Subsampling p-values

Arthur Berg
Penn State University

Timothy L. McMurry
DePaul University

Dimitris N. Politis
University of California, San Diego

April 20, 2010

Abstract

A construction of p-values for hypothesis tests based on subsampling and the related m out of n bootstrap is introduced. The p-values are based on a modification of the usual subsampling hypothesis tests that involves an appropriate centering of the subsampled or bootstrapped test statistics as in the construction of confidence intervals. This modification makes the hypothesis tests more powerful, and as a consequence provides meaningful p-values. The procedure is shown to apply for both i.i.d. and stationary data satisfying mixing conditions. The new p-values have the desired properties of being asymptotically uniform under the null hypothesis and converging to zero under the alternative.

1 Introduction

Let $X_1, \ldots, X_n$ denote a sample of observations taking values in a sample space S with probability law P, where P is assumed to belong to a class of laws $\mathbf{P}$. We consider the problem of testing the null hypothesis H_0 which asserts that $P \in \mathbf{P_0}$ versus the alternative, H_1, that $P \in \mathbf{P_1}$, where $\mathbf{P_0}$ and $\mathbf{P_1}$ are disjoint, and $\mathbf{P} = \mathbf{P_0} \cup \mathbf{P_1}$.

Subsampling confidence intervals were developed in Politis and Romano (1994) while subsampling hypothesis testing was studied in Politis and Romano (1996); Politis, Romano, and Wolf (1999) describes these and many other subsampling procedures in detail, including the related m/n bootstrap, i.e., resampling with smaller resample size; see also Bickel, Götze, and van Zwet (1997). Interestingly, there is little work to date on the construction of p-values for subsampling tests although some literature exists on p-values for the bootstrap with full resample size (e.g. Beran, 1986; Davison and Hinkley, 1997; Loh, 1985; Liu and Singh, 1997; Romano, 1988).

In this paper, we attempt to remedy this situation in providing uselful p-values for subsampling and the m/n bootstrap. At the same time, we revisit the original subsampling tests and show that a test that is essentially based on inverting the subsampling confidence intervals has better power. The next section reviews the framework of subsampling-based inference including the aforementioned modified subsampling test for i.i.d. and dependent data. Section 3 details the p-value construction while Section 4 presents several simulation results; technical proofs have been placed in the last section.

2 Subsampling confidence intervals and hypothesis tests

2.1 Confidence intervals

Subsampling inference starts with a statistic $t_n = t_n(X_1, \ldots, X_n)$ that estimates a parameter $t(P)$. Of interest is the sampling distribution of t_n defined as

$$J_n(x, P) = Prob_P\{\tau_n(t_n(X_1, \ldots, X_n) - t(P)) \leq x\},$$

where τ_n is a normalizing sequence that gives $J_n(x, P)$ a nontrivial limit as $n \to \infty$.

The fundamental idea behind subsampling is that $J_n(x, P)$ can be accurately approximated by the suitably centered and normalized distribution of the same statistic calculated on appropriately chosen subsets of the data of size b. Let $t_{n,b,i}$ denote the statistic calculated on the ith subset of size b, i.e.,

$$t_{n,b,i} := t_b(X_{i_1}, X_{i_2}, \ldots, X_{i_b}).$$

Also let N_n denote the total number of available subsets of the data of size b. Consider the following cases:

1. **I.i.d. data: subsampling.** In this case, the ith subsample is constructed by sampling *without* replacement from the data set $\{X_1, \ldots, X_n\}$ with the purpose of forming a subsample of size b. There are $N_n = \binom{n}{b}$ possible subsamples to choose from but typically this number is huge, so in practice the procedure is implemented by using a large (but not huge) number of randomly selected subsets.

2. **I.i.d. data: b/n bootstrap.** Same as above but sampling *with* replacement from the data set $\{X_1, \ldots, X_n\}$ with the purpose of forming a resample of size b; we will use b/n instead of m/n bootstrap for consistency of notation.

3. **Strictly stationary and strong mixing data: subsampling.** Here, the ith subsample is the sub-series $(X_i, \ldots, X_{i+b-1})$, and $N_n = n - b + 1$.

