

Universitat de Barcelona - Departament d'Estadística

Test d'hipòtesis com a selecció de models.
Una aproximació geomètrica.

Marta Cubedo Culleré
Barcelona, Juny de 2002

Programa de Doctorat de Probabilitats i Estadística. Bieni 1992-1994.

Test d'hipòtesis com a selecció de models.
Una aproximació geomètrica.

Memòria presentada per optar al
grau de Doctora en Matemàtiques
per la Universitat de Barcelona,
per

Marta Cubedo Culleré

Vist i plau
El Director,

Dr. Josep Maria Oller i Sala,
Catedràtic d'Universitat.
Departament d'Estadística.
Universitat de Barcelona.

Barcelona, 28 de Juny de 2002

Als meus pares,

La diferència entre el que fem i el que seríem capaços de fer,
seria suficient per resoldre la major part dels problemes del món.

Gandhi.

Agraïments

Primerament voldria donar les gràcies al Josep Maria Oller, director d'aquesta memòria, per tot el temps que m'ha dedicat, per la seva comprensió i amabilitat, per tot el que he pogut aprendre. Ell ha estat qui m'ha ofert la possibilitat de treballar amb aquesta aproximació geomètrica de l'Estadística tant nova i atractiva per a mí. Recordo amb simpatia els seus plantejaments filosòfics, i són del tot admirables el seu "anar més enllà", la seva actitud perseverant davant els problemes i la seva vessant humana.

També voldria donar les gràcies a tots els meus companys i amics del Departament, especialment al Martín, pels seus consells, perquè sempre té una paraula amable i d'ànim, la qual cosa és necessària en molts moments. A l'Àngel, l'Antonio, el Miquel i el Josep Anton sempre disposats a treballar i donar un cop de mà. Al Jaume i el Sergi els meus primers companys de despatx i Amics "amb majúscules". A la Marian, per les bones estones dintre i fora del Departament, per aquelles apassionants converses de "biberons" i "bolquers" entre classe i classe, i a l'Aurea, per suportar-ho amablement. Als meus companys de promoció: Tere, Ana i Antonio, per compartir tots els bons, i no tant bons, moments passats durant la carrera i encara ara.

Finalment vull donar les gràcies d'una manera molt especial a la meva família. Als meus pares, ells han fet possible que arribés on estic amb el seu esforç, la seva estima i el seu suport incondicional en tot moment. Al meu germà Joan, per confiar tant en mí, per ser el meu millor alumne i per tot el que jo he après amb ell. Al Francesc, per animar-me, fer-me riure i estar amb mí quan més ho necessito. Al Pol, la meva il·lusió, el meu motor, perquè el seu somriure em recorda cada dia les coses importants de la vida.
Gràcies a tots!

Índex

| | | |
|----------|--|-----------|
| 1 | Introducció | 7 |
| 1.1 | Evolució Històrica | 8 |
| 1.2 | Dificultats conceptuals dels tests | 12 |
| 1.3 | Estructura geomètrica dels models estadístics | 18 |
| 1.4 | Algunes noves propietats de la distància de Rao | 21 |
| 2 | Alternatives al contrast d'hipòtesis clàssic | 33 |
| 2.1 | Els Mètodes | 34 |
| 2.1.1 | Regles \mathbf{p} i \mathbf{m} | 34 |
| 2.1.2 | Procediments \mathbf{f} i \mathbf{b} | 36 |
| 2.2 | Relació entre hipòtesis simples i compostes | 38 |
| 2.3 | Distribucions asimptòtiques | 42 |
| 3 | Model Normal | 59 |
| 3.1 | Model normal n -variant amb Σ coneguda | 59 |
| 3.1.1 | mètode – \mathbf{pf} | 60 |
| 3.1.2 | mètode – \mathbf{mf} | 62 |
| 3.1.3 | mètode – \mathbf{pb} | 63 |
| 3.1.4 | mètode – \mathbf{mb} | 64 |
| 3.2 | Comparació dels mètodes | 65 |
| 4 | Una aproximació asimptòtica del mètode–\mathbf{pf} | 69 |
| 4.1 | Cas general amb hipòtesi nul·la simple | 69 |
| 5 | Alguns models uniparamètrics | 75 |
| 5.1 | Model Poisson | 75 |
| 5.2 | Model Gamma | 79 |
| 6 | Model lineal normal | 83 |
| 6.1 | Cas general | 83 |
| 6.2 | Exemple | 90 |

| | | |
|-----------|--|------------|
| 6.2.1 | Cas I | 90 |
| 6.2.2 | Cas II | 98 |
| 6.2.3 | Cas III | 101 |
| 7 | Altres situacions | 105 |
| 7.1 | Quan $\dim \Theta_0 = \dim \Theta$ | 105 |
| 7.1.1 | mètode- pf | 106 |
| 7.1.2 | mètode- mf | 107 |
| 7.1.3 | mètode- pb | 108 |
| 7.1.4 | mètode- mb | 109 |
| 7.1.5 | Un cas particular quan la hipòtesi nul·la no és simple . . . | 109 |
| 8 | Resum de resultats i conclusions | 113 |
| 8.1 | Resum de resultats | 113 |
| 8.2 | Perspectives de la metodologia proposada | 115 |
| 9 | Apèndix | 117 |
| 9.1 | Varietats i funcions C^∞ | 117 |
| 9.2 | Vectors i Espais Tangents | 119 |
| 9.3 | Subvarietats | 122 |
| 9.4 | Corbes | 123 |
| 9.5 | Camps vectorials | 124 |
| 9.6 | Varietats Riemannianes | 125 |
| 9.7 | Varietats amb connexió | 126 |
| 9.8 | Geodèsiques i aplicació exponencial | 127 |
| 9.9 | Longitud d'arc i distància Riemanniana | 129 |
| 10 | Llistat de programes | 131 |

*It is said: "All knowledge is, in the final analysis, history.
All sciences are, in the abstract, mathematics."
I venture to add: "All methods of acquiring knowledge
are essentially through statistics."*

C.R. Rao, *Statistics and Truth*.

Pròleg

L'objectiu general d'aquest treball és posar de relleu alguns problemes relacionats amb la teoria dels tests d'hipòtesis i aportar nous elements que donin uns criteris per a la selecció de models estadístics. Per això usarem principalment, com a mitjà, tècniques geomètriques en el context de l'estadística, i així proposarem noves metodologies, alternatives o complementàries a les tècniques clàssiques, basades en les propietats geomètriques dels models estadístics. En tot aquest treball es tractarà, de forma unificada, el problema de l'especificació del model i els tests d'hipòtesis. Per aclarir una mica més aquesta idea podem dir que es tractaran els tests d'hipòtesis des de la perspectiva de la selecció de models, en el sentit que acceptar una hipòtesi consistirà, essencialment, en acceptar que un model descriurà els trets més rellevants de una part de la realitat *millor*, segons uns criteris prèviament explicitats, que un altre més complex. Per tant els tests d'hipòtesis i la selecció de models es tractaran a través d'unes mateixes eines metodològiques.

De fet, dintre del context dels tests d'hipòtesis, es donen una sèrie de paradoxes que moltes vegades passen desapercebudes, o bé són acceptades, pels estadístics, i no estadístics, que els utilitzen, però que deixen entreveure unes certes mancances a nivell conceptual. Recordem, que en la formulació clàssica dels tests d'hipòtesis, moltes vegades s'accepta una hipòtesi la qual sabem, estrictament parlant, que és falsa. Però prendre aquests tipus de decisions pot ser raonable per diverses raons: doncs un model més senzill ens pot permetre sintetitzar o simplificar una realitat complexa, facilitant, a més, la predicció de nous resultats amb un cert grau de fiabilitat. El que no sembla doncs massa raonable és que aquestes decisions entre hipòtesis es prenguin en base a les seves probabilitats, o probabilitats condicionades, d'error, donat que sabem, en sentit estricte, que són falses.

