En metode til kalibrering af diskriminantanalysens afskæringsværdi

Kai Kristensen

Institut for Statistik og Datalogi, Handelshøjskolen i Århus

SUMMARY: *In this article the application of Fisher's linear discriminant function for classification purposes is considered from a microeconomic point of view focusing on the determination of the cut-off point. The Decision-Maker's utility function (indifference function) specified in the conditional probabilities of correct classification is used as a starting point, whereupon the utility function is maximized subject to the relevant constraint on the conditional probabilities of correct classification. The results are adapted to practice through an approximation of the utility function by help of a Taylor expansion around a set of probabilities calculated from a preliminary estimate of the cut-off point given by the Decision-Maker.*

1. Indledning

Fisher's lineære diskriminantfunktion (Fisher (1936)) har i de senere år vundet relativt stor udbredelse inden for det driftsøkonomiske område ved løsning af taxonomiske problemstillinger. Se eksempelvis Blunch (1968), der indeholder en række afsætningsøkonomiske eksempler, Altman (1968) og Deakin (1972), der er eksempler på finansielle anvendelser og Welker (1974), der kan henføres til det organisatoriske område. Af andre økonomiske anvendelser kan nævnes Tintner (1946) og Adelman and Morris (1968).

I Anderson's version (Anderson (1951) og (1958)) har Fisher's to-gruppe diskriminantfunktion følgende udseende:

$$\begin{aligned} D(\mathbf{x}) &= ln(f_1(\mathbf{x})/f_2(\mathbf{x})) \\ &= \mathbf{x}'\Sigma^{-1}(\mu_1-\mu_2) - ½\,(\mu_1+\mu_2)'\,\Sigma^{-1}\,(\mu_1-\mu_2), \end{aligned} \tag{1}$$

hvor $\mathbf{x}$ er en (pxl) normalfordelt stokastisk vektor, $(\mu_1-\mu_2)$ er differencen mellem de to populationers forventningsvektorer og Σ er den fælles varians/kovariansmatrix.

Ved klassifikatoriske anvendelser heraf i forbindelse med en Bayes' strategi benyttes følgende klassifikationsregel for klassifikation i population 1:

En metode til kalibrering af diskriminantanalysens afskæringsværdi

Kai Kristensen

Institut for Statistik og Datalogi, Handelshøjskolen i Århus

SUMMARY: *In this article the application of Fisher's linear discriminant function for classification purposes is considered from a microeconomic point of view focusing on the determination of the cut-off point. The Decision-Maker's utility function (indifference function) specified in the conditional probabilities of correct classification is used as a starting point, whereupon the utility function is maximized subject to the relevant constraint on the conditional probabilities of correct classification. The results are adapted to practice through an approximation of the utility function by help of a Taylor expansion around a set of probabilities calculated from a preliminary estimate of the cut-off point given by the Decision-Maker.*

1. Indledning

Fisher's lineære diskriminantfunktion (Fisher (1936)) har i de senere år vundet relativt stor udbredelse inden for det driftsøkonomiske område ved løsning af taxonomiske problemstillinger. Se eksempelvis Blunch (1968), der indeholder en række afsætningsøkonomiske eksempler, Altman (1968) og Deakin (1972), der er eksempler på finansielle anvendelser og Welker (1974), der kan henføres til det organisatoriske område. Af andre økonomiske anvendelser kan nævnes Tintner (1946) og Adelman and Morris (1968).

I Anderson's version (Anderson (1951) og (1958)) har Fisher's to-gruppe diskriminantfunktion følgende udseende:

$$\begin{aligned} D(\mathbf{x}) &= ln(f_1(\mathbf{x})/f_2(\mathbf{x})) \\ &= \mathbf{x}'\Sigma^{-1}(\mu_1-\mu_2) - \tfrac{1}{2}(\mu_1+\mu_2)'\Sigma^{-1}(\mu_1-\mu_2), \end{aligned} \tag{1}$$

hvor $\mathbf{x}$ er en (pxl) normalfordelt stokastisk vektor, $(\mu_1\text{-}\mu_2)$ er differencen mellem de to populationers forventningsvektorer og Σ er den fælles varians/kovariansmatrix.

Ved klassifikatoriske anvendelser heraf i forbindelse med en Bayes' strategi benyttes følgende klassifikationsregel for klassifikation i population 1:

$$R_1 = \{\mathbf{x} | D(\mathbf{x}) \geq ln(C_1 q_2 / C_2 q_1)\}, \tag{2}$$

hvor q_i er apriori-sandsynligheden for population nr. i, og C_i er omkostningerne ved at klassificere et element fra population j i population i.

