STATS 300C: Theory of Statistics **Spring 2022**

Lecture 7 — April 19, 2022

Lecturer: Prof. Emmanuel Candès *Editor: Parth Nobel, Scribe: Rahul Kanekar*

Warning: *These notes may contain factual and/or typographic errors. They are based on Emmanuel Candès's course from 2018 to 2022, Lihua Lei's lecture slides from 2022 and scribe notes written by Anav Sood.*

7.1 Outline:

- The PRDS property
- Examples of models with PRDS
- FDR control under PRDS
- A taste of conformal inference
- Conformal p-values are PRDS

7.2 PRDS Property

We begin by defining the notion of an increasing subset

Definition 1 (Increasing Subset of $\mathbb{R}^n$)**.** A subset $D \subset \mathbb{R}^n$ is said to be increasing if for all $x \in D, y \geq x$ implies $y \in D$ where the comparison of y and x is component-wise.

Intuitively, increasing sets ar those that have no boundaries in the North and East directions.

A decreasing set is defined in the same way, only with the opposite relation.
Using this, we define the PRDS property

Definition 2 (PRDS property)**.** A family of random variables $(X_1, ..., X_n)$ is said to be PRDS (Positive Regression Dependence on a subset) on a subset $I_0 \subset \{1, ..., n\}$ if for all $i \in I_0$, the function $\mathbb{P}((X_1, ..., X_n) \in D \mid X_i = x)$ is an increasing function in x for any increasing subset D.

To start, we make some observations about increasing subsets and the PRDS property.

- A set D is increasing iff D^c is decreasing. Thus, $(X_1, ..., X_n)$ is PRDS iff $\mathbb{P}((X_1, ..., X_n) \in D^c \mid X_i = x)$ is decreasing in x for any increasing set D.
- If $(X_1, ..., X_n)$ is PRDS on I_0 and if $Y_i := f_i(X_i)$ for all $1 \leq i \leq n$ where each f_i is strictly increasing or decreasing, then $(Y_1, ..., Y_n)$ is PRDS on I_0 as well. Transformation of this form are called co-monotone transformations. Thus the PRDS property is preserved under co-monotone transformations.

- If $(X_1, ..., X_n)$ is PRDS on I_0 (the set of true nulls), then both $p_i = F_{H_i}(X_i)$ (the right sided p-values) and $p_i = \bar{F}_{H_i}(X_i)$ (the left sided p-values) are PRDS as well. This follows from the fact that the CDF and survival functions are co-monotone transform and hence the p-values are PRDS by the preceeding observation.

7.3 Example of a family with PRDS property

In this section we will prove the following:

Claim: Let $X := (X_1, ..., X_n)$ be a multivariate Gaussian vector with distribution $\mathcal{N}(\mu, \Sigma)$. X is PRDS on $I_0 \iff \Sigma_{ij} \geq 0$ for all $i \in I_0$ and $1 \leq j \leq n$.

Proof. To prove the PRDS property we need to prove that $\mathbb{P}(X_{-i} \in D \mid X_i = x)$ is increasing in x for all increasing sets D, where X_{-i} denotes the vector X with only the ith component removed. WLOG we can assume that $i = 1$ since the argument is same for all $i \in I_0$. This is because we will only use the fact that the we are dealing with a Gaussian with some positive correlations.

We first set up some notation:

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_{-1} \end{pmatrix} \qquad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{1,-1} \\ \Sigma_{-1,1} & \Sigma_{-1,-1} \end{bmatrix}$$

where Σ_{11} is the top left entry of Σ i.e. the variance of $X_1$and μ_1 is the first entry of μ and hence the mean of X_1.

