STUDIES IN HONOUR OF

PENTTI ENSIO MALASKA

ON THE OCCASION OF HIS FIFTIETH
BIRTHDAY

April 11, 1984

Liikkeenjohtotieteen kerho
KVANTTI

Turku 1984

ISBN 951-99537-6-0
Turku 1984

CHEBYSHEV TYPE BOUNDS ON PROBABILITIES FOR A NOMINAL-SCALED RANDOM VARIABLE

Professor
ILKKA VIRTANEN
University of Vaasa
School of Business Studies

ABSTRACT

Much work has been done to find bounds on the probabilities of a random variable in such a case, when information about the distribution is only scantily available. The best known result in this area is Chebyshev's inequality, which can, however, be applied to quantitative variables only. The object of the paper is to develop Chebyshev type bounds for a nominal-scaled variable. The bounds are derived with the following principles: (i) the bounds must presume as little a priori information as possible about the variable, (ii) the amount and type of information presumed must be comparable to the assumptions for Chebyshev's inequality and (iii) the bounds must be fully general, they must be valid for all variables fulfilling the assumptions. The basic information to be used in deriving the bounds consists of the number of classes in the distribution, the mode probability and the entropy of the distribution. The method to be used is based on the geometry of the probability function (pf): the feasible pf's, i.e. the pf's fulfilling the assumptions made, can pass only through a certain region between 0 and 1. The boundaries of this feasible pf-region are found from certain extremal distributions, and hence the required inequalites for the class probabilities can be obtained.

1. BACKGROUND AND NATURE OF THE STUDY

The problem of finding bounds on the probability values of a random variable in such a case, when information about the distribution of the variate is only scantily available, has been the subject of considerable investigation for a period of over hundred years. The best known result in this area is the famous Chebyshev's inequality, which only presumes that the first two moments of the distribution exist. The customary form of this inequality is

(1) $$P\{|X-\mu| \geq k\sigma\} \leq 1/k^2 ,$$

where $\mu = E\{X\}$ is the expectation and $\sigma = D\{X\}$ is the standard deviation of the distribution, and k is an arbitrary positive number. It is well known that the above inequality cannot, without additional information about the distribution, be tightened or improved in its two-sided form.

Some years ago professor Pentti Malaska inspired the author into this problem area by pointing out a spesific geometric method by which one-sided generalizations for inequality (1) might be possible to derive. The succeeded investigations resulted in one-sided inequalities which were more efficient than (1) both in the general case and especially in such cases where some additional information about the distribution was available, see Virtanen (1979 a) and Virtanen (1979 b).

Although inequality (1) is quite general, it can be applied to quantitative variables only. For a qualitative or nominal-scaled variable there exists no result corresponding to (1). In the whole, the concepts and quantities to be used in connection with nominal-scaled variables greatly differ from those used with quantitative variables. One central statistics concerning nominal-scaled variables is entropy. Entropy has been shown to reveal both the dispersion of a variable and (in the case of two- or multidimensional distribution) association between variables, see e.g. Theil (1969), 469-472. As a pioneer in the use of entropy in Finnish business and statistical applications may again be named professor Pentti Malaska (Malaska and Reponen 1979).

The author of this paper has also considered entropy as a statistical index, especially as a measure of statistical dependence (or association), see Astola and Virtanen (1982), Astola and Virtanen (1983), Virtanen (1983), and as a measure of homogeneity in categorical grouping analysis, see Latosaari and Virtanen (1983). In these papers the analogy of the definitions and properties of entropy and coentropy (or joint entropy, entropy of multidimensional distributions) with those of variance and covariance was clearly shown. It became evident that entropy will take the role of variance when a qualitative variable, instead of a quantitative one, is to be handled with.

The lack of Chebyshev type result (1) in the case of a qualitative variable and the promising experiences from the use of entropy as an analogy to variance lead professor Malaska and the author to discuss possibilities to develop Chebyshev type bounds also for a nominal-scaled variable. As a result from a couple of intensive sessions the basic ideas of the present paper were created together. Later on the author developed the ideas further and is alone responsible for the present written form of the paper.

