Extending Subsampling to Sequential Stopping

Wenhao Yang

School of Mathematical Sciences, Peking University

Joint work with Jose Blanchet and Peter Glynn

INFORMS 2026

Warm-up: i.i.d. mean estimation

From stochastic simulation or statistical sampling:

$$X_1, X_2, \ldots \overset{\text{i.i.d.}}{\sim} P, \qquad \mu = \mathbb{E}[X_1].$$

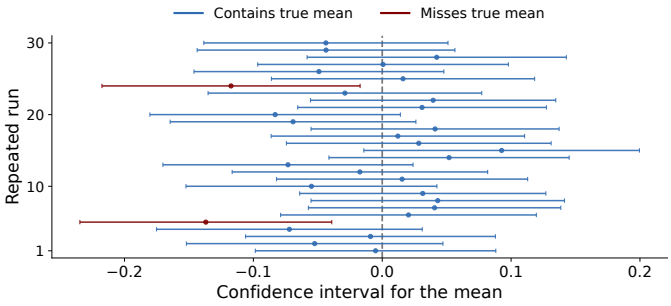
Estimate μ by

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

How accurate is $\bar{X}_n$ after a predetermined n?

Mean estimation setup from AoS Section 2. Assume the mean exists. The next slide starts with the finite-variance case.

Confidence intervals at a fixed sample size



Finite-variance reference

$$\bar{X}_n \pm 1.96 \frac{\widehat{\sigma}_n}{\sqrt{n}}.$$

Here $\widehat{\sigma}_n$ is the sample standard deviation.

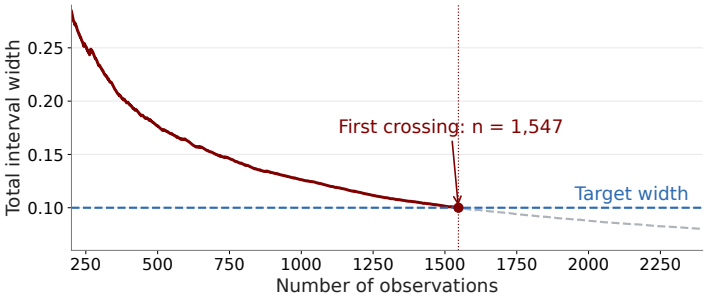
At a fixed budget, the final interval width is random.

Illustration: 30 independent Gaussian samples, $n = 400$, normal-reference intervals using the sample standard deviation. Dashed line: true mean.

Extending to sequential stopping

Choose a target width ε, then let the simulation decide its run length:

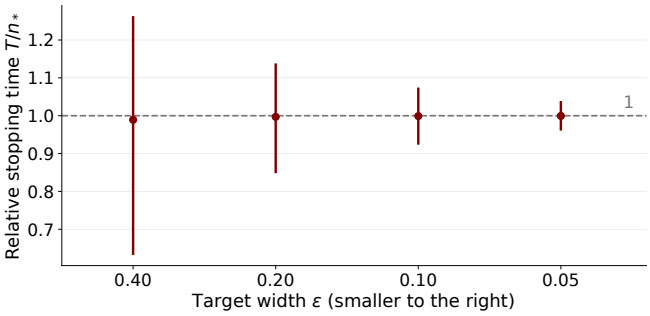
$$T(\varepsilon) = \inf\{n \geq n_0(\varepsilon) : \text{width}(\text{CI}_n) < \varepsilon\}.$$



New question: does the interval still have the target coverage at $T(\varepsilon)$?

Illustrative Gaussian stream. Target total width .10. Gray dashed line: continuation unobserved after stopping.

Why stopping works with finite variance



The scale stabilizes

$$\widehat{\sigma}_n \longrightarrow \sigma.$$

A predictable run length

$$n_*(\varepsilon) = \left(\frac{2 \times 1.96\sigma}{\varepsilon}\right)^2,$$

$$\frac{T(\varepsilon)}{n_*(\varepsilon)} \longrightarrow 1.$$

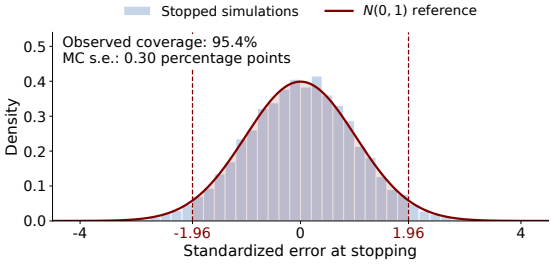
Relative fluctuations in the stopping time vanish as $\varepsilon \downarrow 0$.

