Distribution-uniform anytime-valid sequential inference

Ian Waudby-Smith, Edward H. Kennedy, and Aaditya Ramdas

Carnegie Mellon University
{ian,edward,aramdas}@stat.cmu.edu

Abstract

Are asymptotic confidence sequences and anytime p-values uniformly valid for a nontrivial class of distributions $\mathcal{P}$? We give a positive answer to this question by deriving *distribution-uniform* anytime-valid inference procedures. Historically, anytime-valid methods — including confidence sequences, anytime p-values, and sequential hypothesis tests that enable inference at stopping times — have been justified nonasymptotically. Nevertheless, asymptotic procedures such as those based on the central limit theorem occupy an important part of statistical toolbox due to their simplicity, universality, and weak assumptions. While recent work has derived asymptotic analogues of anytime-valid methods with the aforementioned benefits, these were not shown to be $\mathcal{P}$-*uniform*, meaning that their asymptotics are not uniformly valid in a class of distributions $\mathcal{P}$. Indeed, the anytime-valid inference literature currently has no central limit theory to draw from that is both uniform in $\mathcal{P}$ and in the sample size n. This paper fills that gap by deriving a novel $\mathcal{P}$-uniform strong Gaussian approximation theorem. We apply some of these results to obtain an anytime-valid test of conditional independence without the Model-X assumption, as well as a $\mathcal{P}$-uniform law of the iterated logarithm.

1 Introduction

Some of the simplest and most efficient statistical inference tools are asymptotic ones that rely on large-sample theory such as the central limit theorem (CLT). However, there is a sharp distinction between asymptotics that are only valid for a single distribution P and those that are *uniformly valid* over a large collection of distributions $\mathcal{P}$. To elaborate, consider the classical CLT which states that for independent and identically distributed random variables $X_1, \dots, X_n \sim P$ with mean μ_P and finite variance $\sigma_P^2 < \infty$, their scaled partial sums $\widehat{Z}_n := \sum_{i=1}^n (X_i - \mu_P)/(\sigma_P \sqrt{n})$ are asymptotically standard Gaussian, meaning for any real x, we have $\mathbb{P}_P(\widehat{Z}_n \leqslant x) \to \Phi(x)$ where Φ is the cumulative distribution function (CDF) of a standard Gaussian. However, this is a *distribution-pointwise* statement in the sense that the limit holds for a single $P \in \mathcal{P}$. An unsettling consequence of P-pointwise statements is that no matter how large n is, $|\mathbb{P}_{P'}(\widehat{Z}_n \leqslant x) - \Phi(x)|$ can be far from zero for some $P' \in \mathcal{P}$ — or more informally, asymptotics may be "kicking in" arbitrarily late.

By contrast, *distribution-uniformity* (or more specifically $\mathcal{P}$*-uniformity*) rules out the aforementioned unsettling scenario so that convergence occurs simultaneously for all $P \in \mathcal{P}$. Concretely, consider the difference between P-pointwise versus $\mathcal{P}$-uniform convergence in distribution when written out side-by-side:

$$\underbrace{\sup_{P \in \mathcal{P}} \lim_{n \to \infty} \left| \mathbb{P}_P(\widehat{Z}_n \leqslant x) - \Phi(x) \right| = 0}_{P\text{-pointwise convergence in distribution}} \quad \text{versus} \quad \underbrace{\lim_{n \to \infty} \sup_{P \in \mathcal{P}} \left| \mathbb{P}_P(\widehat{Z}_n \leqslant x) - \Phi(x) \right| = 0,}_{\mathcal{P}\text{-uniform convergence in distribution}} \tag{1}$$

where the essential difference lies in the order of limits and suprema. The initial study of $\mathcal{P}$-uniformity is often attributed to Li [26] and many papers have emphasized its importance in recent years; see Kasy [20], Rinaldo et al. [31], Tibshirani et al. [42], Shah and Peters [38], Kuchibhotla et al. [24], and Lundborg et al. [28]. Note that this literature sometimes refers to distribution-uniformity as "honesty" [26, 24] or simply "uniformity" [20, 31, 42, 38, 28]. We opt for the phrase "distribution-uniform" — or "$\mathcal{P}$-uniform" when we want to specify that uniformity is with respect to $\mathcal{P}$ — since there are many other notions of uniformity throughout probability and statistics, including time-uniformity and quantile-uniformity, both of which will become relevant throughout this paper. We do not use the term "honesty" as it has also been used to refer to other properties of estimators in statistical inference [45, 1] and is sometimes used in the sense of parameter-uniformity [35].

Simultaneously, there is a parallel literature on *time-uniform* (typically called "anytime-valid") inference where the goal is to derive confidence sequences (CSs) — sequences of confidence intervals (CIs) that are uniformly valid for all sample sizes — as well as anytime p-values and sequential hypothesis tests (to be defined more formally later) that can be continuously monitored and adaptively stopped. This literature has historically taken a mostly nonasymptotic approach to inference so that the type-I errors and coverage probabilities hold in finite samples; see the early work of Wald, Robbins, and colleagues [46, 12, 32, 25], as well as the review paper of Ramdas, Grünwald, Vovk, and Shafer [30]

which gives a broad overview of this literature. However, nonasymptotic approaches generally require strong assumptions on the random variables such as lying in a parametric family, *a priori* known bounds on their support, or on their moments. On the other hand, this paper takes an *asymptotic* view of anytime-valid inference where type-I errors and coverage probabilities hold in the limit; see Robbins and Siegmund [33], Waudby-Smith et al. [47], and Bibaut et al. [4]. An advantage of this regime is that the resulting methods take simple, universal forms and allow for substantially weaker conditions (for example, requiring only that absolute moments exist and are finite but for which *a priori* bounds are not known).

To illustrate time-uniformity in the asymptotic regime, suppose that random variables $X_1, \ldots, X_n$ have finite mean μ and variance σ^2 and we would like to derive a CI for μ. A classical asymptotic CI $\dot{C}_n$ has the guarantee that $\limsup_{n\to\infty} \mathbb{P}_P(\mu \notin \dot{C}_n) \leqslant \alpha$, but its asymptotic validity hinges on the sample size n being fixed and pre-specified in advance. By contrast, an asymptotically valid CS $(\bar{C}_k^{(m)})_{k=m}^{\infty}$ can elicit a much stronger property written in juxtaposition with the classical asymptotic CI as follows:

$$\underbrace{\limsup_{n\to\infty} \mathbb{P}_P\left(\mu \notin \dot{C}_n\right) \leqslant \alpha}_{\text{(Asymptotic) fixed-}n\text{ CI}} \quad \text{versus} \quad \underbrace{\limsup_{m\to\infty} \mathbb{P}_P\left(\exists k \geqslant m : \mu \notin \bar{C}_k^{(m)}\right) \leqslant \alpha}_{\text{(Asymptotic) anytime-valid CS}}, \tag{2}$$

where the main difference lies in the fact that the right-hand side probability holds uniformly in $k \geqslant m$ for sufficiently large m. From a practical perspective, the right-hand side permits a researcher to continuously monitor the outcome of an experiment, for example, updating their CIs as each new data point is collected as long as the starting sample size m is sufficiently large. Importantly, these anytime-valid procedures allow for the experiment to stop *as soon as* the researcher has sufficient evidence to reject some null hypothesis (e.g. as soon as $0 \notin \bar{C}_k^{(m)}$ for a null effect of 0). Note that while CSs and anytime p-values are typically studied from a *nonasymptotic* viewpoint, we will henceforth omit the "asymptotic" phrasing when referring to asymptotic procedures such as those in (2) since we are solely interested in asymptotics in this paper (and distribution-uniformity is always trivially satisfied for nonasymptotics).

In this paper, our main goal is to define and derive distribution-uniform anytime-valid tests, p-values, and confidence sequences. However, the time-uniform guarantee in the right-hand side of (2) is a *distribution-pointwise* statement, and to the best of our knowledge, there currently exist no distribution-uniform guarantees for time-uniform asymptotics. The reason for this is subtle and has led us to identify a gap in the probability literature. To elaborate, while fixed-n asymptotics are based on the CLT, Waudby-Smith et al. [47] analyzed asymptotic analogues of nonasymptotic CSs using strong Gaussian approximations (sometimes called "strong invariance principles" or "strong embeddings") such as the seminal results of Strassen [40] and Komlós, Major, and Tusnády [22, 23] — see also Chatterjee [8] and the references therein. Not only have strong approximations not yet been studied from a distribution-uniform perspective, it is not even clear what the right definition of "distribution-uniform strong Gaussian approximation" ought to be. We give both a definition and a corresponding result satisfying it in Section 5, and this serves as a probabilistic foundation for the rest of our statistical results.

1.1 Outline of the paper

Below we outline how the paper will proceed, highlighting our key contributions.

- We begin in Section 2 by defining $\mathcal{P}$-uniform anytime-valid inference in the form of anytime hypothesis tests, anytime p-values, and confidence sequences (Definition 2.1). This definition serves as context for Section 2.1 where we state our main result in Proposition 2.1 (initially without proof). The remaining sections are focused on providing the necessary machinery to prove a stronger version of Proposition 2.1 which is ultimately given in Theorem 3.3.

- Section 2.2 lays some foundations for distribution-, time-, and boundary-uniform central limit theory for centered partial sums, culminating in Proposition 2.2. The results therein are new to the literature even in the distribution-pointwise regime. However, Proposition 2.2 is stated in terms of the true (rather than empirical) variance in standardizing the partial sums, motivating the following section on distribution-uniform almost-sure consistency.
- Section 3 discusses what it means for a sequence of random variables to converge almost-surely *and* uniformly in a class of distributions. The section culminates in a result showing that the empirical variance is a distribution-uniform almost-surely consistent estimator for the true variance and its convergence rate is polynomial in the sample size (Proposition 3.2), which, when combined with Proposition 2.2 from Section 2 yields our main result in Theorem 3.3.
- Section 4 applies the content of the previous sections to the problem of anytime-valid conditional independence testing. We first show that distribution-uniform anytime-valid tests of conditional independence are impossible to derive without imposing structural assumptions, a fact that can be viewed as a time-uniform analogue of the hardness result due to Shah and Peters [38, §2]. We then develop a sequential version of the Generalized Covariance Measure test due to Shah and Peters [38, §3] and show that it distribution- and time-uniformly controls the type-I error (and has nontrivial power) as long as certain regression functions are estimated at sufficiently fast rates. To the best of our knowledge, this is the first anytime-valid test of conditional independence that does not rely on Model-X assumptions.
- Section 5 highlights that all of the preceding results fundamentally rely on a distribution-uniform strong Gaussian approximation (Theorem 5.1) that serves as a (purely probabilistic) foundational piece of our main results and it is the first result of its kind in the literature (to the best of our knowledge). This strong approximation is itself a consequence of a *nonasymptotic* high-probability coupling inequality (Lemma 5.2). Finally, we illustrate how these couplings and approximations give rise to a distribution-uniform law of the iterated logarithm. All three of these results may be of independent interest.

1.2 Notation

Throughout, we will let Ω be a sample space, $\mathcal{F}$ the Borel sigma-algebra, and $\mathcal{P}$ a collection of probability measures so that $(\Omega, \mathcal{F}, P)$ is a probability space for each $P \in \mathcal{P}$. We will often write $(\Omega, \mathcal{F}, \mathcal{P})$ to refer to the collection of probability spaces $(\Omega, \mathcal{F}, P)_{P\in\mathcal{P}}$. Note that $P \in \mathcal{P}$ are defined with respect to the same sample space Ω and sigma-algebra $\mathcal{F}$ but do not need to have a common dominating measure (e.g. $\mathcal{P}$ can consist of infinitely many discrete and continuous distributions as well as their mixtures).

Throughout, we will work with random variables that are defined on the collection of probability spaces $(\Omega, \mathcal{F}, \mathcal{P})$ (unless otherwise specified, as will be the case in Section 5). For any event $A \in \mathcal{F}$, we use $\mathbb{P}_P(A)$ to denote the probability of that event and $\mathbb{E}_P(\cdot)$ to denote the expectation of a random variable with respect to $P \in \mathcal{P}$, meaning for X defined on $(\Omega, \mathcal{F}, P)$,

$$\mu_P \equiv \mathbb{E}_P(X) = \int x \; dP(x). \tag{3}$$

Similarly, $\sigma_P^2 \equiv \mathrm{Var}_P(X)$ will be shorthand for $\mathbb{E}_P(X - \mathbb{E}_P(X))^2$, and so on.

2 What is distribution-uniform anytime-valid inference?

Recalling the $\mathcal{P}$-uniform convergence in distribution guarantee provided in (1), a fixed-n p-value $\dot{p}_n$ defined on $(\Omega, \mathcal{F}, \mathcal{P})$ is said to be $\mathcal{P}_0$-uniform for the null hypothesis $\mathcal{P}_0 \subseteq \mathcal{P}$ if

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_0} \mathbb{P}_P(\dot{p}_n \leqslant \alpha) \leqslant \alpha, \tag{4}$$

and it is easy to see how such a p-value can be constructed given a statistic satisfying the right-hand side of (1). Similarly, Waudby-Smith et al. [47, Definition 2.7] provide a definition of (P-pointwise) time-uniform coverage of asymptotic CSs, which is also implicit in Robbins and Siegmund [33] and Bibaut et al. [4]. Adapting their definition to anytime p-values, we say that $(\bar{p}_k^{(m)})_{k=m}^{\infty}$ has asymptotic time-uniform type-I error control under the null $\mathcal{P}_0$ if

$$\forall P \in \mathcal{P}_0,\ \limsup_{m\to\infty} \mathbb{P}_P(\exists k \geqslant m : \bar{p}_k^{(m)} \leqslant \alpha) \leqslant \alpha. \tag{5}$$

Juxtaposing (4) and (5), we can intuit the right definition of distribution- *and* time-uniform type-I error control, where we simply place a supremum over $\mathcal{P}_0$ inside the limit in (5). We lay this definition out formally alongside corresponding definitions for anytime hypothesis tests, confidence sequences, and sharpness thereof below.

Definition 2.1 ($\mathcal{P}$-uniform anytime-valid statistical inference)**.** *Let $\mathcal{P}$ be a collection of distributions and let $\mathcal{P}_0 \subseteq \mathcal{P}$ be the null hypothesis. We say that $(\bar{\Gamma}_k^{(m)})_{k=m}^{\infty}$ is a $\mathcal{P}_0$-uniform anytime hypothesis test if*

$$\limsup_{m\to\infty} \sup_{P\in\mathcal{P}_0} \mathbb{P}_P\left(\exists k \geqslant m : \bar{\Gamma}_k^{(m)} = 1\right) \leqslant \alpha \tag{6}$$

and that $(\bar{p}_k^{(m)})_{k=m}^{\infty}$ is a $\mathcal{P}_0$-uniform anytime p-value if

$$\limsup_{m\to\infty} \sup_{P\in\mathcal{P}_0} \mathbb{P}_P\left(\exists k \geqslant m : \bar{p}_k^{(m)} \leqslant \alpha\right) \leqslant \alpha. \tag{7}$$

Moreover, we say that $(\bar{C}_k^{(m)})_{k=m}^{\infty}$ is a $\mathcal{P}$-uniform $(1-\alpha)$-confidence sequence for $\theta(P)$ if

$$\limsup_{m\to\infty} \sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\exists k \geqslant m : \theta(P) \notin \bar{C}_k^{(m)}\right) \leqslant \alpha. \tag{8}$$

Finally, we say that all of these procedures are sharp if the limit suprema are limits and the inequalities ($\leqslant \alpha$) are equalities ($= \alpha$).

As one may expect, any $\mathcal{P}$-uniform anytime-valid test, p-value, or CS satisfying Definition 2.1 is also $\mathcal{P}$-uniform for a fixed sample size n in the sense of (4) as well as P-pointwise anytime-valid for any $P \in \mathcal{P}$ in the sense of (5). With Definition 2.1 in mind, we will now derive distribution-uniform anytime hypothesis tests, p-values, and confidence sequences for the mean of independent and identically distributed random variables.

2.1 Our primary goal: Inference for the mean

While Definition 2.1 is a natural extension of distribution-uniform inference to the anytime-valid setting, it is deceptively challenging to derive procedures satisfying Definition 2.1 even for the simplest of statistical problems such as tests for the mean of independent and identically distributed random variables and the main results of this section themselves rely on certain technical underpinnings such as distribution-uniform almost-sure consistency and strong Gaussian approximations. Rather than laboriously discuss these technical details here, let us instead articulate our main goal and result — distribution-uniform anytime inference for the mean — and defer more in-depth discussions to Sections 2.2, 3, and 5.

In many of the results that follow, we will rely on a monotonically increasing function $\Psi : \mathbb{R}^{\geqslant 0} \to [0,1]$ given by

$$\Psi(r) := 1 - 2\left[1 - \Phi(\sqrt{r}) + \sqrt{r}\phi(\sqrt{r})\right]; \quad r \geqslant 0. \tag{9}$$

This function happens to be the cumulative distribution function of a particular probability distribution that we have opted to call the *Robbins-Siegmund distribution* since it was implicitly computed

by Robbins and Siegmund [33] in the context of boundary-crossing probabilities for Wiener processes. For now, we will only rely on the fact that Ψ is invertible and leave more detailed discussions of its properties to Appendix B.1. As we will see shortly, Ψ plays a role in asymptotic anytime-valid inference similar to that of the Gaussian cumulative distribution function in asymptotic fixed-n inference. Indeed, define the process $(\bar{p}_k^{(m)})_{k=m}^{\infty}$ given by

$$\bar{p}_k^{(m)} := 1 - \Psi\left(k\widehat{\mu}_k^2/\widehat{\sigma}_k^2 - \log(k/m)\right) \tag{10}$$

and the intervals $(\bar{C}_k^{(m)}(\alpha))_{k=m}^{\infty}$ given by

$$\bar{C}_k^{(m)}(\alpha) := \widehat{\mu}_k \pm \widehat{\sigma}_k\sqrt{[\Psi^{-1}(1-\alpha) + \log(k/m)]/k}, \tag{11}$$

where $\widehat{\sigma}_k^2 := \frac{1}{k}\sum_{i=1}^{k}(X_i - \widehat{\mu}_k)^2$ is the sample variance. The following result gives conditions under which $(\bar{p}_k^{(m)})_{k=m}^{\infty}$ is a $\mathcal{P}_0$-uniform anytime p-value for the null of $\mu_P = 0$ and $(\bar{C}_k^{(m)}(\alpha))_{k=m}^{\infty}$ is a $\mathcal{P}$-uniform $(1-\alpha)$-CS for μ_P in the sense of Definition 2.1.

Proposition 2.1 (Distribution-uniform anytime-valid inference for the mean)**.** *Let $X_1, X_2, \ldots$ be random variables defined on $(\Omega, \mathcal{F}, \mathcal{P})$, and suppose that for some $\delta > 0$, the $(2+\delta)^{th}$ moment is $\mathcal{P}$-uniformly upper-bounded and the variance is $\mathcal{P}$-uniformly positive, i.e.*

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P|X - \mathbb{E}_P(X)|^{2+\delta} < \infty \quad \textit{and} \quad \inf_{P\in\mathcal{P}} \mathrm{Var}_P(X) > 0. \tag{12}$$

If $\mathcal{P}_0 \subseteq \mathcal{P}$ is a subcollection of distributions so that $\mathbb{E}_P(X) = 0$ for each $P \in \mathcal{P}_0$, then $(\bar{p}_k^{(m)})_{k=m}^{\infty}$ is a sharp $\mathcal{P}_0$-uniform anytime p-value:

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}_0} \mathbb{P}_P\left(\exists k \geqslant m : \bar{p}_k^{(m)} \leqslant \alpha\right) = \alpha, \tag{13}$$

and $(\bar{C}_k^{(m)}(\alpha))_{k=m}^{\infty}$ is a sharp $\mathcal{P}$-uniform CS for the mean:

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\exists k \geqslant m : \mu_P \notin \bar{C}_k^{(m)}(\alpha)\right) = \alpha. \tag{14}$$

Rather than prove Proposition 2.1 directly, we will spend the next few sections laying the groundwork to prove a more general result, culminating in Theorem 3.3. Clearly, one can obtain a sharp $\mathcal{P}_0$-uniform level-α anytime hypothesis test $(\bar{\Gamma}_k^{(m)})_{k=m}^{\infty}$ in the sense of Definition 2.1 from Proposition 2.1 by setting $\bar{\Gamma}_k^{(m)} := \mathbb{1}\{\bar{p}_k^{(m)} \leqslant \alpha\}$ or $\bar{\Gamma}_k^{(m)} := \mathbb{1}\{0 \notin \bar{C}_k^{(m)}\}$. Notice that uniformly bounded $(2+\delta)^{\text{th}}$ moment conditions are precisely what appear in several distribution-uniform central limit theorems [38, 28].

2.2 Time- and $\mathcal{P}$-uniform central limit theory for partial sums

Recall that in the batch (fixed-n, non-sequential) setting, the CLT is typically stated for a single quantile, meaning that the survival function $\mathbb{P}_P(S_n/\sqrt{n} \geqslant x)$ — equivalently, the CDF[1] — of $\sqrt{n}$-scaled normalized partial sums $S_n := \sigma^{-1}\sum_{i=1}^{n}[X_i - \mathbb{E}_P(X)]$ converge to that of a standard Gaussian:

$$\forall x \in \mathbb{R}, \ \lim_{n\to\infty} \left|\mathbb{P}_P(S_n/\sqrt{n} \geqslant x) - [1 - \Phi(x)]\right| = 0. \tag{15}$$

Under no additional assumptions, however, the above holds *quantile*-uniformly [44, Lemma 2.11], meaning

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} \left|\mathbb{P}_P(S_n/\sqrt{n} \geqslant x) - [1 - \Phi(x)]\right| = 0. \tag{16}$$

[1]This discussion is in terms of the survival function $\mathbb{P}_P(S_n/\sqrt{n} \geqslant x)$ instead of the CDF $\mathbb{P}_P(S_n/\sqrt{n} \leqslant x)$ to aid transparent comparisons with boundary-crossing inequalities in Proposition 2.2.

Clearly, (16) is strictly stronger than (15). Particularly relevant to this paper, distribution-uniform *fixed-n* tests and CIs are also stated with quantile-uniformity and their proofs typically rely on this property. Even in the P-pointwise case, however, there is no result showing that an analogous property exists for time-uniform *boundaries* (and it is not clear in what sense such a statement should be formulated). The following theorem provides such a result in both the P-pointwise and $\mathcal{P}$-uniform settings.

Proposition 2.2 (($\mathcal{P}, n, x$)-uniform boundaries for centered partial sums)**.** *Let $X_1, X_2, \ldots$ be random variables defined on probability spaces $(\Omega, \mathcal{F}, \mathcal{P}^\star)$ with finite $(2+\delta)^{th}$ moments, i.e. $\mathbb{E}_P|X - \mathbb{E}_P(X)|^{2+\delta} < \infty$ for every $P \in \mathcal{P}^\star$. Letting $S_n := \sum_{i=1}^n (X_i - \mathbb{E}_P(X_i))/\sigma_P$ be their centered partial sums, we have*

$$\forall P \in \mathcal{P}^\star, \ \lim_{m\to\infty} \sup_{x \geqslant 0} \left| \mathbb{P}_P \left(\exists k \geqslant m : |S_k| / \sqrt{k} \geqslant \sqrt{x + \log(k/m)} \right) - [1 - \Psi(x)] \right| = 0. \tag{17}$$

Furthermore, if $\mathcal{P} \subseteq \mathcal{P}^\star$ is a sub-collection of distributions for which the $(2+\delta)^{th}$ moment is $\mathcal{P}$-uniformly upper-bounded and the variance is $\mathcal{P}$-uniformly positive, then the above limit holds $\mathcal{P}$-uniformly:

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} \sup_{x \geqslant 0} \left| \mathbb{P}_P \left(\exists k \geqslant m : |S_k| / \sqrt{k} \geqslant \sqrt{x + \log(k/m)} \right) - [1 - \Psi(x)] \right| = 0. \tag{18}$$

The proof of Proposition 2.2 in Appendix A.1 relies on our novel distribution-uniform strong Gaussian approximation discussed in Section 5. After some algebraic manipulations, one can see that (18) is equivalent to saying that $\sup_{k\geqslant m}\{S_k^2/(\sigma_P^2 k) - \log(k/m)\}$ converges $\mathcal{P}$-uniformly in distribution to the Robbins-Siegmund distribution, i.e.

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} \sup_{x \geqslant 0} \left| \mathbb{P}_P \left(\sup_{k\geqslant m} \left\{ \frac{S_k^2}{\sigma_P^2 k} - \log(k/m) \right\} \leqslant x \right) - \Psi(x) \right| = 0, \tag{19}$$

and similarly for the P-pointwise case in (17) but with the above limit over m and supremum over $P \in \mathcal{P}$ swapped.

Note that Proposition 2.2 does not quite yield Proposition 2.1 as a direct consequence since the variance σ_P^2 used in the latter is the true (rather than empirical) variance. Moving to a fully empirical version of Proposition 2.2 will require that the variance σ_P^2 is not only consistently estimated but *almost surely* at a *polynomial rate* and for the $\mathcal{P}$-uniform result in (18), we will require that this consistency also holds uniformly in $\mathcal{P}$. But what does it mean for a sequence of random variables (such as a sequence of estimators) to converge to a limit almost surely *and uniformly in* $\mathcal{P}$? The next section provides an answer to this question alongside sufficient conditions for the sample variance to be $\mathcal{P}$-uniformly almost surely consistent.

