8.1 Outline

Agenda: False discovery rate, the PRDS property, conformal inference

1. The PRDS property
2. An example of a PRDS family
3. FDR control under the PRDS property
4. A taste of conformal inference
5. Conformal p-values are PRDS

8.2 The PRDS property

We begin first by defining the notion of increasing/decreasing sets. In the below $x \geq y$ means that $x_i \geq y_i$ for all coordinates.

Definition 1. A set $D \in \mathbb{R}^n$ is called increasing if $x \in D$ and $y \geq x$ implies $y \in D$. (These sets have no boundaries in the North-East directions).

Now we define the PRDS property.

Definition 2. A family of random variables $X = (X_1, \ldots, X_n)$ is PRDS [positive regression dependent on each of a subset] on $I_0$ if for any increasing set and each $i \in I_0$, $\mathbb{P}((X_1, \ldots, X_n) \in D \mid X_i = x)$ is increasing in $x$.

We make a few observations concerning this definition.

- The PRDS property is invariant under co-monotone transformations. If $Y_i = f_i(X_i)$ where the $f_i$’s are either all increasing or all decreasing, then $X$ is PRDS implies that $Y$ is also PRDS.
• $D$ is increasing if and only if $D^c$ is decreasing. Here a decreasing set is a set $D$ such that $x \in D$ and $y \leq x$ implies $y \in D$. As a consequence, we have that a random vector $X$ is PRDS if and only if for any decreasing $D$, $P(X \in D|X_i = x)$ is decreasing in $x$.

• If $\{X_i\}$ is PRDS on $I_0$ (true nulls), then $p_i = F_{H_i}(X_i)$ [right-sided $p$-value] and $p_i = F_{H_i}(X_i)$ [left-sided $p$-value] are both PRDS. For the two-sided test where $p_i = 2F_{H_i}(|X_i|)$ the $p_i$ may not be PRDS as $|X_i|$ is not a monotone transformation.

8.3 An example of a PRDS family

Claim: Consider a multivariate normal Gaussian distribution $X = (X_1, \ldots, X_n) \sim \mathcal{N}(\mu, \Sigma)$. If $\Sigma_{ij} \geq 0$ for all $i \in I_0$ and all $j$ then $(X_1, \ldots, X_n)$ is PRDS over $I_0$. The converse also holds.

Remark: With Gaussian data, PRDS is equivalent to non-negative correlations.

Proof. WLOG suppose that $1 \in I_0$ and consider

$$X = \begin{pmatrix} X_1 \\ X_{(-1)} \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_{(-1)} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{1(-1)} \\ \Sigma_{(-1)1} & \Sigma_{(-1)(-1)} \end{pmatrix}$$

Then the distribution of $X_{(-1)}|X_1 = x$ is given by

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

By our assumption $\Sigma_{(-1)1}$ has all non-negative elements so the conditional mean is non-decreasing in $x$. Note the conditional covariance stays the same. Under these conditions it follows that for a increasing set $D$, $x \leq x' \implies \mathbb{P}(X \in D|X_1 = x) \leq \mathbb{P}(X \in D|X_1 = x')$. To see why note that if $Z \sim \mathcal{N}(\mu_1, \Sigma)$ and $Y \sim \mathcal{N}(\mu_2, \Sigma)$ where $\mu_2 \geq \mu_1$ then letting $v := \mu_2 - \mu_1 \geq 0$ we have that $Z + v \overset{d}{=} Y$. For an increasing set $D$ we have that $Z \in D \implies Z + v \in D$ so $\mathbb{P}(Z \in D) \leq \mathbb{P}(Z + v \in D) = \mathbb{P}(Y \in D)$. Finally, to show $X$ is PRDS on $I_0$ repeat this argument for all $i \in I_0$.

To establish the converse we show that if there is some $i \in I_0$ and $j$ such that $\Sigma_{ij} < 0$ then $X$ is not PRDS. To see this consider the increasing set $D = \{X_j \geq \mu_j\}$. By a similar argument as the above its easy to show that $P(X \in D|X_i = x) = P(X_j \geq \mu_j|X_i = x)$ is strictly decreasing in $x$. 