Bearing the background presented above in mind, I will dedicate this paper to professor Pentti Malaska on the occasion of his fiftieth birthday on the 11th April, 1984. I have written it to honour and thank my inspiring teacher, coworker and colleague for a period of more than fifteen years.

2. PURPOSE AND BASIC ASSUMPTIONS OF THE STUDY

The paper deals with distributions of random variables which are, or at least are considered as, qualitative or nominal-scaled by nature. The basic problem is formulated as follows. What can be said about the distribution of the variable when a certain minimal amount and type of information about the variate is available? What are the feasible distributions under this minimal set of assumptions? How much is the knowledge about the distribution improved, when a certain additional assumption is made? etc.

Or to be more spesific, the object of the paper is to develop "Chebyshev type" bounds, cf. inequality (1), for a nominal-scaled random variable. The bounds must be fully general, i.e. they must be valid for all variables that fulfill the assumptions, and they must presume as little information as possible about the variate in advance, the amount and type of information presumed being comparable to the assumption about the existence of the first two moments in the case of a quantitative variable. The information being used in deriving the bounds consists, besides the natural basic assumption, in the following order of the knowledge about the number of classes, the mode probability and the entropy of the distribution.

The basic assumption (in fact, a natural requirement) concerning the class probabilities of any nominal-scaled random variable is that the probabilities form a distribution, i.e. that the conditions

(2) $$p_c \geq 0\,,\quad c \in C$$

(3) $$\sum_{c \in C} p_c = 1$$

hold (C is the set of the classes, either finite or infinite).

The second basic assumption concerns the numbering of the classes. For a true nominal-scaled variable (the classes of the variable cannot even be ordered) the numbering is arbitrary. Without loss of generality we can therefore assume that the classes have been numbered according to the descending order of the class probabilities. This has no other effect that it makes the notation simpler. We can therefore assume

(4) $$1 \geq p_1 \geq p_2 \geq p_3 \geq \dots \geq 0\,.$$

3. DERIVATION OF THE BOUNDS FOR THE CLASS PROBABILITIES

3.1. Feasible distributions when the number of the classes is given

In this section we assume that the number of the classes is the only additional information available for deriving bounds tighter than conditions (2) - (4) for the class probabilities of the feasible distributions. The following notation will be used: n is the number of classes, p_k^- is the lower bound and p_k^+ the upper bound for the kth class probability, $k = 1,2,\dots,n$.

For the first (i.e. the mode) class we must have $1/n \leq p_1 \leq 1$. The upper bound is evident, the lower bound is due to the way of numbering the classes (condition (4) must hold). We thus have

(5) $$\begin{cases} p_1^- = 1/n \\ p_1^+ = 1\,. \end{cases}$$

For the second class we have $0 \leq p_2 \leq 1/2$. Now the lower bound is evident (it is reached when $p_1 = 1$), the upper bound is a consequence of the inequality $p_2 \leq p_1$ (condition (4)) and equation (3), the bound is reached when $p_1 = p_2 = 1/2$ (and $p_3 = p_4 = \dots = p_n = 0$). On similar lines we obtain in general

$$(6) \quad \begin{cases} p_k^- = 0\,, & k = 2,3,\dots,n \\ p_k^+ = 1/k\,, & k = 2,3,\dots,n\,. \end{cases}$$

Thus, the total region of class probabilities between 0 and 1 is by (5) and (6) divided into two parts: the feasible and non-feasible part (see Fig. 1). From Fig. 1 we can see

