



UNIVERSITAT DE BARCELONA



Departament d'Estadística

Contribucions a les desigualtats en regressió i anàlisi multivariant

Josep M. Duran Rúbies

Contribucions a les desigualtats en regressió i anàlisi multivariant

Memòria presentada, dins del programa de doctorat Probabilitats i Estadística aprovat pel bienni 1994-6, per optar al Títol de Doctor en Matemàtiques per la Universitat de Barcelona

Josep M. Duran Rúbies

Vist i plau
El Director

Dr. Carles M. Cuadras
Catedràtic d'Estadística i IO
Departament d'Estadística
Universitat de Barcelona

A la Rosa Maria
i a la Mare

Vull agrair al doctor Carles M. Cuadras el recolzament de tot tipus que sempre m'ha donat i també pels bons moments amb els quals m'ha obsequiat. Sense ell aquesta memòria no hagués estat possible.

Vull agrair també a la meva dona Rosa Maria, a la meva germana Maria Rosa i al meu nebot Javier el que en moments de desànim m'encoratgessin a seguir.

Finalment agraeixo especialment a la Rosa Maria la gairebé infinita paciència que ha tingut amb mi.

Barcelona, maig 2003.

Contribucions a les desigualtats en regressió i anàlisi multivariant

Maig 2003

Josep M. Duran Rúbies

**Tesi doctoral dirigida per
Carles M. Cuadras**

**Departament d'Estadística
Universitat de Barcelona**

Índex

1	Una desigualtat de la regressió múltiple	11
1.1	Introducció	11
1.2	Regressió múltiple	12
1.3	Desigualtat de Hamilton - Routledge - Cuadras (HRC)	13
1.4	Interpretació per components principals de la desigualtat HRC	17
1.5	Alguns exemples il·lustratius	20
1.6	Regressió en components principals	24
1.7	Influència de l'última component principal	26
1.8	Conclusions	27
2	L'efecte d'augmentar les correlacions simples	29
2.1	Introducció	29
2.2	Variables equicorrelacionades	30
2.3	Exemples per a la desigualtat HRC	31
2.4	Desigualtat HRC forta	34
2.5	Augmentant les correlacions simples no té per què augmentar R^2	36
2.6	Una generalització	40
3	La paradoxa de Simpson	45
3.1	Introducció	45
3.2	Exemples de la paradoxa de Simpson	45
3.3	Anàlisi de Correspondències	50
3.4	Aplicació a la representació de la paradoxa de Simpson	54
4	Desigualtats amb distàncies	63
4.1	Introducció	63

4.2	Dues poblacions	64
4.3	Exemple il·lustratiu	69
4.4	Més de dues poblacions	73
4.5	Generalització	74
4.6	Desigualtat en el MANOVA	74
4.7	Exemple de Fisher	76
4.8	Estudi gràfic de l'exemple de Fisher	79
4.8.1	Variables primera i segona corresponents als sèpals . . .	79
4.8.2	Variables tercera i quarta corresponents als pètals . . .	81
4.9	Representació d'una component principal a l'espai canònic . .	82
4.10	Altres exemples	85
4.11	La paradoxa de Rao	86
5	Densitats a partir de distàncies	93
5.1	Introducció	93
5.2	Variabilitat geomètrica	93
5.3	Desigualtat per a la variabilitat geomètrica	95
5.4	Expansions ortogonals	98
5.5	Funció de proximitat	100
5.6	Representació euclidiana discreta	101
5.7	Transformació de distàncies i funcions de proximitat	103
5.8	Densitats a partir de proximitats	104
5.9	Desigualtat entre la variabilitat geomètrica i l'entropia de Shan- non.	113
5.10	Cas bivariant. Definicions	121
5.10.1	Dissimilaritat bivariant	122
5.10.2	Variabilitat geomètrica bivariant	125
5.10.3	Funció de proximitat bivariant	126
5.11	Densitat conjunta a partir de la funció de proximitat bivariant	128
6	Distància de Wasserstein i altres desigualtats	131
6.1	Introducció	131
6.2	Definició de la distància de Wasserstein	131
6.2.1	Exemples	132
6.3	Classes de Fréchet i correlacions de Hölder	133
6.3.1	Exemple	135

ÍNDIX	4
6.4 Desigualtats per a la distància de Wasserstein	135
6.5 Exemples	137
6.6 Desigualtats per a la variància	139
6.7 Una prova alternativa en el cas logístic	140
6.8 Una desigualtat general	143
7 Desigualtats a partir de la raó de versemblança	145
7.1 Introducció	145
7.2 Teorema de les mitjanes aritmètica i geomètrica	146
7.3 Generalització	149
7.4 Altres desigualtats	151
8 Resum i conclusions	155
8.1 Resum	155
8.2 Conclusions	160

Índex de Figures

1.1	$R^2 = r_{yx_1}^2 + r_{yx_2}^2$. Les variables X_1 i X_2 no estan correlacionades. És com fer dues regressions independents.	14
1.2	Exemple extrem per la desigualtat HRC. $R^2 = 1$ al trobar-se Y en el mateix pla que X_1 i X_2 . D'altra banda: $r_{yx_1}^2 = 0$ i $r_{yx_2}^2 = 1$. Cal notar, també, que $r_{12} = 1$	16
1.3	Exemple per la desigualtat HRC. $r_{yx_1}^2 = 0$ i $r_{yx_2}^2 = 1$. Cal notar, també, que $r_{12} = 0$	17
1.4	Núvol de punts considerant la variable dependent Y i la primera component principal. S'observa la poca relació entre elles i la gran variabilitat de la primera component principal.	22
1.5	Núvol de punts considerant la variable Y i la segona component principal en l'exemple de Hamilton en el que s'observa la clara relació lineal entre elles així com la poca variabilitat de la segona component principal.	22
1.6	Gràfic tridimensional en el que es veu com Y no té tendència lineal al llarg de la recta de regressió de X_2 sobre X_1	23
1.7	Núvol de punts tridimensional en el que es veu com Y té tendència lineal al llarg de la recta de regressió de la segona component principal sobre la primera.	23
2.1	Relació entre el nombre de variables correlacionades amb Y i el coeficient de determinació d'una banda i la suma de $r_{yx_i}^2$ de l'altra.	33
3.1	Paradoxa de Simpson. Probabilitats en l'exemple de l'eficàcia d'un nou tractament. La probabilitat tenint en compte les poblacions i sense tenir-les en compte van en direccions diferents.	48

3.2	Paradoxa de Simpson. Probabilitats en l'exemple d'admissió en tres facultats, A, B i C. Les dades globals van en una direcció descendent i en cadascuna de les facultats la direcció és ascendent.	51
3.3	Representació gràfica de l'AC asimètric per l'exemple de les tres facultats. Representem les ...les amb les coordenades principals i les columnes amb les coordenades estàndar. La proximitat del punt que representa les noies no admeses (n-n) al punt que representa la facultat B, indica una forta demanda a aquesta facultat que és la que més rebutja. (n-s=noies sí admeses, n-n=noies no admeses, N-s=nois sí admesos i N-n=nois no admesos)	56
3.4	Representació gràfica ampliada de l'AC asimètric (representem les ...les amb les coordenades principals i les columnes amb les coordenades estàndar) per l'exemple de les tres facultats. Les noies sol·liciten més l'ingrés a la facultat B que és la que té més percentatge de no admesos, part ombrejada de cada cercle, la qual cosa explica la paradoxa.	57
3.5	Representació gràfica ampliada de l'AC asimètric (representem les ...les amb les coordenades principals i les columnes amb les coordenades estàndar) per a l'exemple del tabac i supervivència després de 20 anys. Els fumadors, ns-f, ss-f estan més prop dels grups d'edat joves, on la supervivència és més gran. La part ratllada del cercle indica la proporció d'individus que no sobreviuen. Els punts ns (no sobreviuen) i ss (sí sobreviuen) són les mitjanes ponderades de ns-f, ns-nf i de ss-f, ss-nf, respectivament. Les barres sobre ns-f i ns-nf indiquen el percentatge de no supervivència de fumadores i no fumadores respectivament.	61
4.1	$\pm_M^2 > \pm_K^2$. La direcció del segment que uneix les mitjanes de les dues poblacions segueix, pràcticament, la direcció de la segona component principal.	65
4.2	Significació del coeficient alfa, $\alpha_i = \cos \mu_i$	68

4.3	El·lipses de concentració pels sis casos presentats per Flury per l'estudi de la discriminació lineal entre dues poblacions. En els casos (4) i (6) s'acompleix la desigualtat (4.2).	70
4.4	Exemple de Fisher sobre ƒors del gènere Iris considerant només les dues variables corresponents als sèpals. La desigualtat (4.12) es verifica per als aparellaments en els quals intervé la població setosa. El segment que uneix la mitjana de la població setosa amb cadascuna de les altres poblacions tendeix a seguir la direcció de la segona component principal (eix curt).	80
4.5	Exemple de Fisher sobre ƒors del gènere Iris considerant només les dues variables corresponents als pètals. La desigualtat (4.12) no es verifica. El segment que uneix les mitjanes de les poblacions segueix la direcció de la primera component principal (eix llarg).	81
4.6	Representació canònica dels individus, les mitjanes de les poblacions i de les dues components principals en l'exemple de Fisher, considerant totes les variables. La desigualtat (4.12) no es verifica. En les poblacions versicolor i virgínica la primera component principal és paral·lela al primer eix canònic, mentre que en la població setosa és la segona component principal la que gairebé és paral·lela al primer eix canònic. Recordem que considerant la població setosa amb cadascuna de les altres dues poblacions la desigualtat (4.2) es verifica, mentre que considerant només les poblacions versicolor i virgínica no es verifica.	83
4.7	Representació canònica dels individus, les mitjanes i les components principals en l'exemple de Fisher considerant les dues variables corresponents als sèpals. La desigualtat (4.12) es verifica. Les segones components principals són gairebé paral·leles al primer eix canònic, la direcció del qual és la de màxima diferència entre poblacions. Aleshores tenim la desigualtat distància de Mahalanobis més gran que distància de K. Pearson.	84

4.8 Representació canònica dels individus, les mitjanes i les components principals en l'exemple de Fisher considerant les dues variables corresponents als pètals. La desigualtat (4.12) no es verifica. Les primeres components principals, màxima variabilitat de les dades, formen un angle agut amb el primer eix canònic, màxima variabilitat de les poblacions. Aleshores tenim que distància de Mahalanobis més petita que distància de K. Pearson. 85

4.9 Exemple en el que la desigualtat (4.12) s'acompleix. La direcció del segment que uneix les mitjanes més aviat segueix la segona component principal. 87

4.10 Exemple dels cranis egipcis trobats en cinc períodes, en el que la desigualtat (4.12) s'acompleix. La direcció principal de les mitjanes segueix gairebé la direcció de la segona component principal (eix petit). 87

4.11 Explicació de la paradoxa de Rao. Només els valors sobre el cantó esquerra-baix i el cantó dreta-dalt (com l'assenyalat amb +), poden donar significació per a x; significació per a y, però no per a (x; y) conjuntament. 91

5.1 Densitat de la distribució uniforme i de la generada per la distància $\frac{p}{jx_i + y_j}$; amb $a = 1$ i $b = 0$: 32787 equivalent a la generada per la distància $jx_i + y_j$ amb $a = 1$ i $b = 0$: 1612. . . 108

5.2 Densitat de la distribució uniforme i de la generada per la distància $\frac{p}{jx_i + y_j}$; amb $a = 1$ i $b = 0$: 33897, equivalent a la generada per la distància $jx_i + y_j$; amb $a = 1$ i $b = 0$: 17231. . . 108

5.3 Densitat de la distribució exponencial $\exp(-\lambda x)$ i de la generada per la distància $jx_i + y_j$, per $\lambda = 8$: 7513. 110

5.4 Densitat de la distribució exponencial $\exp(-\lambda x)$ i de la generada per la distància $jx_i + y_j$; per $\lambda = 1$ i $a = 3$: 1017. 111

5.5 Densitat de la distribució exponencial $\exp(-\lambda x)$ i de la generada per la distància $jx_i + y_j$; per $\lambda = 2$: 2185 i $a = 2$: 1336. . . 111

5.6 Densitat de la distribució exponencial $\exp(-\lambda x)$ i de la generada per la distància $jx_i + y_j$; per $\lambda = 0.5$ i $a = \frac{1}{4}$ 112

5.7 Densitat de la distribució exponencial $\exp(-\lambda x)$ i de la generada per la distància $jx_i + y_j$; per $\lambda = 1$, $a = 1$ i $b = 0$: 9809. . . 112

Índex de Taules

2.1	Exemple amb variables equicorrelacionades. Avaluació de la desigualtat HRC i la condició d'acompliment.	32
2.2	Exemple amb variables equicorrelacionades. Influx de les components principals.	32
2.3	Valors de s pels quals $R^2(s)$ arriba al seu màxim i pels quals l'expressió $\sum_{i=1}^p r_{yz_i}^2 (1 - s_i)$ passa a ser negativa, per diferents valors de p i de $c > 0$	39
3.1	Paradoxa de Simpson. Tant en la població C com en la C^0 , és millor el nou tractament B.	46
3.2	Paradoxa de Simpson. Sense tenir en compte la població, el nou tractament B no beneficia el pacient.	47
3.3	Freqüències d'admissió a tres facultats. En cadascuna de les facultats A, B i C, les noies són beneficiades en l'admissió. . . .	47
3.4	Freqüències d'admissió a tres facultats. Globalment, sense tenir en compte les facultats, els nois són beneficiats en l'admissió. .	47
3.5	Freqüències d'admissió a tres facultats on considerem com una sola variable, sexe - admissió, les variables inicials, sexe i admissió.	55
3.6	El fet de ser fumadora (F) té un efecte positiu sobre la supervivència (NS: no sobrevis, SS: sí sobrevis), després de 20 anys, en les dones d'aquesta mostra.	58
3.7	Taula de freqüències, per l'exemple de la relació entre tabac i supervivència, considerant com una sola variable, supervivència-tabac, les variables inicials supervivència i tabac. En cadascun dels grups d'edat, tret del grup 25-34 anys, el tabac té un efecte negatiu sobre la supervivència, després de 20 anys, en les dones d'aquesta mostra.	58

ÍNDIX DE TAULES	10
4.1 Vectors de mitjanes i matrius de covariància i correlació, així com l'avaluació de les distàncies de Mahalanobis i de K. Pearson pels sis casos de discriminació lineal presentats per Flury. S'observa que la desigualtat (4.2) es verifica en el quart i en el sisè cas.	71
4.2 Valors i vectors propis i coeficients alfa, així com l'avaluació de la condició de compliment de la desigualtat (4.2), pels sis casos de discriminació lineal presentats per Flury. En el quart i en el sisè la condició s'acompleix.	72
4.3 Exemple de Fisher sobre tors del gènere Iris. Vectors de mitjanes per les tres poblacions i pel total de les dades.	76
4.4 Exemple de Fisher. Matrius de dispersió dintre els grups.	77
4.5 Exemple de Fisher. Matrius de dispersió entre els grups.	77
4.6 Avaluació de la desigualtat per dues poblacions en l'exemple de Fisher.	79
5.1 Variabilitat geomètrica per a diferents distribucions i per a la distància arrel quadrada de la distància euclídea ordinària.	96
5.2 Funció de proximitat per a diferents distribucions i per a la distància arrel quadrada de la distància euclidiana ordinària.	101
5.3 Entropia per a diferents distribucions.	114
5.4 Divergència per a diferents distribucions i distàncies.	117
6.1 Dues primeres components principals, valors propis i funcions pròpies per a la variable logística.	141

Capítol 1

Una desigualtat de la regressió múltiple

1.1 Introducció

Les desigualtats en probabilitats, estadística i anàlisi multivariant constitueixen una part fructífera de la teoria matemàtica, doncs permeten acotar certes quantitats, construir tests, o són útils per provar determinats teoremes.

Un exemple del primer tipus és la famosa desigualtat de Cramér-Rao, que proporciona una cota per a la variància d'un estimador no esbiaixat.

Amb les desigualtats de Bonferroni, podem construir tests simultanis, amb aplicacions a les comparacions múltiples de l'anàlisi de la variància.

La desigualtat de Tchebichev permet provar molts resultats de convergència de successions de variables aleatòries, però també trobar intervals de confiança en situacions de poca informació sobre els paràmetres.

Donades dues distribucions univariants, les distribucions bivariants màxima i mínima (cotes de Fréchet) acoten qualsevol distribució bivariant. Anàlogament, qualsevol coeficient de correlació està acotat per les correlacions màxima i mínima de Höfding.

En aquesta memòria presentem, estudiem i il·lustrem diverses desigualtats d'interès en regressió i anàlisi multivariant. No tots els resultats obtinguts són originals, com queda clar en el context. Però sí que és una contribució original la relació entre els resultats, la visió conjunta i algunes il·lustracions.

1.2 Regressió múltiple

Un dels mètodes estadístics més importants consisteix en relacionar una variable dependent Y amb p variables explicatives $X_1; \dots; X_p$, a través d'una funció lineal, segons el model

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \epsilon_i; \quad i = 1; \dots; n; \quad (1.1)$$

on n és el nombre d'individus de la mostra, en els quals hem observat els valors de les variables, $\beta_1; \dots; \beta_p$ són els paràmetres a estimar i ϵ_i és un terme d'error aleatori, que suposem amb mitjana 0 i variància σ^2 .

Matricialment podem escriure el model com

$$Y = X\beta + \epsilon, \quad (1.2)$$

on Y és un vector $n \times 1$ de n observacions de la variable resposta sobre els n individus, X és una matriu $n \times p$ amb les observacions de les p variables explicatives, $\beta = (\beta_0; \dots; \beta_p)'$ és un vector desconegut dels paràmetres de la regressió i ϵ és un vector de termes d'error.¹

L'equació (1.2) és el model de regressió múltiple. El nom "regressió" prové del títol d'un dels primers articles sobre el tema degut a Galton [34].

Suposant $\text{rang}(X) = p$; l'estimació per mínims quadrats ordinaris (OLS) dels paràmetres de regressió, $\beta = (\beta_0; \dots; \beta_p)'$, és

$$\mathbf{b} = (X'X)^{-1} X'Y;$$

Aquest càlcul pot comportar problemes en cas de multicolinealitat, i.e., quan les variables estan molt correlacionades, ja que la matriu $X'X$, (relacionada amb la matriu de correlacions si X és matriu centrada) té un número de condició molt alt en el cas en el que almenys un parell de les variables explicatives $X_1; \dots; X_p$ tinguin correlació propera a 1.

L'anàlisi de regressió múltiple és una de les tècniques estadístiques més utilitzada en la pràctica. O més malament utilitzada, com molt bé explica Rao [57]. De fet algunes vegades es redueix a l'aplicació de fórmules, tant per al càlcul dels paràmetres com per a l'estudi de l'adequació i l'acompliment

¹Algunes vegades el model s'escriu en la forma $y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \epsilon_i$; aleshores la forma matricial seria la mateixa però X seria una matriu $n \times (p + 1)$ amb la primera columna tota d'uns i β passaria a ser un vector $(p + 1) \times 1$.

de les hipòtesis del model. L'ús massiu de l'ordinador potencia encara més aquest fet.

En aquest capítol presentem un resultat desconcertant en un primer cop d'ull però que ens pot avisar d'alguns dels perills de la utilització, a la lleugera, de la regressió múltiple.

1.3 Desigualtat de Hamilton - Routledge - Cuadras (HRC)

Considerem la regressió múltiple entre la variable resposta Y i les variables explicatives $X = (X_1; \dots; X_p)$.

Si observem aquestes variables en n individus escriurem el vector Y i la matriu X en la forma

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}; \quad X = \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1p} \\ X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ X_{n1} & X_{n2} & \dots & X_{np} \end{pmatrix}$$

La descomposició de la suma de quadrats de la variable Y ;

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n e_i^2,$$

on \bar{y} és la mitjana de la mostra de la variable Y ; \hat{y}_i és el valor de la variable Y en l'individu i predit per la regressió i e_i és la diferència entre el valor observat i el valor predit de la variable Y en l'individu i , suggereix utilitzar la següent quantitat, anomenada coeficient de determinació,

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (1.3)$$

per mesurar l'adequació del model de regressió. R^2 és el quadrat del coeficient de correlació múltiple, $R = \text{Corr}(Y; \hat{Y})$.

D'una manera informal, observant (1.3), R^2 és la proporció de variabilitat de la variable Y explicada per les variables $X_1; \dots; X_p$.

Com hem dit, es defineix $R = \sqrt{R^2}$ com el coeficient de correlació múltiple entre Y i $X_1; \dots; X_p$. Una definició alternativa a través del model de regressió es pot trobar en Cuadras [14].

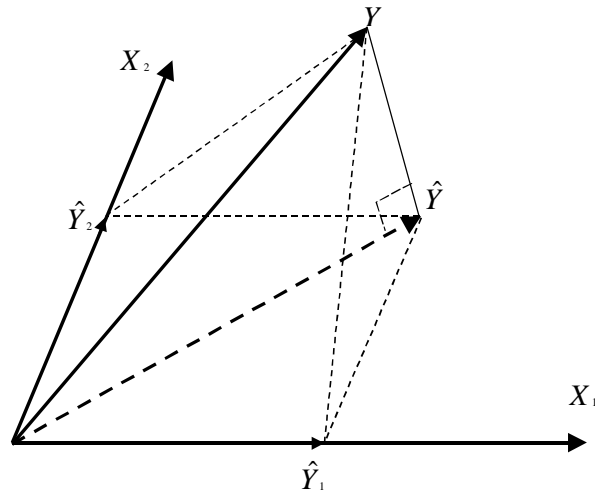


Figura 1.1: $R^2 = r_{yx_1}^2 + r_{yx_2}^2$. Les variables X_1 i X_2 no estan correlacionades. És com fer dues regressions independents.

El seu valor està entre 0, si les variables explicatives no tenen influència en la variable resposta, i 1 si el model de regressió s'ajusta perfectament a les dades.

Si les variables explicatives $X_1; \dots; X_p$ no estan correlacionades s'acompleix la igualtat estricta

$$R^2 = \sum_{i=1}^p r_{yx_i}^2,$$

on r_{yx_i} és el coeficient de correlació simple entre la variable Y i la variable X_i . En aquest cas la regressió múltiple es comporta com si fèssim p regressions simples per separat, explicant cadascuna d'elles part de la variabilitat de la variable Y . En la figura 1.1 podem veure la certesa d'aquesta afirmació, en el cas $p = 2$.

Siguin Y el vector de dades, \hat{Y} el vector estimat per la regressió d' Y sobre X_1 i X_2 conjuntament i \hat{Y}_1 i \hat{Y}_2 els vectors estimats per les regressions de Y

sobre X_1 i X_2 respectivament. Tenim

$$R^2 = \cos^2(Y; \hat{Y}) = \frac{\hat{Y}^0 \hat{Y}}{\hat{Y}^0 \hat{Y}} ;$$

$$r_{yx_1}^2 = \cos^2(Y; X_1) = \frac{\hat{Y}_1^0 \hat{Y}_1}{\hat{Y}^0 \hat{Y}} ;$$

$$r_{yx_2}^2 = \cos^2(Y; X_2) = \frac{\hat{Y}_2^0 \hat{Y}_2}{\hat{Y}^0 \hat{Y}} ;$$

Per tant, si X_1 i X_2 són ortogonals (in correlacionades), tindrem

$$r_{yx_1}^2 + r_{yx_2}^2 = \frac{\hat{Y}_1^0 \hat{Y}_1 + \hat{Y}_2^0 \hat{Y}_2}{\hat{Y}^0 \hat{Y}} = \frac{\hat{Y}^0 \hat{Y}}{\hat{Y}^0 \hat{Y}} = R^2 .$$

És clar que si la condició d'ortogonalitat (in correlació), no s'acomplís, no hi hauria igualtat. Són interessants els gràfics de Hamilton [39] per veure en quin sentit es verifica la desigualtat.

Intuitivament semblaria que, Ott [54], variables explicatives correlacionades només haurien de contenir informació redundant i per tant hauríem de tenir

$$R^2 \leq \sum_{i=1}^k r_{yx_i}^2 .$$

Efectivament, part de la variabilitat de la variable Y explicada per la variable X_i es pot explicar per les altres variables regressores correlacionades amb aquesta.

Però, resulta que la desigualtat contrària també es pot aconseguir. És a dir

$$R^2 > \sum_{i=1}^k r_{yx_i}^2 . \tag{1.4}$$

Hamilton [39], Routledge [58], Cuadras [10], entre altres, presenten i discuteixen aquesta desigualtat, desigualtat que anomenem de Hamilton-Routledge-Cuadras, abreujadament HRC, seguint la terminologia proposada per Tiit [63]. També és interessant el gràfic, figura 1.2, de Kendall i Stuart [47], com un cas extrem per a la desigualtat (1.4). En la figura 1.3 hem modificat aquest últim gràfic, verificant també, però, la desigualtat HRC.

De fet la desigualtat (1.4) succeeix més vegades de les que la intuïció ens podria fer pensar. Així, Shieh [59], resulta ser igualment probable tenir com

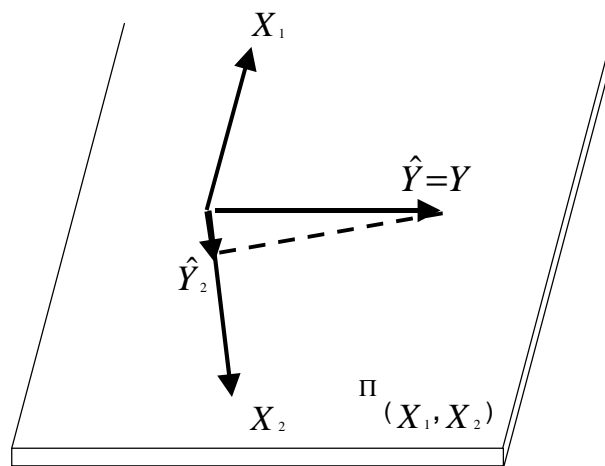


Figura 1.2: Exemple extrem per la desigualtat HRC. $R^2 = 1$ al trobar-se Y en el mateix pla que X_1 i X_2 . D'altra banda: $r_{y x_1}^2 = 0$ i $r_{y x_2}^2 = 1$. Cal notar, també, que $r_{12} = 0$.

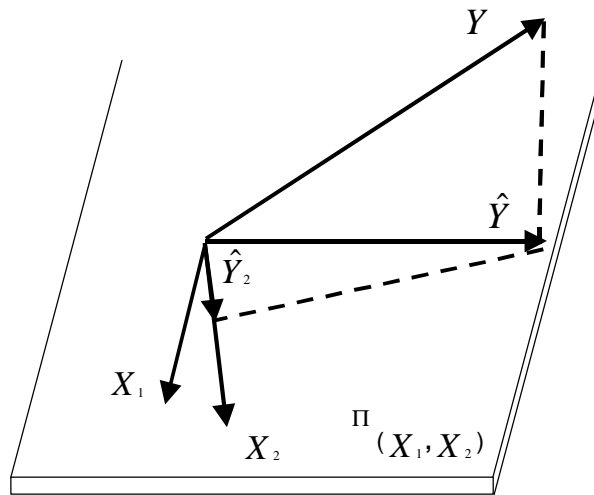


Figura 1.3: Exemple per la desigualtat HRC. $r_{yx_1}^2 = 0$ i $r_{yx_2}^2 > 0$. Cal notar, també, que $r_{12} < 1$.

no tenir la desigualtat HRC en el cas d'una regressió amb dues variables regressores absolutament irrelevantes per descriure la variable resposta. A més, aquest mateix autor, dóna la probabilitat que aparegui la desigualtat HRC en diversos casos, segons diferents valors de r_{12} , β_1 i β_2 , coeficient de correlació entre les dues variables explicatives i els veritables valors dels paràmetres en (1.1) respectivament.

1.4 Interpretació per components principals de la desigualtat HRC

Cuadras [10] interpreta la desigualtat HRC utilitzant l'anàlisi de components principals i demostra que aquesta desigualtat es presenta quan Y està sobretot correlacionada amb les components principals amb variància petita, extretes de la matriu de correlacions R entre les variables explicatives. Aquestes components principals, per tant, poden ser importants per predir Y .

Indiquem $\frac{1}{2}_{yx}^0 = (r_{yx_1}; \dots; r_{yx_p})$. Aleshores

$$r_{yx_1}^2 + \dots + r_{yx_p}^2 = \frac{1}{2}_{yx}^0 \frac{1}{2}_{yx} . \quad (1.5)$$

D'altra banda podem escriure

$$R^2 = \frac{1}{2}_{yx}^0 R^{i-1} \frac{1}{2}_{yx} . \quad (1.6)$$

En efecte, suposant X centrada, tenim

$$\begin{aligned} R^2 &= \text{Corr}^2(Y; \hat{Y}) = \text{Cor}^2(Y; X\mathbf{b}) \\ &= \text{Corr}^2(Y; X(X^0X)^{i-1}X^0Y) \\ &= \frac{\mathbf{1}' \frac{1}{n} Y^0 X (X^0X)^{i-1} X^0 Y}{\sqrt{\mathbf{1}' \frac{1}{n} Y^0 Y} \sqrt{\mathbf{1}' \frac{1}{n} X (X^0X)^{i-1} X^0 Y}} \\ &= \frac{\mathbf{1}' Y^0 X (X^0X)^{i-1} X^0 Y}{(Y^0Y) \mathbf{1}' Y^0 X (X^0X)^{i-1} X^0 Y} \\ &= \frac{Y^0 X (X^0X)^{i-1} X^0 Y}{Y^0 Y} \\ &= \frac{1}{2}_{yx}^0 R^{i-1} \frac{1}{2}_{yx} . \end{aligned}$$

Tenint en compte (1.5) i (1.6) podem escriure la desigualtat (1.4) matricialment en la forma

$$\frac{1}{2}_{yx}^0 R^{i-1} \frac{1}{2}_{yx} > \frac{1}{2}_{yx}^0 \frac{1}{2}_{yx} . \quad (1.7)$$

Cuadras [10] demostra el següent resultat.

Resultat 1 Siguin $Z_1; \dots; Z_p$ les components principals de les variables explicatives obtingudes a partir de la matriu de correlació R: La desigualtat $\frac{1}{2}_{yx}^0 R^{i-1} \frac{1}{2}_{yx} > \frac{1}{2}_{yx}^0 \frac{1}{2}_{yx}$ es verifica si i només si les correlacions simples entre la variable Y i les components principals Z_i amb $\text{var}(Z_i) = \lambda_i < 1$ dominen les correlacions simples entre la variable Y i les components principals Z_i amb

CAPÍTOL 1. UNA DESIGUALTAT DE LA REGRESSIÓ MÚLTIPLE 19

$\text{var}(Z_i) = \sigma_i^2 > 1$ en el sentit que

$$\sum_{\sigma_i < 1} r_{yzi}^2 (1 - \sigma_i^2) > \sum_{\sigma_i > 1} r_{yzi}^2 (\sigma_i^2 - 1):$$

Aquesta condició equival a

$$\sum_{i=1}^p r_{yzi}^2 (1 - \sigma_i^2) > 0.$$

En efecte, obtinguem en primer lloc les components principals de les variables $X_1; \dots; X_p$, variables que suposem estandaritzades sense pèrdua de generalitat ja que només estem interessats en les correlacions. Amb aquesta ...nalitat diagonalitzem la matriu simètrica R ,

$$R = U \alpha U^0, \quad (U^0 U = U U^0 = I_p)$$

on $\alpha = \text{diag}(\sigma_1^2; \dots; \sigma_p^2)$ és la matriu diagonal que conté els valors propis $\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2$ de R i U és la matriu ortogonal que té per columnes els vectors propis corresponents.

La matriu, centrada, XU és la matriu de dades en l'espai de les components principals, que estandaritzada és $Z = XU \alpha^{-1/2}$, tenint en compte que σ_i^2 és la variància de la i -èsima component principal.

Atès que

$$\frac{1}{n} Y^0 X = \frac{1}{n} X^0 Y;$$

$$\frac{1}{n} Y^0 Z = (r_{yz1}, \dots, r_{yzp})^0$$

$$= \frac{1}{n} Z^0 Y;$$

$$R^{-1} = U \alpha^{-1} U^0;$$

la desigualtat (1.7) es pot escriure com

$$\begin{aligned} \frac{1}{n} Y^0 X U_{\alpha} i^{-1} U^0 \frac{1}{n} X^0 Y &> \frac{1}{n} Y^0 X \frac{1}{n} X^0 Y \\ \frac{1}{n} Y^0 X U_{\alpha} i^{-1=2} \alpha i^{-1=2} U^0 \frac{1}{n} X^0 Y &> \frac{1}{n} Y^0 X \frac{1}{n} X^0 Y \\ \frac{1}{n} Y^0 Z \frac{1}{n} (X U_{\alpha} i^{-1=2})^0 Y &> \frac{1}{n} Y^0 X \frac{1}{n} X^0 Y \\ \frac{1}{n} Y^0 Z \frac{1}{n} Z^0 Y &> \frac{1}{n} Y^0 X U_{\alpha} i^{-1=2} \alpha \alpha i^{-1=2} U^0 \frac{1}{n} X^0 Y \\ \frac{1}{n} Y^0 Z \frac{1}{n} Z^0 Y &> \frac{1}{n} Y^0 Z \alpha \frac{1}{n} Z^0 Y , \end{aligned}$$

i, per tant

$$\frac{1}{2} y_z^0 \frac{1}{2} y_z > \frac{1}{2} y_z^0 \alpha \frac{1}{2} y_z ,$$

d'on obtenim

$$\frac{1}{2} y_z^0 (1 - \alpha) \frac{1}{2} y_z \geq 0 ,$$

que equival a

$$\sum_{i=1}^p r_{yzi}^2 (1 - \alpha) \geq 0 . \tag{1.8}$$

1.5 Alguns exemples il·lustratius

En aquesta secció comentem alguns exemples que il·lustren la desigualtat HRC.

Exemple 1.1

Hamilton [39] estudia en extensió un exemple artificial amb $p = 2$, basat en Kendall i Stuart [47], en el que s'obté $R^2 = 0:99992$, $r_{yx_1} = 0:025$, $r_{yx_2} = 0:4341$; la diagonalització de la matriu de correlacions dona els valors propis $\lambda_1 = 1:9$ i $\lambda_2 = 0:1002$ i les correlacions de la variable Y amb les components principals són $r_{yz_1} = 0:22142$ i $r_{yz_2} = 0:97511$.

CAPÍTOL 1. UNA DESIGUALTAT DE LA REGRESSIÓ MÚLTIPLE 21

Amb aquestes dades obtenim

$$\sum_{i=1}^2 r_{yzi}^2 (1 - r_{z_i}^2) = 0:81144 > 0,$$
$$R^2 = 0:99992 > \sum_{i=1}^2 r_{yx_i}^2 = 0:189;$$

verificant-se per tant la desigualtat (1.4), o desigualtat HRC, així com la condició donada en el resultat 1 de la pàgina 19, resumida en l'acompliment de la desigualtat (1.8). S'ha de fer notar que la correlació entre les dues variables explicatives és $r_{x_1x_2} = 0:8998$ i que per tant, malgrat estar fortament correlacionades, no només contenen informació redundant, com demostra el fet d'acomplir-se la desigualtat HRC.

Aquest exemple també ens alerta sobre l'estudi dels diagrames de dispersió entre la variable Y i cadascuna de les variables explicatives X_1 i X_2 .

En les figures 1.4 i 1.5, on es creuen la variable Y amb les components principals, es nota la forta dependència d'aquesta variable Y en la segona component principal. També queda clara la poca variabilitat de la segona component principal comparada amb la de la primera.

Les figures 1.6 i 1.7 mostren com Y no té tendència lineal al llarg de la recta de regressió de X_2 sobre X_1 , en canvi sí que en té en un pla perpendicular a aquesta recta. En canvi al llarg de la recta de regressió de la segona component principal sobre la primera, sí que la variable Y té tendència lineal.

Exemple 1.2

Routledge [58] presenta en detall un exemple real, també per a $p = 2$, en el que (1.4) s'acompleix. Discuteix també els gràfics tridimensionals de la variable Y amb les variables X_1 i X_2 i de la variable Y amb la primera i segona component principal.

Exemple 1.3

Continuant amb el cas $p = 2$, Bertrand i Holder [5] construeixen un exemple amb només 4 dades en el que la desigualtat HRC s'acompleix i citen també un exemple real en el camp de la bioquímica clínica tret de Bertrand, Rudd, Weller i Day [6].

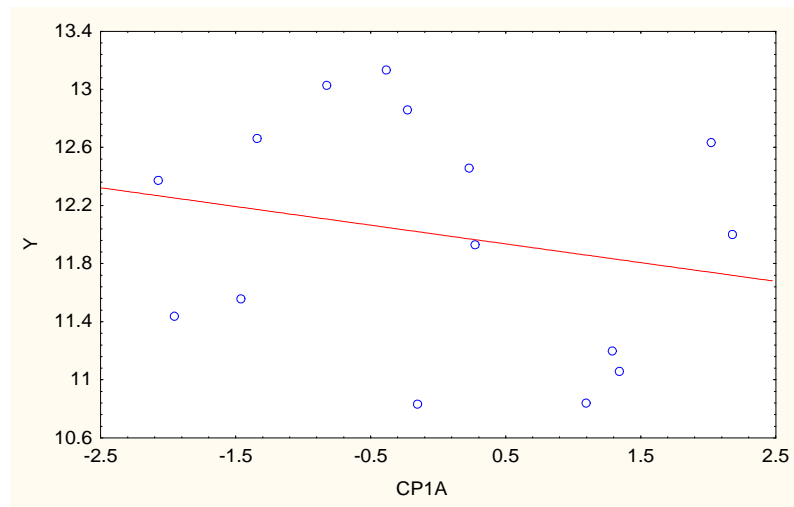


Figura 1.4: Núvol de punts considerant la variable dependent Y i la primera component principal. S'observa la poca relació entre elles i la gran variabilitat de la primera component principal.

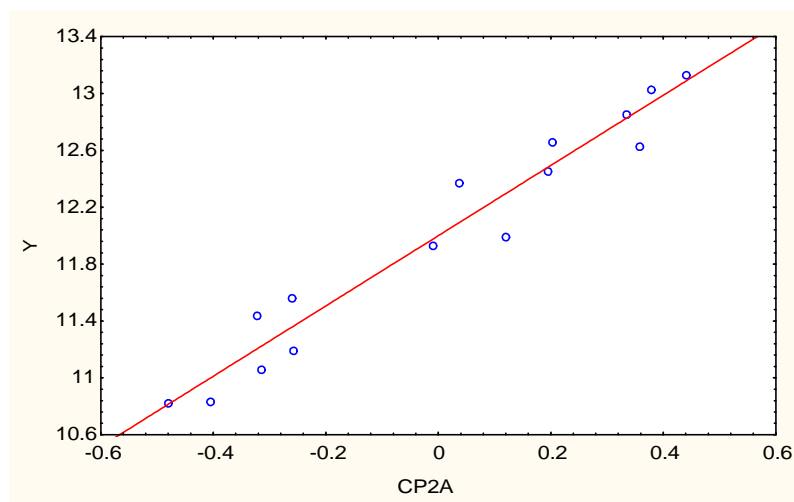


Figura 1.5: Núvol de punts considerant la variable Y i la segona component principal en l'exemple de Hamilton en el que s'observa la clara relació lineal entre elles així com la poca variabilitat de la segona component principal.

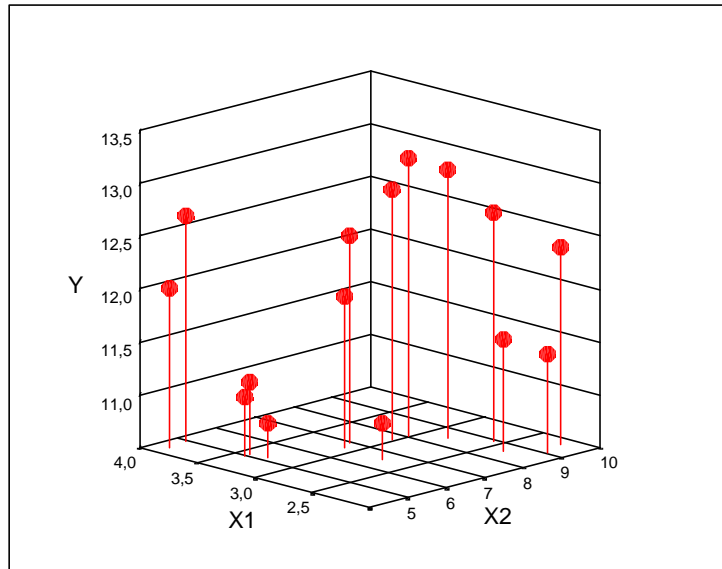


Figura 1.6: Gràfic tridimensional en el que es veu com Y no té tendència lineal al llarg de la recta de regressió de X_2 sobre X_1 .

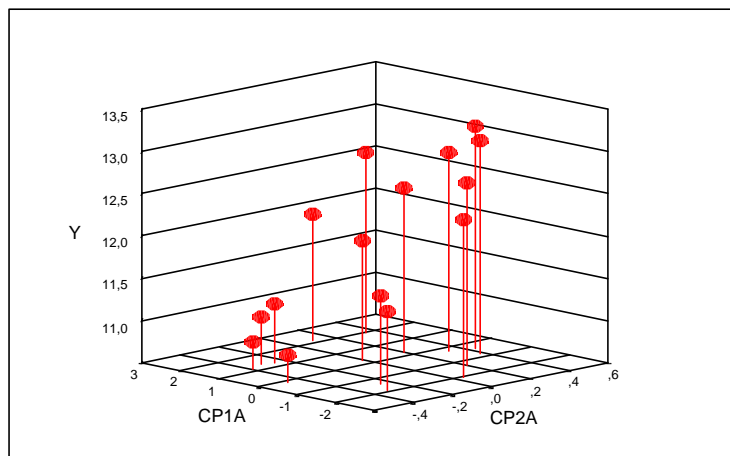


Figura 1.7: Núvol de punts tridimensional en el que es veu com Y té tendència lineal al llarg de la recta de regressió de la segona component principal sobre la primera.

Exemple 1.4

Cuadras [10] estudia un exemple real tret de Jefferies [43] i comentat per Jolliffe [46] i Mardia, Kent i Bibby [51]. A les dades originals, la variable dependent està relacionada amb 13 variables explicatives, però, s'escullen només les variables explicatives X_3, X_6, X_8 i X_{11} trobant-se per tant en el cas $p = 4$. El coeficient de determinació d'aquestes quatre variables és el 95% del coeficient de determinació considerant totes les variables. Els resultats són:

$$\begin{aligned} R^2(4) &= 0.695; \\ r_{yx_1} &= 0.055; \\ r_{yx_3} &= 0.728; \\ r_{yx_6} &= 0.117; \\ r_{yx_8} &= 0.253; \end{aligned}$$

i aleshores la desigualtat (1.4) es verifica.

Les correlacions de la variable Y amb les quatre components principals són

$$0.471; \quad 0.114; \quad 0.045; \quad 0.677;$$

essent

$$\lambda_1 = 1.211; \quad \lambda_2 = 1.059; \quad \lambda_3 = 1.017; \quad \lambda_4 = 0.712;$$

els valors propis de la diagonalització de la matriu R , considerant només les quatre variables esmentades. Es veu, per tant, que la variable Y està més correlacionada amb la quarta component principal, la de variància $\lambda_4 = 0.712$; la més petita. S'acompleix també, com era d'esperar, la desigualtat (1.8)

$$\sum_{i=1}^4 r_{y z_i}^2 (1/\lambda_i) = 0.084389 < 0;$$

condició necessària i suficient a ... de verificar-se la desigualtat HRC.

