EMPIRICAL LIKELIHOOD RATIO TEST WITH DENSITY FUNCTION CONSTRAINTS

Yingxi Liu and Ahmed Tewfik

Electrical and Computer Engineering Department
The University of Texas at Austin
Austin, TX 78712, USA

ABSTRACT

In this work, we study non-parametric hypothesis testing problem with density function constraints. The empirical likelihood ratio test has been widely used in testing problems with moment (in)equality constraints. However, some detection problems cannot be described using moment (in)equalities. We propose a density function constraint along with an empirical likelihood ratio test. This detector is applicable to a wide variety of robust parametric/non-parametric detection problems. Since the density function constraints provide a more exact description of the null hypothesis, the test outperforms many other alternatives such as the empirical likelihood ratio test with moment constraints and robust Kolmogorov-Smirnov test, especially when the alternative hypothesis has a special structure.

Index Terms— empirical likelihood, universal hypothesis testing, goodness-of-fit test, robust detection.

1. INTRODUCTION

The empirical likelihood ratio test was first studied by A. Owen [1–3] to test the validity of moment equalities. It is widely used as a tool for non-parametric detection problems in economics. However, the empirical likelihood ratio test with moment constraints (ELRM) is not powerful enough for some problems which test not only the moment constraints but also many other parameters. This paper studies one such problem which tests density function constraints using an empirical likelihood ratio test (ELRDF). Namely, we consider the problem of testing whether the observations are generated by a density function which is point-wisely bounded. This problem covers a wide range of applications. For example, the robust detection problem started by Huber in [4], where the true density Q is buried in an ϵ-contamination model $Q = (1-\epsilon)P + \epsilon H$, with P the nominal density and H an arbitrary density function, can be treated as a variation of the function density constraint. Huber's test features a clipped version of the likelihood ratio test between the nominal densities that delivers performance which minimizes the worst-case probability of false alarm and miss. In fact, the density function constraint can be applied whenever robustness is needed.

Besides Huber's clipped test, there are other alternative tests that can be applied to test density function constraints. One approach is to use the empirical likelihood ratio test with moment inequalities [5, 6], which is one type of moment constraints. This technique also suffers from the problem of insufficient description using moment inequalities. For example, when the unknown hypothesis contains the null hypothesis (nested hypotheses), ELRM performs only slightly better than flipping a fair coin. Another alternative is the robust version of Kolmogorov-Smirnov (KS) test [7–9]. It is difficult to analytically compare the performance of ELRDF and robust KS test. Numerical examples show that the ELRDF outperforms robust KS test when hypotheses are nested and in low SNR regime. In this work, we also discuss the asymptotic optimality of the ELRDF in the framework of [5, 10–12].

The rest of this paper is organized as follows. Section 2 formulates the ELRDF test and discusses its asymptotic optimality. Section 3 discusses two alternative tests, the ELRM and the robust Kolmogorov-Smirnov test, and compare their performance with ELRDF. Section 4 presents an experimental study of noise uncertainty and tests' performances. Section 5 concludes the paper.

2. DENSITY FUNCTION CONSTRAINED DETECTION PROBLEM

2.1. Problem formulation

Consider a sequence of observations $\mathbf{X} = \{X_i : i = 1, \ldots, n, X_i \in \mathcal{X}\}$ which are i.i.d. generated by probability density f, with cumulative density function (CDF) F. $\mathcal{X} \subseteq \mathbb{R}$ denotes the sample space. Additionally, the empirical CDF with observations $\mathbf{X}$ is denoted as F_e:

$$F_e(x, \mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i \le x\}},$$

where $\mathbf{1}_{\{\cdot\}}$ is the indicator function. Denote $\mathcal{F}_e = \{F_e(x, \mathbf{X}) : \mathbf{X} \in \mathcal{X}^n\}$ as the set of all empirical density functions on the n-dimensional samples space $\mathcal{X}^n$. In the context where $\mathbf{X}$ is provided, we usually write $F_e(x, \mathbf{X})$ simply as $F_e(x)$. Given $\mathbf{X}$, the problem whether F belongs to a certain set of probability densities $\mathcal{F}$ is of interest. This is a universal hypothesis testing problem:

$$\begin{aligned} \mathcal{H}_0 &: F \in \mathcal{F}, \\ \mathcal{H}_1 &: F \notin \mathcal{F}. \end{aligned} \tag{1}$$

