

This is a postprint version of the following published document:

Harsha, K. V., Ravi, J. & Koch, T. (01-09 November 2022). *Second-order asymptotics of Hoeffding-like hypothesis tests* [proceedings]. 2022 IEEE Information Theory Workshop (ITW), Mumbai, India.

DOI: [10.1109/ITW54588.2022.9965931](https://doi.org/10.1109/ITW54588.2022.9965931)

© 2022 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Second-Order Asymptotics of Hoeffding-Like Hypothesis Tests

K. V. Harsha, Jithin Ravi,[†] and Tobias Koch[‡]

[†]Indian Institute of Technology Kharagpur, India

[‡]Universidad Carlos III de Madrid, Leganés, Spain and Gregorio Marañón Health Research Institute, Madrid, Spain.

Emails: kv.harsha1989@gmail.com, jithin@ece.iitkgp.ac.in, koch@tsc.uc3m.es

Abstract—We consider a binary statistical hypothesis testing problem, where n independent and identically distributed random variables Z^n are either distributed according to the null hypothesis P or the alternate hypothesis Q , and only P is known. For this problem, a well-known test is the Hoeffding test, which accepts P if the Kullback-Leibler (KL) divergence between the empirical distribution of Z^n and P is below some threshold. In this paper, we consider Hoeffding-like tests, where the KL divergence is replaced by other divergences, and characterize, for a large class of divergences, the first and second-order terms of the type-II error for a fixed type-I error. Since the considered class includes the KL divergence, we obtain the second-order term of the Hoeffding test as a special case.

I. INTRODUCTION

Statistical hypothesis testing is known to have applications in areas such as information theory, signal processing, and machine learning. The most simple form of hypothesis testing is binary hypothesis testing, where the goal is to determine the distribution of a random variable Z between a null hypothesis P and an alternate hypothesis Q . There can be two types of errors in binary hypothesis testing: The type-I error is the probability of declaring the hypothesis as Q when the true distribution is P . The type-II error is the probability of declaring the hypothesis as P when the true distribution is Q . In general, we are interested in analyzing the trade-off between these two types of errors for a given test, which may have full or only partial access to the distributions P and Q .

When both P and Q are known, the likelihood-ratio test (also known as Neyman-Pearson test [1]) achieves the optimal trade-off between type-I and type-II error. The Neyman-Pearson test is also investigated in an asymptotic setting, where one observes n independent copies of Z and the errors are analyzed asymptotically as n tends to infinity. In this case, we are often interested in the behavior of the type-II error β_n for a fixed type-I error α_n . It is known that, for $\alpha_n \leq \epsilon$, $\epsilon \in (0, 1)$, the corresponding type-II error satisfies [2], [3, Prop. 2.3]

$$-\ln \beta_n = nD_{\text{KL}}(P\|Q) - \sqrt{nV(P\|Q)}Q^{-1}(\epsilon) + O(\ln n) \quad (1)$$

Part of this work was done while K. V. Harsha and J. Ravi were with the Universidad Carlos III de Madrid, Leganés, Spain. K. V. Harsha, J. Ravi, and T. Koch have received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant No. 714161). T. Koch has further received funding from the Spanish Ministerio de Ciencia e Innovación under Grant PID2020-116683GB-C21 / AEI / 10.13039/501100011033.

as $n \rightarrow \infty$, where $D_{\text{KL}}(P\|Q)$ is the Kullback-Leibler (KL) divergence between P and Q [4],

$$V(P\|Q) \triangleq \sum_i P_i \left[\left(\ln \frac{P_i}{Q_i} - D_{\text{KL}}(P\|Q) \right)^2 \right]$$

is the divergence variance, $Q^{-1}(\cdot)$ is the inverse of the tail probability of the standard Normal distribution, and $O(\ln n)$ denotes a term that grows at most as fast as $\ln n$ as $n \rightarrow \infty$. It follows from (1) that the first-order term of $-\ln \beta_n$, sometimes referred to as the *error exponent*, is given by $D_{\text{KL}}(P\|Q)$. Likewise, the second-order term is given by $-\sqrt{V(P\|Q)}Q^{-1}(\epsilon)$.

The case where the test has only partial access to the distributions P and Q is generally studied under the name of *composite hypothesis testing*. Several special cases of composite hypothesis testing have been investigated in the literature. Two such cases are: 1) P is known and Q can be any distribution but P ; 2) P is known and Q belongs to a special class $\mathcal{Q}$. A test proposed by Hoeffding [5], known as the Hoeffding test, is suitable for these two cases. In this test, the null hypothesis P is accepted if the KL divergence between the type T_{Z^n} (empirical distribution) of the observations $Z^n = (Z_1, \dots, Z_n)$ and P is below some threshold c . Otherwise, the alternate hypothesis is accepted.

In [6, Th. III.2], it was shown that the Hoeffding test achieves the same first-order term of $-\ln \beta_n$ as the Neyman-Pearson test. Consequently, not having access to the distribution of the alternate hypothesis does not affect the first-order term. To analyze the second-order term of $-\ln \beta_n$ in the above two cases of composite hypothesis testing, Watanabe proposed a test that is second-order optimal in some sense [7].

