

# Бутстреп-моделирование при построении доверительных интервалов процентилей: применение к контролю качества высокотехнологичной продукции

И. В. Гадолина\*, Н. Г. Лисаченко†

\* Федеральное государственное бюджетное учреждение науки,  
Институт машиноведения им. А.А. Благонравова,  
ул. Бардина, д.4, Москва, Россия, 119334

† Государственный научный центр РФ,  
ОИПП «Технология» им. А.Г.Ромашина,  
Киевское шоссе, д.15, Обнинск, Калужская область, Россия, 249031

**Аннотация.** В качестве альтернативы существующим подходам предложен метод построения базисов (нижних границ доверительных интервалов первой и десятой процентилей распределений) с использованием бутстреп-моделирования. Доверительные интервалы процентилей являются важным интервальным показателем, применяющимся при сертификации высокотехнологичной продукции. Применение метода позволяет построить робастные оценки свободные от недостатков как параметрических, так и непараметрических подходов. На примерах реальных выборок прочностных показателей полимерных композиционных материалов приведено сопоставление значений В-базисов, определенных по применяемому в настоящее время методу и по вновь предложенному.

**Ключевые слова:** процентили, доверительные интервалы, прочность, бутстреп-моделирование.

## 1. Введение

Как говорится в ставшей уже классической книге [1] недостаточно только точечных оценок. Величина оценки меняется от испытания к испытанию и поэтому не может служить устойчивой характеристикой качества изделия. В большой мере ширина доверительных интервалов также зависит от числа испытаний. В связи с этим указывается на целесообразность использования доверительных интервалов для определения границ истинного значения с некоторой вероятностью.

Так, например, при сертификации высокотехнологичной продукции на ряде предприятий необходимо на основании проведенного эксперимента оценить доверительные интервалы некоторых процентилей, в частности, определить так называемые А- и В-базисы, что соответствует нижней границе 95% доверительного интервала 1% и 10% процентили соответственно.

Процентили  $\gamma\%$  – это характеристики выборки, которые выражают ранги элементов массива в виде чисел от 1 до 100,  $\gamma\% \in (0, 100)$ , и являются показателем того, какой процент значений находится ниже

определенного уровня. В какой-то степени сами процентиля уже являются непараметрической интервальной оценкой. Тем не менее, процентиля – это случайная величина, которая определяется по выборке, в связи с чем необходима оценка ее вариабельности. Для вычисления доверительных интервалов значений процентилей, полученных по случайной выборке и используемых для построения А и В базисов, в настоящее время требуется сначала подобрать подходящий вид распределения, для каждого из которых разработаны весьма сложные зависимости [2].

## 2. Основная часть

В настоящей работе в качестве альтернативы методам [2] предлагается метод построения базисов с использованием статистического бутстрепа Бредли Эфрона [3]. Ранее аналогичный подход был рассмотрен в [4], где была исследована точность бутстреп-оценки доверительных интервалов квантилей в зависимости от распределения случайных величин.

Бутстреп применяется для оценки изменчивости различных показателей. Метод оценки базисов основан на многократном моделировании так называемых бутстреп-выборок на базе исходной выборки и опирается на интенсивное использование ЭВМ. Число бутстреп-выборок (обозначается  $R$ ) должно быть велико: в настоящем исследовании применялось  $R = 1000$ . Элементы бутстреп-выборок образуются из элементов исходной, при этом осуществляется случайный выбор с возвращением, а размер каждой бутстреп-выборки равен размеру исходной. Для таких статистик, как  $\gamma\%$  процентиля, математические выражения для дисперсии сложны и их оптимальность строго не доказана. В связи с этим представляет интерес сопоставление интервальных оценок бутстреп  $\gamma\%$  процентиля с применяемыми до настоящего времени методами с целью рассмотрения возможности их внедрения в практику инженерных расчетов.

Исследование проводилось на примере построения доверительных интервалов для реальной задачи оценки вариабельности случайной величины  $\sigma_B$  - предела прочности образца при растяжении в МПа. Выборка включает 30 случайных величин. Среднее значение ... среднеквадратичное отклонение ... Число бутстреп-выборок выбрано  $R = 1000$ . Для каждой  $k = 1, 2, \dots, R$  бутстреп-выборки оценивается 10% процентиля ( $Q_{10}$ ). Таким образом образуется выборка размером  $R$ .

