Joint Coverage Regions: Simultaneous Confidence and Prediction Sets

Edgar Dobriban* & Zhanran Lin*

Department of Statistics & Data Science, The Wharton School, University of Pennsylvania

Wharton UNIVERSITY of PENNSYLVANIA

Introduction and Background

- Modern distribution-free approaches for prediction [2, 3]:
  - model-free and parameter-free;
  - based on exchangeability (or i.i.d.) assumptions and quantiles;
  - finite-sample guarantee;
- For parametric structures? Classical ways to consider:
  - parameters: estimation, confidence intervals;
  - Future observables: prediction, prediction regions;
- Unify these two ideas —— we propose Joint Coverage Regions.

Joint Coverage Regions (JCRs)

Consider a class of distribution $\mathcal{P}$ with a functional $\theta : \mathcal{P} \to \Theta$ for some parameter space Θ. For a random variable Z generated from some distribution $P \in \mathcal{P}$, assume that $o(Z)$ is the observed part of Z.

Definition: We say that J is a $1-\alpha$-joint coverage region (JCR) for (θ, Z) if for all $P \in \mathcal{P}$ we have

$$\mathbb{P}_{Z\sim P}\Big((\theta(P), Z) \in J\,(o(Z))\Big) \geq 1-\alpha.$$

- Goal: Given α, we aim to seek for a $1-\alpha$-JCR J for (θ, Z).
- A generalization of both confidence intervals and prediction sets.

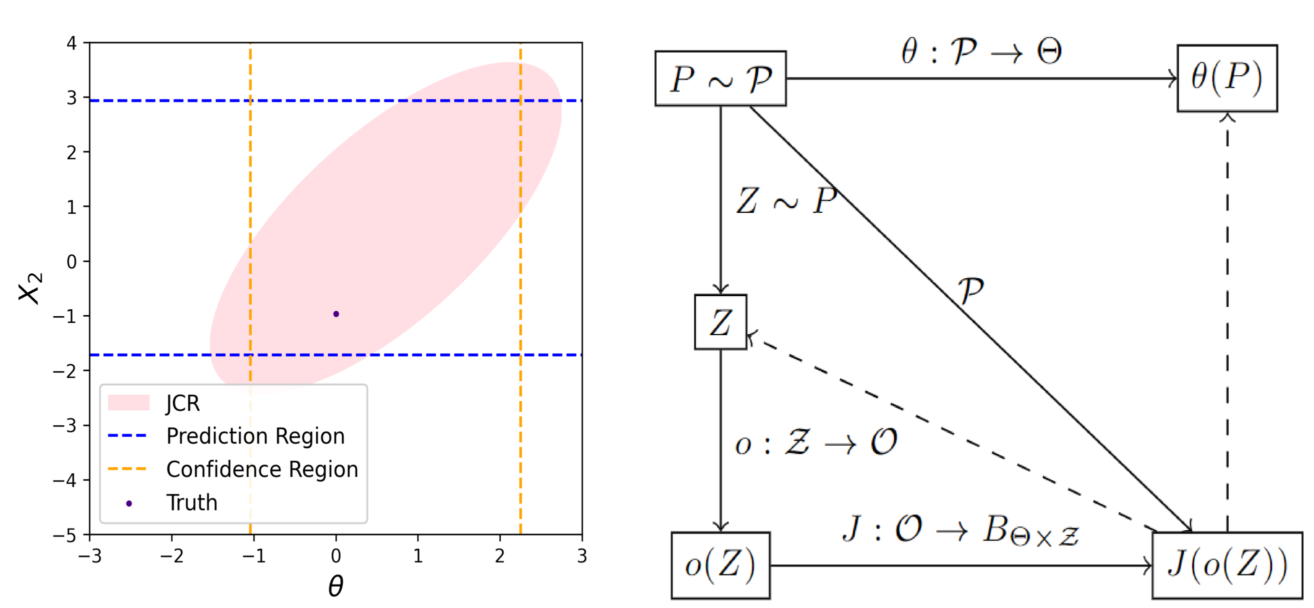

Figure 1. Left: A visualization of the JCR $\{(\theta, X_2) : (X_1-\theta)^2 + (X_2-\theta)^2 \leq \chi^2_{1-\alpha}(2)\}$ under the model $X_1, X_2 \sim \mathcal{N}(\theta, 1)$ with observation $o(X_1, X_2) = x_1$. We show a single trial with $\theta = 0$, $\alpha = 0.1$ and $x_1 = 0.606$. For contrast, we also plot a confidence interval $x_1 \pm q_{1-\alpha/2}$ for θ and a prediction region $x_1 \pm \sqrt{2}q_{1-\alpha/2}$ for X_2. The purple point labeled "Truth" shows the true realization $\theta = 0$ and $x_2 = -0.962$ in this trial. Right: A visual representation of our observation model.

