We thank the reviewers for their thorough comments. One common comment from the reviewers is that they would like to see experiments on more than one data set. If the reviewers find this point to be critical, we would be happy to add another experiment to the supplement in the camera-ready version (we have already completed additional experiments, but left them out because the results were very similar to those on the airfoil data—the theoretical guarantees we provide are exact in finite sample without any distributions assumptions, so in the absence of estimating the likelihood ratio for the weights, there should be no surprises when running our weighted conformal prediction on a new data set).
Below we respond to some of the specific comments and questions in the reviews. Square brackets [...] indicate the reviewer's comments and our replies follow.

Reviewer #2

Thank you for your detailed and positive feedback.
[...analysis of the effects of approximating the likelihood ratio...] Great question; if we have a guarantee on estimating $w(x) = \mathsf{d}\widetilde{P}_X(x)/\mathsf{d}P_X(x)$ in the ℓ_∞ (essential supremum) norm, i.e., if we have bound on

$$\|\hat{w} - w\|_\infty = \text{ess sup}_{x \in \mathbb{R}^d} |\hat{w}(x) - w(x)|,$$

then we can show how to translate this into an approximate coverage guarantee on the conformal prediction intervals. We can include a corollary to this effect in the main paper, with a proof in the supplement. This would of course be a first step: an ℓ_∞ accuracy guarantee may be stringent, but a more careful analysis should possible in more specialized situations, out of the scope of the current paper.
[...classification...] This is a great point. The same methodology extends seamlessly to classification (only the notation changes) and we can point this out in the camera-ready version.
[...code...] We apologize for not including an anonymized link. Our code is publicly available on github and we will add the link as soon as the paper can be de-anonymized.

Reviewer #4

Thank you for your detailed and helpful review.
[...additional experiments...] Please see our comment at the top of the page.
[...analysis of the effects of approximating the likelihood ratio...] Please see our response to Reviewer #2.
[...computational tractability...] Yes, in general the complexity may be combinatorial. However, as mentioned in the paper, fortunately for covariate shift (and a few other scenarios mentioned in the discussion), the calculation is easier.

Reviewer #5

Thank you for your detailed and astute review.
[...transfer learning...] As far as we can tell, this is not directly related the goal of our paper, but as transfer learning is often discussed alongside covariate shift, it would interesting to consider this connection; space permitting we can include a comment in the camera-ready version.
[...importance sampling...] These ideas are intimately connected to our work, and form the basis for much of the methodology in covariate shift, though the modern covariate shift literature simply tends to use different language than the older importance sampling literature in statistics. One could use importance sampling to approximately fix the incompatibility of test and training distributions in conformal prediction (sample from a batch of test points with probabilities that are proportional to their likelihood in the training sample). Our methodology is in essence a *direct* way of doing this, without requiring subsampling, and is preferable since subsampling would introduce additional variance. We can include this interpretation in the camera-ready if accepted.
[...experiments...] Beyond what we mentioned at the top of this page, the experiment is indeed proof-of-concept. But we do believe it shows the method to be practically applicable; we have investigated several angles of the problem, including what happens when we still estimate a likelihood ratio and there is actually no covariate shift (results given in the supplement). Regarding testing whether our method is "competitive", as far as we know, there are no other methods that produce provably valid prediction intervals (with distribution-free, finite-sample coverage guarantees) under covariate shift, so we do not know of competing methods that we might compare against.
[...detect drift incrementally...] This is an interesting question. There is orthogonal work on using conformal prediction to detect a change in distribution: see "Inductive Conformal Martingales for Change-Point Detection". Future work could consider sequential versions of a problem of both detecting covariate shift and producing appropriately corrected prediction intervals, combining the insights of our paper with theirs.
[...generalization to each point coming from different distribution...] While we do do not know of any examples or applications that need this full generality, it is very plausible that the test points lie in (multiple) unknown clusters, with each cluster coming from a different distribution shift. This type of distribution could possibly be detected by mixture methods, correcting for each cluster individually. We are still exploring the full power of the general technique and will continue to do so in future work, but it is outside the scope of this short paper.