Et interessant og i praksis meget betydningsfuldt problem i forbindelse med applikationen af (1) og (2) er evalueringen af de respektive sandsynligheder for korrekt klassifikation. Disse er hhv.:

$$\beta_1 = Pr\ (D(\mathbf{x}) \geq c | 1) = 1 - F_1\ (c) \tag{3}$$

$$\beta_2 = Pr(D(\mathbf{x}) < c | 2) = F_2(c), \tag{4}$$

hvor F_i er diskriminantfunktionens fordelingsfunktion for den i'te population, og c er afskæringsværdien (cfr. højresiden af ulighedstegnet i (2)). Når såvel $D(\mathbf{x})$ som c er kendte, er problemet trivielt, eftersom $D(\mathbf{x})$ da er normalfordelt med parametrene:

$$E(D(\mathbf{x}) | 1) = \tfrac{1}{2}\Delta^2 \tag{5}$$

$$E(D(\mathbf{x}) | 2) = -\tfrac{1}{2}\Delta^2 \tag{6}$$

$$V(D(\mathbf{x}) | 1) = V(D(\mathbf{x}) | 2) = \Delta^2, \tag{7}$$

hvor $\Delta^2 = (\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 - \mu_2)$, dvs. Mahalanobis' generaliserede afstandsfunktion i de teoretiske parametre. I praksis er informationen imidlertid som oftest ufuldstændig, såvel hvad angår diskriminantfunktionen som afskæringsværdien. Dette har initieret en lang række studier på området, hvis fælles kendetegn dog er, at problemerne omkring $D(\mathbf{x})$ behandles indgående, hvorimod c som oftest tages for givet.

Den første behandling af problemet findes i Wald (1944), hvori fordelingen for en variant af (1) søges etableret, når parametrene er ukendte og derfor erstattes af de forventningsrette og konsistente estimater:

$$\hat{\mu}_1 = \bar{\mathbf{x}}_1 = N_1^{-1} \sum_{i=1}^{N_1} \mathbf{x}_{1i} \tag{8}$$

$$\hat{\mu}_2 = \bar{\mathbf{x}}_2 = N_2^{-1} \sum_{i=1}^{N_2} \mathbf{x}_{2i} \tag{9}$$

$$\hat{\Sigma} = \mathbf{S} = (N_1 + N_2 - 2)^{-1} \left(\sum_{i=1}^{N_1} (\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)(\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)' + \sum_{i=1}^{N_2} (\mathbf{x}_{2i} - \bar{\mathbf{x}}_2)(\mathbf{x}_{2i} - \bar{\mathbf{x}}_2)' \right) \quad (10)$$

hvor N_j er stikprøvens størrelse for den j'te population.

Senere er fordelingen for $\hat{D}(\mathbf{x}) = \mathbf{x}'\mathbf{S}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) - \frac{1}{2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$ gjort til genstand for intensive studier med den deraf følgende konklusion, at den nøjagtige fordeling er for kompliceret til at kunne finde anvendelse til numeriske formål, og at man derfor må støtte sig til asymptotiske resultater. Der kan blandt andre henvises til Anderson ((1951), (1973)), Sitgreaves ((1952), (1961)), Bowker and Sitgreaves (1961) og Okamoto ((1963), (1968)). Sidstnævnte bidrag er en generel asymptotisk udvikling for fordelingen for $\hat{D}(\mathbf{x})$ til og med led af ordenen N_1^{-2}, N_2^{-2}, $(N_1N_2)^{-1}$, $(N_1(N_1+N_2-2))^{-1}$, $(N_2(N_1+N_2-2))^{-1}$ og $(N_1+N_2-2)^{-2}$. Det bemærkes, at fordelingsproblemet løses automatisk, når N_1 og N_2 begge er store, eftersom estimaterne (8), (9) og (10) er konsistente. Fordelingen for $\hat{D}(\mathbf{x})$ konvergerer da mod fordelingen for $D(\mathbf{x})$.

Sideløbende med forsøgene på at etablere fordelingen for $\hat{D}(\mathbf{x})$ har man måttet erkende, at begrebet sandsynlighed for korrekt klassifikation ikke længere er entydigt, når parametrene estimeres. Hills (1966) har givet en stringent fremstilling af problemet formuleret i sandsynligheden for at fejlklassificere et individ fra population 1, α_1. Han fremhæver, at foruden den ukendte, sande sandsynlighed, $\alpha_1(D(\mathbf{x}))$, har følgende sandsynligheder interesse:

(1) $\alpha_1(\hat{D}(\mathbf{x})) = Pr(\hat{D}(\mathbf{x}) < c \,|\, 1, \bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \mathbf{S})$ dvs. den faktiske fejlklassifikationssandsynlighed.

(2) $E(\alpha_1(\hat{D}(\mathbf{x}))) = Pr(\hat{D}(\mathbf{x}) < c \,|\, 1)$ dvs. den gennemsnitlige faktiske fejlklassifikationssandsynlighed.

Hertil kommer, at også fordelingen for $\alpha_1(\hat{D}(\mathbf{x}))$ kan være af interesse.

Det indses umiddelbart, at $\alpha_1(\hat{D}(\mathbf{x}))$ er den relevante fejlklassifikationssandsynlighed, når $\hat{D}(\mathbf{x})$ allerede er valgt som klassifikationsinstrument, hvorimod fordelingen for $\alpha_1(\hat{D}(\mathbf{x}))$ og $E(\alpha_1(\hat{D}(\mathbf{x})))$ har interesse på et tidligere trin i planlægningen.