We are particularly interested in the form of $\mathcal{F}$ that is characterized by boundaries of certain CDFs, specifically:

$$\mathcal{F} = \{G : \ F_l(x) \le G(x) \le F_u(x)\}. \tag{2}$$

2.2. Solution

Given $\mathbf{X}$, let $\hat{F}$ be absolutely continuous with respect to F_e. ($\hat{F} \ll F_e$) Let $l(F_e) = \prod_{i=1}^{n} \big(F_e(X_i) - F_e(X_i-)\big) = n^{-n}$ and $l(\hat{F}) = \prod_{i=1}^{n} w_i$, where $w_i = \hat{F}(X_i) - \hat{F}(X_i-)$. The empirical likelihood ratio is defined as:

$$R(\hat{F}, F_e) = \frac{l(\hat{F})}{l(F_e)}. \tag{3}$$

Naturally, $w_i \geq 0$, $\sum_{i=1}^{n} w_i = 1$. Then we can rewrite $R(\hat{F}, F_e)$ as a function of $\vec{w} = [w_1, w_2, \ldots, w_n]^T$:

$$R(\vec{w}, F_e) = \prod_{i=1}^{n} n w_i.$$

In the sequel, we use $R(\hat{F}, F_e)$ and $R(\vec{w}, F_e)$ interchangeably depending on the context. It is known that $l(\hat{F}) \leq l(F_e)$ for all choices of $\vec{w}$ in the probability simplex [3, p. 8]. When $w_i = \frac{1}{n}$ for all i, $l(\hat{F}) = l(F_e)$, then $R(\hat{F}, F_e) \leq 1$. As a first step towards the detection problem, we would like to maximize the empirical likelihood ratio $R(\hat{F}, F_e)$ with respect to $\vec{w}$ when $\hat{F}$ satisfies the boundary conditions:

$$\max_{\vec{w}} \Big\{ R(\vec{w}, F_e) : w_i \geq 0, \sum_{i=1}^{n} w_i = 1, \\ F_l(X_i) \leq \hat{F}(X_i) \leq F_u(X_i) \Big\}.$$

We shall assume without loss of generality that $X_1 < X_2 < \ldots < X_n$. Construct a $(n-1) \times n$ matrix:

$$A = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ 1 & 1 & 0 & \ldots & 0 & 0 \\ 1 & 1 & 1 & \ldots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 1 & \ldots & 1 & 1 & 1 & 0 \end{bmatrix},$$

and let $\vec{F_l} = (F_l(X_1), F_l(X_2), \ldots, F_l(X_{n-1}))^T$, $\vec{F_u} = (F_u(X_1), F_u(X_2), \ldots, F_u(X_{n-1}))^T$. The last constraint is conveniently written as:

$$\vec{F_l} \leq A\vec{w} \leq \vec{F_u}.$$

One should notice that A does not contain a row of all ones since the constraint $\sum_{i=1}^{n} w_i = 1$ will certainly contradict the assertion $F_l(X_n) \leq \sum_{i=1}^{n} w_i \leq F_u(X_n)$. Indeed, one can also drop the constraint $\sum_{i=1}^{n} w_i = 1$ and add an all-one row to the bottom of A. We shall see that at this point, it would not make a dramatic difference to favor one alternative over the other. We formally introduce the empirical likelihood with density function constraints as follows:

$$\max_{\vec{w}} \Big\{ R(\vec{w}, F_e) : w_i \geq 0, \sum_{i=1}^{n} w_i = 1, \\ \vec{F_l} \leq A\vec{w} \leq \vec{F_u} \Big\}. \tag{4}$$

This is a problem with a concave objective function (after taking log operation) and linear constraints. The solution to it is readily available. Let $\vec{w}^*$ be the maximizer and corresponding CDF as F^*. We build the empirical likelihood ratio test with density function constraints on the value of $R(\vec{w}^*, F_e)$:

$$R(\vec{w}^*, F_e) \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} \eta_n, \tag{5}$$

where $0 \leq \eta_n \leq 1$. The test is to say that when the estimated likelihood is close enough to the empirical density, we declare that $\mathcal{H}_0$ is true; and declare $\mathcal{H}_1$ true otherwise.

