Flexible distribution-free conditional predictive bands using density estimators

Proofs for the paper “Flexible distribution-free conditional predictive bands using density estimators”

Definition 5.1. Whenever \( \hat{F} \) is a cdf, \( \hat{F}^{-1} \) refers to the generalized inverse of \( \hat{F} \).

Definition 5.2. \( U_{[a]} \) and \( U_{[a]} \) are the \( \lfloor n^{-1} (na) \rfloor \) empirical quantiles of \( U_1,\ldots,U_n \).

Related to Dist-split

Proof of Theorem 2.2. Let \( U_i = \hat{F}(Y_i | X_i) \). Since \( X_i, Y_i \) are i.i.d. continuous random variables and \( \hat{F} \) is continuous, obtain that \( U_i \) are i.i.d. continuous random variables.

\[
1 - \alpha \leq \mathbb{P} \left( U_{n+1} \in \left[ U_{[0.5a]}; U_{[1-0.5a]} \right] \right) \leq 1 - \alpha + (n+1)^{-1}.
\]

The conclusion follows from noticing that

\[
\mathbb{P} \left( U_{n+1} \in \left[ U_{[0.5a]}; U_{[1-0.5a]} \right] \right) = \mathbb{P} \left( Y_{n+1} \in \left[ \hat{F}^{-1} \left( U_{[0.5a]} \right); \hat{F}^{-1} \left( U_{[1-0.5a]} \right) \right] \right) = \mathbb{P} \left( Y_{n+1} \in C \left( X_{n+1} \right) \right)
\]

Lemma 5.3. Let \( I_1 = \{ i \leq n : \hat{F}(Y_i | X_i) - F(Y_i | X_i) \leq \eta_n^{1/3} \} \) and \( I_2 = \{ 1,\ldots,n \} - I_1 \). Under Assumption 2.3, \( |I_2| = o_p(n) \) and \( |I_1| = n + o_p(n) \).

Proof. Let \( A_n = \left\{ \mathbb{E} \left[ \sup_{y \in \mathbb{Y}} \left( \hat{F}(y | X) - F(y | X) \right)^2 | \hat{F} \right] \geq \eta_n^{1/3} \right\} \) and \( B_n = \{ |\hat{F}(Y | X) - F(Y | X)| \geq \eta_n^{1/3} \} \).

\[
\mathbb{P}(A_n) = \mathbb{E}[\mathbb{P}(B_n | \hat{F}) | A_n] + \mathbb{E}[\mathbb{P}(B_n | \hat{F}) | A_n]
\]

\[
\leq \mathbb{P}(A_n) + \mathbb{E} \left[ \frac{\|E[(\hat{F}(Y | X) - F(Y | X))^2] - \hat{F}(A_n)\|}{\eta_n^{2/3}} \right]
\]

\[
\leq \rho_n + \eta_n^{1/3} = o(1)
\]

Note that \( |I_2| \sim \text{Binomial}(n, \mathbb{P}(B_n)) \). Since \( \mathbb{P}(B_n) = o(1) \), conclude that \( |I_2| = o_p(n) \). That is, \( |I_1| = n + o_p(n) \).

Lemma 5.4. Under Assumption 2.3, if \( U_i = \hat{F}(Y_i | X_i) \), then for every \( a \in (0,1) \), \( U_{[a]} = a + o_p(1) = U_{[a]} \).

Proof. Let \( I_1 \) and \( I_2 \) be as such in Lemma 5.3. Also, let \( G_1, G_1 \) and \( G_0 \) be the empirical quantiles of, respectively, \( \{ U_i : i \in I_1 \} \), \( \{ F_i(Y_i | X_i) : i \in I_1 \} \), and \( \{ F_i(Y_i | X_i) : i \leq n \} \). By definition of \( I_1 \), for every \( a^* \in [0,1] \), \( G_1^{-1}(a^*) = G_1^{-1}(a^*) + o(1) \). Also, \( G_0^{-1}(a^*) = a^* + o_p(1) \). Therefore, since

\[
G_0^{-1} \left( \frac{\lfloor I_1 \lfloor a^* \rfloor}{n} \right) \leq G_1^{-1}(a^*) \leq G_0^{-1} \left( \frac{\lfloor I_1 \lfloor a^* + I_2 \rfloor}{n} \right),
\]

conclude that \( G_1^{-1}(a^*) = a^* + o_p(1) \). Finally, since

\[
\hat{G}^{-1} \left( \frac{na - |I_2|}{|I_1|} \right) \leq U_{[a]} \leq \hat{G}^{-1} \left( \frac{na}{|I_1|} \right),
\]

Conclude that \( U_{[a]} = a + o_p(1) = U_{[a]} \).