The Hoeffding test does not require knowledge about the special class $\mathcal{Q}$ to which Q belongs. Tests that have knowledge of $\mathcal{Q}$ may therefore outperform the Hoeffding test. Such tests include the generalized likelihood-ratio test (GLRT)¹ [8] and test via mismatched divergence [9]. For example, it has been observed that the latter test outperforms the Hoeffding test [9]. However, an analytical comparison between the Hoeffding test and other tests is missing since refined asymptotics of the Hoeffding test are not available.

¹In fact, the Hoeffding test is the GLRT for the case where $\mathcal{Q}$ is the set of all probability distributions different from P .

In this paper, we study the behavior of the second-order term of the Hoeffding test, as well as of Hoeffding-like tests where the KL divergence is replaced by other divergences, in the following referred to as *divergence tests*. For a large class of divergences, we demonstrate that the second-order term is given by $-\sqrt{V(P\|Q)Q_{\chi^2, k-1}^{-1}(\epsilon)}$, where k is the cardinality of the observations and $Q_{\chi^2, k-1}^{-1}(\cdot)$ denotes the inverse of the tail probability of the chi-square distribution with $k-1$ degrees of freedom. This class of divergences includes the KL divergence, so we obtain the second-order term of the Hoeffding test as a special case. A key ingredient in the proof of this result is the well-known fact that, under the null hypothesis, the KL divergence between the type T_{Z^n} and the distribution P converges in probability to the chi-square distribution [9], [10]. However, in our proof, we shall require a stronger, Berry-Esseen-type convergence [11], [12].

The rest of this paper is organized as follows. Section II presents notations and the problem formulation. Section III discusses our main results. Section IV presents the proof of these results. Section V concludes the paper with a summary and discussion. Due to space limitations, some of the proof steps are deferred to the extended version of our paper [13].

II. NOTATIONS AND PROBLEM FORMULATION

A. Notations

Let $f(x)$ and $g(x)$ be two real-valued functions. For $a \in \mathbb{R} \cup \{\infty\}$, we write $f(x) = O(g(x))$ as $x \rightarrow a$ if $\limsup_{x \rightarrow a} \frac{|f(x)|}{|g(x)|} < \infty$. Similarly, we write $f(x) = o(g(x))$ as $x \rightarrow a$ if $\lim_{x \rightarrow a} \frac{|f(x)|}{|g(x)|} = 0$.

B. Divergence and Divergence Test

Let the random variable Z take value in a discrete set $\mathcal{Z} = \{a_1, \dots, a_k\}$, where $k \geq 2$. We denote its probability distribution by a k -length vector $\pi = (\pi_1, \dots, \pi_k)^T$ with entries $\pi_i \triangleq \Pr\{Z = a_i\}$, and we assume that $\pi_i > 0$, $i = 1, \dots, k$. We denote by $\mathcal{P}(\mathcal{Z})$ the set of all such probability distributions. For a length- n sequence z^n , let $T_{z^n} = (T_{z^n}(a_1), \dots, T_{z^n}(a_k))^T$ denote its type. Since probabilities sum to one, any $\pi = (\pi_1, \dots, \pi_k)^T \in \mathcal{P}(\mathcal{Z})$ can be represented by the first $(k-1)$ components of π . With this representation, the set $\mathcal{P}(\mathcal{Z})$ can be identified as a $(k-1)$ -dimensional manifold; see [14] for more details.

Given any two probability distributions $P, Q \in \mathcal{P}(\mathcal{Z})$, one can define a non-negative function $D(P\|Q)$, called a *divergence*, which represents a measure of discrepancy between them. Mathematically, a divergence is defined as follows [14].

Definition 1: Consider two points $P, Q \in \mathcal{P}(\mathcal{Z})$ with coordinates $\mathbf{P} = (P_1, \dots, P_{k-1})^T$ and $\mathbf{Q} = (Q_1, \dots, Q_{k-1})^T$. A divergence $D(P\|Q)$ between P and Q is a smooth function of $\mathbf{P}$ and $\mathbf{Q}$ satisfying the following conditions:

- 1) $D(P\|Q) \geq 0$ for any $P, Q \in \mathcal{P}(\mathcal{Z})$.
- 2) $D(P\|Q) = 0$ if, and only if, $P = Q$.
- 3) The Taylor expansion of D satisfies

$$D(\mathbf{P} + \epsilon\|\mathbf{P}) = \frac{1}{2} \sum_{i,j=1}^{k-1} g_{ij}(\mathbf{P}) \epsilon_i \epsilon_j + O(\|\epsilon\|_2^3)$$

as $\|\epsilon\|_2 \rightarrow 0$, for some $(k-1) \times (k-1)$ -dimensional positive-definite matrix $G = (g_{ij})$ that depends on $\mathbf{P}$ and $\epsilon = (\epsilon_1, \dots, \epsilon_{k-1})^T$, where $\|\cdot\|_p$, $p \geq 1$ is the L^p norm.

A well-known example of a divergence is the f -divergence, which includes the KL divergence and the α -divergence as special cases [15].

