Colloquium Biometricum 38
2008, 67–78

INFLUENCE OF SAMPLE SIZE ON ESTIMATION OF BOOTSTRAP CONFIDENCE INTERVALS FOR MEAN

Andrzej Bochniak

Department of Applied Mathematics and Computer Science
University of Life Sciences in Lublin
Akademicka 13, 20-950 Lublin, Poland
e-mail: andrzej.bochniak@up.lublin.pl

Summary

Bootstrap confidence intervals for mean were examined. Two most popular bootstrap methods were used in computer simulations: percentile and with bias-correction. The bootstrap confidence intervals were compared with standard theoretical intervals in aspects of coverage of true value, length and estimators of intervals bounds. Influence of sample size (from 10 to 100 elements) on considered results was examined. Normal distribution of observations was used in study because it was very known theorems about constructing and correctness of analytical confidence intervals.

Key words and phrases: bootstrap, confidence intervals, mean, normal distribution

Classification AMS 2000: 62F25, 62F40

1. Introduction

In different scientific researches some statistics of examined characteristic is used to compare new methods with classical ones. Sometimes only difference between means is used to compare considered groups. Such comparison informs only about differences between drawn samples from populations. In a case when experimenter wants to compare whole populations, construction of confidence intervals is necessary. Often confidence interval can be calculated analytically on the basis of theoretical equations, but sometimes theory assumptions about considered statistics are not correct.

Nowadays computers can be used to help a scientist in researches. Efron and Tibshirani (1991, 1993) described one of simulation methods called *bootstrap*, which can be used to estimate confidence intervals. It has been successfully used to estimation of different statistics and their confidence intervals since then mainly in such disciplines as medicine, genetics, ecology, agriculture and others (Carpenter and Bithell, 2000; Scheiner and Gurevitch, 1993; Ro and Rannala, 2007; Wolfsegger and Jaki, 2006; Bochniak and Wesołowska-Janczarek, 2005, 2006). The most important advantage of this method is that it does not require the assumption about distribution normality of studied feature, but in some cases it leads to incorrect conclusions (Efron and Tibshirani, 1993; Manichaikul at al., 2006; Schenker 1985; Young 1994).

The aim of this paper is to examine influence of drawn sample size on correctness of bootstrap confidence intervals and estimation of length, lower and upper bounds of interval. It include short description of two mostly used types of bootstrap confidence intervals. The results obtained in simulations by bootstrap methods are later compared with known theoretical results. Due to ability of easy comparison mean value is used in simulations.

2. Construction of bootstrap confidence intervals

Two types of bootstrap confidence intervals were examined: percentile and with bias correction (Efron and Tibshirani, 1991, 1993). They were used to construct confidence intervals for mean in aim to compare results obtained in simulations with known theoretical ones. In some cases the other bootstraps methods are used: bootstrap-t, bootstrap with bias correction and acceleration, non-studentized pivotal, test-inversion, studentized test-inversion.

The percentile bootstrap method is the simplest and most commonly used method to construct bootstrap confidence intervals. The bootstrap technique estimates the precision of a statistic by approximating the unknown sampling distribution in two-step procedure. First, the unknown sampling distribution of values in the population is approximated by a discrete distribution. Then, many bootstrap samples are drawn from this distribution. The unknown sampling distribution is approximated by the distribution of estimates from many bootstrap samples.

In a case when a statistic is applied for two populations there must be two parallel and separate processes of sampling, one for the first population and another one for the second population. Bootstrap samples are drawn from the first and the second discrete distribution respectively to calculate a single estimator for difference between mean values of studied statistics.

There must be at least 1000 bootstrap samples to create confidence interval. Suitable percentiles must be calculated to obtain confidence intervals on required significance level of α in percentile bootstrap method. For example 2.5th and 97.5th percentiles of a bootstrap distribution are used as the bounds of a 0.95 confidence interval. To calculate these percentiles, estimators from 1000 bootstrap replications must be sorted in order from the smallest to the largest. The 2.5th percentile is the average of the 25th and 26th largest values and similarly the 97.5th percentile is the average of 975th and 976th values.
Following symbols are used:

θ – estimated parameter,

$\hat{\theta}$ – estimator of parameter,

$\hat{\theta}_w$ – value of estimator for base sample,

$\hat{\theta}^*$ – estimator obtained by bootstrap method,

$\hat{\theta}^{(\alpha)}$ – percentile of rank α for distribution of estimator values for bootstrap resample.