Des d'aquest punt de vista, i tal com es desenvolupa en la secció 2 del capítol 1, la qüestió bàsica de la inferència estadística no hauria de ser: *Quina és la probabilitat que una afirmació donada sigui certa?*, doncs sabem que cap afirmació és estrictament certa, en el sentit que els models utilitzats en aquest context són només aproximacions de la realitat. Més aviat, la qüestió bàsica hauria de ser: *quant allunyada està una determinada afirmació de la realitat?*, suposant, és clar,

que s'ha definit una distància per mesurar-ho. Dit en altres paraules, una estructura formal per a determinats tipus de raonaments inductius s'hauria de fonamentar més en una mesura de proximitat a la *realitat*, que no pas en una probabilitat. De fet, la probabilitat és una molt bona eina per quantificar la propensió a ocórrer dels esdeveniments observables, doncs parlem en termes de si un esdeveniment ocorre o no, i no pas en termes de si aquest esdeveniment és *vertader* o *fals*. En aquest sentit la probabilitat és una eina bàsica per construir models teòrics de fenòmens observables, però en inferència estadística, segons el que hem comentat anteriorment, sembla més adient utilitzar una distància.

En aquest treball s'ha intentat donar una metodologia alternativa basada en distàncies sobre els models estadístics paramètrics. Com és ben sabut, aquests models, sota certes condicions generals de regularitat, tenen estructura de varietat Riemanniana, de manera que apareix com a distància natural la distància Riemanniana, la qual podem utilitzar com a eina per quantificar l'error estadístic. Cal remarcar que es podria haver utilitzar una altra distància i que aquest no és el tret diferencial d'aquest treball, sinó que, el que es pretén és donar un nou plantejament metodològic dels tests d'hipòtesis, o si més no, obrir un camí alternatiu que pugui evitar algunes de les paradoxes que apareixen en la metodologia clàssica.

En el capítol 2 es proposen quatre mètodes, tot i que, de fet, podríem parlar d'un únic mètode en el sentit que tots ells comparteixen una mateixa filosofia de fons que s'implementa, a efectes pràctics, segons les diferents preferències a l'hora de treballar en estadística: freqüentista o Bayesiana, utilitzant probabilitats o bé moments. Cal tenir en compte també que tots ells es basaran en l'acceptació d'un estimador convenient pel model general, com per exemple l'estimador màxim-versemblant, del vertader valor del paràmetre. Així doncs, i seguint la línia que hem comentat a l'inici d'aquest paràgraf, els quatre mètodes sorgeixen de les combinacions de dues regles i dos procediments. Les regles es defineixen per comparar els errors fets en realitzar les estimacions sota l'espai general de paràmetres i sota el subespai definit per la hipòtesi nul·la, en una regla es comparen probabilitats i en l'altra moments, rebent els noms de regla-*p* i regla-*m*, respectivament. Els dos procediments es defineixen per obtenir una estimació en la mesura de l'error sota el subespai de paràmetres definit per la hipòtesi nul·la, utilitzant una aproximació freqüentista en un cas, i una Bayesiana en l'altre. Aquest dos procediments rebran el nom de procediment-*f* i procediment-*b* respectivament. En aquest segon capítol es tractarà també, la relació entre hipòtesis simples i compostes i les distribucions asimptòtiques d'algunes de les distàncies que apareixen en els mètodes anteriorment descrits.

Des del tercer capítol fins al sisé, es veuran diferents aplicacions dels mètodes a diferents models estadístics coneguts, com és el cas del model normal n -variant amb Σ coneguda, el model Poisson, el model Gamma i el model lineal normal. Notem però, que en el capítol 4 es donarà una aproximació asimptòtica per a un model estadístic paramètric regular n -dimensional general, això es farà utilitzant un dels mètodes abans proposats.

Finalment s'estudiarà el comportament dels mètodes proposats quan no es compleix una de les suposicions bàsiques plantejades en aquest nou desenvolupament metodològic: quan no podem suposar que la probabilitat *a priori* corresponent a la hipòtesi nul·la, o submodel, és zero. Veurem com en aquest cas encara es donen resultats raonables, la qual cosa mostra la robustesa d'aquests nous mètodes.

Capítol 1

Introducció

En molts casos, quan els científics utilitzen un model, són conscients que aquest és només una aproximació de la realitat. Però les raons per utilitzar-lo són diverses: ens permet sintetitzar la variabilitat de les dades mantenint alhora un ajust raonable amb les observacions i facilitant, a més, la predicció de nous resultats amb certa precisió. En un sentit ampli, la modelització fa la realitat més entenedora.

Entre els models possibles es procurarà triar el que millor s'ajusti a la nostra realitat, dintre d'uns nivells acceptables de complexitat. S'estudiarà la selecció d'un model, quan un està inclòs dintre d'un altre, tenint en compte les analogies d'aquest procés de decisió i elecció del model amb la teoria dels tests d'hipòtesis. En aquests models, les observacions apareixen freqüentment com els valors que prenen unes determinades variables aleatòries i les hipòtesis són afirmacions referents a la distribució d'aquestes variables. En aquest context, s'analitzarà també la problemàtica donada quan la selecció d'un model es basa exclusivament en el criteri de la màxima versemblança, doncs aquest criteri ens condueix quasi sempre al model més complex.

Es veurà com l'aplicació dels mètodes clàssics, en algunes situacions, dona lloc a certes paradoxes i com aquestes paradoxes es poden evitar, o almenys es pot obrir un nou camí, utilitzant metodologies alternatives. Aquests nous mètodes es desenvoluparan en alguns casos concrets veient també, en altres, el paral·lelisme amb la teoria clàssica.

1.1 Evolució històrica de la teoria dels tests d'hipòtesis i l'elecció dels models estadístics

Començarem donant el significat d'alguns conceptes bàsics utilitzats dintre del context de la teoria del test d'hipòtesis i l'elecció dels models estadístics.

Explicarem primer el que s'entén per *hipòtesis estadístiques*. Una hipòtesi, quan ve expressada en termes d'una classe de distribucions de probabilitat, esdevé una afirmació que imposa restriccions addicionals sobre la classe de distribucions especificades pels models. Si la hipòtesi especifica completament la distribució, és a dir conté una única distribució de probabilitat, aleshores s'anomena hipòtesi simple; en cas contrari s'anomenarà composta. Donat, a més, que una hipòtesi sota test representa freqüentment l'absència d'un efecte, ens referim a ella com *hipòtesi nul·la*.

La formulació d'hipòtesis i el seu contrast a través de l'observació són passos essencials en el procés científic. Una discussió detallada del seu paper en el desenvolupament de les teories científiques ve donat per Popper [40] l'any 1935.

Tests d'hipòtesis. Un test d'una hipòtesi H és una regla de decisió que especifica, per a cada conjunt possible de valors de les observacions, si acceptar o rebutjar H , havent estat aquests valors particulars observats. Hi ha, per tant, una divisió de tots els possibles conjunts de valors (l'anomenat espai mostral) en dos grups: aquells pels quals la hipòtesi H serà acceptada, (*regió d'acceptació*), i aquells pels quals serà rebutjada, (*regió de rebuig o regió crítica*). En els anomenats *Tests Aleatoritzats* s'introdueix un mecanisme aleatori, dependent de les observacions, per decidir si acceptar o rebutjar H , entenent-se com una generalització del cas anterior.

Els tests estan definits típicament en termes d'un estadístic de test T , encarregat de resumir tota la informació donada per les dades observades i els valors extrems del qual són clarament improbables d'ocórrer si H és certa, però no són sorprenents si H és falsa. Sembla aleshores *natural* rebutjar H quan T prengui valors suficientment grans o suficientment petits. Anomenarem doncs W a la regió crítica o de rebuig.

L'elecció d'un test apropiat no depèn només de la hipòtesi H que es vol acceptar o rebutjar, la qual s'anomenarà *hipòtesi nul·la* i es notarà H_0 , sinó també de les maneres en que aquesta hipòtesi pot ser falsa, això és, de les distribucions de

probabilitat que no satisfan H_0 anomenades *alternatives o hipòtesis alternatives* i que es notaran H_1 .

