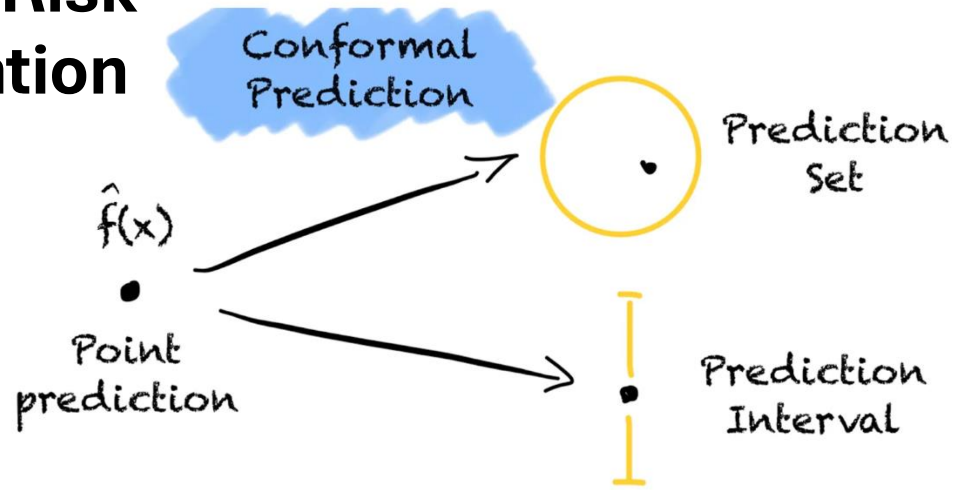
Bayesian Conformal Prediction as a Decision Risk Problem using Bayesian Quadrature Optimisation



Conformal Prediction (CP) is a statistical technique that constructs prediction sets with guaranteed marginal coverage properties for the next observation without any distributional assumptions besides exchangeability.

$$\mathbb{P}\left(y_{\text{new}} \in \mathcal{C}_{\text{cp}}(x_{\text{new}})\right) \geq 1-\alpha.$$

Snell & Griffiths (2025): *Conformal Prediction as Bayesian Quadrature*
- Re-write Conformal Prediction as Decision Risk Problem
- Method: Bayesian Quadrature for integration problem
- But: The non-conformity score remains fixed, independent of θ or λ

$$R(\theta, \lambda) = \int L(\theta, \lambda(z)) f(z \mid \theta)\, dz$$

$$\bar{R}(\lambda) = \sup_{\theta} R(\theta, \lambda) \leq \alpha$$

Fong et al. (2021): *Conformal Bayesian Computation*
- Bayesian Posterior Predictive as non-conformity score
- But: No decision-theoretic framework for the optimal CP set

$$s(x, y) = -\log p(y \mid x, D)$$
$$= -\log \int p(y \mid x, \theta) p(\theta \mid D)$$

Vovk et al. (2005, Theorem 2.10): conformal predictors are at minimum equivalent to conservatively valid predictors. Provides Hints for systematic optimisation

"Can we formulate conformal prediction as a **decision risk problem**, and use this framework to **integrate Bayesian** methods while **preserving distribution-free guarantees**?"

Method & Theory

Rewrite CP into a Decision Risk Minimisation framework

$$R(\theta, \lambda) = \int L(\theta, \lambda(z))\, f(z \mid \theta)\, dz$$

following decision rule λ bounded by worst-case risk constraint

$$\bar{R}(\lambda) = \sup_{\theta} R(\theta, \lambda) \leq \alpha$$

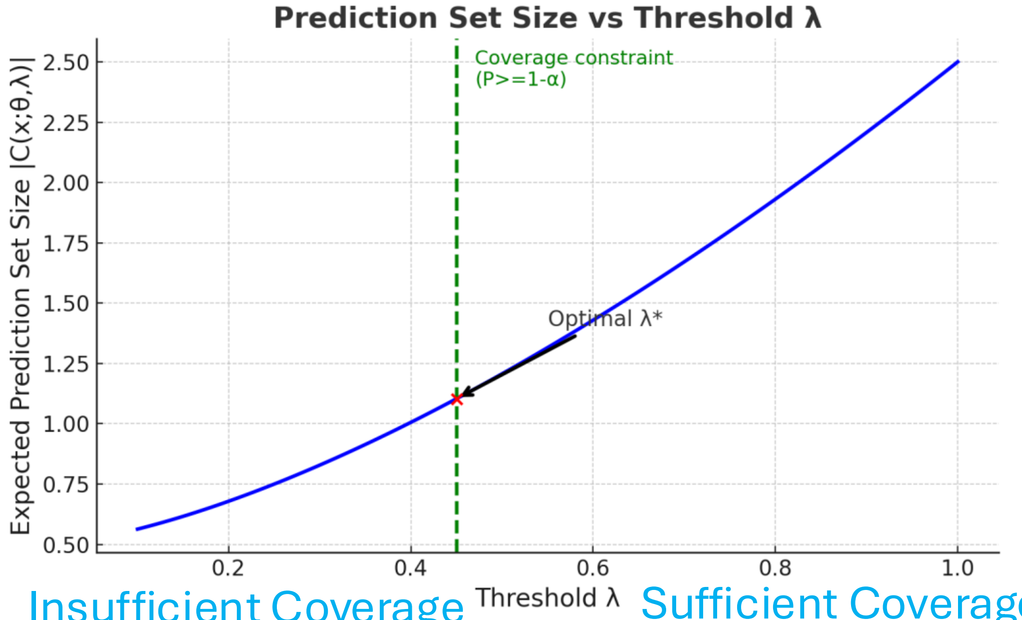
Choose Bayesian non-conformity score and produce prediction set