1.6 Regressió en components principals

Ja comentàvem en la secció 1.1 els problemes que la multicolinealitat entre les variables explicatives $X_1; \dots; X_p$ podia donar en la regressió múltiple. Una

de les maneres d'evitar aquest problema és utilitzar el mètode de la regressió en components principals. Podem veure, per exemple, Jolliffe [46].

Destaquem dues de les principals raons que es presenten per justificar el seu ús. En primer lloc els problemes computacionals deguts a la multicolinealitat de les variables explicatives, problemes que seran superats utilitzant la regressió en components principals al fer servir variables que, per construcció, són incorrelacionades. En segon lloc el fet que totes les variables originals, X_1, \dots, X_p , intervenen en l'estudi, encara que no utilitzem totes les components principals, doncs cadascuna d'aquestes és combinació lineal de les variables explicatives originals. Però precisament és en aquest punt on hi pot haver problemes. El fet que la desigualtat (1.4) sigui una conseqüència de la influència de les components principals amb variància petita ens diu que la regressió en components principals pot fallar també, al eliminar components principals que poden tenir molta importància.

Sobre aquest tema hem de citar, per exemple, Flury [33], que observa que sovint s'utilitzen les primeres components principals i s'ignoren les últimes, és a dir les que tenen variància petita, malgrat que poden ser influents. Quelcom semblant apunten Montgomery and Peck [52]. També trobem, però, autors que ens avisen dels problemes. Així, per exemple, Jolliffe [45] i [46] comenta que a l'hora d'escollir les components hi ha conflicte: d'una banda eliminar les components principals de variància petita per solucionar la multicolinealitat i de l'altra no eliminar les components principals amb alta correlació amb la variable dependent Y . En el seu article dona quatre exemples en els quals les components principals de variància petita són importants a l'hora de predir la variable Y . També Mardia, Kent and Bibby [51] comenten que s'ha de retenir les components principals que tenen més correlació amb la variable dependent. Aquests últims escriuen també que, afortunadament, la tendència de les dades és tal que les components principals amb variància gran són les que expliquen millor la variable Y .

Més recentment Hadi and Ling [38] presenten un exemple en el qual l'última component principal és de fet l'única que va bé per reduir la suma de quadrats residual. En aquest mateix article així com en les puntualitzacions que fa Cuadras [15] al respecte, s'expliquen en quines circumstàncies es poden presentar aquests problemes. Això ho expliquem a la secció següent.

1.7 Influència de l'última component principal

Sigui el model de regressió usual establert per (1.1), matricialment per (1.2). L'estimador OLS de β és

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}.$$

Si les variables estan estandaritzades, podem escriure

$$\mathbf{b} = \mathbf{R}^{-1} \mathbf{1}_{2yx}.$$

Suposem ara que $\mathbf{1}_{2yx} = \frac{1}{n} \mathbf{X}'\mathbf{Y}$ està en la direcció del p -èssim vector propi de \mathbf{R} ,

$$\mathbf{1}_{2yx} = \alpha \mathbf{U}_p; \quad \alpha \neq 0:$$

Aleshores, si $\mathbf{R} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}'$ és la descomposició espectral de la matriu de correlacions, tenim que $\mathbf{R}^{-1} = \mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{U}'$ i per tant

$$\begin{aligned} \mathbf{b} &= \mathbf{U} \left(\sum_{i=1}^p \lambda_i^{-1} \mathbf{U}_i \mathbf{U}_i' \right) (\alpha \mathbf{U}_p) \\ &= \frac{\alpha}{\lambda_p} \mathbf{U}_p; \end{aligned}$$

ja que $\mathbf{U}_p' \mathbf{U}_p = 1$ i $\mathbf{U}_i' \mathbf{U}_p = 0$, per a tot $i \neq p$.

Per tant, si $\mathbf{1}_{2yx}$ està en la direcció del p -èssim vector propi de \mathbf{R} , \mathbf{b} també, i de $\hat{\beta} \cong \mathbf{b}$ podem dir que, aproximadament, també ho estarà $\hat{\beta}$.

Suposem doncs que $\mathbf{U}_p = \mathbf{e}_p$, aleshores

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ &= \mathbf{X}\mathbf{U}\mathbf{U}'\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ &= \mathbf{Z}\mathbf{U}'\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ &= \frac{1}{\lambda_p} \mathbf{Z}\boldsymbol{\beta}_p + \boldsymbol{\varepsilon}, \end{aligned}$$

o també

$$Y = \sum_{s=1}^p Z_s + \epsilon^2,$$

essent Z_p la p -èsima component principal. Aleshores la suma de quadrats residual del model (1.2) és

$$\begin{aligned} R_0^2 &= \sum_{i=1}^n (Y_i - \sum_{s=1}^p Z_{is})^2 \\ &= Y^0 Y - \sum_{i=1}^n \mathbf{b}^0 X_i^0 Y_i \\ &= Y^0 Y - \sum_{s=1}^p Z_s^0 Y, \end{aligned}$$

d'on podem concloure que l'única component principal que reduirà la suma de quadrats residual serà la p -èsima.

Finalment fem notar, que en aquestes condicions

$$R^2 = \sum_{s=1}^p R_s^2 = \sum_{s=1}^p \frac{1}{2} \mathbf{r}_s^0 \mathbf{r}_s^0 = \sum_{s=1}^p \frac{1}{2} \mathbf{r}_s^0 \mathbf{r}_s^0 = \sum_{s=1}^p \frac{1}{2} \mathbf{r}_s^0 \mathbf{r}_s^0$$

atès que $\sum_{s=1}^p \mathbf{r}_s^0 < 1$. Per tant la desigualtat HRC es verifica (Cuadras [15]). D'aquesta manera queda provat el següent resultat.

Resultat 2 Si el vector de correlacions \mathbf{r}_{yx} de la variable Y amb les variables explicatives $X_1; \dots; X_p$; està en la mateixa direcció del p -èsim vector propi de \mathbf{R} , la p -èsima component principal serà l'única que reduirà la suma de quadrats residual. Una condició necessària és que es verifiqui la desigualtat HRC.

1.8 Conclusions

La següent frase de Hamilton [39], podria resumir el capítol:

“No sempre variables correlacionades contenen informació redundant”.

Aquesta frase està implícitament resumida en la desigualtat HRC, (1.7), i ens porta a haver d'anar amb molta cura a l'hora d'utilitzar l'anàlisi de regressió múltiple.

CAPÍTOL 1. UNA DESIGUALTAT DE LA REGRESSIÓ MÚLTIPLE 28

Resumint, alguns d'aquests perills o usos no apropiats de la regressió múltiple són:

- 2 Con...ar en els diagrames de punts X-Y i en les correlacions simples.
Pot ser que $r_{yx_1} \approx r_{yx_2} \approx \dots \approx r_{yx_k} \approx 0$ i en canvi $R^2 \approx 1$. Aquest fenomen apareix quan Y té una forta relació amb les components principals amb variància petita, essent la multicolinealitat una condició suficient però no necessària.
- 2 Descartar variables com a remei per evitar la multicolinealitat.
- 2 Usar el mètode de selecció "forward", el qual ens pot portar a descartar variables amb gran influència conjunta, amb altres variables, sobre Y.
- 2 Ús de la regressió en components principals, al descartar les darreres components principals, les de variància petita, les anomenades "soroll", les quals poden ser importants a l'hora de predir la variable Y:
- 2 Atenció també a l'ús de les tècniques de representació gràfica al llarg dels primers eixos principals, al poder dependre les variables originals a representar, dels últims eixos principals, els quals s'ignoren en la representació per considerar-los com "soroll".

En definitiva: el fet de verificar-se la desigualtat (1.7) adverteix que la variable Y depèn d'aquest "soroll" més que de la part de variabilitat de les dades que estiguem considerant. Tenir en compte aquest fet, pot ajudar a millorar la utilització pràctica de l'anàlisi de regressió múltiple, així com les tècniques de reducció de la dimensió.

Capítol 2

L'efecte d'augmentar les correlacions simples

2.1 Introducció

Siguin $X_1; \dots; X_p; Y$ v.a.'s amb esperances i variàncies finites, definides sobre un mateix espai de probabilitats. En aquest capítol abandonem el model de regressió i ens dediquem a estudiar les correlacions, múltiple i simples, entre Y i $X_1; \dots; X_p$.

Començarem suposant que les correlacions simples són

$$\begin{aligned} r_{yx_i} &> 0; & i = 1; \dots; s; \\ r_{yx_i} &= 0; & i = s + 1; \dots; p; \end{aligned}$$

És a dir, Y només està correlacionada amb s , on $s < p$, de les p variables explicatives i aquesta correlació, a més, la considerem positiva.

Iniciem el capítol amb un altre resultat sorprenent i aparentment poc intuïtiu com és el fet següent, observat per Tiit [62]. Fixat el nombre de variables explicatives i mantenint constant la matriu R , encara que augmentem el nombre s de variables correlacionades positivament amb Y , el coeficient de determinació no té per què augmentar, com semblaria normal al augmentar la influència de les variables explicatives sobre Y : També trobarem exemples per a la desigualtat HRC estudiada en el capítol 1.

2.2 Variables equicorrelacionades

Suposem que les variables explicatives $X_1; \dots; X_p$ estan equicorrelacionades, cas que té interès en certs problemes d'anàlisi multivariant. A més a més, suposem que la correlació de la variable Y amb cadascuna de les variables explicatives $X_1; \dots; X_s$ és 0 ó constant. És a dir

$$R = \begin{pmatrix} 0 & 1 & c & \dots & c \\ c & 1 & \dots & \dots & c \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ c & c & \dots & \dots & 1 \end{pmatrix}; \quad \text{amb } \frac{1}{p_i - 1} < c < 1; \quad (2.1)$$

per garantir que la matriu R sigui definida positiva, i

$$\frac{1}{2}_{yx} = \frac{1}{2}_s = \begin{pmatrix} 0 & r \\ r & 1 \\ \vdots & \vdots \\ 0 & c \\ \vdots & \vdots \\ 0 & c \end{pmatrix}; \quad (2.2)$$

amb $r > 0$.

En aquestes condicions tenim que

$$\frac{1}{2}_s^0 \frac{1}{2}_s = sr^2;$$

$$R^2(s) = \frac{1}{2}_s^0 R^{-1} \frac{1}{2}_s, \quad (2.3)$$

$$= \frac{sr^2 [1 + (p_i - s_i - 1)c]}{[1 + (p_i - 1)c] (1_i - c)},$$

funció que no és creixent en s .

Tiit [62] i [63] i posteriorment Cuadras [12] estudien la possibilitat que es presentin dos resultats sorprenents:

$$R^2(s) > sr^2 \quad \text{i} \quad R^2(s_1) > R^2(s_2) \quad \text{amb } s_1 < s_2.$$

El primer resultat, $R^2(s) > sr^2$, ja ha estat estudiat en el capítol 1. És la desigualtat HRC, de la qual trobarem aquí altres exemples. El segon

CAPÍTOL 2. L'EFECTE D'AUGMENTAR LES CORRELACIONS SIMPLES 31

resultat, que estudiarem més endavant, expressa el fet que augmentant les correlacions amb la variable resposta no té per què augmentar el coeficient de determinació R^2 . Seguint Cuadras [12] donarem una explicació utilitzant components principals.

2.3 Exemples per a la desigualtat HRC

Exemple 2.1

Suposem el model (2.1) i (2.2) amb $p = 8$, $c = 0.5$ i $r = 0.4$. És ben sabut, Cuadras [14], que per una matriu de correlacions de la forma (2.1) els valors propis són

$$\lambda_1 = 1 + (p-1)c; \lambda_2 = \dots = \lambda_p = 1 - c, \quad (2.4)$$

amb vectors propis respectius

$$u_1 = \begin{pmatrix} 0 \\ \frac{1}{p} \\ \vdots \\ \frac{1}{p} \end{pmatrix}, u_2 = \begin{pmatrix} 0 \\ \frac{1}{2} \\ \frac{1}{2} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, u_3 = \begin{pmatrix} 0 \\ \frac{1}{6} \\ \frac{1}{6} \\ \frac{2}{6} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, u_p = \begin{pmatrix} 0 \\ \frac{1}{p(p-1)} \\ \vdots \\ \frac{1}{p(p-1)} \\ \frac{p-1}{p(p-1)} \end{pmatrix}. \quad (2.5)$$

El vector de correlacions de la variable resposta Y amb les components principals és

$$\frac{1}{2} \gamma_Z = \alpha \frac{1}{2} U^0 \frac{1}{2} \gamma_{YX}, \quad (2.6)$$

on U és la matriu que té a les columnes els vectors propis u_i .

Apliquem aquests resultats teòrics a l'exemple que ens ocupa.

$$\lambda_1 = 4.5; \lambda_2 = \dots = \lambda_8 = 0.5,$$

$$u_1 = \begin{pmatrix} 0 \\ \frac{1}{8} \\ \vdots \\ \frac{1}{8} \end{pmatrix}, u_2 = \begin{pmatrix} 0 \\ \frac{1}{2} \\ \frac{1}{2} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, u_3 = \begin{pmatrix} 0 \\ \frac{1}{6} \\ \frac{1}{6} \\ \frac{2}{6} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, u_8 = \begin{pmatrix} 0 \\ \frac{1}{56} \\ \vdots \\ \frac{1}{56} \\ \frac{7}{56} \end{pmatrix}.$$

CAPÍTOL 2. L'EFECTE D'AUGMENTAR LES CORRELACIONS SIMPLES 32

$\frac{1}{2}y_x$	R^2		$r_{yx_i}^2$	$\sum_{i=1}^8 r_{yz_i}^2 (1 - \rho_i)$
(0:4; 0; :::; 0)	0:2844	>	0:16	0:124
(0:4; 0:4; 0; :::; 0)	0:4978	>	0:32	0:170
(0:4; 0:4; 0:4; :::; 0)	0:64	>	0:48	0:160
(0:4; :::; 0:4; 0; 0; 0; 0)	0:7111	>	0:64	0:071
(0:4; :::; 0:4; 0; 0; 0)	0:7111	<	0:8	i 0:089
(0:4; :::; 0:4; 0; 0)	0:64	<	0:96	i 0:320
(0:4; :::; 0:4; 0)	0:4978	<	1:12	j 0:622
(0:4; :::; 0:4)	0:2844	<	1:28	j 0:996

Taula 2.1: Exemple amb variables equicorrelacionades. Avaluació de la desigualtat HRC i la condició d'acompliment.

$\frac{1}{2}y_z$
(0:0667 0:4000 0:2309 0:1633 0:1265 0:1033 0:0873 0:0756)
(0:1333 0 0:4619 0:3266 0:2530 0:2066 0:1746 0:1512)
(0:2 0 0 0:4899 0:3795 0:3098 0:2619 0:2268)
(0:2666 0 0 0 0:5060:41310:34910:3024)
(0:3333 0 0 0 0 0:5164 0:4364 0:3780)
(0:4 0 0 0 0 0 0:5237 0:4536)
(0:4667 0 0 0 0 0 0 0:5292)
(0:5333 0 0 0 0 0 0 0)

Taula 2.2: Exemple amb variables equicorrelacionades. Influència de les components principals.

En la taula 2.1 donem els valors de $\frac{1}{2}y_x$, R^2 , $\sum_{i=1}^8 r_{yx_i}^2$ i $\sum_{i=1}^8 r_{yz_i}^2 (1 - \rho_i)$ en funció del nombre de variables correlacionades amb Y, s. També donem, en la taula 2.2, els vectors $\frac{1}{2}y_z$ segons s.

En els primers quatre casos s'acompleix tant la desigualtat HRC com la desigualtat (1.8), $\sum_{i=1}^8 r_{yz_i}^2 (1 - \rho_i) > 0$, és a dir, hi ha més influència de les components principals amb variància petita. Per tant podem comprovar, un altre cop, com la desigualtat HRC s'acompleix sempre i quan la condició (1.8) es veri...qui. Observant $\frac{1}{2}y_z$ es pot veure també com l'anterior a...rmació va relacionada amb la importància de les components principals de variància més petita d'1, de fet són totes tret de la primera.

CAPÍTOL 2. L'EFECTE D'AUGMENTAR LES CORRELACIONS SIMPLES 33

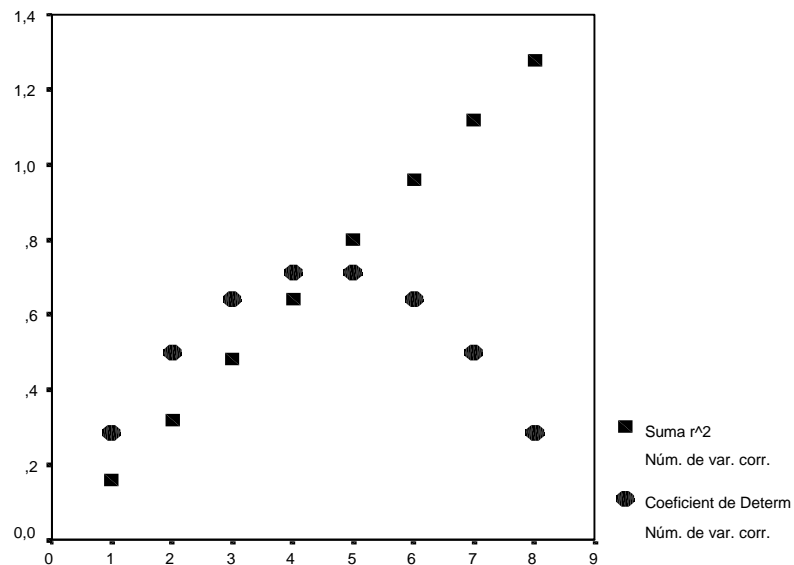


Figura 2.1: Relació entre el nombre de variables correlacionades amb Y i el coeficient de determinació d'una banda i la suma de $r_{yx_i}^2$ de l'altra.

És aclaridor el gràfic de la figura 2.1 en el qual, a part del que acabem de dir, s'observa l'altre resultat sorprenent i poc intuïtiu del qual parlàvem en l'apartat 2.1: encara que augmentem el nombre de variables correlacionades positivament amb Y, de 4 a 5, el coeficient de determinació no augmenta. Si de 5 variables correlacionades amb Y passem a 6, de 6 a 7 o de 7 a 8, aleshores el coeficient de determinació no només no augmenta sinó que disminueix.

Exemple 2.2

Considerem el següent exemple inspirat en Tiit [63]. Tirem un dau i estudiem els següents esdeveniments $A_1 = \{1; 4; 6\}$, $A_2 = \{2; 5; 6\}$ i $A_3 = \{3; 6\}$. Tenim per tant que les probabilitats són: $P(A_i) = \frac{1}{3}$; $i = 1; 2; 3$. Sigui ara el vector aleatori $(X_1; X_2; X_3)$ on cada X_i és igual a l'indicador de l'esdeveniment A_i . Així tenim

$$P(X_i = 1) = \frac{1}{3}; \quad P(X_i = 0) = \frac{2}{3}; \quad i = 1; 2; 3:$$

CAPÍTOL 2. L'EFECTE D'AUGMENTAR LES CORRELACIONS SIMPLES 34

A més a més per a tot i es té

$$E[X_i] = \frac{1}{3}; \text{Var}[X_i] = \frac{2}{9};$$

i per a tot $i \neq j$

$$\text{Cov}[X_i; X_j] = i \frac{1}{9}.$$

Tenim per tant, fent $X_3 = Y$,

$$R = \begin{pmatrix} \tilde{A} & & \\ & 1 & i \frac{1}{2} \\ & i \frac{1}{2} & 1 \end{pmatrix}; \quad \frac{1}{2}_{yx} = \begin{pmatrix} \tilde{A} & \\ & i \frac{1}{2} \end{pmatrix};$$

és a dir, un cas obvi del model de regressió amb variables equicorrelacionades amb $s = p = 2$ i $r = c = i \frac{1}{2}$. En aquestes condicions i per (2.3) tenim

$$R^2 = 1 > \frac{1}{2} \frac{1}{2} = sr^2 = \frac{1}{2};$$

verificant-se la desigualtat HRC.

En general, si suposem $s = p$ i $r = c = i \frac{1}{p}$ tenim, per la fórmula (2.3),

$$R^2 = 1 > \frac{1}{s} \frac{1}{s} = sr^2 = \frac{1}{p};$$

és a dir, s'acompleix la desigualtat HRC.

2.4 Desigualtat HRC forta

Com una generalització de l'exemple comentat en l'apartat 2.3, Tiit [63] enuncia el següent resultat, que podem presentar com una versió forta de la desigualtat HRC.

Resultat 3 En les condicions de l'apartat 2.2, amb $s = p$ i $c = r$, donada una constant k , $1 < k < p$, podem trobar un vector aleatori Y, X_1, \dots, X_p tal que

$$\frac{R^2}{\prod_{i=1}^p r_{yx_i}^2} = k;$$

CAPÍTOL 2. L'EFECTE D'AUGMENTAR LES CORRELACIONS SIMPLES 35

En efecte, cal prendre

$$R = \begin{pmatrix} 1 & \frac{1_i k}{k(p_i - 1)} & \cdots & \frac{1_i k}{k(p_i - 1)} \\ \frac{1_i k}{k(p_i - 1)} & 1 & \cdots & \frac{1_i k}{k(p_i - 1)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1_i k}{k(p_i - 1)} & \frac{1_i k}{k(p_i - 1)} & \cdots & 1 \end{pmatrix},$$

$$\frac{1}{2}y_x = \begin{pmatrix} \frac{1_i k}{k(p_i - 1)} \\ \vdots \\ \frac{1_i k}{k(p_i - 1)} \end{pmatrix} A,$$

verificant la condició donada en l'equació (2.1). En aquest cas, de (2.3) i fent $s = p$; $c = r$; tenim

$$R^2 = \frac{1}{2}y_x^0 R^i \frac{1}{2}y_x$$

$$= \frac{pr^2}{1 + (p - 1)r^2};$$

$$\frac{1}{2}y_x^0 \frac{1}{2}y_x = pr^2;$$

i per tant

$$\frac{R^2}{\sum_{i=1}^p r_{yxi}^2} = \frac{1}{1 + (p - 1)r^2},$$

expressió, que tenint en compte $r = \frac{1_i k}{k(p_i - 1)}$ resulta ser

$$\frac{R^2}{\sum_{i=1}^p r_{yxi}^2} = k,$$

que és el que volíem demostrar.

Observem que en aquest cas la variable Y està relacionada amb la component principal de variància més petita d'1 i en canvi no ho està amb les components principals de variància més gran d'1, verificant-se

$$\sum_{i=1}^p r_{yzi}^2 (1 - \lambda_i) > 0;$$

com es pot comprovar tenint en compte que de (2.4) obtenim

$$\lambda_1 = \lambda_2 = \dots = \lambda_{p-1} = \frac{k(p-1)}{k(p-1)} = 1; \quad \lambda_p = \frac{1}{k} < 1;$$

amb vectors propis els mateixos de (2.5). Cal notar que hem ordenat els valors propis en ordre decreixent. Per tant el vector de correlacions de la variable Y amb les components principals és, de (2.6),

$$\mathbf{r}_{yz} = (0; \dots; 0; \frac{p(k-1)}{k(1-p)} \sqrt{\frac{p}{k}}).$$

Tiit [63] generalitza aquesta situació fent que el vector \mathbf{r}_{yx} sigui $\mathbf{r}_{yx} = (r_{yx_1}, \dots, r_{yx_p})^0$ amb $\sum_{i=1}^p r_{yx_i} = 0$, donant en aquest cas

$$R^2 = \frac{1}{1-c} \sum_{i=1}^p r_{yx_i}^2;$$

i per tant

$$\frac{R^2}{\sum_{i=1}^p r_{yx_i}^2} = \frac{1}{1-c};$$

2.5 Augmentant les correlacions simples no té per què augmentar R^2

En aquesta secció comprovem que l'ajust de la regressió no necessàriament millora afegint variables explicatives correlacionades amb la variable Y; o augmentant les correlacions simples amb Y.

Continuant amb les hipòtesis del model establert per (2.1) i (2.2), el que acabem de dir és, de fet, el segon resultat que presentàvem en l'apartat 2.1,

$$R^2(s_1) > R^2(s_2) \text{ amb } s_1 < s_2;$$

i per al qual hem trobat un exemple en l'apartat 2.3. No és molt clar, com en aquest exemple podem veure, que aquest fet sigui una conseqüència de la influència de les components principals de variància més petita d'1 sobre la variable resposta Y:

Podem dir, en canvi, Cuadras [12], que per $s_2 = p$, Y està correlacionada solament amb la primera component principal i $R^2(s_2) = R^2(s)$ per a qual-sevol s; vegeu taules 2.1 i 2.2.

CAPÍTOL 2. L'EFECTE D'AUGMENTAR LES CORRELACIONS SIMPLES 37

Estudiem teòricament aquest fet. El coeficient de determinació donat en (2.3) es pot descomposar en la forma

$$R^2(s) = r^2 [A(s) + B(s)] ,$$

on

$$A(s) = \frac{s^2}{p[1 + (p_i - 1)c]} > 0;$$

$$B(s) = \frac{s^2}{1 - c} \left[\frac{\mu_1}{s} + \frac{1}{p} \right] > 0;$$

Considerem dos casos: primer, $0 < c < 1$, variables explicatives equicorrelacionades positivament, i segon, $-(p_i - 1) < c < 0$, variables explicatives equicorrelacionades negativament.

Considerem $0 < c < 1$. $A(s)$ és funció creixent en s , és a dir, $A(s_1) < A(s_2)$, amb $s_1 < s_2$, estant relacionat el seu valor amb la variància, $1 + (p_i - 1)c > 1$, de la primera component principal.

$B(s)$ està relacionada amb les darreres components principals, les de variància $1 - c < 1$.

$B(s)$ és creixent entre $s = 0$ i s igual a l'enter més proper al valor $\frac{1 + (p_i - 1)c}{2c}$ i és decreixent entre aquest enter i l'enter, no superior a p , més proper a $\frac{1 + (p_i - 1)c}{c}$.

Segui ara $-(p_i - 1) < c < 0$. $A(s)$ és funció creixent igual que abans, però ara està relacionada amb la variància de la darrera component principal, l'única més petita d'1.

$B(s)$ està relacionada amb les primeres components principals i és decreixent sempre.

En definitiva, per $0 < s_1 < s_2 \leq p$, sempre $A(s_1) < A(s_2)$, però pot ser $B(s_1) > B(s_2)$ i per tant també pot ser $R^2(s_1) > R^2(s_2)$.

Ataquem ara el problema d'una altra manera. En l'exemple citat de l'apartat 2.3, ...gura 2.1, veiem com $R^2(s)$ va augmentant, per disminuir a partir d'un cert s . Podem dir que és quan les components principals de variància petita perden pes en el vector de correlacions $\frac{1}{2}y_z$.

CAPÍTOL 2. L'EFECTE D'AUGMENTAR LES CORRELACIONS SIMPLES 38

Estudiem en l'esquema actual el signe de l'expressió

$$\sum_{i=1}^p r_{yzi}^2 (1 - s_i);$$

el qual donarà resposta a la pregunta que ens acabem de fer.

$$\begin{aligned} \sum_{i=1}^p r_{yzi}^2 (1 - s_i) &= \sum_{i=1}^p \frac{sr^2}{p[1 + (p_i - 1)c]} [1 + (p_i - 1)c] \\ &+ \sum_{i=s+1}^p \frac{sr^2}{i(i-1)(1-c)} [1 - (1-c)] \\ &= s^2 r^2 \frac{i(p_i - 1)c}{p[1 + (p_i - 1)c]} + \frac{c}{1 - c} \sum_{i=1}^p \frac{1}{i} \\ &= sr^2 \frac{c[1 - s + c(p_i - 1)]}{(1 - c)[1 + (p_i - 1)c]}, \end{aligned}$$

de tal manera que el signe de l'expressió $\sum_{i=1}^p r_{yzi}^2 (1 - s_i)$ és el mateix que el de l'expressió

$$i - cs + c + (p_i - 1)c^2;$$

la qual cosa ens diu que serà $\sum_{i=1}^p r_{yzi}^2 (1 - s_i) > 0$ quan

$$s < 1 + (p_i - 1)c;$$

si c és positiu i per tot s si c és negatiu.

Estudiem, per exemple, el cas $c = 0.5$. Tenim

$$\frac{1 + (p_i - 1)c}{2c} = 1 + (p_i - 1)c;$$

és a dir, $B(s)$ assolirà el seu màxim¹ per al mateix s per al qual l'expressió $\sum_{i=1}^p r_{yzi}^2 (1 - s_i)$ canvia a negativa, en altres paraules, quan les components principals de variància més petita d'1 perden pes. En la taula 2.3 podem veure, per diferents valors de p i diferents valors de $c > 0$, per quins valors de s , $R^2(s)$ arriba al seu màxim i per quins valors de s l'expressió $\sum_{i=1}^p r_{yzi}^2 (1 - s_i)$ passa a ser negativa.

¹No només $B(s)$, si no també $R^2(s)$ si s'estudia directament la funció $R^2(s) = \frac{sr^2[1 + (p_i - 1)c]}{[1 + (p_i - 1)c](1 - c)}$.

CAPÍTOL 2. L'EFECTE D'AUGMENTAR LES CORRELACIONS SIMPLES 39

	$p = 2$	$p = 2$	$p = 4$	$p = 4$
c	$\max R^2(s)$	$\sum_{i=1}^p r_{yzi}^2 (1 - \rho_{zi}) < 0$	$\max R^2(s)$	$\sum_{i=1}^p r_{yzi}^2 (1 - \rho_{zi}) < 0$
0:1	s = 2	s = 2	s = 4	s = 2
0:2	s = 2	s = 2	s = 4	s = 2
0:3	s = 2	s = 2	s = 3	s = 2
0:4	s = 2	s = 2	s = 3	s = 3
0:5	s = 1 i 2	s = 2	s = 2 i 3	s = 3
0:6	s = 1	s = 2	s = 2	s = 3
0:7	s = 1	s = 2	s = 2	s = 4
0:8	s = 1	s = 2	s = 2	s = 4
0:9	s = 1	s = 2	s = 2	s = 4

	$p = 8$	$p = 8$	$p = 10$	$p = 10$
c	$\max R^2(s)$	$\sum_{i=1}^p r_{yzi}^2 (1 - \rho_{zi}) < 0$	$\max R^2(s)$	$\sum_{i=1}^p r_{yzi}^2 (1 - \rho_{zi}) < 0$
0:1	s = 8	s = 2	s = 9 i 10	s = 2
0:2	s = 6	s = 3	s = 7	s = 3
0:3	s = 5	s = 4	s = 6	s = 4
0:4	s = 5	s = 4	s = 6	s = 5
0:5	s = 4 i 5	s = 5	s = 5 i 6	s = 6
0:6	s = 4	s = 6	s = 5	s = 7
0:7	s = 4	s = 6	s = 5	s = 8
0:8	s = 4	s = 7	s = 5	s = 9
0:9	s = 4	s = 8	s = 5	s = 10

Taula 2.3: Valors de s pels quals $R^2(s)$ arriba al seu màxim i pels quals l'expressió $\sum_{i=1}^p r_{yzi}^2 (1 - \rho_{zi})$ passa a ser negativa, per diferents valors de p i de $c > 0$.

2.6 Una generalització

Podem fer una generalització del fet que augmentant les correlacions simples no augmenti la múltiple. Siguin Y_1 i Y_2 dues variables resposta i X_1, \dots, X_p variables explicatives. És possible tenir d'una banda

$$R_{Y_1|X}^2 = \sum_{j=1}^p r_{Y_1 X_j}^2 > R_{Y_2|X}^2 = \sum_{j=1}^p r_{Y_2 X_j}^2; \quad (2.7)$$

és a dir, Y_1 està aparentment més correlacionada amb les variables explicatives que Y_2 , i d'altra banda

$$R_1^2 = R_{Y_1|X}^2 < R_2^2 = R_{Y_2|X}^2; \quad (2.8)$$

fet que és degut a la influència sobre Y_i de les components principals amb variància petita.

Com abans, caldria matisar aquest fet en el sentit que el que passa és que Y_2 està més influenciada per les components principals de variància petita que Y_1 . També cal notar que la situació expressada en les dues desigualtats anteriors (2.7) i (2.8) es pot donar tret del cas en què

$$R_1^2 > \sum_{j=1}^p r_{Y_1 X_j}^2 \quad \text{i} \quad R_2^2 < \sum_{j=1}^p r_{Y_2 X_j}^2$$

a la vegada, la qual cosa ens diu, mirant el compliment de la desigualtat HRC, que hi ha influència de les components principals de variància petita en Y_1 però no en Y_2 (Cuadras [12]).

Justi...quem per què la situació plantejada per les desigualtats (2.7) i (2.8) es pugui presentar i donem condicions en termes de les components principals. Suposem

$$\mathbf{y}_{yx} = \begin{bmatrix} r_{y x_1} \\ \vdots \\ r_{y x_p} \end{bmatrix} = \mathbf{R} \mathbf{u},$$

amb \mathbf{u} unitari, on

$$\mathbf{R} = \begin{bmatrix} \tilde{\mathbf{A}} \\ \vdots \\ \mathbf{r}_{y x_i} \end{bmatrix} \mathbf{I}^{\frac{1}{2}} = \begin{bmatrix} \mathbf{I} \\ \vdots \\ \mathbf{r}_{y x_i} \end{bmatrix} \mathbf{I}^{\frac{1}{2}};$$

CAPÍTOL 2. L'EFECTE D'AUGMENTAR LES CORRELACIONS SIMPLES 41

Siguin U_1, \dots, U_p els vectors propis de la matriu R . Podem escriure $u = \sum_{i=1}^p a_i U_i$, al ser $U_1; \dots; U_p$ una base de l'espai \mathbb{R}^p . En aquestes condicions

$$\begin{aligned}
 R^2 &= \frac{1}{2} {}_{yx}^0 R^i \frac{1}{2} {}_{yx} \\
 &= \sum_{i=1}^p a_i U_i R^i \sum_{i=1}^p a_i U_i \\
 &= \sum_{i=1}^p a_i U_i R^i \sum_{i=1}^p a_i U_i \\
 &= \sum_{i=1}^p a_i U_i \sum_{i=1}^p a_i U_i \\
 &= \sum_{i=1}^p a_i^2 U_i
 \end{aligned}$$

D'altra banda

$$\frac{1}{2} {}_{yz} = \sum_{i=1}^p a_i \frac{1}{2} U_i \frac{1}{2} {}_{yx} \quad (2.9)$$

$$\begin{aligned}
 &= \sum_{i=1}^p a_i \frac{1}{2} U_i \frac{1}{2} {}_{yx} \\
 &= \sum_{i=1}^p a_i \frac{1}{2} U_i \frac{1}{2} {}_{yx}
 \end{aligned}$$

Si

$$\frac{1}{2} {}_{y_1x} = \sum_{i=1}^p a_i U_i,$$

amb u proper a la primera component principal, és a dir $a_1 > 0$ i $a_i \geq 0$ per tot $i = 2; \dots; p$;

$$\frac{1}{2} {}_{y_2x} = \sum_{i=1}^p b_i V_i,$$

CAPÍTOL 2. L'EFECTE D'AUGMENTAR LES CORRELACIONS SIMPLES 42

amb v no tan proper a les primeres components principals, podem trobar²

$$R_1^2 = \sum_{i=1}^p \frac{a_i^2}{\lambda_i} < \sum_{i=1}^p \frac{b_i^2}{\lambda_i} = R_2^2,$$

i també

$$\frac{1}{2}_{y_1x}^0 \frac{1}{2}_{y_1x} = \sum_{i=1}^p \lambda_i^{-1} > \sum_{i=1}^p \lambda_i^{-1} = \frac{1}{2}_{y_2x}^0 \frac{1}{2}_{y_2x},$$

fent $\sum_{i=1}^p \lambda_i^{-1} > \sum_{i=1}^p \lambda_i^{-1}$. És a dir, es verifica (2.7) i (2.8) i això ha estat possible gràcies a que Y_2 està més influenciada per les components principals de variància petita que Y_1 , com es pot veure tenint en compte (2.9).

Exemple 2.3

Considerem $p = 4$ variables explicatives $X_1; \dots; X_4$ amb matriu de correlacions

$$R = \begin{pmatrix} 1 & 0.31 & 0.4 & 0.52 \\ 0.31 & 1 & 0.52 & 0.4 \\ 0.4 & 0.52 & 1 & 0.31 \\ 0.52 & 0.4 & 0.31 & 1 \end{pmatrix},$$

i dues variables Y_1 i Y_2 amb vectors de correlacions

$$\frac{1}{2}_{y_1x} = \begin{pmatrix} 0.6 \\ 0.5 \\ 0.4 \\ 0.3 \end{pmatrix} \text{ i } \frac{1}{2}_{y_2x} = \begin{pmatrix} 0.6 \\ 0.5 \\ 0.1 \\ 0.1 \end{pmatrix}$$

que indica clarament que Y_1 està aparentment més correlacionada amb la variable explicativa que Y_2 .

Fent els càlculs adients obtenim

$$R_1^2 = \frac{1}{2}_{y_1x}^0 R^{-1} \frac{1}{2}_{y_1x} = 0.4833;$$

$$R_2^2 = \frac{1}{2}_{y_2x}^0 R^{-1} \frac{1}{2}_{y_2x} = 0.7162;$$

$$\frac{1}{2}_{y_1x}^0 \frac{1}{2}_{y_1x} = \sum_{j=1}^p r_{y_1x_j}^2 = 0.86;$$

$$\frac{1}{2}_{y_2x}^0 \frac{1}{2}_{y_2x} = \sum_{j=1}^p r_{y_2x_j}^2 = 0.63;$$

²Donat $\sum_{i=1}^p \lambda_i^{-1}$, el mínim valor de R^2 es troba quan $\frac{1}{2}_{yx} = \sum_{i=1}^p \lambda_i^{-1} U_1$ que implica $R^2 = \frac{\sum_{i=1}^p \lambda_i^{-1}}{\sum_{i=1}^p \lambda_i^{-1}}$. És a dir, quan Y està correlacionada amb la primera component principal.

CAPÍTOL 2. L'EFECTE D'AUGMENTAR LES CORRELACIONS SIMPLES 43

Tenim per tant el següent esquema

$$R_1^2 < \sum_{j=1}^p r_{y_1 x_j}^2$$

$$R_2^2 > \sum_{j=1}^p r_{y_2 x_j}^2$$

altrament dit

$$R_1^2 < \sum_{j=1}^p r_{y_2 x_j}^2 < R_2^2 < \sum_{j=1}^p r_{y_1 x_j}^2$$

posant de manifest que per a Y_2 la influència de les components principals de variància petita és gran, al veure-se la desigualtat HRC. En tot cas, més influència que per Y_1 , per a la qual no es veu la desigualtat HRC.

També ho podem veure avaluant l'expressió (1.8). En efecte: els valors propis de la matriu R són 0:23, 0:81, 0:57 i 0:39, i per tant

$$\frac{1}{2} y_{1z} = \sum_{i=1}^p \alpha_i \frac{1}{2} U_i^0 y_{1x} = \begin{matrix} 0 & 1 \\ \alpha_1 & 0:60264 \\ \alpha_2 & 0:00000 \\ \alpha_3 & 0:13245 \\ \alpha_4 & 0:32026 \end{matrix} \begin{matrix} C \\ C \\ C \\ A \end{matrix};$$

$$\frac{1}{2} y_{2z} = \sum_{i=1}^p \alpha_i \frac{1}{2} U_i^0 y_{2x} = \begin{matrix} 0 & 1 \\ \alpha_1 & 0:43524 \\ \alpha_2 & 0:05556 \\ \alpha_3 & 0:06622 \\ \alpha_4 & 0:72059 \end{matrix} \begin{matrix} C \\ C \\ C \\ A \end{matrix};$$

que implica

$$\sum_{i=1}^p \alpha_i^2 r_{y_1 z}^2 (1 - \alpha_i^2) = \frac{1}{2} y_{1z}^0 (I_p - \alpha) \frac{1}{2} y_{1z} = 0:3766;$$

$$\sum_{i=1}^p \alpha_i^2 r_{y_2 z}^2 (1 - \alpha_i^2) = \frac{1}{2} y_{2z}^0 (I_p - \alpha) \frac{1}{2} y_{2z} = 0:0862.$$

Aquest exemple està basat en Cuadras [12], el qual dona un segon exemple que il·lustra la idea que tant el mètode de selecció forward com el mètode de la regressió en components principals poden no ser adients.

CAPÍTOL 2. L'EFECTE D'AUGMENTAR LES CORRELACIONS SIMPLES 44

Capítol 3

La paradoxa de Simpson

3.1 Introducció

La paradoxa de Simpson [60] no deixa de ser sorprenent per molt coneguda que sigui i estudiada que hagi estat. Ens alerta contra l'anàlisi de dades de taules de contingència, especialment quan les dades no provenen de grups homogenis.

Aquesta paradoxa consisteix en que pot passar que en cada subpoblació un esdeveniment B inïueixi favorablement en un altre esdeveniment A i, en canvi, en el total de la població, B inïueixi desfavorablement en A.

Comencem aquest capítol comentant dos exemples per il·lustrar la paradoxa. A continuació inclourem una breu descripció de l'anàlisi de correspondències per, ...nalment, aplicar aquesta tècnica a l'estudi d'aquesta paradoxa, incidint especialment en l'aspecte grà...c.

En el capítol 1 interpretàvem la sorprenent desigualtat $R^2 > \sum_{i=1}^p r_{y_i}^2$ utilitzant components principals. El que fem en aquest capítol és utilitzar l'anàlisi de correspondències per estudiar grà...cament les taules de contingència.

3.2 Exemples de la paradoxa de Simpson

Siguin A i B dos esdeveniments. Pot passar simultàniament que tant en la població C com en la població C⁰, l'esdeveniment B sigui favorable a l'esdeveniment A i, en canvi, deixant de banda la població, l'esdeveniment B

	C	C	C ⁰	C ⁰
	B ⁰	B	B ⁰	B
Mort (A ⁰)	950	9000	5000	5
Curat (A)	50	1000	5000	95

Taula 3.1: Paradoxa de Simpson. Tant en la població C com en la C⁰, és millor el nou tractament B.

sigui desfavorable a l'esdeveniment A. És a dir

$$P(A=BC) > P(A=B^0C) \text{ i } P(A=BC^0) > P(A=B^0C^0) ,$$

i en canvi

$$P(A=B) < P(A=B^0) ,$$

essent B⁰ l'esdeveniment contrari a l'esdeveniment B.

És la paradoxa de Simpson.

Cosiderem un parell d'exemples arti...cials però versemblants.

Exemple 3.1

Colin R. Blyth [7] mostra el següent exemple.

Un metge vol provar l'e...càcia d'un nou tractament (B) en front d'un tractament estàndar (B⁰). Obté els resultats de la taula 3.1, tenint en compte que ho prova amb pacients locals (C) i amb pacients de fora de la seva localitat (C⁰).

Observem com, tant en la població local com en la forànea, el nou tractament millora la probabilitat de curar el pacient. En efecte,

$$P(A=BC) = 0:1 > 0:05 = P(A=B^0C) ,$$

$$P(A=BC^0) = 0:95 > 0:5 = P(A=B^0C^0) .$$

En canvi mirant les dades sense diferenciar la població, taula 3.2, haurà de concloure que el nou tractament no bene...cia el pacient. En aquest cas es té

$$P(A=B) = 0:11 < 0:46 = P(A=B^0) .$$

	B ⁰	B
Mort(A ⁰)	5950	9005
Curat(A)	5050	1095

Taula 3.2: Paradoxa de Simpson. Sense tenir en compte la població, el nou tractament B no beneficia el pacient.