The subsampling estimator of $J_n(x, P)$ is defined as follows:

$$L_{n,b}(x) := N_n^{-1} \sum_{i=1}^{N_n} 1\{\tau_b(t_{n,b,i} - t_n) \leq x\}. \tag{1}$$

As shown in Politis and Romano (1994),

$$L_{n,b}(x) - J_n(x, P) \to_P 0 \tag{2}$$

under general conditions including

$$N_n \to \infty, \quad b \to \infty \text{ but } b^k/n \to 0, \tag{3}$$

(see Remark 1 below for the appropriate values of k) and that

$$\tau_n \text{ is such that } \tau_b/\tau_n \to 0 \text{ whenever } b/n \to 0, \tag{4}$$

as well as the following:

Assumption 1. As $n \to \infty$, $J_n(x, P)$ converges weakly to a continuous (in x) limit law $J(x, P)$.

Remark 1. For subsampling, equation (3) is assumed with $k = 1$. For the b/n bootstrap, one needs $k = 2$ if no extra conditions are imposed (Politis and Romano, 1993). However, with some additional structure, it is possible to have $k = 1$ (see Bickel et al., 1997), or even $k = 0$ e.g., when the full resample size bootstrap works and $b = n$.

Strictly speaking, eq. (2) only requires continuity of $J(x, P)$ at x; however, if $J(x, P)$ is continuous and strictly increasing for all x in the region of interest, i.e., in the α-tail, and we can further claim that

$$L_{n,b}^{-1}(\alpha) - J_n^{-1}(\alpha, P) \to 0 \tag{5}$$

in probability where $F^{-1}(\alpha)$ denotes the α-quantile of a distribution F. Eq. (5) is the basis for claiming that subsampling confidence intervals for $t(P)$ based on $L_{n,b}^{-1}(\alpha)$ are asymptotically valid.

2.2 Subsampling hypothesis tests

With great generality, we assume that $t(P) = 0$ under $H_0 : P \in \mathbf{P_0}$, and $t(P) > 0$ under $H_1 : P \in \mathbf{P_1}$. Hence, the null distribution

$$J_n(x, P_0) = Prob_{P_0}\{\tau_n t_n(X_1, \ldots, X_n) \leq x\}$$

is now of interest since the ideal α-level test rejects H_0 in favor of H_1 when $T_n := \tau_n t_n(X_1, \ldots, X_n)$ is bigger than $J_n^{-1}(1 - \alpha, P_0)$.

Using the same subsampling framework described as in the previous subsection, Politis and Romano (1996) defined the subsampling estimator of $J_n(x, P_0)$ as follows:

$$G_{n,b}(x) := N_n^{-1} \sum_{i=1}^{N_n} 1\{\tau_b t_{n,b,i} \leq x\}. \tag{6}$$

Comparing $G_{n,b}(x)$ to $L_{n,b}(x)$ note that the former uses the information that $t(P) = 0$ under H_0 while the latter estimates the centering constant (the statistic t_n) from the data.

Politis et al. (1999) give a proof of the consistency of the test that rejects the null hypothesis at level α when T_n exceeds the $1 - \alpha$ quantile $G_{n,b}^{-1}(1 - \alpha)$. In other words, the test rejects H_0 when

$$N_n^{-1} \sum_{i=1}^{N_n} 1\{T_n \geq T_{n,b,i}\} > 1 - \alpha. \tag{7}$$

where $T_{n,b,i} = \tau_b t_{n,b,i}$. The intuition behind eq. (7) is that under the null hypothesis, the subsampled distribution of the $T_{n,b,i}$ approximates the sampling distribution of T_n, and so we can reject the null hypothesis if T_n is above the $1 - \alpha$ quantile of the subsampled distribution.

Due to its very general applicability, this subsampling test has motivated significant theoretical and practical development; see, for example, Choi (2005); Choi and Chue (2007); Delgado, Rodrıguez-Poo, and Wolf (2001); Gonzalo and Wolf (2005); Lehmann and Romano (2005); Linton, Maasoumi, and Whang (2005); Politis and Romano (2008a,b); Politis, Romano, and Wolf (2001). The shortcoming of this approach is that when the alternative hypothesis is true, $G_{n,b}^{-1}(1-\alpha)$ is not a good approximation of the correct threshold $J_n^{-1}(1 - \alpha, P_0)$. To see this, note that the center of the subsampling distribution of the $T_{n,b,i}$ drifts towards infinity, albeit more slowly than the sampling distribution of T_n. The consequence is a test that–although asymptotically consistent–does not have optimal power, and does not provide an efficient notion of p-value.

