Asymptotic anytime-valid statistical inference

Ian Waudby-Smith

Miller Institute & Department of Statistics
University of California, Berkeley

JSM 2026

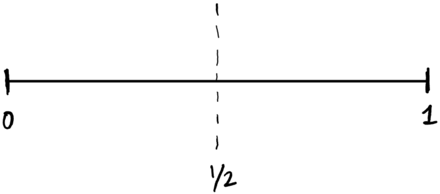
A brief crash course on (nonasymptotic) anytime-valid inference

A toy problem: is a coin biased?

Flip the coin $n = 100$ times.

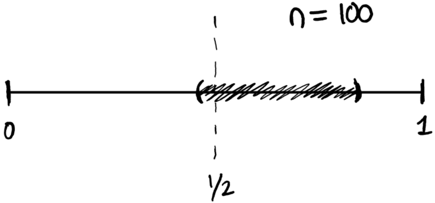
A toy problem: is a coin biased?

Flip the coin $n = 100$ times.



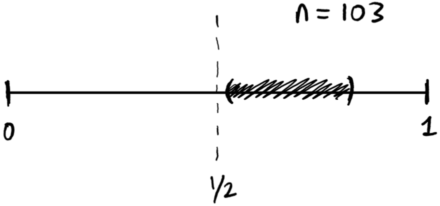
A toy problem: is a coin biased?

Flip the coin $n = 100$ times.



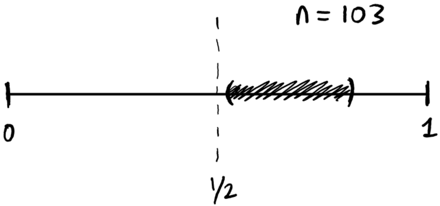
A toy problem: is a coin biased?

Flip the coin $n = 100$ times.



A toy problem: is a coin biased?

Flip the coin $n = 100$ times.

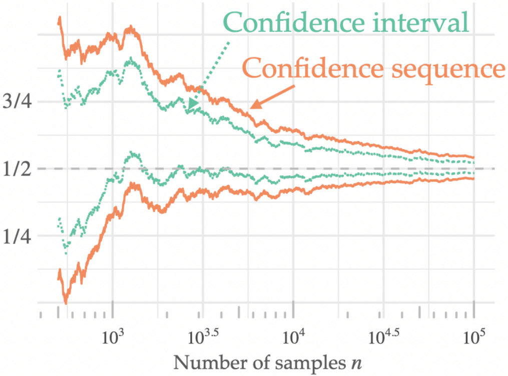


This is tempting, but **inflates type-I error**.

A solution: **Confidence sequences**

(Herbert Robbins & coauthors, 1960s-70s)

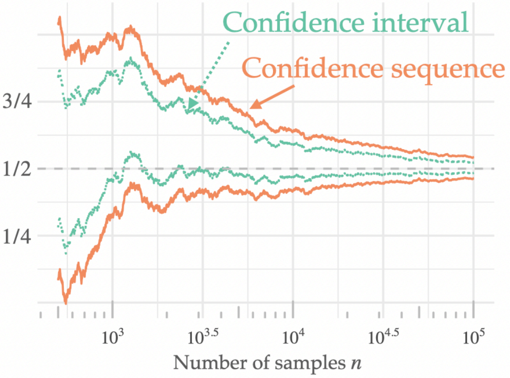
Estimation view



CI: $\forall n \in \mathbb{N},\ P(\theta \in \dot{C}_n) \geqslant 1 - \alpha.$

CS: $P(\forall n \in \mathbb{N},\ \theta \in \bar{C}_n) \geqslant 1 - \alpha.$

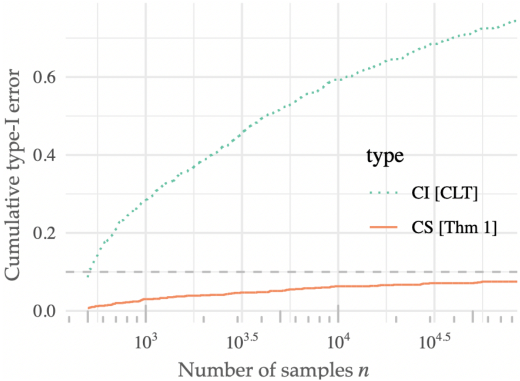
Estimation view



CI: $\forall n \in \mathbb{N},\ P(\theta \in \dot{C}_n) \geqslant 1 - \alpha.$

CS: $P(\forall n \in \mathbb{N},\ \theta \in \bar{C}_n) \geqslant 1 - \alpha.$

Testing view



CI: $\forall n \in \mathbb{N},\ P(\theta \notin \dot{C}_n) \leqslant \alpha.$

CS: $P(\exists n \in \mathbb{N} : \theta \notin \bar{C}_n) \leqslant \alpha.$

Confidence sequences, sequential tests, etc. draw on a rich literature:

(i) Ville, Wald ('30–'40s)

(ii) Robbins, Siegmund, Darling, and Lai ('60s–'80s)

(iii) Vovk, Shafer, Johari, Howard, Kaufmann, Ramdas, Wang, Grünwald, de Heide, Koolen, ter Schure, Orabona, Jun, Duchi, Shekhar, Lindon (2010s-now).

Confidence sequences, sequential tests, etc. draw on a rich literature:

(i) Ville, Wald ('30–'40s)

(ii) Robbins, Siegmund, Darling, and Lai ('60s–'80s)

(iii) Vovk, Shafer, Johari, Howard, Kaufmann, Ramdas, Wang, Grünwald, de Heide, Koolen, ter Schure, Orabona, Jun, Duchi, Shekhar, Lindon (2010s-now).

The vast majority of this work has been *nonasymptotic*:

$$\mathbb{P}(\exists n \in \{1, 2, \dots\} : \mathbf{error}_n) \leqslant \alpha.$$

Statistical problem	Fixed-n test/CI	Anytime-valid test/CS/e-value

Statistical problem	**Fixed-n test/CI**	**Anytime-valid test/CS/e-value**
Parametric inference	Neyman, Pearson, Fisher, Wald, etc.	Robbins & co-authors ('60s, '70s) Grünwald et al. (2024) Wasserman et al. (2020)

Statistical problem	**Fixed-n test/CI**	**Anytime-valid test/CS/e-value**
Parametric inference	Neyman, Pearson, Fisher, Wald, etc.	Robbins & co-authors ('60s, '70s) Grünwald et al. (2024) Wasserman et al. (2020)
Sub-Gaussian (sub-ψ)	Chernoff bounds	Howard et al. (2020, '21)

Statistical problem	**Fixed-n test/CI**	**Anytime-valid test/CS/e-value**
Parametric inference	Neyman, Pearson, Fisher, Wald, etc.	Robbins & co-authors ('60s, '70s) Grünwald et al. (2024) Wasserman et al. (2020)
Sub-Gaussian (sub-ψ)	Chernoff bounds	Howard et al. (2020, '21)
Bounded RVs	Hoeffding, (empirical) Bernstein	Hendriks (2020) W-S & Ramdas (2024) Orabona & Jun (2023)

Statistical problem	**Fixed-n test/CI**	**Anytime-valid test/CS/e-value**
Parametric inference	Neyman, Pearson, Fisher, Wald, etc.	Robbins & co-authors ('60s, '70s) Grünwald et al. (2024) Wasserman et al. (2020)
Sub-Gaussian (sub-ψ)	Chernoff bounds	Howard et al. (2020, '21)
Bounded RVs	Hoeffding, (empirical) Bernstein	Hendriks (2020) W-S & Ramdas (2024) Orabona & Jun (2023)
$\mathbb{E}\|X - \mu\|^{2+\gamma} < C$	Catoni (2012) and others	Wang & Ramdas (2023)

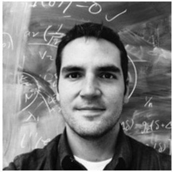
Statistical problem	**Fixed-n test/CI**	**Anytime-valid test/CS/e-value**
Parametric inference	Neyman, Pearson, Fisher, Wald, etc.	Robbins & co-authors ('60s, '70s) Grünwald et al. (2024) Wasserman et al. (2020)
Sub-Gaussian (sub-ψ)	Chernoff bounds	Howard et al. (2020, '21)
Bounded RVs	Hoeffding, (empirical) Bernstein	Hendriks (2020) W-S & Ramdas (2024) Orabona & Jun (2023)
$\mathbb{E}\lvert X-\mu\rvert^{2+\gamma} < C$	Catoni (2012) and others	Wang & Ramdas (2023)
Non-asymptotic inference is provably impossible	**Central limit theorem**	This talk.

Asymptotic anytime-valid inference

Robbins & Siegmund [1970]
Bibaut, Kallus, & Lindon [2022]
W-S et al. [2024, 2026abc]

David Arbour

Adobe Research



Edward H. Kennedy

CMU Statistics

Martin Larsson

CMU Mathematics

Aaditya Ramdas

Stanford Statistics

Ritwik Sinha

Adobe Research

Let us aim to define what is meant by asymptotic anytime-valid inference.

We will look at *doubly indexed* intervals for each $k \geqslant m$:

$$(C_k^{(1)})_{k=1}^{\infty}, (C_k^{(2)})_{k=2}^{\infty}, (C_k^{(3)})_{k=3}^{\infty}, \cdots$$

Let us aim to define what is meant by asymptotic anytime-valid inference.

We will look at *doubly indexed* intervals for each $k \geqslant m$:

$$(C_k^{(m)})_{k=m}^{\infty}; \ \ m \in \mathbb{N}.$$

Let us aim to define what is meant by asymptotic anytime-valid inference.

