LA-UR-10-02355

Approved for public release;
distribution is unlimited.

Title: Hypothesis Testing for Validation and Certification

Author(s): Clint Scovel, Ingo Steinwart

Intended for: Journal of Complexity

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the Los Alamos National Security, LLC for the National Nuclear Security Administration of the U.S. Department of Energy under contract DE-AC52-06NA25396. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

Form 836 (7/06)

This page intentionally left blank.

Hypothesis Testing for Validation and Certification

Clint Scovel
Los Alamos National Laboratory
jcs@lanl.gov

Ingo Steinwart
Universität Stuttgart
Ingo.Steinwart@mathematik.uni-stuttgart.de

April 10, 2010

Abstract

We develop a hypothesis testing framework for the formulation of the problems of 1) the *validation* of a simulation model and 2) using modeling to *certify* the performance of a physical system. These results are used to solve the extrapolative validation and certification problems, namely problems where the regime of interest is different than the regime for which we have experimental data. We use concentration of measure theory to develop the tests and analyze their errors. This work was stimulated by the work of Lucas, Owhadi, and Ortiz [1] where a rigorous method of validation and certification is described and tested. In Remark 2.5 we describe the connection between the two approaches. Moreover, as mentioned in that work these results have important implications in the Quantification of Margins and Uncertainties (QMU) framework. In particular, in Remark 2.6 we describe how it provides a rigorous interpretation of the notion of *confidence* and new notions of *margins* and *uncertainties* which allow this interpretation. Since certain *concentration parameters* used in the above tests may be unkown, we furthermore show, in the last half of the paper, how to derive equally powerful tests which estimate them from sample data, thus replacing the assumption of the values of the concentration parameters with weaker assumptions.

1 Introduction

Validation of simulation models is clearly important and much substantial work has been directed towards it, see e.g. [2, 3, 4, 5, 6, 7, 8] and the references therein. Moreover, the problem appears to go straight to the heart of the philosophy of science (see e.g. [9, 10, 11]). Indeed, [12] assert that validation is impossible, and [1] describe a rigorous method for it. On the other hand, it appears that while all agree that validation is an important and difficult problem, few agree on what the problem actually is. In the words of G. K. Chesterton [13, pg. ix], "It isn't that they can't see the solution. It is that they can't see the problem." In this paper we formulate examples of both the problems of validation and certification as problems of constructing hypothesis tests. A straightforward analysis using concentration of measure theory then provides tests and guarantees on their performance.

Although hypothesis tests have been used in validation before, e.g. in [14, 15], our formulation is quite different. In particular, we formulate null and alternate hypotheses which represent a flexibility in the customer's specification of a performance design threshold. We develop tests that require a clear delineation of assumptions and then use concentration of measure inequalities to analyze the performance of the tests. These results are then used to solve the extrapolative validation and certification problems, namely problems where the deployment regime is different than the experimental regime. This framework is then compared with that of Lucas, Owhadi

and Ortiz [1]. As mentioned in that work, these results also have important implications in the Quantification of Margins and Uncertainties (QMU) framework discussed in detail in [16, 17, 18]. In particular, in Remark 2.6 we discuss how these results provide a rigorous interpretation of the notion of *confidence* and a new notion of *uncertainties* which allow this interpretation. Since certain *concentration parameters* used in the above tests may be unknown, we furthermore show how to derive equally powerful tests which estimate them from sample data, thus replacing the assumption of the values of the concentration parameters with much weaker assumptions. This humble beginning needs to be refined so that it fits better with real applications. It should also incorporate some of the conclusions and structure of the above-mentioned works, but we leave that for the future. Let us now describe our framework and formulations. At a high level we say that *validation* is the assessment of the quality of a model of a physical system and *certification* is using modeling to assess the performance of a physical system. To make these notions more specific, consider the following general framework which will be used for both validation and certification.

Consider the case of a real-valued random variable U that describes the performance of a system and a customer who would like to have a quantitative guarantee on this performance. You inform the customer that you can consider a test of the hypothesis

$$\mathbb{P}\big(U \geq a\big) \geq p$$

where a is the performance design threshold and p is a level of confidence. When pressed to provide the specific values of the parameters a and p the customer may provide values, for example $a = 1000$ and $p = .95$. However, if you then ask him whether $a = 950$ and $p = .93$ would be acceptable, he might respond in the affirmative. Consequently, a more realistic test might be to test

$$\mathbb{P}\big(U \geq A\big) \geq P \tag{1}$$

where A and P are sets instead of real numbers. However, what (1) actually means and how to construct and analyze a test for it are not clear. To resolve this problem, let us introduce some notation. Let $\mathcal{U}$ denote the set of real-valued random variables. For $a \in \mathbb{R}$, $p \in (0,1)$ define the null hypothesis by

$$\mathcal{H}_{a,p} \;:=\; \big\{U \in \mathcal{U} : \mathbb{P}(U \geq a) \geq p\big\}$$

and the alternative by

$$\mathcal{K}_{a,p} \;:=\; \big\{U \in \mathcal{U} : \mathbb{P}(U \geq a) < p\big\}.$$

Consider $a' \leq a$, $p' \leq p$ and suppose that $U \in \mathcal{H}_{a,p} \cap \mathcal{K}_{a',p'}$. Then, since

$$p' > \mathbb{P}(U \geq a') \geq \mathbb{P}(U \geq a) \geq p$$

is a contradiction, we conclude that

$$\mathcal{H}_{a,p} \cap \mathcal{K}_{a',p'} = \emptyset, \quad a' \leq a,\, p' \leq p\,. \tag{2}$$

Therefore, when $a' \leq a$ and $p' \leq p$ we can consider a test of $\mathcal{H}_{a,p}$ against $\mathcal{K}_{a',p'}$. Now let a and p be specified and specify tolerance intervals A and P such that $A \leq a$ and $P \leq p$, where the notation implies that $a \in A$ and $p \in P$. Then by (2) we can define a test of (1) by testing $\mathcal{H}_{a,p}$ against $\mathcal{K}_{a',p'}$ for some $a' \in A$ and $p' \in P$. Given the freedom the tolerance intervals allow in the choice of a' and p', we seek to choose them to our advantage.

Let us first consider the case where $A = \{a\}$, namely there is no tolerance to changing the design criterion. We wish to construct a test of $\mathcal{H}_{a,p}$ against $\mathcal{K}_{a,p'}$ for $p' \in P$. Let $U_i, i = 1,..,n$ be i.i.d. samples from U. We can form a test by composing the sample data $U_i, i = 1,..,n$ with

the indicator function $I_a : \mathbb{R} \to \{0,1\}$ defined by $I_a(u) = 1, u \geq a$ and $I_a(u) = 0, u < a$ to obtain Bernoulli random variables $I_a \circ U_i$. That is, we simply evaluate whether the sample points are greater than or equal to a or not. We form a test of $\mathcal{H}_{a,p}$ against $\mathcal{K}_{a',p'}$ by forming the binomial test of $\mathcal{H}_p$ against $\mathcal{K}_{p'}$ where

$$\mathcal{H}_p \ := \ \{X : \mathbb{P}(X = 1) = r, \, \mathbb{P}(X = 0) = 1 - r, \, r \geq p\}$$

and

$$\mathcal{K}_{p'} \ := \ \{X : \mathbb{P}(X = 1) = r, \, \mathbb{P}(X = 0) = 1 - r, \, r < p'\}.$$

By the Neyman-Pearson Lemma [19, Thm. 3.1] and [19, Thm. 3.2] we know there exists a uniformly most powerful test of $\mathcal{H}_p$ against $\mathcal{K}_{p'}$ (see e. g. [19, Ch. 3]). However, this uniformly most powerful test is characterized through the binomial distribution. The statement of approximate tests with rigorous guarantees on their type I and II errors appears, in principle, to be available but evidently it is no easy matter. Rigorous bounds connecting the binomial distribution to the normal can be found in Feller [20] and to the Poisson distribution in Anderson and Samuels [21]. Guarantees outside of the range of applicability of these results can be found in Slud [22]. Approximations to the optimal test parameters have been derived and studied empirically in Shore [23, 24] and Chernoff [25] has analyzed the asymptotics, in particular when p' is close to p. Although a comprehensive rigorous analysis of this case should be completed, that is not our goal here. Instead we consider the case where $P = \{p\}$, where there is no tolerance to the value p, but a nontrivial tolerance in the design criteria A. That is, we test $\mathcal{H}_{a,p}$ against $\mathcal{K}_{a',p}$ for some $a' \in A$. For simplicity we remove the p from the notation of the hypothesis spaces, that is, from now on $\mathcal{H}_{a,p}$ and $\mathcal{K}_{a',p}$ are denoted by $\mathcal{H}_a$ and $\mathcal{K}_{a'}$ respectively. We will show that reducing the spaces of random variables further allows the development and analysis of efficient tests and that this analysis is quite elementary. The full problem of testing $\mathcal{H}_{a,p}$ against $\mathcal{K}_{a',p'}$ for $a' \in A$ and $p' \in P$ where both tolerance intervals are nontrivial might be accomplished through a combination of the above mentioned analysis and the results herein. To reduce the null and alternative hypothesis sets we will consider random variables U which are generated as $U = F(X)$ by real functions $F : X \to \mathbb{R}$ where each X is a vector random variable. We make assumptions on this set of functions and vector random variables that guarantee the degree of concentration of U about its mean in terms of a concentration parameter $\mathcal{D}$ (all this will be clarified below). We denote by $\mathcal{U}_{\mathcal{D}}$ the resulting space of real-valued random variables and reduce the null and alternative hypothesis spaces accordingly. Having performed this reduction, we will demonstrate how to construct tests of $\mathbb{P}(U \geq A) \geq p$ in terms of $\mathcal{D}$ and describe their type I and II errors. In addition, we observe that if A is large enough compared to $\mathcal{D}$ we can obtain tests with small type I and II errors. We disregard measurability considerations. In many applications, we want to validate a model or certify a physical system in the deployment regime where the real physical system is impossible or expensive to sample. In Section 3 we obtain the first results, as far as we can tell, for this extrapolation problem.

To apply these results to validation, we let $U = F(X)$ denote a measure of a model's fit to a physical system with respect to a quantity of interest. For example if, for the value x of the random variable X, the model predicts the strength of a material to be $s_M(x)$ and the physical system obtains the strength $s_{Ph}(x)$ then we might define $F(x) := \frac{1}{|s_M(x) - s_{Ph}(x)|}$. Then surpassing the performance threshold a is equivalent to $|s_M(x) - s_{Ph}(x)| \leq \frac{1}{a}$. We apply the above mentioned result to obtain a solution to the validation problem of constructing a test of $\mathcal{H}_a$ against $\mathcal{K}_{a'}$ using samples from $F(X)$ which has small type I and II errors. To apply this result to certification, we let $F(X)$ be the performance of the physical system and $M(X)$ be the performance of the physical system predicted by the model. For example, let $F(x) := s_{Ph}(x)$ be the strength of the physical system and $M(x) := s_M(x)$ be the strength of the physical system simulated by the model. We

apply the above mentioned result to obtain a solution to the certification problem of constructing a test of $\mathcal{H}_a$ against $\mathcal{K}_{a'}$ using samples from $F(X)$ and $M(X)$ which has small type I and II errors. Using the above mentioned tests, we observe in a quantitative way the intuitive result that if the validation diameter $\mathcal{D}_{F-M}$ is much smaller than the model diameter $\mathcal{D}_M$, then we need much fewer samples of the real physical system F than the model M to certify the performance of F. In Remark 2.5 we describe the connection to the rigorous validation and certification results of Lucas, Owhadi, and Ortiz [1]. These results generalize easily to other concentration inequalities. In particular, using concentration theorems for non *i.i.d.* sampling we can, with a substantial increase in complexity, obtain good tests when the empirical data are not generated *i.i.d.* or when the components of the random vector X are not independent. These tests and bounds on their performance require knowing the values of the diameter $\mathcal{D}_F$ for validation and $\mathcal{D}_{F-M}$ and $\mathcal{D}_M$ for certification. Since good approximations to these values may not be known in practice, we show, beginning in Section 4, how to estimate them to derive equally powerful tests, replacing the assumption of the values of the concentration parameters with much weaker assumptions. These tests provide validation and certification tests with *estimated diameters*.

