J. Mech. Phys. Solids 183 (2024) 105521

Contents lists available at ScienceDirect

Journal of the Mechanics and Physics of Solids

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

Is a high-throughput experimental dataset large enough to accurately estimate a statistic?

Yifan Zhou [a,1], Sirui Lin [b,1], Xuhui Zhang [b], Hou Wu [a], Jose Blanchet [b,*], Zhigang Suo [c,*], Tongqing Lu [a,*]

[a] Department of Engineering Mechanics, State Key Lab for Strength and Vibration of Mechanical Structures, Xi'an Jiaotong University, Xi'an 710049, China
[b] Department of Management Science and Engineering, Stanford University, 475 Via Ortega, Stanford, CA 94305, USA
[c] John A. Paulson School of Engineering and Applied Sciences, Kavli Institute for Bionano Science and Technology, Harvard University, MA 02138, USA

ARTICLE INFO

Keywords:
Rupture
High-throughput
Bootstrap
Statistics

ABSTRACT

In materials science, experimental datasets are commonly used to estimate various statistics of random variables. This paper focuses on a specific random variable: the rupture stretch of a material. Examples of statistics include average, standard deviation, coefficient of variation, and different quantiles. How accurate is the estimate of such a statistic? The answer depends on the statistic, the size of the experimental dataset, and how much the random variable scatters. Here we demonstrate a procedure to generate a large experimental dataset and use the experimental dataset to estimate the accuracy of various statistics of the rupture stretch. We use a high-throughput experiment to measure the rupture stretches of 160 specimens of a silicone rubber. We then use the bootstrap method to determine the 90 % confidence intervals of several statistics. We find that the experimental dataset accurately estimates the average, standard deviation, and 50 % quantile. However, the experimental dataset does not reliably estimate extremely low or high quantiles. This finding indicates an experimental dataset much larger than 160 specimens is needed to accurately estimate rare-event rupture stretch. We further apply the bootstrap method to an experimental dataset of strengths of 33 specimens of a ceramic. The result indicates that this experimental dataset is too small to accurately estimate the average strength of the ceramic. Our findings demonstrate that the common practice of using small datasets to estimate statistics of material properties is outdated and meaningless. The high-throughput experiment provides a large experimental dataset of rupture stretch, from which the bootstrap method quantifies the accuracy of the estimates of various statistics. The bootstrap method does not require the user to have sophisticated expertise in statistical analysis. Nor does the bootstrap method require the dataset to obey any statistical distribution.

1. Introduction

Experimental measurements are used to determine material properties such as strength. Strength of a material often scatters greatly

* Corresponding authors.
E-mail addresses: jose.blanchet@stanford.edu (J. Blanchet), suo@seas.harvard.edu (Z. Suo), tongqinglu@mail.xjtu.edu.cn (T. Lu).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.jmps.2023.105521
Received 7 October 2023; Received in revised form 1 December 2023; Accepted 7 December 2023
Available online 12 December 2023

from specimen to specimen. For example, when 23 specimens of a fused silica were measured, the strength ranged from about 1 GPa to about 10 GPa (Proctor et al., 1967). As another example, when 40 specimens of a polymer coated silica were measured, the strength ranged from 0.5 to 6 GPa (Kurkjian et al., 1989). Strength of a material is a random variable, commonly characterized using statistics such as average, standard deviation, and quantiles. The scatter of strength is often fit to various statistical distributions, including normal distribution, Weibull distribution, and Gumbel distribution (Basu et al., 2009; Dirikolu and AKTAŞ, 2002; Doremus, 1983; Lu et al., 2002)

To characterize a material with a large scatter in strength, a great number of specimens must be tested under identical conditions. However, the ASTM standard (International, 2015) requires that the strength be measured using 6–10 specimens. The use of such a small number of specimens is understood from a practical limitation. Traditional tests require rupturing specimens one by one, and are time-consuming and labor-intensive. The recent developments of high-throughput experiment methods are helpful to obtain a large experimental dataset. Various of high-throughput experiments have been developed in biology, chemistry, pharmacy, and material science (de Pablo et al., 2019; Hughes et al., 2014; Mennen et al., 2019; Ren et al., 2018; Shevlin, 2017; Soon et al., 2013; Sun et al., 2019). But the high-throughput experiments to measure mechanical properties are few (Darling and Di Carlo, 2015; Tweedie et al., 2005). Our recent paper proposes a high-throughput experiment method to measure rupture stretch of hundreds of specimens simultaneously (Wu et al., 2023; Zhou et al., 2022). The feature of this setup includes (i) a procedure to prepare a large number of specimens; (ii) a kinematic mechanism to apply loads to these specimens simultaneously; and (iii) a method to record rupture of individual specimens by image processing. The large experimental dataset is likely to provide more accurate estimates of various statistics than a small experimental dataset. However, the size of the experimental dataset required for an accurate estimate of a statistic depends on the statistic itself and the variation in the values within the experimental dataset. Whether the dataset of a high-throughput experiment is sufficiently large to accurately estimate a statistic is a subject of investigation.