We will look at *doubly indexed* intervals for each $k \geqslant m$:

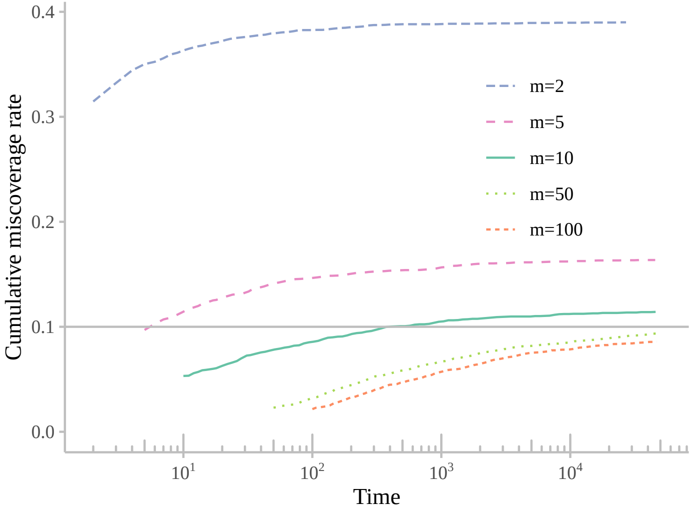
$$(C_k^{(m)})_{k=m}^{\infty}; \; m \in \mathbb{N}.$$

Definition. (W-S et al., [AoS 2024a], Bibaut et al. [2022])

We say $C_k^{(m)}$ has asymptotic anytime $(1-\alpha)$-coverage for μ if

$$\limsup_{m \to \infty} \sup_{P \in \mathcal{P}} P\left(\exists k \geqslant m : \mu \notin C_k^{(m)}\right) \leqslant \alpha.$$

Asymptotic anytime-valid inference in a picture



Example: Let $(X_n)_{n\in\mathbb{N}}$ be i.i.d. with mean μ_P and variance σ_P^2.

Define $\widehat{\mu}_n := \frac{1}{n}\sum_{i=1}^n X_i$ and $\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \widehat{\mu}_n^2$.

Example: Let $(X_n)_{n\in\mathbb{N}}$ be i.i.d. with mean μ_P and variance σ_P^2.

Define $\widehat{\mu}_n := \frac{1}{n}\sum_{i=1}^n X_i$ and $\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \widehat{\mu}_n^2$.

If $\sup_{P\in\mathcal{P}}\mathbb{E}_P|X|^3 < \infty$, then $(C_k^{(m)})_{k=m}^{\infty}$ has asymptotic coverage for μ_P where

$$C_k^{(m)} := \left[\widehat{\mu}_k \pm \widehat{\sigma}_k\sqrt{\frac{\Psi^{-1}(1-\alpha) + \log(k/m)}{k}}\right].$$

Example: Let $(X_n)_{n\in\mathbb{N}}$ be i.i.d. with mean μ_P and variance σ_P^2.

Define $\widehat{\mu}_n := \frac{1}{n}\sum_{i=1}^n X_i$ and $\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \widehat{\mu}_n^2$.

If $\sup_{P\in\mathcal{P}}\mathbb{E}_P|X|^3 < \infty$, then $(C_k^{(m)})_{k=m}^{\infty}$ has asymptotic coverage for μ_P where

$$C_k^{(m)} := \left[\widehat{\mu}_k \pm \widehat{\sigma}_k \sqrt{\frac{\Psi^{-1}(1-\alpha) + \log(k/m)}{k}}\right].$$

That is,

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} P\left(\exists k \geqslant m : \mu_P \notin C_k^{(m)}\right) = \alpha.$$

Example: Let $(X_n)_{n\in\mathbb{N}}$ be i.i.d. with mean μ_P and variance σ_P^2.

Define $\widehat{\mu}_n := \frac{1}{n}\sum_{i=1}^n X_i$ and $\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \widehat{\mu}_n^2$.

If $\sup_{P\in\mathcal{P}} \mathbb{E}_P|X|^3 < \infty$, then $(C_k^{(m)})_{k=m}^\infty$ has asymptotic coverage for μ_P where

$$C_k^{(m)} := \left[\widehat{\mu}_k \pm \widehat{\sigma}_k \sqrt{\frac{\Psi^{-1}(1-\alpha) + \log(k/m)}{k}}\right].$$

That is,

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} P\left(\exists k \geqslant m : \mu_P \notin C_k^{(m)}\right) = \alpha.$$

The CDF Ψ is an anytime-valid analogue of the χ^2 distribution.

Let us provide some intuition for how Ψ appears.

Let us provide some intuition for how Ψ appears.

Typically, one uses the CLT to show that $\sqrt{n}(\widehat{\mu}_n - \mu_{\mathsf{P}})/\widehat{\sigma}_n \to N(0,1)$.

Let us provide some intuition for how Ψ appears.

Typically, one uses the CLT to show that $\sqrt{n}(\widehat{\mu}_n - \mu_{\mathsf{P}})/\widehat{\sigma}_n \to N(0,1)$.

Here, one can instead uniform strong Gaussian couplings (see W-S–Larsson–Ramdas) to show that for a Wiener process $(W(t))_{t \geqslant 0}$,

$$\sup_{k \geqslant m} \left\{ \frac{k(\widehat{\mu}_k - \mu_{\mathsf{P}})^2}{\widehat{\sigma}_k^2} - \log(k/m) \right\} \overset{d}{\approx} \sup_{t \geqslant 1} \left\{ \frac{W(t)^2}{t} - \log(t) \right\},$$

Let us provide some intuition for how Ψ appears.

Typically, one uses the CLT to show that $\sqrt{n}(\widehat{\mu}_n - \mu_{\mathsf{P}})/\widehat{\sigma}_n \to N(0,1)$.

Here, one can instead uniform strong Gaussian couplings (see W-S–Larsson–Ramdas) to show that for a Wiener process $(W(t))_{t\geqslant 0}$,

$$\sup_{k\geqslant m}\left\{\frac{k(\widehat{\mu}_k - \mu_{\mathsf{P}})^2}{\widehat{\sigma}_k^2} - \log(k/m)\right\} \overset{d}{\approx} \sup_{t\geqslant 1}\left\{\frac{W(t)^2}{t} - \log(t)\right\},$$

and that $\Psi(x) = P(\sup_{t\geqslant 1}\left\{W(t)^2/t - \log(t)\right\} \leqslant x)$; $x \geqslant 0$.

Let us provide some intuition for how Ψ appears.

Typically, one uses the CLT to show that $\sqrt{n}(\widehat{\mu}_n - \mu_{\mathsf{P}})/\widehat{\sigma}_n \to N(0,1)$.

Here, one can instead uniform strong Gaussian couplings (see W-S–Larsson–Ramdas) to show that for a Wiener process $(W(t))_{t\geqslant 0}$,

$$\sup_{k\geqslant m}\left\{\frac{k(\widehat{\mu}_k - \mu_{\mathsf{P}})^2}{\widehat{\sigma}_k^2} - \log(k/m)\right\} \overset{d}{\approx} \sup_{t\geqslant 1}\left\{\frac{W(t)^2}{t} - \log(t)\right\},$$

and that $\Psi(x) = P(\sup_{t\geqslant 1}\left\{W(t)^2/t - \log(t)\right\} \leqslant x)$; $x \geqslant 0$.

After finding the constant c_α for which the above crosses c_α with probability α, we obtain the aforementioned confidence sequences.

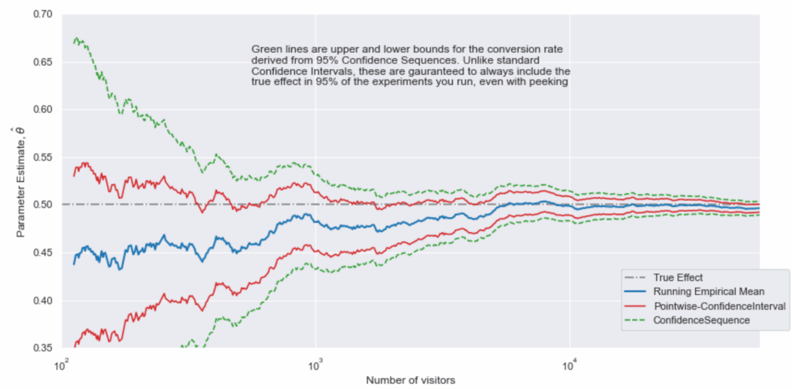
$$C_k^{(m)} := \left[\widehat{\mu}_k \pm \widehat{\sigma}_k \sqrt{\frac{c_\alpha + \log(k/m)}{k}} \right].$$

$$C_k^{(m)} := \left[\widehat{\mu}_k \pm \widehat{\sigma}_k \sqrt{\frac{\Psi(1-\alpha) + \log(k/m)}{k}} \right].$$

Adobe’s Statistical Methodology: Any Time Valid Confidence Sequences

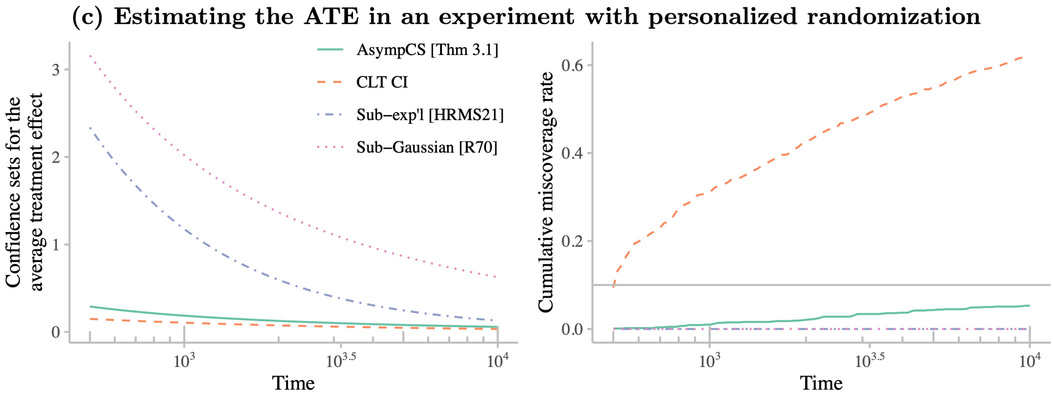
A **Confidence Sequence** is a sequential analog of a **Confidence Interval**, e.g. if you repeat your experiments one hundred times, and calculate an estimate of the mean metric and its associated 95%-Confidence Sequence for every new user that enters the experiment. A 95% Confidence Sequence will include the true value of the metric in 95 out of the 100 experiments that you ran. A 95% Confidence Interval could only be calculated once per experiment in order to give the same 95% coverage guarantee; not with every single new user. Confidence Sequences therefore allow you to continuously monitor experiments, without increasing False Positive error rates.

