Nonparametric Predictive Inference for Bernoulli quantities: two examples

Coolen, F.P.A.
Durham University, Department of Mathematical Sciences
Science Laboratories, South Road
Durham, DH1 3LE, United Kingdom
E-mail: Frank.Coolen@durham.ac.uk

Coolen-Schrijner, P.
Durham University, Department of Mathematical Sciences
Science Laboratories, South Road
Durham, DH1 3LE, United Kingdom
E-mail: Pauline.Schrijner@durham.ac.uk

1. Nonparametric Predictive Inference

Coolen (1998) presented Nonparametric Predictive Inference (NPI) for Bernoulli random quantities, based on a representation of Bernoulli data as outcomes of an experiment similar to that used by Bayes (1763), with Hill's assumption $A_{(n)}$ (Hill 1968, 1988) used to derive direct predictive probabilities for future observations based on available data. The lower and upper probabilities presented by Coolen (1998) have strong internal consistency properties in the theory of interval probability (Augustin and Coolen 2004, Weichselberger 2001). Due to the use of $A_{(n)}$ in deriving these lower and upper probabilities, they fit in a frequentist framework of statistics but can also be interpreted from Bayesian perspective (Hill 1988, Coolen 2006). NPI is also 'perfectly calibrated' in the sense of Lawless and Fredette (2005). In this paper, we briefly give the main results from Coolen (1998), and we illustrate their use in two recently developed applications.

Suppose that we have 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 our inferences 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, let us define $\binom{s+r_0}{s} = 0$. Then the NPI-based upper probability (Coolen 1998) for the event $Y_{n+1}^{n+m} \in R_t$, given data $Y_1^n = s$, for $s \in \{0, \ldots, n\}$, is

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

The corresponding lower probability (Coolen 1998) 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 $R_t^c = \{0, 1, \ldots, m\} \backslash R_t$. This conjugacy property agrees with the fact that these lower and upper probabilities are F-probabilities in Weichselberger's theory of interval probability (Weichselberger 2000, 2001). These lower and upper probabilities also have attractive properties beyond internal consistency, as the interval created by the lower and upper probability for an event A always contains the precise empirical probability for A as based on the observed data, and the lower (upper) probability increases (decreases) as function of n, for constant s/n.

2. Comparison of proportions

To compare two or more independent groups of data, each consisting of a number of successes and failures, for example to compare effectiveness of different treatments, the general upper probability (1) and the conjugacy property can be used to derive upper and lower probabilities on events which directly compare $m \geq 1$ future observations from each of the different treatments, enabling both pairwise and multiple comparisons. For general results on NPI pairwise and multiple comparisons, and more details on the example presented below, we refer to Coolen and Coolen-Schrijner (2007a). We briefly illustrate NPI multiple comparisons in Example 1.

Example 1.
Spiegelhalter, *et al.* (2002) present an analysis of several data sets on mortality in heart operations on children. We use one of those data sets to illustrate NPI multiple comparisons, without comparing it to other sources of information or discussing the quality of the data. This data set consists of the number n_i, for $i = 1, \ldots, 12$, of heart operations on children under 1 year old at 12 medical centres, during the period 1991 until March 1995, and the corresponding number s_i of mortalities. Table 1 also gives s_i/n_i, together with the NPI multiple comparisons results for $m = 10$ and $m = 50$, where '$i > \max j$' denotes the event that the future number of deaths in Centre i is greater than that in all other centres, from m operations at each centre. An aspect of interest in the original study was whether the proportion of mortalities at Centre 1 exceeds those at the other centres. Centre 1 has the highest observed proportion of deaths, and we see that the predictive lower and upper probabilities clearly indicate that, on the basis of our NPI approach, this centre is the most likely one to lead to the highest number of deaths in m future heart operations. For smaller m there is a higher chance of two or more centres leading to the same maximum number of such deaths, hence the differences between the lower and upper probabilities for events '$i > \max j$' and '$i \geq \max j$' tend to decrease for larger m. For $m = 50$, the imprecision is often greater than for $m = 10$, although this is not a general effect, mostly due to the fact that the upper probability for many of these centres gets closer to zero for larger m.

Table 1: Multiple comparisons 12 centres

			$m = 10$		$m = 50$	
Centre	(n_i, s_i)	s_i/n_i	$[\underline{P}, \overline{P}](i > \max j)$	$[\underline{P}, \overline{P}](i \geq \max j)$	$[\underline{P}, \overline{P}](i > \max j)$	$[\underline{P}, \overline{P}](i \geq \max j)$
1	(181,43)	0.2376	[0.177, 0.197]	[0.369, 0.397]	[0.426, 0.482]	[0.526, 0.583]
2	(200,27)	0.1350	[0.033, 0.039]	[0.112, 0.128]	[0.022, 0.032]	[0.041, 0.057]
3	(157,26)	0.1656	[0.061, 0.072]	[0.173, 0.196]	[0.073, 0.098]	[0.114, 0.148]
4	(142,15)	0.1056	[0.017, 0.022]	[0.067, 0.082]	[0.007, 0.011]	[0.014, 0.022]
5	(217,36)	0.1659	[0.060, 0.070]	[0.173, 0.193]	[0.069, 0.089]	[0.110, 0.139]
6	(417,49)	0.1175	[0.021, 0.024]	[0.082, 0.092]	[0.008, 0.011]	[0.017, 0.022]
7	(253,27)	0.1067	[0.016, 0.020]	[0.067, 0.078]	[0.005, 0.008]	[0.011, 0.016]
8	(369,57)	0.1545	[0.048, 0.054]	[0.148, 0.163]	[0.042, 0.054]	[0.073, 0.091]
9	(214,28)	0.1308	[0.030, 0.036]	[0.104, 0.120]	[0.018, 0.026]	[0.034, 0.048]
10	(184,31)	0.1685	[0.064, 0.074]	[0.179, 0.201]	[0.077, 0.101]	[0.121, 0.153]
11	(740,67)	0.0905	[0.009, 0.011]	[0.046, 0.052]	[0.001, 0.002]	[0.003, 0.004]
12	(268,32)	0.1194	[0.022, 0.027]	[0.085, 0.098]	[0.010, 0.014]	[0.020, 0.028]

