INFERENCE USING PHI-DIVERGENCE GOODNESS-OF-FIT TESTS

Nikhil Kundargi, Student Member, IEEE, and Ahmed Tewfik, Fellow, IEEE
University of Texas at Austin

***Abstract*—In this paper we study the inferential use of goodness of fit tests in a non-parametric setting. The utility of such tests will be demonstrated for the test case of spectrum sensing applications in cognitive radios. For the first time, we provide a comprehensive framework for decision fusion of a ensemble of goodness-of-fit testing procedures through an Ensemble Goodness-of-Fit test. Also, we introduce a generalized family of functionals and kernels called Φ-divergences which allow us to formulate goodness-of-fit tests that are parameterized by a single parameter s. The performance of these tests is simulated under gaussian and non-gaussian noise in a MIMO setting. We show that under uncertainty or non-gaussianity in the noise, the performance of non-parametric tests in general, and phi-divergence based goodness-of-fit tests in particular, is significantly superior to that of the energy detector with reduced implementation complexity. Especially important is the property that the false alarm rates of our proposed tests is maintained at a fixed level over a wide variation in the channel noise distributions.**

***Index Terms*—Goodness of Fit tests, Ensemble Tests, Phi Divergence, Spectrum Sensing, Non parametric Inference, Decision Fusion.**

I. Introduction

The problem of signal detection using statistical inference is conventionally treated as that of hypothesis testing for parametric models where the distributions are modeled using known and parametrized probability functions which usually belong to the exponential family of distributions. While such a framework has worked remarkably well in the past, recently, applications have arisen where the hypotheses to be tested are not well defined. Also, the robustness of parametric tests is often inferior to that of equivalent non parametric tests. The penalty incurred for this increase in robustness is a decrease in the power of the test. Goodness of Fit (GoF) tests are a particularly popular and robust class of inferential tests that have been popular for almost a century in the statistical community [1], [2]. Recently, there has been an effort to study the performance of these tests for hypothesis testing applications, and highly encouraging results have been obtained [3]–[5].

The goal of the spectrum sensing problem is to quickly detect if the channel under consideration is vacant and can be used for opportunistic transmission by the secondary user (SU) or if it is occupied by the primary user (PU). In [3], a fast and robust spectrum sensing scheme is proposed using the Kolomogorov-Smirnov (KS) goodness of fit test. In the presence of a Gaussian noise, the KS test has been shown to be highly robust to uncertainty in background noise estimation as compared to the energy detection (ED). Moreover, its performance is significantly better than the ED test in the presence of non-Gaussian noise also, where other spectrum sensing methods often fail. But the KS test can only claim the model that the noise has the estimated noise density can be rejected with a confidence level of α (usually set to 90 - 99.9 %). So, the test statistic can exceed the threshold when the assumed noise model is wrong as often happens in presence of impulsive non-Gaussian noise. Also, [4] recently proposed an Anderson Darling version of goodness of fit test for spectrum sensing but do not apply it to non-gaussian noise models. The problem of reliable detection of gray space transmit opportunities using goodness-of-fit and other non-parametric techniques has been previously studied by the authors in the context of Medium Access layer packet statistics [6]. It has recently been shown that Phi Divergences are the optimal formulations for goodness-of-fit testing [7], [8]. Many previous goodness-of-fit tests that have been proposed can be reduced to be specialized cases of the phi-divergence statistic. We will use these generalizations to come up with a powerful family of goodness of fit statistics.

In this paper, we propose the following a) A generalized framework based on phi-divergences. We show how Phi-divergence goodness-of-fit tests can be selectively tuned to a specific region of the density. b) Extensive simulation results of the performance of these tests under Gaussian and non-Gaussian (impulsive) noise, c) A novel decision fusion method based on the statistical nature of the p-value metric.