Estimation af de nævnte sandsynligheder er behandlet intensivt i litteraturen. For $\alpha_1(\hat{D}(\mathbf{x}))$ kan henvises til Fisher (1936), Smith (1947), Frank et al. (1965), Sorum (1971) og Joy and Tollefson (1975) og for $E(\alpha_1(\hat{D}(\mathbf{x})))$ til

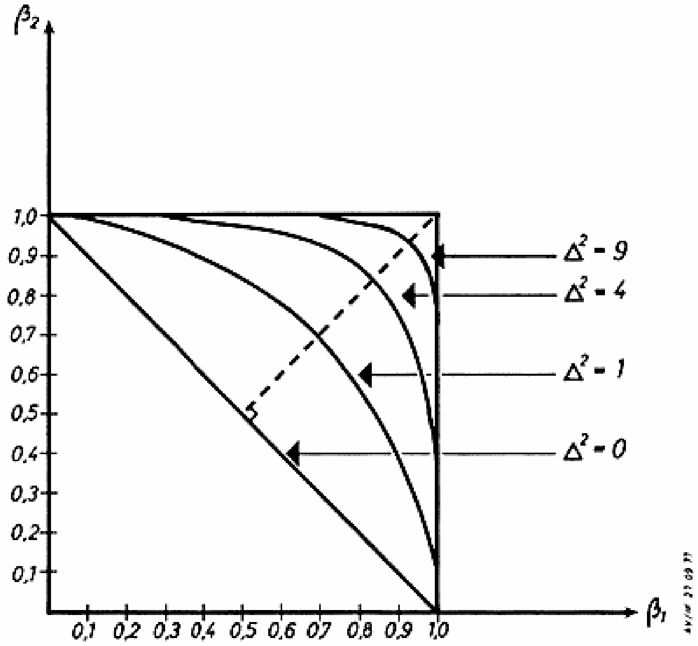
John (1961), Lachenbruch ((1965), (1967), (1968)), Lachenbruch and Mickey (1968), Hills (1966) og Dunn (1971). Endelig har John (1961) givet en asymptotisk udvikling af fordelingen for $\alpha_1(\hat{D}(\mathbf{x}))$. En generel bibliografi er publiceret af Toussaint (1974), ligesom Lachenbruch (1975) giver omfattende henvisninger.

Som tidligere anført tager disse studier afskæringsværdien for givet. Årsagerne hertil kan være mange. Den primære er formentlig, at det inden for mange af de traditionelle anvendelsesområder for diskriminantanalysen er rimeligt at antage $C_1 = C_2$, samtidig med at disse områder (f.eks. det biologiske og det antropologiske) lider under dataknaphed. I driftsøkonomiske analyser er forholdet som regel lige modsat. Således er datafremskaffelsen sjældent et problem, medens det til gengæld ikke vil være forsvarligt at antage, at forholdet mellem omkostningerne er 1. En anvendelse af diskriminantanalysen inden for dette område vil derfor kræve, at fokus vendes mod afskæringsværdien.

Det er den almindelige opfattelse (se f.eks. Lachenbruch (1975)), at det i praksis er meget svært at arbejde med omkostningerne som det direkte grundlag for fastsættelsen af afskæringsværdien, og at man derfor må angribe problemet fra andre sider. Der findes i litteraturen et enkelt forsøg herpå, hvori Anderson (1969) (set med en økonoms øjne) betragter problemet fra en satisficeringssynsvinkel, idet der diskrimineres under hensyntagen til beslutningstagerens satisficeringsniveauer for fejlklassifikationssandsynlighederne. Denne metode lider imidlertid af den svaghed, at den ikke sikrer klassifikation af samtlige elementer, hvorfor der i dette arbejde skal anvises en alternativ fremgangsmåde. Denne fremgangsmåde baseres på en egentlig optimeringstankegang, idet afskæringsværdiens fastsættelse betragtes som et mikroøkonomisk valghandlingsproblem.

2. Afskæringsværdien i mikroøkonomisk belysning

Det er forfatterens erfaring, at det i mangfoldige situationer er lettere at arbejde direkte på de betingede sandsynligheder for korrekt klassifikation, β_1 og β_2, end på omkostninger og apriori-sandsynligheder ved fastsættelse af en rimelig afskæringsværdi. Lad derfor udgangspunktet for det efterfølgende være beslutningstagerens nyttefunktion specificeret i β_1 og β_2, $U(\beta_1, \beta_2)$. Lad endvidere de konkrete parameterværdier være kendt eller estimeret på grundlag af så store stikprøver, at den forannævnte konvergens med rimelighed kan betragtes som opfyldt. Det fremgår nu, at



FIGUR 1. *Bibetingelser som funktion af Mahalanobis'* Δ^2 $(\beta_1 = 1 - \Phi\,(\Phi^{-1}(\beta_2) - \Delta))$.

$$\beta_1 = 1 - F_1(c) = 1 - \Phi((c - \tfrac{1}{2}\Delta^2)/\Delta) \tag{11}$$

$$\beta_2 = F_2(c) = \Phi((c + \tfrac{1}{2}\Delta^2)/\Delta), \tag{12}$$

hvor Φ er standardnormalfordelingsfunktionen. Dette indebærer, da $((c + \tfrac{1}{2}\Delta^2)/\Delta) - \Delta = (c - \tfrac{1}{2}\Delta^2)/\Delta$, at

$$\begin{aligned} \beta_1 &= 1 - F_1(F_2^{-1}(\beta_2)) \\ &= 1 - \Phi\,(\Phi^{-1}(\beta_2) - \Delta), \end{aligned} \tag{13}$$