We now define a divergence test for the following composite hypothesis testing setup: under hypothesis H_0 , the distribution is P ; under hypothesis H_1 , the distribution is anything but P . A Hoeffding-like test or divergence test $T_n^D(r)$ for testing $H_0 : Z^n \sim P^n$ against the alternative $H_1 : Z^n \sim Q^n$ is defined as follows:

Upon observing z^n , if $D(T_{z^n}\|P) < r$ for some $r > 0$, then the null hypothesis P is accepted; else P is rejected.

For $r > 0$, define the acceptance region for P as

$$\mathcal{A}_n^D(r) \triangleq \{z^n \mid D(T_{z^n}\|P) < r\}.$$

Then, the type-I and type-II errors are given by

$$\begin{aligned} \alpha_n(T_n^D(r)) &\triangleq P^n(\mathcal{A}_n^D(r)^c) \\ \beta_n(T_n^D(r)) &\triangleq Q^n(\mathcal{A}_n^D(r)). \end{aligned}$$

Our goal is to analyze the asymptotics of the type-II error when the type-I error satisfies $\alpha_n \leq \epsilon$, $0 < \epsilon < 1$. To this end, we define the first-order term β' and the second-order term β'' of the divergence test as follows:

$$\begin{aligned} \beta' &\triangleq \lim_{n \rightarrow \infty} -\frac{1}{n} \ln \beta_n \\ \beta'' &\triangleq \lim_{n \rightarrow \infty} \frac{-\ln \beta_n - n\beta'}{\sqrt{n}} \end{aligned}$$

if the limits exist. For the Hoeffding test, it is known that $\beta' = D_{\text{KL}}(P\|Q)$. In this paper, we generalize this result to the divergence test for a class of divergences defined in Definition 2. We further obtain β'' for the divergence test.

III. RESULTS

We shall consider the following class of divergences:

Definition 2: Let Ξ denote the class of divergences satisfying the following conditions:

- 1) For any $P, T \in \mathcal{P}(\mathcal{Z})$ and some positive constant η , the second-order Taylor approximation of $D(T\|P)$ around $T = P$ is given by

$$D(T\|P) = \eta d_{\chi^2}(T, P) + O(\|\mathbf{T} - \mathbf{P}\|_2^3) \quad (2)$$

as $\|\mathbf{T} - \mathbf{P}\|_2 \rightarrow 0$, where $\mathbf{T} = (T_1, \dots, T_{k-1})^T$, $\mathbf{P} = (P_1, \dots, P_{k-1})^T$, and d_{χ^2} is the χ^2 -divergence

$$d_{\chi^2}(T, P) \triangleq \sum_{i=1}^k \frac{(T_i - P_i)^2}{P_i}.$$

- 2) For any $P, T \in \mathcal{P}(\mathcal{Z})$ and some positive constant $\tilde{C}$, the divergence D satisfies the Pinsker-type inequality

$$\tilde{C} \|T - P\|_1^2 \leq D(T\|P). \quad (3)$$

3) The tail probability of $\eta^{-1}nD(T_{Z^n}\|P)$ satisfies

$$P^n(\eta^{-1}nD(T_{Z^n}\|P) \geq c) = Q_{\chi^2, k-1}(c) + O(\delta_n) \quad (4)$$

for every $c > 0$ and a positive sequence δ_n satisfying $\delta_n \rightarrow 0$. In (4), $Q_{\chi^2, k-1}(\cdot)$ is the tail probability of a chi-square distribution with $k-1$ degrees of freedom.

As we shall discuss at the end of this section, the class of divergences Ξ is large and includes both the KL divergence and the α -divergence with $-3 \leq \alpha \leq 3$ and $\alpha \neq \pm 1$.

The following theorem characterizes β' and β'' of the divergence test for any divergence D in Ξ .

Theorem 1: Let $D \in \Xi$ and $0 < \epsilon < 1$. Consider the divergence test $T_n^D(r)$ for testing $H_0 : Z^n \sim P^n$ against the alternative $H_1 : Z^n \sim Q^n$, where $P, Q \in \mathcal{P}(\mathcal{Z})$ and $P \neq Q$. Recall that the cardinality of $\mathcal{Z}$ is $k \geq 2$. Then, the type-II error satisfies the following:

Part 1: There exists a threshold value r_n satisfying

$$\alpha_n(T_n^D(r_n)) \leq \epsilon \quad (5)$$

such that, as $n \rightarrow \infty$,

$$-\ln \beta_n(T_n^D(r_n)) \geq nD_{\text{KL}}(P\|Q) - \sqrt{nV(P\|Q)Q_{\chi^2, k-1}^{-1}(\epsilon)} + O(\max\{\delta_n \sqrt{n}, \ln n\}). \quad (6)$$

Part 2: For all $r_n > 0$ satisfying (5), we have as $n \rightarrow \infty$

$$-\ln \beta_n(T_n^D(r_n)) \leq nD_{\text{KL}}(P\|Q) - \sqrt{nV(P\|Q)Q_{\chi^2, k-1}^{-1}(\epsilon)} + O(\max\{\delta_n \sqrt{n}, \ln n\}). \quad (7)$$

In (6) and (7), $Q_{\chi^2, k-1}^{-1}(\cdot)$ is the inverse of $c \mapsto Q_{\chi^2, k-1}(c)$.

