FERMILAB-SLIDES-25-0141-ETD

Unaccounted-for look-elsewhere effect in k-fold cross adaptive anomaly searches

Prasanth Shyamsundar, Fermi National Accelerator Laboratory

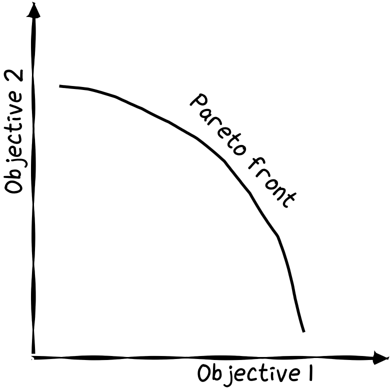
Prasanth Shyamsundar
Fermilab

Nicholas Smith
Fermilab

Manuel Szewc
University of Cincinnati

Breadth–depth tradeoff

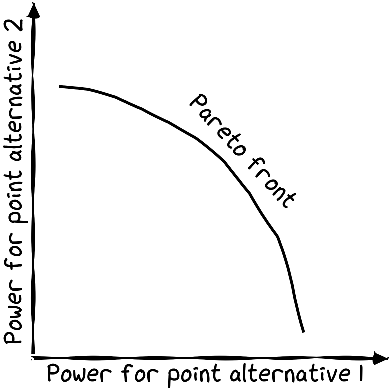
A search focusing on a narrow class of alternative hypotheses will be more sensitive for that class, than a wider search.



This tradeoff is an unavoidable aspect of anomaly detection.

Breadth–depth tradeoff

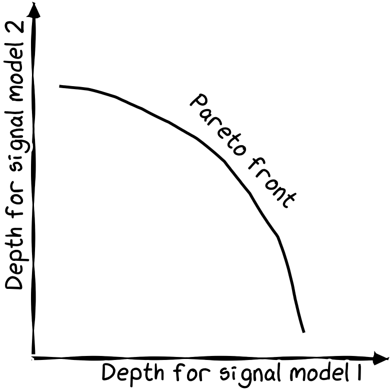
A search focusing on a narrow class of alternative hypotheses will be more sensitive for that class, than a wider search.



This tradeoff is an unavoidable aspect of anomaly detection.

Breadth–depth tradeoff

A search focusing on a narrow class of alternative hypotheses will be more sensitive for that class, than a wider search.



This tradeoff is an unavoidable aspect of anomaly detection.

Adaptive searches and the look-elsewhere effect

- Adaptation stage: $D \longmapsto \Lambda$

  Inference stage: $D \overset{\Lambda}{\longmapsto}$ test statistic, p-value, etc.
- Examples:
  - Likelihood ratio test statistic

$$T \equiv \frac{\sup\limits_{\theta\in\Omega} \mathcal{L}(\theta\,;\,D)}{\mathcal{L}(\theta_0\,;\,D)} = \frac{\mathcal{L}(\Lambda\,;\,D)}{\mathcal{L}(\theta_0\,;\,D)}\,, \qquad \text{where} \quad \Lambda = \arg\sup_{\theta\in\Omega} \mathcal{L}(D\,;\,\theta)$$

  - Largest discrepancy statistic

$$T \equiv \max_b \left[\frac{o_b - e_b}{\sqrt{e_b}}\right] = \left[\frac{o_\Lambda - e_\Lambda}{\sqrt{e_\Lambda}}\right]\,, \qquad \text{where} \quad \Lambda = \arg\max_b \left[\frac{o_b - e_b}{\sqrt{e_b}}\right]$$

  - Train a neural network Λ using observed data. Perform an analysis on the same data with Λ.

Adaptive searches and the look-elsewhere effect (LEE)

- **Local p-value:** p-value computed assuming that Λ was a priori fixed, and wasn't computed from data.

$$\text{Prob}_{\text{null}}\left[\rho_{\text{local}}(D\,;\;\lambda') \leq \alpha\right] \;\leq\; \alpha\,, \qquad \forall\, 0 \leq \alpha \leq 1\,, \forall\, \lambda'\,.$$

- **Global p-value:** Actually valid p-value.

$$\text{Prob}_{\text{null}}\left[\rho_{\text{global}}(D\,;\;\Lambda(D)) \leq \alpha\right] \;\leq\; \alpha\,, \qquad \forall\, 0 \leq \alpha \leq 1\,.$$

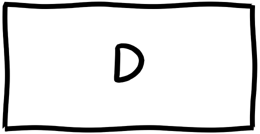
- **Look-elsewhere effect:** Local p-values not being valid global p-values.
- **Unaccounted-for LEE:** When a local p-value is reported as a global one.

Claim: There is a LEE in k-fold cross adaptive searches

(**Opinion:** An adaptive search not having LEE is the special case. Not the other way around.)

