Minimum Entropy Algorithm
for NPI-generated F-probability intervals

Paul Fink
Department of Statistics, LMU Munich

17 December 2012

NPI 1

Underlying model: Multinomial NPI with inference based upon the next observation

(Graphical) foundation on a probability wheel

n observations create n equidistant intervals on the circle, i.e slices on the wheel

Circular-$A_{(n)}$ assumption gives that next observation will fall into any given slice with probability $\frac{1}{n}$

Restrictions on the ordering of the observations onto the wheel.

NPI 2

Let us assume we have n observations of K different classes, with $n_j \geq 0$ for $j = 1, \ldots, K$

Calculation of the class lower/upper probabilities based on one future observation:

$$\max(0, \frac{n_j - 1}{n}) \leq \mathrm{P}(y_{n+1} = c_j) \leq \min(\frac{n_j + 1}{n}, 1)$$

$\Longrightarrow$ set of F-probability intervals

Idea: Starting with the lower probabilities as working 'distribution', then adding mass to classes until it is a probability distribution.

In what way to assign the mass?

Is it optimal?

Minimum entropy algorithm has already been developed for ordinal-NPI by [Crossman, R.J. et al].

Entropy

Contribution of two classes to the complete entropy H:

$$H_1(x_1, x_2) := -\log(x_1)x_1 - \log(x_2)x_2.$$

Entropy H is concave function, so H_1 too.

Mass assignment of m to either x_1 or x_2 or both.

Taking advantage of the concavity:

- $H_1(x_1 + m - c, x_2 + c) \geq H_1(x_1 + m, x_2) = H_1(x_1, x_2 + m)$ for $0 \leq c \leq m$ and $x_1 = x_2$
- $H_1(x_1 + m, x_2) \leq H_1(x_1, x_2 + m)$ for $x_1 > x_2$
- $H_1(x_1 + m, 0) \leq H_1(x_1, m)$ for $x_1 > 0$

Algorithm outline

Starting with lower probabilities, as these mass assignments are at least required.

In each step assigning as much remaining mass as possible to those classes with highest lower probability.

'as much mass as possible' is enforced by the corresponding upper probability or the probability distribution (sum to 1)

Minimum Entropy Algorithm for NPI

Input:	Probability intervals $[l_i, u_i]_1^n$ as generated by the NPI
Output:	A probability distribution $\hat{p} = (\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_n)$

Helping functions:

Sum(x):	returns the sum of the elements of array x
getMaxIndex(x, S):	returns the first index of the maximum value of the array x considering only indices in S

Initialization: S $\leftarrow 1, \ldots, n$

$minEntropyNPI(l, u, \hat{p})\{$
    for $(i = 1$ **to** $n)$ **do** $\{\hat{p}_i \leftarrow l_i\}$
    $mass \leftarrow 1 - Sum(\hat{p})$
    while $(mass > 0)$ **do** $\{$
        $index \leftarrow getMaxIndex(\hat{p}, S)$
        $d \leftarrow u_{index} - \hat{p}_{index}$
        if $(d \leq mass)$ **then** $\{$
            $\hat{p}_{index} \leftarrow u_{index}$
            $S \leftarrow S - \{index\}$
            $mass \leftarrow mass - d$
        $\}$ **else** $\{$
            $\hat{p}_{index} \leftarrow \hat{p}_{index} + mass$
            $mass \leftarrow 0$
        $\}$
    $\}$
$\}$

Properties

Algorithm does not assign unobserved classes any mass.

Step-wise optimal

Minimum entropy distribution complies with the probability wheel

Algorithm gives only 1 minimum entropy distribution. There may be more!!

- negligible as main interest is entropy value, not underlying distribution

Future Prospects

Minimum and maximum entropy create entropy intervals as guarantee and potential

In case of classification trees:
Choosing on a split variable based on comparisons of those intervals
Reasonable opitmality criteria:

- Maximality (only taking the potential into account)
- Interval dominance

Reference

Crossman, R.J., Abellán, J., Augustin, T. and Coolen, F.P.A. (2011) Building Imprecise Classification Trees With Entropy Ranges.*Proceedings of the Seventh International Symposium on Imprecise Probability: Theories and Applications*.