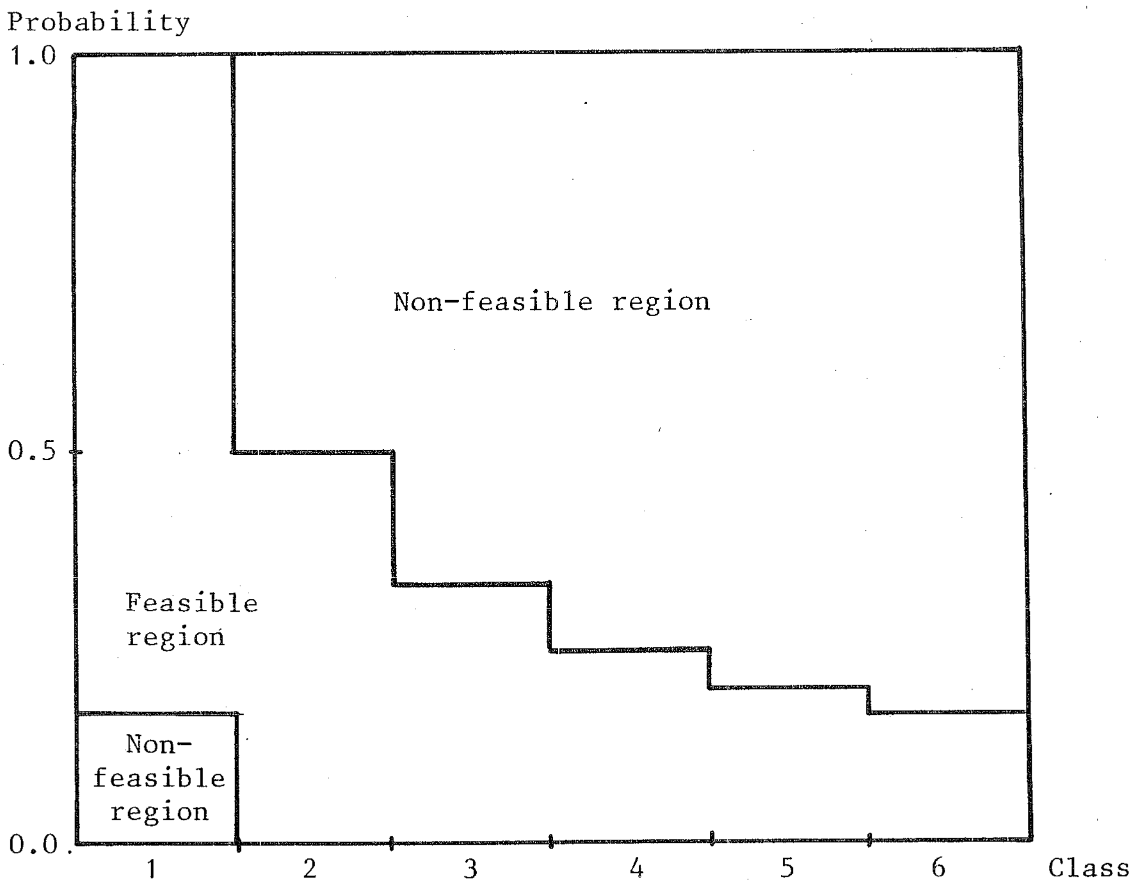


Figure 1. The region of feasible distributions bounded by p_k^- and p_k^+.

that knowledge of the number of classes (together with the numbering assumption (4)) has a strongly shrinking effect on the region of the feasible distributions. However, it is worth to make some comments concerning the feasible region bounded by (5) and (6). First, under the given amount of information the bounds (5) and (6) cannot be improved: all the bounds can be reached (but not violated). Second, the "boundary probabilities" p_k^- and p_k^+ don't compose any distribution, they don't fulfill equation (3). In fact we have (for the non-trivial case $n > 1$)

(7) $$\sum_{k=1}^{n} p_k^- = 1/n < 1 \quad \text{and}$$

(8) $$\sum_{k=1}^{n} p_k^+ = \sum_{k=1}^{n} \frac{1}{k} > 1 .$$

Third, the basic indices (mode probability and entropy) of the feasible distributions are not restricted in any (other than trivial) way. For the mode probability p_0 $(= p_1)$ we have

(9) $$1/n \leq p_0 \leq 1$$

and for entropy H we may easily obtain

(10) $$0 \leq H \leq \ln n .$$

3.2. Feasible distributions when the number of the classes and the mode probability are given

Let the mode probability be known and denoted by p_0. The bounds for the class probabilities become now functions of p_0. Let these bounds (to be derived) be denoted by $p_k^-(p_0)$ and $p_k^+(p_0)$. On the basis of assumption (4) we have $p_1 = p_0$, i.e. for the first class it holds