The difference between confidence sequences and confidence intervals for a single experiment is shown in the animation below:



Other companies mentioning asymptotic confidence sequences in their documentation include Spotify, GrowthBook, LaunchDarkly, and some other A/B testing startups.

Efficiency gains of AsympCSs can be large



Going from a CLT-based CI to an asymptotic CS extends validity from a single m to *all* $k \geqslant m$, even arbitrarily large.

Going from a CLT-based CI to an asymptotic CS extends validity from a single m to *all* $k \geqslant m$, even arbitrarily large.

What if it is known that analysis will never extend beyond some large but finite time N?

Can we have sharper inference by giving up validity beyond N?

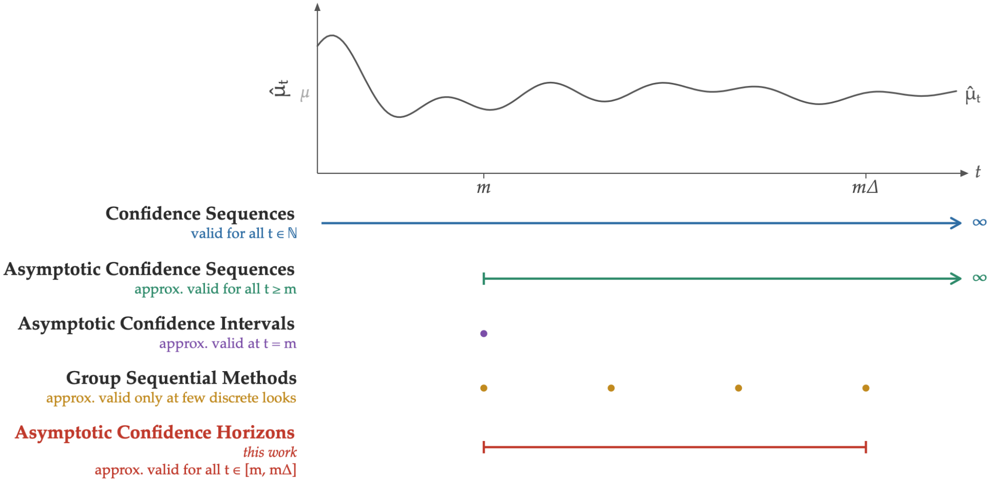
Going from a CLT-based CI to an asymptotic CS extends validity from a single m to *all* $k \geqslant m$, even arbitrarily large.

What if it is known that analysis will never extend beyond some large but finite time N?

Can we have sharper inference by giving up validity beyond N?
Yes, using *confidence horizons*.

Chase Mathis

UC Berkeley Statistics


$\hat{\mu}_t$
μ
$\hat{\mu}_t$
t
m
$m\Delta$
Confidence Sequences
valid for all $t \in \mathbb{N}$
∞
Asymptotic Confidence Sequences
approx. valid for all $t \geq m$
∞
Asymptotic Confidence Intervals
approx. valid at $t = m$
Group Sequential Methods
approx. valid only at few discrete looks
Asymptotic Confidence Horizons
this work
approx. valid for all $t \in [m, m\Delta]$

In the case of a confidence horizon, we show that for $\Delta \geqslant 1$,

$$\max_{m \leqslant k \leqslant m\Delta} \frac{\sqrt{k}|\widehat{\mu}_k - \mu_{\mathsf{P}}|}{\widehat{\sigma}_k} \overset{d}{\approx} \sup_{t \in [1,\Delta]} \frac{|W(t)|}{\sqrt{t}},$$

and the latter has a well-defined distribution function $\Phi(x; \Delta)$, recovering the Gaussian distribution when $\Delta = 1$.

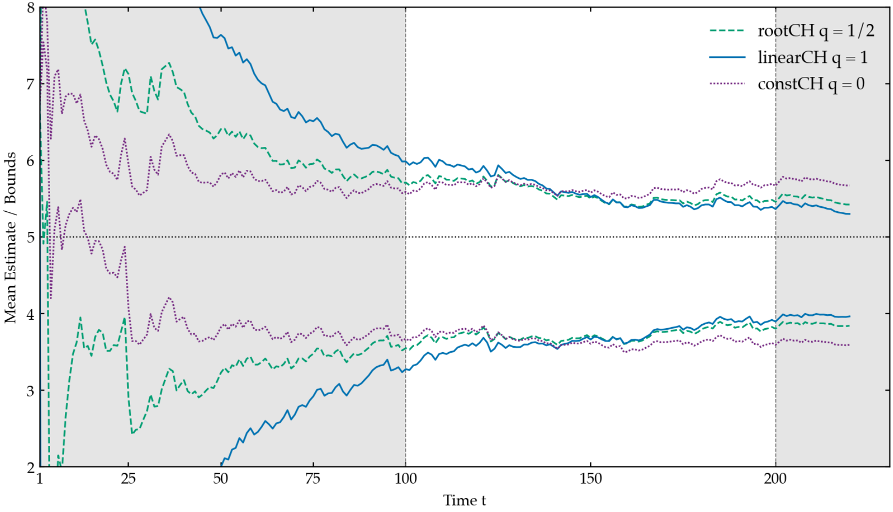
Specifically, if we define

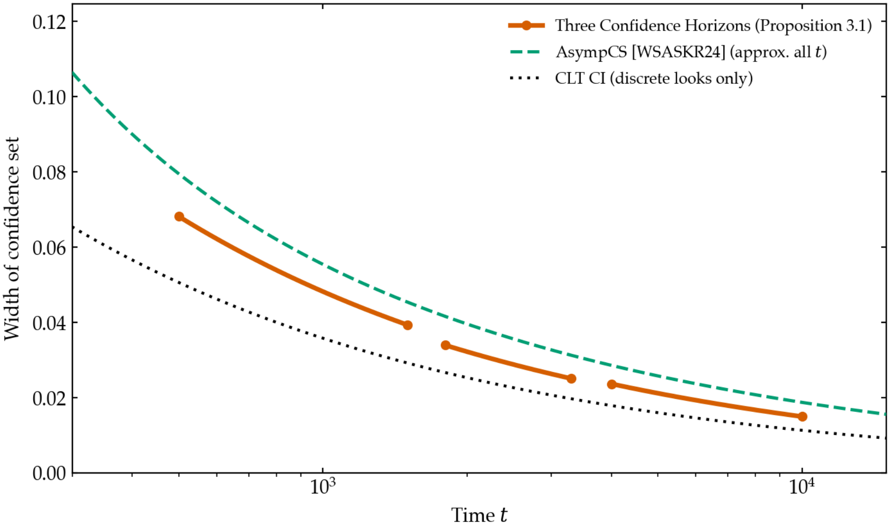
$$C_k = \widehat{\mu}_k \pm \widehat{\sigma}_k \frac{\Phi^{-1}(1-\alpha;\Delta)}{\sqrt{k}},$$

then it holds that

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} P\left(\exists k \in [m, m\Delta] : \mu_P \notin C_k\right) = \alpha.$$

Asymptotic Confidence Horizons





Three Confidence Horizons (Proposition 3.1)
AsympCS [WSASKR24] (approx. all t)
CLT CI (discrete looks only)
Width of confidence set
Time t
0.12
0.10
0.08
0.06
0.04
0.02
0.00
10^3
10^4

Summary

(a) Anytime-valid inference is a paradigm for performing statistical inference at stopping times and with sequentially collected data.

Summary

(a) Anytime-valid inference is a paradigm for performing statistical inference at stopping times and with sequentially collected data.

(b) Asymptotic confidence sequences bring the sharpness and flexibility of CLT-based methods to anytime-valid inference.

Summary

(a) Anytime-valid inference is a paradigm for performing statistical inference at stopping times and with sequentially collected data.

(b) Asymptotic confidence sequences bring the sharpness and flexibility of CLT-based methods to anytime-valid inference.

(c) Confidence horizons provide a way to trade validity past some large sample size N for sharper confidence intervals before N.

Thank you

ianws.com

However, notice that the guarantee

$$\forall P \in \mathcal{P}, \quad \lim_{m \to \infty} P\left(\exists k \geqslant m : \mu_P \notin C_k^{(m)}\right) = \alpha$$

is *distribution-pointwise*.

However, notice that the guarantee

$$\forall P \in \mathcal{P}, \quad \lim_{m \to \infty} P\left(\exists k \geqslant m : \mu_P \notin C_k^{(m)}\right) = \alpha$$

is *distribution-pointwise*.