2.3. Asymptotic optimality

The detector (5) is a partition of $\mathcal{F}_e$, *i.e.*, $\mathcal{F}_e = \Omega_0 \cup \Omega_0^c$, where Ω_0 is explicitly defined as:

$$\Omega_0 = \{F_e : R(F^*, F_e) \geq \eta_n\}.$$

Ω_0 is also referred to as the confidence set. The empirical likelihood ratio test can be interpreted from an information theoretic perspective. Given the observations $\mathbf{X}$, notice that:

$$\begin{aligned} \inf_{F \in \mathcal{F}} D(F_e||F) &= \min_{\vec{w}} \Big\{ \sum_{i=1}^{n} \frac{1}{n} \log \frac{1}{n w_i} : \\ &\qquad w_i \geq 0, \sum_{i=1}^{n} w_i = 1, \vec{F_l} \leq A\vec{w} \leq \vec{F_u} \Big\} \\ &= -\frac{1}{n} \log R(F^*, F_e), \end{aligned}$$

where $D(\cdot||\cdot)$ is the Kullback-Leibler divergence. Then we could express Ω_0 using $\inf_{F \in \mathcal{F}} D(F_e||F)$:

$$\Omega_0 = \{F_e : e^{-n \inf_{F \in \mathcal{F}} D(F_e||F)} \geq \eta_n\}. \tag{6}$$

Let $\hat{\Omega}_0$, $\hat{\Omega}_0^c$ be an arbitrary partition of $\mathcal{F}_e$. The test declares $\mathcal{H}_0$ true if $F_e \in \hat{\Omega}_0$. The error performance of the test is characterized by the *worst-case* probability of false alarm and the probability of miss:

$$\begin{aligned} P_F &= \sup_{F \in \mathcal{F}} F(F_e \notin \hat{\Omega}_0), \\ P_M &= \sup_{F \notin \mathcal{F}} F(F_e \in \hat{\Omega}_0). \end{aligned}$$

In the asymptotic regime, it is customary to study the exponential decay rates of P_F and P_M as the number of samples tends to infinity. Their error exponents are expressed as:

$$\begin{aligned} e_F(\hat{\Omega}_0) &= \lim_{n \to \infty} \inf -\frac{1}{n} \log \sup_{F \in \mathcal{F}} F(F_e \notin \hat{\Omega}_0) \\ &= \lim_{n \to \infty} \inf \inf_{F \in \mathcal{F}} -\frac{1}{n} \log F(F_e \notin \hat{\Omega}_0), \end{aligned}$$

and

$$\begin{aligned} e_M(\hat{\Omega}_0) &= \lim_{n \to \infty} \inf -\frac{1}{n} \log \sup_{F \notin \mathcal{F}} F(F_e \in \hat{\Omega}_0) \\ &= \lim_{n \to \infty} \inf \inf_{F \notin \mathcal{F}} -\frac{1}{n} \log F(F_e \in \hat{\Omega}_0). \end{aligned}$$

The test $\hat{\Omega}_0$ is asymptotically optimal if it solves the generalized Neyman-Pearson problem:

$$\begin{aligned} &\sup_{\Omega} e_M \\ &s.t. : e_F \geq \gamma. \end{aligned} \tag{7}$$

Invoking the Sanov's theorem, the following relationship is obvious:

$$\begin{aligned} e_F(\hat{\Omega}_0) &= \lim_{n \to \infty} \inf \inf_{F \in \mathcal{F}} -\frac{1}{n} \log F(F_e \notin \hat{\Omega}_0) \\ &= \inf_{F_e \notin \hat{\Omega}_0} \inf_{F \in \mathcal{F}} D(F_e||F). \end{aligned}$$

Replacing $\hat{\Omega}_0$ by Ω_0 in (6) yields:

$$\begin{aligned} e_F(\Omega_0) &= \inf_{F_e \notin \Omega_0} \inf_{F \in \mathcal{F}} D(F_e || F) \\ &= -\frac{1}{n} \log \eta_n. \end{aligned}$$

Since the test Ω_0 is the conjectured optimal test, it is desired that all tests be compared on a fair ground. Therefore in the generalized Neyman-Pearson problem (7), we let $\gamma = -\frac{1}{n} \log \eta_n$.

