NeurIPS 2026

ON THE TIGHTNESS AND COMPUTATIONAL TRACTABILITY OF HIGHER-DIMENSIONAL CONFIDENCE SEQUENCES

SPOTLIGHT PAPER

Fabian Denoodt, Sibylle Hess, Joaquin Vanschoren, Christian A. Naesseth

Eindhoven University of Technology

AMOR/e Advanced Models by Open Research & Engineering

University of Amsterdam


A
Add to cart
vs.
B
Buy now
Purchases
Time
Scroll depth


Decide
Website A
Website B
Users over time

Observe  $\cdots$ $\cdots$

Y_1 Y_2 Y_3 Y_t

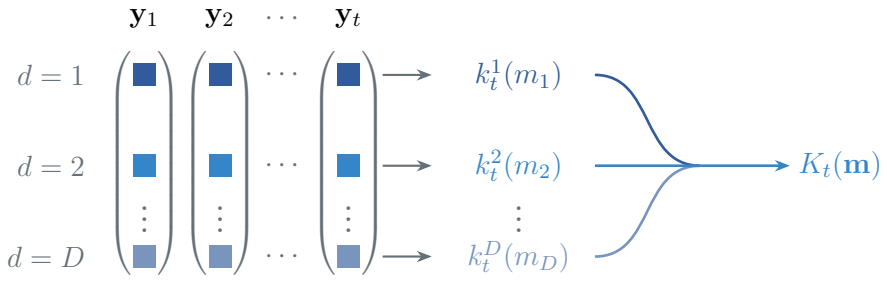
$$\mathbb{E}[Y_t \mid Y_1, \ldots, Y_{t-1}] = \mu$$

A confidence sequence is a sequence of sets $\mathcal{C}_1, \mathcal{C}_2, \ldots$ with

$$\Pr(\forall t \geq 1 : \mu \in \mathcal{C}_t) \geq 1 - \alpha.$$

$$\mathcal{C}_t = \{m \in [0,1] : k_t(m) \leq 1/\alpha\}.$$

Waudby-Smith and Ramdas (2024)



$$\mathcal{C}_t = \{\mathbf{m} : K_t(\mathbf{m}) \leq 1/\alpha\}$$

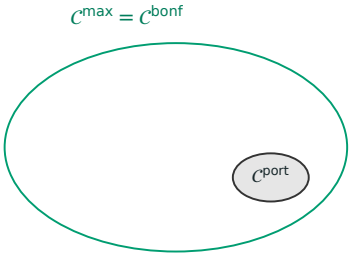
$$\mathcal{C}_t = \{\mathbf{m} : K_t(\mathbf{m}) \leq 1/\alpha\}.$$

Bonferroni / max

$$K_t^{\max}(\mathbf{m}) = \max_d w_d k_t^d(m_d)$$

Portfolio

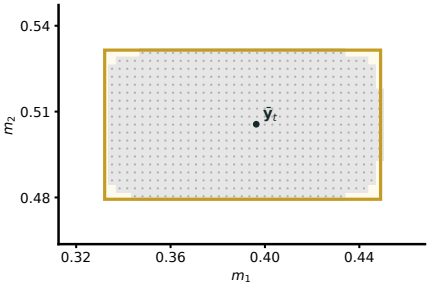
$$K_t^{\text{port}}(\mathbf{m}) = \sum_d w_d k_t^d(m_d)$$

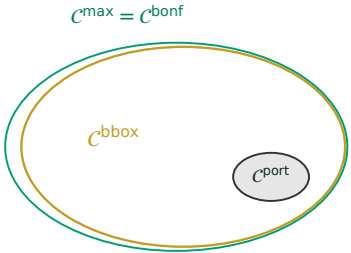


Tractable outer approximations

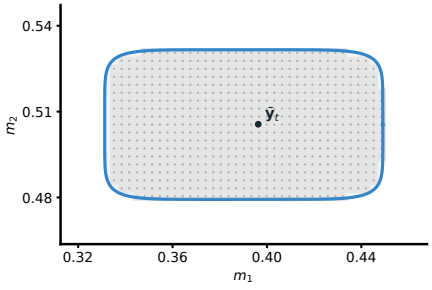
1. Bounding box
2. ℓ_p-ellipsoid
3. Intersection

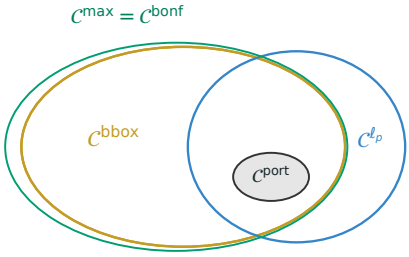
$$\mathcal{C}_t^{\text{bbox}} := \{\mathbf{m} : m_d \in [L_d, U_d], \ \ \forall d\}.$$