2 Validation and Certification with Known Diameters

Let us first describe the concentration parameter $\mathcal{D}$ mentioned above, Let $(\Omega, \mathcal{F}, \mathbb{P})$ denote a probability space and consider a product space $\mathcal{X} = \mathcal{X}^1 \times \cdots \times \mathcal{X}^m$. We call a mapping $X : \Omega \to \mathcal{X}$ a random vector with range $\mathcal{X}$ and will abuse notation by also using the symbol $\mathbb{P}$ for the image probability measure on $\mathcal{X}$. For a function $F : \mathcal{X} \to \mathbb{R}$ we define the partial diameters to be

$$D_j^F = \sup_{x_k = x'_k,\, k \neq j} \big(F(x) - F(x')\big) \quad j = 1, .., m \tag{3}$$

where the supremum is taken over all $x, x' \in \mathcal{X}$ which differ only in their j-th component. Let $\mathcal{D}_F^2 := \sum_{j=1}^m (D_j^F)^2$ define the McDiarmid diameter $\mathcal{D}_F$ of the function F. For a vector random variable X and function $F : \mathcal{X} \to \mathbb{R}$ we consider the random variable $F \circ X : \Omega \to \mathbb{R}$ which we also denote by F. For the random variable F we have McDiarmid's inequality [26, Thm. 3.1, pg. 206]:

Theorem 2.1 *Let $\mathcal{X} = \mathcal{X}^1 \times \cdots \times \mathcal{X}^m$ be a Cartesian product and let $F : \mathcal{X} \to \mathbb{R}$ have the McDiarmid diameter $\mathcal{D}_F$. Then for any product probability measure $\mathbb{P} = \mu_1 \otimes \cdots \otimes \mu_m$ we have*

$$\mathbb{P}(F - \mathbb{E}F \geq r) \leq e^{-\frac{2r^2}{\mathcal{D}_F^2}}$$

If, for $0 < t < 1$, we define

$$r_t \;:=\; \frac{\mathcal{D}_F}{\sqrt{2}} \sqrt{\log t^{-1}}$$

then we have the following useful inequalities:

$$\begin{aligned} \mathbb{P}(F - \mathbb{E}F \geq r_t) &\leq t \\ \mathbb{P}(F - \mathbb{E}F > r_t) &< t \\ \mathbb{P}(F - \mathbb{E}F \leq -r_t) &\leq t \\ \mathbb{P}(F - \mathbb{E}F < -r_t) &< t. \end{aligned}$$

Since this theorem's only dependence on F and X is through the parameter $\mathcal{D}_F$ we can define the subset $\mathcal{U}_{\mathcal{D}} \subset \mathcal{U}$ consisting of all real-valued random variables generated as $U = F(X)$ for some F

and X such that $\mathcal{D}_F \leq \mathcal{D}$. Let $\mathcal{H}_a^{\mathcal{D}} := \mathcal{H}_a \cap \mathcal{U}_{\mathcal{D}}$ and $\mathcal{K}_a^{\mathcal{D}} := \mathcal{K}_a \cap \mathcal{U}_{\mathcal{D}}$ denote null and alternative generated in this way and consider testing $\mathcal{H}_a^{\mathcal{D}}$ against $\mathcal{K}_{a'}^{\mathcal{D}}$. We are now ready to state our main result which we then use to establish both validation and certification results. We describe a test of $\mathcal{H}_a^{\mathcal{D}}$ against $\mathcal{K}_{a'}^{\mathcal{D}}$ for $a - a'$ bounded below in terms of $\mathcal{D}$ and p. Therefore if $[a', a] \subset A$, then the following result provides a test of $\mathbb{P}\big(U \geq A\big) \geq p$, with bounds on its errors. Note that the test is in terms of the value of a function $F' : Y \to \mathbb{R}$ for some random vector Y with the only constraint being $\mathbb{E}F' = \mathbb{E}U$. All tests in this work accept the null $\mathcal{H}_a^{\mathcal{D}}$ by producing $T = 1$ and reject otherwise. Recall that the type I error is defined by $\theta_1(U) := \mathbb{P}(T = 0), U \in \mathcal{H}_a^{\mathcal{D}}$ and the type II error is defined by $\theta_2(U) := \mathbb{P}(T = 1), U \in \mathcal{K}_{a'}^{\mathcal{D}}$.

Theorem 2.2 *Let $0 < p < 1$, $\mathcal{D}, \mathcal{D}' > 0$ and consider $U \in \mathcal{U}_{\mathcal{D}}$. Moreover, consider a vector random variable Y and a function $F' : Y \to \mathbb{R}$ with diameter $\mathcal{D}_{F'} \leq \mathcal{D}'$ such that $\mathbb{E}F' = \mathbb{E}U$. For $0 < t < 1$ define $r_t := \frac{\mathcal{D}}{\sqrt{2}}\sqrt{\log t^{-1}}$ and $r'_t := \frac{\mathcal{D}'}{\sqrt{2}}\sqrt{\log t^{-1}}$. Let $0 < \delta_1, \delta_2 < 1$, and let a and a' satisfy $a - a' \geq r_p + r_{1-p} + r'_{\delta_1} + r'_{\delta_2}$ so that the interval $[a' + r_{1-p} + r'_{\delta_2}, a - r_p - r'_{\delta_1}]$ is nonempty. Let $b \in [a' + r_{1-p} + r'_{\delta_2}, a - r_p - r'_{\delta_1}]$. Then the test T of $\mathcal{H}_a^{\mathcal{D}}$ against $\mathcal{K}_{a'}^{\mathcal{D}}$ defined by*

$$T \;:=\; \begin{cases} 1, & F'(y) \geq b \\ 0, & F'(y) < b \end{cases}$$

satisfies

$$\theta_1(U) < \delta_1\,,$$
$$\theta_2(U) \leq \delta_2\,.$$

The condition $\mathbb{E}F' = \mathbb{E}U$ of Theorem 2.2 can be easily satisfied when i.i.d. samples are available. Therefore, in this case, it is straightforward to use Theorem 2.2 to define tests, with guarantees on their errors, for both *validation* and *certification*.

Corollary 2.3 (Validation) *Let $U = F(X)$ and suppose $\mathcal{D} \geq \mathcal{D}_F$. Let $F(X_i), i = 1, .., n$ be i.i.d. samples of $F(X)$ and define $\langle F \rangle_n := \frac{1}{n}\sum_{i=1}^n F(X_i)$ to be the sample mean. Let $0 < p, \delta_1, \delta_2 < 1$ and for $0 < t < 1$ define $r_t := \frac{\mathcal{D}}{\sqrt{2}}\sqrt{\log t^{-1}}$. Moreover, let a and a' satisfy $a - a' \geq r_p + r_{1-p} + \frac{1}{\sqrt{n}} r_{\delta_1} + \frac{1}{\sqrt{n}} r_{\delta_2}$ so that the interval $[a' + r_{1-p} + \frac{1}{\sqrt{n}} r_{\delta_2}, a - r_p - \frac{1}{\sqrt{n}} r_{\delta_1}]$ is nonempty. Let $b \in [a' + r_{1-p} + \frac{1}{\sqrt{n}} r_{\delta_2}, a - r_p - \frac{1}{\sqrt{n}} r_{\delta_1}]$ and consider the test T of $\mathcal{H}_a^{\mathcal{D}}$ against $\mathcal{K}_{a'}^{\mathcal{D}}$ defined by*

$$T \;:=\; \begin{cases} 1\,, & \langle F \rangle_n \geq b \\ 0\,, & \langle F \rangle_n < b\,. \end{cases}$$

Then we have

$$\theta_1(U) < \delta_1\,,$$
$$\theta_2(U) \leq \delta_2\,.$$

As discussed in the introduction, if $F(X)$ represents a physical system and $M(X)$ a model of that system we can consider how to test the performance of F by decomposing $F(X) = M(X) + \big(F(X) - M(X)\big)$ into the model component and the model deviation component. If the test accepts we obtain certification. We now show how to sample the model and the model deviation to test the performance of the physical system. In particular, the following result shows that if the validation diameter $\mathcal{D}_{F-M}$ is much smaller then the model diameter $\mathcal{D}_M$, then we need much fewer samples of the real physical system F than the model M to certify the performance of F. It is phrased in terms of a general decomposition $F = F_1 + F_2$.

Corollary 2.4 (Certification) *Let $U = F(X)$ where $F := F_1 + F_2$ is the sum of two functions with diameters $\mathcal{D}_{F_1}$ and $\mathcal{D}_{F_2}$. Let $\mathcal{D}, \mathcal{D}_1$, and $\mathcal{D}_2$ satisfy $\mathcal{D}_1 \geq \mathcal{D}_{F_1}$, $\mathcal{D}_2 \geq \mathcal{D}_{F_2}$ and $\mathcal{D} \geq \mathcal{D}_1 + \mathcal{D}_2$. Let $F_1(X_i), i = 1, .., n_1$ be i.i.d. samples of $F_1(X)$ and define $\langle F_1 \rangle_{n_1} := \frac{1}{n_1} \sum_{i=1}^{n_1} F_1(X_i)$ to be the sample mean. Also let $F_2(X_i), i = n_1 + 1, .., n_1 + n_2$ be i.i.d. samples of $F_2(X)$ and define $\langle F_2 \rangle_{n_2} := \frac{1}{n_2} \sum_{i=n_1+1}^{n_1+n_2} F_2(X_i)$ to be the sample mean of the second set of samples. For $0 < t < 1$ define*

$$\rho_t \;:=\; \frac{1}{\sqrt{2}} \sqrt{\frac{\mathcal{D}_1^2}{n_1} + \frac{\mathcal{D}_2^2}{n_2}} \sqrt{\log t^{-1}}$$

and

$$r_t \;:=\; \frac{\mathcal{D}}{\sqrt{2}} \sqrt{\log t^{-1}}$$

Let $0 < \delta_1, \delta_2 < 1$ and suppose that a and a' satisfy $a - a' \geq r_p + r_{1-p} + \rho_{\delta_1} + \rho_{\delta_2}$ so that the interval $[a' + r_{1-p} + \rho_{\delta_2}, a - r_p - \rho_{\delta_1}]$ is nonempty. Let $b \in [a' + r_{1-p} + \rho_{\delta_2}, a - r_p - \rho_{\delta_2}]$ and consider the test T of $\mathcal{H}_a^{\mathcal{D}}$ against $\mathcal{K}_{a'}^{\mathcal{D}}$ defined by

$$T \;:=\; \begin{cases} 1\,, & \langle F_1 \rangle_{n_1} + \langle F_2 \rangle_{n_2} \geq b \\ 0\,, & \langle F_1 \rangle_{n_1} + \langle F_2 \rangle_{n_2} < b\,. \end{cases}$$

Then $U \in \mathcal{U}_{\mathcal{D}}$ and we have

$$\theta_1(U) < \delta_1\,,$$

$$\theta_2 U) \leq \delta_2\,.$$

Moreover, suppose $\mathcal{D}_2 \leq \mathcal{D}_1$, $n_2 \geq \frac{\mathcal{D}_2}{\mathcal{D}_1} n_1$ and define

$$\rho_t \;:=\; \frac{\mathcal{D}_1}{\sqrt{n_1}} \sqrt{\log t^{-1}} \tag{4}$$

in the above conditions on a, a' and the test parameter b. Then for any $\mathcal{D} \geq 2\mathcal{D}_1$ we have $U \in \mathcal{U}_{\mathcal{D}}$, and

$$\theta_2(U) < \delta_1\,,$$

$$\theta_2(U) \leq \delta_2\,.$$

Remark 2.5 (Connection with Lucas, Owhadi and Ortiz [1]) If in Theorem 2.2 we define $b := a' + r_{1-p} + r'_{\delta_2}$ in terms of a performance value a' the condition for acceptance in Theorem 2.2 can be written $\langle F \rangle_n - a' - r'_{\delta_2} \geq r_{1-p}$ which amounts to

$$\frac{\langle F \rangle_n - a' - \frac{\mathcal{D}'}{\sqrt{2}} \sqrt{\log \delta_2{}^{-1}}}{\mathcal{D}} \geq \sqrt{\frac{1}{2} \log (1-p)^{-1}}\,.$$

For the case of validation in Corollary 2.3 we have $\mathcal{D}_{F'} \leq \frac{1}{\sqrt{n}} \mathcal{D}_F$ and so with $p = 1 - \epsilon$, $\delta_2 = \epsilon'$, $\mathcal{D} := \mathcal{D}_F$, and $\mathcal{D}' := \frac{1}{\sqrt{n}} \mathcal{D}_F$ the test of Corollary 2.3 amounts essentially (with a $\sqrt{2}$ better multiplicative factor in last term on the left) to the validation criterion of Lucas, Owhadi, and Ortiz [1, Eqn. 40] for the exact model with single performance measure (their Scenario 3). For certification, we can apply Corollary 2.4 with the choice F_1 as their F and F_2 as their $G - F$. Using the inequality $\mathcal{D}_{F_1+F_2} \leq \mathcal{D}_{F_1} + \mathcal{D}_{F_2}$ and setting $n_1 = n_2$ and $\delta_2 = 2\epsilon'$ we again obtain essentially (in a similar way as mentioned above) the certification criteria of [1, Eqns. 58&59]. Moreover, Corollaries 2.3 and 2.4 show we can interpret the certification criteria of [1] as guarantees that the

type II error is less than δ_2. If we then select a such that $a - a' \geq r_p + r_{1-p} + \rho_{\delta_1} + \rho_{\delta_2}$ we can also assert that the type I error is less than δ_1. In particular, if the design parameter value a' can tolerate being moved so that $a - a' \geq r_p + r_{1-p} + \rho_{\delta_1} + \rho_{\delta_2}$ with δ_1 and δ_2 small, this criterion amounts to a hypothesis test with type I and II errors bounded by δ_1 and δ_2 respectively. In this sense the criteria of [1] appear to correspond with our hypothesis test but with the roles of the hypothesis spaces $\mathcal{H}_a$ and $\mathcal{K}_{a'}$ reversed.

Remark 2.6 (Connection with QMU) For a detailed discussion of the QMU framework please see [16, 17, 18]. In the QMU framework, the confidence is evaluated in terms of a ratio $\frac{M}{U}$ where M is a margin and U is an uncertainty. The National Research Council of the National Academies report [16, Finding 1-1] states that "QMU is a sound and valuable framework that aids the assessment and evaluation of the confidence in the nuclear weapons stockpile." However it also states "There are serious and difficult problems to be resolved in uncertainty quantification, however, including the physical phenomena that are modeled crudely or not at all, the possibility of unknown unknowns, lack of computing power to guarantee the convergence of codes, and insufficient attention to validating experiments. Finally, they state that *"Even if the uncertainties arising from all of the different sources were estimated, their aggregation into an overall uncertainty for a given quantity of interest is a problem that needs further attention."* Although we do not suggest that we can answer all these question now we can make some conclusions along these lines using the discussion of Remark 2.5. For validation, consider the ratio

$$\frac{\langle F \rangle_n - a'}{\mathcal{D}}$$

where the numerator is a "margin" and the denominator is an "uncertainty". The inequality

$$\frac{\langle F \rangle_n - a'}{\mathcal{D}} \geq \sqrt{\frac{1}{2}\log{(1-p)^{-1}}} + \frac{1}{\sqrt{2n}}\sqrt{\log {\delta_2}^{-1}}$$

shows two things. First it shows how we can interpret confidence. That is, if $\mathbb{P}(F \geq a') < p$, namely if the performance is insufficient, then with probability less than δ_2 will we accept the performance as sufficient. Namely, our confidence is δ_2. Moreover, the precise definition of the uncertainty parameter $\mathcal{D}$ shows how this parameter is aggregated so as to maintain the interpretation of the confidence statement. For the certification problem similar comments also apply but we get the added benefit of seeing how modeling uncertainties and validation uncertainties are aggregated and combined and how they influence the number of validation experiments needed compared to the number of modeling runs.