(11) $$p_1^-(p_0) = p_1^+(p_0) = p_0 .$$

If $p_0 = 1$ $(= p_1^+)$ or $p_0 = 1/n$ $(= p_1^-)$, the feasible distribution is uniquely determined. In the first case we have $p_1 = 1$, $p_2 = p_3 = \dots = p_n = 0$, and in the latter case we have the uniform distribution $p_1 = p_2 = \dots = p_n = 1/n$. Therefore, let us in the following consider the non-trivial case $1/n < p_0 < 1$ only. For the second class it must hold

$$(12)\qquad p_2 \leq \min\{p_0, 1-p_0\}$$

(the upper bound is obtained when the whole surplus probability mass $1-p_0$ is located, provided it does not exceed the mode probability p_0, in the second class), and

$$(13)\qquad p_2 \geq \frac{1-p_0}{n-1}$$

(the lower bound is obtained in the case $p_1 = p_0$, $p_2 = p_3 = \dots = p_n = (1-p_0)/(n-1)$). Thus we have

$$(14)\qquad \begin{cases} p_2^-(p_0) = \dfrac{1-p_0}{n-1} \\ p_2^+(p_0) = \min\{p_0, 1-p_0\}\,. \end{cases}$$

In general, the upper bound for the probability p_k is obtained when the probability mass $1-p_0$ is uniformly distributed into the $k-1$ classes $2,3,\dots,k$ (provided this bound does not exceed p_0). The lower bound for this probability is 0, if $p_0 \geq 1/(k-1)$ (the whole probability mass can be located in the first $k-1$ classes), otherwise the bound is obtained by setting $p_1 = p_2 = \dots = p_{k-1} = p_0$ and dividing the remaining probability mass $1-(k-1)p_0$ uniformly into the $n-(k-1)$ classes $k,k+1,\dots,n$. We have thus obtained for $k = 2,3,\dots,n$:

$$(15)\qquad \begin{cases} p_k^-(p_0) = \max\left\{0, \dfrac{1-(k-1)p_0}{n-(k-1)}\right\} \\ p_k^+(p_0) = \min\left\{p_0, \dfrac{1-p_0}{k-1}\right\}. \end{cases}$$

It is evident that

$$(16)\qquad \begin{cases} p_k^-(p_0) \geq p_k^- \\ p_k^+(p_0) \leq p_k^+ \end{cases}$$

for all $1/n \leq p_0 \leq 1$ and for all classes $k = 1,2,\ldots,n$. A fixed mode probability thus shrinks the original region of the feasible distributions. The shrinkage of the feasible region is illustrated in Fig. 2 in the cases $p_0 = 0.8$ and $p_0 = 0.4$ (and with six classes in the distribution).

From (14) we can see that in the non-trivial case $1/n < p_0 < 1$ we have $p_2 \geq p_2^-(p_0) = (1-p_0)/(n-1) > 0$, i.e. the second class cannot be empty. This question can also be considered more generally: under which conditions must the kth class necessarily be non-empty, what is maximum number of empty classes in a feasible distribution etc.

The answer to these questions can be found easily from (15): class k must be non-empty if and only if $p_k^-(p_0) > 0$, i.e. if

