

The upstrap

CIPRIAN M. CRAINICEANU*

Department of Biostatistics, Johns Hopkins University, 615 N. Wolfe St. Baltimore, MD 21205 USA
ccraini1@jhu.edu

ADINA CRAINICEANU

Department of Computer Science, US Naval Academy, 572M Holloway Road, Stop 9F,
Annapolis, MD, 21402, USA

SUMMARY

The bootstrap, introduced in Efron (1979. Bootstrap methods: another look at the jackknife. *The Annals of Statistics* 7, 1–26), is a landmark method for quantifying variability. It uses sampling with replacement with a sample size equal to that of the original data. We propose the upstrap, which samples with replacement either more or fewer samples than the original sample size. We illustrate the upstrap by solving a hard, but common, sample size calculation problem. The data and code used for the analysis in this article are available on GitHub (2018. <https://github.com/ccrainic/upstrap>).

Keywords: Bootstrap; sampling; sample size calculation.

1. ALGORITHM

Consider a data set where the observed vectors of observations at the subject level are $Y_i, i = 1, \dots, n$, θ is a parameter of interest, and $\hat{\theta}_n$ is an estimator of θ . For a fraction $r \in (0, \infty)$, we are interested in estimating the distribution of $\hat{\theta}_{rn}$, the estimator based on a resampled fraction r of the original sample size. The upstrap algorithm is as follows

Data: $Y_i, i = 1, \dots, n$; sample size n ; fraction r ; number of samples B

Result: $\hat{\theta}_{rn}^{(1)}, \dots, \hat{\theta}_{rn}^{(B)}$

$b = 1$;

while $b < B$ **do**

1. Obtain $Y_i^{(b)}, i = 1, \dots, rn$, a sample with replacement of size rn from $Y_i, i = 1, \dots, n$;
2. Obtain $\hat{\theta}_{rn}^{(b)}$, the estimator of θ based on $Y_i^{(b)}, i = 1, \dots, rn$;
3. $b = b + 1$;

end

The upstrap algorithm provides the entire distribution of the estimator for a dataset of size r times the sample size of the original data. The distribution can be used to estimate its characteristics, such as the mean or standard deviation. One might ask about the added value of resampling rn rather than n observations.

*To whom correspondence should be addressed.

	Estimate	Std. error	z value	Pr(> z)
Gender	1.157	0.079	14.61	< 2e-16***
Age	0.037	0.005	7.38	1.61e-13***
BMI	0.138	0.007	18.67	< 2e-16***
HTN	0.758	0.473	1.60	0.109
Age:HTN	-0.009	0.007	-1.22	0.221

We explore this question and illustrate the added benefit for a challenging but routine question in applied regression.

2. SAMPLE SIZE CALCULATION FOR REGRESSION

We now show how to calculate the sample size using the upstrap in a relatively simple regression scenario for which there are no standard methods. Consider the case of a binary regression problem where the outcome is whether or not a person has moderate to severe sleep apnea and the predictors are gender, age, BMI, hypertension status (coded HTN), and hypertension by age interaction. The data comes from the Sleep Heart Health Study (SHHS) (Quan *and others* (1997) and Redline *and others* (1998)) and is publicly available as part of the The National Sleep Research Resource (NSRR) (Dean *and others* (2016)). Moderate to severe sleep apnea is defined as a respiratory disturbance index at $\geq 4\%$ oxygen desaturation (labeled rdi4p in the SHHS dataset) greater or equal to 15. We use data from visit one of the SHHS, which contains 5804 individuals.

The regression results based on these data are shown in the table (intercept not shown) indicating that hypertension (coded HTN) is not significant at the $\alpha = 0.05$ level.

The question that we would like to answer is at what sample size do we expect to identify a hypertension effect on having moderate to severe sleep apnea in this model using the two-sided Wald test at $\alpha = 0.05$ with a power $1 - \beta = 0.8$? The idea is simple. We set a grid of fractions of the sample size; in this case, this grid is $r = 1, 1.2, \dots, 5$ and for every value of r we upstrap $B = 10\,000$ data sets of size rn . For every sample, we conduct the two-sided Wald test for HTN in the model above and reject the null hypothesis of no association if the corresponding p -value is less than 0.05. Figure 1 provides the frequency with which the test for no HTN effect is rejected as a function of the multiplier of the sample size, r .

