

International Prize in Statistics Awarded to Stanford's Bradley Efron

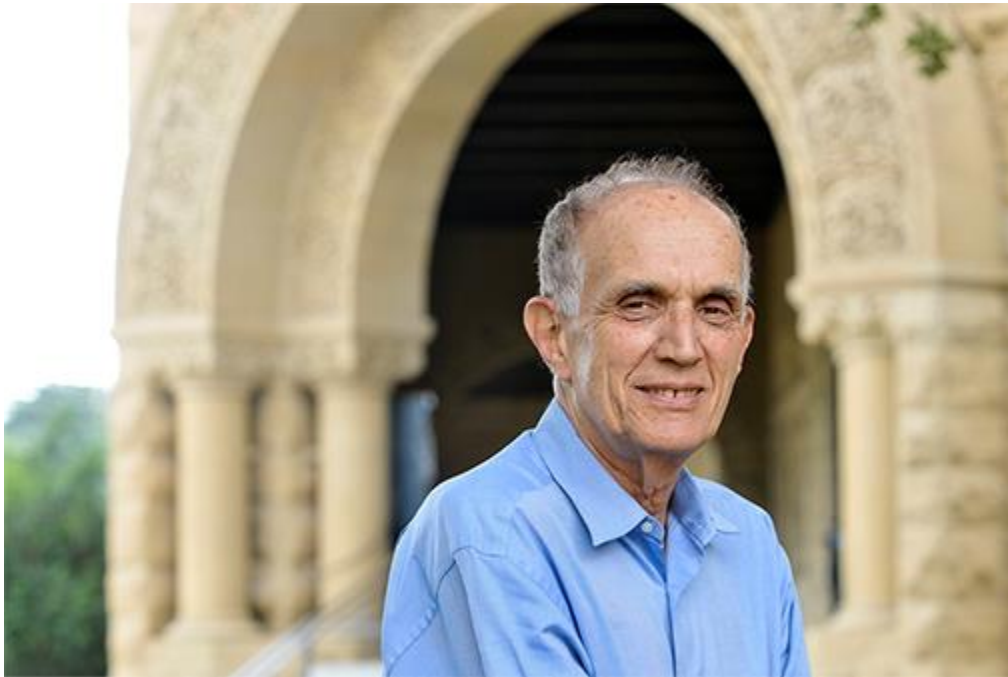


Photo courtesy of Stanford University Department of Statistics

ALEXANDRIA, VA (November 12, 2018): The International Prize in Statistics has been awarded to Bradley Efron, professor of statistics and biomedical data science at Stanford University, in recognition of the "bootstrap," a method he developed in 1977 for assessing the uncertainty of scientific results that has had extraordinary impact across many scientific fields.

With the bootstrap (described by Efron below), scientists are able to learn from limited data in a simple way that enables them to assess the uncertainty of their findings. In essence, it is possible to simulate a potentially infinite number of data sets from an original data set and—in looking at the differences—measure the uncertainty of the result from the original data analysis.

Made possible by computing, the bootstrap powered a revolution that placed statistics at the center of scientific progress. It helped propel statistics beyond techniques that relied on complex mathematical calculations or unreliable

approximations, enabling scientists to assess the uncertainty of their results in more realistic and feasible ways.

"Because the bootstrap is easy for a computer to calculate and is applicable in an exceptionally wide range of situations, the method has found use in many fields of science, technology, medicine and public affairs," says Sir David Cox, inaugural winner of the International Prize in Statistics.

Cornell University and [EPAM Systems Inc.](#) examined research databases worldwide and found that, since 1980, the bootstrap (and multiple variations on the term such as bootstrapping) has been cited in excess of 200,000 documents in more than 200 journals worldwide. Citations are found in fields such as agricultural research, biochemistry, computer science, engineering, immunology, mathematics, medicine, physics and astronomy and the social sciences.

The impact of the bootstrap across research fields as measured by citation

The dataset contains over 200,000 articles from over 200 journals between 1980 and 2018



Agricultural and Biological Sciences	44,476
Arts and Humanities	6,586
Biochemistry, Genetics and Molecular Biology	29,135
Business, Management and Accounting	8,248
Chemical Engineering	1,823
Chemistry	3,734
Computer Science	33,177
Decision Sciences	11,843
Dentistry	71
Earth and Planetary Sciences	9,744
Economics, Econometrics and Finance	12,088
Energy	2,814
Engineering	21,106
Environmental Science	14,563
Health Professions	2,105
Immunology and Microbiology	12,605
Materials Science	3,407
Mathematics	30,072
Medicine	31,022
Multidisciplinary	108
Neuroscience	6,827
Nursing	1,033
Pharmacology, Toxicology and Pharmaceutics	3,331
Physics and Astronomy	13,071
Psychology	9,162
Social Sciences	15,461
Veterinary	926

Courtesy of Cornell University and EPAM Systems Inc.