(17) $$1 - (k-1)p_0 > 0 .$$

Or in other words, if we have

(18) $$p_0 < \frac{1}{k-1},$$

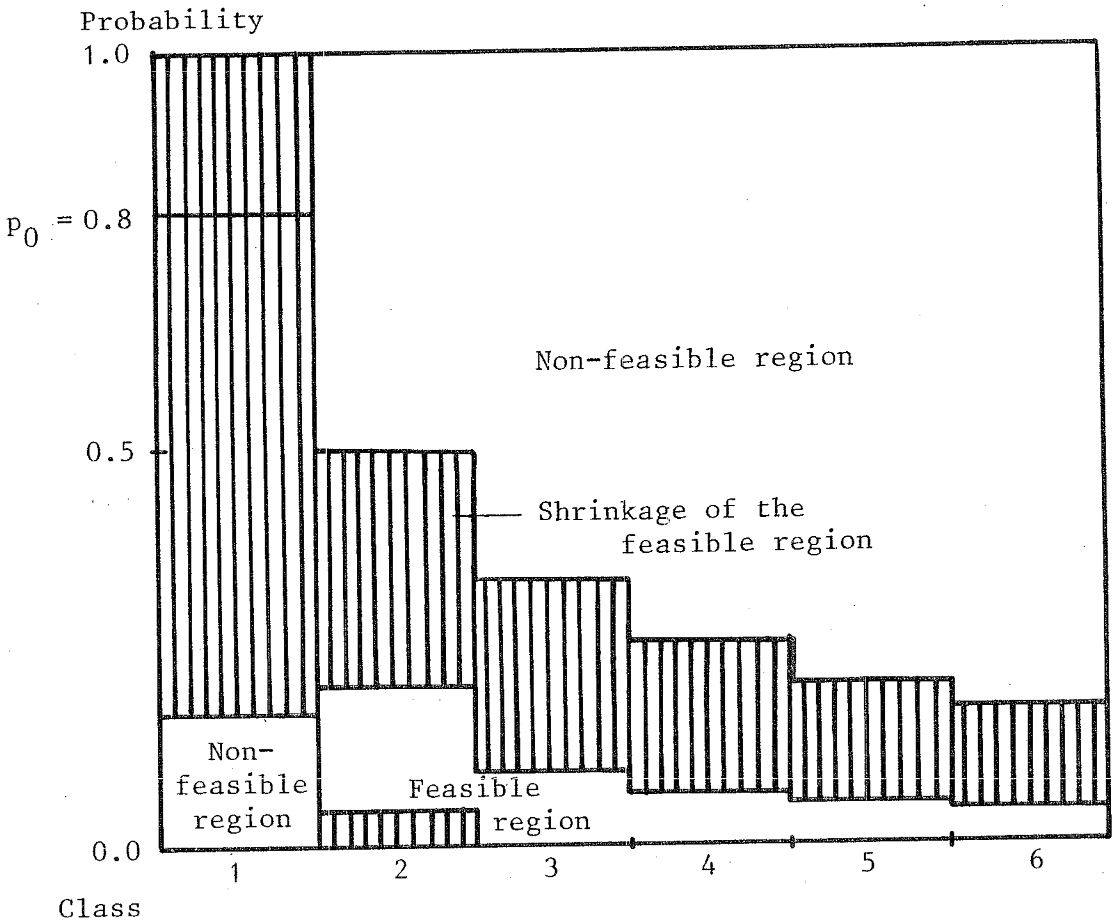
class k cannot be empty. This further implies that none of the classes $1,2,\ldots,k-1$ can either be empty. If, in addition to (18), we also have $p_0 \geq 1/k$, class k is the last class which necessarily must be non-empty. We have thus obtained the following condition for the maximum number of empty classes in a distribution: provided that it holds for the mode probability of a distribution

(19) $$\frac{1}{k} \leq p_0 < \frac{1}{k-1},$$

the distribution may have $n-k$, but no more, empty classes.

Another problem of interest is to consider in which classes a given level, say $\bar{p}$, for the probability may be reached (this problem is not, of course, relevant for those classes, for which we alredy have $p_k^+ < \bar{p}$). The level $\bar{p}$ can no more be reached in class k, if