For example, for multiplier $r = 1$, we obtain the bootstrap p -value, the percent of times the null of no-association is rejected when sampling with replacement a dataset of the same size. For multiplier $r = 2$, we produced $B = 10\,000$ samples with replacement from SHHS data sets with twice the number of subjects $2n = 11\,608$. For each sampled dataset, we ran the model and recorded whether the p -value for HTN was smaller than 0.05. For this sample size, we obtained that the HTN effect was identified in 62% of the samples. We also obtained that the power was equal to 0.794 at the sample size multiplier $r = 3.0$ and 0.817 at multiplier $r = 3.2$, indicating that the power 0.8 would be attained at $3.1 * n \approx 17\,500$ subjects. There are very few methods to estimate the sample size in such examples and we contend that the upstrap is a powerful and general method to conduct such calculations. Similar approaches could be used in many other situations, including sample size calculations for detecting a treatment effect in the context of longitudinal clinical trial data or gene by environment interactions in genomics studies.

One of the limitations of this approach is that it could result in large estimators of sample size that may be impractical in some applications. However, our approach provides the ability to conduct such calculations and support the decision to either initiate or not such a study. As a last point, we consider

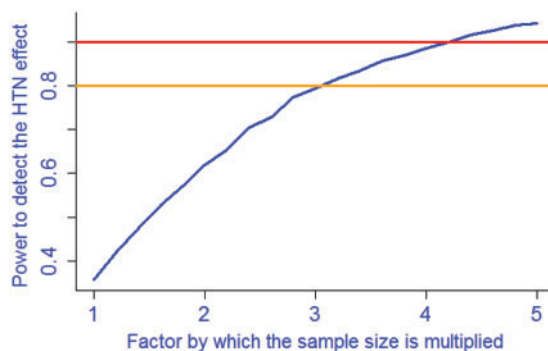


Fig. 1. Power to detect the main effect of HTN (y -axis) as a function of the multiplier, r , of the original sample size (x -axis). Here, $r = 1$ on the x -axis corresponds to the original sample size, $n = 5804$, $r = 2$ corresponds to double the sample size, $2n = 11\,608$, and so on. The model uses moderate to severe sleep apnea (binary variable) as an outcome and gender, age, BMI, HTN, and age by HTN interaction as predictors. Horizontal lines indicate powers equal to 0.8 and 0.9, respectively

that the upstrap is safe to use in all problems where the bootstrap is used. However, more simulations and theoretical work are necessary to establish this assertion.

ACKNOWLEDGMENTS

Conflict of Interest: None declared.

FUNDING

This work was supported by the National Heart, Lung, and Blood Institute, National Institutes of Health (5R01HL123407-04 to C.M.C.); National Institute of Neurological Disorders and Stroke, National Institutes of Health (5R01NS060910-10 to C.M.C.).

REFERENCES

- Dean, D. A., Goldberger, A. L., Mueller, R., Kim, M., Rueschman, M., Mobley, D., Sahoo, S. S., Jayapandian, C. P., Cui, L., Morrical, M. G. and others. (2016). Scaling up scientific discovery in sleep medicine: the National Sleep Research Resource. *Sleep* **39**, 1151–1164.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics* **7**, 1–26.
- GitHub (2018). <https://github.com/ccrainic/upstrap>.
- Quan, S. F., Howard, B. V., Iber, C., Kiley, J. P., Nieto, F. J., O'Connor, G. T., Rapoport, D. M., Redline, S., Robbins, J., Samet, J. M. and others. (1997). The sleep heart health study: design, rationale, and methods. *Sleep* **20**, 1077–1085.
- Redline, S., Sanders, M. H., Lind, B. K., Quan, S. F., Iber, C., Gottlieb, D. J., Bonekat, W. H., Rapoport, D. M., Smith, P. L. and Kiley, J. P. (1998). Methods for obtaining and analyzing unattended polysomnography data for a multicenter study. Sleep Heart Health Research Group. *Sleep* **21**, 759–767.

[Received March 20, 2018; revised August 20, 2018; accepted for publication August 21, 2018]