	A	A	B	B	C	C
	Nois	Noies	Nois	Noies	Nois	Noies
Si	19	20	7	40	10	30
No	31	30	43	210	30	70

Taula 3.3: Freqüències d'admissió a tres facultats. En cadascuna de les facultats A, B i C, les noies són beneficiades en l'admissió.

En el gràfic 3.1 veiem clarament la situació. Les dades vistes globalment van en una direcció ascendent mentre que en cadascuna de les poblacions aquesta direcció es capgira. Retrobarem aquest fet en el capítol 4, en el qual parlarem de desigualtats en distàncies.

Exemple 3.2

Les taules 3.3 i 3.5 descriuen les freqüències absolutes d'admissió, o no, en tres facultats, A, B i C tenint en compte la variable sexe.

D'una banda tenim

$$P(\text{Si} \wedge \text{Noi}) = 0:26 > 0:23 = P(\text{Si} \wedge \text{Noia})$$

	Nois	Noies
Si	36	90
No	104	310

Taula 3.4: Freqüències d'admissió a tres facultats. Globalment, sense tenir en compte les facultats, els nois són beneficiats en l'admissió.

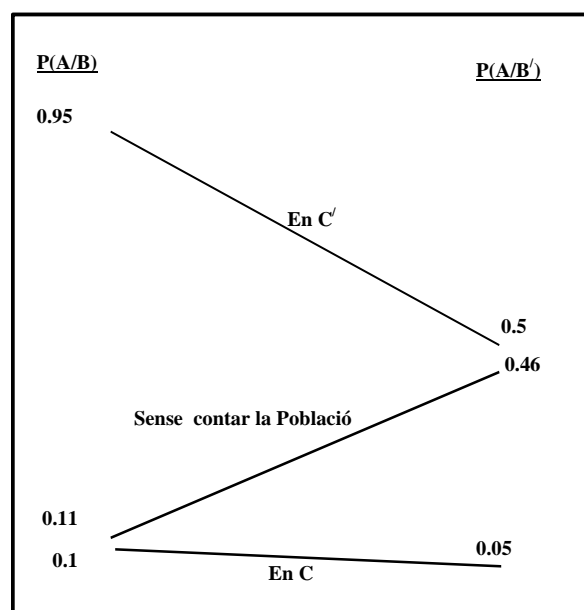


Figura 3.1: Paradoxa de Simpson. Probabilitats en l'exemple de l'eficàcia d'un nou tractament. La probabilitat tenint en compte les poblacions i sense tenir-les en compte van en direccions diferents.

i, d'altra,

$$\begin{aligned} P(\text{Si} \bar{A} \text{Noi} \setminus A) &= 0:38 < 0:40 = P(\text{Si} \bar{A} \text{Noia} \setminus A) , \\ P(\text{Si} \bar{A} \text{Noi} \setminus B) &= 0:14 < 0:16 = P(\text{Si} \bar{A} \text{Noia} \setminus B) , \\ P(\text{Si} \bar{A} \text{Noi} \setminus C) &= 0:25 < 0:30 = P(\text{Si} \bar{A} \text{Noia} \setminus C) , \end{aligned} \quad (3.1)$$

és a dir, la probabilitat (estimada) d'admissió en cada facultat és més gran per a les noies i, en canvi, globalment, la probabilitat d'admissió és més gran per als nois, fet aparentment contradictori.

Però com sempre, les paradoxes no no ho són tant quan es veu realment el que passa.

En primer lloc hem de fer notar que $P(\text{Si} \bar{A} \text{Noi})$ no és la mitjana de

$$P(\text{Si} \bar{A} \text{Noi} \setminus A) , P(\text{Si} \bar{A} \text{Noi} \setminus B) \text{ i } P(\text{Si} \bar{A} \text{Noi} \setminus C) ,$$

com tampoc ho és $P(\text{Si} \bar{A} \text{Noia})$ de

$$P(\text{Si} \bar{A} \text{Noia} \setminus A) , P(\text{Si} \bar{A} \text{Noia} \setminus B) \text{ i } P(\text{Si} \bar{A} \text{Noia} \setminus C) .$$

En realitat són mitjanes ponderades,

$$\begin{aligned} P(\text{Si} \bar{A} \text{Noi}) &= P(A \bar{A} \text{Noi}) P(\text{Si} \bar{A} \text{Noi} \setminus A) + \\ &P(B \bar{A} \text{Noi}) P(\text{Si} \bar{A} \text{Noi} \setminus B) + \\ &P(C \bar{A} \text{Noi}) P(\text{Si} \bar{A} \text{Noi} \setminus C) , \end{aligned}$$

$$\begin{aligned} P(\text{Si} \bar{A} \text{Noia}) &= P(A \bar{A} \text{Noia}) P(\text{Si} \bar{A} \text{Noia} \setminus A) + \\ &P(B \bar{A} \text{Noia}) P(\text{Si} \bar{A} \text{Noia} \setminus B) + \\ &P(C \bar{A} \text{Noia}) P(\text{Si} \bar{A} \text{Noia} \setminus C) . \end{aligned}$$

Si les variables sexe i facultat fossin independents, és a dir

$$P(A \bar{A} \text{Noi}) = P(A \bar{A} \text{Noia}) = P(A) = \frac{1}{2} ,$$

$$P(B \bar{A} \text{Noi}) = P(B \bar{A} \text{Noia}) = P(B) = \frac{1}{3} ,$$

$$P(C \bar{A} \text{Noi}) = P(C \bar{A} \text{Noia}) = P(C) = \frac{1}{3} \text{ i } \frac{1}{2} ,$$

tindríem

$$\begin{aligned} P(\text{Si} \bar{A} \text{Noi}) &= \frac{1}{2} P(\text{Si} \bar{A} \text{Noi} \setminus A) + \frac{1}{3} P(\text{Si} \bar{A} \text{Noi} \setminus B) \\ &+ (\frac{1}{3} \text{ i } \frac{1}{2}) P(\text{Si} \bar{A} \text{Noi} \setminus C) , \end{aligned}$$

$$P(\text{Si} \dot{A} \text{Noia}) = \sum_j P(\text{Si} \dot{A} \text{Noia} \setminus A) + \sum_j P(\text{Si} \dot{A} \text{Noia} \setminus B) \\ + (1 - \sum_j \sum_i \pi_{ij}) P(\text{Si} \dot{A} \text{Noia} \setminus B).$$

Si suposem, per exemple, $P(\text{Si} \dot{A} \text{Noi}) > P(\text{Si} \dot{A} \text{Noia})$, arribaríem a

$$[P(\text{Si} \dot{A} \text{Noi} \setminus A) - P(\text{Si} \dot{A} \text{Noia} \setminus A)] \sum_j \\ + [P(\text{Si} \dot{A} \text{Noi} \setminus B) - P(\text{Si} \dot{A} \text{Noia} \setminus B)] (1 - \sum_j \sum_i \pi_{ij}) >$$

$$(1 - \sum_j \sum_i \pi_{ij}) [P(\text{Si} \dot{A} \text{Noia} \setminus A) - P(\text{Si} \dot{A} \text{Noi} \setminus A)],$$

la qual cosa ens indica que en cap cas poden acomplir-se a la vegada les desigualtats (3.1). És a dir, si hi hagués independència entre les variables sexe i facultat no hi hauria paradoxa. El que passa aquí, en el nostre exemple, és que

$$P(\text{A} \dot{A} \text{Noi}) = 0:36 \neq 0:125 = P(\text{A} \dot{A} \text{Noia}), \\ P(\text{B} \dot{A} \text{Noi}) = 0:36 \neq 0:625 = P(\text{B} \dot{A} \text{Noia}), \\ P(\text{C} \dot{A} \text{Noi}) = 0:29 \neq 0:250 = P(\text{C} \dot{A} \text{Noia}).$$

A més a més

$$P(\text{Si} \dot{A} \text{A}) = 0:39 > 0:29 = P(\text{Si} \dot{A} \text{C}) > 0:16 = P(\text{Si} \dot{A} \text{B}),$$

i justament és en la facultat B, on es dona la mínima acceptació, on hi van més les noies a demanar l'admissió. Podem dir que la paradoxa de Simpson és el resultat de la interacció entre la variable sexe i la variable facultat. Aquesta interacció és produïda pel disseny experimental al ser diferents les proporcions de nois i noies que es presenten a cada facultat. Grà...cament, tornem a veure, ...gura 3.2, que el conjunt de dades van en una direcció descendent mentre que en cada facultat passa el contrari.

La paradoxa de Simpson es pot descriure, i és interessant fer-ho, amb l'ajut de l'anàlisi de correspondències, molt especialment amb els grà...cs propis d'aquesta tècnica d'anàlisi multivariant.

3.3 Anàlisi de Correspondències

L'Anàlisi de Correspondències (AC), té una llarga història que s'inicia el 1935 amb Hirschfeld [42], el 1940 amb Fisher [32] i el 1941 amb Guttman [37]. Ha estat extensament estudiat per Benzécri [4] i Greenacre [36].

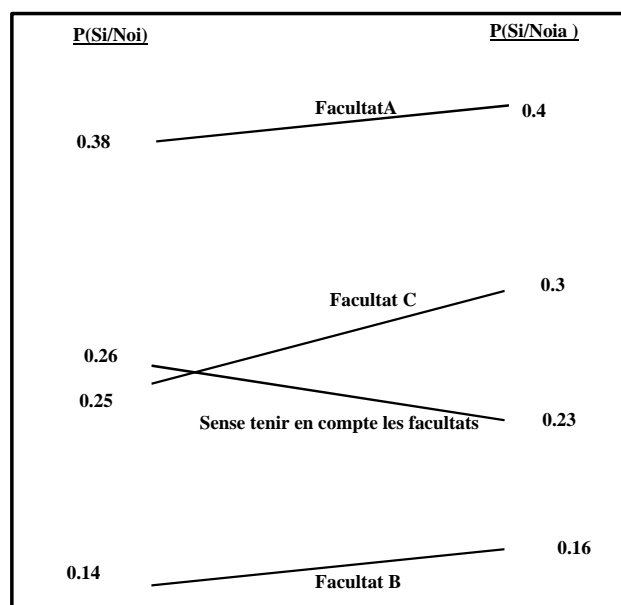


Figura 3.2: Paradoxa de Simpson. Probabilitats en l'exemple d'admissió en tres facultats, A, B i C. Les dades globals van en una direcció descendent i en cadascuna de les facultats la direcció és ascendent.

L'AC és apropiat per representar les categories de les ...les i columnes d'una taula de contingència, essent un dels seus objectius, descriure les relacions existents entre dues variables categòriques i, simultàniament, les relacions entre les categories de cada variable, tot això en un espai de dimensió petita, obtenint, per exemple, una representació geomètrica de les categories d'una variable en relació a la distribució de freqüències relatives de les categories de l'altra variable, utilitzant la distància khi-quadrat.

Suposem que tenim dues variables categòriques A (variable ...la) i B (variable columna) amb I i J categories respectivament, i que han estat observades en $n = \sum_{ij} f_{ij}$ individus, obtenint la taula de contingència $I \in J$

$$\begin{array}{cccccc}
 & B_1 & B_2 & \dots & B_J & \\
 A_1 & f_{11} & f_{12} & \dots & f_{1J} & f_{1t} \\
 A_2 & f_{21} & f_{22} & \dots & f_{2J} & f_{2t} \\
 \vdots & & & \ddots & & \vdots \\
 A_i & f_{i1} & f_{i2} & \dots & f_{iJ} & f_{it} \\
 & f_{t1} & f_{t2} & \dots & f_{tJ} & n
 \end{array} \quad (3.2)$$

on f_{ij} és la freqüència de la intersecció $A_i \cap B_j$; $f_{it} = \sum_j f_{ij}$ són les freqüències de A_i i $f_{tj} = \sum_i f_{ij}$ són les freqüències de B_j . Hem de tenir en compte que la taula (3.2) resumeix la matriu de dades inicial, que típicament és de la forma

$$Z = [X; Y] = \begin{array}{cccccccc}
 1 & 0 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 1 \\
 \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
 i & 0 & 0 & 0 & \dots & 1 & 0 & 1 & \dots & 0 & 0 \\
 \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
 n & 0 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 1 & 0
 \end{array}$$

$A_1 \quad A_2 \quad \dots \quad A_i \quad B_1 \quad B_2 \quad \dots \quad B_J$

on donem el valor 1 quan es presenta una característica i 0 quan no es presenta.

Indiquem per $N = (f_{ij})$ la matriu $I \in J$ amb les freqüències de la taula de contingència. Anomenem matriu de correspondències a

$$P = \frac{1}{n}N;$$

i indiquem per r el vector $I \in 1$ amb els totals marginals de les ...les de P , i per c el vector $J \in 1$ amb els totals marginals de les columnes de P ,

$$r = P1; \quad c = P^01;$$

El problema que tenen les variables categòriques és que no són quantitatives. La quantificació 0 ó 1 anterior és convencional. Assignem doncs a les categories $A_1; \dots; A_I$ de la variable ...la, els valors numèrics $a_1; \dots; a_I$; i a les categories $B_1; \dots; B_J$ de la variable columna, els valors numèrics $b_1; \dots; b_J$. Es a dir, indiquem els vectors

$$a = (a_1; \dots; a_I)^0; \quad b = (b_1; \dots; b_J)^0;$$

i considerem les variables compostes

$$U = Xa; \quad V = Yb;$$

Volem trobar $a; b$ tals que les correlacions entre U i V siguin màximes. Clarament, estem en un problema de correlació canònica, llevat que ara les matrius de covariàncies entre ...les i entre columnes són singulars.

Aplicant la teoria de la correlació canònica tenim, en notació matricial, la solució general

$$A_0 = D_r^{i-1/2}U; \quad B_0 = D_c^{i-1/2}V,$$

on D_r és la matriu diagonal que conté els valors marginals de les ...les i D_c la de les columnes. Com el producte per una constant no altera les correlacions, també obtenim correlacions màximes considerant les matrius

$$A = D_r^{i-1/2}UD_{\sigma}; \quad B = D_c^{i-1/2}VD_{\sigma}; \quad (3.3)$$

on D_{σ} és la matriu diagonal amb, en ordre decreixent, els valors singulars de la descomposició singular

$$D_r^{i-1/2}(P_i \quad rc^0)D_c^{i-1/2} = UD_{\sigma}V^0. \quad (3.4)$$

La representació de ...les i columnes utilitzant les coordenades principals $A; B$ és la solució simètrica. La representació utilitzant les matrius

$$A = D_r^{i-1/2}UD_{\sigma}; \quad B_0 = D_c^{i-1/2}V; \quad (3.5)$$

és a dir, coordenades principals per a les ...les i coordenades estàndar per a les columnes, és la solució asimètrica. Aquesta solució reproduïx millor la dependència entre ...les i columnes.

Per a que les distàncies entre punts-...la i punts-columna tinguin sentit hem de fer intervenir els per...ls de les ...les

$$\left(\frac{p_{i1}}{r_i}, \frac{p_{i2}}{r_i}, \dots, \frac{p_{iJ}}{r_i}\right);$$

matricialment, $Q = D_r^{-1}P$, i els per...ls de les columnes $D_c^{-1}P^0$. La distància entre, per exemple, A_k i A_l calculada a partir de $f_{k\ell}$ i $f_{l\ell}$ només ens donaria la diferència d'individus entre aquestes dues categories, mentre que si la calculem sobre els per...ls de les ...les avaluem les diferències de comportament en relació a les categories de la variable B.

La distància khi-quadrat entre les ...les i i i^0 de N és

$$\chi_{ii^0}^2 = \sum_{j=1}^J \frac{(p_{ij}=r_i - p_{i^0j}=r_{i^0})^2}{c_j};$$

i, anàlogament, podem definir la distància khi-quadrat entre columnes com

$$\chi_{jj^0}^2 = \sum_{i=1}^I \frac{(p_{ij}=c_j - p_{ij^0}=c_{j^0})^2}{r_i};$$

Es demostra que les distàncies euclidianes entre les ...les de la matriu B (matriu A) obtingudes a (3.3), coincideixen amb la distància khi-quadrat $\chi_{jj^0}^2$ (la distància khi-quadrat $\chi_{ii^0}^2$).

3.4 Aplicació a la representació de la paradoxa de Simpson

Exemple 3.3

Després d'aquesta breu descripció de l'AC, apliquem la solució asimètrica d'aquesta tècnica multivariant a l'exemple 3.2, reestructurant les dades en la forma donada en la taula 3.5, en la qual considerem només dues variables, la variable facultat d'una banda i la variable sexe-admissió de l'altra, agrupant en una sola variable les variables inicials sexe i admissió. El mètode asimètric

	A	B	C
Noies si (n-si)	20	40	30
Noies no (n-no)	30	210	70
Nois si (N-si)	19	7	10
Nois no (N-no)	31	43	30

Taula 3.5: Freqüències d'admissió a tres facultats on considerem com una sola variable, sexe - admissió, les variables inicials, sexe i admissió.

de l'AC utilitzat és interessant per examinar les diferències o similituds entre les categories de la variable ...la.

En la ...gura 3.3 es troba la representació grà...ca dels resultats de l'AC.

La proximitat del punt que representa les noies no admeses al punt que representa la facultat B, indica una forta demanda a aquesta facultat que és la que més rebutja.

Aquest últim grà...c el podríem ampliar afegint-li:

- 2 la informació de les freqüències marginals per a les tres facultats, $P(A) = 0:19$, $P(B) = 0:56$ i $P(C) = 0:26$, a través de tres cercles, l'àrea dels quals és proporcional a cadascuna d'aquestes freqüències
- 2 la proporció d'admesos i no admesos, part no ombrejada i sí ombrejada, en cadascuna de les tres facultats representades pels cercles corresponents, $P(\text{Si} \overline{A}) = 0:39$, $P(\text{Si} \overline{B}) = 0:16$ i $P(\text{Si} \overline{C}) = 0:29$.

Tenim el grà...c ampliat en la ...gura 3.4, en el qual s'aprecia grà...cament el que deiem abans, les noies sol·liciten més l'ingrés a la facultat B que és la que té més percentatge de no admesos, la qual cosa explica la paradoxa.

Cal notar que les projeccions, amb origen en el vèrtex A sobre el segment BC, de cadascun dels punts ...la, representen la proporció en que la ...la en qüestió participa en les columnes B i C. El mateix passa amb les altres projeccions amb origen els altres vèrtex, B i C, sobre els altres segments AC i AB. Així, i tal com apareix en el grà...c, els nois no admesos en les facultats B i C són 7 i 10 respectivament, per tant, el segment BC queda repartit de forma inversament proporcional a les freqüències $\frac{7}{7+10}$ i $\frac{10}{7+10}$.

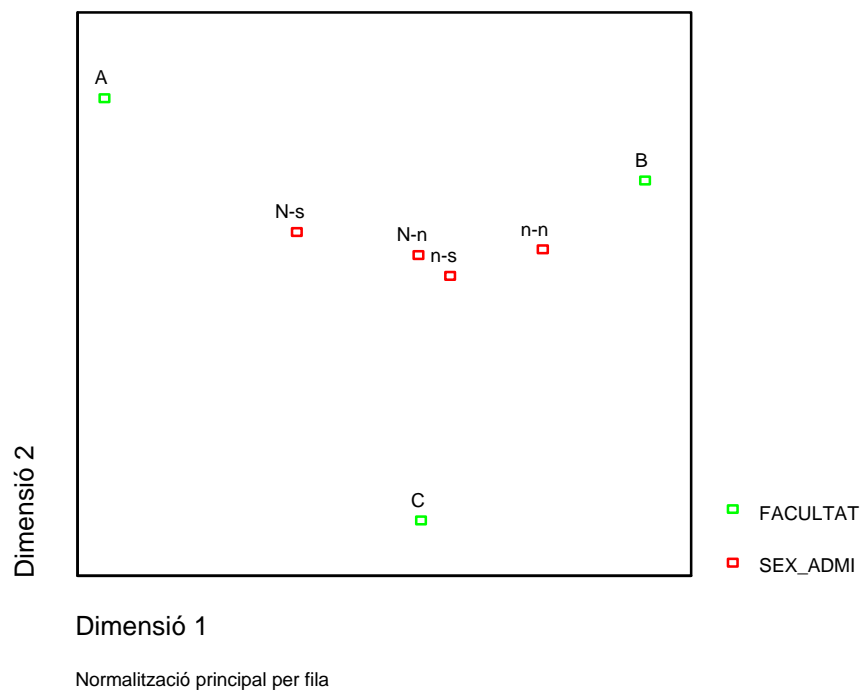


Figura 3.3: Representació gràfica de l'AC asimètric per l'exemple de les tres facultats. Representem les files amb les coordenades principals i les columnes amb les coordenades estàndard. La proximitat del punt que representa les noies no admeses (n-n) al punt que representa la facultat B, indica una forta demanda a aquesta facultat que és la que més rebutja. (n-s=noies sí admeses, n-n=noies no admeses, N-s=nois sí admesos i N-n=nois no admesos)

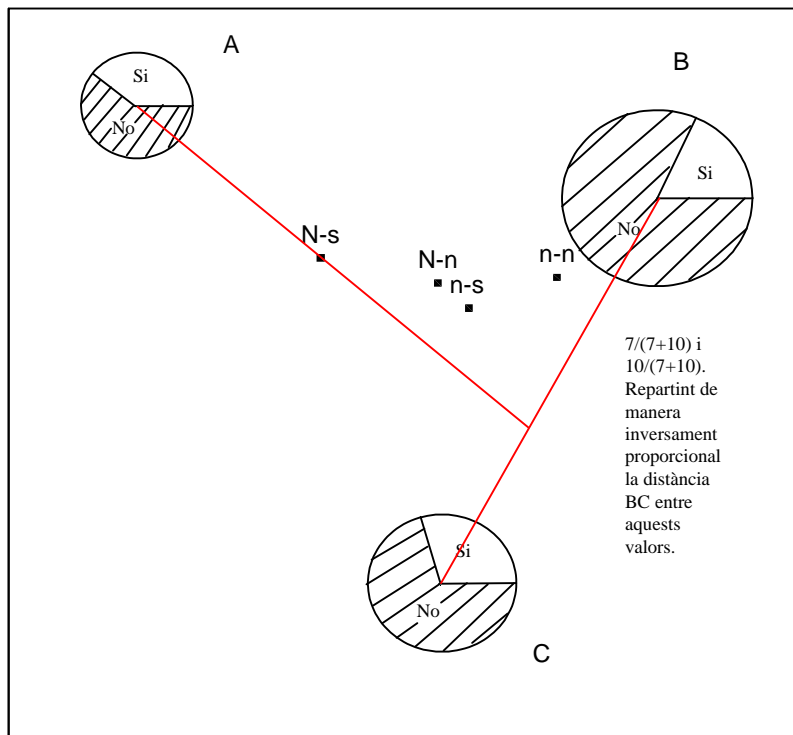


Figura 3.4: Representació gràfica ampliada de l'AC asimètric (representem les files amb les coordenades principals i les columnes amb les coordenades estàndar) per l'exemple de les tres facultats. Les noies sol·liciten més l'ingrés a la facultat B que és la que té més percentatge de no admesos, part ombrejada de cada cercle, la qual cosa explica la paradoxa.

Supervivència	Tabac		Total
	F	NF	
NS	139	230	369
SS	443	502	945
Total	582	732	1314

Taula 3.6: El fet de ser fumadora (F) té un efecte positiu sobre la supervivència (NS: no sobreviu, SS: sí sobreviu), després de 20 anys, en les dones d'aquesta mostra.

	Edat							Total
	18-24	25-34	35-44	45-54	55-64	65-74	75+	
NS-F	2	3	14	27	51	29	13	139
NS-NF	1	5	7	12	40	101	64	230
SS-F	53	121	95	103	64	7	0	443
SS-NF	61	152	114	66	81	28	0	502
Total	117	281	230	208	236	165	77	1314

Taula 3.7: Taula de freqüències, per l'exemple de la relació entre tabac i supervivència, considerant com una sola variable, supervivència-tabac, les variables inicials supervivència i tabac. En cadascun dels grups d'edat, tret del grup 25-34 anys, el tabac té un efecte negatiu sobre la supervivència, després de 20 anys, en les dones d'aquesta mostra.

Exemple 3.4

Considerem, ...nalment, un exemple real estudiat per Appleton et al. [2].

La taula 3.6 conté les dades reals, que classi...quen 1314 dones segons les variables tabac i supervivència. La variable tabac té dues categories: fumadora habitual i no fumadora (F, NF). La variable supervivència també té dues categories: sobreviu (SS) i no sobreviu (NS), segons segueixi viva o no després de 20 anys. L'estudi es va realitzar durant els anys 1972-74 i 1995.

Amb aquestes dades,

$$P(NS\acute{A}F) = 0:2388 < 0:3142 = P(NS\acute{A}NF) ,$$

és a dir, sembla que el tabac té un efecte positiu sobre la supervivència al cap de 20 anys. Considerem, però, la taula 3.7 la qual conté les mateixes

dades de la taula 3.6 introduint la variable edat i considerant solament dues variables, la variable edat i la variable supervivència-tabac, amb els valors NS-F si no sobreviu i és fumadora, NS-NF si no sobreviu i no és fumadora, SS-F si sobreviu i és fumadora i SS-NF si sobreviu i no és fumadora. Ara tenim

$$P(NS|NF) = \begin{matrix} 0 \\ \text{|||||} \\ \text{|||||} \\ \text{|||||} \\ \text{|||||} \\ \text{|||||} \\ \text{|||||} \\ 1 \end{matrix} \begin{matrix} 0:0161 < 0:0364 \\ 0:0318 > 0:0242 \\ 0:0579 < 0:1284 \\ 0:1538 < 0:2077 \\ 0:3306 < 0:4435 \\ 0:7829 < 0:8056 \\ 1 = 1 \end{matrix} \begin{matrix} 0 \\ \text{|||||} \\ \text{|||||} \\ \text{|||||} \\ \text{|||||} \\ \text{|||||} \\ \text{|||||} \\ 1 \end{matrix} = P(NS|F) ,$$

on es veu que l'efecte del tabac és negatiu. Cal notar, però, que en el grup d'edat format per les dones de 25 a 34 anys l'efecte del tabac és lleugerament positiu i que en el grup de més de 75 anys, per raons òbvies, la probabilitat de no sobreviure després de 20 anys és 1 tant en les fumadores com en les no fumadores.

Apliquem a les dades en la forma de la taula 3.7 l'AC. Amb les notacions de la secció 3.3, tenim

$$N = \begin{matrix} 0 & & & & & & 1 \\ \text{|||||} & 2 & 3 & 14 & 27 & 51 & 29 & 13 \\ \text{|||||} & 1 & 5 & 7 & 12 & 40 & 101 & 64 \\ \text{|||||} & 53 & 121 & 95 & 103 & 64 & 7 & 0 \\ & 61 & 152 & 114 & 66 & 81 & 28 & 0 \end{matrix} \begin{matrix} 0 \\ \text{|||||} \\ \text{|||||} \\ \text{|||||} \\ \text{|||||} \\ \text{|||||} \\ \text{|||||} \\ 1 \end{matrix} \begin{matrix} 0 \\ \text{|||||} \\ \text{|||||} \\ \text{|||||} \\ \text{|||||} \\ \text{|||||} \\ \text{|||||} \\ 1 \end{matrix} A'$$

$$P = \begin{matrix} 0 & & & & & & 1 \\ \text{|||||} & 0:002 & 0:002 & 0:011 & 0:021 & 0:039 & 0:221 & 0:010 \\ \text{|||||} & 0:0008 & 0:004 & 0:005 & 0:009 & 0:030 & 0:077 & 0:049 \\ \text{|||||} & 0:040 & 0:092 & 0:072 & 0:078 & 0:049 & 0:005 & 0 \\ & 0:046 & 0:116 & 0:087 & 0:050 & 0:062 & 0:021 & 0 \end{matrix} \begin{matrix} 0 \\ \text{|||||} \\ \text{|||||} \\ \text{|||||} \\ \text{|||||} \\ \text{|||||} \\ \text{|||||} \\ 1 \end{matrix} \begin{matrix} 0 & & & & & & 1 \\ & 0:089 & & & & & \\ \text{|||||} & 0:214 & & & & & \\ \text{|||||} & 0:175 & & & & & \\ \text{|||||} & 0:158 & & & & & \\ \text{|||||} & 0:180 & & & & & \\ \text{|||||} & 0:126 & & & & & \\ & 0:059 & & & & & \end{matrix} \begin{matrix} 0 \\ \text{|||||} \\ \text{|||||} \\ \text{|||||} \\ \text{|||||} \\ \text{|||||} \\ \text{|||||} \\ 1 \end{matrix} A'$$

$$r = \begin{matrix} 0 & & & & & & 1 \\ \text{|||||} & 0:106 & & & & & \\ \text{|||||} & 0:175 & & & & & \\ \text{|||||} & 0:337 & & & & & \\ & 0:382 & & & & & \end{matrix} \begin{matrix} 0 \\ \text{|||||} \\ \text{|||||} \\ \text{|||||} \\ \text{|||||} \\ \text{|||||} \\ \text{|||||} \\ 1 \end{matrix} A', c = \begin{matrix} 0 & & & & & & 1 \\ & 0:089 & & & & & \\ \text{|||||} & 0:214 & & & & & \\ \text{|||||} & 0:175 & & & & & \\ \text{|||||} & 0:158 & & & & & \\ \text{|||||} & 0:180 & & & & & \\ \text{|||||} & 0:126 & & & & & \\ & 0:059 & & & & & \end{matrix} A'$$

Els valors singulars de la descomposició (3.4) són $\lambda_1 = 0:712$, $\lambda_2 = 0:200$ i $\lambda_3 = 0:108$. La variabilitat explicada per les dues primeres dimensions principals és del 97:9%.

La matriu de per...ls de les ...les és

$$Q = \begin{matrix} \text{O} & & & & & & & \text{1} \\ \begin{matrix} \text{m} \\ \text{m} \\ \text{m} \\ \text{m} \\ \text{m} \\ \text{m} \\ \text{m} \\ \text{m} \end{matrix} & \begin{matrix} 0:019 & 0:019 & 0:104 & 0:199 & 0:369 & 0:209 & 0:095 \\ 0:005 & 0:023 & 0:029 & 0:051 & 0:171 & 0:440 & 0:280 \\ 0:119 & 0:273 & 0:214 & 0:231 & 0:145 & 0:015 & 0 \\ 0:120 & 0:304 & 0:228 & 0:131 & 0:162 & 0:055 & 0 \end{matrix} & \begin{matrix} \text{C} \\ \text{C} \\ \text{C} \\ \text{C} \\ \text{C} \\ \text{C} \\ \text{C} \\ \text{C} \end{matrix} \end{matrix}$$

La solució asimètrica (3.5) dona, en les seves dues primeres coordenades de la representació conjunta de ...les i columnes,

$$A = \begin{matrix} \text{O} & & & \text{1} \\ \begin{matrix} \text{m} \\ \text{m} \\ \text{m} \\ \text{m} \\ \text{m} \\ \text{m} \\ \text{m} \\ \text{m} \end{matrix} & \begin{matrix} 0:493 & 0:544 \\ 1:423 & 0:157 \\ 0:462 & 0:029 \\ 0:381 & 0:105 \end{matrix} & \begin{matrix} \text{C} \\ \text{C} \\ \text{C} \\ \text{C} \\ \text{C} \\ \text{C} \\ \text{C} \\ \text{C} \end{matrix} \end{matrix} \quad B_0 = \begin{matrix} \text{O} & & \text{1} \\ \begin{matrix} \text{m} \\ \text{m} \\ \text{m} \\ \text{m} \\ \text{m} \\ \text{m} \\ \text{m} \\ \text{m} \end{matrix} & \begin{matrix} 0:763 & 0:832 \\ 0:738 & 1:022 \\ 0:603 & 0:285 \\ 0:401 & 1:069 \\ 0:181 & 1:572 \\ 1:721 & 0:420 \\ 2:495 & 0:959 \end{matrix} & \begin{matrix} \text{C} \\ \text{C} \\ \text{C} \\ \text{C} \\ \text{C} \\ \text{C} \\ \text{C} \\ \text{C} \end{matrix} \end{matrix}$$

coordenades principals per a les ...les i coordenades estàndar per a les columnes.

La figura 3.5 dona aquesta representació asimètrica ampliada amb la informació de les freqüències marginals per als 7 grups d'edat, àrea de cada cercle, i la proporció de no supervivents i sí supervivents, part no ombrejada i sí ombrejada respectivament, en cadascun dels grups d'edat. Hem afegit també els punts ns (no sobreviuen) i ss (sí sobreviuen) els quals representen les mitjanes ponderades de ns-f, ns-nf i de ss-f, ss-nf, respectivament. Les barres sobre ns-f i ns-nf indiquen el percentatge de no supervivència de fumadores i no fumadores respectivament.

Es pot veure com els fumadors són més pròxims als grups d'edat més joves, on la supervivència és més gran.

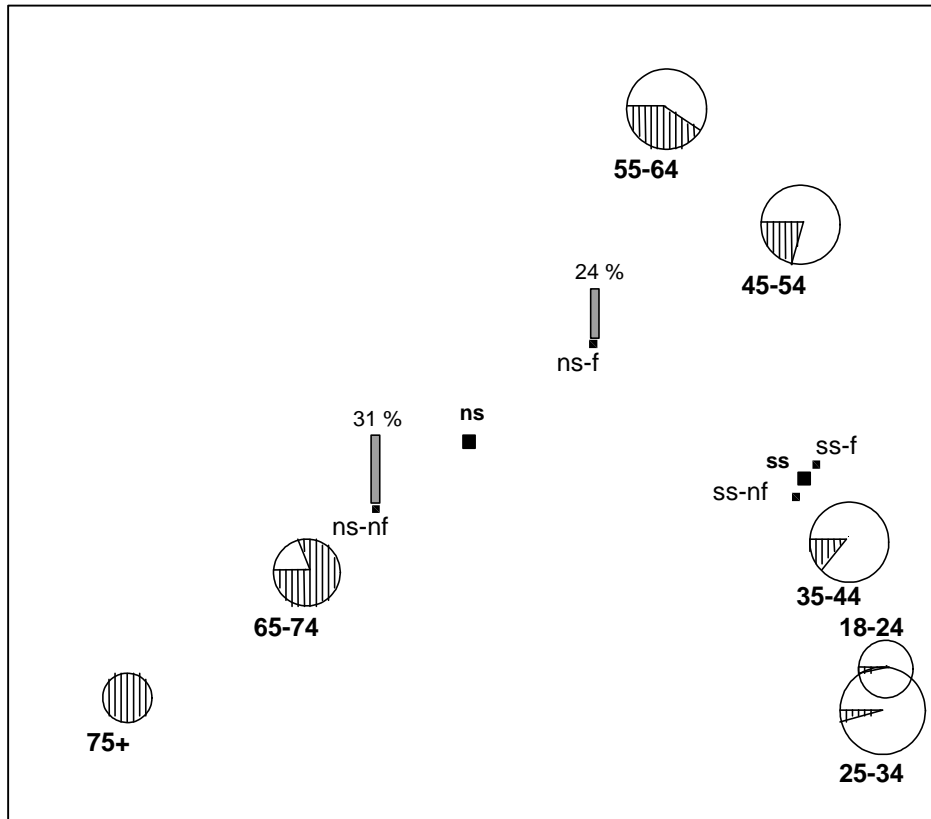


Figura 3.5: Representació gràfica ampliada de l'AC asimètric (representem les ...les amb les coordenades principals i les columnes amb les coordenades estàndar) per a l'exemple del tabac i supervivència després de 20 anys. Els fumadors, ns-f, ss-f estan més prop dels grups d'edat joves, on la supervivència és més gran. La part ratllada del cercle indica la proporció d'individus que no sobreviuen. Els punts ns (no sobreviuen) i ss (sí sobreviuen) són les mitjanes ponderades de ns-f, ns-nf i de ss-f, ss-nf, respectivament. Les barres sobre ns-f i ns-nf indiquen el percentatge de no supervivència de fumadores i no fumadores respectivament.

Capítol 4

Desigualtats amb distàncies

4.1 Introducció

En el capítol 1 hem estudiat la desigualtat

$$\frac{1}{2}_{yx}^0 \mathbf{R}^{-1} \frac{1}{2}_{yx} > \frac{1}{2}_{yx}^0 \frac{1}{2}_{yx} .$$

La forma d'aquesta desigualtat dóna peu a buscar una versió utilitzant distàncies entre dues poblacions, Cuadras, Duran i Fortiana [21].

Aquesta desigualtat expressa el fet que no sempre variables correlacionades contenen informació redundant i, si interpretem $\frac{1}{2}_{yx}^0 \mathbf{R}^{-1} \frac{1}{2}_{yx}$ com una distància de Mahalanobis (variables correlacionades) i $\frac{1}{2}_{yx}^0 \frac{1}{2}_{yx}$ com una distància de K. Pearson (variables incorrelacionades), ens suggereix comparar la distància de Mahalanobis, la qual conté la redundància entre les variables, i la distància de K. Pearson que no la conté. En el present capítol estudiarem la desigualtat que podem expressar dient que la distància de Mahalanobis entre dues poblacions pot ser més gran que la distància de K. Pearson entre elles.

En el moment d'estudiar aquesta desigualtat entre distàncies veurem, com hem vist en el capítol 3 en estudiar els exemples il·lustratius de la paradoxa de Simpson, que les dades es comporten de manera diferent, depenent de si tenim o no en compte les poblacions. Podem dir que la desigualtat en distàncies, Mahalanobis més gran que K. Pearson, és una versió multivariant de la paradoxa de Simpson.

Després de generalitzar a més de dues poblacions aquesta desigualtat, l'estudiarem des del punt de vista de l'anàlisi multivariant de la variància.

Continuarem comentant un exemple clàssic de Fisher, i, per l'estudi gràfic del qual, veurem com representar una component principal en l'espai canònic.

Per acabar el capítol intentarem interpretar la paradoxa de Rao, coneguda com la "maledicció de la dimensionalitat", des del punt de vista de la desigualtat en distàncies, Mahalanobis més gran que K. Pearson.

4.2 Dues poblacions

Siguin $\mu_1 = (\mu_{1j}; \mathbf{S})$ i $\mu_2 = (\mu_{2j}; \mathbf{S})$ dues poblacions multivariants p-dimensionals, essent μ_{1j} i μ_{2j} els vectors de mitjanes respectius i \mathbf{S} la matriu de variàncies-covariàncies en les dues poblacions.

Sigui

$$\pm_M^2 = (\mu_{1j} - \mu_{2j})^0 \mathbf{S}^{-1} (\mu_{1j} - \mu_{2j})$$

el quadrat de la distància de Mahalanobis entre les dues poblacions μ_1 i μ_2 , i, com un cas particular, sigui

$$\pm_K^2 = \sum_{i=1}^p \frac{(\mu_{1i} - \mu_{2i})^2}{s_{ii}^2} = (\mu_{1j} - \mu_{2j})^0 \mathbf{S}_0^{-1} (\mu_{1j} - \mu_{2j})$$

el quadrat de la distància de K. Pearson, on $\mathbf{S}_0 = \text{diag}(\mathbf{S})$.

Com que la distància de Mahalanobis conté la redundància entre les variables, a través dels elements de fora de la diagonal de \mathbf{S}^{-1} ; i la distància de K. Pearson no la conté, a l'intervenir només en aquesta la matriu diagonal \mathbf{S}_0 , podríem esperar tenir la desigualtat

$$\pm_M^2 < \pm_K^2, \quad (4.1)$$

en el sentit que variables incorrelacionades discriminen millor.

Es pot presentar, però, la desigualtat contrària

$$(\mu_{1j} - \mu_{2j})^0 \mathbf{S}^{-1} (\mu_{1j} - \mu_{2j}) = \pm_M^2 > \pm_K^2 = (\mu_{1j} - \mu_{2j})^0 \mathbf{S}_0^{-1} (\mu_{1j} - \mu_{2j}). \quad (4.2)$$

Utilitzant components principals, es pot demostrar el següent resultat, Cuadras i Fortiana [24], el qual podem veure gràficament i intuïtivament en la figura 4.1.

Resultat 4 La desigualtat $\pm_M^2 > \pm_K^2$ es presenta quan μ_{1j} i μ_{2j} segueixen la direcció de l'última component principal.

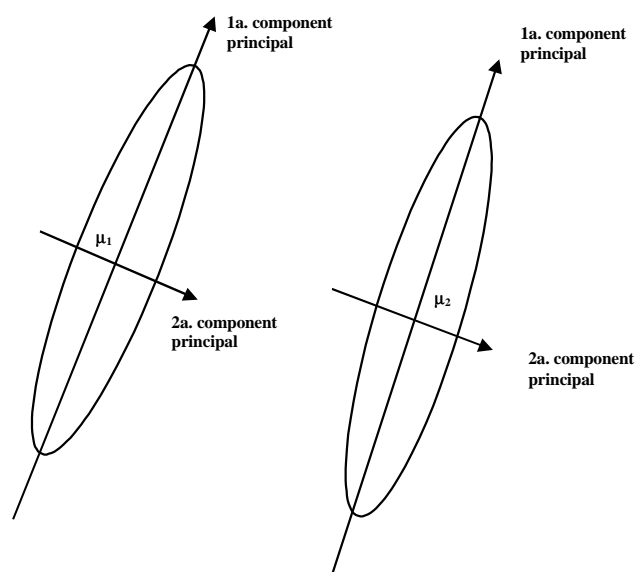


Figura 4.1: $\pm_M^2 > \pm_K^2$. La direcció del segment que uneix les mitjanes de les dues poblacions segueix, pràcticament, la direcció de la segona component principal.

En efecte.

Sigui U_i un vector propi de valor propi λ_i , de la matriu S . Aleshores

$$S U_i = \lambda_i U_i, \quad (4.3)$$

i, per tant

$$S^{-1} U_i = \frac{U_i}{\lambda_i}. \quad (4.4)$$

Suposem que els valors propis de S estan ordenats,

$$\lambda_1 > \lambda_2 > \dots > \lambda_{p-1} > \lambda_p,$$

els quals, tenint en compte que $\mathbf{P} = \mathbf{P}_{i=1}^p U_i = \mathbf{P}_{i=1}^p \lambda_{ii}$, verifiquen

$$\lambda_1 > \lambda_{ii} > \lambda_p > 0, \quad i = 1; \dots; p,$$

essent λ_{ii} la variància de la variable X_i .

Aleshores per a qualsevol vector V , tenim

$$\frac{1}{\lambda_1} V^0 V < V^0 S_0^{-1} V = \frac{V_1^2}{\lambda_{11}} + \dots + \frac{V_p^2}{\lambda_{pp}}, \quad (4.5)$$

$$\frac{1}{\lambda_p} V^0 V > V^0 S_0^{-1} V. \quad (4.6)$$

Sigui $Y_i = U_i^0 X$ la i -èsima component principal, la qual suposem que segueix la direcció de la línia $1_1 1_2$, és a dir suposem que verifiquem $1_1 1_2 = U_i$.