This shortcoming can be remedied while, at the same time, improving the power of the subsampling test through an appropriate centering. To see how, consider the test constructed by inverting the confidence regions obtained by the quantiles of the data-centered subsampling distribution $L_{n,b}(x)$. Such a test rejects H_0 when

$$N_n^{-1} \sum_{i=1}^{N_n} 1\{\tau_n t_n \geq \tau_b (t_{n,b,i} - t_n)\} > 1 - \alpha, \tag{8}$$

i.e., when

$$N_n^{-1} \sum_{i=1}^{N_n} 1\{T_n \geq T_{n,b,i} - \tau_b t_n\} > 1 - \alpha. \tag{9}$$

Under the null hypothesis, $\tau_b t_n \to_P 0$ by Slutsky's theorem, so the above test behaves similarly to test (7) under the null. Under the alternative, however, the subsampling distribution of the $\tau_b(t_{n,b,i} - t_n)$ remains approximately centered at 0 instead of drifting to infinity, and thus test (8) has improved power, as will be formally shown in the next section.

Remark 2. While the tests described in this section appear to be one sided, they can be used to test two sided alternatives through an appropriate choice of test statistic t_n. For example, to test $H_0 : E[X_i] = 0$ versus $H_1 : E[X_i] \neq 0$, the statistic $t_n = |\bar{X}|$ with rate $\tau_n = n^{1/2}$ is appropriate. By the central limit and continuous mapping theorems, $\tau_n t_n$ satisfies the limiting distribution assumptions and produces an asymptotically valid two sided test.

3 Subsampling p-values

The tests in the preceding section, and in all previous subsampling literature, are defined in terms of their critical regions which provide asymptotically consistent decision rules. Nonetheless, in practice p-values are often desired as an approximate measure of the strength of the evidence against the null. In the present section, we describe the construction of p-values for the tests described by equations (7) and (8) and argue that the the centered test (8) is better suited for this purpose than its uncentered counterpart.

For the uncentered test (7), the p-value equals the proportion of the $T_{n,b,i}$ which exceed the observed test statistic T_n. That is, the "uncentered p-value" is

$$p_{uc} = N_n^{-1} \sum_{i=1}^{N_n} 1\{T_n < T_{n,b,i}\}. \tag{10}$$

It is clear from this definition that p_{uc} is the smallest α at which the null hypothesis would not be rejected by subsampling test (7). When the null hypothesis is true, p_{uc} is good measure of the probability of a test statistic being as extreme or more extreme than what was observed. However, when the alternative hypothesis is true, p_{uc} is less adequate. In this situation, the subsampling distribution of the $T_{n,b,i}$ has its center of location at $\tau_b t(P)$ which tends to infinity as $\tau_b \to \infty$. Despite this divergence, the subsampling test (7) has been proven to be consistent, so that p_{uc} tends to zero under the alternative hypothesis. However, this convergence does not occur at an optimal rate.

To intuitively see why, recall the construction of the data-centered subsampling test (8) where the subsampled test statistics are centered at t_n. Since t_n is a consistent estimate of $t(P)$ under either hypothesis, this forces the subsampling distribution $L_{n,b}(x)$ to have its location centered around zero whether or not the null hypothesis is true. For this reason, for the remainder of the paper we focus on the centered subsampling test (8), and investigate its p-value given by

$$p_c = N_n^{-1} \sum_{i=1}^{N_n} 1\{T_n < \tau_b(t_{n,b,i} - t_n)\}. \tag{11}$$

Since the subsampling distribution $L_{n,b}(x)$ stays centered, under the alternative we typically have $t_{n,b,i} - t_n < t_{n,b,i}$, and therefore $p_c < p_{uc}$; this will be made precise in Theorem 1 below.

Remark 3. Because of the above discussion, the p-value defined as p_c in equation (11) is the recommended one for use with subsampling and similarly with b/n bootstrap tests. Note that this notion of p-value is valid *even* in the full-sample size bootstrap case, i.e., bootstrap with $b = n$, when the latter has been shown to be valid in the problem at hand.

Before providing our formal results, we first introduce a final assumption on the dependence structure of the data and the way the subsamples are constructed.

Assumption 2. The data satisfy one of the following two dependence conditions, and the subsampling scheme is chosen accordingly, as discussed at the beginning of Section 2.1.

1. The data $X_1, X_2, \ldots, X_n$ are i.i.d., taking values in a sample space S with common probability law P.

2. $X_1, X_2, \ldots, X_n$ is a sample of stationary observations taking values in a sample space S. Denote the probability law governing the infinite, stationary sequence $\ldots, X_{-1}, X_0, X_1, \ldots$ by P. Further assume that $\alpha_X(m) \to 0$ as $m \to \infty$, where $\alpha_X(\cdot)$ is the α-mixing sequence corresponding to $\{X_t\}$ (see Rosenblatt, 1956).

