University of Southampton

Discussion Papers in Management

A Note on Coarse Classifying in Acceptance Scorecards

Ki Mun Jung
Kyungsung University
South Korea

and

Lyn C Thomas
University of Southampton
UK

February 2004

Number M04-16
ISSN 1356-3548

A Note on Coarse Classifying in Acceptance Scorecards

Ki Mun Jung,

Department of Informational Statistics,

Kyungsung University,

Busan 608-736,

South Korea

Lyn C Thomas

School Of Management,

University of Southampton,

Southampton

Abstract: Traditionally in credit and behavioural scoring one assumes that as all consumers have essentially the same product , its features will not affect whether the consumer defaults or not. Hence one coarse classifies the characteristics concentrating only on the default ratio. As one increasingly customises products and their operational features for each individual, (the very purpose of acceptance scoring), then decisions like whether the customer will accept the product or not must depend on the features offered. This paper investigates how one can deal with this dependency when coarse classifying the characteristics.

Introduction

In classification problems like credit scoring one seeks to relate a consumer's characteristics, such as age, income or years with bank to an outcome variable, such as whether the consumer will default or not. The relationship is almost always non-linear and often non-monotonic. To cope with that one could either seek to fit some non-linear function of age to likelihood of default (by cubic splines for example) or coarse classify the characteristics. The latter, in which one splits the range of the characteristic into a number of separate sets or "bins" and defines a new binary variable for each set so formed, is almost exclusively used in practice. It is easy to understand because the characteristic is defined by a set of scores – one for each bin. It is robust to data changes and it can be used for categorical characteristics like occupancy status as well as continuous ones, so one might put owner occupier and living with parents in one bin and tenant of unfurnished occupancy and tenant of furnished occupancy in another.

Coarse classification revolves about which "bins" or bands (usually called attributes) should be created, i.e. is splitting age into age intervals, 18-25, 26-35, 36-55, over 55 better than splitting into age groups 18-21, 22-30, 31-50 and over 50. This is done by comparing how well these splits discriminate on the dependent variable (whether the consumer defaults or not). Normally what happens in the case of continuous variables is that one splits the characteristics into far more bins than one is expecting to use (maybe groups of deciles or bins each of which have 5% of the population in them) and then combines adjacent bins if they are have similar relationships with the dependent variable. Table 1 shows the situation in an acceptance scorecard (one is trying to estimate which people will accept a financial product being offered to them) of the dependent variable (accept/reject) with age. This was data relating to student bank accounts which explains the unusual age distribution.

Age	16-18	19	20	21-22	23-24	25-26	27-28	29-33	34-39	Above 40	Total
Accept	3	12	8	10	10	6	8	6	7	6	76
Reject	2	13	3	4	3	4	3	3	3	3	41
Total	5	25	11	14	13	10	11	9	10	9	117

Table 1. Numbers of offers accepted and rejected in 10 age bands.

Three statistic – the χ^2 statistic, the F information statistic and Somer's D concordance statistic- are commonly used to describe how good a particular coarse classification is at differentiating the tow outcomes on the dependent variable. The most commonly used is the χ^2 statistic which wants the splits that are most unlikely to support the hypothesis that all the attributes draw from populations which have the same proportion of accepts in them. Using that statistic here, let a_i and r_i be the number of accepts and rejects with attribute i and let a and r be the total number of accepts and rejects. Let $\hat{a}_i = (a_i + r_i)a/(a+r)$ and $\hat{r}_i = (a_i + r_i)r/(a+r)$ be the expected number of accept and rejects with attribute i if the proportion of accepts is the same for all attributes. Then the χ^2 statistic is defined by

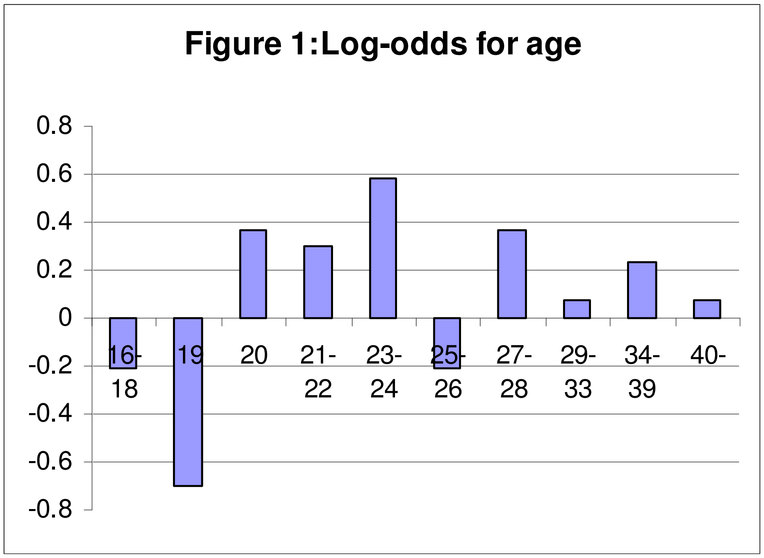
$$\chi^2 = \sum_i \frac{(a_i - \hat{a}_i)^2}{\hat{a}_i} + \frac{(r_i - \hat{r}_i)^2}{\hat{r}_i}.$$