3 Almost-sure consistency and time-uniform asymptotics

In Proposition 2.2, we stated a $\mathcal{P}$-, time-, and boundary-uniform convergence result for centered partial sums, but this depended on those partial sums $S_n(P) := \sum_{i=1}^n (X_i - \mu_P)/\sigma_P$ being weighted by the true standard deviation σ_P. The natural next step is to replace the true variance σ_P^2 by an *empirical* variance $\widehat{\sigma}_n^2$ so that the results of Proposition 2.2 still hold with $\widehat{\sigma}_n^2$ in place of σ_P^2, thereby providing tools that can be used to derive $\mathcal{P}$-uniform anytime-valid tests, p-values, and confidence sequences. However, the conditions that must be placed on $\widehat{\sigma}_n^2$ are different from what one may encounter in a classical asymptotic inference analysis — indeed we will require $\widehat{\sigma}_n^2$ to be $\mathcal{P}$-uniformly almost-surely consistent for $\sigma_P^2 := \mathbb{E}_P(X - \mathbb{E}_P X)^2$ at a faster-than-logarithmic rate. However, the notion of $\mathcal{P}$-uniform almost-sure convergence is not commonly encountered in the statistical literature, so this section is dedicated to reviewing this.

3.1 What is $\mathcal{P}$-uniform almost-sure consistency?

Recall the classical notion of convergence in P-probability for a single $P \in \mathcal{P}$ and its natural extension to $\mathcal{P}$-uniform convergence in probability. That is, a sequence of random variables $Y_1, Y_2, \ldots$ defined on a probability space $(\Omega, \mathcal{F}, P)$ is said to converge *in probability* to 0 (or $Y_n = \dot{o}_P(1)$ for short) if for any $\varepsilon > 0$,

$$\sup_{P \in \mathcal{P}} \lim_{n \to \infty} \mathbb{P}_P(|Y_n| \geqslant \varepsilon) = 0, \tag{20}$$

and that this convergence holds uniformly in $\mathcal{P}$ (or $Y_n = \dot{o}_{\mathcal{P}}(1)$ for short) if

$$\lim_{n \to \infty} \sup_{P \in \mathcal{P}} \mathbb{P}_P(|Y_n| \geqslant \varepsilon) = 0. \tag{21}$$

The extension of (20) to (21) is very natural, but at first glance, an analogous extension for almost-sure convergence is less obvious. Indeed, recall that a sequence of random variables $Y_1, Y_2, \ldots$ is said to converge P-almost surely to 0 for every $P \in \mathcal{P}$ if

$$\forall P \in \mathcal{P},\ \mathbb{P}_P\left(\lim_{n \to \infty} |Y_n| = 0\right) = 1. \tag{22}$$

It is not immediately obvious what the "right" notion of $\mathcal{P}$-uniform almost-sure consistency ought to be since taking an infimum over $P \in \mathcal{P}$ of the above probabilities does not change the statement of (22) whatsoever. Intuitively, it is not possible to simply swap limits and suprema in (22) as was done when (20) was extended to (21). However, it *is* possible to make such a leap when using an equivalent definition of almost-sure convergence using an idea attributable to Chung [11] and Beck and Giesy [3]. To elaborate, it is a well-known fact that for any $P \in \mathcal{P}$,

$$\mathbb{P}_P\left(\lim_{n \to \infty} |Y_n| = 0\right) = 1 \quad \text{if and only if} \quad \forall \varepsilon > 0,\ \lim_{m \to \infty} \mathbb{P}_P\left(\sup_{k \geqslant m} |Y_k| \geqslant \varepsilon\right) = 0,^{2} \tag{23}$$

and for this reason, instead of writing $Y_n = o_{\text{a.s.}}(1)$ as a shorthand for P-almost-sure convergence, we write $Y_n = \bar{o}_P(1)$ with the overhead bar $\bar{o}$ to emphasize time-uniformity and the subscript o_P to emphasize the distribution P that this convergence is with respect to. As such, a natural notion of $\mathcal{P}$-uniform almost-sure convergence is one that places a supremum over $P \in \mathcal{P}$ in the right-hand limit of (23) which we make precise in the following definition.

Definition 3.1 ($\mathcal{P}$-uniform almost-sure convergence [11, 48])**.** *We say that a sequence of random variables $Y_1, Y_2, \ldots$ defined on the probability spaces $(\Omega, \mathcal{F}, \mathcal{P})$ converges $\mathcal{P}$-uniformly and almost surely to 0 if*

$$\forall \varepsilon > 0, \quad \lim_{m \to \infty} \sup_{P \in \mathcal{P}} \mathbb{P}_P\left(\sup_{k \geqslant m} |Y_k| \geqslant \varepsilon\right) = 0, \tag{24}$$

and we write $Y_n = \bar{o}_{\mathcal{P}}(1)$ for short, where the overhead bar $\bar{o}$ emphasizes time-uniformity and the subscript $o_{\mathcal{P}}$ emphasizes $\mathcal{P}$-uniformity. Finally, we write $Y_n = \bar{o}_{\mathcal{P}}(r_n)$ for a monotonically nonincreasing sequence $(r_n)_{n=1}^{\infty}$ if $r_n \cdot Y_n = \bar{o}_{\mathcal{P}}(1)$.

The expression in (24) initially appeared in a paper by Chung [11] in a proof of a $\mathcal{P}$-uniform strong law of large numbers, and later in a more explicit form by Beck and Giesy [3]. Table 1 summarizes the four notions of convergence $\dot{o}_P(\cdot)$, $\bar{o}_P(\cdot)$, $\dot{o}_{\mathcal{P}}(\cdot)$, and $\bar{o}_{\mathcal{P}}(\cdot)$ and the implications between them.

Adapting (24) to the discussion of consistency in parameter estimation, however, requires some additional care since the parameter of interest may itself depend on the distribution $P \in \mathcal{P}$. That is, let $(\widehat{\theta}_n)_{n=1}^{\infty}$ be a sequence of estimators and for each $P \in \mathcal{P}$, let $\theta(P) \in \mathbb{R}$ be a real-valued parameter. We will consider $\widehat{\theta}_n$ to be a $\mathcal{P}$-uniformly consistent estimator for $\theta \equiv \{\theta(P)\}_{P \in \mathcal{P}}$ if

$$\lim_{m \to \infty} \sup_{P \in \mathcal{P}} \mathbb{P}_P\left(\sup_{k \geqslant m} |\widehat{\theta}_k - \theta(P)| \geqslant \varepsilon\right) = 0, \tag{25}$$

[2] For a short proof of this fact, see Waudby-Smith et al. [47, Section B.3]

Table 1: Four notions of convergence with implications between them. Recall that $\bar{o}_P(\cdot)$ is equivalent to P-a.s. convergence. Clearly, if a sequence of random variables converges with respect to one of the four cells below, it also does so with respect to the cell above and/or to the left of it. This section is concerned with the strongest of the four, found in the bottom right cell with the **bolded** frame: $\mathcal{P}$-uniform almost-sure convergence.

	P-**pointwise**		$\mathcal{P}$-**uniform**
In probability	$\dot{o}_P(\cdot)$	$\Longleftarrow$	$\dot{o}_{\mathcal{P}}(\cdot)$
	$\Uparrow$		$\Uparrow$
Almost surely	$\bar{o}_P(\cdot)$	$\Longleftarrow$	$\boxed{\bar{o}_{\mathcal{P}}(\cdot)}$

and as a shorthand, we will write $\widehat{\theta}_n - \theta = \bar{o}_{\mathcal{P}}(1)$. Similarly to Definition 3.1, we write $\widehat{\theta}_n - \theta = \bar{o}_{\mathcal{P}}(r_n)$ if $r_n \cdot (\widehat{\theta}_n - \theta) = \bar{o}_{\mathcal{P}}(1)$.

Following the relationship between $\dot{o}_P$ and $\dot{O}_P$ notation in the fixed-n in-probability setting, we now provide an analogous definition of time- and $\mathcal{P}$-uniform stochastic boundedness. To the best of our knowledge, this definition is new to the literature.

Definition 3.2. *We say that a sequence of random variables $Y_1, Y_2, \ldots$ defined on $(\Omega, \mathcal{F}, \mathcal{P})$ is time- and $\mathcal{P}$-uniformly stochastically bounded if for any $\delta > 0$, there exists some $C \equiv C(\delta) > 0$ and $M \equiv M(C, \delta) > 1$ so that for all $m \geqslant M$,*

$$\sup_{P \in \mathcal{P}} \mathbb{P}_P \left(\exists k \geqslant m : |X_k| > C \right) < \delta, \tag{26}$$

and we write $Y_n = \bar{O}_{\mathcal{P}}(1)$ as a shorthand for the above. Similar to Definition 3.1, we write $Y_n = \bar{O}_{\mathcal{P}}(r_n)$ if $r_n \cdot Y_n = \bar{O}_{\mathcal{P}}(1)$.

Note that we do *not* refer to Definition 3.2 as $\mathcal{P}$-uniform "almost sure" boundedness since even in the P-pointwise case, almost-sure boundedness and time-uniform stochastic boundedness are not equivalent despite the relationship in (23) for almost-sure and time-uniform *convergence*. A related condition has also appeared in the context of conditional local independence testing as in Christgau et al. [10]. As one may expect, there is a calculus of $\bar{o}_{\mathcal{P}}(\cdot)$ and $\bar{O}_{\mathcal{P}}(\cdot)$ analogous to that for $\dot{o}_{\mathcal{P}}(\cdot)$ and $\dot{O}_{\mathcal{P}}(\cdot)$. We lay this out formally in the following lemma, but the proofs are routine and can be found in Appendix A.2.

Lemma 3.1 (Calculus of $\bar{O}_{\mathcal{P}}(\cdot)$ and $\bar{o}_{\mathcal{P}}(\cdot)$)**.** *Let $Y_1, Y_2, \ldots$ be random variables defined on $(\Omega, \mathcal{F}, \mathcal{P})$. Let $(a_n)_{n=1}^{\infty}$ and $(b_n)_{n=1}^{\infty}$ be positive and monotonically nonincreasing sequences. Then we have the following basic implications:*

$$Y_n = \bar{o}_{\mathcal{P}}(a_n) \implies Y_n = \bar{O}_{\mathcal{P}}(a_n) \tag{27}$$

$$Y_n = \bar{o}_{\mathcal{P}}(a_n)\bar{O}_{\mathcal{P}}(b_n) \implies Y_n = \bar{o}_{\mathcal{P}}(a_n b_n) \tag{28}$$

$$Y_n = \bar{O}_{\mathcal{P}}(a_n)\bar{O}_{\mathcal{P}}(b_n) \implies Y_n = \bar{O}_{\mathcal{P}}(a_n b_n) \tag{29}$$

$$Y_n = \bar{o}_{\mathcal{P}}(a_n) + \bar{O}_{\mathcal{P}}(a_n) \implies Y_n = \bar{O}_{\mathcal{P}}(a_n) \tag{30}$$

$$Y_n = \bar{o}_{\mathcal{P}}(a_n) + \bar{o}_{\mathcal{P}}(b_n) \implies Y_n = \bar{o}_{\mathcal{P}}(\max\{a_n, b_n\}). \tag{31}$$

Furthermore, (31) holds with $\bar{o}_{\mathcal{P}}(\cdot)$ replaced by $\bar{O}_{\mathcal{P}}(\cdot)$ on both sides. Finally, if $Y_n = \bar{O}_{\mathcal{P}}(a_n)$ and $a_n/b_n \to 0$, then $Y_n = \bar{o}_{\mathcal{P}}(b_n)$.

The calculus provided in Lemma 3.1 will appear frequently throughout the proofs of our main results. In the next section, we discuss $\mathcal{P}$-uniform, almost-sure, polynomial-rate variance estimation and its implications for deriving an empirical version of Proposition 2.2.

3.2 $\mathcal{P}$-uniform almost-sure variance estimation

In Section 2.2, we alluded to the fact that arriving at a fully empirical version of Proposition 2.2 would require $\mathcal{P}$-uniform almost-surely consistent estimation of the variance $\sigma^2 := \mathbb{E}(X - \mathbb{E}X)^2$ at a faster-than-logarithmic rate. With Definition 3.1 and the expression (24) in mind, we now provide sufficient conditions for this consistency.

Proposition 3.2 ($\mathcal{P}$-uniform almost-surely consistent variance estimation)**.** *Consider the same setup as in Proposition 2.2 where $(X_n)_{n=1}^\infty$ have $\mathcal{P}$-uniformly upper-bounded $(2+\delta)^{th}$ moments and $\mathcal{P}$-uniformly positive variances. Then the sample variance $\widehat{\sigma}_n^2 := \frac{1}{n}\sum_{i=1}^n (X_i - \widehat{\mu}_n)^2$ is a $\mathcal{P}$-uniformly almost-surely consistent estimator of the variance σ^2 at a polynomial rate, meaning there exists $\beta > 0$ so that*

$$\widehat{\sigma}_n^2 = \sigma^2 + \bar{o}_{\mathcal{P}}(n^{-\beta}), \tag{32}$$

or more formally, for all $\varepsilon > 0$, we have

$$\lim_{m\to\infty} \mathbb{P}_P\left(\sup_{k\geqslant m} k^\beta |\widehat{\sigma}_k^2 - \sigma_P^2| \geqslant \varepsilon\right) = 0. \tag{33}$$

Proposition 3.2 is an immediate consequence of Waudby-Smith, Larsson, and Ramdas [48] combined with the de la Vallée Poussin criterion for uniform integrability (see Chong [9] and Hu and Rosalsky [19]).

3.3 The main result: $(\mathcal{P}, n, \alpha)$-uniform statistical inference

Pairing together Proposition 2.2 and Proposition 3.2, we obtain the following ($\mathcal{P}$-uniform) anytime p-values and CSs whose type-I errors converge to the nominal level $\alpha \in (0,1)$ uniformly in α. We present this in the following result on distribution-, time-, and α-uniform — or $(\mathcal{P}, n, \alpha)$-uniform for short — statistical inference. This is our main result and it implies both Proposition 2.1 and Proposition 2.2 as special cases.

Theorem 3.3 ($(\mathcal{P}, n, \alpha)$-uniform statistical inference)**.** *Let $X_1, X_2, \ldots$ be defined on $(\Omega, \mathcal{F}, \mathcal{P})$ and suppose that for some $\delta > 0$, the $(2+\delta)^{th}$ moment is $\mathcal{P}$-uniformly upper-bounded and the variance is $\mathcal{P}$-uniformly positive. Recall the definitions of $(\bar{p}_k^{(m)})_{k=m}^\infty$ and $(\bar{C}_k^{(m)}(\alpha))_{k=m}^\infty$ from Proposition 2.1:*

$$\bar{p}_k^{(m)} := 1 - \Psi\left(k\widehat{\mu}_k^2/\widehat{\sigma}_k^2 - \log(k/m)\right) \tag{34}$$

$$\text{and} \quad \bar{C}_k^{(m)}(\alpha) := \widehat{\mu}_k \pm \widehat{\sigma}_k\sqrt{[\Psi^{-1}(1-\alpha) + \log(k/m)]/k}. \tag{35}$$

Let $\mathcal{P}_0 \subseteq \mathcal{P}$ be a subcollection of distributions so that $\mathbb{E}_P(X) = 0$ for each $P \in \mathcal{P}_0$. Then the time-uniform type-I error of $(\bar{p}_k^{(m)})_{k=m}^\infty$ and the time-uniform miscoverage of $(\bar{C}_k^{(m)})_{k=m}^\infty$ converge to $\alpha \in (0,1)$ uniformly in α, meaning

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}_0} \sup_{\alpha\in(0,1)} \left|\mathbb{P}_P\left(\exists k \geqslant m : \bar{p}_k^{(m)} \leqslant \alpha\right) - \alpha\right| = 0, \quad \text{and} \tag{36}$$

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} \sup_{\alpha\in(0,1)} \left|\mathbb{P}_P\left(\exists k \geqslant m : \mathbb{E}_P(X) \notin \bar{C}_k^{(m)}(\alpha)\right) - \alpha\right| = 0. \tag{37}$$

The full proof of Theorem 3.3 can be found in Appendix A.3. As alluded to at the beginning of Section 2, its proof relies on a $\mathcal{P}$-uniform strong Gaussian approximation theorem discussed in Section 5. Before that, we will discuss how the results derived thus far can be used to conduct distribution-uniform anytime-valid tests of conditional independence.

4 Illustration: Sequential conditional independence testing

In this section, we aim to derive anytime-valid tests for the null hypothesis, $X \perp\!\!\!\perp Y \mid Z$ given $\mathbb{R} \times \mathbb{R} \times \mathbb{R}^d$-valued triplets $(X_n, Y_n, Z_n)_{n=1}^{\infty}$ on probability spaces $(\Omega, \mathcal{F}, \mathcal{P})$. Several works on conditional independence testing operate under the so-called "Model-X" assumption where the conditional distribution of $X \mid Z$ is known exactly [6]. We do not work under the Model-X assumption in this illustration. It is well-known that testing for conditional independence is much simpler under Model-X, and indeed the recent works of Duan et al. [13], Shaer et al. [37], and Grünwald et al. [17] derive powerful anytime-valid tests in that paradigm. Borrowing a quote from the recent work of Grünwald et al. [17], the authors write *"it is an open question to us how to construct general sequential tests of conditional independence without the [Model-X] assumption"*. This section gives an answer to this question, deriving tests that draw inspiration from the batch tests found in Shah and Peters [38] — a pair of authors we will henceforth refer to as S&P. Before giving a brief refresher on batch conditional independence testing and the main results of S&P, let us review some basic concepts in weak regression consistency since nuisance function estimation will form key conditions for our results.

4.1 Prelude: weak regression consistency

An important part of conditional independence testing (in both batch and sequential settings as we will see) is the ability to consistently estimate certain regression functions. Recall that the (potentially random) squared $L_2(P)$ risk of a regression estimator $\widehat{f}_n : \mathbb{R}^d \to \mathbb{R}$ for a function $f : \mathbb{R}^d \to \mathbb{R}$ is given by

$$\|\widehat{f}_n - f\|^2_{L_2(P)} := \int_{z \in \mathbb{R}^d} \left(\widehat{f}_n(z) - f(z)\right)^2 dP(z). \tag{38}$$

Importantly, if sample splitting is used to construct $\widehat{f}_n$, the norm $\|\cdot\|_{L_2(P)}$ is to be interpreted as conditional on that "training" data. Recall from Györfi et al. [18, Definition 1.1] that a regression estimator $\widehat{f}_n : \mathbb{R}^d \to \mathbb{R}$ is P-weakly consistent for a function $f : \mathbb{R}^d \to \mathbb{R}$ in $L_2(P)$ at a rate of r_n if its expected $L_2(P)$ risk vanishes at that rate, meaning

$$\mathbb{E}_P \|\widehat{f}_n - f\|_{L_2(P)} = o(r_n), \tag{39}$$

and hence we will say that $\widehat{f}_n$ is $\mathcal{P}$-weakly consistent at the rate r_n if the above convergence occurs uniformly in the class of distributions $\mathcal{P}$:

$$\sup_{P \in \mathcal{P}} \mathbb{E}_P \|\widehat{f}_n - f\|_{L_2(P)} = o(r_n). \tag{40}$$

At times, we may omit $L_2(P)$ from the norm $\|\cdot\|_{L_2(P)}$ in (39) and write $\|\cdot\|$ when the norm is clear from context.

4.2 A brief refresher on batch conditional independence testing

Given $\mathbb{R} \times \mathbb{R} \times \mathbb{R}^d$-valued triplets $(X_i, Y_i, Z_i)_{i=1}^n$ from some distribution in a class $\mathcal{P}$, the problem of conditional independence testing is concerned with the null

$$H_0 : X \perp\!\!\!\perp Y \mid Z \quad \text{versus the alternative} \quad H_1 : X \not\perp\!\!\!\perp Y \mid Z. \tag{41}$$

As alluded to before, without the Model-X assumption, powerful tests for the conditional independence null H_0 in (41) are *impossible* to derive (even in the batch and asymptotic settings) unless additional distributional or structural assumptions are imposed [S&P, §2]. Indeed, S&P show that even in the bounded setting where $(X, Y, Z) \sim P \in \mathcal{P}^\star$ take values in $[0,1] \times [0,1] \times [0,1]$, any test with distribution-uniform type-I error control under H_0 is powerless against *any* alternative in H_1.

Formally, if $\mathcal{P}_0{}^\star \subset \mathcal{P}^\star$ is the subset of distributions satisfying H_0 (and hence $\mathcal{P}_1^\star := \mathcal{P}^\star \backslash \mathcal{P}_0{}^\star$ satisfies H_1), then

$$\underbrace{\sup_{P \in \mathcal{P}_1^\star} \limsup_{n\to\infty} \mathbb{P}_P\left(\dot{\Gamma}_n = 1\right)}_{\text{Best-case } \mathcal{P}_1^\star\text{-pointwise power}} \leqslant \underbrace{\limsup_{n\to\infty} \sup_{P \in \mathcal{P}_0^\star} \mathbb{P}_P\left(\dot{\Gamma}_n = 1\right)}_{\text{Worst-case } \mathcal{P}_0^\star\text{-uniform type-I error}}. \tag{42}$$

As a consequence of (42), one cannot derive a more powerful test than the trivial one that ignores all of the data $(X_i, Y_i, Z_i)_{i=1}^n$ and randomly outputs 1 with probability α.

Despite the rather pessimistic result in (42), S&P derive the Generalized Covariance Measure (GCM) test which manages to achieve nontrivial power while still uniformly controlling the type-I error. The caveat here is that they are controlling the type-I error in a restricted (but nevertheless rich and nonparametric) class of nulls $\mathcal{P}_0 \subseteq \mathcal{P}_0^\star$, and the restriction they impose is that certain nuisance functions are sufficiently estimable, a requirement commonly appearing in other literatures including semiparametric functional estimation [21, 2]. Let us now review the key aspects of their test. S&P introduce the estimated residuals $R_{i,n}$ for each $i \in [n]$:

$$R_{i,n} := \{X_i - \widehat{\mu}_n^x(Z_i)\}\{Y_i - \widehat{\mu}_n^y(Z_i)\} \tag{43}$$

where $\widehat{\mu}_n^x(z)$ and $\widehat{\mu}_n^y(z)$ are estimates of the regression functions $\mu^x(z) := \mathbb{E}(X \mid Z = z)$ and $\mu^y(z) := \mathbb{E}(Y \mid Z = z)$. For the remainder of the discussion on batch conditional independence testing, we will assume that $\widehat{\mu}_n^x(Z_i)$ and $\widehat{\mu}_n^y(Z_i)$ are constructed from an independent sample (e.g. through sample-splitting or cross-fitting, in which case we may assume access to $2n$ triplets of (X, Y, Z)) for mathematical simplicity, but S&P do not always suggest doing so. However, we will not dwell on arguments for or against sample splitting here. From the residuals in (43), they construct the test statistic $\mathrm{G\dot{C}M}_n$ taking the form

$$\mathrm{G\dot{C}M}_n := \frac{1}{n\widehat{\sigma}_n} \sum_{i=1}^n R_{i,n} \tag{44}$$

where $\widehat{\sigma}_n^2 := \frac{1}{n}\sum_{i=1}^n R_{i,n}^2 - \left(\frac{1}{n}\sum_{i=1}^n R_{i,n}\right)^2$ and they show that if the regression functions (μ^y, μ^x) are estimated sufficiently fast (and under some other mild regularity conditions) then $\sqrt{n}\mathrm{G\dot{C}M}_n$ has a standard Gaussian limit, enabling asymptotic (fixed-n) inference. We formally recall a minor simplification of their main result here. Consider the following three assumptions for a class of distributions $\mathcal{P}_0$.

Assumption GCM-1 (Product regression error decay)**.** *The weak convergence rate of the average of product residuals is faster than $n^{-1/2}$, i.e.*

$$\sup_{P \in \mathcal{P}_0} \|\mu^x - \widehat{\mu}_n^x\|_{L_2(P)} \cdot \|\mu^y - \widehat{\mu}_n^y\|_{L_2(P)} = o(n^{-1/2}). \tag{45}$$

Assumption GCM-2 ($\mathcal{P}_0$-uniform regularity of regression errors)**.** *Letting $\xi^x := \{X - \mu^x(Z)\}$ and $\xi^y := \{Y - \mu^y(Z)\}$ denote the true residuals, the variances of $\{\widehat{\mu}_n^x(Z) - \mu^x(Z)\}\xi^y$ and $\{\widehat{\mu}_n^y(Z) - \mu^y(Z)\}\xi^x$ are $\mathcal{P}_0$-uniformly vanishing, i.e.*

$$\sup_{P \in \mathcal{P}_0} \mathrm{Var}_P\left(\{\widehat{\mu}_n^x(Z) - \mu^x(Z)\} \cdot \xi^y\right) = o(1) \tag{46}$$

$$\text{and} \quad \sup_{P \in \mathcal{P}_0} \mathrm{Var}_P\left(\{\widehat{\mu}_n^y(Z) - \mu^y(Z)\} \cdot \xi^x\right) = o(1). \tag{47}$$

Assumption GCM-3 ($\mathcal{P}_0$-uniformly bounded moments)**.** *The true product residuals defined above have $\mathcal{P}_0$-uniformly upper-bounded $(2+\delta)^{th}$ moments for some $\delta > 0$ and uniformly lower-bounded second moments:*

$$\sup_{P \in \mathcal{P}_0} \mathbb{E}_P\left|\xi^x \xi^y\right|^{2+\delta} < \infty \tag{48}$$

$$\text{and} \quad \inf_{P \in \mathcal{P}_0} \mathrm{Var}_P(\xi^x \xi^y) > 0. \tag{49}$$

With these three assumptions in mind, we are ready to recall a simplified version of Shah and Peters [38, Theorem 6].