Теорема:

Если представить в виде вариационного ряда  $R$  значений случайной величины процентиля  $Q_{10}$ , то значения членов вариационного ряда с индексами  $LOW$  и  $UP$  ограничат  $\alpha\%$  доверительный интервал для случайной величины  $Q_{10}$ . Значения индексов вариационного ряда вычисляются по формулам:

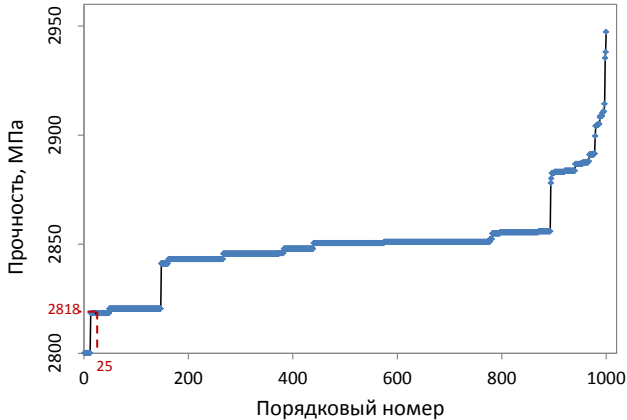


Рис. 1. Вариационный ряд значений 10% процентилей с указанием нижней 95% доверительной границы

$$LOW = \text{целая часть} \left[ \frac{1-\alpha}{2} R \right]$$

$$UP = \text{целая часть} \left[ \frac{1+\alpha}{2} R \right]$$

Следствие:

Для нижней границы 95% доверительного интервала 10% процентиля Q10 при индексе  $LOW = 25$  значение В-базиса составляет 2518 МПа.

Рисунок 1 поясняет данное рассуждение. Граница бутстреп показана штриховой линиями.

Преимуществом оценки с применением бутстреп-моделирования является их большая интуитивная понятность, а также свойство робастности.

Основные статистические характеристики и значения В-базисов для некоторых выборок случайных значений прочности на растяжение  $\sigma_B$ , МПа для пяти поставок продукции приведены в табл. 1. Показаны В-базисы, построенные по предлагаемому методу и по формулам [2]. Незначительные расхождения значений, вероятно, объясняются несоответствием распределений случайных величин нормальному закону, что предполагается в методе построения по формулам [2]. В этом смысле метод, основанный на моделировании бутстреп, возможно, является более надежным.

Таблица 1

Данные по случайной величине предела прочности при растяжении  $\sigma_B$ , МПа, по поставкам препрега HexPly

Номер поставки	Среднее значение	Станд. откл.	10% процентиль	В-базис бутстреп	В-базис [2]
7	3010	140	2820	2790	2690
8	2960	80	2850	2820	2780
9	2720	150	2510	2580	2400
10	2850	170	2630	2580	2460
11	2730	130	2580	2500	2410

### 3. Заключение

Разработан метод для построения доверительных интервалов процентилей случайной выборки в непараметрической постановке с использованием моделирования бутстреп. На примере выборок по значениям прочности некоторых поставок продукции показано, что значения нижней границы 95% доверительного интервала 10% процентили прочности (В-базис), полученные по формулам [2] и при моделировании бутстреп близки (расхождения составляют от 1% до 7%). В дальнейшем предполагается провести апробирование разработанного метода на примере других физических величин. Из-за малого размера выборок ( $n = 30$ ) по рассмотренным методам не удалось построить А-базисы. Для этой цели, возможно, понадобится разработка методов на основе математического моделирования.

### Литература

1. *Беляев Ю. К., Гнеденко Б. Г., Соловьев А. Д.* Математические методы в теории надежности. — М.: Наука, 1965.
2. *Ллойд Э., Ледерман У.* Справочник по прикладной статистике. — М.: Финансы и статистика, 1989.
3. *Диаконис П., Эфрон Б.* Статистические методы с интенсивным использованием ЭВМ. // В мире науки, 1983, No. 7. — С. 60–73.
4. *Yvonne H. S, Lee S.* Iterated smoothed bootstrap confidence intervals for population quantiles. // The Annals of Statistics, 2005, Vol. 33, No. 1. — P. 437–462.

UDC 519.2

## **Bootstrap-modelling for building confidence intervals for percentiles: high-tech production quality control application**

**I. V. Gadolina\***, **N. G. Lisachenko†**

*\* Federal budget - funded research*

*Mechanical Engineering Research Institute of the Russian Academy of Sciences  
Ul. Bardina 4, Moscow, 119334, Russia*

*† State Research Center of Russian Federation,  
ORPE «Technologiya» named after A.G.Romashin,  
Kievskoye Shosse 15, Obninsk, Kaluga Region, 249031, Russia*

As an alternative to currently used approaches the method of construction of basises (condence intervals for percentiles) using bootstrap simulation is proposed. Confidence intervals for percentiles are an important interval indicator, used for certification of high-tech production. The application of this method will allow building robust estimation free from the disadvantages of parametric and non-parametric approaches. The comparison of B-basises, defined by the used up to now and newly proposed methods examples of real samples of the strength parameters of the materials.

**Keywords:** percentiles, confidence intervals, robustness, bootstrap simulation.