En metode til klassificering af kundeemner.

Af NIELS BLUNCH*)

1. Klassificeringsproblemet. – 2. Definition af diskriminantfunktionen. – 3. Hvad er kravet til en klassificeringsregel? – 4. Klassificering af ét element i en af to mulige populationer med vilkårlige, kendte tæthedsfunktioner. – 5. Den multinormale fordeling. – 6. Klassificering i en af to mulige multinormale fordeling med kendte parametre. – 7. Klassificering i en af to mulige multinormale fordelinger, når parametrene skal estimeres. – 8. Anvendelser af diskriminantanalyse inden for afsætningsøkonomien.

1. Klassificeringsproblemet

Enhver, der beskæftiger sig med afsætningsøkonomiske problemstillinger, det være sig som praktiker eller teoretiker, vil med korte mellemrum stå over for spørgsmålet: Hvorledes skelner jeg mellem forskellige grupper af efterspørgere? En besvarelse af dette spørgsmål vil nemlig være forudsætningen for en rationel løsning af en lang række afsætningsøkonomiske beslutningsproblemer.

Som eksempler på beslutningssituationer, hvor dette klassificeringsproblem opstår, kan nævnes følgende tre eksempler:
Eksempel 1. Ved salg af kostbare forbrugsvarer (f. eks. swimming pools) eller produktionsmidler (f. eks. maskiner), hvor salgsindsatsen i stort omfang er af personlig art, er det værd på forhånd – d. v. s. før man indleder salgsarbejdet over for den enkelte kundemulighed – at kunne bedømme sandsynligheden for, at netop denne kundemulighed vil optræde som *effektiv* efterspørger af varen, idet dette vil være afgørende for, om det med *de store omkostninger, der er forbundet med den enkelte salgsindsats,* vil kunne betale sig at foretage henvendelse til netop denne kundemulighed. Dette må foruden af købssandsynligheden afhænge af omkostningerne ved salgsindsatsen og af varens dækningsbidrag.

*) cand. merc., amanuensis ved Institut for Markedsøkonomi, Handelshøjskolen i Århus.

En metode til klassificering af kundeemner.

Af Niels Blunch*)

1. Klassificeringsproblemet. – 2. Definition af diskriminantfunktionen. – 3. Hvad er kravet til en klassificeringsregel? – 4. Klassificering af ét element i en af to mulige populationer med vilkårlige, kendte tæthedsfunktioner. – 5. Den multinormale fordeling. – 6. Klassificering i en af to mulige multinormale fordeling med kendte parametre. – 7. Klassificering i en af to mulige multinormale fordelinger, når parametrene skal estimeres. – 8. Anvendelser af diskriminantanalyse inden for afsætningsøkonomien.

1. Klassificeringsproblemet

Enhver, der beskæftiger sig med afsætningsøkonomiske problemstillinger, det være sig som praktiker eller teoretiker, vil med korte mellemrum stå over for spørgsmålet: Hvorledes skelner jeg mellem forskellige grupper af efterspørgere? En besvarelse af dette spørgsmål vil nemlig være forudsætningen for en rationel løsning af en lang række afsætningsøkonomiske beslutningsproblemer.

Som eksempler på beslutningssituationer, hvor dette klassificeringsproblem opstår, kan nævnes følgende tre eksempler:

Eksempel 1. Ved salg af kostbare forbrugsvarer (f. eks. swimming pools) eller produktionsmidler (f. eks. maskiner), hvor salgsindsatsen i stort omfang er af personlig art, er det værd på forhånd – d. v. s. før man indleder salgsarbejdet over for den enkelte kundemulighed – at kunne bedømme sandsynligheden for, at netop denne kundemulighed vil optræde som *effektiv* efterspørger af varen, idet dette vil være afgørende for, om det med *de store omkostninger, der er forbundet med den enkelte salgsindsats,* vil kunne betale sig at foretage henvendelse til netop denne kundemulighed. Dette må foruden af købssandsynligheden afhænge af omkostningerne ved salgsindsatsen og af varens dækningsbidrag.

*) cand. merc., amanuensis ved Institut for Markedsøkonomi, Handelshøjskolen i Århus.

