

DOI: 10.26896/1028-6861-2017-83-11-73-77

УДК (UDC) 519.2

РАЗРАБОТКА МЕТОДА ПОСТРОЕНИЯ ДОВЕРИТЕЛЬНЫХ ИНТЕРВАЛОВ ДЛЯ ПРОЦЕНТИЛЕЙ СЛУЧАЙНОЙ ВЫБОРКИ ПРОЧНОСТИ КОМПОЗИТОВ С ПРИМЕНЕНИЕМ БУТСТРЕП-МОДЕЛИРОВАНИЯ

© И. В. Гадолина¹, Н. Г. Лисаченко²

¹ Институт машиноведения им. А. А. Благонравова РАН, Москва, Россия; e-mail: gadolina@mail.ru

² АО «ОНПП «Технология» им. А. Г. Ромашина» Госкорпорации «Ростех», Москва, Россия; e-mail: lisachenkong@technologiya.ru

Статья поступила 14 апреля 2017 г.

Предложен метод построения базисов (доверительных интервалов для процентиля) с использованием бутстреп-моделирования в качестве альтернативы применяемым в настоящее время подходам. Бутстреп-моделирование — это метод численного моделирования распределений на основе многократного воспроизведения данных без привлечения информации о законах распределения. Поскольку характеристики прочности являются случайными величинами, необходимо статистическое оценивание с построением интервальных данных, чему посвящена настоящая работа. В качестве иллюстрации и пояснения метода показан пример распределения среднего для выборки значений прочности на сдвиг композитов с применением бутстреп-моделирования. Для построения доверительных интервалов для процентиля распределений значений прочностных характеристик в настоящее время рекомендуется отнести распределение к одному из известных законов (нормальному, логнормальному или Вейбулла), а непараметрический подход обычно дает консервативные (заниженные) результаты, а потому нежелателен, что объясняет необходимость разработки альтернативного метода. Проведено сопоставление *B*-базисов, определенных по вновь предложенному и по применяемому до настоящего времени методам, на примерах реальных выборок прочностных показателей композитных материалов. Рассмотрены конкретные примеры показателей прочности по сдвигу и растяжению образцов, изготовленных из препрега HexPly методом автоклавного формования.

Ключевые слова: процентиля; доверительные интервалы; композиты; прочность; бутстреп-моделирование.

A METHOD OF CONSTRUCTING CONFIDENCE INTERVALS FOR PERCENTILES OF COMPOSITE STRENGTH RANDOM VARIABLE USING BOOTSTRAP SIMULATION

© I. V. Gadolina and N. G. Lisachenko

Submitted April 14, 2017.

A method is proposed for constructing basic sets (confidence intervals for percentiles) using bootstrap simulation as an alternative to currently used approaches. Bootstrap simulation is a method for numerical modeling of distributions based on multiple data reproduction the without using any information regarding the distribution laws. Since the strength characteristics are random variables, statistical estimation with construction of the interval characteristics is required. This is the goal of the study. An illustrative example of constructing the confidence intervals for mean strength value using bootstrap-modeling is considered. To construct the confidence intervals for percentiles of the distributions of the strength characteristics we recommend to assign the distribution to one of the currently known (normal, lognormal distribution or Weibull) laws, unlike the existing non-parametric approach that generally gives a conservative (too low) and thus undesirable results, which is the reason for developing a new approach. A comparison of *B*-bases, determined by the newly proposed and traditional method is carried out on real samplings of the strength characteristics of composite materials. The specific examples of strength parameters for shear and tensile strength of the specimens made of prepreg HexPly (composite materials, semiproducts) using an autoclave molding method are presented.

Keywords: percentiles; confidence intervals; composites; durability; bootstrap simulation.