Lemma 5.5. Let \( U_i = \hat{F}(Y_i | X_i) \). Under Assumptions 2.3 and 2.4,

\[
\hat{F}_{[1-0.5a]}(U_{[1-0.5a]} | X_n) + o_p(1)
\]

Proof. In order to prove the first equality, it is enough to show that \( \hat{F}_{[1-0.5a]}(U_{[1-0.5a]} | X_n) + o_p(1) \) and that \( \hat{F}_{[0.5a]}(U_{[0.5a]} | X_n) + o_p(1) \) is the first part follows from Lemma 5.4 and the continuity of \( \hat{F}(y | x) \) (Assumption 2.4). For the second part, note, that if \( \sup_{y \in \mathbb{Y}} \hat{F}(y | x) - F(y | x) < \eta_n \), then, for every \( a^* \), \( \hat{F}(y | x) - F(y | x) < \eta_n \). Using this observation, the proof of the second part follows from Assumption 2.4, and observing that \( U_{[0.5a]} = 0.5a + o_p(1) \) (Lemma 5.4) and \( \mathbb{P}(\sup_{y \in \mathbb{Y}} \hat{F}(y | x) - F(y | x) \leq \eta_n) = o(1) \) (Assumption 2.3).

The proof for the \( 1 - .5 \alpha \) quantile is analogous to the one for the \( .5 \alpha \) quantile.

Proof of Theorem 2.5. Follows directly from Lemma 5.5.

Related to CD-split

Proof Theorem 3.3. Let \( \{i_1,\ldots,i_{n1}\} = \{ i : X_i \in A(X_{n1}) \} \), \( U_i = \hat{F}(Y_i | X_i) \), for \( I = 1,\ldots,n_I \), and \( U_{n1+I} = \hat{F}(Y_{n1+I} | X_{n1+I}) \). Since \( X_i, Y_i \), \( i = 1,\ldots,n_I \), \( X_{n1+I}, Y_{n1+I} \) are i.i.d. random variables, obtain that \( U_i \) are i.i.d. random variables conditional on the event \( X_{n1+I} = A(X_{n1}) \) and on \( i_1,\ldots,i_{n1} \). Therefore,

\[
1 - \alpha \leq \mathbb{P} \left( U_{n1+I} \geq U_{[a]} | X_{n1} = A(X_{n1}) \right)
\]

The conclusion follows from the fact that \( Y_{n1} = 1 \Rightarrow U_{n1+1} \geq U_{[1-a]} \) and because this holds for every sequence \( i_1,\ldots,i_{n1} \).

Proof of Theorem 3.8. Let \( U_i := f(Y_i | X_i) \), \( i = 1,\ldots,m, \)

\( U_{n1+I} := f(Y_{n1+I} | X_{n1+I}) \), and \( W := \{ x_1,\ldots,x_m, x_{n1} \} \). If \( g_{x_i} \) for every \( i = 1,\ldots,m \), then \( U_1,\ldots,U_m, U_{n1+1} \) are i.i.d. conditional on \( W \). Indeed, for every \( t \in \mathbb{R} \),

\[
\mathbb{P}(U_i \geq t | W) = \mathbb{P}(f(Y_i | X_i) \geq t) = \mathbb{P}(f(Y_{n1+1} | X_{n1+1}) \geq t) = \mathbb{P}(U_{n1+1} \geq t | X_{n1}),
\]

where the next-to-last equality follows from the definition of the profile of the density.

For every \( K \in \mathbb{R} \), let \( Q(K) := \{ i : f(Y_i | X_i) \geq K \} \). Because \( U_i \)'s are conditionally independent and identically distributed, then \( Q(K) | W \sim \text{Binomial}(m, \mathbb{P}(f(Y_i | X_i) \geq K)) \). It follows that \( Q(K) / m \to \mathbb{P}(f(Y_i | X_i) \geq K) \). In
particular, \( Q(t^*)/m \xrightarrow{m \to \infty} a.s. 1 - \alpha \). Now, by definition
\[ Q(T_m)/m \xrightarrow{m \to \infty} a.s. 1 - \alpha. \]
Conclude that \( T_m \xrightarrow{m \to \infty} T^* \).

\[ \square \]

Proof of Theorem 3.9. Item (i) was already shown as part of the proof of Theorem 3.8. To show (ii), assume that 
\[ t^*(x_a, \alpha) = t^*(x_b, \alpha) \] for every \( \alpha \in (0, 1) \). Now, notice that \( t^*(x_a, \alpha) \) is such that 
\[ g_{x_a}(t^*(x_a, \alpha)) = 1 - \alpha. \]
Conclude that 
\[ g_{x_a}(t^*(x_a, \alpha)) = g_{x_b}(t^*(x_b, \alpha)) \] for every \( \alpha \in (0, 1) \). Now, because \( \hat{f} \) is continuous, 
\[ \{ t^*(x_a, \alpha) : \alpha \in (0, 1) \} = \text{Im}(\hat{f}(x_a)). \] Thus, \( g_{x_a} = g_{x_b} \), and therefore \( x_a \sim x_b \). \[ \square \]