In general case in bootstrap methods confidence interval has a form of

$$(\hat{\theta}_L, \hat{\theta}_U) = \left(\theta^{*(\alpha_L)}, \theta^{*(\alpha_U)}\right) \tag{2.1}$$

where:

$\hat{\theta}_L$ – estimator of lower bound of interval,

$\hat{\theta}_U$ – estimator of upper bound of interval,

α_L, α_U – ranks for lower and upper percentiles.

In percentile method bounds of estimated confidence intervals at required confidence level 1-α is defined by percentiles of rank:

$$\alpha_L = \tfrac{\alpha}{2} \tag{2.2}$$

$$\alpha_U = 1 - \tfrac{\alpha}{2}$$

Percentile method assumes that median M^* of distribution of estimator values for all bootstrap samples is equal to value $\hat{\theta}_w$ of estimator for original sample. This case is presented on Fig. 1a. where an arrow shows median of bootstrap sample distribution and parameter value for base sample as well.

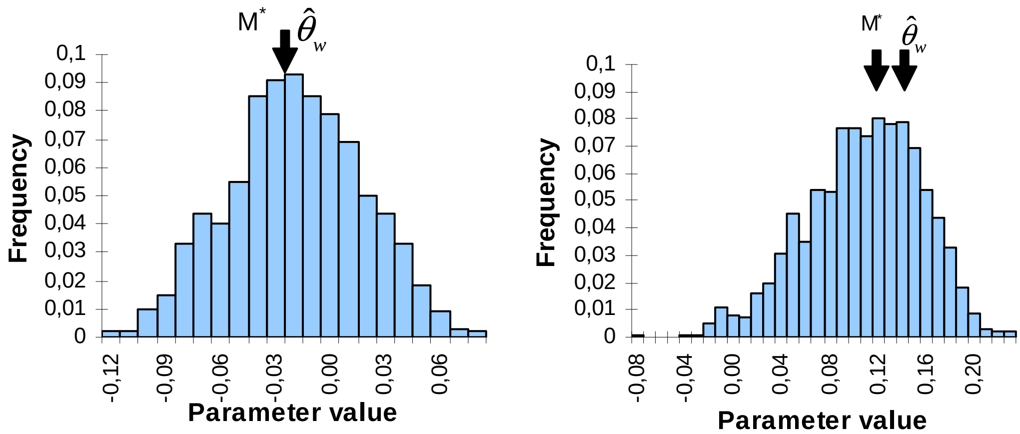

Fig. 1. An example of estimator values distribution for bootstrap samples in which median M^* and parameter value $\hat{\theta}_w$ for original sample are equal (left) and different (right)

The situation described above does not always take place. Sometimes influence of a few disturbing values or distribution of parameter can cause that estimators of parameter for bootstrap samples are error biased. Such situation is presented on Fig. 1b where median of distribution of bootstrap samples parameter is less from parameter value for base sample. This causes that generated percentile confidence interval is moved to less values.

Bootstrap method with bias correction (Efron and Tibshirani 1993) permits to improve precision of percentile method in such cases. This method adds two steps after estimation of distribution of considered parameter for bootstrap samples:

– calculating the fraction of bootstrap samples for which parameter value is less than parameter calculated for original sample,
– calculating new interval bounds to comply bias correction.

Let F be fraction of parameter estimators for bootstrap samples which are less than value of parameter for original sample. Then

$$z_0 = \Phi^{-1}(F), \tag{2.3}$$

where Φ^{-1} signifies the inverse cumulative normal distribution N(0,1). Percentiles estimating confidence interval with bias correction have the following ranks:

$$\alpha_L = \Phi\left(2z_0 + z^{\left(\frac{\alpha}{2}\right)}\right) \tag{2.4}$$

$$\alpha_U = \Phi\left(2z_0 + z^{\left(1-\frac{\alpha}{2}\right)}\right),$$

where:

$z^{(\alpha)}$ - percentile of rank α for standard normal distribution.

