Nonparametric Detection Using Empirical Distributions and Bootstrapping

Martin Gölz[†], Visa Koivunen[†] and Abdelhak Zoubir[⋆]
Aalto University[†], Finland and TU Darmstadt[⋆], Germany
Email: {visa.koivunen}@aalto.fi

***Abstract*—This paper addresses the problem of decision making when there is no or very vague knowledge about the probability models associated with the hypotheses. Such scenarios occur for example in Internet of Things (IoT), environmental surveillance and data analytics. The probability models are learned from the data by empirical distributions that provide an accurate approximation of the true model. Hence, the approach is fully nonparametric. The bootstrap method is employed to approximate the distribution of the decision statistic. The actual test is based on the Anderson-Darling test that is shown to perform reliably even if the empirical distributions differ only slightly. The proposed detector allows controlling Type I and II error levels without specifying explicit probability models or performing tedious large sample analysis. It is also proved that the test can achieve the specified power. Numerical simulations validate the results.**

I. Introduction

In a variety of signal processing applications and tasks, one needs to make decisions at a desired error level. Prime examples include radar, wireless communication, IoT, environmental surveillance, biomedical applications as well as data analytics. Decision making is commonly formulated as a binary hypothesis testing problem, and a decision about the presence or absence of a signal is made based on an observed data set, usually recorded by a sensor. A test statistic is computed and compared to a threshold value to choose between two hypotheses. To achieve the desired performance and find appropriate thresholds, explicit assumptions on the underlying probability models are made. Unfortunately, these assumptions on probability models may not be valid in practice, the models may be very complicated or it may not be feasible to specify an explicit model. Furthermore, large sample analysis establishing asymptotic Gaussianity is often needed to obtain quantitative information about the performance of the designed inference method. One may avoid many of these problems by resorting to approximate models or nonparametric techniques instead. Empirical distributions obtained via bootstrapping can be used as a basis for reliable statistical inference. This approach allows for controlling the Type I and Type II error levels in decision making as well as for producing information about confidence intervals, bias and variances. The inference can be completely nonparametric since the required empirical distributions can be learned from the data. This is particularly important in applications where the observations may be made in very different operational environments and data in different sensors might obey different probability models.

In [1], a variety of examples of bootstrap applications in signal processing is given, including a hypothesis test for a parameter being smaller or larger than a given threshold. A parametric detection scheme using the bootstrap to obtain a test statistic's distribution is introduced in [2]. A method of combining sequential analysis and bootstrapping for parametric signal detection in cognitive radio is provided in [3]. An overview of nonparametric bootstrap distance tests from a statistics point of view can be found in [4].

In this paper, a fully nonparametric hypothesis testing method is proposed. The method approximates the underlying probability models using bootstrapping. It assumes the availability of some training data when there is only noise and interference present. Consequently, the distribution of the test statistic under hypothesis H_0 can be approximated using the empirical distribution obtained by bootstrapping. The detection task at hand is to decide whether the observed data comes from the same or a different distribution. The proposed test compares the empirical cumulative distribution functions (EDFs). The actual test is based on the Anderson-Darling [5] test that performs highly reliably even if the maximum difference between the distributions is not large but they differ slightly in many places, including the tails. The proposed method allows for achieving the desired detection performance without specifying explicit probability models or performing tedious large sample analysis. Simulations demonstrate its capability to distinguish between different distributions and that it achieves the specified power. In addition, we demonstrate that only a rather small amount of observed data is required to come to reliable decisions. Potential applications include hypothesis testing in an IoT and environmental sensing, where establishing explicit probability models for massive numbers of sensors is not feasible, but at the same time one needs decision making algorithms that fulfill the desired performance criteria in terms of Type I and Type II error probabilities.

This paper is organized as follows. In Section II, the system model for hypothesis testing is provided. A brief description of the employed bootstrapping method is given as well. In Section III, the proposed bootstrapped detection algorithm and Anderson-Darling test are described in detail. The test statistic is derived and the asymptotic validity of the proposed test is proved. Finally, in Section IV, simulation results demonstrating the capability to distinguish between a variety of distributions and achieving the desired error levels in detection are provided.