Many modern asymptotic methods aim to be *distribution-uniform*, in the sense of

$$\lim_{m \to \infty} \sup_{P \in \mathcal{P}} P\left(\exists k \geqslant m : \mu_P \notin C_k^{(m)}\right) = \alpha.$$

How were we able to show that

$$C_k^{(m)} := \left[\widehat{\mu}_k \pm \widehat{\sigma}_k \sqrt{\frac{\Psi^{-1}(1-\alpha) + \log(k/m)}{k}}\right]?$$

satisfies the $\mathcal{P}$-uniform coverage guarantee?

How were we able to show that

$$C_k^{(m)} := \left[\widehat{\mu}_k \pm \widehat{\sigma}_k \sqrt{\frac{\Psi^{-1}(1-\alpha) + \log(k/m)}{k}} \right]?$$

satisfies the $\mathcal{P}$-uniform coverage guarantee?

After some algebraic juggling, this can be reduced to showing that

$$\lim_{m \to \infty} \sup_{P \in \mathcal{P}} \sup_{x \geqslant 0} \left| P \left(\sup_{k \geqslant m} \left\{ \frac{k \cdot (\widehat{\mu}_k - \mu_P)^2}{\widehat{\sigma}_k^2} - \log(k/m) \right\} \leqslant x \right) - \Psi(x) \right| = 0.$$

How were we able to show that

$$C_k^{(m)} := \left[\widehat{\mu}_k \pm \widehat{\sigma}_k \sqrt{\frac{\Psi^{-1}(1-\alpha) + \log(k/m)}{k}} \right] ?$$

satisfies the $\mathcal{P}$-uniform coverage guarantee?

After some algebraic juggling, this can be reduced to showing that

$$\sup_{k \geqslant m} \left\{ \text{statistic}_k^{(m)} \right\} \xrightarrow{\mathcal{P}} \Psi(x)$$

How were we able to show that

$$C_k^{(m)} := \left[\widehat{\mu}_k \pm \widehat{\sigma}_k \sqrt{\frac{\Psi^{-1}(1-\alpha) + \log(k/m)}{k}} \right]?$$

satisfies the $\mathcal{P}$-uniform coverage guarantee?
After some algebraic juggling, this can be reduced to showing that

$$\sup_{k \geqslant m} \left\{ \text{statistic}_k^{(m)} \right\} \xrightarrow{\mathcal{P}} \Psi(x)$$

$\to$ reduces time- and $\mathcal{P}$-uniform coverage / type-I error control to *$\mathcal{P}$-uniform convergence in distribution* to this d.f. Ψ.

Where does a distribution like Ψ arise?

Let $(W(t))_{t\geqslant 0}$ be a Wiener process (1-dim Brownian motion).

Where does a distribution like Ψ arise?

Let $(W(t))_{t \geqslant 0}$ be a Wiener process (1-dim Brownian motion).
Take the following transformation:

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

Where does a distribution like Ψ arise?

Let $(W(t))_{t \geqslant 0}$ be a Wiener process (1-dim Brownian motion).
Take the following transformation:

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

Then $\mathbb{P}(\zeta \leqslant x) = \Psi(x)$; $x \geqslant 0$.
(see Robbins & Siegmund (1970) or W-S et al., 2024a).

Some related lines of work

$\mathcal{P}$-uniform strong laws of large numbers

$\mathcal{P}$-uniform KMT approximation

$\mathcal{P}$-uniform asymptotic anytime-valid inference

Sequential conditional independence testing

Some related lines of work

(W-S-Larsson-Ramdas)

$\mathcal{P}$-uniform strong laws of large numbers

$\mathcal{P}$-uniform KMT approximation

$\mathcal{P}$-uniform asymptotic anytime-valid inference

Sequential conditional independence testing

(W-S-Kennedy-Ramdas)

Supplementary slides

$\mathcal{P}$-uniform anytime-valid inference

Distribution-uniform anytime-valid inference and the Robbins-Siegmund distributions

Ian Waudby-Smith[†], Edward H. Kennedy[‡], and Aaditya Ramdas[‡]

[†] University of California, Berkeley
[‡] Carnegie Mellon University
ianws@berkeley.edu {edward,aramdas}@stat.cmu.edu

Brief detour: **uniform** vs **pointwise** asymptotics

Mathematically, the distinction is quite simple.

Mathematically, the distinction is quite simple.

$$P\text{-pointwise:}\quad \sup_{P\in\mathcal{P}} \limsup_{n\to\infty} P\left(\mu \notin C_n\right) \leqslant \alpha.$$

$$\mathcal{P}\text{-uniform:}\quad \limsup_{n\to\infty} \sup_{P\in\mathcal{P}} P\left(\mu \notin C_n\right) \leqslant \alpha.$$

($\mathcal{P}$-uniform $\implies$ P-pointwise for all $P \in \mathcal{P}$.)

Mathematically, the distinction is quite simple.

$$P\text{-pointwise:} \quad \sup_{P \in \mathcal{P}} \limsup_{n \to \infty} P\left(\mu \notin C_n\right) \leqslant \alpha.$$

$$\mathcal{P}\text{-uniform:} \quad \limsup_{n \to \infty} \sup_{P \in \mathcal{P}} P\left(\mu \notin C_n\right) \leqslant \alpha.$$

($\mathcal{P}$-uniform $\implies$ P-pointwise for all $P \in \mathcal{P}$.)

Many refer to distribution-uniformity as "honesty".

→ See Li (1989), Kuchibhotla, Balakrishnan, and Wasserman (2022, '23)

→ Also, Robins et al. (2003) or §1.2.4 in Tsybakov for pointwise "pathologies".

What is unsettling about pointwise asymptotics?

Thought experiment: Suppose we want type-I error at most $\alpha + \varepsilon$. Take $n \in \{1, 2, \dots\}$ to be *as large as you like*.

What is unsettling about pointwise asymptotics?

Thought experiment: Suppose we want type-I error at most $\alpha + \varepsilon$. Take $n \in \{1, 2, \dots\}$ to be *as large as you like.*

$$\textbf{Pointwise:} \quad \exists P^{(n)} \in \mathcal{P} \ \text{ so that } \ P^{(n)}\left(\mu \notin C_n\right) > \alpha + \varepsilon.$$

What is unsettling about pointwise asymptotics?

Thought experiment: Suppose we want type-I error at most $\alpha + \varepsilon$. Take $n \in \{1, 2, \ldots\}$ to be *as large as you like.*

Pointwise: $\exists P^{(n)} \in \mathcal{P}$ so that $P^{(n)}(\mu \notin C_n) > \alpha + \varepsilon$.

Uniform: $\forall P \in \mathcal{P}$, we have $P(\mu \notin C_n) < \alpha + \varepsilon$.

A guarantee like

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} P\left(\mu \notin C_n\right) \leqslant \alpha$$

allows us to articulate what distributional properties certain asymptotic approximations are *insensitive* to.

</detour>

From here onwards, let's take uniformity as a desideratum
(remembering that pointwise asymptotics are obtained for free).

From here onwards, let's take uniformity as a desideratum
(remembering that pointwise asymptotics are obtained for free).

This will lead to strictly stronger statistical guarantees and motivate some interesting explorations in probability (latter half of the talk).

Let $(X_n)_{n\in\mathbb{N}}$ be i.i.d. on $(\Omega, \mathcal{F}, P)_{P\in\mathcal{P}}$ where $\mathcal{P}$ is a collection of distributions for which

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P|X - \mu_P|^{2+\delta} < \infty \quad \text{for some } \delta > 0.$$

Let $(X_n)_{n\in\mathbb{N}}$ be i.i.d. on $(\Omega, \mathcal{F}, P)_{P\in\mathcal{P}}$ where $\mathcal{P}$ is a collection of distributions for which

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P |X - \mu_P|^{2+\delta} < \infty \quad \text{for some } \delta > 0.$$

Theorem (W-S et al., 2024b): $\mathcal{P}$-uniform anytime-valid inference.

Given the confidence set from a few slides ago,

$$C_k^{(m)} := \left[\widehat{\mu}_k \pm \widehat{\sigma}_k \sqrt{\frac{\Psi^{-1}(1-\alpha) + \log(k/m)}{k}} \right],$$

Let $(X_n)_{n\in\mathbb{N}}$ be i.i.d. on $(\Omega, \mathcal{F}, P)_{P\in\mathcal{P}}$ where $\mathcal{P}$ is a collection of distributions for which

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P |X - \mu_P|^{2+\delta} < \infty \quad \text{for some } \delta > 0.$$

Theorem (W-S et al., 2024b): $\mathcal{P}$-uniform anytime-valid inference.

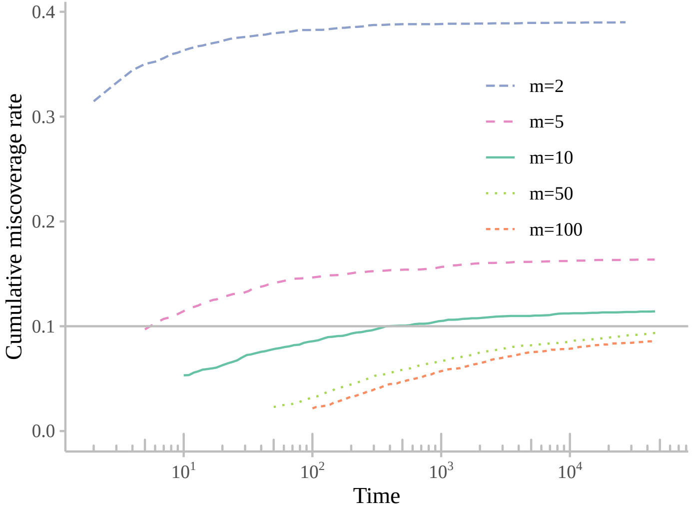
Given the confidence set from a few slides ago,

$$C_k^{(m)} := \left[\widehat{\mu}_k \pm \widehat{\sigma}_k \sqrt{\frac{\Psi^{-1}(1-\alpha) + \log(k/m)}{k}} \right],$$

we now have asymptotic $(1-\alpha)$-coverage *uniformly* in $\mathcal{P}$:

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} P\left(\exists k \geqslant m : \mu_P \notin C_k^{(m)} \right) = \alpha.$$

Asymptotic anytime-valid inference in a picture



But how were we able to show that

$$C_k^{(m)} := \left[\widehat{\mu}_k \pm \widehat{\sigma}_k \sqrt{\frac{\Psi^{-1}(1-\alpha) + \log(k/m)}{k}} \right]?$$

satisfies the $\mathcal{P}$-uniform coverage guarantee?