Theorem 1. *Let b satisfy (3) with k as discussed in Remark 1, and let τ_b and τ_n satisfy (4). Then, under Assumptions 1–2, if $P \in \mathbf{P_1}$ the centered p-value, p_c given by equation (11) is not bigger than p_{uc} given by equation (10) with probability tending to one, and if the test statistic t_n is nonnegative with probability 1, then $p_c \leq p_{uc}$ with probability 1 for all $P \in \mathbf{P}$, whether $P \in \mathbf{P_0}$ or $P \in \mathbf{P_1}$ (under the null and alternative hypotheses).*

Theorem 1 provides formal justification for our claim that the data-centered subsampling test (8) is more powerful than the uncentered test (7). The recommendation of not explicitly imposing the null hypothesis on the test statistic in question has also been noted in the related framework of bootstrap hypothesis testing in regression (Paparoditis and Politis, 2005).

The next theorem establishes that, as desired, the centered p-value has a limiting uniform distribution under the null hypothesis, and that $p_c \to_P 0$ under the alternative.

Theorem 2. *Under the conditions of Theorem 1,*

i. if $P \in \mathbf{P_0}$, then $Prob_P\{p_c \leq x\} \to_P x$ for $0 < x < 1$.

ii. if $P \in \mathbf{P_1}$, then $p_c \to_P 0$.

4 Simulations

We ran five simulation experiments in order to test our methods. In each case, we calibrated the subsampling procedure by choosing a block size b which achieved the desired $\alpha = 0.05$ under the null hypothesis for the centered version of the test; see Chapter 9 in Politis et al. (1999) for further discussion. Each test was repeated 1000 times.

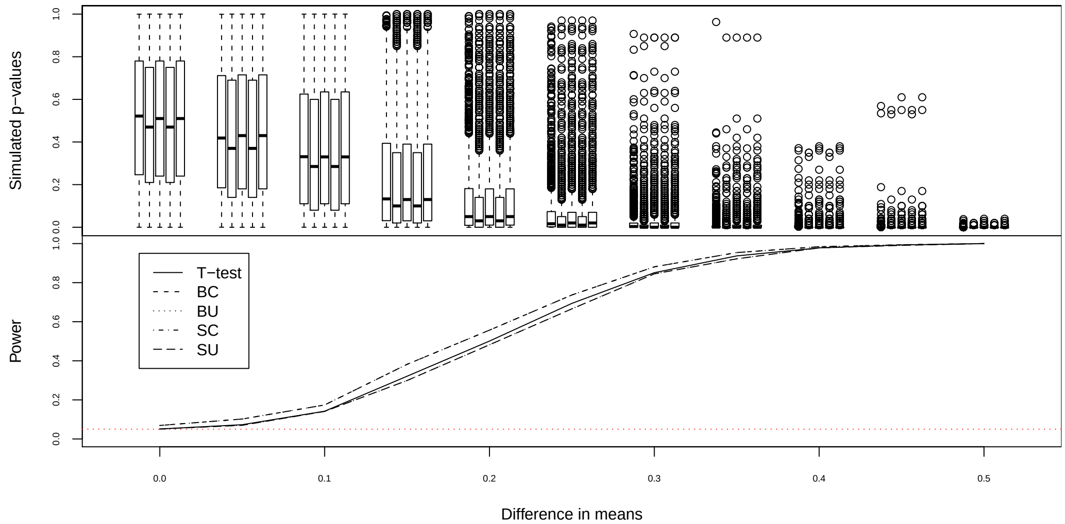

Figure 1: One sample tests for the mean of normal data.

The results of each experiment are presented as a two panel plot. The bottom panel shows empirical power curves, and the top panel shows boxplots of the simulated p-values for each of five types of hypothesis tests at each considered true value of the parameter in question. Each cluster of five box plots shows the distributions of p-values at one value of the parameter. The boxes within a cluster correspond to the five following tests in order from left to right: a classical test; the centered b/n bootstrap; uncentered b/n bootstrap; centered subsampling; and finally uncentered subsampling. The x-axis position of the middle box in each cluster indicates the true value of the parameter.

Experiment 1 (t-test): For the first experiment, we simulated $n = 100$ observations from a $N(\mu, 1)$ distribution for $\mu = 0.0, 0.05, \ldots, 0.5$. We tested the hypothesis $H_0 : \mu = 0$ versus the alternative $H_1 : \mu \neq 0$ using the test statistic $t_n = |\bar{X}|$, the absolute value of the sample mean. The chosen block size was $b = 1$; in this setting a constant block size will provide asymptotically correct results because the limiting sampling distribution and the finite sample sampling distribution are the same. Results are shown in Figure 1. The centered subsampling and bootstrap tests are slightly biased towards rejecting the null hypothesis, but are extremely competitive with the classical, and optimal, t-test. The two uncentered tests are only slightly less powerful than the t-test.