$$s(x, y, \theta) = -\log \int p(y \mid x, \theta)\, p(\theta \mid \mathcal{D})\, d\theta,$$

$$C(x; \theta, \lambda) = \{y \in \mathcal{Y} : s(x, y, \theta) \leq \lambda\}$$

Our Decision Risk Problem:

$$R(\lambda) = \iiint L(y, C(x; \theta, \lambda))\, p(y \mid x, \theta)\, p(\theta \mid \mathcal{D})\, p(x)\, dx\, dy\, d\theta.$$



Insufficient Coverage Sufficient Coverage

Goal: **Smallest possible sets** while **guaranteeing coverage**

We turn CP into a decision problem:

Decision Risk $R(\lambda)$

$$\lambda^* = \arg\min_{\lambda} \boxed{\mathbb{E}_{\theta \sim p(\theta|D)}\, \mathbb{E}_x \left[|C(x; \theta, \lambda)|\right]}$$

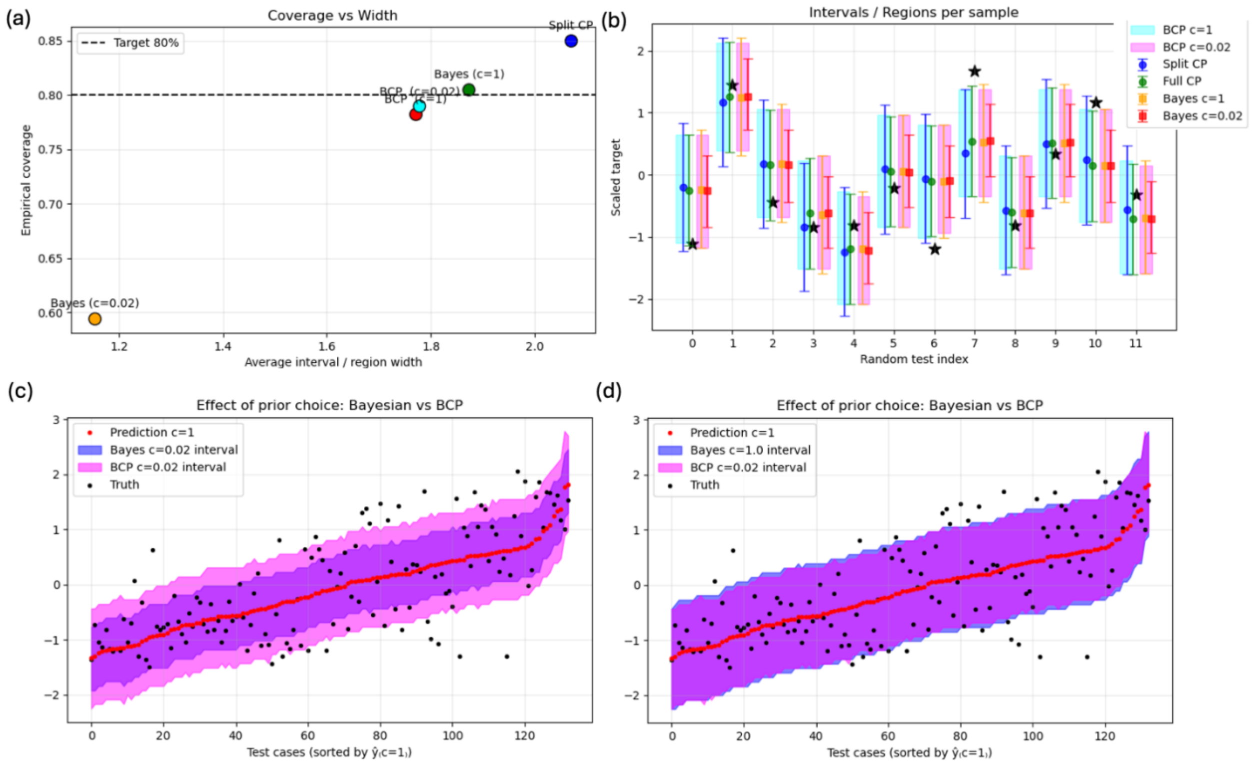
s.t.

