodin IN. High order approximation for the coverage probability by a confident set centered at the positive-part James-Stein estimator. Statist Probab Lett. 2009;79:1823–1828.
[9] Kazimi C, Brownstone D. Bootstrap confidence bands for shrinkage estimators. J Econometrics. 1999;90:99–127.
[10] Lehmann EL, Casella G. Theory of point estimation. 2nd ed. New York: Springer-Verlag; 1998.
[11] Hwang JT, Casella G. Minimax confidence sets for the mean of a multivariate normal distribution. Ann Statist. 1982;10:868–881.