Anytime-valid, Bayes-assisted, Prediction-Powered Inference

Valentin Kilian*, Stefano Cortinovis*, François Caron

Department of Statistics, University of Oxford,

*Equal contribution. Order decided by coin toss.

Our contributions

- Prediction-Powered Inference (PPI) [Angelopoulos et al., 2023] leverages an ML model to derive tight fixed-time confidence sets.
- We extend the PPI framework to the **anytime-valid setting**.
- We can leverage **prior information** on the accuracy of the ML model to obtain tighter anytime-valid confidence sets.

Prediction-Powered Inference (PPI)

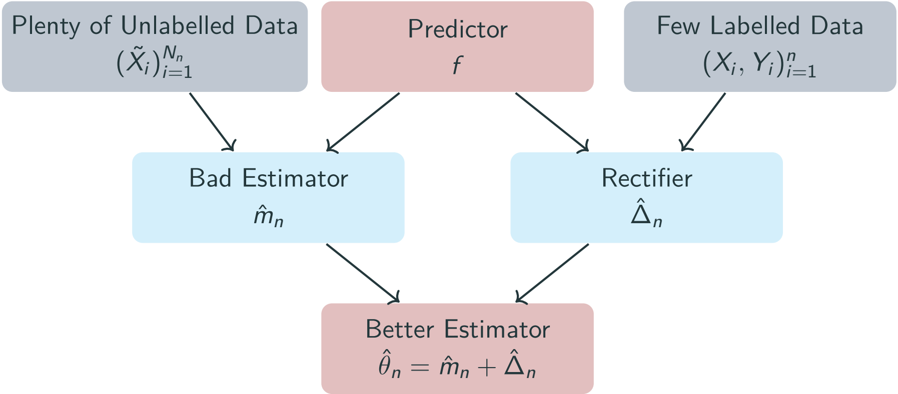
Given an input/output pair $(X, Y) \sim \mathbb{P} = \mathbb{P}_X \times \mathbb{P}_{Y|X}$, consider the goal of estimating

$$\theta^\star = \underset{\theta \in \mathbb{R}}{\arg\min} \;\; \mathbb{E}[\ell_\theta(X, Y)],$$

where $\ell_\theta(x, y)$ is a convex loss function parameterised by $\theta \in \mathbb{R}$.

To simplify the notation we focus from now on mean estimation ($\theta^\star = \mathbb{E}[Y]$). We refer to our article for the general case.

Prediction-Powered Inference (PPI)



Confidence Sequences v.s. Confidence Intervals

Confidence Interval (CI):
For all $t \geq 1$, $\Pr\left(\mu \in \mathcal{C}_{\alpha,t}\right) \geq 1 - \alpha$.

→ CLT → **Asymptotic CI**

Confidence Sequence (CS)
$\Pr(\mu \in \mathcal{C}_{\alpha,t}$ **for all** $\boldsymbol{t \geq 1}) \geq 1 - \alpha$.

→ KMT → **Asymptotic CS** (AsympCS)

Anytime-valid, Bayes-assisted, Prediction-Powered Inference

Under some realistic integrability assumptions we have:

Theorem (AsympCS for m_θ)

Let $(\widehat{\sigma}_n^f)^2$ be the sample variance of $(f(\widetilde{X}_i))_{i=1}^{N_n}$. Let $\delta \in (0,1)$. For any $\rho > 0$,

$$\mathcal{R}_{\delta,n} = \left[\widehat{m}_n \pm \frac{\widehat{\sigma}_n^f}{\sqrt{N_n}} \sqrt{\left(1 + \frac{1}{N_n \rho^2}\right) \log\left(\frac{N_n \rho^2 + 1}{\delta^2}\right)}\right].$$

forms a $(1-\delta)$–AsympCS with approximation rate $1/\sqrt{n \log n}$ for m.

Anytime-valid, Bayes-assisted, Prediction-Powered Inference

Under some realistic integrability assumptions we have:

Theorem (Bayes-assisted AsympCS for Δ_θ)

Let $(\widehat{\sigma}_n^\Delta)^2$ be the sample variance of $(Y_i - f(X_i))_{i=1}^n$. Assume that $|\frac{n}{N_n} - r| = O(1/n^{1-a})$ for some $r \in [0,1]$. ***For any continuous proper prior π,***

$$\mathcal{T}_{\kappa,n} = \left[\widehat{\Delta}_n \pm \frac{\widehat{\sigma}_n^\Delta}{\sqrt{n}} \sqrt{\log\left(\frac{n}{2\pi\kappa^2 \boldsymbol{\eta_n(\widehat{\Delta}_n/\widehat{\sigma}_n^\Delta)^2}} \right)} \right],$$

where $\kappa \in (0,1)$ and $\eta_t : \mathbb{R} \to (0, \sqrt{t/(2\pi)})$ is defined as

$$\boldsymbol{\eta_t(z) = \int_{-\infty}^{\infty} \mathcal{N}(z; \zeta, 1/t)\, \pi(\zeta)\, d\zeta},$$

forms a $(1-\kappa)$–AsympCS with approximation rate $1/\sqrt{n \log n}$ for Δ.

Anytime-valid, Bayes-assisted, Prediction-Powered Inference

Finally we get for $\alpha = \delta + \kappa$:

$$\mathcal{C}_{\alpha,n} = \Bigg[\widehat{\theta}_n \pm \Bigg\{ \frac{\widehat{\sigma}_n^{\Delta}}{\sqrt{n}} \sqrt{\log\left(\frac{n}{2\pi\kappa^2 \boldsymbol{\eta_n(\widehat{\Delta}_n/\widehat{\sigma}_n^{\Delta})}}\right)} + \frac{\widehat{\sigma}_n^{f}}{\sqrt{N_n}} \sqrt{\frac{1+N_n\rho^2}{N_n\rho^2}\log\left(\frac{N_n\rho^2+1}{\delta^2}\right)} \Bigg\}\Bigg]$$

Experiments

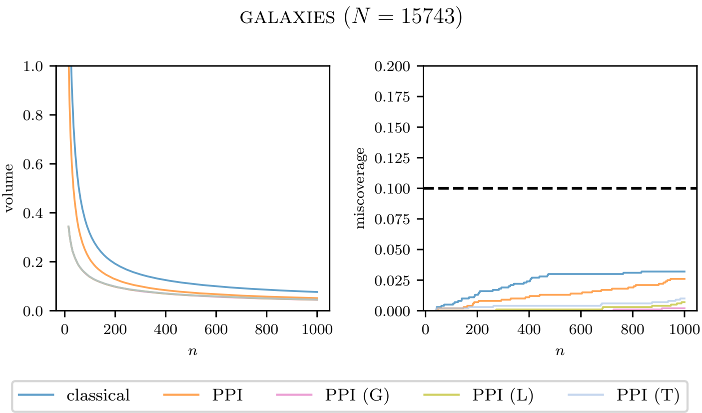


Figure 1: Real data study. Average interval volume and cumulative miscoverage rate over 1000 repetitions for the GALAXIES dataset.