der ses at være den relevante bibetingelse, hvorunder nyttefunktionen skal maksimeres med det formål at bestemme optimalkombinationen for de nævnte sandsynligheder og hermed den optimale afskæringsværdi. Det fremgår, at bibetingelsen er en funktion af Δ^2 med et forløb som skitseret i figur nr. 1. Lad herefter $M_i = \delta U(\beta_1, \beta_2)/\delta\beta_i$, dvs. den partielle grænsenytte for den i'te sandsynlighed for korrekt klassifikation. Det ønskede maksimum for nyttefunktionen kan da findes af ligningen:

$$\frac{dU(\beta_1, \beta_2)}{d\beta_2} = M_1 \frac{\delta\beta_1}{\delta\beta_2} + M_2 = 0, \tag{14}$$

der under hensyntagen til bibetingelsen (13) kan ændres til:

$$M_1 \left(- \frac{\delta F_1 (F_2^{-1}(\beta_2))}{\delta F_2^{-1}(\beta_2)} \frac{\delta F_2^{-1}(\beta_2)}{\delta\beta_2} \right) + M_2 = 0 \tag{15}$$

$$\Leftrightarrow M_1 \left(-f_1(c) \frac{1}{f_2(c)} \right) + M_2 = 0,$$

hvoraf fremgår, at den optimale afskæringsværdi skal findes af udtrykket:

$$f_1(c)M_1 = f_2(c)M_2 \tag{16}$$
$$\Leftrightarrow f_1(c)/f_2(c) = M_2/M_1,$$

hvor f_i ligesom i (15) angiver den normale tæthedsfunktion for den i'te population. Da $f_1(c) \propto exp(-\frac{1}{2}(c-\frac{1}{2}\Delta^2)^2/\Delta^2)$, og da $f_2(c) \propto exp(-\frac{1}{2}(c+\frac{1}{2}\Delta^2)^2/\Delta^2)$, fremgår det ved indsættelse, at den optimale afskæringsværdi optræder ved:

$$exp(c) = M_2/M_1$$
$$\Leftrightarrow c = ln(M_2/M_1). \tag{17}$$

Idet M_1/M_2 benævnes det marginale substitutionsforhold mellem de betingede sandsynligheder for korrekt klassifikation, kan optimalbetingelsen hermed formuleres som følger: *I optimumspunktet er afskæringsværdien lig med minus den naturlige logaritme til det marginale substitutionsforhold mellem de betingede sandsynligheder for korrekt klassifikation.*

Da $M_2/M_1 = -d\beta_1/d\beta_2$ for $dU(\beta_1,\beta_2) = 0$, fremgår det, at optimum også kan udtrykkes ved ligningen:

$$exp(c) = -E_{\beta_2}(\beta_1)\beta_1/\beta_2, \tag{18}$$

hvor $E_{\beta_2}(\beta_1) = (\delta ln\beta_1/\delta ln\beta_2) = (\delta\beta_1/\delta\beta_2)\ (\beta_2/\beta_1)$ angiver elasticiteten af β_1 med hensyn til β_2. Da imidlertid $\beta_1/\beta_2 \rightarrow 1$ for $\Delta^2 \rightarrow \infty$, fremgår det, at for store værdier af Δ^2 kan fastsættelsen af den optimale afskæringsværdi med rimelighed foretages udelukkende på grundlag af kendskab til elasticiteten i indifferenskurverne, hvilket indebærer visse praktiske fordele, eftersom denne ofte udviser en bemærkelsesværdig konstans. Formelt fås:

$$c = ln\,(-E\beta_2\,(\beta_1)) \text{ for } \Delta^2 \to \infty. \tag{19}$$

Når situationen $\Delta^2 \to \infty$ tages op til speciel overvejelse, er årsagen den, at gevinsten ved indførelse af en klassifikationsregel vil være positivt samvarierende med Δ^2. Da indførelsen endvidere fra en driftsøkonomisk synsvinkel repræsenterer et tilfælde af total tilpasning, må gevinsten foruden variable omkostninger tillige kunne dække implementeringsomkostningerne, der som oftest er af en ikke uvæsentlig størrelsesorden. Dette har som konsekvens, at man i det praktiske arbejde ofte konfronteres med relativt store værdier af Δ^2.

Den anvendte fremgangsmåde ved optimeringen svarer til en anvendelse af den velkendte kæderegel for differentiation på strukturen: $U(\beta_1,\beta_2)$ hvor $\beta_1 = g_1(c)$ og $\beta_2 = g_2(c)$. I den hensigt at opnå et resultat analogt til efterspørgselsteoriens Gossen-betingelse, der som bekendt siger, at den vejede grænsenytte i optimum skal være ens i alle anvendelser og lig med indkomstens grænsenytte, må denne struktur omformes til: $U(\beta_1,\beta_2)$ hvor $\beta_1{=}g_{11}(c,\Delta)$ og $\beta_2{=}g_{21}(c,\Delta)$, da Δ for nærværende afstikker grænser for valget af sandsynligheder, der er analoge til de grænser, indkomsten afstikker for konsumentens varevalg. Idet φ betegner standardnormaltætheden for den i'te population, fører differentiation heraf til følgende ligningssystem:

$$\begin{aligned}\delta U/\delta\Delta &= M_1(\delta\beta_1/\delta\Delta) + M_2(\delta\beta_2/\delta\Delta) \\ &= (M_1\varphi_1 - M_2\varphi_2)\,((c+\tfrac{1}{2}\Delta^2)/\Delta^2) + M_2\varphi_2\end{aligned} \tag{20}$$

