Nonparametric Predictive Inference (An Introduction)

Frank Coolen and Tahani Coolen-Maturi

SIPTA Summer School 2010
Durham University
3 September 2010

Hill's assumption $A_{(n)}$ (Hill, 1968)

- $X_1, \ldots, X_n, X_{n+1}$ are real-valued and exchangeable random quantities
- $x_1 < x_2 < \ldots < x_n$ are the ordered observed values of $X_1, \ldots, X_n$ (and let $x_0 = -\infty$ and $x_{n+1} = \infty$)
- For X_{n+1}, $A_{(n)}$ is given by

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

Nonparametric predictive inference (NPI)

- NPI is based on Hill's assumption $A_{(n)}$
- Let $\mathcal{B}$ be the Borel σ-field over $\mathbb{R}$. For any element $B \in \mathcal{B}$, lower probability $\underline{P}(.)$ and upper probability $\overline{P}(.)$ for the event $X_{n+1} \in B$, based on the intervals $I_j = (x_{j-1}, x_j)$ $(j = 1, 2, \ldots, n+1)$ created by n real-valued non-tied observations, and the assumption $A_{(n)}$, are

$$\underline{P}(X_{n+1} \in B) = \frac{1}{n+1} |\{j : I_j \subseteq B\}|$$

$$\overline{P}(X_{n+1} \in B) = \frac{1}{n+1} |\{j : I_j \cap B \neq \emptyset\}|$$

Illustration example

$n = 4$

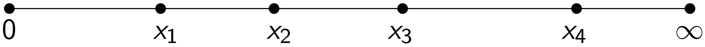
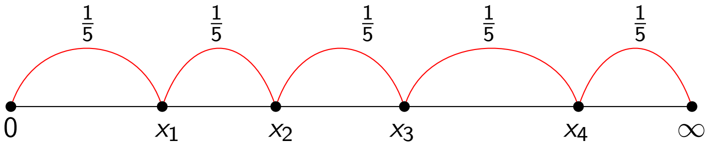


Illustration example

$n = 4$



$P(X_5 \in (0, x_1)) = \frac{1}{5}$ $\qquad$ $P(X_5 \in (x_4, \infty)) = \frac{1}{5}$

$P(X_5 \in (x_i, x_{i+1})) = \frac{1}{5}$, $i = 1, 2, 3$

Illustration example

$n = 4$

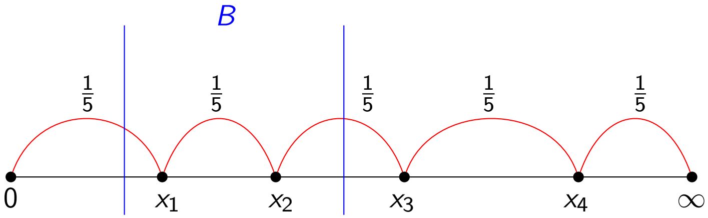
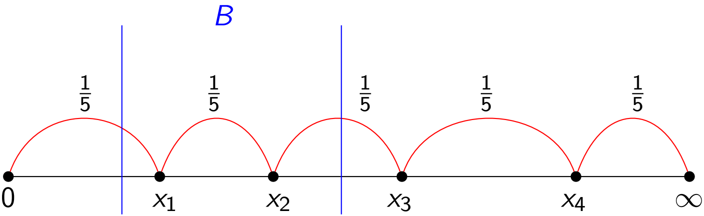


Illustration example

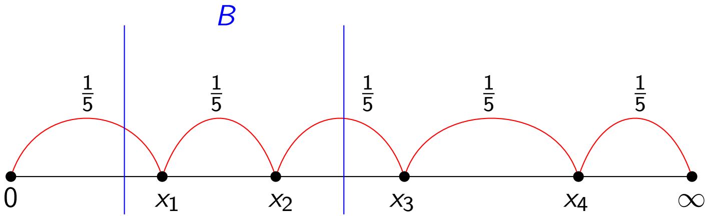
$n = 4$



$\underline{P}[X_5 \in B] = \frac{1}{5}$ $\overline{P}[X_5 \in B] = \frac{3}{5}$

Illustration example

$n = 4$



$\underline{P}[X_5 \in B] = \frac{1}{5}$ $\overline{P}[X_5 \in B] = \frac{3}{5}$

$Imprecision = \overline{P} - \underline{P} = \frac{2}{5} = 0.4$

NPI for m future observations

- We are interested in $m \geq 1$ future observations, X_{n+i} for $i = 1, \ldots, m$.
- We link the data and future observations via Hill's assumption $A_{(n)}$, actually via $A_{(n+m-1)}$ (which implies $A_{(n+k)}$ for all $k = 0, 1, \ldots, m-2$).
- Let $S_j = \#\{X_{n+i} \in I_j,\ i = 1, \ldots, m\}$, then inferences about these m future observations, assuming $A_{(n+m-1)}$, can be based on the following probabilities, for any $(s_1, \ldots, s_{n+1})$ with non-negative integers s_j with $\sum_{j=1}^{n+1} s_j = m$