ISBN 978-0-9928626-7-1 © EURASIP 2017

II. System model

In this section, we describe the system model and give a brief overview on the bootstrap method and how it is used to approximate distributions for hypothesis testing. The detection task is formulated as a binary hypothesis testing problem with the following two hypotheses:

- H_0: The data comes from the nominal distribution.
- H_1: The data comes from some other distribution

In this paper the nominal distribution, instead of being known, is approximated using bootstrapping to learn from the data and form empirical distributions. The classical detection tasks with only noise under H_0 and a present signal under H_1 are special cases of this model, in particular if assuming a random signal.

A binary detector declares which of the two hypotheses is in place. Commonly this is done by comparing a test statistic τ to a threshold value. The proposed algorithm uses a nonparametric test statistic τ from two samples X and Y. The value of τ is compared to its distribution under H_0. If the value is *significant*, i.e., sufficiently far at the tail of the empirical distribution under H_0, we accept H_1, otherwise we accept H_0. In this paper, the distribution of τ under H_0 is not obtained by making explicit assumptions about the distribution but by applying the bootstrap to training data. The training data should obey the nominal distribution such that there is only ambient noise and interference present. The EDF of τ is then approximated using the bootstrap without making any assumptions on underlying probability models. Hence, our approach is completely non-parametric.

Our binary hypothesis test is based on two data sets (samples) X and Y. X is called *ambient sample* and Y *observation sample*. Data set X is *reference* or *training data* acquired when there is no signal present. The observation sample Y contains data recorded when the detector is in operational use with the goal to detect the presence or absence of a signal.

The empirical distribution under H_0 is learned as follows. In the training phase, the sensor is exposed to ambient noise in its location when there is no signal present. During this period, the sensor records a *training data set* I, following a distribution F. The training phase is carried out only once, if the ambient noise is stationary. Changes in the noise statistics may require retraining and new approximation of empirical and test statistic distributions. When the sensor is in actual operational use performing a detection task, we randomly sample X from I.

When recording the observation sample Y, we want to determine if the data comes from a different distribution. In the absence of a signal, the underlying distribution G of X is similar to F, since again only noise is being recorded. Our hypothesis testing problem can hence be reformulated as

- H_0: Both, X and Y, follow the same distribution $F = G$.
- H_1: X and Y follow different distributions $F \neq G$.

Any hypothesis testing procedure is subject to errors described in terms of *probability of false alarm* $\alpha = p_{fa}$, and *probability of missed detection* p_{md}. The probability of detection, i.e. the power of the test, is thus $p_d = 1 - p_{md}$. In our approach, the decision is based on a nonparametric test statistic τ. Its value is computed exclusively from the data in X and Y, without any explicit assumptions on the underlying distributions. To finally come to a decision, we evaluate

$$|\tau| < q_{1-\alpha}, \tag{1}$$

where $q_{1-\alpha}$ denotes the $(1-\alpha)$-quantile obtained from the theoretical distribution of τ under H_0, using the probability of false alarm α. If (1) holds, H_0 is accepted, otherwise, H_1 is chosen. This provides a probability of $1-\alpha$ of detecting H_0 when it is true. Since the theoretical distribution of the test statistic τ is not known and and no distributional assumptions are made, we approximate it using nonparametric bootstrapping and empirical distributions. Furthermore, the theoretical quantile value in (1) cannot be used. Instead, its empirical counterpart $q^*_{1-\alpha}$, obtained from the empirical distribution is employed. The bootstrap allows for finding such percentile points with high accuracy. It is important to ensure that one is still able to achieve the desired performance levels in terms of p_{md} and $\alpha = p_{fa}$. The proof that the specified error probabilities still hold is given in Section III-C.