Theoretical Constructions

- **Pivot**: Suppose that L is a pivot, in the sense that when $Z \sim P$ for $P \in \mathcal{P}$, the distribution Q of $L(\theta(P), Z)$ is known and does not depend on P. Let S be a measurable set such that $Q(S) \geq 1-\alpha$. Then, we have a $1-\alpha$-JCR
$$J(o^*) = \{(\theta, z) : o(z) = o^*,\ L(\theta, z) \in S\}\,.$$
An exact and informative pivot might be hard to derive in some cases.
- **Conditional pivots**: If L has distribution $Q(v)$ conditionally on $V(\theta(P), Z) = v$, and S satisfies that for a.e. v, $Q(S(v)) \geq 1-\alpha$, we have $1-\alpha$-JCR:
$$J(o^*) = \{(\theta, z) : o(z) = o^*,\ L(\theta, z) \in S(V(\theta, z))\}\,.$$
- **Test-Statistics Approach and Randomization.** For fixed $K \geq 1$, we sample
$$M = (M_1, \ldots, M_K),\ M_i \overset{\text{iid}}{\sim} m(Q(V(\theta, z))) \text{ for } i \in [K]$$
for some test-statistic m. Then, we can let $\alpha' = \lfloor (K+1)\alpha \rfloor / K$, and construct the randomized JCR by
$$J(o^*) = \left\{(\theta, z) : o(z) = o^*,\ m(L(\theta, z)) > q_{\alpha'}(\{M\}), M \sim m(Q(V(\theta, z))^K\right\},$$
where $q_\alpha(M)$ denotes the α-quantile of the distribution M. This is a special case of conditional pivots.
- **Group Invariance.** Suppose that for some *invariant function* $I : \Theta \times \mathcal{Z} \to \mathcal{I}$ and for some group $\mathcal{G}$ acting on $\mathcal{I}$ we have for all $P \in \mathcal{P}$ and all $g \in \mathcal{G}$ that when $Z \sim P$,
$$gI(\theta_P, Z) =_d I(\theta_P, Z).$$

We assume that $\mathcal{G}$ is a compact group having a left Haar measure U. Then we get a uniform distribution conditional on the orbit under $\mathcal{G}$.

Theorem: Under the assumptions above, we sample $G_1, \ldots, G_K$ i.i.d. from U. Define
$$J(o^*) = \left\{(\theta, z) : m(I(\theta, z)) > q_{\alpha'}\big(m(G_1 I(\theta, z)), \ldots, m(G_K I(\theta, z))\big), o(z) = o^*\right\},$$
where $\alpha' = \lfloor \alpha(K+1) \rfloor / K$. Then J is a $1-\alpha$ joint coverage region:
$$\mathbb{P}_{Z, G_1, \ldots, G_K}\big((\theta_P, Z) \in J(o(Z))\big) \geq 1-\alpha.$$

Group Invariance in Statistics [1]: spherical noises (orthogonal invariant); i.i.d./exchangeability (permutation invariant); the sign-flip transforms $\{\pm 1\}^n$ (symmetry); etc.

A Case Study and Simulation

Consider a multiple inference problem with two random variables X_1, X_2 that satisfy $X_1 \sim \mathcal{N}(\theta, 1)$, $X_2 \sim X_1 + \mathcal{N}(\theta, 1)$. Suppose that we only observe x_1, i.e., $o(X_1, X_2) = x_1$ and we aim to

(1) construct a valid confidence region C_α for θ;

(2) construct a valid prediction region T_α for X_2;

while controlling the probability of miscoverage.

- Classical approaches:
$$C_\alpha : \theta \in x_1 \pm q_{\alpha/2} \text{ and } T_\alpha : X_2 \in 2x_1 \pm \sqrt{2}q_{1-\alpha/2}.$$
- **Limitations:** (1) Due to multiplicity, we need an additional correction, e.g., the Bonferroni correction. (2) Using the same data x_1 for both tasks may fail to avoid the multiplicity correlation, which may make the FWER mis-calibrated.
- JCR Approach: we aim to construct joint coverage region J_α and control the miscoverage rate $\mathbb{P}(I_{12} \neq 0)$, where $I_{12} = I((\theta, X_2) \notin J_\alpha)$ as a joint indicator.
- Specifically, we may consider the JCR
$$J_\alpha = \{(\theta, X_2) : X_2 - 2\theta \in [\sqrt{2}q_{\alpha/2}, \sqrt{2}q_{1-\alpha/2}]\}, \tag{1}$$
which covers θ and X_2 simultaneously with $\mathbb{P}(I_{12} \neq 0) = \mathbb{P}((\theta, X_2) \notin J_\alpha) \leq \alpha$. Formally, since $X_1 - \theta \sim \mathcal{N}(0, 1)$, $X_2 - X_1 - \theta \sim \mathcal{N}(0, 1)$, we have the pivot $X_2 - 2\theta \sim \mathcal{N}(0, 2)$, so that
$$\mathbb{P}((\theta, X_2) \notin J_\alpha) = \mathbb{P}(X_2 - 2\theta \notin [\sqrt{2}q_{\alpha/2}, \sqrt{2}q_{1-\alpha/2}]) = \alpha.$$
Of course, we may also use the pivot $X_2 - X_1 - \theta \sim \mathcal{N}(0, 1)$, defining
$$J'_\alpha(x_1) = \{(\theta, X_2) : X_2 - x_1 - \theta \in [q_{\alpha/2}, q_{1-\alpha/2}]\}. \tag{2}$$
- We can further intersect the JCRs in (1), (2) with confidence regions for θ to obtain bounded JCRs. For instance, we can intersect
$$J_{\alpha/2} \text{ or } J'_{\alpha/2} \text{ with } C_{\alpha/2} = \{\theta \in x_1 \pm q_{\alpha/4}\}$$
to yield a slightly conservative region with coverage rate over $1-\alpha$. Figure 2 shows the regions $J_\alpha, J'_\alpha, C_{\alpha/2}, T_{\alpha/2}$ as defined. as well as the intersections $J_{\alpha/2} \cap C_{\alpha/2}$, $J'_{\alpha/2} \cap C_{\alpha/2}$. It is shown that JCR approach better captures problem structure, as well as getting smaller and more reasonable region. Numerical simulation is shown in Table 1 to compare coverage.