Den almindelige fremgangsmåde i sådanne tilfælde er, at købssandsynligheden antages at være en funktion af en række egenskaber ved kundemuligheden f. eks. i swimming pool-tilfældet hans indtægt, hans sociale stilling, hans haves størrelse, etc. Man søger da på grundlag af en bedømmelse af en række af sådanne sociale og økonomiske variable at klassificere kundemuligheden som effektiv efterspørger eller ikke effektiv efterspørger.
Eksempel 2. Når et forsikringsselskab modtager en forsikringsbegæring, søger man på grundlag af de i begæringen givne oplysninger at placere forsikringstageren i én af flere risikoklasser som grundlag for præmieberegningen, idet man hverken ønsker, at forsikringstageren kommer til at betale for lille præmie i forhold til risikoen, hvilket vil være en dårlig forretning for selskabet, eller at han skal betale så stor en præmie, at han forsikrer i et andet selskab eller undlader at forsikre.
Eksempel 3. Når en reklamekampagne eller anden salgsfremmende indsats for en mærkevare skal planlægges, er det af væsentlig betydning for såvel udformningen af budskabet som for mediavalget, at man er i stand til på grundlag af socio-økonomiske variable at skelne mellem grupper af efterspørgere. Eksempler på sådanne grupperinger er

a) Købere af en vare opdelt efter det oftest købte varemærke.
b) Loyale kontra mindre loyale købere af et varemærke.
c) Købere af en vare opdelt efter den type forretning, hvori den enkelte køber fortrinsvis foretager sine indkøb.

Det fælles træk ved disse og utallige andre problemstillinger af lignende art, der forekommer inden for afsætningsøkonomien, er, at man er interesseret i på grundlag af måling af en række *variable* (f. eks. indtægt, alder etc.), der er tilknyttet et *element* (f. eks. en kundemulighed), at klassificere dette element som tilhørende *én* af flere i forvejen definerede *populationer*. Skal det have nogen mening at benytte sådanne målinger på det enkelte element som grundlag for en klassificering, må man naturligvis have gjort sig nogle tanker om, hvilke måleresultater man med rimelighed kan forvente, hvis elementet tilhører en bestemt population. Dette kan lidt mere formelt udtrykkes derved, at sandsynligheden – i form af en tæthedsfunktion – for fremkomsten af ethvert muligt sæt af måleresultater på det enkelte element, betinget af at elementet stammer fra populationen π_r, er kendt for alle r. Foretages der måling af p forskellige variable ved hvert element, er de til hvert univers knyttede tæthedsfunktioner altså p-dimensionale, medens den foretagne observation kan betragtes som en søjlevektor i p dimensioner:

$$X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ x_p \end{pmatrix} \tag{1}$$

Store bogstaver betegner her som overalt i det følgende vektorer eller matricer.

Vi betragter altså elementet karakteriseret ved vektoren X som en – i statistisk forstand – tilfældig observation fra én af de givne populationer. Spørgsmålet er da: Når en vektor X er observeret, fra hvilken population stammer da det pågældende element?

Inden for en videnskab som antropologien har man ganske tilsvarende problemer, idet man dér ønsker på grundlag af måling af fysiske egenskaber ved et individ at placere det racemæssigt. Allerede i 1920'erne begyndte man at angribe dette problem med det statistiske værktøj og en lang udvikling har ført til fremkomsten af en speciel teknik *multipel diskriminant-analyse,* der er anvendelig ved løsningen af dette problem.

Siden da har teknikken efterhånden fundet anvendelse inden for den psykologiske og sociologiske forskning, medens den har været ret ukendt uden for disse ret snævre fagområder. Der er imidlertid ingen grund til at den multiple diskriminantanalyse ikke også skulle kunne finde anvendelse inden for de mange andre områder, hvor klassificeringsspørgsmålet spiller en lignende rolle.

Bemærk, at antallet og arten af populationer er fastlagt i forvejen. Det er ikke hensigten at finde de mest hensigtsmæssige klassificeringskriterier, men udelukkende at finde frem til den mest hensigtsmæssige klassificering af et eller flere elementer i en af flere *i forvejen* definerede populationer.

2. *Definition af diskriminantfunktionen*

Antag, at et element enten kan tilhøre populationen π_1 eller populationen π_2. Klassificeringen afhænger af den observerede vektor X. Vi ønsker at opstille en regel, således at elementer, der er karakteriseret ved en bestemt mængde vektorer X, vil blive klassificeret som tilhørende π_1 og ellers som tilhørende π_2.

Vi kan forestille os de enkelte elementer afbildet som punkter i et p-

dimensionalt rum. Observationerne falder da mere eller mindre som to punktklynger – eller som to p-dimensionale tæthedsfunktioner i det kontinuerte tilfælde, som vi herefter vil holde os til.

Da de to tæthedsfunktioner (eller punktklynger) imidlertid er overlappende (i modsat tilfælde ville klassificeringen jo give sig selv, når X er konstateret), ønsker vi at trække en *grænse* i det p-dimensionale rum, eller – mere nøjagtigt udtrykt – vi ønsker at opdele rummet i to *områder* således, at hvis en observation falder i området R_1, klassificeres den som tilhørende π_1, og falder den i R_2, klassificeres den som tilhørende π_2. Da tæthedsfunktionerne som sagt er overlappende, er det umuligt at undgå, at nogle observationer vil blive anbragt på den »gale« side af grænsen, men vort formål er at finde en procedure for valg af R_1 og R_2, der er mest mulig »hensigtsemæssig«.