Even though the proposed detector uses empirical quantities, it is an asymptotically valid algorithm satisfying specified error probabilities for large sample sizes. For smaller samples however, it may be subject to empirical inaccuracies. It is desirable to design algorithms that perform reliably in the face of inaccuracies even for small sample sizes. Two guidelines in [6] allow for reducing these inaccuracies when implementing a bootstrapped hypothesis test. First the importance of resampling assuming H_0 to achieve a small p_{md} is emphasized. The second guideline, which only provides small improvements according to the authors, recommends to use a test statistic that is free from the influence of its variance in order to achieve a better accuracy. When describing algorithm and results, we point out how these guidelines are followed.

A. The Bootstrap

The bootstrap [7] is a powerful statistical method that allows for obtaining distributions of an estimator of interest θ by approximating them with empirical distributions. The lack of knowledge on the probability models is compensated by computational capabilities. The idea is to resample with replacement a large number B of replica data sets from the original data set. The resampled data sets are of the same size as the original data set. Finally, statistics of interest are computed using each bootstrap replica data set. The bootstrapped distribution of the desired statistic is given by its B bootstrap values. A detailed introduction to the implementation of the bootstrap in signal processing is provided in [1].

III. Bootstrapped detection algorithm

A. Test statistic

In this section, we propose three nonparametric test statistics for the bootstrapped empirical distributions: The Kolmogorov-Smirnov test statistic (KS), the Cramér-von Mises test statistic (CvM) and the Anderson-Darling test statistic (AD). Their formulas are provided for the special case considered in our algorithm where the distributions of two samples are compared.

The widely known KS statistic evaluates the maximum difference between two distributions. For our problem of a two sample test, we have to compare the EDF F_n of sample X to G_m, the EDF of observed data Y. This results in

$$\tau_{\mathrm{KS}_2} = \sup|F(x) - G(x)| \approx \sup|F_n(x) - G_m(x)|, \qquad (2)$$

where F and G represent the CDFs of samples X and Y, respectively, which are approximated by the samples' EDFs F_n and G_m. The KS statistic is formed by finding the maximum distance between the EDFs. Hence, it is expected to provide good results if there is high discrepancy between the values of the CDFs $F(x)$ and $G(x)$ for some x. If there is no significant peak difference between the CDFs, the KS test may lack power.

An alternative test statistic is given in [8] with the goal of capturing more subtle differences between distributions. The two sample version of the CvM-criterion compares EDFs F_n and G_m, by

$$\tau_{\mathrm{CvM}_2} = \frac{nm}{n+m}\int_{-\infty}^{\infty}[F_n(x) - G_m(x)]^2 dH_{n+m}(x), \qquad (3)$$

with $H_{n+m} = \frac{n}{n+m}F_n(x) + \frac{m}{n+m}G_m(x)$ being the EDF of the combination of X and Y. Instead of finding a maximum, we integrate the squared deviations of the distributions. The test is thus less sensitive to the peak difference in the distributions. It rather considers the total sum of squared deviations between the EDFs and is expected to thereby provide higher power than KS tests.

The third test statistic is the AD statistic, proposed in [5]. It introduces a weighting function $\psi(x) \geq 0$. In the case of $\psi(x) = 1$, the AD and the CvM statistic are similar. The two sample version is found in [9] and given by

$$\tau_{\mathrm{AD}_2} = \frac{nm}{n+m}\int_{-\infty}^{\infty}\psi(x)[F_n(x) - G_m(x)]^2 dH_{n+m}(x). \qquad (4)$$

$\psi(x)$ allows for emphasizing certain parts of the distribution, for example the tails, in quantifying the differences between the distributions. For smaller sample sizes there are usually few observations from the distribution's tails. Consequently, the differences between the tails of the EDFs are not necessarily captured unless they are emphasized. In this paper,

$$\psi(x) = \frac{1}{H_{n+m}(x)(1 - H_{n+m}(x))},$$

which was also tested in [5]. It represents the reciprocal of the variance of $\sqrt{n}[F_n(x) - G_m(x)]$ and thereby puts higher weight to the tails of the distributions than the other statistics considered in this paper. As in (3), H_{n+m} denotes the EDF of the combination of X and Y.