$$P(\bigcap_{j=1}^{n+1} \{S_j = s_j\}) = \binom{n+m}{n}^{-1}$$

Comparing two independent groups, X and Y

We have two independent groups X and Y:

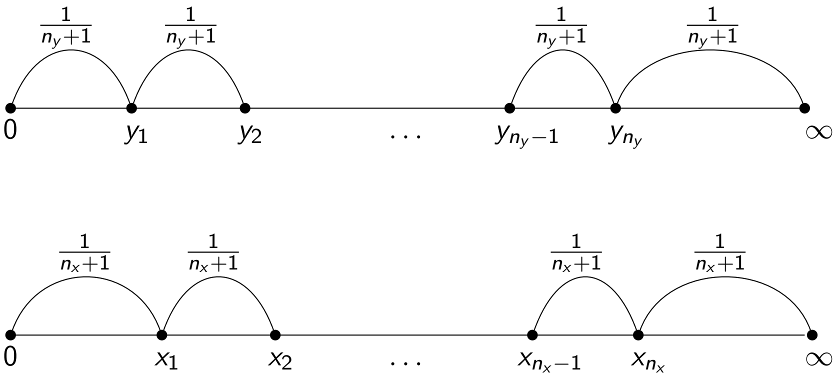
$$x_1 < x_2 < \ldots < x_{n_x} \qquad \text{and} \qquad y_1 < y_2 < \ldots < y_{n_y}$$

The classical methods test $H_0 : F_X = F_Y$.

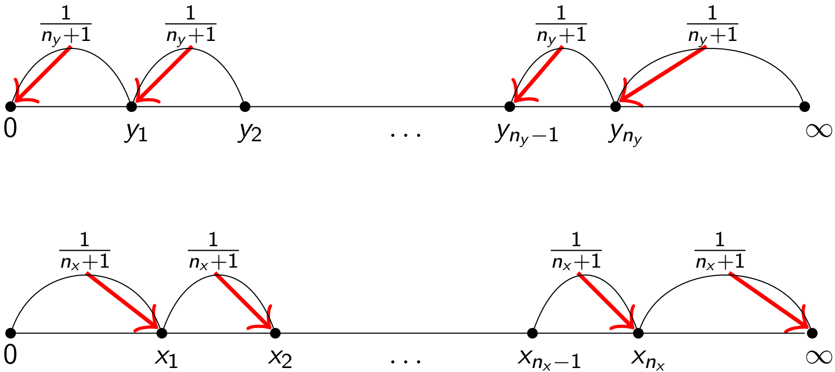
For complete data, Coolen (1996) introduced NPI to compare two independent groups depending on $A_{(n)}$. This is given via the lower and upper probabilities

$$\underline{P}(X_{n_x+1} < Y_{n_y+1}) \qquad \overline{P}(X_{n_x+1} < Y_{n_y+1})$$

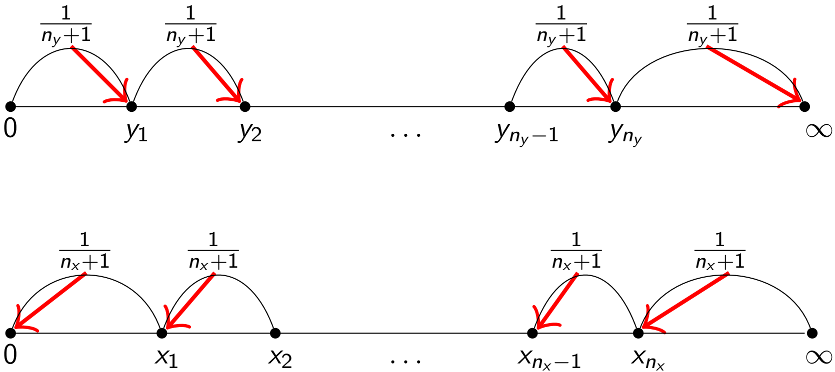
Illustration



Lower Probability, $\underline{P}(X_{n_x+1} < Y_{n_y+1})$



Upper Probability, $\overline{P}(X_{n_x+1} < Y_{n_y+1})$



Example

We use data on birthweights for 12 male and 12 female babies as presented by Dobson (1983).