Proof: See Section IV. ■

Since the sequence δ_n in (6) and (7) vanishes as $n \rightarrow \infty$, it follows that, for the optimal threshold value r_n ,

$$-\ln \beta_n(T_n^D(r_n)) = nD_{\text{KL}}(P\|Q) - \sqrt{nV(P\|Q)Q_{\chi^2, k-1}^{-1}(\epsilon)} + o(\sqrt{n}). \quad (8)$$

Hence, Theorem 1 characterizes the first and second-order term of the divergence test for any $D \in \Xi$. We hasten to add that (7) is a converse within the class of divergence tests. There may be other hypothesis tests that only require knowledge of the distribution of the null hypothesis and that achieve a higher second-order performance than (8).

Clearly, the divergence test cannot outperform the Neyman-Pearson test. The following lemma implies that the second-order term β'' of the divergence test is *strictly* smaller than the second-order term of the Neyman-Pearson test.

Lemma 2: For every $0 < \epsilon < 1$ and $\eta = 1, 2, \dots$, we have $\sqrt{Q_{\chi^2, \eta}^{-1}(\epsilon)} > Q^{-1}(\epsilon)$.

Proof: Let Y_i , $i = 1, \dots, \eta$ be independent and identically distributed standard Normal random variables. Then, for $y \geq 0$,

$$Q(y) < \Pr(Y_1^2 \geq y^2) \leq \Pr\left(\sum_{i=1}^{\eta} Y_i^2 \geq y^2\right) = Q_{\chi^2, \eta}(y^2) \quad (9)$$

where the first inequality follows because

$$\Pr(Y_1^2 \geq y^2) = \Pr(|Y_1| \geq y) = 2Q(y)$$

and the second inequality follows because $Y_1^2 \leq \sum_{i=1}^{\eta} Y_i^2$ with probability one. Let $Q_{\chi^2, \eta}(y^2) = \epsilon$. Since $y \mapsto Q(y)$ is strictly decreasing, we obtain from (9) that there exists a $y' < y$ such that $Q(y') = \epsilon$. It follows that $Q^{-1}(\epsilon) = y'$ is strictly smaller than $\sqrt{Q_{\chi^2, \eta}^{-1}(\epsilon)} = y$. ■

Following similar steps as in the proof of Lemma 2, it can be shown that $\sqrt{Q_{\chi^2, k-1}^{-1}(\epsilon)}$ is increasing in k . Thus, as the cardinality of $\mathcal{Z}$ increases, the performance of the divergence test degrades. The same observation has been made in [9] for the Hoeffding test by analyzing the mean and variance of $D_{\text{KL}}(T_{Z^n}\|P)$ in the limit as $n \rightarrow \infty$.

For certain divergences in Ξ , we can obtain more precise asymptotics than (8) by tightening the $o(\sqrt{n})$ term. To this end, we shall first discuss the three conditions stated in Definition 2.

Condition 1: When D is the f -divergence D_f , it follows from a Taylor-series expansion of $D_f(T\|P)$ around $T = P$ that D_f satisfies (2) with $\eta = \frac{f''(1)}{2}$ [16, Th. 4.1]. Since both the α -divergence D_α and the KL divergence D_{KL} belong to the f -divergence class with $f''(1) = 1$, we obtain that D_α and D_{KL} satisfy (2) with $\eta = \frac{1}{2}$ [15].

Condition 2: It is well-known that the KL divergence D_{KL} satisfies Pinsker's inequality [4, Lemma 11.6.1]. From [17, Th. 3], it further follows that Pinsker's inequality can be generalized to many f -divergences under some conditions on f . In particular, it follows from [17, Cor. 6] that the α -divergence satisfies a Pinsker-type inequality when $-3 \leq \alpha \leq 3$, $\alpha \neq \pm 1$.

Condition 3: This condition is a Berry-Esseen-type convergence of the statistic $\eta^{-1}nD(T_{Z^n}\|P)$ to the chi-square distribution with $k-1$ degrees of freedom. It can be shown that, for both the α -divergence with $\alpha \neq 1$ and the KL divergence, (4) holds with $\eta = \frac{1}{2}$ and $\delta_n = \frac{1}{\sqrt{n}}$. See the extended version of this paper [13] for more details.

We conclude that D_{KL} and D_α with $-3 \leq \alpha \leq 3$ and $\alpha \neq \pm 1$ satisfy all three conditions in Definition 2 with $\delta_n = \frac{1}{\sqrt{n}}$ and thus belong to the class Ξ . We thus have the following corollary to Theorem 1.