8.4 FDR control under the PRDS property

Theorem 1. (Benjamini & Yekutieli (2001)): If the joint distribution of the statistics (or joint dist. of the $p$-values) is PRDS on the set of true nulls $\mathcal{H}_0$, then the Benjamini-Hochberg procedure $BH(\alpha)$ controls the FDR at level $\alpha n_0/n$. ($BH(\alpha)$ may become conservative under positive dependence since $FDR = \alpha n_0/n$ under independence).
Remark: An important feature of this theorem is that it does not make explicit assumptions on the dependence structure among the non-null hypothesis. This is good from the point of view of applications where we typically have some knowledge of the phenomenon under the null but know (and are willing to assume) very little of the phenomenon under the alternative. Unfortunately, we typically don’t know much how about how the non-nulls depend on the nulls so it’s generally not known whether the statistics arising in a particular application are PRDS.

A consequence of the PRDS property is that for \( t \leq t' \),

\[
P(D|p_i \leq t) \leq P(D|p_i \leq t')
\]

if \( i \) is a null and the set \( D \) is the event that the null \( p_j \) are in an increasing set. For a proof of this fact, see Appendix A. We use this result in the proof of the Theorem below.

Proof. [E. Candés, R. Foygel Barber ]

We know that

\[
FDR = \mathbb{E} \left( \sum_{i \in \mathcal{H}_0} \frac{V_i}{1 \vee R} \right), \quad V_i = 1\{	ext{reject } H_i\}
\]

Recall that in the independent case: \( \mathbb{E} V_i/(1 \vee R) = \alpha/n \). Now we want to show that \( \mathbb{E} V_i/(1 \vee R) \leq \alpha/n \). This immediately implies that the FDR is at most \( \alpha n_0/n \). Set \( \alpha_k = \alpha k/n \) and note that for \( i \in \mathcal{H}_0 \)

\[
\frac{V_i}{1 \vee R} = \sum_{k \geq 1} \frac{1\{p_i \leq \alpha_k\}1\{R = k\}}{k}
\]

\[
= \sum_{k \geq 1} \frac{1\{p_i \leq \alpha_k\}(1\{R \leq k\} - 1\{R \leq k - 1\})}{k}
\]

\[
= \sum_{k=1}^{n-1} \left[ \frac{1\{p_i \leq \alpha_k\}}{k} - \frac{1\{p_i \leq \alpha_{k+1}\}}{k+1} \right] 1\{R \leq k\} + \frac{1\{R \leq n\}1\{p_i \leq \alpha\}}{n} \quad \text{(Integration by parts)}
\]

Since \( R \leq n \) always and \( p_i \sim U(0, 1) \) we know that

\[
\mathbb{E} \left( \frac{1\{R \leq n\}1\{p_i \leq \alpha\}}{n} \right) = \frac{\alpha}{n}
\]

Thus it suffices to show that

\[
\mathbb{E} \left( \sum_{k=1}^{n-1} \left[ \frac{1\{p_i \leq \alpha_k\}}{k} - \frac{1\{p_i \leq \alpha_{k+1}\}}{k+1} \right] 1\{R \leq k\} \right) \leq 0
\]
To see that this is the case note that for each $k$ we have

$$
\mathbb{E}\left( \left[ \frac{1}{k} \mathbb{1}\{p_i \leq \alpha_k\} - \frac{1}{k+1} \mathbb{1}\{p_i \leq \alpha_{k+1}\} \right] \mathbb{1}\{R \leq k\} \right) = \frac{\mathbb{P}(p_i \leq \alpha_k, R \leq k) - \mathbb{P}(p_i \leq \alpha_{k+1}, R \leq k)}{k+1}
$$

$$= \frac{\mathbb{P}(R \leq k | p_i \leq \alpha_k) \mathbb{P}(p_i \leq \alpha_k)}{k} - \frac{\mathbb{P}(R \leq k | p_i \leq \alpha_{k+1}) \mathbb{P}(p_i \leq \alpha_{k+1})}{k+1}
$$

$$= \frac{\alpha}{n} \left( \mathbb{P}(R \leq k | p_i \leq \alpha_k) - \mathbb{P}(R \leq k | p_i \leq \alpha_{k+1}) \right)
$$

where we’re able to apply the PRDS property since \{R \leq k\} is the event that the $p_i$ are in an increasing set (if we increase $p_i$, $R$ is non-increasing as we can only make fewer rejections) and $\alpha_{k+1} \geq \alpha_k$. \qed

**Example:** Suppose $X \sim \mathcal{N}(\mu, \Sigma)$ where $\Sigma_{ij} \geq 0$. If we wish to test $H_{0i} : \mu_i = 0$ vs $H_{1i} : \mu_i > 0$, our test statistics are PRDS (and thus our $p$-values will also be PRDS). The same conclusion holds if we wish to test $H_{0i} : \mu_i = 0$ vs $H_{1i} : \mu_i < 0$. However, for testing against the alternative $\mu_i \neq 0$, the test statistics $|X_i|$ no longer have the PRDS property.

