Confidence sets with the asymptotically constant coverage probability centered at the positive part James-Stein estimator

by **Andrei Volodin**

Unitl July 1st, 2009: Department of Mathematics and Statistics
University of Regina, Regina
Saskatchewan, Canada, S4S 4G2.
E-mail: andrei@math.uregina.ca
Web page: http://www.math.uregina.ca/~volodin/

After July 1st, 2009: **School of Mathematics and Statistics University of Western Australia**
35 Stirling Highway
Crawley 6009 WA Australia

Abstract

The asymptotic expansions presented in [1], [2] for the coverage probability by -confidence set centered at the James-Stein estimator, show that this probability depends on the noncentrality parameter (the sum of the squares of the means of normal distributions we are estimating). In this paper we establish how this expansions can be used for the construction of the confidence region with constant confidence level, which is asymptotically (or) equal to . The accuracy of the results obtained is shown by the Monte-Carlo statistical simulations.

Key Words: Confidence sets, positive part James-Stein estimator, multivariate normal distribution, coverage probability, asymptotical expansions, second order asymptotic

1 Introduction

In this paper we address the confidence estimation problem of the mean vector of the -variate normal distribution with independent components and equal unit variances . Let be the sample mean vector that is calculated from samples of equal size of marginal distributions. The confidence set has the given confidence coefficient , if is defined as the quantile of the central chi-square distribution with degrees of freedom according to the formula , where is the chi-square distribution function.

The confidence set possesses the minimax property, but there exist other minimax sets which may provide improved coverage probability for all values of the noncentrality parameter for . In this paper we consider one of these sets

centered at the positive modification of the James-Stein estimator [4]

(we restrict to be greater than 2). Subsequently, many studies followed by various authors. We are mostly interested in the results connected with asymptotic investigations of the coverage probability of the true value of vector , and we refer to the bibliography presented in paper [1].

In [1], 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 depends on the values of the vector via the parameter ; in fact it is a decreasing function of , with , where

where

The second term is represented as a double integral and it is established that for , and for . It is important to note that we have the general formula that deals with both cases of the asymptotic behaviour of for the coverage probability by the confidence set centered at the positive-part James-Stein estimator.

In paper [2], the second order asymptotic expansions of coverage probabilities for and were established. Numerical illustrations for the same values of and as in paper [1] show that for and the third term of the asymptotic has the magnitude of order , but the general picture is such that the third term makes the approximation even worse, especially for small values of . Therefore, the initial approximations (1) might be interpreted from the practical point of view as approximations of the order

Note that James-Stein estimators are applied in the case if it is known that the true values of the means are concentrated close to some mutual value . This is the reason why we call these estimates as shrinkage. Without loss of generality in this paper it is assumed that and . If the concentration point and the variance is different from these values, in all formulae we should substitute the vector by . All properties of the shrinkage estimations with a big amount of examples can be found in Chapter 5 of monograph [5].

In this paper we show how using the approximation (1), it is possible to construct a confidence region with constant, asymptotically equal (or) to the fixed confidence level . The method of constructions owns some ideas from paper [3]. 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 consists of the random function

of the vector parameter and vector statistic with components

The transformation of to a region with asymptotically constant coverage probability is based on the fact that is a strictly increasing function of the variable , if the expression under the root in (1) is positive. This is simple to establish rewriting the function in the form

where .

The region of real values for the function is defined by the inequality

Note that and for , that is, the complex values of the function belong to the regions of small values of and .

Introduce the subregion of the parametric space .

Theorem. *Let . Then for , the coverage probability of the true value of the parameter by subregion the following asymptotic equalities are true if , and if .*

The region is a part of the region and covers the region

that is,

Proof. The first statement of the theorem concerning the confidence region follows directly from the results of [1] (see formula (1)). Really,

The region

corresponds to the maximum value as a function of for each fixed value of (recall that is strictly decreasing function of). The region corresponds the minimal value of for . These two remarks prove the inclusion relations (2).