Vor klassificeringsprocedure vil altså få form af en *funktion* af den observerede vektor X, således at denne funktion antager vidt forskellige værdier alt efter, om det pågældende element er placeret i R_1 eller R_2.

En sådan funktion kaldes en *diskriminantfunktion*. Vi vil her for en nemheds skyld kun beskæftige os med *lineære diskriminantfunktioner*, men andre funktionstyper kan naturligvis også tænkes.

Diskriminantfunktionen vil kunne skrives:

$$D = a_1x_1 + a_2x_2 + \ldots + a_ix_i + \ldots + a_px_p \qquad (2)$$

Proceduren vil da være den, at de på et element foretagne målinger indsættes i udtrykket (2).

Hvis $D \geqslant b$ klassificeres elementet som tilhørende π_1.
Hvis $D < b$ » » » » π_2.

Vor fastlæggelse af områderne R_1 og R_2 og dermed vor klassificeringsprocedure vil altså afhænge af »vægtene« $a_1, a_2 \ldots a_i \ldots a_p$ samt af konstanten b. Det gælder altså om at fastlægge disse størrelser »hensigtsmæssigt«.

Bemærk, at ligningen $D = b$ fremstiller den plan, der deler det p-dimensionale rum i de to halvrum R_1 og R_2.

Vi har i dette afsnit skitseret problemløsningen i tilfældet med kun to populationer og vil også i det følgende begrænse os hertil. En generalisering til r populationer ($r > 2$) medfører ikke principielt nye problemer, kun mere regnearbejde, idet vi må arbejde med flere diskriminantfunktioner.[1])

[1]) Der henvises til (1) i litteraturlisten.

3. Hvad er kravet til en klassificeringsregel?

I almindelighed (d. v. s. i ortodoks »objektiv« statistisk teori) plejer man at udforme klasificeringsproceduren (diskriminantfunktionen) på en sådan måde, at de betingede sandsynligheder for de to typer af fejlplaceringer skal være så små som mulige og lige store.

Kaldes de to tæthedsfunktioner $f_1(X)$ og $f_2(X)$, kan dette lidt mere formelt formuleres:

$$\text{minimér} \quad \int_{R_2} f_1(X)dX \tag{3}$$

under bibetingelsen

$$\int_{R_2} f_1(X)dX = \int_{R_1} f_2(X)dX \tag{4}$$

Det er imidlertid ikke vanskeligt at forestille sig situationer, i hvilke ovennævnte kriterium kan føre til uhensigtsmæssige resultater. F. eks. medfører denne procedure, at det forventede antal fejlklassificeringer ved klassificering af et bestemt antal elementer, f. eks. 10, er størst for de elementer, der tilhører den største population. Ligeledes kan man også let forestille sig situationer, hvor ulemperne ved den ene art fejlklassificeringer er betydelig større end ved den anden.

Disse bemærkninger afslører, at det kan være hensigtsmæssigt, at klassificeringsreglen (= diskriminantfunktionen) foruden at være en funktion af vektoren X, også afhænger af:

1) *a priori-sandsynlighederne for, at et element skal tilhøre hver af de to populationer.* Hvis der a priori – d. v. s. før målingerne foretages – er større sandsynlighed for, at et vilkårligt element tilhører den ene population frem for den anden, vil man (alt andet lige) være mest tilbøjelig til at klassificere elementet som tilhørende den population, der har størst a priori-sandsynlighed.

2) *forholdet mellem offeromkostningerne ved de to former for fejlklassificering.* Idet man (alt andet lige) vil være mest tilbøjelig til at klassificere elementet i den population, hvor omkostningerne, dersom det senere skulle vise sig, at klassificeringen er forkert, er mindst.

Set fra et *beslutningsteoretisk* synspunkt må tankegangen derfor være følgende:

Hvis det har nogen mening at klassificere et element som tilhørende en af to (eller evt. flere) populationer, må det være, fordi der vil blive truffet forskellige *beslutninger* m. h. t. elementets »behandling«, alt efter om det klassificeres som tilhørende π_1 eller π_2. Ved en fejlklassificering vil man derfor udsætte elementet for en forkert »behandling«, d. v. s. en behandling, der er forskellig fra den under de givne forhold optimale. Herved pådrager man sig nogle offeromkostninger (større omkostninger eller mindre indtægter), som ville være undgået, hvis den optimale behandling var foretaget.

I tilfældet med potentiel efterspørger/ikke potentiel efterspørger vil en fejlklassificering kunne medføre enten

1) at der indledes salgsarbejde over for en konsument, som ikke vil være potentiel efterspørger, hvorved man pådrager sig forgæves salgsomkostninger,

eller

2) at man undlader at foretage salgsarbejde over for en potentiel efterspørger og derved går glip af et muligt dækningsbidrag.