(20) $$p_k^+(p_0) < \bar{p},$$



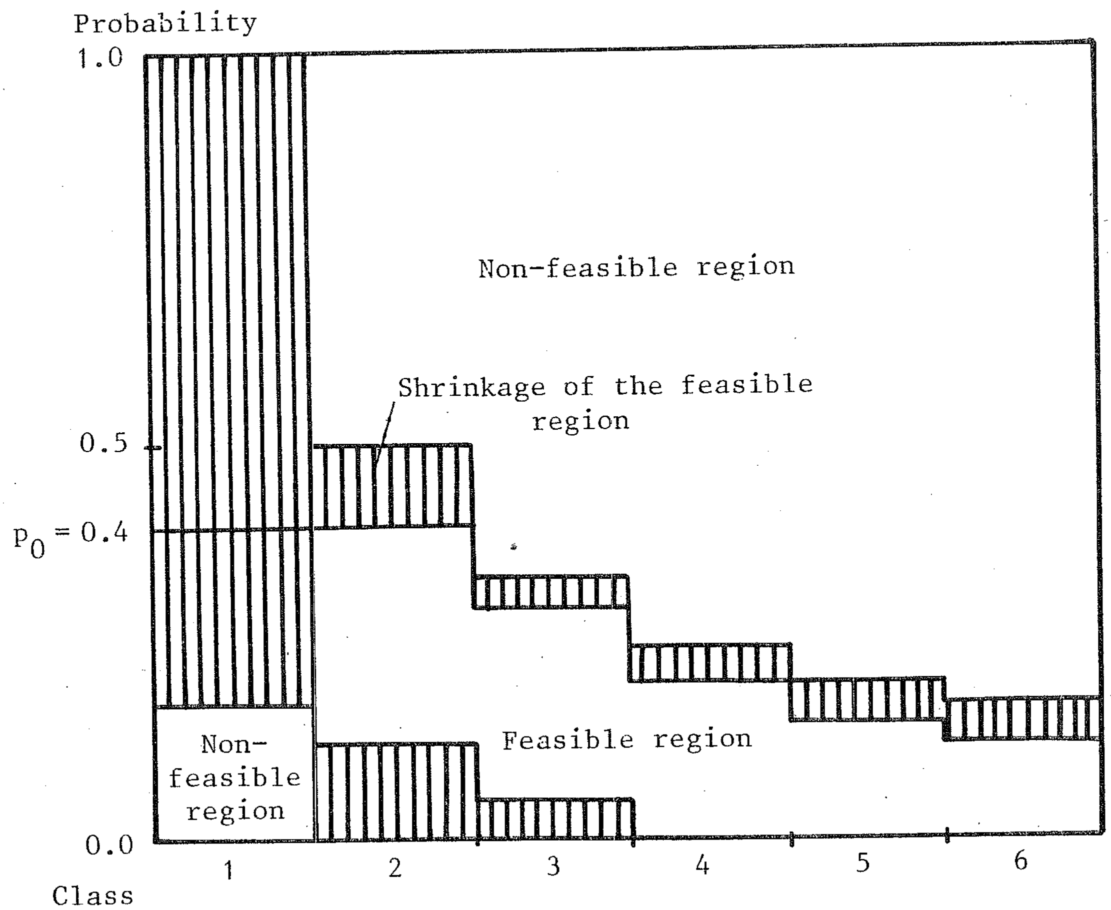


Figure 2. The shrinkage of the region of feasible distributions due to fixing the mode probability.

or, using (15), if

(21) $$\frac{1-p_0}{k-1} < \bar{p}\,.$$

As far as the mode probability is concerned, this gets the condition

(22) $$p_0 > 1-(k-1)\,\bar{p}$$

to imply that the class probability p_k (and the class probabilities $p_{k+1},\ldots,p_n$ either) cannot reach the level $\bar{p}$.

When only the number of the classes is fixed (= n), entropy can vary between 0 and ln n in the family of the feasible distributions. Fixing the mode probability sets, however, restrictions for the possible values of entropy. With given p_0 (= p_1), the maximum entropy is obtained when $p_1 = p_0$ and the remaining probability mass $1-p_0$ is uniformly distributed into the classes 2,3,...,n. The upper bound for the possible values of entropy is the following:

(23) $$H^+(p_0) = -\,p_0 \ln p_0 - \sum_{k=2}^{n} \frac{1-p_0}{n-1} \ln \frac{1-p_0}{n-1}$$

$$= -\,p_0 \ln p_0 - (1-p_0) \ln (1-p_0) + (1-p_0) \ln (n-1)\,.$$

The minimum entropy, on the other hand, is obtained when the probability mass is centralized into as few classes as possible and the rest classes are left empty. Bearing the constraints due to assumptions in mind we notice that the distribution having the minimum entropy is the following: $p_1 = p_2 = \ldots = p_r = p_0$, $p_{r+1} = p'$, $p_{r+2} = \ldots = p_n = 0$, where we have denoted

