

Beyond Predicting Responses: Conformal Inference for Latent Distributional Parameters

Minxing Zheng Wenbin Zhou Shixiang Zhu

Carnegie Mellon University



Motivation: Hidden Target Parameter

The observed response is often not the ultimate object of interest:

- air-quality sensor → **pollutant concentration**;
- user rating → **user preference**;
- wildfire count → **ignition risk**.

We observe a context X and a random response Y , while the instance-specific parameter θ governing the law of Y remains hidden.

Two fundamental challenges

- No latent labels:** θ is unobserved in both history and future data.
- Non Identifiability:** one response can be compatible with multiple latent parameters.

Problem Setup

Hierarchical latent variable model

$$X \sim P_X, \quad \theta \mid X = x \sim \pi(\cdot \mid x) \quad (\text{mixing distribution}),$$

$$Y \mid (X = x, \theta) \sim P_\theta(\cdot \mid x) \quad (\text{forward model}).$$

Goal: **uncertainty set for the latent parameter**

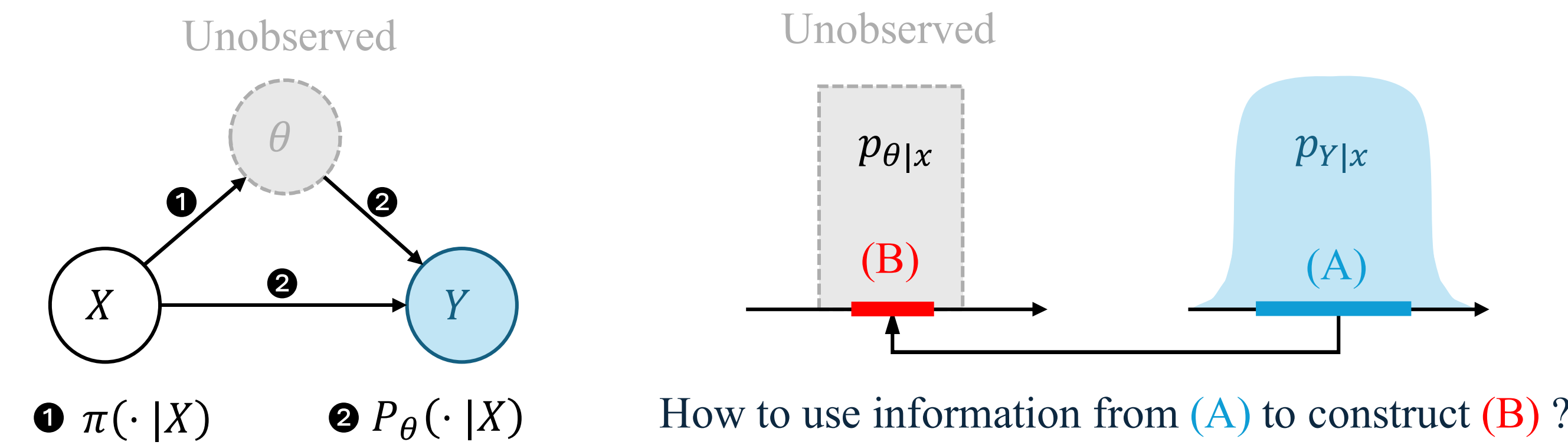
Given data $\mathcal{D}_n = \{(X_i, Y_i)\}_{i=1}^n$ (θ_i unobserved), build an uncertainty set $\mathcal{U}_\alpha(X_{n+1}; \mathcal{D}_n) \subseteq \Theta$ for a future θ_{n+1} satisfying

$$\mathbb{P}\{\theta_{n+1} \in \mathcal{U}_\alpha(X_{n+1}; \mathcal{D}_n)\} \geq 1 - \alpha$$

for a new instance X_{n+1} and a target level $\alpha \in (0, 1)$. A *prediction set for the random, instance-specific θ_{n+1}* — not a confidence region for a fixed population parameter.

What we assume — and what we do not

- (i) mixing distribution $\pi(\cdot \mid x)$ **unknown**;
- (ii) forward family $\{P_\theta(\cdot \mid x)\}_{\theta \in \Theta}$ known or evaluable.