The larger the value of χ^2, the better is the coarse classifying in that it less likely that the different attributes have the same distribution of accepts and rejects. In the above example for ease of understanding, we only look at splits of age into two attributes, so there are 9 possible splits into two intervals by combining adjacent "bins". Table 2 gives the χ^2 values for these 9 splits and shows that splitting between 19 and 20 is optimal.

Split	Chi-square value
16-18, Above 19	0.0564
16-19, Above 20	3.9652
16-20, Above 21	2.1766
16-22, Above 23	1.1205
16-24, Above 25	0.2115
16-26, Above 27	0.4694
16-28, Above 29	0.1360
16-33, Above 34	0.1196
16-39, Above 40	0.0125

Table 2. Chi square value for age splits

Even in this case if one wanted to look at all possible splits there would be $2^9 = 512$ different combinations to consider and in most problems this is not a realistic thing to do. A crude but quite effective method of dealing with this problem is to use subjective judgment based on the ratios of accepts to rejects in the different bands. In this one combines adjacent bands if they have somewhat similar ratios, and one always considers non-adjacent intervals of a characteristic as separate attributes even if their ratios are very close. Normally one computes the log odds for each band i.e. $\log(a_i r / r_i a)$ for band i and so for the bands given in Table 1 one would have the following diagram



Looking at this one is still led to the 16-19,20 and over split or if one wanted to put in two splits possibly 16-19, 20-24, 25 and over.

This is the norm for coarse classifying as outlined in McNab and Wynn[1] or Thomas, Crook and Edelman[2]. However recently there has been a move to customization of offers and differential customer relationships which means that different versions of a product are offered to different applicants (or the credit rules and limits applied to the borrowers are different). Clearly the choice of the version of the product offered must affect the likelihood of the person accepting that product for otherwise there would be no point in offering different versions of the product. Hence this choice affects the accept reject ratio and so the choice of bands to arrive at in coarse classifying the characteristics. The purpose of this note is to investigate possible ways of dealing with this problem using as an illustration the example outlined above. Although we concentrate on modelling the likelihood of acceptance when there are different versions of the product offered, an identical analysis would need to be performed to model the likelihood of a consumer defaulting if different consumers are subject to different credit regimes.

Ways of coarse classifying in acceptance scorecards

Customer relationship management in financial organisation means trying to ensure that the service provided to the customer meet their needs and their aspirations. This begins with the choice of product to offer them as increasing a financial organisation has a generic product – credit card, mortgage or current account with overdraft facility- which has a number of features that can be varied from customer to customer. Thus for the credit card this would be the overdraft limit, the interest rate charged, whether there is an annual fee, what free air miles of other points are available For a current account this would be the overdraft limit, the interest rate charged on overdrafts, the interest rate paid when the account is in credit, what free gifts are given on joining, and whether there is no fee on foreign exchange transactions.

The objective of an acceptance scorecard is to estimate the probability $p(\mathbf{x},\mathbf{o})$ that a consumer with characteristics $\mathbf{x}$ is likely to accept a product with offer characteristics $\mathbf{o}$. One can do this by taking the result (they accepted/they rejected) of the offers made to a sample of previous customers. This is very similar to credit scoring (estimating how likely customers with characteristics $\mathbf{x}$ are likely to default in a given period) and so one can use exactly the same methodology as is used in credit scoring including the idea of coarse classifying the characteristics. However it is clear that whether a consumer accepts the product offered depends heavily on the features of the particular offer made to them as well as on their characteristics. This is substantially different to what traditionally was assumed in credit scoring namely that the lending product features like the credit limit did not affect the probability of default. So if we believe there is this interaction between the outcome variable and the offer characteristics how do we coarse classify a consumer characteristic like age?

To answer this question we concentrate on the example given in Table 1, which gives the relationship between age and acceptance/rejection of the offer of a student account in a sample of 117 students. The student account offer had several features – what was the overdraft limit (no interest was charged if inside this limit), the interest paid when the account was in credit, whether or not there was a no fee on foreign exchange transactions, whether there was an offer of a credit card with the account and the free gift offered initially. A questionnaire of the students confirmed that the overdraft limit was by far the most important of these features and in this case there were five levels of overdraft limit offered - £1000, £1250, £1400, £1800 and £2000.

