MVP: Practical Adversarial Multivalid Conformal Prediction

George Noarov, University of Pennsylvania

With Osbert Bastani, Varun Gupta, Chris Jung, Ramya Ramalingam, and Aaron Roth

Prediction Sets and Conformal Prediction

- Traditionally: given features $x \in \mathcal{X}$, produce accurate point estimate for label $y_x \in \mathcal{Y}$
- A different perspective: create a *prediction set* $T(x) \subseteq \mathcal{Y}$ that contains y_x with probability 0.9:

$$\Pr_{(x,y_x)}[y_x \in T(x)] = 0.9 \text{ ("valid 0.9 marginal coverage")}$$

- **Conformal prediction**: A widely adopted paradigm for building prediction sets:

1. Pre-train a *conformal score function* $s(x,y) \in \mathbb{R}$: higher values $\Rightarrow$ more disagreement between x, y
2. Given x, compute a *threshold* q and output prediction set $T(x) = \{y: s(x,y) \leq q\}$

- Conformal guarantees: exchangeable dataset $\Rightarrow$ valid 0.9 coverage on test data, no matter the score

Our contribution: MVP (MultiValid Prediction)

Vanilla Conformal Prediction

- Offline (batch) setting: a separate training/calibration set and a test set
- Requires I.I.D. or exchangeable data
- Marginal coverage guarantees

Our Method: MVP

- Online setting: data revealed sequentially, used both for training and testing
- Works even for adversarial data
- **MultiValid coverage**: Stronger than marginal:
  - Valid coverage on arbitrary feature space regions
  - Threshold Calibration (validity conditional on the predicted threshold)

MultiValidity ⇒ Group Conditional Coverage

- Given a group collection $\mathcal{G} = \{G_1, G_2, \dots, G_n\}$ where each $G_i \subseteq \mathcal{X}$ (groups can overlap)
  - If $x \in \mathcal{X}$ are individuals and $y \in Y$ their credit scores, groups G_i could be demographic groups
  - If $x \in \mathcal{X}$ encode market data and $y \in Y$ represent stock volatility, groups G_i could be market events
- MultiValid coverage ⇒ valid 0.9 coverage conditional on $x \in G_i$ for all i
- Ensures that no group receives unfairly bad coverage