B. Detailed Description of the Algorithm

The training phase has to be carried out once before using the sensor. During training, data is assumed to originate from the unknown distribution corresponding to H_0. During the actual detection task, only the observation and the decision phase need to be executed.

During the training phase, besides recording the training data sample I of length $s >> n$, we also bootstrap the distribution of the test statistic τ under H_0. Thus, the bootstrap needs to be applied only once during the whole process, as long as the ambient noise statistics do not change.

We resample a large number of B replica data sets from I and split each resample into a bootstrap representation of X and Y. Since H_0 is assumed to be true, Y, X and I follow the same distribution F. Assuming H_0 when bootstrapping fulfill s guideline 1 in [6]. Then, we compute τ for every bootstrap representation of X and Y, obtaining in total B values of the test statistic. Those form the empirical distribution of τ under H_0, which is used later in the decision phase. We set the percentile $q^*_{1-\alpha}$ to the $\lceil(1-\alpha)\cdot B\rceil$ largest value of the bootstrapped test statistic.

After completing the training, the sensor is used for the actual detection task. Hence, it records the observation sample Y of length m. Based on Y, we want to determine if the distribution of the data has changed or if the observed data obeys the same distribution as the training data. Note that it is desirable to choose m as small as possible in order to keep the process of recording observation data as short as possible.

To decide between H_0 and H_1, we randomly sample X of size $n << s$ from the training data sample and compute the test statistic τ. Finally, if Eq. (1) holds, we decide in favor of H_0, otherwise, the alternate hypothesis H_1 is accepted. In other words, we accept H_1 if τ differs sufficiently from the value suggested by the empirical distribution of τ under H_0.

The proposed bootstrapped detector is summarized in Algorithm 1.

C. Validity and Consistency

In this section, we prove that the proposed bootstrap detection method approximates the threshold $q_{1-\alpha}$ in (1).

Proposition 1: An approximation is valid if the probability of τ crossing the approximating quantile $q^*_{1-\alpha}$ asymptotically reaches the theoretical probability of missed detection α, i. e.

$$P_{F,G}\{|\tau| > q^*_{1-\alpha}\} \to \alpha, \text{ as } \min\{n,m\} \to \infty. \qquad (5)$$

Proof 1: Let us use the definition of $S_n(x)$ in [4],

$$S_n(x) = \sqrt{n}[F_n(x) - F(x)], \qquad (6)$$

with F denominating the theoretical underlying CDF and F_n its empirical approximation. Since we do not want to decide if a sample's EDF is similar to an assumed theoretical CDF, but compare the EDFs of two observed samples, we replace F by the EDF $G_m(x)$ of the second sample and obtain

$$S_{n_2}(x) = \sqrt{n}[F_n(x) - G_m(x)]. \qquad (7)$$

As shown in [4], (5) holds for a KS-type of test statistic. Additionally, it is stated in the second last paragraph of Section 1 in [4] that the proof remains valid for continuous functions of $S_n(x)$. The CvM statistic, as defined in (3), is a squared and integrated version of $S_{n_2}(x)$. AD is a

ISBN 978-0-9928626-7-1 © EURASIP 2017

Algorithm 1 Bootstrapped Signal Detector

Training Phase

Step 1.) Obtain a training datasample $I = I_1, \ldots, I_s$ of length s from an unknown distribution function F.

Step 2.) Resample a large amount B of samples $Z^{*b}, b = 1, \ldots, B$ of length $n+m$, from I.

for $b = 1, \ldots, B$ **do**

 Step 3.) Define $X^{b*} = Z_1^{b*}, \ldots Z_n^{b*}$ and $Y^{b*} = Z_{n+1}^{b*}, \ldots, Z_{n+m}^{b*}$.