I dette tilfælde vil det ikke være forbundet med større vanskeligheder at bestemme de af en fejlklassificering betingede offeromkostninger, medens det i andre tilfælde kan være et særdeles vanskeligt problem at formulere disse omkostninger i kr. Dette er heller ikke nødvendigt. Da – som vi senere skal bevise – det kun er forholdet mellem de to arter af omkostninger, der har betydning, kræves det kun, at de er ensbenævnte udtryk for graden af »uønskethed« af forkerte beslutninger som følge af forkerte klassificeringer.

Betegner $P(\pi_1)$ og $P(\pi_2)$ a priori-sandsynlighederne for, at et tilfældigt element skal tilhøre π_1 henholdsvis π_2, og benyttes betegnelserne $C(1|2)$ og $C(2|1)$ for offeromkostningerne ved at klassificere et element tilhørende π_2 som tilhørende π_1 henholdsvis offeromkostningerne ved at klassificere et element, der tilhører π_1 som tilhørende π_2, kan problemets *offeromkostnings-matrix* skitseres som følger:

Tabel 1
Offeromkostnings-matrix

Beslutning \ Population	π_1	π_2
	$P(\pi_1)$	$P(\pi_2)$
π_1	0	$C(1\|2)$
π_2	$C(2\|1)$	0

Det må være et rimeligt krav til klassificeringsreglen, at den på en eller anden måde minimerer disse omkostninger, således at de ved gentagen brug af klassificeringsreglen bliver så små som mulige. Dette kan lidt mere præcist udtrykkes som følger:

Vi ønsker at fastlægge vor klassificeringsregel på en sådan måde, at de forventede offeromkostninger ved fejlklassificeringer minimeres, hvilket formelt kan udtrykkes som følger:

$$\min_{R_1,R_2}\left[C(2|1)\, P(\pi_1)\int_{R_2} f_1(X)\, dX + C(1|2)\, P(\pi_2)\int_{R_1} f_2(X)\, dX\right] \quad (5)$$

En procedure, der fører til dette resultat, kaldes en *Bayes' procedure.* Vi vil senere se, at en procedure, der bygger på kriteriet (3–4) under visse, meget specielle forudsætninger vil være en Bayes procedure, men ellers vil vi benytte det Bayes'ianske synspunkt, som det er formuleret i (5).

4. *Klassificering af et element i en af to mulige populationer med vilkårlige, kendte tæthedsfunktioner*

Vort formål er at vælge R_1 og R_2 således, at (5) minimeres. For enhver klassificeringsregel $R^* = (R_1^*, R_2^*)$ er de forventede offeromkostninger som følge af fejlklassificering givet ved

$$C(2|1)\, P(\pi_1)\int_{R_2^*} f_1(X)\,dX + C(1|2)\, P(\pi_2)\int_{R_1^*} f_2(X)\,dX \quad (6)$$

af

$$\int_{R_2^*} f_2(X)\,dX + \int_{R_1^*} f_2(X)\,dX = 1 \quad (7)$$

fås

$$\int_{R_1^*} f_2(X)\,dX = 1 - \int_{R_2^*} f_2(X)\,dX \quad (8)$$

som indsat i (6) giver

$$C(2|1)\, P(\pi_1)\int_{R_2^*} f_1(X)\,dX + C(1|2)\, P(\pi_2)\left[1 - \int_{R_2^*} f_2(X)\,dX\right] \quad (9)$$

$$= \int_{R_2^*}\{C(2|1)P(\pi_1)f_1(X) - C(1|2)P(\pi_2)f_2(X)\}dX + C(1|2)P(\pi_2) \quad (10)$$

Vi ønsker at minimere (10). Da $C(1|2)$ og $P(\pi_2)$ er konstanter, opnås dette tydeligvis ved at fastlægge R_2^* således at det indeholder alle de punkter, for hvilke $C(2|1)\, P(\pi_1)\, f_1(X) - C(1|2)\, P(\pi_2)\, f_2(X)$ er negativ. Punkter, for hvilke dette udtryk er positivt tillægges da R_1^*.