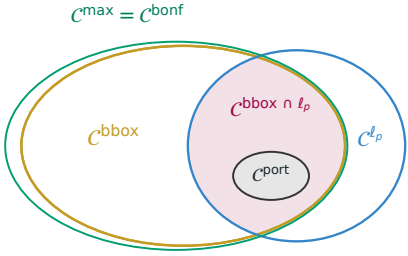


$$\mathcal{C}_t^{\ell_p} := \left\{ \mathbf{m} : \sum_d \left| \frac{m_d - a_d}{r_d} \right|^p \leq 1 \right\}.$$



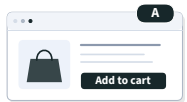


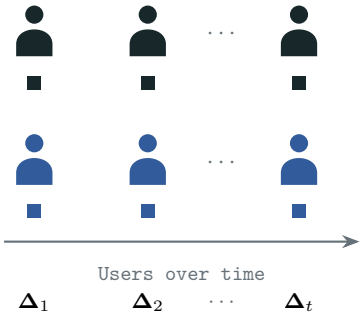
$$\mathcal{C}_t^{\text{bbox} \cap \ell_p} := \mathcal{C}_t^{\text{bbox}} \cap \mathcal{C}_t^{\ell_p}.$$



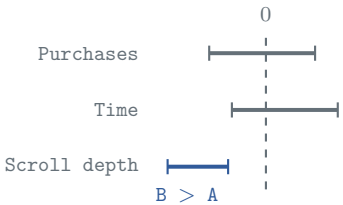
Results

MULTI-METRIC A/B TESTING - USECASE 1





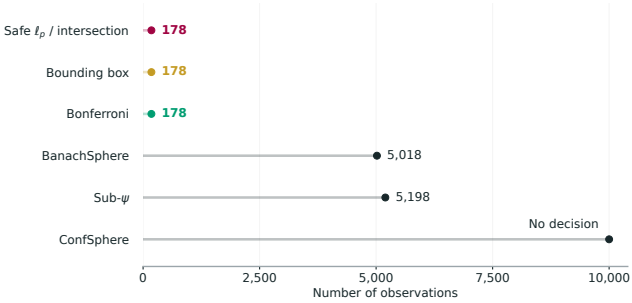
$$\blacksquare = \begin{pmatrix} \text{Purchases} \\ \text{Time} \\ \text{Scroll depth} \\ \vdots \end{pmatrix}$$



$$\Delta_t = \blacksquare - \blacksquare$$

MULTI-METRIC A/B TESTING - USECASE 1






A
Add to cart
B
Buy now
Safe ℓ_p / intersection
356
Bounding box
398
Bonferroni
399
BanachSphere
8,052
Sub-ψ
6,650
ConfSphere
No decision
0
2,500
5,000
7,500
10,000
Number of observations

COMPARING TWO ML MODELS - USECASE 2

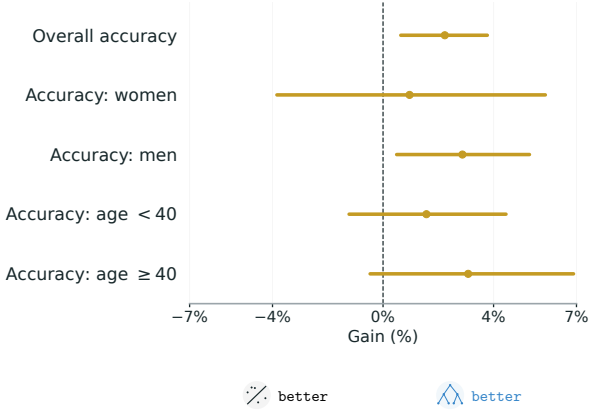


Logistic regression

vs



Boosted trees



Bounding box

At least as tight as Bonferroni, at the same complexity.

Ellipsoid and intersection

Close to the portfolio CS, at much lower complexity.

Earlier decisions

Multi-metric A/B tests and ML model comparisons.

Fabian Denoodt[1], Sibylle Hess[1],
Joaquin Vanschoren[1], Christian A. Naesseth[2]

TU/e [1]Eindhoven University of Technology

AMOR/e
Advanced Models by
Open Research & Engineering

[2]University of Amsterdam

f.l.m.denoodt@tue.nl

Website

Project page