Corollary 3: Let $0 < \epsilon < 1$. Consider the divergence test $T_n^D(r)$ with $D = \{D_{\text{KL}}, D_\alpha\}$, $-3 \leq \alpha \leq 3$, $\alpha \neq \pm 1$ for testing $H_0 : Z^n \sim P^n$ against the alternative $H_1 : Z^n \sim Q^n$, where $P, Q \in \mathcal{P}(\mathcal{Z})$ and $P \neq Q$. Recall that the cardinality of $\mathcal{Z}$ is $k \geq 2$. Then, the type-II error of $T_n^D(r_n)$ minimized over all threshold values r_n satisfying (5) can be characterized as

$$-\ln \beta_n(T_n^D(r_n)) = nD_{\text{KL}}(P\|Q) - \sqrt{nV(P\|Q)Q_{\chi^2, k-1}^{-1}(\epsilon)} + O(\ln n), \quad \text{as } n \rightarrow \infty.$$

When D is the KL divergence, the divergence test is the Hoeffding test. Hence, the above corollary recovers the second-order term of the Hoeffding test as a special case.

IV. PROOF OF THEOREM 1

A. Proof of Part 1

From (4), it follows that there exist $M_0 > 0$ and $N_0 \in \mathbb{N}$ such that, for $n \geq N_0$,

$$|P^n(\eta^{-1}nD(T_{Z^n}\|P) \geq c) - Q_{\chi^2, k-1}(c)| \leq M_0\delta_n. \quad (10)$$

For $0 < \epsilon < 1$, let

$$r_n = \frac{\eta}{n} Q_{\chi^2, k-1}^{-1}(\epsilon - M_0\delta_n). \quad (11)$$

Then, from (10), it follows that $\alpha_n(T_n^D(r_n)) \leq \epsilon$, $n \geq N_0$.

We next consider the type-II error. To this end, we define

$$A_D(r') \triangleq \{T \in \mathcal{P}(\mathcal{Z}) \mid D(T\|P) < r'\}, \quad r' > 0. \quad (12)$$

Then, for $n \geq N_0$,

$$\begin{aligned} \beta_n(T_n^D(r_n)) &= \sum_{\tilde{P} \in \mathcal{P}_n \cap A_D(r_n)} Q^n(\mathcal{T}(\tilde{P})) \\ &\leq \sum_{\tilde{P} \in \mathcal{P}_n \cap A_D(r_n)} \exp\{-nD_{\text{KL}}(\tilde{P}\|Q)\} \end{aligned} \quad (13)$$

where the last step follows from [4, Th. 11.1.4]. In (13), $\mathcal{P}_n$ is the set of types with denominator n and $\mathcal{T}(\tilde{P})$ is the type class of $\tilde{P}$, i.e., the set of sequences z^n with type $\tilde{P}$. We next derive a lower bound on $D_{\text{KL}}(\tilde{P}\|Q)$ for $\tilde{P} \in \mathcal{P}_n \cap A_D(r_n)$. To this end, we use the following auxiliary results.

By (3), for every $T \in A_D(r)$, we have $\|\mathbf{T} - \mathbf{P}\|_1 \leq \tilde{k}_1\sqrt{r}$ for some constant $\tilde{k}_1$. Since any two norms on a finite-dimensional Euclidean space are equivalent, this implies that there exists a constant $C_0 > 0$ such that $\|\mathbf{T} - \mathbf{P}\|_2 \leq C_0\sqrt{r}$. Thus, for any $T \in A_D(r_n)$ with r_n in (11),

$$\|\mathbf{T} - \mathbf{P}\|_2 = O(1/\sqrt{n}). \quad (14)$$

Next define

$$A_{\chi^2}(r') \triangleq \{T \in \mathcal{P}(\mathcal{Z}) \mid d_{\chi^2}(T, P) < r'\}, \quad r' > 0$$

and denote by $\bar{A}_{\chi^2}(r')$ the closure of $A_{\chi^2}(r')$. To proceed further, we use the following lemmas.

Lemma 4: Let r_n be given in (11). Then, there exist $M > 0$ and $N_1 \geq N_0$ such that

$$A_D(r_n) \subseteq \bar{A}_{\chi^2}\left(\frac{r_n}{\eta} + \frac{M}{\eta n^{3/2}}\right), \quad n \geq N_1. \quad (15)$$

Proof: See [13, Lemma 4]. ■

Lemma 5: For any two probability distributions Q and T , the second-order Taylor approximation of $D_{\text{KL}}(T\|Q)$ around the null hypothesis $T = P$ is given by

$$\begin{aligned} D_{\text{KL}}(T\|Q) &= D_{\text{KL}}(P\|Q) + \sum_{i=1}^k (T_i - P_i) \ln\left(\frac{P_i}{Q_i}\right) \\ &\quad + \frac{1}{2}d_{\chi^2}(T, P) + O(\|\mathbf{T} - \mathbf{P}\|_2^3), \quad \text{as } \|\mathbf{T} - \mathbf{P}\|_2 \rightarrow 0. \end{aligned} \quad (16)$$

Proof: See [13, Lemma 3]. ■

Next, let us consider the function

$$\ell(\Gamma) \triangleq \sum_{i=1}^k (\Gamma_i - P_i)\alpha_i, \quad \alpha_i \triangleq \ln\left(\frac{P_i}{Q_i}\right), \quad \Gamma \in \mathcal{P}(\mathcal{Z}). \quad (17)$$

Then, for $n \geq N_1$ and $\tilde{P} \in \mathcal{P}_n \cap A_D(r_n)$, we have that

$$\ell(\tilde{P}) \geq \min_{\Gamma \in \mathcal{P}_n \cap A_D(r_n)} \ell(\Gamma) \geq \min_{\Gamma \in \bar{A}_{\chi^2}\left(\frac{r_n}{\eta} + \frac{M}{\eta n^{3/2}}\right)} \ell(\Gamma) \quad (18)$$

where the second inequality follows from (15).