Per la igualtat (4.4), podem escriure

$$S^{-1} (1_1 1_2) = \frac{1_1 1_2}{\lambda_i},$$

i, per tant,

$$\pm_M^2 = (1_1 1_2)^0 S^{-1} (1_1 1_2) = \frac{(1_1 1_2)^0 (1_1 1_2)}{\lambda_i}.$$

Ara, de les desigualtats (4.5) i (4.6) tenim

$$(1_1 1_2)^0 S^{-1} (1_1 1_2) = \pm_M^2 < \pm_K^2 = (1_1 1_2)^0 S_0^{-1} (1_1 1_2),$$

si $\lambda_i = \lambda_1$, i

$$(1_1 1_2)^0 S^{-1} (1_1 1_2) = \pm_M^2 > \pm_K^2 = (1_1 1_2)^0 S_0^{-1} (1_1 1_2),$$

si $s_i = s_p$, que és el que volíem veure.

Més en general, trobarem la desigualtat (4.2), quan $\mu_{i_1}^2$ segueixi pràcticament la direcció de les components principals de variància petita. Això és el que enunciem en el següent resultat.

Resultat 5 La desigualtat $\sum_{i=1}^M s_i^2 > \sum_{k=1}^K s_k^2$ es verificarà sempre que

$$\sum_{i=1}^M s_i \mu_{i_1}^2 > 0,$$

essent $\mu_i = U_i^0 (x_{i_1} - x_{i_2})$ el producte escalar entre el vector diferència de mitjanes $x = x_{i_1} - x_{i_2}$ i el i -èssim vector propi de S .

En efecte, sense pèrdua de generalitat, suposem dues poblacions multivariants $x_1 = (0; R)$ i $x_2 = (1; R)$; on R és la matriu de correlacions. (En tot cas sempre podem suposar $x = x_{i_1} - x_{i_2}$ i en cas que les variables no tinguessin variància 1, sempre podríem seguir la demostració prenent $\frac{1}{\sqrt{1-k}}$).

La desigualtat (4.2) la podem escriure, per tant,

$$x^0 R x^1 > x^0 1,$$

i també

$$x^0 R x^1 - I_p > 0. \tag{4.7}$$

Sigui $R = U \alpha U^0$, amb $U^0 U = U U^0 = I_p$, la descomposició espectral de R , que també podem escriure

$$R = \sum_{i=1}^M s_i U_i U_i^0.$$

De la mateixa manera $R x^1 = U \alpha x^1 U^0$, i

$$R x^1 = \sum_{i=1}^M s_i x_{i_1} U_i U_i^0.$$

Així,

$$\begin{aligned} R x^1 - I_p &= \sum_{i=1}^M s_i x_{i_1} U_i U_i^0 - \sum_{i=1}^M U_i U_i^0 \\ &= \sum_{i=1}^M s_i (x_{i_1} - 1) U_i U_i^0. \end{aligned}$$

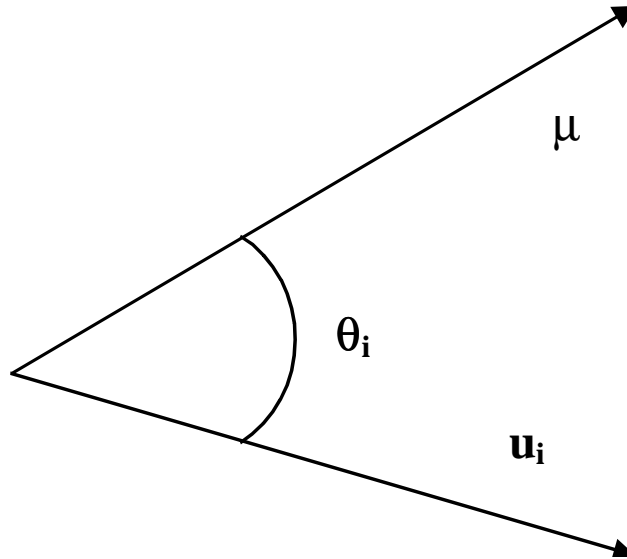


Figura 4.2: Significació del coeficient alfa, $\alpha_i = \cos \mu_i$.

La desigualtat (4.7), i per tant la (4.2), queda ara

$$\sum_{i=1}^p \alpha_i^2 \|U_i\|^2 > 0, \tag{4.8}$$

és a dir

$$\sum_{i=1}^p \alpha_i^2 > 0. \tag{4.9}$$

Definim per a tot $i = 1; \dots; p$, el coeficient $\alpha_i = \|U_i\|^{-1}$ que podria considerar-se quelcom semblant a una correlació, vegeu la figura 4.2. Tenim per tant

$$\alpha_i = \cos \mu_i = \frac{\langle U_i, k \rangle}{\|U_i\| \|k\|} = \cos \mu_i,$$

considerant també $\|k\| = 1$, sense pèrdua de generalitat, ja que sempre podem fer $\frac{1}{\|k\|}$.

Vectorialment seria

$$\alpha_i = \langle U_i, k \rangle.$$

Les desigualtats (4.8) i (4.9) queden, com volíem demostrar,

$$\sum_{i=1}^p \alpha_i^2 > 0, \tag{4.10}$$

que matricialment és

$$\mathbb{R}^1 \alpha_i^{-1} \mathbb{I}_p \mathbb{R}^0 > 0,$$

respectivament.

Atès que $\sum_{i=1}^p \alpha_i^2 = 1$; d'una manera informal podem dir que la desigualtat (4.2) es complirà quan α_i^2 sigui pròxim a 1 per a $\alpha_i < 1$ i α_i^2 pròxim a 0 per a $\alpha_i > 1$. També informalment podem dir que (4.10) s'acomplirà quan l'angle entre $\mathbb{1}$ i U_i sigui "petit" (mateixa direcció) per als vectors propis que donen les components principals Y_i amb variància més petita d'1; i sigui "gran" (direcció orthogonal) per als vectors propis que donen les components principals Y_i amb variància més gran d'1.

Podem afirmar, per tant, que en aquesta situació, la direcció geomètrica que segueixen les mitjanes de les dues poblacions és diferent a la direcció principal que segueixen les variables observades, direcció que ens ve donada per la primera component principal.

4.3 Exemple il·lustratiu

Flury [33], en l'estudi del discriminador lineal per dues variables i dues poblacions, presenta sis casos típics, els quals podem veure en la figura 4.3.

Els valors dels paràmetres utilitzats per Flury, suposant $\mu_2 = 0$, estan a la taula 4.1, en la qual apareixen α_1 , S , R i els valors de les distàncies de Mahalanobis i de K. Pearson.

En el primer cas i en el cinquè les distàncies entre les dues poblacions coincideixen al ser $S = S_0$.

En els casos 2 i 3 es verifica la desigualtat (4.1), la que generalment podríem esperar tenir, doncs S conté informació redundant. En la figura 4.3 s'observa com en aquests casos el vector $\mathbb{1}$ i la segona component principal van en direccions perpendiculars. De fet podem calcular l'angle $\mu_2 = \arccos \frac{\mathbb{1}^T u_2}{\|\mathbb{1}\| \|u_2\|} = 90^\circ$.

En els casos quart i sisè la situació és diferent. Es verifica la desigualtat (4.2) i en el quart cas s'aprecia com els vectors $\mathbb{1}$ i u_2 estan en una direcció més propera. De fet, en el quart cas $\mu_2 = 45^\circ$ i en el sisè $\mu_2 = 0^\circ$.

En la taula 4.2 hi ha els valors i vectors propis corresponents a la descomposició espectral de la matriu de correlacions R , així com els valors dels coeficients α_1 i α_2 i l'avaluació de la desigualtat (4.10). Aquesta última

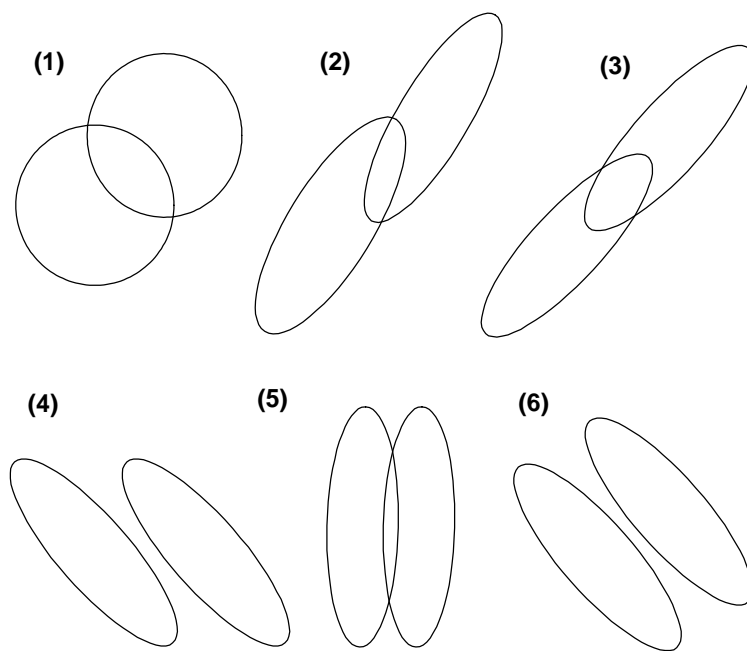


Figura 4.3: El·lipses de concentració pels sis casos presentats per Flury per l'estudi de la discriminació lineal entre dues poblacions. En els casos (4) i (6) s'acompleix la desigualtat (4.2).

Cas	\bar{A}^{-1}	$\bar{A}^{-1} S$	$\bar{A}^{-1} R$	\pm_M^2		\pm_K^2
1	$\begin{pmatrix} 2 \\ \bar{A} \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ \bar{A} & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ \bar{A} & 0 \end{pmatrix}$	8	=	8
2	$\begin{pmatrix} 2 \\ \bar{A} \end{pmatrix}$	$\begin{pmatrix} 2 & 2 \\ \bar{A} & 2 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.8165 \\ \bar{A} & 0.8165 \end{pmatrix}$	2	<	3:3
3	$\begin{pmatrix} 2 \\ \bar{A} \end{pmatrix}$	$\begin{pmatrix} 2 & 1.9 \\ \bar{A} & 1.9 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.95 \\ \bar{A} & 0.95 \end{pmatrix}$	2:05	<	4
4	$\begin{pmatrix} 2 \\ \bar{A} \end{pmatrix}$	$\begin{pmatrix} 2 & 1.6 \\ \bar{A} & 1.6 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.8 \\ \bar{A} & 0.8 \end{pmatrix}$	10	>	2
5	$\begin{pmatrix} 2 \\ \bar{A} \end{pmatrix}$	$\begin{pmatrix} 2 & 0 \\ \bar{A} & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ \bar{A} & 0 \end{pmatrix}$	2	=	2
6	$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 2 & 1.8 \\ \bar{A} & 1.8 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.9 \\ \bar{A} & 0.9 \end{pmatrix}$	10	>	1

Taula 4.1: Vectors de mitjanes i matrius de covariància i correlació, així com l'avaluació de les distàncies de Mahalanobis i de K. Pearson pels sis casos de discriminació lineal presentats per Flury. S'observa que la desigualtat (4.2) es verifica en el quart i en el sisè cas.

Cas	s_1	s_2	$\tilde{A} u_1$	$\tilde{A} u_2$	\tilde{r}_1	\tilde{r}_2	\tilde{P}^3 $\frac{1}{s_1} \tilde{r}_1 + \frac{1}{s_2} \tilde{r}_2$
1	1	1	$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$	2	2	0
2	1:82	0:18	$\begin{pmatrix} 0:707 \\ 0:707 \end{pmatrix}$	$\begin{pmatrix} 0:707 \\ 0:707 \end{pmatrix}$	2:83	0	$3:5959 < 0$
3	1:95	0:05	$\begin{pmatrix} 0:707 \\ 0:707 \end{pmatrix}$	$\begin{pmatrix} 0:707 \\ 0:707 \end{pmatrix}$	2:83	0	$3:8974 < 0$
4	1:8	0:2	$\begin{pmatrix} 0:707 \\ 0:707 \end{pmatrix}$	$\begin{pmatrix} 0:707 \\ 0:707 \end{pmatrix}$	1:41	1:41	$7:1110 > 0$
5	1	1	$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$	2	0	0
6	1:9	0:1	$\begin{pmatrix} 0:707 \\ 0:707 \end{pmatrix}$	$\begin{pmatrix} 0:707 \\ 0:707 \end{pmatrix}$	0	1:41	$17:5260 > 0$

Taula 4.2: Valors i vectors propis i coeficients alfa, així com l'avaluació de la condició de compliment de la desigualtat (4.2), pels sis casos de discriminació lineal presentats per Flury. En el quart i en el sisè la condició s'acompleix.

columna il·lustra el resultat 5. Cal fer notar que els valors dels coeficients α_1 i α_2 no estan normalitzats. Per considerar-los com $\alpha_i = \cos \mu_i$ s'hauria de dividir els valors donats en la taula 4.2 per les corresponents normes.

4.4 Més de dues poblacions

Considerem ara el cas de $g > 2$ poblacions.

Siguin $\mathbf{x}_i = (\mathbf{x}_i, \mathbf{S})$, amb $i = 1, \dots, g$; g poblacions multivariants.

Una generalització natural de la noció de distància (al quadrat) entre dues poblacions és la noció de variabilitat geomètrica, Cuadras, Fortiana i Oliva [25], entre g poblacions, definida per

$$V = \frac{1}{2g^2} \sum_{i,j=1}^g \pm_{ij}^2,$$

on \pm_{ij}^2 és la distància (al quadrat) entre les poblacions \mathbf{x}_i i \mathbf{x}_j .

Una generalització de la desigualtat (4.2) tindrà, per tant, la forma

$$\frac{1}{2g^2} \sum_{i,j=1}^g (\mathbf{x}_i - \mathbf{x}_j)^0 \mathbf{S}^{-1} (\mathbf{x}_i - \mathbf{x}_j) > \frac{1}{2g^2} \sum_{i,j=1}^g (\mathbf{x}_i - \mathbf{x}_j)^0 \mathbf{S}_0^{-1} (\mathbf{x}_i - \mathbf{x}_j). \quad (4.11)$$

Definint de manera natural la variabilitat geomètrica de Mahalanobis, i simbolitzant-la V_M , i la variabilitat geomètrica de K. Pearson, V_K , la desigualtat (4.11) la podem escriure de forma reduïda

$$V_M > V_K.$$

L'acompliment d'aquesta desigualtat indica que les mitjanes de les poblacions segueixen una direcció diferent a la de les dades observades globalment. Més concretament, les mitjanes segueixen la direcció de les components principals, obtingudes de \mathbf{S} , amb variància petita. De fet és una generalització del que es deia en l'apartat 4.2 al parlar de l'acompliment de la desigualtat (4.2) per a dues poblacions.

Un exemple que il·lustra aquesta desigualtat en el cas $g = 3$, és l'exemple de Fisher [31], clàssic en anàlisi multivariant, el qual comentarem amb més detall en l'apartat 4.7. En l'estudi gràfic d'aquest exemple, apartat 4.8, es veu clar el que hem comentat en el paràgraf anterior.

4.5 Generalització

Suposem ara $X = (X_1; \dots; X_p)$ amb densitat $f(x; \mu)$, on $\mu \in E$ és un vector q -dimensional i considerem la hipòtesi nul·la $H_0: \mu \in E_0 \subset E$.

La raó de versemblança per aquest test és

$$\lambda = \frac{L(\hat{\mu}^a)}{L(\hat{\mu})},$$

on $L(\hat{\mu}^a)$ és la funció de versemblança per l'estimador màxim versemblant restringit $\hat{\mu}^a$ de μ i $L(\hat{\mu})$ és la funció de versemblança de l'estimador màxim versemblant $\hat{\mu}$ de μ .

Es pot demostrar, Rao [56], la igualtat asimptòtica

$$-2 \log \lambda \stackrel{a}{\approx} V_{\mu^a}^0 F_{\mu^a}^{-1} V_{\mu^a},$$

on $V_{\mu^a} = \frac{1}{n} \sum_{i=1}^n Z_i(\mu^a)$, essent $Z_i(\mu^a)$ els "eficient scores" d'una mostra $x_1; \dots; x_n$, i F_{μ^a} la matriu d'informació de Fisher.

Tenint en compte, a més a més, que una generalització natural de la distància de Mahalanobis és precisament la distància de Rao, $Z^0 F^{-1} Z$, una generalització de les desigualtats que estem estudiant, (1.4), (4.2) i (4.11), és

$$-2 \log \lambda > -2 \log \lambda_1 - \dots - 2 \log \lambda_p, \quad (4.12)$$

on λ_i són les raons de versemblança per a cadascuna de les p variables estudiades. Cuadras [11] ja presentava aquesta desigualtat fent notar que el seu compliment indica una anormal connexió entre model i dades, ortogonalitat entre la direcció que segueix el model i la principal variabilitat de les dades, la qual ve donada per la primera component principal.

4.6 Desigualtat en el MANOVA

Estudiem la desigualtat (4.12) en l'anàlisi multivariant de la variància (MANOVA).

En l'anàlisi multivariant, quan ens trobem amb dades corresponents a un grup de variables mesurades en dues o més poblacions, utilitzem moltes vegades, a part de l'anàlisi canònica de poblacions i l'anàlisi discriminant, la tècnica estadística coneguda com MANOVA.

La formulació de l'anàlisi multivariant de la variància, en el cas d'un factor, és com segueix, vegeu, per exemple, Cuadras [14].

Suposem p variables observades, $Y_1; \dots; Y_p$ en g grups o poblacions. L'observació de la variable Y_u en l'individu h de la població i és

$$y_{ihu} = \mu_{iu} + e_{ihu}, \text{ amb } i = 1; \dots; g; \quad h = 1; \dots; n_i; \quad u = 1; \dots; p,$$

on $e_{ih} = (e_{ih1}; \dots; e_{ihp})^0$ són $N_p(0; \mathbf{S})$ i independents, i $\mu_i = (\mu_{i1}; \dots; \mu_{ip})^0$ és el vector de mitjanes que representa l'efecte de cada població en el model. De fet amb aquestes notacions i posant $y_{ih} = (y_{ih1}; \dots; y_{ihn})^0$ podríem escriure el model en la forma

$$y_{ih} = \mu_i + e_{ih}, \text{ amb } i = 1; \dots; g; \quad h = 1; \dots; n_i.$$

Una altra manera de presentar aquest disseny és considerar $y_{ih}; h = 1; \dots; n_i$, com una mostra aleatòria d'una $N_p(\mu_i; \mathbf{S})$; $i = 1; \dots; g$.

És usual voler contrastar la hipòtesi d'igualtat de mitjanes en les g poblacions, contra la hipòtesi alternativa de desigualtat d'algun parell de mitjanes, $H_0: \mu_1 = \dots = \mu_g$ vs. $H_1: \mu_i \neq \mu_j$; per a alguna $i \neq j$.

Per avaluar la desigualtat (4.12) necessitem la raó de versemblança.

Sabem, Cuadras [14] o Mardia, Kent i Bibby [51], que el criteri de la raó de versemblança és¹

$$\alpha = \frac{\det(W)}{\det(T)},$$

on

$$W = \prod_{h=1}^{n_i} \prod_{i=1}^g (y_{ih} - \bar{y}_i)(y_{ih} - \bar{y}_i)^0,$$

$$T = \prod_{h=1}^{n_i} \prod_{i=1}^g (y_{ih} - \bar{y})(y_{ih} - \bar{y})^0,$$

són les matrius de dispersió dintre dels grups i total respectivament, amb

$$\bar{y}_i = \frac{\sum_{h=1}^{n_i} y_{ih}}{n_i},$$

$$\bar{y} = \frac{\sum_{h=1}^{n_i} \sum_{i=1}^g y_{ih}}{N},$$

$$N = \sum_{i=1}^g n_i.$$

¹Tinguem present que $\alpha_{\text{Wilks}}^{n-2} = \alpha$.

Variable	Setosa	Versicolor	Virgínica	Total
Longitud sèpals	5:006	5:936	6:588	5:843
Amplada sèpals	3:428	2:770	2:974	3:057
Longitud pètals	1:462	4:260	5:550	3:757
Amplada pètals	0:246	1:326	2:026	1:199

Taula 4.3: Exemple de Fisher sobre ƒors del gènere Iris. Vectors de mitjanes per les tres poblacions i pel total de les dades.

Es demostra que $T = W + B$, on $B = \sum_{i=1}^g n_i (\bar{y}_i - \bar{y})(\bar{y}_i - \bar{y})^0$ és la matriu de dispersió entre grups.²

Per tant, el criteri de la raó de versemblança el podem escriure també en la forma

$$\alpha = \frac{\det(W)}{\det(W + B)} .$$

Rebutjarem la hipòtesi nul·la quan obtinguem valors signi...cativament petits de α .

4.7 Exemple de Fisher

Un cop feta aquesta petita introducció, podem passar a veure l'acompliment de la desigualtat (4.12) en un exemple, clàssic en anàlisi multivariant, degut a Fisher[31] i comentat per, entre altres, Anderson[1], Cuadras [14], Johnson i Wichern [44], Mardia, Kent i Bibby [51] i Morrison [53].

En aquest estudi sobre ƒors del gènere Iris, Fisher va observar 4 variables, longitud i amplada de sèpals i pètals ($p = 4$), en 3 espècies ($g = 3$), Virgínica, Versicolor i Setosa, agafant 50 individus de cadascuna d'elles.

Els resultats mostrals necessaris per calcular les diferents raons de versemblança i avaluar la desigualtat (4.12) apareixen a les taules 4.3, 4.4 i 4.5.

En aquestes dues últimes taules, 4.4 i 4.5, les matrius corresponents a totes les variables estan descomposades en quatre blocs per diferenciar les

²Matricialment tenim $\bar{Y}_i = \frac{1}{n_i} \mathbf{1}_{n_i}^0 Y_i$; $W = \sum_{i=1}^g Y_i^0 H_{n_i} Y_i$, $T = Y^0 H_n Y$ i $B = \sum_{i=1}^g n_i (\bar{Y}_i - \bar{Y})(\bar{Y}_i - \bar{Y})^0$.

		O \tilde{A}		I \tilde{A}		1	
$W =$	\tilde{A}	W_{11}	W_{12}	38:956	13:630	24:703	5:645
		W_{21}	W_{22}	13:630	16:962	8:148	4:808
				24:703	8:148	27:322	6:284
				5:645	4:808	6:284	6:156

Taula 4.4: Exemple de Fisher. Matrius de dispersió dintre els grups.

		O \tilde{A}		I \tilde{A}		1	
$B =$	\tilde{A}	B_{11}	B_{12}	63:212	19:953	165:17	71:278
		B_{21}	B_{22}	19:953	11:3452	57:23	22:932
				165:17	57:23	436:73	186:69
				71:278	22:932	186:69	80:413

Taula 4.5: Exemple de Fisher. Matrius de dispersió entre els grups.

variables corresponents als sèpals de les corresponents als pètals, per aquest ordre.

En tot aquest estudi hem de fer notar que, Cuadras [14], el test d'homogeneïtat entre les matrius de covariàncies resulta signi...catiu.

Avaluem la desigualtat (4.12) en diferents casos. Primer considerant les quatre variables, per després estudiar per separat les variables corresponents al sèpals i als pètals. Els comentaris es faran en l'apartat 4.8, en estudiar grà...cament aquest exemple.

Per a les quatre variables tenim

$$\alpha = \frac{\det(W)}{\det(W + B)} = 0:02344,$$

mentre que considerant-les individualment es té

$$\alpha_1 = \frac{w_{11}}{(w_{11} + b_{11})} = 0:38129 ,$$

$$\alpha_2 = \frac{w_{22}}{(w_{22} + b_{22})} = 0:59922 ,$$

$$\alpha_3 = \frac{w_{33}}{(w_{33} + b_{33})} = 0:05863 ,$$

$$\alpha_4 = \frac{w_{44}}{(w_{44} + b_{44})} = 0:07112 ,$$

on $\alpha_i = \alpha(x_i)$, $i = 1; \dots; 4$, corresponen a les quatre variables estudiades.

Per tant trobem en aquest cas la desigualtat

$$i \ 2 \log \alpha = 7:5066 < 13:913 = i \ 2 \log \alpha_1 \ i \ \dots \ i \ 2 \log \alpha_4,$$

i no s'acompleix (4.12).

Considerem, però, per separat, les dues variables corresponents als sèpals i als pètals, és a dir, les dues primeres variables d'una banda i les dues últimes de l'altra i tornem a avaluar la desigualtat (4.12).

Tenim en el primer cas, considerant les variables corresponents als sèpals,

$$\alpha_{12} = \frac{\det(W_{12})}{\det(W_{12} + B_{12})} = 0:16654,$$

d'on

$$i \ 2 \log \alpha_{12} = 3:585 > 2:9526 = i \ 2 \log \alpha_1 \ i \ 2 \log \alpha_2.$$

En aquest cas, per tant, sí que s'acompleix la desigualtat (4.12).

En el segon cas, considerant solament les dues últimes variables, les corresponents als pètals, els resultats serien

$$\alpha_{34} = \frac{\det(W_{34})}{\det(W_{34} + B_{34})} = 0:0438 ,$$

d'on

$$i \ 2 \log \alpha_{34} = 6:2562 < 10:96 = i \ 2 \log \alpha_3 \ i \ 2 \log \alpha_4.$$

Per tant el que tenim en aquest cas és la desigualtat contrària.

També podem veure com aquest exemple il·lustra la desigualtat (4.11) com ja avançàvem en la secció 4.4. Per a les variables dels sèpals, per a les quals s'acompleix la desigualtat (4.2) tenim

$$S = \begin{matrix} \mathbf{\tilde{A}} & \mathbf{!} \\ 0:2650 & 0:0927 \\ 0:0927 & 0:1154 \end{matrix} ,$$

essent els vectors de mitjanes els corresponents de la taula 4.3. Es veri...ca, per tant

$$V_M = 8:49 > 4:49 = V_K .$$

Per últim, en la taula 4.6 està avaluada la desigualtat (4.2) per a cada parell de poblacions. La desigualtat (4.2) no es veri...ca al considerar les dues variables corresponents al pètals. En els altres casos, considerant totes les variables o només les variables corresponents als sèpals, la desigualtat es veri...ca sempre que una de les dues poblacions sigui la espècie setosa.

	Totes les variables
Setosa-Versicolor	92:9 = $\pm_M^2 > \pm_K^2 = 75:0$
Setosa-Virgínica	195:1 = $\pm_M^2 > \pm_K^2 = 184:1$
Versicolor-Virgínica	14:2 = $\pm_M^2 < \pm_K^2 = 16:6$
	Dues variables corresponents als sèpals
Setosa-Versicolor	19:9 = $\pm_M^2 > \pm_K^2 = 8:0$
Setosa-Virgínica	21:5 = $\pm_M^2 > \pm_K^2 = 11:1$
Versicolor-Virgínica	1:3 = $\pm_M^2 < \pm_K^2 = 1:7$
	Dues variables corresponents als pètals
Setosa-Versicolor	62:5 = $\pm_M^2 < \pm_K^2 = 67:0$
Setosa-Virgínica	131:6 = $\pm_M^2 < \pm_K^2 = 173:0$
Versicolor-Virgínica	10:1 = $\pm_M^2 < \pm_K^2 = 14:9$

Taula 4.6: Avaluació de la desigualtat per dues poblacions en l'exemple de Fisher.

4.8 Estudi gràfic de l'exemple de Fisher

4.8.1 Variables primera i segona corresponents als sèpals

Hem vist en l'apartat 4.7 com les desigualtats (4.11) i (4.12) s'acompleixen per a les variables longitud i amplada dels sèpals mesurades en les tres poblacions considerades: setosa, versicolor i virgínica.

En la figura 4.4 observem com, al representar els individus i les dues components principals de cada població, eix llarg per a la primera component principal i eix curt per a la segona, tallant-se en la mitjana de cada població, en l'espai de les dues variables considerades ara, les corresponents als sèpals, la direcció que segueixen les mitjanes considerant les poblacions setosa i versicolor i també setosa i virgínica, és la de les segones components principals. De fet, es verifica la desigualtat (4.2) en ambdós casos, com podem veure a la taula 4.6. La direcció que segueixen les mitjanes considerant les poblacions versicolor i virgínica és la de les primeres components principals. En aquest cas la desigualtat (4.2) no es verifica, taula 4.6.

En general, considerant les variables corresponents als sèpals i les tres poblacions, podem dir que les mitjanes no segueixen la direcció de les dades

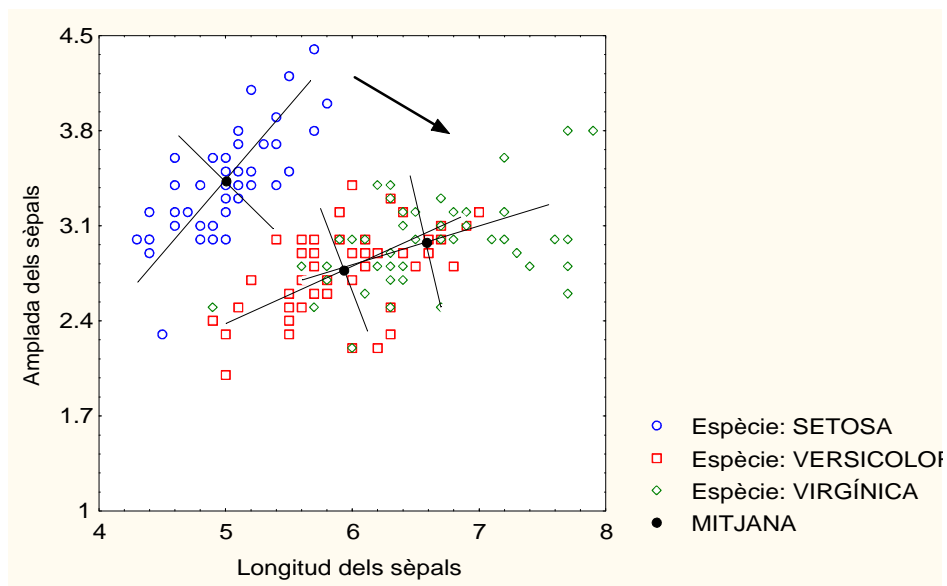


Figura 4.4: Exemple de Fisher sobre ƒors del gènere *Iris* considerant només les dues variables corresponents als sèpals. La desigualtat (4.12) es verifica per als aparellaments en els quals intervé la població setosa. El segment que uneix la mitjana de la població setosa amb cadascuna de les altres poblacions tendeix a seguir la direcció de la segona component principal (eix curt).

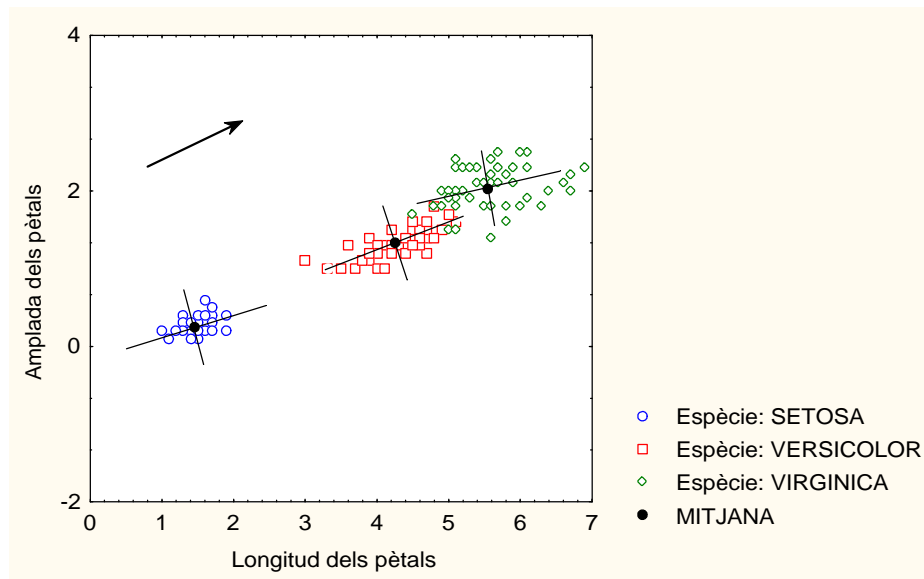


Figura 4.5: Exemple de Fisher sobre ƒors del gènere Iris considerant només les dues variables corresponents als pètals. La desigualtat (4.12) no es veri...ca. El segment que uneix les mitjanes de les poblacions segueix la direcció de la primera component principal (eix llarg).

considerades totes en conjunt.

Com ja es deia en l'apartat 4.4, les mitjanes de les poblacions segueixen una direcció diferent a la direcció del conjunt de totes les dades, ...gura 4.4.

4.8.2 Variables tercera i quarta corresponents als pètals

Hem vist també en l'apartat 4.7 com les desigualtats (4.11) i (4.12) no s'a-compleixen per a les variables longitud i amplada dels pètals mesurades en les tres poblacions considerades: setosa, versicolor i virgínica.

En la ...gura 4.5 observem com les mitjanes segueixen clarament la direcció de les primeres components principals, les de variància més gran, la direcció de les quals és la direcció al llarg de la qual hi ha màxima dispersió dels individus. En aquest cas, per a tots els aparellaments possibles de dues poblacions, entre les tres considerades, la desigualtat (4.2) no es veri...ca, taula 4.6.

Tot plegat conforma l'afirmació de Cuadras [11], en el sentit que el model segueix una direcció ortogonal a la variabilitat principal de les dades quan la desigualtat (4.12) s'acompleix, indicant una anormal connexió entre model i dades.

4.9 Representació d'una component principal a l'espai canònic

Si volem representar els individus i les components principals de cada població, considerant les quatre variables estudiades per Fisher, hem d'utilitzar la representació a l'espai canònic.

A continuació s'explica com obtenir la representació de les components principals a l'espai canònic.

Siguin X , i x_1, x_2, \dots, x_p , la matriu de dades i les variables observables, respectivament.

Considerem també

$$z_1 = t_1 x_1 + \dots + t_p x_p,$$

la primera component principal i V , la matriu de transformació canònica.

Podem representar les variables x_i ; $i = 1; \dots; p$, per les matrius de dades, punts alineats en l'espai dels individus,

$$x_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{matrix} r_{1i} \\ r_{2i} \\ \vdots \\ r_{ni} \end{matrix} \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{matrix} 1 \\ \vdots \\ 1 \end{matrix} A'$$

on $r_{1i}, r_{2i}, \dots, r_{ni}$ és la partició

$$r_i = r_{1i} \cdot r_{1i} \cdot \dots \cdot r_{ni} \cdot R_i,$$

essent r_i i R_i els valors mínim i màxim que pren la variable x_i .

En aquestes condicions podem representar qualsevol component principal, en particular la primera z_1 , per la matriu de dades en l'espai dels individus

$$e_1 = t_1 x_1 + \dots + t_p x_p = \begin{pmatrix} 0 \\ \vdots \\ t_1^{r_{11}} \\ \vdots \\ t_1^{r_{n1}} \end{pmatrix} \begin{matrix} t_1^{r_{12}} & t_1^{r_{12}} & \vdots \\ t_1^{r_{21}} & t_1^{r_{22}} & \vdots \\ \vdots & \vdots & \vdots \\ t_1^{r_{n1}} & t_1^{r_{n2}} & \vdots \end{matrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} A'$$

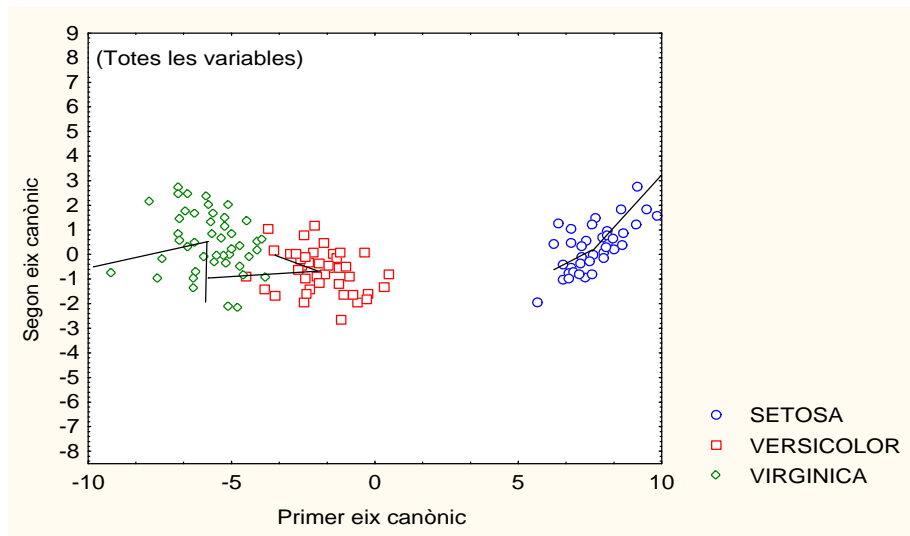


Figura 4.6: Representació canònica dels individus, les mitjanes de les poblacions i de les dues components principals en l'exemple de Fisher, considerant totes les variables. La desigualtat (4.12) no es verifica. En les poblacions versicolor i virgínica la primera component principal és paral·lela al primer eix canònic, mentre que en la població setosa és la segona component principal la que gairebé és paral·lela al primer eix canònic. Recordem que considerant la població setosa amb cadascuna de les altres dues poblacions la desigualtat (4.2) es verifica, mentre que considerant només les poblacions versicolor i virgínica no es verifica.

$$\text{on } n = \sum_{i=1}^p n_i .$$

Finalment la representació de la primera component principal sobre l'espai canònic és la successió de punts alineats, ...les de la matriu

$$e_1 V .$$

Apliquem aquest resultat a l'estudi conjunt de les quatre variables estudiades per Fisher.

Recordem que l'espai canònic diferencia grups i que els eixos tenen màxim poder de discriminació en relació a la distància de Mahalanobis. És a dir, l'anàlisi canònica representa les poblacions de manera òptima, fent la distància de Mahalanobis igual a la distància euclidiana, al llarg d'uns eixos

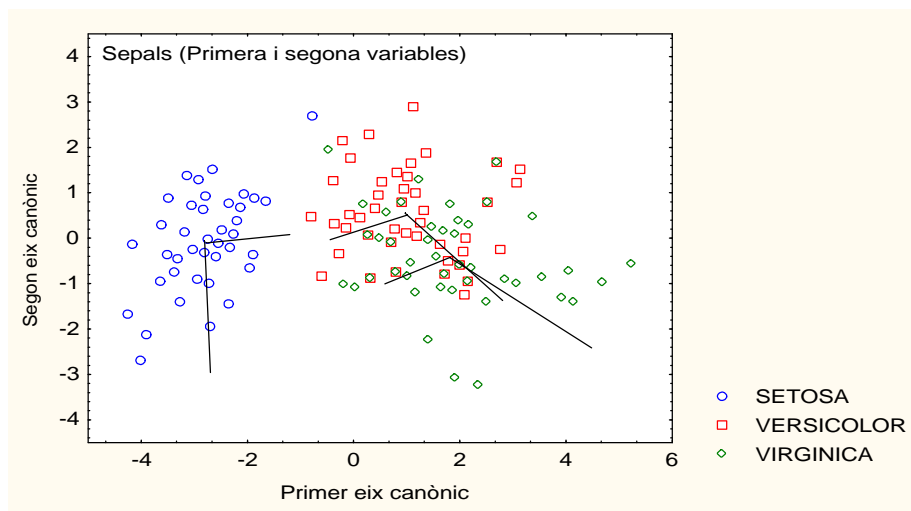


Figura 4.7: Representació canònica dels individus, les mitjanes i les components principals en l'exemple de Fisher considerant les dues variables corresponents als sèpals. La desigualtat (4.12) es verifica. Les segones components principals són gairebé paral·leles al primer eix canònic, la direcció del qual és la de màxima diferència entre poblacions. Aleshores tenim la desigualtat distància de Mahalanobis més gran que distància de K. Pearson.

ortogonals, anomenats canònics. El primer eix canònic és la direcció de màxima diferència entre grups.

En la figura 4.6 estan representats a l'espai canònic tots els individus de les tres poblacions considerant les quatre variables, així com les components principals. Recordem, secció 4.7, que al considerar les quatre variables la desigualtat (4.12) no es verifica. En les poblacions versicolor i virgínica la primera component principal és paral·lela al primer eix canònic, mentre que en la població setosa és la segona component principal la que gairebé és paral·lela al primer eix canònic. Recordem que considerant la població setosa amb cadascuna de les altres dues poblacions la desigualtat (4.2) es verifica, mentres que considerant només versicolor i virgínica no es verifica, taula 4.6.

Per veure el que està passant representem en l'espai canònic tots els individus de les tres poblacions i les seves components principals però considerant només les variables corresponents als sèpals en primer lloc, figura 4.7, i només les variables corresponents als pètals en segon lloc, figura 4.8. Recordem, sec-

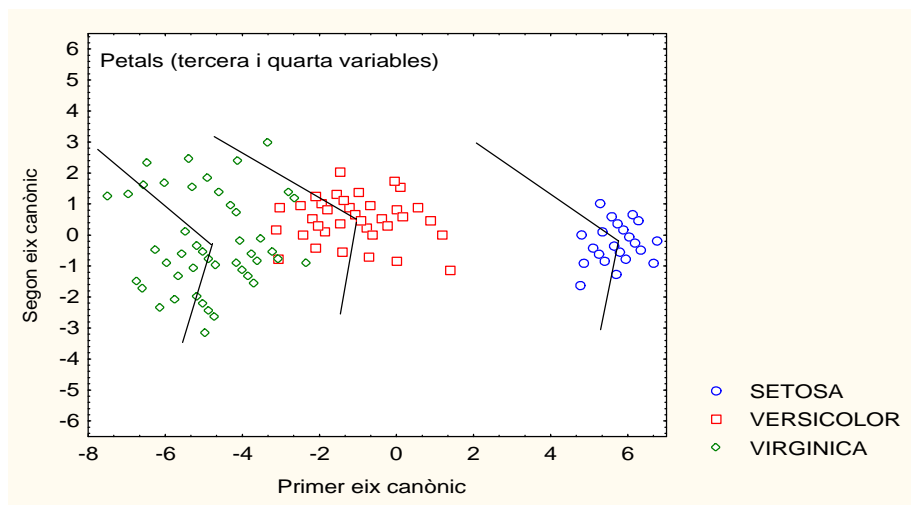


Figura 4.8: Representació canònica dels individus, les mitjanes i les components principals en l'exemple de Fisher considerant les dues variables corresponents als pètals. La desigualtat (4.12) no es verifica. Les primeres components principals, màxima variabilitat de les dades, formen un angle agut amb el primer eix canònic, màxima variabilitat de les poblacions. Aleshores tenim que distància de Mahalanobis més petita que distància de K. Pearson.

ció 4.7, que la desigualtat (4.12) es verifica en el primer cas i no en el segon.

Ara la situació és clara. En la figura 4.7 les segones components principals són gairebé paral·leles al primer eix canònic, en canvi en la figura 4.8 són les primeres components principals les que tenen una direcció més semblant a la del primer eix.

Podem dir, per tant, que l'acompliment de les desigualtats (4.2), (4.11) i (4.12) va lligada al fet que la màxima variabilitat de les dades, representada per les components principals, no tingui la mateixa direcció que la màxima variabilitat de les poblacions representada pel primer eix canònic.

4.10 Altres exemples

A continuació fem referència a dos exemples en els quals també es verifica la desigualtat (4.12).

Exemple 4.1

Mardia, Kent i Bibby [51], estudien les mesures per dues variables en dues poblacions, dues espècies relacionades d'escarabats petits, *Chaetocnema concinna* i *Chaetocnema heikertingeri* habitualment confoses una amb l'altra. La primera variable és la mida en micròmetres de les primeres articulacions de les dues primeres potes. La segona variable és la corresponent suma per a les segones articulacions.