II. Inference Problem Formulation

Consider the scenario where a unlicensed cognitive radio is trying to detect the presence of a licensed primary user via spectrum sensing. We model the general case where both the primary and secondary users have multiple antennas. Specifically, the MIMO channel is created by M_T transmit antennas and M_R receive antennas. The disrete-time baseband MIMO channel with fading at a given cognitive radio is

$$\mathbf{y}[n] = \sum_{p=1}^{P} \sum_{l=0}^{L-1} \mathbf{H}_p[n,l] \mathbf{s}_p[n-l] = \mathbf{v}[n], \tag{1}$$

where $\mathbf{y}[n]$ is the received signal after sampling, P is the number of PUs transmitting over the sensed channel, the multipath delay in number of symbol intervals is L, $\mathbf{H}_p[n,l] \in \mathcal{C}^{M_T \times M_R}$ is the complex MIMO channel tap matrix. Also $\mathbf{s}_p[n-l] \in \mathcal{C}^{M_T}$ is the signal vector received at the cognitive radio antennas at time n and $\mathbf{v}[n] \in \mathcal{C}^{M_R}$ is the noise vector. For the special case of frequency flat fading with a block transmission of size T symbols, eq (1) simplifies to,

$$\mathbf{Y} = \sqrt{\frac{E_s}{M_T}} \mathbf{HS} + \mathbf{V}. \tag{2}$$

Here, $\mathbf{Y}$ is the $M_R \times T$ block of received signal vectors, E_s is the total average energy available at the transmitter over a single symbol period, $\mathbf{H}$ is the MIMO channel matrix. Note that the channel noise $\mathbf{V} \triangleq [\mathbf{v}[n], n = 1, ..., T]$ can take on arbitrary distributions.

The problem is structured such that the misclassification of a occupied channel as vacant is heavily penalized. This leads to the following formulation of the hypothesis testing inference problem.

$$\begin{aligned} &\mathcal{H}_0 : \text{Only background noise present} \\ &\mathcal{H}_1 : \text{Primary user signal + Noise present} \end{aligned} \tag{3}$$

Radio Frequency Interference (RFI) caused due to impulsive noise is a severe issue in modern spectrum sensing applications. The non-parametric nature of the godness-of-fit test becomes advantageous when there is such type of added RFI front end noise that is non-Gaussian in nature. The performance of conventional detection techniques degrades severely under such conditions, whereas we show that GoF tests are robustly able to handle such noise. We follow the bivariate Middleton Class A noise model as proposed in [9]–[11], which also explicitly accounts for the standard thermal noise through an additive gaussian component. Narrowband impulsive noise is modeled as a series of independent events that are identically distributed. Also, the in-phase and quadrature components are modeled as i.i.d.

III. Phi-Divergence based Goodness-of-Fit Tests

A. Goodness-of-Fit Procedures

A goodness-of-fit test is a procedure for testing how well a certain distribution fits a given observation [1], [2]. To be more specific, consider a continuous random variable $\mathbf{X}$ with distribution $\mathcal{F}(x)$ and let $X_1, X_2, ..., X_n$ be a random sample of independent and identically distributed random variables each following distribution $\mathcal{F}(x)$, with order statistics $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$. To implement a goodness-of-fit test, we modify eq (3) to

$$\begin{aligned} &\mathcal{H}_0 : \mathcal{F}(x) = \mathcal{F}_0(x) \text{ (Null hypothesis)}, \\ &\mathcal{H}_1 : \mathcal{F}(x) \neq \mathcal{F}_0(x) \text{ (Alternative hypothesis)}. \end{aligned} \tag{4}$$

Here $\mathcal{F}_0(x)$ is the hypothesized null distribution function to be tested. The alternative hypothesis is transformed into a composite hypothesis that is defined as the complement of the null hypothesis. The *Empirical Distribution Function* (edf) of $\mathbf{X}$ is defined as

$$\mathcal{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(X_i < x), \qquad -\infty < x < \infty \tag{5}$$

where $\mathbb{I}(.)$ is the indicator function that evaluates to 1 if the condition in the braces is true, and is 0 otherwise. Also, the probability integral transformation theorem is stated below. The edf as defined in eq (5) combined with the probability integral transformation theorem leads directly to a number of powerful goodness-of-fit tests. See [12] for a generalized proof of the theorem.