Two-stage adaptive search

- If $\Gamma(D)$ and $\Lambda(D)$ are independent and $\rho_{\text{local}}(\Gamma(D)\ ;\ \Lambda(D))$ is a valid local p-value, then it is a valid global p-value.
- This leads to the two-stage adaptive search:



- Does not overcome the breadth-depth tradeoff.
- Helps avoid computing of a trials factor or performing pseudo-expts to estimate p-values. **Hard to quantify the LEE of a neural network.**

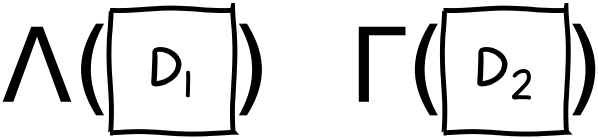
Two-stage adaptive search

- If $\Gamma(D)$ and $\Lambda(D)$ are independent and $\rho_{\text{local}}(\Gamma(D)\ ;\ \Lambda(D))$ is a valid local p-value, then it is a valid global p-value.
- This leads to the two-stage adaptive search:

- Does not overcome the breadth-depth tradeoff.
- Helps avoid computing of a trials factor or performing pseudo-expts to estimate p-values. **Hard to quantify the LEE of a neural network.**

Two-stage adaptive search

- If $\Gamma(D)$ and $\Lambda(D)$ are independent and $\rho_{\text{local}}(\Gamma(D)\ ;\ \Lambda(D))$ is a valid local p-value, then it is a valid global p-value.
- This leads to the two-stage adaptive search:

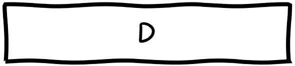


- Does not overcome the breadth-depth tradeoff.
- Helps avoid computing of a trials factor or performing pseudo-expts to estimate p-values. **Hard to quantify the LEE of a neural network.**

Fermilab

k-fold cross adaptive search (1805.02664 [hep-ph], 1902.02634 [hep-ph])

- Split datasets into k independent subsets.
- Train k different NNs $\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k}$, each time leaving one datasubset out.
- "Test" each NN on the corresponding holdout set.
- **Combine into one analysis, in some manner.**
- Rationale: LEE free search without sacrificing event-count (p-values are computed assuming that $\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k}$ are a priori fixed).

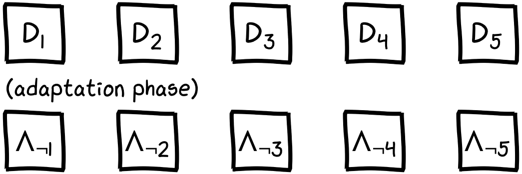


k-fold cross adaptive search (1805.02664 [hep-ph], 1902.02634 [hep-ph])

- Split datasets into k independent subsets.
- Train k different NNs $\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k}$, each time leaving one datasubset out.
- "Test" each NN on the corresponding holdout set.
- **Combine into one analysis, in some manner.**
- Rationale: LEE free search without sacrificing event-count (p-values are computed assuming that $\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k}$ are a priori fixed).

k-fold cross adaptive search (1805.02664 [hep-ph], 1902.02634 [hep-ph])

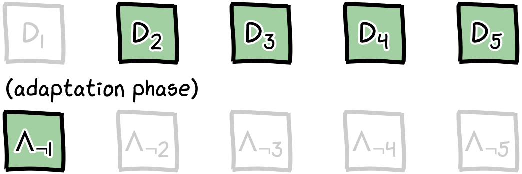
- Split datasets into k independent subsets.
- Train k different NNs $\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k}$, each time leaving one datasubset out.
- "Test" each NN on the corresponding holdout set.
- **Combine into one analysis, in some manner.**
- Rationale: LEE free search without sacrificing event-count (p-values are computed assuming that $\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k}$ are a priori fixed).



Fermilab

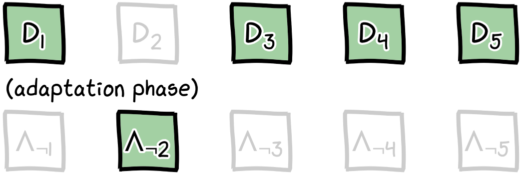
k-fold cross adaptive search (1805.02664 [hep-ph], 1902.02634 [hep-ph])

- Split datasets into k independent subsets.
- Train k different NNs $\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k}$, each time leaving one datasubset out.
- "Test" each NN on the corresponding holdout set.
- **Combine into one analysis, in some manner.**
- Rationale: LEE free search without sacrificing event-count (p-values are computed assuming that $\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k}$ are a priori fixed).



k-fold cross adaptive search (1805.02664 [hep-ph], 1902.02634 [hep-ph])

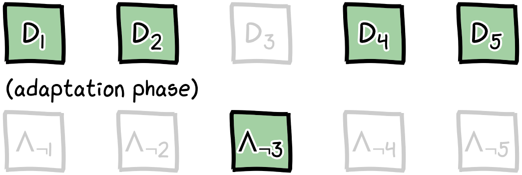
- Split datasets into k independent subsets.
- Train k different NNs $\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k}$, each time leaving one datasubset out.
- "Test" each NN on the corresponding holdout set.
- **Combine into one analysis, in some manner.**
- Rationale: LEE free search without sacrificing event-count (p-values are computed assuming that $\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k}$ are a priori fixed).