The following Conjecture is open.

**Conjecture:** Suppose $X \sim \mathcal{N}(\mu, \Sigma)$ for general covariance matrix $\Sigma$ and we want to test the null $H_{0i} : \mu_i = 0$. Using the $BH(\alpha)$ procedure on $p$-values obtained from the individual 2-sided tests will control the $FDR$ at level $\alpha$.

### 8.5 A taste of conformal inference

Prior to discussing another interesting application of the PRDS property, we’ll take a brief detour and discuss conformal inference and some of its applications in machine learning. Some notable pioneers in developing conformal inference and introducing related ideas to the field of statistics include Vladmimir Voyk, Jing Lei, and Larry Wasserman.

As society allows predictive machine learning models to influence decision making in sensitive settings, it becomes increasingly important to get a reliable gauge on such predictive models’ uncertainty. For example, suppose we’re training a machine learning model to predict students’ future GPAs. A traditional model may output the prediction 3.94 for a particular set of input features. We instead would prefer a model which outputs $3.94 \pm ??$ where ?? characterizes a reasonable range the future GPA will likely lie in. This motivates the following definition.

**Definition 3.** Suppose $(X, Y) \sim P$ is a sample from some joint distribution. A $1 - \alpha$ prediction interval for $Y$ is a set $C(X)$ such that $P(Y \in C(X)) \geq 1 - \alpha$. 

---

8-4
For the typical supervised learning problem, we observe training data \( \{(X_i, Y_i)\}_{i=1}^n \). We will try and use the observed data to make a \( 1 - \alpha \) prediction interval \( C(X_{n+1}) \) for \( Y_{n+1} \) where \( (X_{n+1}, Y_{n+1}) \) is a test point not observed during the training process. Our goal will be to make such a valid prediction interval in the finite sample setting with minimal distributional assumptions.

### 8.5.1 Constructing prediction intervals

As a start, consider the following naive (and incorrect) approach to constructing a \( 1 - \alpha \) prediction interval.

**Algorithm 1: Naive \( 1 - \alpha \) prediction interval**

**Input:**
- Datapoints \( (X_1, Y_1), \ldots, (X_n, Y_n) \sim P \subseteq (\mathbb{R}^d, \mathbb{R}) \).
- Test point \( (X_{n+1}, Y_{n+1}) \sim P \).
- Regression algorithm \( \hat{\mu} \).
- Coverage level \( 1 - \alpha \in (0, 1) \).

**Process:**
- Fit regression function \( \hat{\mu} : \mathbb{R}^d \to \mathbb{R} \) on \( \{(X_i, Y_i) : i \in \{1, \ldots, n\}\} \).
- For \( i \in \{1, \ldots, n\} \) set \( r_i = |Y_i - \hat{\mu}(X_i)| \).
- Let \( q \) be the \([ (n + 1)(1 - \alpha) ]\)-th smallest value of \( \{r_i : i \in \{1, \ldots, n\}\} \).

**Output:**
- Prediction interval \( C(X_{n+1}) = [\hat{\mu}(X_{n+1}) - q, \hat{\mu}(X_{n+1}) + q] \)

While perhaps intuitive, the Algorithm 1 is incorrect. We can easily fit a regression model that perfectly fits our observed data. In this case, for the test point \( X_{n+1} \) the algorithm will return the singleton \( \{\hat{\mu}(X_{n+1})\} \). Under typical noise assumptions this certainly is not a \( 1 - \alpha \) prediction interval. In essence, the sizes of the training residuals are biased downwards (overfitting) and the naive algorithm does nothing to account for this. A concrete example of this is shown in Figure 8.1.

