

The bootstrap method in the anisotropy analysis

A V Karelin, S A Voronov and V V Malakhov

National Research Nuclear University MEPhI (Moscow Engineering Physics Institute),
Kashirskoe highway 31, Moscow, 115409, Russia

E-mail: karelin@hotmail.ru

Abstract. In this article we discuss the application of the bootstrap method for the large-scale cosmic rays anisotropy study in the PAMELA experiment with the calorimeter. This method was used to obtain the statistical uncertainties for the dipole phase and amplitude in the case when the collected in the experiment number of events was not sufficient enough for the direct observation of the anisotropy.

1. Introduction

The large-scale anisotropy of the high-energy cosmic rays is seen in various ground-based experiments [1-5]. In the first order it has a dipole form. The first attempt to observe the anisotropy within the same energy range (as in the ground-based measurements) in the satellite was made with a calorimeter in the PAMELA experiment [6]. The collected number of events still was not sufficient enough to observe the anisotropy directly (just to install the upper limit at 10^{-3} level) [7]. Nevertheless the angular resolution (0.3 degree in the average) [8] allowed to make 72 correlated measurements in a right ascension scale with the help of the integration technique [9]. The hypothesis about the existence of the dipole in the experimental data provided the possibility to construct this dipole and obtain the amplitude and phase of it. Here we consider the calculation of the statistical uncertainties for this measurements with the bootstrap technique [10].

2. The bootstrap method

The bootstrap method is described in detail in [11, 12]. It is widely used for the calculation of the statistical errors when their direct evaluation is difficult or not realizable at all. This method is based on the simulation of the large number of data sets directly from the real experimental set of the events. The simulated data sets are used to estimate the statistical fluctuations of the parameters for the investigated phenomenon. The simulation consist of the random selections of the events from experimental data set with returning them back to this experimental set. In other words, a new data set is organized by the random event selection from the experimental data set without taking into account whether some events already have been taken to create this particular new data set or not. Thus, any value (event) from the initial set of the experimental data can be taken once, many times or can not be taken at all. The each simulated data set contains the same number of events as the original experimental one. When the simulation of a large number of the new data sets is done then interested parameters are calculating for each of them. After this it is easy to get distributions of these parameters which actually



are the probability density distribution functions for the parameters of interest. This function determines the statistical uncertainties of the measurable parameters.

It should be emphasized that this method does not provide an information about the most probable value of the parameter of interest, but only allows one to obtain the distribution of the parameter values which is similar to the distribution obtained by the multiple (repeated many times) measurements in the experiment.

The distribution of the relative occurrence frequency of the parameter value obtained by this method provides complete information about the actual distribution without any assumptions or a priori knowledge about the sampling distribution of this parameter and, what is more, takes into account the various effects caused by the experimental equipment.

Justification of this method lies in the fact that the experimental data already themselves contain all available information about the actual distribution. And, therefore, the best way to create a model for the measured parameter is to create a large number of "copies" of the experimental distribution. The bootstrap method is widely used in the gamma-ray astronomy for the identification of point sources [13], moreover it has been used in the statistical analysis for the selection of positrons highly contaminated by the background protons in the PAMELA experiment [14]. Using the calorimeter data for the identification of three groups of cosmic ray particles - electrons, protons, and a mixture of positrons and protons the statistical uncertainties for the energy spectrum of positrons have been obtained via bootstrap method.