With this notation, the required conditional distribution can be written as follows:

$$X_{-1} \mid X_1 = x \sim \mathcal{N}(\mu_{-1} + \Sigma_{-1,1}\Sigma_{11}^{-1}(x - \mu_1), \Sigma_{-1,-1} - \Sigma_{-1,1}\Sigma_{11}^{-1}\Sigma_{1,-1})$$

Since $\Sigma_{j1} \geq 0$ for all j, we have that $\Sigma_{-1,1} \geq 0$ entrywise. As a result, the conditional mean is non-decreasing in x. Further, the conditional covariance matrix is the same for all x. Hence for any $x \geq y \in \mathbb{R}$, if μ_x, μ_y denote the conditional means of X_{-1} then $\mu_x \geq \mu_y$ entrywise. In particular, if U, V are random variables having the conditional distribution of X_{-1} given $X_1 = x$ and y respectively, then $U \sim w + V$ where $w := \mu_x - \mu_y \geq 0$. In particular, We now have the following chain of arguments:

$$\begin{aligned}
\mathbb{P}(X \in D \mid X_1 = x) &= \mathbb{P}((x, X_{-1}) \in D \mid X_1 = x) && \\
&= \mathbb{P}((x, U) \in D) && U \sim X_{-1} \mid X_1 = x \\
&\geq \mathbb{P}((y, U - w) \in D) && x \geq y, w \geq 0 \text{ and } D \text{ is increasing} \\
&= \mathbb{P}((y, V) \in D) && \text{Since } V \sim U - w \\
&= \mathbb{P}(y, X_{-1} \in \mid X_1 = y) && \\
&= \mathbb{P}((X_1, X_{-1}) \in D \mid X_1 = y) && \\
&= \mathbb{P}(X \in D \mid X_1 = y) &&
\end{aligned}$$

This shows that the conditional probability is a non-decreasing function and hence the family is PRDS.

Conversely, if we want to prove that PRDS implies that all correlations are non-negative, we can proceed by contradiction. Assume that there is some $\Sigma_{j1} < 0$. Then the conditional distribution expression given before still holds which gives that

$$X_j \mid X_1 = x \sim \mathcal{N}(\mu_j + \Sigma_{j1}\Sigma_{11}^{-1}(x - \mu_1), \sigma_j^2)$$

where σ_j^2 is a consant since the covariance does not change with x. The conditional mean is seen to be a strictly decreasing function of x since $\Sigma_{j1} < 0$ and $\Sigma_{11} > 0$ which gives that the conditional probability of the event $\{X_j \geq \mu_j\}$ is strictly decreasing in x. Since the set $\{x_j \geq \mu_j\}$ is increasing, we have a contradiction to the PRDS property.

□

7.4 FDR control under PRDS

Theorem 1. (Benjamini and Yekutieli, 2001) If the joint distribution of the p-values $p_1, .., p_n$ is PRDS on the set of true nulls $\mathcal{H}_0$, then the BH procedure controls the FDR at level $\frac{\alpha n_0}{n}$.

Remark 1. As noted before, PRDS property translates from statistics to one sided p-values. Hence, to apply the above theorem, we can simply check PRDS property on the statistics itself.

Remark 2. The good thing about this theorem is that it asserts FDR control without assuming any dependence structure on the non-null p-values. This is desirable since we usually don't know about the structure of the non-null p-values. However, it does assume the PRDS property which involves knowing how the non-nulls relate to the true nulls which is generally not well known. Thus the theorem is difficult to apply in practice.

We will use the following proposition proved in Appendix A.

Proposition 1. If the p-values are PRDS on the set of true nulls, then the function $\mathbb{P}((p_1, ..., p_n) \in D \mid p_i \leq t)$ is non-decreasing in t for D an increasing set.

Proof. (Proof of Theorem 1) We assume WLOG that $H_{01}, ..., H_{0n_0}$ are the true nulls. We know that

$$\text{FDR} = \mathbb{E}\left(\sum_{i=1}^{n_0} \frac{V_i}{R \vee 1}\right)$$

where $V_i = \mathbf{1}(\text{reject } \mathcal{H}_{0i})$ and R is the number of rejections. Note that it is enough to show $\mathbb{E}\left(\frac{V_i}{R\vee 1}\right) \leq \frac{\alpha}{n}$ for $i = 1, ..., n_0$.