From Response to Latent-Parameter Coverage

Standard conformal prediction

Given a prediction model $\hat{f} : \mathcal{X} \rightarrow \mathcal{Y}$, standard conformal prediction (CP) constructs a valid set for the future response Y_{n+1} . Given calibration data $\mathcal{D}_n = \{(X_i, Y_i)\}_{i=1}^n$ and target miscoverage level $\gamma \in (0, 1)$, compute nonconformity scores $S_i = S(X_i, Y_i)$ and quantile $\hat{q}_{1-\gamma} = S_{(\lceil (n+1)(1-\gamma) \rceil)}$. Then

$$C_\gamma(x; \mathcal{D}_n) = \{y : S(x, y) \leq \hat{q}_{1-\gamma}\}.$$

Under exchangeability,

$$\mathbb{P}\{Y_{n+1} \in C_\gamma(X_{n+1}; \mathcal{D}_n)\} \geq 1 - \gamma.$$

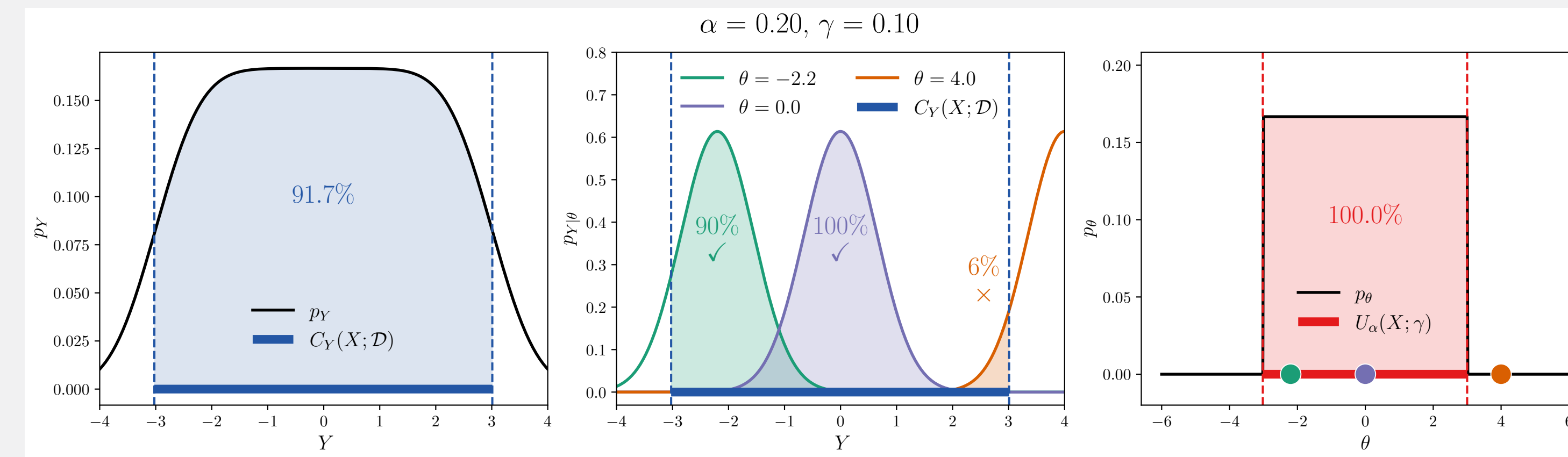
What does each approach additionally require?

Method	Target	Prior-free	Label-free	Finite-sample
Standard CP	Y_{n+1}	✓	✓	✓ [†]
Bayes & EmpBayes	θ_{n+1}	✗	✓	✗
Noisy CP	θ_{n+1}	✓	✗ [‡]	✗
LatentCP	θ_{n+1}	✓	✓	✓

[†]covers Y , not θ ; [‡]inverse estimation.

LatentCP is the only row that targets θ_{n+1} and clears all three columns.