Theorem (S&P: $\mathcal{P}_0$-uniform validity of the GCM test)**.** *Suppose $(X_i, Y_i, Z_i)_{i=1}^n$ are $\mathbb{R}\times\mathbb{R}\times\mathbb{R}^d$-valued random variables on the probability spaces $(\Omega, \mathcal{F}, \mathcal{P})$ and let $\mathcal{P}_0 \subset \mathcal{P}$ be the collection of distributions in $\mathcal{P}$ satisfying the conditional independence null H_0 and Assumptions GCM-1, GCM-2, and GCM-3. Then,*

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}_0} \sup_{x\in\mathbb{R}} \left| \mathbb{P}_P(\sqrt{n}\dot{\mathrm{GCM}}_n \leqslant x) - \Phi(x) \right| = 0. \tag{50}$$

and hence the function given by $\Gamma_k^{(m)} := \mathbb{1}\left\{|\sqrt{n}\dot{\mathrm{GCM}}_n| \geqslant \Phi^{-1}(1-\alpha/2)\right\}$ is a $\mathcal{P}_0$-uniform level-α test.

We will now shift our focus to *sequential* conditional independence testing with anytime-valid type-I error guarantees. Before deriving an explicit test, we first demonstrate in Proposition 4.1 that the hardness of conditional independence testing highlighted in (42) has a similar analogue in the anytime-valid regime.

4.3 On the hardness of anytime-valid conditional independence testing

As mentioned in Section 4.2, S&P illustrated the fundamental hardness of conditional independence testing by showing that unless additional restrictions are placed on the null hypothesis $\mathcal{P}_0^\star$, any $\mathcal{P}_0^\star$-uniformly valid (fixed-n) test is powerless against any alternative, i.e.

$$\underbrace{\sup_{P\in\mathcal{P}_1^\star} \limsup_{n\to\infty} \mathbb{P}_P\left(\dot{\Gamma}_n = 1\right)}_{\text{Best-case } \mathcal{P}_1^\star\text{-pointwise power}} \leqslant \underbrace{\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_0^\star} \mathbb{P}_P\left(\dot{\Gamma}_n = 1\right)}_{\text{Worst-case } \mathcal{P}_0^\star\text{-uniform type-I error}}. \tag{42 revisited}$$

Does an analogous result hold if $\dot{\Gamma}_n$ is replaced by an anytime-valid hypothesis test $\bar{\Gamma}_k^{(m)}$ as in Definition 2.1? The following proposition gives an answer to this question, confirming that anytime-valid conditional independence testing is fundamentally hard in a sense similar to (42).

Proposition 4.1 (Hardness of anytime-valid conditional independence testing)**.** *Suppose $(X_n, Y_n, Z_n)_{n=1}^\infty$ are $[0,1]^3$-valued triplets on the probability spaces $(\Omega, \mathcal{F}, \mathcal{P}^\star)$ where $\mathcal{P}^\star$ consists of all distributions supported on $[0,1]^3$. Let $\mathcal{P}_0^\star \subseteq \mathcal{P}^\star$ be the subset of distributions satisfying the conditional independence null H_0 and denote $\mathcal{P}_1^\star := \mathcal{P}^\star \backslash \mathcal{P}_0^\star$. Then for any potentially randomized test $(\bar{\Gamma}_k^{(m)})_{k=m}^\infty$,*

$$\sup_{P\in\mathcal{P}_1^\star} \limsup_{m\to\infty} \mathbb{P}_P\left(\exists k \geqslant m : \bar{\Gamma}_k^{(m)} = 1\right) \leqslant \limsup_{m\to\infty} \sup_{P\in\mathcal{P}_0^\star} \mathbb{P}_P\left(\exists k \geqslant m : \bar{\Gamma}_k^{(m)} = 1\right). \tag{51}$$

In other words, no $\mathcal{P}_0^\star$-uniform anytime-valid test can have power against any alternative in $\mathcal{P}_1^\star$ at any $\{m, m+1, \dots\}$-valued stopping time no matter how large m is.

The proof can be found in Appendix A.4. It should be noted that Proposition 4.1 is not an immediate consequence of S&P's fixed-n hardness result in (42) since while it is true that the time-uniform *type-I error* in the right-hand side of (51) is always larger than its fixed-n counterpart, the time-uniform *power* in the left-hand side of (51) is typically much larger than the fixed-n power. Indeed, while an important facet of hypothesis testing is to find tests with power as close to 1 as possible, the time-uniform power of anytime-valid tests is typically *equal to* 1, and such tests are sometimes referred to explicitly as "tests of power 1" for this reason [34]. This should not be surprising since the ability to reject at any stopping time (data-dependent sample size) larger than m introduces a great deal of flexibility. The fact that this flexibility is insufficient to overcome $\mathcal{P}_0^\star$-uniform control of the time-uniform type-I error is what makes Proposition 4.1 nontrivial.

Using the techniques of Section 2, we will now derive an anytime-valid analogue of S&P's GCM test with similar distribution-uniform guarantees, allowing the tests and p-values to be continuously monitored and adaptively stopped.

4.4 SeqGCM: The sequential generalized covariance measure test

We will now lay out the assumptions required for our SeqGCM test to have distribution-uniform anytime-validity. Similar to our discussion of the batch GCM test in the previous section, we will assume that for each n, $\widehat{\mu}_n^x$ and $\widehat{\mu}_n^y$ are trained from an independent sample. This can be achieved easily by supposing that at each time n, we observe pairs $(X_1^{(n)}, Y_1^{(n)}, Z_1^{(n)}), (X_2^{(n)}, Y_2^{(n)}, Z_2^{(n)})$ where the first is used for training $(\widehat{\mu}_i^x, \widehat{\mu}_i^y)_{i=n}^{\infty}$ and the second is used for evaluating $\{X_n - \widehat{\mu}_n^x(Z_n)\} \cdot \{Y_n - \widehat{\mu}_n^y(Z_n)\}$.

Recall that in S&P's GCM test, the test statistic $\dot{\mathrm{GCM}}_n := \frac{1}{n}\sum_{i=1}^n R_{i,n}/\widehat{\sigma}_n^2$ was built from the product residuals $R_{i,n}$ that were defined in (43) as

$$R_{i,n} := \{X_i - \widehat{\mu}_n^x(Z_i)\}\{Y_i - \widehat{\mu}_n^y(Z_i)\}. \tag{52}$$

In particular, note that the regression estimators $\widehat{\mu}_n^x$ and $\widehat{\mu}_n^y$ are trained *once* on a held-out sample of size n and then evaluated on $Z_1, \ldots, Z_n$, which is perfectly natural in the batch setting. By contrast, we will evaluate the product residual

$$R_n := \{X_n - \widehat{\mu}_n^x(Z_n)\}\{Y_n - \widehat{\mu}_n^y(Z_n)\} \tag{53}$$

to arrive at the test statistic

$$\overline{\mathrm{GCM}}_n := \frac{1}{n\widehat{\sigma}_n}\sum_{i=1}^n R_i, \tag{54}$$

where we will abuse notation slightly and redefine $\widehat{\sigma}_n^2 := \frac{1}{n}\sum_{i=1}^n R_i^2 - \left(\frac{1}{n}\sum_{i=1}^n R_i\right)^2$. The main difference between (52) and (53) is that in the latter case, the index for regression estimators $(\widehat{\mu}_n^x, \widehat{\mu}_n^y)$ is the same as those on which these functions are evaluated. Notice that while $\overline{\mathrm{GCM}}_n$ is more amenable to online updates than $\dot{\mathrm{GCM}}_n$, it does less to exploit the most up-to-date regression estimates. Nevertheless, as we will see shortly, it is still possible to control the distribution- and time-uniform asymptotic behavior of $\overline{\mathrm{GCM}}_n$ under *weak* regression consistency conditions on $(\widehat{\mu}_n^x, \widehat{\mu}_n^y)$. This is in contrast to some earlier work of Waudby-Smith et al. [47, Section 3] that also considered asymptotic time-uniform inference with nuisance estimation (focusing on the problem of average treatment effect estimation), but relied on *strong* regression consistency conditions. It should be noted that the weak consistency rates we impose here are polylogarithmically faster than those considered by Waudby-Smith et al. [47]. The key technique that will allow us to derive *strong* convergence behavior of certain sample averages of nuisances from *weak* consistency of regression functions is a distribution-uniform strong law of large numbers due to Waudby-Smith, Larsson, and Ramdas [48, Theorem 2]. This will be discussed further after the statement of Theorem 4.2.

Since the assumptions required for our SeqGCM test are similar in spirit to those of S&P's batch GCM test (Assumptions GCM-1, GCM-2, and GCM-3) we correspondingly name them "Assumptions SeqGCM-1 and SeqGCM-2" and underline certain keywords to highlight their differences (we do not need to make additional moment assumptions beyond those found in Assumption GCM-3, and thus there is no "SeqGCM-3" to introduce).

Assumption SeqGCM-1 (Product regression error decay)**.** *The weak convergence rate of the product of average squared residuals is no slower than* $\underline{(n\log^{2+\delta} n)^{-1/2}}$ *for some* $\delta > 0$*, i.e.*

$$\sup_{P\in\mathcal{P}_0} \|\widehat{\mu}_n^x - \mu^x\|_{L_2(P)} \cdot \|\widehat{\mu}_n^y - \mu^y\|_{L_2(P)} = O\left(\frac{1}{\sqrt{n\log^{2+\delta}(n)}}\right), \tag{55}$$

Assumption SeqGCM-2 ($\mathcal{P}_0$-uniform regularity of regression errors)**.** *Both* $\mathrm{Var}\left(\{\widehat{\mu}_n^x(Z) - \mu^x(Z)\}\cdot\xi_n^y\right)$ *and* $\mathrm{Var}\left(\{\widehat{\mu}_n^y(Z) - \mu^y(Z)\}\cdot\xi_n^x\right)$ *are* $\mathcal{P}_0$*-uniformly vanishing to 0 no slower than* $\underline{1/(\log n)^{2+\delta}}$ *for some*

$\delta > 0$, *i.e.*

$$\sup_{P \in \mathcal{P}_0} \mathrm{Var}_P\left(\{\widehat{\mu}_n^x(Z) - \mu^x(Z)\} \cdot \xi^y\right) = O\left(\frac{1}{(\log n)^{2+\delta}}\right) \tag{56}$$

$$\textit{and} \quad \sup_{P \in \mathcal{P}_0} \mathrm{Var}_P\left(\{\widehat{\mu}_n^y(Z) - \mu^y(Z)\} \cdot \xi^x\right) = O\left(\frac{1}{(\log n)^{2+\delta}}\right). \tag{57}$$

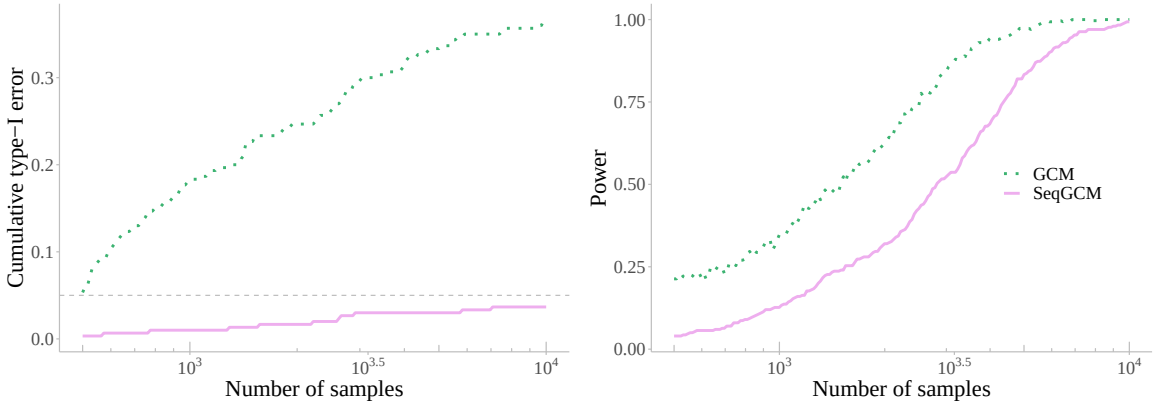

Figure 1: Empirical cumulative type-I error rates and power for the fixed-n GCM test of S&P versus the sequential GCM test (SeqGCM) in Theorem 4.2 with a target type-I error of $\alpha = 0.05$ in a simulated conditional independence testing problem. Notice that in the left-hand side plot, the type-I error rate for the GCM starts at around $\alpha = 0.05$ but steadily grows as more samples are collected. By contrast, the SeqGCM test remains below $\alpha = 0.05$ for all $k \geqslant m = 300$. In the right-hand side plot, we see that the power of the GCM test is higher than that of SeqGCM. This is unsurprising given that SeqGCM has a stronger (time-uniform) type-I error guarantee, but both have power near 1 after 10,000 samples.

With Assumptions SeqGCM-1, SeqGCM-2, and GCM-3 in mind, we are ready to state the $\mathcal{P}_0$-uniform type-I error guarantees of the SeqGCM test.

Theorem 4.2 ($\mathcal{P}_0$-uniform type-I error control of the SeqGCM)**.** *Suppose $(X_i, Y_i, Z_i)_{i=1}^{\infty}$ are $\mathbb{R} \times \mathbb{R} \times \mathbb{R}^d$-valued triplets defined on the probability spaces $(\Omega, \mathcal{F}, \mathcal{P})$ and let $\mathcal{P}_0 \subseteq \mathcal{P}$ be a collection of distributions in $\mathcal{P}$ satisfying the conditional independence null H_0 and Assumption SeqGCM-1, SeqGCM-2, and GCM-3. Define*

$$\bar{p}_{k,m}^{\mathrm{GCM}} := 1 - \Psi\left(k(\overline{\mathrm{GCM}}_k)^2 - \log(k/m)\right). \tag{58}$$

Then $(\bar{p}_{k,m}^{\mathrm{GCM}})_{k=m}^{\infty}$ forms a $\mathcal{P}_0$-uniform anytime p-value for the conditional independence null:

$$\lim_{m \to \infty} \sup_{P \in \mathcal{P}_0} \sup_{\alpha \in (0,1)} \left|\mathbb{P}_P\left(\exists k \geqslant m : \bar{p}_{k,m}^{\mathrm{GCM}} \leqslant \alpha\right) - \alpha\right| = 0. \tag{59}$$

The proof can be found in Appendix A.5 and uses the results from the previous sections combined with the distribution-uniform strong laws of large numbers (SLLNs) for independent but non-identically distributed random variables due to Waudby-Smith, Larsson, and Ramdas [48, Theorem 2]. The latter is crucial to analyzing the (uniform) almost sure convergence properties of sample averages with online regression estimators under weak consistency assumptions (SeqGCM-1 and SeqGCM-2).

To give some intuition as to when Assumption SeqGCM-1 may be satisfied, suppose that μ^x and μ^y are d-dimensional and Hölder s-smooth [18, §3.2]. Note that the minimax rate for estimating such functions in the resulting class of distributions $\mathcal{P}(s)$ is given by

$$\inf_{\widehat{\mu}_n^x} \sup_{P \in \mathcal{P}(s)} \mathbb{E}_P \|\widehat{\mu}_n^x - \mu^x\|_{L_2(P)}^2 \asymp n^{-2s/(2s+d)}, \tag{60}$$

and similarly for μ^y. In particular, if $d < 2s$ so that the dimension is not too large relative to the smoothness, then minimax-optimal local polynomial estimators $\widehat{\mu}_n^x$ and $\widehat{\mu}_n^y$ for μ^x and μ^y can be constructed and will be $\mathcal{P}(s)$-weakly consistent at rates of $o\left((n \log^{2+\delta} n)^{-1/4}\right)$. In this case, Assumption SeqGCM-1 (and Assumption GCM-1) will be satisfied as long as $\mathcal{P}_0 \subseteq \mathcal{P}(s)$. More broadly, any regression algorithms can be used to construct $\widehat{\mu}_n^x$ and $\widehat{\mu}_n^y$ (e.g. using random forests, neural networks, nearest neighbors, etc.) and they can further be selected via cross-validation or aggregated [5, 43].

The left-hand side plot of Fig. 1 demonstrates how the SeqGCM test controls the type-I error rate under the null uniformly over time while the standard GCM test fails to. The right-hand side plot compares their empirical power under one alternative.

5 Distribution-uniform strong Gaussian approximation

In this section, we both articulate what it means for a strong (almost-sure) coupling to be "$\mathcal{P}$-uniform" and then provide such a coupling in the form of a strong Gaussian approximation in Theorem 5.1. Before that, however, let us give a brief historical overview of weak and strong Gaussian approximations in the P-pointwise setting to contextualize and motivate the result to come. Given iid random variables $(X_1, \dots, X_n)$ with mean μ and finite variance σ^2 on a probability space $(\Omega, \mathcal{F}, P)$, the CLT states that standardized partial sums $S_n := \sum_{i=1}^n (X_i - \mu)/\sigma$ converge in distribution to a standard Gaussian with CDF $\Phi(x)$ after $\sqrt{n}$-rescaling:

$$\forall x \in \mathbb{R}, \ \lim_{n \to \infty} \mathbb{P}_P(S_n/\sqrt{n} \leqslant x) \to \Phi(x). \tag{61}$$

Note that (61) is only a statement about the *distribution* of S_n, but a stronger statement can be made in terms of a *coupling* between S_n and a partial sum of iid Gaussians [16, Eq. (1.2)]. Concretely, one can define a new probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P})$ containing random vectors $((\widetilde{X}_1, Y_1), (\widetilde{X}_2, Y_2), \dots, (\widetilde{X}_n, Y_n))$ where $(Y_1, \dots, Y_n)$ are marginally standard Gaussian and $(\widetilde{X}_1, \dots, \widetilde{X}_n)$ have the same marginal distribution as $(X_1, \dots, X_n)$ so that

$$\widetilde{S}_n - G_n = \dot{o}_{\widetilde{P}}(\sqrt{n}), \tag{62}$$

where $\widetilde{S}_n := \sum_{i=1}^n \widetilde{X}_i$ and $G_n := \sum_{i=1}^n Y_i$ and without loss of generality, we may simply write $S_n - G_n = \dot{o}_P(\sqrt{n})$. Indeed, we could have simply started with a probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P})$ rich enough to describe (X, Y) jointly and for this reason, some authors write "without loss of generality" to refer to this probability space construction [41]. Crucially, $Y_1, \dots, Y_n$ are independent of each other, but the random variables S_n and G_n are highly dependent, and clearly (62) $\implies$ (61).

For the purposes of obtaining a *time-uniform* guarantee, however, neither (61) nor (62) are sufficient since they only hold for a single sample size n, and naive union bounds over $n \in \mathbb{N}$ are not sharp enough to remedy the issue. Fortunately, there do exist analogues of (62) that hold *almost-surely* and hence uniformly for all n simultaneously. The study of such results — *strong* Gaussian approximations — began with the seminal results of Strassen [40] who used the Skorokhod embedding [39] to obtain an almost-sure analogue of (62) but with an iterated logarithm rate:

$$S_n - G_n = o_{\text{a.s.}}(\sqrt{n \log \log n}), \tag{63}$$

where $o_{\text{a.s.}}(\cdot)$ denotes P-a.s. convergence. As noted in (23), the above is equivalent to saying that for any $\varepsilon > 0$, we have $\lim_{m \to \infty} \mathbb{P}_P(\exists k \geqslant m : |S_k - G_k| > \varepsilon) = 0$, and we write this as

$$S_n - G_n = \bar{o}_P(\sqrt{n \log \log n}), \tag{64}$$

following the notation laid out in Section 3. Improvements to the iterated logarithm rate in (63) and (64) were made by Strassen [41] under higher moment assumptions, with the optimal rates uncovered in the famous papers by Komlós, Major, and Tusnády [22, 23] and Major [29].

Here, we do not focus on attaining optimal coupling rates since error rates incurred from estimation of nuisances (such as the variance) typically dominate them and optimal rates would not change our main statistical results in any meaningful way (much like they do not "improve" CLT-based confidence intervals). However, we do highlight the fact that the results of Strassen [40, 41], Komlós, Major, and Tusnády [22, 23], Major [29], and every other work on strong approximation to our knowledge only hold P-a.s. for a fixed P, and hence are not $\mathcal{P}$-uniform in any sense. We will now define "distribution-uniform strongly coupled processes" in Definition 5.1 and subsequently provide one such coupling in Theorem 5.1.

Definition 5.1 ($(\mathcal{P}, n)$-uniformly coupled stochastic processes)**.** *For each probability measure P in a collection $\mathcal{P}$, let $(S_n(P))_{n=1}^{\infty}$ be a stochastic process defined on the probability space $(\Omega, \mathcal{F}, P)$. Let $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P}(P))_{P\in\mathcal{P}}$ be a new collection of probability spaces containing stochastic processes $(\widetilde{S}_n(P))_{n=1}^{\infty}$ and $(G_n)_{n=1}^{\infty}$ so that $(\widetilde{S}_n(P))_{n=1}^{\infty}$ has the same distribution as $(S_n(P))_{n=1}^{\infty}$ for each $P \in \mathcal{P}$. We say that $(S_n(P))_{n=1}^{\infty}$ and $(G_n)_{n=1}^{\infty}$ are $(\mathcal{P}, n)$-uniformly coupled at a rate of r_n if for every $\varepsilon > 0$,*

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} \mathbb{P}_{\widetilde{P}(P)}\left(\exists k \geqslant m : \frac{|\widetilde{S}_k(P) - G_k|}{r_k} \geqslant \varepsilon\right) = 0, \tag{65}$$

and we write $S_n - G_n = \bar{o}_{\mathcal{P}}(r_n)$ as a shorthand for (65).

Since time-uniform convergence with high probability and almost-sure convergence — denoted by $o_{\text{a.s.}}(\cdot)$ and $\bar{o}_P(\cdot)$ respectively — are equivalent, observe that Definition 5.1 reduces to the standard notion of P-a.s. strong approximation when $\mathcal{P} = \{P\}$ is a singleton. To avoid repeating the technicalities of constructing a new probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P}(P))$ with equidistributed random variables and so on, some authors in the strong approximation literature refer to this procedure as "the construction" [14, 15] and they will say that "there exists a construction such that $S_n - G_n = o_{\text{a.s.}}(r_n)$" as a shorthand. We henceforth adopt and extend this convention to the $\mathcal{P}$-uniform setting by writing "there exists a construction such that $S_n - G_n = \bar{o}_{\mathcal{P}}(r_n)$". Let us now give a strong Gaussian approximation for partial sums of random variables with finite $(2+\delta)^{\text{th}}$ finite absolute moments.

Theorem 5.1 (Distribution-uniform strong Gaussian approximation)**.** *Let $(X_n)_{n=1}^{\infty}$ be independent and identically distributed random variables defined on the collection of probability spaces $(\Omega, \mathcal{F}, \mathcal{P})$ with means $\mu_P := \mathbb{E}_P(X)$ and variances $\sigma_P^2 := \mathbb{E}_P(X - \mu_P)^2$. If X has $q > 2$ uniformly upper-bounded moments, and a uniformly positive variance, i.e.*

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P|X - \mu_P|^q < \infty \quad \text{and} \quad \inf_{P\in\mathcal{P}} \sigma_P^2 > 0, \tag{66}$$

then there exists a construction with independent standard Gaussians $(Y_n)_{n=1}^{\infty} \sim N(0,1)$ so that

$$\left|\sum_{i=1}^{n} \frac{X_i - \mu_P}{\sigma_P} - \sum_{i=1}^{n} Y_i\right| = \bar{o}_{\mathcal{P}}(n^{1/q}\log^{2/q}(n)). \tag{67}$$

We remark that Theorem 5.1 is a purely probabilistic result that may be of interest outside of statistical inference altogether. To the best of our knowledge, Theorem 5.1 serves as the first *distribution-uniform strong Gaussian approximation* in the literature. Note that the rate (67) is optimal up to a factor of $\log^{2/q}(n)$ compared to the best rate possible in the P-pointwise setting and in fact $\log^{2/q}(n)$ can be replaced by any $f(n)^{1/q}$ as long as $\sum_{n=1}^{\infty}(nf(n))^{-1} < \infty$. As we alluded to before, improvements to this rate would *not* advance the statistical inference goals of this paper. The reason behind this is that strong approximation rates are often dominated by the rates of errors incurred from estimating nuisance functions such as the variance (which is often of order $\sqrt{\log\log n/n}$ or slower). Nevertheless, in future work we will explore rate-optimal analogues of Theorem 5.1 in a thorough study of distribution-uniform strong approximations but we keep the current version here because it is sufficient for the current paper's objectives.

In fact, the strong approximation of Theorem 5.1 is a corollary of the following more general *nonasymptotic* high-probability strong Gaussian coupling inequality for independent (but not necessarily identically distributed) random variables that depends on features of the distribution of X in transparent ways.