Fermilab

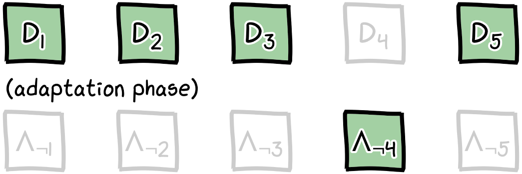
k-fold cross adaptive search (1805.02664 [hep-ph], 1902.02634 [hep-ph])

- Split datasets into k independent subsets.
- Train k different NNs $\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k}$, each time leaving one datasubset out.
- "Test" each NN on the corresponding holdout set.
- **Combine into one analysis, in some manner.**
- Rationale: LEE free search without sacrificing event-count (p-values are computed assuming that $\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k}$ are a priori fixed).



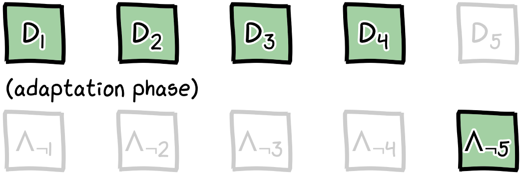
k-fold cross adaptive search (1805.02664 [hep-ph], 1902.02634 [hep-ph])

- Split datasets into k independent subsets.
- Train k different NNs $\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k}$, each time leaving one datasubset out.
- "Test" each NN on the corresponding holdout set.
- **Combine into one analysis, in some manner.**
- Rationale: LEE free search without sacrificing event-count (p-values are computed assuming that $\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k}$ are a priori fixed).



k-fold cross adaptive search (1805.02664 [hep-ph], 1902.02634 [hep-ph])

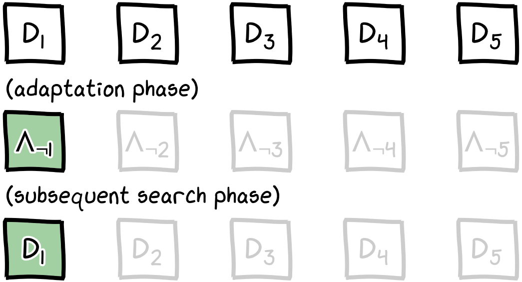
- Split datasets into k independent subsets.
- Train k different NNs $\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k}$, each time leaving one datasubset out.
- "Test" each NN on the corresponding holdout set.
- **Combine into one analysis, in some manner.**
- Rationale: LEE free search without sacrificing event-count (p-values are computed assuming that $\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k}$ are a priori fixed).



Fermilab

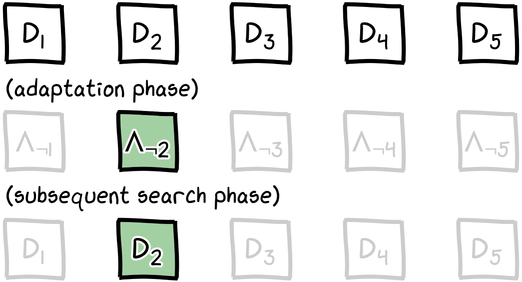
k-fold cross adaptive search (1805.02664 [hep-ph], 1902.02634 [hep-ph])

- Split datasets into k independent subsets.
- Train k different NNs $\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k}$, each time leaving one datasubset out.
- "Test" each NN on the corresponding holdout set.
- **Combine into one analysis, in some manner.**
- Rationale: LEE free search without sacrificing event-count (p-values are computed assuming that $\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k}$ are a priori fixed).



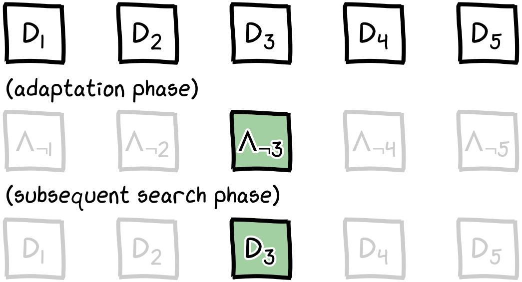
k-fold cross adaptive search (1805.02664 [hep-ph], 1902.02634 [hep-ph])

- Split datasets into k independent subsets.
- Train k different NNs $\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k}$, each time leaving one datasubset out.
- "Test" each NN on the corresponding holdout set.
- **Combine into one analysis, in some manner.**
- Rationale: LEE free search without sacrificing event-count (p-values are computed assuming that $\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k}$ are a priori fixed).