3. System reliability

We consider inference for reliability of a k-out-of-m system, consisting of m exchangeable components such that the system functions if at least k of these components function. We assume that n such components, also exchangeable with the m in the system, have been tested, of which s functioned satisfactorily. The general upper probability (1) and the conjugacy property can be used to derive lower and upper probabilities for the event $Y_{n+1}^{n+m} \geq k$, as this event corresponds to successful functioning of a k-out-of-m system. We also denote these lower and upper probabilities for the event that the k-out-of-m system functions successfully, by $\underline{P}(m : k|n, s)$ and $\overline{P}(m : k|n, s)$, respectively. For detailed presentation of this NPI approach for reliability of such systems, see Coolen and Coolen-Schrijner (2007b). For the NPI lower and upper probabilities for successful functioning of a k-out-of-m system, the following relation holds

$$(2) \qquad \overline{P}(m : k|n, s) = \underline{P}(m : k|n, s + 1)$$

The result (2) can obviously be used to reduce computational effort, if upper and lower probabilities are required for all possible values of s. We would also like to emphasize the elegance of this equality, as it implies that the intervals created by corresponding lower and upper probabilities of successful system functioning, for $s = 0, 1, \ldots, n$, form a partition of the interval $[0, 1]$. We illustrate this approach in Examples 2 and 3.

Example 2.
Consider a series system with 10 exchangeable components (so $k = m = 10$), and the only information available is the result of a test of 2 components, also exchangeable with the 10 to be used in the system. For the three possible values of the number of successes in the tests, $s = 0, 1, 2$, the NPI lower and upper probabilities for successful functioning of the system are $[\underline{P}, \overline{P}](10 : 10|2, 0) = [0, \frac{1}{66}]$, $[\underline{P}, \overline{P}](10 : 10|2, 1) = [\frac{1}{66}, \frac{1}{6}]$ and $[\underline{P}, \overline{P}](10 : 10|2, 2) = [\frac{1}{6}, 1]$. These values illustrate property (2), and the value 0 (1) of the lower (upper) probability for the case $s = 0$ ($s = 2$) reflects that in this case there is no strong evidence that the components can actually function (fail).

Example 3.
If one aims at testing to demonstrate high reliability, one may only allow the release of a system for practical use if testing of components revealed zero failures. In risk assessment, it is attractive to consider reliability requirements in terms of the lower probability of successful functioning of the system, given the test results on the components. Table 2 presents the minimum number of zero-failure tests required to achieve a chosen value for the lower probability of successful system functioning, for $k = 8$ and m varying from 8 to 12, so one has the option of building redundancy into the system. The requirement considered is $\underline{P}(m : 8|n, n) \geq p$ for different values of p. The main conclusion from Table 2 is that, in order to demonstrate high reliability via zero-failure testing, one requires quite a large number of successful tests, yet this number can be substantially reduced by building in redundancy.

Table 2: Reliability demonstration: zero-failure tests vs redundancy

	$m = 8$	9	10	11	12
$p = 0.75$	24	9	6	4	4
0.80	32	11	7	5	4
0.85	46	14	8	6	5
0.90	72	19	11	8	6
0.95	153	30	16	11	9
0.99	792	77	33	21	15

4. Concluding remarks

In this paper, only NPI for Bernoulli quantities has been illustrated. NPI has been developed and applied for more situations, including lifetime data with right-censored observations and multinomial data. Such methods typically required slight variations to Hill's assumption $A_{(n)}$, together with a different assumed data representation for the latter case. In addition to several statistical inferential procedures, NPI methods have also been presented for applications in Reliability and in Operations Research, for example providing decision support methods for replacement of technical units, which are fully adaptive to process data (Coolen-Schrijner and Coolen 2004). Coolen (2006) provides a short overview with further references, and also presents NPI as an attractive alternative to so-called 'objective Bayesian methods'.

NPI needs to be developed further in order to enhance its practical use. In particular, methods for dealing with covariates must be developed, together with NPI for multivariate data. Generally speaking, such methods will require careful development of appropriate exchangeability assumptions, to then be used in a post-data manner in the same spirit as Hill's $A_{(n)}$, which underlies NPI.

REFERENCES

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

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions, Royal Society of London* **53**, 370-418; **54**, 296-325.

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

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

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

Coolen, F.P.A. and Coolen-Schrijner, P. (2007b). Nonparametric predictive inference for voting systems. In: *Proceedings MMR 2007*, Glasgow, July 2007, to appear.

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.

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

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: Oxford University Press.

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

Spiegelhalter, D.J., Aylin, P., Best, N.G., Evans, S.J.W. and Murray, G.D. (2002). Commissioned analysis of surgical performance using routine data: lessons from the Bristol inquiry. *Journal of the Royal Statistical Society, Series A* **165**, 191-231.

Weichselberger, K. (2000). The theory of interval-probability as a unifying concept for uncertainty. *International Journal of Approximate Reasoning* **24**, 149-170.

Weichselberger, K. (2001). *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I. Intervallwahrscheinlichkeit als umfassendes Konzept.* Heidelberg: Physika.