First International Workshop on
Functional and Operatorial Statistics.
Toulouse, June 19-21, 2008

K-sample Subsampling

Dimitris N. Politis* and Joseph P. Romano

University of California—San Diego and Stanford University
dpolitis@ucsd.edu and romano@stanford.edu

Abstract

The problem of subsampling in two-sample and K-sample settings is addressed where both the data and the statistics of interest take values in general spaces. We show the asymptotic validity of subsampling confidence intervals and hypothesis tests in the case of independent samples, and give a comparison to the bootstrap in the K-sample setting.

1 Introduction

Subsampling is a statistical method that is most generally valid for nonparametric inference in a large-sample setting. The applications of subsampling are numerous starting from i.i.d. data and regression, and continuing to time series, random fields, marked point processes, etc. ; see Politis, Romano and Wolf (1999) for a review and list of references.

Interestingly, the two-sample and K-sample settings have not been explored yet in the subsampling literature ; we attempt to fill this gap here. So, consider K independent datasets : $\underline{X}^{(1)}, \ldots, \underline{X}^{(K)}$ where $\underline{X}^{(k)} = (X_1^{(k)}, \ldots, X_{n_k}^{(k)})$ for $k = 1, \ldots, K$. The random variables $X_j^{(k)}$ take values in an arbitrary space[1] $\mathbf{S}$; typically, $\mathbf{S}$ would be $\mathbf{R}^d$ for some d, but $\mathbf{S}$ can very well be a function space. Although the dataset $\underline{X}^{(k)}$ is independent of $\underline{X}^{(k')}$ for $k \neq k'$, there may exist some dependence *within* a dataset. For conciseness, we will focus on the case of independence within samples here ; the general case will be treated in a follow-up bigger exposition.

Thus, in the sequel we will assume that, for any k, $X_1^{(k)}, \ldots, X_{n_k}^{(k)}$ are i.i.d. An example in the i.i.d. case is the usual two-sample set-up in biostatistics where d 'features' (body characteristics, gene expressions, etc.) are measured on a group of patients, and then again measured on a control group. The probability law associated with such a K-sample experiment is specified by $P = (P_1, \ldots, P_K)$, where P_k is the underlying probability of the kth sample ; more formally, the joint distribution of all the observations is the product measure $\prod_{k=1}^{K} P_k^{n_k}$. The goal is inference (confidence regions, hypothesis tests, etc.)

[1] Actually, one can let S vary with k as well, but we do not pursue this here for lack of space.

regarding some parameter $\theta = \theta(P)$ that takes values in a general normed linear space $\mathbf{B}$ with norm denoted by $||\cdot||$. Denote $\mathbf{n} = (n_1, \ldots, n_K)$, and let $\hat{\theta}_{\mathbf{n}} = \hat{\theta}_{\mathbf{n}}(\underline{X}^{(1)}, \ldots, \underline{X}^{(K)})$ be an estimator of θ. It will be assumed that $\hat{\theta}_{\mathbf{n}}$ is consistent as $\min_k n_k \to \infty$. In general, one could also consider the case where the number of samples K tends to ∞ as well.

Let $g : \mathbf{B} \to \mathbf{R}$ be a continuous function, and let $J_{\mathbf{n}}(P)$ denote the sampling distribution of the "root" $g[\tau_{\mathbf{n}}(\hat{\theta}_{\mathbf{n}} - \theta(P))]$ under P, with corresponding cumulative distribution function

$$J_{\mathbf{n}}(x, P) = Prob_P\{g[\tau_{\mathbf{n}}(\hat{\theta}_{\mathbf{n}} - \theta(P))] \leq x\} \tag{1}$$

where $\tau_{\mathbf{n}}$ is a normalizing sequence; in particular, $\tau_{\mathbf{n}}$ is to be thought of as a fixed function of $\mathbf{n}$ such that $\tau_{\mathbf{n}} \to \infty$ when $\min_k n_k \to \infty$. As an example, $g(\cdot)$ might be a continuous function of the norm $||\cdot||$ or a projection operator.

As in the one-sample case, the basic assumption that is required for subsampling to work is existence of a *bona fide* large-sample distribution, i.e.,

Assumption 1.1 *There exists a nondegenerate limiting law $J(P)$ such that $J_{\mathbf{n}}(P)$ converges weakly to $J(P)$ as $\min_k n_k \to \infty$.*

The α–quantile of $J(P)$ will be denoted by $J^{-1}(\alpha, P) = \inf\{x : J(x, P) \geq \alpha\}$. In addition to Assumption 1.1, we will use the following mild assumption.