(24) $$r = \left[1/p_0\right],$$

i.e. r is the greatest integer not exceeding $1/p_0$, and

(25) $$p' = 1 - rp_0 = 1 - \left[1/p_0\right] p_0\,.$$

We have obtained the following lower bound for entropy of the feasible distributions:

(26) $$H^{-}(p_0) = - rp_0 \ln p_0 - p' \ln p' .$$

For example, the entropies of the feasible distributions in Fig. 2 vary between 0.500 and 0.694 (the case $p_0 = 0.8$) or between 1.055 and 1.639 (the case $p_0 = 0.4$). These can be compared with the basic case (mode probability not fixed), where the bounds are 0 and 1.792, respectively.

3.3. Feasible distributions when the number of the classes and the entropy of the distribution are given

In this section we assume that the information available in advance consists of the knowledge of the number of classes and of the value of the entropy (but not of the mode probability). With n classes in the distribution, the given value of entropy, say H_0, must lie between 0 and $\ln n$. The problem now becomes: with n classes in the distribution and with given entropy H_0, $0 \leq H_0 \leq \ln n$, what is the family of the feasible distributions, i.e. what are the bounds on the class probabilities defining this family? Or for each class k, $k = 1,2,\ldots,n$, we seek bounds (as tight as possible) on class probabilities p_k

(27) $$p_k^{-}(H_0) \leq p_k \leq p_k^{+}(H_0)$$

such that the entropy condition

(28) $$- \sum_{k=1}^{n} p_k \ln p_k = H_0$$

(and also the basic assumptions (2) - (4)) hold.

Due to the non-linearity of equation (28) the problem cannot be solved analytically. But a numerical solution can be obtained quite straightforwardly. The general form of the whole algorithm will not be presented here, only the basic ideas of the procedure and some numerical results are shown.

The lower and upper bounds for the class probabilities are based on certain extremal distributions (as, in fact, was the case when the mode probability was fixed). The extremal distributions consist of two or three groups of class probabilities: in n_h first classes we have a high class probability p_h, i.e. $p_1 = p_2 = ... = p_{n_h} = p_h$, in the following n_l classes we have a low class probability p_l, i.e. $p_{n_h+1} = ... = p_{n_h+n_l} = p_l$, and the last $n_0 = n - n_h - n_l$ classes are empty (the group of empty classes can exist, however, only when H_0 is small enough to make it possible to locate the whole probability mass into fewer than n classes; $H_0 \leq \ln(n_h + n_l)$). The extremal distributions are determined so that p_h and p_l differ as much as possible, (28) must be valid, however. Searching through all the possible extremal distributions, the bounds (27) are obtained.

Table 1 gives some numerical results on the subject. We have chosen $n = 5$. This implies that we must have $0 \leq H_0 \leq \ln 5 = 1.609$. We have computed the bounds for three possible values of entropy, viz. $H_0 = 1.4$, $H_0 = 1$ and $H_0 = 0.3$. Let us consider, for example, the extremal distributions determining the boundaries for the third class. Take first the case $H_0 = 1.4$. The upper bound for p_3 is found from the distribution $p_1 = p_2 = p_3 = 0.297$, $p_4 = p_5 = 0.055$: $p_3^+(H_0) = 0.297$. The distribution $p_1 = p_2 = 0.360$, $p_3 = p_4 = p_5 = 0.093$ gives the lower bound for p_3: $p_3^-(H_0) = 0.093$. In the case $H_0 = 1$ the extremal distributions for class 3 are $p_1 = p_2 = 0.461$, $p_3 = p_4 = p_5 = 0.026$ (gives $p_3^-(H_0) = 0.026$) and $p_1 = 0.550$, $p_2 = p_3 = 0.225$, $p_4 = p_5 = 0$ (gives $p_3^+(H_0) = 0.225$). For $H_0 = 0.3$ we find $p_1 = 0.911$, $p_2 = 0.089$, $p_3 = p_4 = p_5 = 0$ (gives $p_3^-(H_0) = 0$) and $p_1 = 0.930$, $p_2 = p_3 = 0.035$, $p_4 = p_5 = 0$ (gives $p_2^+(H_0) = 0.035$). From the values of the table we see that the region of the feasible distributions with the value of entropy fixed is strongly affected by that fixed value.