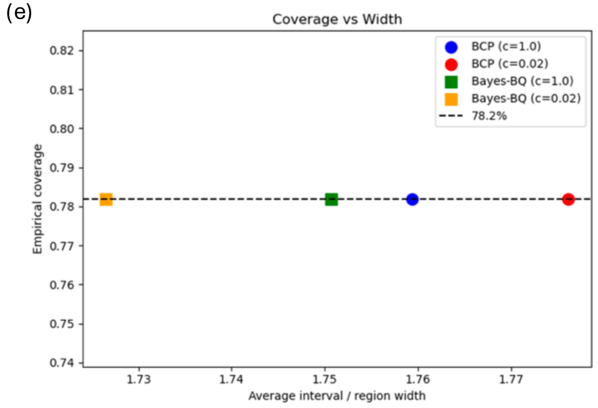
$$\mathbb{P}_{\theta \sim p(\theta|D)}(Y \in C(X; \theta, \lambda)) \geq 1 - \alpha$$

Minimizing its expectation reduces decision risk while integrating over model uncertainty

Thus, minimizing expected set size |C(x;θ,λ)| is equivalent to minimizing risk.

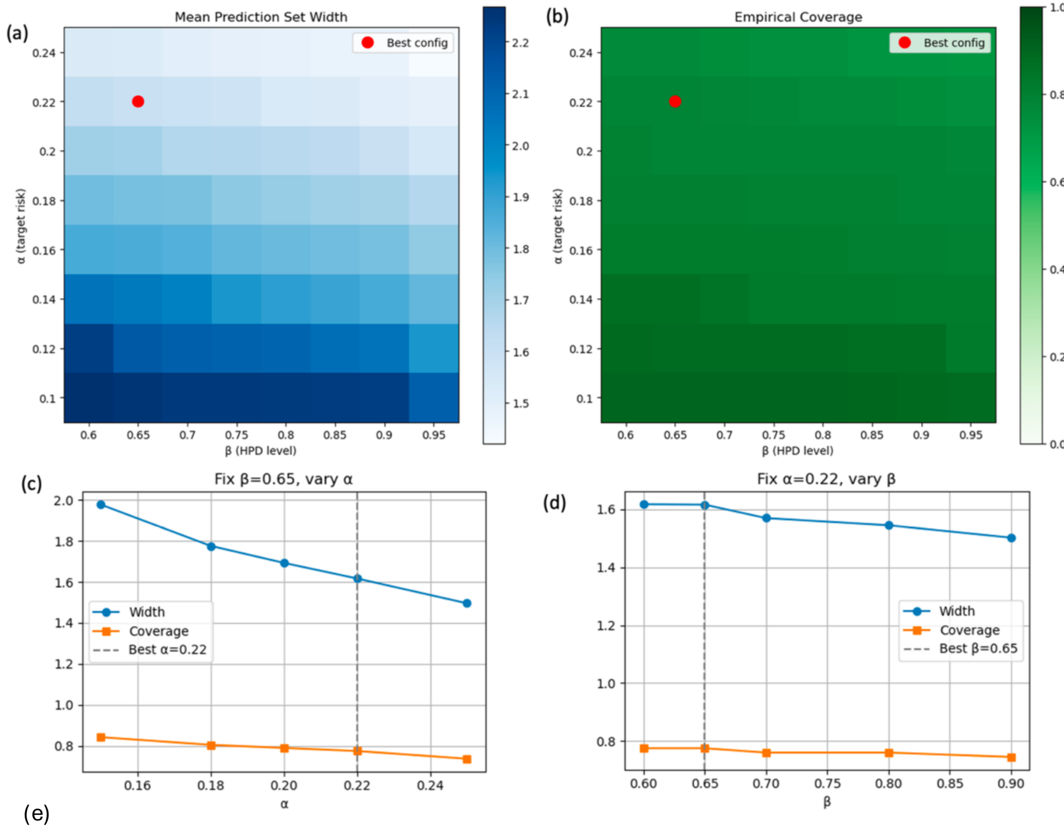
Results & Conclusion

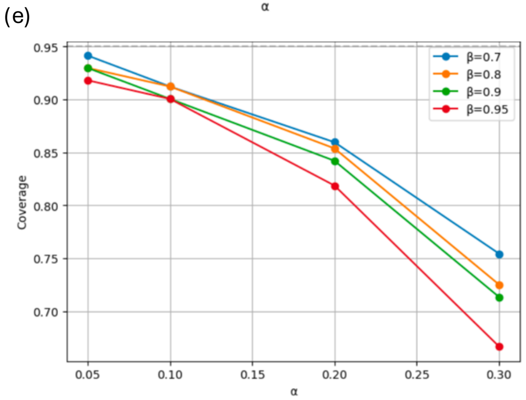




Sparse Regression

Method	Prior(c)	Width	Coverage
BCP	0.02	1.7761	0.7820
BQ (optimized)	0.02	1.6651	0.7820
BCP	1.00	1.7507	0.7820
BQ (optimized)	1.00	1.6874	0.7895





Breast Cancer Classification

Method	Coverage	Prediction set size
BCP	0.910	0.919
BQ-optimized	0.901	0.889