Lemma 5.2 (Strong Gaussian coupling inequality)**.** *Let $(X_n)_{n=1}^\infty$ be independent random variables on the probability space $(\Omega, \mathcal{F}, P)$. Suppose that for some $q \geqslant 2$, we have $\mathbb{E}_P|X_k - \mathbb{E}_P X_k|^q < \infty$ for each $k \in \mathbb{N}$. Let $f(\cdot)$ be a positive and increasing function so that $\sum_{n=1}^\infty (nf(n))^{-1} < \infty$ and*

$$\sum_{k=1}^{\infty} \frac{\mathbb{E}_P|X_k - \mathbb{E}_P X_k|^q / \sigma_k^q}{kf(k)} < \infty, \tag{68}$$

where $\sigma_k^2 := \mathrm{Var}_P(X_k)$. Then one can construct a probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P}(P))$ rich enough to define $(\widetilde{X}_n, Y_n)_{n=1}^\infty$ where $\widetilde{X}_n$ and X_n are equidistributed for each n and $(Y_n)_{n=1}^\infty$ are marginally independent standard Gaussians so that for any $\varepsilon > 0$,

$$\mathbb{P}_{\widetilde{P}(P)}\left(\exists k \geqslant m : \left|\frac{\sum_{i=1}^k (\widetilde{X}_i - Y_i)}{k^{1/q} f(k)^{1/q}}\right| > \varepsilon\right) \leqslant \frac{C_{q,f}}{\varepsilon^q}\left\{\sum_{k=2^{m-1}}^{\infty} \frac{\mathbb{E}_P|X_k - \mathbb{E}_P X_k|^q/\sigma_k^q}{kf(k)} + \frac{1}{2^m}\sum_{k=1}^{2^{m-1}-1} \frac{\mathbb{E}_P|X_k - \mathbb{E}_P X_k|^q/\sigma_k^q}{kf(k)}\right\}, \tag{69}$$

where $C_{q,f}$ is a constant that depends only on q and f.

Instantiating Lemma 5.2 in the identically distributed case with $q = 2 + \delta$ for some $\delta > 0$ and taking suprema over $P \in \mathcal{P}$ on both sides of (69) yields Theorem 5.1. The proofs of Lemma 5.2 and Theorem 5.1 can be found in Appendix A.7.

A straightforward consequence of Lemma 5.2 and Theorem 5.1 is that the law of the iterated logarithm holds uniformly in a class of distributions with uniformly bounded $(2+\delta)^{\text{th}}$ moments.

Corollary 5.3 (A $\mathcal{P}$-uniform law of the iterated logarithm)**.** *Suppose $(X_n)_{n=1}^\infty$ are defined on probability spaces $(\Omega, \mathcal{F}, \mathcal{P})$ where $\mathcal{P}$ is a collection of distributions such that*

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P|X - \mathbb{E}_P X|^{2+\delta} < \infty \quad \text{and} \quad \inf_{P\in\mathcal{P}} \mathrm{Var}_P(X) > 0 \tag{70}$$

for some $\delta > 0$. Then,

$$\sup_{k\geqslant n} \frac{|\sum_{i=1}^k (X_i - \mathbb{E}_P(X))|}{\sqrt{2\mathrm{Var}_P(X) k \log\log k}} = 1 + \bar{o}_{\mathcal{P}}(1). \tag{71}$$

A proof of Corollary 5.3 is provided in Appendix A.6 and follows from Theorem 5.1 combined with Kolmogorov's P-pointwise law of the iterated logarithm.

6 Summary & discussion

We gave a definition of "distribution-uniform anytime-valid inference" as a time-uniform analogue of distribution-uniform fixed-n inference and then derived explicit hypothesis tests, p-values, and confidence sequences satisfying that definition. Our methods relied on a novel boundary for centered partial sums that is uniformly valid in a class of distributions, in time, and in a family of boundaries. Along the way, we discussed what it meant for a sequence of random variables to converge distribution-uniformly almost-surely, and provided definitions for distribution- and time-uniform stochastic boundedness alongside a calculus for manipulating sequences with these types of asymptotics. At their core, all of our results relied on a novel strong Gaussian approximation that allows a partial sum process to

be tightly coupled with an implicit Gaussian process uniformly in time and in a class of distributions. We believe this is the first result of its kind in the literature. Zooming out, we believe that this strong Gaussian approximation forms the tip of the iceberg for distribution-uniform strong laws. In future work, we plan to study these problems in depth.

Acknowledgments. IW-S thanks Tudor Manole and Rajen Shah for insightful discussions. The authors acknowledge support from NSF grants IIS-2229881 and DMS-2310718.

References

[1] Susan Athey, Julie Tibshirani, and Stefan Wager. Generalized random forests. *The Annals of Statistics*, 47(2):1148–1178, 2019. 2

[2] Sivaraman Balakrishnan, Edward H Kennedy, and Larry Wasserman. The fundamental limits of structure-agnostic functional estimation. *arXiv preprint arXiv:2305.04116*, 2023. 12

[3] Anatole Beck and Daniel P Giesy. P-uniform convergence and a vector-valued strong law of large numbers. *Transactions of the American Mathematical Society*, 147(2):541–559, 1970. 8

[4] Aurélien Bibaut, Nathan Kallus, and Michael Lindon. Near-optimal non-parametric sequential tests and confidence sequences with possibly dependent observations. *arXiv preprint arXiv:2212.14411*, 2022. 3, 5, 41

[5] Leo Breiman. Stacked regressions. *Machine learning*, 24(1):49–64, 1996. 16

[6] Emmanuel Candes, Yingying Fan, Lucas Janson, and Jinchi Lv. Panning for gold:'model-x'knockoffs for high dimensional controlled variable selection. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 80(3):551–577, 2018. 11

[7] Tapas Kumar Chandra. de la Vallée Poussin's theorem, uniform integrability, tightness and moments. *Statistics & Probability Letters*, 107:136–141, 2015. 33

[8] Sourav Chatterjee. A new approach to strong embeddings. *Probability Theory and Related Fields*, 152(1-2):231–264, 2012. 3

[9] Kong-Ming Chong. On a theorem concerning uniform integrability. *Publ Inst Math (Beograd)(NS)*, 25(39):8–10, 1979. 10, 33

[10] Alexander Mangulad Christgau, Lasse Petersen, and Niels Richard Hansen. Nonparametric conditional local independence testing. *The Annals of Statistics*, 51(5):2116–2144, 2023. 9

[11] Kai Lai Chung. The strong law of large numbers. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, volume 2, pages 341–353. University of California Press, 1951. 8

[12] D. A. Darling and Herbert E. Robbins. Confidence sequences for mean, variance, and median. *Proceedings of the National Academy of Sciences of the United States of America*, 58 1:66–8, 1967. 2

[13] Boyan Duan, Aaditya Ramdas, and Larry Wasserman. Interactive rank testing by betting. In *Conference on Causal Learning and Reasoning*, pages 201–235. PMLR, 2022. 11

[14] Uwe Einmahl. Strong invariance principles for partial sums of independent random vectors. *Annals of Probability*, 15(4):1419–1440, 1987. 17

[15] Uwe Einmahl. Extensions of results of Komlós, Major, and Tusnády to the multivariate case. *Journal of multivariate analysis*, 28(1):20–68, 1989. 17

[16] Uwe Einmahl. A new strong invariance principle for sums of independent random vectors. *Journal of Mathematical Sciences*, 163(4):311–327, 2009. 16

[17] Peter Grünwald, Alexander Henzi, and Tyron Lardy. Anytime-valid tests of conditional independence under model-x. *Journal of the American Statistical Association*, pages 1–12, 2023. 11

[18] László Györfi, Michael Kohler, Adam Krzyzak, and Harro Walk. *A distribution-free theory of nonparametric regression*, volume 1. Springer, 2002. 11, 15

[19] Tien-Chung Hu and Andrew Rosalsky. A note on the de la Vallée Poussin criterion for uniform integrability. *Statistics & probability letters*, 81(1):169–174, 2011. 10, 33

[20] Maximilian Kasy. Uniformity and the delta method. *Journal of Econometric Methods*, 8(1): 20180001, 2018. 2

[21] Edward H Kennedy. Semiparametric doubly robust targeted double machine learning: a review. *arXiv preprint arXiv:2203.06469*, 2022. 12

[22] János Komlós, Péter Major, and Gábor Tusnády. An approximation of partial sums of independent rv's, and the sample df. i. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 32(1-2):111–131, 1975. 3, 16, 17, 42

[23] János Komlós, Péter Major, and Gábor Tusnády. An approximation of partial sums of independent rv's, and the sample df. ii. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 34(1):33–58, 1976. 3, 16, 17, 42

[24] Arun Kumar Kuchibhotla, Sivaraman Balakrishnan, and Larry Wasserman. Median regularity and honest inference. *Biometrika*, 110(3):831–838, 2023. 2

[25] Tze Leung Lai. On confidence sequences. *The Annals of Statistics*, 4(2):265–280, 1976. 2

[26] Ker-Chau Li. Honest confidence regions for nonparametric regression. *The Annals of Statistics*, 17(3):1001–1008, 1989. 2

[27] M Lifshits. Lecture notes on strong approximation. *Pub. IRMA Lille*, 53(13), 2000. 38

[28] Anton Rask Lundborg, Rajen D Shah, and Jonas Peters. Conditional independence testing in Hilbert spaces with applications to functional data analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(5):1821–1850, 2022. 2, 6

[29] Péter Major. The approximation of partial sums of independent rv's. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 35(3):213–220, 1976. 16, 17

[30] Aaditya Ramdas, Peter Grünwald, Vladimir Vovk, and Glenn Shafer. Game-theoretic statistics and safe anytime-valid inference. *Statistical Science*, 2023. 2

[31] Alessandro Rinaldo, Larry Wasserman, and Max G'Sell. Bootstrapping and sample splitting for high-dimensional, assumption-lean inference. *The Annals of Statistics*, 47(6):3438–3469, 2019. 2

[32] Herbert Robbins. Statistical methods related to the law of the iterated logarithm. *The Annals of Mathematical Statistics*, 41(5):1397–1409, 1970. 2

[33] Herbert Robbins and David Siegmund. Boundary crossing probabilities for the Wiener process and sample sums. *The Annals of Mathematical Statistics*, pages 1410–1429, 1970. 3, 5, 6, 41, 42

[34] Herbert Robbins and David Siegmund. The expected sample size of some tests of power one. *The Annals of Statistics*, 2(3):415–436, 1974. 13

[35] James Robins and Aad van der Vaart. Adaptive nonparametric confidence sets. *Annals of statistics*, 34(1):229–253, 2006. 2

[36] Aleksandr Ivanovich Sakhanenko. Estimates in an invariance principle. *Matematicheskie Trudy*, 5:27–44, 1985. 38

[37] Shalev Shaer, Gal Maman, and Yaniv Romano. Model-X sequential testing for conditional independence via testing by betting. In *International Conference on Artificial Intelligence and Statistics*, pages 2054–2086. PMLR, 2023. 11

[38] Rajen D Shah and Jonas Peters. The hardness of conditional independence testing and the generalised covariance measure. *The Annals of Statistics*, 48(3):1514–1538, 2020. 2, 4, 6, 11, 12, 13, 14, 15, 29, 33

[39] A.B. Skorokhod. *Research on the theory of random processes*. 1961. 16

[40] Volker Strassen. An invariance principle for the law of the iterated logarithm. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 3(3):211–226, 1964. 3, 16, 17

[41] Volker Strassen. Almost sure behavior of sums of independent random variables and martingales. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 3, page 315. University of California Press, 1967. 16, 17

[42] Ryan J Tibshirani, Alessandro Rinaldo, Rob Tibshirani, and Larry Wasserman. Uniform asymptotic inference and the bootstrap after model selection. *The Annals of Statistics*, 46(3):1255–1287, 2018. 2

[43] Alexandre B Tsybakov. Optimal rates of aggregation. In *Learning theory and kernel machines*, pages 303–313. Springer, 2003. 16

[44] Aad W van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000. 6, 22, 45

[45] Stefan Wager and Susan Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242, 2018. 2

[46] Abraham Wald. Sequential tests of statistical hypotheses. *The Annals of mathematical statistics*, 16(2):117–186, 1945. 2

[47] Ian Waudby-Smith, David Arbour, Ritwik Sinha, Edward H Kennedy, and Aaditya Ramdas. Time-uniform central limit theory and asymptotic confidence sequences. *arXiv preprint arXiv:2103.06476*, 2021. 3, 5, 8, 14, 41, 42

[48] Ian Waudby-Smith, Martin Larsson, and Aaditya Ramdas. Distribution-uniform strong laws of large numbers. *arXiv preprint arXiv:2402.00713*, 2024. 8, 10, 14, 15, 33, 34, 35, 36

A Proofs of the main results

In the proofs to come, we will make extensive use of the notions of convergence in Table 1, especially $\bar{o}_{\mathcal{P}}(\cdot)$ and $\bar{O}_{\mathcal{P}}(\cdot)$. However, some of our terms will be converging or asymptotically bounded with respect to different indices — e.g. there may be two sequences $(X_n)_{n=1}^{\infty}$ and $(Y_k)_{k=1}^{\infty}$ with indices n and k that are diverging to ∞ not necessarily together (e.g., imagine $k = n^2$). Writing $X_n = \bar{o}_{\mathcal{P}}(r_n)$ and $Y_k = \bar{o}_{\mathcal{P}}(r_k)$ is unambiguous, for example, but when no rate is specified, we will remove ambiguity with respect to indices n or k by saying $X_n = \bar{o}_{\mathcal{P}}^{(n)}(1)$ and $Y_k = \bar{o}_{\mathcal{P}}^{(k)}(1)$.

A.1 Proof of Proposition 2.2

Proposition 2.2 (($\mathcal{P}, n, x$)-uniform boundaries for centered partial sums)**.** *Let $X_1, X_2, \ldots$ be random variables defined on probability spaces $(\Omega, \mathcal{F}, \mathcal{P}^\star)$ with finite $(2+\delta)^{th}$ moments, i.e. $\mathbb{E}_P|X - \mathbb{E}_P(X)|^{2+\delta} < \infty$ for every $P \in \mathcal{P}^\star$. Letting $S_n := \sum_{i=1}^n (X_i - \mathbb{E}_P(X_i))/\sigma_P$ be their centered partial sums, we have*

$$\forall P \in \mathcal{P}^\star, \quad \lim_{m\to\infty} \sup_{x \geqslant 0} \left| \mathbb{P}_P \left(\exists k \geqslant m : |S_k| / \sqrt{k} \geqslant \sqrt{x + \log(k/m)} \right) - [1 - \Psi(x)] \right| = 0. \tag{17}$$

Furthermore, if $\mathcal{P} \subseteq \mathcal{P}^\star$ is a sub-collection of distributions for which the $(2+\delta)^{th}$ moment is $\mathcal{P}$-uniformly upper-bounded and the variance is $\mathcal{P}$-uniformly positive, then the above limit holds $\mathcal{P}$-uniformly:

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} \sup_{x \geqslant 0} \left| \mathbb{P}_P \left(\exists k \geqslant m : |S_k| / \sqrt{k} \geqslant \sqrt{x + \log(k/m)} \right) - [1 - \Psi(x)] \right| = 0. \tag{18}$$

Proof. Let $\underline{\sigma}^2 > 0$ be a uniform lower bound on $\inf_{P\in\mathcal{P}} \mathrm{Var}_P(X)$. Writing out $\sup_{k\geqslant m}\{S_k^2/\sigma_P^2 k - \log(k/m)\}$ and invoking the strong Gaussian coupling of Theorem 5.1, we have on a potentially enriched probability space a partial sum $G_n := \sum_{i=1}^n Y_i$ of standard Gaussians $Y_1, \ldots, Y_n \sim N(0,1)$ so that for some $q = 2 + \delta/2$ (say),

$$\begin{aligned}
\sup_{k\geqslant m}\{S_k^2/\sigma_P^2 k - \log(k/m)\} &= \sup_{k\geqslant m} \left\{ \left(\sigma_P G_k + \bar{o}_{\mathcal{P}}\left(k^{1/q}\right)\right)^2 / (\sigma_P^2 k) - \log(k/m) \right\} && (72)\\
&= \sup_{k\geqslant m} \left\{ \frac{\sigma_P^2 G_k^2 + \bar{O}_{\mathcal{P}}(k^{1/q}\sqrt{k \log\log k}) + \bar{o}_{\mathcal{P}}(k^{2/q})}{\sigma_P^2 k} - \log(k/m) \right\} && (73)\\
&= \sup_{k\geqslant m} \left\{ \frac{G_k^2}{k} + \frac{1}{\underline{\sigma}^2}\bar{O}_{\mathcal{P}}\left(\sqrt{\frac{\log\log k}{k^{1-2/q}}}\right) + \frac{1}{\underline{\sigma}^2}\bar{o}_{\mathcal{P}}\left(\frac{k^{2/q}}{k}\right) - \log(k/m) \right\} && (74)\\
&= \sup_{k\geqslant m} \left\{ \frac{G_k^2}{k} - \log(k/m) + \bar{o}_{\mathcal{P}}^{(k)}(1) \right\}, && (75)
\end{aligned}$$

where (73) expands the square and applies the ($\mathcal{P}$-uniform) law of the iterated logarithm (Corollary 5.3) to $(G_n)_{n=1}^{\infty}$, (74) uses the $\mathcal{P}$-uniform lower-boundedness of the variance, and (75) consolidates the $\bar{o}_{\mathcal{P}}(\cdot)$ terms. Now, notice that $\sup_{k\geqslant m}\{G_k^2/k - \log(k/m)\}$ converges *uniformly* to the Robbins-Siegmund distribution (Lemma B.2) since the distribution of the supremum does not depend on any measure P. That is,

$$\forall x \geqslant 0, \quad \lim_{m\to\infty} \sup_{P\in\mathcal{P}} \left| \mathbb{P}_P \left(\sup_{k\geqslant m} \left\{ \frac{G_k^2}{k} - \log(k/m) \right\} \leqslant x \right) - \Psi(x) \right| = 0. \tag{76}$$

Applying van der Vaart [44, Lemma 2.11] and using the fact that Ψ is continuous, we have that the above also holds uniformly in $x \geqslant 0$:

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} \sup_{x\geqslant 0} \left| \mathbb{P}_P \left(\sup_{k\geqslant m} \left\{ \frac{G_k^2}{k} - \log(k/m) \right\} \leqslant x \right) - \Psi(x) \right| = 0. \tag{77}$$

Some algebraic manipulations will reveal that the above is equivalent to the desired result:

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} \sup_{x\geqslant 0} \left| \mathbb{P}_P \left(\exists k \geqslant m : |S_k| / \sqrt{k} \geqslant \sqrt{x + \log(k/m)} \right) - [1 - \Psi(x)] \right| = 0, \tag{78}$$

which completes the proof. □

A.2 Proof of Lemma 3.1

Lemma 3.1 (Calculus of $\bar{O}_{\mathcal{P}}(\cdot)$ and $\bar{o}_{\mathcal{P}}(\cdot)$)**.** *Let $Y_1, Y_2, \ldots$ be random variables defined on $(\Omega, \mathcal{F}, \mathcal{P})$. Let $(a_n)_{n=1}^{\infty}$ and $(b_n)_{n=1}^{\infty}$ be positive and monotonically nonincreasing sequences. Then we have the following basic implications:*

$$Y_n = \bar{o}_{\mathcal{P}}(a_n) \implies Y_n = \bar{O}_{\mathcal{P}}(a_n) \tag{27}$$
$$Y_n = \bar{o}_{\mathcal{P}}(a_n)\bar{O}_{\mathcal{P}}(b_n) \implies Y_n = \bar{o}_{\mathcal{P}}(a_n b_n) \tag{28}$$
$$Y_n = \bar{O}_{\mathcal{P}}(a_n)\bar{O}_{\mathcal{P}}(b_n) \implies Y_n = \bar{O}_{\mathcal{P}}(a_n b_n) \tag{29}$$
$$Y_n = \bar{o}_{\mathcal{P}}(a_n) + \bar{O}_{\mathcal{P}}(a_n) \implies Y_n = \bar{O}_{\mathcal{P}}(a_n) \tag{30}$$
$$Y_n = \bar{o}_{\mathcal{P}}(a_n) + \bar{o}_{\mathcal{P}}(b_n) \implies Y_n = \bar{o}_{\mathcal{P}}(\max\{a_n, b_n\}). \tag{31}$$

Furthermore, (31) holds with $\bar{o}_{\mathcal{P}}(\cdot)$ replaced by $\bar{O}_{\mathcal{P}}(\cdot)$ on both sides. Finally, if $Y_n = \bar{O}_{\mathcal{P}}(a_n)$ and $a_n/b_n \to 0$, then $Y_n = \bar{o}_{\mathcal{P}}(b_n)$.

Proof of (27) Suppose that $Y_n = \bar{o}_{\mathcal{P}}(a_n)$. We want to show that for any δ, there exists $C \equiv C(\delta)$ and $M \equiv M(\delta)$ so that for all $m \geqslant M$,

$$\textbf{Goal:} \quad \sup_{P\in\mathcal{P}} \mathbb{P}_P \left(\sup_{k\geqslant m} |a_k^{-1} Y_k| \geqslant C \right) < \delta. \tag{79}$$

Proof. This is immediate from the definition of $\bar{o}_{\mathcal{P}}(a_n)$. Indeed, fix any $\varepsilon > 0$ and choose $M \equiv M(\varepsilon)$ so that for any $m \geqslant M$,

$$\sup_{P\in\mathcal{P}} \mathbb{P}_P \left(\sup_{k\geqslant m} |a_k^{-1} Y_k| \geqslant \varepsilon \right) < \delta. \tag{80}$$

Identifying C with ε completes the proof. □

Proof of (28). Suppose that $Y_n = A_n B_n$ with $A_n = \bar{o}_{\mathcal{P}}(a_n)$ and $B_n = \bar{o}_{\mathcal{P}}(b_n)$. We want to show that $a_n^{-1} b_n^{-1} Y_n = \bar{o}_{\mathcal{P}}(1)$. More formally, our goal is to show that for arbitrary $\varepsilon, \delta > 0$, there exists $M \equiv M(\varepsilon, \delta) \geqslant 1$ so that for all $m \geqslant M$,

$$\textbf{Goal:} \quad \sup_{P\in\mathcal{P}} \mathbb{P}_P \left(\exists k \geqslant m : |a_k^{-1} b_k^{-1} Y_k| \geqslant \varepsilon \right) < \delta. \tag{81}$$

Proof. Choose M sufficiently large so that for all $m \geqslant M$,

$$\sup_{P\in\mathcal{P}} \mathbb{P}_P \left(\sup_{k\geqslant m} |a_k^{-1} A_k| \geqslant \sqrt{\varepsilon/2} \right) < \delta \quad \text{and} \quad \sup_{P\in\mathcal{P}} \mathbb{P}_P \left(\sup_{k\geqslant m} |b_k^{-1} B_k| \geqslant \sqrt{\varepsilon/2} \right) < \delta. \tag{82}$$

Then, writing out the equation in (81), we have that

$$\sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\exists k \geqslant m : |a_k^{-1} b_k^{-1} Y_k| \geqslant \varepsilon\right) \tag{83}$$

$$\leqslant \sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\exists k \geqslant m : |a_k^{-1} A_k||b_k^{-1} B_k| \geqslant \varepsilon\right) \tag{84}$$

$$\leqslant \sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\exists k \geqslant m : |a_k^{-1} A_k||b_k^{-1} B_k| \geqslant \varepsilon \;\middle|\; \sup_{k\geqslant m} |a_k^{-1} A_k| < \sqrt{\varepsilon/2} \text{ and } \sup_{k\geqslant m} |b_k^{-1} B_k| < \sqrt{\varepsilon/2}\right) + \tag{85}$$

$$\sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\sup_{k\geqslant m} |a_k^{-1} A_k| < \sqrt{\varepsilon/2} \text{ and } \sup_{k\geqslant m} |b_k^{-1} B_k| < \sqrt{\varepsilon/2}\right) \tag{86}$$

$$\leqslant \underbrace{\sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\exists k \geqslant m : \varepsilon/2 \geqslant \varepsilon\right)}_{=0} + \tag{87}$$

$$\underbrace{\max\left\{\sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\sup_{k\geqslant m} |a_k^{-1} A_k| < \sqrt{\varepsilon/2}\right),\ \mathbb{P}\left(\sup_{k\geqslant m} |b_k^{-1} B_k| < \sqrt{\varepsilon/2}\right)\right\}}_{<\delta} \tag{88}$$

$$< \delta, \tag{89}$$

which completes the proof. □

Proof of (29). Suppose that $Y_n = A_n B_n$ with $A_n = \bar{O}_{\mathcal{P}}(a_n)$ and $B_n = \bar{O}_{\mathcal{P}}(b_n)$. Our goal is to show that for any $\delta > 0$, there exists some $C \equiv C(\delta)$ and $M \equiv M(C, \delta)$ so that

$$\textbf{Goal:} \quad \sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\sup_{k\geqslant m} |a_n^{-1} b_n^{-1} Y_n| > C\right) < \delta. \tag{90}$$

Proof. Fix $\delta > 0$. Let C_a, M_a, C_b, M_b be sufficiently large so that for all $m \geqslant \max\{M_a, M_b\}$,

$$\sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\sup_{k\geqslant m} |a_k^{-1} A_k| \geqslant M_a\right) < \delta \quad \text{and} \quad \sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\sup_{k\geqslant m} |b_k^{-1} B_k| \geqslant M_b\right) < \delta. \tag{91}$$

Now, set $C = C_a C_b + 1$. Then,

$$\sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\sup_{k\geqslant m} |a_k^{-1} b_k^{-1} Y_k| \geqslant C\right) \tag{92}$$

$$\leqslant \sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\sup_{k\geqslant m} |a_k^{-1} A_k||b_k^{-1} B_k| \geqslant C\right) \tag{93}$$

$$\leqslant \sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\sup_{k\geqslant m} C_a C_b \geqslant C\right) + \sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\sup_{k\geqslant m} |a_k^{-1} A_k| > C_a \text{ and } |b_k^{-1} B_k| > C_b\right) \tag{94}$$

$$\leqslant \underbrace{\sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\sup_{k\geqslant m} C_a C_b \geqslant C_a C_b + 1\right)}_{=0} + \tag{95}$$

$$\underbrace{\max\left\{\sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\sup_{k\geqslant m} |a_k^{-1} A_k| > C_a\right),\ \sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\sup_{k\geqslant m} |b_k^{-1} B_k| > C_b\right)\right\}}_{<\delta}, \tag{96}$$

which completes the proof. □

Proof of (30). Suppose $Y_n = A_n + A'_n$ with both $A_n = \bar{o}_{\mathcal{P}}(a_n)$ and $A'_n = \bar{O}_{\mathcal{P}}(a_n)$. The goal is to show that for every $\delta > 0$, there exists $C > 0$ and $M \geqslant 1$ so that for all $m \geqslant M$,

$$\textbf{Goal:} \quad \sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\sup_{k\geqslant m} a_k^{-1} |Y_k| > C\right) < \delta. \tag{97}$$

Proof. Fix $\delta > 0$. Let C' and M' be so that $\sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\sup_{k\geqslant m} a_k^{-1}|A'_k| > C'\right) < \delta/2$. Fix any $\varepsilon \in (0, C')$ and let $M^\star$ be so that $\sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\sup_{k\geqslant m} a_k^{-1}|A_k| \geqslant \varepsilon\right) < \delta/2$ for all $m \geqslant M^\star$. Choose $M > \max\{M', M^\star\}$. Then, for all $m \geqslant M$,

$$\begin{aligned}
&\sup_{P\in\mathcal{P}} \mathbb{P}\left(\sup_{k\geqslant m} a_k^{-1}|A_k + A'_k| \geqslant C'\right) && (98)\\
\leqslant\ &\sup_{P\in\mathcal{P}} \mathbb{P}\left(\sup_{k\geqslant m} a_k^{-1}|A_k| + a_k|A'_k| \geqslant C'\right) && (99)\\
\leqslant\ &\sup_{P\in\mathcal{P}} \mathbb{P}\left(\sup_{k\geqslant m} a_k^{-1}|A_k| \geqslant C'\right) + \sup_{P\in\mathcal{P}} \mathbb{P}\left(\sup_{k\geqslant m} a_k^{-1}|A'_k| \geqslant C'\right) && (100)\\
\leqslant\ &\sup_{P\in\mathcal{P}} \mathbb{P}\left(\sup_{k\geqslant m} a_k^{-1}|A_k| \geqslant \varepsilon\right) + \sup_{P\in\mathcal{P}} \mathbb{P}\left(\sup_{k\geqslant m} a_k^{-1}|A'_k| \geqslant C'\right) && (101)\\
<\ &\delta, && (102)
\end{aligned}$$

which completes the proof. □

Proof of (31). Suppose $Y_n = A_n + B_n$ with $A_n = \bar{o}_{\mathcal{P}}(a_n)$ and $B_n = \bar{o}_{\mathcal{P}}(b_n)$. The goal is to show that for every $\varepsilon, \delta > 0$, there exists $M \geqslant 1$ so that for all $m \geqslant M$,

$$\textbf{Goal:}\quad \sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\sup_{k\geqslant m} c_k^{-1}|Y_k| \geqslant \varepsilon\right) < \delta, \tag{103}$$

where $c_k = \max\{a_k, b_k\}$.