In the field of mechanics of materials, several statistical methods have been used to test whether an experimental dataset is large enough to estimate various statistics of material properties. Examples include F-test, T-test, A-D test, etc. (Wang et al., 2022; Wu et al., 2023). Here we apply the bootstrap method to the dataset generated by the high-throughput experiment, and determine the confidence interval of the estimate of each statistic (Diaconis and Efron, 1983; Efron and Tibshirani, 1991). The bootstrap method does not require the user to have sophisticated expertise in statistical analysis. Nor does the bootstrap method require the dataset to obey any statistical distribution. This method has been used to evaluate various statistics of material properties (Edwards et al., 2012; Young et al., 2007). When employing the bootstrap method to construct confidence intervals, it is essential to have a sufficiently large dataset. If the dataset is too small, for instance, less than 10, the resulting confidence interval may become too wide, making it challenging to draw

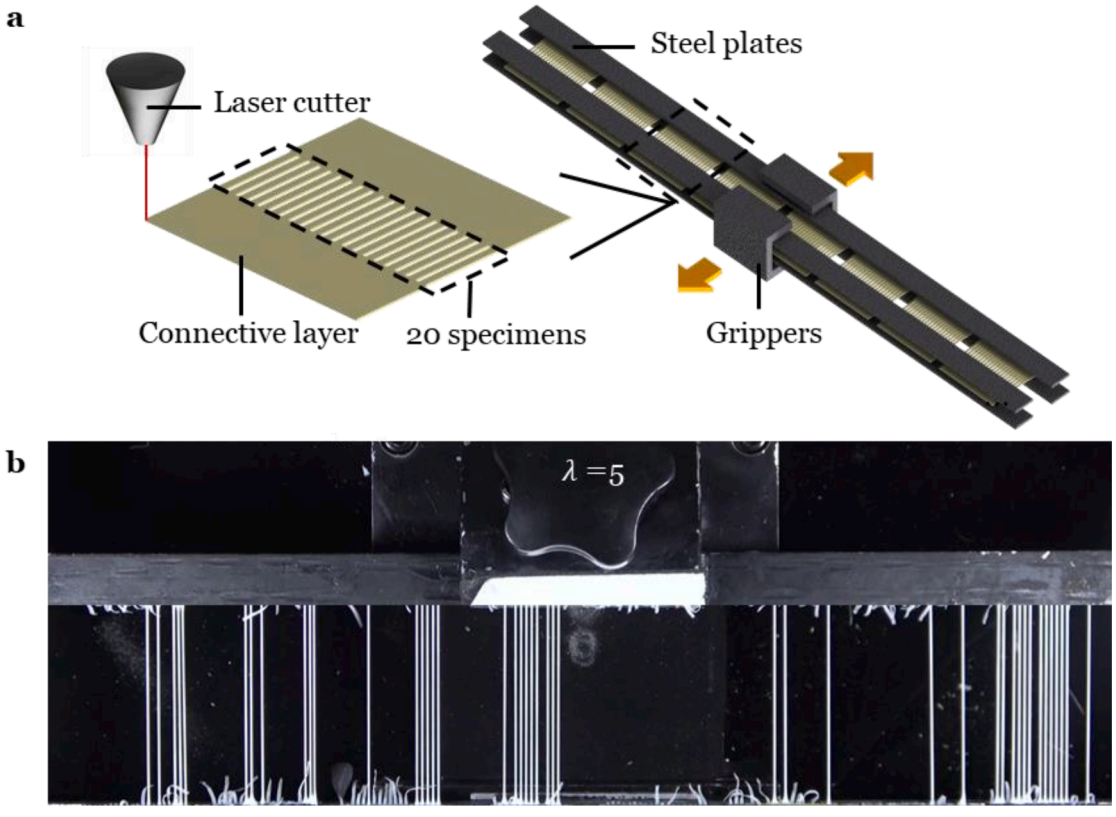

Fig. 1. High-throughput rupture experiment. (a) A silicone rubber sheet containing 20 specimens is prepared using a laser cutter. A total of eight sheets containing 160 specimens are pulled simultaneously to the same stretch. (b) A snapshot of the experiment at a stretch of $\lambda = 5$.

meaningful conclusions. Nevertheless, obtaining a large dataset of strength is nearly impossible without high-throughput experiments.

Here we combine the high-throughput experiment and bootstrap method to characterize the statistics of rupture stretch of a material. We conduct a high-throughput experiment to simultaneously stretch 160 specimens of a silicone rubber and record their individual rupture stretches. We then apply the bootstrap method to compute the 90 % confidence interval for each statistic, including the average, standard deviation, coefficient of variation, and various quantiles. Our findings indicate that the experimental dataset can accurately estimate the average, standard deviation, and 50 % quantile. However, it does not accurately estimate extremely low or high quantiles. This finding indicates an experimental dataset much larger than 160 specimens is needed to accurately estimate rare-event rupture stretch. We further apply the bootstrap method to an existing experimental dataset of 33 specimens of a ceramic. The result indicates that this experimental dataset is too small to accurately estimate even the average strength of the ceramic.

The combination of high-throughput experiment and bootstrap analysis is synergistic. The former tests a large number of specimens, and the latter calculates the confidence interval of any statistic. The high-throughput experiment and the bootstrap analysis work well together because, for any statistic, a sufficiently large experimental dataset can generate a narrow bootstrap confidence interval. Furthermore, if the bootstrap analysis shows that the confidence interval of a statistic is too wide to satisfy the requirement of an application, additional high-throughput experiments can be conducted to provide a large enough dataset. The bootstrap analysis quantifies the size of the experimental dataset required for an accurate estimate of a statistic. For example, when estimating extreme quantiles, the bootstrap analysis suggests caution, and the high-throughput experiment offers an opportunity to address this issue by creating a sufficiently large dataset. It is hoped that the iteration of high-throughput experiment and bootstrap analysis will be adopted to characterize the statistics of materials. Our findings demonstrate that the common practice of using small datasets to estimate statistics of material properties is outdated and meaningless.