3. Computer simulations

Computer simulations were carried out to estimate correctness of bootstrap confidence intervals. The observations were randomly drawn from population with standard normal distribution $N(0,1)$. Intervals of significance level $\alpha=0.05$ were constructed for samples of size 10, 20, 30, 40, 50, 70 and 100. One thousand bootstrap samples were drawn from original sample for construction of single confidence interval (percentile and with bias correction as well) as described in the previous section.

Confidence intervals constructed for mean by bootstrap methods were compared with standard theoretical intervals constructed for samples from normal distribution with unknown mean μ and unknown standard deviation σ. In a case of sample size n and confidence level α confidence interval for mean is well known and given by:

$$\bar{x} - t\left(1-\tfrac{1}{2}\alpha, n-1\right)\frac{s}{\sqrt{n}} < \mu < \bar{x} + t\left(1-\tfrac{1}{2}\alpha, n-1\right)\frac{s}{\sqrt{n}} \tag{3.1}$$

where

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ - mean of the sample,

$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$ - standard deviation of the sample,

$t(\alpha,\nu)$ – fractile of rank α of Student's *t* distribution with ν degrees of freedom.

The following problems were considered for constructed intervals:

– percentage coverage of mean real value equal 0 with comparison of standard analytical intervals coverage; confidence intervals were constructed for 10000 drawn samples of sizes mentioned above,
– comparison of average length for different intervals,
– relative difference of length for pairs of considered confidence intervals,
– distribution of length difference for constructed intervals for the same sample in relationship with analytical values,
– calculating estimators of intervals bounds in dependence of mean and standard deviation of drawn samples,
– determining tendency of relation between bounds and mean as well as standard deviation.

Simulations were made in Microsoft Excel 2003 PL Professional associated with usage of Visual Basic for Applications in Excel for programming necessary functions to random drawing of numbers, constructing intervals and saving collected results which were graphically work out later.

4. Results of computer simulations

The bootstrap confidence intervals are not perfect. As one can see (Fig. 2) coverage of mean true value, which is 0 for standard normal distribution from which observations were drawn, does not have required level. Percentage coverage of intervals should oscillate about 95% for confidence level α=0.05. Standard theoretical intervals (eq. 3.1) satisfy this requirement, but bootstrap intervals have always less coverage although it converges to 95% as $n \to \infty$. It seems that sample size of 30 elements must be used at least to calculate quite correct bootstrap intervals (with coverage aprox. 93.52%).

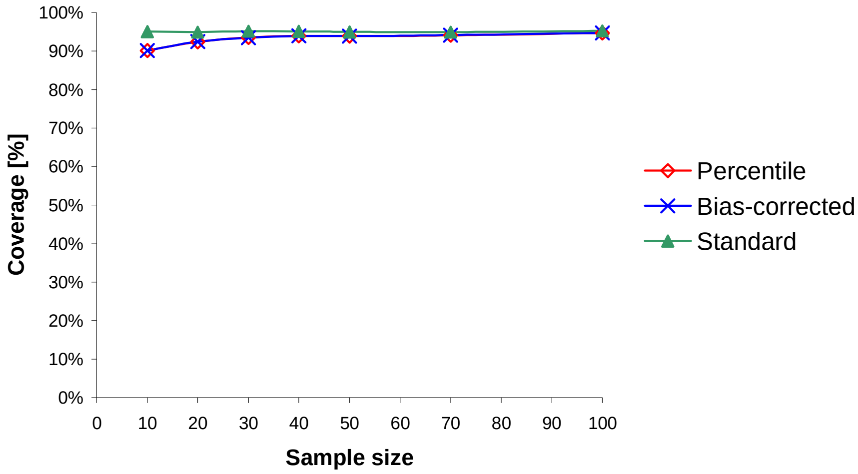

Fig. 2. Percentage coverage of true value by confidence intervals

Table 1. Coverage error for bootstrap confidence intervals

Sample size	Percentile	Bias-corrected	$\frac{1}{n}\cdot 100\%$
10	4.86%	4.89%	10.00%
20	2.63%	2.56%	5.00%
30	1.48%	1.51%	3.33%
40	1.10%	1.07%	2.50%
50	1.14%	1.11%	2.00%
70	0.90%	0.87%	1.43%
100	0.34%	0.34%	1.00%

This fact is consistent with theorems that coverage error for two-sided confidence intervals is of order $O\left(\frac{1}{n}\right)\cdot 100\%$ or sometimes even smaller – under assumption that $\left(\hat{\theta}-\theta\right)/\hat{\sigma}$ is smooth function of sample moments with asymptotic normal distribution (Hall, 1992). The exact values of coverage error obtained in simulations with comparison to theoretical limiting function $\frac{1}{n}\cdot 100\%$ are shown in Table 1.