Theorem 1. *Probability Integral Transformation*
Let a random variable $\mathbf{X}$ have a distribution $\mathcal{F}(x)$. If $\mathcal{F}$ is continuous, the random variable Z produced by a transformation $Z = \mathcal{F}(X)$ has a uniform probability distribution over the interval $0 \leq z \leq 1$.

B. Phi-Divergences

In this section we introduce the framework of Phi-Divergences for the purpose of inference between two hypotheses. Let $\phi(x)$ be a convex function with domain $x \in [0, \infty)$ and range $\Re \cup \{\infty\}$. Then the supremum and integral versions of the Jager-Wellner Generalized Phi Divergences will be defined as

$$\mathcal{S}_n(\phi) = \sup_x K_\phi\big(\mathcal{F}_n(x), x\big), \tag{6}$$

$$\mathcal{T}_n(\phi) = \int_0^1 K_\phi\big(\mathcal{F}_n(x), x\big)\, dx. \tag{7}$$

For ease of computation, we restrict ourselves to the following special class of ϕ-functions using the parameter $s \in \Re$

$$\phi_s(x) = \begin{cases} [1 - x - sx - x^s]/[s(1-s)] & s \neq 0, 1, \\ x(\log x - 1) + 1 & s = 1, \\ \log(1/x) + x - 1 & s = 0. \end{cases} \tag{8}$$

Also, with the kernel function used in eq (6) defined as in eq (9), we get the family of ϕ_s-divergences.

$$K_s(u, v) = v\phi_s(u/v) + (1 - v)\phi_s([1 - u]/[1 - v]). \tag{9}$$

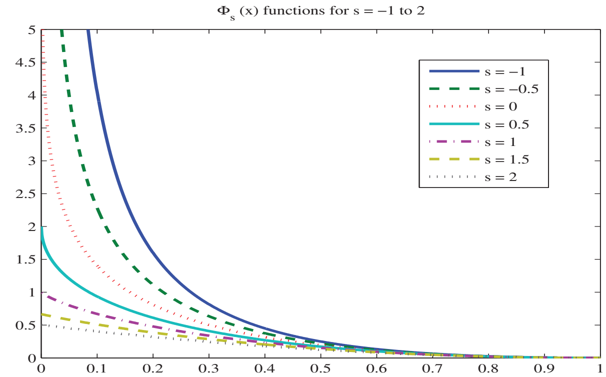

Fig. 1: Φ function plots

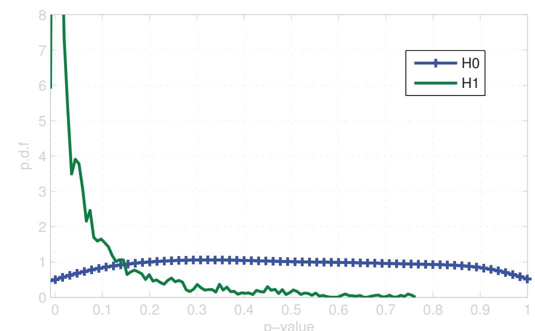

Fig. 2: Distribution of p-values under H_1 and H_0 for SNR = -2 dB, test size = 50

After substituting $\phi_s(x)$ as per eq (8) into eq (9), we get,

$$K_s(u, v) = \frac{1}{s(1-s)}\left(1 - u^s v^{1-s} - (1-u)^s (1-v)^{1-s}\right). \tag{10}$$

Here, ϕ_s is continuous in $s \,\forall\, x \in (0, 1)$ and K_s is continuous in $s \,\forall\, (u, v) \in (0, 1)^2$. For each s, we obtain a unique goodness-of-fit test.