Inferential Region	Coverage Rate	CPCIs
$C_{\alpha/2} \times T_{\alpha/2}$	91.93%	[91.38%, 92.46%]
J_α	89.91%	[89.30%, 90.49%]
$J_{\alpha/2} \cap C_{\alpha/2}$	90.23%	[89.63%, 90.81%]

Table 1. To validate coverage, we take $\theta = 0$ and run $10,000$ independent trials. In each trial, we record the following events: $(\theta, X_2) \in C_{\alpha/2} \times T_{\alpha/2}$, $(\theta, X_2) \in J_\alpha$, $(\theta, X_2) \in J_{\alpha/2} \cap C_{\alpha/2}$. We compute the coverage rates and their corresponding Clopper-Pearson CIs (CPCIs) for $\alpha = 0.1$.

As expected, the region J_α, $J_{\alpha/2} \cap C_{\alpha/2}$ yield desirable coverage, while the intersection JCR $C_{\alpha/2} \times T_{\alpha/2}$ is more so.

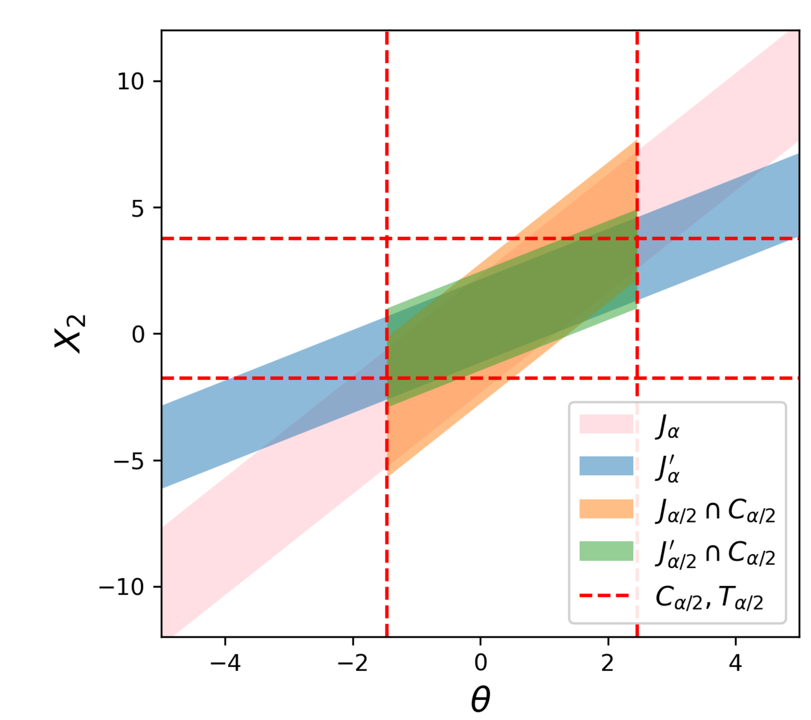

Figure 2. A visualization for $J_\alpha, C_{\alpha/2}, T_{\alpha/2}$ for a single trial with observation $x_1 = 0.5$.

References

[1] Morris L Eaton.
Group invariance applications in statistics.
IMS, 1989.

[2] Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan J Tibshirani, and Larry Wasserman.
Distribution-free predictive inference for regression.
Journal of the American Statistical Association, 113(523):1094–1111, 2018.

[3] Glenn Shafer and Vladimir Vovk.
A tutorial on conformal prediction.
Journal of Machine Learning Research, 9(Mar):371–421, 2008.