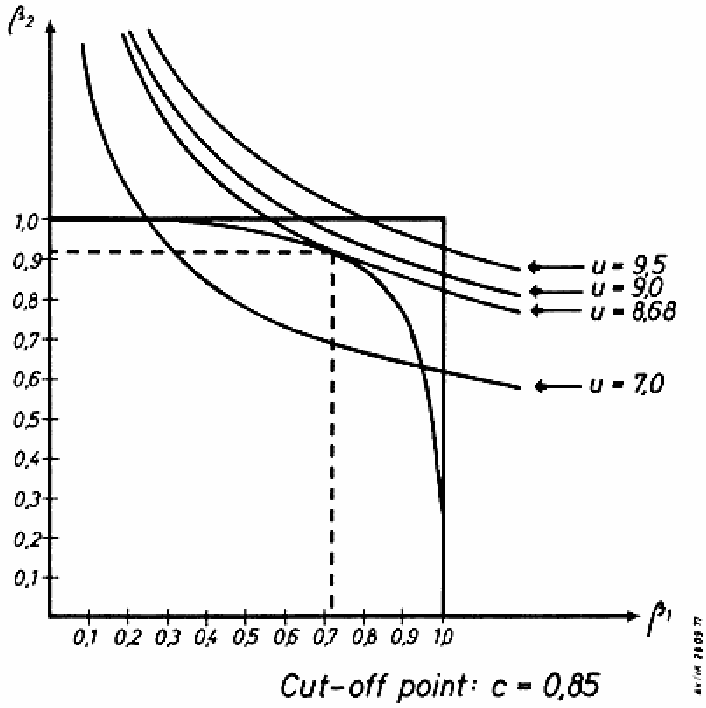
$$\delta U/\delta c = (M_2\varphi_2 - M_1\varphi_1)\Delta^{-1}. \tag{21}$$

Sættes (21) lig med 0 følger, at (20) kan gives nedenstående fremstilling:

$$\delta U/\delta\Delta = M_2\varphi_2 = M_1\varphi_1, \tag{22}$$

der gør det muligt at formulere følgende til Gossen-betingelsen analoge optimalbetingelse: *Den optimale afskæringsværdi findes, hvor de med standardnormaltæthederne vejede grænsenytter af de betingede sandsynligheder for korrekt klassifikation er ens og lig med grænsenytten af Mahalanobis' Δ.*

Alternativt kunne vi have nået dette resultat ved følgende Lagrange-formulering af vort optimeringsproblem: Maksimér $U(\beta_1,\beta_2)$ når $\Phi^{-1}(\beta_2) - \Phi^{-1}(1-\beta_1) = \Delta$. Resultatet heraf er $M_1\varphi_1 = M_2\varphi_2 = \lambda$, hvor λ er Lagrange-multiplikatoren. Men sidstnævnte er netop lig med den ændring, man opnår i kriteriefunktionen, når konstantleddet i bibetingelsen ændres infinitesimalt (Johansen (1962)), hvoraf resultatet følger.



FIGUR 2. *Bestemmelse af optimalkombination for en Cobb-Douglas nyttefunktion med* $A = 10$, $a = 0{,}25$ *og* $b = 0{,}75$. $\Delta^2 = 4$.

Som beregningseksempler antages i det efterfølgende, at beslutningstagerens nyttefunktion er af Cobb-Douglas resp. CES-typen. Dvs.:

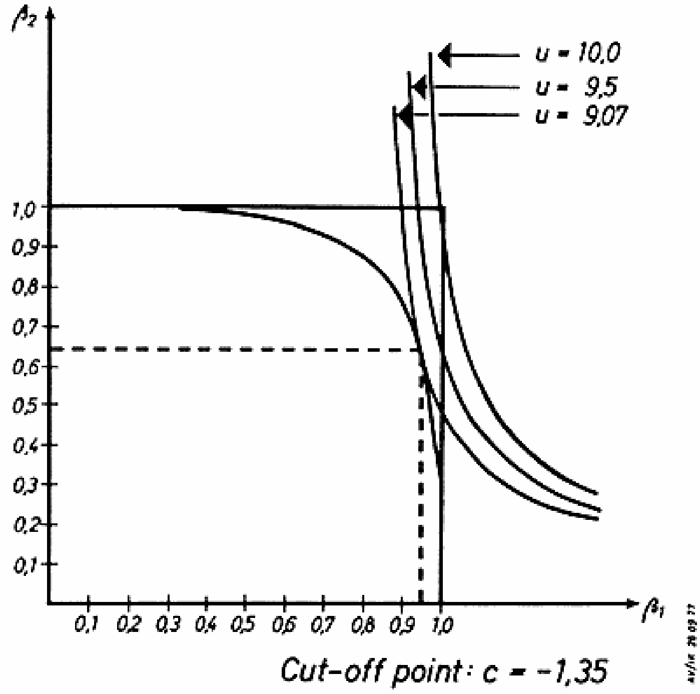
$$U_1(\beta_1, \beta_2) = A\beta_1{}^a\beta_2{}^b \tag{23}$$

$$U_2(\beta_1, \beta_2) = B(C\beta_1{}^{-\varrho} + D\beta_2{}^{-\varrho})^{-(1/\varrho)}, \tag{24}$$

hvor $A>0$, $0<a<1$, $0<b<1$, $B>0$, $C>0$, $D>0$ og $-1<\varrho<0$ eller $\varrho>0$.