Experiment 2 (χ^2 independence test): In the second simulation, we ran a simple test for independence on a 2×2 table with sample size $n = 100$. Pairs of Bernoulli random variables (X, Y) were generated from the model defined by $P(X = 1) = 1/2$, and $P(Y = 1|X = 1) = p_y$ and $P(Y = 1|X = 0) = 1 - p_y$, so the null hypothesis of independence is satisfied for $p_y = 1/2$ and not satisfied otherwise. In simulation we considered values $p_y = 0.5, 0.525, \ldots, 0.75$. The test statistic was taken to be $t_n = X^2/n$, where X^2 is the standard χ^2 test statistic. The results are compared with the classical χ^2 test. In this case, the subsampling tests appear more powerful than their bootstrap counterparts, and the centered subsampling test may

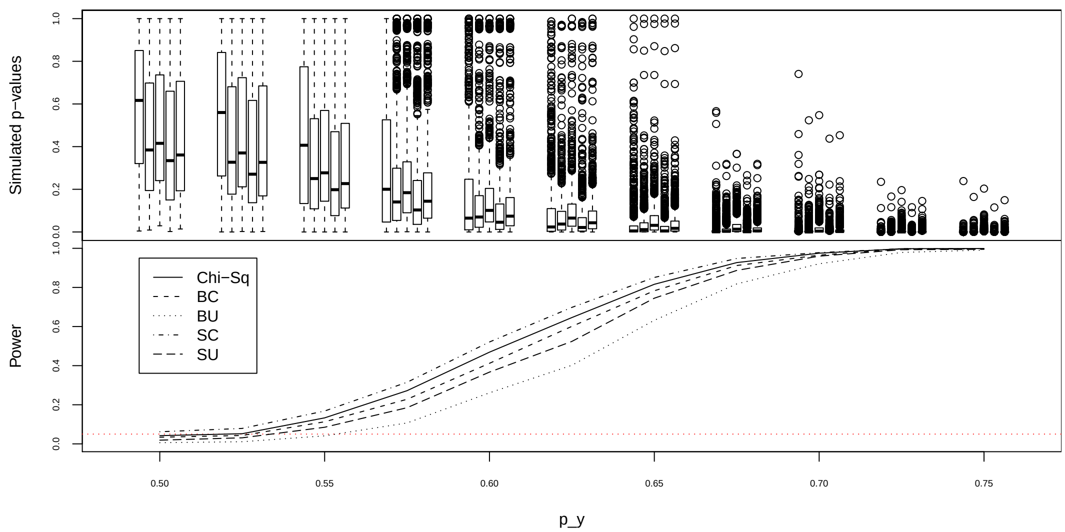

Figure 2: Tests for independence using the χ^2 test statistic.

be somewhat more powerful than the classical test (which is slightly biased under H_0); see Figure 2.

Experiment 3 (regression coefficient): For the third experiment, we generated $n = 100$ pairs (X, Y) with X distributed $N(0, 1)$ and $Y = \beta X + \epsilon$, with ϵ independent of X and having a t-distribution on 5 degrees of freedom. We tested whether the slope of the fitted linear regression was statistically significant by subsampling/bootstrapping pairs (X_i, Y_i). The simulation was run for $\beta = 0.0, 0.05, \ldots, 0.5$, and we used the test statistic $t_n = |\hat{\beta}|$, where $\hat{\beta}$ is the slope of the linear regression fitted with intercept. The results are shown in Figure 3. The centered subsampling and bootstrap tests are as powerful as the classical alternative.

Experiment 4 (Wilcoxon test): For the fourth experiment we generated $n = 100$ observations from a Cauchy distribution and performed a two sided test for the null hypothesis that the location of the median was 0 based on Wilcoxon's signed rank test statistic, W. We used the test statistic $t_n = |W/[n(n+1)/4]-1|$, which asymptotically satisfies the limiting distribution assumptions with rate $\tau_n = n^{1/2}$ (see DasGupta, 2008, Chapter 24). The simulations were run with true simulated medians $0.0, 0.1, \ldots, 1.0$. The two centered tests showed no loss in efficiency as compared with the Wilcoxon procedure. The subsampled test statistic t_b has a discrete sampling distribution on a relatively small number of values; this discreteness is clearly visible in the p-value boxplots in Figure 4.

