arxiv.org/abs/2507.17831

On Focusing Statistical Power for Searches and Measurements in Particle Physics

James Carzon[1], Aishik Ghosh[2,3], Rafael Izbicki[4], Ann B Lee[1], Luca Masserano[1], Daniel Whiteson[2]

[1]Department of Statistics and Data Science, Carnegie Mellon University
[2]Department of Physics and Astronomy, University of California Irvine
[3]Physics Division, Lawrence Berkeley National Laboratory
[4]Department of Statistics, Federal University of São Carlos

What? Shorter confidence intervals than with likelihood ratio tests
Why? Accelerated time-to-discovery; precise physical measurements
How? Auxilliary knowledge via *focus functions*

Searches and Measurements

The binned analysis set-up

- Likelihoods in particle physics are often resolved as histograms:

$$p(\mathcal{D};\mu,\boldsymbol{\alpha},\boldsymbol{\beta}) = \prod_{i=1}^{d} f(N_i;\mu\alpha_i+\beta_i)$$

for *signal strength* $\mu > 0$, Poisson PMF $f(N;\lambda)$.

Neyman construction of confidence intervals

- Searches and measurements in particle physics are often performed as simple-versus-composite hypothesis tests,

$$H_0 : \mu = \mu_0, \qquad \text{vs.} \qquad H_1 : \mu \neq \mu_0. \qquad (\mu_0 \in \Theta)$$

- The Neyman construction of a $(1-\alpha)100\%$ confidence interval compares e.g. the generalized likelihood ratio statistic (LRS),

$$T(\mathcal{D};\mu_0) = -2\log\left(\frac{p(\mathcal{D};\mu_0)}{\sup_{\mu\in\Theta} p(\mathcal{D};\mu)}\right), \qquad (1)$$

against sampling quantiles, e.g. via Wilks' theorem,

$$\hat{C}_{\mu_0} = \chi^2_{1-\alpha}(\mathrm{df} = \dim(\Theta) - \dim(\Theta_0)).$$

- The test is evaluated across Θ:

$$I(\mathcal{D}) = \{\mu_0 \in \Theta : T(\mathcal{D};\mu_0) < \hat{C}_{\mu_0}\}.$$

Upshot:
- The LRS is *not optimal* in general for interval length.
- Without Wilks', calibration via Monte Carlo (MC) is *expensive*.

Focusing Statistical Power

I. The focused test statistic (FTS)

Idea: Incorporate auxilliary knowledge into the test statistic,

$$T_f(\mathcal{D};\mu_0) := -2\log\left(\frac{p(\mathcal{D};\mu_0)}{\left[\int f(d\mu)p(\mathcal{D};\mu)\right]}\right) \qquad (2)$$

- *Focus function*, f: A way of specifying regions of the parameter space where test power should be focused.
- We use Gaussian focus functions, $f(\mu) = N(\mu; m, s)$, truncated to Θ.

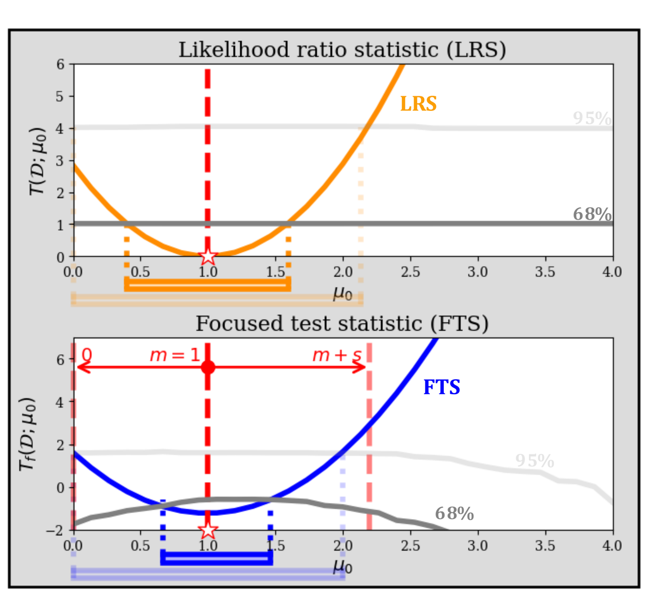

Fig 1. Construction of intervals

II. Quantile regression (QR)-based calibration

Idea: Estimate critical value surface with machine learning

- QR with 9,000 pseudo-experiments (PEs) is competitive with MC with 10^6 PEs.
- For FTS, QR is again more sample-efficient.