Assumption 1.2 *As $\min_k n_k \to \infty$, $\tau_{\mathbf{b}}\|\hat{\theta}_{\mathbf{n}} - \theta(P)\| = o_P(1)$.*

Assumptions 1.1 and 1.2 are implied by the following assumption, as long as $\tau_{\mathbf{b}}/\tau_{\mathbf{n}} \to 0$.

Assumption 1.3 *As $\min_k n_k \to \infty$, the distribution of $\tau_{\mathbf{n}}(\hat{\theta}_{\mathbf{n}} - \theta(P))$ under P converges weakly to some distribution (on the Borel σ-field of the normed linear space $\mathbf{B}$).*

Here, weak convergence is understood to be taken in the modern sense of Hoffmann-Jorgensen; see Section 1.3 of van der Vaart and Wellner (1996). That Assumption 1.3 implies both Assumptions 1.1 and 1.2 follows by the Continuous Mapping Theorem; see Theorem 1.3.6 of van der Vaart and Wellner (1996).

2 Subsampling hypothesis tests in K samples

For $k = 1, \ldots, K$, let $\mathcal{S}_k$ denote the set of all size b_k (unordered) subsets of the dataset $\{X_1^{(k)}, \ldots, X_{n_k}^{(k)}\}$ where b_k is an integer in $[1, n_k]$. Note that the set $\mathcal{S}_k$ contains $Q_k = \binom{n_k}{b_k}$ elements that are enumerated as $S_1^{(k)}, S_2^{(k)}, \ldots, S_{Q_k}^{(k)}$. A K–fold subsample is then constructed by choosing one element from each super-set $\mathcal{S}_k$ for $k = 1, \ldots, K$. Thus, a typical K–fold subsample has the form : $S_{i_1}^{(1)}, S_{i_2}^{(2)}, \ldots, S_{i_K}^{(K)}$, where i_k is an integer in $[1, Q_k]$ for $k = 1, \ldots, K$. It is apparent that the number of possible K–fold subsamples is $Q = \prod_{k=1}^{K} Q_k$. So a subsample value of the general statistic $\hat{\theta}_{\mathbf{n}}$ is

$$\hat{\theta}_{\mathbf{i},\mathbf{b}} = \hat{\theta}_{\mathbf{b}}(S_{i_1}^{(1)}, \ldots, S_{i_K}^{(K)}) \tag{2}$$

where $\mathbf{b} = (b_1, \ldots, b_K)$ and $\mathbf{i} = (i_1, \ldots, i_K)$. The subsampling distribution approximation to $J_{\mathbf{n}}(P)$ is defined by

$$L_{\mathbf{n},\mathbf{b}}(x) = \frac{1}{Q} \sum_{i_1=1}^{Q_1} \sum_{i_2=1}^{Q_2} \cdots \sum_{i_K=1}^{Q_K} 1\{g[\tau_{\mathbf{b}}(\hat{\theta}_{\mathbf{i},\mathbf{b}} - \hat{\theta}_{\mathbf{n}})] \leq x\}. \tag{3}$$

The distribution $L_{\mathbf{n},\mathbf{b}}(x)$ is useful for the construction of subsampling confidence sets as discussed in Section 3. For hypothesis testing, however, we instead let

$$G_{\mathbf{n},\mathbf{b}}(x) = \frac{1}{Q} \sum_{i_1=1}^{Q_1} \sum_{i_2=1}^{Q_2} \cdots \sum_{i_K=1}^{Q_K} 1\{g[\tau_{\mathbf{b}}(\hat{\theta}_{\mathbf{i},\mathbf{b}})] \le x\}, \tag{4}$$

and consider the general problem of testing a null hypothesis H_0 that $P = (P_1, \ldots, P_k) \in \mathbf{P}_0$ against H_1 that $P \in \mathbf{P}_1$. The goal is to construct an asymptotically valid null distribution based on some test statistic of the form $g(\tau_{\mathbf{n}}\hat{\theta}_{\mathbf{n}})$, whose distribution under P is defined to be $G_{\mathbf{n}}(P)$ (with c.d.f. $G_n(\cdot, P)$). The subsampling critical value is obtained as the $1-\alpha$ quantile of $G_{\mathbf{n},\mathbf{b}}(\cdot)$, denoted $g_{\mathbf{n},\mathbf{b}}(1-\alpha)$. We will make use of the following assumption.

