Cornell University Library

We gratefully acknowledge supp
the Simons Fo
and member ins

arXiv.org > stat > arXiv:1106.2125

Statistics > Methodology

Bootstrapping data arrays of arbitrary order

Art B. Owen, Dean Eckles

(Submitted on 10 Jun 2011 (v1), last revised 27 Sep 2012 (this version, v3))

In this paper we study a bootstrap strategy for estimating the variance of a mean taken over large multifactor crossed random effects data sets. We apply bootstrap reweighting independently to the levels of each factor, giving each observation the product of independently sampled factor weights. No exact bootstrap exists for this problem [McCullagh (2000) Bernoulli 6 285-301]. We show that the proposed bootstrap is mildly conservative, meaning biased toward overestimating the variance, under sufficient conditions that allow very unbalanced and heteroscedastic inputs. Earlier results for a resampling bootstrap only apply to two factors and use multinomial weights that are poorly suited to online computation. The proposed reweighting approach can be implemented in parallel and online settings. The results for this method apply to any number of factors. The method is illustrated using a 3 factor data set of comment lengths from Facebook.

Submission history

From: Art B. Owen [view email]
[v1] Fri, 10 Jun 2011 17:26:33 GMT (28kb,D)
[v2] Wed, 25 Jan 2012 01:28:14 GMT (31kb,D)
[v3] Thu, 27 Sep 2012 11:07:10 GMT (177kb)

Which authors of this paper are endorsers?

Link back to: arXiv, form interface, contact.