 Step 4.) Obtain the test statistic τ^{b*} using the bootstrap samples X^{b*} and Y^{b*}.

end for

Step 5.) Obtain the the empirical $(1-\alpha)$-quantile $q_{1-\alpha}^*$ from the empirical distribution of τ.

Observation Phase

Step 6.) Obtain an observation sample $Y = Y_1, \ldots, Y_m$ of length m from an unknown distribution function G.

Decision Phase

Step 7.) Sample the initialization sample $X = X_1, \ldots, X_n$ of length $n << s$ from I.

Step 8.) Compute test statistic τ for X and Y.

if $|\tau| < q_{1-\alpha}^*$ **then**

 Step 9.) Accept H_0.

else

 Step 9.) Accept H_1.

end if

weighted version of CvM, in our case using the continuous weight $\psi(x) = \frac{1}{H_{n+m}(x)(1-H_{n+m}(x))}$. Hence, both CvM and AD are continuous functions of $S_{n_2}(x)$. By showing that CvM and AD are continuous functions of $S_{n_2}(x)$, the two sample version of $S_n(x)$, the proof in [4] is also valid for these two statistics. Hence, Proposition 1 holds for all three considered test statistics.

Consequently, our devised algorithm can be used with all three test statistics. Furthermore, in Section 2.5 of [4], the consistency of the procedure is considered. For a consistent test, p_{md} converges to zero under H_1, when sample length grows. According to [4], $P_1\{\tau > C^\alpha\} = 1$ as $n \to \infty$, hence the procedure is consistent.

IV. Simulation Results

We tested the proposed algorithm for a variety of scenarios and distributions. In this section, some of the results demonstrating its performance are presented and pros and cons are studied. We compare the results for different test statistics and recommend which one to use.

We select $\alpha = 5\%$ for all scenarios. The goal is to correctly accept H_0 very close to or above 95% in the case that the data for both samples originate from the same distribution. Since we use an empirical data set, slight deviations from the desired 95% are allowed and all those results are accepted as being exact that do not deviate from the desired level by more than 1%. Hence, under H_0, an acceptance level of at least 94% has to be achieved. On the other hand, a low p_{md} is of interest. Hence, when data sets X and Y obey different distributions, the acceptance rate for H_0 should be very low. The aim is to keep it below 5% in our simulations. We set the number of bootstrap resamples $B = 1000$ and use sample lengths of $n = 500$ and $m = 100$, as long as not stated otherwise.

A. Choice of Test Statistic

The detection performance in distinguishing between different distributions is studied: Gaussian against Exponential distributions and Rice against Rayleigh distributions. The parameters of the distributions are chosen to shape the CDFs as similar as possible, as illustrated in Figure 1. Parameter values are given in the parentheses after distribution names, i.e., $\mathcal{N}(1,1)$ denotes a Gaussian distribution with $\mu = \sigma = 1$.

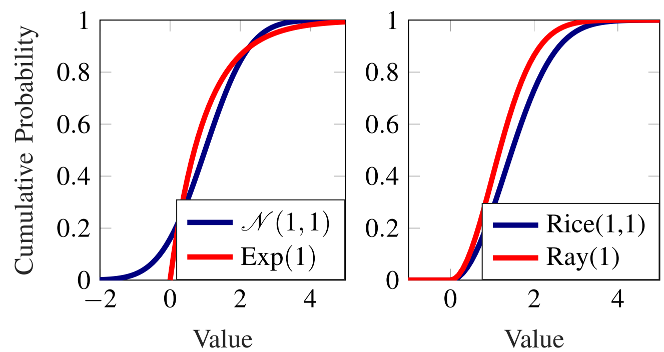

Fig. 1. The CDFs used in the numerical simulations. The parameters are chosen such that the distributions have similar shapes.