Proof. Fix $\varepsilon, \delta > 0$. Let M be so that $\sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\sup_{k\geqslant m} a_k|A_k| > \varepsilon\right) < \delta/2$ and $\sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\sup_{k\geqslant m} b_k|B_k| > \varepsilon\right) < \delta/2$ for all $m \geqslant M$. Then, for all $m \geqslant M$,

$$\begin{aligned}
&\sup_{P\in\mathcal{P}} \mathbb{P}\left(\sup_{k\geqslant m} c_k^{-1}|A_k + B_k| \geqslant \varepsilon\right) && (104)\\
\leqslant\ &\sup_{P\in\mathcal{P}} \mathbb{P}\left(\sup_{k\geqslant m} c_k^{-1}|A_k| + c_k^{-1}|B_k| \geqslant \varepsilon\right) && (105)\\
\leqslant\ &\sup_{P\in\mathcal{P}} \mathbb{P}\left(\sup_{k\geqslant m} a_k^{-1}|A_k| + b_k^{-1}|B_k| \geqslant \varepsilon\right) && (106)\\
\leqslant\ &\sup_{P\in\mathcal{P}} \mathbb{P}\left(\sup_{k\geqslant m} a_k^{-1}|A_k| \geqslant \varepsilon\right) + \sup_{P\in\mathcal{P}} \mathbb{P}\left(\sup_{k\geqslant m} b_k^{-1}|B_k| \geqslant \varepsilon\right) && (107)\\
<\ &\delta, && (108)
\end{aligned}$$

which completes the proof. □

Proof that if $Y_n = \bar{O}_{\mathcal{P}}(a_n)$ and $a_n/b_n \to 0$, then $Y_n = \bar{o}_{\mathcal{P}}(b_n)$. Let $\varepsilon, \delta > 0$. The goal is to show that there exists $M \equiv M(\varepsilon, \delta) \geqslant 1$ so that for all $m \geqslant M$,

$$\textbf{Goal:}\quad \sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\sup_{k\geqslant m} b_k^{-1}|Y_k| \geqslant \varepsilon\right) < \delta. \tag{109}$$

Proof. Let $C > 0$ and $M_1 \geqslant 1$ be constants so that $\sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\sup_{k\geqslant m} a_k^{-1}|Y_k| \geqslant C\right) < \delta$ for all $m \geqslant M_1$. Moreover, choose $M_2 \geqslant 1$ so that $a_k/b_k < \varepsilon/C$ for all $k \geqslant M_2$. Set $M := \max\{M_1, M_2\}$.

Then, for all $m \geqslant M$,

$$\begin{aligned}
&\sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\sup_{k\geqslant m} b_k^{-1}|Y_k| \geqslant \varepsilon\right) && (110)\\
=\ &\sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\exists k \geqslant m : b_k^{-1}|Y_k| \geqslant \varepsilon\right) && (111)\\
=\ &\sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\exists k \geqslant m : a_k^{-1}|Y_k| \geqslant (b_k/a_k)\varepsilon\right) && (112)\\
\leqslant\ &\sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\exists k \geqslant m : a_k^{-1}|Y_k| \geqslant (C/\not{\varepsilon})\cdot\not{\varepsilon}\right) && (113)\\
=\ &\sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\exists k \geqslant m : a_k^{-1}|Y_k| \geqslant C\right) && (114)\\
<\ &\delta, && (115)
\end{aligned}$$

which completes the proof. □

A.3 Proof of Theorem 3.3

Theorem 3.3 ($(\mathcal{P}, n, \alpha)$-uniform statistical inference)**.** *Let $X_1, X_2, \ldots$ be defined on $(\Omega, \mathcal{F}, \mathcal{P})$ and suppose that for some $\delta > 0$, the $(2+\delta)^{th}$ moment is $\mathcal{P}$-uniformly upper-bounded and the variance is $\mathcal{P}$-uniformly positive. Recall the definitions of $(\bar{p}_k^{(m)})_{k=m}^{\infty}$ and $(\bar{C}_k^{(m)}(\alpha))_{k=m}^{\infty}$ from Proposition 2.1:*

$$\bar{p}_k^{(m)} := 1 - \Psi\left(k\widehat{\mu}_k^2/\widehat{\sigma}_k^2 - \log(k/m)\right) \tag{34}$$

$$\text{and}\quad \bar{C}_k^{(m)}(\alpha) := \widehat{\mu}_k \pm \widehat{\sigma}_k\sqrt{[\Psi^{-1}(1-\alpha) + \log(k/m)]/k}. \tag{35}$$

Let $\mathcal{P}_0 \subseteq \mathcal{P}$ be a subcollection of distributions so that $\mathbb{E}_P(X) = 0$ for each $P \in \mathcal{P}_0$. Then the time-uniform type-I error of $(\bar{p}_k^{(m)})_{k=m}^{\infty}$ and the time-uniform miscoverage of $(\bar{C}_k^{(m)})_{k=m}^{\infty}$ converge to $\alpha \in (0,1)$ uniformly in α, meaning

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}_0} \sup_{\alpha\in(0,1)} \left|\mathbb{P}_P\left(\exists k \geqslant m : \bar{p}_k^{(m)} \leqslant \alpha\right) - \alpha\right| = 0, \quad \text{and} \tag{36}$$

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} \sup_{\alpha\in(0,1)} \left|\mathbb{P}_P\left(\exists k \geqslant m : \mathbb{E}_P(X) \notin \bar{C}_k^{(m)}(\alpha)\right) - \alpha\right| = 0. \tag{37}$$

Proof. Throughout, denote $S_n := \sum_{k=1}^n (X_k - \mathbb{E}_P(X))$. The proof is broken up into two steps. The first (and main) step of the proof shows that $\sup_{k\geqslant m}\{S_k^2/(\widehat{\sigma}_k^2 k) - \log(k/m)\}$ converges $(\mathcal{P}, x)$-uniformly to the Robbins-Siegmund distribution Ψ. The second step of the proof uses the first to show how such convergence is equivalent to $\bar{p}_k^{(m)}$ and $\bar{C}_k^{(m)}(\alpha)$ forming distribution-uniform anytime-valid p-values and confidence sequences, respectively, in the senses of Definition 2.1.

Step 1: Establishing the asymptotic distribution of $\sup_{k\geqslant m}\{S_k^2/(\widehat{\sigma}_k^2 k) - \log(k/m)\}$. First, notice that by Proposition 3.2,

$$|\widehat{\sigma}_n^2 - \sigma^2| = \bar{o}_{\mathcal{P}_0}(1/\log n). \tag{116}$$

Letting $\underline{\sigma}^2 > 0$ be the $\mathcal{P}$-uniform lower-bound on the variance so that $\inf_{P\in\mathcal{P}} \sigma_P^2 \geqslant \underline{\sigma}^2$, we therefore have

$$\begin{aligned}
\frac{1}{\widehat{\sigma}_n^2} &= \frac{1}{\sigma_P^2 + \bar{o}_{\mathcal{P}}(1/\log n)} && (117)\\
&= \frac{1}{\sigma_P^2(1 + \underline{\sigma}^{-2}\cdot\bar{o}_{\mathcal{P}}(1/\log n))} && (118)\\
&= \frac{1}{\sigma_P^2(1 + \bar{o}_{\mathcal{P}}(1/\log n))}. && (119)
\end{aligned}$$

Now, let γ_n be the $1 + \bar{o}_{\mathcal{P}}(1/\log n)$ term in the above denominator so that $\widehat{\sigma}_n^{-2} = \sigma_P^{-2}\gamma_n^{-1}$. Writing out $\sup_{k \geqslant m}\{S_k^2/(\widehat{\sigma}_k^2 k) - \log(k/m)\}$ and using the above, we then have

$$\sup_{k \geqslant m}\{S_k^2/\widehat{\sigma}_n^2 k - \log(k/m)\} = \sup_{k \geqslant m}\left\{\frac{S_k^2}{\sigma_P^2 \gamma_k k} - \log(k/m)\right\} \tag{120}$$

$$= \sup_{k \geqslant m}\left\{\left(\frac{S_k^2}{\sigma_P^2 k} - \gamma_k \log(k/m)\right)\frac{1}{\gamma_k}\right\} \tag{121}$$

$$= \sup_{k \geqslant m}\left\{\left(\frac{S_k^2}{\sigma_P^2 k} - \log(k/m) + \log(k/m)\cdot \bar{o}_{\mathcal{P}}(1/\log k)\right)\frac{1}{\gamma_k}\right\} \tag{122}$$

$$= \sup_{k \geqslant m}\left\{\left(\frac{S_k^2}{\sigma_P^2 k} - \log(k/m) + \bar{o}_{\mathcal{P}}^{(k)}(1)\right)\frac{1}{1 + \bar{o}_{\mathcal{P}}^{(k)}(1)}\right\}, \tag{123}$$

where (123) uses the fact that $k/m \leqslant k$ for any $m \geqslant 1$. We will now justify why the above converges $\mathcal{P}$- and quantile-uniformly to the Robbins-Siegmund distribution $\Psi(\cdot)$. First, by Lemma B.1, we have that $\sup_{k \geqslant m}\{S_k^2/\sigma^2 k - \log(k/m)\}$ converges $\mathcal{P}$- and quantile-uniformly in distribution to Ψ as $m \to \infty$. That is,

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} \sup_{x \geqslant 0}\left|\mathbb{P}_P\left(\sup_{k \geqslant m}\left\{\frac{S_k^2}{\sigma_P^2 k} - \log(k/m)\right\} \leqslant x\right) - \Psi(x)\right| = 0. \tag{124}$$

By the fact that $(\mathcal{P}, n, x)$-uniform convergence to Lipschitz CDFs is preserved under additive $\bar{o}_{\mathcal{P}}(1)$-perturbations (Lemma B.4) and the fact that $\Psi(\cdot)$ is Lipschitz (Lemma B.3), we have that

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} \sup_{x \geqslant 0}\left|\mathbb{P}_P\left(\sup_{k \geqslant m}\left\{\frac{S_k^2}{\sigma_P^2 k} - \log(k/m) + \bar{o}_{\mathcal{P}}^{(k)}(1)\right\} \leqslant x\right) - \Psi(x)\right| = 0. \tag{125}$$

Finally, using the fact that $(\mathcal{P}, n, x)$-uniform convergence in distribution is preserved under multiplicative $(1 + \bar{o}_{\mathcal{P}}(1))^{-1}$-perturbations (Lemma B.5), we have that

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} \sup_{x \geqslant 0}\left|\mathbb{P}_P\left(\sup_{k \geqslant m}\left\{\left(\frac{S_k^2}{\sigma_P^2 k} - \log(k/m) + \bar{o}_{\mathcal{P}}^{(k)}(1)\right)\frac{1}{1 + \bar{o}_{\mathcal{P}}^{(k)}(1)}\right\} \leqslant x\right) - \Psi(x)\right| = 0. \tag{126}$$

Step 2: Establishing validity of $\bar{p}_k^{(m)}$ and $\bar{C}_k^{(m)}(\alpha)$. Writing out the definition of $\bar{p}_k^{(m)}$, we have for any $P \in \mathcal{P}_0$ and $\alpha \in (0,1)$,

$$\mathbb{P}_P\left(\exists k \geqslant m : \bar{p}_k^{(m)} \leqslant \alpha\right) \tag{127}$$

$$= \mathbb{P}_P\left\{\exists k \geqslant m : 1 - \Psi\left(k\widehat{\mu}_k^2/\widehat{\sigma}_k^2 - \log(k/m)\right) \leqslant \alpha\right\} \tag{128}$$

$$= \mathbb{P}_P\left\{\exists k \geqslant m : \Psi\left(k\widehat{\mu}_k^2/\widehat{\sigma}_k^2 - \log(k/m)\right) \geqslant 1 - \alpha\right\} \tag{129}$$

$$= \mathbb{P}_P\left(\exists k \geqslant m : k\widehat{\mu}_k^2/\widehat{\sigma}_k^2 - \log(k/m) \geqslant \Psi^{-1}(1-\alpha)\right) \tag{130}$$

$$= \mathbb{P}_P\left(\sup_{k \geqslant m}\left\{k\widehat{\mu}_k^2/\widehat{\sigma}_k^2 - \log(k/m)\right\} \geqslant \Psi^{-1}(1-\alpha)\right). \tag{131}$$

Recalling that $x \mapsto \Psi(x)$ is a bijection between $\mathbb{R}^{\geqslant 0}$ and $[0,1)$ and invoking Step 1, we have the desired result:

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}_0} \sup_{\alpha\in(0,1)}\left|\mathbb{P}_P\left(\exists k \geqslant m : \bar{p}_k^{(m)} \leqslant \alpha\right) - \alpha\right| = 0, \tag{132}$$

completing the justification for $\bar{p}_k^{(m)}$. Moving on to $\bar{C}_k^{(m)}(\alpha)$, we have for any $P \in \mathcal{P}$ and any $\alpha \in (0,1)$ that

$$\mathbb{P}_P\left(\exists k \geqslant m : \mathbb{E}_P(X) \notin \bar{C}_k^{(m)}(\alpha)\right) \tag{133}$$

$$= \mathbb{P}_P\left(\exists k \geqslant m : \mathbb{E}_P(X) \notin \left(\widehat{\mu}_k \pm \widehat{\sigma}_k\sqrt{[\Psi^{-1}(1-\alpha) + \log(k/m)]/k}\right)\right) \tag{134}$$

$$= \mathbb{P}_P\left(\exists k \geqslant m : \left|\sum_{i=1}^{k}(X_i - \mathbb{E}_P(X))\right| \geqslant \widehat{\sigma}_k\sqrt{k[\Psi^{-1}(1-\alpha) + \log(k/m)]}\right) \tag{135}$$

$$= \mathbb{P}_P\left(\exists k \geqslant m : S_k^2/\widehat{\sigma}_k^2 k - \log(k/m) \geqslant \Psi^{-1}(1-\alpha)\right) \tag{136}$$

$$= \mathbb{P}_P\left(\sup_{k \geqslant m}\left\{S_k^2/\widehat{\sigma}_k^2 k - \log(k/m)\right\} \geqslant \Psi^{-1}(1-\alpha)\right), \tag{137}$$

and thus we have that

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} \sup_{\alpha\in(0,1)} \left|\mathbb{P}_P\left(\exists k \geqslant m : \mathbb{E}_P(X) \notin C_k^{(m)}(\alpha)\right) - \alpha\right| = 0 \tag{138}$$

via the same reasoning as was used for the anytime p-value. This completes the proof.

□

A.4 Proof of Proposition 4.1

Proposition 4.1 (Hardness of anytime-valid conditional independence testing)**.** *Suppose* $(X_n, Y_n, Z_n)_{n=1}^{\infty}$ *are* $[0,1]^3$*-valued triplets on the probability spaces* $(\Omega, \mathcal{F}, \mathcal{P}^\star)$ *where* $\mathcal{P}^\star$ *consists of all distributions supported on* $[0,1]^3$*. Let* $\mathcal{P}_0^\star \subseteq \mathcal{P}^\star$ *be the subset of distributions satisfying the conditional independence null* H_0 *and denote* $\mathcal{P}_1^\star := \mathcal{P}^\star \backslash \mathcal{P}_0^\star$*. Then for any potentially randomized test* $(\bar{\Gamma}_k^{(m)})_{k=m}^{\infty}$*,*

$$\sup_{P\in\mathcal{P}_1^\star} \limsup_{m\to\infty} \mathbb{P}_P\left(\exists k \geqslant m : \bar{\Gamma}_k^{(m)} = 1\right) \leqslant \limsup_{m\to\infty} \sup_{P\in\mathcal{P}_0^\star} \mathbb{P}_P\left(\exists k \geqslant m : \bar{\Gamma}_k^{(m)} = 1\right). \tag{51}$$

In other words, no $\mathcal{P}_0^\star$*-uniform anytime-valid test can have power against any alternative in* $\mathcal{P}_1^\star$ *at any* $\{m, m+1, \dots\}$*-valued stopping time no matter how large* m *is.*

Proof. Suppose for the sake of contradiction that there exists a potentially randomized test $(\bar{\Gamma}_k^{(m)})_{k=m}^{\infty}$ so that for some $\alpha \in (0,1)$, we have both

$$\limsup_{m\to\infty} \sup_{P\in\mathcal{P}_0^\star} \mathbb{P}_P\left(\exists k \geqslant m : \bar{\Gamma}_k^{(m)} = 1\right) \leqslant \alpha \tag{139}$$

and

$$\sup_{P\in\mathcal{P}_1^\star} \limsup_{m\to\infty} \mathbb{P}_P\left(\exists k \geqslant m : \bar{\Gamma}_k^{(m)} = 1\right) > \alpha. \tag{140}$$

Then there must exist $\varepsilon > 0$ so that we can always find m_1 arbitrarily large and nevertheless satisfy

$$\sup_{P\in\mathcal{P}_1^\star} \mathbb{P}_P\left(\exists k \geqslant m_1 : \bar{\Gamma}_k^{(m_1)} = 1\right) > \alpha + \varepsilon. \tag{141}$$

Furthermore, by (139), there exists $m_0 \geqslant 1$ large enough so that for all $m \geqslant m_0$,

$$\sup_{P\in\mathcal{P}_0^\star} \mathbb{P}_P\left(\exists k \geqslant m : \bar{\Gamma}_k^{(m)} = 1\right) < \alpha + \varepsilon. \tag{142}$$

In particular, choose some $m_1 \geqslant m_0$ so that (141) holds. Notice that the events

$$A_M := \{\bar{\Gamma}_k^{(m_1)} = 1 \text{ for some } m_1 \leqslant k \leqslant M\} \tag{143}$$

are nested for $M = m_1, m_1+1, \ldots$ and that $A_M \to A := \{\exists k \geqslant m_1 : \bar{\Gamma}_k^{(m_1)} = 1\}$ as $M \to \infty$. Consequently, there must exist some $M^\star$ such that

$$\sup_{P \in \mathcal{P}_1^\star} \mathbb{P}_P \left(\max_{m_1 \leqslant k \leqslant M^\star} \bar{\Gamma}_k^{(m_1)} = 1 \right) > \alpha + \varepsilon. \tag{144}$$

On the other hand, notice that by virtue of being a $\mathcal{P}_0^\star$-uniform anytime valid test and the fact that $m_1 \geqslant m_0$, we have that $\max_{m_1 \leqslant k \leqslant M^\star} \bar{\Gamma}_k^{(m_1)}$ uniformly controls the type-I error under $\mathcal{P}_0^\star$, i.e.

$$\sup_{P \in \mathcal{P}_0^\star} \mathbb{P}_P \left(\max_{m_1 \leqslant k \leqslant M^\star} \bar{\Gamma}_k^{(m_1)} = 1 \right) \leqslant \sup_{P \in \mathcal{P}_0^\star} \mathbb{P}_P \left(\exists k \geqslant m_1 : \bar{\Gamma}_k^{(m_1)} = 1 \right) < \alpha + \varepsilon. \tag{145}$$

Combining the above with the hardness result of Shah and Peters [38, Theorem 2] applied to the test $\max_{m_1 \leqslant k \leqslant M^\star} \bar{\Gamma}_k^{(m_1)}$, we have that

$$\sup_{P \in \mathcal{P}_1^\star} \mathbb{P}_P \left(\max_{m_1 \leqslant k \leqslant M^\star} \bar{\Gamma}_k^{(m_1)} = 1 \right) < \alpha + \varepsilon, \tag{146}$$

contradicting (144), and thus completing the proof of Proposition 4.1. □

A.5 Proof of Theorem 4.2

Theorem 4.2 ($\mathcal{P}_0$-uniform type-I error control of the SeqGCM)**.** *Suppose $(X_i, Y_i, Z_i)_{i=1}^\infty$ are $\mathbb{R} \times \mathbb{R} \times \mathbb{R}^d$-valued triplets defined on the probability spaces $(\Omega, \mathcal{F}, \mathcal{P})$ and let $\mathcal{P}_0 \subseteq \mathcal{P}$ be a collection of distributions in $\mathcal{P}$ satisfying the conditional independence null H_0 and Assumption SeqGCM-1, SeqGCM-2, and GCM-3. Define*

$$\bar{p}_{k,m}^{\mathrm{GCM}} := 1 - \Psi\left(k(\overline{\mathrm{GCM}}_k)^2 - \log(k/m)\right). \tag{58}$$

Then $(\bar{p}_{k,m}^{\mathrm{GCM}})_{k=m}^\infty$ forms a $\mathcal{P}_0$-uniform anytime p-value for the conditional independence null:

$$\lim_{m \to \infty} \sup_{P \in \mathcal{P}_0} \sup_{\alpha \in (0,1)} \left| \mathbb{P}_P \left(\exists k \geqslant m : \bar{p}_{k,m}^{\mathrm{GCM}} \leqslant \alpha \right) - \alpha \right| = 0. \tag{59}$$

Before proceeding with the proof, notice that the estimated residual R_i can be written as

$$R_i = \xi_i + b_i + \nu_i \tag{147}$$

where $\xi_i := \xi_i^x \cdot \xi_i^y$ is a true product residual with

$$\xi_i^x := \{X_i - \mu^x(Z_i)\} \quad \text{and} \quad \xi_i^y := \{Y_i - \mu^y(Z_i)\}, \tag{148}$$

b_i is a product regression error term given by

$$b_i := \{\widehat{\mu}_i^x(Z_i) - \mu^x(Z_i)\} \{\widehat{\mu}_i^y(Z_i) - \mu^y(Z_i)\}, \tag{149}$$

and $\nu_i := \nu_i^{x,y} + \nu_i^{y,x}$ is a cross-term where

$$\nu_i^{x,y} := \{\widehat{\mu}_i^x(Z_i) - \mu^x(Z_i)\} \xi_i^y, \quad \text{and} \tag{150}$$

$$\nu_i^{y,x} := \{\widehat{\mu}_i^y(Z_i) - \mu^y(Z_i)\} \xi_i^x. \tag{151}$$

Furthermore, define their averages as $\bar{b}_n := \frac{1}{n} \sum_{i=1}^n b_i$ and similarly for $\bar{\nu}_n^{x,y}$, $\bar{\nu}_n^{y,x}$, and $\bar{\xi}_n$. We may at times omit the argument (Z_i) from $\widehat{\mu}_i^x(Z_i) \equiv \widehat{\mu}_i^x$ or $\mu^x(Z_i) \equiv \mu^x$ etc. when it is clear from context. With these shorthands in mind, we are ready to prove Theorem 4.2.