We have used McDiarmid's inequality Theorem 2.1 as the model for concentration in this paper, but that is not necessary. All that was needed is a concentration parameter $\mathcal{D}_F$ which scales a certain way with sampling. In particular, concentration theorems that do not require i.i.d. sampling, for example the martingale difference inequality [26, Thm. 3.14, Page 224], can be applied to derive results similar, but more complex, to those obtained. Another example of a concentration theorem is the following for Lipschitz functions, [27, Cor. 1.17]:

Theorem 2.7 *Let $\mathcal{X} = \mathcal{X}^1 \times \cdots \times \mathcal{X}^m$ be the Cartesian product of metric spaces (X_i, d_i) with diameters $D_i, i = 1, .., m$ and let $\mathcal{D}_{\mathcal{X}}^2 := \sum_{i=1}^m D_i^2$. Let $F : \mathcal{X} \to \mathbb{R}$ be Lipschitz with respect the ℓ_1 metric $d := \sum_{i=1}^m d_i$ with Lipschitz constant $|F|$. Then for any product probability measure $\mathbb{P} = \mu_1 \otimes \cdots \otimes \mu_m$ we have*

$$\mathbb{P}(F - \mathbb{E}F \geq r) \leq e^{-\frac{r^2}{2|F|^2\mathcal{D}_{\mathcal{X}}^2}}.$$

The following, easy to prove, proposition shows that the previous results also apply using the Lipschitz concentration Theorem 2.7.

Proposition 2.8 *Consider the concentration result and notation of Theorem 2.7. Then Theorem 2.2 and Corollaries 2.3 and 2.4 hold with $\mathcal{D}$ replaced by $2|F|\mathcal{D}_{\mathcal{X}}$.*

3 Extrapolative Validation and Certification

In this section we consider when we want to validate a model or certify a physical system in a regime where the real physical system is impossible or expensive to sample. That is, suppose we wish to validate or certify a random variable $\hat{F}(\hat{X})$ which is expensive or impossible to sample but are able to sample a related random variable $F(X)$. When samples from $\hat{F}(\hat{X})$ are unavailable we have the following validation result in terms of the Kolmogorov distance

$$d(F,\hat{F}) \;:=\; \max_b \left|\mathbb{P}(F \le b) - \mathbb{P}(\hat{F} \le b)\right| \tag{5}$$

between two random variables F and $\hat{F}$. The corresponding certification result is very similar, but we omit it for brevity.

Theorem 3.1 (Extrapolative Validation) *Let $U = F(X)$ have McDiarmid diameter $\mathcal{D}_F$. Let $F(X_i), i = 1,..,n$ be i.i.d. samples of $F(X)$ and define $\langle F\rangle_n := \frac{1}{n}\sum_{i=1}^n F(X_i)$ to be the sample mean. Let $0 < p < 1$, $0 < \delta_p < \min(p, 1-p)$ and suppose that $\hat{F} : \hat{\mathcal{X}} \to \mathbb{R}$ satisfies*

$$d(F,\hat{F}) \le \delta_p\,.$$

Let $0 < \delta_1, \delta_2 < 1$ and for $0 < t < 1$ define $r_{\mathcal{H}} := \frac{1}{\sqrt{2}}\sqrt{\log(p-\delta_p)^{-1}} + \frac{1}{\sqrt{2n}}\sqrt{\log\delta_1^{-1}}$ and $r_{\mathcal{K}} := \frac{1}{\sqrt{2}}\sqrt{\log(1-p-\delta_p)^{-1}} + \frac{1}{\sqrt{2n}}\sqrt{\log\delta_2^{-1}}$. Then if $a - a' \ge \mathcal{D}_F(r_{\mathcal{H}} + r_{\mathcal{K}})$ the test of $\{\mathbb{P}(\hat{F} \ge a) \ge p\}$ versus $\{\mathbb{P}(\hat{F} \ge a') < p\}$ defined by

$$T \;:=\; \begin{cases} 1\,, & \langle F\rangle_n \ge a - \mathcal{D}_F r_{\mathcal{H}} \\ 0\,, & \langle F\rangle_n < a - \mathcal{D}_F r_{\mathcal{H}} \end{cases}$$

satisfies

$$\theta_1 \le \delta_1\,,$$
$$\theta_2 \le \delta_2\,.$$

When samples from $\hat{F}$ are available but more expensive than samples from F, we can use the sample data to estimate the Kolmogorov distance between F and $\hat{F}$ and then corporate the estimate in the test as discussed in Section 4 and afterwords. The following estimate is efficient in the sense that it uses the concentration of the Kolmogorov-Smirnov statistic of Dvoretzky, Kiefer and Wolfowitz [28] improved to have a tight constant by Massart [29] (see also [30, Thm. 12.9]) as follows: Let n i.i.d samples be taken from F and let $\mathbb{P}_n$ denote its empirical measure and let $n' \le n$ i.i.d samples be taken from $\hat{F}$ and let $\mathbb{P}_{n'}$ denote its empirical measure. Then the Dvoretzky, Kiefer and Wolfowitz Theorem states that

$$\mathbb{P}^n\Big(\sup_{b\in\mathbb{R}}\left|\mathbb{P}_n(F \le b) - \mathbb{P}(F \le b)\right| > \varepsilon\Big) \le 2e^{-2n\varepsilon^2}$$

and

$$\mathbb{P}^{n'}\Big(\sup_{b\in\mathbb{R}}\left|\mathbb{P}_{n'}(\hat{F} \le b) - \mathbb{P}(\hat{F} \le b)\right| > \varepsilon\Big) \le 2e^{-2n'\varepsilon^2}.$$

Let us define

$$d_{n,n'}(F,\hat{F}) \;:=\; \sup_{b\in\mathbb{R}} \left|\mathbb{P}_n(F\le b) - \mathbb{P}_{n'}(\hat{F}\le b)\right|$$

as an estimator of the Kolmogorov distance $d(F,\hat{F})$ defined in (5). Then since

$$\begin{aligned} |d(F,\hat{F}) - d_{n,n'}(F,\hat{F})| &= \left|\sup_{b\in\mathbb{R}} \left|\mathbb{P}(F\le b) - \mathbb{P}(\hat{F}\le b)\right| - \sup_{b\in\mathbb{R}} \left|\mathbb{P}_n(F\le b) - \mathbb{P}_{n'}(\hat{F}\le b)\right|\right| \\ &\le \sup_{b\in\mathbb{R}} \left|\mathbb{P}_n(F\le b) - \mathbb{P}(F\le b)\right| + \sup_{b\in\mathbb{R}} \left|\mathbb{P}_{n'}(\hat{F}\le b) - \mathbb{P}(\hat{F}\le b)\right| \end{aligned}$$

we use $n' \le n$ to conclude by a simple union bound that

$$\begin{aligned} & \mathbb{P}^{n+n'}\big(|d(F,\hat{F}) - d_{n,n'}(F,\hat{F})| > \varepsilon\big) \\ \le\; & \mathbb{P}^{n}\big(\sup_{b\in\mathbb{R}} \left|\mathbb{P}_n(F\le b) - \mathbb{P}(F\le b)\right| > \frac{\varepsilon}{2}\big) + \mathbb{P}^{n'}\big(\sup_{b\in\mathbb{R}} \left|\mathbb{P}_{n'}(\hat{F}\le b) - \mathbb{P}(\hat{F}\le b)\right| > \frac{\varepsilon}{2}\big) \\ \le\; & 4e^{-\frac{1}{2}n'\varepsilon^2}\,. \end{aligned}$$

That is, we have

$$\mathbb{P}^{n+n'}\big(|d(F,\hat{F}) - d_{n,n'}(F,\hat{F})| > \varepsilon\big) \le 4e^{-\frac{1}{2}n'\varepsilon^2}\,.$$

whose confidence form is

$$\mathbb{P}^{n+n'}\left(\left|d(F,\hat{F}) - d_{n,n'}(F,\hat{F})\right| > \sqrt{\frac{2\ln 4 + 2\ln\delta^{-1}}{n'}}\right) \le \delta\,. \tag{6}$$

This estimation inequality (6) can be used, along the lines of Section 4 and afterword, to prove a version of Theorem 3.1 where the estimate $d_{n,n'}(F,\hat{F})$ is used instead of the Kolmogorov distance $d(F,\hat{F})$. Moreover, since the test and its performance depend logarithmically on this estimate, we should be able to obtain good tests where n' is *much smaller* than n. In particular, we should be able to obtain good tests *if* the Kolmogorov distance is small enough- instead of by assuming that it is so. However, for brevity, we do not complete this program here but move to the estimation of diameters in validation and certification tests.

4 Estimation of Diameters in Hypothesis Tests

The validation and certification results, Corollaries 2.3 and 2.4, require the value of the diameter $\mathcal{D}_F$ for validation and $\mathcal{D}_{F-M}$ and $\mathcal{D}_M$ for certification. In principle the modeling and domain experts should have much to say about bounding these values. However, sample data should also say something about them. With the eventual goal of combining expert knowledge about the relevant diameters with information from sample data, we now proceed to describe how sample data can be used to estimate these diameters. This will be accomplished through an estimation procedure and the introduction of "higher order" concentration parameters. To that end, we now invert the concentration theorem to its "confidence version" so that the diameters appear inside the probability statement. This allows the comparison of the diameter with an estimable parameter and a mechanism for incorporating estimates of these parameters in the concentration theorems and therefore into the definitions of tests and the analysis of their performance.

4.1 Diameters in Concentration Theorems

By a simple function inversion, McDiarmid's inequality can be written

$$\mathbb{P}\Big(F - \mathbb{E}F \geq f(\mathcal{D}_F, \delta)\Big) \leq \delta \tag{7}$$

where $f(r, \delta) := \frac{r}{\sqrt{2}}\sqrt{\log \delta^{-1}}$. This inversion was used in the proof of the main Theorem 2.2. The following two lemmas reformulate those parts of Theorem 2.2 which we will use as basic building blocks for developing validation and certification tests with estimated diameters.

Lemma 4.1 *Let $0 < p < 1$ and $a, a' \in \mathbb{R}$ and consider the functions $f_H : \mathbb{R}^3 \to \mathbb{R}$ and $f_K : \mathbb{R}^3 \to \mathbb{R}$ defined by*

$$\begin{aligned} f_H(r, r', \delta) &:= \frac{r}{\sqrt{2}}\sqrt{\log p^{-1}} + \frac{r'}{\sqrt{2}}\sqrt{\log \delta^{-1}} - a\,, \\ f_K(r, r', \delta) &:= \frac{r}{\sqrt{2}}\sqrt{\log (1-p)^{-1}} + \frac{r'}{\sqrt{2}}\sqrt{\log \delta^{-1}} + a'\,. \end{aligned}$$

Then for all $0 < \delta < 1$ and all $F, F' \in \mathcal{U}$ which satisfy $\mathbb{E}F' = \mathbb{E}F$, we have

$$\begin{aligned} \mathbb{P}(F' \leq -f_H(\mathcal{D}_F, \mathcal{D}_{F'}, \delta) | F \in \mathcal{H}_a) &\leq \delta\,, \\ \mathbb{P}(F' \geq f_K(\mathcal{D}_F, \mathcal{D}_{F'}, \delta) | F \in \mathcal{K}_{a'}) &\leq \delta\,. \end{aligned}$$

The following simple lemma shows how to use the results of Lemma 4.1 to construct hypothesis tests with controlled errors. It is formulated in terms of the primary variable F, a test variable F', and a vector $\vec{F}$ of auxiliary variables.

Lemma 4.2 *Let $\mathcal{H}, \mathcal{K} \subset \mathcal{U}$ be null and alternative hypothesis spaces and let $k \in \mathbb{N}$ and $0 < \delta_1, \delta_2 < 1$. Consider functions $g_K, g_H : \mathcal{U}^k \to \mathbb{R}$ such that for all $F, F' \in \mathcal{U}$ there exists a vector $\vec{F}$ of auxiliary random variables $F^j, j = 1, .., k$ such that*

$$\mathbb{P}\Big(F' \leq -g_H(\vec{F}) \big| F \in \mathcal{H}\Big) \leq \delta_1,$$

$$\mathbb{P}\Big(F' \geq g_K(\vec{F}) \big| F \in \mathcal{K}\Big) \leq \delta_2.$$

We call any such vector $\vec{F}$ admissible for F, F'. Now suppose $F, F' \in \mathcal{U}$ and consider any admissible and vector $\vec{F}$. Consider the test T of $F \in \mathcal{H}$ against $F \in \mathcal{K}$ defined by

$$T := \begin{cases} 1\,, & F' > -g_H(\vec{F}) \\ 0\,, & F' \leq -g_H(\vec{F}) \end{cases} \tag{8}$$

Then if $g_H(\vec{F}) + g_K(\vec{F}) \leq 0$ we have

$$\theta_1(T) \leq \delta_1,$$

$$\theta_2(T) \leq \delta_2.$$

In general, concentration theorems can be used to establish results like Lemma 4.1 and then Lemma 4.2 can be used to establish a test and bound its errors. In particular, we see how the main theorem, Theorem 2.2, with test point fixed at the right-hand side of the interval, can then

be obtained by a combined application of Lemmas 4.1 and 4.2: first apply Lemma 4.1 and then apply Lemma 4.2 with $\mathcal{H} := \mathcal{H}_a, \mathcal{K} := \mathcal{H}_{a'}, \vec{F} := (F, F')$ and

$$g_H(\vec{F}) = g_H(F, F') \ := \ f_H(\mathcal{D}_F, \mathcal{D}_{F'}, \delta_1)\,,$$

$$g_K(\vec{F}) = g_K(F, F') \ := \ f_K(\mathcal{D}_F, \mathcal{D}_{F'}, \delta_2)\,.$$