But how were we able to show that

$$C_k^{(m)} := \left[\widehat{\mu}_k \pm \widehat{\sigma}_k \sqrt{\frac{\Psi^{-1}(1-\alpha) + \log(k/m)}{k}}\right]?$$

satisfies the $\mathcal{P}$-uniform coverage guarantee?

After some algebraic juggling, this can be reduced to showing that

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} \sup_{x\geqslant 0} \left| P\left(\sup_{k\geqslant m} \left\{ \frac{k\cdot(\widehat{\mu}_k - \mu_P)^2}{\widehat{\sigma}_k^2} - \log(k/m) \right\} \leqslant x \right) - \Psi(x) \right| = 0.$$

But how were we able to show that

$$C_k^{(m)} := \left[\widehat{\mu}_k \pm \widehat{\sigma}_k \sqrt{\frac{\Psi^{-1}(1-\alpha) + \log(k/m)}{k}} \right]?$$

satisfies the $\mathcal{P}$-uniform coverage guarantee?

After some algebraic juggling, this can be reduced to showing that

$$\sup_{k \geqslant m} \left\{ \text{statistic}_k^{(m)} \right\} \xrightarrow{\mathcal{P}} \Psi(x)$$

But how were we able to show that

$$C_k^{(m)} := \left[\widehat{\mu}_k \pm \widehat{\sigma}_k \sqrt{\frac{\Psi^{-1}(1-\alpha) + \log(k/m)}{k}} \right]?$$

satisfies the $\mathcal{P}$-uniform coverage guarantee?
After some algebraic juggling, this can be reduced to showing that

$$\sup_{k \geqslant m} \left\{ \text{statistic}_k^{(m)} \right\} \xrightarrow{\mathcal{P}} \Psi(x)$$

→ reduces time- and $\mathcal{P}$-uniform coverage / type-I error control to $\mathcal{P}$-*uniform convergence* *in distribution* to this d.f. Ψ.

Ψ: The Robbins-Siegmund distribution

Ψ: The Robbins-Siegmund distribution

Consider the function

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

where Φ is the Gaussian CDF and ϕ is the Gaussian density.

Ψ: The Robbins-Siegmund distribution

Consider the function

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

where Φ is the Gaussian CDF and ϕ is the Gaussian density.

Fact: Ψ *is a continuous CDF on* $x \geqslant 0$*.*

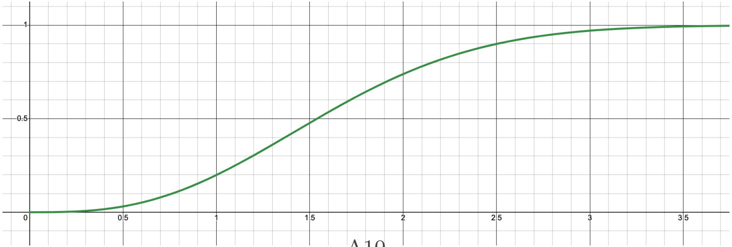
Ψ: The Robbins-Siegmund distribution

Consider the function

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

where Φ is the Gaussian CDF and ϕ is the Gaussian density.

Fact: Ψ *is a continuous CDF on* $x \geqslant 0$.



Where does a distribution like Ψ arise?

Let $(W(t))_{t \geqslant 0}$ be a Wiener process (1-dim Brownian motion).

Where does a distribution like Ψ arise?

Let $(W(t))_{t \geqslant 0}$ be a Wiener process (1-dim Brownian motion).
Take the following transformation:

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

Where does a distribution like Ψ arise?

Let $(W(t))_{t \geqslant 0}$ be a Wiener process (1-dim Brownian motion).
Take the following transformation:

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

Then $\mathbb{P}(\zeta \leqslant x) = \Psi(x)$; $x \geqslant 0$.
(see Robbins & Siegmund (1970) or W-S et al., 2024a).

How can we formally relate the distributions of

$$\sup_{k \geqslant m} \left\{ \frac{k \cdot (\widehat{\mu}_k - \mu)^2}{\widehat{\sigma}_k^2} - \log(k/m) \right\} \quad \text{and} \quad \sup_{t \geqslant 1} \left\{ \frac{W(t)^2}{t} - \log(t) \right\}$$

for large m?

The answer relies on some new strong Gaussian approximation theory.

An overview of this line of work

$\mathcal{P}$-uniform strong laws of large numbers

$\mathcal{P}$-uniform KMT approximation

~~$\mathcal{P}$-uniform asymptotic anytime-valid inference~~✓

Sequential conditional independence testing

$\mathcal{P}$-uniform Komlós-Major-Tusnády (KMT) approximation

(Or strong Gaussian approximation; strong invariance principles)

Nonasymptotic and distribution-uniform Komlós-Major-Tusnády approximation

Ian Waudby-Smith[†], Martin Larsson[‡], and Aaditya Ramdas[‡]

[†]University of California, Berkeley
[‡]Carnegie Mellon University

A crash course on strong Gaussian approximation

A crash course on strong Gaussian approximation

Suppose $(X_n)_{n\in\mathbb{N}}$ are mean-zero i.i.d. on $(\Omega, \mathcal{F}, P)$ with $\mathbb{E}_P|X|^q < \infty$; $q > 2$.
While the CLT allows us to conclude

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i/\sigma \xrightarrow{d} N(0,1),$$

more can be said.

A crash course on strong Gaussian approximation

Suppose $(X_n)_{n\in\mathbb{N}}$ are mean-zero i.i.d. on $(\Omega, \mathcal{F}, P)$ with $\mathbb{E}_P|X|^q < \infty$; $q > 2$.
While the CLT allows us to conclude

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i/\sigma \xrightarrow{d} N(0,1),$$

more can be said.
There exists an implicit $(Y_n)_{n\in\mathbb{N}}$ of i.i.d. $N(0,\sigma^2)$ such that

$$\left| \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} Y_i \right| = o\left(n^{1/q}\right) \quad P\text{-a.s.}$$

(Strassen '64, '67; Komlós-Major-Tusnády '75, '76, Major '76).

A crash course on strong Gaussian approximation

A statement like

$$\left| \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} Y_i \right| = o\left(n^{1/q}\right) \quad P\text{-a.s.}$$

was used in Bibaut et al. (2022) and W-S et al. (2024a) to derive <u>P-pointwise</u> anytime-valid inference guarantees like

$$\forall P \in \mathcal{P}, \quad \lim_{m \to \infty} P\left(\exists k \geqslant m : \mu \notin C_k^{(m)}\right) = \alpha.$$

However, the results of Strassen, KMT, and the surrounding literature is fundamentally P-pointwise.

However, the results of Strassen, KMT, and the surrounding literature is fundamentally P-pointwise.

Let us now generalize the notion of strong approximations to classes of distributions $\mathcal{P}$ and show exactly when $o(n^{1/q})$ rates are attainable.

(Our results are actually *nonasymptotic*, but I will omit those details.)

Definition (W-S et al., 2025): $\mathcal{P}$-uniform strong approximation

We say $\sum_{i=1}^n X_i$ satisfies a $\mathcal{P}$-uniform strong approximation on $(\Omega, \mathcal{F}, P)_{P\in\mathcal{P}}$ with a rate of r_n if

$$\forall \varepsilon > 0, \quad \lim_{m\to\infty} \sup_{P\in\mathcal{P}} P\left(\sup_{k\geqslant m} \left| \frac{\sum_{i=1}^k X_i - \sum_{i=1}^k Y_i}{r_k} \right| \geqslant \varepsilon \right) = 0.$$

Shorthand notation:

$$\sum_{i=1}^n X_i - \sum_{i=1}^n Y_i = \bar{o}_{\mathcal{P}}(r_n).$$

Definition (W-S et al., 2025): $\mathcal{P}$-uniform strong approximation

We say $\sum_{i=1}^{n} X_i$ satisfies a $\mathcal{P}$-uniform strong approximation on $(\Omega, \mathcal{F}, P)_{P \in \mathcal{P}}$ with a rate of r_n if

$$\forall \varepsilon > 0, \quad \lim_{m \to \infty} \sup_{P \in \mathcal{P}} P\left(\sup_{k \geqslant m} \left| \frac{\sum_{i=1}^{k} X_i - \sum_{i=1}^{k} Y_i}{r_k} \right| \geqslant \varepsilon \right) = 0.$$

Shorthand notation:

$$\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} Y_i = \bar{o}_{\mathcal{P}}(r_n).$$

Note when $\mathcal{P} = \{P\}$: $\quad \sup_{k \geqslant m} |Z_k| = o_P(1)$ *if and only if* $Z_n = o(1)$ P-a.s.