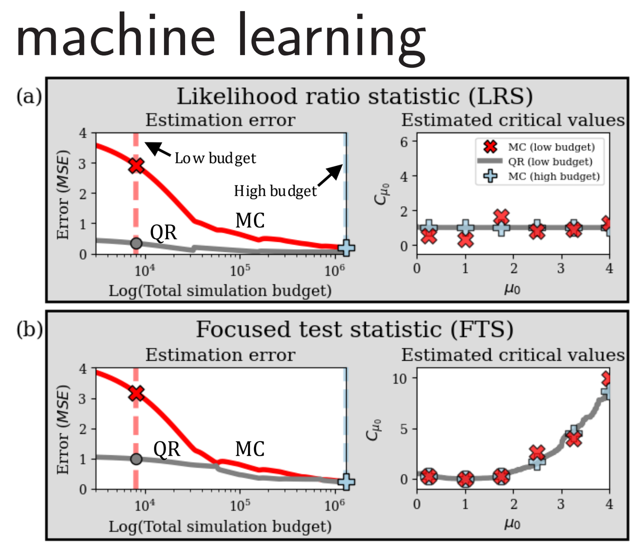

Fig 2. MSE of critical value estimates via MC (red) and quantile regression (QR, grey) on 300 test points.

Case Studies

I. Measurement of Higgs coupling via its signal strength

- **Goal:** Measure Higgs boson-to-tau lepton coupling (one-to-one with μ)
- We use the 2014 HiggsML benchmark data [1]
- We compare

$$f_{\text{wide}} = N(1.0, 2.4),$$
$$f_{\text{narrow}} = N(1.0, 1.2)$$

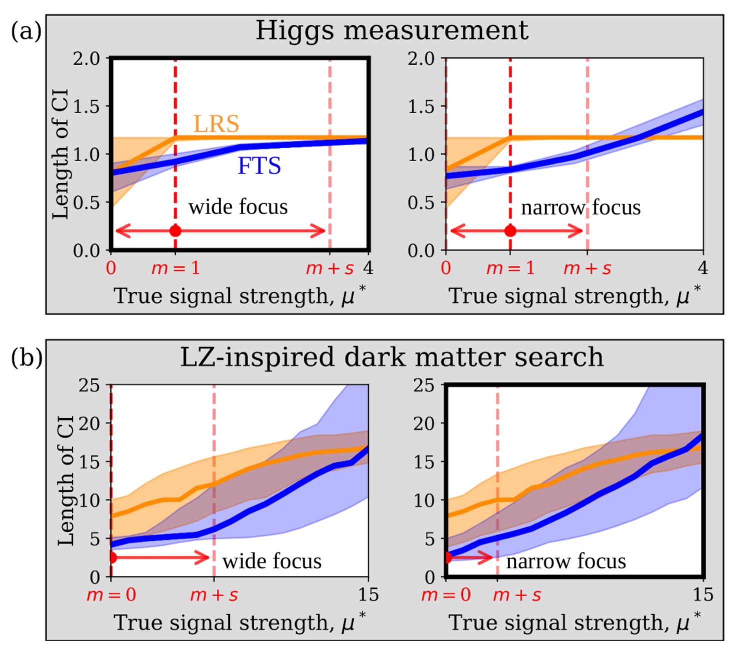

Fig 3. Median interval lengths

II. Search for weakly interacting massive particles (WIMPs)

- **Goal:** Search for dark matter based on LZ experiment [2]
- Synthetic data mimics LZ data release with NEST code [3]
- $\mu^* = 1 \to 1:1{,}200$ signal-background ratio
- Default hypothesis: $\mu^* = 0$