To compute the minimum in (18), we define the following quantities:

$$\mathcal{I} \triangleq \{i = 1, \dots, k \mid \alpha_i - D_{\text{KL}}(P\|Q) > 0\} \quad (19)$$

$$\tau \triangleq \max_{j \in \mathcal{I}} \alpha_j - D_{\text{KL}}(P\|Q). \quad (20)$$

We then have the following result.

Lemma 6: Let $0 < \sqrt{\tilde{r}} < \frac{\sqrt{V(P\|Q)}}{\sqrt{V(P\|Q)}}$. Then, the probability distribution Γ^* that minimizes $\ell(\tilde{T})$ over $\bar{A}_{\chi^2}(\tilde{r})$ is given by

$$\Gamma_i^* = P_i + \frac{\sqrt{\tilde{r}}(D_{\text{KL}}(P\|Q) - \alpha_i)P_i}{\sqrt{V(P\|Q)}}, \quad i = 1, \dots, k. \quad (21)$$

Moreover,

$$\min_{\Gamma \in \bar{A}_{\chi^2}(\tilde{r})} \ell(\Gamma) = \ell(\Gamma^*) = -\sqrt{V(P\|Q)}\tilde{r}. \quad (22)$$

Proof: See [13, Lemma 5]. ■

For any $T \in A_D(r_n)$ with r_n in (11), (14) and (16) yield

$$D_{\text{KL}}(T\|Q) = D_{\text{KL}}(P\|Q) + \ell(T) + \frac{1}{2}d_{\chi^2}(T, P) + O(n^{-3/2}).$$

Consequently, there exist $M_2 > 0$ and $\tilde{N}_2 \in \mathbb{N}$ such that, for all $n \geq \tilde{N}_2$ and $T \in A_D(r_n)$,

$$D_{\text{KL}}(T\|Q) \geq D_{\text{KL}}(P\|Q) + \ell(T) - \frac{M_2}{n^{3/2}} \quad (23)$$

where we have also used that $d_{\chi^2}(T, P)$ is non-negative. Let

$$r'_n = \frac{r_n}{\eta} + \frac{M}{\eta n^{3/2}}. \quad (24)$$

Since r'_n vanishes as $n \rightarrow \infty$, we can choose $\tilde{N}$ such that, for $n \geq \tilde{N}$, r'_n satisfies $\sqrt{r'_n} < \tau^{-1}\sqrt{V(P\|Q)}$ with τ defined in (20). Then, it follows from (18) and (22) that, for $\tilde{P} \in \mathcal{P}_n \cap A_D(r_n)$ and $n \geq N_2 \triangleq \max\{N_1, \tilde{N}_2, \tilde{N}\}$, (23) can be further lower-bounded as

$$D_{\text{KL}}(\tilde{P}\|Q) \geq D_{\text{KL}}(P\|Q) - \sqrt{V(P\|Q)}r'_n - \frac{M_2}{n^{3/2}}. \quad (25)$$

Applying (25) to (13), and using that $|\mathcal{P}_n| \leq (n+1)^{|\mathcal{Z}|}$ [4, Th. 11.1.1], the type-II error can then be upper-bounded as

$$\begin{aligned} \beta_n(T_n^D(r_n)) &\leq (n+1)^{|\mathcal{Z}|} \exp\left\{-nD_{\text{KL}}(P\|Q)\right. \\ &\quad \left.+ n\sqrt{V(P\|Q)}r'_n + \frac{M_2}{\sqrt{n}}\right\}, \quad n \geq N_2. \end{aligned} \quad (26)$$

By (24) and (11),

$$\sqrt{r'_n} \leq \frac{1}{\sqrt{n}}\sqrt{Q_{\chi^2, k-1}^{-1}(\epsilon)} + O(1/n) + O(\delta_n/\sqrt{n}). \quad (27)$$

Consequently, taking logarithms on both sides of (26), and using (27), we obtain that

$$\begin{aligned} \ln \beta_n(T_n^D(r_n)) &\leq -nD_{\text{KL}}(P\|Q) + \sqrt{nV(P\|Q)}Q_{\chi^2, k-1}^{-1}(\epsilon) \\ &\quad + O(\max\{\delta_n\sqrt{n}, \ln n\}). \end{aligned} \quad (28)$$

This proves Part 1 of Theorem 1.