Assumption 2.1 *If $P \in \mathbf{P}_0$, there exists a nondegenerate limiting law $G(P)$ such that $G_{\mathbf{n}}(P)$ converges weakly to $G(P)$ as $\min_k n_k \to \infty$.*

Let $G(\cdot, P)$ denote the c.d.f. corresponding to $G(P)$. Let $G^{-1}(1-\alpha, P)$ denote a $1-\alpha$ quantile of $G(P)$. The following result gives the consistency of the procedure under H_0, and under a sequence of contiguous alternatives; for the definition contiguity see Section 12.3 of Lehmann and Romano (2007). One could also obtain a simple consistency result under fixed alternatives.

Theorem 2.1 *Suppose Assumption 2.1 holds. Also, assume that, for each $k = 1, \ldots, K$, we have $b_k/n_k \to 0$, and $b_k \to \infty$ as $\min_k n_k \to \infty$.*
(i) Assume $P \in \mathbf{P}_0$. If $G(\cdot, P)$ is continuous at its $1-\alpha$ quantile $G^{-1}(1-\alpha, P)$, then

$$g_{\mathbf{n},\mathbf{b}} \xrightarrow{P} G^{-1}(1-\alpha, P) \tag{5}$$

and

$$Prob_P\{g(\tau_{\mathbf{n}}\hat{\theta}_{\mathbf{n}}) > g_{\mathbf{n},\mathbf{b}}(1-\alpha)\} \to \alpha \quad \text{as } \min_k n_k \to \infty \ . \tag{6}$$

(ii) Suppose, that for some $P = (P_1, \ldots, P_k) \in \mathbf{P}_0$, $P_{k,n_k}^{n_k}$ is contiguous to $P_k^{n_k}$ for $k = 1, \ldots K$. Then, under such a contiguous sequence, $g(\tau_{\mathbf{n}}\hat{\theta}_{\mathbf{n}})$ is tight. Moreover, if it converges in distribution to some random variable T and $G(\cdot, P)$ is continuous at $G^{-1}(1-\alpha, P)$, then the limiting power of the test against such a sequence is $P\{T > G^{-1}(1-\alpha, P)\}$.

Proof : To prove (i), let x be a continuity point of $G(\cdot, P)$. We claim

$$G_{\mathbf{n},\mathbf{b}}(x) \xrightarrow{P} G(x, P) \ . \tag{7}$$

To see why, note that $E[G_{\mathbf{n},\mathbf{b}}(x)] = G_{\mathbf{b}}(x, P) \to G(x, P)$. So, by Chebychev's inequality, to show (7), it suffices to show $Var[G_{\mathbf{n},\mathbf{b}}(x)] \to 0$. To do this, let $d = d_{\mathbf{n}}$ be the greatest integer $\le \min_k(n_k/b_k)$. Then, for $j = 1, \ldots, d$, let $\bar{\theta}_{j,\mathbf{b}}$ be equal to the statistic $\hat{\theta}_{\mathbf{b}}$ evaluated at the data set where the observations from the kth sample are $(X^{(k)}_{b_k(j-1)+1}, X^{(k)}_{b_k(j-1)+2}, \ldots, X^{(k)}_{b_k(j-1)+b_k})$. Then, set

$$\bar{G}_{\mathbf{n},\mathbf{b}}(x) = d^{-1} \sum_{j=1}^{d} 1\{g(\tau_{\mathbf{b}}\bar{\theta}_{j,\mathbf{b}}) \le x\} \ .$$

By construction, $\bar{G}_{\mathbf{n},\mathbf{b}}(x)$ is an average of i.i.d. 0–1 random variables with expectation $G_{\mathbf{b}}(x,P)$ and variance that is bounded above by $1/(4d_{\mathbf{n}}) \to 0$. But, $G_{\mathbf{n},\mathbf{b}}(x)$ has smaller variance than $\bar{G}_{\mathbf{n},\mathbf{b}}(x)$. This last statement follows by a sufficiency argument from the Rao-Blackwell Theorem; indeed,

$$G_{\mathbf{n},\mathbf{b}}(x) = E[\bar{G}_{\mathbf{n},\mathbf{b}}(x) | \hat{P}^{(k)}_{n_k},\ k = 1,\ldots,K] \ ,$$

where $\hat{P}^{(k)}_{n_k}$ is the empirical measure in the kth sample. Since these empirical measures are sufficient, it follows that

$$Var(G_{n,b}(x)) \leq Var(\bar{G}_{n,b}(x)) \to 0 \ .$$

Thus, (7) holds. Then, (5) follows by Lemma 11.2.1(ii) of Lehmann and Romano (2005). Application of Slutsky's Theorem yields (6).