Изучение прочностных и упругих свойств композитов вызывает научный и практический интерес [1]. Исследовали образцы, изготовленные методом автоклавного формования из препрега марки HexPly на основе уг-

леродной ленты. Для сертификации продукции определяли прочностные характеристики: 1) предел прочности при растяжении (σ_b , МПа); 2) модуль упругости при растяжении (E , ГПа); 3) прочность межслойного

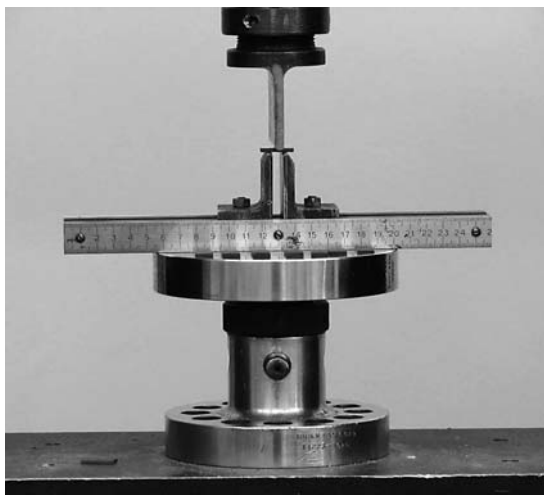


Рис. 1. Приспособление с образцом препрега для испытания на межслойный сдвиг методом трехточечного изгиба

сдвига при нормальной температуре (τ_{20} , МПа); 4) прочность межслойного сдвига при повышенной температуре (τ_{120} , МПа). Приспособление с установленным образцом для испытаний на межслойный сдвиг показано на рис. 1.

При сертификации высокотехнологичной продукции, к которой относятся композиты из препрега марки HexPly, необходимо на основании проведенного эксперимента оценить доверительные интервалы для ряда перцентилей, в частности, определить так называемые A - и B -базисы [2], являющиеся нижними границами 95 %-го доверительного интервала, — 1 %-й и 10 %-й перцентилем соответственно. Перцентили γ (%) — это характеристики выборки, которые выражают ранги элементов массива в виде чисел от 1 до 100 — γ (%) \in (0, 100) и являются показателем того, какой процент значений находится ниже определенного уровня. В более общем случае используют квантили $q \in$ (0, 1), $q = \gamma/100$. Математически квантиль определяют следующим образом. Пусть имеются независимые и одинаково распределенные случайные величины $X = \{x_1, x_2, \dots, x_n\}$, для которых существует функция распределения F с плотностью $f = F'$. Определим q -й квантиль популяции, такой, что $F^{-1}(q) \in \inf\{x \in R: F(x) \geq q\}$. Значение квантили $q = 0,1$ (или, что то же самое, 10-й перцентилем) указывает на то, что 10 % значений располагается ниже этого уровня. Квантиль (перцентиль) — это случайная величина, которая определяется по выборке, в связи с чем необходима оценка ее вариабельности. Для вычисления доверительных интервалов для значений перцентилей, полученных по случайной выборке, используемых для построения базисов, требуется сначала подобрать подходящие виды распределений, для каждого из которых разработаны весьма сложные зависимости. Так, например, для расчета базисов в предпо-

ложении нормального распределения случайной величины в [2] предлагаются формулы

$$B = \bar{x} - k_B s,$$

$$A = \bar{x} - k_A s, \quad (1)$$

где \bar{x} — среднее; s — среднеквадратическое отклонение, k_B и k_A — соответствующие объему выборки коэффициенты допуска. Значения этих коэффициентов приведены в таблицах или могут быть вычислены с погрешностью не более 0,2 % по следующим формулам:

$$k_B = 1,282 + \exp\left(0,958 - 0,520 \ln(n) + \frac{3,19}{n}\right),$$

$$k_A = 2,326 + \exp\left(1,340 - 0,522 \ln(n) + \frac{3,87}{n}\right), \quad (2)$$

где n — размер выборки.

Исследовали распределение прочностных характеристик образцов, изготовленных методом автоклавного формования из препрега марки HexPly на основе углеродной ленты, отобранного от нескольких поставок продукции. Каждая поставка включает в себя обычно от пяти до семи партий препрега. Проведен анализ стабильности прочностных характеристик [3] и сделан вывод о влиянии номера поставки на эти характеристики, т.е. наблюдаемая нестабильность не позволяет объединять большое количество данных в одну выборку. В связи с этим в данной работе были проанализированы статистические характеристики для нескольких поставок отдельно. Исследование причин нестабильности продукции выходит за рамки рассматриваемой темы и нами не проводилось.