Glynn and Whitt (1992). I.i.d. finite nonzero variance, ordinary sample SD. Minimum length: $n_0(\varepsilon) \to \infty$, $\varepsilon^2 n_0(\varepsilon) \to 0$.
Plot: 5,000 Gaussian runs per width; lines: central 95%, dots: medians.

Asymptotically valid coverage

The standardized error keeps its normal limit at termination:

$$\frac{\sqrt{T(\varepsilon)}(\bar{X}_{T(\varepsilon)} - \mu)}{\widehat{\sigma}_{T(\varepsilon)}} \Rightarrow N(0,1).$$



The terminal interval

$$\bar{X}_{T(\varepsilon)} \pm \frac{\varepsilon}{2}$$

has coverage

$$\mathbb{P}\{\mu \in \mathrm{CI}_{T(\varepsilon)}\} \to .95$$

as $\varepsilon \downarrow 0$.

Glynn and Whitt (1992), under the preceding i.i.d. conditions. Illustration: 5,000 stopped Gaussian runs, target width $.05$, $n_0(\varepsilon) = \max\{2, \lceil 2/\varepsilon \rceil\}$. Dashed lines: asymptotic reference quantiles ± 1.96.

Sum-up: the finite-variance case

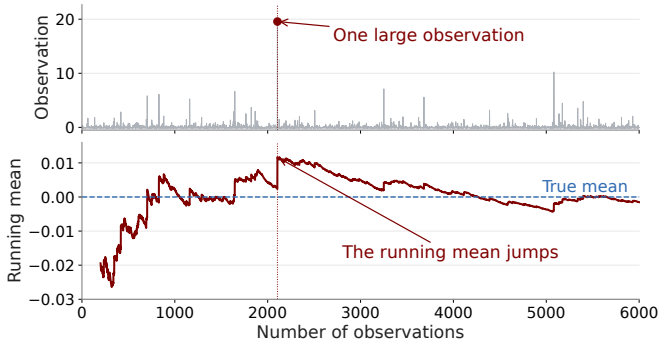
For i.i.d. observations with $0 < \sigma^2 < \infty$:

	Fixed sample size	**Sequential stopping**
Choose first	Sample size n	Target width ε
Run length	Fixed n	Random $T(\varepsilon)$
Reported interval	$\bar{X}_n \pm 1.96 \dfrac{\widehat{\sigma}_n}{\sqrt{n}}$	$\bar{X}_{T(\varepsilon)} \pm \dfrac{\varepsilon}{2}$
Interval width	Random	Exactly ε
Coverage limit	95% as $n \to \infty$	95% as $\varepsilon \downarrow 0$

Classical i.i.d. finite-variance setting. Sequential column reports the target-width interval from the preceding slide. Ordinary sample SD; $n_0(\varepsilon) \to \infty$, $\varepsilon^2 n_0(\varepsilon) \to 0$. Glynn and Whitt (1992).

A rare observation can move the running mean

Pareto tails with $1 < \alpha < 2$: finite mean, infinite variance.



Rare observations affect both the estimation error and the stopping decision.

Illustrative centered Pareto stream, tail index 1.8, noise scale 0.1. This is a simulated example, not a paper experiment.

A stable CLT for heavy-tailed means

For i.i.d. observations with $\mathbb{E}[X_i] = \mu$:

Finite variance

$0 < \mathrm{Var}(X_1) = \sigma^2 < \infty.$

$$\sqrt{n}\,(\bar{X}_n - \mu) \Rightarrow N(0, \sigma^2).$$

The error has scale $n^{-1/2}$.

Infinite variance

Assume stable domain of attraction, with $1 < \alpha < 2$.

$$n^{1-1/\alpha}\ell(n)(\bar{X}_n - \mu) \Rightarrow Z_\alpha.$$

Z_α is α-stable. $\ell(n)$ is slowly varying.

Pareto example, $\alpha = 1.8$: the error has scale $n^{-4/9}$.

Both the convergence rate and the limiting quantiles change.

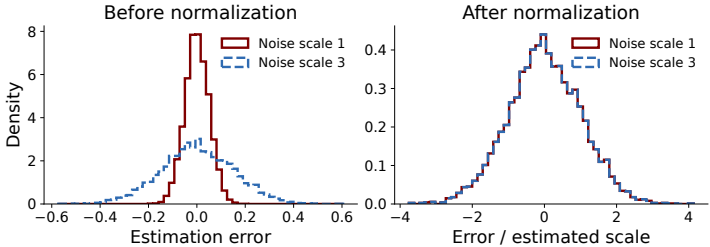
AoS Section 2, stable domain-of-attraction condition and stable CLT. For pure Pareto tails, constants can be absorbed into Z_α. The stable limit's scale and skewness depend on the observation law.