The asymptotic optimality of the test (6) can be argued in a similar way as in [5]. To rigorously establish the asymptotic optimality of the test Ω_0, one needs to follow several steps. Firstly, one shall notice that the test is asymptotically consistent. Namely, when the null hypothesis is true, $P\{F_e \in \Omega_0\} \stackrel{n\to\infty}{\longrightarrow} 1$. This is true according to Glivenko-Cantelli theorem [13, 14]: the empirical distribution uniformly converges to the true distribution. Secondly, one shall argue that for an alternative test $\hat{\Omega}_0$ that satisfies:

$$e_F(\hat{\Omega}_0) \geq \gamma + \delta,$$

$\forall \delta > 0$, it follows that

$$e_M(\Omega_0) \geq e_M(\hat{\Omega}_0) + \kappa(\delta),$$

$\forall \kappa(\delta) > 0$. For detailed techniques, readers are referred to [5, 10].

3. OTHER COMPETITIVE DETECTORS

The problem (1) is of wide interest as a non-parametric detection problem. Several detectors proposed in the past could also be considered as possible solutions. In this section, the empirical likelihood ratio test with moment constraints and the robust KS test will be discussed and compared with ELRDF.

3.1. Empirical likelihood ratio test with moment constraints

The empirical likelihood ratio test with moment constraints (ELRM) is closely related to our proposed detector. The difference is that the density function constraints are replaced by constraints on the moments such as the mean, variance, etc. For example, suppose $l = \int x dF_l(x)$ and $u = \int x dF_u(x)$. One can consider the following alternative for the original optimization problem (4):

$$\max_{\vec{w}} \Big\{ \prod_{i=1}^{n} n w_i : w_i \geq 0, \sum_{i=1}^{n} w_i = 1, \\ u \leq \sum_{i=1}^{n} w_i X_i \leq l \Big\}. \tag{8}$$

If the detection problem is to decide whether the underlying probability has a mean that falls onto the interval $[u, l]$, the optimal solution to (8) also enjoys the asymptotic optimality in the generalized Neyman-Pearson sense. Nevertheless, the moment constraint alone cannot provide an exact description of the set $\mathcal{F}$ in (2). Although in theory, a density function is determined when all of its moments are determined, it is impractical to have a long list of moment conditions.

Another case in which the ELRM would fail is when the hypotheses are nested. For example, in a communication system where the receiver performs uncorrelated signal detection. The hypotheses that a signal is transmitted and that no signal present are nested. Consider a numerical example. The null distribution is a mixture of normal distribution $\mathcal{N}(0, 1)$ with unknown means uniformly located in the interval $[-0.1, 0.1]$. The alternative distribution is also a mixture of normal distribution with unknown means located in $[-1, 1]$. Sample size is 20. The receiver operational characteristic (ROC) curve shown in Figure 1 indicates that ELRDF outperforms ELRM.

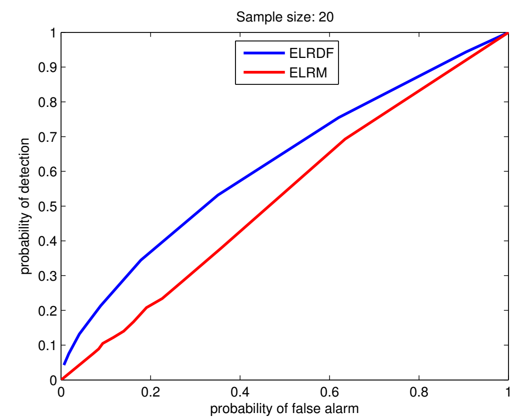

Fig. 1: ROC ELRDF vs ELRM for nested hypotheses.