Framework Overview



LatentCP: Calibrate, Invert and Aggregate

Step 1 — response-space conformal set

For $\gamma \in (0, \alpha)$, construct the response-space conformal set $C_\gamma(x; \mathcal{D}_n)$ satisfying

$$\mathbb{P}\{Y_{n+1} \in C_\gamma(X_{n+1}; \mathcal{D}_n)\} \geq 1 - \gamma.$$

Step 2 — forward compatibility and set inversion

For each θ , measure how much probability its forward law assigns to the response set:

$$p_\gamma(\theta, x) := \mathbb{P}_{Y \sim P_\theta(\cdot \mid x)}\{Y \in C_\gamma(x; \mathcal{D}_n)\},$$

Keep θ only when it is sufficiently compatible, i.e., it puts at most γ/α of its probability outside C_γ .

$$\mathcal{U}_\alpha(x; \gamma) := \left\{ \theta \in \Theta : p_\gamma(\theta, x) \geq 1 - \frac{\gamma}{\alpha} \right\}$$

Step 3 — Multilevel Miscoverage Aggregation

Any choice of $\gamma \in (0, \alpha)$ gives a valid set, but the size of $\mathcal{U}_\alpha(x; \gamma)$ can vary dramatically with γ . We consider a randomized miscoverage level $\gamma \sim \nu$ defined over $\Gamma \subset (0, \alpha)$. Define the normalized incompatibility score:

$$e_\gamma(\theta, x) := \frac{1 - p_\gamma(\theta, x)}{\gamma}, \quad e_\nu(\theta, x) = \int_\Gamma e_\gamma(\theta, x) d\nu(\gamma).$$

and consider the latent set $\mathcal{U}_\alpha(x; \nu)$ that aggregates over γ and seek to minimize its expected size:

$$\mathcal{U}_\alpha(x; \nu) := \left\{ \theta \in \Theta : e_\nu(\theta, x) \leq \frac{1}{\alpha} \right\}, \quad \nu^* \in \arg \min_\nu \mathbb{E}[|\mathcal{U}_\alpha(X_{n+1}; \nu)|].$$

$\nu = \delta_\gamma$ recovers the single-level set.

Theoretical Guarantees

Validity theorem. If $\{(X_i, \theta_i)\}_{i=1}^{n+1}$ are exchangeable with $Y_i \sim P_{\theta_i}(\cdot \mid X_i)$ conditionally independent, and ν is fixed independently of the calibration sample and the test point, then

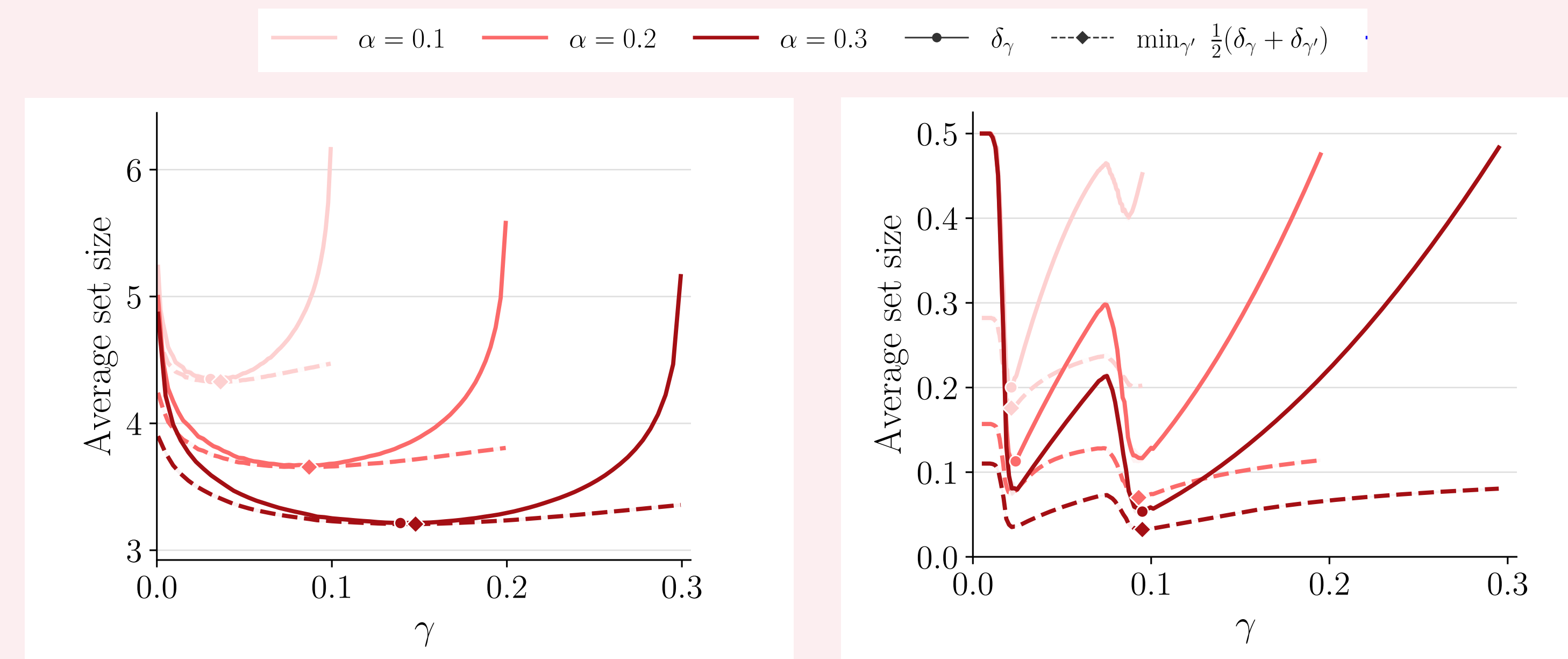
$$\mathbb{P}\{\theta_{n+1} \in \mathcal{U}_\alpha(X_{n+1}; \nu)\} \geq 1 - \alpha.$$