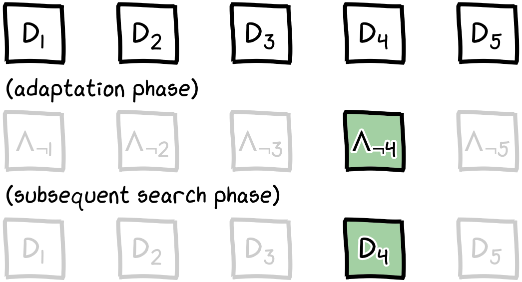
k-fold cross adaptive search (1805.02664 [hep-ph], 1902.02634 [hep-ph])

- Split datasets into k independent subsets.
- Train k different NNs $\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k}$, each time leaving one datasubset out.
- "Test" each NN on the corresponding holdout set.
- **Combine into one analysis, in some manner.**
- Rationale: LEE free search without sacrificing event-count (p-values are computed assuming that $\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k}$ are a priori fixed).



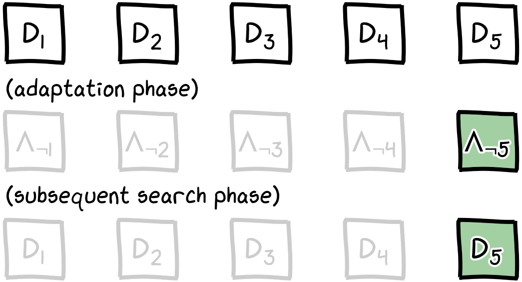
k-fold cross adaptive search (1805.02664 [hep-ph], 1902.02634 [hep-ph])

- Split datasets into k independent subsets.
- Train k different NNs $\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k}$, each time leaving one datasubset out.
- “Test” each NN on the corresponding holdout set.
- **Combine into one analysis, in some manner.**
- Rationale: LEE free search without sacrificing event-count (p-values are computed assuming that $\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k}$ are a priori fixed).



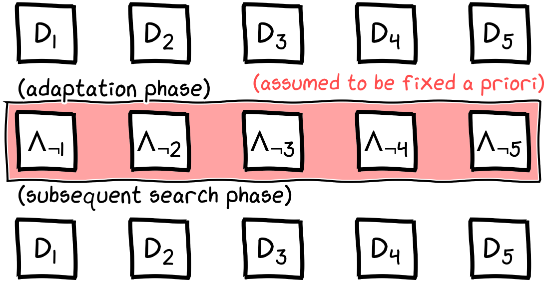
k-fold cross adaptive search (1805.02664 [hep-ph], 1902.02634 [hep-ph])

- Split datasets into k independent subsets.
- Train k different NNs $\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k}$, each time leaving one datasubset out.
- "Test" each NN on the corresponding holdout set.
- **Combine into one analysis, in some manner.**
- Rationale: LEE free search without sacrificing event-count (p-values are computed assuming that $\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k}$ are a priori fixed).



k-fold cross adaptive search (1805.02664 [hep-ph], 1902.02634 [hep-ph])

- Split datasets into k independent subsets.
- Train k different NNs $\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k}$, each time leaving one datasubset out.
- "Test" each NN on the corresponding holdout set.
- **Combine into one analysis, in some manner.**
- Rationale: LEE free search without sacrificing event-count (p-values are computed assuming that $\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k}$ are a priori fixed).



Rest of the talk...

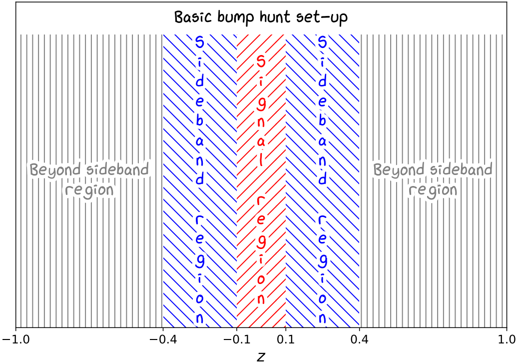
- Discuss why **one should expect** a LEE.
- Show the results of a study.

Rest of the talk...

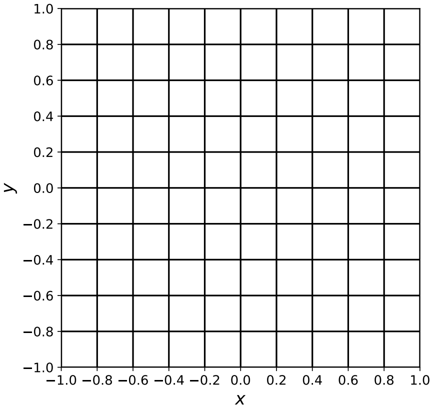
- ~~Discuss why **one should expect** a LEE.~~
- Show the results of a study.
- Discuss why **there is** a LEE.

An example analysis

- Event description: $(x, y, z) \in [-1, 1] \times [-1, 1] \times [-1, 1]$
- **Null hypothesis:** Total number of events $\sim$ Poisson(10^5).
  Events are uniformly distributed (IID).
- Bump bunt in z. Event selection in (x, y).