Relations between the regions (shaded area) and for are represented on Fig.1. The left picture corresponds to , the right picture to . For the given values of , realizations of and as independent normal random variables were obtained and their corresponding regions (shaded area plus the white spot in the middle that corresponds to negative values of the under root expression in) and are calculated. The figures provide projections on the plane of levels of these regions when (left figure) and when (right figure).

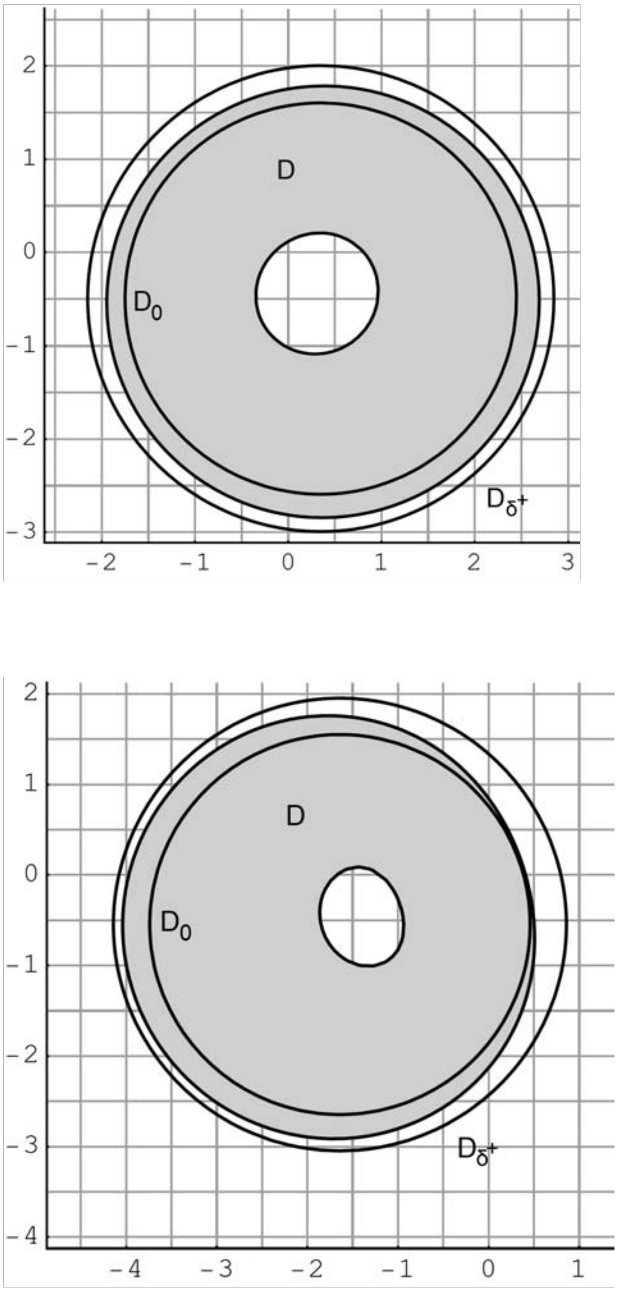


Fig. 1

Therefore, the region corresponds to the maximum gain in the size of confidence regions that were constructed on the principle of constant (al least asymptotically) coverage probability. If the true value , then the confidence coefficient of this region equals exactly to the given level . The region is a ball of the radius . With respect to the region , the radius is smaller in time. Since for the quantile , then with the growth of the number of components of the vector we observe, where are independent standard normal, the confidence region shrinks to a point.

3 Graphical illustrations of the coverage probabilities

Coverage probability by the region of the true value of the parametric vector depends on only through the values of the parametric function . Therefore and the plots of these functions for the values and are presented on Fig.2 (), Fig.3 (), and Fig.5 (). Fig.4 presents similar plots of when and the sample size . For a comparison on Fig.6 we present () the coverage probabilities of by the James-Stein confidence region (thick lines) and their approximation (thing lines). Calculations were conducted by the Monte-Carlo method with replications for each fixed values of and . The values of were chosen identical and equal to because the coverage probabilities depend on the coordinates of vector only through the symmetric functions and . With high probability guaranteed, the accuracy of calculations of the coverage probability values is of the order 0.002.