2. High-throughput rupture experiment

2.1. Experimental dataset

We have developed a high-throughput rupture experiment (Wu et al., 2023; Zhou et al., 2022). Here we use a variation of this experimental setup to pull 160 silicone rubber specimens. Starting with a silicone rubber sheet, we use a laser cutter (Universal Laser Systems, Inc.) to cut 20 specimens, each having a dumbbell shape (Fig. 1a). The central part of each specimen has the size of $12 \times 1 \times 0.1mm$. The upper and bottom of the specimens remain connected, so that all the 20 specimens can be glued to steel plates simultaneously. The steel plates hold a total of eight sheets of specimens. The steel plates are gripped and pulled by a home-made tensile machine at a rate of $30mm/min$ (Fig. 1b). Define the stretch λ by the current spacing between the two grippers divided by the initial spacing. We videotape the experiment using a camera (SONY FDR-AX60). As the tensile machine pulls the two grippers, the specimens gradually rupture. For example, when the applied stretch is $\lambda = 5$, a total of 104 specimens have ruptured (Fig. 1b). Following the method described in our previous papers, we process the video to detect the rupture stretches of the specimens. The fraction of ruptured specimens, F, is plotted as a function of stretch, λ (Fig. 2). Each data point corresponds to a ruptured specimen. The plot consists of the experimental dataset of rupture stretches of the 160 specimens, $\{\lambda^1, \lambda^2, \lambda^3, .., \lambda^{160}\}$.

2.2. Is the experimental dataset identically and independently distributed?

When a large number of dice are rolled simultaneously, it is commonly expected that the individual dice behave independently but similarly. In statistics, the dice are said to be independent and identically distributed (iid). However, in a high-throughput experiment,

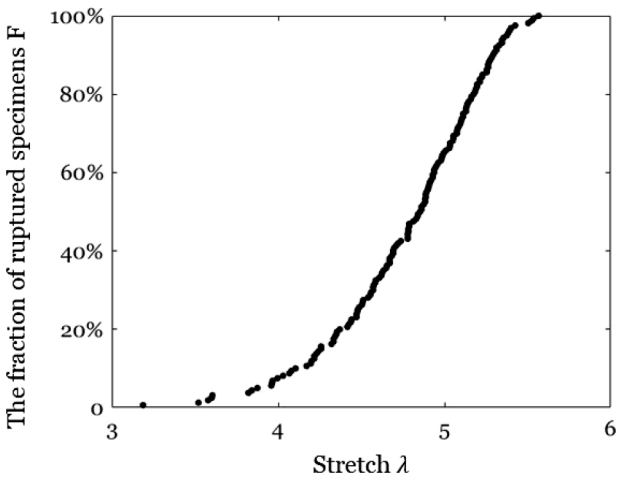

Fig. 2. The fraction of ruptured specimens as a function of applied stretch.

such as the one described in this paper, the specimens may not be iid. The non-iid behavior can be caused by various issues in the experimental design (Wu et al., 2023). Examples include insufficient rigidity of the plates, insufficient frictional force between the grippers and the plates, and inconsistent preparation of the specimens. These issues can be mitigated by improving the experimental design.

Whether an experimental dataset is iid is tested as follows. Divide the experimental dataset into several subsets. Each subset contains a large number of specimens and generates a cumulative distribution function (cdf). If the cdfs of all the subsets are sufficiently close, the experimental dataset is said to be iid. One way to measure the similarity between cdfs of multiple subsets of specimens is the Anderson-Darling test (Scholz and Stephens, 1987). In our experiment, 160 specimens are fabricated in eight sheets, each sheet containing 20 specimens. We divide the experimental dataset into two subsets, with each subset containing four sheets of specimens. From the cdfs of the subsets, the Anderson-Darling test calculates a P-value, which measures how likely the observed difference between cdfs is due to chance, rather than bias in experiments. We test the null hypothesis H_0 that the experimental dataset is iid. We calculate the P-value using the function module "AnDarksamtest" in MATLAB and select a significance level of $\alpha = 0.01$. If $P < \alpha$, we reject H_0 and conclude that the experimental dataset is not iid. If $P > \alpha$, we cannot reject H_0 and there is insufficient evidence to conclude that the experimental dataset is not iid. Our experimental dataset gives $P = 0.037$, so we cannot reject H_0. We will regard that our experimental dataset is iid.

2.3. Several statistics of the experimental dataset

Given an experimental dataset of rupture stretch $\{\lambda^1, \lambda^2, \lambda^3, .., \lambda^n\}$, where n is the number of specimens, we can calculate various statistics. The most frequently used statistic in rupture is the average rupture stretch, which is defined by

$$\mu = \frac{1}{n}\sum_{i=1}^{n}\lambda^i. \tag{1}$$

The average rupture stretch of the 160 specimens is 4.77.
The standard deviation of rupture stretch is defined by

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\lambda^i - \mu\right)^2}. \tag{2}$$

The standard deviation of rupture stretch of the 160 specimens is 0.46.
The coefficient of variation of rupture stretch is defined by

$$CV = \sigma/\mu. \tag{3}$$

The coefficient of variation of rupture stretch of the 160 specimens is 0.096.
The α-quantile is defined by

$$F(\lambda) = \alpha. \tag{4}$$

Because the function $F(\lambda)$ is discrete, for a given value of α, the dataset may not yield a value of rupture stretch λ. In that case one chooses the smallest value of λ such that $F(\lambda) > \alpha$. The 50 % quantile of rupture stretch is the stretch of the 80th ruptured specimen and is 4.85. The 10 % quantile of rupture stretch is the stretch of the 16th ruptured specimen and is 4.10. The 1.25 % quantile of rupture stretch is the stretch of the 2nd ruptured specimen and is 3.52.