However, what is important here is that we are now in a position define tests which use estimates of $f_H(\mathcal{D}_F, \mathcal{D}_{F'}, \delta_1)$ and $f_K(\mathcal{D}_F, \mathcal{D}_{F'}, \delta_2)$. Since we see no efficient way of estimating the McDiarmid diameter $\mathcal{D}_F$ of a function F but we do know something about the estimation of the usual diameter D_G of a function G defined by

$$D_G \ := \ \sup_{x,x' \in \mathcal{X}} \big(G(x) - G(x')\big),$$

we ask whether we can estimate the McDiarmid diameter by estimating the usual diameters of a set of auxiliary set of functions. To that end we first introduce a relationship between the McDiarmid diameter and the usual diameters of a set of auxiliary observables. These latter diameters we will then estimate using extreme value estimators in Section 4.2. Now, ignoring for the moment the question of the attainment of suprema, if we define

$$F^j(x_j) \ := \ F(x_1^*, .., x_{j-1}^*, x_j, x_{j+1}^*, .., x_m^*)$$

where

$$\begin{aligned} & (x_1^*, .., x_{j-1}^*, x_{j+1}^*, .., x_m^*) \\ := \quad & \arg \max_{x_1,..,x_{j-1},x_{j+1},..,x_m} \max_{x_j, x_j'} \Big(F(x_1, .., x_{j-1}, x_j, x_{j+1}, .., x_m) - F(x_1, .., x_{j-1}, x_j', x_{j+1}, .., x_m)\Big) \end{aligned}$$

it follows that

$$\mathcal{D}_F^2 = \sum_{j=1}^m D_{F^j}^2\,.$$

Namely the McDiarmid diameter is a function of diameters. However, this relation will only be of use to us if the functions $F^j, j = 1, .., k$ are observable, namely, they can be evaluated. Now suppose we are in possession of a set $F^j, j = 1, .., k$ auxiliary observables and let $\vec{D}$ denote the vector of their diameters. Suppose we also have functions g_H and g_K such that

$$f_H(\mathcal{D}_F, \mathcal{D}_{F'}, \delta) \le g_H(\vec{D}, \delta)\,, \quad 0 < \delta < 1,$$

$$f_K(\mathcal{D}_F, \mathcal{D}_{F'}, \delta) \le g_K(\vec{D}, \delta)\,, \quad 0 < \delta < 1.$$

Then since Lemma 4.1 asserts that for all $0 < \delta < 1$ we have

$$\begin{aligned} \mathbb{P}^n\Big(F' \le -f_H\big(g(D), g'(D), \delta\big) \big| \mathcal{H}\Big) \ & \le \delta\,, \\ \mathbb{P}^n\Big(F' \ge f_K\big(g(D), g'(D), \delta\big) \big| \mathcal{K}\Big) \ & \le \delta, \end{aligned}$$

it follows easily that for all $0 < \delta < 1$ we have

$$\mathbb{P}^n\Big(F' \le -g_H\big(\vec{D}, \delta\big) \big| \mathcal{H}\Big) \ \le \delta\,, \tag{9}$$

$$\mathbb{P}^n\Big(F' \ge g_K\big(\vec{D}, \delta\big) \big| \mathcal{K}\Big) \ \le \delta, \tag{10}$$

Consequently, we can apply Lemma 4.2 to obtain tests defined instead in terms of the estimable functions $g_H(\vec{D},\delta)$ and $g_K(\vec{D},\delta)$. Most importantly, *the inequalities (9) remain valid with the vector of diameters $\vec{D}$ replaced by the vector of essential diameters.* When the essential diameter is *much* smaller than the given diameter, this difference can often offset the looseness corresponding to the error associated with estimating the essential diameter using the empirical diameter.

Let us now give the first important example of auxiliary observables. In this case, they will be none other than the functions F, F' themselves, but will require the introduction of new functions, $c_F, c_{F'}$ of F and F' which will have to be approximately known. To that end, define a coefficient of the separability c_F of the function F with respect to the m components of $\mathcal{X} := \prod_{j=1}^m \mathcal{X}^j$ as follows:

Definition 4.3 Let $\mathcal{X} := \prod_{j=1}^m \mathcal{X}^j$ be a product and consider a function $F : \mathcal{X} \to \mathbb{R}$, its diameter D_F, and its McDiarmid diameter $\mathcal{D}_F$. We define the coefficient of separability c_F with respect to the product $\mathcal{X}$ to be

$$c_F \;:=\; \frac{\mathcal{D}_F}{D_F}.$$

With this definition it is clear that if we define $g_H(D_F, D_{F'}, \delta) := f_H(c_F D_F, c_{F'} D_{F'}, \delta)$ and $g_K(D_F, D_{F'}, \delta) := f_K(c_F D_F, c_{F'} D_{F'}, \delta)$, where we suppress the dependency on $c_F, c_{F'}$, we have

$$f_H(\mathcal{D}_F, \mathcal{D}_{F'}, \delta) = g_H(D_F, D_{F'}, \delta)\,, \quad 0 < \delta < 1,$$

$$f_K(\mathcal{D}_F, \mathcal{D}_{F'}, \delta) = g_K(D_F, D_{F'}, \delta)\,, \quad 0 < \delta < 1$$

and therefore

$$\begin{aligned} \mathbb{P}^n\Big(F' \le -g_H\big(D_F, D_{F'}, \delta\big)\big|\mathcal{H}\Big) &\le \delta\,, \\ \mathbb{P}^n\Big(F' \ge g_K\big(D_F, D_{F'}, \delta\big)\big|\mathcal{K}\Big) &\le \delta, \end{aligned}$$

Although we have now introduced a new function c_F which will have to be known or well bounded, this function has nice properties, which we now describe, which make assuming its value a weaker assumption than assuming the value of a McDiarmid diameter. Let us say that a map $\phi : \prod_{j=1}^m \mathcal{X}^j \to \prod_{j=1}^m \mathcal{X}'^j$ is a diagonal bijection if it is a product map $\phi = \prod_{j=1}^m \phi_j$ such that $\phi_j : \mathcal{X}^j \to \mathcal{X}'^j$ is a bijection for all $j = 1, .., m$. The following lemma shows that $F \mapsto c_F$ is a bounded invariant under non-singular affine transformations $F \mapsto aF + b$ of the function F and a diagonal bijective invariant.

Lemma 4.4 *The mapping $F \mapsto c_F$ is a diagonal bijective invariant. Moreover, we have*

$$c_{aF+b} = c_F, \quad a, b \in \mathbb{R}, a \neq 0$$

and

$$\frac{1}{\sqrt{m}} \le c_F \le \sqrt{m}.$$

In Example 7.2 in the Appendix we describe the attainment of the the extreme case $c_F = \frac{1}{\sqrt{m}}$ and $c_F = \sqrt{m}$: roughly, the lower bound is attained for functions which are separable in the m components and the upper bound is obtained for a function related to the Euclidean metric[1].

[1] In personal communication, L. Gurvits has demonstrated that nontrivial lower bounds may not exist when $\mathcal{X}$ is not a product. For example it is easy to construct cases where the partial diameters are all zero and the diameter is not.

Lemma 4.2 states that conditions such as

$$\mathbb{P}^n\Big(F' \leq -g_H(D)\big|\mathcal{H}\Big) \quad \leq \delta_1\,, \tag{11}$$

$$\mathbb{P}^n\Big(F' \geq g_K(D)\big|\mathcal{K}\Big) \quad \leq \delta_2. \tag{12}$$

and $g_H(D) + g_K(D) \leq 0$ for functions of the *essential* diameter vector of auxiliary observables are sufficient to develop a good test. In the above analysis this was accomplished by knowledge about the coefficients of separability $c_F, c_{F'}$ which allowed us to use the functions F and F' as their own auxiliary observables. However relations such as (11), where the inequalities are in terms of the usual diameters, can be obtained through other concentration inequalities. For example, if we instead appeal to the Lipschitz concentration Theorem 2.7, it is easy to obtain inequalities (11) from Lemma 4.1 with $\mathcal{D}_F$ replaced by $2|F|\mathcal{D}_X$. However, it is easy to show that

$$\mathcal{D}_F \leq |F|\mathcal{D}_{\mathcal{X}}$$

indicating that McDiarmid's Theorem 2.1 provides a superior concentration guarantee. On the other hand, since $\mathcal{D}_{\mathcal{X}}$ is a sum of diameters, and the supremum

$$|F| \;:=\; \sup_{x\neq x'} \frac{F(x)-F(x')}{d(x,x')}$$

can be estimated by the empirical Lipschitz coefficient

$$|\hat{F}| \;:=\; \sup_{X_i\neq X_{i'}, i,i'=1,..,n} \frac{F(X_i)-F(X_{i'})}{d(X_i,X_{i'})},$$

it follows that $|F|$ and $\mathcal{D}_{\mathcal{X}}$ can be estimated from sample data using Corollary 4.9 in Section 4.2. However, this will require the random variable X to be observable. Consequently, when $\mathcal{D}_F$ has no readily apparent auxiliary observables (such as when no knowledge of the coefficient of separability c_F is available), and X is observable, using the Lipschitz concentration Theorem 2.7 may prove fruitful.

4.2 Concentration of Empirical Quantiles

Since we will be concerned with the effects of estimating essential diameters using sample data, we now describe results, of independent interest, concerning the concentration of empirical quantiles about distributional quantiles and show how to use them to bound empirical diameters with respect to essential diameters. Let X be a real random variable with probability measure $\mathbb{P}$ and recall its distribution function $\mathbb{F}(\xi) := \mathbb{P}(X \leq \xi)$. For $0 < p < 1$ define the quantiles $\xi_p := \mathbb{F}^{-1}(p) := \inf\{\xi : \mathbb{F}(\xi) \geq p\}$. We will use important properties of $\mathbb{F}$ and $\mathbb{F}^{-1}$ listed in Theorem 7.1 in the Appendix. Moreover, let $X_i, i = 1, n$ be i.i.d samples from X. Let $\mathbb{P}_n$ denote the corresponding empirical measure, denote by $\mathbb{F}_n$ its corresponding distribution function, and let $\hat{\xi}_p := \inf\{\xi : \mathbb{F}_n(\xi) \geq p\}$ denote the empirical quantiles. We will use the following improvement of a theorem of Serfling [31, Thm. 2.3.2].

Theorem 4.5 *Let $0 < q < 1$ and suppose that $\xi > \xi_q$. Then with $\delta_1 := \mathbb{F}(\xi) - q$ we have*

i) $\mathbb{P}^n(\hat{\xi}_q > \xi) \leq e^{-2n\delta_1^2}$

ii) $\mathbb{P}^n(\hat{\xi}_q > \xi) \leq e^{-\frac{n\delta_1^2}{2(1-\mathbb{F}(\xi))+\frac{2}{3}\delta_1}}$

iii) $\mathbb{P}^n(\hat{\xi}_q > \xi) \leq e^{-\frac{n\delta_1^2}{2\mathbb{F}(\xi)}}$

On the other hand suppose that $\xi < \xi_q$. Then with $\delta_2 := q - \mathbb{F}(\xi)$ we have

i) $\mathbb{P}^n(\hat{\xi}_q < \xi) \leq e^{-2n\delta_2^2}$

ii) $\mathbb{P}^n(\hat{\xi}_q < \xi) \leq e^{-\frac{n\delta_2^2}{2\mathbb{F}(\xi)+\frac{2}{3}\delta_2}}$

iii) $\mathbb{P}^n(\hat{\xi}_q < \xi) \leq e^{-\frac{n\delta_2^2}{2(1-\mathbb{F}(\xi))}}$

Theorem 4.5 now gives us a good tool to compare empirical diameters with quantiles.

Theorem 4.6 *Let X be a real random variable and let $X_i, i = 1, .., n$ i.i.d. samples. For $0 < p < 1$, we have*

i) Let $D_n := \sup_{i=1,..,n} X_i - \inf_{i=1,..,n} X_i$ denote the empirical range. Then we have

$$\mathbb{P}^n\Big(D_n < \xi_p - \xi_{1-p}\Big) \leq 2e^{-\frac{1}{2}n(1-p)}.$$

ii) Suppose X is a non-negative random variable and let $S_n := \sup_{i=1,..,n} X_i$ denote the empirical supremum. Then we have

$$\mathbb{P}^n\Big(S_n < \xi_p\Big) \leq e^{-\frac{1}{2}n(1-p)}.$$

We now show how to use Theorem 4.6 to bound the the empirical diameters in terms of essential diameters. To that end, let $X_- := ess\inf X$ and $X_+ := ess\sup X$. Then the essential diameter is $D := X_+ - X_-$. We introduce a *tail function* quantifying the behavior of a random variable near its range limit.

Definition 4.7 Let the tail function τ^X corresponding to X be defined by

$$\tau^X(\epsilon) := \sup_{t:\, \xi_{1-t} - \xi_t \geq \frac{D}{1+\epsilon},} t \qquad \epsilon > 0 \tag{13}$$

$$t : \xi_{1-t} - \xi_t \geq \frac{D}{1+\epsilon}, \tag{14}$$

Roughly speaking the function $\tau^X(\epsilon)$ is such that the set obtained by eliminating the right and left tails of mass τ is at least $\frac{1}{1+\epsilon}$ as large as the diameter. Characterization of tail behaviors lies as the heart of the theory of the limiting behavior of extreme order statistics (see e.g. Arov and Bobrov[32], Pickands [33], Barndorff-Neilsen [34]) and will no doubt be useful when the diameters are unbounded, but since we concern ourselves with the bounded case here, the tail function (13) appear sufficient to our needs. The following proposition provides a lower bound for $\tau(\epsilon)$ in terms of the distribution function for X.

Proposition 4.8 *Let X be a real random variable and suppose that $X_- := ess\inf X$ and $X_+ := ess\sup X$ are finite. Then in terms of the essential diameter $D := X_+ - X_-$, we have*

$$\tau^X(\varepsilon) \geq \min\Big(\mathbb{F}(X_- + \frac{\varepsilon}{2(1+\varepsilon)}D), 1 - \mathbb{F}(X_+ - \frac{\varepsilon}{2(1+\varepsilon)}D)\Big), \quad \varepsilon > 0.$$

As an elementary application, consider the case where the tails are not too thin. That is suppose for some $\kappa > 0$ we have $\mathbb{F}(X_+ - x) < 1 - \left(\frac{x}{D}\right)^\kappa, 0 < x < D$ and $\mathbb{F}(X_- + x) \geq \left(\frac{x}{D}\right)^\kappa, 0 < x < D$. We conclude from Proposition 4.8 that

$$\tau^X(\epsilon) \geq \left(\frac{\epsilon}{2(1+\epsilon)}\right)^\kappa \geq \left(\frac{\epsilon}{4}\right)^\kappa$$

which for ϵ small is $t^X(\epsilon) \gtrapprox \left(\frac{\epsilon}{2}\right)^\kappa$. For non-negative random variables we proceed similarly to definition (13) and define

$$\tau_+^X(\epsilon) := \sup \delta \tag{15}$$

$$\xi_{1-\delta} \geq \tfrac{X_+}{1+\epsilon} \tag{16}$$

Similar arguments used in the proof of Proposition 4.8 imply that $\tau_+^X(\epsilon) \geq 1 - \mathbb{F}(\frac{X_+}{1+\epsilon})$. We are now in a position to compare the empirical diameter with the essential diameter using the tail function τ.