$$f_{\text{wide}} = N(0, 6.0),$$
$$f_{\text{narrow}} = N(0, 3.0)$$

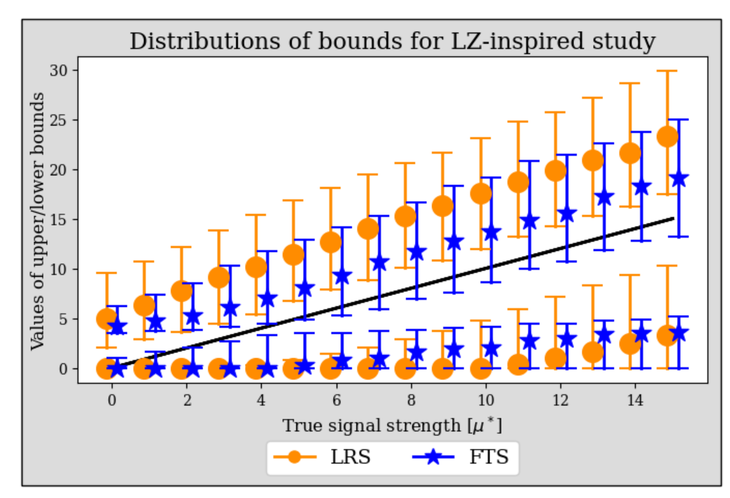

Fig 4. Distributions of 68% confidence interval upper and lower bounds

Main finding

- Shorter interval lengths with FTS compared to LRS

Setting		Test statistic		
Experiment	μ^*	LRS	FTS-wide	FTS-narrow
Higgs (mass)	1.0	1.08 (2.01)	0.94 (1.89)	**0.85 (1.79)**
Higgs (vis. mass)	1.0	1.26 (2.31)	1.10 (2.17)	**0.98 (2.08)**
LZ-inspired	0.0	5.99 (13.87)	4.68 (11.47)	**3.89 (11.29)**
LZ-inspired	1.0	7.08 (15.18)	5.29 (12.82)	**4.68 (12.75)**

Next Steps

I. Nuisance constraint sets

Idea: Construct $S_\gamma(\mathcal{D};\mu_0)$, a valid $(1-\gamma)$-level confidence set for ν at fixed $\mu = \mu_0$ for pre-specified $\gamma \in (0,\alpha)$ [4, 5].

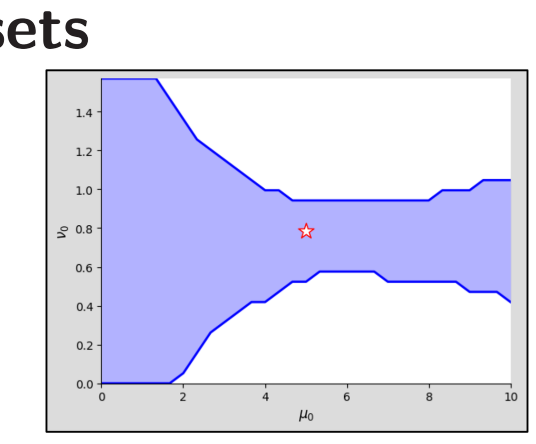
Fig 5. An example of a nuisance constraint set.

II. Rejection probability function

Idea: Estimate CDF of statistic for amortized testing:

$$F(t;\mu,\nu) = \mathbb{P}(T(\mathcal{D};\mu,\nu) \geq t)$$

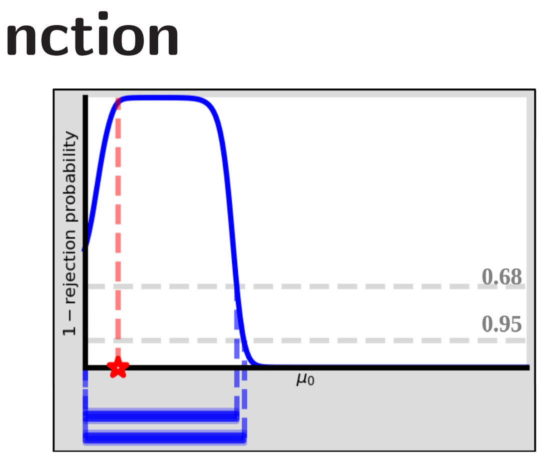
Fig 6. 1−probability of rejection under a statistic T.

III. Fast, nuisance-aware confidence intervals

$$I(D) = \{\mu_0 : \sup_{\nu\in S_\gamma(D;\mu_0)} F(T(D;\mu_0,\nu)) > \alpha - \gamma\}$$

References

(1) ATLAS Collaboration, *Dataset from the ATLAS Higgs Boson Machine Learning Challenge*, 2014.
(2) J. Aalbers et al., *Phys. Rev. Lett.*, 2025, **135**, 011802.
(3) J. Brodsky et al., *NESTCollaboration/nest: Geant4 Integration Fixes and Updates*, 2019.
(4) L. Masserano et al., Proceedings of the 41st International Conference on Machine Learning, ed. R. Salakhutdinov et al., PMLR, 2024, vol. 235, pp. 34987–35012.
(5) M. Stanley et al., *Confidence intervals for functionals in constrained inverse problems via data-adaptive sampling-based calibration*, en, arXiv:2502.02674 [stat], 2025.

This material was prepared using Bridges-2 at PSC through allocation MTH250002 from the ACCESS program and was partially supported by the U.S. National Science Foundation under Grant No. DMS-2053804.

Code: github.com/lee-group-cmu/focus-paper

Contact: jcarzon@andrew.cmu.edu