Vi definerer altså R_1 og R_2 som følger[1]):

[1]) Punkter, hvor $C(2|1)\ P(\pi_1)f_1(X) - C(1|2)\ P(\pi_2)f_2(X) = 0$, er her arbitrært tillagt R_1.

$$R_1\colon\ C(2|1)\,P(\pi_1)\,f_1(X)-C(1|2)\,P(\pi_2)\,f_2(X)\geqslant 0$$
$$R_2\colon\ C(2|1)\,P(\pi_1)\,f_1(X)-C(1|2)\,P(\pi_2)\,f_2(X)<0 \qquad (11)$$

Det vil ofte være mere hensigtsmæssigt at arbejde med forholdet mellem tæthedsfunktionerne

$$D=\frac{f_1(X)}{f_2(X)} \qquad (12)$$

I så fald kan (11) skrives:

$$R_1\colon\ D=\frac{f_1(X)}{f_2(X)}\geqslant\frac{C(1|2)\,P(\pi_2)}{C(2|1)\,P(\pi_1)}$$
$$R_2\colon\ D=\frac{f_1(X)}{f_2(X)}<\frac{C(1|2)\,P(\pi_2)}{C(2|1)\,P(\pi_1)} \qquad (13)$$

Det skal sluttelig nævnes, at den klassiske regel (3)–(4), som det umiddelbart vil fremgå ved analog anvendelse af de her benyttede udledninger, fører frem til klassificeringsreglen

$$R_1\colon\ D=\frac{f_1(X)}{f_2(X)}\geqslant 1$$
$$R_2\colon\ D=\frac{f_1(X)}{f_2(X)}<1 \qquad (14)$$

der kun vil være en Bayes' procedure under forudsætningen

$$C(1|2)\,P(\pi_2)=C(2|1)\,P(\pi_1). \qquad (15)$$

Specielt er forudsætningen opfyldt for

$$P(\pi_1)=P(\pi_2)=0{,}5 \text{ og } C(2|1)=C(1|2), \qquad (16)$$

der formodentlig mere eller mindre bevidst ligger bag anvendelsen af den klassiske regel.

Udtrykkene (13) giver en formel løsning af klassificeringsproblemet uafhængig af tæthedsfunktionerne $f_1(X)$ og $f_2(X)$.

Af hensyn til løsningen af problemerne omkring anvendelsen af statistiske testprocedurer i forbindelse med diskriminantanalyse er det naturligvis påkrævet at kende disse tæthedsfunktioner, idet deres type og parametre dog eventuelt kan estimeres på grundlag af stikprøver fra hver af de mulige populationer. Disse problemer er imidlertid kun nogenlunde grundigt gennemarbejdede for så vidt tæthedsfunktionerne er multinormale. Denne forudsætning er også gjort i de indtil nu publicerede eksempler på anvendelse af diskriminantanalyse på marketingproblemer. Vi vil derfor i det følgende give en kort introduktion til den multinormale fordeling og dernæst skitsere beregningen af diskriminantfunktionen i det multinormale tilfælde.

5. *Den multinormale fordeling*

Den éndimensionale normalfordelings tæthedsfunktion kan som bekendt skrives som følger:

$$f(x) = ce^{-\frac{1}{2}\sigma^{-2}(x-\mu)^2} = ce^{-\frac{1}{2}(x-\mu)\sigma^{-2}(x-\mu)}, \tag{17}$$

hvor μ er fordelingens middelværdi og σ^2 dens varians, medens c er en konstant, der er valgt således, at integralet over $f(x)$ bliver 1. Den statistiske variabel x siges at være normal $N(\mu,\sigma^2)$.

Tæthedsfunktionen for den p-dimensionale normalfordeling har en analog form.

Skalar-variablen x erstattes med en søjlevektor:

$$X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \cdot \\ \cdot \\ \cdot \\ x_p \end{pmatrix} \tag{18}$$

Skalar-parametren μ, der angiver fordelingens middelværdi, erstattes med søjlevektoren:

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \cdot \\ \cdot \\ \cdot \\ \mu_p \end{pmatrix} \tag{19}$$

der angiver middelværdierne i de p marginalfordelinger.

I stedet for parametren σ^2 benyttes *co-variansmatricen*

$$V = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \dots & \sigma_{2p} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \dots & \sigma_{3p} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_{p1} & \sigma_{p2} & \sigma_{p3} & \dots & \sigma_{pp} \end{pmatrix} \tag{20}$$

hvor σ_{ij} for $i \neq j$ angiver co-variansen mellem den i'te og den j'te variabel, medens σ_{ij} for $i = j$ angiver variansen i den i'te marginalfordeling. Skrives $\sigma_{ij} = \sigma_i^2$ for alle $i = j$, kan (20) skrives:

$$V = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \dots & \sigma_{2p} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \dots & \sigma_{3p} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_{p1} & \sigma_{p2} & \sigma_{p3} & \dots & \sigma_p^2 \end{pmatrix} \tag{21}$$

Endelig erstattes udtrykket

$$\sigma^{-2}(x-\mu)^2 = (x-\mu)\,\sigma^{-2}(x-\mu) \tag{22}$$

med

$$(X-\mu)^T V^{-1}(X-\mu) \tag{23}$$

hvor T angiver, at vektoren er transponeret.

Den p-dimensionale normalfordelings tæthedsfunktion kan således skrives:

$$f(X) = c' e^{-\frac{1}{2}(X-\mu)^T V^{-1}(X-\mu)}, \tag{24}$$

hvor c' er en konstant, der er valgt således, at integralet over $f(X)$ bliver 1.