Corollary 4.9 *Let X be a real random variable and let $X_i, i = 1, .., n$ be i.i.d. samples from X. Then*

i) Let $D_n := \sup_{i=1,..,n} X_i - \inf_{i=1,..,n} X_i$ denote the empirical range and let τ be define by (13). Then for all $\varepsilon > 0$ we have

$$\mathbb{P}^n\left(D_n < \frac{D}{1+\epsilon}\right) \leq 2e^{-\frac{n\tau(\epsilon)}{2}}.$$

ii) Suppose X is a non-negative random variable and let $S_n := \sup_{i=1,..,n} X_i$ denote the empirical supremum and let τ_+ be define by (15). Then for all $\varepsilon > 0$ we have

$$\mathbb{P}^n\left(S_n < \frac{X_+}{1+\epsilon}\right) \leq e^{-\frac{n\tau_+(\epsilon)}{2}}.$$

4.3 Estimation in Hypothesis Tests

Lemma 4.2 and the discussion thereafter shows that when $f_H(D) + f_K(D) \leq 0$ (thus determining a relationship between the performance thresholds a, a' and the diameter D) the test of Lemma 4.2 of $\mathcal{H}$ against $\mathcal{K}$ has type I error not greater than δ_1 and type II error not greater than δ_2. However, when good upper bounds on D are not known and thus it is not known if $f_H(D) + f_K(D) \leq 0$, these results may be of limited value. To resolve this situation we use sample data to estimate D and use the estimate to test the condition $f_H(D) + f_K(D) \leq 0$. To develop validation and certification tests along the lines above will involve sequential tests. The type of test we consider we call a stop option hypothesis test:

Definition 4.10 For $i = 1, 2$ consider a null hypothesis $\mathcal{H}_i$ and alternative $\mathcal{K}_i$ of sets of real random variables, and a test T_i of $\mathcal{H}_i$ against $\mathcal{K}_i$. Define the reduced hypothesis spaces

$$\mathcal{H}_{2\epsilon} := (\mathcal{K}_1 \cup \mathcal{H}_1) \cap \mathcal{H}_2,$$

$$\mathcal{K}_{2\epsilon} := (\mathcal{K}_1 \cup \mathcal{H}_1) \cap \mathcal{K}_2.$$

We define the stop option test $T_1 \blacktriangleleft T_2$ which first implements T_1 and if the outcome is acceptance, to use T_2 to test $\mathcal{H}_{2\epsilon}$ against $\mathcal{K}_{2\epsilon}$:

$$T_1 \blacktriangleleft T_2 := \begin{cases} 0 & T_1 = 0 \\ (1,0) & T_1 = 1, T_2 = 0 \\ (1,1) & T_1 = 1, T_2 = 1 \end{cases}$$

All types of errors for the test $T_1 \blacktriangleleft T_2$ can be controlled by the following three types of errors:

$$\theta_1(T_1 \blacktriangleleft T_2) \;:=\; \mathbb{P}\big(T_1 = 0 \big| \mathcal{H}_1\big)$$

$$\theta_{11}(T_1 \blacktriangleleft T_2) \;:=\; \mathbb{P}\big(\{T_1 = 1, T_2 = 0\} \big| \mathcal{K}_1 \cup (\mathcal{H}_1 \cap \mathcal{H}_2)\big)$$
$$\theta_{12}(T_1 \blacktriangleleft T_2) \;:=\; \mathbb{P}\big(\{T_1 = 1, T_2 = 1\} \big| \mathcal{K}_1 \cup (\mathcal{H}_1 \cap \mathcal{K}_2)\big)$$

Since

$$\mathcal{K}_1 \cup (\mathcal{H}_1 \cap \mathcal{H}_2) = (\mathcal{K}_1 \cup \mathcal{H}_1) \cap \mathcal{H}_2 = \mathcal{H}_{2\epsilon}\,,$$
$$\mathcal{K}_1 \cup (\mathcal{H}_1 \cap \mathcal{K}_2) = (\mathcal{K}_1 \cup \mathcal{H}_1) \cap \mathcal{K}_2 = \mathcal{K}_{2\epsilon}\,,$$

it follows that

$$\theta_{11}(T_1 \blacktriangleleft T_2) = \mathbb{P}\big(\{T_1 = 1, T_2 = 0\} \big| \mathcal{H}_{2\epsilon})\big)\,,$$
$$\theta_{12}(T_1 \blacktriangleleft T_2) = \mathbb{P}\big(\{T_1 = 1, T_2 = 1\} \big| \mathcal{K}_{2\epsilon})\big)\,.$$

Consequently, the stop option test $T_1 \blacktriangleleft T_2$ converts tests of $\mathcal{H}_i$ against $\mathcal{K}_i, i = 1, 2$ into a test of $\mathcal{H}_1$ against $\mathcal{K}_1$ and if accepted then tests $\mathcal{H}_{2\epsilon}$ against $\mathcal{K}_{2\epsilon}$. Since

$$\begin{aligned}
\mathbb{P}\big(T_1 = 1 \big| \mathcal{K}_1\big) &= \mathbb{P}\big(\{T_1 = 1, T_2 = 0\} \big| \mathcal{K}_1\big) + \mathbb{P}\big(\{T_1 = 1, T_2 = 1\} \big| \mathcal{K}_1\big) \\
&\leq \mathbb{P}\big(\{T_1 = 1, T_2 = 0\} \big| \mathcal{K}_1 \cup (\mathcal{H}_1 \cap \mathcal{H}_2)\big) + \mathbb{P}\big(\{T_1 = 1, T_2 = 1\} \big| \mathcal{K}_1 \cup (\mathcal{H}_1 \cap \mathcal{K}_2)\big) \\
&= \theta_{11} + \theta_{12},
\end{aligned}$$

it follows that if all the errors $\theta_1, \theta_{11}, \theta_{12}$, are small, then given $\mathcal{K}_1$ with high probability we reject $\mathcal{H}_1$ and stop, and given $\mathcal{H}_1$ with high probability we accept $\mathcal{H}_1$ and test well on the second test T_2 when applied to the reduced null hypothesis $\mathcal{H}_{2\epsilon}$ against the reduced alternative $\mathcal{K}_{2\epsilon}$. Finally, we note that we can also define the errors to be conditional errors as in the conditional hypothesis testing framework analyzed in [35]. In the applications of this paper, one can show that given $\mathcal{H}_1$ the conditional errors are roughly the same as above, and given $\mathcal{K}_1$, the conditional errors are not good. However, in this case with high probability the first test will reject and stop.

We now proceed to implement the stop option test in the validation and certification setting. To simplify the analysis in the following theorem, instead of first testing $\mathcal{H}_1 = \{f_H(D) + f_K(D) \leq 0\}$ against $\mathcal{K}_1 = \{f_H(D) + f_K(D) > 0\}$, (where D is the essential diameter vector) we test $\mathcal{H}_1 = \{f_H((1+\epsilon)D) + f_K((1+\epsilon)D) \leq 0\}$ against $\mathcal{K}_1 = \{f_H(D) + f_K(D) > 0\}$. Also observe that this result is stated in terms of auxiliary variables which are sampled concomitantly with the sampling of the primary variable F. More general situations can be easily addressed.

Theorem 4.11 *Let $\mathcal{H}_2$ and $\mathcal{K}_2$ denote null and alternate hypothesis spaces of real random variables. Let X be a random variable with range $\mathcal{X}$ and probability law $\mathbb{P}$, and let $F : \mathcal{X} \to \mathbb{R}$ and $F' : \mathcal{X}^n \to \mathbb{R}$. Consider non-observable i.i.d. samples $X_i, i = 1, .., n$ and observable $F'(X_1, .., X_n)$. In addition, let k be a positive integer and let $F^j : \mathcal{X} \to \mathbb{R}, j = 1, .., k$ be a collection of auxiliary observables with essential diameters $D_{F^j}, j = 1, .., k$. Let $D := \langle D_{F^j} \rangle_{j=1,..,k}$ denote the corresponding vector of essential diameters,*

$$\hat{D}_{F^j} \;:=\; \sup_{i=1,..,n} F^j(X_i) - \inf_{i=1,..,n} F^j(X_i)$$

denote the empirical diameters, and $\hat{D} := \langle \hat{D}_{F^j} \rangle_{j=1,..,k}$ the vector of empirical diameters. Let $f_H : \mathbb{R}^k \to \mathbb{R}$ and $f_K : \mathbb{R}^k \to \mathbb{R}$ be non-decreasing functions such that

$$\begin{aligned}
\mathbb{P}^n(F' \leq -f_H(D) | \mathcal{H}_2) &\leq \delta_1, \\
\mathbb{P}^n(F' \geq f_K(D) | \mathcal{K}_2) &\leq \delta_2.
\end{aligned}$$

Let $\varepsilon > 0$ and define

$$\begin{aligned}\mathcal{H}_1 &= \{f_H((1+\epsilon)D) + f_K((1+\epsilon)D) \le 0\},\\ \mathcal{K}_1 &= \{f_H(D) + f_K(D) > 0\}.\end{aligned}$$

and define the test T_1 of $\mathcal{H}_1$ against the alternative $\mathcal{K}_1$ by

$$T_1 \;:=\; \begin{cases} 1\,, & f_H((1+\epsilon)\hat{D}) + f_K((1+\epsilon)\hat{D}) \le 0 \\ 0\,, & f_H((1+\epsilon)\hat{D}) + f_K((1+\epsilon)\hat{D}) > 0\,. \end{cases} \tag{17}$$

Moreover, define the test T_2 of $\mathcal{H}_2$ against $\mathcal{K}_2$ by

$$T_2 \;:=\; \begin{cases} 1\,, & F'(X_1,..,X_n) > -f_H((1+\epsilon)\hat{D}) \\ 0\,, & F'(X_1,..,X_n) \le -f_H((1+\epsilon)\hat{D})\,. \end{cases} \tag{18}$$

Finally, consider the stop option test $T_1 \blacktriangleleft T_2$ and it associated errors θ_1, θ_{11}, and θ_{12}, defined in Definition 4.10. Then if we define

$$\Delta \;:=\; \sum_{j=1}^{k} \mathbb{P}^n\Big((1+\epsilon)\hat{D}_{F^j} < D_{F^j}\Big),$$

we have

$$\begin{aligned}\theta_1 &= 0\\ \theta_{12} &\le \delta_1 + \Delta\\ \theta_{12} &\le \delta_2 + \Delta\,.\end{aligned}$$

Moreover, for $j = 1,..,k$, let τ^j denote the tail functions of F^j defined in (13), and let

$$n_j^\epsilon(\delta) \;:=\; \frac{2\log\frac{2k}{\delta}}{\tau^j(\epsilon)}\,. \tag{19}$$

Then if $n \ge \max\big(n_j^\epsilon(\delta_1), n_j^\epsilon(\delta_2)\big), j = 1,..,k$ we have

$$\begin{aligned}\theta_1 &= 0\\ \theta_{11} &\le 2\delta_1\\ \theta_{12} &\le 2\delta_2\,.\end{aligned}$$

The constraint $n \ge n^\epsilon(\delta)$ is logarithmic in δ^{-1} with multiplier $\frac{2}{\tau(\epsilon)}$. If $\tau(\epsilon)$ is not too small then this is a weak constraint. For fixed sample sizes, this relation can be used to determine a lower bound on the size of ϵ which can be used.

5 Validation and Certification with Estimated Diameters

We now present tests for validation and certification using estimated diameters. They show that if the coefficients of separability c are approximately known then the validation and certification Corollaries 2.3 and 2.4, using the estimated diameters, are still good.

Corollary 5.1 (Validation with Estimated Diameters) *Let $a' \leq a$, $0 < p < 1$ and let*

$$\mathcal{H}_2 = \{\mathbb{P}(F \geq a) \geq p\}$$

$$\mathcal{K}_2 = \{\mathbb{P}(F \geq a') < p\}.$$

With the assumptions of Theorem 4.11, let

$$F'(X_1, .., X_n) \ := \ \langle F \rangle_n$$

be the sample mean. Let $c \geq c_F$ be a known constant and consider $F : \mathcal{X} \to \mathbb{R}$ with essential diameter D as the only auxiliary observable. Let τ be the tail function of F defined in (13). Let $\varepsilon > 0$ and let $n^\epsilon(\delta)$ be defined in (19) with $k = 1$. Let $0 < \delta_1, \delta_2 < 1$ and define

$$f_H(r) \ := \ \frac{cr}{\sqrt{2}}\sqrt{\log p^{-1}} + \frac{cr}{\sqrt{2n}}\sqrt{\log \delta_1^{-1}} - a$$

$$f_K(r) \ := \ \frac{cr}{\sqrt{2}}\sqrt{\log (1-p)^{-1}} + \frac{cr}{\sqrt{2n}}\sqrt{\log \delta_2^{-1}} + a'$$

and let

$$\hat{D} \ := \ \sup_{i=1,..,n} F(X_i) - \inf_{i=1,..,n} F(X_i)$$

denote the empirical diameter. Moreover, consider the stop option test $T_1 \blacktriangleleft T_2$ and its associated errors θ_1, θ_{11}, and θ_{12}, as in Theorem 4.11. Then for $n \geq \max(n^\epsilon(\delta_1), n^\epsilon(\delta_2))$ we have

$$\begin{aligned} \theta_1 &= 0 \\ \theta_{11} &\leq 2\delta_1 \\ \theta_{12} &\leq 2\delta_2 \,. \end{aligned}$$

For certification, we now address the case where $F = F_1 + F_2$, where F_1 can represent the model and F_2 represent the difference between the physical system and the model.