Male (X)	2625	2628	2795	2847	2925	2968
	2975	3163	3176	3292	3421	3473
Female (Y)	2412	2539	2729	2754	2817	2875
	2935	2991	3126	3210	3231	3317

$$\underline{P}(X_{13} > Y_{13}) = \frac{86}{169} = 0.509$$

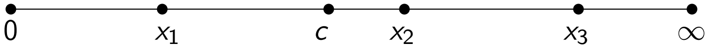
$$\overline{P}(X_{13} > Y_{13}) = \frac{111}{169} = 0.657.$$

Right-censored data

- Very common in lifetime data, e.g. in survival analysis and reliability studies
- Actual event of interest for a unit (e.g. death or failure) is not observed, only that this event has not happened by an observed time
- We assume that the censoring mechanism is independent of the residual lifetime at the time of censoring, hence that every unit still in the experiment had the same probability of being the unit censored ('non-informative censoring')

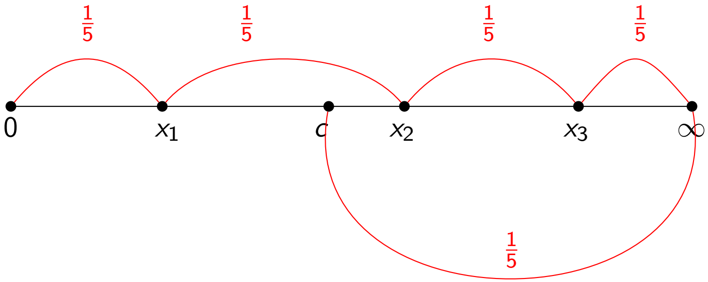
rc-$A_{(n)}$ assumption (Coolen & Yan, 2004)

$n = 4$



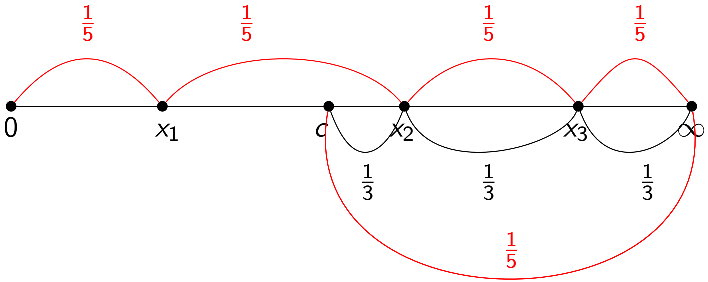
rc-$A_{(n)}$ assumption (Coolen & Yan, 2004)

$n = 4$



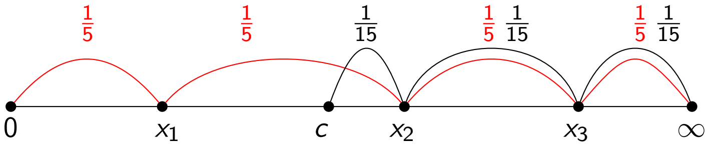
rc-$A_{(n)}$ assumption (Coolen & Yan, 2004)

$n = 4$



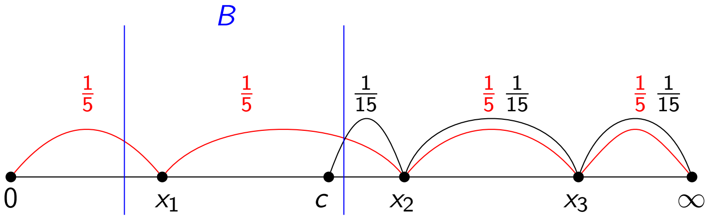
rc-$A_{(n)}$ assumption (Coolen & Yan, 2004)

$n = 4$



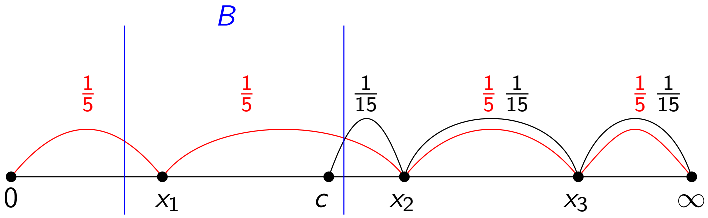
rc-$A_{(n)}$ assumption (Coolen & Yan, 2004)