Per especificar un test del tipus anterior completament, és encara necessari seleccionar un o uns determinats valors crítics. Aquesta selecció es fa normalment sobre les bases de les següents consideracions: els valors pels quals la hipòtesi nul·la serà rebutjada podrien ocórrer fins i tot si aquesta hipòtesi fos certa; això hauria de ser, no obstant, molt improbable. Una mesura de com són de sorprenents aquests valors sota H_0 és la probabilitat d'observar-los quan H_0 és certa. Aquesta probabilitat, $P_{H_0}(W)$, s'anomena *nivell de significació o probabilitat d'error de primera espècie* del test, que es nota també freqüentment per α i el qual voldrem que prengui valors petits tals com: 0.01 ó 0.05. El valor o valors crítics es determinen doncs a partir de l'equació:

$$P_{H_0}(W) \leq \alpha$$

Un altre requisit que és convenient en un test és mantenir un nivell de significació baix però controlant que la probabilitat de rebutjar H_0 , quan sigui falsa, sigui alta. Aquesta probabilitat, $P_{H_1}(W)$, s'anomena *potència del test*, on H_1 representa la hipòtesi alternativa. Al valor $P_{H_1}(W^C)$ se l'anomena *probabilitat d'error de segona espècie*. Per a més detalls es pot veure Lehmann [29] i Neyman i Pearson [33], entre altres.

Per començar amb l'evolució històrica de la teoria dels tests d'hipòtesis, direm que uns dels primers tests en aparèixer van ser els anomenats *tests de significació*, que s'usen en situacions on el fenomen que s'està estudiant té una component d'atzar, la qual pot aparèixer de manera natural o ser introduïda artificialment mitjançant simulació. Aquests tests proporcionen les fronteres que decideixen en quin grau els resultats inusuals poden ser atribuïts a l'atzar. Per obtenir aquestes fronteres s'usa un model que descriu els efectes de l'atzar. Si aquest és l'única font dels resultats observats, aleshores el model mateix ha de ser suficient per explicar dits resultats. Els tests de significació decideixen, doncs, la suficiència o no del model en aquest sentit. La hipòtesi que ens diu que el model és *suficient* s'anomena normalment *hipòtesi nul·la*, mentre que no es parla explícitament d'hipòtesi alternativa.

Per dur a terme aquests tests de significació, són necessaris uns objectes adequats capaços de resumir les dades observades, els estadístics de test. Una vegada avaluat aquest estadístic a partir de les observacions, es calcula la probabilitat, sota la hipòtesi nul·la, que un resultat sigui *pitjor* que l'observat, és a dir, el *p-valor* o probabilitat de la cua corresponent. Quan més petita sigui aquesta probabilitat, més forta serà l'evidència contra la hipòtesi nul·la.

La primera publicació on s'usa un raonament semblant a un test de significació s'atribueix a John Arbuthnot (1710). En un estudi sobre el sexe dels nadons en Londres durant 82 anys, conclou que l'atzar no és suficient per explicar que durant aquests anys el nombre de nadons de sexe masculí hagi dominat al femení. Però tot i que és un exemple clar del problema de trobar una distribució que s'ajusti als efectes produïts per l'atzar, Arbuthnot va arribar a aquesta conclusió sobre les bases d'una simple generalització inductiva dels 82 anys observats. No va utilitzar l'obtenció d'un nivell de significació petit en el seu raonament.

En el 1812, Laplace [28], amb la seva *Théorie Analytique des Probabilités*, va presentar un tipus de raonament que podria ser anomenat test de significació, usant com a distribució de l'estadístic de test la distribució normal.

L'any 1900, Karl Pearson [39] tracta el problema de mesurar el bon ajust entre les dades observades i altres distribucions diferents a la normal. Va definir una funció estadística anomenada χ^2 a partir de variables normals i incorrelacionades, de manera que el valor d'aquesta funció χ^2 augmentés a mesura que les variables es desviaven dels valors esperats. Cal notar que el test de Pearson és un test per a mostres grans, ja que necessita un gran nombre d'observacions per usar les propietats de normalitat asimptòtica.

Sota el nom de Student [47], W. S. Gosset proposa l'any 1908 la primera solució a un cert tipus de tests per a mostres petites. Modelant l'atzar a partir de la distribució normal, Student va aconseguir definir una distribució de probabilitat exacta quan la desviació estàndard havia de ser aproximada per la desviació estàndard mostral. Com a resultat es va obtenir la distribució t de Student.

A principis de l'any 1920 R. A. Fisher va estendre els resultats de Student afegint a la seva teoria l'estimació dels paràmetres en la construcció del test. Usant les dades per estimar els valors dels paràmetres va aconseguir reduir els graus de llibertat. Fisher [15] va demostrar que en molts casos es pot trobar una distribució, generalment asimptòtica, per l'estadístic de test sota la hipòtesi nul·la, estimant els paràmetres lliures a partir del mètode de la màxima-versemblança. Però Fisher mai va fer una teoria de construcció del test de significació, malgrat que ell va inventar o desenvolupar la majoria dels tests més usats i fins i tot alguns tests no paramètrics bàsics. Fisher no estava d'acord amb l'ús de la probabilitat inversa, usada a l'hora de determinar els valors crítics per a les diferents distribucions dels estadístics de test, i de fet no li atreïa la teoria de la decisió. Però, tot i que no estava d'acord amb l'ús de la probabilitat inversa, entesa en un sentit Bayesià, sentia la necessitat d'usar alguna eina semblant donat que va intentar desenvolupar

un substitut: la probabilitat fiducial.

Dels tests de significació, només preocupats per la hipòtesi nul·la es va passar a la teoria del contrast d'hipòtesis o tests d'hipòtesis que es coneix actualment, en els quals s'especifica completament el tipus d'hipòtesi alternativa que convé en cada cas, donant lloc a les diferents classes de tests.

L'any 1928, J. Neyman i E. S. Pearson van començar a desenvolupar una metodologia del contrast d'hipòtesis definint els errors de primera i segona espècie i investigant un principi intuïtiu per a la construcció de tests que els va portar al test de la raó de versemblança. El principi de la raó de versemblança va ser extraordinàriament fructífer, d'ell se'n derivaven, de forma exacta o aproximada, molts dels altres tests coneguts, incloent-hi el de la χ^2 de Pearson.

Encara quedava però, una important qüestió per resoldre: quin és el millor test? Neyman i Pearson havien formulat implícitament, en el 1928, el seu criteri: fixats els percentatges d'errors de primera espècie, minimitzar els errors de segona espècie. Cap a l'any 1933, s'havien establert totes les definicions claus per permetre una presentació ordenada que tenia com a eix central el lema de Neyman-Pearson, el qual desenvolupa uns criteris per a contrastar una hipòtesi simple front una alternativa també simple, estenent-se en alguns casos a simple front alternativa composta. Tot el que es refereix al contrast d'hipòtesis ha estat desenvolupat sobre aquestes bases des de llavors. Per a més detalls veure Neyman i Pearson [33], Neyman [32], Pearson [37] i [38].

Un altre problema ha aparegut en la història recent dels tests d'hipòtesis. Els resultats d'un test són *rebutjar* la hipòtesi nul·la o *no rebutjar* la hipòtesi nul·la. Però la pregunta que queda és: quines hipòtesis poden ser legítimament *acceptades* o provisionalment agafades com a resultat d'un test? Suposem que un test doni com a resultat rebutjar la hipòtesi nul·la, justifica això la acceptació d'una altra hipòtesi sobre l'existència d'una font sistemàtica que dona lloc a les observacions? Suposem, al contrari, que un test doni un resultat fortament no significatiu, justifica això l'acceptació de la hipòtesi nul·la? Hi ha fortes raons per pensar o creure que no.

Un test no significatiu ha d'assenyalar que la hipòtesi nul·la podria ser una font adequada per explicar les dades, però sense suposicions addicionals, aquest resultat no ha de provar necessàriament que la hipòtesi nul·la sigui l'única explicació possible per obtenir aquestes observacions. La confusió que envolta aquest resultat pot ser eliminada d'alguna manera si es realitza un canvi subtil en la interpre-

tació dels tests. Per exemple, el test χ^2 de Pearson anomenat anteriorment, usa nivells de significació com a mesura d'ajust entre hipòtesis i dades. Resultats no significatius indicarien un bon ajust, i no la certesa, de la hipòtesi nul·la. De fet en aquesta línia anirà dirigit tot aquest treball. Com a bibliografia addicional sobre l'història de contrast d'hipòtesis podem destacar Fisher, R.A. [17], Kendall, M.G. i Plackett, R.L. [27].