In the aim to find the reason of that fact following analyses were made in which length of constructed intervals was examined. Bootstrap confidence intervals turn out to be a little shorter with reference to theoretical intervals. It is shown in Table 2. Changes of average length of confidence intervals caused by sample size can be seen there as well.

Table 2. Average length of confidence intervals

	Type of interval		
Sample size	Standard	Percentile	Bias-corrected
10	1.399	1.143	1.148
20	0.923	0.841	0.843
30	0.740	0.696	0.697
50	0.565	0.545	0.545
100	0.396	0.389	0.389

Relative differences between length of confidence intervals for all pairs of considered types are presented in Table 3.

Table 3. Relative differences between length of confidence intervals in %

	Pairs of intervals		
Sample size	Standard – percentile	Standard – bias corrected	Bias corrected – percentile
10	5 to 27%	7 to 28%	-6 to 16%
20	-1 to 19%	-3 to 18%	-7 to 9%
30	-5 to 17%	-5 to 16%	-6 to 7%
50	-8 to 13%	-9 to 14%	-6 to 7%
100	-9 to 12%	-10 to 13%	-6 to 7%

The larger is sample size, the less differences between length of standard and bootstrap intervals are and in general case bootstrap intervals are shorter than theoretical ones (because difference is mainly in positive range). Differences between both bootstrap confidence intervals are slight and almost equally (except sample size 10 – see Table 3) distributed around 0. It informs that bootstrap estimator for sample mean is only little biased and percentile method is sufficient for calculating bootstrap confidence interval for mean. An example of approximated density functions for distribution of length differences between standard and percentile interval for different sample sizes are shown in Fig. 3.

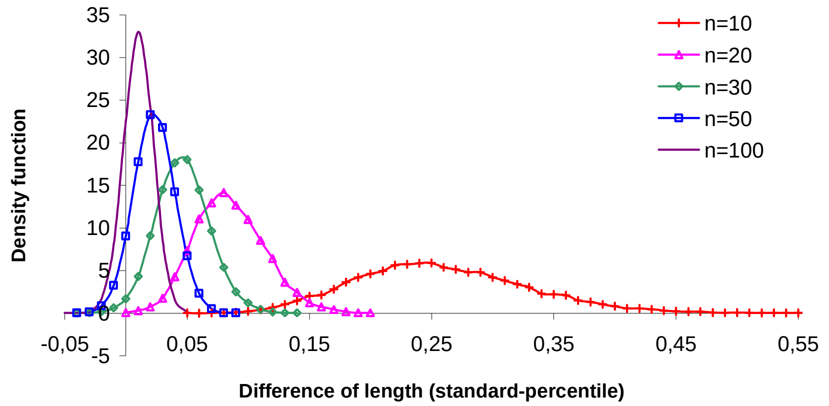

Fig. 3. Approximated density function for distribution of length differences between standard and percentile confidence intervals