Proof of Theorem 4.2. Note that by some simple algebraic manipulations, it suffices to show that $\sup_{k \geqslant m} \left\{ k\overline{\mathrm{GCM}}_k^2 - \log(k/m) \right\}$ converges $\mathcal{P}_0$-uniformly to the Robbins-Siegmund distribution as $m \to \infty$. Begin by writing $\overline{\mathrm{GCM}}_n$ as

$$\overline{\mathrm{GCM}}_n := \frac{1}{n\widehat{\sigma}_n^2} \sum_{i=1}^{n} R_i \tag{152}$$

$$\equiv \widehat{\sigma}_n^{-1} \left(\bar{\xi}_n + \bar{\nu}_n + \bar{b}_n \right) \tag{153}$$

and through a direct calculation, notice that our squared GCM statistic can be written as

$$\overline{\mathrm{GCM}}_n^2 = \frac{\bar{\xi}_n^2 + 2\bar{\xi}_n(\bar{\nu}_n + \bar{b}_n) + (\bar{\nu}_n + \bar{b}_n)^2}{\widehat{\sigma}_n^2} \tag{154}$$

$$= \underbrace{\frac{\bar{\xi}_n^2}{\widehat{\sigma}_n^2}}_{(i)} + \underbrace{\frac{2\bar{\xi}_n(\bar{\nu}_n + \bar{b}_n)}{\widehat{\sigma}_n^2}}_{(ii)} + \underbrace{\frac{(\bar{\nu}_n + \bar{b}_n)^2}{\widehat{\sigma}_n^2}}_{(iii)}. \tag{155}$$

In the discussion to follow, we analyze these three terms separately (in Steps 1, 2, and 3, respectively) and combine them to yield the desired result in Step 4.

Step 1: Analyzing (i)**.** In Lemma A.1, we show that under the assumptions of Theorem 4.2, the estimator $\widehat{\sigma}_n^2 := \frac{1}{n}\sum_{i=1}^n R_i^2$ is $\mathcal{P}_0$-uniformly consistent for $\mathrm{Var}(\xi) \equiv \mathbb{E}(\xi^2)$ at a rate faster than $1/\log n$, meaning

$$\widehat{\sigma}_n^2 - \mathrm{Var}(\xi) = \bar{o}_{\mathcal{P}_0}(1/\log n). \tag{156}$$

Invoking Assumption GCM-3, let $\underline{\sigma}^2$ be a uniform lower bound on the variance. Then, for any $P \in \mathcal{P}_0$,

$$(i) \equiv \frac{\bar{\xi}_n^2}{\widehat{\sigma}_n^2} \tag{157}$$

$$= \frac{\bar{\xi}_n^2}{\sigma_P^2 + \bar{o}_{\mathcal{P}_0}(1/\log n)} \tag{158}$$

$$= \frac{\bar{\xi}_n^2}{\sigma_P^2 \cdot (1 + \sigma_P^{-2} \cdot \bar{o}_{\mathcal{P}_0}(1/\log n))} \tag{159}$$

$$= \frac{\bar{\xi}_n^2}{\sigma_P^2 \cdot (1 + \underline{\sigma}^{-2} \cdot \bar{o}_{\mathcal{P}_0}(1/\log n))} \tag{160}$$

$$= \frac{\bar{\xi}_n^2}{\sigma_P^2 \cdot (1 + \bar{o}_{\mathcal{P}_0}(1/\log n))}. \tag{161}$$

The final form of (i) above will be used later in Step 4 of the proof.

Step 2: Analyzing (ii)**.** In Lemmas A.2 and A.3, we show that under the assumptions of Theorem 4.2, $\bar{b}_n = \bar{o}_{\mathcal{P}_0}(1/\sqrt{n \log\log n})$ and $\bar{\nu}_n = \bar{o}_{\mathcal{P}_0}(1/\sqrt{n \log\log n})$, respectively. Recall that $\widehat{\sigma}_n^2 - \mathbb{E}(\xi^2) = \bar{o}_{\mathcal{P}_0}(1/\log n)$ by Lemma A.1. Furthermore, we have by the uniform law of the iterated logarithm in Corollary 5.3 that $\bar{\xi}_n = \bar{O}_{\mathcal{P}_0}(\sqrt{\log\log n/n})$. Combining these four convergence results together with the calculus outlined in Lemma 3.1, we have

$$(ii) = \frac{2\bar{\xi}_n \cdot (\bar{\nu}_n + \bar{b}_n)}{\widehat{\sigma}_n^2} \tag{162}$$

$$= \frac{\bar{O}_{\mathcal{P}_0}(\sqrt{\log\log n/n}) \cdot \bar{o}_{\mathcal{P}_0}(1/\sqrt{n \log\log n})}{\underline{\sigma}^2 \cdot (1 + \bar{o}_{\mathcal{P}_0}(1))} \tag{163}$$

$$= \bar{o}_{\mathcal{P}_0}(1/n). \tag{164}$$

Step 3: Analyzing (iii)**.** Again by Lemmas A.2 and A.3 and the calculus of Lemma 3.1, we have that

$$(iii) \leqslant \left| \frac{(\bar{\nu}_n + \bar{b}_n)^2}{\widehat{\sigma}_n^2} \right| \tag{165}$$

$$= \left| \frac{\bar{o}_{\mathcal{P}_0}(1/n)}{\underline{\sigma}^2 \cdot (1 + \bar{o}_{\mathcal{P}_0}(1))} \right| \tag{166}$$

$$= \bar{o}_{\mathcal{P}_0}(1/n). \tag{167}$$

Step 4: Putting (i)**–**(iii) **together.** Writing out $\overline{\mathrm{GCM}}_n^2$ and noting the forms of (i), (ii), and (iii) displayed above, we have that for any $P \in \mathcal{P}_0$,

$$\overline{\mathrm{GCM}}_n^2 = \underbrace{\frac{\bar{\xi}_n^2}{\widehat{\sigma}_n^2}}_{(i)} + \underbrace{\frac{2\bar{\xi}_n(\bar{\nu}_n + \bar{b}_n)}{\widehat{\sigma}_n^2}}_{(ii)} + \underbrace{\frac{(\bar{\nu}_n + \bar{b}_n)^2}{\widehat{\sigma}_n^2}}_{(iii)} \tag{168}$$

$$= \frac{\bar{\xi}_n^2/\sigma_P^2}{1 + \bar{o}_{\mathcal{P}_0}(1/\log n)} + \bar{o}_{\mathcal{P}_0}(1/n). \tag{169}$$

Similar to the proof of Theorem 3.3, let γ_n be the $1 + \bar{o}_{\mathcal{P}_0}(1/\log n)$ denominator of the first term above. Then for any $P \in \mathcal{P}_0$ and any $x \geqslant 0$,

$$\mathbb{P}_P\left(\sup_{k \geqslant m} \left\{ k\overline{\mathrm{GCM}}_k^2 - \log(k/m) \right\} \leqslant x \right) \tag{170}$$

$$= \mathbb{P}_P\left(\sup_{k \geqslant m} \left\{ k\left(\frac{\bar{\xi}_k^2/\sigma_P^2}{\gamma_k} + \bar{o}_{\mathcal{P}_0}^{(k)}(1) \right) - \log(k/m) \right\} \leqslant x \right) \tag{171}$$

$$= \mathbb{P}_P\left(\sup_{k \geqslant m} \left\{ \frac{k}{\gamma_k}\left(\bar{\xi}_k^2/\sigma_P^2 + \gamma_k \cdot \bar{o}_{\mathcal{P}_0}^{(k)}(1) - \gamma_k \log(k/m) \right) \right\} \leqslant x \right) \tag{172}$$

$$= \mathbb{P}_P\left(\sup_{k \geqslant m} \left\{ \frac{k}{1 + \bar{o}_{\mathcal{P}_0}^{(k)}(1)}\left(\bar{\xi}_k^2/\sigma_P^2 + \bar{o}_{\mathcal{P}_0}^{(k)}(1) - \log(k/m) \right) \right\} \leqslant x \right), \tag{173}$$

and hence similar to the proof of Theorem 3.3, we apply Proposition 2.1, Lemma B.4, and Lemma B.5 in succession to arrive at the desired result:

$$\lim_{m \to \infty} \sup_{P \in \mathcal{P}_0} \sup_{x \geqslant 0} \left| \mathbb{P}_P\left(\exists k \geqslant m : k\overline{\mathrm{GCM}}_k^2 - \log(k/m) \geqslant x \right) - [1 - \Psi(x)] \right| = 0, \tag{174}$$

which completes the proof of Theorem 4.2. □

Lemma A.1 ($\mathcal{P}_0$-uniformly strongly consistent variance estimation)**.** *Let* $\widehat{\sigma}_n^2$ *be the sample variance of* R_i*:*

$$\widehat{\sigma}_n^2 := \frac{1}{n}\sum_{i=1}^n R_i^2 - \left(\frac{1}{n}\sum_{i=1}^n R_i \right)^2. \tag{175}$$

Then,

$$\widehat{\sigma}_n^2 - \mathbb{E}(\xi^2) = \bar{o}_{\mathcal{P}_0}\left(\frac{1}{\log n} \right). \tag{176}$$

Proof of Lemma A.1. First, consider the following decomposition:

$$R_i^2 = \left[\xi_i^x \xi_i^y + \xi_i^x \left\{\mu^y - \widehat{\mu}_i^y\right\} + \xi_i^y \left\{\mu^x - \widehat{\mu}_i^x\right\} + \left(\widehat{\mu}_i^x - \mu^x\right)\left(\widehat{\mu}_i^y - \mu^y\right)\right]^2 \tag{177}$$

$$= \xi_i^2 + \tag{178}$$

$$\underbrace{2(\xi_i^x)^2 \xi_i^y \left\{\mu^y - \widehat{\mu}_i^y\right\} + 2(\xi_i^y)^2 \xi_i^x \left\{\widehat{\mu}_i^x - \mu^x\right\}}_{\mathrm{I}_i} + \tag{179}$$

$$\underbrace{4\xi_i \left\{\mu^x - \widehat{\mu}_i^x\right\}\left\{\mu^y - \widehat{\mu}_i^y\right\}}_{\mathrm{II}_i} + \tag{180}$$

$$\underbrace{2\xi_i^x \left\{\mu^y - \widehat{\mu}_i^y\right\}^2 \left\{\mu^x - \widehat{\mu}_i^x\right\} + 2\xi_i^y \left\{\mu^x - \widehat{\mu}_i^x\right\}^2 \left\{\mu^y - \widehat{\mu}_i^y\right\}}_{\mathrm{III}_i} + \tag{181}$$

$$\underbrace{\left\{\mu^x - \widehat{\mu}_i^x\right\}^2 \left\{\mu^y - \widehat{\mu}_i^y\right\}^2}_{\mathrm{IV}_i}. \tag{182}$$

Letting $\bar{\mathrm{I}}_n := \frac{1}{n}\sum_{i=1}^n \mathrm{I}_i$ and similarly for $\bar{\mathrm{II}}_n$, $\bar{\mathrm{III}}_n$, and $\bar{\mathrm{IV}}_n$, we have that

$$\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n \xi_i^2 + \bar{\mathrm{I}}_n + \bar{\mathrm{II}}_n + \bar{\mathrm{III}}_n + \bar{\mathrm{IV}}_n - (\bar{R}_n)^2 \tag{183}$$

and we will separately show that $\bar{\mathrm{I}}_n$, $\bar{\mathrm{II}}_n$, $\bar{\mathrm{III}}_n$, $\bar{\mathrm{IV}}_n$, and $(\bar{R}_n)^2$ are all $\bar{o}_{\mathcal{P}_0}(1/\log n)$.

Step 1: Convergence of $\bar{\mathrm{I}}_n$. By the Cauchy-Schwarz inequality, we have that

$$\frac{1}{n}\sum_{i=1}^n (\xi_i^x)^2 \xi_i^y \left\{\mu^y - \widehat{\mu}_i^y\right\} \leqslant \underbrace{\sqrt{\frac{1}{n}\sum_{i=1}^n (\xi_i^x \xi_i^y)^2}}_{(\star)} \cdot \underbrace{\sqrt{\frac{1}{n}\sum_{i=1}^n (\xi_i^x)^2 \left\{\mu^y - \widehat{\mu}_i^y\right\}^2}}_{(\dagger)}. \tag{184}$$

Now, writing $\xi_i := \xi_i^x \xi_i^y$, notice that

$$(\star) \equiv \frac{1}{n}\sum_{i=1}^n \xi_i^2 \tag{185}$$

$$\leqslant \left(\frac{1}{n}\sum_{i=1}^n \xi_i^2 - \mathbb{E}(\xi_i^2)\right) + \mathbb{E}(\xi_i^2) \tag{186}$$

$$= \bar{o}_{\mathcal{P}_0}(1) + \mathbb{E}\left[\left(|\xi_i|^{2+\delta}\right)^{\frac{2}{2+\delta}}\right] \tag{187}$$

$$\leqslant \bar{o}_{\mathcal{P}_0}(1) + \left(\mathbb{E}|\xi_i|^{2+\delta}\right)^{\frac{2}{2+\delta}} \tag{188}$$

$$\leqslant \bar{O}_{\mathcal{P}_0}(1), \tag{189}$$

where the last line follows from Assumption GCM-3. Moreover, by Lemma A.4, we have that $(\dagger) = \bar{o}_{\mathcal{P}_0}(1/\log n)$, and hence by Lemma 3.1, $\bar{\mathrm{I}}_n \leqslant (\star)\cdot(\dagger) = \bar{o}_{\mathcal{P}_0}(1/\log n)$.

Step 2: Convergence of $\bar{\mathrm{II}}_n$. Again by Cauchy-Schwarz, we have

$$\frac{1}{n}\sum_{i=1}^n \xi_i^x \xi_i^y \{\mu^x - \widehat{\mu}_i^x\}\{\mu^y - \widehat{\mu}_i^y\} \leqslant \sqrt{\frac{1}{n}\sum_{i=1}^n (\xi_i^x)^2 \{\mu^y - \widehat{\mu}_i^y\}^2} \cdot \sqrt{\frac{1}{n}\sum_{i=1}^n (\xi_i^y)^2 \{\mu^x - \widehat{\mu}_i^x\}^2}, \tag{190}$$

and hence again by Lemma A.4, we have $\bar{\mathrm{II}}_n = \bar{o}_{\mathcal{P}_0}(1/\log n)$.

Step 3: Convergence of $\overline{\mathrm{III}}_n$. Following Shah and Peters [38, Section D.1] and using the inequality $2|ab| \leqslant a^2 + b^2$ for any $a, b \in \mathbb{R}$, we have

$$\frac{2}{n}\sum_{i=1}^{n} \xi_i^x \{\mu^y - \widehat{\mu}_i^y\}^2 \{\mu^x - \widehat{\mu}_i^x\} \tag{191}$$

$$\leqslant \frac{1}{n}\sum_{i=1}^{n} (\xi_i^x)^2 \{\mu^y - \widehat{\mu}_i^y\}^2 + \frac{1}{n}\sum_{i=1}^{n} \{\mu^y - \widehat{\mu}_i^y\}^2 \{\mu^x - \widehat{\mu}_i^x\}^2, \tag{192}$$

and hence by Lemmas A.4 and A.2, we have $\overline{\mathrm{III}}_n = \bar{o}_{\mathcal{P}_0}(1/\log n)$.

Step 4: Convergence of $\overline{\mathrm{IV}}_n$. First, notice that

$$\overline{\mathrm{IV}}_n := \frac{1}{n}\sum_{i=1}^{n} \{\mu^x - \widehat{\mu}_i^x\}^2 \cdot \{\mu^y - \widehat{\mu}_i^y\}^2 \tag{193}$$

$$\leqslant n \cdot \frac{1}{n}\sum_{i=1}^{n} \{\mu^x - \widehat{\mu}_i^x\}^2 \cdot \frac{1}{n}\sum_{i=1}^{n} \{\mu^y - \widehat{\mu}_i^y\}^2 . \tag{194}$$

Applying Lemmas A.2 and 3.1, we have that $\overline{\mathrm{IV}}_n = \bar{o}_{\mathcal{P}_0}(1/\log n)$.

Step 5: Convergence of $(\bar{R}_n)^2$ to 0. We will show that $(\bar{R}_n)^2 = \bar{o}_{\mathcal{P}_0}(1/\log n)$. Using the decomposition in (147) at the outset of the proof of Theorem 4.2, we have that

$$\bar{R}_n := \bar{\xi}_n + \bar{b}_n + \bar{\nu}_n. \tag{195}$$

Therefore, we can write its square as

$$(\bar{R}_n)^2 = (\bar{\xi}_n)^2 + 2\bar{\xi}_n \cdot \left(\bar{b}_n + \bar{\nu}_n\right) + (\bar{b}_n + \bar{\nu}_n)^2. \tag{196}$$

By Assumption GCM-3, we have that there exists a $\delta > 0$ so that $\sup_{P \in \mathcal{P}_0} \mathbb{E}_P |\xi|^{2+\delta} < \infty$. By the de la Vallée-Poussin criterion for uniform integrability [9, 19, 7], we have that the $(1+\delta)^{\text{th}}$ moment of ξ is uniformly integrable:

$$\lim_{m \to \infty} \sup_{P \in \mathcal{P}_0} \mathbb{E}_P \left(|\xi|^{1+\delta} \mathbb{1}\{|\xi|^{1+\delta} \geqslant m\}\right) = 0. \tag{197}$$

By Waudby-Smith et al. [48, Theorem 1], we have that $\bar{\xi}_n = \bar{o}_{\mathcal{P}_0}\left(n^{1/(1+\delta)-1}\right)$, and in particular,

$$\bar{\xi}_n = \bar{o}_{\mathcal{P}_0}\left(1/\sqrt{\log n}\right). \tag{198}$$

Using Lemma 3.1, we observe that

$$(\bar{\xi}_n)^2 = \bar{o}_{\mathcal{P}_0}\left(1/\log n\right), \tag{199}$$

and hence it now suffices to show that $\bar{b}_n + \bar{\nu}_n = \bar{o}_{\mathcal{P}_0}(1/\log n)$. Indeed, by Lemmas A.2 and A.3, we have that $\bar{b}_n = \bar{o}_{\mathcal{P}_0}(1/\sqrt{n \log\log n})$ and $\bar{\nu}_n = \bar{o}_{\mathcal{P}_0}(1/\sqrt{n \log\log n})$, respectively. Putting these together, we have

$$(\bar{R}_n)^2 = (\bar{\xi}_n)^2 + 2\bar{\xi}_n \cdot \left(\bar{b}_n + \bar{\nu}_n\right) + (\bar{b}_n + \bar{\nu}_n)^2 = \bar{o}_{\mathcal{P}_0}(1/\log n), \tag{200}$$

completing the argument for Step 5.

Step 6: Convergence of $\widehat{\sigma}_n^2$ to $\mathbb{E}(\xi^2)$. Putting Steps 1–5 together, notice that

$$\widehat{\sigma}_n^2 - \mathbb{E}(\xi^2) = \frac{1}{n}\sum_{i=1}^n \xi_i^2 - \mathbb{E}(\xi^2) + \bar{\mathrm{I}}_n + \bar{\mathrm{II}}_n + \bar{\mathrm{III}}_n + \bar{\mathrm{IV}}_n - (\bar{R}_n)^2 \tag{201}$$

$$= \frac{1}{n}\sum_{i=1}^n \xi_i^2 - \mathbb{E}(\xi^2) + \bar{o}_{\mathcal{P}_0}(1/\log n). \tag{202}$$

Now, since $\sup_{P\in\mathcal{P}} \mathbb{E}_P|\xi^2|^{1+\delta/2} < \infty$, ξ^2 we have by the de la Vallée criterion for uniform integrability that ξ^2 has a $\mathcal{P}_0$-uniformly integrable $(1+\delta/4)^{\text{th}}$ moment meaning that

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}_0} \mathbb{E}_P\left[(\xi^2)^{1+\delta/4}\mathbb{1}\left\{(\xi^2)^{1+\delta/4} > m\right\}\right] = 0, \tag{203}$$

and hence by Waudby-Smith et al. [48, Theorem 1(a)], we have that

$$\frac{1}{n}\sum_{i=1}^n \xi_i^2 - \mathbb{E}(\xi^2) = \bar{o}_{\mathcal{P}}\left(n^{1/(1+\delta/4)-1}\right), \tag{204}$$

and in particular, $\frac{1}{n}\sum_{i=1}^n \xi_i^2 - \mathbb{E}(\xi^2) = \bar{o}_{\mathcal{P}}(1/\log n)$, so that $\widehat{\sigma}_n^2 - \mathbb{E}(\xi^2) = \bar{o}_{\mathcal{P}}(1/\log n)$, completing the proof. □

Lemma A.2 (Convergence of the average bias term)**.** *Under Assumption SeqGCM-1, we have that*

$$\bar{b}_n \equiv \frac{1}{n}\sum_{i=1}^n b_i = \bar{o}_{\mathcal{P}}\left(1/\sqrt{n\log\log n}\right). \tag{205}$$

Proof. Under Assumption SeqGCM-1, we have that

$$\sup_{P\in\mathcal{P}_0} \|\widehat{\mu}_n^x - \mu^x\|_{L_2(P)} \cdot \|\widehat{\mu}_n^y - \mu^y\|_{L_2(P)} = O\left(\frac{1}{\sqrt{n\log^{2+\delta}(n)}}\right), \tag{206}$$

and hence let $C_{\mathcal{P}_0} > 0$ be a constant depending only on $\mathcal{P}_0$ so that

$$\sup_{P\in\mathcal{P}_0} \|\widehat{\mu}_n^x - \mu^x\|_{L_2(P)} \cdot \|\widehat{\mu}_n^y - \mu^y\|_{L_2(P)} \leqslant \frac{C_{\mathcal{P}_0}}{\sqrt{(n+1)\log^{2+\delta/2}(n+1)\log\log(n+1)}} \tag{207}$$

for all n sufficiently large. Consider the following series for all $k \geqslant m$ for any $m \geqslant 3$

$$\sup_{P\in\mathcal{P}_0} \sum_{k=m}^{\infty} \frac{\mathbb{E}_P\left|\left\{\widehat{\mu}_{k-1}^x(Z_k) - \mu^x(Z_k)\right\}\cdot\left\{\widehat{\mu}_{k-1}^y(Z_k) - \mu^y(Z_k)\right\}\right|}{\sqrt{k/\log\log k}} \tag{208}$$

$$\leqslant \sup_{P\in\mathcal{P}_0} \sum_{k=m}^{\infty} \frac{\left\|\widehat{\mu}_{k-1}^x - \mu^x\right\|_{L_2(P)} \cdot \left\|\widehat{\mu}_{k-1}^y - \mu^y\right\|_{L_2(P)}}{\sqrt{k/\log\log k}} \tag{209}$$

$$= \sum_{k=m}^{\infty} \frac{C_{\mathcal{P}_0}}{\sqrt{k\log^{2+\delta/2}(k)\cancel{\log\log k}}\cdot\sqrt{k/\cancel{\log\log k}}} \tag{210}$$

$$= \sum_{k=m}^{\infty} \frac{C_{\mathcal{P}_0}}{k\log^{1+\delta/4}(k)}, \tag{211}$$

and since $(k\log^{1+\delta/4}(k))^{-1}$ is summable for any $\delta > 0$, we have that the above vanishes as $m \to \infty$, hence

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}_0} \sum_{k=m}^{\infty} \frac{\mathbb{E}_P\left|\left\{\widehat{\mu}_{k-1}^x(Z_k) - \mu^x(Z_k)\right\}\cdot\left\{\widehat{\mu}_{k-1}^y(Z_k) - \mu^y(Z_k)\right\}\right|}{\sqrt{k/\log\log k}} = 0. \tag{212}$$

Applying Waudby-Smith et al. [48, Theorem 2], we have that

$$\bar{b}_n \equiv \frac{1}{n}\sum_{k=1}^{n}\left\{\widehat{\mu}^x_{k-1}(Z_k)-\mu^x(Z_k)\right\}\cdot\left\{\widehat{\mu}^y_{k-1}(Z_k)-\mu^y(Z_k)\right\} = \bar{o}_{\mathcal{P}_0}\left(\frac{1}{\sqrt{n\log\log n}}\right), \tag{213}$$

which completes the proof. □

Lemma A.3 (Convergence of average cross-terms)**.** *Suppose that for some $\delta > 0$, and some independent Z with the same distribution as Z_n,*

$$\sup_{P\in\mathcal{P}_0}\mathbb{E}_P\left[(\{\widehat{\mu}^x_n(Z_n)-\mu^x(Z_n)\}\xi^y_n)^2\right] = O\left(\frac{1}{(\log n)^{2+\delta}}\right). \tag{214}$$

Then,

$$\frac{1}{n}\sum_{i=1}^{n}\nu_i^{x,y} = \bar{o}_{\mathcal{P}_0}(1/\sqrt{n\log\log n}), \tag{215}$$

with an analogous statement holding when x and y are swapped in the above condition and conclusion.