B. Proof of Part 2

Define for $0 < \epsilon < 1$

$$r_n^\epsilon \triangleq \inf \{r > 0 \mid P^n(D(T_{Z^n} \| P) \geq r) \leq \epsilon\}. \quad (29)$$

We have the following results:

Lemma 7: For M_0 and δ_n given in (10), we have

$$r_n^\epsilon = \frac{\eta}{n} Q_{\chi^2, k-1}^{-1}(\epsilon) + O(\delta_n/n). \quad (30)$$

Proof: See [13, Lemma 6]. $\blacksquare$

Lemma 8: There exist $M'_1 > 0$ and $N'_1 \in \mathbb{N}$ such that

$$\bar{A}_{\chi^2} \left(\frac{r_n^\epsilon}{\eta} - \frac{M'_1}{\eta n^{3/2}} \right) \subseteq A_D(r_n^\epsilon), \quad n \geq N'_1. \quad (31)$$

Proof: See [13, Lemma 7]. $\blacksquare$

For any $T \in A_D(r_n^\epsilon)$, it follows from (3) and (16) that $\|\mathbf{T} - \mathbf{P}\|_2 = O(1/\sqrt{n})$ and that there exist $M'_2 > 0$ and $N'_2 \in \mathbb{N}$ such that, for $n \geq N'_2$,

$$|D_{\text{KL}}(T \| Q) - D_{\text{KL}}(P \| Q) - \ell(T) - \frac{1}{2} d_{\chi^2}(T, P)| \leq \frac{M'_2}{n^{3/2}}. \quad (32)$$

If $r < r_n^\epsilon$ then, by definition of r_n^ϵ , the type-I error $\alpha_n(T^D)$ exceeds ϵ . Hence, such a threshold violates (5). We thus assume without loss of optimality that $r \geq r_n^\epsilon$. In this case,

$$\beta_n(T_n^D(r)) \geq Q^n(D(T_{Z^n} \| P) < r_n^\epsilon).$$

Together with Lemma 8, this implies that, for $n \geq N'_1$, the type-II error can be lower-bounded as

$$\beta_n(T_n^D(r)) \geq \sum_{\tilde{P} \in \mathcal{P}_n \cap \bar{A}_{\chi^2}(\bar{r}_n)} Q^n(\mathcal{T}(\tilde{P})) \quad (33)$$

where

$$\bar{r}_n = \frac{r_n^\epsilon}{\eta} - \frac{M'_1}{\eta n^{3/2}}. \quad (34)$$

By (30), $\bar{r}_n$ vanishes as $n \rightarrow \infty$. We can therefore find an $\tilde{N}'$ such that, for $n \geq \tilde{N}'$, we have $\sqrt{\bar{r}_n} < \tau^{-1} \sqrt{V(P \| Q)}$ with τ defined in (20). To lower-bound the type-II error further, let us consider the following lemma.

Lemma 9: Consider the minimizing probability distribution Γ^* given in (21) in Lemma 6, namely

$$\Gamma_i^* = P_i + \frac{\sqrt{\bar{r}_n}(D_{\text{KL}}(P \| Q) - \alpha_i)P_i}{\sqrt{V(P \| Q)}}, \quad i = 1, \dots, k \quad (35)$$

with $\bar{r}_n$ given in (34), α_i given in (17), and $n \geq \tilde{N}'$. Then, there exist $\tilde{N} \in \mathbb{N}$ and a type distribution $T_n^* \in \mathcal{P}_n \cap \bar{A}_{\chi^2}(\bar{r}_n)$ such that

$$|n\ell(\Gamma^*) - n\ell(T_n^*)| \leq \kappa, \quad n \geq \tilde{N} \quad (36)$$

for some constant $\kappa > 0$.

Proof: See [13, Lemma 8]. $\blacksquare$

From (33), the type-II error of the divergence test can be lower-bounded as

$$\begin{aligned} \beta_n(T_n^D(r)) &\geq Q^n(\mathcal{T}(T_n^*)) \\ &\geq \frac{1}{(n+1)^{|\mathcal{Z}|}} \exp\{-nD_{\text{KL}}(T_n^* \| Q)\}, \quad n \geq \tilde{N}_1 \end{aligned} \quad (37)$$

where $\tilde{N}_1 \triangleq \max\{N'_1, N'_2, \tilde{N}\}$, and $T_n^* \in \mathcal{P}_n \cap \bar{A}_{\chi^2}(\bar{r}_n)$ is a type distribution satisfying (36). The second inequality in (37) follows from [4, Th. 11.1.4].

We next upper-bound $D_{\text{KL}}(T_n^* \| Q)$. To this end, we use (32) and (36), and that $\ell(\Gamma^*) = -\sqrt{V(P, Q)}\bar{r}_n$, to obtain

$$\begin{aligned} nD_{\text{KL}}(T_n^* \| Q) &\leq nD_{\text{KL}}(P \| Q) + n\ell(T_n^*) + \frac{1}{2}nd_{\chi^2}(T_n^*, P) + \frac{M'_2}{\sqrt{n}} \\ &\leq nD_{\text{KL}}(P \| Q) - n\sqrt{V(P, Q)}\bar{r}_n + \kappa + \frac{n\bar{r}_n}{2} + \frac{M'_2}{\sqrt{n}} \end{aligned}$$

for $n \geq \tilde{N}_1$. Substituting the above equation in (37) and taking logarithms, we get for $n \geq \tilde{N}_1$

$$\begin{aligned} \ln \beta_n(T_n^D(r)) &\geq -|\mathcal{Z}| \ln(n+1) - nD_{\text{KL}}(P \| Q) \\ &\quad + n\sqrt{V(P \| Q)}\bar{r}_n - \kappa - \frac{n\bar{r}_n}{2} - \frac{M'_2}{\sqrt{n}}. \end{aligned} \quad (38)$$

