

# The upstrap

Ciprian M. Crainiceanu\*

February 7, 2018

Bootstrap [2] 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.

## 1 Algorithm

Consider a data set where the observed vector of observations at the subject level are  $Y_i$ ,  $i = 1, \dots, n$ ,  $\theta$  is a parameter of interest, and  $\hat{\theta}_n$  is an estimator of  $\theta$ . For a fraction  $r \in (0, \infty)$  we are interested in estimating the distribution of  $\hat{\theta}_{rn}$ , the estimator based on the 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  $\theta$  based on  $Y_i^{(b)}, i = 1, \dots, rn$ ;
3.  $b = b + 1$ ;

**end**

## 2 Sample size calculation for regression

We now show how to calculate the sample size using upstrapping 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, and hypertension by age interaction. The data comes from the Sleep Heart Health Study (SHHS) [3, 4] and is publicly available as part of the The National Sleep Research Resource (NSRR) [1]. Moderate to severe sleep apnea is defined as a Respiratory Disturbance Index at  $\geq 4\%$

---

\*Professor, Department of Biostatistics, Johns Hopkins University, 615 N. Wolfe St. Baltimore, MD 21205 USA. E-mail: ccrainic@jhsph.edu

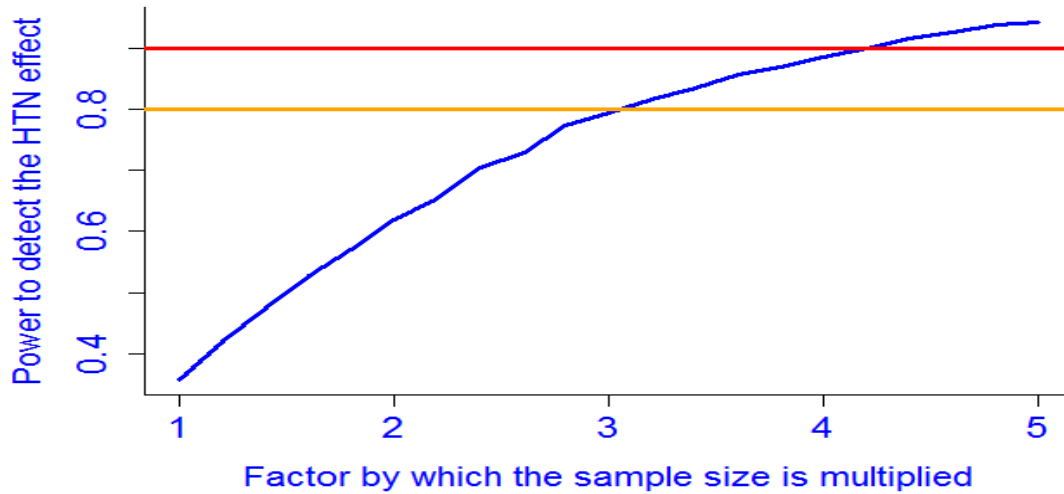


Figure 1: Power to detect the HTN effect in an upstrap (sample with replacement of the data with a larger sample size as the original data) as a function of the multiplier  $r$  of the original sample size.

oxygen desaturation (rdi4p) greater or equal to 15 (citations). We use data from visit 1 of the SHHS, which contains 5804 individuals.

The regression results based on these data are shown below indicating that hypertension (coded HTNDerv\_s1) is not significant at the  $\alpha = 0.05$  level.

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-8.710927	0.421770	-20.653	< 2e-16	***
gender	1.156897	0.079181	14.611	< 2e-16	***
age_s1	0.036892	0.005000	7.378	1.61e-13	***
bmi_s1	0.138369	0.007412	18.669	< 2e-16	***
HTNDerv_s1	0.758377	0.473129	1.603	0.109	
age_s1:HTNDerv_s1	-0.008748	0.007146	-1.224	0.221	

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 probability  $1 - \beta = 0.8$ . 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  $B = 10000$  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  $r = 2$  we produced  $B = 10000$  samples with replacement from SHHS data sets with twice the number of subjects  $2n = 11608$ . For each data 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 estimating a fixed effect (e.g. treatment) using longitudinal data in the context of a clinical trials or the sample size necessary to detect gene by gene and gene by environment interactions in genomics studies.

## References

- [1] D.A. Dean, A L. Goldberger, R. Mueller, M. Kim, M. Rueschman, D. Mobley, S.S. Sahoo, C.P. Jayapandian, L. Cui, M.G. Morrical, S. Surovec, G.Q. Zhang, and S. Redline. Scaling Up Scientific Discovery in Sleep Medicine: The National Sleep Research Resource. *Sleep*, 39(5):1151–1164, 2016.
- [2] B. Efron. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26, 1979.
- [3] S.F. Quan, B.V. Howard, C. Iber, J.P. Kiley, F.J. Nieto, G.T. O’Connor, D.M. Rapoport, S. Redline, J. Robbins, J.M. Samet, , and P.W. Wahl. The Sleep Heart Health Study: design, rationale, and methods. *Sleep*, 20(12):1077–1085, 1997.
- [4] S. Redline, M. H. Sanders, B.K. Lind, S.F. Quan, C. Iber, D.J. Gottlieb, W.H. Bonekat, D.M. Rapoport, P.L. Smith, and J.P. Kiley. Methods for obtaining and analyzing unattended polysomnography data for a multicenter study. Sleep Heart Health Research Group. *Sleep*, 21(7):759767, 1998.