To prove (ii), we know that $g_{\mathbf{n},\mathbf{b}} \xrightarrow{P} G^{-1}(1-\alpha, P)$ under P. Contiguity forces the same convergence under the sequence of contiguous alternatives. The result follows by Slutsky's Theorem. $\diamond$

3 Subsampling confidence sets in K samples

Let $c_{\mathbf{n},\mathbf{b}}(1-\alpha) = \inf\{x : L_{\mathbf{n},\mathbf{b}}(x) \geq 1-\alpha\}$ where $L_{\mathbf{n},\mathbf{b}}(x)$ was defined in (3).

Theorem 3.1 *Assume Assumptions 1.1 and 1.2, where g is assumed uniformly continuous. Also assume that, for each $k = 1,\ldots,K$, we have $b_k/n_k \to 0$, $\tau_{\mathbf{b}}/\tau_{\mathbf{n}} \to 0$, and $b_k \to \infty$ as $\min_k n_k \to \infty$.*
(i) Then, $L_{\mathbf{n},\mathbf{b}}(x) \xrightarrow{P} J(x,P)$ for all continuity points x of $J(\cdot,P)$.
(ii) If $J(\cdot,P)$ is continuous at $J^{-1}(1-\alpha,P)$, then the event

$$\{g[\tau_{\mathbf{n}}(\hat{\theta}_{\mathbf{n}} - \theta(P))] \leq c_{\mathbf{n},\mathbf{b}}(1-\alpha)\} \tag{8}$$

has asymptotic probability equal to $1-\alpha$; therefore, the confidence set $\{\theta :\ g[\tau_{\mathbf{n}}(\hat{\theta}_{\mathbf{n}} - \theta)] \leq c_{\mathbf{n},\mathbf{b}}(1-\alpha)\}$ has asymptotic coverage probability $1-\alpha$.

Proof : Assume without loss of generality that $\theta(P) = 0$ (in which case $J_{\mathbf{n}}(P) = G_{\mathbf{n}}(P)$). Let x be a continuity point of $J(\cdot,P)$. First, we claim that

$$L_{\mathbf{n},\mathbf{b}}(x) - G_{\mathbf{n},\mathbf{b}}(x) \xrightarrow{P} 0 \ . \tag{9}$$

Given $\epsilon > 0$, there exists $\delta > 0$, so that $|g(x) - g(x')| < \epsilon$ if $\|x - x'\| < \delta$. But then, $|g[\tau_{\mathbf{b}}(\hat{\theta}_{\mathbf{i},\mathbf{b}} - \hat{\theta}_{\mathbf{n}})] - g(\tau_{\mathbf{b}}\hat{\theta}_{\mathbf{i},\mathbf{b}})| < \epsilon$ if $\|\tau_{\mathbf{b}}\hat{\theta}_{\mathbf{n}}\| < \delta$; this latter event has probability tending to one. It follows that, for any fixed $\epsilon > 0$,

$$G_{\mathbf{n},\mathbf{b}}(x-\epsilon) \leq L_{\mathbf{n},\mathbf{b}}(x) \leq G_{\mathbf{n},\mathbf{b}}(x+\epsilon)$$

with probability tending to one. But, the behavior of $G_{\mathbf{n},\mathbf{b}}(x)$ was given in Theorem 2.1. Letting $\epsilon \to 0$ through continuity points of $J(\cdot,P)$ yields (9) and (i). Part (ii) follows from Slutsky's Theorem. $\diamond$

Remark. The uniform continuity assumption for g can be weakened to continuity if Assumptions 1.1 and 1.2 are replaced by Assumption 1.3. However, the proof is much more complicated and relies on a K-sample version of Theorem 7.2.1 of Politis, Romano and Wolf (1999).