Corollary 5.2 (Certification with Estimated Diameters) *Let $a' \leq a$, $0 < p < 1$ and let*

$$\mathcal{H}_2 = \{\mathbb{P}(F \geq a) \geq p\}$$

$$\mathcal{K}_2 = \{\mathbb{P}(F \geq a') < p\}.$$

With the assumptions of Theorem 4.11, let $F = F_1 + F_2$. Let $n = n_1 + n_2$ and in terms of the observation $F_1(X_i), i = 1, .., n_1$ and $F_2(X_i), i = n_1 + 1, n_1 + n_2$ define

$$F'(X_1, .., X_n) \ := \ \langle F_1 \rangle_{n_1} + \langle F_2 \rangle_{n_2}.$$

Let $c_1 \geq c_{F_1}$ and $c_2 \geq c_{F_2}$ be known constants and consider F_1 with essential diameter D_1 and F_2 with essential diameter D_2 as auxiliary observables, with diameter vector $D := (D_1, D_2)$. Let $\tau^j, j = 1, 2$ be the tail functions, defined in (13), of F_1 and F_2 respectively. Let $\varepsilon > 0$ and let $n_j^\epsilon(\delta), j = 1, 2$ be defined in (19) with $k = 2$. Let $\delta_1, \delta_2 < 1$ and define

$$f_H(s_1, s_2) \ := \ (c_1 s_1 + c_2 s_2)\sqrt{\log p^{-1}} + \sqrt{\left(\frac{c_1^2 s_1^2}{2n_1} + \frac{c_2^2 s_2^2}{2n_2}\right)\log \delta_1^{-1}} - a$$

$$f_K(s_1,s_2) \;:=\; (c_1 s_1 + c_2 s_2)\sqrt{\log(1-p)^{-1}} + \sqrt{\Big(\frac{c_1^2 s_1^2}{2n_1} + \frac{c_2^2 s_2^2}{2n_2}\Big)\log \delta_2^{-1}} + a'$$

and let

$$\hat{D}_1 \;:=\; \sup_{i=1,..,n_1} F_1(X_i) - \inf_{i=1,..,n_1} F_2(X_i)$$

$$\hat{D}_2 \;:=\; \sup_{i=n_1+1,..,n_1+n_2} F_2(X_i) - \inf_{i=n_1+1,..,n_1+n_2} F_2(X_i)$$

denote the empirical diameters with empirical diameter vector $\hat{D} := (\hat{D}_1, \hat{D}_2)$*. Moreover, consider the stop option test* $T_1 \blacktriangleleft T_2$ *and its associated errors* θ_1, θ_{11}*, and* θ_{12}*, as in Theorem 4.11. Then for* $n_j \geq \max\left(n_j^{\epsilon}(\delta_1), n_j^{\epsilon}(\delta_2)\right), j = 1,2$ *we have*

$$\begin{aligned} \theta_1 &= 0 \\ \theta_{11} &\leq 2\delta_1 \\ \theta_{12} &\leq 2\delta_2\,. \end{aligned}$$

6 Proofs

Proof of Theorem 2.2: To begin we first prove the following simple lemma that quantifies how the mass constraints of the null $\mathcal{H}_a^{\mathcal{D}}$ or the alternative $\mathcal{K}_a^{\mathcal{D}}$ imply a constraint on the value of $\mathbb{E}U$.

Lemma 6.1 *Let* $0 < p < 1$*,* $a \in \mathbb{R}$ *and* $\mathcal{D} > 0$*. For* $0 < t < 1$ *let* $r_t := \frac{\mathcal{D}}{\sqrt{2}}\sqrt{\log t^{-1}}$*. Suppose* $U \in \mathcal{H}_a^{\mathcal{D}}$*, then* $\mathbb{E}U \geq a - r_p$*. Suppose* $U \in \mathcal{K}_a^{\mathcal{D}}$*, then* $\mathbb{E}U \leq a + r_{1-p}$*.*

Proof: Let $U \in \mathcal{H}_a^D$ and suppose to the contrary that $\mathbb{E}U < a - r_p$. Then we have

$$p \leq \mathbb{P}(U \geq a) \leq \mathbb{P}(U > \mathbb{E}U + r_p) = \mathbb{P}(U - \mathbb{E}U > r_p) < p$$

which is a contradiction, thus establishing the first assertion. Now let $U \in \mathcal{K}_a^D$ and suppose to the contrary that $\mathbb{E}U > a + r_{1-p}$. Then

$$p > \mathbb{P}(U \geq a) \geq \mathbb{P}(U > \mathbb{E}U - r_{1-p}) = \mathbb{P}(U - \mathbb{E}U > -r_{1-p}) = 1 - \mathbb{P}(U - \mathbb{E}U \leq -r_{1-p}) \geq p$$

which is a contradiction, thus establishing the second assertion. ■

The confidence version of the following result essentially completes the proof of Theorem 2.2.

Lemma 6.2 *With the assumptions of Theorem 2.2 let* $a, a' \in \mathbb{R}$ *satisfy* $a - a' \geq r_p + r_{1-p}$ *so that the interval* $[a' + r_{1-p}, a - r_p]$ *is nonempty. Let* $b \in [a' + r_{1-p}, a - r_p]$ *and consider the test* T *of* $\mathcal{H}_a^{\mathcal{D}}$ *against* $\mathcal{K}_{a'}^{\mathcal{D}}$ *defined by*

$$T \;:=\; \begin{cases} 1, & F'(y) \geq b \\ 0, & F'(y) < b\,. \end{cases}$$

Then we have

$$\theta_1(U) < e^{-\frac{2((a-r_p)-b)^2}{\mathcal{D}'^2}},$$

$$\theta_2(U) \leq e^{-\frac{2(b-(a'+r_{1-p}))^2}{\mathcal{D}'^2}}.$$

Proof: Suppose $U \in \mathcal{H}_a^D$. Then by Lemma 6.1 we have

$$b \leq b - (a - r_p) + \mathbb{E}U = b - (a - r_p) + \mathbb{E}F'$$

so that by Theorem 2.1 applied to F', we conclude that the Type I error satisfies

$$\theta_1(U) = \mathbb{P}(F'(Y) < b) \leq \mathbb{P}(F' - \mathbb{E}F' < b - (a - r_p)) < e^{-\frac{2((a-r_p)-b)^2}{D_{F'}^2}} \leq e^{-\frac{2((a-r_p)-b)^2}{D'^2}}.$$

Similarly, if we suppose $U \in \mathcal{K}_{a'}^D$, by Lemma 6.1 we have

$$b \geq b - (a' + r_{1-p}) + \mathbb{E}U = b - (a' + r_{1-p}) + \mathbb{E}F'.$$

Consequently, the Type II error satisfies

$$\theta_2(U) = \mathbb{P}(F'(Y) \geq b) \leq \mathbb{P}(F' - \mathbb{E}F' \geq b - (a' + r_{1-p})) \leq e^{-\frac{2(b-(a'+r_{1-p}))^2}{D_{F'}^2}} \leq e^{-\frac{2(b-(a'+r_{1-p}))^2}{D'^2}}.$$

■

We are now ready to complete the proof of Theorem 2.2. It is easy to show that a, a' and b satisfy the assumptions of Lemma 6.2. Since it follows from the assumptions that $(a - r_p) - b \geq r'_\delta$ and $b - (a' + r_{1-p}) \geq r'_{\delta'}$ the assertion follows from Lemma 6.2. ■

Proof of Corollary 2.3: It is not hard to see that $\mathbb{E}F' = \mathbb{E}F = \mathbb{E}U$. Let the vector variable X have J components and index the J^n components of $\prod_{i=1}^n \mathcal{X}_i$ by the map $i, j \mapsto k_{ij} = J(i-1)+j, j = 1, .., J, i = 1, .., n$. First observe that if we define $F' : \prod_{i=1}^n \mathcal{X}_i \to \mathbb{R}$ by $F'(\prod_{i=1}^n X_i) := \langle F \rangle_n$ we have $\mathbb{E}F' = \mathbb{E}F = \mathbb{E}U$. Moreover, for all $j = 1, .., J, i = 1, .., n$ we have $D_{k_{ij}}^{F'} \leq \frac{1}{n} D_j^F, j = 1, .., J, i = 1, .., n$ and therefore

$$\mathcal{D}_{F'}^2 = \sum_{j=1,..,J, i=1,..,n} (D_{k_{ij}}^{F'})^2 \leq \frac{1}{n^2} \sum_{j=1,..,J, i=1,..,n} (D_j^F)^2 = \frac{1}{n} \sum_{j=1,..,J} (D_j^F)^2 = \frac{1}{n} \mathcal{D}_F^2$$

and so we conclude that $\mathcal{D}_{F'} \leq \frac{\mathcal{D}_F}{\sqrt{n}} \leq \frac{\mathcal{D}}{\sqrt{n}}$. The assertion follows from Theorem 2.2. ■

Proof of Corollary 2.4: As in the proof of Corollary 2.3 index the $J^{n_1+n_2}$ components of $\prod_{i=1}^{n_1+n_2} \mathcal{X}_i$ by the map $i, j \mapsto k_{ij} = J(i-1) + j, j = 1, .., J, i = 1, .., n_1 + n_2$. Since $D_j^F \leq D_j^{F_1} + D_j^{F_2}, j = 1, .., J$ it easily follows the triangle inequality in ℓ_2 and the assumptions that $\mathcal{D}_F \leq \mathcal{D}_{F_1} + \mathcal{D}_{F_2} \leq \mathcal{D}_1 + \mathcal{D}_2 \leq \mathcal{D}$. Consequently, $U \in \mathcal{U}_D$ and we can apply Theorem 2.2. To that end observe that $F' : \prod_{i=1}^{n_1+n_2} \mathcal{X}_i \to \mathbb{R}$ defined by $F'(\prod_{i=1}^{n_1+n_2} X_i) := \langle F_1 \rangle_{n_1} + \langle F_2 \rangle_{n_2}$ satisfies $\mathbb{E}F' = \mathbb{E}F_1 + \mathbb{E}F_2 = \mathbb{E}F = \mathbb{E}U$. Moreover, it follows that $D_{k_{ij}}^{F'} \leq \frac{1}{n_1} D_j^{F_1}$ if $1 \leq i \leq n_1$ and $D_{k_{ij}}^{F'} \leq \frac{1}{n_2} D_j^{F_2}$ if $n_1 + 1 \leq i \leq n_1 + n_2$. Consequently we conclude that

$$\begin{aligned} \mathcal{D}_{F'}^2 &= \sum_{j=1,..,J, 1 \leq i \leq n_1} (D_{k_{ij}}^{F'})^2 + \sum_{j=1,..,J, n_1+1 \leq i \leq n_1+n_2} (D_{k_{ij}}^{F'})^2 \\ &\leq \frac{1}{n_1^2} \sum_{j=1,..,J, 1 \leq i \leq n_1} (D_j^{F_1})^2 + \frac{1}{n_2^2} \sum_{j=1,..,J, n_1+1 \leq i \leq n_1+n_2} (D_j^{F_2})^2. \end{aligned}$$

Therefore, since

$$\sum_{j=1,..,J, 1 \leq i \leq n_1} (D_j^{F_1})^2 = n_1 \mathcal{D}_{F_1}^2 \leq n_1 \mathcal{D}_1^2$$

and

$$\sum_{j=1,..,J,n_1+1\leq i\leq n_1+n_2} (D_j^{F_2})^2 = n_2\mathcal{D}_{F_2}^2 \leq n_2\mathcal{D}_2^2$$

we conclude that

$$\mathcal{D}_{F'} \leq \sqrt{\frac{\mathcal{D}_1^2}{n_1} + \frac{\mathcal{D}_2^2}{n_2}}\,.$$

Consequently Theorem 2.2 implies the first assertion. For the second observe that $\mathcal{D}_1 \leq \mathcal{D}_2$ implies that $\mathcal{D}_F \leq \mathcal{D}_1 + \mathcal{D}_2 \leq 2\mathcal{D}_1$ which implies $U \in \mathcal{U}_D$. Moreover, setting $n_2 \geq \frac{\mathcal{D}_2}{\mathcal{D}_1} n_1$ implies that $\mathcal{D}_{F'} \leq \sqrt{\frac{\mathcal{D}_1^2}{n_1} + \frac{\mathcal{D}_2^2}{n_2}} \leq \mathcal{D}_1\sqrt{\frac{2}{n_1}}$. Theorem 2.2 then implies the assertion. ■

Proof of Lemma 3.1: The assumption $d(F, \hat{F}) \leq \delta_p$ implies that if $\mathbb{P}(\hat{F} \geq a) \geq p$ that $\mathbb{P}(F \geq a) \geq p - \delta_p$ and if $\mathbb{P}(\hat{F} \geq a') < p$ that $\mathbb{P}(F \geq a') < p + \delta_p$. The result then follows from Corollary 2.3. ■

Proof of Lemma 4.2: The first assertion is trivial and since $g_H(\vec{F}) + g_K(\vec{F}) \leq 0$ the second follows from

$$\theta_2 = \mathbb{P}(F' > -g_H(\vec{F})|\mathcal{K}) \leq \mathbb{P}(F' \geq g_K(\vec{F})|\mathcal{K}) \leq \delta_2.$$

■

Proof of Lemma 4.4: The first assertion follows from the fact that both $F \mapsto \mathcal{D}_F$ and $F \mapsto D_F$ are diagonal bijective invariants. The second assertion follows from the fact that both $F \mapsto \mathcal{D}_F$ and $F \mapsto D_F$ are invariant under $F \mapsto F + b, b \in \mathbb{R}$ and both transform through scaling $F \mapsto aF$ by $\mathcal{D}_{aF} = |a|\mathcal{D}_F$ For the second, let x, x' approximately achieve the supremum in D_F to accuracy ϵ. That is $F(x) - F(x') \geq D_F - \epsilon$. Then using the product nature of $\mathcal{X}$ we find that

$$\begin{aligned} F(x) - F(x') &= F(x_1, .., x_m) - F(x_1', .., x_m') \\ &= F(x_1, x_2, .., x_m) - F(x_1', x_2, .., x_m) + F(x_1', x_2, .., x_m) - F(x_1', x_2', .., x_m) + ... \\ &\leq \sum_{i=1}^m D_j^F \end{aligned}$$

and so conclude that $D_F \leq \sum_{j=1}^m D_j^F + \epsilon$. Since ϵ is arbitrary we then conclude that

$$D_F \leq \sum_{j=1}^m D_j^F \leq \sqrt{m}\sqrt{\sum_{j=1}^m \left(D_j^F\right)^2} = \sqrt{m}\mathcal{D}_F$$

from which we conclude that

$$c_F = \frac{\mathcal{D}_F}{D_F} \geq \frac{1}{\sqrt{m}}.$$

On the other hand, since $D_F \geq D_j^F, j = 1, .., m$ we obtain

$$D_F^2 \geq \frac{1}{m}\sum_{j=1}^m \left(D_j^F\right)^2 = \frac{1}{m}\mathcal{D}_F^2$$

and conclude that $c_F = \frac{\mathcal{D}_F}{D_F} \leq \sqrt{m}$.