Proof. We will only prove the result for $\nu_i^{x,y}$ but the same argument goes through for $\nu_i^{y,x}$. Appealing to (214), let $C_{\mathcal{P}_0}$ be a constant so that

$$\sup_{P\in\mathcal{P}}\mathbb{E}_P\left[(\{\widehat{\mu}^x_n(Z_n)-\mu^x(Z_n)\}\xi^y_n)^2\right] \leqslant \frac{C_{\mathcal{P}_0}}{(\log n)^{2+\delta}}. \tag{216}$$

Then notice that for all m sufficiently large

$$\begin{aligned}
&\sup_{P\in\mathcal{P}}\sum_{k=m}^{\infty}\frac{\mathbb{E}_P\left[(\{\widehat{\mu}^x_k(Z_k)-\mu^x(Z_k)\}\xi^y_k)^2\right]}{k/\log\log k} && (217)\\
\leqslant\ &\sup_{P\in\mathcal{P}}\sum_{k=m}^{\infty}\frac{C_{\mathcal{P}_0}}{k(\log k)^{2+\delta}/\log\log k} && (218)\\
\leqslant\ &\sup_{P\in\mathcal{P}}\sum_{k=m}^{\infty}\frac{C_{\mathcal{P}_0}}{k(\log k)^{1+\delta}} && (219)\\
=\ &0, && (220)
\end{aligned}$$

and hence

$$\lim_{m\to\infty}\sup_{P\in\mathcal{P}}\sum_{k=m}^{\infty}\frac{\mathbb{E}_P\left[\left(\{\widehat{\mu}^x_{k-1}(Z_k)-\mu^x(Z_k)\}\xi^y_k\right)^2\right]}{k/\log\log k} = 0. \tag{221}$$

By Waudby-Smith et al. [48, Theorem 2], we have that

$$\frac{1}{n}\sum_{i=1}^{n}\nu_i^{x,y} = \bar{o}_{\mathcal{P}}(1/\sqrt{n\log\log n}), \tag{222}$$

completing the proof. □

Lemma A.4 (Convergence of average squared cross-terms)**.** *Under Assumption SeqGCM-2, we have that*

$$\frac{1}{n}\sum_{i=1}^{n}(\nu_i^{x,y})^2 \equiv \frac{1}{n}\sum_{i=1}^{n}(\xi^x_i)^2\{\mu^y-\widehat{\mu}^y_i\}^2 = \bar{o}_{\mathcal{P}_0}\left(1/\log n\right). \tag{223}$$

An analogous statement holds with ξ^x_n replaced by ξ^y_n and $\{\mu^y(Z_n)-\widehat{\mu}^y_n(Z_n)\}$ replaced by $\{\mu^x(Z_n)-\widehat{\mu}^x_n(Z_n)\}$.

Proof. Using Assumption SeqGCM-2, let $C_{\mathcal{P}_0} > 0$ be a constant so that

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P\left[(\xi_n^x)^2\{\mu^y(Z_n) - \widehat{\mu}_{n-1}^y(Z_n)\}^2\right] \leqslant \frac{C_{\mathcal{P}_0}}{(\log n)^{2+\delta}}. \tag{224}$$

Therefore, we have that

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}_0} \sum_{k=m}^{\infty} \frac{\mathbb{E}_P\left[(\xi_k^x)^2\{\mu^y(Z_k) - \widehat{\mu}_{k-1}^y(Z_k)\}^2\right]}{k(\log k)^{-1}} \tag{225}$$

$$\leqslant \lim_{m\to\infty} \sum_{k=m}^{\infty} \frac{C_{\mathcal{P}_0}}{k(\log k)^{2+\delta-1}} \tag{226}$$

$$= \lim_{m\to\infty} \sum_{k=m}^{\infty} \frac{C_{\mathcal{P}_0}}{k(\log k)^{1+\delta}} \tag{227}$$

$$= 0. \tag{228}$$

Combining the above with Waudby-Smith et al. [48, Theorem 2], we have that

$$\frac{1}{n}\sum_{i=1}^{n}(\xi_i^x)^2\{\mu^y - \widehat{\mu}_i^y(Z_i)\}^2 = \bar{o}_{\mathcal{P}_0}\left(1/\log n\right), \tag{229}$$

completing the proof. □

A.6 Proof of Corollary 5.3

Corollary 5.3 (A $\mathcal{P}$-uniform law of the iterated logarithm)**.** *Suppose $(X_n)_{n=1}^{\infty}$ are defined on probability spaces $(\Omega, \mathcal{F}, \mathcal{P})$ where $\mathcal{P}$ is a collection of distributions such that*

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P|X - \mathbb{E}_P X|^{2+\delta} < \infty \quad \textit{and} \quad \inf_{P\in\mathcal{P}} \mathrm{Var}_P(X) > 0 \tag{70}$$

for some $\delta > 0$. Then,

$$\sup_{k\geqslant n} \frac{|\sum_{i=1}^{k}(X_i - \mathbb{E}_P(X))|}{\sqrt{2\mathrm{Var}_P(X)k\log\log k}} = 1 + \bar{o}_{\mathcal{P}}(1). \tag{71}$$

Proof. This is a consequence of our distribution-uniform strong Gaussian coupling given in Theorem 5.1. Letting $\mu_P := \mathbb{E}_P(X)$ and $\sigma_P := \sqrt{\mathrm{Var}_P(X)}$ to reduce notational clutter, note that by Theorem 5.1, we have that there exists a construction with a sequence of standard Gaussians $Y_1, Y_2, \ldots$ such that

$$\sum_{i=1}^{n}(X_i - \mu_P)/\sigma_P = \sum_{i=1}^{n} Y_i + \bar{o}_{\mathcal{P}}(n^{1/q}(\log n)^{2/q}), \tag{230}$$

where $q := 2 + \delta$, or more formally that for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}} \mathbb{P}_P\left(\exists k \geqslant n : \frac{|\sum_{i=1}^{k}(X_i - \mu_P)/\sigma_P - \sum_{i=1}^{k} Y_i|}{k^{1/q}(\log k)^{2/q}} > \varepsilon\right) = 0. \tag{231}$$

Now, by the law of the iterated logarithm, we also have that

$$\sup_{\ell\geqslant n} \frac{|\sum_{i=1}^{\ell} Y_i|}{\sqrt{2\ell\log\log\ell}} = 1 + \bar{o}_P(1), \tag{232}$$

for each $P \in \mathcal{P}$, and since Y has the same distribution on every element of $P \in \mathcal{P}$, the above also holds with $\bar{o}_P(1)$ replaced by $\bar{o}_{\mathcal{P}}(1)$. Now, to prove the final result, we have that

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}} \mathbb{P}_P \left(\exists k \geqslant n : \left| \sup_{\ell \geqslant k} \frac{|\sum_{i=1}^{\ell}(X_i - \mu_P)|}{\sqrt{2\sigma_P^2 \ell \log\log \ell}} - 1 \right| > \varepsilon \right) \tag{233}$$

$$\leqslant \lim_{n\to\infty} \sup_{P\in\mathcal{P}} \mathbb{P}_P \left(\exists k \geqslant n : \left| \sup_{\ell \geqslant k} \left\{ \frac{|\sum_{i=1}^{\ell}(X_i - \mu_P)/\sigma_P - \sum_{i=1}^{\ell} Y_i|}{\sqrt{2\ell \log\log \ell}} + \frac{|\sum_{i=1}^{\ell} Y_i|}{\sqrt{2\ell \log\log \ell}} \right\} - 1 \right| > \varepsilon \right) \tag{234}$$

$$\leqslant \lim_{n\to\infty} \sup_{P\in\mathcal{P}} \mathbb{P}_P \left(\exists k \geqslant n : \left| \sup_{\ell \geqslant k} \left\{ \varepsilon/2 + \frac{|\sum_{i=1}^{\ell} Y_i|}{\sqrt{2\ell \log\log \ell}} \right\} - 1 \right| > \varepsilon/2 \right) + \tag{235}$$

$$\lim_{n\to\infty} \sup_{P\in\mathcal{P}} \mathbb{P}_P \left(\exists k \geqslant n : \frac{|\sum_{i=1}^{k}(X_i - \mu_P)/\sigma_P - \sum_{i=1}^{k} Y_i|}{\sqrt{2k \log\log k}} > \varepsilon/2 \right) \tag{236}$$

$$\leqslant \lim_{n\to\infty} \sup_{P\in\mathcal{P}} \mathbb{P}_P \left(\exists k \geqslant n : \left| \sup_{\ell \geqslant k} \left\{ \frac{|\sum_{i=1}^{\ell} Y_i|}{\sqrt{2\ell \log\log \ell}} \right\} - 1 \right| > \varepsilon/2 \right) + \tag{237}$$

$$\underbrace{\lim_{n\to\infty} \sup_{P\in\mathcal{P}} \mathbb{P}_P \left(\exists k \geqslant n : \frac{|\sum_{i=1}^{k}(X_i - \mu_P)/\sigma_P - \sum_{i=1}^{k} Y_i|}{k^{1/q}(\log k)^{2/q}} > \varepsilon/2 \right)}_{=0} \tag{238}$$

$$= \lim_{n\to\infty} \sup_{P\in\mathcal{P}} \mathbb{P}_P \left(\exists k \geqslant n : \left| \sup_{\ell \geqslant k} \left\{ \frac{|\sum_{i=1}^{\ell} Y_i|}{\sqrt{2\ell \log\log \ell}} \right\} - 1 \right| > \varepsilon/2 \right) \tag{239}$$

$$= \sup_{P\in\mathcal{P}} \lim_{n\to\infty} \mathbb{P}_P \left(\exists k \geqslant n : \left| \sup_{\ell \geqslant k} \left\{ \frac{|\sum_{i=1}^{\ell} Y_i|}{\sqrt{2\ell \log\log \ell}} \right\} - 1 \right| > \varepsilon/2 \right) \tag{240}$$

$$= 0, \tag{241}$$

where the second inequality follows from Theorem 5.1 and the third follows from the triangle inequality and the fact that $k^{1/q}(\log k)^{2/q} \leqslant 2k \log\log k$ for all k sufficiently large. The second-last equality follows from the fact that the probability does not depend on features of the distribution P and the last equality follows from the P-pointwise law of the iterated logarithm. □

A.7 Proof of Lemma 5.2 and Theorem 5.1

Lemma 5.2 (Strong Gaussian coupling inequality)**.** *Let $(X_n)_{n=1}^{\infty}$ be independent random variables on the probability space $(\Omega, \mathcal{F}, P)$. Suppose that for some $q \geqslant 2$, we have $\mathbb{E}_P|X_k - \mathbb{E}_P X_k|^q < \infty$ for each $k \in \mathbb{N}$. Let $f(\cdot)$ be a positive and increasing function so that $\sum_{n=1}^{\infty}(nf(n))^{-1} < \infty$ and*

$$\sum_{k=1}^{\infty} \frac{\mathbb{E}_P|X_k - \mathbb{E}_P X_k|^q/\sigma_k^q}{kf(k)} < \infty, \tag{68}$$

where $\sigma_k^2 := \mathrm{Var}_P(X_k)$. Then one can construct a probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P}(P))$ rich enough to define $(\widetilde{X}_n, Y_n)_{n=1}^{\infty}$ where $\widetilde{X}_n$ and X_n are equidistributed for each n and $(Y_n)_{n=1}^{\infty}$ are marginally independent standard Gaussians so that for any $\varepsilon > 0$,

$$\mathbb{P}_{\widetilde{P}(P)} \left(\exists k \geqslant m : \left| \frac{\sum_{i=1}^{k}(\widetilde{X}_i - Y_i)}{k^{1/q} f(k)^{1/q}} \right| > \varepsilon \right) \leqslant \frac{C_{q,f}}{\varepsilon^q} \left\{ \sum_{k=2^{m-1}}^{\infty} \frac{\mathbb{E}_P|X_k - \mathbb{E}_P X_k|^q/\sigma_k^q}{kf(k)} + \frac{1}{2^m} \sum_{k=1}^{2^{m-1}-1} \frac{\mathbb{E}_P|X_k - \mathbb{E}_P X_k|^q/\sigma_k^q}{kf(k)} \right\}, \tag{69}$$

where $C_{q,f}$ is a constant that depends only on q and f.

First, we need the following result due to Lifshits [27, Theorem 3.3] which is itself a refinement of an inequality due to Sakhanenko [36].

Lemma A.5 (Sakhanenko-Lifshits inequality)**.** *Let $X_1, X_2, \cdots : \Omega \to \mathbb{R}$ be independent mean-zero random variables on a probability space $(\Omega, \mathcal{F}, P)$ and let $q \geqslant 2$. Then one can construct a new probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P})$ rich enough to contain $(\widetilde{X}_n, Y_n)_{n=1}^{\infty}$ so that $(X_1, X_2, \dots)$ and $(\widetilde{X}_1, \widetilde{X}_2, \dots)$ are equidistributed and $(Y_1, Y_2, \dots)$ are standard Gaussian random variables so that*

$$\mathbb{E}_P \left(\max_{1 \leqslant k \leqslant n} \left| \sum_{i=1}^{k} X_i/\sigma_P(X_i) - \sum_{i=1}^{k} Y_i \right| \right)^q \leqslant C_q \sum_{i=1}^{n} \frac{\mathbb{E}_P|X_i|^q}{\sigma_P(X_i)^q} \tag{242}$$

where C_q is a constant depending only on q.

Notice that $\sigma_P(X_i) = \sigma_{\widetilde{P}}(X_i)$ and $\mathbb{E}_P|X_i|^q = \mathbb{E}_{\widetilde{P}}|X_i|^q$ in the above lemma so we may use them interchangeably.

Proof of the main result

Proof of Lemma 5.2. Throughout the proof, we will use σ_i in place of $\sigma_P(X_i)$ whenever the distribution P is clear from context. We will also let $S_k(P)$ and G_k be the partial sums given by

$$S_k(P) := \sum_{i=1}^{k} (X_i - \mathbb{E}_P(X_i))/\sigma_P(X_i) \quad \text{and} \quad G_k := \sum_{i=1}^{k} Y_i. \tag{243}$$

For any $P \in \mathcal{P}$, we appeal to Lemma A.5 and let $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P})$ be a construction so that for any n,

$$\mathbb{E}_{\widetilde{P}} \left(\max_{1 \leqslant k \leqslant n} |S_k - G_k| \right)^q \leqslant C_q \sum_{k=1}^{n} \frac{\mathbb{E}_P|X_k - \mathbb{E}_P(X_k)|^q}{\sigma_P(X_k)^q}, \tag{244}$$

By Markov's inequality, we have that for any $z > 0$,

$$\mathbb{P}_{\widetilde{P}} \left(\max_{1 \leqslant k \leqslant n} |S_k(P) - G_k| > z \right) \leqslant C_q \frac{\sum_{k=1}^{n} \mathbb{E}_P|X_k - \mathbb{E}_P(X_k)|^q/\sigma_k^q}{z^q}, \tag{245}$$

noting that the right-hand side does not depend on the new probability space, but only on the original P. Defining $\Delta_k \equiv \Delta_k(P) := S_k(P) - G_k$, we have that for any k and n,

$$\max_{\mathcal{D}(n-1) \leqslant k < \mathcal{D}(n)} \{\Delta_k\} = \underbrace{\max_{\mathcal{D}(n-1) \leqslant k < \mathcal{D}(n)} \left\{\Delta_k - \Delta_{\mathcal{D}(n-1)-1}\right\}}_{(a)} + \underbrace{\Delta_{\mathcal{D}(n-1)-1}}_{(b)}, \tag{246}$$

where $\mathcal{D}(n) := 2^n$ are exponentially spaced demarcation points that will become important in the arguments to follow. We will proceed by separately bounding (a) and (b) time-uniformly with high-probability.

Step 1: Bounding (a) time-uniformly with high probability. Let $a_k := k^{1/q} f(k)^{1/q}$. By (245) applied to $\Delta_k - \Delta_{\mathcal{D}(n-1)-1} \equiv \sum_{i=\mathcal{D}(n-1)}^{k} (X_i - \mathbb{E}_P(X_i))/\sigma_i$ with $z := \varepsilon a_{\mathcal{D}(n-1)}$, we have that

$$\mathbb{P}_{\widetilde{P}} \left(\max_{\mathcal{D}(n-1) \leqslant k < \mathcal{D}(n)} \{\Delta_k - \Delta_{\mathcal{D}(n-1)-1}\} > \varepsilon a_{\mathcal{D}(n-1)} \right) \tag{247}$$

$$\leqslant C_q \sum_{k=\mathcal{D}(n-1)}^{\mathcal{D}(n)-1} \frac{\mathbb{E}_P|X_k - \mathbb{E}_P(X_k)|^q/\sigma_k^q}{a_{\mathcal{D}(n-1)}^q \varepsilon^q} \tag{248}$$

$$\leqslant \varepsilon^{-q} C_q \sum_{k=\mathcal{D}(n-1)}^{\mathcal{D}(n)-1} \frac{\mathbb{E}_P|X_k - \mathbb{E}_P(X_k)|^q/\sigma_k^q}{a_k^q}. \tag{249}$$

Union bounding over $n = m, m+1, \ldots$ we have that

$$\mathbb{P}_{\tilde{P}}\left(\exists n \geqslant m \text{ and } k \in \{\mathcal{D}(n-1), \ldots, \mathcal{D}(n)-1\} : \{\Delta_k - \Delta_{\mathcal{D}(n-1)-1}\} > \varepsilon a_k\right) \tag{250}$$

$$\leqslant \mathbb{P}_{\tilde{P}}\left(\exists n \geqslant m : \max_{\mathcal{D}(n-1) \leqslant k < \mathcal{D}(n)} \{\Delta_k - \Delta_{\mathcal{D}(n-1)-1}\} > \varepsilon a_{\mathcal{D}(n-1)}\right) \tag{251}$$

$$\leqslant \sum_{n=m}^{\infty} \varepsilon^{-q} C_q \sum_{k=\mathcal{D}(n-1)}^{\mathcal{D}(n)-1} \frac{\mathbb{E}_P|X_k - \mathbb{E}_P(X_k)|^q/\sigma_k^q}{a_k^q}. \tag{252}$$

$$\leqslant \varepsilon^{-q} C_q \sum_{n=\mathcal{D}(m-1)}^{\infty} \frac{\mathbb{E}_P|X_k - \mathbb{E}_P(X_k)|^q}{a_k^q}. \tag{253}$$

$$\tag{254}$$

Step 2: Bounding (b) time-uniformly with high probability. Applying (245) to $\Delta_{\mathcal{D}(n-1)-1}$ with $z = \varepsilon a_{\mathcal{D}(n-1)}$, we have that

$$\mathbb{P}_{\tilde{P}}\left(|\Delta_{\mathcal{D}(n-1)-1}| > \varepsilon a_{\mathcal{D}(n-1)}\right) \tag{255}$$

$$\leqslant C_q \sum_{k=1}^{\mathcal{D}(n-1)-1} \frac{\mathbb{E}_P|X_k - \mathbb{E}_P(X_k)|^q/\sigma_k^q}{a_{\mathcal{D}(n-1)}^q \varepsilon^q} \tag{256}$$

$$\leqslant \frac{C_q}{\varepsilon^q a_{\mathcal{D}(n-1)}^q} \sum_{k=1}^{\mathcal{D}(n-1)-1} \mathbb{E}_P|X_k - \mathbb{E}_P(X_k)|^q/\sigma_k^q. \tag{257}$$

Union bounding again over $n = m, m+1, \ldots$, we have

$$\mathbb{P}_{\tilde{P}}\left(\exists n \geqslant m \text{ and } k \in \{\mathcal{D}(n-1), \ldots, \mathcal{D}(n)-1\} : |\Delta_{\mathcal{D}(n-1)-1}| > \varepsilon a_k\right) \tag{258}$$

$$\leqslant \mathbb{P}_{\tilde{P}}\left(\exists n \geqslant m \text{ and } k \in \{\mathcal{D}(n-1), \ldots, \mathcal{D}(n)-1\} : |\Delta_{\mathcal{D}(n-1)-1}| > \varepsilon a_{\mathcal{D}(n-1)}\right) \tag{259}$$

$$= \mathbb{P}_{\tilde{P}}\left(\exists n \geqslant m : |\Delta_{\mathcal{D}(n-1)-1}| > \varepsilon a_{\mathcal{D}(n-1)}\right) \tag{260}$$

$$\leqslant \sum_{n=m}^{\infty} \frac{C_q}{\varepsilon^q a_{\mathcal{D}(n-1)}^q} \sum_{k=1}^{\mathcal{D}(n-1)-1} \mathbb{E}_P|X_k - \mathbb{E}_P(X_k)|^q/\sigma_k^q. \tag{261}$$

Step 3: Union bounding over the results from Steps 1 and 2. Putting Steps 1 and 2 together, we have the following time-uniform crossing inequality for $|\Delta_k|$:

$$\mathbb{P}_{\tilde{P}}(\exists k \geqslant m : |\Delta_k| > 2\varepsilon a_k) \tag{262}$$

$$\leqslant \mathbb{P}_{\tilde{P}}(\exists k \geqslant m : |\Delta_k - \Delta_{\mathcal{D}(n-1)-1}| + |\Delta_{\mathcal{D}(n-1)-1}| > a_k + a_k) \tag{263}$$

$$\leqslant \varepsilon^{-q} C_q \left[\sum_{n=\mathcal{D}(m-1)}^{\infty} \frac{\mathbb{E}_P|X_k - \mathbb{E}_P(X_k)|^q}{a_k^q} + \sum_{n=m}^{\infty} \frac{1}{a_{\mathcal{D}(n-1)}^q} \sum_{k=1}^{\mathcal{D}(n-1)-1} \frac{\mathbb{E}_P|X_k - \mathbb{E}_P(X_k)|^q}{\sigma_k^q}\right] \tag{264}$$

$$\leqslant \varepsilon^{-q} C_q \left[\sum_{n=\mathcal{D}(m-1)}^{\infty} \frac{\mathbb{E}_P|X_k - \mathbb{E}_P(X_k)|^q}{a_k^q} + \sum_{n=m}^{\infty} \frac{1}{a_{\mathcal{D}(n-1)}^q} \sum_{k=1}^{\mathcal{D}(n-1)-1} \frac{\mathbb{E}_P|X_k - \mathbb{E}_P(X_k)|^q}{\sigma_k^q}\right]. \tag{265}$$

Letting $\rho_k^q := \mathbb{E}_P|X_k - \mathbb{E}_P(X_k)|^q$ and further simplifying the above expression so that it does not depend on the demarcation points $\mathcal{D}(n)$, we have

$$\mathbb{P}_{\tilde{P}}(\exists k \geqslant m : |\Delta_k| > 2\varepsilon a_k) \tag{266}$$

$$\leqslant \varepsilon^{-q} C_q \left[\sum_{k=\mathcal{D}(m-1)}^{\infty} \frac{\rho_k^q/\sigma_k^q}{a_k^q} + \sum_{n=m}^{\infty} \frac{1}{a_{\mathcal{D}(n-1)}^q} \sum_{k=1}^{\mathcal{D}(n-1)-1} \rho_k^q/\sigma_k^q \right] \tag{267}$$

$$\leqslant \varepsilon^{-q} C_q \left[\sum_{k=\mathcal{D}(m-1)}^{\infty} \frac{\rho_k^q/\sigma_k^q}{kf(k)} + \sum_{n=m}^{\infty} \frac{1}{\mathcal{D}(n-1)f(\mathcal{D}(n-1))} \sum_{k=1}^{\mathcal{D}(n-1)-1} \rho_k^q/\sigma_k^q \right] \tag{268}$$

$$\leqslant \varepsilon^{-q} C_q \left[\sum_{k=\mathcal{D}(m-1)}^{\infty} \frac{\rho_k^q/\sigma_k^q}{kf(k)} + \sum_{n=m}^{\infty} \frac{1}{\mathcal{D}(n-1)} \sum_{k=1}^{\mathcal{D}(n-1)-1} \frac{\rho_k^q/\sigma_k^q}{f(\mathcal{D}(k))} \right] \tag{269}$$

$$= \varepsilon^{-q} C_q \left[\sum_{k=2^{m-1}}^{\infty} \frac{\rho_k^q/\sigma_k^q}{kf(k)} + \sum_{n=m}^{\infty} \frac{1}{2^{n-1}} \sum_{k=1}^{\mathcal{D}(n-1)-1} \frac{\rho_k^q/\sigma_k^q}{f(2^k)} \right] \tag{270}$$

$$\leqslant \varepsilon^{-q} C_q \left[\sum_{k=2^{m-1}}^{\infty} \frac{\rho_k^q/\sigma_k^q}{kf(k)} + \sum_{n=m}^{\infty} \frac{C_f^{-1}}{2^{n-1}} \sum_{k=1}^{2^n} \frac{\rho_k^q/\sigma_k^q}{kf(k)} \right], \tag{271}$$

where (268) follows from the definition of $a_k := k^{1/q} f(k)^{1/q}$, (269) follows from the fact that f is increasing and that $\mathcal{D}(k) := 2^k$, (270) follows from the definition of $\mathcal{D}(\cdot)$, and (271) from the fact that $f(k) \geqslant C_f \log(k)$ for all $k \geqslant 1$ and some constant C_f depending only on f (if this were not true, then $\sum_{k=1}^{\infty}[kf(k)]^{-1}$ would not be summable).

The final result follows from observing that $\sum_{k=1}^{2^n} \rho_k^q/(\sigma_k^q kf(k)) \leqslant \sum_{k=1}^{\infty} \rho_k^q/(\sigma_k^q kf(k))$ and absorbing constants only depending on q and f into $C_{q,f}$:

$$\mathbb{P}_{\tilde{P}}(\exists k \geqslant m : |\Delta_k| > \varepsilon a_k) \tag{272}$$

$$\leqslant 2^q \varepsilon^{-q} C_q \left[\sum_{k=2^{m-1}}^{\infty} \frac{\rho_k^q/\sigma_k^q}{kf(k)} + \sum_{n=m}^{\infty} \frac{C_f^{-1}}{2^{n-1}} \sum_{k=1}^{2^n} \frac{\rho_k^q/\sigma_k^q}{kf(k)} \right] \tag{273}$$

$$\leqslant 2^q \varepsilon^{-q} C_q C_f^{-1} \left[C_f \sum_{k=2^{m-1}}^{\infty} \frac{\rho_k^q/\sigma_k^q}{kf(k)} + 2^{-m} \left(\sum_{k=1}^{2^{m-1}-1} \frac{\rho_k^q/\sigma_k^q}{kf(k)} + \sum_{k=2^{m-1}}^{\infty} \frac{\rho_k^q/\sigma_k^q}{kf(k)} \right) \right] \tag{274}$$

$$= 2^q \varepsilon^{-q} C_q C_f^{-1} \left[(C_f + 2^{-m}) \sum_{k=2^{m-1}}^{\infty} \frac{\rho_k^q/\sigma_k^q}{kf(k)} + \frac{1}{2^m} \sum_{k=1}^{2^{m-1}-1} \frac{\rho_k^q/\sigma_k^q}{kf(k)} \right] \tag{275}$$

$$\leqslant \varepsilon^{-q} C_{q,f} \left[\sum_{k=2^{m-1}}^{\infty} \frac{\rho_k^q/\sigma_k^q}{kf(k)} + \frac{1}{2^m} \sum_{k=1}^{2^{m-1}-1} \frac{\rho_k^q/\sigma_k^q}{kf(k)} \right], \tag{276}$$

which completes the proof □

Let us now show how Theorem 5.1 is a consequence of the above.