3.2. Robust Kolmogorov-Smirnov test

The Kolmogorov-Smirnov test is a popular non-parametric test. Its robust version is directly applicable to our problem with density function constraint $\mathcal{F}$ in (2). The test statistic of the robust KS test has the following form:

$$D_n = \inf_{F \in \mathcal{F}} \sup_x |F_e(x) - F(x)|. \tag{9}$$

The test compares the statistic with a constant:

$$\sqrt{n} D_n \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} \gamma. \tag{10}$$

It is difficult to analytically compare the performance of ELRDF and robust KS test. We consider an example where the null hypothesis is a mixture of normal distributions with unknown mean located in $[-2, 2]$. The alternative hypothesis is a normal distribution located at 2.02 and 2.2. When the alternative hypothesis is located at 2.2, the robust KS test is slightly better for small probability of false alarm. When the alternative hypothesis is located at 2.02, ELRDF outperforms the robust KS test. We draw a conclusion without rigorous proof that the ELRDF performs better than robust KS test in low SNR regime. It could be interpreted heuristically as follows. When the alternative hypothesis is very close to the null hypothesis, the robust KS test would fail since it is too close to the boundaries F_l and F_u. But the ELRDF computes the "distance" between the empirical density and the estimated density, which is located inside the boundaries of F_l and F_u. Therefore, ELRDF still can discriminate the hypotheses with reasonable probability.

4. EXPERIMENTAL RESULT

4.1. Uncertainty of noise distribution

In this section, we examine the uncertainty of noise distribution where the noise samples are obtained from a software-defined radio device. When a large amount of noise samples are examined, they appear to follow the same Gaussian distribution. However, when a

Table 1: Percentage of test failures of direct and mean shifted K-S test.

Data sets	1	2	3	4	5	6
Direct K-S test	0.0255	0.0252	0.0270	0.0261	0.0260	0.0286
Mean shifted K-S test	3.5217×10^{-4}	1.9234×10^{-4}	0	0	1.8871×10^{-4}	1.7986×10^{-4}

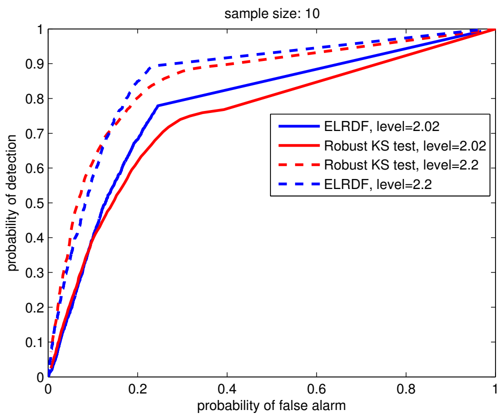

Fig. 2: ROC ELRDF vs robust KS test.

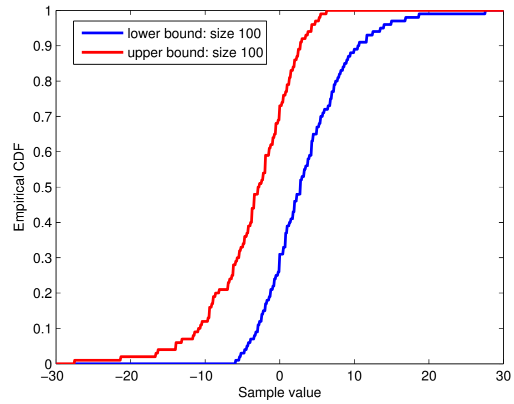

Fig. 3: Uncertainty region of experimental noise samples.

small portion of the samples are examined, they appear to follow another Gaussian distribution. This observation is supported by results of Kolmogorov-Smirnov test [8, 9]. Specifically, we study six data sets with 2 million samples each, which is used to generate an estimation of Gaussian distribution $\mathcal{N}(m, \sigma^2)$. For every 500 samples, the KS test of significance level 0.01 with critical value $\frac{1.63}{\sqrt{500}}$ is performed to test whether the 500 samples are generated by $\mathcal{N}(m, \sigma^2)$ or not. If the samples are generated by the same distribution F, this percentage of failed tests should not exceed the significance level 0.01. However, from the second row in Table 1, we notice that the percentages of the test failures exceeds the significance level by a noticeable margin. This result indicates that the noise samples are not generated by the same Gaussian distribution. To compare, a batch of mean shifted K-S tests are conducted, where in each test, the mean m of the null hypothesis is shifted to the mean of each 500-sample group. We notice that the failure percentages drop significantly for all the data sets (third row in Table 1). This indicates that there exist shifts in the mean for the Gaussian distributions while generating the samples. The empirical uncertainty region of noise sample distribution is plot in Figure 3 for finite sample size of 100.