IV. Robust Fusion of Goodness of Fit Tests

Phi-divergences are universal hypothesis tests and their behavior is controlled only in one direction, i.e testing for $\mathcal{H}_0$. The behavior of the test in rejecting $\mathcal{H}_0$ when $\mathcal{H}_1$ holds, i.e in the alternative hypothesis regime, is called the power of the test procedure (P_D). The power changes depending on the type of distribution we are testing for. As a result, there is no 'uniformly most powerful' test from such a non-parametric setting, and different tests will be locally the most powerful, for a given set of alternate distributions. Depending on the structure of the distribution under the alternate hypothesis, the different goodness-of-fit tests vary greatly in power. Thus it is not possible to recommend a single test as an omnibus test over a range of SNRs as its performance may be surpassed by another test under different operating conditions. Recognizing this fact, we propose to apply a battery of tests for uniformly spaced s using a novel Thresholded Extreme Value (TEV) method that we call the Ensemble Goodness-of Fit (EG) test.

A. Ensemble of $\mathbf{\Phi}$-Divergence Test (EP) test

In statistics, the p-value of a test is defined as the tail integral of the density of the test statistic. Consider a GoF test with a statistic $T(\mathbf{X})$.

Let $Z_T(t)$ and $W_T(t)$ be the cumulative distributions of T under the null hypothesis ($\mathcal{H}_0$) and alternative hypothesis ($\mathcal{H}_1$) respectively. Then for a given observed test statistic $T(\mathbf{X}) = \tau$, the p-value is,

$$\rho(\tau) = \sup_{\theta \in \Omega_0} \mathrm{P}(T(\mathbf{X}) \geq \tau). \tag{11}$$

For the special case of a simple null hypothesis,

$$\rho(\tau) = \mathrm{P}(T > \tau | \mathcal{H}_0) = 1 - Z_T(\tau) = \int_{\tau}^{+\infty} \mathbf{z}_T(t)\, dt. \tag{12}$$

The p-value acts as an indicator of the confidence of the decision reached by theGoF test under a set of specific operating conditions. A low p-value indicates that we are highly unsure about rejecting the null hypothesis while a high p-value indicates that we are highly confident in rejecting the null hypothesis to be true. The principle advantage here is that the p-value is obtained via an implicit probability integral transformation and hence the p-values are distributed uniformly over $[0, 1]$ under the null hypothesis. This allows us to compare the outputs of different types of goodness-of-fit tests that have mismatched ranges of their test statistics on a standardized $[0, 1]$ interval.

Theorem 2. *The distribution of p-values for the null hypothesis is uniform over [0,1] for any test sample size.*

Our approach is based on the observation in [13] that the distribution of the p-value under *both* hypothesis is essential to formulate a threshold. Out of the ensemble of GoF tests,we accept the decision of the test that has the most extreme p-value, i.e , it has the highest confidence metric to support the decision. We will illustrate the process using the distribution of the p-values for SNR = 0 case, for the setup described in Results section. The distribution of p-values for the ensemble of tests is plotted in figure 2. The uniformity of the p-values under the null hypothesis is clearly seen, also the distribution under the alternative is highly skewed towards 0. The level of significance α to reject the null hypothesis is easily noted to be p-value $= \alpha$. We will calculate a *second order P-value* defined as the upper tail integral of the alternate p-value distribution and threshold it at level β. Note that this differs from conventional p-value use in that the threshold depends via β on the distribution under the alternate hypothesis. After these rejection regions have been defined, the remaining range of p-values is subject to a *randomized test* (see Chp.3 , Lehmann [14] for a review of randomized testing). This test randomly decides the outcome after normalizing with a predetermined prior distribution of the two hypothesis in the randomization region. The Kolmogorov Smirnov test, Anderson-Darling test e.t.c are individually sensitive to changes only in certain regions of the distribution and sacrifice high local power. in order to attain medium power over the complete support of the c.d.f. In contrast, the $\mathbf{\Phi}$-Divergence tests have been designed to be selective to changes in *different* regions of the c.d.f, and this selectivity is controlled via the tuning parameter s. Hence, the ensemble demonstrates a rake like property by being highly sensitive to changes over the complete support of the c.d.f. Thus, the Ensemble Goodness-of Fit (EG) test based on $\mathbf{\Phi}$-Divergences is implemented as follows