Variablerne X siges at være normale $N(\mu, V)$. Det ses umiddelbart, at for $p = 1$ fås den velkendte tæthedsfunktion for den éndimensionale normalfordeling (17).

6. *Klassificering i en af to mulige multinormale fordelinger med kendte parametre*

Der er givet to multinormale fordelinger $N(\mu^{(1)},V)$ og $N(\mu^{(2)},V)$ med ens co-variansmatricer.

Den r'te tæthedsfunktion ($r = 1,2$) er:

$$f_r(X) = c' e^{-\frac{1}{2}(X-\mu^{(r)})^T V^{-1} (X-\mu^{(r)})}, \tag{25}$$

og forholdet mellem tæthedsfunktionerne er

$$\frac{f_1(X)}{f_2(X)} = \frac{e^{-\frac{1}{2}(X-\mu^{(1)})^T V^{-1} (X-\mu^{(1)})}}{e^{-\frac{1}{2}(X-\mu^{(2)})^T V^{-1} (X-\mu^{(2)})}}, \tag{26}$$

hvilket kan skrives:

$$\frac{f_1(X)}{f_2(X)} = e^{-\frac{1}{2}[(X-\mu^{(1)})^T V^{-1} (X-\mu^{(1)}) - (X-\mu^{(2)})^T V^{-1} (X-\mu^{(2)})]} \tag{27}$$

Området R_1, for klassificering i π_1, er den mængde af vektorer X, for hvilke forholdet $f_1(X)/f_2(X) \geqslant k$, hvor k vælges hensigtsmæssigt, d. v. s. i overensstemmelse med det formål, der forfølges med den pågældende klassificeringsregel (jfr. afsnit 3).

Da logaritmefunktionen er monotont voksende, kan vi omskrive (27) i logaritmer og får da:

$$-\tfrac{1}{2}[(X-\mu^{(1)})^T - V^{-1}(X-\mu^{(1)}) - (X-\mu^{(2)})^T V^{-1}(X-\mu^{(2)})] \geqslant \ln k \tag{28}$$

eller

$$X^T V^{-1}(\mu^{(1)} - \mu^{(2)}) - \tfrac{1}{2}(\mu^{(1)} + \mu^{(2)})^T V^{-1}(\mu^{(1)} - \mu^{(2)}) \geqslant \ln k \tag{29}$$

Udtrykket $\frac{1}{2}(\mu^{(1)} + \mu^{(2)})$ er gennemsnittet af middeltalsvektorerne i de to fordelinger. Kaldes denne størrelse $\bar{\mu}$ kan (29) skrives:

$$X^T V^{-1}(\mu^{(1)} - \mu^{(2)}) - \bar{\mu}^T V^{-1}(\mu^{(1)} - \mu^{(2)}) \geqslant \ln k \tag{30}$$

Da sidste led på venstre side er en konstant, kan vi benytte første led alene som vor *diskriminantfunktion*. Det ses, at diskriminantfunktionen er en lineær funktion af den observerede vektor X. Vi skriver da (30) som

$$D = X^T V^{-1}(\mu^{(1)} - \mu^{(2)}) \geqslant \ln k + \bar{\mu}^T V^{-1}(\mu^{(1)} - \mu^{(2)}) \tag{31}$$

Som tidligere nævnt, må konstanten k ansættes »hensigtsmæssigt«. F. eks. vil en antagelse af det »ortodokse« kriterium for optimalitet for klassificeringsreglen: minimering af de *betingede* sandsynligheder for fejlklassificeringer føre til $k = 1$ (if. (14)), hvoraf følger, at $\ln k = 0$. (31) kan da skrives:

$$D = X^T V^{-1}(\mu^{(1)} - \mu^{(2)}) \geqslant \bar{\mu}^T V^{-1}(\mu^{(1)} - \mu^{(2)}) \tag{32}$$

Med andre ord: Såfremt diskriminantfunktionen ved indsættelse af den observerede vektor X antager en værdi, der er større end eller lig med den værdi, den tager ved indsættelse af vektoren $\bar{\mu}$, der angiver gennemsnittet af $\mu^{(1)}$ og $\mu^{(2)}$, er det observerede element beliggende i R_1 og klassificeres derfor som tilhørende π_1.

I alle de praktiske anvendelser inden for marketingområdet, som er offentliggjort, er benyttet $k = 1$, medens det – i de fleste tilfælde – betydelig mere realistiske beslutningsteoretiske oplæg if. (13) må føre til at

$$k = \frac{C(1|2)\,P(\pi_2)}{C(2|1)\,P(\pi_1)} \tag{33}$$

indsættes i (31).