Each of the above statistics is calculated from the dataset obtained in our experiment. Of course, this experiment can be conducted

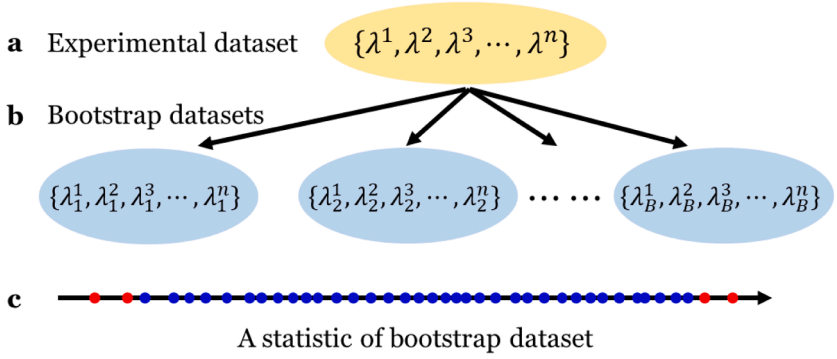

Fig. 3. Bootstrap method. (a) An experimental dataset of n specimens is obtained: $\{\lambda^1, \lambda^2, \lambda^3, .., \lambda^n\}$. (b) B bootstrap datasets are generated from the experimental dataset: $\{\lambda_1^1, \lambda_1^2, \lambda_1^3, .., \lambda_1^n\}$, $\{\lambda_2^1, \lambda_2^2, \lambda_2^3, .., \lambda_2^n\}$, ..., $\{\lambda_B^1, \lambda_B^2, \lambda_B^3, .., \lambda_B^n\}$. (c) For illustration purposes, B is set as 40. The statistic of each bootstrap dataset is represented by a dot on a line. The lowest and highest 5 % statistics (i.e., the leftmost and rightmost two dots) are colored red. The four red dots are removed and the remaining range represents the bootstrap estimate of the 90 % confidence interval of the statistic.

repeatedly to provide rupture stretches of more than 160 specimens. The larger dataset can also be used to calculate the statistic. Will the statistic calculated using the experimental dataset of 160 specimens be an accurate estimate of the statistic of larger datasets? The accuracy of estimates of the above statistics will be gauged by the bootstrap method.

3. Bootstrap method

The bootstrap method is a technique to quantify the uncertainty associated with a statistic estimated from a large experimental dataset (Diaconis and Efron, 1983; Efron and Tibshirani, 1991). Let an experimental dataset of n specimens be $\lambda^1, \lambda^2, .., \lambda^n$ (Fig. 3a). From this experimental dataset, we can estimate various statistics. Examples include average, standard deviation, coefficient of variation, and different quantities. Imagine that the experiment could be conducted with infinite specimens. Such a hypothetical experiment would provide true values of the statistics. How much does each statistic estimated from the experiment of n specimens differ from the true value of the statistic? To answer this question, the bootstrap method uses the experimental dataset of n specimens to calculate a confidence interval of the true value of the statistic.

Specifically, the bootstrap method generates B datasets, called bootstrap datasets. Each bootstrap dataset is generated by randomly selecting n specimens from the experimental dataset with equal probability, allowing repetition. Denote the Bth bootstrap dataset by $\lambda_B^1, \lambda_B^2, \cdots, \lambda_B^n$ (Fig. 3b). For example, let $B = 40$. Each of the 40 bootstrap datasets estimates a value of a statistic. This procedure results in 40 values of the statistic, which are plotted as dots on a line (Fig. 3c). The lowest 5 % of the 40 values are the leftmost two dots, colored red. The highest 5 % of the 40 values are the rightmost two dots, also colored red. After removing the four red dots, the range of the remaining 36 blue dots is the 90 % confidence interval of the true value of the statistic.

In summary, the bootstrap method invokes the variability in a large, but finite experimental dataset. The method makes no assumption on the distribution of the hypothetical experiment of infinite specimens. By repetitively resampling the experimental dataset, the bootstrap method generates a confidence interval of the true value of the statistic.

4. Accuracy of average rupture stretch estimated from experimental dataset

4.1. Bootstrap datasets

As we describe above, the high-throughput experiment generates rupture stretches of 160 specimens (Fig. 2). From the experimental dataset we generate 40 bootstrap datasets. For each bootstrap dataset, we randomly select 160 rupture stretches with replacement from the experimental dataset. Thus, the bootstrap specimens may contain repetitions of actual specimens. We represent each bootstrap dataset by the fraction of ruptured bootstrap specimens as a function of stretch, $F(\lambda)$.

Two representative bootstrap datasets are plotted on the $F - \lambda$ plane (Fig. 4). The two functions are alike but have some differences. An inspection shows that each of the 40 bootstrap datasets is somewhat different from the experimental dataset. This observation simply means that each bootstrap dataset contains some repetitive specimens.

4.2. Average rupture stretch

Each of the 40 bootstrap datasets of rupture stretches produces a bootstrap average. We represent the 40 bootstrap averages as 40 dots on a line (Fig. 5). Fig. 5a and b show the same 40 dots in different scales. The bootstrap averages scatter in the interval [4.69,

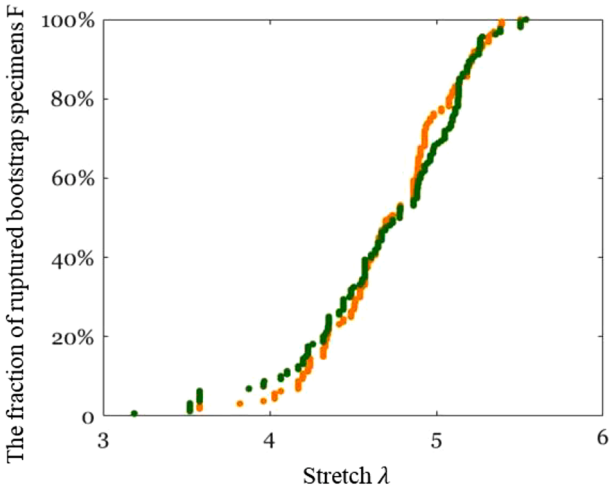

Fig. 4. Bootstrap datasets. For two representative bootstrap datasets, the fractions of ruptured bootstrap specimens are plotted as functions of stretch, $F(\lambda)$.