En el gràfic 4.9 s'observa com les segones components principals de cada població segueixen gairebé la direcció de les mitjanes de les poblacions.

La direcció de les dades i el model són ortogonals.

Exemple 4.2

Arenas, Cuadras i Fortiana [3], estudien les dades corresponents a les mesures de l'alçada basibregmàtica i l'alçada nasal, en mil·límetres, de trenta cranis egipcis per cadascun dels cinc períodes estudiats, exemple degut a Thomson and Randall-Maciver, 1905, (veure Manly [49]). En el gràfic 4.10, és interessant observar les direccions de les components principals en referència a les mitjanes de cada període.

4.11 La paradoxa de Rao

Rao [55] comenta unes dades que donarien lloc a una paradoxa, coneguda com la "maledicció de la dimensionalitat". Un estudi d'aquesta paradoxa es deu a Healy [41].

Rao presenta les següents mides mostrals, mitjanes i desviacions típiques de dues variables x (longitud del fèmur), y (longitud del húmer), obtingudes sobre dues poblacions (Anglo-indis, Indis),

Mitjanes	x	y
$n_1 = 27$	460.4	345.1
$n_2 = 20$	444.3	333.2
Diferència	16.1	11.9
Desv. típiques	23.7	18.2

Les desviacions típiques són estimacions sobre les 47 dades. El coeficient de

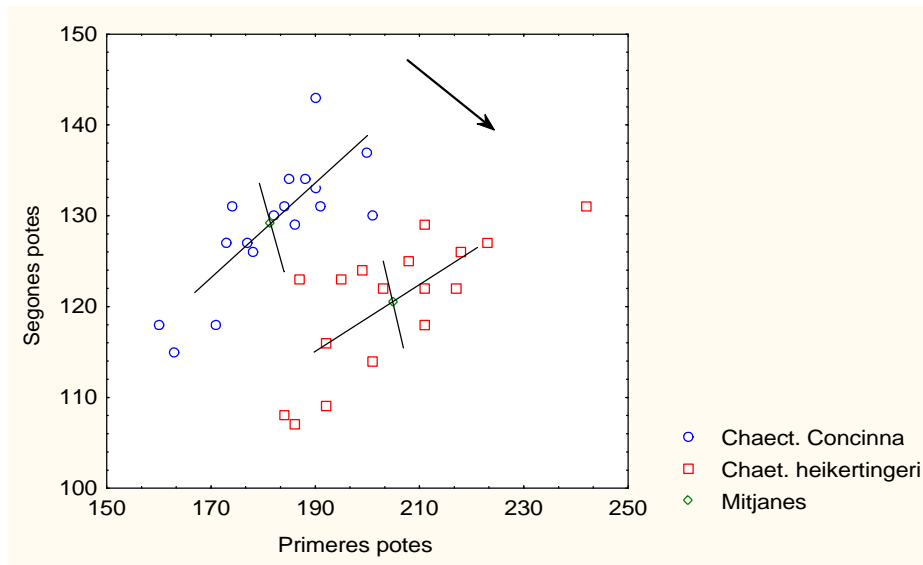


Figura 4.9: Exemple en el que la desigualtat (4.12) s'acompleix. La direcció del segment que uneix les mitjanes més aviat segueix la segona component principal.

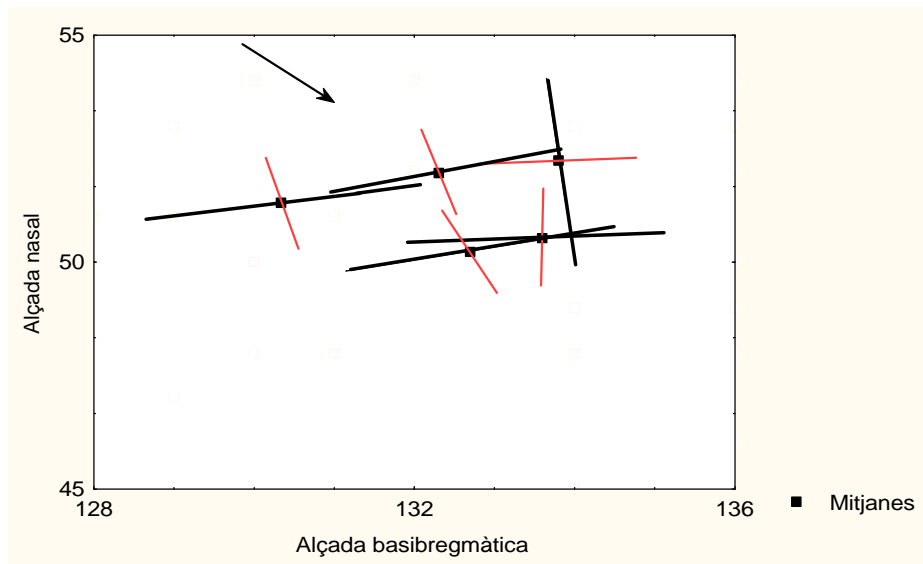


Figura 4.10: Exemple dels cranis egipcis trobats en cinc períodes, en el que la desigualtat (4.12) s'acompleix. La direcció principal de les mitjanes segueix gairebé la direcció de la segona component principal (eix petit).

correlació entre x ; y és

$$r = 0:8675$$

i l'estimació conjunta de la matriu de covariàncies és

$$S = \begin{pmatrix} \tilde{A} & \\ & ! \end{pmatrix} = \begin{pmatrix} 561:7 & 374:2 \\ 374:2 & 331:24 \end{pmatrix} :$$

Suposant normalitat i homogeneïtat de variàncies-covariàncies, volem comparar les dues poblacions. Ho farem primer amb les dues variables per separat (test univariant) i després amb les dues variables alhora (test bivariant).

Primera variable. El test t (45 graus de llibertat) dona

$$t = \frac{r}{\frac{n_1 n_2}{n_1 + n_2}} \frac{(\bar{x}_1 - \bar{x}_2)}{s_1} = 2:303, \quad (p = 0:0260).$$

Segona variable. El test t (45 graus de llibertat) dona

$$t = \frac{r}{\frac{n_1 n_2}{n_1 + n_2}} \frac{(\bar{y}_1 - \bar{y}_2)}{s_2} = 2:216, \quad (p = 0:0318).$$

Dues variables. El test T^2 de Hotelling, on $p = 2$ i $D = (\bar{x}_1 \quad \bar{x}_2; \bar{y}_1 \quad \bar{y}_2)^0$ dona

$$\frac{n_1 + n_2 - 1}{(n_1 + n_2 - 2)p} \frac{1}{n_1 n_2} D^0 S^{-1} D = 2:69 \gg F_{n_1 + n_2 - 1, p}^p \quad (p = 0:079)$$

que segueix una F amb 2 i 44 graus de llibertat.

Per tant, els tests univariants són significatius, però el test bivariant no ho és, contradint la creença que un test multivariant hauria de ser més significatiu que un test univariant.

Intentem interpretar aquesta paradoxa des del punt de vista de la desigualtat (4.1). El quadrat de la distància de Mahalanobis sobre les dues

variables és

$$\begin{aligned} \pm_M^2 &= \frac{D^0 S_1^{-1} D}{3} \begin{matrix} \tilde{A} \\ 561:7 & 374:2 \\ 374:2 & 331:24 \end{matrix} \begin{matrix} ! \\ i \\ 1 \\ \tilde{A} \\ ! \end{matrix} \\ &= \frac{16:1}{11:9} \begin{matrix} 561:7 & 374:2 \\ 374:2 & 331:24 \end{matrix} \begin{matrix} 16:1 \\ 11:9 \end{matrix} \\ &= 0:4783 . \end{aligned}$$

El quadrat de la distància de K. Pearson és

$$\begin{aligned} \pm_K^2 &= \frac{D^0 S_0^{-1} D}{3} \begin{matrix} \tilde{A} \\ 561:7 & 0 \\ 0 & 331:24 \end{matrix} \begin{matrix} ! \\ i \\ 1 \\ \tilde{A} \\ ! \end{matrix} \\ &= \frac{16:1}{11:9} \begin{matrix} 561:7 & 0 \\ 0 & 331:24 \end{matrix} \begin{matrix} 16:1 \\ 11:9 \end{matrix} \\ &= 0:8890 . \end{aligned}$$

Evidentment,

$$\pm_M^2 < \pm_K^2 :$$

Per tant: el vector que uneix les dues poblacions tendeix a seguir la direcció de la primera component principal.

En efecte, el primer vector propi de la matriu S amb mòdul unitat és

$$U_1 = (0:8045; 0:5940)^0$$

i la primera component principal és

$$Y_1 = 0:8045x + 0:5940y:$$

El vector de diferències és $D = (16:1; 11:9)$: El cosinus de l'angle entre els dos vectors dóna

$$\cos \mu = \frac{D \uparrow U}{\|D\| \|U\|} = 1$$

que correspon a l'angle $\mu = 0$: En aquest cas, sorprenentment, el vector diferència segueix exactament la direcció de la primera component principal.

També veiem que la distància bivariant gairebé no augmenta al passar d'una a dues variables. En efecte, la distància bivariant hem vist que és 0:4783 , mentre que les distàncies univariants són

$$\pm_1^2 = \frac{(x_1 - \bar{x}_1)^2}{s_1^2} = 0:4614; \quad \pm_2^2 = \frac{(y_1 - \bar{y}_1)^2}{s_2^2} = 0:4270:$$

Seguidament, interpretem geomètricament la paradoxa de Rao.

Per al test bivariant, amb un nivell de significació 0.05, acceptarem la hipòtesi nul·la si el vector diferència $D = (\bar{x}, \bar{y})$, amb $\bar{x} = \bar{x}_1$ i \bar{x}_2 i $\bar{y} = \bar{y}_1$ i \bar{y}_2 ; pertany a l'el·lipse

$$5.62 \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} \bar{A} \\ \bar{A} \end{pmatrix} \begin{pmatrix} 561:7 & 374:2 \\ 374:2 & 331:24 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 3:2;$$

on 3.2 és el punt crític per a una F amb 2 i 44 graus de llibertat. Així, no hi ha significació si x i y varien

$$0:040369x^2 + 0:09121xy + 0:068456y^2 = 3:2;$$

Anàlogament, per a la primera variable, la diferència \bar{x} ha de verificar

$$\frac{\bar{r}}{n_1 + n_2} \frac{n_1 n_2 (\bar{x}_1 - \bar{x}_2)}{s_1} = 2;$$

on 2 és el punt crític per a una t amb 45 graus de llibertat. Això ens dona les rectes

$$0:143\bar{x} = 2; \quad 0:143\bar{x} = -2;$$

Per a la segona variable obtenim les rectes

$$0:1862\bar{y} = 2; \quad 0:1862\bar{y} = -2;$$

La figura 4.11 il·lustra aquestes regions. Podem veure que només el cantó inferior-esquerra i el cantó superior-dret pot donar significació per a x; significació per a y, però no per a (x; y) conjuntament. La diferència, assenyalada amb una +, cau just sobre la primera component principal al cantó dret-dalt de la figura.

També veiem que seguint la direcció de la segona component principal

$$Y_2 = 0:5940x + 0:8045y,$$

els cantons esquerra-dalt i dreta-baix, podem trobar significació per a cada variable per separat, però aleshores hi haurà també significació per a les dues variables conjuntament. Tot això ens porta al següent resultat.

Resultat 6 En el cas bivariant, la paradoxa de Rao es dona si el vector diferència entre les dues poblacions tendeix a seguir la direcció de la primera component principal, i no es pot donar si segueix la direcció de la segona component principal. Per tant, no es pot donar si $\pm \frac{D_1}{M} > \pm \frac{2}{K}$.

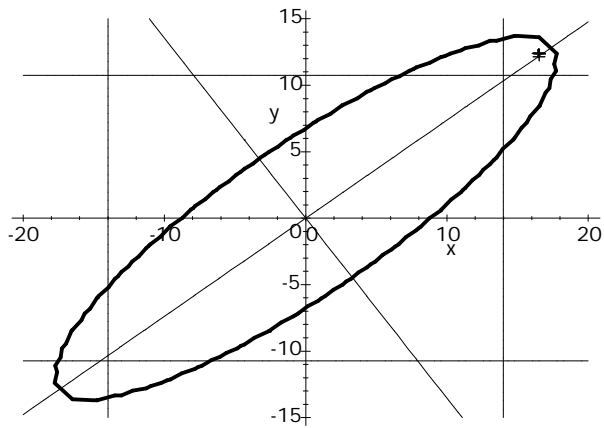


Figura 4.11: Explicació de la paradoxa de Rao. Només els valors sobre el cantó esquerra-baix i el cantó dreta-dalt (com l'assenyalat amb +), poden donar signi...cació per a x; signi...cació per a y, però no per a (x; y) conjuntament.

Nota: La frase "tendeix a seguir la direcció" vol dir que els vectors directors formen un angle agut, i queda precisada pel compliment de la desigualtat.

Capítol 5

Densitats a partir de distàncies

5.1 Introducció

Cuadras, Atkinson i Fortiana [19] consideren la possibilitat de construir funcions de densitat a partir de distàncies i estudien mesures de divergència entre aquesta densitat i la vertadera densitat. Això pot tenir aplicació en anàlisi discriminant al poder utilitzar distàncies en lloc de funcions de densitat.

Així mateix Cuadras [16] introdueix la noció de distància conjunta definida a partir de dues distàncies, diguem, marginals, per analogia amb la construcció de distribucions de probabilitat amb marginals donades.

En aquest capítol començarem definint les nocions de variabilitat geomètrica i funció de proximitat per a variables aleatòries, introduïdes per Cuadras i Fortiana [22] i Cuadras, Fortiana i Oliva [25].

Tractarem després el tema de la construcció de funcions de densitat en el cas univariant per acabar amb l'extensió natural al cas bivariant.

5.2 Variabilitat geomètrica

En la secció 4.4 havíem definit la variabilitat geomètrica com una generalització natural de la noció de distància al cas de g poblacions multivariants. En aquesta secció definim la noció de variabilitat geomètrica en el cas d'una variable aleatòria.

Sigui X una variable aleatòria definida en un espai de probabilitat $(\Omega; \mathcal{A}; P)$ amb valors a $E \subseteq \mathbb{R}$ i amb funció de distribució $F(x)$ absolutament contínua

i densitat $f(x)$ respecte la mesura de Lebesgue.

Sigui també $\pm : E \times E \rightarrow \mathbb{R}$ una distància, és a dir, per a $x, y, z \in E$:

1. $\pm(x; x) = 0$;
2. $\pm(x; y) = \pm(y; x) \geq 0$;
3. $\pm(x; y) + \pm(y; z) \geq \pm(x; z)$;

La variabilitat geomètrica de la variable aleatòria X respecte la distància \pm , $V_{\pm}(X)$, es defineix com

$$\begin{aligned} V_{\pm}(X) &= \frac{1}{2} \int_{E \times E} \int_{E \times E} \pm^2(x; y) dF(x) dF(y) \\ &= \frac{1}{2} \int_{E \times E} \pm^2(x; y) f(x) f(y) dx dy, \end{aligned}$$

sempre que aquesta integral existeixi.

Nota: La definició de variabilitat geomètrica també s'aplica a una dissimilaritat, és a dir, a una \pm tal que $\pm(x; y) = \pm(y; x) \geq \pm(x; x) = 0$:

La quantitat $2V_{\pm}(X)$ és l'esperança de la funció distància al quadrat, avaluada en dues variables aleatòries independents i igualment distribuïdes.

De fet, podem entendre $V_{\pm}(X)$ com una generalització de la variància de la variable aleatòria X . En efecte, la variabilitat geomètrica de la variable aleatòria X respecte la distància

$$\pm(x; y) = |x - y|,$$

és

$$\begin{aligned} V_{\pm}(X) &= \frac{1}{2} \int_{E \times E} \int_{E \times E} (x - y)^2 f(x) f(y) dx dy \\ &= \frac{1}{2} \int_{E \times E} (x - y)^2 f(x) f(y) dx dy \\ &= \text{var}(X). \end{aligned}$$

Considerem ara la funció $u : E \times E \rightarrow [0; 1]$ definida per

$$u(t; x) = \begin{cases} \frac{1}{2} & \text{si } t < x \\ 0 & \text{si } t \geq x \end{cases}$$

Sigui el procés estocàstic $X = \{X_t\}_{t \in E}$, amb $X_t = u(t; X)$, per $t \in E$. És a dir, per qualsevol $t \in E$, X_t és l'indicador de l'esdeveniment $[X > t] \in \mathcal{A}$, i es distribueix com una variable aleatòria Bernoulli amb probabilitat $p = 1 - F(t)$.

La funció de covariància d'aquest procés ve donada per

$$\text{Cov}(X) = K(s; t) = \min\{F(s), F(t)\} - F(s)F(t), \quad (5.1)$$

amb traça

$$\begin{aligned} \text{tr}(K) &= \int_E K(s; s) ds \\ &= \int_E F(s)(1 - F(s)) ds. \end{aligned}$$

Cuadras i Fortiana [22] estableixen el següent resultat.

Resultat 7 Sigui la variable aleatòria X tal que $\lim_{s \rightarrow \infty} sF(s) = 0$, i considerem la distància $\pm(x; y) = \frac{p}{jx - yj}$. Aleshores tenim que la variabilitat geomètrica de la variable aleatòria X respecte aquesta distància és

$$V_{\pm}(X) = \text{tr}(K). \quad (5.2)$$

Per tant, la variabilitat geomètrica de la variable X respecte la distància $\pm(x; y) = \frac{p}{jx - yj}$, és la traça del procés X . És, en definitiva, una mesura de la dispersió de la variable X .

En la taula 5.1 podem veure les variabilitats geomètriques per diferents distribucions en el cas de les distàncies $\pm(x; y) = \frac{p}{jx - yj}$, cas en el qual coincidirà amb la $\text{var}(X)$ com hem vist abans, i $\pm(x; y) = \frac{p}{jx - yj}$.

5.3 Desigualtat per a la variabilitat geomètrica

El següent resultat ens dona una cota inferior i una cota superior per a la variabilitat geomètrica d'una variable.

Resultat 8 Sigui X una variable aleatòria, amb mitjana μ i variància σ^2 donades en un espai de probabilitat $(\Omega; \mathcal{A}; P)$ amb valors a $E \subseteq \mathbb{R}$

Distribució	Variabilitat geomètrica	Variància
Qualsevol [f(x) = F'(x)]	$\int_E (1 - F(t)) F'(t) dt$	var(X)
Normal $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$\frac{3}{4}$	$\frac{3}{4}\sigma^2$
Uniforme [1]	$\frac{1}{6}$	$\frac{1}{12}$
Exponencial [$\lambda \exp(-\lambda x)$]	$\frac{1}{2\lambda}$	$\frac{1}{\lambda^2}$
Logística $\frac{f(x)}{2} (e^{-x} + e^{i-x})^{-2}$	$\frac{1}{2}$	$\frac{1}{12}$
Laplace $\frac{1}{2} e^{-\lambda x }$	$\frac{3}{4}$	$\frac{2}{\lambda^2}$

Taula 5.1: Variabilitat geomètrica per a diferents distribucions i per a la distància arrel quadrada de la distància euclídea ordinària.

i amb funció de distribució $F(x)$ absolutament contínua i densitat $f(x)$ respecte la mesura de Lebesgue. Sigui $V_{\pm}(X)$ la variabilitat geomètrica d'aquesta variable respecte la distància $\pm(x; y) = \sqrt{|x - y|}$. Aleshores

$$\frac{1}{6} \leq \inf_{x \in E} \frac{f(x)}{f(x)} \cdot V_{\pm}(X) \leq \frac{3}{4} \quad (5.3)$$

En efecte, d'una banda tenim que, Grané [35],

$$\frac{1}{2}^+(X; U) = \frac{3}{4} V_{\pm}(X), \quad (5.4)$$

essent $\frac{1}{2}^+(X; U)$ la correlació màxima de Hörding, secció 6.3, entre les variables X i U , aquesta última distribuïda uniformement en $[0; 1]$. Podem observar que aquesta equació és equivalent a

$$\frac{1}{2}(X; F(X)) = \frac{3}{4} V_{\pm}(X) :$$

En particular tenim que

$$\frac{3}{4} V_{\pm}(X) .$$

D'altra banda també es verifica, Cuadras [13], la desigualtat

$$\operatorname{ess\,inf}_{x \in E} \frac{f(x)^{1/2}}{g(x)^{3/4}} \cdot \frac{\sigma_X^{3/4}}{\sigma_Y^{1/4}} \geq \frac{1}{2} d_{\pm}(X; Y), \quad (5.5)$$

on Y és una altra variable aleatòria amb densitat $g(y)$ i essent σ_X i σ_Y les desviacions típiques de les variables X i Y respectivament.

Com $g(y) = 1$ i $\sigma_Y = \frac{1}{\sqrt{12}}$, suposant Y distribuïda uniformement en $[0; 1]$, i fent $\sigma_X = \frac{1}{6}$, tenim

$$\operatorname{ess\,inf}_{x \in E} f(x)g \cdot \frac{\sigma_X^{3/4}}{\sigma_Y^{1/4}} \geq \frac{1}{2} d_{\pm}(X; Y),$$

i, tenint en compte la igualtat (5.4) podem escriure

$$\frac{\sigma_X^{3/4}}{\sigma_Y^{1/4}} \geq \frac{1}{2} d_{\pm}(X; Y) \cdot \frac{\sigma_Y^{1/4}}{\sigma_X^{3/4}} = \frac{1}{6} \frac{\sigma_Y^{1/4}}{\sigma_X^{3/4}} V_{\pm}(X) = 6 V_{\pm}(X),$$

és a dir

$$\frac{1}{6} \operatorname{ess\,inf}_{x \in E} f(x)g \cdot V_{\pm}(X).$$

Finalment, considerant un altre cop la igualtat (5.4) obtenim (5.3).

Quan la variància és finita, la variabilitat geomètrica també ho és i està acotada tal com indiquen les desigualtats anteriors.

Exemple 5.1

En el cas en que X estigui uniformement distribuïda en $[0; 1]$ tenim, al ser $f(x) = 1$,

$$\frac{1}{6} \cdot V_{\pm}(X) \cdot \frac{1}{6},$$

la qual cosa concorda amb el valor de la variabilitat geomètrica per a la distància $d_{\pm}(x; y) = \frac{1}{2} |x - y|$ d'una variable aleatòria distribuïda uniformement en $[0; 1]$.

Exemple 5.2

Considerem la variable aleatòria exponencial X amb funció de densitat $f(x) = \lambda \exp(-\lambda x)$, amb $\lambda > 0$ i $x > 0$. Recordem que $\operatorname{var}(X) = \frac{1}{\lambda^2}$.

Per a la distància $\pm(x; y) = \int_{jx}^{jy} \dots$ sabem que la variabilitat geomètrica és, (5.2),

$$\begin{aligned} V_{\pm}(X) &= \int_{jx}^{jy} (1 - F(x)) F(x) dx \\ &= \int_{jx}^{jy} (\exp(-jx)) (1 - \exp(-jx)) dx \\ &= \frac{1}{2j}. \end{aligned}$$

La desigualtat (5.3) es verifica, ja que

$$\frac{1}{6} < 0 \cdot \frac{1}{2j} \cdot \frac{1}{3} < 0 \cdot \frac{1}{2j} \cdot \frac{1}{3}.$$

5.4 Expansions ortogonals

Considerarem en tota aquesta secció la distància $\pm(x; y) = \int_{jx}^{jy} \dots$.

Sigui X una variable aleatòria, amb mitjana i variància finites, definida en un espai de probabilitat $(\Omega; \mathcal{A}; P)$ amb valors a $E = [a; b] \subset \mathbb{R}$, funció de distribució $F(x)$ absolutament contínua i densitat $f(x)$ respecte la mesura de Lebesgue. Per (5.3) la variabilitat geomètrica també serà finita. Suposarem en tot aquest apartat que es verifica la condició $\lim_{y \rightarrow a} yF(y) = 0$.

Sabem per Cuadras i Fortiana [22] que existeix una successió completa, $\{f_n\}_{n \in \mathbb{N}}$ de funcions pròpies ortonormals, solucions de

$$\int_E K(s; t) \tilde{A}_n(t) dt = \lambda_n \tilde{A}_n(s) \quad \text{amb } s \in E,$$

i la corresponent successió de valors propis, $\{\lambda_n\}_{n \in \mathbb{N}}$, essent $K(s; t)$ la covariància del procés X definida en (5.1).

Així mateix fent

$$\begin{aligned} h_n(x) &= \int_a^x \tilde{A}_n(t) dt \quad \text{amb } x \in E, \\ X_n &= h_n(X) \quad \text{amb } n = 1, 2, \dots, \end{aligned}$$

es té que:

1. Les variables X_n són incorrelacionades, és a dir

$$\text{Cov}(X_i; X_j) = 0 \quad \text{per a } i \neq j.$$

2. $\text{Var}(X_i) = \sigma_i^2$ per a tot i .

3. $V_{\pm}(X) = \text{tr}(K) = \sum_{i=1}^{\infty} \sigma_i^2$.

4. Si X, X^0 segueixen la mateixa distribució i $X_n = h_n(X), X_n^0 = h_n(X^0)$, $n \geq 1$, aleshores $\sum_{i=1}^{\infty} \sigma_i^2(X; X^0)$ es pot expandir en la forma

$$\sum_{i=1}^{\infty} \sigma_i^2(X; X^0) = \sum_{n=1}^{\infty} \sigma_n^2(X_n; X_n^0).$$

5. $h_n(x), n \geq 1$, són solucions de l'equació diferencial

$$\sigma_n^2 h'' + (h_{i-1})' f = 0, \tag{5.6}$$

amb la condició $h(a) = h^0(a) = 0$, essent $\sigma_n^2 = \text{var}[h(X)], \sigma_n^2 = E[h(X)]$.

6. La variable aleatòria X és pot expandir en la forma (convergència en mitjana quadràtica)

$$X = \mu_0 + \sum_{i=1}^{\infty} b_i (X_i - \mu_i),$$

essent $\mu_0 = E[X], \mu_i = E[h_i(X)] = E[X_i]$ i verificant-se que la constant b_i és igual a $h_i(b)$.

7. Podem considerar que cada X_n és una component principal del procés estocàstic X definit en la secció 5.2. Així mateix, aquestes components X_n , poden interpretar-se com les dimensions principals per a la distància $\pm(x; y) = \sum_{j=1}^{\infty} \sigma_j^2(x_j - y_j)^2$.

Una condició per a l'existència d'aquestes components és que la traça de K i la variabilitat geomètrica siguin finites. De fet, per a la distància $\pm(x; y) = \sum_{j=1}^{\infty} \sigma_j^2(x_j - y_j)^2$ ambdues quantitats coincideixen, (5.2). A més recordem que, (5.3), si la variància és finita també ho és la variabilitat geomètrica.

Exemple 5.3

Si X és una variable aleatòria amb distribució exponencial, $f(x) = \lambda \exp(-\lambda x)$, $x > 0$, $\lambda > 0$, les solucions de l'equació diferencial (5.6) són de la forma, Cuadras i Fortiana [22],

$$h_i(x) = \frac{\lambda^{2i}}{\Gamma(i)^2} J_0(\lambda x) \exp\left(-\frac{\lambda x}{2}\right) J_0(\lambda x), \quad i \in \mathbb{N},$$

essent $J_0(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(n!)^2} \left(\frac{x}{2}\right)^{2n}$ la funció de Bessel de primera espècie i d'ordre 0, J_i la i -èsima arrel positiva de la funció de Bessel de primera espècie i ordre 2, és a dir, la i -èsima arrel positiva de l'equació $J_1(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!(n+1)!} \left(\frac{x}{2}\right)^{2n+1} = 0$.

Cuadras, Fortiana i Lahlou estudien detalladament les expansions ortogonals de les distribucions uniforme, exponencial, logística, Pareto i Laplace. Vegeu Cuadras i Fortiana [22] i Cuadras i Lahlou ([26], [27]).

5.5 Funció de proximitat

Sigui \mathcal{X} una població representada per una variable aleatòria X amb valors a $E \subseteq \mathbb{R}$ i funció de densitat $f(x)$, respecte la mesura de Lebesgue.

Sigui $x \in E$ un valor de la variable aleatòria X . La funció de proximitat de x a la població \mathcal{X} , respecte una distància \pm , és

$$\hat{A}_{\pm}^2(x) = \int_E \pm^2(x; y) f(y) dy \mid V_{\pm}(X).$$

Hem escrit $\hat{A}_{\pm}^2(x)$ per conveniència de notació, atès que en general és positiva. Ho serà, per exemple, si \pm és una distància euclidiana. Una condició per a que \pm sigui euclidiana és que \pm^2 sigui condicionalment de...nida negativa, és a dir,

$$\sum_{i,j=1}^n a_i a_j \pm^2(x_i; x_j) \leq 0$$

per a qualsevol nombres reals $a_1; \dots; a_n$ amb $\sum_{i=1}^n a_i = 0$ i $x_1; \dots; x_n \in E$:

Nota: La definició de funció de proximitat també s'aplica a una dissimilaritat, és a dir, a una \pm tal que $\pm(x; y) = \pm(y; x)$ i $\pm(x; x) = 0$: Aleshores escriurem

$$\hat{A}_{\pm}^{\pi}(x) = \int_E \pm^2(x; y) f(y) dy \mid V_{\pm}(X):$$

Distribució	Funció de proximitat
Qualsevol [f(x) = F'(x)]	$\int_a^b (F(y))^2 dy$ i $\int_x^b (2F(y) - 1) dy$
Normal N(μ, σ ²)	$2f(x) + \frac{1}{\sigma} \int_{-\infty}^x \frac{f(x) - f(y)}{\sigma} dy$ i $\frac{1}{\sigma} \int_x^{\infty} \frac{f(x) - f(y)}{\sigma} dy$
Uniforme en [0; 1]	x^2 i $x + \frac{1}{3}$
Exponencial [λ exp(-λx)]	$x + (2 \exp(-λx) - 1) \frac{1}{λ}$ i $\frac{1}{2λ}$
Logística $\frac{f}{2} (e^{-x} + e^{x})^{-2}$	$\frac{1}{3} x + \frac{1}{3} \ln(e^{2x} + 1)$ i $\frac{1}{2}$
Laplace $\frac{f}{2} e^{- x }$	$\frac{1}{4} + \frac{1}{2} e^{- x }$

Taula 5.2: Funció de proximitat per a diferents distribucions i per a la distància arrel quadrada de la distància euclidiana ordinària.

Per citar un exemple, considerem la distància $\pm(x; y) = \sqrt{|x - y|}$, si X es distribueix segons una distribució exponencial de paràmetre λ, la funció de proximitat és

$$A_{\pm}^2(x) = x + (2 \exp(-λx) - 1) \frac{1}{λ} \text{ i } \frac{1}{2λ}$$

essent E[X] = $\frac{1}{λ}$.

En la taula 5.2 podem veure les funcions de proximitat per diferents distribucions en el cas de la distància $\pm(x; y) = \sqrt{|x - y|}$.

5.6 Representació euclidiana discreta

Sigui (E; ±) i L un espai de Hilbert real i separable amb producte escalar $\langle \cdot; \cdot \rangle$ i norma $\|u\| = \sqrt{\langle u; u \rangle}$. Anomenarem representació euclidiana discreta d'una distància ±, a l'aplicació $\tilde{A} : E \rightarrow L$

$$x \mapsto \tilde{A}(x) = (x_1; x_2; \dots; x_n; \dots),$$

tal que

$$\pm^2(x; y) = k \tilde{A}(x) \cdot \tilde{A}(y) k^2 = \sum_{i=1}^n (x_i - y_i)^2,$$

Es ben conegut que aquesta aplicació existeix si E és ...nit i \pm és una distància euclidiana, Mardia et al. [51].

Per exemple, Cuadras i Fortiana [22], l'aplicació

$$x \mapsto \tilde{A}(x) = \frac{\rho - \mu}{\frac{1}{4}} \cos(\frac{1}{4}x) ; \frac{1}{2} \cos(\frac{2}{4}x) ; \frac{1}{3} \cos(\frac{3}{4}x) :::$$

és una representació euclidiana discreta de la variable aleatòria X , distribuïda uniformement en $[0; 1]$, per a la distància $\pm(x; y) = \sqrt{\sum_{j=1}^p (x_j - y_j)^2}$.

Nota: Farem ús d'aquesta representació discreta de \pm en la construcció de densitats bivariants. Aquí utilitzarem especialment la variabilitat geomètrica i la funció de proximitat, que tenen, quan la representació existeix, una interpretació interessant. Cuadras, Fortiana i Oliva [25] proven el següent resultat.

Resultat 9 Si existeix una representació euclidiana d'una distància \pm , $x \mapsto \tilde{A}(x)$ tal que $E \sum_{i=1}^n k \tilde{A}(X) k^2 < 1$, indicant $\tilde{A}(X) = (X_1, X_2, \dots)$ i $E[\tilde{A}(X)] = (E[X_1], E[X_2], \dots)$; aleshores s'acompleix:

1. $V_{\pm}(X) = E \sum_{i=1}^n k \tilde{A}(X) k^2 - k E[\tilde{A}(X)] k^2$;
2. $\hat{A}_{\pm}^2(x) = k \tilde{A}(x) \cdot E[\tilde{A}(X)] k^2$;

La funció de proximitat $\hat{A}_{\pm}^2(x)$ pot ser interpretada com una mena de distància al quadrat entre una observació x i un vector mitjà associat a la població que representa X .

Exemple 5.4

Un exemple senzill es dona en el cas $\pm(x; y) = |x - y|$. Com hem vist abans, secció 5.2, per aquesta distància tenim

$$V_{\pm}(X) = E \sum_{i=1}^n X_i^2 - (E[X])^2 = \text{var}(X),$$

i per tant

$$\begin{aligned}
 \hat{A}_{\pm}^2(x) &= \int_{Z^E} \pm^2(x; y) f(y) dy + V_{\pm}(X) \\
 &= \int_{Z^E} (x - y)^2 f(y) dy + \text{var}(X) \\
 &= x^2 + E[X^2] - 2xE[X] + \text{var}(X) \\
 &= x^2 - 2xE[X] + (E[X])^2 \\
 &= (x - E[X])^2.
 \end{aligned}$$

En aquest cas la representació euclidiana \tilde{A} de la distància $\pm(x; y) = |x - y|$ és la identitat. Així, en el cas uniforme obtenim $\hat{A}_{\pm}^2(x) = |x - \frac{1}{2}|^2$ i anàlogament en els altres casos.

5.7 Transformació de distàncies i funcions de proximitat

Sigui \pm una distància. Considerem la transformació

$$\pm \rightarrow \mathbb{E},$$

amb

$$\mathbb{E}^2(x; y) = \begin{cases} a\pm^2(x; y) + b & \text{si } x \neq y \\ 0 & \text{si } x = y \end{cases} \quad \text{amb } a, b \geq 0.$$

Aquesta transformació lineal de la distància al quadrat, coneguda en Escalament Multidimensional, conserva la propietat triangular. En particular, si la distància és euclidiana, la transformada també ho és.

Es verifica:

1. $V_{\mathbb{E}}(X) = aV_{\pm}(X) + \frac{b}{2}$
2. $\hat{A}_{\mathbb{E}}^2(x) = a\hat{A}_{\pm}^2(x) + \frac{b}{2}$

En efecte.

$$\begin{aligned}
 V_{\mathbb{E}}(X) &= \frac{1}{2} \int_{Z^E \times Z^E} \mathbb{E}^2(x; y) f(x) f(y) dx dy \\
 &= \frac{1}{2} \int_{Z^E \times Z^E} [a\pm^2(x; y) + b] f(x) f(y) dx dy
 \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{2} \int_{E \times E} a_{\pm^2}(x; y) f(x) f(y) dx dy + \int_{E \times E} b f(x) f(y) dx dy \\
 &= a V_{\pm}(X) + \frac{b}{2}. \\
 \hat{A}_{\pm}^2(x) &= \int_{E^2} a_{\pm^2}(x; y) f(y) dy \int_{E} V_{\pm}(X) \\
 &= \int_{E^2} [a_{\pm^2}(x; y) + b] f(y) dy \int_{E} a V_{\pm}(X) + \frac{b}{2} \\
 &= \int_{E^2} a_{\pm^2}(x; y) f(y) dy + \int_{E^2} b f(y) dy \int_{E} a V_{\pm}(X) + \frac{b}{2} \\
 &= a \int_{E^2} a_{\pm^2}(x; y) f(y) dy \int_{E} V_{\pm}(X) + b \int_{E^2} f(y) dy \int_{E} V_{\pm}(X) + \frac{b}{2} \\
 &= a \hat{A}_{\pm}^2(x) + \frac{b}{2}.
 \end{aligned}$$

5.8 Densitats a partir de proximitats

Cuadras, Fortiana i Oliva [25] presenten un mètode molt general d'anàlisi discriminant. Aquest mètode suposa, implícitament, la construcció d'una densitat de probabilitat a partir d'una distància (o dissimilaritat). Motivats per aquesta construcció, estudiem algunes densitats que sortiren a partir de distàncies.

Una definició natural de densitat de proximitat generada per la distància \pm és

$$f_{\pm}(x) = \exp \left[- \frac{1}{a} \hat{A}_{\pm}^2(x) \right],$$

la qual es justifica intuïtivament si tenim present que l'exponent, la funció de proximitat $\hat{A}_{\pm}^2(x)$, és, secció 5.6, com una mena de distància al quadrat entre x i un vector mitjà associat a la població, el mateix que tenim en l'exponent de la distribució normal, per exemple.

En general, però, ens podem trobar amb que $f_{\pm}(x)$ no sigui una funció de densitat, és a dir, que

$$\int_{E} \exp \left[- \frac{1}{a} \hat{A}_{\pm}^2(x) \right] dx \notin 1.$$

No obstant sempre podem trobar valors adients a i b , per als quals la

distància transformada

$$\mathbb{E}^2(x; y) = a \pm^2(x; y) + b,$$

proporcioni una funció de densitat de proximitat

$$f_{\mathbb{E}}(x) = \exp \int_1^x A_{\mathbb{E}}^2(x) dx,$$

que compleixi

$$\int_E \exp \int_1^x A_{\mathbb{E}}^2(x) dx = 1, \tag{5.7}$$

i que per tant sigui una funció de densitat de probabilitat sobre E.

En conseqüència, sempre que sigui necessari podrem considerar que

$$f_{\pm}(x) = \exp \int_1^x A_{\pm}^2(x) dx$$

és una funció de densitat. En cas contrari transformem \pm en \mathbb{E} .

Nota: Evidentment podríem de...nir

$$f_{\mathbb{E}}(x) = \frac{1}{\int_E f_{\pm}(x) dx} f_{\pm}(x);$$

però això és equivalent a considerar

$$\mathbb{E}^2(x; y) = \pm^2(x; y) + 2 \ln \int_E f_{\pm}(x) dx;$$

i aquesta solució no és més que un cas particular de l'anterior.

Comencem il·lustrant una densitat construïda partint d'una \pm que coincideix amb f:

Distribució normal

Sigui la distribució $N(1; \frac{3}{4}^2)$ amb $\frac{3}{4}^2 > 0$ tal que $\ln(\frac{3}{4} \sqrt{2\pi}) < 0$ i la distància

$$\pm(x; y) = \frac{|x - y|}{\frac{3}{4}}.$$

De

$$\frac{1}{2} E(X - Y)^2 = \frac{3}{4}^2; \quad E(X - x)^2 = (x - 1)^2 + \frac{3}{4}^2;$$

fàcilment deduïm que

$$V_{\pm}(X) = 1; \quad A_{\pm}^2(x) = (x - 1)^2 = \frac{3}{4}^2;$$

Aleshores $\exp(-\frac{1}{2} \hat{A}_{\pm}^2(x))$ no és una densitat, però si prenem $a = 1/2$; $b = \frac{1}{2 \ln(\frac{3}{4} \sqrt{\frac{1}{2}})}$ i definim la distància (al quadrat)

$$d(x; y)^2 = \frac{1}{2} \frac{(x - y)^2}{\frac{3}{4}^2} - \ln(\frac{3}{4} \sqrt{\frac{1}{2}}) \quad \text{si } x \leq y;$$

llavors

$$f(x; 1; \frac{3}{4}^2) = f_{\pm}(x):$$

El cas $\ln(\frac{3}{4} \sqrt{\frac{1}{2}}) > 0$; que donaria una "distància" negativa, es discuteix a la secció 5.9. Malgrat això, veurem que la construcció és també possible.

Distribució uniforme

Sigui la variable aleatòria X distribuïda uniformement en $[0; 1]$. Considerarem dos casos segons la distància \pm escollida.

2 Considerem la distància $\pm(x; y) = \sqrt{|x - y|}$.

Tenim, taula 5.2, que la funció de proximitat és

$$\hat{A}_{\pm}^2(x) = x^2 - x + \frac{1}{3}.$$

Comprovem que la funció

$$\begin{aligned} f_{\pm}(x) &= \exp\left(-\frac{1}{2} \hat{A}_{\pm}^2(x)\right) \\ &= \exp\left(-\frac{1}{2} \left(x^2 - x + \frac{1}{3}\right)\right) \\ &= \exp\left(-\frac{1}{2} x^2 + \frac{1}{2} x - \frac{1}{6}\right), \end{aligned}$$

no és una funció de densitat. En efecte,

$$\begin{aligned} &\int_0^1 \exp\left(-\frac{1}{2} x^2 + \frac{1}{2} x - \frac{1}{6}\right) dx \\ &= \frac{1}{2} \sqrt{\frac{1}{2}} e^{-\frac{1}{6}} \operatorname{erf}\left(\frac{x - \frac{1}{2}}{\frac{1}{2}}\right) \Big|_0^1 \\ &= \sqrt{\frac{1}{4}} e^{-\frac{1}{6}} \operatorname{erf}\left(\frac{1}{2}\right) \\ &= 0.8488, \end{aligned}$$

essent $\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$ la funció error¹.

Transformem la distància en la forma

$$d^2(x; y) = a|x - y| + b,$$

la qual cosa transforma la funció de proximitat, com hem vist en la secció 5.7, en

$$\begin{aligned} A_{\frac{a}{2}}^2(x) &= aA_{\frac{a}{2}}^2(x) + \frac{b}{2} \\ &= a \left(x^2 + x + \frac{1}{3} \right) + \frac{b}{2}, \end{aligned}$$

i imposem (5.7). Tenim així

$$\begin{aligned} 1 &= \int_0^z \frac{1}{\sqrt{a}} \exp\left(-\frac{1}{2a} \left(x^2 + x + \frac{1}{3}\right) + \frac{b}{2}\right) dx \\ &= \frac{1}{\sqrt{a}} e^{\frac{1}{2a} \left(\frac{1}{3} + \frac{b}{2}\right)} \int_0^z \frac{1}{\sqrt{a}} \exp\left(-\frac{1}{2a} \left(x^2 + x + \frac{1}{3}\right)\right) dx. \end{aligned}$$

La solució és de la forma

$$b = \frac{1}{2} \ln \frac{1 - \frac{1}{\sqrt{a}} \int_0^z \frac{1}{\sqrt{a}} \exp\left(-\frac{1}{2a} \left(x^2 + x + \frac{1}{3}\right)\right) dx}{\frac{1}{\sqrt{a}} \int_0^z \frac{1}{\sqrt{a}} \exp\left(-\frac{1}{2a} \left(x^2 + x + \frac{1}{3}\right)\right) dx} + \frac{1}{6} a.$$

Podríem considerar el cas, molt més interessant, en el qual la transformació de la distància sigui, només, un canvi d'escala i no una translació. És a dir, fem $b = 0$.