I Training phase: Calculate the distribution of p-values of all the test statistics under Null and Alternate hypotheses. Calculate the rejection region and the randomization region.

II Test Phase: For the test sample, obtain test statistics for each test in the ensemble, and the corresponding p-values.

III If the test p-value falls within a rejection region, pick the corresponding hypothesis.

IV Otherwise, randomly pick the hypothesis using the prior distribution and the magnitude of p-value.

V If *any* test rejects the null hypothesis, then the EP test decides in favor of $\mathcal{H}_1$.

The consistently superior performance of this test is shown via experimetal simulations in section V.

V. Results

We consider a setup that consists of a single primary user in a frequency flat fading environment with a block transmission/reception of size T symbols per block. The primary transmitter and the secondary receiver have 2 antennas each, i.e, $N_T, N_R = 2$. Thus the test size is $N = N_R \times T$ complex samples. Quadrature Phase Shift Keying modulation is employed and the noise is circularly complex gaussian with a spatial correlation coefficient of 0.2 between the two antennas. At the initiation of the testing period in the absence of the primary signal, a sequence of training samples is obtained with a duration of $N = 100$ complex samples. These training samples are used to estimate the distribution of the test samples under the null hypothesis using a gaussian kernel. After the training phase, each goodness-of-fit test statistic is evaluated using equation (10). Also, the corresponding p-value is calculated using the distribution of the test statistic that is approximated with high accuracy using the Noe recursion relations [7], [15]. Only the p-values are used in later processing and act as the summary statistic. For individual goodness-of-fit tests, the level of significance α for the test is decided *a priori*, that corresponds to the probability of falsely rejecting the null hypothesis (P_{FA}). If the p-value is less than α, the null hypothesis is rejected.

A. Ensemble Goodness-of Fit (EG) test based on $\mathbf{\Phi}$-Divergences

The EG test thresholds are implemented as per the algorithm described in Section IV-A for the following specifications. The ensemble consists of 7 $\mathbf{\Phi}$-divergence tests in addition to the Kolmogorov-Smirnov test for s $= -1, -0.5, 0, 0.5, 1, 1, 5, 2$. This ensemble is found to comprehensively improve on the performance of each individual test, including the KS test, but is not the only possible configuration. Any number of tests can be chosen to form the ensemble. The performance of the Ensemble Goodness-of-Fit test is shown in figure 3 for a test sample size of $N = 50$, i.e, the received block size $T = 25$. The level of significance for Probability of False Alarm is set as $\alpha = 0.05$ and the second order threshold for Probability of Missed Detection is set to be $\beta = 0.1$ using the method described in Section IV-A. The p-value corresponding to β is calculated empirically for each SNR level and noise distribution. The interim region is setup as the randomized testing region. Figure (3a) and figure (3d) plot the P_D and P_{FA} performance in presence of Gaussian noise. The Signal-to-Noise Ratio is varied from -12 dB to 0 dB. As the noise realization for the energy detector follows the model exactly, its performance is better than that of the goodness-of-fit tests. Also, the EG test uniformly outperforms the KS test alone. This behaviour is because of the rake like selectivity property of the test mentioned in Section IV-A, and can be seen through all operating conditions. Thus, the EG test is consistently upper bounds the performance of the KS test. Figure (3b) and figure (3e) plot the P_D and P_{FA} performance in presence of Middleton Class A noise mixture distribution. The Middleton noise has a $\Gamma = 0.5$ at both antennas where Γ controls the ratio of the Gaussian noise component to the Non-Gaussian noise component in the mixture distribution. For such a mixture with a dominant Gaussian noise, the power of the GoF as measured by the P_D is now better than that for Gaussian noise