"While statistics offers no magic pill for quantitative scientific investigations, the bootstrap is the best statistical pain reliever ever produced," says Xiao-Li Meng, Whipple V. N. Jones Professor of Statistics at Harvard University. "It has saved countless scientists and researchers the headache of finding a way to assess uncertainty in complex problems by providing a simple and practical way to do so in many seemingly hopeless situations."

"The bootstrap was a quantum leap in statistical methodology that has enabled researchers to improve the lives of people everywhere," says Sally Morton, dean of

and professor of statistics in the Virginia Tech College of Science. "Efron is a statistical poet of enormous beauty, applicability and impact."

Efron will accept the prize next summer at the [2019 World Statistics Congress](#) in Kuala Lumpur.

Background

Efron was born in May 1938 to Russian immigrants and grew up in St. Paul, Minnesota. He credits his salesman father, Miles, for cultivating in him a love of math and science, in part through baseball and bowling scoring. "He kept track of these things," says Efron, "so I grew up with a lot of numbers around me and that was very helpful—I was training to be a statistician without realizing it."

He won a national merit scholarship the year it was first introduced and went to Caltech. It was more than an intellectually eye-opening experience. "I'd never seen a mountain or an ocean," says Efron.

Initially, he thought he was going to become a mathematician, but he realized abstract mathematics was not where his interests or talents lay. Enrolling in a PhD program at Stanford, he switched to statistics. "I remember when going into statistics that first year I thought, 'This will be pretty easy, I've dealt with math and that's supposed to be hard.' But statistics was much harder for me at the beginning than any other field. It took years before I felt really comfortable."

Statistics, by this point (the early 1950s), had been deeply immersed in decision theory—and the formal mathematical principles that explained inference. It was abstract, highly mathematical—and not especially concerned with applied problems. This was about to change under the synergistic influence of data analysis and computing. Statistics suddenly became deeply relevant to scientific research because it could answer questions previously unanswerable.

Statisticians grappled with the problems of outliers in data sets, limited data and multiple unknowns, developing techniques through computers that were able to overcome the calculative limits of human thought. It was in this milieu that Efron was inspired by and built on the work of John Tukey, David Cox and Rupert Miller to create the bootstrap.

The name was inspired by the 18th-century fictional character Baron Munchausen and a variation of a story in which he pulls himself out of a swamp by his own

bootstraps. In a similar vein, [the statistician or scientist could now use their own data to assess the uncertainty in their own data](#). Initially, Efron's paper on the bootstrap was rejected for publication because it didn't have enough theorems—[so he added some at the end](#).

"The truth is, I didn't think it was anything wonderful when I did it," says Efron. "But it was one of those lucky ideas that is better than it seems at first view." It was a tool many scientists could use, and use easily, especially as personal computing provided the power to do the number crunching. And it worked.

The Bootstrap, Simplified

Suppose you want to know the average household income in your city. You can't afford a complete census, so you randomly sample 100 households, record the 100 incomes and take their average, say \$29,308. That sounds precise, but you would like some estimate of how accurate it really is. A straightforward, but impractical, approach would be to take several more random samples of 100 households, compute the average each time and see how much the averages differed from each other.

The bootstrap lets you approximate this impractical approach using only the original sample's data. A bootstrap data set is a random sample of size 100 drawn from the original 100 incomes. You can imagine writing each of the original incomes on a slip of paper, putting the slips into a hat and randomly drawing a slip out. Record the number, put the slip back into the hat and repeat this process 99 more times. The result would be a bootstrap data set, and we can make as many bootstrap data sets as we wish, each time taking their average. Let's say we do 250 of them, giving 250 bootstrap averages. The variability of the 250 averages is the bootstrap estimate of accuracy for the original estimate \$29,308.

The same idea can be applied to find the accuracy of any statistic, say the median income instead of the average or, perhaps, something much more complicated, which makes the bootstrap ideal for the often elaborate statistical methods of modern scientific practice.

More About Brad Efron

Efron is Max H. Stein Professor of Humanities and Sciences, professor of statistics and professor of biostatistics with the department of biomedical data science in the school of medicine; he serves as co-director of the mathematical and computational sciences program. He has held visiting faculty appointments at Harvard, UC Berkeley and Imperial College, London. A recipient of a 2005 National Medal of

Science for his contributions to theoretical and applied statistics, especially the bootstrap sampling technique, he was awarded the Guy Medal in Gold by the Royal Statistical Society in 2014. He served in 2004 as president of the American Statistical Association.

[Stanford University profile](#)