4.87]. The lowest 5 % averages (i.e., the leftmost two dots) and the highest 5 % averages (i.e., the rightmost two dots) are colored red. We remove the four red dots, and the remaining interval [4.74, 4.84] is the bootstrap estimate of the 90 % confidence interval of the average rupture stretch.

As we noted above, the average rupture stretch of the experimental dataset of 160 specimens is 4.77. The bootstrap method estimates the width of the 90 % confidence interval of the average rupture stretch to be 0.1. The ratio of the two numbers, $0.1/4.77 = 0.02$, gives a measure of the size of the 90 % confidence interval relative to the average stretch. This accuracy is likely to satisfy most applications, and one might say that the experimental dataset of 160 specimens is large enough to estimate this particular statistic: average rupture stretch.

4.3. The effect of the number of specimens

In engineering practice, the average rupture stretch is often estimated using an experimental dataset of few specimens, e.g., fewer than ten specimens (ASTM D638-14, 2015). To determine the accuracy of these estimates, we examine experiment datasets of various numbers of specimens. From the actual experimental dataset of 160 specimens, we create 16 virtual experimental datasets by randomly selecting rupture data without replacement, each containing n = 10, 20, 30, …, or 160 specimens. For each virtual experimental dataset, we generate 40 bootstrap datasets and calculate the average rupture stretch μ for each. These bootstrap averages are plotted on the μ-n plane (Fig. 6). As the number of specimens in the virtual experimental dataset increases, the scatter of the bootstrap averages decreases. We also calculate the 90 % confidence interval of the estimated average rupture stretch for each of the 16 virtual datasets. For a virtual experiment dataset with ten specimens, the 90 % confidence interval of the average rupture stretch is [4.46, 4.96], which is wider than that of the experimental dataset of 160 specimens [4.74, 4.84]. Whether a confidence interval is narrow enough depends on applications. If the bootstrap method estimates a confidence interval is too wide to be acceptable in an application, one must conduct an additional high-throughput experiment of more specimens.

4.4. The effect of the number of bootstrap datasets

We can examine how the number of bootstrap datasets affects the 90 % confidence interval. We test B values of 20, 40, 60, and 80. Each bootstrap dataset provides a bootstrap average μ. As before, for each value of B, we plot the average of each bootstrap dataset as a dot on a line, and mark the lowest 5 % and the highest 5 % dots in red (Fig. 7). We remove the red dots, and the remaining interval is the bootstrap estimate of the 90 % confidence interval of the average rupture stretch. When B is greater than 20, the 90 % confidence interval remains almost unchanged. In the remainder of the paper, we will fix the number of bootstrap datasets to $B = 40$.

5. The accuracy of the standard deviation of rupture stretch estimated from the experimental dataset

The standard deviation of rupture is another commonly used statistic. Fig. 8 shows 40 dots representing the 40 bootstrap standard deviations of rupture stretch. These values scatter in the range [0.39, 0.48]. After removing the four red dots, the remaining range [0.40, 0.47] is the bootstrap estimate for the 90 % confidence interval of the standard deviation of rupture stretch. The width of the 90 % confidence interval is 0.07. The standard deviation for the experimental dataset of 160 specimens is 0.46. The ratio of these two numbers, $0.07/0.46 = 0.15$, provides a measure of the size of the confidence interval relative to the standard deviation. This implies that, with 90 % confidence, the interval width is about 15 % of the standard deviation of the experimental dataset. This level of accuracy is likely to be sufficient for most applications, indicating that a dataset of 160 specimens is large enough to estimate this particular statistic: standard deviation.

Fig. 9 shows 40 dots representing the 40 bootstrap coefficients of variation. These values scatter in the range [0.091, 0.112]. The bootstrap estimate for the 90 % confidence interval of the coefficient of variation is [0.095, 0.111] with the width of 0.016. The coefficient of variation for the experimental dataset of 160 specimens is 0.096. The ratio of these two numbers, $0.016/0.096 = 0.17$.

6. The accuracy of the quantile rupture stretch estimated from the experimental dataset

We use the bootstrap method to determine the accuracy of three quantiles: the 50 %, 10 %, and 1.25 % quantile rupture stretches.

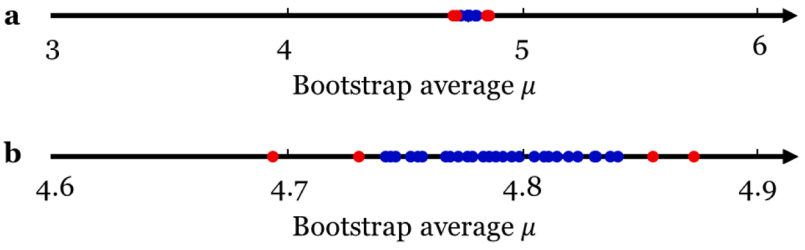

Fig. 5. Bootstrap estimate for the 90 % confidence interval of the average rupture stretch. (a) The data points are plotted using the same interval [3,6] as used in Figs. 2 and 4. (b) The data points are plotted in the interval [4.6, 4.9].

