ARTICLE IN PRESS

Advanced Powder Technology xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

Advanced Powder Technology

journal homepage: www.elsevier.com/locate/apt

Original Research Paper

An application of bootstrap method for analysis of particle size distribution

Tatsushi Matsuyama

Faculty of Science and Engineering, Soka University, 1-236 Tangi-cho, Hachioji, Tokyo 192-8577, Japan

ARTICLE INFO

Article history:
Received 4 November 2017
Received in revised form 23 February 2018
Accepted 1 March 2018
Available online xxxx

Keywords:
Particle diameter
Particle size
Particle size distribution
Confidential interval
Evaluation of uncertainty

ABSTRACT

The Bootstrap method, which is used widely in statistics, is a very powerful method that can be applied to the analysis of particle size distribution. For a number based measurement, this method can estimate the statistic uncertainty or confidence interval for any statistical quantities of interest from the distribution, with very simple protocol and without any parametric assumptions. This paper gives a demonstration to introduce the method to the community of particle size analysis.

1. Introduction

In particle size analysis (PSD analysis), a common question is how many particles should be counted for a precise analysis. A straightforward answer is that it depends on the required precision. Also nowadays the question can be rephrased to a requirement to evaluate the uncertainty (Type A evaluation of measurement uncertainty) due to finite sample size, or limited number of particles, resulting in a statistical fluctuation.

A sophisticated mathematic theory to answer the question was given by Masuda and Gotoh [1], but it was limited to the case to evaluate confidence interval of median size for lognormal distribution.

Bootstrap method is a statistic method to evaluate or to estimate such uncertainty or confidence interval of statistical quantities, which was introduced by Efron [2] at first and now is widely accepted and used [3]. This method is simple, powerful and absolutely suitable to use in the evaluation of the particle size distribution analysis with the following features:

(1) Nonparametric: there is no necessity to assume any mathematical expression or model to represent a particle size distribution.
(2) Flexibility: not only the mean value or median of a distribution, but also any kind of statistical quantities of interest, e.g., any percentile or standard deviation can be evaluated for confidence interval.
(3) Simplicity: because the method is based on resampling, the algorithm is very simple, and no complicated mathematical calculation is required. The analysis is performed only with a single measurement set.

Despite these useful features, the bootstrap method is not yet widely used in the community of PSD analysis, where few published applications have been found [4]. Therefore, this paper aims to demonstrate the use of the bootstrap method in PSD analysis in order to advocate its use.

2. Algorithm of the bootstrap method

The bootstrap method is based on resampling protocol, with replacement. Fig. 1 shows an illustration of its basic scheme. As the algorithm shows, in the bootstrap method the statistical variation due to limited sampling from population is estimated by an approximation with resampling from an originally measured sample, instead of whole population. Although a couple of enhancements or variations have been developed, a simple version of the algorithm of the bootstrap method is as follows:

(1) Given a set of size measurement of N particles, $\boldsymbol{x} = \{x_1, x_2, x_3, \ldots, x_N\}$

E-mail address: tatsushi@t.soka.ac.jp

https://doi.org/10.1016/j.apt.2018.03.002
0921-8831/

Please cite this article in press as: T. Matsuyama, An application of bootstrap method for analysis of particle size distribution, Advanced Powder Technology (2018), https://doi.org/10.1016/j.apt.2018.03.002

Nomenclature

a	upper bound of confidence interval
b	lower bound of confidence interval
B	number of bootstrap iterations
$F(x)$	lognormal distribution of cumulative form
$F_n(x)$	normal distribution of cumulative form
N	sample size (number of particles)
sd	sample standard deviation
x	particle size
$\boldsymbol{x}$	a measured sample as a set of particle size, $\boldsymbol{x} = \{x_1, x_2, x_3, \ldots, x_N\}$
x_{50}	median size
x_{50}^B	median size by bootstrap estimation
x_{50}^P	median size of population
x_{90}	90th percentile size
x_{90}^P	90th percentile size of population
$\boldsymbol{x}^B$	bootstrap sample generated by resampling from the measured sample, $\boldsymbol{x}$, with replacement
x_g	geometric average of population
α	parameter giving $100(1-\alpha)\%$ confidence level; $\alpha = 0.025$ for 95% confidence level.
δ	half width of confidence interval (of 95% in this paper), $\delta = (\delta_+ + \delta_-)/2$
δ_+	upper width of confidence interval,
δ_-	lower width of confidence interval,
θ	statistical quantities of interest
$\boldsymbol{\theta}^*$	a set of θ^*, generated by repeated bootstrap sampling
$\bar{\theta}^*$	arithmetic average of $\boldsymbol{\theta}^*$
$\theta^*_{\alpha/2}$	$100(\alpha/2)\%$ percentile value of $\boldsymbol{\theta}^*$
μ	$\mu = log x_g$
σ	$\sigma = log \sigma_g$
σ_g	geometric standard deviation