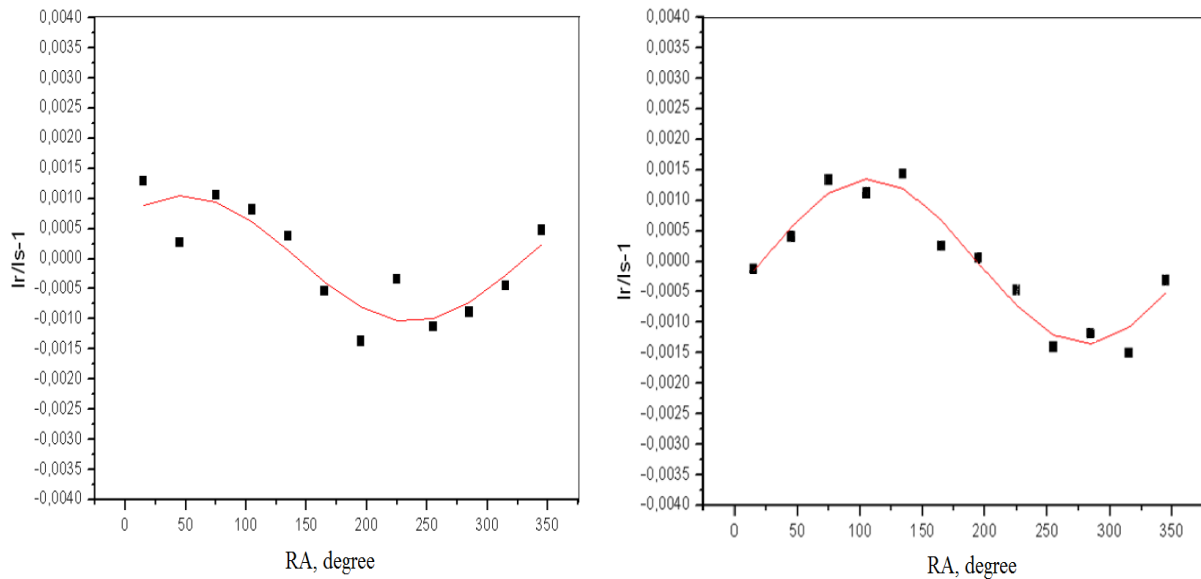


Figure 1. The $Ir/Is-1$ distributions obtained by the bootstrap method. Ir - the real intensity, Is - simulated by shafting method background intensity. RA - right ascension.

3. The experimental data proceeding

The bootstrap method has been used for one more independent assessment of the statistical uncertainties of the amplitude and phase of the dipole anisotropy obtained with the PAMELA calorimeter [9]. The experimental set of events contains about 600 thousand coordinates (right ascension only) measured in the equatorial coordinate system. For the simulation of the new samples a standard random number generator was used. However, as the measured effect of the anisotropy is rather small ($\sim 10^{-3}$), and all random generators are pseudo-random generators to a greater or lesser degree, the following procedure was introduced increasing the "randomness"

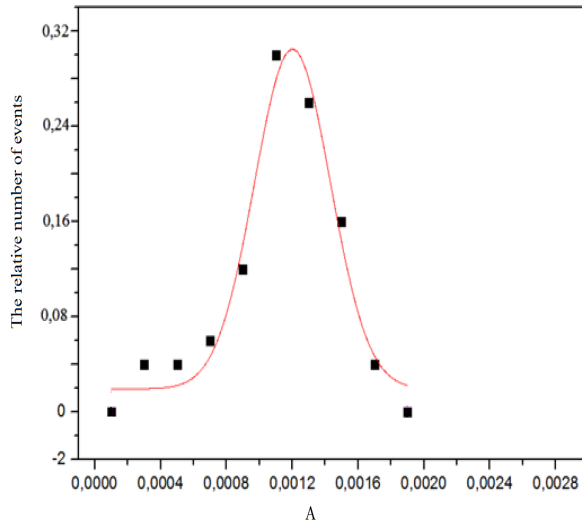


Figure 2. The distribution of the amplitudes A of the dipoles obtained with the bootstrap method.