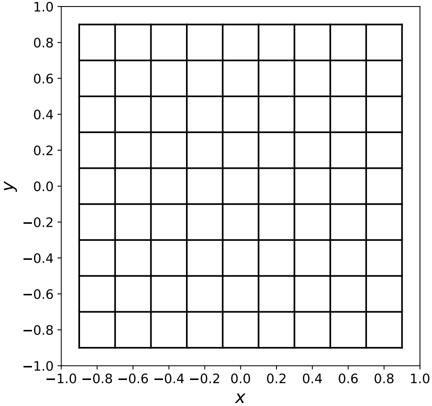


An example analysis



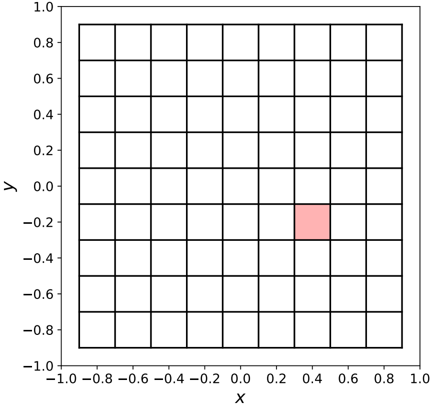
- Bin the $x - y$ space: 10×10 bins, shifted 9×9 bins.
- Identify the bin b with the largest $\frac{\mathrm{SR}_b - r\,\mathrm{SB}_b}{\sqrt{\mathrm{SR}_b + r^2\,\mathrm{SB}_b}}$, where $r = 1/3$.
- Anomaly score = $-$ (distance from center of this bin).
- Use sideband and beyond sideband data to pick the threshold on anomaly score for accepting $\approx 10\%$ of the background data.
- k-fold CAS:
  - Choose $k = 5$.
  - Train a classifier on $k - 1$ datasubsets and use on holdout subset. Repeat k times.
  - Merge all accepted events and perform bump hunt.
  - Compute the local p-value and report it as global.

An example analysis



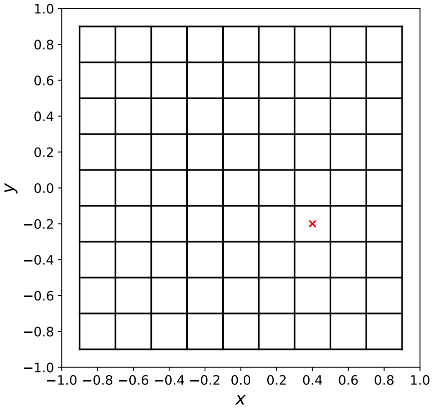
- Bin the $x - y$ space: 10×10 bins, shifted 9×9 bins.
- Identify the bin b with the largest $\dfrac{\mathrm{SR}_b - r\,\mathrm{SB}_b}{\sqrt{\mathrm{SR}_b + r^2\,\mathrm{SB}_b}}$, where $r = 1/3$.
- Anomaly score = $-$ (distance from center of this bin).
- Use sideband and beyond sideband data to pick the threshold on anomaly score for accepting $\approx 10\%$ of the background data.
- k-fold CAS:
  - Choose $k = 5$.
  - Train a classifier on $k - 1$ datasubsets and use on holdout subset. Repeat k times.
  - Merge all accepted events and perform bump hunt.
  - Compute the local p-value and report it as global.

An example analysis



- Bin the $x - y$ space: 10×10 bins, shifted 9×9 bins.
- Identify the bin b with the largest $\dfrac{\mathrm{SR}_b - r\,\mathrm{SB}_b}{\sqrt{\mathrm{SR}_b + r^2\,\mathrm{SB}_b}}$, where $r = 1/3$.
- Anomaly score = $-$ (distance from center of this bin).
- Use sideband and beyond sideband data to pick the threshold on anomaly score for accepting $\approx 10\%$ of the background data.
- k-fold CAS:
  - Choose $k = 5$.
  - Train a classifier on $k-1$ datasubsets and use on holdout subset. Repeat k times.
  - Merge all accepted events and perform bump hunt.
  - Compute the local p-value and report it as global.

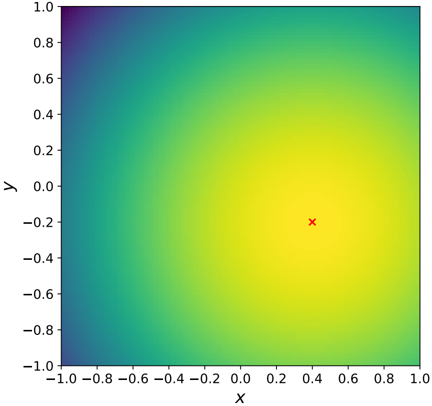
An example analysis



- Bin the $x - y$ space: 10×10 bins, shifted 9×9 bins.
- Identify the bin b with the largest $\dfrac{\mathrm{SR}_b - r\,\mathrm{SB}_b}{\sqrt{\mathrm{SR}_b + r^2\,\mathrm{SB}_b}}$, where $r = 1/3$.
- Anomaly score = $-$ (distance from center of this bin).
- Use sideband and beyond sideband data to pick the threshold on anomaly score for accepting $\approx 10\%$ of the background data.
- k-fold CAS:
  - Choose $k = 5$.
  - Train a classifier on $k - 1$ datasubsets and use on holdout subset. Repeat k times.
  - Merge all accepted events and perform bump hunt.
  - Compute the local p-value and report it as global.