Theorem (W-S et al., 2025): $\mathcal{P}$-uniform KMT approximation, Part I.

Suppose $(X_n)_{n\in\mathbb{N}}$ are i.i.d. mean-zero and unit variance on $(\Omega, \mathcal{F}, P)_{P\in\mathcal{P}}$. Then

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} \mathbb{E}_P\left[|X|^q \mathbb{1}\{|X|^q > m\}\right] = 0$$

if and only if

$$\sum_{i=1}^n X_i - \sum_{i=1}^n Y_i = \bar{o}_{\mathcal{P}}(n^{1/q}).$$

Theorem (W-S et al., 2025): $\mathcal{P}$-uniform KMT approximation, Part I.

Suppose $(X_n)_{n\in\mathbb{N}}$ are i.i.d. mean-zero and unit variance on $(\Omega, \mathcal{F}, P)_{P\in\mathcal{P}}$. Then

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} \mathbb{E}_P \left[|X|^q \mathbb{1}\{|X|^q > m\} \right] = 0$$

if and only if

$$\sum_{i=1}^n X_i - \sum_{i=1}^n Y_i = \bar{o}_{\mathcal{P}}(n^{1/q}).$$

Note: Uniform integrability $\iff$ $\sup_{P\in\mathcal{P}} \mathbb{E}_P \left[\varphi(|X|^q) \right] < \infty$ for some $\varphi(x)/x \to \infty$.

Theorem (W-S et al., 2025): $\mathcal{P}$-uniform KMT approximation, Part I.

Suppose $(X_n)_{n\in\mathbb{N}}$ are i.i.d. mean-zero and unit variance on $(\Omega, \mathcal{F}, P)_{P\in\mathcal{P}}$. Then

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} \mathbb{E}_P \left[|X|^q \mathbb{1}\{|X|^q > m\} \right] = 0$$

if and only if

$$\sum_{i=1}^n X_i - \sum_{i=1}^n Y_i = \bar{o}_{\mathcal{P}}(n^{1/q}).$$

Note: Uniform integrability $\iff \sup_{P\in\mathcal{P}} \mathbb{E}_P \left[\varphi(|X|^q) \right] < \infty$ for some $\varphi(x)/x \to \infty$.

Corollary (KMT '76): $\mathbb{E}_P|X|^q < \infty \iff \sum_{i=1}^n X_i - \sum_{i=1}^n Y_i = o(n^{1/q})$ P-a.s.

An overview of this line of work

$\mathcal{P}$-uniform strong laws of large numbers

~~$\mathcal{P}$-uniform KMT approximation~~✓

~~$\mathcal{P}$-uniform asymptotic anytime-valid inference~~✓

Sequential conditional independence testing

An overview of this line of work

$\mathcal{P}$-uniform strong laws of large numbers

~~$\mathcal{P}$-uniform KMT approximation~~✓

~~$\mathcal{P}$-uniform asymptotic anytime-valid inference~~✓

Sequential conditional independence testing

Returning back to the convergence in distribution result,

$$\sup_{k \geqslant m} \left\{ \text{statistic}_k^{(m)} \right\} \xrightarrow{\mathcal{P}} \Psi(x)$$

$\square$

Returning back to the convergence in distribution result,

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} \sup_{x\geqslant 0} \left| P\left(\sup_{k\geqslant m} \left\{ \frac{k\cdot(\widehat{\mu}_k - \mu_P)^2}{\sigma_P^2} - \log(k/m) \right\} \leqslant x \right) - \Psi(x) \right| = 0.$$

$\square$

Returning back to the convergence in distribution result,

$$\lim_{m \to \infty} \sup_{P \in \mathcal{P}} \sup_{x \geqslant 0} \left| P\left(\sup_{k \geqslant m} \left\{ \frac{k \cdot (\widehat{\mu}_k - \mu_P)^2}{\sigma_P^2} - \log(k/m) \right\} \leqslant x \right) - \Psi(x) \right| = 0.$$

The proof proceeds something like this:

$\square$

Returning back to the convergence in distribution result,

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} \sup_{x\geqslant 0} \left| P\left(\sup_{k\geqslant m} \left\{ \frac{k\cdot(\widehat{\mu}_k - \mu_P)^2}{\sigma_P^2} - \log(k/m) \right\} \leqslant x \right) - \Psi(x) \right| = 0.$$

The proof proceeds something like this:

(a) $\widehat{\mu}_k := \frac{1}{k}\sum_{i=1}^{k} X_i \approx \frac{1}{k}\sum_{i=1}^{k} Y_i$ Gaussians uniformly in $\mathcal{P}$ *and* $k \geqslant m$.

$\square$

Returning back to the convergence in distribution result,

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} \sup_{x\geqslant 0} \left| P\left(\sup_{k\geqslant m} \left\{ \frac{k\cdot(\widehat{\mu}_k - \mu_P)^2}{\sigma_P^2} - \log(k/m) \right\} \leqslant x \right) - \Psi(x) \right| = 0.$$

The proof proceeds something like this:

(a) $\widehat{\mu}_k := \frac{1}{k}\sum_{i=1}^k X_i \approx \frac{1}{k}\sum_{i=1}^k Y_i$ Gaussians uniformly in $\mathcal{P}$ *and* $k \geqslant m$.

(b) $\sum_{i=1}^k (Y_i - \mu_P)/\sigma_P \approx W(k)$ for a Wiener process $(W(t))_{t\geqslant 0}$.

$\square$

Returning back to the convergence in distribution result,

$$\lim_{m \to \infty} \sup_{P \in \mathcal{P}} \sup_{x \geqslant 0} \left| P\left(\sup_{k \geqslant m} \left\{ \frac{k \cdot (\widehat{\mu}_k - \mu_P)^2}{\sigma_P^2} - \log(k/m) \right\} \leqslant x \right) - \Psi(x) \right| = 0.$$

The proof proceeds something like this:

(a) $\widehat{\mu}_k := \frac{1}{k} \sum_{i=1}^{k} X_i \approx \frac{1}{k} \sum_{i=1}^{k} Y_i$ Gaussians uniformly in $\mathcal{P}$ *and* $k \geqslant m$.

(b) $\sum_{i=1}^{k} (Y_i - \mu_P)/\sigma_P \approx W(k)$ for a Wiener process $(W(t))_{t \geqslant 0}$.

(c) Scale-invariance: $\sup_{k \geqslant m} \left\{ (W(k))^2/k - \log(k/m) \right\} \overset{d}{=} \sup_{t \geqslant 1} \left\{ (W(t))^2/t - \log(t) \right\}$.

□

Returning back to the convergence in distribution result,

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} \sup_{x\geqslant 0} \left| P\left(\sup_{k\geqslant m} \left\{ \frac{k\cdot(\widehat{\mu}_k - \mu_P)^2}{\sigma_P^2} - \log(k/m) \right\} \leqslant x \right) - \Psi(x) \right| = 0.$$

The proof proceeds something like this:

(a) $\widehat{\mu}_k := \frac{1}{k}\sum_{i=1}^k X_i \approx \frac{1}{k}\sum_{i=1}^k Y_i$ Gaussians uniformly in $\mathcal{P}$ *and* $k \geqslant m$.

(b) $\sum_{i=1}^k (Y_i - \mu_P)/\sigma_P \approx W(k)$ for a Wiener process $(W(t))_{t\geqslant 0}$.

(c) Scale-invariance: $\sup_{k\geqslant m} \left\{(W(k))^2/k - \log(k/m)\right\} \stackrel{d}{=} \sup_{t\geqslant 1} \left\{(W(t))^2/t - \log(t)\right\}$.

(d) Recall that $\zeta := \sup_{t\geqslant 1} \left\{(W(t))^2/t - \log(t)\right\} \sim \Psi$. $\square$

However, there's one last missing piece. The statement

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} \sup_{x\geqslant 0} \left| P\left(\sup_{k\geqslant m} \left\{ \frac{k\cdot(\widehat{\mu}_k - \mu_P)^2}{\widehat{\sigma}_k^2} - \log(k/m) \right\} \leqslant x \right) - \Psi(x) \right| = 0$$

involves $\widehat{\sigma}_k^2$ <u>*not*</u> σ_P^2.

However, there's one last missing piece. The statement

$$\lim_{m \to \infty} \sup_{P \in \mathcal{P}} \sup_{x \geqslant 0} \left| P\left(\sup_{k \geqslant m} \left\{ \frac{k \cdot (\widehat{\mu}_k - \mu_P)^2}{\widehat{\sigma}_k^2} - \log(k/m) \right\} \leqslant x \right) - \Psi(x) \right| = 0$$

involves $\widehat{\sigma}_k^2$ <u>*not*</u> σ_P^2.

In a nutshell, we need that

$$\widehat{\sigma}_n^2 - \sigma_P^2 = o(n^{-\beta})$$

almost surely and $\mathcal{P}$-uniformly for some $\beta > 0$.

An overview of this line of work

$\mathcal{P}$-uniform strong laws of large numbers

~~$\mathcal{P}$-uniform KMT approximation~~✓

~~$\mathcal{P}$-uniform asymptotic anytime-valid inference~~✓

Sequential conditional independence testing

$\mathcal{P}$-uniform strong laws of large numbers

Distribution-uniform strong laws of large numbers

Ian Waudby-Smith[†], Martin Larsson[‡], and Aaditya Ramdas[‡]

[†]University of California, Berkeley
[‡]Carnegie Mellon University

A.N. Kolmogorov's SLLN (1930):

X is integrable $\mathbb{E}_P|X| < \infty$ **if and only if**

$$\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}_P X = o(1) \quad P\text{-a.s.}$$

This is a P-*pointwise* statement.