1.2 Algunes dificultats conceptuals dels tests d'hipòtesis

Qualsevol argument científic es refereix a una part de la realitat i pretén establir afirmacions que hi estiguin d'acord. Però totes les afirmacions es fan amb un llenguatge de referència, el qual és intrínsecament limitat per ajustar-se fidelment a la realitat. En paraules de L. Wittgenstein [48], *Els límits del meu llenguatge signifiquen els límits del meu món*, com pot veure's en el seu *Tractatus logico-philosophicus*.

Tot i aquesta limitació, és convenient fer discussions sota un context general dins el qual aquestes afirmacions es consideraran certes, sinó no seria possible expressar cap argument, amb la pretensió de ser cert. Considerem, per exemple, la representació conceptual d'una direcció, a l'espai físic, mitjançant una recta real. És clar que quan més ric sigui el llenguatge, més possibilitat hi haurà que s'ajusti millor a la realitat, però moltes vegades, en l'estudi d'una part de realitat donada, és útil introduir algunes simplificacions conceptuals per representar algunes característiques bàsiques del problema d'una manera més entenedora, encara que aquesta simplificació suposi allunyar-se una mica de la realitat. Sota aquest punt de vista la qüestió bàsica en inferència estadística no seria *quina és la probabilitat que una determinada afirmació sigui certa?*, donat que, quasi segurament, no és estrictament certa, la qüestió bàsica és *quant allunyada està aquesta afirmació de la vertadera?*, suposant que tenim un procediment per quantificar aquesta llunyania, per exemple, una distància. En altres paraules, per estudiar formalment certs tipus de raonaments inductius ens hauríem de basar més en una mesura de *proximitat a la veritat* que en la *probabilitat d'error*. Observem que, sota aquesta aproximació, la estricta falsedat d'una afirmació feta en aquest llenguatge de referència no invalida aquestes consideracions, donat que pressuposem que aquest llenguatge és prou ric per *ajustar-se suficientment* a la realitat per a un propòsit donat.

En aquest treball considerarem problemes de test d'hipòtesis en el context dels models estadístics paramètrics quan la dimensió corresponent a la hipòtesi nul·la és estrictament més petita que la corresponent a la dimensió del model total. Intentarem resoldre aquests tipus de problemes a partir de mètodes que ens proporcionin un significat precís a les idees que hem comentat anteriorment, és a dir, donar raons clares de per què algunes vegades triem el model més simple (hipòtesi nul·la) encara que aquesta no sigui certa.

Dintre del context dels models estadístics paramètrics, un problema de contrast d'hipòtesis estarà definit per una part Θ_0 de l'espai de paràmetres Θ , $\Theta_0 \subset \Theta$, corresponent a la hipòtesi nul·la. Considerem un cas freqüent, que ja hem anomenat anteriorment: quan els graus de llibertat sota la hipòtesi nul·la són estrictament més petits que els corresponents a l'espai total de paràmetres. Geomètricament parlant, Θ serà una varietat n -dimensional, real, C^∞ i Θ_0 una subvarietat r -dimensional estrictament inclosa en Θ amb $0 \leq r < n$, això és: $H_0 : \theta \in \Theta_0$ front $H_1 : \theta \notin \Theta_0$, on θ representa el paràmetre corresponent a la vertadera funció de probabilitat. Generalment Θ és un conjunt obert de \mathbb{R}^n i en aquest cas s'acostuma a usar els mateixos símbols, θ , per notar punts i coordenades, és a dir, agafant com a sistema de coordenades la identitat. Per a més detalls, veure l'apèndix.

Amb l'objectiu de construir un test, una aproximació ingènua, si no féssim cas d'algunes de les anteriors observacions, podria ser plantejar el problema de la forma següent: donada una mostra de mida k , quina és la *probabilitat* (des d'una perspectiva freqüentista) que H_0 sigui certa? i decidir entre les hipòtesis tenint en compte aquesta probabilitat.

Per respondre a aquesta pregunta hauríem de conèixer la *distribució a priori* sobre Θ , anomenem-la Π , i seguir una aproximació Bayesiana estàndard. El primer problema seria l'elecció de la distribució a priori, suposant que volguéssim mantenir un significat freqüentista de la probabilitat a posteriori. Aquesta elecció dependria de cada anàlisi estadístic en particular, una vegada determinades *completament* les condicions experimentals sota les quals, la població estadística *actual*, ha estat seleccionada. Però, malauradament, aquesta informació no és normalment coneguda, malgrat la seva possible existència en qualsevol situació particular, doncs normalment, cada mostra s'ha generat partir d'unes condicions experimentals reals de les quals no en som conscients.

Tot i aquest problema, en la majoria dels casos és raonable suposar que la mesura a priori en la varietat Θ és tal que, donat qualsevol sistema de coordenades $\phi : U \subset \Theta \rightarrow \mathbb{R}^n$, la corresponent mesura induïda en \mathbb{R}^n (o en $\phi(U)$) és absolutament

contínua respecte la mesura de Lebesgue en \mathbb{R}^n , o equivalentment, respecte la mesura Riemanniana dV corresponent a la mètrica informativa. Per a més detalls veure Jeffreys [24]. Notem que la suposició anterior és compatible amb la *quasi* total absència de coneixement a priori sobre la vertadera mesura de probabilitat, *menys coneixement*, de fet, que la suposició de qualsevol a priori no informativa. Per aclarir una mica això, considerem els següents exemples:

- Si considerem Θ un obert de \mathbb{R}^n i agafem un sistema de coordenades donat per la identitat, en tot Θ , amb la hipòtesi anterior aconseguirem que la distribució de probabilitat *a priori* "tingui densitat" respecte la mesura de Lebesgue en \mathbb{R}^n .
- Si considerem Θ una varietat Riemanniana i com a mesura, dV , el volum Riemannià donat per la mètrica informativa, la distribució *a priori* serà absolutament contínua respecte dV . És a dir, si Π és la probabilitat a priori, existeix una funció positiva $h \in L^1(dV)$, anomenada densitat a priori, tal que $d\Pi = h dV$. Dit d'una altra manera, identificant els punts de la varietat, $\theta \in \Theta$ amb les seves coordenades $(\theta_1, \dots, \theta_n) \in \mathbb{R}^n$ tenim que $dV = \sqrt{g(\theta)} d\theta$, on $g(\theta) = \det(g_{ij}(\theta))$, essent $g_{ij}(\theta)$ els coeficients de la matriu d'informació de Fisher sota la parametrització $(\theta_1, \dots, \theta_n)$ i $d\theta$ la mesura de Lebesgue sobre \mathbb{R}^n , en l'apartat 1.3 es fa una descripció més detallada sobre la construcció de la mètrica informativa a partir de la matriu d'informació de Fisher i per a més detalls es pot veure Amari [3], Hicks [23] i Spivak [46]. Així finalment

$$d\Pi = h(\theta) \sqrt{g(\theta)} d\theta$$

Sota aquesta hipòtesi tenim clarament, (donat que Θ_0 és r -dimensional i estrictament inclosa en Θ), que la probabilitat *a priori* de la hipòtesi nul·la és zero, $\Pi(H_0) = 0$, i la probabilitat *a posteriori*, $P(H_0|x) = 0$ també és zero!, doncs Θ_0 és un conjunt de mesura Riemanniana zero.

És més, des d'un punt de vista purament aplicat, *l'experiència mostra* que en problemes de contrast d'hipòtesi d'aquest tipus en els quals, $\dim \Theta_0 < \dim \Theta$, una vegada fixat el nivell de significació del test, *la hipòtesi nul·la serà (quasi) sempre rebutjada* quan es treballi amb un *grandària mostral suficientment gran* i sempre que no canviïn les condicions experimentals sota les quals mostregem. Això no és sorprenent si es té en compte que aquests models en general, i en particular el submodel proposat corresponent a la hipòtesi nul·la, són només aproximacions de la realitat, i aleshores, estrictament parlant i des d'un punt de vista lògic, incorrectes o falsos. Aquestes consideracions han estat observades a la literatura per diversos autors, es pot veure per exemple Berger [7].

