NONPARAMETRIC PREDICTIVE INFERENCE[1]

Frank P.A. Coolen[2]
Department of Mathematical Sciences, Durham University, UK
E-mail: frank.coolen@durham.ac.uk
Tahani Coolen-Maturi[3]
Department of Mathematical Sciences, Durham University, UK
E-mail: tahani.maturi@durham.ac.uk

1 Overview

Nonparametric Predictive Inference (NPI) is a statistical methodology based on Hill's assumption $A_{(n)}$ [30], which gives a direct conditional probability for a future observable random quantity, conditional on observed values of related random quantities [4, 7]. Suppose that $X_1, \ldots, X_n, X_{n+1}$ are exchangeable real-valued random quantities. Let the ordered observed values of $X_1, \ldots, X_n$ be denoted by $x_{(1)} < x_{(2)} < \ldots < x_{(n)}$, and, for ease of notation, let $x_{(0)} = -\infty$ and $x_{(n+1)} = \infty$ (or define these as other known bounds for the quantities X_j). For a future observation X_{n+1}, based on the n ordered observations $x_{(1)}, \ldots, x_{(n)}$, the assumption $A_{(n)}$ [30] is

$$P(X_{n+1} \in (x_{(j-1)}, x_{(j)})) = \frac{1}{n+1} \quad \text{for } j = 1, 2, \ldots, n+1$$

Note that, in case ties may occur, this can be dealt with in several ways, e.g. by breaking the ties or by using closed intervals $[x_{(j-1)}, x_{(j)}]$ instead of the open intervals in the above probabilities for X_{n+1}. $A_{(n)}$ does not assume anything else, and is a post-data assumption related to exchangeability. Hill [31] discusses $A_{(n)}$ in detail. Inferences based on $A_{(n)}$ are frequentist, predictive and nonparametric,

[1]To appear in: **International Encyclopedia of Statistical Science** (2nd Edition), M. Lovrig (Editor), Springer, Berlin, 2024.

[2]Frank Coolen is Professor of Statistics at the Department of Mathematical Sciences, Durham University (UK). He serves on the editorial boards of, amongst others, *Journal of Statistical Theory and Practice*, *Communications in Statistics*, *Journal of Risk and Reliability*, and *Quality and Reliability Engineering International*. His main research is on foundations and methods of Statistics and Reliability, in particular he has been developing Nonparametric Predictive Inference (NPI) since the mid-90s. Together with collaborators and students, Frank has published over 300 journal and conference papers, a majority of these on NPI or other statistical methods that use imprecise probabilities.

[3]Tahani Coolen-Maturi is Professor of Statistics at the Department of Mathematical Sciences, Durham University (UK). She is Associate Editor for *Journal of Statistical Theory and Practice* and *Mathematical Methods of Statistics*, and was co-editor of two special issues for the former journal. Her main focus is on developing statistical methodologies for real-world applications, such as reliability, finance, and medical applications. Over the past two decades, she has made significant contributions to the development of Nonparametric Predictive Inference (NPI). Tahani has published over 80 journal and conference papers with her collaborators and students, mainly on NPI and related methods to quantify uncertainty via imprecise probability.

and can be considered suitable if there is hardly any knowledge about the random quantity of interest, other than the n observations, or if one does not want to use such information, e.g. to study effects of additional assumptions underlying other statistical methods. The assumption $A_{(n)}$ is not sufficient to derive precise probabilities for many events of interest, but corresponding optimal bounds for probabilities for any event of interest involving X_{n+1} can be derived using De Finetti's Fundamental Theorem of Probability [26]. These bounds are lower and upper probabilities in the theory of imprecise probability [5, 35, 36], and as such they have strong consistency properties [4].

NPI is a framework of statistical theory and methods that use these $A_{(n)}$-based lower and upper probabilities, and also considers several variations of $A_{(n)}$ which are suitable for different inferences. For example, NPI has been presented for Bernoulli data [6], multinomial data [9] and right-censored data [16]. NPI enables inferences for $m \geq 1$ future observations, with their interdependence explicitly taken into account, and based on sequential assumptions $A_{(n)}, \ldots, A_{(n+m-1)}$ [3]. NPI provides a solution to some explicit goals formulated for objective (Bayesian) inference, which cannot be obtained when using precise probabilities [7]. NPI is exactly calibrated [32], which is a strong consistency property in frequentist statistics, and it never leads to results that are in conflict with inferences based on empirical probabilities.