$n = 4$



rc-$A_{(n)}$ assumption (Coolen & Yan, 2004)

$n = 4$



$\underline{P}[X_5 \in B] = 0$ $\overline{P}[X_5 \in B] = \frac{7}{15}$

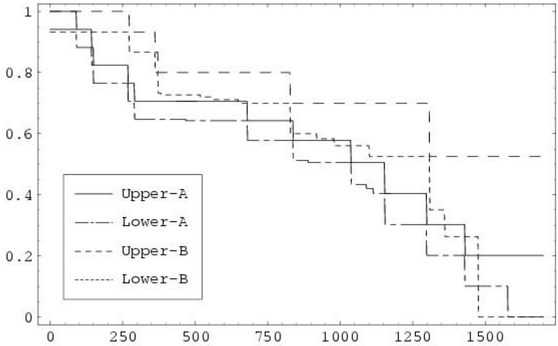
$Imprecision = \overline{P} - \underline{P} = \frac{7}{15} = 0.4667$

Example

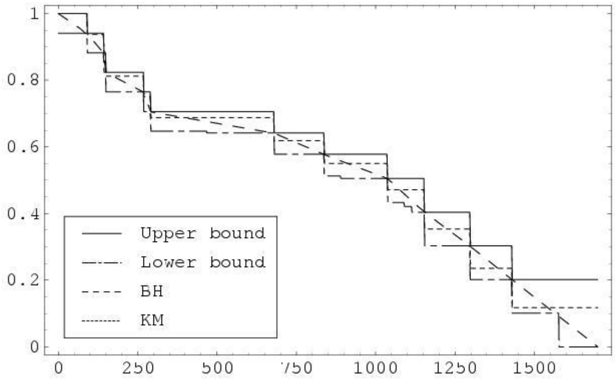
Table: Cervical cancer survival data ($> t$ indicates right-censoring at t)

Control - A		New therapy - B	
90	>890	272	>919
142	1037	362	>978
150	>1090	373	>1100
269	>1113	>383	1307
291	1153	>519	>1360
>468	1297	>563	>1476
680	1429	>650	
837	>1577	827	

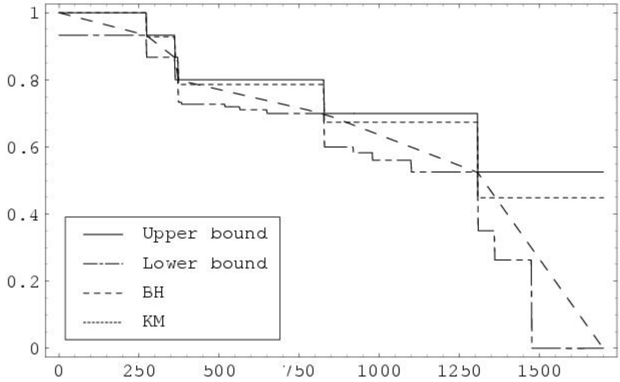
rc-$A_{(n)}$ assumption (Coolen & Yan, 2004)



rc-$A_{(n)}$ assumption (Coolen & Yan, 2004)



rc-$A_{(n)}$ assumption (Coolen & Yan, 2004)



NPI for Bernoulli data

- $n + m$ exchangeable Bernoulli trials (each 'success' or 'failure')
- Y_j^l: the number of successes in trials j to l
- Data on first n trials: $Y_1^n = s$
- Interest in Y_{n+1}^{n+m}
- FC: 'Low structure imprecise predictive inference for Bayes' problem' *Stat. & Prob. Letters 36, 1998, 349-357*.

- $R_t = \{r_1, \ldots, r_t\}$, with $1 \leq t \leq m+1$ and $0 \leq r_1 < r_2 < \ldots < r_t \leq m$, (define $\binom{s+r_0}{s} = 0$).