Així doncs, des d'aquesta perspectiva, quin és el significat filosòfic de l'acceptació de la hipòtesi nul·la, donat que sabem, sota febles suposicions, que és quasi segurament falsa?

Què volem dir quan acceptem hipòtesis tals com:

- El pes, d'una població donada, està normalment distribuït.
- Ambdues poblacions normals tenen la mateixa variància.
- Aquest dau és *perfecte*, això és: $p_1 = p_2 = \dots = p_6 = \frac{1}{6}$?

En tots aquests casos *sabem* que la hipòtesi nul·la és quasi segurament falsa, *independentment* de qualsevol mostra en particular.

De la discussió anterior se'n dedueix clarament que la *probabilitat d'error* és *no informativa* respecte el rebuig o acceptació de la hipòtesi nul·la. A més, quin és el sentit de fer el *nivell de significació baix* ($\alpha = 0.05$, $\alpha = 0.01$, o més petit) si aquest només representa la probabilitat *condicional* d'error *corresponent a un esdeveniment de probabilitat zero*?

Si només es vol provar que *la hipòtesi nul·la és aproximadament certa* és més necessària una mesura d'ajust entre hipòtesis i dades que una mesura de la *propensió* de que una hipòtesi sigui certa, més com s'utilitzava el test de bondat d'ajust χ^2 a començaments de segle. Una discussió interessant sobre hipòtesis científiques es pot trobar a Good [19]. Altres discussions i crítiques interessants sobre la interpretació dels tests d'hipòtesis es poden trobar a Rollal [44], i a Harlow, Mulaik i Steiger [30].

Notem que els procediments estàndards, tals com fixar un nivell de significació (d'acord o no amb el grandària mostral), o calcular les probabilitats de les cues, apareixen de forma natural quan considerem el clàssic problema de Neymann-Pearson: les hipòtesis nul·la i alternativa són *simples*. En aquest cas, qualsevol a priori raonable dona lloc a una probabilitat a posteriori positiva per a la hipòtesi nul·la i una manera adequada de minimitzar la probabilitat d'error, sota la quasi total absència de coneixement sobre la distribució a priori, és maximitzar la potència d'un test una vegada fixat un nivell de significació. Però quan l'alternativa és una hipòtesi composta *no comptable*, actuar de la mateixa manera, per analogia, no sembla prou justificat doncs no clarifica massa bé quin és l'avantatge d'acceptar la hipòtesi nul·la.

Un altre aspecte important es planteja quan els problemes de test d'hipòtesis es veuen com la selecció d'un model apropiat, quan un model està inclòs dintre de l'altre. Aquest és, de fet, un problema lleugerament diferent des d'un punt de vista lògic, però l'aproximació anterior ens pot ajudar a veure també el camí a seguir. Per altra banda, els mètodes estadístics que es desenvoluparan seran aplicables tant en la *selecció del model* com en els *tests d'hipòtesis*.

Des d'aquesta perspectiva, quan es tenen diversos models paramètrics cada un inclòs en l'altre, la selecció del model basada *exclusivament* en el criteri de la màxima-versemblança no resulta convenient donat que sempre ens condueix al model més complex, *independentment* de la grandària mostral, la qual cosa va en contra del sentit comú. Això porta a diversos autors a proposar algunes penalitzacions a la complexitat del model, veure Akaike [2], Good i Gaskins [20]. Però tot i així tampoc no acaben de funcionar.

Agafem, per exemple, el criteri de Akaike, *Akaike Information Criteria*, on es considera la següent quantitat:

$$AIC = -2 \ln(\Lambda) + k$$

on Λ és la raó de versemblança i k el nombre de paràmetres del model. El model amb menor *AIC* és considerat el millor en el sentit de minimitzar l'error quadràtic mig predictiu. No obstant Schwartz (1978), entre altres autors, va puntualitzar que el criteri d'Akaike no és consistent en el sentit que no selecciona el vertader model, amb probabilitat propera a un, quan el grandària mostral tendeix a infinit.

Altres criteris alternatius, com el *Bayesian Information Criteria*

$$BIC = -2 \ln(\Lambda) + k \ln(n)$$

han estat proposats però, en general, aquests criteris o bé minimitzen l'error quadràtic mig predictiu, o bé són consistents, però no ambdues coses a la vegada.

Una tasca que hauria de dur-se a terme abans de desenvolupar qualsevol mètode estadístic de selecció del model seria fer explícites les avantatges de proposar inicialment un model estadístic. Una frase que sintetitza aquesta idea és *All models are wrong but some are useful*, citada en una discussió sobre la robustesa en la construcció d'un model científic per G.E.P. Box l'any 1979 en el llibre *Robustness in Statistics*, pp. 201-236 (London, Academic Press).

En aquest sentit, podríem considerar la selecció del model com un primer pas en

el problema de la estimació i avaluar, a partir d'una mostra prèviament obtinguda, què passaria si treballéssim en el futur amb el model seleccionat. Caldria enfocar l'atenció en l'habilitat dels models estadístics per *facilitar l'estimació* dels mecanismes probabilístics en base a l'informació d'una mostra particular que hagi estat ja obtinguda. Aquest punt de vista està relacionat i pot ser compatible amb altres aproximacions a la selecció del model: aquelles basades en la selecció del model que produeix les millors prediccions per a futures observacions del mateix model que ha generat el conjunt de dades donat. Es pot veure com a referència Geisser i Eddy [18].

En altres paraules, seria convenient intentar clarificar alguns dels problemes estadístics abans esmentats amb l'ajuda de les eines de l'estimació puntual. Per assolir aquest propòsit es necessita avaluar el rendiment d'un estimador estadístic. Recordem que, implícitament, fem servir la paraula *estimador* com a sinònim d'*estimador del mecanisme probabilístic*, això és, donada una mostra proposem una probabilitat a través de l'estimació d'un paràmetre. La bondat d'un estimador, considerat així, respecte la mesura de probabilitat associada a un experiment estadístic hauria d'estar basada en la magnitud dels errors corresponents a la probabilitat estimada dels diferents esdeveniments de l'àlgebra d'observacions. Una manera raonable de proporcionar una mesura quantitativa escalar d'aquests errors és a través d'una distància entre mesures de probabilitat, distància que hauria de ser interpretada com una mesura que avalués com són de diferents dues probabilitats i que per tant, hauria de ser independent de la parametrització del model.

Hi ha diverses distàncies naturals que es poden utilitzar, però donat que sota certes condicions de regularitat els models estadístics tenen estructura d'espai mètric, si considerem la distància Riemanniana induïda per la matriu d'informació de Fisher, també coneguda com distància de Rao, és aquesta distància un candidat natural per mesurar les desviacions de les estimacions, veure com a referències també Rao [42], Burbea i Rao [10], Cencov [11] i Oller i Corcuera [35]. Seria també natural usar la *distància de Hellinger* o la *distància de Bhattacharyya* i, fins i tot, algunes divergències naturals com la de *Kullback-Leibler*, etc., les quals donen totes, localment, la mètrica informativa. En aquest treball s'usarà la distància Riemanniana, principalment per la seva pròpia bellesa i les seves bones propietats dintre de les varietats paramètriques, però cal remarcar que es podria haver utilitzat qualsevol altra distància.

1.3 Estructura geomètrica dels models estadístics

En estadística sovint es tracta a una família paramètrica de distribucions de probabilitat com un model estadístic. Considerem les següents definicions i notacions prèvies:

- χ : espai mostral, freqüentment \mathbb{R} o \mathbb{R}^k .
- \mathcal{a} : σ -àlgebra dels subconjunts de χ , freqüentment l'àlgebra de Borel de \mathbb{R} o \mathbb{R}^k .
- μ una mesura positiva sobre (χ, \mathcal{a}) .
- P_θ , $\theta \in \Theta$: mesura de probabilitat en \mathcal{a} .
- Θ : espai de paràmetres, una varietat real C^∞ n -dimensional. Sovint Θ és un conjunt obert de \mathbb{R}^n i, en aquest cas, s'acostuma a usar el mateix símbol (θ) per a notar punts i coordenades. Així doncs notarem $\phi(\theta) = (\theta_1, \dots, \theta_n) \in \mathbb{R}^n$.