The first two columns of Table I name the tested distributions. F denotes the underlying CDF of the larger ambient sample X, whereas G refers to the CDF of the smaller observation sample Y. The remaining columns give the acceptance rates of H_0 for the different test statistics: Kolmogorov-Smirnov, Cramér-von Mises and Anderson-Darling.

In the first four lines in Table I, Exponential and Gaussian distributions are compared. In lines one and two, X and Y follow the same distributions, hence, an acceptance rate of 95% should be reached. AD meets the required rate for both cases, the other two τ provide values close to the desired percentage. However, we note that the acceptance level when using CvM in line two falls off by more than 1%.

The results for testing Rice against Rayleigh in lines five to eight confirm the previous observations: In lines five and six, $F = G$ and, again, AD achieves the desired rates. This time, in line six, KS falls off by more than 1%.

Finally, it remains to look at the cases when X and Y are drawn from different distributions. In lines three and four, both, KS and AD test statistic, deliver good probability of missed detection levels. The AD test, however, provides higher power in both cases. The CvM test statistic falls off and fails to provide the desired power level. Lines seven and eight present the results for Rice and Rayleigh distributions. Again, AD provides low levels of p_{md}. The results for CvM are close to acceptable and KS is inaccurate.

ISBN 978-0-9928626-7-1 © EURASIP 2017

TABLE I
PERCENTAGE OF ACCEPTANCE OF H_0 FOR $n = 500$ AND $m = 100$. THE AD TEST PERFORMS THE BEST OVERALL.

F	G	KS	CvM	AD
Exp., $\lambda = 1$	Exp., $\lambda = 1$	94.28%	94.75%	95.65%
Gaus., $\mu = \sigma = 1$	Gaus., $\mu = \sigma = 1$	95%	93.8%	95.94%
Exp., $\lambda = 1$	Gaus., $\mu = \sigma = 1$	5.16%	13.69%	0.09%
Gaus., $\mu = \sigma = 1$	Exp., $\lambda = 1$	4.98%	14.65%	4.3%
Rice, $\nu = 1, \sigma = 1$	Rice, $\nu = 1, \sigma = 1$	95.42%	95.08%	95.56%
Ray., $\sigma = 1$	Ray., $\sigma = 1$	93.74%	94.71	95.53%
Rice, $\nu = 1, \sigma = 1$	Ray., $\sigma = 1$	9.73%	6.67%	5.89%
Ray., $\sigma = 1$	Rice $\nu = 1, \sigma = 1$	9.3%	5.94%	3.74%

In all considered cases, also those not included in this paper, AD provides either the best or close to the best results in terms of acceptance rate and missed detection level compared to the other tested statistics. For some examples, CvM shows better behavior than KS, for others, KS provides better values.

Since the values produced by AD are either better than or close to the best provided by the other statistics, using AD is recommended for our algorithm.

Finally, the weighting function of the Anderson-Darling statistic can explain the high accuracy of acceptance rate when H_0 is true. While KS and CvM sometimes fail to satisfy the required 95%-level of acceptance in Table I, AD always achieves the required level of correctness. Having chosen $\psi(x) = \frac{1}{H_{n+m}(x)(1-H_{n+m}(x))}$, we normalize S_{n_2} in (7) to form the test statistic. This fulfills guideline 2 for bootstrapped hypothesis tests in [6], which leads to a higher accuracy in terms of level of acceptance.

B. Problems with Laplace against Gaussian Distributions

Distinguishing Laplace and Gaussian distributions is considered. P_d is very low, however, we are able to present a solution on how to fix the lacking power of the procedure.

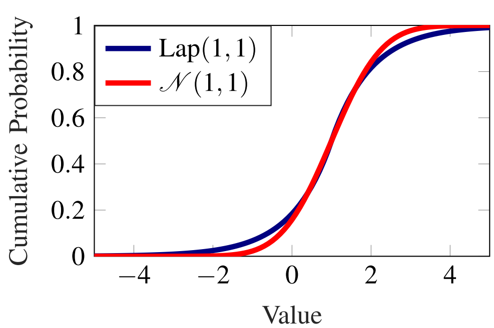

Fig. 2. Laplace and Gaussian CDFs. Both are very similar with a slight difference in the tails.