Disse funktionstyper er valgt som illustration, fordi begge besidder egenskaber, der ofte antages at være til stede hos nyttefunktioner. Begge er således kvasikonkave og har dermed konvekse niveaumængder (se fig. 2 og 3). Endvidere er begge homogene; Cobb-Douglas funktionen af graden a+b og CES funktionen af graden 1. Endelig gælder der for Cobb-Douglas funktionen, at elasticiteten i dens niveauflader (indifferenskurver) er konstant (jfr. tidligere).

For Cobb-Douglas funktionen ((23)) er $d\beta_1/d\beta_2 = -(b/a)\ (\beta_1/\beta_2)$, og optimalpunktet bestemmes følgelig af:



FIGUR 3. *Bestemmelse af optimalkombination for en CES nyttefunktion med* $B = 10$, $C = 0{,}9$ og $D = 0{,}1$. $\Delta^2 = 4$.

$$c = ln((b/a)\ (\beta_1/\beta_2)), \tag{25}$$

som løses ved hjælp af iteration. Bemærk i øvrigt, at $E_{\beta_2}(\beta_1) = -(b/a)$, hvorfor optimalpunktet for store værdier af Δ^2 kan beregnes direkte:

$$c = ln(b/a) \text{ for } \Delta^2 \to \infty. \tag{26}$$

I figur nr. 2 er beregningerne udført for $A = 10$, $a = .25$, $b = .75$ og $\Delta^2 = 4$. Resultatet er, at den optimale værdi af c er .85, hvortil svarer $\beta_1 = .719$ og $\beta_2 = .924$. U-værdien er 8.68.

For CES funktionen bestemmes optimalpunktet af:

$$c = ln((D/C)\ (\beta_1/\beta_2)^{(1+\varrho)}), \tag{27}$$

som ligeledes løses ved iteration. I figur nr. 3 er beregningerne udført for $B = 10$, $C = .90$, $D = .10$, $\varrho = 1$ og $\Delta^2 = 4$. Det fremgår, at den optimale c-værdi er -1.35 hvortil svarer $\beta_1 = .953$ og $\beta_2 = .627$. U-værdien er 9.07.

3. Fastsættelse af afskæringsværdien i praksis

I den praktiske situation, hvor beslutningstagerens nyttefunktion er ukendt, må approximative metoder tages i anvendelse. Rent konkret anmodes beslutningstageren om at give et indledende skøn over afskæringsværdien. Lad dette skøn være c_0. Hertil svarer følgende sandsynligheder for korrekt klassifikation:

$$\beta_1^* = \int_{c_0}^{\infty} f_1 (D(\mathbf{x}))\, dD(\mathbf{x}) \tag{28}$$

$$\beta_2^* = \int_{-\infty}^{c_0} f_2 (D(\mathbf{x}))\, dD(\mathbf{x}) \tag{29}$$

I det ved β_1^* og β_2^* dannede punkt Taylor-udvikles nyttefunktionen, $U(\beta_1,\beta_2)$, og erstattes af første ordens leddene. Dvs. at tangentplanet i punktet (β_1^*,β_2^*) benyttes som substitut for den sande nyttefunktion. Idet $U_t(\beta_1,\beta_2)$ betegner tangentplanet fås:

$$\begin{aligned} U_t(\beta_1,\beta_2) &= U(\beta_1^*,\beta_2^*) + (\beta_1-\beta_1^*)\,[M_1]_{(\beta_1^*,\beta_2^*)} + \\ &(\beta_2-\beta_2^*)\,[M_2]_{(\beta_1^*,\beta_2^*)}. \end{aligned} \tag{30}$$

Ved omformning fås heraf:

$$\begin{aligned} &(U_t(\beta_1,\beta_2)-U(\beta_1^*,\beta_2^*) + \beta_1^*[M_1]_{(\beta_1^*,\beta_2^*)} + \beta_2^*[M_2]_{(\beta_1^*,\beta_2^*)})/[M_2]_{(\beta_1^*,\beta_2^*)} \\ &\quad = \beta_1[M_1/M_2]_{(\beta_1^*,\beta_2^*)} + \beta_2. \end{aligned} \tag{31}$$

Tangentplanet kan naturligvis måles i nye enheder uden tab af generalitet, og der arbejdes derfor videre på højresiden af (31):

$$\begin{aligned} U_t^*(\beta_1,\beta_2) &= \beta_1[M_1/M_2]_{(\beta_1^*,\beta_2^*)} + \beta_2 \\ &= H\beta_1 + \beta_2, \end{aligned} \tag{32}$$

hvor $H = [M_1/M_2]_{(\beta_1^*,\beta_2^*)}$ dvs. det marginale substitutionsforhold i det ved beslutningstageren givne punkt. Minimering heraf under bibetingelsen (13) fører til resultatet:

$$c = -\,ln(H), \tag{33}$$

hvilket kan indses umiddelbart af resultaterne i afsnit 2 (specielt (17)). Imidlertid er H ukendt og må erstattes af et skøn. Dette skøn opnås ved at anmode beslutningstageren om at angive en ny kombination af sandsynligheder for korrekt klassifikation, som efter hans opfattelse *er beliggende på samme isonyttekurve (indifferenskurve) som den ved c_0 givne kombination.* Lad denne kombination være $(\beta_1^{**}, \beta_2^{**})$, hvorefter skønnet over H, $\hat{H}$, kan beregnes til:

$$\hat{H} = -(\beta_2^{*} - \beta_2^{**}) \,/\, (\beta_1^{*} - \beta_1^{**}). \tag{34}$$