We considered three main approaches to the problem – ignore the connection of acceptance with overdraft limit, consider each overdraft limit separately and try to find the best coarse classification overall or coarse classify the impact of the overdraft limit and then consider each bin of overdraft limits separately and try to find the best coarse classification over all these bins.

Ignoring the connection leads to the analysis in the introduction where one takes the numbers of accepts and rejects in each bin irrespective of what overdraft limit was offered to each consumer. This has the advantage of ease of calculation but it is counter intuitive to say that the overdraft limit offered is not important when the whole purpose is to calculate the change in probabilities that changing the level of the overdraft limit will have on the acceptance probability. The second approach means looking at the consumers offered each overdraft limit separately and determining for each such subgroup the chi-square values for each different split combination and then averaging these over the groups in some suitable way. The problem in this example (and in many others) is that the numbers in some of the offer bands are so small that when one starts splitting them by age bands the numbers are too small for the χ^2 statistic to have any statistical meaning. In this case one was trying to calculate χ^2 values on 10 cells when there were less than 20 cases in a subgroup in total. Even with much larger samples than the one used here if there are large numbers of combinations of offer features this problem will persist.

The third approach was to initially bin the overdraft limit into attributes which were consistent in the way they affected the accept/reject decision, just as was done for age in the introduction. This suggested splitting overdraft limit into two groups (£1000 and £1250) and (£1400 and above). There were 65 observations in the first group and 52 in the second group. Now we wish to determine which is the best split on age which works best over these two bands of overdraft limit? One could of course have different age bands for the two different overdraft limit groups as is done when the population is segmented but it was felt this would not be sensible if there were more than two or three overdraft limit bands, which might well be the case in more realistic examples. The way we chose to measure the effectiveness of age splits using the χ^2 values was as follows. Let n, n_1 and n_2 be the number of observations in the whole of the original sample, the number in the first overdraft subgroup and the number in the second overdraft subgroup respectively. Let χ_1^2 and χ_2^2 be the Chi-square values for a split on subgroup1 and subgrop2 respectively. Then the weighted mean of the Chi-square values is defined by

$$\chi_{wm}^2 = (n_1/n)\chi_1^2 + (n_2/n)\chi_2^2 .$$

Table 3 shows the Chi-square value for each split in the two subgroups of the overdraft limit and the weighted mean of the Chi-square values for each split. From Table 3 the best split (again only single splits were considered) was to divide age into 16-19 and 20 and over. It is interesting that this best split is the same as when the original whole sample was used. However this is not always the case. If for example we had split overdraft limit into three groups - £1000 only, £1250 only and £1400 or more then the best one split on age is to divide into classes 16-28 and above 29 and over.

No.	Splits	£1000-£1250	£1400 or more	Weighted mean of
		Chi-square values	Chi-square values	Chi-square values
1	16-18, Above 19	1.1849	0.0384	0.6753
2	16-19, Above 20	5.3565	0.0739	3.0087
3	16-20, Above 21	2.3214	0.0168	1.2971
4	16-22, Above 23	0.7753	0.0566	0.4559
5	16-24, Above 25	1.0317	0.6017	0.8406
6	16-26, Above 27	2.0518	0.7242	1.4618
7	16-28, Above 29	2.2196	2.4352	2.3154
8	16-33, Above 34	2.4074	2.6365	2.5092
9	16-39, Above 40	0.9758	2.0377	1.0033

Table 3. Chi-square values with two subgroups of credit limits

Using monotone property in coarse classifying

One can get more robust results on the appropriate splits in coarse classifying variables if one is willing to accept the idea of a monotone "utility" function on the continuous offer features such as the credit limit. By this we mean that those who rejected an overdraft limit of a given amount will also reject an identical but offer with an even smaller overdraft limit. This seems a very reasonable assumption to make. The converse assumption is slightly more contentious, namely that if some one accepts an offer of an account with a certain overdraft limit they will also accept the identical offer if the overdraft limit is raised. The question here is whether some consumers wish to restrict their ability to get into too much debt. If we take both assumptions to hold ,we say the monotone property exists.