Предлагаем в качестве альтернативы методам [2] использовать метод построения базисов с применением статистического бутстрепа.

Остановимся кратко на описании метода статистического бутстрепа. Его предложил в 1977 г. математик Бредли Эфрон [4]. Статистический бутстреп — это способ получения робастных оценок среднеквадратических ошибок и доверительных интервалов, применяется для оценки изменчивости различных показателей. Метод основан на многократном моделировании так называемых бутстреп-выборок на основе исходной выборки и на использовании ЭВМ. Число бутстреп-выборок (R) должно быть велико: в настоящем исследовании $R = 100 - 1000$. При быстрой работе компьютеров на вычисления затрачиваются доли секунды. Размер каждой бутстреп-выборки соответствует размеру исходной (n), а элементы бутстреп-выборок образуются из элементов исходной, при этом осуществляется случайный выбор с возвращением. Для статистик, для которых существуют точные математические выражения для анализа изменчивости, на ряде примеров показано удовлетворительное соответствие

бутстреп- и классических оценок. К настоящему времени уже накоплен значительный опыт применения статистического бутстрепа к инженерным задачам (см., например, [5]). С другой стороны, математики предупреждают о чрезмерном увлечении этим методом [6]: «...там, где эконометрическая теория хорошо развита, где найдены методы анализа данных в том или ином смысле близкие оптимальным, бутстрепу делать нечего».

В нашем случае для таких статистик, как, например, процентиль, математические выражения для дисперсии сложны и их оптимальность строго не доказана. В связи с этим представляет интерес сопоставление интервальных оценок бутстреп процентиля с применяемыми до настоящего времени методами в целях рассмотрения возможности первых внедрения в практику инженерных расчетов. Вопросам оценки доверительных интервалов для квантилей с помощью бутстреп-моделирования посвящены работы [7, 8]. В [7] предложен метод сглаживания доверительных интервалов для квантилей с помощью, в частности, ядерной оценки. В [8] исследуется точность бутстреп-оценки доверительных интервалов квантилей в зависимости от распределения случайных величин.

Для пояснения процедуры построения доверительных интервалов с применением бутстреп-моделирования вначале рассмотрим пример их построения для среднего значения прочности на сдвиг τ_{20} , МПа. Этот пример приведен в связи с тем, что теоретически распределение среднего хорошо изучено и для выборки, данные которой не противоречат нормальному закону [3], получены теоретические доверительные интервалы по критерию Стьюдента [9]. Таким образом, мы можем сопоставить теоретические доверительные интервалы с теми, что нам дает бутстреп. Выборка (исходная) размером $n = 30$ элементов приведена в табл. 1 (данные не противоречат гипотезе о нормальном распределении [3]).

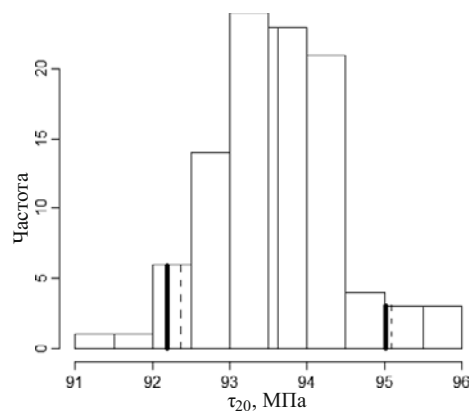


Рис. 2. Распределение средних по 100 бутстреп-выборкам на основе исходной: пунктирные линии — 90 %-ные бутстреп доверительные границы среднего; жирные линии — то же по критерию Стьюдента [9]

Характеристики исходной выборки: математическое ожидание (τ_{20}) = 93,61 МПа; среднеквадратическое отклонение (τ_{20}) = 4,502 МПа.

По правилам формируется R бутстреп-выборок. В качестве примера в табл. 2 показана смоделированная k -я бутстреп-выборка, построенная на основе исходной.