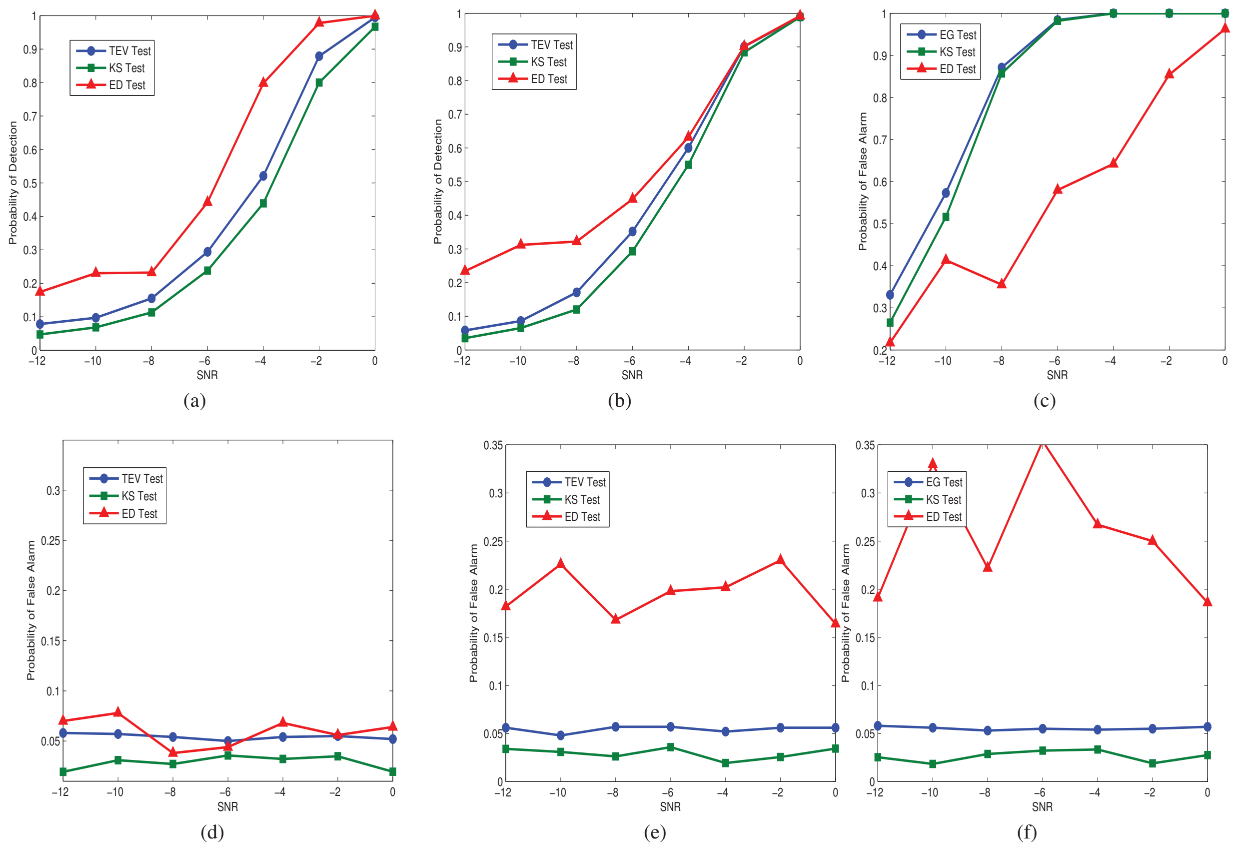

Fig. 3: Performance of the various tests for test size = 50. Fig 3a : P_D for Gaussian noise, Fig 3b : P_D for Non-Gaussian noise with $\Gamma = 0.5$, Fig 3d : P_{FA} for Gaussian noise, Fig 3e : P_{FA} for non-Gaussian noise, Fig 3f : P_{FA} for Non-Gaussian noise with $\Gamma = 0.1$