Tenim, doncs,

$$\begin{aligned} 1 &= \int_0^z \frac{1}{\sqrt{a}} \exp\left(-\frac{1}{2a} \left(x^2 + x + \frac{1}{3}\right)\right) dx \\ &= \frac{1}{\sqrt{a}} \int_0^z \frac{1}{\sqrt{a}} \exp\left(-\frac{1}{2a} \left(x^2 + x + \frac{1}{3}\right)\right) dx, \end{aligned}$$

que no té solució llevat de la trivial² $a = 0$.

¹ $\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt = 2G\left(\frac{z^2}{2}\right) - 1,$

on $G(z)$ és la funció de distribució de la normal estàndar.

² $\lim_{a \rightarrow 0} \frac{1}{\sqrt{a}} \int_0^z \frac{1}{\sqrt{a}} \exp\left(-\frac{1}{2a} \left(x^2 + x + \frac{1}{3}\right)\right) dx = 1:$

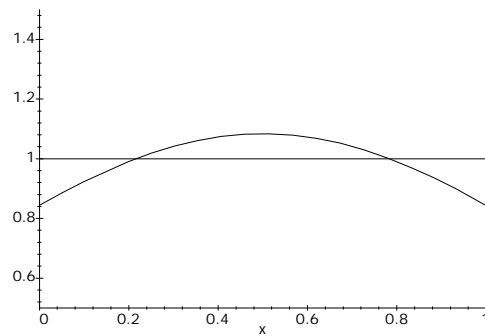


Figura 5.1: Densitat de la distribució uniforme i de la generada per la distància $\sqrt{|x_i - y_j|}$; amb $a = 1$ i $b = 0.32787$ equivalent a la generada per la distància $|x_i - y_j|$ amb $a = 1$ i $b = 0.1612$:

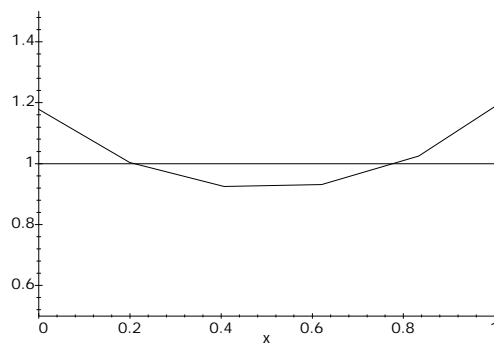


Figura 5.2: Densitat de la distribució uniforme i de la generada per la distància $\sqrt{|x_i - y_j|}$; amb $a = 1$ i $b = 0.33897$, equivalent a la generada per la distància $|x_i - y_j|$; amb $a = 1$ i $b = 0.17231$:

² Considerem la distància $d_{\pm}(x; y) = |x_i - y_j|$.

En la taula 5.2 tenim que la funció de proximitat és

$$A_{\pm}^2(x) = \frac{\mu}{2} \left(x_i - \frac{1}{2} \right)^2.$$

Comprovem que la funció

$$\begin{aligned} f_{\pm}(x) &= \exp\left(-\frac{\mu}{2} A_{\pm}^2(x)\right) \\ &= \exp\left(-\frac{\mu}{4} \left(x_i - \frac{1}{2}\right)^2\right), \end{aligned}$$

no és una funció de densitat. En efecte,

$$\begin{aligned} \int_0^1 \exp\left(-\frac{\mu}{4} \left(x_i - \frac{1}{2}\right)^2\right) dx &= \frac{\sqrt{\pi}}{2} \operatorname{erf}\left(\frac{\sqrt{\mu}}{2}\right) \\ &= 0.92256. \end{aligned}$$

Transformem la distància en la forma

$$d_{\pm}^2(x; y) = a(x_i - y)^2 + b,$$

la qual cosa transforma la funció de proximitat, com hem vist en la secció 5.7, en

$$\begin{aligned} A_{\pm}^2(x) &= a A_{\pm}^2(x) + \frac{b}{2} \\ &= a \left(x_i - \frac{1}{2}\right)^2 + \frac{b}{2}, \end{aligned}$$

i imposem (5.7). Tenim així

$$\begin{aligned} 1 &= \int_0^1 \exp\left(-\frac{\mu}{2} \left(a \left(x_i - \frac{1}{2}\right)^2 + \frac{b}{2}\right)\right) dx \\ &= \frac{\sqrt{\pi}}{2} \frac{e^{-\frac{\mu b}{4}}}{\sqrt{a}} \operatorname{erf}\left(\frac{\sqrt{\mu}}{2} \frac{1}{\sqrt{a}}\right), \end{aligned}$$

essent la solució en aquest cas general

$$b = \frac{2}{\mu} \ln \frac{\sqrt{\pi}}{\operatorname{erf}\left(\frac{\sqrt{\mu}}{2} \frac{1}{\sqrt{a}}\right)}.$$

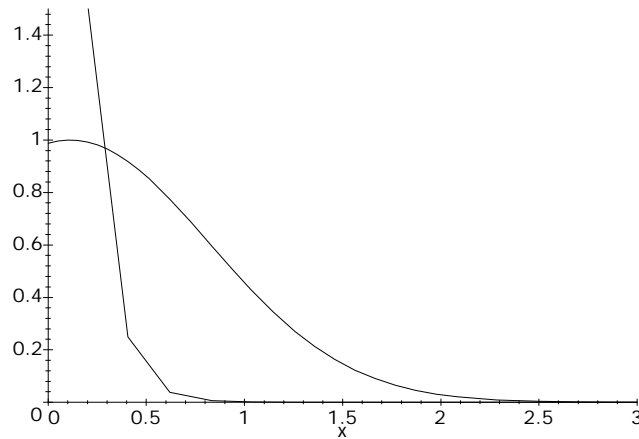


Figura 5.3: Densitat de la distribució exponencial $\exp(-\lambda x)$ i de la generada per la distància $|x_i - y_j|$, per $\lambda = 8.7513$.

Considerem el cas $b = 0$. Tenim

$$\begin{aligned}
 1 &= \int_0^{\infty} \lambda \exp(-\lambda x) \frac{1}{2} dx \\
 &= \frac{\lambda}{2} \int_0^{\infty} \exp(-\lambda x) dx
 \end{aligned}$$

i l'única solució és la trivial³.

En les figures 5.1 i 5.2 podem veure dos exemples.

Distribució exponencial

Considerem la distància $d_{\pm}(x; y) = |x_i - y_j|$ i la variable aleatòria exponencial X amb funció de densitat $f(x) = \lambda \exp(-\lambda x)$.

En la taula 5.2 tenim que la funció de proximitat és

$$A_{\pm}^2(x) = \lambda \int_0^x \exp(-\lambda t) dt$$

³

$$\lim_{\lambda \rightarrow 0} \frac{\lambda}{2} \int_0^{\infty} \exp(-\lambda x) dx = 1.$$

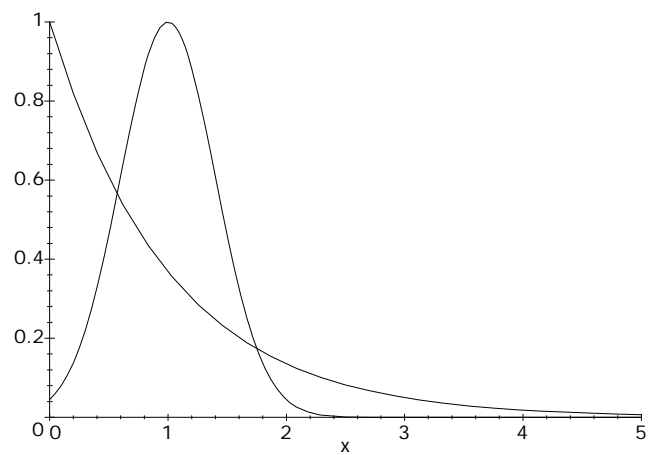


Figura 5.4: Densitat de la distribució exponencial $\exp(-x)$ i de la generada per la distància $|x_i - y_j|$; per $\alpha = 1$ i $a = 3:1017$.

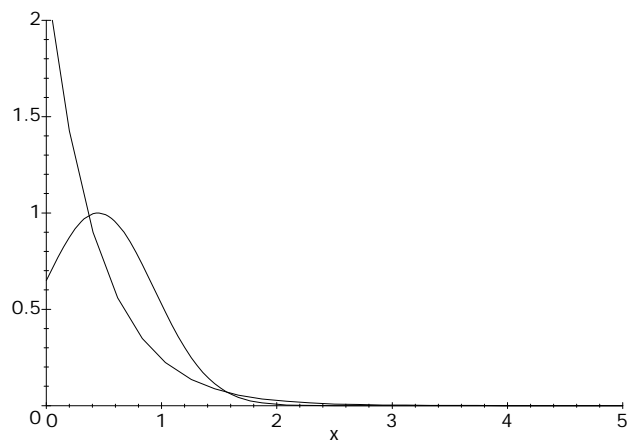


Figura 5.5: Densitat de la distribució exponencial $\exp(-x)$ i de la generada per la distància $|x_i - y_j|$; per $\alpha = 2:2185$ i $a = 2:1336$.

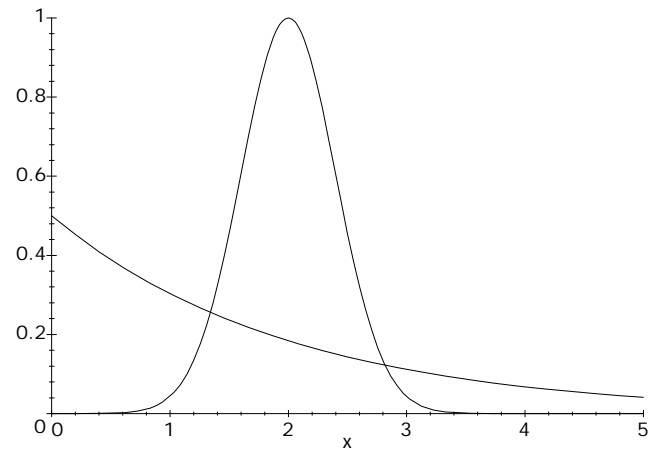


Figura 5.6: Densitat de la distribució exponencial $\exp(-\lambda x)$ i de la generada per la distància $|x_i - y_j|$; per $\lambda = 0.5$ i $a = 1/4$.

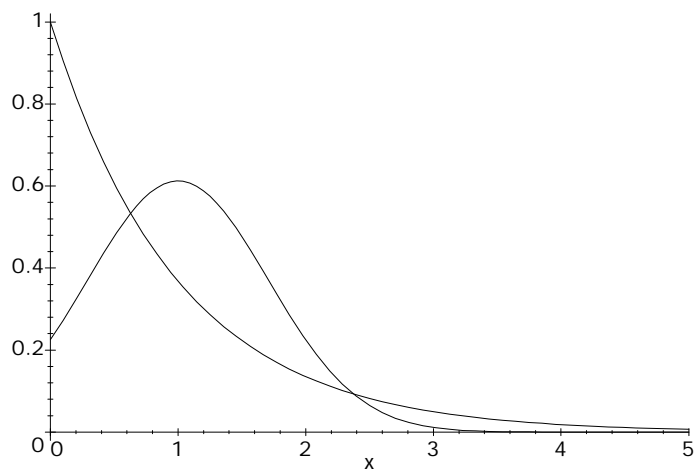


Figura 5.7: Densitat de la distribució exponencial $\exp(-\lambda x)$ i de la generada per la distància $|x_i - y_j|$; per $\lambda = 1$, $a = 1$ i $b = 0.9809$.

Comprovem si la funció

$$f_{\pm}(x) = \exp\left[-\frac{\mu}{\sigma} A_{\pm}^2(x)\right] \\ = \exp\left[-\mu \left(x \pm \frac{1}{\sigma}\right)^2\right],$$

és una funció de densitat.

$$\int_0^{\infty} \exp\left[-\mu \left(x \pm \frac{1}{\sigma}\right)^2\right] dx \\ = \frac{1}{2} \sqrt{\frac{\pi}{\mu}} + \frac{1}{2} \sqrt{\frac{\pi}{\mu}} \operatorname{erf}\left(\frac{\mu}{\sigma}\right),$$

expressió que per $\sigma = 1$ dóna la unitat.

Suposem, ara, $\sigma \neq 1$. Tenim que, fent el canvi habitual,

$$A_{\pm}^2(x) = a A_{\pm}^2\left(\frac{x}{a}\right) + \frac{b}{2} \\ = a \left(\frac{x}{a} \pm \frac{1}{\sigma}\right)^2 + \frac{b}{2},$$

i imposant

$$\int_0^{\infty} \exp\left[-\mu \left(\frac{x}{a} \pm \frac{1}{\sigma}\right)^2 - \frac{\mu b}{2}\right] dx = 1,$$

dóna la solució

$$b = \frac{2 \ln 2 \sqrt{\frac{\pi}{\mu}}}{1 + \operatorname{erf}\left(\frac{\mu}{\sigma}\right)}.$$

Considerant el cas $b = 0$. Tenim

$$\int_0^{\infty} \exp\left[-\mu \left(\frac{x}{a} \pm \frac{1}{\sigma}\right)^2\right] dx = \frac{1}{2} \sqrt{\frac{\pi}{\mu}} \left[1 + \operatorname{erf}\left(\frac{\mu}{\sigma}\right)\right].$$

Per a diferents valors del paràmetre σ i imposant que la integral anterior sigui 1, obtindríem el corresponent valor de a .

En les figures 5.3, 5.4, 5.5, 5.6 i 5.7 podem veure alguns exemples.

5.9 Desigualtat entre la variabilitat geomètrica i l'entropia de Shannon.

El següent resultat, Cuadras, Atkinson i Fortiana [19] i Cuadras i Fortiana [23], ens assegura que l'entropia de Shannon, $H(f) = \int_{\mathbb{R}} f(x) \ln f(x) dx$,

Distribució de probabilitat	Entropia de Shannon
Qualsevol	$\int_{\mathbb{R}} f(x) \ln f(x) dx$
Normal $\frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{x^2}{2\sigma^2})$	$\frac{1}{2} \ln 2 + \frac{1}{2} \ln \frac{1}{\sigma} + \ln \frac{1}{\sigma} + \frac{1}{2}$
Uniforme [0; 1]	0
Exponencial [$\lambda \exp(-\lambda x)$]	$\lambda \ln \lambda + 1$
Logística $\frac{1}{2} \frac{e^{-x}}{(e^{-x} + e^x)^2}$	$2 \lambda \ln 2 - \lambda \ln \lambda$
Laplace $\frac{1}{2} e^{-\lambda x }$	$\frac{1}{2} \ln 2 - \frac{1}{2} \ln \lambda$
Pareto $\frac{\lambda}{x^{\lambda+1}}$	$\lambda \ln \lambda - \lambda \ln \lambda$

Taula 5.3: Entropia per a diferents distribucions.

la qual, mesura també la variabilitat de la variable X , és una cota inferior de la variabilitat geomètrica. En la taula 5.3 podem veure l'entropia de Shannon per a diferents distribucions.

Resultat 10 Sigui X una variable aleatòria, amb mitjana i variància finites, definida en un espai de probabilitat $(\Omega; \mathcal{A}; P)$ amb valors a $E \subseteq \mathbb{R}$ i amb funció de distribució $F(x)$ absolutament contínua i densitat $f(x)$ respecte la mesura de Lebesgue. Sigui també \pm una distància que genera la funció de densitat $f_{\pm}(x) = \exp(-\int \hat{A}_{\pm}^2(x))$, essent $\hat{A}_{\pm}^2(x)$ la funció de proximitat respecte la distància \pm . Si $I(f; f_{\pm})$ és la divergència de Kullback-Leibler, es verifica:

1. $V_{\pm}(X) \leq H(f)$.
2. $I(f; f_{\pm}) = V_{\pm}(X) - H(f) \leq 0$:

En efecte, partint de la definició de la divergència de Kullback-Leibler tenim

$$\begin{aligned}
 I(f; f_{\pm}) &= \int_E f(x) \ln \frac{f(x)}{f_{\pm}(x)} dx \\
 &= \int_E f(x) \ln \frac{f(x)}{\exp(-\int \hat{A}_{\pm}^2(x))} dx \\
 &= \int_E f(x) \ln f(x) + \int \hat{A}_{\pm}^2(x) dx \\
 &= \int_E f(x) \ln f(x) dx + \int_E f(x) \hat{A}_{\pm}^2(x) dx \\
 &= -H(f) + V_{\pm}(X).
 \end{aligned}$$

Però, $I(f; f_{\pm}) \leq 0$, verificant-se la igualtat si $f(x) = f_{\pm}(x)$, Ellis [30]. Per tant

$$I(f; f_{\pm}) = -H(f) + V_{\pm}(X) \leq 0,$$

i en particular

$$V_{\pm}(X) \leq H(f). \tag{5.8}$$

Com que la divergència de Kullback-Leibler, $I(f; f_{\pm})$ mesura la divergència entre les funcions $f(x)$ i $f_{\pm}(x)$ tenim que

$$V_{\pm}(X) = H(f), \quad f(x) = f_{\pm}(x).$$

És a dir:

Resultat 11 Les densitats vertadera, $f(x)$, i generada per la distància \pm , $f_{\pm}(x)$, seran idènticament iguals si la variabilitat de la variable X respecte la distància \pm i l'entropia de Shannon de la funció $f(x)$ són iguals.

En particular podem dir que si una densitat f té entropia negativa o nul·la, aleshores no pot ser generada usant el procediment estudiat en la secció 5.8. En la taula 5.4 podem veure la divergència de Kullback-Leibler per a diferents distribucions en el cas de les distàncies $\pm(x; y) = |x - y|$ i $\pm(x; y) = \sqrt{|x - y|}$.

Cuadras, Atkinson i Fortiana [19] demostren, també, que sempre existeix, però, una funció simètrica $s(x; y)$ que, suplint la funció distància $\pm^2(x; y)$, proporciona una funció de proximitat $A_s^2(x)$, monotonament relacionada⁴ amb la funció de densitat $f(x)$ i tal que $V_s(X) = H(f)$, la qual cosa assegura que f és idènticament igual a la funció de densitat de proximitat generada per s , f_s .

Un exemple de funció simètrica alternativa és $s(x; y) = \frac{1}{2} (\ln f(x) + \ln f(y))$. En efecte,

$$\begin{aligned} V_s(X) &= \frac{1}{2} \int \int s(x; y) f(x) f(y) dx dy \\ &= \frac{1}{2} \int \int (\ln f(x) + \ln f(y)) f(x) f(y) dx dy \\ &= \frac{1}{2} \int \ln f(x) f(x) dx + \frac{1}{2} \int \ln f(y) f(y) dy \\ &= \int \ln f(x) f(x) dx = H(f). \end{aligned}$$

La funció de proximitat queda

$$A_s^2(x) = \int s(x; y) f(y) dy + V_s(X)$$

⁴ $A_s^2(x_1) < A_s^2(x_2)$ si i només si $f(x_1) > f(x_2)$ i per tant $f_s(x_1) > f_s(x_2)$ si i només si $f(x_1) > f(x_2)$, essent $f_s(x) = \exp\left(\frac{1}{2} A_s^2(x)\right)$.

Distància	Distribució de probabilitat	Divergència de Kullback-Leibler
$\pm(x; y) = jx \mid yj$	Qualsevol	$I(f; f_{\pm}) = V_{\pm}(X) \mid H(f)$
	Normal $N(0; \frac{3}{4}^2)$	$\frac{3}{4}^2 \mid \frac{1}{2} \ln 2 \mid \frac{1}{2} \ln \frac{1}{4} \mid \ln \frac{3}{4} \mid \frac{1}{2}$
	Uniforme $[0; 1]$	$\frac{1}{12}$
	Exponencial $[\text{exp}(j \text{exp}(x))]$	$\frac{1}{\text{exp}^2} + \ln \text{exp} \mid 1$
	Logística $\frac{f}{2} (e^{-x} + e^{i \cdot x})^{-2}$	$\frac{1}{12} \frac{1}{2} \mid 2 + \ln 2 + \ln \frac{1}{2}$
	Laplace $\frac{f}{2} e^{-j \cdot x}$	$\frac{2}{3} \mid \frac{1}{2} \ln 2 + \frac{1}{2} \ln \frac{1}{2}$
Pareto $\frac{f}{x^{\text{exp}+1}}$	$\frac{\text{exp}}{(2 \text{exp} - 1)(\text{exp} - 1)} + \frac{j \text{exp} - 1 + \text{exp} \ln \text{exp}}{\text{exp}}$	
Distància	Distribució de probabilitat	Divergència de Kullback-Leibler
$\pm(x; y) = \frac{p}{jx \mid yj}$	Normal	$\frac{3}{4} \mid \frac{1}{2} \ln 2 \mid \frac{1}{2} \ln \frac{1}{4} \mid \ln \frac{3}{4} \mid \frac{1}{2}$
	Uniforme	$\frac{1}{6}$
	Exponencial	$\frac{1}{2 \text{exp}} + \ln \text{exp} \mid 1$
	Logística	$\frac{1}{2} \mid 2 + \ln 2 + \ln \frac{1}{2}$
	Laplace	$\frac{3}{4} \mid \frac{1}{2} \ln 2 + \frac{1}{2} \ln \frac{1}{2}$
	Pareto	$\frac{\text{exp}}{(2 \text{exp} - 1)(\text{exp} - 1)} + \frac{j \text{exp} - 1 + \text{exp} \ln \text{exp}}{\text{exp}}$

Taula 5.4: Divergència per a diferents distribucions i distàncies.

$$\begin{aligned}
 &= \int_0^1 \int_0^1 (j \ln f(x) + i \ln f(y)) f(y) dy + \int_0^1 \int_0^1 f(x) \ln f(x) dx \\
 &= \int_0^1 f(y) \ln f(x) dy + \int_0^1 f(y) \ln f(y) + \int_0^1 f(x) \ln f(x) dx \\
 &= \int_0^1 \ln f(x) dx.
 \end{aligned}$$

Finalment, la densitat de proximitat generada per a la funció simètrica $s(x; y)$ és

$$\begin{aligned}
 f_s(x) &= \exp \left(\int_0^1 A_s^2(x) \right) \\
 &= \exp(\ln f(x)) \\
 &= f(x).
 \end{aligned}$$

Aquesta construcció, però, no és única. En el següent exemple ho podem veure.

Distribució exponencial

La densitat exponencial $f_{\theta}(x) = \theta \exp(-\theta x)$, $\theta > 0$ i $x > 0$, té entropia

$$\begin{aligned}
 H(f_{\theta}) &= \int_0^{\infty} f_{\theta}(x) \ln f_{\theta}(x) dx \\
 &= \int_0^{\infty} \theta \exp(-\theta x) \ln(\theta \exp(-\theta x)) dx \\
 &= 1 - \ln \theta,
 \end{aligned}$$

la qual, per $\theta = e$ serà més petita o igual a zero i per $\theta < e$ serà positiva.

Així per exemple, per $\theta = 4$ tenim que $H(f_4) = 1 - \ln 4 < 0$, i per tant en cap cas pot existir una distància \pm a través de la qual arribem a $f_{\pm} \sim f_{\theta}$.

Si utilitzem, però, la funció simètrica alternativa: $S(x; y) = 1 + 16xy + 2 \ln \frac{1}{4}$, tenim

$$\begin{aligned}
 V_S(X) &= \frac{1}{2} \int_0^{\infty} \int_0^{\infty} (1 + 16xy + 2 \ln \frac{1}{4}) (4 \exp(-4x)) (4 \exp(-4y)) dx dy \\
 &= 1 - \ln 4,
 \end{aligned}$$

que es igual a l'entropia de Shannon. La divergència és per tant 0 i calculant la funció de proximitat

$$\begin{aligned}
 A_S^2(x) &= \int_0^{\infty} (1 + 16xy + 2 \ln \frac{1}{4}) (4 \exp(-4y)) dy + (1 - 2 \ln 2) \\
 &= \int_0^{\infty} \ln 4 + 4x,
 \end{aligned}$$

tenim que

$$\begin{aligned} f_s(x) &= \exp[-(2 \ln 2 + 4x)] \\ &= \exp(\ln 4 - 4x) \\ &= 4e^{-4x}. \end{aligned}$$

En general, per a la funció exponencial $f_\theta(x) = \theta \exp(-\theta x)$ la funció simètrica

$$s(x; y) = 1 + \theta^2 xy + 2 \ln \frac{1}{\theta},$$

dóna una variabilitat

$$\begin{aligned} V_s(X) &= \int_0^\infty \int_0^\infty s(x; y) f(x) f(y) dx dy \\ &= \int_0^\infty \int_0^\infty (1 + \theta^2 xy + 2 \ln \frac{1}{\theta}) (\theta \exp(-\theta x)) (\theta \exp(-\theta y)) dx dy \\ &= 1 + \ln \frac{1}{\theta}, \end{aligned}$$

que és igual a l'entropia de Shannon.

La funció de proximitat queda

$$\begin{aligned} A_s^2(x) &= \int_0^\infty s(x; y) f(y) dy + V_s(X) \\ &= \int_0^\infty (1 + \theta^2 xy + 2 \ln \frac{1}{\theta}) (\theta \exp(-\theta y)) dy + (1 + \ln \frac{1}{\theta}) \\ &= \theta x + \ln \frac{1}{\theta}. \end{aligned}$$

Finalment, la densitat de proximitat generada per la funció simètrica $s(x; y)$ és

$$\begin{aligned} f_s(x) &= \exp[-A_s^2(x)] \\ &= \exp(-\theta x - \ln \frac{1}{\theta}) \\ &= f_\theta(x). \end{aligned}$$

Distribució uniforme

Com l'entropia d'una variable aleatòria distribuïda uniformement en $[0; 1]$ és

$$\begin{aligned} H(f) &= \int_0^1 f(x) \ln f(x) dx \\ &= \int_0^1 1 \ln(1) dx \\ &= 0, \end{aligned}$$

mai una distància podrà generar una funció de densitat idènticament igual a la funció de densitat vertadera.

En la secció 5.8 hem vist com per a la distància $\pm(x; y) = |x - y|$ i la variable aleatòria X distribuïda uniformement en $[0; 1]$, la funció $f_{\pm}(x)$ és funció de densitat quan

$$b = \frac{1}{2} \ln \frac{1 - \frac{1}{2a}}{1 + \frac{1}{2a}}.$$

La millor aproximació serà la que faci mínim el valor de la divergència $I(f; f_{\pm})$, que en el nostre cas serà

$$\begin{aligned} I(f; f_{\pm}) &= H(f) + V_{\pm}(X) \\ &= H(f) + a V_{\pm}(X) + \frac{b}{2} \\ &= \frac{1}{2} \ln \frac{1 - \frac{1}{2a}}{1 + \frac{1}{2a}}. \end{aligned}$$

Quant menor sigui el valor del paràmetre a , més pròxima estarà la funció f_{\pm} de f .

Cas general

Hem vist que relaxant la condició de distància i considerant només funcions simètriques $s(x; y)$; també podem construir funcions de densitat. El cas més general seria considerar

$$s(x; y) = | \ln f(x) - \ln f(y) |;$$

però només té interès teòric, doncs té poc sentit construir una densitat a partir d'una funció simètrica de la vertadera densitat.

Partim novament d'una distància \pm i considerem la transformació

$$f_{\pm}^a(x; y) = \begin{cases} a \pm^2(x; y) + b & \text{si } x \neq y \\ 0 & \text{si } x = y. \end{cases} \quad a > 0; b \in \mathbb{R};$$

on ara b pot ser negatiu. Per tant perdem en aquest cas les propietats d'una distància i per aquest motiu hem escrit f_{\pm}^a en comptes de f_{\pm}^2 : La transformació de la variabilitat geomètrica i la funció de proximitat és la mateixa

$$V_{\mathbb{E}}(X) = aV_{\pm}(X) + \frac{b}{2}; \quad \hat{A}_{\mathbb{E}}^a(x) = a\hat{A}_{\pm}^2(x) + \frac{b}{2};$$

llevat que ara poden ser quantitats negatives. Això no és inconvenient per a construir una densitat trobant constant adients $a; b$ i aplicant

$$f_{\mathbb{E}}(x) = \exp \left\{ -\frac{1}{2} \hat{A}_{\mathbb{E}}^a(x) \right\} :$$

Exemple: Podem reconstruir la densitat normal $N(1; \frac{3}{4}^2)$ considerant

$$\mathbb{E}^a(x; y) = \frac{1}{2} \frac{(x - y)^2}{\frac{3}{4}^2} - \ln \left(\frac{3}{4} \sqrt{\frac{3}{4}^2} \right) \quad \text{si } x \in y;$$

encara que sigui $\ln \left(\frac{3}{4} \sqrt{\frac{3}{4}^2} \right) > 0$: Aleshores la densitat estaria generada per una funció simètrica que no és distància. D'altra banda, recordem que generar la densitat f a partir d'una distància mai és possible si $H(f)$ és negativa.

5.10 Cas bivariant. Definicions

Siguin X i Y dues variables aleatòries definides sobre un mateix espai de probabilitats $(\Omega; \mathcal{A}; P)$ amb valors a $S(X); S(Y) \subseteq \mathbb{R}$. Considerem el parell $(X; Y)$ amb valors a $S(X) \times S(Y) \subseteq \mathbb{R} \times \mathbb{R}$, funció de distribució $F(x; y)$ i funció de densitat $f(x; y)$.

Recordem que la funció de distribució bivariant de $(X; Y)$ és una funció contínua per la dreta en \mathbb{R}^2 ,

$$F(x; y) = P(X \leq x; Y \leq y);$$

tal que compleixi les condicions:

1. $F(-\infty; y) = F(x; -\infty) = 0$:
2. $F(+\infty; +\infty) = 1$:
3. $F(b; d) - F(a; d) - F(b; c) + F(a; c) \geq 0$, per a qualsevol rectangle $(a; b) \subseteq (c; d)$:

Indicarem les distàncies, variabilitats geomètriques i funcions de proximitat, relatives a X i Y per \pm_1 i \pm_2 , $V_{\pm_1}(X) = V_1$, $V_{\pm_2}(X) = V_2$, $\hat{A}_{\pm_1}^2(x) = \hat{A}_1^2(x)$ i $\hat{A}_{\pm_2}^2(x) = \hat{A}_2^2(x)$, respectivament.

Siguin $w; w^0 \in \mathbb{R}^2$ -: Suposem que existeix, seccions 5.4 i 5.6, una successió ortogonal de variables aleatòries X_1, X_2, \dots , tal que

$$\begin{aligned} \sum_{i=1}^{\infty} (x_i - x_i^0)^2 &= \|x - x^0\|^2, \\ V_1 &= \sum_{i=1}^{\infty} \sigma_i^2 = \sum_{i=1}^{\infty} \text{Var}(X_i), \end{aligned}$$

on $X(w) = x$, $X(w^0) = x^0$, $X_i(w) = x_i$ i $X_i(w^0) = x_i^0$, $i = 1; 2; \dots$

Anàlogament, suposem que existeix una successió ortogonal de variables aleatòries Y_1, Y_2, \dots , tal que

$$\begin{aligned} \sum_{i=1}^{\infty} (y_i - y_i^0)^2 &= \|y - y^0\|^2, \\ V_2 &= \sum_{i=1}^{\infty} \sigma_i^2 = \sum_{i=1}^{\infty} \text{Var}(Y_i), \end{aligned}$$

on $Y(w) = y$, $Y(w^0) = y^0$, $Y_i(w) = y_i$ i $Y_i(w^0) = y_i^0$, $i = 1; 2; \dots$

El que tenim són dues representacions euclidianes ortogonals de les variables aleatòries X i Y ,

$$x = (x_1; x_2; \dots)^T \quad x^0 = (x_1^0; x_2^0; \dots)^T,$$

i anàlogament

$$y = (y_1; y_2; \dots)^T \quad y^0 = (y_1^0; y_2^0; \dots)^T.$$

És a dir

$$\begin{aligned} \pm_1^2(x; x^0) &= \sum_{i=1}^{\infty} (x_i - x_i^0)^2, \\ \pm_2^2(y; y^0) &= \sum_{i=1}^{\infty} (y_i - y_i^0)^2. \end{aligned}$$

5.10.1 Dissimilaritat bivariant

Siguin $S(X); S(Y)$ els suports de les v.a.s $X; Y$: Volem definir una dissimilaritat sobre $S(X) \times S(Y)$ a partir de les distàncies \pm_1 i \pm_2 , definides en $S(X)$

i $S(Y)$ respectivament, tenint en compte la possible informació redundant deguda a la dependència.

Considerem la matriu $R = (\rho_{ij})$ amb les correlacions entre les variables X_i i Y_j .

Siguin $P = (x; y)$ i $P^0 = (x^0; y^0)$ dues observacions de $(X; Y)$.

Definició 1 Definim la dissimilaritat conjunta entre P i P^0 , $\pm_{12}(P; P^0)$, com

$$\begin{aligned} \pm_{12}^2(P; P^0) &= \pm_1^2(x; x^0) + \pm_2^2(y; y^0) - \sum_{i,j=1}^X \rho_{ij} (x_i - x_i^0)(y_j - y_j^0) \quad (5.9) \\ &= \pm_1^2(x; x^0) + \pm_2^2(y; y^0) - (x - x^0)^T R (y - y^0). \end{aligned}$$

El terme $(x - x^0)^T R (y - y^0)$ conté la dependència entre les variables X i Y , és a dir representa la interacció entre aquestes variables.

Al construir una dissimilaritat bivariant conjunta, a més de tenir en compte la redundància o dependència entre les variables, s'haurien d'acomplir certes propietats, intuïtives, que la definició (5.9) satisfà.

Propietats de la dissimilaritat bivariant

1. És simètrica i no negativa

$$\pm_{12}^2(P; P^0) = \pm_{12}^2(P^0; P) \geq 0;$$

En efecte, si $(x - x^0)^T R (y - y^0) \leq 0$ és evident. En cas contrari

$$\begin{aligned} \pm_{12}^2(P; P^0) &= kx - x^0k^2 + ky - y^0k^2 - (x - x^0)^T R (y - y^0) \\ &\geq kx - x^0k^2 + ky - y^0k^2 - 2(x - x^0)^T (y - y^0) \\ &= (x - x^0 - (y - y^0))^T (x - x^0 - (y - y^0)) \geq 0; \end{aligned}$$

2. Si la distància definida per a una de les variables és zero, aquesta distància no influeix en la dissimilaritat conjunta. És a dir,

$$\begin{aligned} \pm_1 = 0 & \Rightarrow \pm_{12} = \pm_2, \\ \pm_2 = 0 & \Rightarrow \pm_{12} = \pm_1. \end{aligned}$$

En efecte, si $\pm_1^2(x; x^0) = 0$ per qualsevol x, x^0 , aleshores

$$\pm_1^2(x; x^0) = \sum_{i=1}^X (x_i - x_i^0)^2 = 0,$$

el que signi...ca

$$\sum_{i,j=1}^{\infty} \frac{1}{2} \frac{(x_i - x_i^0)(y_j - y_j^0)}{\sqrt{(x_i - x_i^0)^2 + (y_j - y_j^0)^2}} = 0,$$

i per tant

$$\begin{aligned} \pm_{12}^2(P; P^0) &= 0 + \pm_2^2(y; y^0) + 0 \\ &= \pm_2^2(y; y^0). \end{aligned}$$

3. Si les dues distàncies són iguals, la dissimilaritat és també igual a la distància comuna. És a dir,

$$X \text{ i } Y \text{ són idènticament iguals, i } \pm_{12}^2 = \pm_1^2 = \pm_2^2.$$

En efecte, al ser X i Y idènticament iguals i recordant que X_1, X_2, \dots és una successió ortogonal de variables aleatòries, tenim que $\frac{1}{2} \frac{(x_i - x_i^0)(y_j - y_j^0)}{\sqrt{(x_i - x_i^0)^2 + (y_j - y_j^0)^2}} = 1$ per a $i = j$ i $\frac{1}{2} \frac{(x_i - x_i^0)(y_j - y_j^0)}{\sqrt{(x_i - x_i^0)^2 + (y_j - y_j^0)^2}} = 0$ per a $i \neq j$. Per tant

$$\begin{aligned} \pm_{12}^2(P; P^0) &= \pm_1^2(x; x^0) + \pm_1^2(y; y^0) + \sum_{i=1}^{\infty} \frac{(x_i - x_i^0)(y_i - y_i^0)}{\sqrt{(x_i - x_i^0)^2 + (y_i - y_i^0)^2}} \\ &= \pm_1^2(x; x^0) + \pm_1^2(y; y^0) + \pm_1^2(x; x^0) \\ &= \pm_1^2(y; y^0). \end{aligned}$$

4. Si les variables són independents, la dissimilaritat conjunta ha de sumar els efectes de les dues distàncies (al quadrat). És a dir

$$X \text{ i } Y \text{ són estocàsticament independents } \Rightarrow \pm_{12}^2 = \pm_1^2 + \pm_2^2.$$

En efecte, cal tenir present que si X i Y són estocàsticament independents aleshores $R = 0$, essent R la matriu amb les correlacions entre els X_i i Y_j .

Les dues primeres propietats a...rmen que la dissimilaritat (al quadrat) conjunta no canvia quan una de les distàncies és zero o quan les dues distàncies (al quadrat) són idèntiques. La quarta propietat és el Teorema de Pitàgores aplicat a distàncies. En general, la dissimilaritat \pm_{12} conté la redundància entre \pm_1 i \pm_2 . D'ara endavant suposarem $S(X) = S(Y) = E \frac{1}{2} R$:

5.10.2 Variabilitat geomètrica bivariant

Definició 2 La variabilitat geomètrica bivariant, $V_{\pm 12}(X; Y)$, del parell $(X; Y)$ respecte les distàncies \pm_1 i \pm_2 que escriurem V_{12} , és

$$V_{12} = V_1 + V_2 + \sum_{i,j=1}^{\infty} \frac{1}{2} \rho_{\pm_1 \pm_2}^{ij}$$

essent $\sigma_i^2 = \text{var}(X_i)$; $\sigma_j^2 = \text{var}(Y_j)$:

La definició es dedueix de

$$E[(X_i - X_i^0)(Y_j - Y_j^0)] = 2 \text{cov}[X_i; Y_j].$$

En efecte,

$$\begin{aligned} V_{12} &= \frac{1}{2} \int \int \sum_{i,j=1}^{\infty} \rho_{\pm_1 \pm_2}^{ij} (P; P^0) f(x; y) f(x^0; y^0) dP dP^0 \\ &= \frac{1}{2} \int \int \sum_{i,j=1}^{\infty} \rho_{\pm_1}^{ij} (x; x^0) f(x; y) f(x^0; y^0) dP dP^0 + \\ &\quad + \frac{1}{2} \int \int \sum_{i,j=1}^{\infty} \rho_{\pm_2}^{ij} (y; y^0) f(x; y) f(x^0; y^0) dP dP^0 + \\ &\quad + \frac{1}{2} \int \int \sum_{i,j=1}^{\infty} \rho_{\pm_1 \pm_2}^{ij} (x_i - x_i^0)(y_j - y_j^0) f(x; y) f(x^0; y^0) dP dP^0 \\ &= \frac{1}{2} \int \int \sum_{i,j=1}^{\infty} \rho_{\pm_1}^{ij} (x; x^0) f(x; y) dy f(x^0; y^0) dy^0 dx dx^0 + \\ &\quad + \frac{1}{2} \int \int \sum_{i,j=1}^{\infty} \rho_{\pm_2}^{ij} (y; y^0) f(x; y) dx f(x^0; y^0) dx^0 dy dy^0 + \\ &\quad + \frac{1}{2} \sum_{i,j=1}^{\infty} \rho_{\pm_1 \pm_2}^{ij} \int \int (x_i - x_i^0)(y_j - y_j^0) f(x; y) f(x^0; y^0) dP dP^0 \\ &= \frac{1}{2} \sum_{i,j=1}^{\infty} \rho_{\pm_1}^{ij} \int \int (x; x^0) f(x) f(x^0) dx dx^0 + \frac{1}{2} \sum_{i,j=1}^{\infty} \rho_{\pm_2}^{ij} \int \int (y; y^0) f(y) f(y^0) dy dy^0 + \\ &\quad + \frac{1}{2} \sum_{i,j=1}^{\infty} \rho_{\pm_1 \pm_2}^{ij} E[(X_i - X_i^0)(Y_j - Y_j^0)] \\ &= V_1 + V_2 + \frac{1}{2} \sum_{i,j=1}^{\infty} \rho_{\pm_1 \pm_2}^{ij} 2 \text{cov}[X_i; Y_j] \\ &= V_1 + V_2 + \sum_{i,j=1}^{\infty} \rho_{\pm_1 \pm_2}^{ij} \end{aligned}$$

Propietats de la variabilitat geomètrica bivariant

1.

$$X \text{ i } Y \text{ són idènticament iguals, i } \rho_{\pm 1} = \rho_{\pm 2} \Rightarrow V_{12} = V_1 = V_2.$$

En efecte, si X i Y són idènticament iguals i $\rho_{\pm 1} = \rho_{\pm 2}$ aleshores $V_1 = V_2$ i $\rho_{ij} = 1$ per $i = j$ i $\rho_{ij} = 0$ per $i \neq j$, per tant

$$\begin{aligned} V_{12} &= V_1 + V_2 \sum_{i,j=1}^p \rho_{ij}^2 \\ &= 2V_1 \sum_{i,j=1}^p \rho_{ij}^2 \end{aligned}$$

Tenint en compte que $V_1 = \sum_{i=1}^p \sigma_i^2$, podem escriure

$$V_{12} = 2V_1 \sum_{i=1}^p \rho_{ii}^2 = V_1$$

2. En general

$$V_{12} \leq V_1 + V_2.$$

Cal tenir present que $\rho_{ij}^2 \leq \rho_{ii} \rho_{jj} = 1$ per a qualsevol valor de i i de j .

3.

$$X \text{ i } Y \text{ són estocàsticament independents } \Rightarrow V_{12} = 0.$$

En efecte, cal tenir present que si X i Y són estocàsticament independents aleshores $R = 0$.

5.10.3 Funció de proximitat bivariant

Definició 3 La funció de proximitat $A_{\pm 12}^2(x; y) = A_{12}^2(x; y)$, per a la dissimilaritat \pm_{12} , és

$$\begin{aligned} A_{12}^2(x; y) &= A_1^2(x) + A_2^2(y) - \sum_{i,j=1}^p \rho_{ij} (x_i - E[X_i]) (y_j - E[Y_j]) \\ &= A_1^2(x) + A_2^2(y) - (x - \mathbf{1}_x)^T R (y - \mathbf{1}_y), \end{aligned}$$

on

$$\begin{aligned} (x_i - 1_x)^T &= (x_1 - E[X_1]; x_2 - E[X_2]; \dots)^T, \\ (y_i - 1_y)^T &= (y_1 - E[Y_1]; y_2 - E[Y_2]; \dots)^T, \end{aligned}$$

i R és la matriu amb les correlacions entre les variables X_i i Y_j .