Видно, что некоторые значения повторяются, например, номерам 1 и 11 соответствуют $\tau_{20} = 107,8$ МПа. Некоторые элементы из исходной выборки не вошли ни разу в k -ю выборку (например, элемент с номером 1: $\tau_{20} = 86,1$ МПа), но они, возможно, войдут в бутстреп-выборку с номером $k + 1$.

Характеристики k -й бутстреп-выборки: математическое ожидание (τ_{20}) = 93,99 МПа; среднеквадратическое отклонение (τ_{20}) = 5,137 МПа. Видно, что эти характеристики для исходной и k -й выборок отличаются, хотя и незначительно. Для каждой $k = 1, 2, \dots, R$ бутстреп-выборки можно определить свои характеристики, которые отражают изменчивость точечных оценок [4, 5].

Таблица 1. Прочность на сдвиг при нормальной температуре (исходная выборка)

Номер по порядку	1	2	3	4	5	6	7	8	9	10
τ_{20} , МПа	86,1	89,7	97,1	95,9	93,7	94,4	90,6	93,6	96,4	88,8
Номер по порядку	11	12	13	14	15	16	17	18	19	20
τ_{20} , МПа	101	89,3	85,9	92,8	94,3	91,5	90,6	91,5	92,8	90,1
Номер по порядку	21	22	23	24	25	26	27	28	29	30
τ_{20} , МПа	97,2	94,9	91,5	93,8	96	95,9	89,3	99,8	95,9	107,8

Таблица 2. Бутстреп-выборка (k -я) на основе выборки, показанной в табл. 1

Номер по порядку	1	2	3	4	5	6	7	8	9	10
τ_{20} , МПа	107,8	93,6	90,1	99,8	92,8	97,1	89,3	90,5	91,5	89,3
Номер по порядку	11	12	13	14	15	16	17	18	19	20
τ_{20} , МПа	107,8	96,4	93,6	94,4	93,8	90,1	91,5	89,3	101,0	92,8
Номер по порядку	21	22	23	24	25	26	27	28	29	30
τ_{20} , МПа	89,3	88,8	89,7	101,0	97,2	93,7	93,8	89,3	89,7	93,8

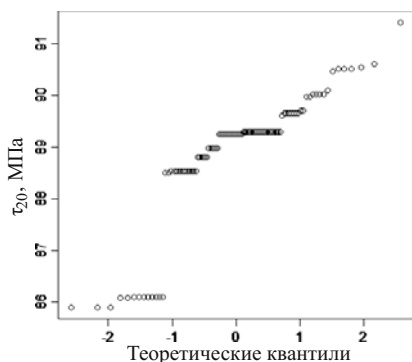


Рис. 3. Распределение 10 %-ной процентиля по бутстреп-выборкам

Для построения доверительных интервалов с применением бутстреп-моделирования для каждой $k = 1, 2, \dots, R$ бутстреп-выборки оценим среднее и получим выборку значений размером R . Гистограмма среднего показана на рис. 2. Видно, что гистограмма имеет колоколообразную форму, характерную для нормального распределения. Среднее значение по бутстреп-выборкам: математическое ожидание $(\tau_{20}) = 93,603$ МПа (обозначено на рис. 2 тонкой линией). Данное значение близко к среднему по исходной выборке (93,61 МПа).

Для построения α (%) доверительных интервалов для среднего R значений средних располагают в вариационный ряд. Индексы членов вариационного ряда (LOW — нижний и UP — верхний), ограничивающих доверительный интервал α [$\alpha \in (0, 1)$], определяются из выражений:

$$\text{LOW} = \text{целая часть} \left[\frac{1-\alpha}{2} R \right], \quad (3)$$

$$\text{UP} = \text{целая часть} \left[\frac{1+\alpha}{2} R \right], \quad (4)$$

которые для $R = 100$ и $\alpha = 0,9$ составят: $\text{LOW} = 5$ и $\text{UP} = 95$. Формулы (3), (4) справедливы также и для других случаев построения доверительных интервалов. Члены вариационного ряда бутстреп τ_{20} с этими индексами ограничивают 90 %-й доверительный интервал для среднего. Границы бутстреп показаны пунктирными линиями на рис. 2. Здесь же для сравнения жирными линиями отмечены доверительные интервалы, построенные по формулам математической

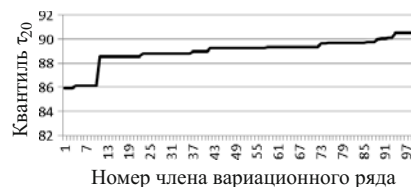


Рис. 4. Вариационный ряд значений 10 %-ных процентов τ_{20} (МПа)

статистики по критерию Стьюдента [9]. Видно, что доверительные интервалы близки.