Observe that if k rejections are made, then $\mathcal{H}_i$ is rejected iff $p_i \leq \frac{\alpha k}{n}$. Hence, $V_i = \mathbf{1}(p_i \leq \frac{\alpha R}{n})$. Using this we get that

$$\mathbb{E}\left(\frac{V_i}{R \vee 1}\right) = \sum_{k=1}^{n} \frac{\mathbb{P}\left(p_i \leq \frac{\alpha k}{n}, R = k\right)}{k}$$

We now have for any true null $\mathcal{H}_i$,

$$\begin{aligned}
\sum_{k=1}^n \frac{\mathbb{P}\left(p_i \le \frac{\alpha k}{n}, R = k\right)}{k} &= \sum_{k=1}^n \mathbb{P}\left(p_i \le \frac{\alpha k}{n}\right) \frac{\mathbb{P}(R = k \mid p_i \le \frac{\alpha k}{n})}{k} \\
&\le \sum_{k=1}^n \frac{\alpha k}{n} \frac{\mathbb{P}(R = k \mid p_i \le \frac{\alpha k}{n})}{k} && \text{(Superuniformity under null)} \\
&= \frac{\alpha}{n} \sum_{k=1}^n \mathbb{P}\left(R = k \mid p_i \le \frac{\alpha k}{n}\right)
\end{aligned}$$

Hence, it suffices to show that $\sum_{k=1}^n \mathbb{P}\left(R = k \mid p_i \le \frac{\alpha k}{n}\right) \le 1$. Observe that $\{R \le k\}$ is an increasing event i.e. it can be written as $\{(p_1, ..., p_n) \in D\}$ for some increasing set. This is because increasing all p-values increases the p-value at each rank. Hence, any ranked p-value above the threshold remains above it's threshold i.e. we accept at least as many as before and hence, do not reject more hypotheses. Using this, we get that

$$\begin{aligned}
\sum_{k=1}^n \mathbb{P}\left(R = k \mid p_i \le \frac{\alpha k}{n}\right) &= \mathbb{P}(R \le n \mid p_i \le \alpha) - \mathbb{P}\left(R \le 0 \mid p_i \le \frac{\alpha}{n}\right) \\
&\quad + \sum_{k=1}^{n-1} \left(\mathbb{P}\left(R \le k \mid p_i \le \frac{\alpha k}{n}\right) - \mathbb{P}\left(R \le k \mid p_i \le \frac{\alpha(k+1)}{n}\right)\right)
\end{aligned}$$

Note that each of the summands in the summation in the second expression are non-positive since $\mathbb{P}(\{R \le k\} \mid p_i \le x)$ is increasing in x. Also, $\mathbb{P}\left(R \le 0 \mid p_i \le \frac{\alpha}{n}\right) \ge 0$ which implies that

$$\sum_{k=1}^n \mathbb{P}\left(R = k \mid p_i \le \frac{\alpha k}{n}\right) \le \mathbb{P}(R \le n \mid p_i \le \alpha) \le 1.$$

As stated before, this proves the upper bound on the FDR.

□

Example 1. If we want to test $H_{0i} : \mu_i = 0$ versus $H_{1i} : \mu_i > 0$ for $X \sim \mathcal{N}(\mu, \Sigma)$ with $\Sigma_{ij} \ge 0$ for all i, j then our one sided test statistics will result in PRDS p-values. As a result, we can test using BH and still control FDR. Similarly we can test with FDR control when the alternative is $\mu < 0$. However for the two sided hypotheses, the statistics we use i.e. $|X_i|$ do not form a co-monotone transformation and hence the resulting p-values might not be PRDS. In fact, determining if using the BH procedure on two-sided p-values controls FDR is an open problem.