alone, and matches that of the Energy detector at SNR > -4 dB. Another critical point to note here is the false alarm events, given by P_{FA}, are significantly higher for the energy detector. Thus the design parameters are violated for the ED test while the EG test and the KS test still satisfy the P_{FA} design constraints.

VI. Conclusion

We have shown that nonparametric goodness-of-fit tests show robust performance under non-gaussian noise in contrast to parametric tests like the energy detector. A systematic approach to design a nonparametric test for the particular set of alternative hypotheses is proposed via using the Φ-divergence based GoF tests for a tuning parameter s. Also, the Ensemble GoF test that comprises of these Φ-divergence tests is shown through extensive simulations to consistently outperform the individual GoF tests, including the Kolmogorov-Smirnov test.

References

[1] R. D'Agostino and M. Stephens, *Goodness-of-fit techniques*. CRC, 1986.

[2] M. Stephens, "EDF statistics for goodness of fit and some comparisons," *Journal of the American Statistical Association*, vol. 69, no. 347, pp. 730–737, 1974.

[3] G. Zhang, X. Wang, Y. Liang, and J. Liu, "Fast and robust spectrum sensing via Kolmogorov-Smirnov test," *Communications, IEEE Transactions on*, vol. 58, no. 12, pp. 3410–3416, 2010.

[4] H. Wang, E. Yang, Z. Zhao, and W. Zhang, "Spectrum sensing in cognitive radio using goodness of fit testing," *Wireless Communications, IEEE Transactions on*, vol. 8, no. 11, pp. 5427–5430, 2009.

[5] A. Glen, L. Leemis, and D. Barr, "Order statistics in goodness-of-fit testing," *Reliability, IEEE Transactions on*, vol. 50, no. 2, pp. 209–213, 2001.

[6] N. Kundargi and A. Tewfik, "ProTOMAC: Proactive Transmit Opportunity Detection at the MAC Layer for Cognitive Radios," in *IEEE International Conference on Communications (ICC), 2010*.

[7] L. Jager and J. Wellner, "Goodness-of-fit tests via phi-divergences," *The Annals of Statistics*, vol. 35, no. 5, pp. 2018–2053, 2007.

[8] N. Cressie, "Multinomial goodness-of-fit tests," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 46, no. 3, pp. 440–464, 1984.

[9] D. Middleton, "Non-Gaussian noise models in signal processing for telecommunications: New methods an results for class A and class B noise models," *Information Theory, IEEE Transactions on*, vol. 45, no. 4, pp. 1129–1149, 1999.

[10] A. Chopra, K. Gulati, B. Evans, K. Tinsley, and C. Sreerama, "Performance bounds of MIMO receivers in the presence of radio frequency interference," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 2817–2820.

[11] L. Liu and M. Amin, "Performance analysis of GPS receivers in non-Gaussian noise incorporating precorrelation filter and sampling rate," *Signal Processing, IEEE Transactions on*, vol. 56, no. 3, pp. 990–1004, 2008.

[12] J. Angus, "The probability integral transform and related results," *SIAM review*, vol. 36, no. 4, pp. 652–654, 1994.

[13] R. Donahue, "A Note on Information Seldom Reported Via the P Value." *The American Statistician*, vol. 53, no. 4, 1999.

[14] E. Lehmann and J. Romano, *Testing statistical hypotheses*. Springer Verlag, 2005.

[15] M. Noe, "The calculation of distributions of two-sided Kolmogorov-Smirnov type statistics," *The Annals of Mathematical Statistics*, vol. 43, no. 1, pp. 58–64, 1972.