■

Proof of Theorem 4.5: Consider the first set of assertions. According to the proof of [31, Thm. 2.3.2] we have

$$\mathbb{P}^n(\hat{\xi}_q > \xi) = \mathbb{P}^n\Big(\sum_{i=1}^n V_i - \sum_{i=1}^n \mathbb{E}V_i > n\delta_1\Big)$$

where $\delta_1 := \mathbb{F}(\xi) - q$ and $V_i := I(X_i > \xi)$. Consequently we obtain $\mathbb{E}V_i = 1 - \mathbb{F}(\xi), i = 1, .., n$. Applying Hoeffding's inequality [26, Eqn. 2.4] establishes the first assertion. The second assertion follows from [26, Thm. 2.3b] which he attributes to [36] in the binomial case. The last assertion follows from [26, Thm. 2.3c] by the change of variables $V_i' := 1 - V_i$, which he also attributes to [36] in the binomial case.

For the second set of assertions, observe that according to the proof of [31, Thm. 2.3.2] we have

$$\mathbb{P}^n(\hat{\xi}_q < \xi) \leq \mathbb{P}^n\Big(\sum_{i=1}^n W_i - \sum_{i=1}^n \mathbb{E}W_i > n\delta_2\Big)$$

where $\delta_2 := q - \mathbb{F}(\xi)$ and $W_i := I(X_i \leq \xi)$. Consequently we obtain $\mathbb{E}W_i = \mathbb{F}(\xi), i = 1, .., n$. The assertions then follow in the same way as in the first set but with the role of $\mathbb{F}$ switched with $1 - \mathbb{F}$. ■

Proof of Theorem 4.6: Since the first assertion is is clearly true when $\xi_p - \xi_{1-p} \leq 0$ we can assume $\xi_p - \xi_{1-p} > 0$ and $\frac{1}{2} < p < 1$. First observe that for $c, \alpha \in \mathbb{R}$ we have

$$\mathbb{P}^n\Big(\hat{\xi}_{p'} - \hat{\xi}_{1-p'} < c(\xi_p - \xi_{1-p})\Big) \leq \mathbb{P}^n\Big(\hat{\xi}_{p'} < c\xi_p + \alpha\Big) + \mathbb{P}^n\Big(\hat{\xi}_{1-p'} > c\xi_{1-p} + \alpha\Big). \tag{20}$$

We will address each term on the right-hand side separately using Theorem 4.5. For $\epsilon > 0$ let $p' > 1 - \frac{1}{n}$ so that we have the identities $\hat{\xi}_{p'} = \sup_{i=1,..,n} X_i$ and $\hat{\xi}_{1-p'} = \sup_{i=1,..,n} X_i$. Since $p' > p$ it follows that $\xi_p \leq \xi_{p'}$ and consequently $\xi_p - \epsilon < \xi_{p'}$. Moreover, a similar argument shows that $\xi_{1-p} + \epsilon > \xi_{1-p'}$. Consequently, if we define

$$c_\epsilon \;:=\; 1 - \frac{2\epsilon}{\xi_p - \xi_{1-p}}$$

and

$$\alpha_\epsilon \;:=\; \frac{\xi_p(\xi_{1-p} + \epsilon) - \xi_{1-p}(\xi_p - \epsilon)}{\xi_p - \xi_{1-p}}$$

it follows that $c_\epsilon < 1$ and

$$c_\epsilon\xi_p + \alpha_\epsilon = \xi_p - \epsilon < \xi_{p'}$$

$$c_\epsilon\xi_{1-p} + \alpha_\epsilon = \xi_{1-p} + \epsilon > \xi_{1-p'}.$$

Consequently we can apply Part 2iii of Theorem 4.5 to the first term on the right-hand side of (20) to obtain

$$\mathbb{P}^n\Big(\sup_{i=1,..,n} X_i < c_\epsilon\xi_p + \alpha_\epsilon\Big) = \mathbb{P}^n\Big(\sup_{i=1,..,n} X_i < \xi_p - \epsilon\Big) = \mathbb{P}^n\Big(\hat{\xi}_{p'} < \xi_p - \epsilon\Big) \leq e^{-\frac{n\delta_2^2}{2(1-\mathbb{F}(\xi_p-\epsilon))}}$$

where $\delta_2 := p' - \mathbb{F}(\xi_p - \epsilon)$. Letting $p' \mapsto 1$ we obtain

$$\mathbb{P}^n\Big(\sup_{i=1,..,n} X_i < c_\epsilon\xi_p + \alpha_\epsilon\Big) \leq e^{-\frac{1}{2}n\left(1-\mathbb{F}(\xi_p-\epsilon)\right)}.$$

Since $\mathbb{F}(\xi_p - \epsilon) \leq \mathbb{F}(\xi_p-) \leq p$ we then conclude that

$$\mathbb{P}^n\Big(\sup_{i=1,..,n} X_i < c_\epsilon \xi_p + \alpha_\epsilon\Big) \leq e^{-\frac{1}{2}n(1-p)}. \tag{21}$$

For the second term on the right of (20) we apply Part 1iii of Theorem 4.5 to obtain

$$\mathbb{P}^n\Big(\inf_{i=1,..,n} X_i > c_\epsilon \xi_{1-p} + \alpha_\epsilon\Big) = \mathbb{P}^n\Big(\inf_{i=1,..,n} X_i > \xi_{1-p} + \epsilon\Big) = \mathbb{P}^n\Big(\hat{\xi}_{1-p'} > \xi_{1-p} + \epsilon\Big) \leq e^{-\frac{n\delta_1^2}{2\mathbb{F}(\xi_{1-p}+\epsilon)}}$$

where $\delta_1 := \mathbb{F}(\xi_{1-p} + \epsilon) - 1 + p'$. Letting $p' \mapsto 1$ and using $F(\xi_{1-p} + \epsilon) \geq F(\xi_{1-p}) \geq 1 - p$ we obtain

$$\mathbb{P}^n\Big(\inf_{i=1,..,n} X_i < c_\epsilon \xi_p + \alpha_\epsilon\Big) \leq e^{-\frac{1}{2}n(1-p)}. \tag{22}$$

We combine the inequalities (21) and (22) with (20) to obtain

$$\mathbb{P}^n\Big(D_n < c_\epsilon(\xi_p - \xi_{1-p})\Big) \leq 2e^{-\frac{1}{2}n(1-p)}.$$

Since $c_\epsilon \uparrow 1$ as $\epsilon \downarrow 0$ the first assertion of the theorem follows (see e.g. [37, Thm. 1.2.7]).

For the second assertion, observe that it is clearly true when $\xi_p = 0$ so we can assume $\xi_p > 0$. Now observe that the proof of Equation (21) actually proved that

$$\mathbb{P}^n\Big(\sup_{i=1,..,n} X_i < \xi_p - \epsilon\Big) \leq e^{-\frac{1}{2}n(1-p)}.$$

Since $(\xi_p - \epsilon) \uparrow \xi_p$ as $\epsilon \downarrow 0$ the second assertion follows.

■

Proof of Proposition 4.8: Let $\delta_+ := 1 - \mathbb{F}(X_1 - \frac{\varepsilon}{2(1+\varepsilon)}D)$ and $\delta_- := \mathbb{F}(X_- + \frac{\varepsilon}{2(1+\varepsilon)}D)$ and define $\delta^* = \min(\delta_-, \delta_+) - \delta'$ with $\delta' > 0$. Then since Lemma 7.1asserts that $\mathbb{F}^{-1}(t) \leq x$ if and only if $t \leq \mathbb{F}(x)$ it follows that $\mathbb{F}^{-1}(\mathbb{F}(X_+ - \frac{\varepsilon}{2(1+\varepsilon)}D) + \delta') > X_+ - \frac{\varepsilon}{2(1+\varepsilon)}D$ and therefore

$$\xi_{1-\delta^*} \geq \xi_{1-\delta_+ + \delta'} = \mathbb{F}^{-1}(1 - \delta_+ + \delta') = \mathbb{F}^{-1}(\mathbb{F}(X_+ - \frac{\varepsilon}{2(1+\varepsilon)}D) + \delta') \geq X_+ - \frac{\varepsilon}{2(1+\varepsilon)}D.$$

Moreover, since

$$\xi_{\delta^*} \leq \xi(\delta_-) = \mathbb{F}^{-1}(\delta_-) = \mathbb{F}^{-1}(\mathbb{F}(X_- + \frac{\varepsilon}{2(1+\varepsilon)}D)) \leq X_- + \frac{\varepsilon}{2(1+\varepsilon)}D$$

we conclude that

$$\xi_{1-\delta^*} - \xi_{\delta^*} \geq \big(1 - \frac{2\varepsilon}{2(1+\varepsilon)}\big)D = \frac{D}{1+\varepsilon}.$$

The assertion then follows by letting $\delta' \mapsto 0$. ■

Proof of Corollary 4.9: We will only prove the first assertion since the proof of the second is essentially the same. Since the assertion is trivially true when $D = 0$ we can assume this not the case. Then since for $0 < p \leq \delta(\epsilon)$ we have

$$\xi_{1-p} - \xi_p = \frac{\xi_{1-p} - \xi_p}{D} D \geq \frac{D}{1+\epsilon}$$

the assertion follows from from Theorem 4.6. ■

Proof of Theorem 4.11: Since $\hat{D} \leq D$ with probability 1 we find that

$$\begin{aligned}
\theta_1 &= \mathbb{P}^n\big(T_1 = 0 \big| \mathcal{H}_1\big) \\
&= \mathbb{P}^n\big(f_H((1+\epsilon)\hat{D}) + f_K((1+\epsilon)\hat{D}) > 0 \big| f_H((1+\epsilon)D) + f_K((1+\epsilon)D) \leq 0\big) \\
&\leq \mathbb{P}^n\big(f_H((1+\epsilon)D) + f_K((1+\epsilon)D) > 0 \big| f_H((1+\epsilon)D) + f_K((1+\epsilon)D) \leq 0\big) \\
&= 0
\end{aligned}$$

establishing the first assertion. For the second and third assertions we use following lemma.

Lemma 6.3 *With the assumptions of Theorem 4.11 let $F' : \mathcal{X}^n \to \mathbb{R}$ and let $f : \mathbb{R}^k \to \mathbb{R}$ be non-decreasing. Then we have*

$$\mathbb{P}^n\Big(F' \geq f\big((1+\epsilon)\hat{D}\big)\Big) \leq \mathbb{P}^n\Big(F' \geq f(D)\Big) + \sum_{j=1}^{k} \mathbb{P}^{n_j}\Big((1+\epsilon)\hat{D}_{F^j} < D_{F^j}\Big). \tag{23}$$

Moreover, for $n_j \geq n_j^{\epsilon}(\delta), j = 1, .., k$ we have

$$\sum_{j=1}^{k} \mathbb{P}^n\Big((1+\epsilon)\hat{D}_{F^j} < D_{F^j}\Big) \leq \delta$$

and therefore

$$\mathbb{P}^n\Big(F' \geq f\big((1+\epsilon)\hat{D}\big)\Big) \leq \mathbb{P}^n\Big(F' \geq f\big(D\big)\Big) + \delta$$

Proof: By the monotonicity of f it follows that

$$\Big\{F - f\big((1+\epsilon)\hat{D}\big) \geq 0\Big\} \tag{24}$$

$$\subset \Big\{F - f(D) \geq 0\Big\} \cup \cup_{j=1}^{k}\Big\{(1+\epsilon)\hat{D}_{F^j} < D_{F^j}\Big\}. \tag{25}$$

Consequently, we obtain the first assertion:

$$\begin{aligned}
\mathbb{P}^n\Big(F' \geq f\big((1+\epsilon)\hat{D}\big)\Big) &\leq \mathbb{P}^n\Big(F' \geq f(D)\Big) + \sum_{j=1}^{k} \mathbb{P}^n\Big((1+\epsilon)\hat{D}_{F^j} < D_{F^j}\Big) \\
&= \mathbb{P}^n\Big(F' \geq f(D)\Big) + \sum_{j=1}^{k} \mathbb{P}^n\Big((1+\epsilon)\hat{D}_{F^j} < D_{F^j}\Big).
\end{aligned}$$

For the second, observe that by Corollary 4.9 we find that for each j we have

$$\mathbb{P}^n\Big((1+\epsilon)\hat{D}_{F^j} < D_{F^j}\Big) = \mathbb{P}^n\Big((1+\epsilon)\hat{D}_{F^j} < D_{F^j}\Big) \leq 2e^{-\frac{n\tau^j(\epsilon)}{2}}.$$

Since the assumption $n \geq n_j^{\epsilon}(\delta)$, defined in (19), implies that

$$2e^{-\frac{n\tau^j(\epsilon)}{2}} \leq \frac{\delta}{k}, \quad j = 1, .., k$$

the second assertion follows from the first. ■

We now proceed to the second and third assertions of Theorem 4.11. Observe that

$$\begin{aligned}\theta_{11} &= \mathbb{P}^n\big(T_1 = 1, T_2 = 0\big|\mathcal{H}_{2\epsilon}\big)\\ &\leq \mathbb{P}^n\big(T_2 = 0\big|\mathcal{H}_2\big)\\ &= \mathbb{P}^n\big(F' \leq -f_H((1+\epsilon)\hat{D})\big|\mathcal{H}_2\big)\end{aligned}$$

Since $\Delta = \sum_{j=1}^k \mathbb{P}^n\Big((1+\epsilon)\hat{D}_{F^j} < D_{F^j}\Big)$, Lemma 6.3 (Equation 23) applied to $-F'$ then shows that

$$\theta_{11} \leq \mathbb{P}^n\big(F' \leq -f_H((1+\epsilon)\hat{D})\big|\mathcal{H}_2\big) \leq \mathbb{P}^n\big(F' \leq -f_H(D)\big|\mathcal{H}_2\big) + \Delta$$

thus establishing the second assertion. Since $T_1 = 1$ implies that $f_H((1+\epsilon)\hat{D}) + f_K((1+\epsilon)\hat{D}) \leq 0$ we find that

$$\begin{aligned}\theta_{12} &= \mathbb{P}^n\big(T_1 = 1, T_2 = 1\big|\mathcal{K}_{2\epsilon}\big)\\ &= \mathbb{P}^n\big(T_1 = 1, F' > -f_H((1+\epsilon)\hat{D})\big|\mathcal{K}_{2\epsilon}\big)\\ &\leq \mathbb{P}^n\big(T_1 = 1, F' > f_K((1+\epsilon)\hat{D})\big|\mathcal{K}_{2\epsilon}\big)\\ &\leq \mathbb{P}^n\big(F' \geq f_K((1+\epsilon)\hat{D})\big|\mathcal{K}_2\big)\end{aligned}$$

As in the previous case, Lemma 6.3 then shows that

$$\theta_{12} \leq \mathbb{P}^n\big(F' \geq f_K((1+\epsilon)\hat{D})\big|\mathcal{K}_2\big) \leq \mathbb{P}^n\big(F' \geq f_K(D)\big|\mathcal{K}_2\big) + \Delta$$

thus establishing the third assertion.