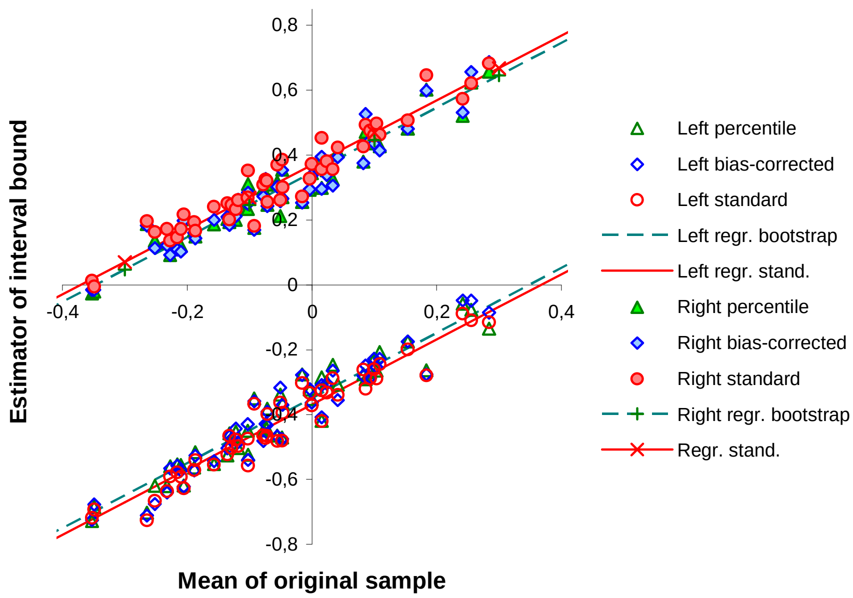

Fig. 4. Dependency of confidence intervals bounds on mean value of original sample

Difference and tendency of lower and upper bound estimators of confidence intervals were examined as well in dependency of mean and standard deviation of original sample from which bootstrap samples were generated. Such analysis for sample size n=30 are shown in Fig. 4 and Fig. 5. Bottom points and lines on each charts relate to estimators for lower bound of intervals and similarly upper points and lines apply to estimators for upper bounds. For readability only values for small fraction of samples are visible here. One can see that theoretical intervals in general include bootstrap ones, because their lower estimators are less and upper ones are larger from matching bounds of bootstrap confidence intervals.

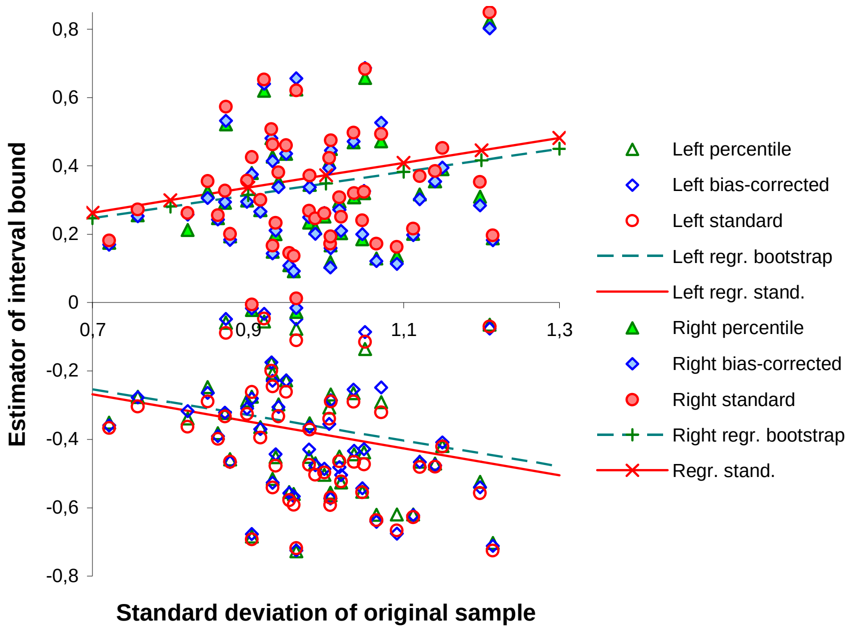

Fig. 5. Dependency of confidence intervals bounds on standard deviation value of base sample

Straight lines visible on both charts show linear regression functions for these dependencies approximated for all 10000 data by least squared method. They inform about parallelism in dependency of bounds for bootstrap and theoretical intervals on mean and standard deviation of original sample. These lines also confirm that standard confidence intervals in general include bootstrap ones because lines for bootstrap methods lie between lines for theoretical ones. Coefficients of linear functions for percentile and bias-corrected methods are approximately equal and functions are almost covering each other.

Straight lines for lower and upper bound estimators are also parallel to each other in the case of dependency on mean value and symmetrical in regards to an axis OX in a case of dependency on standard deviation. In this case estimators are more scattered from approximated straights.

For other sample sizes situation is similar to described above, but ranges for achieved mean and standard deviations for drawn samples are less for larger samples. Regression straight lines for all sizes are parallel for mean and they have less angle of slope to axis OX for larger sizes in dependency on standard deviation.