- NPI upper probability:

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

- NPI lower probability:

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

NPI for Bernoulli data

Centre	(n_i, s_i)	s_i/n_i	order	Centre	(n_i, s_i)	s_i/n_i	order
1	(181,43)	0.2376	1	7	(253,27)	0.1067	10
2	(200,27)	0.1350	6	8	(369,57)	0.1545	5
3	(157,26)	0.1656	4	9	(214,28)	0.1308	7
4	(142,15)	0.1056	11	10	(184,31)	0.1685	2
5	(217,36)	0.1659	3	11	(740,67)	0.0905	12
6	(417,49)	0.1175	9	12	(268,32)	0.1194	8

Table 1. Heart operations mortality data.

m :	1	3	5	10	50	250
$\overline{P}(1 > 11)$	0.220	0.457	0.578	0.730	0.964	1.000
$\underline{P}(1 > 11)$	0.215	0.447	0.566	0.716	0.957	0.999
$\Delta(1 > 11)$	0.005	0.010	0.012	0.014	0.007	0.001
$\overline{P}(1 \geq 11)$	0.931	0.882	0.875	0.894	0.981	1.000
$\underline{P}(1 \geq 11)$	0.930	0.878	0.870	0.887	0.976	1.000
$\Delta(1 \geq 11)$	0.001	0.004	0.005	0.007	0.005	0.000
$\overline{P}(3 > 5)$	0.143	0.277	0.335	0.394	0.478	0.527
$\underline{P}(3 > 5)$	0.137	0.264	0.318	0.369	0.426	0.441
$\Delta(3 > 5)$	0.006	0.013	0.017	0.025	0.052	0.086
$\overline{P}(3 \geq 5)$	0.863	0.736	0.682	0.631	0.573	0.558
$\underline{P}(3 \geq 5)$	0.858	0.724	0.666	0.606	0.522	0.473
$\Delta(3 \geq 5)$	0.005	0.012	0.016	0.025	0.051	0.085
$\overline{P}(3 > 4)$	0.153	0.327	0.421	0.536	0.763	0.902
$\underline{P}(3 > 4)$	0.146	0.311	0.397	0.502	0.707	0.846
$\Delta(3 > 4)$	0.007	0.016	0.024	0.034	0.056	0.056
$\overline{P}(3 \geq 4)$	0.913	0.828	0.795	0.777	0.835	0.915
$\underline{P}(3 \geq 4)$	0.907	0.813	0.775	0.749	0.788	0.864
$\Delta(3 \geq 4)$	0.006	0.015	0.020	0.028	0.047	0.051

Table 2. Some pairwise comparisons between centres.

	$m = 10$		$m = 50$	
i	$[\underline{P}, \overline{P}](i > \max_{j \neq i} j)$	$[\underline{P}, \overline{P}](i \geq \max_{j \neq i} j)$	$[\underline{P}, \overline{P}](i > \max_{j \neq i} j)$	$[\underline{P}, \overline{P}](i \geq \max_{j \neq i} j)$
1	[0.177, 0.197]	[0.369, 0.397]	[0.426, 0.482]	[0.526, 0.583]
2	[0.033, 0.039]	[0.112, 0.128]	[0.022, 0.032]	[0.041, 0.057]
3	[0.061, 0.072]	[0.173, 0.196]	[0.073, 0.098]	[0.114, 0.148]
4	[0.017, 0.022]	[0.067, 0.082]	[0.007, 0.011]	[0.014, 0.022]
5	[0.060, 0.070]	[0.173, 0.193]	[0.069, 0.089]	[0.110, 0.139]
6	[0.021, 0.024]	[0.082, 0.092]	[0.008, 0.011]	[0.017, 0.022]
7	[0.016, 0.020]	[0.067, 0.078]	[0.005, 0.008]	[0.011, 0.016]
8	[0.048, 0.054]	[0.148, 0.163]	[0.042, 0.054]	[0.073, 0.091]
9	[0.030, 0.036]	[0.104, 0.120]	[0.018, 0.026]	[0.034, 0.048]
10	[0.064, 0.074]	[0.179, 0.201]	[0.077, 0.101]	[0.121, 0.153]
11	[0.009, 0.011]	[0.046, 0.052]	[0.001, 0.002]	[0.003, 0.004]
12	[0.022, 0.027]	[0.085, 0.098]	[0.010, 0.014]	[0.020, 0.028]

Table 3. Multiple comparisons between centres.

Further comments and research challenges

- NPI has been presented for many problems in Statistics, Reliability, Risk and OR
- NPI is never in disagreement with inferences based on empirical probabilities, so one could call NPI 'objective'
- A particularly nice NPI model for multinomial data has been presented (FC and Thomas Augustin)
- Main challenges are to develop NPI for multi-dimensional random quantities, including use of co-variates and multivariate statistics
- Most importantly: NPI has helped us to get better understanding of foundations of statistics with imprecise probabilities

References

For more information and references:

www.npi-statistics.com