Den praktiske bestemmelse af afskæringsværdien finder således sted ved anvendelse af:

$$c = -\ln(\hat{H}). \tag{35}$$

Hvis det opnåede resultat ligger fjernt fra c_0, kan der rejses tvivl om approximationens validitet, og proceduren bør derfor gentages. Principielt bør proceduren gennemløbes, til løsningen er stabil, og det indses derfor, at metoden alternativt kan betragtes som et kalibreringsinstrument for en traditionel Bayes-løsning.

I Altman (1968) anvendes diskriminantanalyse til at skelne mellem konkursramte (P_2) og ikke-konkursramte (P_1) virksomheder på basis af følgende nøgletal:

$$\begin{aligned}
&X_1\text{: Working capital/aktiver (\%)}\\
&X_2\text{: Tilbageholdt indtjening/aktiver (\%)}\\
&X_3\text{: Indtjening før renter og skat/aktiver (\%)}\\
&X_4\text{: Egenkapitalens markedsværdi/gæld (\%)}\\
&X_5\text{: Omsætning/aktiver.}
\end{aligned} \tag{36}$$

Resultatet er følgende diskriminantfunktion:

$$D(\mathbf{x}) = -2.365 + .012X_1 + .014X_2 + .033X_3 + .006X_4 + .999X_5, \tag{37}$$

med $\Delta^2 = 5.31$, idet der ses bort fra konvergensproblemerne.

Antag som et eksempel på den ovenfor anførte metodes anvendelse, at en beslutningstager vil benytte (37) i forbindelse med sin kreditgivning, idet han som udgangspunkt benytter sig af den traditionelle afskæringsværdi på 0.

I den hensigt at kalibrere afskæringsværdien beregnes den til c_0 hørende kombination af sandsynligheder for korrekt klassifikation til:

$$(\beta_1^*, \beta_2^*) = (.875, .875). \tag{38}$$

Beslutningstageren oplyser, at han for at være indifferent forlanger, at et fald i β_2 på .05 skal kompenseres med en stigning i β_1 på .08. I nytteteoretiske termer følger heraf, at $(.875, .875) \sim (.955, .825) = (\beta_1^{**}, \beta_2^{**})$. Skønnet over parameteren H kan derefter beregnes til:

$$\begin{aligned}\hat{H} &= -(.875 - .825) \,/\, (.875 - .955) \\ &= .05/.08 = .625\end{aligned} \tag{39}$$

jf. (34). Af (35) følger da:

$$\begin{aligned}c &= -\ln(.625) \\ &= .47,\end{aligned} \tag{40}$$

hvorefter de faktiske sandsynligheder for korrekt klassifikation kan beregnes til $\beta_1 = .829$ og $\beta_2 = .915$.

Kreditgiverens beslutningsregel bliver herefter, at der ydes kredit, når $D(\mathbf{x})$ er større end eller lig .47, og at der gives afslag, når $D(\mathbf{x})$ er mindre end .47.

Litteratur

ADELMAN, I. and MORRIS, C. T. 1968. Performance Criteria for Evaluating Economic Development Potential: An Operational Approach. *Quarterly Journal of Economics*, vol. LXXXII, pp. 260–280.

ALTMAN, E. 1968. Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy. *Journal of Finance*, pp. 589–610.

ANDERSON, J. A. 1969. Discrimination between k Populations with Constraints on the Probabilities of Misclassification. *Journal of the Royal Statistical Society*, B31, p. 123.

ANDERSON, T. W. 1951. Classification by Multivariate Analysis. *Psychometrika*, vol. 16, pp. 31–50.

ANDERSON, T. W. 1958. An Introduction to Multivariate Statistical Analysis. New York.

I den hensigt at kalibrere afskæringsværdien beregnes den til c_0 hørende kombination af sandsynligheder for korrekt klassifikation til:

$$(\beta_1^*, \beta_2^*) = (.875, .875). \tag{38}$$

Beslutningstageren oplyser, at han for at være indifferent forlanger, at et fald i β_2 på .05 skal kompenseres med en stigning i β_1 på .08. I nytteteoretiske termer følger heraf, at $(.875, .875) \sim (.955, .825) = (\beta_1^{**}, \beta_2^{**})$. Skønnet over parameteren H kan derefter beregnes til:

$$\begin{aligned} \hat{H} &= -(.875 - .825) \,/\, (.875 - .955) \\ &= .05/.08 = .625 \end{aligned} \tag{39}$$

jf. (34). Af (35) følger da:

$$\begin{aligned} c &= -ln(.625) \\ &= .47, \end{aligned} \tag{40}$$

hvorefter de faktiske sandsynligheder for korrekt klassifikation kan beregnes til $\beta_1 = .829$ og $\beta_2 = .915$.

Kreditgiverens beslutningsregel bliver herefter, at der ydes kredit, når $D(\mathbf{x})$ er større end eller lig .47, og at der gives afslag, når $D(\mathbf{x})$ er mindre end .47.