The following algorithm, motivated by the work in Gammerman et al. (1998), addresses this issue by splitting the data into a training set (in-sample data) and a hold-out set (out-of-sample data). To get an idea of the distribution of residuals our model will generate on
We generate \(n=100\) i.i.d. \((X_i, Y_i)\) pairs by sampling \(X_i \sim \text{Unif}([0, \frac{1}{100}, \frac{2}{100}, \ldots, 1])\), \(\epsilon_i \sim \mathcal{N}(0, 10)\), and letting \(Y_i = 1 + X_i + 50X_i^2 + \epsilon_i\). We fit an over-parameterized model (polynomial regression with degree \(p = 15\)) to half the data, which is our in-sample data, and examine the distribution of the in-sample and out-of-sample residuals.

future data, we only examine the residuals from the held-out data.

Algorithm 2: Correct \(1 - \alpha\) prediction interval

**Input:**
- Split sizes \(n_1 + n_2 = n\).
- Datapoints \((X_1, Y_1), \ldots, (X_n, Y_n) \sim P \subseteq (\mathbb{R}^d, \mathbb{R})\).
- Test point \((X_{n+1}, Y_{n+1}) \sim P\).
- Regression algorithm \(\hat{\mu}\).
- Coverage level \(1 - \alpha \in (0, 1)\).

**Process:**
- Randomly split \(\{1, \ldots, n\}\) into disjoint \(\mathcal{I}_1\) and \(\mathcal{I}_2\) with \(|\mathcal{I}_1| = n_1\) and \(|\mathcal{I}_2| = n_2\).
- Fit regression function \(\hat{\mu} : \mathbb{R}^d \to \mathbb{R}\) on \(\{(X_i, Y_i) : i \in \mathcal{I}_1\}\).
- For \(i \in \mathcal{I}_2\) set \(r_i = |Y_i - \hat{\mu}(X_i)|\).
- Let \(q\) be the \(\lceil((|\mathcal{I}_2| + 1)(1 - \alpha))\rceil\)-th smallest value of \(\{r_i : i \in \mathcal{I}_2\}\).

**Output:**
- Prediction interval \(C(X_{n+1}) = [\hat{\mu}(X_{n+1}) - q, \hat{\mu}(X_{n+1}) + q]\)

The above algorithm returns a valid \(1 - \alpha\) prediction interval under the assumption that our data \(\{(X_i, Y_i)\}_{i=1}^{n+1}\) are exchangeable. Recall \(\{Z_i\}_{i=1}^n\) being exchangeable means for any permutation \(\pi : \{1, \ldots, n\} \to \{1, \ldots, n\}\), \((Z_1, \ldots, Z_n) \overset{d}{=} (Z_{\pi(1)}, \ldots, Z_{\pi(n)})\). A proof of this theorem, stated below, will be given in the next sub-section.

**Theorem 2.** (Papadopoulos, Proedrou, Vovk & Gammerman (2002)) Under the assumption that \(\{(X_i, Y_i)\}_{i=1}^{n+1}\) are exchangeable, if \(C(X_{n+1})\) is the output of Algorithm 2 for datapoints \(\{(X_i, Y_i)\}_{i=1}^n\) and test point \((X_{n+1}, Y_{n+1})\) then \(P(Y_{n+1} \in C(X_{n+1})) \geq 1 - \alpha\).  

\(^1\)Technically we only prove the Theorem under the assumption that w.p. 1 no two residuals \(\hat{\mu}(X_i) - Y_i\) have the same absolute value.
Remark: It’s immediate that i.i.d. random variables are exchangeable. As a typical example of a collection of exchangeable random variables that are not i.i.d, consider \( Z_i = \rho Z + \epsilon_i \) for \( i \in \{1, \ldots, n\} \), where \( Z \) is some fixed random variable and the \( \epsilon_i \) are i.i.d.

8.5.2 Conformal \( p \)-values

Suppose we have data \( Z_1, \ldots, Z_n \) and \( Z_{n+1} \) drawn exchangeably from \( P \). We want to get an idea of if \( Z_{n+1} \) “conforms” to previous observations.