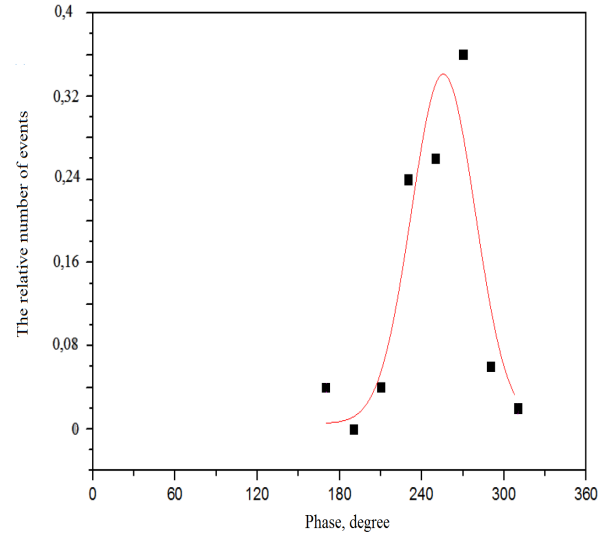


Figure 3. The distribution of the phases of the dipoles obtained with the bootstrap method.

of the results. The entire amount of data was broken down into one hundred equal parts, each of which was in turn divided into 10 parts. At first a number was randomly generated from 1 to 100, then another number from 1 to 10 and then the event was randomly chosen within that smallest area.

Thus it was created 100 new samples of the data. Each sample contained a number of simulated events exactly equal to the experimental value. Each data set produced by the bootstrap method was underwent the same procedure to obtain the dipole anisotropy as it had been done with the real experimental data. The background data sets were obtained with the shuffling technique [15]. The typical examples of the dipoles obtained by the bootstrap method are shown in figure 1. The distributions of the amplitudes A and phases are shown in figures 2 and 3.

4. The results

As a result of the bootstrap method the distributions were obtained for the amplitude (see figure 2) and phase (see figure 3) of the dipole. The distributions were approximated by Gaussian curves. The approximation values for σ were obtained:

$$\sigma_A = 2.310^{-4}, \quad (1)$$

$$\sigma_{phase} = 20^\circ. \quad (2)$$

σ_A and σ_{phase} are the estimation of the statistical uncertainties for the experimentally obtained values of these parameters. These resulting values of statistical errors are close to the values obtained as a result of the approximation of the experimental dipole by the harmonic function [16].

5. Acknowledgments

Authors wishing to acknowledge the financial support of the Russian Science Foundation grant N 14-12-00373.

References

- [1] Amenomori M *et al.* 2005 *Astrophys. J.* **626** 29
- [2] Abdo A A *et al.* 2009 *Astrophys. J.* **698** 2121-30
- [3] Abbasi R *et al.* 2011 *Astrophys. J.* A **740** 16
- [4] BenZvi S 2014 *ASTRA Proc.* **1** 33
- [5] Lidvansky A S *et al.* 2007 *Proceedings of the 30th ICRC* (Mexico: Yucatan Autonoma Univ.)
- [6] Adriani O *et al.* 2014 *Physics Reports* **544** 4 323
- [7] Voronov S A and Karelin A V 2016 *Journal of Physics: Conference Series* **678** 1
- [8] Karelin A V *et al.* 2016 *Proceedings of the ICPPA-2016* (Moscow: NRNU MEPhI)
- [9] Karelin A V *et al.* 2015 *JETP Letters* **101**(5) 295
- [10] Efron B 1979 *Ann. Statistics* **7** 1
- [11] Barbe P and Bertail P 1995 *The Weighted Bootstrap* (New York: Springer)
- [12] Efron B 1982 *The Bootstrap, Jackknife, and Other Resampling Plans* (Philadelphia: SIAM)
- [13] Simpson G and Mayer-Hasselwander H 1986 *Astron. Astrophys.* **162** 340
- [14] Adriani O *et al.* 2010 *Astroparticle Physics* **34** 1
- [15] Karelin A V *et al.* 2015 *ASTRA Proc.* **2** 35
- [16] Karelin A V *et al.* 2016 *Journal of Physics: Conference Series* **675** 3