Distribution-free in π ; finite n ; no latent labels. With $\nu = \delta_\gamma$ this covers the single-level construction.

Sandwich proposition. For $\Gamma = \{\gamma_1, \dots, \gamma_K\} \subset (0, \alpha)$,

$$\bigcap_k \mathcal{U}_\alpha(x; \gamma_k) \subseteq \mathcal{U}_\alpha(x; \nu) \subseteq \bigcup_k \mathcal{U}_\alpha(x; \gamma_k).$$

Aggregation combines complementary evidence.



Geometry of the Latent Sets

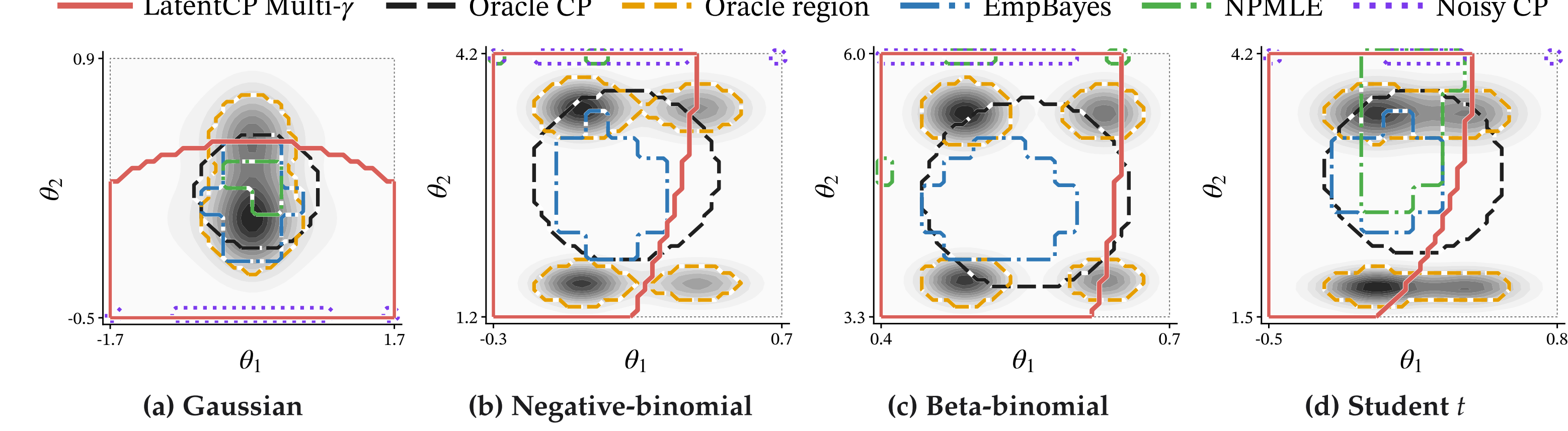


Figure: Gray shading is the true latent density (evaluation only). LatentCP is the only method valid across all four cases.

