Immediate prediction under exchangeability & representation insensitivity
Gert de Cooman, Enrique Miranda & Erik Quaeghebeur

1 The setting

Sampling: A subject makes a fixed number \(N > 0\) of successive observations, represented by random variables \(X_1, \ldots, X_N\). For example, when drawing coloured balls without replacement from an urn, \(X_i\) designates the unknown colour of the \(i\)-th ball.

Immediate prediction: The subject in some way uses zero or more observations \(X_1, \ldots, X_N\) made so far (\(N \geq 1\)), to predict, or make inferences about, the value of the next observation \(X_{N+1}\).

Families of predictive lower previsions: The subject can determine, beforehand, a finite and non-empty set of possible values, or categories, for the random variables \(X_1, \ldots, X_N\). For each \(n\) and each sequence \(x = (x_1, \ldots, x_n)\) in \(X^n\), she can give a predictive lower prevision \(P_n^L(x)\) for \(X_n\), given the values \((X_1, \ldots, X_{n-1}, x_n)\) of the previous observations. It is defined on the set of all gambles \(f\) on \(X^n\).

\[ P_n^L(x) = \mathbb{E}^L(x) = \mathbb{E}^L(f) = \mathbb{E}^L(f|x) = \mathbb{E}^L(f|X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = x_n) \]

An \(X^n\) family \(\sigma^n\) of predictive lower previsions is the set formed for all possible observations:

\[ \sigma^n = \{ P_n^L(x) | x \in X^n \} \text{ and } \mathbb{E}^L(f|X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = x_n) \]

Precise predictive families are those that only contain precise linear previsions. Systems of predictive lower previsions: The inferences or predictions of a predictive \(n\)-family may depend on the actual choice of \(f\), made. So we let our subject consider predictive families for all conceivable choices of \(f\). We collect these families in a system \(\sigma^n\) of predictive lower previsions:

\[ \sigma^n = \{ P_n^L(x) | x \in X^n \} \text{ is a finite and non-empty set} \]

Immediate prediction under exchangeability: The system \(\sigma^n\) satisfies exchangeability if it is possible to 

\[ \sigma^n = \{ P^L_n(x) | x \in X^n \} \text{ is exchangeable} \]


2 Requirements & Assumptions

3 Some results

From sequences of observations to count vectors: In any regularly exchangeable predictive system, the predictive lower previsions \(P_f^n(x)\) only depend on the sequence of observations \(x\) through its count vector \(m(x)\), with

\[ m(x) = (k \in \{1, \ldots, n\} : x_k = f) \]

All predictive lower previsions for given sequences with the same count vector \(m(x)\) can therefore be written as \(P_{m(x)}^L(f)\).

\[ P_{m(x)}^L(f) = P_{m(x)}^L(f(x)) = P_{m(x)}^L(f_{m(x)}(x)) \]

Specificity (optional): Specificity is a requirement that works between predictive lower previsions for a different number of observations related by pooling.

An exchangeable predictive system is specific if it holds for all gambles \(f\) and all non-trivial events \(A\) in \(X^n\) containing a non-zero number \(m_A\) of observations, it holds that

\[ \{ P_{m(x)}^L(f) | x \in X^n \} \text{ is specific} \]

\[ P_{m(x)^*}^L(f) = P_{m(x)^*}^L(f(x)) = P_{m(x)^*}^L(f_{m(x)}(x)) \]

So, for regularly exchangeable predictive systems, count vectors are a sufficient statistic. From now on, we only consider (possibly non-exchangeable) predictive systems for which this is the case.

Selected references


Acknowledgements

Jean-Marie Bernard, Frank Coolen, Thomas Augustin, and the reviewers.

3 Some more requirements

Representation insensitivity: Representation insensitivity is a property that connects between predictive lower previsions for the same count vector.

\[ P_{m(x)}^L(f) = P_{m(x)}^L(f(x)) = P_{m(x)}^L(f_{m(x)}(x)) \]

5 More results

The lower probability function: With any predictive system we associate a map \(\lambda\) defined for all \(x\) and \(f\), by

\[ \lambda(x, f) = \mathbb{E}^L(f) = \mathbb{E}^L(f|X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = x_n) \]

For representation insensitive systems it fully characterizes all predictive lower probabilities \(\mathbb{P}(\cdot|x)\). \(\lambda\) is therefore called the lower probability function, \(\lambda\).

\[ \lambda(x, f) = \mathbb{E}^L(f) = \mathbb{E}^L(f|X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = x_n) \]

It allows us to draw intuitively appealing conclusions, which are valid in any coherent representation insensitive system:

(i) The lower probability of observing an event that has not/always been observed before is zero/one;
(ii) If \(x\) remains fixed, then both the lower and upper probability of observing \(A\) again do not decrease if \(m_A\) increases;
(iii) If systems that are also regularly exchangeable if \(m_A\) remains the same as \(x\) increases, then the lower probability for observing \(A\) again does not increase.

Immediate prediction under exchangeability & representation insensitivity: To start: all the \(P_f^n\) in a representation insensitive and exchangeable predictive system must be vacuous.

A subject that is too conservative to learn uses the regularly exchangeable vacuous predictive system \(\Psi\). All its predictive lower previsions are vacuous, so \(P_f^n = \mathbb{E}^L(f) = \mathbb{E}^L(f|X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = x_n) = \mathbb{E}^L(f|x)\).

A subject that believes that categories unobserved in the past remain so in the future, uses the (not regularly) exchangeable Haldane predictive system \(\Phi\). For \(f\), all its predictive lower previsions are linear and strongly tied to the observations:

\[ P_f^n = \mathbb{E}^L(f) = \mathbb{E}^L(f|X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = x_n) = \mathbb{E}^L(f|x) \]

Other systems can be formed as convex mixtures of the two extreme ones above. We define mixing predictive systems \(\sigma^n\) with \(\{0, 1\}\) bounded mixing sequence \(z(0) = z(1) = 0\). Note that implicitly \(\varepsilon = 0\). Representation insensitivity is retained after mixing: a sufficient condition for regular exchangeability is the reformulated ‘useful inequality’ \(\varepsilon(n, \cdot) \leq \frac{m_A(n, \cdot)}{\mathbb{E}^L(f)}\).

The lower probability function of a mixing system is given by \(\varepsilon(n, x, \cdot) = \frac{m_A(n, \cdot)}{\mathbb{E}^L(f)}\). As \(z(0) = \varepsilon(0, \cdot) = \varepsilon(1, \cdot) = 0\), a mixing system can be defined by the lower probability of observing any non-trivial event that has been observed once in \(n\) trials: It is the most conservative system with these lower probabilities.

The Imprecise Dirichlet-Multinomial Model: Any mixing system that is specific or for which the ‘useful equality’ holds, is uniquely characterized by some \(c > 0\) such that \(\varepsilon(n, \cdot) = \frac{c}{n}\) and

\[ \varepsilon(n, \cdot) = \frac{m_A(n, \cdot)}{\mathbb{E}^L(f)} \]

This regularly exchangeable representation insensitive predictive system is related to the imprecise Dirichlet-Multinomial model with hyper-parameter \(x\).