Nonparametric Predictive Inference*

Frank P.A. Coolen
Durham University, UK
frank.coolen@durham.ac.uk

1 Overview

Nonparametric predictive inference (NPI) is a statistical method based on Hill's assumption $A_{(n)}$ [13], which gives a direct conditional probability for a future observable random quantity, conditional on observed values of related random quantities [1, 3]. Suppose that $X_1, \ldots, X_n, X_{n+1}$ are continuous and exchangeable random quantities. Let the ordered observed values of $X_1, \ldots, X_n$ be denoted by $x_{(1)} < x_{(2)} < \ldots < x_{(n)} < \infty$, and let $x_{(0)} = -\infty$ and $x_{(n+1)} = \infty$ for ease of notation. For a future observation X_{n+1}, based on n observations, $A_{(n)}$ [13] is

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

$A_{(n)}$ does not assume anything else, and is a post-data assumption related to exchangeability. Hill [14] discusses $A_{(n)}$ in detail. Inferences based on $A_{(n)}$ are predictive and nonparametric, 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. $A_{(n)}$ is not sufficient to derive precise probabilities for many events of interest, but it provides optimal bounds for probabilities for all events of interest involving X_{n+1}. These bounds are lower and upper probabilities in the theories of imprecise probability [17] and interval probability [18], and as such they have strong consistency properties [1]. 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, multinomial data and lifetime data with right-censored observations. NPI enables inferences for $m \geq 1$ future observations, with their interdependence explicitly taken into account, and based on sequential assumptions $A_{(n)}, \ldots, A_{(n+m-1)}$. NPI provides a solution to some explicit goals formulated for objective (Bayesian) inference, which cannot be obtained when using precise probabilities [3]. NPI is also exactly calibrated [15], which is a strong consistency property, and it never leads to results that are in conflict with inferences based on empirical probabilities.

NPI for Bernoulli random quantities [2] 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 enable 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

*Contribution to *International Encyclopedia of Statistical Sciences (Springer)* 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

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

The corresponding NPI lower probability is derived via the conjugacy property

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

where $R_t^c = \{0, 1, \ldots, m\} \backslash R_t$.

For multinomial data, a latent variable representation via segments of a probability wheel has been presented, together with a corresponding adaptation of $A_{(n)}$ [5]. 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 [9]. This method provides an attractive predictive alternative to the well-known Kaplan-Meier estimate (KME) for such data.

2 Applications

Many applications of NPI have been presented in the literature. These include solutions to problems in Statistics, Reliability and Operational Research. 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 [8]. 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 [12]. 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 [10].

NPI has been applied for comparisons of multiple groups of proportions data [6], 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 [11, 16]. 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 [4]. 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 decisions, which are often deemed attractive in reliability and risk analysis.

3 Challenges

Development of NPI is gathering momentum, inferential problems for which NPI solutions have recently been presented or are being developed include aspects of medical diagnosis with the use of ROC curves, robust classification, inference on competing risks, quality control and acceptance sampling. Main research challenges for NPI include its generalization for multidimensional data, which is similarly challenging for NPI as for general nonparametric methods due to the lack of a unique natural ordering of the data. NPI theory and methods that enable information from covariates to be taken into account also provide interesting and challenging research opportunities. A research monograph introducing NPI theory, methods and applications is currently in development [7], further information is available from www.npi-statistics.com.

References

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

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

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

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

[5] 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.

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

[7] Coolen, F.P.A. and Coolen-Schrijner, P. *Nonparametric Predictive Inference*. Wiley, Chichester, to appear.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

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

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

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

[16] 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.

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

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