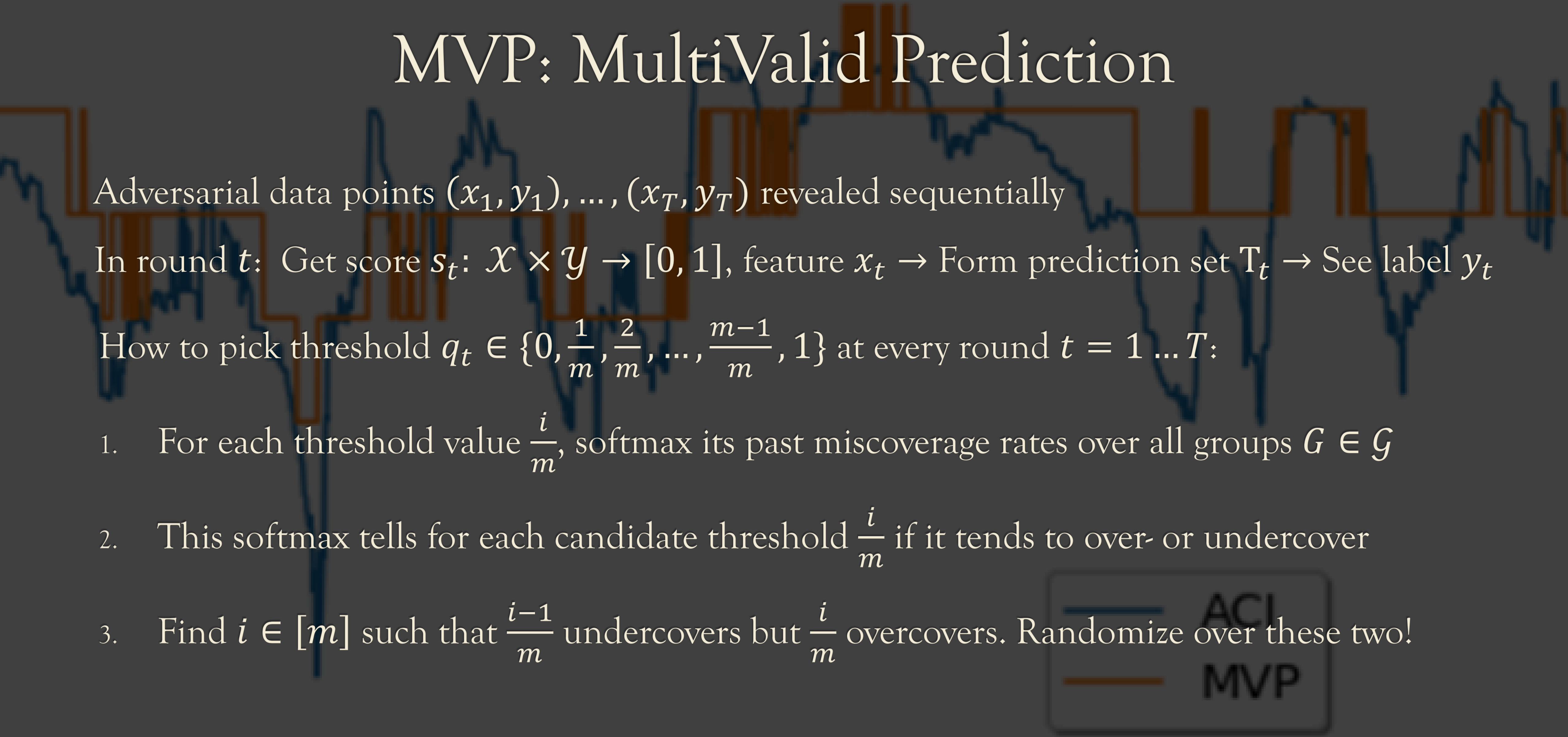
MVP: MultiValid Prediction

Adversarial data points $(x_1, y_1), \ldots, (x_T, y_T)$ revealed sequentially

In round t: Get score $s_t\colon \mathcal{X} \times \mathcal{Y} \to [0,1]$, feature x_t → Form prediction set T_t → See label y_t

How to pick threshold $q_t \in \{0, \frac{1}{m}, \frac{2}{m}, \ldots, \frac{m-1}{m}, 1\}$ at every round $t = 1 \ldots T$:

1. For each threshold value $\frac{i}{m}$, softmax its past miscoverage rates over all groups $G \in \mathcal{G}$
2. This softmax tells for each candidate threshold $\frac{i}{m}$ if it tends to over- or undercover
3. Find $i \in [m]$ such that $\frac{i-1}{m}$ undercovers but $\frac{i}{m}$ overcovers. Randomize over these two!



Empirical Performance

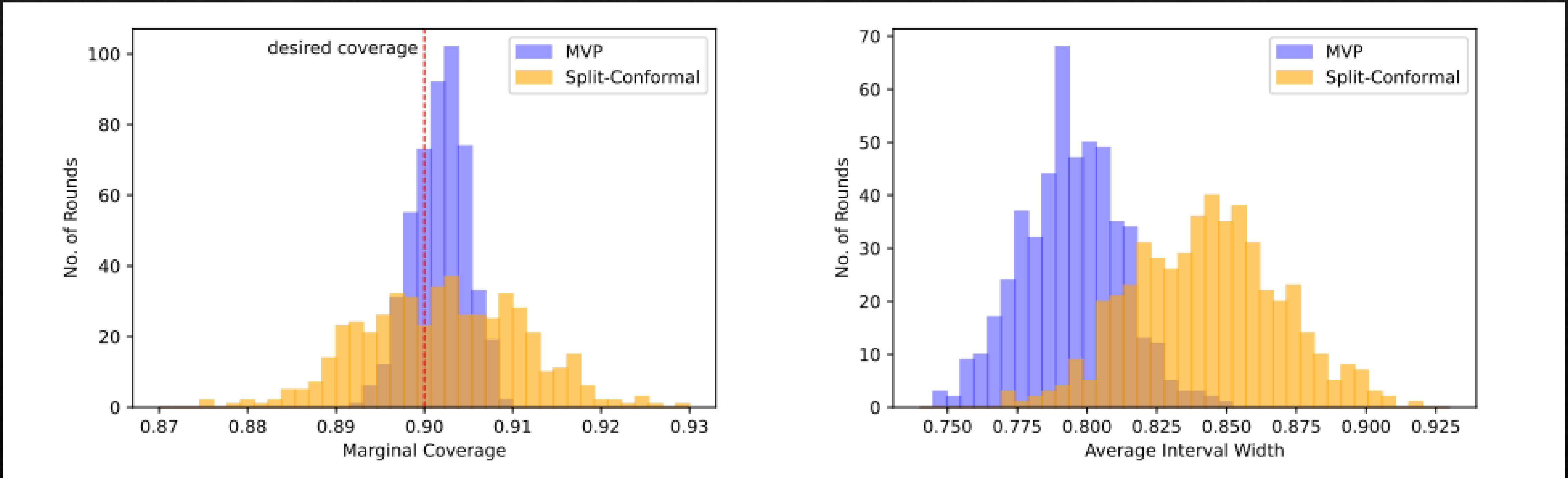
Strong coverage guarantees on various kinds of data:

- IID/Exchangeable data
- Covariate shift
- Time series
- Adversarial data

Matches/exceeds performance of existing methods "on their turf":

- Split conformal prediction [Lei et al.]
- Conformal prediction under covariate shift [Tibshirani et al.]
- Conservative nonoverlapping group-conditional coverage [Foygel Barber et al.]
- ACI [Gibbs and Candes]

Empirical Performance



Thanks!

Practical Adversarial Multivalid Conformal Prediction

Osbert Bastani, Varun Gupta, Christopher Jung,
Georgy Noarov, Ramya Ramalingam, Aaron Roth