Homework 4

Advanced Topics in Statistical Learning, Spring 2023

Due Friday April 14 at 5pm

1 Basic fact about CDFs and quantiles [14 points]

In this exercise, we'll walk through a number of basic but important facts about quantiles and cumulative distribution functions (CDFs). Let F be a CDF, of the form

$$F(x) = \mathbb{P}(X \leq x), \quad x \in \mathbb{R},$$

for some real-valued random variable X. Let Q be the corresponding quantile function,

$$Q(t) = \inf\{x : F(x) \geq t\}, \quad t \in [0,1].$$

(This is often denoted as $Q = F^{-1}$, even when the inverse of F does not exist in the usual sense.) We note that F is always nondecreasing and right-continuous; the latter says, for any x,

$$F(x) = \lim_{y \to x^+} F(y)$$

(where $y \to x^+$ means that y approaches x from the right).

(a) Prove that for any x and any t, [3 pts]

$$F(x) \geq t \iff Q(t) \leq x.$$

This is sometimes called the *Galois inequality* for the quantile function. Hint: one direction follows from the definition of Q, and the other is a consequence of right-continuity of F.

(b) Use part (a) to prove that if $U \sim \mathrm{Unif}(0,1)$, then $Q(U)$ is distributed according to F (which means it has CDF F). [2 pts]

(c) Use part (a) to prove that for any t, [2 pts]

$$F(Q(t)) \geq t,$$

with equality if and only if t is in the range of F.

(d) Use parts (b) and (c) to prove that if X is distributed according to F, then $F(X)$ is sub-uniform, which means that for any t, [3 pts]

$$\mathbb{P}(F(X) \leq t) \leq t.$$

with equality $\mathbb{P}(F(X) \leq t) = t$ if and only if t is in the closure of the range of F. Hint: you may start by replacing X with $Q(U)$, for $U \sim \mathrm{Unif}(0,1)$, as they have the same distribution.

(e) Give a concrete worked example to show when equality fails in the result in part (d). [2 pts]

(f) We can always achieve equality in part (d) via auxiliary randomization. Define

$$F^*(x; v) = \lim_{y \to x^-} F(y) + v \cdot \Big(F(x) - \lim_{y \to x^-} F(y)\Big),$$

where $y \to x^-$ means that y approaches x from the left. Show empirically, by revisiting your example in part (e), that for $V \sim \mathrm{Unif}(0,1)$, independent of X, and for any t, [2 pts]

$$\mathbb{P}(F^*(X; V) \leq t) = t.$$

Bonus: prove it!

2 Calibration-conditional beta coverage [15 points]

Let D_1 and D_2 be an arbitrary partition of $\{1, \dots, n\}$ into sets of sizes $n_1 = |D_1|$ and $n_2 = |D_2|$. Recall, in lecture, we stated that for the split conformal prediction band $\hat{C}_n$, computed on a proper training points (X_i, Y_i), $i \in D_1$ and a calibration points (X_i, Y_i), $i \in D_2$, it holds that

$$\mathbb{P}\Big(Y_{n+1} \in \hat{C}_n(X_{n+1}) \,\Big|\, (X_i, Y_i),\, i = 1, \dots, n\Big) \sim \mathrm{Beta}(k_\alpha, n_2 + 1 - k_\alpha), \tag{1}$$

where $k_\alpha = \lceil (1-\alpha)(n_2 + 1) \rceil$. This assumes that (X_i, Y_i), $i = 1, \dots, n+1$ were all i.i.d. draws from some arbitrary distribution P. In what follows, you will work towards a proof of the result in (1).

(a) Let $U_1, \dots, U_n$ be i.i.d. from $\mathrm{Unif}(0, 1)$, and denote their *order statistics* by

$$U_{(1)} \le U_{(2)} \le \cdots \le U_{(n)}.$$

Fix any $1 \le k \le n$. Prove that the density $f_{(k)}$ of $U_{(k)}$ is given by [4 pts]

$$f_{(k)}(x) = n \binom{n-1}{k-1} x^{k-1} (1-x)^{n-k}.$$

Hint: calculate the probability that $U_{(k)}$ lies in $[x, x+\epsilon]$, then send $\epsilon \to 0$.

(b) From the result in part (a), argue that [1 pt]

$$U_{(k)} \sim \mathrm{Beta}(k, n+1-k).$$

(You just need to recall the form of the beta distribution; you don't need to do any hard calculations here.)

(c) Now let $X_1, \dots, X_{n+1}$ be i.i.d. according to F, and assume that F is continuous. Prove that [3 pts]

$$\mathbb{P}\big(X_{n+1} \le X_{(k)} \,|\, X_1, \dots, X_n\big) \sim \mathrm{Beta}(k, n+1-k).$$

Hint: let $U_i = F(X_i)$, $i = 1, \dots, n+1$. Use Q1 part (d) to argue that each $U_i \sim \mathrm{Unif}(0, 1)$. Then use the result from part (b) of this question.

(d) Use part (c) to show that the split conformal prediction result (1) holds whenever the distribution of each conformity score $V(X_i, Y_i)$ is continuous. [2 pts]

(e) Carry out a simulation to empirically verify (1) for split conformal prediction. You may fix α, but examine a few different values of the calibration set size n_2. [5 pts]

(f) As a bonus, show that the result in part (c) still holds when the CDF F is not necessarily continuous, and therefore the result in (1) holds in general, even when each $R_i = V(X_i, Y_i)$ does not have a continuous distribution, provided we break ties uniformly at random.

Hint: using Q1 part (b), argue that you can replace each X_i with $Q(U_i)$, where Q denotes the quantile function associated with F, and U_i, $i = 1, \dots, n$ are i.i.d. from $\mathrm{Unif}(0, 1)$.

3 X-conditional coverage: impossible! [18 points]

In lecture, we briefly discussed an impossibility result for X-conditional coverage in a distribution-free setting. This exercise investigates X-conditional coverage more deeply from theoretical and practical angles.

(a) Recite and prove any of the impossibility results in Lei and Wasserman (2014); Vovk (2012); Barber et al. (2021). Note: it's perfectly fine for your solution to follow directly from what's in any of those papers—no new arguments are needed—but make sure to be fully precise in stating the result and providing its proof, all in your own words. [8 pts]

(b) Conduct a simulation to show that, nonetheless, different conformity scores can give rise to appreciably different results in terms of approximate conditional coverage in practice. By "approximate" here, we mean some kind of local coverage, where we average over a small neighborhood of a point x.

Start with the absolute residual as your conformity score, and then compare its approximate conditional coverage performance to at least one other conformity score (a number of choices are available), where you have designed the latter to do better in your simulation. Note: split conformal is fine to use throughout. [10 pts]

References

Rina Foygel Barber, Emmanuel J. Candès, Aaditya Ramdas, and Ryan J. Tibshirani. The limits of distribution-free conditional predictive inference. *Information and Inference*, 10(2):455–482, 2021.

Jing Lei and Larry Wasserman. Distribution-free prediction bands for non-parametric regression. *Journal of the Royal Statistical Society: Series B*, 76(1):71–96, 2014.

Vladimir Vovk. Conditional validity of inductive conformal predictors. *Asian Conference on Machine Learning*, 2012.