Experiment 5 (Kolmogorov-Smirnov): In the final experiment we generated $n = 100$ observations from the mixture distribution

$$f_X(x) = \left(\frac{1}{2}\phi(x-\mu) + \frac{1}{2}\phi(x+\mu)\right),$$

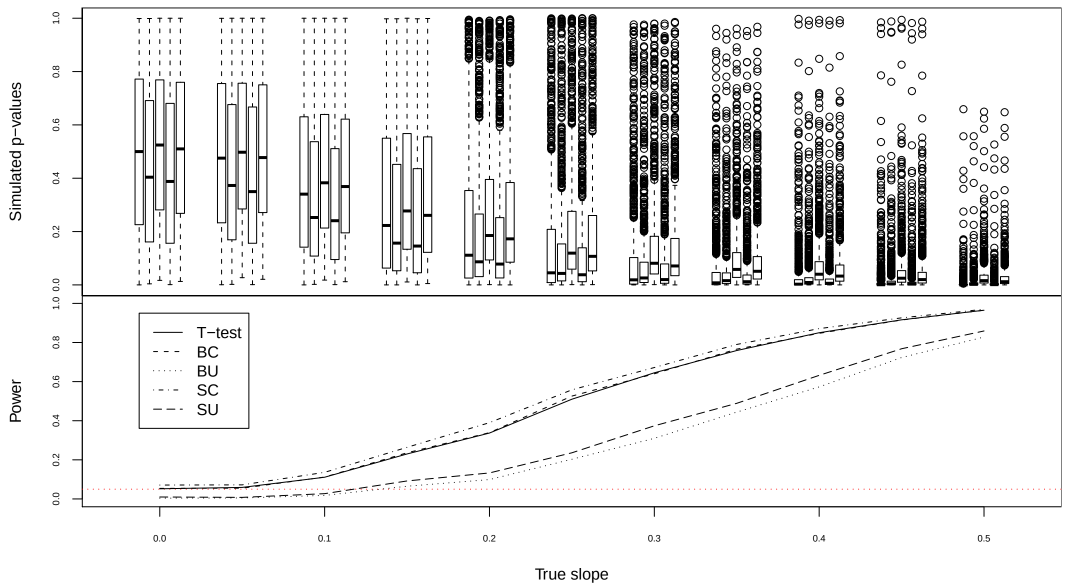

Figure 3: Tests for the significance of a regression coefficient.

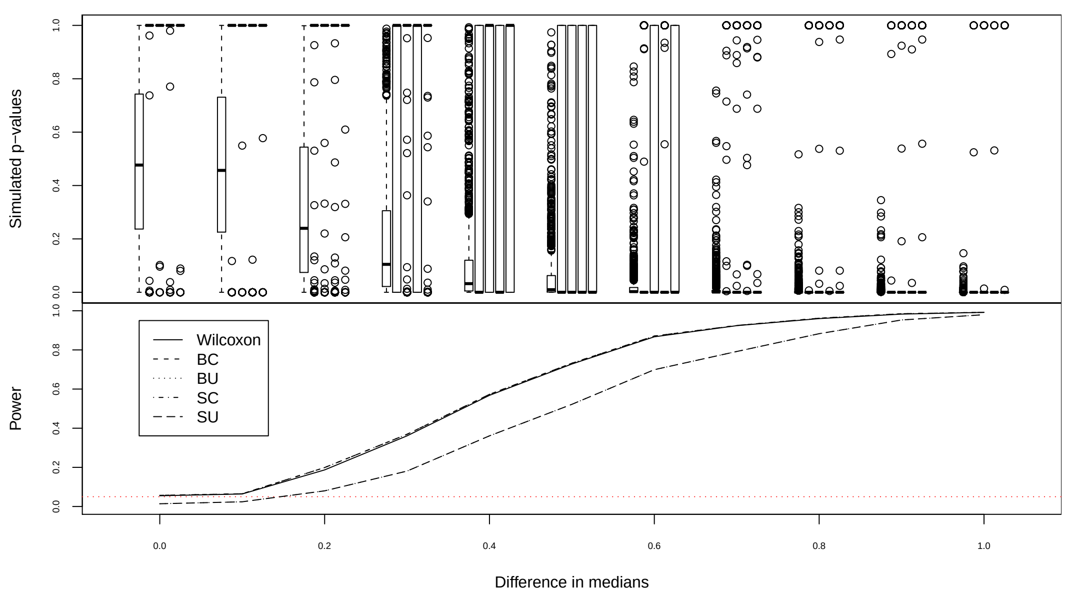

Figure 4: Signed rank test on Cauchy data.