Amb les notacions habituals, la definició es dedueix de

$$\begin{aligned} \hat{A}_{12}^2(x; y) &= \int_{E \in E} \int_{E \in E} (P; P^0) f(x^0; y^0) dP^0; V_{12} \\ &= \hat{A}_1^2(x) + \hat{A}_2^2(y) + \sum_{i,j=1}^n \frac{1}{2} \int_{E \in E} (x_i - x_i^0) (y_j - y_j^0) f(x^0; y^0) dP^0 \\ &\quad + \sum_{i,j=1}^n \frac{1}{2} \rho_{ij} \int_{E \in E} (x_i - x_i^0) (y_j - y_j^0) f(x^0; y^0) dP^0 \\ &= \hat{A}_1^2(x) + \hat{A}_2^2(y) + \sum_{i,j=1}^n \frac{1}{2} \text{cov}[X_i; Y_j] \int_{E \in E} (x_i - x_i^0) (y_j - y_j^0) f(x^0; y^0) dP^0 \\ &= \hat{A}_1^2(x) + \hat{A}_2^2(y) + \sum_{i,j=1}^n \frac{1}{2} \int_{E \in E} (X_i - E[X_i]) (Y_j - E[Y_j]) f(x; y) dP \\ &\quad + \sum_{i,j=1}^n \frac{1}{2} \int_{E \in E} (x_i - x_i^0) (y_j - y_j^0) f(x^0; y^0) dP^0 \\ &= \hat{A}_1^2(x) + \hat{A}_2^2(y) + \sum_{i,j=1}^n \frac{1}{2} (x_i - E[X_i]) (y_j - E[Y_j]). \end{aligned}$$

Propietats de la funció de proximitat bivariant

1. La funció \hat{A}_{12}^2 és no negativa i la podem expressar com

$$\hat{A}_{12}^2(x; y) = \sum_{i,j=1}^n (x_i - 1_x)^T R (y_i - 1_y) \geq 0;$$

En efecte, l'expressió surt d'aplicar el resultat 9. Aleshores, si $(x_i - 1_x)^T R (y_i - 1_y) \geq 0$ la no negativitat és evident. En cas contrari

$$\begin{aligned} \hat{A}_{12}^2(x; y) &= \sum_{i,j=1}^n (x_i - 1_x)^T R (y_i - 1_y) \\ &\geq \sum_{i,j=1}^n (x_i - 1_x)^T R (y_i - 1_y) - 2 \sum_{i,j=1}^n (x_i - 1_x)^T R (y_i - 1_y) \\ &= \sum_{i,j=1}^n (x_i - 1_x)^T R (y_i - 1_y) \geq 0; \end{aligned}$$

2. X i Y són idènticament iguals i $\pm_1 = \pm_2$)

$$\hat{A}_{12}^2(x; y) = \hat{A}_1^2(x) = \hat{A}_2^2(y).$$

En efecte, si X i Y són idènticament iguals i $\pm_1 = \pm_2$ aleshores $\hat{A}_1^2(x) = \hat{A}_2^2(y)$ i $\frac{1}{2}i_{ij} = 1$ per $i = j$ i $\frac{1}{2}i_{ij} = 0$ per $i \neq j$, per tant

$$\begin{aligned} \hat{A}_{12}^2(x; y) &= \hat{A}_1^2(x) + \hat{A}_2^2(y) - \sum_{i,j=1}^n \frac{1}{2}i_{ij} (x_i - E[X_i]) (y_j - E[Y_j]) \\ &= 2\hat{A}_1^2(x) - \sum_{i=1}^n (x_i - E[X_i]) (x_i - E[X_i]) \\ &= 2\hat{A}_1^2(x) - \sum_{i=1}^n (x_i - E[X_i])^2, \end{aligned}$$

i tenint en compte el resultat 9

$$\hat{A}_{12}^2(x; y) = 2\hat{A}_1^2(x) - \hat{A}_1^2(x) = \hat{A}_1^2(x).$$

3. X i Y són estocàsticament independents)

$$\hat{A}_{12}^2(x; y) = \hat{A}_1^2(x) + \hat{A}_2^2(y) : \quad (5.10)$$

En efecte, cal tenir present, un altre cop, que si X i Y són estocàsticament independents aleshores la matriu de les correlacions definida en 5.10.1 és $R = 0$.

5.11 Densitat conjunta a partir de la funció de proximitat bivariant

Sigui $f(x; y)$ la densitat conjunta del parell $(X; Y)$. Sigui $f_1(x) = \exp(-\hat{A}_1^2(x))$ i $f_2(y) = \exp(-\hat{A}_2^2(y))$ les densitats generades per les distàncies \pm_1 i \pm_2 , respectivament. Recordem que sempre podem suposar que són densitats, secció 5.8. Suposarem també que, malgrat puguin ser resultat de transformar una distància, \pm_1 i \pm_2 admeten una representació euclidiana en els termes de la secció 5.10.

Considerem la distància \pm tal que

$$\pm^2 = \pm_1^2 + \pm_2^2,$$

que correspon al cas de suposar X i Y estocàsticament independents. Aleshores la funció de proximitat és positiva, i és, (5.10),

$$\hat{A}_{12}^2(x; y) = \hat{A}_1^2(x) + \hat{A}_2^2(y),$$

i, per tant

$$\begin{aligned} f_{\pm}(x; y) &= \exp\left\{-\frac{1}{2}[\hat{A}_{12}^2(x; y)]\right\} = \exp\left\{-\frac{1}{2}[\hat{A}_1^2(x) + \hat{A}_2^2(y)]\right\} \\ &= \exp\left\{-\frac{1}{2}\hat{A}_1^2(x)\right\} \exp\left\{-\frac{1}{2}\hat{A}_2^2(y)\right\} \\ &= f_1(x) f_2(y), \end{aligned}$$

és la densitat generada per la distància \pm .

En el cas general, considerem la dissimilaritat bivariant, definida en (5.9). Indicant $\bar{x} = x - x^0$; $\bar{y} = y - y^0$; la densitat de proximitat generada per \pm_{12} és

$$f_{\pm_{12}}(x; y) = \exp\left\{-\frac{1}{2}[\hat{A}_1^2(x) + \hat{A}_2^2(y) + \bar{x}^T R \bar{y}]\right\}.$$

En general, però, ens podem trobar que

$$\begin{aligned} &\int_{E \times E} \exp\left\{-\frac{1}{2}[\hat{A}_1^2(x) + \hat{A}_2^2(y) + \bar{x}^T R \bar{y}]\right\} dx dy \\ &= \int_{E \times E} \exp\left\{-\frac{1}{2}\hat{A}_1^2(x)\right\} \exp\left\{-\frac{1}{2}\hat{A}_2^2(y)\right\} \exp\left\{-\frac{1}{2}\bar{x}^T R \bar{y}\right\} dx dy \\ &= \int_{E \times E} f_{\pm_1}(x) f_{\pm_2}(y) \exp\left\{-\frac{1}{2}\bar{x}^T R \bar{y}\right\} dx dy \\ &= \int_{E \times E} \exp\left\{-\frac{1}{2}\bar{x}^T R \bar{y}\right\} f_{\pm}(x; y) dx dy \\ &\in 1, \end{aligned}$$

i per tant $f_{\pm_{12}}(x; y)$ podria no ser una densitat.

En aquest cas podem considerar la dissimilaritat

$$\pm_{12}^{\alpha}(P; P^0)_{\alpha} = \pm_1^2(x; x^0) + \pm_2^2(y; y^0) + \alpha (x - x^0)^T R (y - y^0), \quad (5.11)$$

on $\alpha > 0$ és una constant, que suposarem que existeix, tal que

$$\int_{E \times E} \exp\left\{-\frac{1}{2\alpha} \bar{x}^T R \bar{y}\right\} f_{\pm}(x; y) dx dy = 1.$$

Amb aquest canvi la variabilitat geomètrica serà

$$V_{12(\alpha)} = V_1 + V_2 + \alpha \sum_{i,j=1}^2 \frac{1}{2} p_{ij} - p_{\bar{x}} p_{\bar{y}},$$

i la funció de proximitat (que podria ser negativa)

$$\hat{A}_{12}^{\pm}(x; y)_{\otimes} = \hat{A}_1^2(x) + \hat{A}_2^2(y) - \otimes \bar{x}^T R \bar{y}.$$

Aleshores $\pm_{12}^{\pm}(P; P^0)_{\otimes}$ proporciona una funció de proximitat que genera una densitat

$$f_{\pm_{12}}(x; y)_{\otimes} = \exp \left\{ - \left[\hat{A}_1^2(x) + \hat{A}_2^2(y) - \otimes \bar{x}^T R \bar{y} \right] \right\},$$

tal que

$$V_{\pm} = V_1 + V_2 - V_{12(\otimes)} - H(f),$$

amb igualtat si $f = f_{\pm_{12}(\otimes)}$.

En efecte, partint de la definició de la divergència de Kullback-Leibler i tenint en compte que aquesta sempre és més gran o igual que zero, tenim

$$\begin{aligned} I(f; f_{\pm_{12}(\otimes)}) &= \int_{\mathcal{E}} \int_{\mathcal{E}} f(x; y) \ln \frac{f(x; y)}{f_{\pm_{12}(\otimes)}(x; y)} dx dy \\ &= \int_{\mathcal{E}} \int_{\mathcal{E}} f(x; y) \ln \frac{f(x; y)}{\exp \left\{ - \left[\hat{A}_1^2(x) + \hat{A}_2^2(y) - \otimes \bar{x}^T R \bar{y} \right] \right\}} dx dy \\ &= - H(f) - \int_{\mathcal{E}} \int_{\mathcal{E}} f(x; y) \ln \exp \left\{ - \left[\hat{A}_1^2(x) + \hat{A}_2^2(y) - \otimes \bar{x}^T R \bar{y} \right] \right\} dx dy \\ &= - H(f) - \int_{\mathcal{E}} \int_{\mathcal{E}} f(x; y) \left[\hat{A}_1^2(x) + \hat{A}_2^2(y) - \otimes \bar{x}^T R \bar{y} \right] dx dy \\ &= - H(f) + V_{12(\otimes)} \geq 0. \end{aligned}$$

Acabem de provar el següent resultat.

Resultat 12 La densitat construïda a partir de \pm_{12}^{\pm} , la dissimilaritat conjunta (5.11) considerant la redundància de les variables, és més propera a la densitat verdadera f que la densitat construïda a partir de la distància \pm , definida suposant independència estocàstica entre les variables X i Y .

Capítol 6

Distància de Wasserstein i altres desigualtats

6.1 Introducció

La distància de Wasserstein és una mesura de discrepància entre dues funcions de distribució, que té relació amb les correlacions màxima i mínima de Hoeffding. Aquesta distància té aplicacions en tests d'ajust d'unes dades a una distribució i per comparar la forma d'una distribució en relació a una altra. En aquest capítol estudiem algunes desigualtats relacionades amb aquesta distància, i uns altres tipus de desigualtats per a la variància d'una funció d'una variable aleatòria, que tenen aspectes semblants a la desigualtat de Cramér-Rao.

6.2 Definició de la distància de Wasserstein

Definició 4 Siguin $F(x)$ i $G(x)$ dues funcions de distribució absolutament contínues. Definim la distància de Wasserstein entre aquestes dues distribucions, $W(F; G)$, per l'expressió

$$W(F; G) = \int_0^1 |F^{-1}(t) - G^{-1}(t)| dt^{1/2}.$$

Aquesta definició té sentit tenint en compte que si X és una variable aleatòria definida en un espai de probabilitat $(\Omega; \mathcal{A}; P)$ amb valors a $E \subseteq \mathbb{R}$ i

amb funció de distribució $F(x)$ absolutament contínua, aleshores la variable aleatòria $F(X) = T$ està uniformement distribuïda en $(0; 1)$; i recíprocament, si T és una funció de distribució uniforme en $(0; 1)$, aleshores $X = F^{-1}(T)$ té la funció $F(x)$ per funció de distribució. En efecte,

$$\begin{aligned} T(t) &= P(T \leq t) \\ &= P(F(X) \leq t) \\ &= P(X \leq F^{-1}(t)) \\ &= F(F^{-1}(t)) = t. \end{aligned}$$

D'altra banda

$$\begin{aligned} P(X \leq x) &= P(F^{-1}(T) \leq x) \\ &= P(T \leq F(x)) = F(x). \end{aligned}$$

6.2.1 Exemples

Siguin $F(x)$ i $G(x)$ dues distribucions uniformes en $(a; b)$ i $(c; d)$ respectivament. En aquest cas el quadrat de la distància de Wasserstein és

$$\begin{aligned} W^2(F; G) &= \int_0^1 |F^{-1}(t) - G^{-1}(t)|^2 dt \\ &= \int_0^1 (a + (b-a)t - (c + (d-c)t))^2 dt \\ &= \frac{1}{3} (a-c)^2 + (b-d)(a+c) + (b-d)^2, \end{aligned} \tag{6.1}$$

essent 0 en el cas $a = c$ i $b = d$.

Siguin, ara, $F(x)$ i $G(x)$ dues distribucions, la primera uniformement distribuïda en $(0; 1)$ i la segona, exponencial de paràmetre 1. Aquí tenim

$$\begin{aligned} W^2(F; G) &= \int_0^1 (t + \ln(1-t))^2 dt \\ &= \frac{5}{6}. \end{aligned} \tag{6.2}$$

En el cas general de dues distribucions $F(x)$ i $G(x)$ la primera uniforme en $(a; b)$ i la segona, exponencial de paràmetre α , obtenim

$$\begin{aligned} W^2(F; G) &= \int_0^1 \left(a + (b-a)t - \frac{1}{\alpha} \ln(1-t) \right)^2 dt \\ &= \frac{2(a^2 + b^2)\alpha^2 + a\alpha(2b\alpha - 3) - 9b\alpha + 12}{6\alpha^2}. \end{aligned} \tag{6.3}$$

Si el que volem calcular és la distància de Wasserstein per dues distribucions, $F(x)$ i $G(x)$, exponencials de paràmetres θ i τ respectivament, obtindrem

$$\begin{aligned} W^2(F; G) &= \int_0^1 \left| \frac{1}{\theta} \ln(1-t) - \frac{1}{\tau} \ln(1-t) \right|^2 dt \quad (6.4) \\ &= \frac{2(\theta - \tau)^2}{\theta^2 + \tau^2}, \end{aligned}$$

essent 0 en el cas trivial $\theta = \tau$.

6.3 Classes de Fréchet i correlacions de Hoërding

Recordem breument els conceptes de classe de Fréchet i de correlacions de Hoërding i les seves propietats més importants.

Siguin $X; Y$ dues variables aleatòries definides en un espai de probabilitat $(\Omega; \mathcal{A}; P)$. Considerem la funció de distribució bivariant $H(x; y)$ definida en la secció 5.10. Sigui també les marginals univariants $F; G$ de H ,

$$F(x) = P(X \leq x), \quad G(y) = P(Y \leq y);$$

Definició 5 Anomenem classe de Fréchet $\mathcal{F}(F; G)$ a la família de distribucions bivariants $H(x; y)$, amb marginals univariants $F(x)$ i $G(y)$.

La funció $H^0(x; y) = F(x)G(y)$ pertany a $\mathcal{F}(F; G)$ i correspon al cas per al qual X i Y són estocàsticament independents.

Altres funcions de $\mathcal{F}(F; G)$ són $H^-(x; y) = \max\{F(x) + G(y) - 1; 0\}$ i $H^+(x; y) = \min\{F(x); G(y)\}$ anomenades respectivament cota inferior i cota superior de Fréchet.

Resultat 13 Es verifica:

- 1. $H^-(x; y) \leq H(x; y) \leq H^+(x; y)$ per a qualsevol $H(x; y) \in \mathcal{F}(F; G)$;
- 2. $H = H^-$ si i només si s'acompleix la relació funcional $F(X) + G(Y) = 1$.
- 3. $H = H^+$ si i només si s'acompleix la relació funcional $F(X) = G(Y)$.

CAPÍTOL 6. DISTÀNCIA DE WASSERSTEIN I ALTRES DESIGUALTATS 134

Definició 6 Suposem que $E(X)$; $E(Y)$ i $E(XY)$ existeixen. Sigui $\rho = \rho(X; Y)$ el coeficient de correlació entre les dues variables. Anomenem correlacions de Höfding a les mínima i màxima correlacions ρ^- ; ρ^+ ; que podem tenir per a qualsevol $H \in \mathcal{H}(F; G)$.

Resultat 14 Es verifica:

$$\rho^- \leq \rho \leq \rho^+;$$

$$\rho^- = \rho^+ \text{ si i només si } H = H^-;$$

$$\rho = \rho^+ \text{ si i només si } H = H^+;$$

Per calcular els coeficients de Höfding, en el cas de variables aleatòries X i Y , amb mitjana 0 i variància 1, és interessant expressar aquests coeficients en la forma

$$\rho^- = \int_0^1 F^{-1}(p)G^{-1}(1-p)dp, \quad \rho^+ = \int_0^1 F^{-1}(p)G^{-1}(p)dp;$$

El següent resultat, Mardia [50], ens dona la correlació entre dues funcions utilitzant les funcions de distribució conjunta i marginals.

Resultat 15 Sigui X i Y dues variables aleatòries amb funció de distribució conjunta $H(x; y)$ i marginals $F; G$, absolutament contínues. Podem expressar la correlació entre les dues variables aleatòries en la forma

$$\rho(X; Y) = \frac{1}{\sqrt{f_X(x)}\sqrt{f_Y(y)}} \int_{\mathbb{R}^2} (H - FG) dx dy.$$

Aquest resultat, degut a Höfding, es pot considerar una conseqüència del següent, més general, Cuadras[18], el qual ens permet expressar la covariància en termes de les funcions de distribució i no de les funcions de densitat.

Resultat 16 Sigui X i Y dues variables aleatòries amb funció de distribució conjunta $H(x; y)$ i marginals $F; G$, absolutament contínues. Suposem $\phi(x)$ i $\psi(y)$ dues funcions tals que $E(\phi(X)) < +\infty$; $E(\psi(Y)) < +\infty$ i $E(\phi(X)\psi(Y)) < +\infty$. Aleshores podem escriure la covariància entre $\phi(X)$ i $\psi(Y)$ en termes de les funcions de distribució en la forma

$$\text{cov}(\phi(X), \psi(Y)) = \int_{\mathbb{R}^2} (\phi(x)\psi(y) - F(x)G(y)) dH(x; y). \quad (6.5)$$

6.3.1 Exemple

Calculem la correlació màxima de Höfding, $\frac{1}{2}^+$, entre dues distribucions $F(x)$ i $G(x)$ la primera uniforme en $(a; b)$ i la segona, exponencial de paràmetre θ .

Acabem de veure que $\frac{1}{2} = \frac{1}{2}^+$ si i només si $H = H^+(x; y) = \min\{F(x); G(y)\}$, per tant

$$\begin{aligned} \frac{1}{2}^+(X; Y) &= \frac{1}{\int_a^b \int_0^\infty \min\left\{\frac{x-a}{b-a}; e^{-\theta y}\right\} dx dy} \\ &= \frac{1}{\int_a^b \int_0^\infty \min\left\{\frac{x-a}{b-a}; e^{-\theta y}\right\} dx dy} \\ &= \frac{1}{\int_a^b \int_0^\infty \min\left\{\frac{x-a}{b-a}; e^{-\theta y}\right\} dx dy} \\ &= \frac{1}{\int_a^b \int_0^\infty \min\left\{\frac{x-a}{b-a}; e^{-\theta y}\right\} dx dy} \\ &= \frac{1}{\int_a^b \int_0^\infty \min\left\{\frac{x-a}{b-a}; e^{-\theta y}\right\} dx dy} \end{aligned}$$

on $\theta = \frac{1}{b-a}$.

6.4 Desigualtats per a la distància de Wasserstein

Barrio et al. [29] donen una expressió alternativa per a la distància de Wasserstein entre les distribucions F i G la qual, amb els conceptes definits en la secció 6.3, en particular amb la correlació màxima de Höfding $\frac{1}{2}^+(F; G)$, es pot escriure com

$$W^2(F; G) = \frac{1}{12} + \frac{1}{12} \left(\frac{1}{2}^+(F; G) + \left(\frac{1}{2} - \frac{1}{2}^+(F; G) \right)^2 \right), \quad (6.6)$$

essent $\frac{1}{2}$ i $\frac{1}{2}^+(F; G)$ les mitjanes de les distribucions F i G i $\frac{1}{12}$ i $\frac{1}{12}$ les seves variàncies respectives¹. Vegeu també Cuadras ([17]).

¹En el cas de dues distribucions $F(x)$ i $G(x)$ la primera uniforme en $(a; b)$ i la segona, exponencial de paràmetre θ , tenim

$$W^2(F; G) = \frac{(b-a)^2}{12} + \frac{1}{12} \left(\frac{(b-a)^2}{12} + \frac{1}{\theta^2} + \frac{1}{2} + \frac{a+b}{2} + \frac{1}{\theta} \right)$$

CAPÍTOL 6. DISTÀNCIA DE WASSERSTEIN I ALTRES DESIGUALTATS 136

Siguin les distribucions F i G absolutament contínues i amb densitats f i g respectivament. Suposem que $\frac{3}{4}_1 \cdot \frac{3}{4}_2$.

Com $0 < \frac{1}{2}^+ (F; G) \cdot 1$ tenim

$$2\frac{3}{4}_1\frac{3}{4}_2\frac{1}{2}^+ (F; G) \cdot 2\frac{3}{4}_1\frac{3}{4}_2$$

i, de (6.6), obtenim una cota inferior per $W^2 (F; G)$;

$$\begin{aligned} W^2 (F; G) &\geq \frac{3}{4}_1^2 + \frac{3}{4}_2^2 \text{ i } 2\frac{3}{4}_1\frac{3}{4}_2 + (1 - \frac{1}{2} \text{ i } 1 - \frac{1}{2})^2 \\ &= (\frac{3}{4}_1 \text{ i } \frac{3}{4}_2)^2 + (1 - \frac{1}{2} \text{ i } 1 - \frac{1}{2})^2. \end{aligned}$$

D'altra banda, de la desigualtat (5.5) i fent $\frac{1}{2}_0 = \text{ess inf}_{x \in E} \frac{f(x)}{g(x)}$ tenim

$$\frac{1}{2}_0 = \text{ess inf}_{x \in E} \frac{\frac{1}{2} f(x)}{g(x)} \cdot \frac{\frac{3}{4}_1\frac{3}{4}_2}{\frac{3}{4}_2^2} \frac{1}{2}^+ (F; G), \tag{6.7}$$

d'on

$$\text{i } 2\frac{3}{4}_2^2 \frac{1}{2}_0 \geq \text{i } 2\frac{3}{4}_1\frac{3}{4}_2\frac{1}{2}^+ (F; G),$$

i per tant una cota superior per $W^2 (F; G)$ és

$$W^2 (F; G) \leq \frac{3}{4}_1^2 + \frac{3}{4}_2^2 \text{ i } 2\frac{3}{4}_2^2 \frac{1}{2}_0 + (1 - \frac{1}{2} \text{ i } 1 - \frac{1}{2})^2.$$

Tenim, en definitiva, les desigualtats

$$(\frac{3}{4}_1 \text{ i } \frac{3}{4}_2)^2 + (1 - \frac{1}{2} \text{ i } 1 - \frac{1}{2})^2 \cdot W^2 (F; G) \leq \frac{3}{4}_1^2 + \frac{3}{4}_2^2 (1 - 2\frac{1}{2}_0) + (1 - \frac{1}{2} \text{ i } 1 - \frac{1}{2})^2.$$

En el cas en el qual $\frac{1}{2}_0 = \frac{\frac{3}{4}_1}{\frac{3}{4}_2}$ (distribució normal, Cuesta et al. [28]) les desigualtats anteriors es converteixen en igualtats.

Una altra expressió per a la distància de Wasserstein en la que entra la variabilitat geomètrica, $V_{\pm} (X)$, respecte la distància $\pm (x; y) = \frac{1}{2} \sqrt{|x - y|}$, és

$$\begin{aligned} W^2 (F; U) + V_{\pm} (X) &= \frac{3}{4}_1^2 + \frac{1}{12} + \frac{1}{2} (1 - \frac{1}{2} \text{ i } \frac{1}{2})^2, \tag{6.8} \\ &= \frac{1}{6} \frac{2b^2a^2 + 2ba^2 + 2a^2}{a^2} \text{ i } \frac{9b \text{ i } 3a + 12}{6}, \end{aligned}$$

que coincideix, obviament, amb (6.3).

la qual es dedueix de les fórmules (5.4) i (6.6), considerant $G = U$ distribuïda uniformement en $(0; 1)$. En efecte,

$$\begin{aligned} W^2(F; U) &= \frac{1}{12} + \frac{1}{12} \int_0^1 (F(x) - x)^2 dx + \left(\int_0^1 F(x) dx - \frac{1}{2} \right)^2 \\ &= \frac{1}{12} + \frac{1}{12} \int_0^1 (F(x) - x)^2 dx + \left(\int_0^1 F(x) dx - \frac{1}{2} \right)^2 \\ &= \frac{1}{12} + \frac{1}{12} \int_0^1 (F(x) - x)^2 dx + \left(\int_0^1 F(x) dx - \frac{1}{2} \right)^2, \end{aligned}$$

d'on surt la igualtat (6.8).

També podem obtenir, immediatament, la desigualtat

$$\frac{1}{12} + \frac{1}{12} \int_0^1 (F(x) - x)^2 dx \geq \frac{1}{6} \inf_{x \in E} f(x)g,$$

de les expressions (5.3) i (6.8), que es convertiren en igualtats si i només si F fos una distribució uniformement distribuïda en $(0; 1)$.

$W^2(F_n; F)$ és una bona mesura de la bondat d'ajust, Barrio et al. [29], essent F_n la funció de distribució empírica i F la funció teòrica.

6.5 Exemples

Cuadras i Cuadras [20] calculen la correlació màxima de Hoëding entre la variable aleatòria X , de mitjana μ i variància σ^2 , i una mostra x_1, \dots, x_N de mitjana \bar{x} i variància s^2 , per a les distribucions uniforme, exponencial i logística. Utilitzant aquests resultats podem calcular la distància de Wasserstein per aquests casos.

Distribució uniforme

Sigui U una variable aleatòria que segueix la distribució uniforme $[0; 1]$, i $u_1 < \dots < u_N$ una mostra ordenada. En aquest cas la correlació màxima de Hoëding és

$$\frac{1}{2N} \left(\frac{1}{s} \sum_{i=1}^N u_i(2i-1) - \frac{\bar{x}}{s} \right)^2,$$

on \bar{x} i s són la mitjana i la desviació típica mostrals. Per tant, de l'expressió (6.6), obtenim per a la distància de Wasserstein el valor

$$W^2(F; G) = \frac{1}{12} + \frac{1}{12} \int_0^1 (F(x) - x)^2 dx + \left(\int_0^1 F(x) dx - \frac{1}{2} \right)^2$$

$$\begin{aligned}
 &= s^2 + \frac{1}{12} \sum_{i=1}^N (2i-1) \frac{1}{N^2} \sum_{i=1}^N u_i(2i-1) \frac{1}{2} + \bar{x} \frac{1}{2} \\
 &= s^2 + \bar{x}^2 + \frac{1}{3} \sum_{i=1}^N \frac{1}{N^2} u_i(2i-1).
 \end{aligned}$$

Distribució exponencial

Si considerem la variable aleatòria X seguint la distribució exponencial de mitjana 1 i una mostra de mida N , aleshores la correlació màxima de Hoeffding és

$$\frac{1}{2N} \sum_{i=1}^N x_i \log(N-i)^{N-i} - \log(N+1-i)^{N+1-i} + x_N \log(N);$$

on \bar{x} i s són la mitjana i la desviació típica mostrals. La distància de Wasserstein és

$$\begin{aligned}
 W^2(F; G) &= \frac{1}{4} + \frac{1}{4} \sum_{i=1}^N (2i-1) \frac{1}{N^2} \sum_{i=1}^N x_i A_i + (\bar{x} - 1)^2 \\
 &= s^2 + 1 + \frac{1}{N} \sum_{i=1}^N x_i A_i + (\bar{x} - 1)^2 \\
 &= s^2 + (\bar{x} - 1)^2 + \frac{1}{N} \sum_{i=1}^N x_i A_i,
 \end{aligned}$$

on $A_i = \log(N-i)^{N-i} - \log(N+1-i)^{N+1-i} + x_N \log(N)$.

Distribució logística

En el cas de la distribució logística de mitjana 0 i variància $\frac{1}{3}$ tenim

$$\frac{1}{2N} \sum_{i=1}^N \frac{x_i [A_i - B_i]}{s^2};$$

on

$$\begin{aligned}
 A_i &= (N-i) \log(N-i) + i \log(i), \\
 B_i &= (N-i+1) \log(N-i+1) + (i-1) \log(i-1);
 \end{aligned}$$

La distància de Wasserstein és, per tant

$$W^2(F; G) = \frac{1}{4} + \frac{1}{4} \sum_{i=1}^N (2i-1) \frac{1}{N^2} \sum_{i=1}^N x_i A_i + (\bar{x} - 1)^2$$

$$\begin{aligned}
 &= s^2 + \frac{1}{3} \sum_{i=1}^n \frac{2}{N} x_i [A_i - B_i] + \bar{x}^2 \\
 &= s^2 + \bar{x}^2 + \frac{1}{3} \sum_{i=1}^n \frac{2}{N} x_i [A_i - B_i].
 \end{aligned}$$

6.6 Desigualtats per a la variància

Chernoff [9] demostra que, per una variable aleatòria X amb distribució normal de mitjana 0 i variància 1 i per qualsevol funció absolutament contínua G , es té

$$E \left(G'(X) \right)^2 \leq \text{var} (G(X)), \tag{6.9}$$

sempre que $\text{var} (G(X)) < +1$. Si la funció G és lineal, la desigualtat anterior passa a ser una igualtat.

Cacoullos [8], extén la desigualtat (6.9), en la que tenim una cota superior per a la variància d'una funció de la variable aleatòria normal estandar, donant també una cota inferior per aquesta variància $\text{var} (G(X))$, en la forma

$$[E (G'(X))]^2 \leq \text{var} (G(X)),$$

per tant podem escriure, per una variable aleatòria $X \sim N(0, 1)$ i una funció G absolutament contínua, les desigualtats

$$[E (G'(X))]^2 \leq \text{var} (G(X)) \leq E \left(G'(X) \right)^2, \tag{6.10}$$

obtenint igualtats si la funció G és lineal. Cacoullos es basa en les desigualtats de Cramér-Rao i de Cauchy-Schwarz per demostrar aquestes desigualtats.

Les desigualtats (6.10) prenen, per una variable aleatòria X distribuïda segons una normal de mitjana μ i variància σ^2 , la forma

$$\sigma^2 [E (G'(X))]^2 \leq \text{var} (G(X)) \leq \sigma^2 E \left(G'(X) \right)^2. \tag{6.11}$$

Cacoullos [8] dona també desigualtats d'aquest tipus, relacionant la $E (G'(X))$ amb la $\text{var} (G(X))$, per variables aleatòries amb distribució exponencial, de Poisson i binomial.

Klaassen [48] generalitza les desigualtats (6.11) a qualsevol variable aleatòria. Així, per una variable aleatòria X , binomial de paràmetres n i p , es té

$$\sigma^2 [E ((n - X) G'(X))]^2 \leq \text{var} (G(X)) \leq \sigma^2 E \left((n - X) G'(X) \right)^2.$$

Per una variable aleatòria Poisson de paràmetre λ , tenim

$$\lambda [E(G^0(X))]^2 \cdot \text{var}(G(X)) \cdot \frac{1}{\lambda} E(G^0(X))^2,$$

i en el cas de la distribució logística de mitjana 0 i variància $\frac{1}{12}$ tenim

$$\frac{3}{4} [E(G^0(X))]^2 \cdot \text{var}(G(X)) \cdot \frac{1}{2} E(G^0(X))^2.$$

6.7 Una prova alternativa en el cas logístic

Considerem una variable aleatòria Z seguint la distribució logística amb funció de densitat

$$f(z) = \frac{e^{-z}}{(1 + e^{-z})^2}, \quad -\infty < z < +\infty, \quad (6.12)$$

i, per tant, funció de distribució

$$F(z) = \frac{1}{1 + e^{-z}}, \quad -\infty < z < +\infty.$$

Sigui $K(x; y)$ la covariància del procés X definida a (5.1),

$$K(x; y) = \min_{F(x); F(y)} \int_I F(x)F(y) dx dy,$$

on I és el suport. El teorema de Mercer diu que si R és un nucli continu, simètric i semi-definit positiu, aleshores es compleix:

$$R(s; t) = \sum_{i=1}^{\infty} \lambda_i \tilde{A}_i(s) \tilde{A}_i(t)$$

amb convergència absoluta i uniforme en les dues variables s i t juntes, on $\{\tilde{A}_i\}$ és una base ortonormal de funcions pròpies, essent \tilde{A}_j la funció pròpia amb valor propi λ_j .

Si F és continua K compleix les condicions del teorema de Mercer i per tant:

$$K(x; y) = \sum_{j=1}^{\infty} \lambda_j \tilde{A}_j(x) \tilde{A}_j(y); \quad x, y \in I,$$

on \tilde{A}_j és la funció pròpia amb valor propi λ_j .

CAPÍTOL 6. DISTÀNCIA DE WASSERSTEIN I ALTRES DESIGUALTATS 141

i	$h_i(Z)$	λ_i	\tilde{A}_i
1	$\frac{1}{6}F(Z)$	$\frac{1}{2}$	$\frac{1}{6}f$
2	$\frac{1}{30}f(Z)$	$\frac{1}{6}$	$\frac{1}{30}(2F_i - f)$

Taula 6.1: Dues primeres components principals, valors propis i funcions pròpies per a la variable logística.

En la secció 5.4, expandint X en termes de v.a.'s incorrelacionades $h_n(X)$; hem vist que

$$h_n(x) = \int_{-1}^x \tilde{A}_n(t) dt; \quad x \in I. \quad (6.13)$$

En el cas que ens ocupa, Z variable logística estàndard, tenim

$$h_n(Z) = \frac{1}{n(n+1)} [L_n(2F(Z) - 1) - (2F(Z) - 1)^n], \quad n \geq 1, \quad (6.14)$$

on $L_n(Z)$ són els polinomis de Legendre² sobre $(0; 1)$. Els valors propis són

$$\lambda_n = \frac{1}{n(n+1)}. \quad (6.15)$$

Així de les expressions (6.14), (6.15) i (6.13) obtenim, per exemple, les dues primeres components principals, així com els respectius valors i funcions pròpies. La taula 6.1 mostra els resultats.

Sigui X una variable aleatòria logística general amb variància $\frac{3}{4}$. No hi ha pèrdua de generalitat suposant que la mitjana és $\mu = 0$. La funció de densitat és

$$f(x) = \frac{1}{\sqrt{3}} \frac{e^{-x/\sqrt{3}}}{1 + e^{-x/\sqrt{3}}}, \quad -1 < x < +1,$$

on $\sigma = \frac{\sqrt{3}}{2}$.

² $L_1(x) = \frac{1}{2}(2x - 1)$; $L_2(x) = \frac{1}{2}(6x^2 - 6x + 1)$; $L_3(x) = \frac{1}{2}(20x^3 - 30x^2 + 12x - 1)$ etc., complint les relacions:

$$\int_{-1}^1 L_m(x)L_n(x)dx = \begin{cases} 0; & \text{per } m \neq n; \\ 1 & \text{per } m = n. \end{cases}$$

CAPÍTOL 6. DISTÀNCIA DE WASSERSTEIN I ALTRES DESIGUALTATS 142

Considerem la logística estàndard Z , amb funció de densitat definida en (6.12) i, per tant, $\text{var}(Z) = \frac{1}{3}$. Així podem considerar

$$X = \alpha Z,$$

amb $\alpha = \frac{\sqrt{3}}{2}$.

Sigui G una funció absolutament contínua. Del resultat 16 i tenint en compte que $\int_{\mathbb{R}} \tilde{A}_i(x) f(x) dx = 1, \quad i = 1$, $0, \quad i > 1$, al ser $f(x)$ la primera funció pròpia, podem escriure

$$\begin{aligned} \text{cov}(F(Z), G(Z)) &= \int_{\mathbb{R}^2} (\min_{j=1, \dots, n} \tilde{A}_j(x); F(y)) f(x) G'(y) dy \\ &= \int_{\mathbb{R}^2} \sum_{j=1}^n \tilde{A}_j(x) \tilde{A}_j(y) f(x) G'(y) dy \\ &= \int_{\mathbb{R}^2} \sum_{j=1}^n \tilde{A}_j(x) \tilde{A}_j(y) f(x) dx G'(\alpha z) dy \\ &= \int_{\mathbb{R}^2} (\sum_{j=1}^n \tilde{A}_j(x) \tilde{A}_j(y) + \dots) f(x) dx G'(\alpha z) dy \\ &= \int_{\mathbb{R}^2} \frac{1}{2} f(x) f(y) + \dots G'(\alpha z) dy \\ &= \int_{\mathbb{R}} \frac{1}{2} f(y) G'(\alpha z) dy \\ &= \frac{\alpha}{2} E[G'(\alpha z)]. \end{aligned}$$

Per la desigualtat de Cauchy-Schwarz tenim

$$(\text{cov}(F(Z), G(Z)))^2 \leq \text{var}(F(Z)) \text{var}(G(\alpha Z)),$$

$$\frac{\alpha^2}{4} (E[G'(\alpha Z)])^2 \leq \text{var}(F(Z)) \text{var}(G(\alpha Z)),$$

i considerant que $\alpha = \frac{\sqrt{3}}{2}$ i que $\text{var}(F(Z)) = \frac{1}{12}$, al ser $F(Z)$ uniforme en $(0; 1)$, tenim la desigualtat

$$\frac{3^2}{4} \text{var}(X) [E(G'(X))]^2 \leq \text{var}(G(X)),$$

vàlida per a una variable general $X = \alpha Z$.

6.8 Una desigualtat general

Sigui G una funció derivable. Suposem que G^0 és de quadrat integrable. Al formar les funcions pròpies \tilde{A}_n de K un sistema ortonormal complet, tenim que

$$G^0 = a_1 \tilde{A}_1 + \tilde{A}^?,$$

essent \tilde{A}_1 la primera funció pròpia de K i $\tilde{A}^?$ una funció ortogonal a \tilde{A}_1 respecte al producte escalar $\langle a(x); b(x) \rangle = \int_a^b a(x)b(x)dx$: Aleshores

$$\begin{aligned} a_1 &= \int_a^b \tilde{A}_1(x) G^0(x) dx \\ &= h_1(b) E[G^0(Y)], \end{aligned}$$

on $h_1(b) = \int_a^b \tilde{A}_1(x) dx$ i Y és la variable aleatòria amb densitat

$$f_1(y) = \frac{\tilde{A}_1(y)}{h_1(b)} \quad a < y < b.$$

Recordem que \tilde{A}_1 és una funció positiva. A més

$$\int_a^b f_1(y) dy = \frac{1}{h_1(b)} \int_a^b \tilde{A}_1(y) dy = 1.$$

D'altra banda, integrant G^0 obtenim

$$G = a_1 h_1 + \odot;$$

on $h_1(x) = \int_a^x \tilde{A}_1(t) dt$; $\odot(x) = \int_a^x \tilde{A}^?(t) dt$; vegeu (6.13). Llavors

$$\text{var}(G(X)) = a_1^2 \text{var}(h_1(X)) + \text{var}(\odot(X)) + 2a_1 \text{cov}(h_1(X); \odot(X)):$$

Tenint en compte que la distribució de $(X; X)$ és $\min\{F(x); F(y)\}$; aplicant la fórmula (6.5) que dona la covariància entre funcions:

$$\begin{aligned} \text{cov}(h_1(X); \odot(X)) &= \int_a^b \int_a^b K(x; y) dh_1(x) d\odot(x) \\ &= \int_a^b \int_a^b K(x; y) \tilde{A}_1(x) dx \tilde{A}^?(y) dy \\ &= \int_a^b \tilde{A}_1(y) \tilde{A}^?(y) dy = 0: \end{aligned}$$

Conseqüentment

$$\begin{aligned} \text{var}(G(X)) &\leq a_1^2 \text{var}(h_1(X)) \\ &= h_1(b)^2 \text{var}(h_1(X)) (E[G^0(Y)])^2: \end{aligned}$$

CAPÍTOL 6. DISTÀNCIA DE WASSERSTEIN I ALTRES DESIGUALTATS 144

amb igualtat si G és h_1 :

Hem obtingut el següent resultat.

Resultat 17 Sigui X una variable aleatòria absolutament contínua amb funció de distribució F i suport $(a; b)$. Suposem que existeix la descomposició

$$K(x; y) = \sum_{j=1}^{\infty} \tilde{A}_j(x) \tilde{A}_j(y);$$

on $K(x; y) = \min_{F(x); F(y)} \int_0^1 F(x)F(y) g_i$; $F(x)F(y)$; \tilde{A}_n és el sistema ortonormal complet de funcions pròpies i $\lambda_1, \lambda_2, \dots$ els corresponents valors propis.

Sigui Y una variable aleatòria amb funció de densitat

$$f_1(y) = \frac{\tilde{A}_1(y)}{h_1(b)} \quad a < y < b.$$

Si G és una funció derivable i $G(X)$ té variància finita, aleshores es verifica la desigualtat

$$\text{var}(G(X)) \leq h_1(b)^2 \text{var}[h_1(X)] [EG^0(Y)]^2;$$

amb igualtat si G és h_1 :

La desigualtat per a la distribució logística surt també com una conseqüència d'aquesta desigualtat general, atès que en el cas logístic tenim

$$f_1 = f$$

i la distribució de Y és la mateixa que la de X .

Capítol 7

Desigualtats a partir de la raó de versemblança

7.1 Introducció

El test de la raó de versemblança és utilitzat per contrastar hipòtesis compostes del tipus

$$\begin{aligned} H_0 &: (\mu_1; \dots; \mu_k) \in \Omega \\ H_1 &: (\mu_1; \dots; \mu_k) \in \Omega^c \end{aligned}$$

essent $\mu_1; \dots; \mu_k$ els paràmetres de la funció $f(x; \mu_1; \dots; \mu_k)$, funció de densitat d'una variable aleatòria X : Ω és una regió paramètrica k -dimensional i Ω^c un subconjunt d' Ω .

Aquest test parteix de la raó de versemblança

$$\alpha = \frac{L_3(\mathbf{b})}{L(\mathbf{b})}$$

on

$$L(\mathbf{b}) = \sup_{\mu \in \Omega} L(\mathbf{x}; \mu) \quad \text{i} \quad L_3(\mathbf{b}) = \sup_{\mu \in \Omega^c} L(\mathbf{x}; \mu)$$

essent $L = f(x_1; \mu) \dots f(x_n; \mu)$ la funció de versemblança.

Mitjançant la propietat d'acotació de la raó de versemblança

$$0 < \alpha < 1, \tag{7.1}$$

Stefanski [61] demostra les conegudes desigualtats entre les mitjanes aritmètica, geomètrica i harmònica

$$\frac{\sum_{i=1}^k a_i}{k} \geq \sqrt[k]{\prod_{i=1}^k a_i} \geq \frac{k}{\sum_{i=1}^k \frac{1}{a_i}}. \quad (7.2)$$

Utilitza el test

$$H_0 : \mu_1 = \dots = \mu_k \quad (7.3)$$

$$H_1 : \text{no tots els } \mu_i \text{ iguals, } i = 1; \dots; k,$$

sobre els paràmetres de k variables aleatòries independents, $X_1; \dots; X_k$, amb funcions de densitat

$$f_i(x; \mu_i) = \mu_i \exp(-\mu_i x) \quad ; x > 0; \mu_i > 0; i = 1; \dots; k, \quad (7.4)$$

$$f_i(x; \mu_i) = \frac{\mu_i}{x^2} \exp\left(-\frac{\mu_i}{x}\right) \quad ; x > 0; \mu_i > 0; i = 1; \dots; k, \quad (7.5)$$

respectivament.