NPI for Bernoulli random quantities [6] is based on a latent variable representation of Bernoulli data as real-valued outcomes of an experiment in which there is a completely unknown threshold value, such that outcomes to one side of the threshold are successes and to the other side failures. The use of $A_{(n)}$ together with lower and upper probabilities enables inference without a prior distribution on the unobservable threshold value, as is needed in Bayesian statistics where this threshold value is typically represented by a parameter. Suppose that there is a sequence of $n + m$ exchangeable Bernoulli trials, each with 'success' and 'failure' as possible outcomes, and data consisting of s successes in n trials. Let Y_1^n denote the random number of successes in trials 1 to n, then a sufficient representation of the data for NPI is $Y_1^n = s$, due to the assumed exchangeability of all trials. Let Y_{n+1}^{n+m} denote the random number of successes in trials $n + 1$ to $n + m$. Let $R_t = \{r_1, \ldots, r_t\}$, with $1 \leq t \leq m + 1$ and $0 \leq r_1 < r_2 < \ldots < r_t \leq m$, and, for ease of notation, define $\binom{s+r_0}{s} = 0$. Then the NPI upper probability for the event $Y_{n+1}^{n+m} \in R_t$, given data $Y_1^n = s$, for $s \in \{0, \ldots, n\}$, is [6]

$$\overline{P}(Y_{n+1}^{n+m} \in R_t | Y_1^n = s) = \binom{n+m}{n}^{-1} \sum_{j=1}^{t} \left[\binom{s+r_j}{s} - \binom{s+r_{j-1}}{s} \right] \binom{n-s+m-r_j}{n-s}$$

The corresponding NPI lower probability can be derived via the conjugacy property [6]

$$\underline{P}(Y_{n+1}^{n+m} \in R_t | Y_1^n = s) = 1 - \overline{P}(Y_{n+1}^{n+m} \in R_t^c | Y_1^n = s)$$

where $R_t^c = \{0, 1, \ldots, m\} \backslash R_t$. The counting method that leads to these NPI lower and upper probabilities is explained in detail by Aboalkhair et al [2].

For multinomial data, a latent variable representation based on the idea of a probability wheel has been presented, together with a corresponding adaptation of $A_{(n)}$ [9]. For data including right-censored observations, as often occur in lifetime data analysis, NPI is based on a variation of $A_{(n)}$ which effectively uses a similar exchangeability assumption for the future lifetime of a right-censored unit at its moment of censoring [16]. This method provides an attractive predictive alternative to the well-known Kaplan-Meier estimate for such data.

2 Applications

Many applications of NPI have been presented in the literature. These include solutions to problems in Statistics, Risk and Reliability, Operational Research and Finance. For example, NPI methods for multiple comparisons of groups of real-valued data are attractive for situations where such comparisons are naturally formulated in terms of comparison of future observations from the different groups [14]. NPI provides a frequentist solution to such problems which does not depend on counterfactuals, which play a role in hypothesis testing and are often criticized by opponents of frequentist statistics. An important advantage of the use of lower and upper probabilities is that one does not need to add assumptions to data which one feels are not justified. A nice example occurs in precedence testing, where experiments to compare different groups may be terminated early in order to save costs or time [25]. In such cases, the NPI lower and upper probabilities are the sharpest bounds corresponding to all possible orderings of the not-fully observed data. NPI provides an attractive framework for decision support in a wide range of problems where the focus is naturally on a future observation. For example, NPI methods for replacement decisions of technical units are powerful and fully adaptive to process data [23].

NPI has been applied for comparisons of multiple groups of proportions data [13], where the number m of future observations per group plays an interesting role in the inferences. Effectively, if m increases the inferences tend to become more imprecise, while imprecision tends to decrease if the number of observations in the data set increases. NPI for Bernoulli data has also been implemented for system reliability, with particularly attractive algorithms for optimal redundancy allocation [24, 33]. NPI for multinomial data enables inference if the number of outcome categories is not known, and explicitly distinguishes between defined and undefined categories for which no observations are available yet [8]. Typically, if outcome categories have not occurred yet, the NPI lower probability of the next observation falling in such a category is zero, but the corresponding NPI upper probability is positive and depends on whether or not the category is explicitly defined, on the total number of categories or whether this number is unknown, and on the number of categories observed so far. Such NPI upper probabilities can be used to support cautious decision making, which is often attractive in reliability and risk analysis.

NPI has been introduced for assessing the accuracy of a classifier's ability to discriminate between two groups of binary data [19], and for diagnostic tests with ordinal observations [27] and with real-valued observations [20]. NPI has been presented for three-group ROC analysis, with real-valued observations, to assess the ability of a diagnostic test to discriminate among three ordered classes or groups [21]. NPI has also been developed for three-group ROC analysis with ordinal outcomes [17] and to derive an optimal linear combination of biomarkers subject to limits of detection [18].