Theorem 5.1 (Distribution-uniform strong Gaussian approximation)**.** *Let $(X_n)_{n=1}^{\infty}$ be independent and identically distributed random variables defined on the collection of probability spaces $(\Omega, \mathcal{F}, \mathcal{P})$ with means $\mu_P := \mathbb{E}_P(X)$ and variances $\sigma_P^2 := \mathbb{E}_P(X - \mu_P)^2$. If X has $q > 2$ uniformly upper-bounded moments, and a uniformly positive variance, i.e.*

$$\sup_{P \in \mathcal{P}} \mathbb{E}_P|X - \mu_P|^q < \infty \quad \textit{and} \quad \inf_{P \in \mathcal{P}} \sigma_P^2 > 0, \tag{66}$$

then there exists a construction with independent standard Gaussians $(Y_n)_{n=1}^{\infty} \sim N(0,1)$ *so that*

$$\left|\sum_{i=1}^{n} \frac{X_i - \mu_P}{\sigma_P} - \sum_{i=1}^{n} Y_i\right| = \bar{o}_{\mathcal{P}}(n^{1/q} \log^{2/q}(n)). \tag{67}$$

Proof. The proof of Theorem 5.1 amounts to analyzing the $\mathcal{P}$-uniform tail behavior of the probability bound in Lemma 5.2. Indeed, for each $P \in \mathcal{P}$, let $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P}(P))$ be the construction that yields

$$\mathbb{P}_{\widetilde{P}}\left(\exists k \geqslant m : \left|\frac{\sum_{i=1}^{k}(X_i - \mu_P)/\sigma_P - \sum_{i=1}^{k} Y_i}{k f(k)}\right| > \varepsilon\right) \tag{277}$$

$$\leqslant \varepsilon^{-q} C_{q,f}\left[\sum_{k=2^{m-1}}^{\infty} \frac{\rho_P^q/\sigma_P^q}{k f(k)} + \frac{1}{2^m} \sum_{k=1}^{2^{m-1}-1} \frac{\rho_P^q/\sigma_P^q}{k f(k)}\right], \tag{278}$$

where $\rho_P^q := \mathbb{E}_P|X - \mathbb{E}_P X|^q$ and $\sigma_P^2 := \mathbb{E}_P(X - \mathbb{E}_P)^2$. Let $\bar{\rho} < \infty$ be the uniform upper bound so that $\sup_{P\in\mathcal{P}} \rho_P^q \leqslant \bar{\rho}^q$ and $\underline{\sigma}^2 > 0$ be the uniform lower bound on the variance so that $\inf_{P\in\mathcal{P}} \sigma_P^2 \geqslant \underline{\sigma}^2$. Replacing the above finite sum by its infinite extension and taking suprema over P on both sides, we have that

$$\sup_{P\in\mathcal{P}} \mathbb{P}_{\widetilde{P}(P)}\left(\exists k \geqslant m : \left|\frac{\sum_{i=1}^{k}(X_i - \mu_P)/\sigma_P - \sum_{i=1}^{k} Y_i}{k f(k)}\right| > \varepsilon\right) \tag{279}$$

$$\leqslant \frac{\bar{\rho}^q C_{q,f}}{\underline{\sigma}^q \varepsilon^q}\left[\underbrace{\sum_{k=2^{m-1}}^{\infty} \frac{1}{k f(k)}}_{(\star)} + \underbrace{\frac{1}{2^m}\sum_{k=1}^{\infty} \frac{1}{k f(k)}}_{(\dagger)}\right]. \tag{280}$$

Now, $(\star) \to 0$ and $(\dagger) \to 0$ as $m \to \infty$, both of which follow from the fact that $1/[k f(k)]$ is summable. Instantiating the above for $f(k) := \log^2(k)$ completes the proof. □

B Additional theoretical discussions and results

B.1 The Robbins-Siegmund distribution

Fundamental to this paper is a probability distribution that describes the supremum of a transformed Wiener process with a delayed start (see Lemma B.1). As far as we can tell, the distribution was first (implicitly) discovered by Robbins and Siegmund [33] and as such we refer to it as the *Robbins-Siegmund distribution*. In this section, we provide its cumulative distribution function (CDF) Ψ and show how the suprema of scaled Wiener processes have this distribution Ψ. The Robbins-Siegmund distribution has also been implicitly used in Waudby-Smith et al. [47] and Bibaut et al. [4] for the sake of P-pointwise anytime-valid inference.

Definition B.1 (The Robbins-Siegmund distribution)**.** *We say that a nonnegative random variable* $\mathfrak{R}$ *follows the Robbins-Siegmund (R-S) distribution if its CDF is given by*

$$\Psi(r) := 1 - 2\left[1 - \Phi(\sqrt{r}) + \sqrt{r}\phi(\sqrt{r})\right]; \quad r \geqslant 0, \tag{281}$$

where Φ *and* ϕ *are the CDF and density of a standard Gaussian, respectively.*

The following lemma demonstrates how the supremum of a transformed Wiener process follows the Robbins-Siegmund distribution.

Lemma B.1. *Let $(W(t))_{t\geqslant 0}$ be a standard Wiener process and define*

$$\mathfrak{R} := \sup_{t\geqslant 1}\left\{\frac{W(t)^2}{t} - \log t\right\}. \tag{282}$$

Then, $\mathfrak{R}$ follows the Robbins-Siegmund distribution given in Definition B.1.

Proof. Rather than derive its CDF for a given r, we will derive the survival function $\mathbb{P}(\mathfrak{R}\geqslant a^2)$ for any $a\geqslant 0$, showing that $\mathbb{P}(\mathfrak{R}\geqslant a^2) = 1-\Psi(a^2)$ as given in Definition B.1, which will yield the desired result.

$$\begin{aligned}
\mathbb{P}(\mathfrak{R}\geqslant a^2) &= \mathbb{P}\left(\exists t\geqslant 1 : W(t)^2/t - \log t \geqslant a^2\right) && (283)\\
&= \mathbb{P}\left(\exists t\geqslant 1 : |W(t)| \geqslant \sqrt{t\left[a^2+\log t\right]}\right) && (284)\\
&= 2[1-\Phi(a)+a\phi(a)] \equiv 1-\Psi(a^2). && (285)
\end{aligned}$$

where the last line follows from Robbins and Siegmund [33] but with their value of τ set to 1. Alternatively, a different proof found in Waudby-Smith et al. [47, Lemma A.14] yields the desired result. □

The following lemma demonstrates that appropriately scaled discrete Gaussian partial sums converge to the Robbins-Siegmund distribution.

Lemma B.2 (Transformed Gaussian partial sums converge to the Robbins-Siegmund distribution)**.** *Let G_k be a sum of iid Gaussian random variables with mean zero and variance σ^2. Then,*

$$\sup_{k\geqslant m}\left\{\frac{G_k^2}{k\sigma^2} - \log(k/m)\right\} \xrightarrow{d} \Psi \quad \text{as } m\to\infty. \tag{286}$$

Proof. Since G_k is a sum of iid Gaussian random variables with mean zero and variance σ^2, we have by Komlós, Major, and Tusnády [22, 23] that $G_k = \sigma W(k) + \bar{O}_P(\log k)$ where $(W(t))_{t\geqslant 0}$ is a standard Wiener process. We will now show that $\sup_{k\geqslant n}\left\{G_k^2/k\sigma^2 - \log(k/n)\right\}$ converges to the Robbins-Siegmund distribution.

$$\begin{aligned}
&\sup_{k\geqslant n}\left\{\frac{G_k^2}{k\sigma^2} - \log(k/n)\right\} && (287)\\
=\ &\sup_{k\in[n,\infty)}\left\{\frac{(W(k)+O(\log k))^2}{k\sigma^2} - \log(k/n) + \bar{O}_P\left(\frac{\log(k)\sqrt{k\log\log k}}{k+1}\right) + O\left(\log\left[\frac{n+1}{n}\right]\right)\right\} && (288)\\
=\ &\sup_{k\in[n,\infty)}\left\{\frac{W(k)^2}{k\sigma^2} + \bar{O}_P\left(\frac{\log k}{k}\cdot\sqrt{k\log\log k}\right) - \log(k/n)\right\} + \bar{o}_P(1) && (289)\\
=\ &\sup_{k\in[n,\infty)}\left\{\frac{W(k)^2}{k\sigma^2} - \log(k/n)\right\} + \bar{o}_P(1) && (290)\\
=\ &\sup_{tn\in[n,\infty)}\left\{W(tn)^2/tn\sigma^2 - \log(tn/n)\right\} + \bar{o}_P(1) && (291)\\
=\ &\sup_{t\in[1,\infty)}\left\{nW(t)^2/tn\sigma^2 - \log(t)\right\} + \bar{o}_P(1) && (292)\\
=\ &\sup_{t\in[1,\infty)}\left\{W(t)^2/t\sigma^2 - \log(t)\right\} + \bar{o}_P(1), && (293)
\end{aligned}$$

where (288) results from the discrete-to-continuous overshoot in $1/k$ and $\log(k/n)$ when taking a supremum over $k\in[n,\infty)$ instead of over $k\in\{n,n+1,\dots\}$ and (292) follows from elementary properties of the Wiener process. It follows that

$$\sup_{k\geqslant n}\left\{\frac{G_k^2}{k\sigma^2} - \log(k/n)\right\} \xrightarrow{d} \Psi \quad \text{as } n\to\infty. \tag{294}$$

□

The following lemma establishes that the Robbins-Siegmund distribution has a Lipschitz CDF.

Lemma B.3. *The cumulative distribution function $\Psi(r)$ of a Robbins-Siegmund-distributed random variable is L-Lipschitz with $L \leqslant 1/4$. In other words,*

$$\sup_{r \geqslant 0} \left| \frac{d}{dr} \Psi(r) \right| \leqslant 1/4. \tag{295}$$

Proof. Clearly, it suffices to show that $1 - \Psi(r)$ is L-Lipschitz. Defining $f(r) := 1 - \Psi(r)$, we have that

$$\begin{aligned} f(r) &:= 2(1 - \Phi(\sqrt{r}) + \sqrt{r}\phi(\sqrt{r})) && (296) \\ &= 2 - 2\Phi(\sqrt{r}) + 2\sqrt{r}\phi(\sqrt{r}). && (297) \end{aligned}$$

A direct calculation reveals that

$$\begin{aligned} f'(r) &= -\phi(\sqrt{r})\frac{1}{\sqrt{r}} + \frac{1}{\sqrt{r}}\phi(\sqrt{r}) + \sqrt{r}\phi'(\sqrt{r})\frac{1}{\sqrt{r}} && (298) \\ &= \phi'(\sqrt{r}) && (299) \\ &= \frac{-\sqrt{r}}{\sqrt{2\pi}} \exp\left\{-r/2\right\}, && (300) \end{aligned}$$

from which it is easy to check that $\sup_{r \geqslant 0} |f'(r)| \leqslant 1/4$, completing the proof. □

B.2 Uniform convergence of perturbed random variables

Throughout many of our proofs, we rely on facts about convergence of random variables under $\mathcal{P}$-uniformly small perturbations. Similar results are common in the proofs of $\mathcal{P}$-uniform fixed-n central limit theorems but are only discussed in the context of Gaussian limiting distributions and for time-pointwise convergence. We show here that similar results hold for *Robbins-Siegmund* limiting distributions (in fact, for any continuous and Lipschitz distribution) under time- and $\mathcal{P}$-uniformly small perturbations to random variables inside suprema over time.

Lemma B.4 (Time-uniform closure under additive $\bar{o}_{\mathcal{P}}(1)$-perturbations)**.** *Let $((A_{k,m})_{k=m}^{\infty})_{m=1}^{\infty}$ be a doubly indexed sequence of random variables on $(\Omega, \mathcal{F}, \mathcal{P})$. Let $Z \sim F(z)$ with where the CDF F is L-Lipschitz and does not depend on $P \in \mathcal{P}$. Suppose that*

$$\lim_{m \to \infty} \sup_{P \in \mathcal{P}} \sup_{z \in \mathbb{R}} \left| \mathbb{P}_P \left(\sup_{k \geqslant m} \{A_{k,m}\} \geqslant z \right) - \mathbb{P}_P(Z \geqslant z) \right| = 0. \tag{301}$$

If $R_n = \bar{o}_{\mathcal{P}}(1)$, then

$$\lim_{m \to \infty} \sup_{P \in \mathcal{P}} \sup_{z \in \mathbb{R}} \left| \mathbb{P}_P \left(\sup_{k \geqslant m} \{A_{k,m} + R_k\} \geqslant z \right) - \mathbb{P}_P(Z \geqslant z) \right| = 0. \tag{302}$$

Proof. Let $\varepsilon > 0$ be any positive constant. Using (301) and the fact that $R_n = \bar{o}_{\mathcal{P}}(1)$, let M be large enough so that for all $m \geqslant M$, we have

$$\sup_{P \in \mathcal{P}} \sup_{z \in \mathbb{R}} \left| \mathbb{P}_P \left(\sup_{k \geqslant m} \{A_{k,m}\} \geqslant z \right) - \mathbb{P}_P(Z \geqslant z) \right| < \varepsilon \tag{303}$$

and so that

$$\sup_{P \in \mathcal{P}} \sup_{z \in \mathbb{R}} \mathbb{P}_P \left(\sup_{k \geqslant m} |R_k| \geqslant \varepsilon \right) < \varepsilon. \tag{304}$$

Then, writing out $\mathbb{P}_P(\sup_{k\geqslant m}\{A_{k,m}+R_k\}\geqslant z)$ for any $P\in\mathcal{P}$, $z\in\mathbb{R}$, and $m\geqslant M$, we find the following upper bound,

$$\mathbb{P}_P\left(\sup_{k\geqslant m}\{A_{k,m}+R_k\}\geqslant z\right) \tag{305}$$

$$=\mathbb{P}_P\left(\sup_{k\geqslant m}\{A_{k,m}+R_k\}\geqslant z\;\middle|\;\sup_{k\geqslant m}|R_k|<\varepsilon\right)\underbrace{\mathbb{P}_P(|R_n|<\varepsilon)}_{\leqslant 1}+ \tag{306}$$

$$\mathbb{P}_P\left(\sup_{k\geqslant m}\{A_{k,m}+R_k\}\geqslant z\;\middle|\;\sup_{k\geqslant m}|R_k|\geqslant\varepsilon\right)\underbrace{\mathbb{P}_P\left(\sup_{k\geqslant m}|R_k|\geqslant\varepsilon\right)}_{\leqslant\varepsilon} \tag{307}$$

$$\leqslant\mathbb{P}_P\left(\sup_{k\geqslant m}\{A_{k,m}\}\geqslant z-\varepsilon\right)+\varepsilon, \tag{308}$$

and via a similar argument, the corresponding lower bound,

$$\mathbb{P}_P\left(\sup_{k\geqslant m}\{A_{k,m}+R_k\}\geqslant z\right) \tag{309}$$

$$\geqslant\mathbb{P}_P\left(\sup_{k\geqslant m}\{A_{k,m}\}\geqslant z+\varepsilon\right)-\varepsilon. \tag{310}$$

Using first the upper bound, we thus have that

$$\mathbb{P}_P\left(\sup_{k\geqslant m}\{A_{k,m}+R_k\}\geqslant z\right)-\mathbb{P}_P(Z\geqslant z) \tag{311}$$

$$\leqslant\mathbb{P}_P\left(\sup_{k\geqslant m}\{A_{k,m}\}\geqslant z-\varepsilon\right)-\mathbb{P}_P(Z\geqslant z)+\varepsilon \tag{312}$$

$$\leqslant\mathbb{P}_P(Z\geqslant z-\varepsilon)-\mathbb{P}_P(Z\geqslant z)+2\varepsilon \tag{313}$$

$$=1-F(z-\varepsilon)-(1-F(z))+2\varepsilon \tag{314}$$

$$=F(z)-F(z-\varepsilon)+2\varepsilon \tag{315}$$

$$\leqslant(L+2)\varepsilon, \tag{316}$$

where the last line used the fact that F is L-Lipschitz for some $L>0$. Similarly,

$$\mathbb{P}_P\left(\sup_{k\geqslant m}\{A_{k,m}+R_k\}\geqslant z\right)-\mathbb{P}_P(Z\geqslant z) \tag{317}$$

$$\geqslant\;-(L+2)\varepsilon, \tag{318}$$

Putting the two together, we have that

$$\left|\mathbb{P}_P\left(\sup_{k\geqslant m}\{A_{k,m}+R_k\}\geqslant z\right)-\mathbb{P}_P(Z\geqslant z)\right|\leqslant(L+2)\varepsilon, \tag{319}$$

and since L neither depends on z nor on P, we have that

$$\left|\mathbb{P}_P\left(\sup_{k\geqslant m}\{A_{k,m}+R_k\}\geqslant z\right)-\mathbb{P}_P(Z\geqslant z)\right|\leqslant(L+2)\varepsilon. \tag{320}$$

Since ε was arbitrary, it follows that

$$\lim_{n\to\infty}\sup_{P\in\mathcal{P}}\sup_{z\in\mathbb{R}}\left|\mathbb{P}_P\left(\sup_{k\geqslant m}\{A_{k,m}+R_k\}\geqslant z\right)-\mathbb{P}_P(Z\geqslant z)\right|=0, \tag{321}$$

which completes the proof.

□

Lemma B.5 (Time-uniform closure under multiplicative $\bar{o}_{\mathcal{P}}(1)$-perturbations)**.** *Let $((A_{k,m})_{k=m}^{\infty})_{m=1}^{\infty}$ be a doubly indexed sequence of random variables on $(\Omega, \mathcal{F}, \mathcal{P})$. Let $Z \sim F(z)$ with CDF F not depending on $P \in \mathcal{P}$. Suppose that*

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} \sup_{x\in\mathbb{R}} \left| \mathbb{P}_P \left(\sup_{k\geqslant m} \{A_{k,m}\} \leqslant x \right) - F(x) \right| = 0 \tag{322}$$

and suppose that $R_n = \bar{o}_{\mathcal{P}}(1)$. Then,

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} \sup_{x\in\mathbb{R}} \left| \mathbb{P}_P \left(\sup_{k\geqslant m} \left\{ \frac{A_{k,m}}{1+R_k} \right\} \leqslant x \right) - F(x) \right| = 0. \tag{323}$$

Proof. The proof proceeds in four steps. In First, we ensure that the CDF of $\sup_{k\geqslant m} A_{k,m}$ is $(\mathcal{P}, x)$-uniformly close to F. Second, we use a result of van der Vaart [44] and Slutsky's theorem to justify why deterministically perturbed continuous random variables converge quantile-uniformly in distribution. Third, we use the fact that $R_n = \bar{o}_{\mathcal{P}}(1)$ to ensure that R_n is $\mathcal{P}$- and time-uniformly smaller than a certain radius. The fourth and final steps puts these results together to arrive at the desired result.

Let $\varepsilon > 0$ be arbitrary. Our goal is to show that there exists M sufficiently large so that for all $m \geqslant M$,

$$\textbf{Goal:} \quad \sup_{P\in\mathcal{P}} \sup_{x\in\mathbb{R}} \left| \mathbb{P}_P \left(\sup_{k\geqslant m} \left\{ \frac{A_{k,m}}{1+R_k} \right\} \leqslant x \right) - F(x) \right| < 2\varepsilon. \tag{324}$$

(Here, the multiplication by 2 is only for algebraic convenience later on.)

Step 1: Ensuring that the CDF of $\sup_{k\geqslant m} A_{k,m}$ is $(\mathcal{P}, x)$-uniformly close to $F(x)$. By the assumption in (322), choose M_1 large enough so that whenever $m \geqslant M_1$, we have

$$\sup_{P\in\mathcal{P}} \sup_{x\in\mathbb{R}} \left| \mathbb{P}_P \left(\sup_{k\geqslant m} \{A_{k,m}\} \leqslant x \right) - F(x) \right| < \varepsilon \tag{325}$$

Step 2: CDFs of deterministically perturbed random variables are close to F. Letting $X \sim F$ be a continuous random variable with CDF F, note that

$$\frac{X}{1+h} \xrightarrow{d} X \tag{326}$$

as $h \to 0$ by Slutsky's theorem. Consequently, by van der Vaart [44, Lemma 2.11] combined with the fact that $F(x)$ is continuous in $x \in \mathbb{R}$, we have that

$$\lim_{h\to 0} \sup_x |F(x(1+h)) - F(x)| = \lim_{h\to 0} \sup_x \left| \mathbb{P}\left(\frac{X}{1+h} \leqslant x \right) - \mathbb{P}(X \leqslant x) \right| \tag{327}$$

$$= 0. \tag{328}$$

As such, let $h_2 > 0$ be a positive number so that whenever $|h| \leqslant h_2$,

$$\sup_{x\in\mathbb{R}} |F(x(1+h)) - F(x)| < \varepsilon. \tag{329}$$

Step 3: Ensuring that R_n is $\mathcal{P}$- and time-uniformly close to 0. Given the assumption that $R_n = \bar{o}_{\mathcal{P}}(1)$, choose M_3 large enough so that for all $m \geqslant M_3$, we have

$$\sup_{P\in\mathcal{P}} \mathbb{P}_P \left(\sup_{k\geqslant m} |R_k| \geqslant h_2 \right) < \varepsilon, \tag{330}$$

where h_2 is as in Step 2.

Step 4: Putting Steps 1–3 together to obtain the final bound. Set $M = \max\{M_1, M_3\}$. First, consider the following upper bound on $\mathbb{P}_P\left(\sup_{k\geqslant m}\{A_{k,m}/(1+R_k)\} \leqslant x\right)$ for any $m \geqslant M$:

$$\mathbb{P}_P\left(\sup_{k\geqslant m}\left\{\frac{A_{k,m}}{1+R_k}\right\} \leqslant x\right) \tag{331}$$

$$= \mathbb{P}_P\left(\sup_{k\geqslant m}\left\{\frac{A_{k,m}}{1+R_k}\right\} \leqslant x \;\middle|\; \sup_{k\geqslant m}|R_k| < h_2\right) \underbrace{\mathbb{P}_P\left(\sup_{k\geqslant m}|R_k| < h_2\right)}_{\leqslant 1} + \tag{332}$$

$$\mathbb{P}_P\left(\sup_{k\geqslant m}\left\{\frac{A_{k,m}}{1+R_k}\right\} \leqslant x \;\middle|\; \sup_{k\geqslant m}|R_k| \geqslant h_2\right) \underbrace{\mathbb{P}_P\left(\sup_{k\geqslant m}|R_k| \geqslant h_2\right)}_{\leqslant \varepsilon} \tag{333}$$

$$\leqslant \mathbb{P}_P\left(\sup_{k\geqslant m}\left\{\frac{A_{k,m}}{1+h_2}\right\} \leqslant x\right) + \varepsilon. \tag{334}$$

By a similar argument, we have that for all $m \geqslant M$,

$$\mathbb{P}_P\left(\sup_{k\geqslant m}\left\{\frac{A_{k,m}}{1+R_k}\right\} \leqslant x\right) \tag{335}$$

$$\geqslant \mathbb{P}_P\left(\sup_{k\geqslant m}\left\{\frac{A_{k,m}}{1-h_2}\right\} \leqslant x\right) - \varepsilon. \tag{336}$$

Keeping these upper and lower bounds in mind, we have that

$$\mathbb{P}_P\left(\sup_{k\geqslant m}\left\{\frac{A_{k,m}}{1+R_k}\right\} \leqslant x\right) - F(x) \tag{337}$$

$$\leqslant \mathbb{P}_P\left(\sup_{k\geqslant m}\{A_{k,m}\} \leqslant x(1+h_2)\right) - F(x) + \varepsilon \tag{338}$$

$$\leqslant F(x(1+h_2)) - F(x) + \varepsilon \tag{339}$$

$$\leqslant 2\varepsilon. \tag{340}$$

and

$$\mathbb{P}_P\left(\sup_{k\geqslant m}\left\{\frac{A_{k,m}}{1+R_k}\right\} \leqslant x\right) - F(x) \tag{341}$$

$$\geqslant \mathbb{P}_P\left(\sup_{k\geqslant m}\{A_{k,m}\} \leqslant x(1-h_2)\right) - F(x) - \varepsilon \tag{342}$$

$$\geqslant F(x(1-h_2)) - F(x) - \varepsilon \tag{343}$$

$$\geqslant -2\varepsilon. \tag{344}$$

Putting these upper and lower bounds on the difference of probabilities together and noting that their bounds do not depend on $P \in \mathcal{P}$ nor on $x \in \mathbb{R}$, we have

$$\sup_{P\in\mathcal{P}}\sup_{x\in\mathbb{R}}\left|\mathbb{P}_P\left(\sup_{k\geqslant m}\left\{\frac{A_{k,m}}{1+R_k}\right\} \leqslant x\right) - F(x)\right| \leqslant 2\varepsilon, \tag{345}$$

which completes the proof.

□