Aquest capítol el dediquem a la recerca d'altres desigualtats utilitzant aquest mateix procediment. En primer lloc donarem una demostració del teorema de les mitjanes aritmètica i geomètrica¹.

7.2 Teorema de les mitjanes aritmètica i geomètrica

Siguin $X_i; i = 1; \dots; k$, k variables aleatòries independents amb funció de densitat, exponencial, (7.4).

Considerem el test (7.3). Sigui $x_{i_1}; \dots; x_{i_{n_i}}, i = 1; \dots; k$, mostres aleatòries simples de les variables $X_i, i = 1; \dots; k$.

La funció de versemblança és

$$L = \prod_{i=1}^{n_1} \mu_1 \exp(-\mu_1 x_{1j}) \dots \prod_{i=1}^{n_k} \mu_k \exp(-\mu_k x_{kj}).$$

¹Hardy, Littlewood i Pólya [40] expressen així la desigualtat (7.8).

CAPÍTOL 7. DESIGUALTATS A PARTIR DE LA RAÓ DE VERSEMBLANÇA 147

Sota H_0 tenim

$$\ln L_0 = \sum_{i=1}^k n_i \ln \sum_{j=1}^k x_{ij} \prod_{i=1}^k \prod_{j=1}^k x_{ij}^{-n_i},$$

per tant, de

$$\frac{\partial \ln L_0}{\partial x_{ij}} = \frac{1}{x_{ij}} \sum_{i=1}^k n_i \prod_{i=1}^k \prod_{j=1}^k x_{ij}^{-n_i} = 0,$$

obtenim

$$x_{ij} = \frac{P_k}{P_k} \frac{P_k^{-1} n_i}{\sum_{j=1}^k x_{ij}}.$$

El màxim de L sota H_0 és

$$\begin{aligned} L(\mathbf{b}) &= \prod_{i=1}^k \frac{P_k^{-1} n_i}{\sum_{j=1}^k x_{ij}} \exp \left(\sum_{i=1}^k \sum_{j=1}^k x_{ij} \prod_{i=1}^k \prod_{j=1}^k x_{ij}^{-n_i} \right) \\ &= \prod_{i=1}^k \frac{P_k^{-1} n_i}{\sum_{j=1}^k x_{ij}} \exp \left(\sum_{i=1}^k n_i \right). \end{aligned}$$

De la mateixa manera, sota H_1 tenim

$$\ln L_1 = n_1 \ln \sum_{j=1}^k x_{1j} + \dots + n_k \ln \sum_{j=1}^k x_{kj},$$

i de

$$\frac{\partial \ln L_1}{\partial x_{ij}} = \frac{n_i}{x_{ij}} \prod_{j=1}^k x_{ij}^{-n_i} = 0,$$

obtenim

$$x_{ij} = \frac{P_k^{-1} n_i}{\sum_{j=1}^k x_{ij}} = \frac{1}{x_i}.$$

El màxim de L sota H_1 és

$$L(\mathbf{b}) = \prod_{i=1}^k \frac{P_k^{-1} n_i}{x_i} \exp \left(\sum_{i=1}^k n_i \right).$$

La raó de versemblança queda

$$\alpha = \frac{\prod_{i=1}^k \frac{P_k^{-1} n_i}{\sum_{j=1}^k x_{ij}} \exp \left(\sum_{i=1}^k \sum_{j=1}^k x_{ij} \prod_{i=1}^k \prod_{j=1}^k x_{ij}^{-n_i} \right)}{\prod_{i=1}^k \frac{P_k^{-1} n_i}{x_i} \exp \left(\sum_{i=1}^k n_i \right)}$$

$$\mu = \frac{\prod_{i=1}^k \prod_{j=1}^{n_i} x_{ij}}{\prod_{i=1}^k \bar{x}_i^{n_i}} \quad (7.6)$$

Tenint en compte la desigualtat (7.1), podem escriure (7.6) en la forma

$$\mu = \frac{\prod_{i=1}^k \frac{1}{\bar{x}_i^{n_i}} \cdot \frac{\prod_{j=1}^{n_i} x_{ij}}{\bar{x}_i^{n_i}}}{\prod_{i=1}^k \bar{x}_i^{n_i}}$$

Si posem $s_i = \prod_{j=1}^{n_i} x_{ij}$; obtenim el següent resultat.

Resultat 18 Siguin n_1, \dots, n_k nombres enters positius i s_1, \dots, s_k nombres reals positius. S'acompleix la següent desigualtat

$$\left(\frac{s_1}{n_1}\right)^{n_1} \dots \left(\frac{s_k}{n_k}\right)^{n_k} \leq \left(\frac{s_1 + \dots + s_k}{n_1 + \dots + n_k}\right)^{n_1 + \dots + n_k};$$

amb igualtat si els $s_i = n_i$ són iguals.

Fent els canvis

$$n_i = b_i > 0; \quad \frac{\prod_{j=1}^{n_i} x_{ij}}{n_i} = a_i > 0; \quad i = 1; \dots; k, \quad (7.7)$$

la desigualtat anterior és un cas particular de la desigualtat més general (Hardy et al., [40])

$$a_1^{b_1} \dots a_k^{b_k} \leq \frac{a_1 b_1 + \dots + a_k b_k}{b_1 + \dots + b_k}^{b_1 + \dots + b_k}; \quad (7.8)$$

on $a_1, \dots, a_k; b_1, \dots, b_k$ són nombres reals positius, amb igualtat si $a_1 = \dots = a_k$.

De (7.6) també podem obtenir la desigualtat

$$\frac{\prod_{i=1}^k a_i^{a_i}}{b_1 + b_2 + \dots + b_k}^{a_1 + a_2 + \dots + a_k} \leq \frac{\prod_{i=1}^k a_i^{a_i}}{b_1^{a_1} \dots b_k^{a_k}}, \quad (7.9)$$

on a_1, \dots, a_k són enters i b_1, \dots, b_k són nombres reals positius, amb igualtat si $\frac{a_1}{b_1} = \dots = \frac{a_k}{b_k}$, fent els canvis

$$n_i = a_i > 0; \quad x_{ij} = b_j > 0; \quad i = 1; \dots; k, \quad (7.10)$$

invertint els termes i tenint en compte la desigualtat (7.1). Com veurem més endavant, aquesta desigualtat també val si a_1, \dots, a_k són reals positius.

De fet si haguéssim partit de k variables aleatòries independents amb funció de densitat (7.5) en lloc de (7.4) arribaríem a les mateixes desigualtats fent els canvis adients.

També, partint de $X_i; i = 1; \dots; k$, k variables aleatòries independents amb funció de densitat Poisson

$$f_i(x; \lambda_i) = \exp(-\lambda_i) \frac{\lambda_i^x}{x!}; x = 0; 1; 2; \dots; i = 1; \dots; k,$$

arribem a les mateixes desigualtats (7.8) i (7.9). Si en aquest cas féssim $\lambda_1 = \dots = \lambda_k = 1$ obtindríem la desigualtat

$$x^{kx} \cdot x_1^{x_1} \dots x_k^{x_k},$$

la qual podríem expressar, per $k = 3$, en la forma més coneguda

$$a^a b^b c^c \leq \frac{\mu^{a+b+c}}{3} \prod_{a+b+c}^1,$$

amb $a; b; c > 0$, o, en forma més general

$$a_1^{a_1} \dots a_k^{a_k} \leq \frac{\mu^{a_1 + \dots + a_k}}{k} \prod_{a_1 + \dots + a_k}^1,$$

on $a_1; \dots; a_k$ són enters positius.

7.3 Generalització

Stefanski [61] considera, també, la generalització del procediment comentat en la secció 7.1, a variables aleatòries amb funció de densitat del tipus

$$f_i(x; \mu_i) = \exp\left(-\frac{x \mu_i}{\lambda} - \frac{b(\mu_i)}{\lambda} + c(x; \lambda)\right); \lambda > 0; i = 1; \dots; k, \quad (7.11)$$

on b és una funció convexa dues vegades diferenciable i c una funció bivariant. És a dir, considera la generalització a la família de distribucions exponencial.

Considerant la hipòtesi

$$\begin{aligned} H_0 &: \mu_1 = \dots = \mu_k \\ H_1 &: \text{no tots els } \mu_i \text{ iguals, } i = 1; \dots; k, \end{aligned} \quad (7.12)$$

CAPÍTOL 7. DESIGUALTATS A PARTIR DE LA RAÓ DE VERSEMBLANÇA 150

la raó de versemblança és

$$\alpha = \frac{\exp \left\{ \sum_{i=1}^k \bar{x}_i \left[\sum_{j=1}^k \frac{1}{b_j} \ln \frac{b_j}{b_j + x_{ij}} \right] \right\}}{\exp \left\{ \sum_{i=1}^k \bar{x}_i \left[\sum_{j=1}^k \frac{1}{b_j} \ln \frac{b_j}{b_j + x_{ij}} \right] \right\}}$$

la qual ens porta a la desigualtat

$$\bar{x}_i \left[\sum_{j=1}^k \frac{1}{b_j} \ln \frac{b_j}{b_j + x_{ij}} \right] \geq \frac{1}{k} \sum_{i=1}^k \bar{x}_i \left[\sum_{j=1}^k \frac{1}{b_j} \ln \frac{b_j}{b_j + x_{ij}} \right] \quad (7.13)$$

La funció exponencial (7.4) és un cas particular de la funció (7.11), fent

$$\mu_i = \bar{x}_i; \quad \lambda = 1; \quad c = 0; \quad b(t) = \ln \frac{1}{1-t} \quad (7.14)$$

Amb els canvis (7.14), la desigualtat (7.13) es transforma en

$$\frac{\sum_{i=1}^k a_i}{k} \geq \frac{1}{k} \sum_{i=1}^k a_i$$

Un altre cas particular de la funció de densitat (7.11) correspon al cas de k variables aleatòries independents amb distribució normal de mitjana μ_i i igual desviació típica σ . Els canvis a fer són

$$\mu_i = \bar{x}_i; \quad \lambda = \frac{1}{2\sigma^2}; \quad c(x; \lambda) = \frac{x^2}{2\lambda} \ln \frac{1}{1 - 2\lambda x^2} \quad i \quad b(t) = \frac{1}{2} t^2, \quad (7.15)$$

i la desigualtat que dóna és la coneguda

$$\sum_{i=1}^k (x_{ij} - \bar{x}_i)^2 \geq 0.$$

Si per contrastar les hipòtesis (7.12) agaféssim k mostres de mida n_1, \dots, n_k ; de les k variables de...nides per les funcions de densitat (7.11), la raó de versemblança seria

$$\alpha = \frac{\exp \left\{ \sum_{i=1}^k \bar{x}_i \left[\sum_{j=1}^k \frac{1}{b_j} \ln \frac{b_j}{b_j + x_{ij}} \right] \right\}}{\exp \left\{ \sum_{i=1}^k \bar{x}_i \left[\sum_{j=1}^k \frac{1}{b_j} \ln \frac{b_j}{b_j + x_{ij}} \right] \right\}}$$

la qual ens porta, tenint en compte (7.1), a la desigualtat

$$b^{0i-1} \frac{\prod_{i=1}^k \prod_{j=1}^{n_i} x_{ij}}{\prod_{i=1}^k n_i} \leq b^{0i-1} \frac{\prod_{i=1}^k \prod_{j=1}^{n_i} x_{ij}}{\prod_{i=1}^k n_i} \quad (7.16)$$

Fent els canvis (7.14) i (7.7), de la desigualtat (7.16) sortiria la desigualtat (7.8) i amb els canvis (7.14) i (7.10), la desigualtat (7.9).

En el cas de k variables aleatòries independents amb distribució normal de mitjana μ_i i igual desviació típica σ , amb els canvis (7.15) i mostres de mida n_1, \dots, n_k , obtindríem

$$\frac{\prod_{i=1}^k \prod_{j=1}^{n_i} x_{ij}}{\prod_{i=1}^k n_i} \leq \frac{\prod_{i=1}^k \prod_{j=1}^{n_i} x_{ij}}{\prod_{i=1}^k n_i}$$

la qual podríem expressar, amb els canvis $n_i = b_i > 0$; $\prod_{j=1}^{n_i} x_{ij} = a_i > 0$; $i = 1, \dots, k$,

$$\frac{(a_1 + a_2 + \dots + a_k)^2}{b_1 + b_2 + \dots + b_k} \leq \frac{a_1^2}{b_1} + \dots + \frac{a_k^2}{b_k}$$

on a_1, \dots, a_k són nombres reals positius i b_1, \dots, b_k són enters.

De fet, escollint diferents funcions convexes $b(t)$, de (7.16) obtindríem altres desigualtats, com per exemple

$$(a_1 + \dots + a_k)^{\frac{p}{p-1}} \leq k^{\frac{1}{p-1}} \left(a_1^{\frac{p}{p-1}} + \dots + a_k^{\frac{p}{p-1}} \right)^{\frac{1}{p}}$$

que sortiria de $b(t) = \frac{1}{p} t^p$, $p > 1$ i $t > 0$.

Finalment, de $b(t) = 1 + t^2$ obtindríem

$$\frac{\bar{x}}{1 + \bar{x}^2} \leq \frac{1}{k} \sum_{i=1}^k \frac{x_i}{1 + x_i^2}$$

7.4 Altres desigualtats

En les seccions 7.2 i 7.3 hem fet referència a variables aleatòries distribuïdes segons una distribució exponencial, normal i Poisson com una generalització

CAPÍTOL 7. DESIGUALTATS A PARTIR DE LA RAÓ DE VERSEMBLANÇA 152

de la família exponencial definida per la funció de densitat (7.11). Hem trobat desigualtats a través de la propietat (7.1) de la raó de versemblança aplicada a contrastar un test del tipus (7.3).

Bé directament, com hem fet a la secció 7.2, o com a conseqüència de (7.13) o (7.16) podem trobar desigualtats partint de variables aleatòries distribuïdes, per exemple, segons una Bernoulli, una binomial o una Pascal.

Distribucions Bernoulli i binomial

Partint de k variables independents distribuïdes segons una Bernoulli

$$f_i(x; \theta_i) = \theta_i^x (1 - \theta_i)^{1-x}; \quad x = 0, 1; \quad 0 < \theta_i < 1 \quad i = 1, \dots, k,$$

que com la binomial, pertany també a la família exponencial de distribucions, i tenint en compte els canvis

$$\mu_i = \ln \frac{\theta_i}{1 - \theta_i}; \quad \eta_i = 1; \quad c(x; \theta_i) = 0 \quad i \quad b(t) = \ln(1 + \exp(t)), \quad (7.17)$$

i per tant

$$b^0(t) = \frac{\exp(t)}{1 + \exp(t)} \quad b^{0,1}(x) = \ln \frac{x}{1 - x},$$

arribem, de (7.13), a la desigualtat

$$\bar{x} \ln \frac{\bar{x}}{1 - \bar{x}} - \sum_{i=1}^k \theta_i \ln \frac{\theta_i}{1 - \theta_i} \geq \frac{\sum_{i=1}^k x_i \ln \frac{x_i}{1 - x_i} - \sum_{i=1}^k \theta_i \ln \frac{\theta_i}{1 - \theta_i}}{k}. \quad (7.18)$$

De la mateixa manera, considerant la distribució binomial i modificant els canvis (7.17) amb $b(t) = n \ln(1 + \exp t)$ i $c(x; \theta) = \ln \binom{n}{x}$, tindríem la desigualtat

$$\bar{x} \ln \frac{\bar{x}}{n - \bar{x}} - n \ln \frac{\theta}{1 - \theta} \geq \frac{\sum_{i=1}^k x_i \ln \frac{x_i}{n - x_i} - n \ln \frac{\theta}{1 - \theta}}{k}.$$

Distribució Pascal o geomètrica

Siguin k variables aleatòries independents amb funció de distribució

$$f_i(x; \theta_i) = \theta_i (1 - \theta_i)^{x-1}; \quad x = 1, 2, \dots; \quad 0 < \theta_i < 1 \quad i = 1, \dots, k.$$

CAPÍTOL 7. DESIGUALTATS A PARTIR DE LA RAÓ DE VERSEMBLANÇA 153

Com aquesta distribució també pertany a la família exponencial considerant

$$\mu_i = \ln(1 - \lambda_i) \quad ; \quad \lambda_i = 1 \quad ; \quad c(x; \lambda) = 0 \quad i \quad b(t) = \ln \frac{\exp(t)}{1 - \exp(t)}$$

i, per tant

$$b^0(t) = \frac{1}{1 - \exp(t)} \quad b^{0i-1}(x) = \ln \frac{x - 1}{x}$$

fent servir la desigualtat (7.13) obtindríem

$$\bar{x} \ln \frac{\bar{x} - 1}{\bar{x}} \leq \ln(\bar{x} - 1) \cdot \frac{\sum_{i=1}^k x_i \ln \frac{x_i - 1}{x_i} - \sum_{i=1}^k \ln(x_i - 1)}{k}$$

Distribució de Cauchy i distribució de Laplace

Considerant 2 variables aleatòries independents amb funció de distribució de Cauchy

$$f_i(x; \lambda_i) = \frac{1}{\lambda_i \left(1 + \frac{x^2}{\lambda_i^2}\right)^2} \quad ; \quad -\lambda_i < x < \lambda_i \quad ; \quad \lambda_i > 0, \quad i = 1; 2,$$

o bé 2 variables aleatòries independents amb funció de distribució de Laplace

$$f_i(x; \lambda_i) = \frac{1}{\lambda_i} \exp(-|x|/\lambda_i) \quad ; \quad -\lambda_i < x < \lambda_i \quad ; \quad \lambda_i > 0, \quad i = 1; 2,$$

obtenim, respectivament, les desigualtats

$$\frac{(x_1 - x_2)^2}{(x_1 + x_2)^2} \leq 0, \quad \frac{4x_1x_2}{(x_1 + x_2)^2} \exp\left(-\frac{x_1x_2}{x_1x_2}\right) \leq 1,$$

la qual es transforma també en: $(x_1 - x_2)^2 \leq 0$, fent que $x_1 > 0$ i $x_2 > 0$.

Considerant k variables distribuïdes segons Laplace tindríem un altre cop, considerant, $x_i > 0$, $i = 1; \dots; k$,

$$\frac{\sum_{i=1}^k x_i}{k} \geq \left(\prod_{i=1}^k x_i\right)^{\frac{1}{k}}$$

la coneguda desigualtat entre les mitjanes aritmètica i geomètrica. Per tant no obtenim desigualtats noves.

CAPÍTOL 7. DESIGUALTATS A PARTIR DE LA RAÓ DE VERSEMBLANÇA154

Capítol 8

Resum i conclusions

8.1 Resum

En aquesta memòria hem estudiat diverses desigualtats d'interès, principalment, en regressió i anàlisi multivariant.

En el capítol primer hem presentat un resultat desconcertant en un primer cop d'ull però que ens pot avisar d'alguns dels perills de la utilització, a la lleugera, de la regressió múltiple, com, per exemple, confiar en els diagrames de punts X-Y i en les correlacions simples, descartar variables com a remei per evitar la multicolinealitat, usar el mètode de selecció "forward", ús de la regressió en components principals i també l'ús de les tècniques de representació gràfica al llarg dels primers eixos principals. Aquest resultat és la desigualtat de Hamilton-Routledge-Cuadras, abreujadament HRC,

$$R^2 > \sum_{i=1}^k r_{yx_i}^2,$$

on R^2 és el quadrat del coeficient de correlació múltiple i r_{yx_i} és el coeficient de correlació simple entre la variable dependent Y i la variable explicativa X_i : El fet de verificar-se la desigualtat HRC adverteix que la variable resposta depèn de l'anomenat "soroll", les darreres components principals, les de variància petita, més que de la part de variabilitat de les dades que estiguem considerant. Tenir en compte aquest fet, pot ajudar a millorar la utilització pràctica de l'anàlisi de regressió múltiple, així com les tècniques de reducció de la dimensió.

La desigualtat HRC ens diu que no sempre variables correlacionades contenen informació redundant.

Hem interpretat la desigualtat HRC per components principals, donant com a condició per a que es veri...qui,

$$\sum_{i=1}^p r_{y_i}^2 (1 - \lambda_i) > 0;$$

on λ_i és el i -èssim valor propi de la matriu de correlació R : Aquesta condició la podem expressar dient que la desigualtat HRC es veri...carà, en situacions pràctiques i parlant de manera informal, quan les correlacions simples entre la variable dependent i les components principals amb variància petita siguin grans, mentre que les correlacions simples entre la variable dependent i les components principals amb variància gran siguin petites.

També hem vist com l'acompliment de la desigualtat HRC és una condició necessària per a que la p -èssima component principal sigui l'única component que redueixi la suma de quadrats residual.

Deixant el model de regressió múltiple i considerant només correlacions, en el capítol segon hem estudiat un altre resultat sorprenent i aparentment poc intuïtiu com és el fet següent. Fixat el nombre de variables explicatives i mantenint constant la matriu R , encara que augmentem el nombre de variables correlacionades positivament amb la variable dependent, el coeficient de determinació no té per que augmentar, com semblaria normal al augmentar la influència de les variables explicatives sobre aquella. Hem interpretat, també, aquest fet per components principals i hem estudiat una generalització. Essent Y_1 i Y_2 dues variables resposta i X_1, \dots, X_p variables explicatives és possible tenir d'una banda

$$\sum_{j=1}^p r_{y_1 x_j}^2 > \sum_{j=1}^p r_{y_2 x_j}^2;$$

és a dir, Y_1 està aparentment més correlacionada amb les variables explicatives que Y_2 , i d'altra banda

$$R_1^2 < R_2^2;$$

fet que és degut a la influència sobre Y_i de les components principals amb variància petita.

En aquest capítol segon també hem trobat una versió forta per a la desigualtat HRC en el sentit que

$$\sum_{i=1}^p R^2_{yx_i} = k;$$

essent k , $1 < k < p$, una constant donada.

Un exemple al ...nal del capítol il·lustra la idea, ja comentada en el capítol primer, que tant el mètode de selecció forward com el mètode de la regressió en components principals poden no ser adients.

En el capítol tercer examinem la paradoxa de Simpson per a la interpretació grà...ca de la qual incloem una breu descripció de l'anàlisi de correspondències. El que fem en aquest capítol és utilitzar l'anàlisi de correspondències per estudiar grà...cament les taules de contingència. Hem ampliat els grà...cs habituals de la solució asimètrica de l'anàlisi de correspondències per interpretar millor la paradoxa. Estudiant els exemples il·lustratius de la paradoxa de Simpson hem vist que les dades es comporten de manera diferent, depenent de si tenim o no en compte les poblacions.

Aquest fet el retrobem en el capítol quart al estudiar la desigualtat entre distàncies, Mahalanobis més gran que K. Pearson,

$$(\mathbf{1}_1 \text{ i } \mathbf{1}_2)^0 \mathbf{S}^{-1} (\mathbf{1}_1 \text{ i } \mathbf{1}_2) = \pm_M^2 > \pm_K^2 = (\mathbf{1}_1 \text{ i } \mathbf{1}_2)^0 \mathbf{S}_0^{-1} (\mathbf{1}_1 \text{ i } \mathbf{1}_2),$$

essent $\mathbf{1}_1$ i $\mathbf{1}_2$ els vectors de mitjanes respectius de les poblacions p -dimensionals $\mathbf{1}_1 = (\mathbf{1}_1; \mathbf{S})$ i $\mathbf{1}_2 = (\mathbf{1}_2; \mathbf{S})$; amb \mathbf{S} la matriu de variàncies-covariàncies en les dues poblacions i on $\mathbf{S}_0 = \text{diag}(\mathbf{S})$.

Podem dir que la desigualtat en distàncies, Mahalanobis més gran que K. Pearson, és una versió multivariant de la paradoxa de Simpson.

És aquest fet el que ha motivat incloure en aquesta memòria la paradoxa de Simpson.

Si la desigualtat HRC la escrivim en la forma

$$\mathbf{1}_{yx}^0 \mathbf{R}^{-1} \mathbf{1}_{yx} > \mathbf{1}_{yx}^0 \mathbf{1}_{yx},$$

on $\mathbf{1}_{yx}^0 = (r_{yx_1}; \dots; r_{yx_p})$; sembla adient buscar una altra versió utilitzant distàncies entre dues poblacions.

Aleshores en aquest capítol quart comparem la distància de Mahalanobis, la qual conté la redundància entre les variables, i la distància de K. Pearson

que no la conté. La desigualtat entre distàncies, $\pm_M^2 > \pm_K^2$; expressa, com la desigualtat HRC, el fet que no sempre variables correlacionades contenen informació redundant.

En el cas de dues poblacions hem vist com la desigualtat $\pm_M^2 > \pm_K^2$ es presenta quan la direcció del segment que uneix les mitjanes de les dues poblacions segueix, pràcticament, la direcció de la segona component principal. Més en general, trobarem la desigualtat $\pm_M^2 > \pm_K^2$, quan $\mathbf{h}_1, \mathbf{h}_2$ segueixi pràcticament la direcció de les components principals de variància petita. Podem afirmar, per tant, que la direcció geomètrica que segueixen les mitjanes de les dues poblacions és diferent a la direcció principal que segueixen les variables observades, direcció que ens ve donada per la primera component principal.

Després de generalitzar a més de dues poblacions aquesta desigualtat, la qual cosa ens ha permès afirmar que el seu acompliment indica una anormal connexió entre model i dades, ortogonalitat entre la direcció que segueix el model i la principal variabilitat de les dades, la qual ve donada per la primera component principal, l'hem estudiat des del punt de vista de l'anàlisi multivariant de la variància.

A continuació hem estudiat de manera exhaustiva un exemple, clàssic en anàlisi multivariant, degut a Fisher, per al que es verifica la desigualtat distància de Mahalanobis més gran que distància de K. Pearson.

Hem posat especial èmfasi en l'estudi de l'aspecte gràfic de l'exemple de Fisher i per fer-ho hem hagut, al voler representar els individus i les components principals de cada població considerant les quatre variables estudiades per Fisher, d'utilitzar la representació en l'espai canònic. Hem explicat com obtenir la representació de les components principals en l'espai canònic.

Per acabar aquest capítol quart hem estudiat la paradoxa de Rao coneguda també com la "maledicció de la dimensionalitat", i hem vist com aquesta no es pot donar si es verifica la desigualtat $\pm_M^2 > \pm_K^2$:

El capítol cinquè l'hem dedicat a la possibilitat de construir, en el cas univariant i la seva extensió natural al cas bivariant, funcions de densitat a partir de distàncies o dissimilaritats i estudiar mesures de divergència entre aquesta densitat i la vertadera densitat. Això pot tenir aplicació en anàlisi discriminant al poder utilitzar distàncies en lloc de funcions de densitat.

En el desenvolupament del capítol hem considerat diverses desigualtats,

començant amb la que fa referència a la variabilitat geomètrica, noció que com la de funció de proximitat, han estat recordades al principi del capítol. Aquesta desigualtat ens dona una cota inferior i una cota superior per a la variabilitat geomètrica d'una variable i es presenta en la forma

$$\frac{1}{6} \operatorname{ess\,inf}_{x \in E} f(x) \cdot V_{\pm}(X) \cdot \frac{\sigma^2}{3},$$

on $V_{\pm}(X)$ és la variabilitat geomètrica de la variable aleatòria X ; amb mitjana i variància, σ^2 ; ...nites i densitat $f(x)$; respecte la distància $\pm(x; y) = \int |x - y| f(x) f(y) dx dy$:

Després de donar algun exemple de construcció de densitats a partir de la funció de proximitat en el cas d'una distribució uniforme i d'una exponencial, hem passat a comentar la desigualtat entre la variabilitat geomètrica i l'entropia de Shannon, desigualtat que ens assegura que l'entropia de Shannon, $H(f) = - \int_E f(x) \ln f(x) dx$, la qual, mesura també la variabilitat de la variable X , és una cota inferior de la variabilitat geomètrica.

Finalment hem fet extensives aquestes nocions al cas bivariant trobant una nova desigualtat que ens permet afirmar que la densitat construïda a partir de la dissimilaritat conjunta considerant la redundància de les variables, és més propera a la densitat vertadera que la densitat construïda a partir de la distància definida suposant la independència estocàstica de les variables X i Y .

En el capítol sisè hem estudiat algunes desigualtats relacionades amb la distància de Wasserstein, i uns altres tipus de desigualtats per a la variància d'una funció d'una variable aleatòria, que tenen aspectes semblants a la desigualtat de Cramér-Rao.

En la primera part del capítol es donen definicions per a la distància de Wasserstein entre dues distribucions $F; G$, $W^2(F; G)$; en les que apareixen la correlació màxima de Höfding, així com les mitjanes μ_1 i μ_2 de les distribucions i les seves variàncies σ_1^2 i σ_2^2 i la variabilitat geomètrica, $V_{\pm}(X)$ de la variable aleatòria X respecte la distància \pm . A partir d'aquí sorgeixen desigualtats del tipus

$$(\sigma_1^2 + \sigma_2^2)^2 + (\mu_1 - \mu_2)^2 \cdot W^2(F; G) \cdot \frac{\sigma_1^2 + \sigma_2^2}{2} + (\mu_1 - \mu_2)^2,$$

on $\frac{1}{2} = \text{ess inf}_{x \in E} \frac{f(x)}{g(x)}$; $f(x)$ i $g(x)$ densitats respectives de F i G , i també

$$\frac{1}{4} + \frac{1}{12} \leq \frac{1}{3} + \frac{1}{2} \cdot W^2(F; G) \leq \frac{1}{4} + \frac{1}{12} + \frac{1}{2} \leq \frac{1}{6} \text{ess inf}_{x \in E} f(x)g(x),$$

essent U uniforme en $(0; 1)$:

En la segona part d'aquest capítol donem desigualtats per a la variància, per qualsevol distribució, basades en les desigualtats de Cramér-Rao i de Cauchy-Schwarz. Així en el cas de la distribució logística de mitjana 0 i variància $\frac{\pi^2}{12}$ tenim

$$\frac{3}{4} \frac{1}{\pi^2} [E(G(X))]^2 \leq \text{var}(G(X)) \leq \frac{1}{2} E(G(X))^2.$$

Finalment donem una prova alternativa d'aquesta desigualtat.

El capítol setè i darrer l'hem dedicat a la recerca de desigualtats utilitzant la propietat d'acotació de la raó de versemblança. En primer lloc hem donat una demostració del teorema de les mitjanes aritmètica i geomètrica,

$$a_1^{b_1} \cdots a_k^{b_k} \leq \frac{a_1 b_1 + \cdots + a_k b_k}{b_1 + \cdots + b_k}^{b_1 + \cdots + b_k},$$

on $a_1, \dots, a_k; b_1, \dots, b_k$ són nombres reals o enters positius, segons els casos.

Hem obtingut també expressions semblants com

$$\frac{a_1^{a_1 + a_2 + \cdots + a_k}}{b_1 + b_2 + \cdots + b_k} \leq \frac{a_1}{b_1} \cdots \frac{a_k}{b_k},$$

$$a_1^{a_1} \cdots a_k^{a_k} \leq \frac{a_1 + \cdots + a_k}{k}^{a_1 + \cdots + a_k},$$

essent $k > 1$:

La part final del capítol ha estat dedicada a trobar desigualtats semblants utilitzant distribucions diverses, Bernoulli, binomial, Pascal, geomètrica, de Cauchy i de Laplace.

8.2 Conclusions

Podríem resumir els principals resultats i conclusions d'aquesta memòria en els següents punts:

- ² La desigualtat de Hamilton-Routledge-Cuadras, que ens diu que no sempre variables correlacionades contenen informació redundant i ens pot avisar d'alguns dels perills de la utilització, a la lleugera, de la regressió múltiple, es verifica si les correlacions simples entre la variable dependent i les components principals amb variància petita siguin grans, mentre que les correlacions simples entre la variable dependent i les components principals amb variància gran siguin petites. De manera més precisa:

$$\sum_{\lambda_i < 1} r_{yZ_i}^2 (1 - \lambda_i) > \sum_{\lambda_i > 1} r_{yZ_i}^2 (\lambda_i - 1):$$

El seu acompliment és una condició necessària per a que la p-èsima component principal sigui l'única component que redueixi la suma de quadrats residual.

- ² Essent Y_1 i Y_2 dues variables resposta i X_1, \dots, X_p variables explicatives és possible tenir, d'una banda, $\sum_{j=1}^p r_{y_1 x_j}^2 > \sum_{j=1}^p r_{y_2 x_j}^2$, i, d'altra banda, $R_1^2 < R_2^2$; fet que és degut a la influència sobre Y_i de les components principals amb variància petita i que ens avisa que el mètode de la regressió en components principals pot no ser adient.
- ² La desigualtat entre distàncies, Mahalanobis més gran que K. Pearson, versió multivariant de la paradoxa de Simpson i que expressa, com la desigualtat de Hamilton-Routledge-Cuadras, el fet que no sempre variables correlacionades contenen informació redundant, s'acomplirà quan la direcció geomètrica que segueixen les mitjanes de dues poblacions coincideix pràcticament amb la direcció de les components principals de variància petita. El seu acompliment indica una anormal connexió entre model i dades.
- ² La paradoxa de Rao coneguda també com la "maledicció de la dimensionalitat" no es pot donar si es verifica la desigualtat distància de Mahalanobis més gran que distància de K. Pearson.
- ² La densitat construïda a partir de la dissimilaritat conjunta considerant la redundància de les variables, és més propera a la densitat vertadera que la densitat construïda a partir de la distància definida suposant la independència estocàstica de les variables.

- ² Hem obtingut algunes desigualtats per a la distància de Wasserstein en funció de la variància, la mitjana i la densitat. També és possible relacionar la variància d'una funció amb l'esperança de la seva derivada.
- ² Finalment, la raó de versemblança aplicada a diversos models estadístics ens permet provar algunes desigualtats clàssiques, com la que relaciona les mitjanes aritmètica, geomètrica i harmònica i altres menys conegudes.

Bibliogra...a

- [1] E. Anderson. The irises of the gaspé peninsula. *Bulletin of the American Iris Society*, 59:2–5, 1939.
- [2] D. R. Appleton, J. M. French, and M. P. J. Vanderpump. Ignoring a covariate: A example of Simpson’s paradox. *The American Statistician*, 50:340–341, 1996.
- [3] C. Arenas, C. M. Cuadras, and J. Fortiana. *Multicua*. Departament d’Estadística, UB, 1993.
- [4] J. P. Benzecri. *L’Analyse des données*. Dunod, 1973.
- [5] P. V. Bertrand and R. L. Holder. A quirk in multiple regression: The whole regression can be greater than the sum of its parts. *The Statistician*, 37:371–374, 1988.
- [6] P. V. Bertrand, B. T. Rudd, P. H. Weller, and A. J. Day. Free cortisol and creatinine in urine of healthy children. *Clinical Chemistry*, 33:2047–2051, 1987.
- [7] C. R. Blyth. On Simpson’s paradox and the sure-thing principle. *Journal of the American Statistical Association*, 67:364–366, 1972.
- [8] T. Cacoullos. On upper and lower bounds for the variance of a function of a random variable. *The Annals of Probability*, 10:799–809, 1982.
- [9] H. Chernoff. A note on an inequality involving the normal distribution. *The Annals of Probability*, 9:533–535, 1981.
- [10] C. M. Cuadras. Interpreting an inequality in multiple regression. *The American Statistician*, 47(4):256–258, 1993.

- [11] C. M. Cuadras. Discussion on "The relation between theory and applications in statistics" by D.R.Cox. *TEST*, 4(2):207–261, 1995.
- [12] C. M. Cuadras. Increasing the correlations with the response variable may not increase the coefficient of determination: A PCA interpretation. *New Trends in Probability and Statistics*, 3:75–83, 1995.
- [13] C. M. Cuadras. A distribution with given marginals and given regression curve. In: L. Ruschendorf, B. Schweizer and M. D. Taylor (eds.). *Distributions with Fixed Marginals and Related Topics*, pp. 76-83. 1996.
- [14] C. M. Cuadras. *Métodos de Anàlisis Multivariante*. EUB, Barcelona, 1996.
- [15] C. M. Cuadras. Some cautionary notes on the use of principal components regression (Revisited). *The American Statistician*, 52(4):371, 1998.
- [16] C. M. Cuadras. Multidimensional dependencies in classification and ordination. In: *Analyses Multidimensionnelles des Données*. K. Fernández, A. Morineau, eds. CISIA-CERESTA, Saint-Mandé (France), 1998.
- [17] C. M. Cuadras. Discussion on "Contributions of empirical and quantile processes to the asymptotic theory of goodness-of-fit tests" by E. del Barrio, J. A. Cuesta-Albertos and C. Matrán. *Test*, 9(1):1–96, 2000.
- [18] C. M. Cuadras. On the covariance between functions. *Journal of Multivariate Analysis*, 81:19–27, 2002.
- [19] C. M. Cuadras, R. Atkinson, and J. Fortiana. Probability densities from distances and discrimination. *Statistics & Probability Letters*, 33:405–411, 1997.
- [20] C. M. Cuadras and D. Cuadras. Orthogonal expansions and distinction between logistic and normal. In: *Goodness-of-fit Tests and Validity Models*. C. Huber-Carol, N. Balakrishnan, M.S. Nikulin, M. Mesbah, eds., pp. 325-338, Birkhauser, Boston, 2001.

- [21] C. M. Cuadras, J. M. Duran, and J. Fortiana. On an inequality in multivariate analysis. XXIV Congreso Nacional de Estadística e Investigación Operativa, 1998.
- [22] C. M. Cuadras and J. Fortiana. A continuous metric scaling solution for a random variable. *Journal of Multivariate Analysis*, 52:1–14, 1995.
- [23] C. M. Cuadras and J. Fortiana. Continuous scaling in bivariate copula. in: *Distribution with given marginals and moment problems*. 1997.
- [24] C. M. Cuadras and J. Fortiana. Geometrical aspects in multivariate statistical analysis with applications. 1998.
- [25] C. M. Cuadras, J. Fortiana, and F. Oliva. The proximity of an individual to a population with applications in discriminant analysis. *J. of Classification*, 14:117–136, 1997.
- [26] C. M. Cuadras and Y. Lahlou. Some orthogonal expansions for the logistic distribution. *Comm. Stat.-Theor. Meth.*, 29(12):2643–2663, 2000.
- [27] C. M. Cuadras and Y. Lahlou. Principal components of the pareto distribution. In: *Distributions with given marginals and statistical modelling* (C.M. Cuadras, J. Fortiana and J.A. Rodríguez-Lallena, Eds.), pp. 43-50, Kluwer Ac. Pub, Dordrecht, 2002.
- [28] J. Cuesta-Albertos, C. Matrán-Bea, and A. Tuero-Diaz. On lower bounds for the l^2 -wasserstein metric in a Hilbert space. *J. of Theoretical Probability*, 9:263–283, 1996.
- [29] E. del Barrio, J. Cuesta-Albertos, and C. Matrán-Bea. Contributions of empirical and quantile processes to the asymptotic theory of goodness-of-fit tests. *Test*, 9, 2000.
- [30] R. Ellis. *Entropy, Large Deviations, and Statistical Mechanics*. Springer-Verlag, 1985.
- [31] R. A. Fisher. The use of multiple measurements in taxonomic problems. *Ann. Eugen.*, 7:179–188, 1936.

- [32] R. A. Fisher. The precision of discriminant functions. *Ann. Eugen.*, 10:422–429, 1940.
- [33] B. Flury. *A ...rst course in Multivariate Statistics*. Springer, 1997.
- [34] F. Galton. Regression toward mediocrity in heredity stature. *Journal of Anthropological Institute*, 15:246–263, 1885.
- [35] A. Grané. Bondad de ajuste basada en la descomposición de la correlación máxima. Tesis Doctoral. Fac. de Matemáticas, Universidad de Barcelona, 1999.
- [36] M. J. Greenacre. *Theory and Applications of Correspondence Analysis*. Academic Press, Inc., London, 1984.
- [37] L. Guttman. The quanti...cation of a class of attributes: A theory and method of scale construction. In: P. Horst (ed.). *The prediction of personal adjustment*, New York., 1941.
- [38] A. Hadi and R. Ling. Some cautionary notes on the use of principal components regression. *The American Statistician*, 52:15–19, 1998.
- [39] D. Hamilton. Sometimes $r^2 > r_{yx_1}^2 + r_{yx_2}^2$. correlated variables are not always redundant. *The American Statistician*, 41:129–132, 1987.
- [40] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge University Press, London, 1952.
- [41] M. Healy. Rao's paradox concerning multivariate tests of signi...cance. *Biometrics*, 25:411–413, 1969.
- [42] H. O. Hirschfeld. A connection between correlation and contingency. *Proceedings of the Cambridge Philosophical Society*, 31:520–524, 1935.
- [43] J. N. Jeffers. Two case studies in the application of principal component analysis. *Applied statistics*, 16:225–236, 1967.
- [44] R. A. Johnson and D. W. Wichern. *Applied Multivariate Statistical Analysis*. Prentice-Hall, Inc., 1992.

- [45] I. Jolliffe. A note on the use of principal components in regression. *Applied Statistics*, 31:300–302, 1982.
- [46] I. T. Jolliffe. *Principal Component Analysis*. New York: Springer-Verlag, 1986.
- [47] M. G. Kendall and A. Stuart. *The advanced Theory of Statistics*, volume 2. New York: Hafner Publishing, 1973.
- [48] C. A. J. Klaassen. On an inequality of Chernoff. *The Annals of Probability*, 13:966–974, 1985.
- [49] B. F. J. Manly. *Multivariate statistical methods: a primer*. 1986.
- [50] J. Mardia. *Families of bivariate distributions*. London: Charles Griffin, 1970.
- [51] J. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis*. London: Academic Press, 1979.
- [52] D. Montgomery and E. Peck. *Introduction to Linear Regression Analysis*. John Wiley and Sons, INC., 1992.
- [53] D. Morrison. *Multivariate Statistical Methods*. Mc Graw Hill, New York, 2 edition, 1976.
- [54] L. Ott. *An Introduction to Statistical Methods and Data Analysis*. Boston: Duxbury Press, 1984.
- [55] C. R. Rao. *Advanced statistical methods in biometric research*. Wiley, New York, 1952.
- [56] C. R. Rao. *Linear Statistical Inference and its Applications*. Wiley Eastern Limited, 1984.
- [57] C. R. Rao. *Estadística y verdad. Aprovechando el azar*. PPU, 1994.
- [58] R. D. Routledge. When stepwise regression fails: Correlated variables some of which are redundant. *International Journal of Mathematical Education in Science and Technology*, 21:403–410, 1990.

- [59] G. Shieh. The inequality between the coefficient of determination and the sum of squared simple correlations coefficients. *The American Statistician*, 55(2):121–124, 2001.
- [60] E. H. Simpson. The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society*, 13:238–241, 1951.
- [61] L. Stefanski. A note on the arithmetic-geometric-harmonic mean inequalities. *The American Statistician*, 50(3):246–247, 1996.
- [62] E. Tiit. "Formal computation of regression parameters". *Proceedings Sixth Symposium COMSTAT 1984*. Physica-Verlag, Vienna, 1984.
- [63] E. Tiit. Discussion on the paper of C. M. Cuadras "Increasing the correlations with the response variable may not increase the coefficient of determination: A PCA interpretation". *New Trends in Probability and Statistics*, 3:85–92, 1995.