Classification methods based on NPI have been shown to perform well [1], while NPI applied to option pricing has shown interesting differences from the classical theory, due to the natural adaptation of NPI to data [28, 29]. Recently, NPI has been applied as a natural framework to study reproducibility of statistical inferences [12]. Several of these applications have required the use of NPI for multiple future observations, based on the frequentist statistics property of NPI that all orderings of m future real-valued observations among n data observations are equally likely. As this involves consideration of $\binom{n+m}{m}$ orderings, going through all of these is impossible except in some cases, where either corresponding NPI lower and upper probabilities for events of interest can be derived analytically, or when both n and m are small. For most cases of practical interest one needs an alternative method for implementation of NPI. Two solutions to this computational problem have been presented. First, sampling of the orderings of future data among the data observations can provide a solution [15], this will lead to estimates of the NPI lower and upper probabilities. Secondly, a bootstrap method based on NPI can be used [11]. In this method, based on the assumption $A_{(n)}$, first one interval from the partition of the real line, based on the observed data, is selected, followed by the selection of one future observation from that interval. Next, this future observation is added to the data and the process is repeated to draw a second future observation, and so on, until the required number of future observations has been drawn. Repeated application leads to multiple NPI-bootstrap samples, which can then be used for the predictive inference of interest. This is fully in line with the analytic NPI approach in that it keeps all orderings of future observations and data observations equally likely to occur, but it differs from the analytic NPI approach in the sense that inferences are no longer imprecise. Both these computational methods have been successfully applied in studies of reproducibility of statistical hypothesis tests [10, 11, 12, 15, 34].

3 Further Developments and Challenges

Due to the explicitly predictive nature of NPI, it is particularly suitable for problems which are naturally formulated in terms of future observations. Due to the manner in which statistical theory and methods have been developed historically, hypothesis testing or estimation are the established approaches for most statistical problems. But practical problems often have a predictive nature, hence there is scope to develop NPI for a wide range of applications. A main research challenge for NPI is its generalization to multi-dimensional random quantities. Coolen-Maturi et al

[22] present one solution by combining NPI for the marginals with an estimated copula to take the dependence structure into account. NPI for regression models is a further important challenge, research into this has recently been initiated. Further information on the development and application of NPI is available from *www.npi-statistics.com.*

References

[1] Abellan, J., Baker, R.M., Coolen, F.P.A., Crossman, R.J. and Masegosa, A.R. (2014). Classification with decision trees from a nonparametric predictive inference perspective. *Computational Statistics & Data Analysis*, **71**, 789–802.

[2] Aboalkhair, A.M., Coolen, F.P.A. and MacPhee, I.M. (2013). Nonparametric predictive reliability of series of voting systems *European Journal of Operational Research* **226**, 77–84.

[3] Arts, G.R.J., Coolen, F.P.A. and van der Laan, P. (2004). Nonparametric predictive inference in statistical process control. *Quality Technology & Quantitative Management* **1**, 201-216.

[4] Augustin, T. and Coolen, F.P.A. (2004). Nonparametric predictive inference and interval probability. *Journal of Statistical Planning and Inference* **124**, 251–272.

[5] Augustin, T. Coolen, F.P.A., de Cooman, G. and Troffaes, M.C.M. (Eds) (2014). *Introduction to Imprecise Probabilities.* Wiley, Chichester.

[6] Coolen, F.P.A. (1998). Low structure imprecise predictive inference for Bayes' problem. *Statistics & Probability Letters* **36**, 349–357.

[7] Coolen, F.P.A. (2006). On nonparametric predictive inference and objective Bayesianism. *Journal of Logic, Language and Information* **15**, 21–47.

[8] Coolen, F.P.A. (2007). Nonparametric prediction of unobserved failure modes. *Journal of Risk and Reliability* **221**, 207–216.

[9] Coolen, F.P.A. and Augustin, T. (2009). A nonparametric predictive alternative to the Imprecise Dirichlet Model: the case of a known number of categories. *International Journal of Approximate Reasoning* **50**, 217–230.

[10] Coolen, F.P.A. and Bin Himd, S. (2014). Nonparametric predictive inference for reproducibility of basic nonparametric tests. *Journal of Statistical Theory and Practice* **8**, 591-–618.

[11] Coolen, F.P.A. and Bin Himd, S. (2020). Nonparametric predictive inference bootstrap with application to reproducibility of the two-sample Kolmogorov–Smirnov test. *Journal of Statistical Theory and Practice* **14**, 26.

[12] Coolen, F.P.A. and Coolen-Maturi, T. (2024). Statistical reproducibility. In: *International Encyclopedia of Statistical Science*, M. Lovric (Ed.). Springer, Heidelberg.

[13] Coolen, F.P.A. and Coolen-Schrijner, P. (2007). Nonparametric predictive comparison of proportions. *Journal of Statistical Planning and Inference* **137**, 23–33.