In that case one can add fictional results to the sample. So someone who accepted an offer with an overdraft limit of £1250, can also be assumed to have accepted the same offer when the overdraft limit was £1400, or £1800 or £2000. Someone who rejected an offer when the overdraft limit was £1400 will also be assumed to reject the offer if the overdraft limit was £1250 or £1000. Thus theoretically one could have accept/reject decisions for all the consumers in the sample at all overdraft limits. The reality of course is that this will not happen but it will ensure that the accept rate is likely to increase as the overdraft limit is increased. Using this monotone property for the data in Table 1 leads to the revised data set given in Table 4. Note that this idea increases the apparent size of the sample four fold.

	16-18	19	20	21-22	23-24	25-26	27-28	28-33	34-39	Above 40	Total
Accepts	8	49	32	44	30	19	36	26	29	32	305
Rejects	6	26	9	11	6	6	5	6	9	7	91
Total	14	75	41	55	36	25	41	32	38	39	396

Table 4. Numbers of accepts and rejects using monotone property

We can now consider using the three approaches suggested for coarse classification in the previous section but with this enhanced data set. One feels that the first approach of using the whole sample as is has more validity now as some of the impact of the overdraft limit on the accept/reject decision has now been included in the enhancement of the data set. The chi-square values for each binary split using this approach is given in Table 5 which is the equivalent of

Table 2 with the monotone property included. In this case the same split 16-19 and 20 and over turns out to be the most discriminating.

Split	Chi-square value
16-18, Above 19	3.2399
16-19, Above 20	10.9201
16-20, Above 21	8.0097
16-22, Above 23	5.1592
16-24, Above 25	3.0112
16-26, Above 27	3.3832
16-28, Above 29	0.6644
16-33, Above 34	0.2615
16-39, Above 40	0.6187

Table 5. Chi-Square value for age splits with monotone property assumed

Using the monotone property overcomes one of the difficulties of the second approach of creating subgroups for each different overdraft limit separately in that the numbers in each subgroup are considerably enhanced. It also gives some connections between the separate subgroups. However if one has a lot of different possible overdraft limits then one still has calculate the chi-squared values for each possible split on each overdraft limit subgroup and that can be a considerable set of calculations. It may still be better to first split the overdraft limits into a few subsets. (One uses the enhanced sample to do this) and then apply the weighted mean approach to the splits on each of the subsets or bands of overdraft limits created. Doing this in this case meant one still chose the best two subgroup split to be overdraft limits of £1000 and £1250 in one group and

£1400, £1800 and £2000 in the other group. Calculating the weighted mean of the chi-square values as was done in Table 3 led to Table 6 when the monotone property had been assumed.

No.	Splits	£1000-£1250 Chi-square values	£1400 or more Chi-square values	Weighted mean of Chi-square values
1	16-18, Above 19	4.0368	0.7727	2.8004
2	16-19, Above 20	16.4425	0.1535	10.2724
3	16-20, Above 21	8.6137	0.9960	5.7282
4	16-22, Above 23	3.0949	1.9536	2.6626
5	16-24, Above 25	4.6518	0.0309	2.9015
6	16-26, Above 27	4.8795	0.0654	3.0560
7	16-28, Above 29	5.1779	1.7219	3.8688
8	16-33, Above 34	4.4678	2.2231	3.6175
9	16-39, Above 40	1.6445	2.6273	2.0168

Table 6. Chi-square values for two subgroups of credit limits when monotone property is used

Thus in this example it is reassuring to see that the four ways suggested all end up with the same split. This is certainly not always the case.

Conclusions

This note identifies a new problem which arises when one seeks to expand the use of credit scoring techniques to other applications such as acceptance scoring. It occurs because instead of seeking just to estimate, $p(\mathbf{x})$, the probability that the relevant outcome (default, acceptance) will occur to or be chosen by a consumer with personal characteristics **x**, one now wishes to estimate

$p(\mathbf{x},\mathbf{o})$ where the **o** are features of the product which strongly affect whether or not this relevant outcome will occur. Ignoring this feature can invalidate the assumption which underlie the idea of coarse classifying the variables **x** so as to allow for a simple but not necessarily monotone connection between **x** and the probability $p(\mathbf{x})$.

Computational analysis on an acceptance scorecard problem suggests that if it seems reasonable that the particular offer feature would have the monotone property (credit limits and interest rate levels seem to have this property) this should be used. Whether it is used or not, the approach which initially bands credit limits into a small number of subgroups depending on their relationship to the outcome variable and then seeks to find the splits with the best weighted mean Chi-square value over these subgroups appears to be a robust way of coarse classifying the other variables.

References

1] McNab H., Wynn A., Principles and Practice of Consumer Credit Risk Management, CIB Publishing, Canterbury, 2000

2] Thomas L.C., Edelman D.B., Crook J.N., *Credit Scoring and its Applications,* SIAM, Philadelphia, 2002.