Fermilab

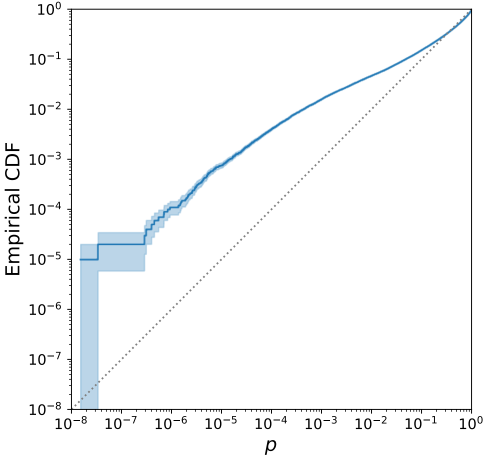
An example analysis



- Bin the $x - y$ space: 10×10 bins, shifted 9×9 bins.
- Identify the bin b with the largest $\dfrac{\mathrm{SR}_b - r\,\mathrm{SB}_b}{\sqrt{\mathrm{SR}_b + r^2\,\mathrm{SB}_b}}$, where $r = 1/3$.
- Anomaly score = $-$ (distance from center of this bin).
- Use sideband and beyond sideband data to pick the threshold on anomaly score for accepting $\approx 10\%$ of the background data.
- k-fold CAS:
  - Choose $k = 5$.
  - Train a classifier on $k - 1$ datasubsets and use on holdout subset. Repeat k times.
  - Merge all accepted events and perform bump hunt.
  - Compute the local p-value and report it as global.

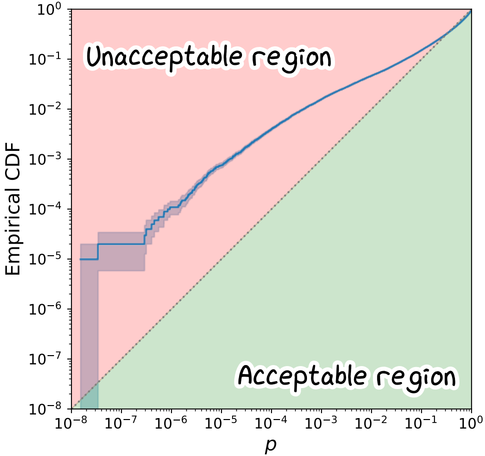
Result

Empirical CDF of local p-value under the null hypothesis with Wald error band (using 10^5 pseudo experiments)



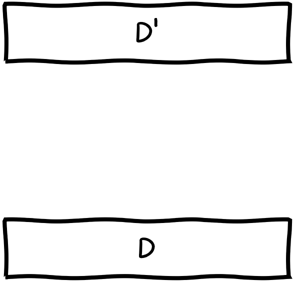
Result

Empirical CDF of local p-value under the null hypothesis with Wald error band (using 10^5 pseudo experiments)



Why is there a LEE?

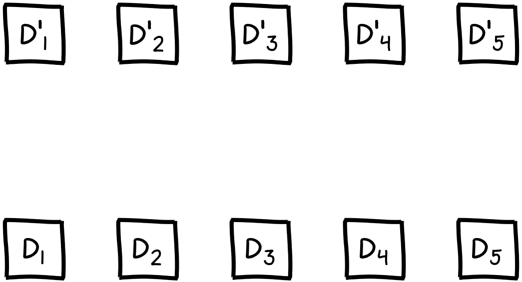
- Because $(\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k})$ is not independent of $(D_1, \ldots, D_k)$ in k-fold CAS.
- Consider the following modification of k-fold CAS, which uses twice the amount of data. Call it the k-fold Independent Adaptive Search (IAS).



- $(\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k})$ and $(D_1, \ldots, D_k)$ have the same dist. under CAS and IAS.
- $(\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k})$ is **independent** of $(D_1, \ldots, D_k)$ under k-fold IAS.

Why is there a LEE?

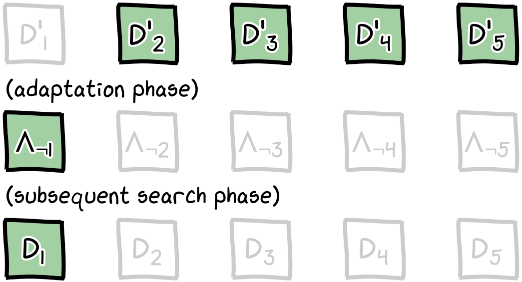
- Because $(\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k})$ is not independent of $(D_1, \ldots, D_k)$ in k-fold CAS.
- Consider the following modification of k-fold CAS, which uses twice the amount of data. Call it the k-fold Independent Adaptive Search (IAS).



- $(\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k})$ and $(D_1, \ldots, D_k)$ have the same dist. under CAS and IAS.
- $(\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k})$ is **independent** of $(D_1, \ldots, D_k)$ under k-fold IAS.