Self-normalization removes an unknown scale

Estimating α, $\ell(n)$ and the parameters of Z_α is difficult.

Use the sample standard deviation $\widehat{\sigma}_n$ to self-normalize:

$$R_n = \frac{\sqrt{n}\,(\bar{X}_n - \mu)}{\widehat{\sigma}_n}.$$



The self-normalized limit still needs the right quantiles.

Motivation: AoS Section 2. $\widehat{\sigma}_n$ is the sample standard deviation. Gaussian illustration at fixed $n = 400$, with noise scales 1 and 3. This illustrates scale cancellation, not universal normality.

Fixed-time and stopped limits

Fixed sample size

$$R_n \xrightarrow{d} W_\alpha, \qquad n \to \infty.$$

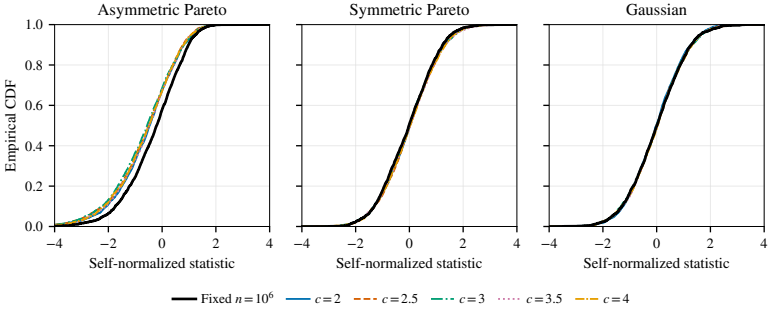
At the stopping time

$$R_{T(\varepsilon)} = \frac{\sqrt{T(\varepsilon)}\big(\bar{X}_{T(\varepsilon)} - \mu\big)}{\widehat{\sigma}_{T(\varepsilon)}} \xrightarrow{d} W_{\mathrm{stop}}, \qquad \varepsilon \downarrow 0.$$

The fixed-time and stopped limiting laws can differ.

AoS Section 2 and Theorem 3.2, under the stated stable-domain, joint process-limit and stopping regularity conditions. $T(\varepsilon)$ uses a fixed positive width multiplier and an admissible minimum run length.

Fixed and stopped distributions can differ



- Asymmetric Pareto: the fixed and stopped curves differ clearly.
- The five asymmetric comparisons all reject equality in the paper's KS tests.

AoS Figure 3 and Table 1. Fixed $n = 10^6$; stopped $\varepsilon = .01$; 2,000 replications per group. Gaussian and symmetric comparisons do not reject.

The key observation

For any fixed $c > 0$:

$$T_c(\varepsilon) = \inf\left\{ n \geq n_0(\varepsilon) : c\,\frac{\widehat{\sigma}_n}{\sqrt{n}} < \varepsilon \right\}.$$

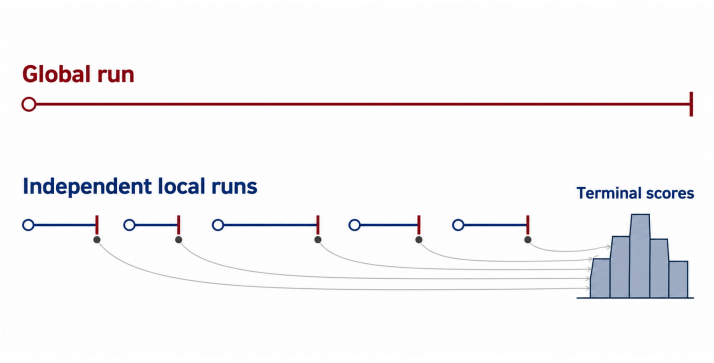
irrelevant

$$R_{T_c(\varepsilon)} \xrightarrow{d} W_{\text{stop}}, \qquad \varepsilon \downarrow 0.$$

Shorter stopped runs can learn the same terminal quantiles.

AoS Theorem 3.2, under joint self-similarity and the stated stopping regularity conditions. The coefficient changes the run length but not the limiting terminal distribution.

The key idea: calibration runs must also stop



- Run many independent, shorter versions of the stopping procedure.
- Collect their standardized errors at their own termination times.

Conceptual illustration. Independent local simulations restart from fresh streams alongside the global simulation.

Procedure summary

Global target: ε. Local tolerance: ε^r, $0 < r < 1$.

1. **Simulate.**
   Advance the global run and one independent, coarser local run.

2. **Update calibration.**
   At each local stop, save its terminal mean and scale, then restart.
   Update terminal quantiles.

3. **Stop at the requested precision.**
   End the global run when its calibrated interval width is below ε.

AoS Algorithm 1. Local runs use coefficient 1.