7.5 Conformal p-values and PRDS

As we saw in the previous section, the Benjamini-Hochberg procedure controls the FDR under the PRDS condition. However for quite some time there weren't many generic examples of models that produced PRDS p-values. The two main examples over the last 20 years were the Gaussian model with non-negative correlations (Section 7.3) and the two sided multivariate t-test with independent p-values.

In the past few years, there have been more examples of PRDS values:

- One-sided test with convex combinations of log-concave statistics [2]
- One-sided test with recursive order statistics [2]
- Conformal p-values [1]

The next section gives a brief description of conformal inference and conformal p-values followed by a short proof of conformal p-values being PRDS.

7.6 A taste of conformal inference

As machine learning becomes more prevalent in decision making, it is increasingly important to gauge what the margin of error on the prediction is. In conformal inference, we seek prediction algorithms that returns a range of values—within which it predicts the output will lie—rather than simply a point estimate. In other words, we are looking for something along the lines of a confidence interval rather than a single value. This is particularly crucial when the machine learning algorithm is being used in setting such as predicting GPA's which can influence fairly important decisions. An estimate of a student' potential GPA as $[3.35, 3.45]$ is a more powerful prediction than $[2.8, 4.0]$ though the midpoint of both intervals is 3.4.

A naive (and incorrect) way to build a confidence region would be to fit the machine learning model using a training set, obtain an estimator $\hat{\mu}$ and then find the $\lceil (n+1)(1-\alpha) \rceil$ smallest element of the residuals $\{|\hat{\mu}(x) - y|\}_{(x,y)}$ where (x, y) ranges across the training data. We can then define $\hat{C}(X) = [\hat{\mu}(X) - q, \hat{\mu}(X) + q]$ for X independent of the training data. However, the residuals on the training set will underestimate the residuals on a general input X and the set $\hat{C}(X)$ will not cover Y with a high probability.

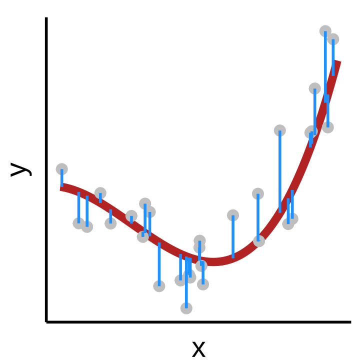

Figure 7.1: Training data residuals

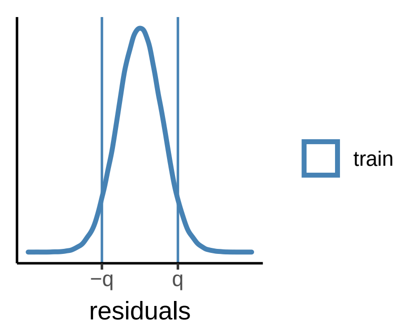

Figure 7.2: Histogram of training data residuals

The drawback of this idea is that we are calculating residuals on inputs where the algorithm has already been fit to do a good job. The residuals are bound to be less because we have selected $\hat{\mu}$ so that they are minimized. The residuals on the test set independent of the training set will be higher than the residuals on the training set. The following picture makes this clearer.

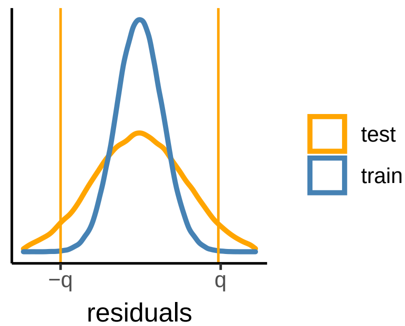

Figure 7.3: Comparison of residuals

What we need is a measure of how the residuals of $\hat{\mu}$ behave on an independent set. This is where the idea of a calibration set comes in.

A calibration set is a data set, independent of the training set on which we calculate the residuals. Generally the model will be fit on a part of the data set with the other data held out. The calibration set will be this held out data set. It turns out that using the quantiles of the residuals on the calibration set and estimating in the way described above does give a confidence interval with high coverage.