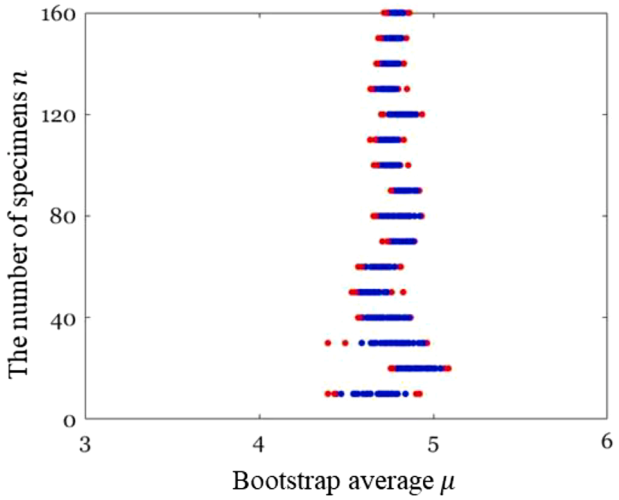

Fig. 6. What would the outcome be if we had a small dataset from our experiment? Our actual experiment generates a dataset of rupture stretches of 160 specimens. From this dataset, we create 16 separate virtual datasets without replacement, each containing 10, 20, 30, …, 160 specimens. For each virtual dataset of n specimens, we generate 40 bootstrap datasets. The average rupture stretch μ of each bootstrap dataset is represented as a dot on the μ-n plane. As the number of specimens increases, the scatter of the bootstrap averages decreases.

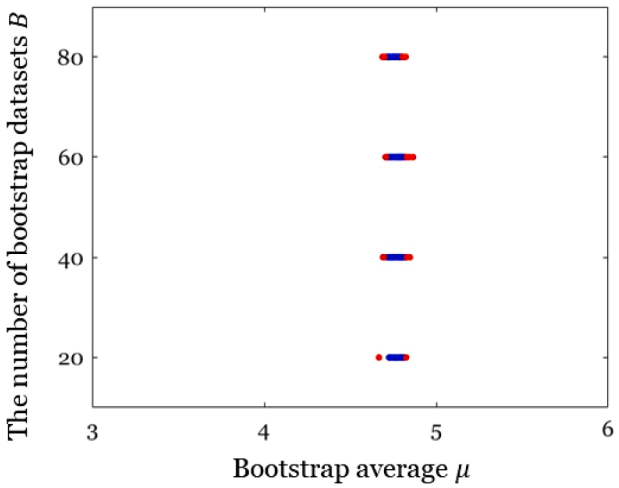

Fig. 7. The bootstrap estimate of the 90 % confidence interval is nearly unchanged when the number of bootstrap datasets becomes large enough.

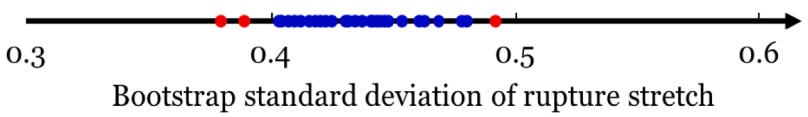

Fig. 8. The 90 % confidence interval of the standard deviation of rupture stretch using the bootstrap method.

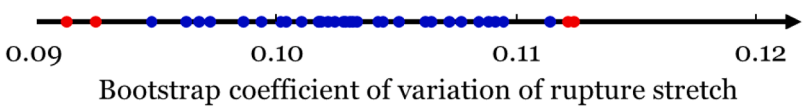

Fig. 9. The 90 % confidence interval of the coefficient of variation of rupture stretch using the bootstrap method.

Fig. 10a shows 40 dots representing the 40 bootstrap 50 % quantile rupture stretches. These values scatter in the range [4.64, 4.9]. The 90 % confidence interval is [4.67, 4.89]. The width of the 90 % confidence interval is 0.22. The experimental dataset of 160 specimens has a 50 % quantile rupture stretch of 4.85. The ratio of the two numbers, $0.22/4.85 = 0.045$. This level of accuracy indicates that an experimental dataset of 160 specimens is large enough to estimate this particular statistic: the 50 % quantile rupture stretch.

The bootstrap 10 % and 1.25 % quantile rupture stretches are plotted in Fig. 10b and 10c, respectively. For extremely low or high quantiles such as the 1.25 % quantile rupture stretch, there is often repetition among the bootstrap values, resulting in only 5 distinct dots in Fig. 11c. As a result, the calculated 90 % confidence interval of the 1.25 % quantile rupture stretch is not reliable and the dataset cannot provide accurate estimates for extremely low or high quantiles.

7. Bootstrap average strength of a brittle solid

The strength of brittle solids such as ceramics can vary greatly. An experiment of 33 specimens found that the average rupture strength was 860 MPa (Basu et al., 2009). How much does the average rupture strength estimated from the experiment of 33 specimens differ from the true value of the average rupture strength? To answer this question, from the experimental dataset of 33 specimens, 40 bootstrap datasets are generated, each producing an average rupture strength. These bootstrap averages are plotted as 40 dots on a line (Fig. 11). The lowest and highest 5 % bootstrap averages, colored in red, are removed to leave a range of [770 MPa, 960 MPa]. This range is the bootstrap estimate for the 90 % confidence interval of the true average rupture strength. The width of the 90 % confidence interval is 190 MPa. The average rupture strength of the experimental dataset is 860 MPa. The ratio of these two numbers, $190/860 = 0.22$, indicates the size of the confidence interval relative to the average rupture strength. This ratio for the ceramic is about an order of magnitude larger than that for silicone rubber. This difference may reflect that ceramic is more flaw sensitive than silicone rubber, or the experimental dataset of 33 specimens is still too small to accurately estimate the average rupture strength. We then generate a virtual experiment dataset of 33 silicone rubber specimens from the original 160 specimens, and find that the 90 % confidence interval of average rupture stretch is [4.73, 4.93]. The width of the 90 % confidence interval is 0.2. The average rupture stretch of the virtual experimental dataset is 4.83. The ratio between these two numbers is 0.04, much smaller than that for the ceramic. This comparison confirms that the ceramic is more flaw sensitive than rubber, and requires a larger experimental dataset to accurately estimate the average rupture strength than silicone rubber does.