Why is there a LEE?

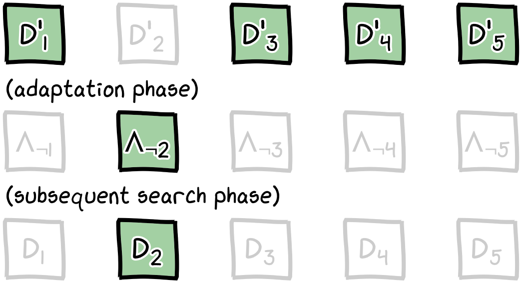
- Because $(\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k})$ is not independent of $(D_1, \ldots, D_k)$ in k-fold CAS.
- Consider the following modification of k-fold CAS, which uses twice the amount of data. Call it the k-fold Independent Adaptive Search (IAS).



- $(\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k})$ and $(D_1, \ldots, D_k)$ have the same dist. under CAS and IAS.
- $(\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k})$ is **independent** of $(D_1, \ldots, D_k)$ under k-fold IAS.

Why is there a LEE?

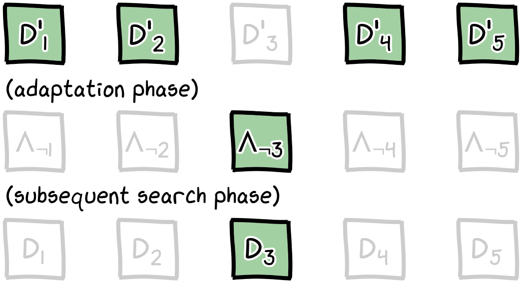
- Because $(\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k})$ is not independent of $(D_1, \ldots, D_k)$ in k-fold CAS.
- Consider the following modification of k-fold CAS, which uses twice the amount of data. Call it the k-fold Independent Adaptive Search (IAS).



- $(\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k})$ and $(D_1, \ldots, D_k)$ have the same dist. under CAS and IAS.
- $(\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k})$ is **independent** of $(D_1, \ldots, D_k)$ under k-fold IAS.

Why is there a LEE?

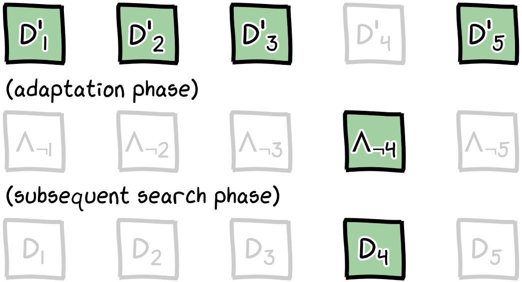
- Because $(\Lambda_{\neg 1}, \dots, \Lambda_{\neg k})$ is not independent of $(D_1, \dots, D_k)$ in k-fold CAS.
- Consider the following modification of k-fold CAS, which uses twice the amount of data. Call it the k-fold Independent Adaptive Search (IAS).



- $(\Lambda_{\neg 1}, \dots, \Lambda_{\neg k})$ and $(D_1, \dots, D_k)$ have the same dist. under CAS and IAS.
- $(\Lambda_{\neg 1}, \dots, \Lambda_{\neg k})$ is **independent** of $(D_1, \dots, D_k)$ under k-fold IAS.

Why is there a LEE?

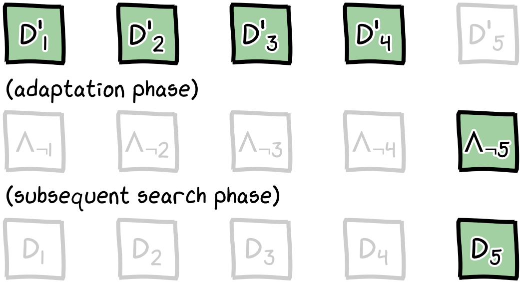
- Because $(\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k})$ is not independent of $(D_1, \ldots, D_k)$ in k-fold CAS.
- Consider the following modification of k-fold CAS, which uses twice the amount of data. Call it the k-fold Independent Adaptive Search (IAS).



- $(\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k})$ and $(D_1, \ldots, D_k)$ have the same dist. under CAS and IAS.
- $(\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k})$ is **independent** of $(D_1, \ldots, D_k)$ under k-fold IAS.

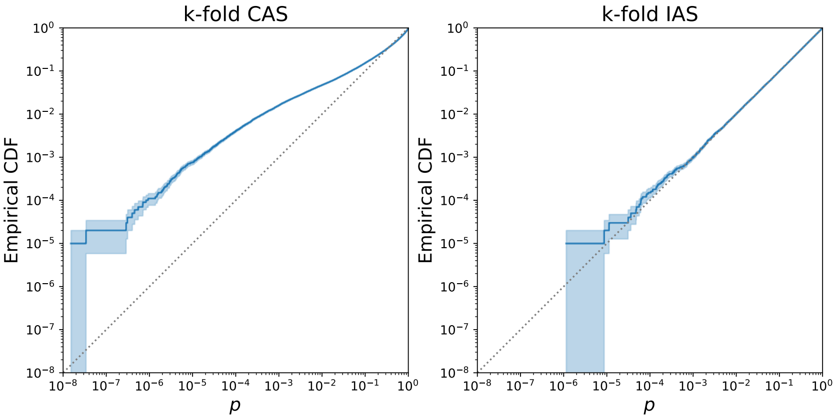
Why is there a LEE?