Coverage vs. Efficiency

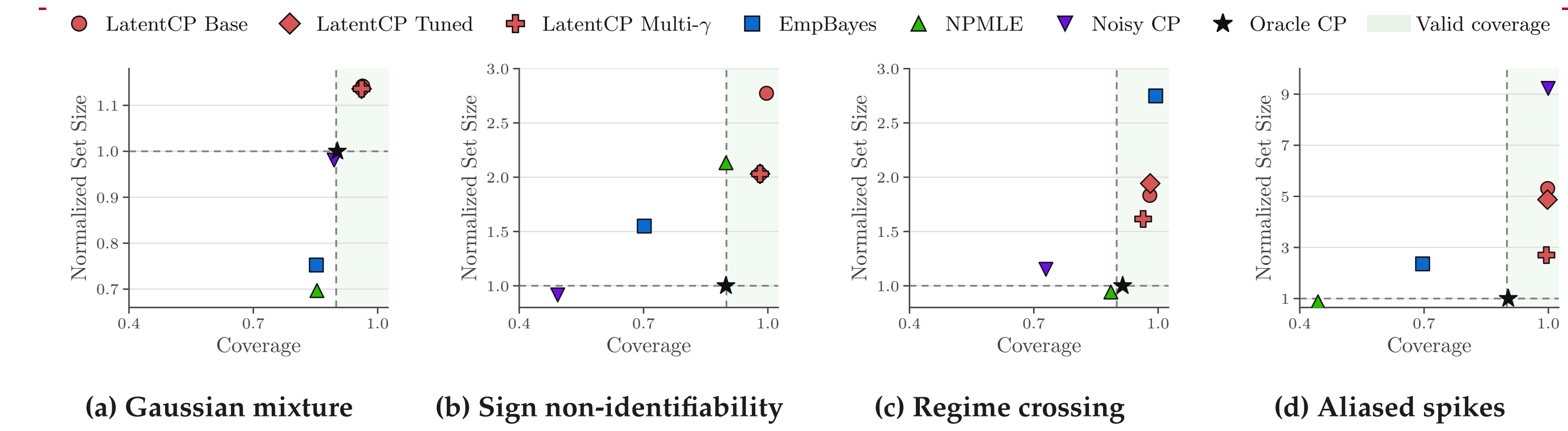


Figure: Coverage versus normalized set size at $1 - \alpha = 0.9$. All LatentCP variants attain nominal coverage across four inverse-problem stress tests. Among valid methods, optimized multi- γ aggregation yields the smallest sets.

California Wildfires Application

For spatial unit s and period t , the fire count follows the forward model

$$Y_{s,t} \mid (X_{s,t}, \lambda_{s,t}) \sim \text{Poisson}(\Delta_{s,t} \lambda_{s,t}),$$

where $\lambda_{s,t}$ is the latent fire intensity and $\Delta_{s,t}$ is exposure.

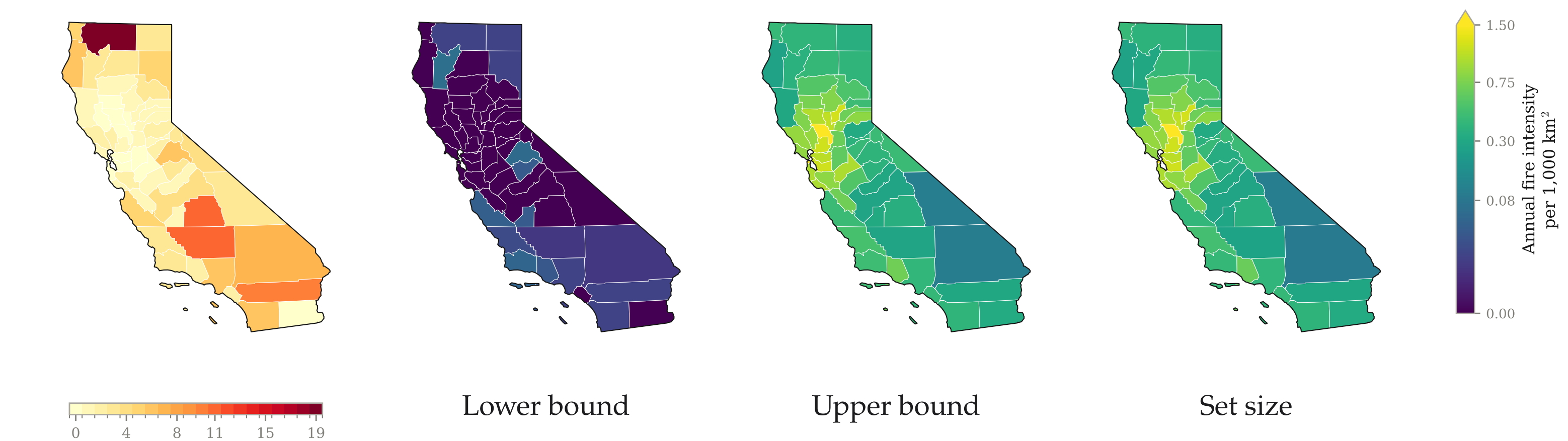


Figure: Upper and lower endpoints and set widths reveal spatial variation in latent fire-intensity uncertainty.