Also for both samples following a Laplace distribution, the desired α level is clearly met, as the simulations show. Unfortunately, since the CDFs are very similar (see Figure 2), the level of p_{md} is almost always above 90% for KS, CvM and AD. Apparently, the information provided by the observed samples X and Y is not sufficient.

TABLE II
PROBABILITY OF ACCEPTING H_0 WITH INCREASED SAMPLE SIZES $n = 2500$ AND $m = 500$. ONLY AD PROVIDES SATISFYING RESULTS.

F	G	KS	CvM	AD
Lap., $\mu = 1 = 1$	Lap., $\mu = b = 1$	93.49%	95.95%	94.49%
Lap., $\mu = b = 1$	Gaus., $\mu = \sigma = 1$	63.35%	76.89%	0.35%
Gaus., $\mu = \sigma = 1$	Lap., $\mu = b = 1$	63.42%	66.49%	1.58%

Hence, the sample lengths n and m need to be increased. To find out whether the proposed detector can at all deliver a powerful test for this example, we repeated the simulations with $n = 2500$ and $m = 500$. The results in Table II show very good properties for AD. The other statistics still fail to fulfill the desired power properties. This is to be expected, since the differences between Laplace and Gaussian are in the tails, as it can be seen in Figure 2. This again confirms our choice of AD as preferred τ. By increasing the sample length, the consistency of the method is restored.

V. CONCLUSION

In this paper an algorithm for fully non-parametric hypothesis testing using empirical cumulative distributions is proposed. No explicit assumptions on the probability models are made. Instead, bootstrapping is used to approximate the EDFs. The distributions of test statistics under hypothesis H_0 are learned from training data using the bootstrap. By using Anderson-Darling type test statistics on approximated EDFs, highly reliable performance is achieved, even if the distributions differ only slightly in their tails. We also proved that the proposed test achieves the desired error levels in decision making without relying on large sample analysis and Gaussianity. Simulations demonstrate the reliable performance of the proposed detection algorithm.

REFERENCES

[1] A. M. Zoubir and B. Boashash, "The bootstrap and its application in signal processing," *Signal Processing Magazine, IEEE*, vol. 15, pp. 56–76, Jan. 1998.
[2] H.-T. Ong and A. M. Zoubir, "Bootstrap-based detection of signals with unknown parameters in unspecified correlated interference," *IEEE Transactions on Signal Processing*, vol. 51, pp. 135–141, Jan 2003.
[3] F. Y. Suratman, *Spectrum Sensing in Cognitive Radio: Bootstrap and Sequential Detection Approaches*. PhD thesis, Technische Universität, Darmstadt, February 2014.
[4] J. P. Romano, "A bootstrap revival of some nonparametric distance tests," *Journal of the American Statistical Association*, vol. 83, no. 403, pp. 698–708, 1988.
[5] T. W. Anderson and D. A. Darling, "Asymptotic theory of certain goodness of fit criteria based on stochastic processes," *Ann. Math. Statist.*, vol. 23, pp. 193–212, 06 1952.
[6] P. Hall and S. R. Wilson, "Two guidelines for bootstrap hypothesis testing," *Biometrics*, vol. 47, no. 2, pp. 757–762, 1991.
[7] B. Efron, "Bootstrap methods: Another look at the jackknife," *Ann. Statist.*, vol. 7, pp. 1–26, 01 1979.
[8] T. W. Anderson, "On the distribution of the two-sample Cramer-von Mises criterion," *Ann. Math. Statist.*, vol. 33, pp. 1148–1159, 09 1962.
[9] A. N. Pettitt, "A two-sample Anderson–Darling rank statistic," *Biometrika*, vol. 63, no. 1, pp. 161–168, 1976.

ISBN 978-0-9928626-7-1 © EURASIP 2017