In general, we may also try to approximate the distribution of a studentized root of the form $g(\tau_{\mathbf{n}}[\hat{\theta}_{\mathbf{n}} - \theta(P)]/\hat{\sigma}_{\mathbf{n}}$, where $\hat{\sigma}_{\mathbf{n}}$ is some estimator which tends in probability to some finite nonzero constant $\sigma(P)$. The subsampling approximation to this distribution is

$$L^{+}_{\mathbf{n},\mathbf{b}}(x) = \frac{1}{Q}\sum_{i_1=1}^{Q_1}\sum_{i_2=1}^{Q_2}\cdots\sum_{i_K=1}^{Q_K} 1\{g[\tau_{\mathbf{b}}(\hat{\theta}_{\mathbf{i},\mathbf{b}} - \hat{\theta}_{\mathbf{n}})]/\hat{\sigma}_{\mathbf{i},\mathbf{b}} \leq x\}, \tag{10}$$

where $\hat{\sigma}_{\mathbf{i},\mathbf{b}}$ is the estimator $\hat{\sigma}_{\mathbf{b}}$ computed from the $\mathbf{i}$th subsampled data set. Also let $c^{+}_{\mathbf{n},\mathbf{b}}(1-\alpha) = \inf\{x : L^{+}_{\mathbf{n},\mathbf{b}}(x) \geq 1-\alpha\}$.

Theorem 3.2 *Assume Assumptions 1.1 and 1.2, where g is assumed uniformly continuous. Let $\hat{\sigma}_{\mathbf{n}}$ satisfy $\hat{\sigma}_{\mathbf{n}} \xrightarrow{P} \sigma(P) > 0$. Also assume that, for each $k = 1, \ldots, K$, we have $b_k/n_k \to 0$, $\tau_{\mathbf{b}}/\tau_{\mathbf{n}} \to 0$, and $b_k \to \infty$ as $\min_k n_k \to \infty$.*
(i) Then, $L_{\mathbf{n},\mathbf{b}}(x) \xrightarrow{P} J(x \cdot \sigma(P), P)$ if $J(\cdot, P)$ is continuous at $x\sigma(P)$.
(ii) If $J(\cdot, P)$ is continuous at $J^{-1}(1-\alpha, P)/\sigma(P)$, then the event

$$\{g[\tau_{\mathbf{n}}(\hat{\theta}_{\mathbf{n}} - \theta(P))]/\hat{\sigma}_{\mathbf{n}} \leq c^{+}_{\mathbf{n},\mathbf{b}}(1-\alpha)\} \tag{11}$$

has asymptotic probability equal to $1-\alpha$.

4 Random subsamples and the K–sample bootstrap

For large values of n_k and b_k, $Q = \prod_k \binom{n_k}{b_k}$ can be a prohibitively large number ; considering *all* possible subsamples may be impractical and, thus, we may resort to Monte Carlo. To define the algorithm for generating random subsamples of sizes $b_1, \ldots, b_K$ respectively, recall that subsampling in the i.i.d. single-sample case is tantamount to sampling *without* replacement from the original dataset ; see e.g. Politis et al. (1999, Ch. 2.3). Thus, for $m = 1, \ldots, M$, we can generate the mth joint subsample as $\underline{X}^{(1)}_m, \underline{X}^{(2)}_m, \ldots, \underline{X}^{(K)}_m$ where $\underline{X}^{(k)}_m = \{X^{(k)}_{I_1}, \ldots, X^{(k)}_{I_{b_k}}\}$, and $I_1, \ldots, I_{b_k}$ are b_k numbers drawn randomly *without* replacement from the index set $\{1, 2, \ldots, n_k\}$. Note that the random indices drawn to generate $\underline{X}^{(k)}_m$ are independent to those drawn to generate $\underline{X}^{(k')}_m$ for $k \neq k'$.

Thus, a randomly chosen subsample value of the statistic $\hat{\theta}_{\mathbf{n}}$ is given by $\hat{\theta}_{m,\mathbf{b}} = \hat{\theta}_{\mathbf{b}}(\underline{X}^{(1)}_m, \ldots, \underline{X}^{(K)}_m)$, with corresponding subsampling distribution defined as

$$\tilde{L}_{\mathbf{n},\mathbf{b}}(x) = \frac{1}{M}\sum_{m=1}^{M} 1\{g[\tau_{\mathbf{b}}(\hat{\theta}_{m,\mathbf{b}} - \hat{\theta}_{\mathbf{n}})] \leq x\}. \tag{12}$$

The following corollary shows that $\tilde{L}_{\mathbf{n},\mathbf{b}}(x)$ and its $1-\alpha$ quantile $\tilde{c}_{\mathbf{n},\mathbf{b}}(1-\alpha)$ can be used for the construction of large-sample confidence regions for θ ; its proof is analogous to the proof of Corollary 2.1 of Politis and Romano (1994).