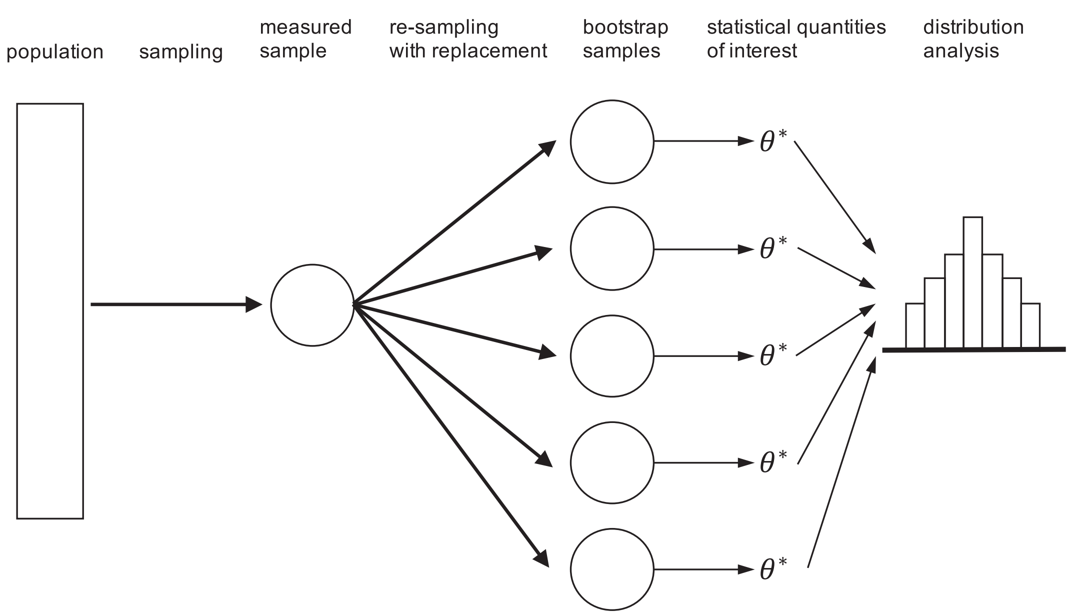

Fig. 1. Schematic illustration of the procedure of bootstrap method.

(2) Resample, with replacement, randomly from the originally measured sample $\boldsymbol{x}$ to produce bootstrap sample, $\boldsymbol{x}^B = \{x_1^*, x_2^*, x_3^*, \ldots, x_N^*\}$, of sample size N, which is the same number as the originally measured sample. Note that a same datum is allowed to be drawn multiple times.
(3) Calculate any statistical quantities of interest (mean value, median of a distribution, any percentile, standard deviation and so on), θ^*, from the bootstrap sample $\boldsymbol{x}^B$.
(4) Repeat (2)–(3) for B times to produce a set of $\boldsymbol{\theta}^* = \{\theta_1^*, \theta_2^*, \theta_3^*, \ldots, \theta_B^*\}$.
(5) The arithmetic average $\bar{\theta}^* = \frac{\sum_{b=1}^{B} \theta_b^*}{B}$ gives the most likely estimation for θ of the population.
(6) The sample standard deviation of $\boldsymbol{\theta}^*$, as $Sd = \sqrt{\frac{\sum_{b=1}^{B} (\theta_b - \bar{\theta})^2}{B-1}}$, is the uncertainty of the measurement.
(7) To determine confidence interval, (a, b), 'percentile method' is adopted in this paper as one of the simplest methods, among a couple of methods proposed [3]. With the method, the upper and lower bounds of the confidential interval for $100(1-\alpha)\%$ confidence level are given as,

$$\begin{cases} a = \theta^*_{\alpha/2} \\ b = \theta^*_{1-\alpha/2} \end{cases} \tag{1}$$

where $\theta^*_{\alpha/2}$ and $\theta^*_{1-\alpha/2}$ are $100(\alpha/2)\%$ and $100(1-\alpha/2)\%$ percentile value of $\boldsymbol{\theta}^*$, respectively. For 95% confidence level, $\alpha = 0.05$, for instance. Also from these the lower and upper widths of the confidence interval are given as,

$$\begin{cases} \delta_- = \bar{\theta}^* - \theta^*_{\alpha/2} \\ \delta_+ = \theta^*_{1-\alpha/2} - \bar{\theta}^* \end{cases} \tag{2}$$

Please cite this article in press as: T. Matsuyama, An application of bootstrap method for analysis of particle size distribution, Advanced Powder Technology (2018), https://doi.org/10.1016/j.apt.2018.03.002