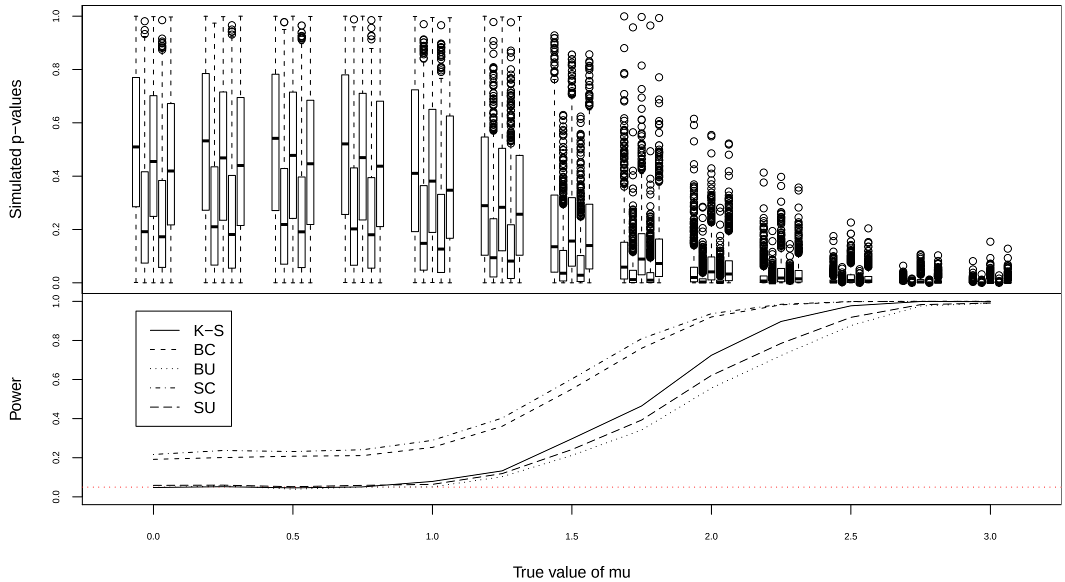

Figure 5: Kolmogorov-Smirnov tests on mean 0 variance 1 mixtures of normal data.

where ϕ is the standard normal density, and we then scaled the data by $(1+\mu^2)^{-1/2}$ so that the mean and variance were always 0 and 1 respectively. We considered $\mu = 0.0, 0.25, \ldots, 3.0$. We then tested the hypothesis that the data were $N(0,1)$ using the Kolmogorov-Smirnov test statistic $t_n = \sup_x |F_n(x) - \Phi(x)|$. This is the only test we ran in which subsampling performed poorly compared with the classical alternative. No choice of block size produced simulated α for the centered tests close to the desired $\alpha = 0.05$, so we chose the block size to give the uncentered tests the desired α. The centered tests were quite powerful but heavily biased. The uncentered tests were somewhat less powerful than the classical test.

5 Technical proofs

Proof of Theorem 1: Recall that under the alternative hypothesis, $t_n \to_P t(P) > 0$. Therefore, when the alternative hypothesis is true, $t_{n,b,i} > t_{n,b,i} - t_n$ with probability approaching one. The consequence is that, with probability approaching one,

$$1\{T_n < \tau_b(t_{n,b,i} - t_n)\} \leq 1\{T_n < T_{n,b,i}\}.$$

Hence,

$$p_c = N_n^{-1} \sum_{i=1}^{N_n} 1\{T_n < \tau_b(t_{n,b,i} - t_n)\} \leq N_n^{-1} \sum_{i=1}^{N_n} 1\{T_n < T_{n,b,i}\} = p_{uc}.$$

If $t_n \geq 0$ with probability one, then the above inequalities hold with probability one under both the null and alternative hypotheses. $\square$

Proof of Theorem 2: The proofs of both parts rely on Theorems 2.2.1 and 3.2.1 in Politis et al. (1999), which establish consistency of subsampling confidence intervals for i.i.d. and mixing data, and proves uniform convergence in probability of $L_{n,b}(x)$ to $J_n(x,P)$. We begin with part (i).

By equations (11) and (1), $p_c = 1 - L_{n,b}(T_n)$. Let E_n be the event that $\sup_x |L_{n,b}(x) - J_n(x,P)| < \epsilon$. By Theorem 2.2.1 and Theorem 3.2.1 in Politis et al. (1999), E_n has probability tending to one.