7. *Klassificering i en af to mulige multinormale fordelinger, når parametrene skal estimeres*

Der er givet en stikprøve $X_1^{(1)}$, $X_2^{(1)}$ .... $X_{n_1}^{(1)}$ fra π_1 ($X^{(1)}$ er $N(\mu^{(1)}, V)$) og en stikprøve $X_1^{(2)}$, $X_2^{(2)}$ .... $X_{n_2}^{(2)}$ fra π_2 ($X^{(2)}$ er $N(\mu^{(2)}, V)$). På grundlag af denne information ønskes et element med tilhørende målinger X klassificeret som tilhørende enten π_1 eller π_2.

I modsætning til det i forrige afsnit behandlede tilfælde kender vi ikke parametrene i de to fordelinger, men må estimere dem på grundlag af de to stikprøver.

Maksimum-likelihood-estimatorerne for $\mu^{(1)}$ og $\mu^{(2)}$ er

$$X^{(1)} = \frac{1}{n_1} \sum_{\alpha=1}^{n_1} X_\alpha^{(1)} \text{ og } X^{(2)} = \frac{1}{n_2} \sum_{\alpha=1}^{n_2} X_\alpha^{(2)}), \tag{35}$$

medens maksimum-likelihood-estimatoren for den fælles co-varians-matrix er:

$$\hat{V} = \frac{1}{n_1 + n_2 - 2} (\hat{W}^{(1)} + \hat{W}^{(2)}), \tag{35}$$

hvor

$$\hat{W}^{(1)} = \sum_{\alpha=1}^{n_1} (X_\alpha^{(1)} - X^{(1)})(X_\alpha^{(1)} - X^{(1)})^T \tag{36}$$

og

$$\hat{W}^{(2)} = \sum_{\alpha=1}^{n_1} (X_\alpha^{(2)} - X^{(2)})(X_\alpha^{(2)} - X^{(2)})^T \tag{37}$$

Ved indsættelse af (34) og (35) i (31) fås:

$$\hat{D} = X^T \hat{V}^{-1} (X^{(1)} - X^{(2)}) \geqslant \ln\ k + \bar{\bar{X}}^T \hat{V}^{-1} (X^{(1)} - X^{(2)}) \qquad (38)$$

hvor

$$\bar{\bar{X}} = \tfrac{1}{2} (X^{(1)} + X^{(2)}) \qquad (39)$$

$\hat{D}$ kan da benyttes som en tilnærmelse til den eksakte (men ukendte) diskriminantfunktion D.

Det vides ikke, om $\hat{D}$ er optimal, men det må antages, at tilnærmelsen er »god«, da fordelingen for $\hat{D}$ konvergerer mod fordelingen for D, når såvel $n_1 \to \infty$ som $n_2 \to \infty$, forudsat, at populationerne er ubegrænsede (eller der udtrækkes med tilbagelægning).

8. *Anvendelser af diskriminantanalyse inden for afsætningsøkonomien*

Der er gennem de senere år fremkommet en del referater af praktiske anvendelser af diskriminantanalysen på afsætningsøkonomiske problemstillinger. Disse er hovedsagelig offentliggjort i amerikanske tidsskrifter, medens vi endnu har til gode at se et eksempel på teknikkens anvendelse i en dansk virksomhed.

Nedenstående artikler må betragtes som repræsentative med hensyn til de problemer man på teknikkens nuværende stade angriber med diskriminantanalyse. Numrene henviser til litteraturlisten.

Banks (L. 3) har benyttet diskriminantanalyse til at studere forskellene i forbrugernes præferancer for forskellige varemærker på basis af forskelle i produkternes egenskaber.

Buck (L. 4) er interesseret i at fastslå den relative vægt, der kan tillægges elleve forskellige socio-økonomiske variable som indikatorer for en husstands besiddelse af køleskab, og for dens besiddelse af gas- henholdsvis el-køkken, idet hans problem er at reducere antallet af forklarende variable.

Bucklin (L. 5) har undersøgt forbrugernes valg af shopping-center som funktion af socio-økonomiske og demografiske kriterier.

Claycamp (L. 6) har studeret de sociale og økonomiske forskelle mellem sparere, der benytter forskellige opsparingsformer.

Evans (L. 7) har studeret, i hvilket omfang socio-økonomiske og psykologiske faktorer bestemmer en husstands valg af bilmærke.

Frank, Masy & Morrison (L. 8) har behandlet problemet: Er det muligt på grundlag af en husstands socio-økonomiske egenskaber og dens købevaner at forudsige dens »villighed« til at acceptere et nyintroduceret varemærke? Det konkrete tilfælde drejede sig om introduktionen af en færdig-

malet kaffe på Chicagomarkedet, og forfatterne benyttede »Chicago Tribune«s forbrugerpanel til fremskaffelse af de nødvendige data.
King (L. 9) har benyttet diskriminantanalyse til at vurdere forskellige faktorers værdi som mål for et sælgerdistrikts »godhed«.
Massy (L. 10) har studeret lytternes valg af radioprogrammer baseret på socio-økonomiske og demografiske kriterier.