5. Conclusions

Bootstrap methods are sometimes used to construct confidence intervals for examined statistics. Computer simulations made for mean show that for small sample size ($n<30$) these methods may be not reliable because they do not enough cover true value of estimated parameter for whole population. They have less length in comparison to standard theoretical intervals. On the other side bootstrap methods do not require some assumption about distribution of examined statistics, so they can be applied to constructing confidence intervals in such cases. There are theorems for specific situations and basic statistics, such as mean or standard deviation, when bounds of confidence intervals can be calculated analytically by equations. However sometimes assumptions about normal distribution of examined characteristic is on the edge of acceptance or simply that distribution is unknown and applying known equations may lead to bigger error than bootstrap methods offer. Of course some simulations are required before usage these methods to experimental data in all new cases to see correctness of application of bootstrap methods to specific situation.

Two bootstrap methods: percentile and bias-corrected give almost the same intervals having similar coverage and length. It results from symmetrical distribution of mean. In such cases more complicated in calculating bias-corrected method is not necessary (Efron and Tibshirani 1993). Bootstrap confidence intervals have parallel tendency for dependency on mean and standard deviation of original sample from which bootstrap intervals are generated.

References

Bochniak A., Wesołowska-Janczarek M. (2005). Wykorzystanie metody bootstrapu do badania wpływu pola magnetycznego na własności mechaniczne źdźbeł zbóż. *Inżynieria Rolnicza*. Nr 14 (74), 37-44.

Bochniak A., Wesołowska-Janczarek M. (2006). Porównanie metod bootstrapu zastosowanych do oceny wpływu różnych sposobów odchwaszczania na zdolności konkurencyjne pszenicy. *Colloquium Biometryczne* 36, 171-184.

Carpenter J., Bithell J. (2000). Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in Medicine*, vol. 19, 1141-1164.

Efron B., Tibshirani R.J. (1993). *An introduction to the Bootstrap*. Chapman & Hall. New York.

Efron B., Tibshirani R.J. (1991). Statistical data analysis in the computer age. *Science* 253, 390-395.

Hall P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, New York.

Manichaikul A., Dupuis J., Sen S., Broman K.W. (2006). Poor Performance of Bootstrap Confidence Intervals for the Location of a Quantitative Trait Locus. *Genetics*, Vol. 174, 481-489.

Ro S. and Rannala B. (2007). Inferring Somatic Mutation Rates Using the Stop-Enhanced Green Fluorescent Protein Mouse. *Genetics*, Vol. 177, 9-16.

Scheiner S.M., Gurevitch J. (1993). *Design and Analysis of Ecological Experiments*. Chapman & Hall, New York.

Schenker N. (1985). Qualms about bootstrap confidence intervals. *Journal of American Statistical Association*, vol. 80, 390.

Wolfsegger M.J., Jaki T. (2006). Simultaneous confidence intervals by iteratively adjusted alpha for relative effects in the one-way layout. *Statistics and Computing*, Vol. 16 (1), 15-23.

Young G.A. (1994). Bootstrap – more than a stab in the dark? *Statistical Science*, vol. 9, 382-415.

WPŁYW LICZEBNOŚCI PRÓBY NA OSZACOWANIE BOOTSTRAPOWYCH PRZEDZIAŁÓW UFNOŚCI DLA ŚREDNIEJ

Streszczenie

Metodą symulacyjną badano bootstrapowe przedziały ufności dla średniej z próby. W symulacjach uwzględniono dwie najbardziej popularne metody bootstrapowe: percentylową i z poprawą obciążenia. Otrzymane tymi metodami przedziały porównano ze standardowymi teoretycznymi przedziałami dla średniej. W badaniach uwzględniono następujące aspekty: pokrycie rzeczywistej wartości średniej dla populacji, estymatory lewego i prawego końca przedziału oraz jego długości. W symulacjach sprawdzano wpływ liczebności próby (od 10 do 100 elementów) na otrzymane wyniki. Próby do badań, w celu łatwego porównania z teoretycznymi przedziałami oraz poprawności bootstrapowych przedziałów ufności, były losowane ze standaryzowanego rozkładu normalnego.

Słowa kluczowe: bootstrap, przedziały ufności, średnia, rozkład normalny

Klasyfikacja AMS 2000: 62F25, 62F40