What the procedure delivers

The requested precision at termination

$$\text{total interval width} < \varepsilon.$$

Coverage calibrated for that termination time

$$\mathbb{P}\{\mu \in \widehat{\mathrm{CI}}(\varepsilon)\} \longrightarrow 1 - \delta \qquad \text{as } \varepsilon \downarrow 0.$$

- **Model-agnostic:** the same procedure applies directly to the finite-variance case.

AoS Theorem 3.3. Coverage is asymptotic. Finite-accuracy performance depends on the model, scaling and tuning.

Application: stopping averaged SGD

One-dimensional quadratic objective

$$\theta_k = \theta_{k-1} - \eta_k g_k, \qquad g_k = H(\theta_{k-1} - \theta^*) + \xi_k.$$

Polyak average

$$\bar{\theta}_n = \frac{1}{n}\sum_{k=1}^{n} \theta_k.$$

Mean squared gradient

$$\widehat{v}_n = \frac{1}{n}\sum_{k=1}^{n} g_k^2.$$

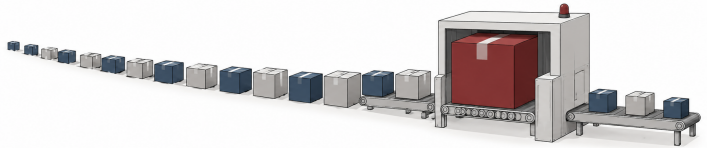
Self-normalization at termination: $T = T(\varepsilon)$

$$\frac{\sqrt{T}\,(\bar{\theta}_T - \theta^*)}{\sqrt{\widehat{v}_T}} \xrightarrow{d} W_{\text{stop}}^{\text{SGD}}, \qquad \varepsilon \downarrow 0.$$

Our procedure chooses $T(\varepsilon)$**:** the calibrated interval for θ^* has width below ε.

AoS Section 5.2 and Theorems 3.2–3.3. I.i.d. mean-zero stable-domain noise, $1 < \alpha < 2$; $H > 0$, $\eta_k = c_\eta k^{-\varrho}$, $1/\alpha < \varrho < 1$, $0 < c_\eta H < 1$.

One large service time creates many long waits



- A service time of order x creates an excursion with height and duration of order x.
- Its contribution to cumulative delay is therefore of order x^2.

Queue mechanism from AoS Section 5.3. Image is a conceptual illustration, not simulation output.

The usual std normalizer can fail

Light-traffic M/G/1 with Pareto service

Fixed load $\rho = 0.1$, service tail index $\alpha = 3.6$. Waiting times W_i have finite variance.

Mean error

$$n^{4/9}(\bar{W}_n - \mu_W) \xrightarrow{d} Z_{1.8}.$$

The error fluctuates on scale

$$n^{-4/9}.$$

Sample standard deviation

$$\widehat{\sigma}_n \xrightarrow{p} \sigma_W > 0.$$

$$\underbrace{\frac{\widehat{\sigma}_n}{\sqrt{n}}}_{\text{usual i.i.d. error proxy}} \sim \sigma_W n^{-1/2}.$$

The two rates do not match: $n^{-4/9}$ versus $n^{-1/2}$.

Analytic example at fixed low utilization. Rates from AoS Section 5.3 and the queue scaling discussion. For pure Pareto service, constants are absorbed into the stable limit $Z_{1.8}$.

Sectioning

Compare m independent cumulative paths $S_j(t)$ with their average $\bar{S}(t)$.

$$V_{\text{sec}}(t) = \frac{1}{m-1} \sum_{j=1}^{m} \left[S_j(t) - \bar{S}(t) \right]^2.$$

Self-normalizing scale for $\bar{S}(t)/t$:

$$\frac{\sqrt{V_{\text{sec}}(t)}}{t}.$$

Sectioning: variation across independent simulation runs. AoS Section 4.

Coverage: Std vs Sectioning

Model	Std (%)	Sectioning (%)
i.i.d. symmetric Pareto	92.4	91.2
i.i.d. asymmetric Pareto	94.6	96.0
i.i.d. Gaussian	95.8	93.0
SGD, symmetric noise	96.0	92.2
SGD, asymmetric noise	96.6	92.4
M/G/1 queue	90.8	93.2

AoS Tables 2, 4 and 6. $r = .50$, width $.01$, 500 runs, target 95%. Std: SoS, using gradient squares for SGD. Sectioning: $m = 5$. All displayed runs terminated. Queue Std is diagnostic.

Takeaways

- **A model-agnostic framework for sequential stopping**
- **Alternative normalizers when standard deviation fails**
- **Applications to stochastic approximation and queueing systems**

Thank you!