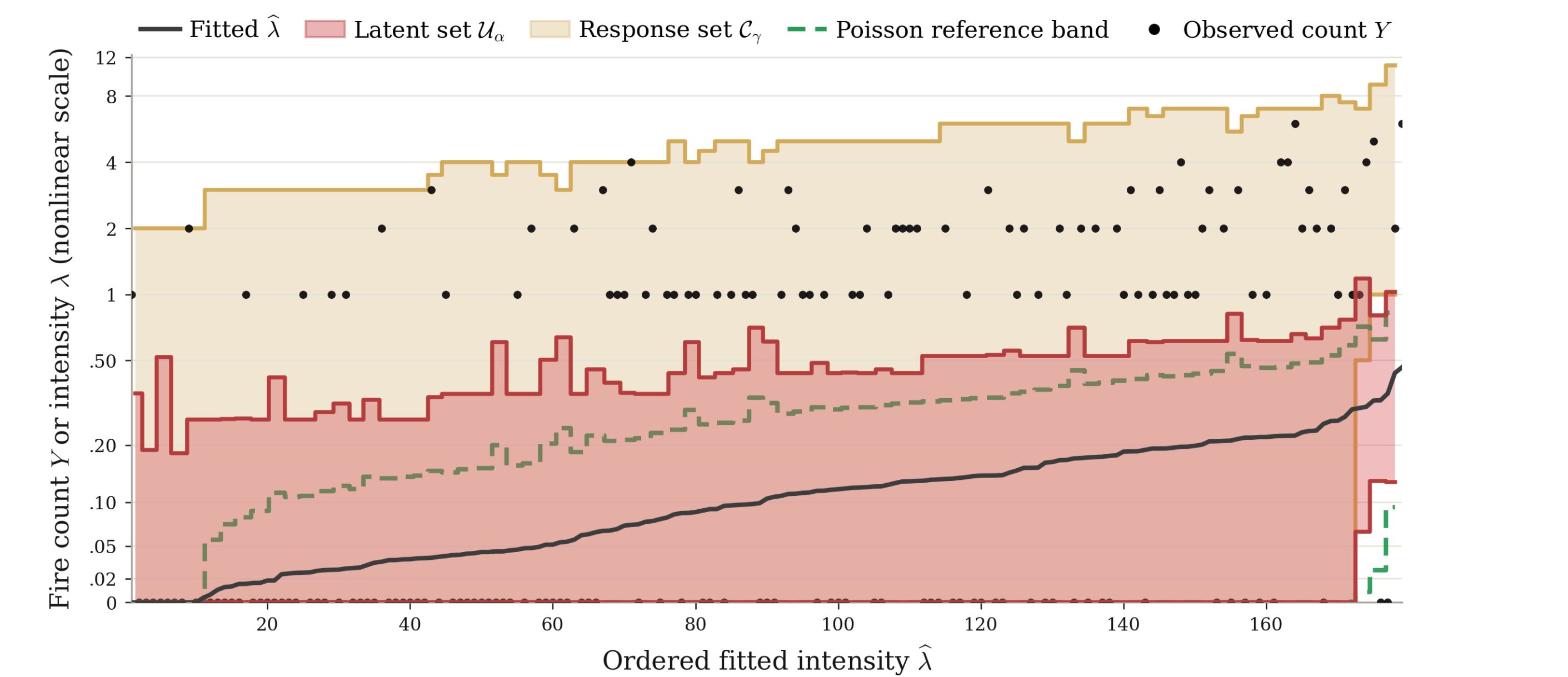


Figure: Response and latent uncertainty sets across spatial cells, ordered by fitted intensity. C_γ targets future fire counts, whereas $\mathcal{U}_\alpha$ targets latent fire intensity.