Kai Lai Chung's SLLN (1951):

*"For use in **statistical** applications, Professor **Wald** raised the question of the uniformity of the SLLN with respect to a family of distributions."*

If X is $\mathcal{P}$-*uniformly integrable*, i.e.

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} \mathbb{E}_P\left\{|X|\mathbb{1}\{|X| > m\}\right\} = 0,$$

Then

$$\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}X = \bar{o}_{\mathcal{P}}(1).$$

However, both Kolmogorov's and Chung's SLLNs say that

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}_P(X)) \to 0 \quad \text{a.s.},$$

whether P-**pointwise** or $\mathcal{P}$-**uniformly**.

However, both Kolmogorov's and Chung's SLLNs say that

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}_P(X)) \to 0 \quad \text{a.s.},$$

whether P-**pointwise** or $\mathcal{P}$-**uniformly**.

(i) Can anything be said about *how fast* this converges?

(ii) If $\mathbb{E}_P|X| = \infty$ (e.g. Cauchy), can we control how fast this *diverges*?

Marcinkiewicz-Zygmund (**M-Z**) SLLN (1937):

$\mathbb{E}_P|X|^q < \infty$; $q \in [1, 2)$ **if and only if**

$$\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}_P X = o\left(n^{1/q-1}\right) \quad P\text{-a.s.}$$

Moreover, for $q \in (0, 1)$, $\mathbb{E}_P|X|^q < \infty$ **if and only if**

$$\frac{1}{n}\sum_{i=1}^{n} X_i = o\left(n^{1/q-1}\right) \quad P\text{-a.s.}$$

Like Kolmogorov's SLLN, this is P-pointwise. What about a $\mathcal{P}$-uniform generalization?

Theorem (W-S et al. 2024c): A $\mathcal{P}$-uniform M-Z SLLN.

For $q \in [1, 2)$,

$$\lim_{m \to \infty} \sup_{P \in \mathcal{P}} \mathbb{E}_P \left[|X - \mathbb{E}_P X|^q \mathbb{1}\{|X - \mathbb{E}_P X|^q \geqslant m\} \right] = 0$$

if and only if

$$\frac{1}{n} \sum_{i=1}^{n} X_i - \mathbb{E} X = \bar{o}_{\mathcal{P}}(n^{1/q - 1}),$$

and similarly for $q \in (0, 1)$ but with $\mathbb{E}_P X$ replaced by 0.

Again, unif. integrability $\iff \sup_{P \in \mathcal{P}} \mathbb{E}_P \left[\varphi(|X - \mathbb{E}_P X|^q) \right] < \infty$ for $\varphi(x)/x \to \infty$.

This unifies the SLLNs of Kolmogorov ('30), M-Z ('37), and Chung ('51).

An overview of this line of work

~~$\mathcal{P}$-uniform strong laws of large numbers~~✓

~~$\mathcal{P}$-uniform KMT approximation~~✓

~~$\mathcal{P}$-uniform asymptotic anytime-valid inference~~✓

Sequential conditional independence testing

An overview of this line of work

~~$\mathcal{P}$-uniform strong laws of large numbers~~✓

~~$\mathcal{P}$-uniform KMT approximation~~✓

~~$\mathcal{P}$-uniform asymptotic anytime-valid inference~~✓

Sequential conditional independence testing

Returning back to the convergence in distribution result,

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} \sup_{x\geqslant 0} \left| P\left(\sup_{k\geqslant m} \left\{ \frac{k\cdot(\widehat{\mu}_k - \mu_P)^2}{\widehat{\sigma}_k^2} - \log(k/m) \right\} \leqslant x \right) - \Psi(x) \right| = 0$$

we have that $\widehat{\sigma}_k^2 - \sigma_P^2 = \bar{o}_{\mathcal{P}}(n^{-\beta})$ as long as $\sup_{P\in\mathcal{P}} \mathbb{E}_P|X - \mu_P|^{2+\delta} < \infty$, for example.

Theorem (W-S et al., 2024b): $\mathcal{P}$-uniform anytime-valid inference.

Suppose $\sup_{P \in \mathcal{P}} \mathbb{E}_P |X - \mu_P|^{2+\delta} < \infty$.

Then for the sequence of intervals

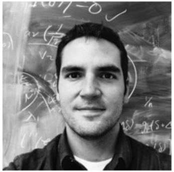
$$C_k^{(m)} := \left[\widehat{\mu}_k \pm \widehat{\sigma}_k \sqrt{\frac{\Psi^{-1}(1-\alpha) + \log(k/m)}{k}} \right],$$

we have

$$\lim_{m \to \infty} \sup_{P \in \mathcal{P}} P\left(\exists k \geqslant m : \mu_P \notin C_k^{(m)} \right) = \alpha.$$

At the risk of oversimplification, this is a drop-in anytime-valid replacement for CLT-based confidence intervals (and tests, etc.).

Sequential conditional independence testing



Edward H. Kennedy

CMU Statistics

Aaditya Ramdas

CMU Statistics

Given i.i.d. $\mathbb{R} \times \mathbb{R} \times \mathbb{R}^d$-valued triplets $(X_n, Y_n, Z_n)_{n \in \mathbb{N}}$, we want to test

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

Given i.i.d. $\mathbb{R} \times \mathbb{R} \times \mathbb{R}^d$-valued triplets $(X_n, Y_n, Z_n)_{n \in \mathbb{N}}$, we want to test

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

Under the so-called **Model-X** assumption (Candès et al. 2018):

$$\left\{ \textbf{Model-X:} \quad X \mid Z \quad \text{is known exactly} \right\},$$

some very satisfying nonasymptotic anytime-valid tests exist.
(Duan-Ramdas-Wasserman '22, Shaer et al. '23, Grünwald-Henzi-Lardy '23).

What about moving beyond **Model-X**?

What about moving beyond **Model-X**?

Paraphrasing a quote from Grünwald, Henzi, and Lardy (2023):

> *"the anytime-valid tests in this paper are highly tailored to the Model-X assumption, and it is an open question to us as to how to construct general sequential tests of conditional independence without it."*

What about moving beyond **Model-X**?

Paraphrasing a quote from Grünwald, Henzi, and Lardy (2023):

> *"the anytime-valid tests in this paper are highly tailored to the Model-X assumption, and it is an open question to us as to how to construct general sequential tests of conditional independence without it."*

We provide one answer to this open question.

The Hardness of Conditional Independence Testing and the Generalised Covariance Measure

Rajen D. Shah*
University of Cambridge, UK
r.shah@statslab.cam.ac.uk

Jonas Peters†
University of Copenhagen, Denmark
jonas.peters@math.ku.dk

March 25, 2022

First, let's discuss their hardness result.

Why do prior works assume that $X \mid Z$ was known (**Model-X**)?

Suppose $(X_i, Y_i, Z_i)_{i=1}^n$ are supported on the unit cube

$$\mathcal{P}^\star = \{\text{dist'ns on } [0,1]^3\}.$$

Why do prior works assume that $X \mid Z$ was known (**Model-X**)?

Suppose $(X_i, Y_i, Z_i)_{i=1}^n$ are supported on the unit cube

$$\mathcal{P}^\star = \{\text{dist'ns on } [0,1]^3\}.$$

We are interested in the null

$$\mathcal{P}_0^\star := \{P \in \mathcal{P}^\star : X \perp\!\!\!\perp_P Y \mid Z\} \quad \text{vs} \quad \mathcal{P}_1^\star := \mathcal{P}^\star \backslash \mathcal{P}_0^\star.$$

Why do prior works assume that $X \mid Z$ was known (**Model-X**)?

Suppose $(X_i, Y_i, Z_i)_{i=1}^n$ are supported on the unit cube

$$\mathcal{P}^\star = \{\text{dist'ns on } [0,1]^3\}.$$

We are interested in the null

$$\mathcal{P}_0^\star := \{P \in \mathcal{P}^\star : X \perp\!\!\!\perp_P Y \mid Z\} \quad \text{vs} \quad \mathcal{P}_1^\star := \mathcal{P}^\star \backslash \mathcal{P}_0^\star.$$

Theorem (Shah & Peters '20): Hardness of cond. independence testing

Let $\dot{\Gamma}_n$ be *any* $\mathcal{P}_0^\star$-uniform hypothesis test.

$$\underbrace{\sup_{P \in \mathcal{P}_1^\star} \lim_{n \to \infty} P\left(\dot{\Gamma}_n \text{ rejects}\right)}_{\text{Best-case power}} \leqslant \underbrace{\lim_{n \to \infty} \sup_{P \in \mathcal{P}_0^\star} P\left(\dot{\Gamma}_n \text{ rejects}\right)}_{\text{Worst-case type-I error}} \leqslant \alpha.$$

Why do prior works assume that $X \mid Z$ was known (**Model-X**)?
Suppose $(X_i, Y_i, Z_i)_{i=1}^n$ are supported on the unit cube

$$\mathcal{P}^\star = \{\text{dist'ns on } [0,1]^3\}.$$

We are interested in the null

$$\mathcal{P}_0^\star := \{P \in \mathcal{P}^\star : X \perp\!\!\!\perp_P Y \mid Z\} \quad \text{vs} \quad \mathcal{P}_1^\star := \mathcal{P}^\star \backslash \mathcal{P}_0^\star.$$

Theorem (Shah & Peters '20): Hardness of cond. independence testing

Let $\dot{\Gamma}_n$ be *any* $\mathcal{P}_0^\star$-uniform hypothesis test.

$$\underbrace{\sup_{P \in \mathcal{P}_1^\star} \lim_{n \to \infty} P\left(\dot{\Gamma}_n \text{ rejects}\right)}_{\text{Best-case power}} \leqslant \underbrace{\lim_{n \to \infty} \sup_{P \in \mathcal{P}_0^\star} P\left(\dot{\Gamma}_n \text{ rejects}\right)}_{\text{Worst-case type-I error}} \leqslant \alpha.$$

In words: "*The **most powerful** uniform test is the one that ignores* $(X_i, Y_i, Z_i)_{i=1}^n$ *and rejects with probability* α."