So, with probability tending to one,

$$1 - J_n(T_n, P) - \epsilon < p_c < 1 - J_n(T_n, P) + \epsilon.$$

By the above, we can estimate the distribution of p by from the distribution of $1 - J_n(T_n, P)$. Since $J_n(x, P)$ is the distribution of T_n and since it itself has a continuous limit law $J(x, P)$, $J_n(T_n, P)$ tends to the uniform distribution on $[0, 1]$. Therefore, with probability tending to one,

$$\begin{array}{rcccl} Prob_P\{1 - J_n(T_n, P) + \epsilon < x\} & \leq & Prob_P\{p_c < x\} & \leq & Prob_P\{1 - J_n(T_n, P) - \epsilon < x\} \\ x - \epsilon & \leq & Prob_P\{p_c < x\} & \leq & x + \epsilon. \end{array}$$

To prove (ii), recall that p_{uc} has the property of being the smallest α at which the null hypothesis would not be rejected using the uncentered test. By Theorems 2.6.1 and 3.5.1 in Politis et al. (1999), the uncentered subsampling test rejects the null hypothesis with probability approaching one at any fixed $\alpha > 0$. Therefore, $p_{uc} \to_P 0$. By Theorem 1, $p_{uc} < p_c$ with probability approaching one, so the result follows. □

References

R. Beran. Simulated power functions. *The Annals of Statistics*, 14(1):151–173, 1986.

P.J. Bickel, F. Götze, and W.R. van Zwet. Resampling fewer than n observations: gains, losses, and remedies for losses. *Statistica Sinica*, 7:1–32, 1997.

I. Choi. Subsampling vector autoregressive tests of linear constraints. *Journal of Econometrics*, 124(1):55–89, 2005.

I. Choi and T.K. Chue. Subsampling hypothesis tests for nonstationary panels with applications to exchange rates and stock prices. *Journal of Applied Econometrics*, 22(2), 2007.

A. DasGupta. *Asymptotic theory of statistics and probability*. Springer Verlag, New York, 2008.

A.C. Davison and D.V. Hinkley. *Bootstrap methods and their application*. Cambridge Univ Pr, 1997.

M.A. Delgado, J.M. Rodrıguez-Poo, and M. Wolf. Subsampling inference in cube root asymptotics with an application to Manski's maximum score estimator. *Economics Letters*, 73(2):241–250, 2001.

J. Gonzalo and M. Wolf. Subsampling inference in threshold autoregressive models. *Journal of Econometrics*, 127(2):201–224, 2005.

E.L. Lehmann and J.P. Romano. *Testing Statistical Hypotheses*. Springer Verlag, 2005.

O. Linton, E. Maasoumi, and Y.J. Whang. Consistent testing for stochastic dominance under general sampling schemes. *The Review of Economic Studies*, 72(3):735–765, 2005.

R.Y. Liu and K. Singh. Notions of limiting P values based on data depth and bootstrap. *Journal of the American Statistical Association*, 92(437):266–277, 1997.

W.Y. Loh. A new method for testing separate families of hypotheses. *Journal of the American Statistical Association*, 80(390):362–368, 1985.

E. Paparoditis and D.N. Politis. Bootstrap hypothesis testing in regression models. *Statistics and Probability Letters*, 74(4):356–365, 2005.

D.N. Politis and J.P. Romano. Estimating the distribution of a Studentized statistic by subsampling. *Bulletin of the International Statistical Institute, 49th Session*, 2:315–316, 1993.

D.N. Politis and J.P. Romano. Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics*, 22(4):2031–2050, 1994.

D.N. Politis and J.P. Romano. Subsampling for econometric models – comments on bootstrapping time series models. *Econometric Rev.*, 15(2):169–176, 1996.

D.N. Politis and J.P. Romano. K-sample Subsampling. *Functional and Operational Statistics, Springer Verlag, Heidelberg*, pages 247–254, 2008a.

D.N. Politis and J.P. Romano. K-sample subsampling in general spaces: The case of independent time series. *Journal of Multivariate Analysis*, 2008b.

D.N. Politis, J.P. Romano, and M. Wolf. *Subsampling.* Springer Verlag, 1999.

D.N. Politis, J.P. Romano, and M. Wolf. On the asymptotic theory of subsampling. *Statistica Sinica*, 11(4):1105–1124, 2001.

J.P. Romano. A bootstrap revival of some nonparametric distance tests. *Journal of the American Statistical Association*, 83(403):698–708, 1988.

M. Rosenblatt. A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America*, 42(1):43–47, 1956.