Així doncs anomenarem *model estadístic paramètric* a la terna:

$$\{(\chi, \mathcal{a}, \mu); \Theta; f\}$$

on (χ, \mathcal{a}, μ) és un espai de mesura, Θ , també anomenat *espai de paràmetres*, és una varietat, i f , la *funció model*, és una funció mesurable $f : \chi \times \Theta \rightarrow \mathbb{R}$ tal que $f(x, \theta) \geq 0$ i $P_\theta(dx) = f(x, \theta)\mu(dx)$ és una mesura de probabilitat sobre (χ, \mathcal{a}) , $\forall \theta \in \Theta$, i essent μ una mesura positiva sobre l'espai mesurable (χ, \mathcal{a}) . Ens referirem a μ com la *mesura de referència* del model. A més suposarem que f satisfà les següents condicions de regularitat:

- (i) Θ és una varietat connexa i Hausdorff.
- (ii) Si fixem x , la funció real sobre Θ , $\theta \mapsto f(x, \theta)$, és una funció de classe C^k , amb k convenient a cada cas particular que considerem.
- (iii) Fixat θ , si identifiquem les " n " funcions, de x , següents amb les seves respectives classes d'equivalència dintre de $L^2(P_\theta)$:

$$\frac{\partial}{\partial \theta_i} \log f(x, \theta) \quad i = 1, \dots, n$$

aquestes són linealment independents.

(iv) Els moments de les variables aleatòries

$$\frac{\partial}{\partial \theta_i} \log f(\cdot, \theta)$$

existeixen fins a l'ordre k necessari.

(v) Les derivades parcials $\frac{\partial}{\partial \theta_i}$, o en general fins a ordre k , $\frac{\partial^k}{\partial \theta_{i_1} \cdots \partial \theta_{i_k}}$, on $i_1, \dots, i_k \in \{1, \dots, n\}$, i la integració de $f(x, \theta)$ respecte la mesura μ es poden intercanviar de la següent manera:

$$\begin{aligned} \frac{\partial}{\partial \theta_i} \int_{\mathcal{X}} f(x, \theta) \mu(dx) &= \int_{\mathcal{X}} \frac{\partial}{\partial \theta_i} f(x, \theta) \mu(dx) \\ \frac{\partial^k}{\partial \theta_{i_1} \cdots \partial \theta_{i_k}} \int_{\mathcal{X}} f(x, \theta) \mu(dx) &= \int_{\mathcal{X}} \frac{\partial^k}{\partial \theta_{i_1} \cdots \partial \theta_{i_k}} f(x, \theta) \mu(dx) \end{aligned}$$

Quan es satisfacin totes aquestes condicions, direm que el model estadístic paramètric és *regular*, i en aquest cas l'espai Θ tindrà una estructura natural de varietat Riemanniana, donada per la seva mètrica informativa. Per a més detalls, veure Amari [3], Atkinson i Mitchell [4], Barndorff-Nielsen [5], Burbea i Rao [10], Burbea [8] i Oller [34], entre altres.

Cal notar que donat un model estadístic $\{(\mathcal{X}, a, \mu); \Theta; f\}$, tal i com l'acabem de definir, podem establir les següents condicions suficients sobre la funció de densitat $f(x, \theta)$ i poder derivar sota el signe integral:

Proposició 1.3.1 *Suposem que Θ és un obert de \mathbb{R}^n i l'aplicació $(x, \theta) \rightarrow f(x, \theta)$ de $\mathcal{X} \times \Theta$ en $\overline{\mathbb{R}}$, amb $\theta \in \Theta$, compleix*

- 1) $\theta \rightarrow f(x, \theta)$ és finita i C^r μ -q.p.t. $x \in \mathcal{X}$
- 2) Per a cada $\theta \in \Theta$ existeixen r funcions integrables $g_k(x)$ $k = 1, \dots, r$, en un entorn de θ tal que

$$\left| \frac{\partial^k f(x, \theta)}{\partial \theta_{i_1} \cdots \partial \theta_{i_k}} \right| \leq g_k(x) \quad k = 1, \dots, r \quad i_1, \dots, i_k \in \{1, \dots, n\}$$

Aleshores

$$\frac{\partial^k}{\partial \theta_{i_1} \cdots \partial \theta_{i_k}} \int_{\mathcal{X}} f(x, \theta) \mu(dx) = \int_{\mathcal{X}} \frac{\partial^k}{\partial \theta_{i_1} \cdots \partial \theta_{i_k}} f(x, \theta) \mu(dx)$$

La demostració es pot trobar a qualsevol text d'anàlisi matemàtica que contingui un mínim de teoria de la mesura, i està basada en els teoremes de convergència dominada de Lebesgue, veure per exemple Dieudonné [13].

Observació 1.3.1 *Si Θ és una varietat de \mathbb{R}^n , la proposició anterior també és certa.*

En efecte, com que Θ és una varietat sobre \mathbb{R}^n , tot conjunt obert $U \subset \Theta$ s'identifica homeomòrficament amb un entorn obert de \mathbb{R}^n . Aleshores, usant com a sistema de coordenades la identitat $\phi = Id$ i fent un abús de notació a l'hora d'identificar els punts $\theta \in U$ amb les seves coordenades en \mathbb{R}^n , $\phi(\theta) = (\theta_1, \dots, \theta_n)$, podem aplicar la proposició anterior. ■

Dintre de la varietat Θ de \mathbb{R}^n , i identificant punts amb coordenades, podem considerar la matriu $n \times n$, $G = (g_{ij})$, definida per

$$g_{ij}(\theta) = \int_{\mathcal{X}} \frac{\partial}{\partial \theta_i} \log f(x, \theta) \frac{\partial}{\partial \theta_j} \log f(x, \theta) f(x, \theta) \mu(dx) \quad i, j = 1, \dots, n \quad (1.1)$$

Sota les condicions de regularitat definides anteriorment, aquesta matriu anomenada *matriu d'informació de Fisher*, és simètrica i estrictament definida positiva. A més, els coeficients de G , g_{ij} , defineixen les components d'un camp tensorial de segon ordre covariant, simètric i definit positiu, que podem utilitzar per definir l'element mètric diferencial

$$ds^2 = \sum_{i=1}^n \sum_{j=1}^n g_{ij}(\theta) d\theta_i d\theta_j$$

el qual proporciona a Θ l'estructura de varietat Riemanniana, que en ser connexa i Hausdorff, esdevé espai mètric, veure Hicks [23] o Spivak [46], entre d'altres, i la corresponent distància anomenada en general distància Riemanniana, en aquest context també s'anomena distància de Rao, veure Rao [43] i Burbea i Rao [10], de la qual són conegudes les següents propietats:

- És invariant sota transformacions *suficients* del model estadístic.
- És invariant sota canvis en la mesura de referència.
- És invariant sota reparametrizacions, és a dir, difeomorfismes entre Θ i Θ' .

- És invariant sota canvis de coordenades en la varietat paramètrica Θ .
- És, llevat d'una constant multiplicativa, localment proporcional a totes les φ -divergències, $D_\varphi(P, Q) = \int_{\mathcal{X}} \varphi\left(\frac{f}{g}\right) g d\mu$, amb φ convexa i suficientment "suau", P i Q dues probabilitats i on μ és una mesura que domina P i Q amb $f = \frac{dP}{d\mu}$, $g = \frac{dQ}{d\mu}$.

Com a conclusió podem dir que: en els models estadístics paramètrics regulars, les mesures probabilístiques indueixen una mètrica Riemanniana en la varietat de paràmetres Θ , la qual passa a ser doncs, una varietat Riemanniana.

1.4 Algunes noves propietats de la distància de Rao

En aquesta secció es desenvoluparan alguns resultats interessants que s'obtenen a partir de la distància de Rao i les seves propietats. Aquests resultats poden ser útils per clarificar el procés de selecció d'una distància. Per això considerarem les següents definicions i notacions prèvies.

Sigui $\{(\mathcal{X}, a, \mu); \Theta; f\}$ un model estadístic paramètric que compleix les condicions de regularitat definides a l'apartat anterior.