Corollary 4.1 *Assume the conditions of Theorem 3.1. As $M \to \infty$, parts (i) and (ii) of Theorem 3.1 remain valid with $\tilde{L}_{\mathbf{n},\mathbf{b}}(x)$ and $\tilde{c}_{\mathbf{n},\mathbf{b}}(1-\alpha)$ instead of $L_{\mathbf{n},\mathbf{b}}(x)$ and $c_{\mathbf{n},\mathbf{b}}(1-\alpha)$.*

Similarly, hypothesis testing can be conducted using the notion of random subsamples. To describe it, let $\tilde{g}_{\mathbf{n},\mathbf{b}}(1-\alpha) = \inf\{x : \tilde{G}_{\mathbf{n},\mathbf{b}}(x) \geq 1-\alpha\}$ where

$$\tilde{G}_{\mathbf{n},\mathbf{b}}(x) = \frac{1}{M}\sum_{m=1}^{M} 1\{g[\tau_{\mathbf{b}}(\hat{\theta}_{m,\mathbf{b}})] \leq x\}. \tag{13}$$

Corollary 4.2 *Assume the conditions of Theorem 2.1. As $M \to \infty$, parts (i) and (ii) of Theorem 2.1 remain valid with $\tilde{G}_{\mathbf{n},\mathbf{b}}(x)$ and $\tilde{g}_{\mathbf{n},\mathbf{b}}(1-\alpha)$ instead of $G_{\mathbf{n},\mathbf{b}}(x)$ and $g_{\mathbf{n},\mathbf{b}}(1-\alpha)$.*

The bootstrap in two-sample settings is often used in practical work ; see Hall and Martin (1988) or van der Vaart and Wellner (1996). In the i.i.d. set-up, resampling and (random) subsampling are very closely related since, as mentioned, they are tantamount to sampling *with* vs. *without* replacement from the given i.i.d. sample. By contrast to subsampling, however, no general validity theorem is available for the bootstrap *unless* a smaller resample size is used ; see Politis and Romano (1993).

Actually, the general validity of K-sample bootstrap that uses a resample size b_k for sample k follows from the general validity of subsampling as long as $b_k^2 << n_k$. To state it, let $J^*_{\mathbf{n},\mathbf{b}}(x)$ denote the bootstrap (pseudo-empirical) distribution of $g[\tau_{\mathbf{b}}(\hat{\theta}^*_{\mathbf{n},\mathbf{b}} - \hat{\theta}_{\mathbf{n}})]$ where $\hat{\theta}^*_{\mathbf{n},\mathbf{b}}$ is the statistic $\hat{\theta}_{\mathbf{b}}$ computed from the bootstrap data. Similarly, let $c^*_{\mathbf{n},\mathbf{b}}(1-\alpha) = \inf\{x : J^*_{\mathbf{n},\mathbf{b}}(x) \geq 1-\alpha\}$. The proof of the following corollary parallels the discussion in Section 2.3 of Politis et al. (1999).

Corollary 4.3 *Under the additional condition $b_k^2/n_k \to 0$ for all k, Theorem 3.1 is valid as stated with $J^*_{\mathbf{n},\mathbf{b}}(x)$ and $c^*_{\mathbf{n},\mathbf{b}}(1-\alpha)$ in place of $L_{\mathbf{n},\mathbf{b}}(x)$ and $c_{\mathbf{n},\mathbf{b}}(1-\alpha)$ respectively.*

Références

[1] Hall, P. and Martin, M. (1988). On the bootstrap and two-sample problems, *Australian Journal of Statistics*, vol. 30A, 179-192.

[2] Lehmann, E.L. and Romano, J. (2005). *Testing Statistical Hypotheses*, 3rd edition, Springer, New York.

[3] Politis, D.N., and Romano, J.P. (1993). Estimating the Distribution of a Studentized Statistic by Subsampling, in *Bulletin of the International Statistical Institute, 49th Session,* Firenze, August 25 - September 2, 1993, Book 2, pp.315-316.

[4] Politis, D.N., and Romano, J.P. (1994). Large sample confidence regions based on subsamples under minimal assumptions, *Ann. Statist.*, vol. 22, 2031-2050.

[5] Politis, D., Romano, J. and Wolf, M. (1999). *Subsampling,* Springer, New York.

[6] van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes*, Springer, New York.