Theorem 2. (Papadoupoulos, Proedrou, Vovk, Gammerman '02, [3]) Let $\{(X_i, y_i)\}_{i=1}^n$ be the calibration set and q be the $\lceil (n+1)(1-\alpha) \rceil$ smallest element of $|y_i - \hat{\mu}(x_i)|$ on the calibration set. Define $\hat{C}(X) = [\hat{\mu}(X_{n+1}) - q, \hat{\mu}(X_{n+1}) + q]$. Then

$$\mathbb{P}(Y_{n+1} \in \hat{C}(X_{n+1})) \geq 1 - \alpha.$$

The proof of the above result used the fact that if we are given i.i.d (or exchangeable) statistics $T_1, ..., T_n$ and if $r_1, ..., r_n$ denote the rankings of these statistics, then all rankings are equally likely. Using this idea, we can define conformal p-values. The next section will be dedicated to proving that the conformal p-values are indeed p-values and that they have the PRDS property.

7.7 PRDS property of conformal p-values

We first introduce the idea of conformity scores and how they can be used to obtain p-values.

Suppose we have a real valued function s and $Z_1, ..., Z_{n+1}$ are drawn exchangeably from a distribution P. We define a *conformal p-value* as follows:

$$p(z) = \frac{1 + |\{1 \leq i \leq n : s(Z_i) \leq s(z)\}|}{n+1}.$$

The function s is called a conformity score and it is decided beforehand, perhaps in the course of fitting a model. It is a measure of how much a new observation conforms to previous data. A high score means high conformity, a low score means low conformity. The goal of defining the conformal p-value is that we now use it to test if new data points are from the same distribution as $Z_1, ..., Z_n \sim P$. The set $\{s(Z_1), ..., s(Z_n)\}$ gives an idea of how well data from P conform to previous data. If a new data point Z_{n+1} conforms much more poorly than $Z_1, ..., Z_n$ to previous data, that counts as evidence against it being from P. Thus, conformal p-values can be used to test for equality of distributions without any knowledge of the actual underlying distributions.

The following result makes this intuition rigorous.

Theorem 3. Assuming that $s(Z_1), ..., s(Z_n)$ are almost surely distinct, then $p(Z_{n+1})$ is a valid p-value i.e. $p(Z_{n+1})$ is uniformly distributed over $\left\{\frac{1}{n+1}, ..., \frac{n}{n+1}, 1\right\}$.

Given $Z_1, ..., Z_n \sim P$ and $Z_1, ..., Z_n \perp\!\!\!\perp Z_{n+1}, ..., Z_{n+m}$, we can now define the conformal p-values $p_1 = p(Z_{n+1}), ..., p_m = p(Z_{n+m})$. The p-value p_i can be used to test the hypothesis $H_{0,i} : Z_{n+i} \sim P$ as shown in Theorem 3.

If we use $s(x, y) = -|y - \hat{\mu}(x)|$ as our conformity score then inverting the p-values actually gives the predictive intervals from the previous section. We give short proof of this below. If $Z_{n+1} = (X, Y)$, then

$$\begin{aligned}
Y \notin \hat{C}(X) &\iff |Y - \hat{\mu}(X)| \leq \lceil (n+1)(1-\alpha) \rceil \text{smallest value of } |Y_i - \hat{\mu}(X_i)| \\
&\iff |\{1 \leq i \leq n : s(Z_i)\}| \leq \lfloor (n+1)\alpha \rfloor - 1 \\
&\iff p(Z_{n+1}) \leq \frac{\lfloor (n+1)\alpha \rfloor}{n+1} \\
&\implies p(Z_{n+1}) \leq \alpha
\end{aligned}$$

This proves the correspondence between p-values and predictive intervals.

Finally, we have a recent result from [1] that shows that conformal p-values are PRDS.