- Because $(\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k})$ is not independent of $(D_1, \ldots, D_k)$ in k-fold CAS.
- Consider the following modification of k-fold CAS, which uses twice the amount of data. Call it the k-fold Independent Adaptive Search (IAS).



- $(\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k})$ and $(D_1, \ldots, D_k)$ have the same dist. under CAS and IAS.
- $(\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k})$ is **independent** of $(D_1, \ldots, D_k)$ under k-fold IAS.

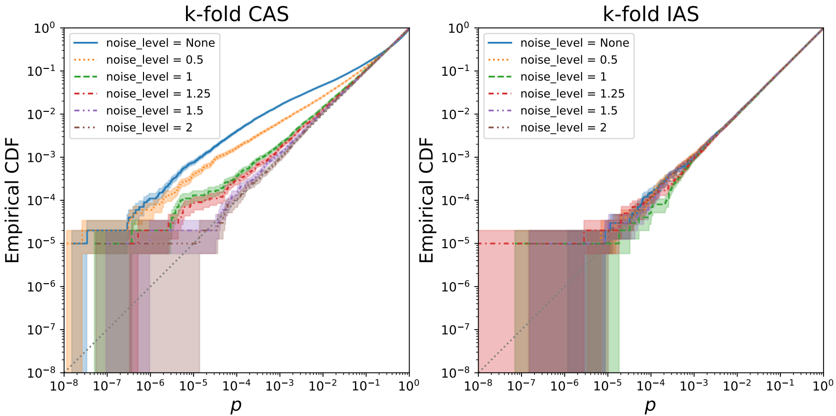
CAS–IAS comparison



A useful question: Given that all the $\Lambda_{\neg i}$-s have picked out the same bin (say, out of sheer luck) what is the null distribution of the bump-hunt test statistic under IAS? What is it under CAS?

Fermilab

CAS–IAS comparison



Add random noise in the training as a regularization technique to reduce dependence on data? **This will (a) only delay the onset of LEE to lower p-values and (b) make the adaptation stage noise-driven at higher p-values.**

Summary and Outlook

- k-fold CAS has a look-elsewhere effect.
- This is underappreciated in the literature.
  (There is one investigation into this issue in the literature in 1902.02634 [hep-ph], with only 10^3 pseudo experiments.)
- This can have far-reaching consequences:
  - k-fold CAS has become ubiquitous in the anomaly detection literature.
  - The reported sensitivity of techniques that use k-fold CAS can be overestimated.
  - The reported significances of apparent anomalies can be overestimated.
- This is a short talk, but we'll have a lot more content in the paper...
  - The mechanisms that convert the dependence between $(\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k})$ and $(D_1, \ldots, D_k)$ into a LEE.
  - How noisy-features problem can be explained as a direct consequence of the breadth–depth tradeoff.
  - How to probe really low p-values without doing a lot of pseudo-experiments.
  - Characterizing Pareto-optimal tests for composite hypothesis testing, etc.

Summary and Outlook

Thank you! Questions?

- k-fold CAS has a look-elsewhere effect.
- This is underappreciated in the literature.
  (There is one investigation into this issue in the literature in 1902.02634 [hep-ph], with only 10^3 pseudo experiments.)
- This can have far-reaching consequences:
  - k-fold CAS has become ubiquitous in the anomaly detection literature.
  - The reported sensitivity of techniques that use k-fold CAS can be overestimated.
  - The reported significances of apparent anomalies can be overestimated.
- This is a short talk, but we'll have a lot more content in the paper...
  - The mechanisms that convert the dependence between $(\Lambda_{\neg 1}, \ldots, \Lambda_{\neg k})$ and $(D_1, \ldots, D_k)$ into a LEE.
  - How noisy-features problem can be explained as a direct consequence of the breadth–depth tradeoff.
  - How to probe really low p-values without doing a lot of pseudo-experiments.
  - Characterizing Pareto-optimal tests for composite hypothesis testing, etc.

Acknowledgments

These slides were prepared using the resources of the Fermi National Accelerator Laboratory (Fermilab), a U.S. Department of Energy, Office of Science, Office of High Energy Physics HEP User Facility. Fermilab is managed by Fermi Forward Discovery Group, LLC, acting under Contract No. 89243024CSC000002.

PS is supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics QuantISED program under the following grants:

- "HEP Machine Learning and Optimization Go Quantum", Award Number 0000240323
- "DOE QuantiSED Consortium QCCFP-QMLQCF", Award Number DE-SC0019219