Приведенный выше пример носил ознакомительный характер. Нас же интересуют доверительные интервалы для процентов. Вернемся к примеру случайной исходной выборке τ_{20} . Рассмотрим построение $\alpha = 0,95$ доверительных интервалов процента $\alpha = 10$ %.

В программном комплексе [10], который использовали для моделирования бутстреп и построения графиков в данной работе, приводятся несколько уточненных алгоритмов оценки процентов. По умолчанию использовали алгоритм программы № 3. Для построения B -базиса по каждой из R (в данном примере $R = 100$) бутстреп-выборок вычисляли 10 %-ную процентилю $Q(0,1)$. На рис. 3 показано распределение процентиля $Q(0,1)$ на нормальной вероятностной бумаге (теоретические квантили нормального распределения показаны по горизонтали). График существенно отклоняется от прямой линии, что свидетельствует о неправомерности описания распределения процентов нормальным законом.

В несколько другой форме, возможно, более подходящей для понимания алгоритма, это же распределение показано на рис. 4.

Значения индексов вариационного ряда (см. рис. 4) для определения B -базиса определяются по формулам (3), (4). Но поскольку для $\alpha = 0,95$ при $R = 100$ индексы получаются не целыми, то возьмем значения для индексов 2 и 3 и вычислим их среднее. В данном случае оно составляет 85,9 МПа, что согласуется со значением 85,6 МПа, оцененным по формулам (1), (2). Преимуществом оценки с применением бутстреп-моделирования является их большая интуитивная понятность, а также свойство робастности.

Для увеличения точности оценок далее в исследовании использовали число бутстреп-выборок $R = 1000$. Основные статистические характеристики и

Таблица 3. Данные по прочности на растяжение σ_b , МПа, препрега марки HexPlay

Номер поставки	Среднее	Среднеквадратическое отклонение	10 %-й процентиль	B -базис	
				бутстреп	[2]
7	3013,3	142,6	2817,0	2789	2693
8	2962,5	84,9	2850,5	2818	2780
9	2719,3	154,6	2512,2	2580	2361
10	2848,9	172,3	2630,2	2580	2463
11	2729,7	134,7	2579,3	2501	2403

значения B -базисов для выборок случайных значений прочности на растяжение σ_v (МПа) для пяти поставок приведены в табл. 3. Здесь представлены B -базисы, построенные по предлагаемому методу и по формулам (1), (2). Незначительные расхождения значений, возможно, объясняются несоответствием распределений случайных величин нормальному закону, что предполагается в методе построения по формулам (1), (2). В этом смысле метод, основанный на моделировании бутстреп, возможно, является более надежным.

Поскольку число элементов в выборках невелико ($n \approx 30$), что характерно для имеющихся в распоряжении исследователей данных по прочности углепластиков, по исходной выборке по разработанному методу, как и по формулам (1), (2), построить доверительные интервалы для 1 %-й процентиля (A -базиса) не представляется возможным. Для построения A -базиса по выборкам малого размера, возможно, потребуется разработка метода оценки базиса с экстраполяцией распределений по некоторым известным законам распределений. На основании анализа литературы по вопросам распределения прочности композиционных материалов и на основании результатов авторов [3] возможно использовать следующие законы: 1) нормальный; 2) двухпараметрический Вейбулла; 3) логарифмически-нормальный.