Litteratur

ADELMAN, I. and MORRIS, C. T. 1968. Performance Criteria for Evaluating Economic Development Potential: An Operational Approach. *Quarterly Journal of Economics,* vol. LXXXII, pp. 260–280.

ALTMAN, E. 1968. Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy. *Journal of Finance,* pp. 589–610.

ANDERSON, J. A. 1969. Discrimination between k Populations with Constraints on the Probabilities of Misclassification. *Journal of the Royal Statistical Society,* B31, p. 123.

ANDERSON, T. W. 1951. Classification by Multivariate Analysis. *Psychometrika,* vol. 16, pp. 31–50.

ANDERSON, T. W. 1958. An Introduction to Multivariate Statistical Analysis. New York.

ANDERSON, T. W. 1973. Asymptotic Evaluation of the Probabilities of Misclassification by Linear Discriminant Functions. I T. Cacoullos, ed. (1973), pp. 17–35.

ANDERSON, T. W. 1973. An Asymptotic Expansion of the Distribution of the »Studentized« Classification Statistic W. *Annals of Statistics*, 1, pp. 964–972.

BLUNCH, N. 1968. En metode til klassificering af kundeemner. *Erhvervsøkonomisk Tidsskrift*, nr. 2, pp. 81–95.

BOWKER, A. H. and SITGREAVES, R. 1961. An Asymptotic Expansion of the Distribution Function of the W-classification Statistic. I H. Solomon, ed. (1961), pp. 293–310.

CACOULLOS, T., ed. 1973. Discriminant Analysis and Applications. New York.

DEAKIN, E. B. 1972. A Discriminant Analysis of Predictors of Business Failure. *Journal of Accounting Research*, pp. 167–179.

DUNN, O. J. 1971. Some Expected Values for Probabilities of Correct Classification in Discriminant Analysis. *Technometrics*, 13, p. 345.

FISHER, R. A. 1936. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 7, pp. 179–188.

FRANK, R. E., MASSY, W. F. and MORRISON, D. G. 1965. Bias in Multiple Discriminant Analysis. *Journal of Marketing Research*, 2, pp. 250–258.

HILLS, M. 1966. Allocation Rules and their Error Rates. *Journal of the Royal Statistical Society*, B28, pp. 1–31.

JOHANSEN, L. 1962. *Notat om tolkningen af Lagrange-multiplikatorer*. Memo fra Socialøkonomisk Institut. Oslo.

JOHN, S. 1961. Errors in Discrimination. *Annals of Mathematical Statistics*, 32, pp. 1125–1144.

JOY, O. M. and TOLLEFSON, J. O. 1975. On the Financial Applications of Discriminant Analysis. *Journal of Financial and Quantitative Analysis*, pp. 723–739.

LACHENBRUCH, P. A. 1965. *Estimation of Error Rates in Discriminant Analysis*. Ph. D. dissertation. University of California at Los Angeles.

LACHENBRUCH, P. A. 1967. An Almost Unbiased Method of Obtaining Confidence Intervals for the Probability of Misclassification in Discriminant Analysis. *Biometrics*, 23, pp. 639–645.

LACHENBRUCH, P. A. 1968. On Expected Values of Probabilities of Misclassification in Discriminant Analysis, Necessary Sample Size, and a Relation with the Multiple Correlation Coefficient. *Biometrics*, 24, p. 823.

LACHENBRUCH, P. A. 1975. Discriminant Analysis. New York.

LACHENBRUCH, P. A. and MICKEY, M. R. 1968. Estimation of Error Rates in Discriminant Analysis. *Technometrics*, 10, p. 1.

OKAMOTO, M. 1963. An Asymptotic Expansion for the Distribution of the Linear Discriminant Function. *Annals of Mathematical Statistics*, 34, pp. 1286–1301.

OKAMOTO, M. 1968. Correction to: An Asymptotic Expansion for the Distribution of the Linear Discriminant Function. *Annals of Mathematical Statistics*, 39, pp. 1358–1359.

SITGREAVES, R. 1952. On the Distribution of Two Random Matrices used in Classification Procedures. *Annals of Mathematical Statistics*, 23, pp. 263–270.

SITGREAVES, R. 1961. Some Results on the Distribution of the W-classification Statistic. I H. Solomon, ed. (1961), pp. 241–251.

SMITH, C. A. B. 1947. Some Examples of Discrimination. *Annals of Eugenics,* 18, pp. 272–283.

SOLOMON, H., ed. 1961. *Studies in Item Analysis and Prediction.* Stanford University Press.

SORUM, M. 1971. Estimating the Conditional Probability of Misclassification. *Technometrics,* 13, p. 333.

TINTNER, G. 1946. Some Applications of Multivariate Analysis to Economic Data. *Journal of the American Statistical Association,* vol. 41, pp. 472–500.

TOUSSAINT, G. T. 1974. Bibliography on Estimation of Misclassification. *IEEE Transactions on Information Theory,* vol. IT-20, pp. 472–479.

WALD, A. 1944. On a Statistical Problem Arising in the Classification of an Individual in One of Two Groups. *Annals of Mathematical Statistics,* 15, pp. 145–162.

WELKER, R. B. 1974. Discriminant Analysis as an Aid to Employee Selection. *The Accounting Review,* pp. 514–523.