[14] Coolen, F.P.A. and van der Laan, P. (2001). Imprecise predictive selection based on low structure assumptions. *Journal of Statistical Planning and Inference* **98**, 259–277.

[15] Coolen, F.P.A. and Marques, F.J. (2020). Nonparametric predictive inference for test reproducibility by sampling future data orderings. *Journal of Statistical Theory and Practice* **14**, 62.

[16] Coolen, F.P.A. and Yan, K.J. (2004). Nonparametric predictive inference with right-censored data. *Journal of Statistical Planning and Inference* **126**, 25–54.

[17] Coolen-Maturi, T. (2017). Three-group ROC predictive analysis for ordinal outcomes. *Communications in Statistics – Theory and Methods* **46**, 9476–9493.

[18] Coolen-Maturi, T. (2017). Predictive inference for best linear combination of biomarkers subject to limits of detection. *Statistics in Medicine* **36**, 2844–2874.

[19] Coolen-Maturi, T., Coolen-Schrijner, P. and Coolen, F.P.A. (2012). Nonparametric predictive inference for binary diagnostic tests. *Journal of Statistical Theory and Practice* **6**, 665–680.

[20] Coolen-Maturi, T., Coolen-Schrijner, P. and Coolen, F.P.A. (2012). Nonparametric predictive inference for diagnostic accuracy. *Journal of Statistical Planning and Inference* **142**, 1141–1150.

[21] Coolen-Maturi, T., Elkhafifi, F.F. and Coolen, F.P.A. (2014). Three-group ROC analysis: A nonparametric predictive approach. *Computational Statistics & Data Analysis* **78**, 69–81.

[22] Coolen-Maturi, T., Coolen, F.P.A. and Muhammad, N. (2016). Predictive inference for bivariate data: Combining nonparametric predictive inference for marginals with an estimated copula. *Journal of Statistical Theory and Practice* **10**, 515–538.

[23] Coolen-Schrijner, P. and Coolen, F.P.A. (2004). Adaptive age replacement based on nonparametric predictive inference. *Journal of the Operational Research Society* **55**, 1281–1297.

[24] Coolen-Schrijner, P., Coolen, F.P.A. and MacPhee, I.M. (2008). Nonparametric predictive inference for systems reliability with redundancy allocation. *Journal of Risk and Reliability* **222**, 463–476.

[25] Coolen-Schrijner P., Maturi, T.A. and Coolen, F.P.A. (2009). Nonparametric predictive precedence testing for two groups. *Journal of Statistical Theory and Practice* **3**, 273–287.

[26] De Finetti, B. (1974). *Theory of Probability* (2 volumes). Wiley, Chichester.

[27] Elkhafifi, F.F. and Coolen, F.P.A. (2012). Nonparametric predictive inference for accuracy of ordinal diagnostic tests. *Journal of Statistical Theory and Practice* **6**, 681–697.

[28] He, T., Coolen, F.P.A. and Coolen-Maturi, T. (2019). Nonparametric predictive inference for European option pricing based on the binomial tree model *Journal of the Operational Research Society* **70**, 1692–1708.

[29] He, T., Coolen, F.P.A. and Coolen-Maturi, T. (2021). Nonparametric predictive inference for American option pricing based on the binomial tree model. *Communications in Statistics - Theory and Methods* **50**, 4657–4684.

[30] Hill, B.M. (1968). Posterior distribution of percentiles: Bayes' theorem for sampling from a population. *Journal of the American Statistical Association* **63**, 677–691.

[31] Hill, B.M. (1988). De Finetti's theorem, induction, and $A_{(n)}$ or Bayesian nonparametric predictive inference (with discussion). In: *Bayesian Statistics* (Eds. J.M. Bernardo, et al.), 3, pp. 211–241. Oxford University Press.

[32] Lawless, J.F. and Fredette, M. (2005). Frequentist prediction intervals and predictive distributions. *Biometrika* **92**, 529–542.

[33] MacPhee, I.M., Coolen, F.P.A. and Aboalkhair, A.M. (2009). Nonparametric predictive system reliability with redundancy allocation following component testing. *Journal of Risk and Reliability* **223**, 181–188.

[34] Simkus, A., Coolen, F.P.A., Coolen-Maturi, T., Karp, N.A. and Bendtsen, C. (2022). Statistical reproducibility for pairwise t-tests in pharmaceutical research. *Statistical Methods in Medical Research* **31**, 673–688.

[35] Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities.* Chapman & Hall, London.

[36] Weichselberger, K. (2001). *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I. Intervalwahrscheinlichkeit als umfassendes Konzept* (in German). Physika, Heidelberg.