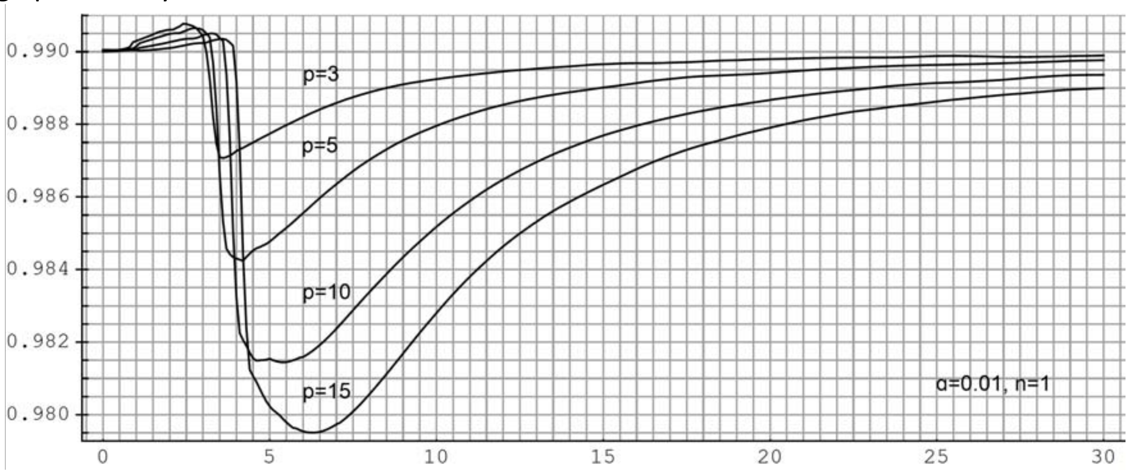
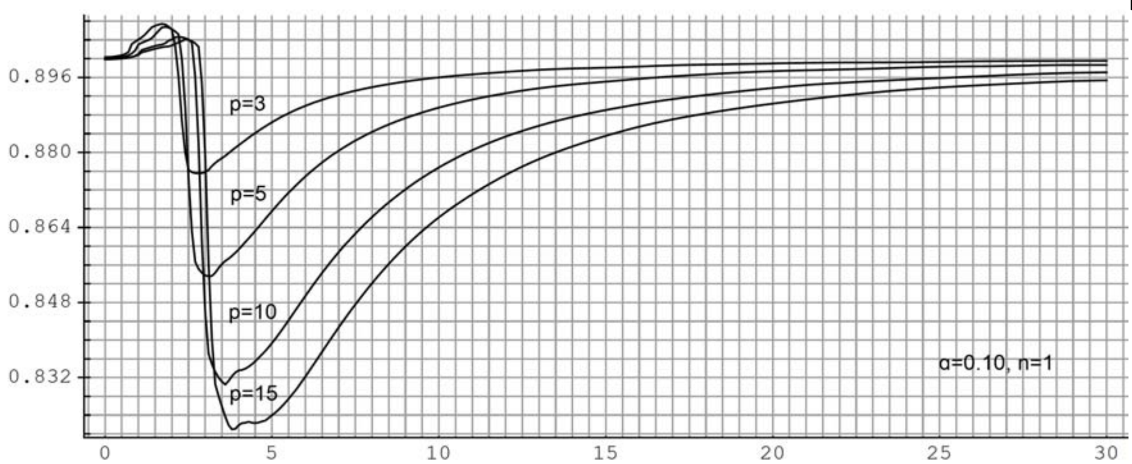
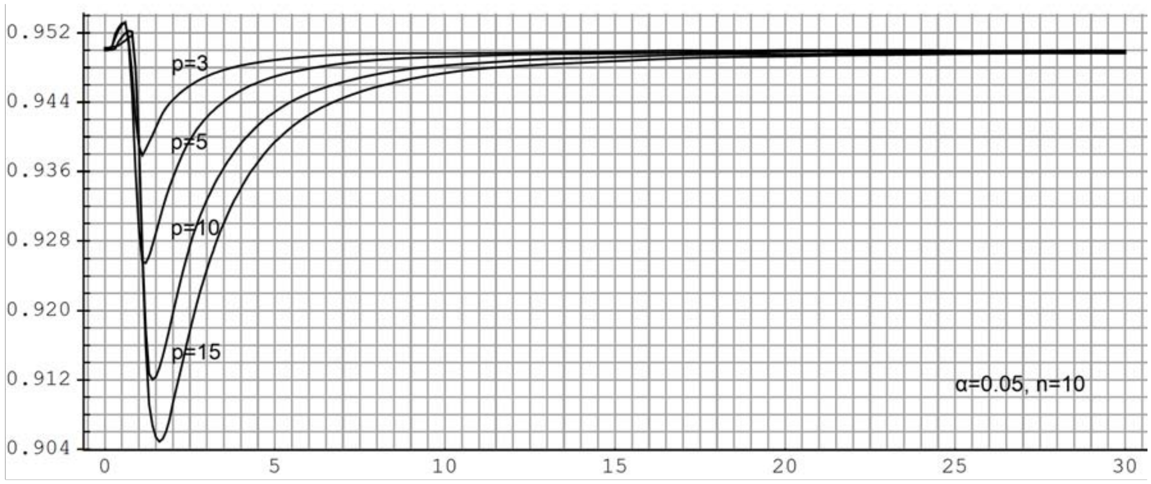