By (30) and (34),

$$\sqrt{\bar{r}_n} = \frac{1}{\sqrt{n}} \sqrt{Q_{\chi^2, k-1}^{-1}(\epsilon)} + O(1/n) + O(\delta_n/\sqrt{n}). \quad (39)$$

It follows that (38) can be written as

$$\begin{aligned} -\ln \beta_n(T_n^D(r)) &\leq nD_{\text{KL}}(P \| Q) - \sqrt{nV(P \| Q)Q_{\chi^2, k-1}^{-1}(\epsilon)} \\ &\quad + O(\max\{\delta_n\sqrt{n}, \ln n\}). \end{aligned}$$

This proves Part 2 of Theorem 1.

V. CONCLUSIONS

For the divergence test and for a large class of divergences, we established the first-order and the second-order terms of the type-II error given that the type-I error is upper-bounded by a given value. The divergence test does not require the knowledge of the distribution of the alternate hypothesis, hence, it is suitable for problems where we only have access to the distribution of the null hypothesis. The class of divergences considered in this paper includes well-known divergences such as the KL divergence and the α -divergence, and our divergence test specializes to the Hoeffding test if the chosen divergence is the KL divergence. It is well-known that the Hoeffding test is first-order optimal in the sense that its first-order term is equal to the first-order term of the Neyman-Pearson test. However, our results demonstrate that the second-order term of the divergence test for the class of divergences considered in this paper, and therefore also of the Hoeffding test, is strictly smaller than the second-order term of the Neyman-Pearson test. The question whether there exists a test that only requires the knowledge of the distribution of the null hypothesis and that achieves a higher second-order performance than the Hoeffding test is yet to be explored.

ACKNOWLEDGMENT

The authors would like to thank Shun Watanabe for his helpful comments during the initial phase of this work and for pointing out relevant literature.

REFERENCES

- [1] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, vol. 231, no. 694-706, pp. 289–337, Feb. 1933.
- [2] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
- [3] V. Y. F. Tan, "Asymptotic estimates in information theory with non-vanishing error probabilities," *Foundations and Trends® in Communications and Information Theory*, vol. 11, no. 1-2, pp. 1–184, 2014.
- [4] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley-Interscience, 2006.
- [5] W. Hoeffding, "Asymptotically optimal tests for multinomial distributions," *The Annals of Mathematical Statistics*, pp. 369–401, Apr. 1965.
- [6] I. Csiszár, "The method of types," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2505–2523, Oct. 1998.
- [7] S. Watanabe, "Second-order optimal test in composite hypothesis testing," in *Proc. 2018 International Symposium on Information Theory and Its Applications (ISITA)*, Singapore, Oct. 2018, pp. 722–726.
- [8] O. Zeitouni, J. Ziv, and N. Merhav, "When is the generalized likelihood ratio test optimal?" *IEEE Transactions on Information Theory*, vol. 38, no. 5, pp. 1597–1602, Sep. 1992.
- [9] J. Unnikrishnan, D. Huang, S. P. Meyn, A. Surana, and V. V. Veeravalli, "Universal and composite hypothesis testing via mismatched divergence," *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1587–1603, Mar. 2011.
- [10] S. S. Wilks, "The large-sample distribution of the likelihood ratio for testing composite hypotheses," *The Annals of Mathematical Statistics*, vol. 9, no. 1, pp. 60–62, 1938.
- [11] T. R. Read, "Closer asymptotic approximations for the distributions of the power divergence goodness-of-fit statistics," *Annals of the Institute of Statistical Mathematics*, vol. 36, pp. 59–69, Dec. 1984.
- [12] J. K. Yarnold, "Asymptotic approximations for the probability that a sum of lattice random vectors lies in a convex set," *The Annals of Mathematical Statistics*, pp. 1566–1580, Oct. 1972.
- [13] K. V. Harsha, J. Ravi, and T. Koch, "Second-order asymptotics of Hoeffding-like hypothesis tests," [arXiv:2205.05631 \[cs.IT\]](https://arxiv.org/abs/2205.05631), May 2022.
- [14] S.-I. Amari, *Information geometry and its applications*. Springer, 2016, vol. 194.
- [15] S.-I. Amari and A. Cichocki, "Information geometry of divergence functions," *Bulletin of the Polish Academy of Sciences. Technical Sciences*, vol. 58, no. 1, pp. 183–195, 2010.
- [16] I. Csiszár and P. C. Shields, *Information Theory and Statistics: A Tutorial*. Now Publishers Inc, 2004.
- [17] G. L. Gilardoni, "On Pinsker's and Vajda's type inequalities for Csiszár's f -divergences," *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5377–5386, Nov. 2010.