Definim el conjunt

$$\mathcal{M}(\mathcal{X}, a) = \{X : \mathcal{X} \rightarrow \mathbb{R} : \text{funcions mesurables sobre } a\}$$

Considerem també els següents conjunts L i S

$$L = \left\{ X \in \mathcal{M}(\mathcal{X}, a) : \int X^2(x) f(x, \theta) \mu(dx) < \infty, \forall \theta \in \Theta \right\}$$

$$S = \left\{ X : X \in L, \frac{\partial}{\partial \theta_i} \int_{\mathcal{X}} X(x) f(x, \theta) \mu(dx) = \int_{\mathcal{X}} X(x) \frac{\partial f(x, \theta)}{\partial \theta_i} \mu(dx), \forall \theta \in \Theta \right\}$$

També podem definir el següent subespai

$$H_\theta = \left\langle \frac{\partial}{\partial \theta_1} \log f(\cdot, \theta), \dots, \frac{\partial}{\partial \theta_n} \log f(\cdot, \theta) \right\rangle \subset L$$

on les derivades parcials estan avaluades en un cert punt θ fixat prèviament.

Usarem també la següent notació

$$\Phi_X(\theta) = E_\theta(X) = \int_{\mathcal{X}} X(x) f(x; \theta) \mu(dx)$$

amb

$$\text{cov}_\theta(X, Y) = E_\theta((X - E_\theta(X))(Y - E_\theta(Y))) = E_\theta(XY) - E_\theta(X)E_\theta(Y)$$

$$\text{var}_\theta(X) = \text{cov}_\theta(X, X)$$

Designarem per $T_\theta(\Theta)$ l'espai tangent a θ i els seus elements seran vectors tangents a la varietat en el punt θ . Considerem també, $\langle \cdot, \cdot \rangle_\theta$ el producte escalar a $T_\theta(\Theta)$ induït per la mètrica Riemanniana i $\| \cdot \|_\theta$ la norma Euclidiana associada. Per a més detalls veure l'apèndix.

Lema 1.4.1 *Sigui $X \in S$, aleshores es compleix la següent igualtat*

$$\frac{\partial \Phi_X}{\partial \theta_i} = \text{cov}_\theta\left(X, \frac{\partial \log f}{\partial \theta_i}\right)$$

Demostració:

Donat que $X \in S$, sabem que podem passar les derivades dintre del signe integral i tenint en compte, a més, les propietats de la funció logaritme podem considerar la següent cadena d'igualtats

$$\begin{aligned} \frac{\partial \Phi_X}{\partial \theta_i} &= \frac{\partial}{\partial \theta_i} \int_{\mathcal{X}} X(x) f(x, \theta) \mu(dx) \\ &= \int_{\mathcal{X}} X(x) \frac{\partial f(x, \theta)}{\partial \theta_i} \mu(dx) \\ &= \int_{\mathcal{X}} X(x) \frac{\partial \log f(x, \theta)}{\partial \theta_i} f(x, \theta) \mu(dx) \\ &= \text{cov}_\theta\left(X, \frac{\partial \log f}{\partial \theta_i}\right) \end{aligned}$$

L'última desigualtat donat que $E_\theta\left(\frac{\partial \log f}{\partial \theta_i}\right) = 0$, $i = 1, \dots, n$.

■

Per clarificar els resultats que veurem a continuació, enunciem primer una proposició en un context general d'àlgebra lineal, que aplicarem després al nostre cas particular.

Proposició 1.4.1 *Sigui E un espai vectorial amb una forma bilineal simètrica i semidefinida positiva $\langle \cdot, \cdot \rangle$ i $H \subset E$ de manera que $H = \langle u_1, \dots, u_n \rangle$ amb $\{u_1, \dots, u_n\}$ una base de H de manera que $\langle \cdot, \cdot \rangle > 0$ sobre aquest subespai. Donat qualsevol $X \in E$ existeix $U \in H$, únic, de forma que*

$$\|X - U\|^2 = \langle X - U, X - U \rangle$$

és mínima, i aquesta U l'anomenarem $U = \pi(X)$, projecció ortogonal de X sobre H . Aleshores X es pot descompondre en

$$X = U + V, \quad U \in H, \quad V \in H^\perp$$

amb $V = X - U$ i $V \in H^\perp$ en el sentit que $\langle U, V \rangle = 0$, dit d'una altra manera

$$H^\perp = \{V \in E : \langle U, V \rangle = 0, \forall U \in H\}$$

Demostració:

La demostració d'aquest resultat es pot veure en qualsevol llibre d'àlgebra lineal bàsica, per exemple Queisane [41].

■

Corol·lari 1.4.1 *En el nostre cas particular, considerem l'espai vectorial n -dimensional de funcions*

$$H_\theta = \left\langle \frac{\partial}{\partial \theta_1} \log f(\cdot, \theta), \dots, \frac{\partial}{\partial \theta_n} \log f(\cdot, \theta) \right\rangle \subset L$$

Aleshores, donat que el producte escalar de $L^2(P_\theta)$, permet definir en L una forma bilineal simètrica i semidefinida positiva, mitjançant la identificació dels elements de L amb les corresponents classes d'equivalència en $L^2(P_\theta)$, i donat que el model estadístic satisfà les condicions de regularitat definides anteriorment, sabem que sobre H_θ aquesta forma bilineal serà definida positiva i la projecció ortogonal $\pi_\theta : L \rightarrow H_\theta$ estarà ben definida amb $U = \pi_\theta(X)$ i $V = X - U \in H_\theta^\perp$ de manera que $\text{cov}_\theta(U, V) = 0$.

Notem que U i V depenen de θ , però prescindim dels subíndexs per no fer tant carregosa la notació.

Observem també que el producte escalar en H_θ definit per aquesta forma bilineal es pot notar de la següent manera, donats dos elements $W, Z \in H_\theta$ i tenint en compte que $E_\theta \left(\frac{\partial}{\partial \theta_i} \log f \right) = 0$, $i = 1, \dots, n$

$$\langle W, Z \rangle_{H_\theta} = E_\theta(W Z) = \text{cov}_\theta(W, Z)$$

i

$$\|W\|_{H_\theta}^2 = \text{var}_\theta(W)$$

Lema 1.4.2 *Sigui $X \in S$ tal que*

$$X = U + V, \quad U = \pi_\theta(X) \in H_\theta, \quad V = X - U \in H_\theta^\perp$$

aleshores l'aplicació definida per:

$$\begin{aligned} \mathfrak{J} : (H_\theta, \langle \cdot, \cdot \rangle_{H_\theta}) &\longrightarrow (T_\theta(\Theta), \langle \cdot, \cdot \rangle_\theta) \\ z^1 \frac{\partial \log f}{\partial \theta_1} + \dots + z^n \frac{\partial \log f}{\partial \theta_n} &\longmapsto z^1 \frac{\partial}{\partial \theta_1} + \dots + z^n \frac{\partial}{\partial \theta_n} \end{aligned}$$

és una isometria entre espais vectorials tal que

- a) $\text{grad}(\Phi_X) = \mathfrak{J}(\pi_\theta(X)) = \mathfrak{J}(U)$
- b) $\|\text{grad}(\Phi_X)\|_\theta^2 = \text{var}_\theta(U)$
- c) $\text{var}_\theta(X) = \|\text{grad}(\Phi_X)\|_\theta^2 + \text{var}_\theta(V)$

Demostració:

Suposarem donat un sistema de coordenades a Θ i farem un abús de notació, ja comentat anteriorment, identificant els punts $\theta \in \Theta$ amb les seves coordenades, per a més detalls veure l'apèndix. Cal observar que, per construcció de l'aplicació \mathfrak{J} i tal com s'han definit els espais H_θ i $T_\theta(\Theta)$, tenim

$$\begin{aligned} \left\langle \frac{\partial \log f}{\partial \theta_i}, \frac{\partial \log f}{\partial \theta_j} \right\rangle_{H_\theta} &= E_\theta \left(\frac{\partial \log f}{\partial \theta_i} \frac{\partial \log f}{\partial \theta_j} \right) = g_{ij}(\theta) \\ &= \left\langle \frac{\partial}{\partial \theta_i}, \frac{\partial}{\partial \theta_j} \right\rangle_\theta = \left\langle \mathfrak{J} \left(\frac{\partial \log f}{\partial \theta_i} \right), \mathfrak{J} \left(\frac{\partial \log f}{\partial \theta_j} \right) \right\rangle_\theta \end{aligned}$$

Així doncs com el producte escalar, i per tant el tensor mètric, dels dos espais coincideix tindrem que \mathfrak{J} és una isometria.