Table 1. The bounds on class probabilities with given number of classes and with given entropy as a function of the entropy.

Number of the class	Bounds with given number of classes (n=5)		Bounds with given number of classes (n=5) and given entropy H_0					
			$H_0 = 1.400$		$H_0 = 1.000$		$H_0 = 0.300$	
(k)	p_k^-	p_k^+	$p_k^-(H_0)$	$p_k^+(H_0)$	$p_k^-(H_0)$	$p_k^+(H_0)$	$p_k^-(H_0)$	$p_k^+(H_0)$
1	0.200	1.000	0.249	0.490	0.430	0.712	0.911	0.942
2	0	0.500	0.128	0.360	0.072	0.461	0.014	0.089
3	0	0.333	0.093	0.297	0.026	0.225	0	0.035
4	0	0.250	0.055	0.249	0	0.111	0	0.021
5	0	0.200	0.002	0.128	0	0.072	0	0.014

The next step in increasing the amount of available information would be the case where both the mode probability and the entropy of the distribution are known in advance. The knowledge about the mode probability fixes the first class probability $p_1 = p_0$, bounds for the other classes are obtained considering the remaining probability mass $1 - p_0$ on similar lines as above. The procedure is, however, here omitted.

4. CONCLUSIONS

We have considered the problem of deriving bounds for the class probabilities of a nominal-scaled random variable under minimal a priori information. We have tried to maintain certain analogies to the famous Chebyshev's inequality, especially in the amount and type of available information. It turned out that the assumption concerning the knowledge of the mode probability and of the number of the classes may be regarded as a substitute for the assumption about the existence of the first two moments in Chebyshev's inequality. Under this assumption, the bounds for the class probabilites were obtained in analytical form. The bounds became functions of both the mode probability (location parameter) and the number of classes (range or dispersion parameter). It was also possible to take entropy as a priori information. A fixed entropy has a strong effect on the shape of the feasible distributions, it must, in this connection, be regarded as a shape parameter. As a function of fixed entropy the bounds can be derived only numerically. The paper also contains some examples of this numerical procedure.

REFERENCES

Astola, J. - Virtanen, I. (1982). Entropy correlation coefficient, a measure of statistical dependence for categorized data. Proceedings of the University of Vaasa, Discussion Papers 44.

Astola, J. - Virtanen, I. (1983). A measure of overall statistical dependence based on the entropy concept. Proceedings of the University of Vaasa, Research Papers 91.

Latosaari, E. - Virtanen, I. (1983). Entropy as a measure of homogeneity in categorical grouping analysis. Proceedings of the University of Vaasa, Research Papers 96 (in Finnish, English summary).

Malaska, P. - Reponen, T. (1979). Problem areas in management in the 1980's, in Conference of the Researchers of the Turku School of Economics 1979. Publications of the Turku School of Economics, Series A-7:1979 (in Finnish).

Theil, H. (1969). On the use of information theory concepts in the analysis of financial statements. Management Science 15:9, 459-480.

Virtanen, I. (1979 a). One-sided bounds of the Tchebycheff type for some general classes of distributions. Lappeenranta University of Technology, Department of Physics and Mathematics, Research Report 1/1979 (in Finnish, English summary).

Virtanen, I. (1979 b). One-sided bounds of the Tchebycheff type for some general classes of distributions, in Twelffh European Meeting of Statisticians (EMS-1979) Abstracts.

Virtanen, I. (1983). An entropy based correlation coefficient for categorized data, in Meristö, T. - Heinistö, K. (eds.): Essays in Management Studies. Visit to Boston 1983, 15-18.