I alle de i dette afsnit nævnte tilfælde er det forudsat, at normalitetskravet er opfyldt, således at diskriminantfunktionen får den i formel (31) afsnit 6 og (38) afsnit 7 nævnte form. I en del af eksemplerne er denne forudsætning dog ret problematisk, hvorfor de statistiske tests må tages med alt mulig forbehold.

Litteratur

En teoretisk behandling af den »ortodokse« diskriminantanalyse findes i

(1) *M. G. Kendall:*
A Course in Multivariate Analysis.
Griffin & Co. 1965. London.

En beslutningsteoretisk synsmåde, hvori også indgår en optimering af antallet af målinger på det enkelte element, er anlagt i

(2) *Paul E. Green:*
Bayesian Classification Procedures in Analyzing Customer Chararteristics.
Journal of Marketing Research, Vol. 1 (1964) May, pp. 44–50.

Som exempler på praktiske anvendelser af linear diskriminantanalyse kan nævnes

(3) *S. Banks:*
The Relationship Between Preference and Purchase of Brands.
Journal of Marketing, Vol. 15 (1950), pp. 145–157.

(4) *S. F. Buck:*
A New Look at Old Data.
Papers, ESOMAR Congress 1966.

(5) *Louis P. Bucklin:*
The Concept of Mass in Intra-Urban Shopping.
Journal of Marketing, Vol. 31 (1667), October, pp. 37–42.

(6) *H. J. Claycamp:*
Characteristics of Owners of Thrift Deposits in Commercial Banks and Savings and Loan Associations.
Journal of Marketing Research, Volä. 2 (1965), pp. 163–170.

(7) *F. B. Evans:*
Psychological and Objective Factors in the Prediction of Brand Choice, Ford Versus Chevrolet.
Journal of Busiess, Vol. 32 (1959), pp. 340–369.

malet kaffe på Chicagomarkedet, og forfatterne benyttede »Chicago Tribune«s forbrugerpanel til fremskaffelse af de nødvendige data.
King (L. 9) har benyttet diskriminantanalyse til at vurdere forskellige faktorers værdi som mål for et sælgerdistrikts »godhed«.
Massy (L. 10) har studeret lytternes valg af radioprogrammer baseret på socio-økonomiske og demografiske kriterier.

I alle de i dette afsnit nævnte tilfælde er det forudsat, at normalitetskravet er opfyldt, således at diskriminantfunktionen får den i formel (31) afsnit 6 og (38) afsnit 7 nævnte form. I en del af eksemplerne er denne forudsætning dog ret problematisk, hvorfor de statistiske tests må tages med alt mulig forbehold.

Litteratur

En teoretisk behandling af den »ortodokse« diskriminantanalyse findes i

(1) *M. G. Kendall:*
A Course in Multivariate Analysis.
Griffin & Co. 1965. London.

En beslutningsteoretisk synsmåde, hvori også indgår en optimering af antallet af målinger på det enkelte element, er anlagt i

(2) *Paul E. Green:*
Bayesian Classification Procedures in Analyzing Customer Chararteristics.
Journal of Marketing Research, Vol. 1 (1964) May, pp. 44–50.

Som exempler på praktiske anvendelser af lineær diskriminantanalyse kan nævnes

(3) *S. Banks:*
The Relationship Between Preference and Purchase of Brands.
Journal of Marketing, Vol. 15 (1950), pp. 145–157.

(4) *S. F. Buck:*
A New Look at Old Data.
Papers, ESOMAR Congress 1966.

(5) *Louis P. Bucklin:*
The Concept of Mass in Intra-Urban Shopping.
Journal of Marketing, Vol. 31 (1667), October, pp. 37–42.

(6) *H. J. Claycamp:*
Characteristics of Owners of Thrift Deposits in Commercial Banks and Savings and Loan Associations.
Journal of Marketing Research, Volä. 2 (1965), pp. 163–170.

(7) *F. B. Evans:*
Psychological and Objective Factors in the Prediction of Brand Choice, Ford Versus Chevrolet.
Journal of Busiess, Vol. 32 (1959), pp. 340–369.

(8) *Frank, Massy and Morrison:*
The Determinants of Innovative Behaviour With Respect to a Branded Frequently Purchased Food Product.
Proceedings of American Marketing Association, December 1964.

(9) *W. R. King:*
Marketing Expansion – A Statistical Analysis.
Management Science, Vol. 9 (1963), pp. 563–573.

(10) *W. F. Massy:*
Discriminant Analysis of Audience Characteristics.
Journal of Advertising Research, Vol. 5 (1965), pp. 39–48.