4.2. Example: testing a constant

We consider an example where the alternative distribution has a constant boost of the noise and the empirical uncertainty region studied in previous subsection. However, this information is not available and we consider a universal test. As the closest alternative, the performance of the robust KS test is studied as a comparison. Figure 4 plots the ROC curves of the two tests for sample size 10 and constant level at 3 and 5. The plots indicates that ELRDF outperforms the robust KS test.

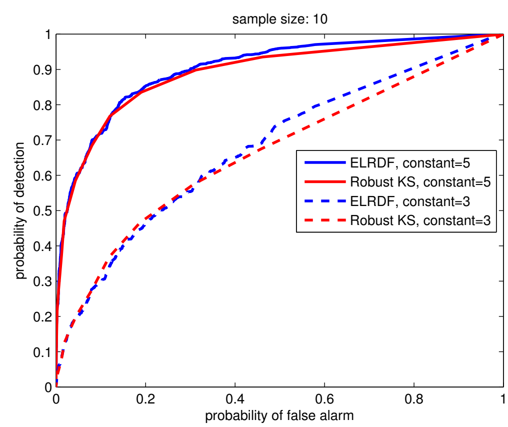

Fig. 4: Performance comparison of detecting a constant: ELRDF vs robust KS test.

5. CONCLUSION

This work proposes a novel empirical likelihood ratio test with density function constraints (ELRDF). This test is applicable to many applications in robust parametric/non-parametric detection problems. By providing an exact description of uncertainty using density function constraints, this test delivers better performance compared to empirical likelihood ratio test with moment constraints (ELRM). A detection problem with nested hypotheses is provided as an illustration. More importantly, the paper shows with a numerical and an experimental examples where the alternative hypothesis is a constant boost of the noise that ELRDF outperforms the robust KS test especially at the low SNR regime. It also discusses the asymptotic optimality of this test. The following step of this paper is to establish a rigorous proof of the asymptotic optimality and the asymptotic distribution of test statistics.

6. REFERENCES

[1] A. Owen, "Empirical likelihood ratio confidence intervals for a single functional," *Biometrika*, vol. 75, no. 2, pp. 237–249, 1988.

[2] ——, "Empirical likelihood ratio confidence regions," *The Annals of Statistics*, vol. 18, no. 1, pp. 90–120, 1990.

[3] ——, *Empirical likelihood*. Chapman & Hall/CRC, 2001, vol. 92.

[4] P. Huber, "A robust version of the probability ratio test," *The Annals of Mathematical Statistics*, vol. 36, no. 6, pp. 1753–1758, 1965.

[5] I. Canay, "El inference for partially identified models: large deviations optimality and bootstrap validity," *Journal of Econometrics*, vol. 156, no. 2, pp. 408–425, 2010.

[6] A. Rosen, "Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities," *Journal of Econometrics*, vol. 146, no. 1, pp. 107–117, 2008.

[7] J. Unnikrishnan, S. Meyn, and V. Veeravalli, "On thresholds for robust goodness-of-fit tests," in *Information Theory Workshop (ITW), 2010 IEEE*. IEEE, 2010, pp. 1–4.

[8] F. Massey Jr, "The Kolmogorov-Smirnov test for goodness of fit," *Journal of the American statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.

[9] H. Lilliefors, "On the Kolmogorov-Smirnov test for normality with mean and variance unknown," *Journal of the American Statistical Association*, vol. 62, no. 318, pp. 399–402, 1967.

[10] O. Zeitouni and M. Gutman, "On universal hypotheses testing via large deviations," *Information Theory, IEEE Transactions on*, vol. 37, no. 2, pp. 285–290, 1991.

[11] J. Qin and J. Lawless, "Empirical likelihood and general estimating equations," *The Annals of Statistics*, pp. 300–325, 1994.

[12] Y. Kitamura, A. Santos, and A. Shaikh, "On the asymptotic optimality of empirical likelihood for testing moment restrictions," *Econometrica*, vol. 80, no. 1, pp. 413–423, 2012.

[13] R. Dudley, *Uniform central limit theorems*. Cambridge Univ Press, 1999, vol. 23.

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