Definition 4. A conformity score \( s(z) \in \mathbb{R} \) is used to measure how much \( z \) corresponds to previous observations (high score \( \rightsquigarrow \) it conforms, low score \( \rightsquigarrow \) it does not conform). For a particular conformity score and observed data \( Z_1, \ldots, Z_n \), conformal \( p \)-values are given by

\[
p(z) = \frac{1 + \#\{i : s(Z_i) \leq s(z)\}}{n + 1}
\]

Theorem 3. Assuming that the conformity scores of the observed data \( \{s(Z_i)\}_{i=1}^{n+1} \) have no ties w.p. 1 then the conformal \( p \)-value \( p(Z_{n+1}) \) is a valid \( p \)-value. More specifically \( p(Z_{n+1}) \) is uniformly distributed over \( \{1/(n + 1), 2/(n + 1), \ldots, 1\} \).

Proof. Since the \( \{Z_i\}_{i=1}^{n+1} \) are exchangeable so are the \( \{s(Z_i)\}_{i=1}^{n+1} \). This immediately implies that \( s(Z_{n+1}) \)'s rank amongst the \( s(Z_1), \ldots, s(Z_n) \) is uniformly distributed on \( \{1, \ldots, n + 1\} \).

With this new knowledge we can revisit the prediction interval problem from above and give a proof for Theorem 2.

Proof. Note we assume that \( \{Z_i\}_{i=1}^{n+1} := \{(X_i, Y_i)\}_{i=1}^{n+1} \) are exchangeable. Consider the score function \( s(z) = -|\hat{\mu}(x) - y| \). As in Algorithm 2 let \( \{Z_i\}_{i \in I_1} \) be our training set, \( \{Z_i\}_{i \in I_2} \) be our hold-out set, and \( Z_{n+1} \) test point. Note that implicitly \( s \) and \( \hat{\mu} \) are functions of \( \{Z_i\}_{i \in I_1} \).

When we’re being careful in indicating this, we’ll denote them by \( s_{z_1, \ldots, z_{n_1}} \) and \( \hat{\mu}_{z_1, \ldots, z_{n_1}} \).

We claim that the \( \{s(Z_i)\}_{i \in I_2 \cup \{n+1\}} \) are exchangeable. To see why, consider permutations \( \pi : \{1, \ldots, n + 1\} \rightarrow \{1, \ldots, n + 1\} \) that fix the first \( n_1 \) points. It suffices to show that for all such permutations

\[
(s(Z_{n+1}), \ldots, s(Z_{n+1})) \overset{d}{=} (s(Z_{\pi(n+1)}), \ldots, s(Z_{\pi(n+1)}))
\]

From the exchangeability of the \( \{Z_i\}_{i=1}^{n+1} \) it follows that

\[
(s_{z_1, \ldots, z_{n_1}}(Z_{n+1}), \ldots, s_{z_1, \ldots, z_{n_1}}(Z_{n+1}))
\]

\[
= (-|Y_{n+1} - \hat{\mu}_{z_1, \ldots, z_{n_1}}(X_{n+1})|, \ldots, -|Y_{n+1} - \hat{\mu}_{z_1, \ldots, z_{n_1}}(X_{n+1})|)
\]

\[
\overset{d}{=} (-|Y_{\pi(n+1)} - \hat{\mu}_{z_{\pi(n+1)}, \ldots, z_{\pi(n+1)}}(X_{\pi(n+1)})|, \ldots, -|Y_{\pi(n+1)} - \hat{\mu}_{z_{\pi(n+1)}, \ldots, z_{\pi(n+1)}}(X_{\pi(n+1)})|)
\]

\[
= (-|Y_{\pi(n+1)} - \hat{\mu}_{z_1, \ldots, z_{n_1}}(X_{\pi(n+1)})|, \ldots, -|Y_{\pi(n+1)} - \hat{\mu}_{z_1, \ldots, z_{n_1}}(X_{\pi(n+1)})|)
\]