Таким образом, разработан метод для построения доверительных интервалов для 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 %-го доверительного интервала 10 %-й процентиля прочности (B -базис), полученные по формулам [2] и при моделировании бутстреп, близки. Представляет интерес дальнейшее исследование возможности применения статистического бутстрепа для анализа точности экспериментально полученных прочностных характеристик высокотехнологичной продукции.

ЛИТЕРАТУРА

1. Смердов А. А., Думанский А. М., Таирова Л. П. Комплексные экспериментальные исследования деформативных и прочностных свойств композитов для отсеков и обтекателей ракет-носителей / Вестник МГТУ им. Н. Э. Баумана. Сер. Машиностроение. 2012. С. 124 – 136.
2. Composite Materials Handbook - 17 (CMH-17). SAE International on behalf of CMH-17, Wichita State University, March 2012. Chapter 8. Statistical Method. P. 552 – 712.
3. Лисаченко Н. Г., Попов А. Г., Гадолина И. В. Анализ стабильности прочностных свойств современных углепластиков: сб. тр. конференции / Деформирование и разрушение композиционных материалов DFCMS-2016. ИМАШ РАН, Москва, Россия, 18 – 20 октября 2016 г. С. 74 – 76.
4. Диаконис П., Эфрон Б. Статистические методы с интенсивным использованием ЭВМ / В мире науки. 1983. № 7. С. 60 – 73.
5. Адлер Ю. П., Гадолина И. В., Ляндрес М. Н. Бутстреп-моделирование при построении доверительных интервалов по цензурированным выборкам / Заводская лаборатория. 1987. Т. 53. № 10. С. 90 – 94.
6. Орлов А. И. Эконометрика: Учебник. — М.: Экзамен, 2002. — 576 с.
7. By Yvonne H. S., Lee S. Iterated smoothed bootstrap confidence intervals for population quantiles / The Annals of Statistics. 2005. Vol. 33. N 1. P. 437 – 462.
8. Вешняков Б. В., Кибзун А. И. Применение метода бутстрепа для оценивания функции квантили / Автоматика и телемеханика. 2007. Вып. 1. С. 46 – 60.
9. Степнов М. Н., Зинин А. В. Прогнозирование характеристик материалов и элементов конструкций. — М.: Инновационное машиностроение, 2016. — 391 с.
10. Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>.

REFERENCES

1. Smerdov A. A., Dumanskii A. M., Tairova L. P. A comprehensive experimental study of the deformation and strength properties of composites for compartments and fairings of launch vehicles / Vestnik MG TU im. N. É. Bauman. Ser. Mashinost. 2012. P. 124 – 136 [in Russian].
2. Composite Materials Handbook - 17 (CMH-17). SAE International on behalf of CMH-17, Wichita State University, March 2012. Chapter 8. Statistical Method. P. 552 – 712.
3. Lisachenko N. G., Popov A. G., Gadolina I. V. The analysis of the stability of the strength properties of modern carbon fiber reinforced plastics: proc. of the conf. / Deformation and fracture of composite materials DFCMS-2016. IMASh RAN, Moscow, Russia, October 18 – 20, 2016. P. 74 – 76 [in Russian].
4. Diaconis P., Efron B. Computer-intensive method in Statistic / Sci. American. 1983. Vol. 248. N 5. P. 116 – 131.
5. Adler Yu. P., Gadolina I. V., Lyandres M. N. Bootstrap-modeling for confidence-interval building for censored sets / Zavod. Lab. 1987. Vol. 53. N 10. P. 90 – 94 [in Russian].
6. Orlov A. I. Econometry. — Moscow: Ékzamen, 2002. — 576 p. [in Russian].
7. By Yvonne H. S., Lee S. Iterated smoothed bootstrap confidence intervals for population quantiles / The Annals of Statistics. 2005. Vol. 33. N 1. P. 437 – 462.
8. Veshnyakov B. V., Kibzun A. I. Application of bootstrap method for quantile function estimation / Avtomat. Telemekh. 2007. Issue 1. P. 46 – 60 [in Russian].
9. Stepnov M. N., Zinin A. V. Prediction of materials and machins parts characteristics. — Moscow: Innovatsionnoe mashinostroenie, 2016. — 391 p. [in Russian].
10. Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>.