Theorem 4. If the distribution of $s(Z_i)$ is continuous for all $1 \leq i \leq n$, then the conformal p-values are PRDS on the set of true nulls.

Proof. Once again, it is enough to show the monotonicity of conditional probability only with respect to the first p-value. Assume $H_1, ..., H_{n_0}$ are the true nulls. Let $S_j := s(Z_j)$ for $1 \leq j \leq n + m$.

Note that $(p_2, ..., p_m)$ is a deterministic function of p_1 and $W_1 := (S_{(1)}, ...S_{(n+1)}, S_{n+2}, ..., S_{n+m})$ where W_1 denotes the order statistics. This is because, given the ranking of $S_1, ..., S_{n+1}$ and also the p-value p_1, we can figure out the rank of S_{n+1} among $S_1, ..., S_{n+1}$. From this we find the sorted order of the statistics $S_1, ..., S_n$. The p-values $p_2, ..., p_m$ depend only on $S_{n+2}, ..., S_{n+m}$ and the sorted order of $S_1, ..., S_n$. Hence, we can calculate $p_2, ..., p_m$ from p_1

and W_1. We can now say that $(p_2, ..., p_m) = G(p_1, W_1)$.

We now show that the function $G(p_1, W_1)$ is increasing in p_1. For this, we will show that increasing p_1 and keeping W_1 fixed does not decrease any of the p-values.

$$\begin{aligned}
p_1 = \frac{k}{n+1} &\implies \{S_1, ..., S_n\} = \{S_{(1)}, ..., S_{(k-1)}, S_{(k+1)}, ..., S_{(n+1)}\} \\
&\implies p_j = \frac{1}{n+1}\left\{1 + \sum_{i \neq k, i \leq n+1} \mathbf{1}(S_{n+j} \geq S_{(i)})\right\} \\
&\implies p_j = \frac{1}{n+1}\left\{1 - \mathbf{1}(S_{n+j} \geq S_{(k)}) + \sum_{i=1}^{n+1} \mathbf{1}(S_{n+j} \geq S_{(i)})\right\}
\end{aligned}$$

Note that the last term $\left\{1 - \mathbf{1}(S_{n+j} \geq S_{(k)}) + \sum_{i=1}^{n+1} \mathbf{1}(S_{n_j} \geq S_{(i)})\right\}$ increases with p_1 provided everything else is fixed. Hence, $G(p_1, W_1)$ is increasing in p_1.

Finally, we note that p_1 is independent of W_1 under the null. The null here means that $Z_{n+1} \sim P$. In that case, we know that p_j is uniform over $\{\frac{0}{n+1}, ..., 1\}$. Also, given $\{S_{(1)}, ..., S_{(n+1)}\}$ the value of S_{n+1} is uniform over this set since the random variables are i.i.d under the null and hence, p_1 has the same distribution even after conditioning which shows that $p_1 \perp\!\!\!\perp (S_{(1)}, ..., S_{(n+1)})$. Since $Z_{n+1} \perp\!\!\!\perp Z_{n+2}, ..., Z_{n+m}$ by assumption, we have that $p_1 \perp\!\!\!\perp W_1$. This enables the following calculation. For any increasing set D,

$$\begin{aligned}
\mathbb{P}((p_2, ..., p_m) \in D \mid p_1 = x) &= \mathbb{P}(G(p_1, W_1) \in D \mid p_1 = x) \\
&= \mathbb{E}_{W_1}(\mathbb{P}(G(p_1, W_1) \mid p_1 = x, W_1)) \\
&= \mathbb{E}_{W_1}(\mathbf{1}(G(x, W_1) \in D)) \qquad\qquad (p_1 \perp\!\!\!\perp W_1)
\end{aligned}$$

Since $G(p_1, W_1)$ is increasing in p_1 and D is an increasing set, $\mathbf{1}(G(x, W_1) \in D)$ is increasing in x which implies the PRDS property.