8. Discussion

Large datasets have been accumulating in materials science. For example, Guo et al. (2021) have used the few-shot learning algorithm and genetic algorithm to learn 54 datasets of five-element alloys simulated by molecular dynamics, and found the composition that optimizes stiffness and critical resolved shear stress. Wang et al. (2022) have integrated statistical learning with domain knowledge, and learned the strength law from 123 datasets obtained by the three-point bending test of concrete. Xiong et al. (2020a,b) have developed machine learning models for mechanical properties of various materials. Buehler (2023) have developed artificial intelligence models for mechanical and material design. These methods may also be used to analyze large datasets generated by high-throughput experiments.

Several methods have been used to analyze small datasets. For example, the data augmentation method makes transformations to existing dataset to increase the amount of data and obtain better results. Transfer learning uses the existing knowledge of solving similar problems to accelerate the study of a current problem. These methods have not been used to calculate the confidence interval of estimates of material properties. A powerful approach to treat small datasets is to use domain knowledge. The approach is especially effective when the small datasets have great uncertainty. This theme is outside the scope of the present paper, and will be pursued in subsequent work. By contrast, the bootstrap method requires no domain knowledge.

9. Conclusion

In summary, we have conducted a high-throughput experiment to generate an experimental dataset of rupture stretches of 160 specimens. We use the experimental dataset to estimate various statistics, including average, standard deviation, coefficient of variation, and different quantiles. The accuracy of each statistic is then gauged using the bootstrap method. For each statistic, the bootstrap method calculates the 90 % confidence interval. Our findings indicate that the experimental dataset can accurately estimate the average, standard deviation, coefficient of variation, and 50 % quantile of the rupture stretch. However, the experimental dataset is too small to accurately estimate extremely low or high quantiles. We further apply the bootstrap method to an experimental dataset of the strength of 33 specimens of a ceramic. The result indicates that this experimental dataset is too small to accurately estimate the average strength of the ceramic. The bootstrap method provides a quantitative measure of the size of the experimental dataset required for an accurate estimation of a statistic. If the bootstrap analysis shows that the confidence interval of a statistic is too wide to satisfy the requirement of an application, additional high-throughput experiments can be conducted to provide a large enough dataset. This procedure iterates until the dataset reaches the necessary size to meet the required accuracy of the application.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used CHATGPT in order to improve readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

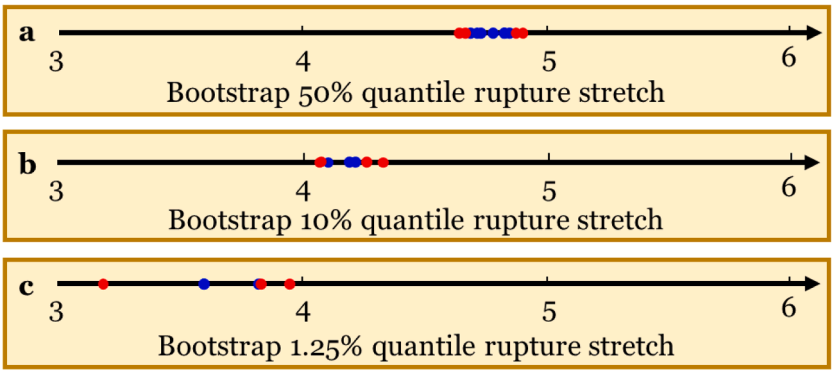

Fig. 10. The 90 % confidence interval of three quantiles, estimated using the bootstrap method. (a) the 50 % quantile rupture stretch, (b) the 10 % quantile rupture stretch, and (c) the 1.25 % quantile rupture stretch.

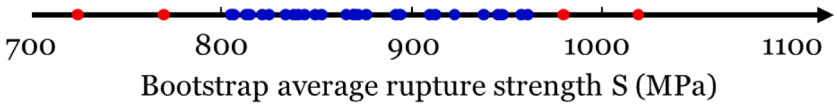

Fig. 11. Bootstrap estimate of the 90 % confidence interval of the average rupture strength of a ceramic.

CRediT authorship contribution statement

Yifan Zhou: Data curation, Methodology, Validation, Writing – original draft. **Sirui Lin:** Methodology, Validation, Writing – review & editing. **Xuhui Zhang:** Methodology. **Hou Wu:** Data curation, Writing – review & editing. **Jose Blanchet:** Funding acquisition, Methodology, Resources, Supervision, Writing – review & editing. **Zhigang Suo:** Conceptualization, Funding acquisition, Resources, Supervision, Writing – review & editing. **Tongqing Lu:** Conceptualization, Funding acquisition, Resources, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The work at Xi'an Jiaotong University is supported by NSFC (No. 11922210). The work at Stanford and Harvard are supported by the Air Force Office of Scientific Research under award number FA9550-20-1-0397.