Anytime-valid conditional independence testing is similarly hard.

Theorem (W-S et al., 2024b): Hardness of anytime-valid cond. indep. testing

Let $\bar{\Gamma}_n^{(m)}$ be *any* $\mathcal{P}_0^\star$-uniform anytime test.

$$\underbrace{\sup_{P \in \mathcal{P}_1^\star} \lim_{m \to \infty} P\left(\exists k \geqslant m : \bar{\Gamma}_k^{(m)} \text{ rejects}\right)}_{\text{Best-case anytime power}} \leqslant \underbrace{\lim_{m \to \infty} \sup_{P \in \mathcal{P}_0^\star} P\left(\exists k \geqslant m : \bar{\Gamma}_k^{(m)} \text{ rejects}\right)}_{\text{Worst-case anytime type-I error}} \leqslant \alpha.$$

Despite their pessimistic result, Shah & Peters show that there is still hope.

Despite their pessimistic result, Shah & Peters show that there is still hope. Using $(X_n, Y_n, Z_n)_{n\in\mathbb{N}}$, define

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

where $\widehat{\mu}_n^x$ and $\widehat{\mu}_n^y$ are estimates of $\mu^x(Z) := \mathbb{E}(X \mid Z)$ and $\mu^y(Z) := \mathbb{E}(Y \mid Z)$.

Despite their pessimistic result, Shah & Peters show that there is still hope. Using $(X_n, Y_n, Z_n)_{n\in\mathbb{N}}$, define

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

where $\widehat{\mu}_n^x$ and $\widehat{\mu}_n^y$ are estimates of $\mu^x(Z) := \mathbb{E}(X \mid Z)$ and $\mu^y(Z) := \mathbb{E}(Y \mid Z)$. Define the *generalized covariance measure* (GCM) statistic:

$$\dot{\mathrm{GCM}}_n := \frac{1}{n}\sum_{i=1}^{n} R_{i,n}/\widehat{\sigma}_n.$$

Intuition: if $\widehat{\mu}_n^x \approx \mu^x$ and $\widehat{\mu}_n^y \approx \mu^y$, then

$$R_{i,n} \approx \{X_i - \mu^x(Z_i)\} \cdot \{Y_i - \mu^y(Z_i))\}, \quad \mathbb{E}_{H_0}(R_{i,n}) = 0.$$

Theorem (Shah & Peters '20): The GCM test

Suppose $\sup_{P\in\mathcal{P}_0} \mathbb{E}_P|\{X_i - \mu^x(Z_i)\}\cdot\{Y_i - \mu^y(Z_i))\}|^{2+\delta} < \infty$ and

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

$$\text{Then} \quad \lim_{n\to\infty} \sup_{P\in\mathcal{P}_0} \sup_{\alpha\in(0,1)} |P\left(p_n \leqslant \alpha\right) - \alpha| = 0,$$

where $p_n := 1 - \Phi(\sqrt{n}\dot{\mathrm{GCM}}_n)$.

In particular, $\dot{\Gamma}_n := \mathbb{1}\{|\sqrt{n}\dot{\mathrm{GCM}}| \geqslant \Phi^{-1}(1-\alpha/2)\}$ is a $\mathcal{P}_0$-uniform level-α test for conditional independence with nontrivial power!

The *sequential* generalized covariance measure (SeqGCM)

Define $\overline{\mathrm{GCM}}_k := \frac{1}{k}\sum_{i=1}^{k}\{X_i - \widehat{\mu}_i^x(Z_i)\} \cdot \{Y_i - \widehat{\mu}_i^y(Z_i))\}/\widehat{\sigma}_k$.

The *sequential* generalized covariance measure (SeqGCM)

Define $\overline{\mathrm{GCM}}_k := \frac{1}{k}\sum_{i=1}^{k}\{X_i - \widehat{\mu}_i^x(Z_i)\} \cdot \{Y_i - \widehat{\mu}_i^y(Z_i))\}/\widehat{\sigma}_k$.

Theorem (W-S et al., 2024b): The SeqGCM test

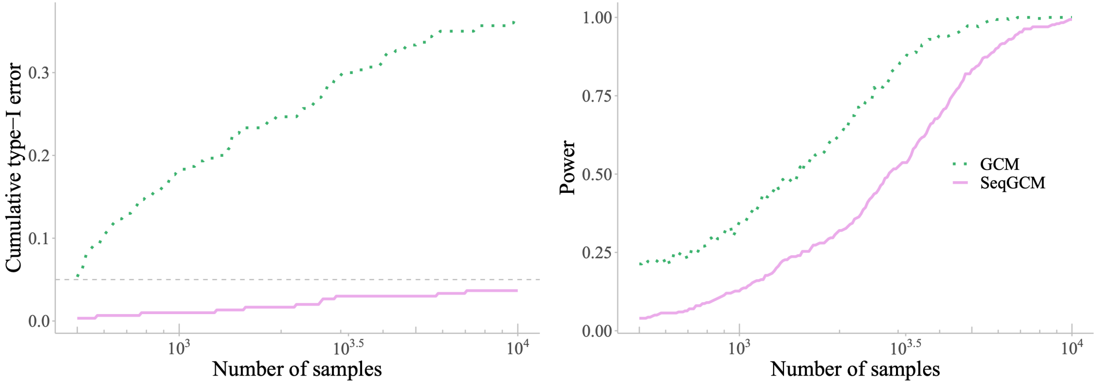
Under the same conditions as the GCM test (Shah & Peters, 2020), suppose

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

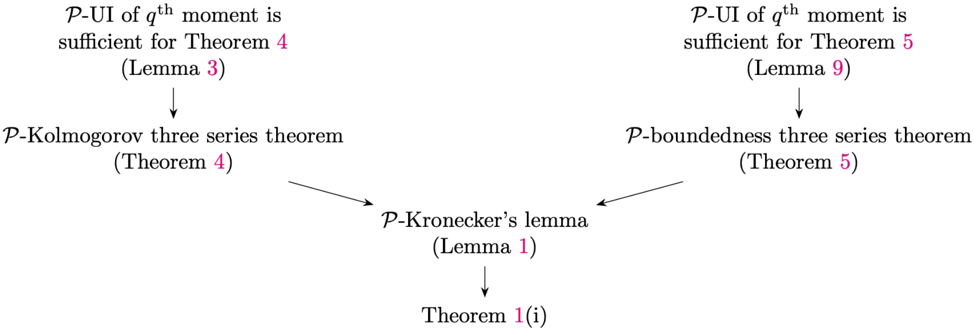
$$\text{Then,} \quad \lim_{m\to\infty} \sup_{P\in\mathcal{P}_0} P\left(\exists k \geqslant m : \overline{p}_k^{(m)} \leqslant \alpha\right) = \alpha,$$

where $\overline{p}_k^{(m)} := 1 - \Psi(k\overline{\mathrm{GCM}}_k^2 - \log(k/m))$.

GCM vs SeqGCM: type-I error and power



	Kolmogorov (1937)	**Chung (1951)**	**M-Z (1937)**	**This work**
$q = 1$	✓	✓	✓	✓
$q \in (0, 2)$			✓	✓
$\mathcal{P}$-uniform		✓		✓
"if and only if"	✓		✓	✓


$\mathcal{P}$-UI of q^{th} moment is sufficient for Theorem 4 (Lemma 3)
$\mathcal{P}$-Kolmogorov three series theorem (Theorem 4)
$\mathcal{P}$-UI of q^{th} moment is sufficient for Theorem 5 (Lemma 9)
$\mathcal{P}$-boundedness three series theorem (Theorem 5)
$\mathcal{P}$-Kronecker's lemma (Lemma 1)
Theorem 1(i)

A word on uniform integrability

Uniform integrability of the q^{th} moment,

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} \mathbb{E}_P\left[|X - \mathbb{E}_P X|^q \mathbb{1}\{|X - \mathbb{E}_P X|^q \geqslant m\}\right] = 0$$

is equivalent to the $\varphi(|\cdot|^q)^{\text{th}}$ moment being $\mathcal{P}$-uniformly bounded:

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P\left[\varphi(|X - \mathbb{E}_P X|^q)\right] < \infty,$$

for some $\varphi(\cdot) \geqslant 0$ so that $\displaystyle\lim_{y\to\infty} \frac{\varphi(y)}{y} = \infty$.

(due to Charles Jean de la Vallée-Poussin)

A $\mathcal{P}$-uniform SLLN for non-identically distributed random variables

Theorem (W-S et al., 2024c): A $\mathcal{P}$-uniform SLLN for **non-iid** RVs

Let $(X_n)_{n\in\mathbb{N}}$ be independent RVs on $(\Omega, \mathcal{F}, P)_{P\in\mathcal{P}}$. Suppose

$$\lim_{m\to\infty} \sup_{P\in\mathcal{P}} \sum_{k=m}^{\infty} \frac{\mathbb{E}_P|X_k - \mathbb{E}_P X_k|^q}{a_k^q} = 0$$

for some $a_n \nearrow \infty$ and some $q \in [1, 2]$. Then,

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}X_i) = \bar{o}_{\mathcal{P}}(a_n/n),$$

This is a $\mathcal{P}$-uniform generalization of the usual independent SLLN (see §IX of Petrov (1975)).