$\square$

Appendix A

Claim: If $\mathbf{p} = (p_1, ..., p_n)$ is PRDS on the set of true nulls, then for any increasing set D, $\mathbb{P}((p_1, ..., p_n) \in D \mid p_i \leq t)$ is non-decreasing in t for any i such that H_i is null.

Proof. For any t, $\mathbb{P}(\mathbf{p} \in D \mid p_i \leq t) = \frac{\mathbb{P}(\mathbf{p} \in D, p_i \leq t)}{\mathbb{P}(p_i \leq t)}$. For $t' > t$, we have that

$$\mathbb{P}(\mathbf{p} \in D \mid p_i \leq t') = \frac{\mathbb{P}(\mathbf{p} \in D, p_i \leq t) + \mathbb{P}(\mathbf{p} \in D, p_i \in (t, t'])}{\mathbb{P}(p_i \leq t) + \mathbb{P}(p_i \in (t, t'])}$$

It suffices to show that

$$\frac{\mathbb{P}(\mathbf{p} \in D, p_i \leq t)}{\mathbb{P}(p_i \leq t)} \leq \frac{\mathbb{P}(\mathbf{p} \in D, p_i \in (t, t'])}{\mathbb{P}(p_i \in (t, t'])}$$

The last statement is because for any positive number a, b, c, d, $\frac{a}{b} \leq \frac{a+c}{b+d} \iff \frac{a}{b} \leq \frac{c}{d}$. Note that if F_i denotes the distribution of p_i then

$$\begin{aligned}
\mathbb{P}(\mathbf{p} \in D, p_i \leq t) &= \int_0^t \mathbb{P}(\mathbf{p} \in D \mid p_i = s) dF_i(s) \\
&\leq \int_0^t \mathbb{P}(\mathbf{p} \in D \mid p_i = t) dF_i(s) && \text{(By PRDS)} \\
&= \mathbb{P}(\mathbf{p} \in D \mid p_i = t)\mathbb{P}(p_i \leq t) \\
&\implies \frac{\mathbb{P}(\mathbf{p} \in D, p_i \leq t)}{\mathbb{P}(p_i \leq t)} \leq \mathbb{P}(\mathbf{p} \in D \mid p_i = t)
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
\mathbb{P}(\mathbf{p} \in D, p_i \in (t, t']) &= \int_t^{t'} \mathbb{P}(\mathbf{p} \in D \mid p_i = s) dF_i(s) \\
&\geq \int_t^{t'} \mathbb{P}(\mathbf{p} \in D \mid p_i = t) dF_i(s) && \text{(By PRDS)} \\
&= \mathbb{P}(\mathbf{p} \in D \mid p_i = t)\mathbb{P}(p_i \in (t, t']) \\
&\implies \frac{\mathbb{P}(\mathbf{p} \in D, p_i \in (t, t'])}{\mathbb{P}(p_i \in (t, t'])} \geq \mathbb{P}(\mathbf{p} \in D \mid p_i = t)
\end{aligned}$$

Hence,

$$\frac{\mathbb{P}(\mathbf{p} \in D, p_i \in (t, t'])}{\mathbb{P}(p_i \in (t, t'])} \geq \frac{\mathbb{P}(\mathbf{p} \in D, p_i \leq t)}{\mathbb{P}(p_i \leq t)}$$

which completes the proof. $\square$

Bibliography

[1] Stephen Bates, Emmanuel Candès, Lihua Lei, Yaniv Romano, and Matteo Sesia. Testing for outliers with conformal p-values. *arXiv preprint arXiv:2104.08279*, 2021.

[2] Jackson H Loper, Lihua Lei, William Fithian, and Wesley Tansey. Smoothed nested testing on directed acyclic graphs. *arXiv preprint arXiv:1911.09200*, 2019.

[3] Harris Papadopoulos, Kostas Proedrou, Volodya Vovk, and Alex Gammerman. Inductive confidence machines for regression. In *European Conference on Machine Learning*, pages 345–356. Springer, 2002.