Fig. 2

Fi

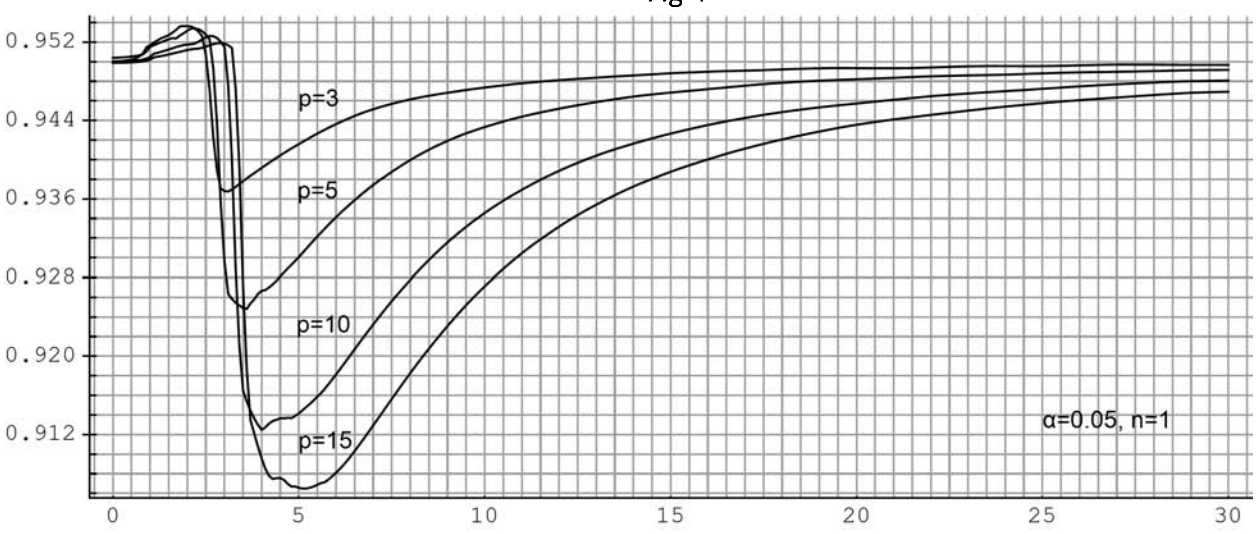


g. 3



0.952
0.944
0.936
0.928
0.920
0.912
0.904
0
5
10
15
20
25
30
p=3
p=5
p=10
p=15
α=0.05, n=10

Fig 4



0.952
0.944
0.936
0.928
0.920
0.912
0
5
10
15
20
25
30
p=3
p=5
p=10
p=15
α=0.05, n=1