\[
= (s_{z_1, \ldots, z_{n_1}}(Z_{\pi(n+1)}), \ldots, s_{z_1, \ldots, z_{n_1}}(Z_{\pi(n+1)}))
\]
Then letting
\[ p(z) = \frac{1 + \#\{i \in I_2 : s(Z_i) \leq s(z)\}}{|I_2| + 1} \]
give our conformal p-values, it follows that
\[ Y_{n+1} \not\in C(X_{n+1}) \iff \]
s\( (Z_{n+1}) \) is strictly less than at least \(|I_2| + 1 \) many of the \( \{s(Z_i) : i \in I_2\} \)
\[ p(Z_{n+1}) \leq \frac{1 + |I_2| - [(|I_2| + 1)(1 - \alpha)]}{1 + |I_2|} = 1 - \frac{[(|I_2| + 1)(1 - \alpha)]}{|I_2| + 1} \]
Since \( p(Z_{n+1}) \leq 1 - \frac{[(|I_2| + 1)(1 - \alpha)]}{|I_2| + 1} \implies p(Z_{n+1}) \leq \alpha \). The same proof as for Theorem 3 tells us this happens with probability \( \leq \alpha \) so long as there are no ties in the \( s(Z_k) \) w.p. 1.

\[ \square \]

8.6 Conformal p-values are PRDS

Suppose we’re in a situation where we’ve observed training data \( Z_1, \ldots, Z_n \overset{i.i.d.}{\sim} P \) and we further receive some testing data \( Z_{n+1}, \ldots, Z_{n+m} \) which are drawn independently, but not all necessarily from \( P \). We want to detect which \( Z_{n+1}, \ldots, Z_{n+m} \) are not drawn from \( P \). Namely we want to test \( H_0 : Z_{n+i} \sim P, \ 1 \leq i \leq m \). We may find ourselves in a situation where it’s too restrictive to control the FWER, but we don’t want to waste time investigating too many false rejections. Ideally we’d like to control the false discovery rate. The above work tells us that the \( i \)th conformal p-value
\[ p_i = \frac{1 + \#\{1 \leq j \leq n : s(Z_j) \leq s(Z_{n+i})\}}{n + 1}, \quad i = 1, \ldots, m \]
is a valid p-value under \( H_0 \) so long as we don’t have ties in the \( s(Z_k) \) w.p. 1. Still, the \( \{p_i\}_{i \in \mathbb{N}_0} \) will be highly dependent on one another through the \( Z_1, \ldots, Z_n \). The following exciting result, however, tells us that we can still apply Benjamini-Hochberg to control the false discovery rate.

**Theorem 4.** (Bates, Cand´ es, Romano, Sesia (2021)) The above conformal p-values are PRDS!

### Appendix A

**Claim:** Let \( (X_1, \ldots, X_n) \) be PRDS on the set \( I_0 \) and \( D \) be an increasing set. Then for \( i \in I_0 \), \( P(X \in D|X_i \leq t) \) is non-decreasing in \( t \).

**Proof.** It suffices to show that for \( t' > t \) that \( \mathbb{P}(X \in D|X_i \leq t) \leq \mathbb{P}(X \in D|t < X_i \leq t') \) as if this is the case then
\[
\mathbb{P}(X \in D|X_i \leq t') = \mathbb{P}(X_i \leq t'|X_i \leq t')\mathbb{P}(X \in D|X_i \leq t) + P(X_i > t|X_i \leq t')\mathbb{P}(X \in D|t < X_i \leq t') \\
\geq \mathbb{P}(X_i \leq t'|X_i \leq t')\mathbb{P}(X \in D|X_i \leq t) + P(X_i > t|X_i \leq t')\mathbb{P}(X \in D|X_i \leq t) \\
= \mathbb{P}(X_i \in D|X_i \leq t)
\]
To see that this indeed is the case let \( \mu \) be the conditional law of \( X_i \) given \( X_i \leq t \) and \( \tilde{\mu} \) be the conditional law of \( X_i \) given \( t < X_i \leq t' \). Note that \( \mu \) is supported on \((-\infty, t]\) and \( \tilde{\mu} \) is supported on \((t, t']\). From the PRDS property of \( X \) we know that \( \mu \) a.s. \( \mathbb{P}(X \in D | X_i = s) \leq \mathbb{P}(X \in D | X_i = t) \) and \( \tilde{\mu} \) a.s. \( \mathbb{P}(X \in D | X_i = t) \leq \mathbb{P}(X \in D | X_i = s) \), so

\[
\mathbb{P}(X \in D | X_i \leq t) = \int \mathbb{P}(X \in D | X_i = s)d\mu \\
\leq \mathbb{P}(X \in D | X_i = t) \\
\leq \int \mathbb{P}(X \in D | X_i = s)d\tilde{\mu} \\
= \mathbb{P}(X \in D | t < X_i \leq t')
\]
Bibliography