The last set of assertions follows by observing that the assumption $n \geq \max\big(n_j^\epsilon(\delta_1), n_j^\epsilon(\delta_2)\big), j = 1, .., k$ and Lemma 6.3 implies that $\Delta \leq \min(\delta_1, \delta_2)$.

■

Proof of Corollary 5.1: Since $\mathbb{E}F' = \mathbb{E}F$, Lemma 4.1 implies that

$$\mathbb{P}^n\Big(F' \leq -f'_H(\mathcal{D}_F, \mathcal{D}_{F'}, \delta_1)\big|\mathcal{H}_2\Big) \leq \delta_1\,,$$

$$\mathbb{P}^n\Big(F' \geq f'_K(\mathcal{D}_F, \mathcal{D}_{F'}, \delta_2)\big|\mathcal{K}_2\Big) \leq \delta_2\,,$$

where

$$\begin{aligned}f'_H(r_1, r_2, \delta) &:= \frac{r_1}{\sqrt{2}}\sqrt{\log\delta^{-1}} + \frac{r_2}{\sqrt{2}}\sqrt{\log p^{-1}} - a\,,\\ f'_K(r_1, r_2, \delta) &:= \frac{r_1}{\sqrt{2}}\sqrt{\log\delta^{-1}} + \frac{r_2}{\sqrt{2}}\sqrt{\log(1-p)^{-1}} + a'\,.\end{aligned}$$

By Definition 4.3 we have $\mathcal{D}_F \leq cD$. Moreover, the proof of Corollary 2.3 shows that

$$\mathcal{D}_{F'} \leq \frac{1}{\sqrt{n}}\mathcal{D}_F \leq \frac{cD}{\sqrt{n}}.$$

Consequently, we have $f'_H(\mathcal{D}_F, \mathcal{D}_{F'}, \delta_1) \leq f_H(D)$ and $f'_K(\mathcal{D}_F, \mathcal{D}_{F'}, \delta_2) \leq f_K(D)$ and therefore

$$\mathbb{P}^n\Big(F' \leq -f_H(D)\big|\mathcal{H}_2\Big) \leq \delta_1\,,$$

$$\mathbb{P}^n\Big(F' \geq f_K(D)\big|\mathcal{K}_2\Big) \leq \delta_2\,,$$

The assertion then follows from Theorem 4.11.

■

Proof of Corollary 5.2: As in the proof of Corollary 5.1, since $\mathbb{E}F' = \mathbb{E}F$, Lemma 4.1 implies that

$$\mathbb{P}^n\Big(F' \leq -f'_H(\mathcal{D}_F, \mathcal{D}_{F'}, \delta_1)\big|\mathcal{H}_2\Big) \leq \delta_1\,,$$

$$\mathbb{P}^n\Big(F' \geq f'_K(\mathcal{D}_F, \mathcal{D}_{F'}, \delta_2)\big|\mathcal{K}_2\Big) \leq \delta_2\,,$$

where

$$f'_H(r_1, r_2, \delta) \;:=\; \frac{r_1}{\sqrt{2}}\sqrt{\log p^{-1}} + \frac{r_2}{\sqrt{2}}\sqrt{\log \delta^{-1}} - a\,,$$

$$f'_K(r_1, r_2, \delta) \;:=\; \frac{r_1}{\sqrt{2}}\sqrt{\log (1-p)^{-1}} + \frac{r_2}{\sqrt{2}}\sqrt{\log \delta^{-1}} + a'\,.$$

By Definition 4.3 we have $\mathcal{D}_{F_1} \leq c_1 D_1$ and $\mathcal{D}_{F_2} \leq c_2 D_2$. Therefore it follows that

$$\mathcal{D}_F = \mathcal{D}_{F_1+F_2} \leq \mathcal{D}_{F_1} + \mathcal{D}_{F_2} \leq c_1 D_1 + c_2 D_2.$$

Moreover, the proof of Corollary 2.4 implies that

$$\mathcal{D}_{F'} \leq \sqrt{\frac{\mathcal{D}_{F_1}^2}{n_1} + \frac{\mathcal{D}_{F_2}^2}{n_2}} \leq \sqrt{\frac{c_1^2 D_1^2}{n_1} + \frac{c_2^2 D_2^2}{n_2}}.$$

Consequently, we have $f'_H(\mathcal{D}_F, \mathcal{D}_{F'}, \delta_1) \leq f_H(D)$ and $f'_K(\mathcal{D}_F, \mathcal{D}_{F'}, \delta_2) \leq f_K(D)$ and therefore

$$\mathbb{P}^n\Big(F' \leq -f_H(D)\big|\mathcal{H}_2\Big) \leq \delta_1\,,$$

$$\mathbb{P}^n\Big(F' \geq f_K(D)\big|\mathcal{K}_2\Big) \leq \delta_2\,,$$

The assertion then follows from Theorem 4.11.

■

7 Appendix

The following Lemma from [31, Lem. 1.1.4 & Sec. 2.3] lists important properties of the distribution function $\mathbb{F}(x) := \mathbb{P}(X \leq x)$ and its corresponding quantile function $\mathbb{F}^{-1}(t) := \inf\{x : \mathbb{F}(x) \geq t\}$.

Lemma 7.1 *Let $\mathbb{F}$ be a distribution function. Then $\mathbb{F}$ is right continuous and the function $\mathbb{F}^{-1}, 0 < t < 1$ is non-decreasing, left continuous and satisfies*

i) $\mathbb{F}^{-1}(\mathbb{F}(x)) \leq x, -\infty < x < \infty\,.$

ii) $\mathbb{F}(\mathbb{F}^{-1}(t)) \geq t \geq \mathbb{F}(\mathbb{F}^{-1}(t)-), 0 < t < 1\,.$

iii) $\mathbb{F}(x) \geq t$ *if and only if* $x \geq \mathbb{F}^{-1}(t)$.

Example 7.2 (Extreme values of c_F) Let $F(x) := \sum_{j=1}^m F_j(x_j)$. Then since $F(x) - F(x') = \sum_{j=1}^m \big(F_j(x_j) - F_j(x'_j)\big)$ it follows that

$$D_F = \sum_{j=1}^m D_j^{F_j}.$$

Moreover, since $D_j^F = D_j^{F_j}, j = 1, .., m$ we obtain

$$\mathcal{D}_F^2 = \sum_{j=1}^{m} \left(D_j^{F_j}\right)^2$$

and therefore

$$c_F^2 = \frac{\sum_{j=1}^{m} \left(D_j^{F_j}\right)^2}{\left(\sum_{j=1}^{m} D_j^{F_j}\right)^2}.$$

In particular, when $D_j^{F_j} = D_1^{F_1}, j = 1, , m$ we obtain $c_F = \frac{1}{\sqrt{m}}$. On the other hand, let $\mathcal{X} := [0,1]^m \subset \mathbb{R}^m$ and let $F(x) := \|x\|, \|x\| \leq 1$ and $F(x) := 0, \|x\| > 1$ where $\|x\|$ is the Euclidean norm of x. Then it is easy to see that $D_F = 1$, $D_j^F = 1, j = 1, .., m$ and therefore $\mathcal{D}_F^2 = m$. Consequently in this case we obtain $c_F = \sqrt{m}$.

Acknowledgments

We gratefully acknowledge several illuminating discussions with Nicholas Hengartner regarding this work.

References

[1] L.J. Lucas, H. Owhadi, and M. Ortiz. Rigorous verfication, validation, uncertainty quantification and certification through concentration-of-measure inequalities. *Comput. Methods Appl. Mech. Engrg.*, 197:4591–4609, 2008.

[2] W. L. Oberkampf and T. Trucano. Verification and validation in computational fluid dynamics. *Progr. Aerospace Sci.*, 38:209–272, 2002.

[3] W. L. Oberkampf, T. Trucano, and C. Hirsch. Verification, validation and predictive capability in computational engineering and physis. *Appl. Mech. Rev.*, 57(5):345–385, 2004.

[4] D. Eardley (Study Leader). Quantification of margins and uncertainties. *Technical Report JSR-04-330, JASON, The MITRE Corporation, 7515 Colshire Dr., McLean Virginia, 22102*, 20052005.

[5] O. Balci. Principles and techniques of simulation validation, verification, and testing. In C. Alexopoulos, K. Kang, W. R. Lilegdon, and D. Goldsman, editors, *Proceedings 1995 Winter Simulation Conference*, pages 147–154. ACM, 1993.

[6] R. G. Sargent. Verifying and validating simulation models. In J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain, editors, *Proceedings 1996 Winter Simulation Conference*, pages 55–64. ACM, 1996.

[7] R. G. Sargent. Validation and verification of simulation models. In R. G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, editors, *Proceedings 2004 Winter Simulation Conference*, pages 17–28. ACM, 2004.

[8] R. G. Sargent. Verification and validation of simulation models. In M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, editors, *Proceedings of the 2005 Winter Simulation Conference*, pages 130–143, 2005.

[9] T. H. Naylor and J. M. Finger. Verification of computer simulation models. *Management Science*, 14(2):B92–B101, 1967.

[10] G. B. Kleindorfer and R. Ganeshan. The philosophy of science and validation in simulation. In G. W. Evans, M. Mollaghasemi, and E. C. Russel, editors, *Proceedings 1993 Winter Simulation Conference*, pages 50–57. ACM, 1993.

[11] G. B. Kleindorfer, L. O'Neill, and R. Ganeshan. Validation in simulation: Various positions in the philosophy of science. *Management Science*, 44(8):1087–1099, 1998.

[12] N. Oreskes, K. Shrader-Frechette, and K. Belitz. Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, 263:641–646, 1994.

[13] V. Barbu and Th. Precupanu. *Convexity and Optimization in Banach Spaces*. D. Reidel Publishing Company, Dordrecht, 1986.

[14] O. Balci and R. G. Sargent. A methodology for cost-risk analysis in the statistical validation of simulation models. *Comm. ACM*, 24(4):190–197, 1981.

[15] J. P. C. Keijnen. Satistical validation of simulation models. *European Journal of Operational Research*, 87:21–34, 1995.

[16] Committee on the Evaluation of Quantification of Margins and Uncertainties Methodology for Assessing and Certifying the Reliability of the Nuclear Stockpile. *Evaluation of Quantification of Margins and Uncertainties Methodology for Assessing and Certifying the Reliability of the Nuclear Stockpile*. National Academies Press, Washington, DC, 2009.

[17] M. Pilch, T. G. Trucano, and J. C. Helton. Ideas underlying quantification of margins and uncertainties (QMU):a white paper. Sandia Nat. Lab. Report SAND2006-5001, `http://www.stanford.edu/group/uq/docs/qmu_ideas.pdf`, 2006.

[18] D. Sharp and M. Wood-Schultz. QMU and nuclear weapons certification. *Los Alamos Science*, 28:47–53, 2003.

[19] E. L. Lehmann. *Testing Statistical Hypotheses*. Chapman and Hall, New York, second edition, 1994.

[20] W. Feller. On the normal approximation to the binomial distribution. *Annals of Mathematical Statistics*, 16(4):319–329, 1945.

[21] T.W. Anderson and S.M. Samuels. Some inequalities among binomial and poisson probabilities. *Proc. Fifth Berkeley Symp. Math. Statist.*, 1:1–12, 1965.

[22] E.V. Slud. Distribution inequalities for the binomial law. *Annals of Probability*, 5(3):404–412, 1977.

[23] H. Shore. Approximate closed form expressions for the decision variables of some tests related to the binomial distribution. *The Statistician*, 35:471–478, 1986.

[24] H. Shore. Simple general approximations for a random variable and its inverse distribution function base on linear tranformations of a nonskewed variate. *SIAM J. Sci. Stat. Comput.*, 7(1):1–23, 1986.

[25] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics*, 23(4):493–507, 1952.

[26] C. McDiarmid. Concentration. In M. Habib, C. McDiarmid, R. Ramirez-Alfonsin, and B. Reed, editors, *Probabilistic Methods for Algorithmic Discrete Mathematic*, pages 195–248. Springer, 1991.

[27] M. Ledoux. *The Concentration of Measure Phenomenon*. American Mathematical Society, Providence, RI, 2001.

[28] A. Dvoretzky, J. Kiefer, and J. Wolfowitz. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.*, 27:642–669, 1956.

[29] P. Massart. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Ann. Probab.*, 18:1269–1283, 1990.

[30] L. Devroye, L. Györfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition*. Springer, New York, 1997.

[31] R.J. Serfling. *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, New York, 1980.

[32] Z. Arov and A.A. Bobrov. The extreme terms of a sample and their role in the sum of independent variables. *Theory of Probability and Its Applications*, 5(4):377–306, 1960.

[33] J. Pickands III. Statistical inference using extreme order statistics. *Annals of Statistics*, 3(1):119–131, 1975.

[34] O. Barndorff-Neilsen. On the limit behavior of extreme order statistics. *Inst. Math. Stat.*, 34(3):992–1002, 1963.

[35] J. Kiefer. Admissibility of conditional confidence procedures. *Annals of Statistics*, 4(5):836–865, 1976.

[36] D. Angluin and L. Valiant. Fast probabilistic algorithms for hamiltonian circuits and matchings. *J. Computer and System Sciences*, 18:155–193, 1979.

[37] R.B. Ash. *Real Analysis and Probability*. Academic Press, New York, 1972.