Fig. 5

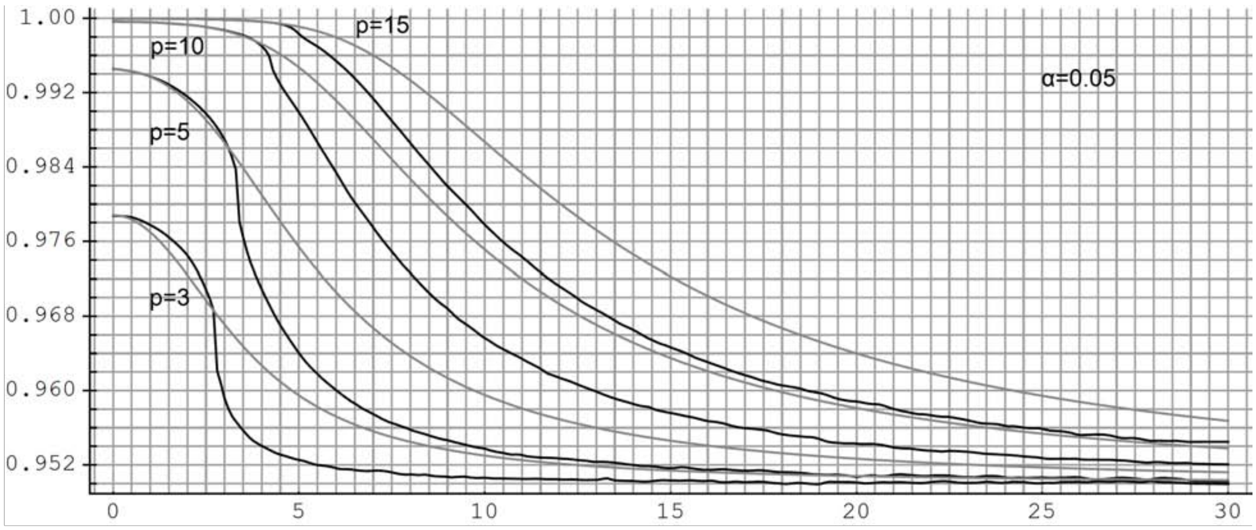


Fig. 6

The graphical illustrations presented provide the following conclusions.

The coverage probability constancy corresponding to the nominal can be guaranteed only in the neighborhood of the shrinkage point . Approximately up to the values , the coverage probability is insignificantly higher than the nominal. ``A catastrophe'' (the sharp slump of coverage probability) happens after these values of , and as plots on Fig.5 and Fig.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.

An increase in the sample size (see Fig.4), shrinks the region of small values of the coverage probability, but their minimum value stays practically the same. On the other hand, the region of constant, corresponding to the nominal tail values of is moved towards smaller values.

The maximum value of the difference strongly depends on the values of and increases with a growth of . Thus, for , the actual coverage probability is close to , while for it reduces to .

Therefore, we could recommend to use the confidence region only in the case when the true values of the parametric vector are close to the shrinkage point, otherwise it is better to use confidence sets centered by the James-Stein estimator or by the sample means.

References

1. S. E. Ahmed, A. K. MD. E. Saleh, A. I. Volodin, I. N. Volodin *Asymptotic expansion of the coverage probability of James-Stein estimators,* Theory Probab. Appl., 51(2007), pp. 683-695.

2. S. E. Ahmed, A. I. Volodin, I. N. Volodin *Asymptotic higher order for coverage probability by confident set centered at the positive-part James-Stein estimator,* Submitted to the Statistics and Probability Letters.

3. J. T. Hwang, G. Casella *Minimax confidence sets for the mean of a multivariate normal distribution,* Ann. Statist., 10(1982), pp. 868-881.

4. W. James, C. Stein *Estimation with quadratic loss,* Proc. Fourth Berkeley Symp. Math. Statist. Probab., 1(1961), pp. 361 -380.

5. E. L. Lehmann, G. Casella *Theory of Point Estimation,* 2nd ed., Springer-Verlag, New York, 1998.

6. C. Stein *Confidence sets for the mean of a multivariate normal distribution,* J. Roy. Statist. Soc., Ser. B, (24)1962, pp. 265-296.