References

Basu, B., Tiwari, D., Kundu, D., Prasad, R., 2009. Is Weibull distribution the most appropriate statistical strength distribution for brittle materials? Ceram. Int. 35, 237–246.
Buehler, M.J., 2023. MeLM, a generative pretrained language modeling framework that solves forward and inverse mechanics problems. J. Mech. Phys. Solids 181, 105454.
Darling, E.M., Di Carlo, D., 2015. High-throughput assessment of cellular mechanical properties. Annu. Rev. Biomed. Eng. 17, 35–62.
de Pablo, J.J., Jackson, N.E., Webb, M.A., Chen, L.Q., Moore, J.E., Morgan, D., Jacobs, R., Pollock, T., Schlom, D.G., Toberer, E.S., 2019. New frontiers for the materials genome initiative. NPJ Comput. Mater. 5, 41.
Diaconis, P., Efron, B., 1983. Computer-intensive methods in statistics. Sci. Am. 248, 116–131.
Dirikolu, M.H., Aktaş, A., 2002. Statistical analysis of fracture strength of composite materials using Weibull distribution. Turk. J. Eng. Environ. Sci. 26, 45–48.
Doremus, R., 1983. Fracture statistics: a comparison of the normal, Weibull, and Type I extreme value distributions. J. Appl. Phys. 54, 193–198.
Edwards, D.J., León, R.V., Young, T.M., Guess, F.M., Crookston, K.A., 2012. Comparison of two wood plastic composite extruders using bootstrap confidence intervals on measurements of sample failure data. Qual. Eng. 25, 23–33.
Efron, B., Tibshirani, R., 1991. Statistical data analysis in the computer age. Science 253, 390–395.
Guo, T., Wu, L., Li, T., 2021. Machine learning accelerated, high throughput, multi-objective optimization of multiprincipal element alloys. Small 17, e2102972.

Hughes, J.R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., De Gobbi, M., Taylor, S., Gibbons, R., Higgs, D.R., 2014. Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. Nat. Genet. 46, 205–212.
International, A., 2015. ASTM D638-14, Standard Test Method for Tensile Properties of Plastics. ASTM International.
Kurkjian, C.R., Krause, J.T., Matthewson, M.J., 1989. Strength and fatigue of silica optical fibers. J. Lightwave Technol. 7, 1360–1370.
Lu, C., Danzer, R., Fischer, F.D., 2002. Fracture statistics of brittle materials: weibull or normal distribution. Phys. Rev. E Stat. Nonlinear Soft Matter Phys. 65, 067102.
Mennen, S.M., Alhambra, C., Allen, C.L., Barberis, M., Berritt, S., Brandt, T.A., Campbell, A.D., Castañón, J., Cherney, A.H., Christensen, M., 2019. The evolution of high-throughput experimentation in pharmaceutical development and perspectives on the future. Org. Process Res. Dev. 23, 1213–1242.
Proctor, B., Whitney, I., Johnson, J., 1967. The strength of fused silica. Proc. R. Soc. Lond. Ser. A Math. Phys. Sci. 297, 534–557.
Ren, F., Ward, L., Williams, T., Laws, K.J., Wolverton, C., Hattrick-Simpers, J., Mehta, A., 2018. Accelerated discovery of metallic glasses through iteration of machine learning and high-throughput experiments. Sci. Adv. 4, eaaq1566.
Scholz, F.W., Stephens, M.A., 1987. K-sample Anderson–darling tests. J. Am. Stat. Assoc. 82, 918–924.
Shevlin, M., 2017. Practical high-throughput experimentation for chemists. ACS Med. Chem. Lett. 8, 601–607.
Soon, W.W., Hariharan, M., Snyder, M.P., 2013. High-throughput sequencing for biology and medicine. Mol. Syst. Biol. 9, 640.
Sun, S., Hartono, N.T., Ren, Z.D., Oviedo, F., Buscemi, A.M., Layurova, M., Chen, D.X., Ogunfunmi, T., Thapa, J., Ramasamy, S., 2019. Accelerated development of perovskite-inspired materials via high-throughput synthesis and machine-learning diagnosis. Joule 3, 1437–1451.
Tweedie, C.A., Anderson, D.G., Langer, R., Van Vliet, K.J., 2005. Combinatorial material mechanics: high-throughput polymer synthesis and nanomechanical screening. Adv. Mater. 17, 2599–2604.
Wang, J.H., Jia, J.N., Sun, S., Zhang, T.Y., 2022. Statistical learning of small data with domain knowledge-sample size-and pre-notch length-dependent strength of concrete. Eng. Fract. Mech. 259, 108160.
Wu, H., Zhang, X., Zhou, Y., Blanchet, J., Suo, Z., Lu, T., 2023. Detection and reduction of systematic bias in high-throughput rupture experiments. J. Mech. Phys. Solids 174, 105249.
Xiong, J., Shi, S.Q., Zhang, T.Y., 2020a. A machine-learning approach to predicting and understanding the properties of amorphous metallic alloys. Mater. Des. 187, 108378.
Xiong, J., Zhang, T., Shi, S., 2020b. Machine learning of mechanical properties of steels. Sci. China Technol. Sci. 63, 1247–1255.
Young, T.M., Perhac, D.G., Guess, F.M., León, R.V., 2007. Bootstrap confidence intervals for percentiles of reliability data for wood plastic composites. For. Prod. J. Personal communication.
Zhou, Y., Zhang, X., Yang, M., Pan, Y., Du, Z., Blanchet, J., Suo, Z., Lu, T., 2022. High-throughput experiments for rare-event rupture of materials. Matter 5, 654–665.