Donat que $U = \pi_\theta(X)$ tenim que $\text{cov}_\theta(U, V) = 0$, i també

$$\text{var}_\theta(X) = \text{var}_\theta(U) + \text{var}_\theta(V) \quad (1.2)$$

Per altra banda, usant el lema 1.4.1

$$\frac{\partial \Phi_X}{\partial \theta_i} = \text{cov}_\theta\left(X, \frac{\partial \log f}{\partial \theta_i}\right) = \text{cov}_\theta\left(U, \frac{\partial \log f}{\partial \theta_i}\right)$$

l'última igualtat donat que

$$\text{cov}_\theta\left(V, \frac{\partial \log f}{\partial \theta_i}\right) = 0, \quad i = 1, \dots, n$$

A més, tenint en compte que U es pot expressar com

$$U = \lambda^1 \frac{\partial \log f}{\partial \theta_1} + \dots + \lambda^n \frac{\partial \log f}{\partial \theta_n}$$

per a unes funcions escalars convenientes (dependents de θ) $\lambda^1, \dots, \lambda^n$, tenim

$$\frac{\partial \Phi_X}{\partial \theta_i} = \sum_{j=1}^n \lambda^j \text{cov}_\theta\left(\frac{\partial \log f}{\partial \theta_j}, \frac{\partial \log f}{\partial \theta_i}\right) = \sum_{j=1}^n \lambda^j g_{ij}$$

on g_{ij} són els coeficients de la matriu d'informació de Fisher, és a dir, les components del tensor mètric.

Per altra banda la i -ena component de $\text{grad}(\Phi_X)$, respecte les bases del camp vectorial $\partial/\partial\theta_1, \dots, \partial/\partial\theta_n$, sobre θ , ve donada per

$$\begin{aligned} \text{grad}(\Phi_X)^i &= \sum_{k=1}^n g^{ik} \frac{\partial \Phi_X}{\partial \theta_k} = \sum_{k=1}^n g^{ik} \left(\sum_{j=1}^n \lambda^j g_{kj} \right) \\ &= \sum_{j=1}^n \left(\sum_{k=1}^n g^{ik} g_{kj} \right) \lambda^j \\ &= \sum_{j=1}^n \delta_i^j \lambda^j = \lambda^i \end{aligned}$$

essent δ_i^j les funcions delta de Kronecker, per tant

$$\text{grad}(\Phi_X) = \mathfrak{J}(U)$$

i a més, en ser \mathfrak{J} una isometria

$$\|\text{grad}(\Phi_X)\|_\theta^2 = \|U\|_{H_\theta}^2 = \text{var}_\theta(U) \quad (1.3)$$

donat que $E_\theta(U) = 0$.

Aleshores combinant les equacions (1.2) i (1.3), obtenim l'últim resultat resultat buscat

$$\text{var}_\theta(X) = \|\text{grad}(\Phi_X)\|_\theta^2 + \text{var}_\theta(V)$$

■

Una conseqüència immediata dels resultats anteriors és el següent

Corol·lari 1.4.2 Si $X \in S$ és tal que $\text{var}_\theta(X) \leq 1$, $\forall \theta \in \Theta$, aleshores

$$\|\text{grad}(\Phi_X)\|_\theta^2 \leq 1$$

Demostració:

$$\|\text{grad}(\Phi_X)\|_\theta^2 = \text{var}_\theta(X) - \text{var}_\theta(V) \leq \text{var}_\theta(X) \leq 1$$

■

Teorema 1.4.1 Donat $\theta_0 \in \Theta$, i $v_{\theta_0} \in T_{\theta_0}(\Theta)$ ($v_{\theta_0} \neq 0$), designem per $v_{\theta_0}(\Phi_X)$ l'acció del vector sobre la funció Φ_X . Definim l'aplicació

$$\begin{aligned} g : S &\longrightarrow \mathbb{R} \\ X &\longmapsto v_{\theta_0}(\Phi_X) \end{aligned}$$

Aleshores:

$$\max_{X \in S: \text{var}_{\theta_0}(X)=1} g(X) = \sqrt{\langle v, v \rangle_{\theta_0}} = \|v\|_{\theta_0}$$

i una variable aleatòria X que fa màxim el valor anterior ve donada per

$$X = \frac{v^1}{\|v\|_{\theta_0}} \frac{\partial \log f(\cdot, \theta)}{\partial \theta_1} \Big|_{\theta_0} + \dots + \frac{v^n}{\|v\|_{\theta_0}} \frac{\partial \log f(\cdot, \theta)}{\partial \theta_n} \Big|_{\theta_0}$$

on v^1, \dots, v^n són les components de v_{θ_0} , respecte la base de $T_{\theta_0}(\Theta)$,

$$\frac{\partial}{\partial \theta_1} \Big|_{\theta_0}, \dots, \frac{\partial}{\partial \theta_n} \Big|_{\theta_0}.$$

Demostració:

Tenint en compte que

$$v_{\theta_0}(\Phi_X) = v^1 \frac{\partial \Phi_X}{\partial \theta_1} \Big|_{\theta_0} + \dots + v^n \frac{\partial \Phi_X}{\partial \theta_n} \Big|_{\theta_0}$$

i segons el Lema (1.4.1)

$$\frac{\partial \Phi_X}{\partial \theta_i} = \text{cov}_{\theta}(X, \frac{\partial \log f}{\partial \theta_i})$$

resulta

$$v_{\theta_0}(\Phi_X) = \sum_{i=1}^n v^i \text{cov}_{\theta_0}(X, \frac{\partial \log f(\cdot, \theta)}{\partial \theta_i} \Big|_{\theta_0})$$

Però donat que $X \in L \subset S$, podem usar la següent descomposició:

$$X = U + V, \quad U \in H_{\theta_0}, \quad V \in H_{\theta_0}^{\perp}$$

verificant $E_{\theta_0}(U) = 0$, doncs

$$U = U^1 \frac{\partial \log f(\cdot, \theta)}{\partial \theta_1} \Big|_{\theta_0} + \dots + U^n \frac{\partial \log f(\cdot, \theta)}{\partial \theta_n} \Big|_{\theta_0}$$

També es compleix que $\text{cov}_{\theta_0}(U, V) = 0$; així tenim que, utilitzant el conveni de la suma d'índexs repetits,

$$v_{\theta_0}(\Phi_X) = g_{ij}(\theta_0) v^i U^j$$

Per la desigualtat de Cauchy-Schwarz el producte escalar $v_{\theta_0}(\Phi_X)$ serà màxim quan

$$U^j = \lambda v^j \quad \lambda > 0$$

per tant

$$\text{var}_{\theta_0}(U) = \lambda^2 g_{ij}(\theta_0) v^i v^j = \text{var}_{\theta_0}(X) - \text{var}_{\theta_0}(V) = 1 - \text{var}_{\theta_0}(V)$$

$$\lambda^2 = \frac{1 - \text{var}_{\theta_0}(V)}{g_{ij}(\theta_0) v^i v^j} = \frac{1 - \text{var}_{\theta_0}(V)}{\|v\|_{\theta_0}^2}$$

i tenim

$$\begin{aligned} v_{\theta_0}(\Phi_X) &= \frac{\sqrt{1 - \text{var}_{\theta_0}(V)}}{\|v\|_{\theta_0}} g_{ij}(\theta_0) v^i v^j \\ &= \frac{\sqrt{1 - \text{var}_{\theta_0}(V)}}{\|v\|_{\theta_0}} \|v\|_{\theta_0}^2 = \sqrt{1 - \text{var}_{\theta_0}(V)} \|v\|_{\theta_0} \end{aligned}$$

i això és màxim quan $\text{var}_{\theta_0}(V) = 0$, per tant

$$\max_{X \in S: \text{var}_{\theta_0}(X)=1} g(X) = \|v\|_{\theta_0}$$

Si prenem $V = 0$, la variable X que fa màxim aquest valor és

$$X = \frac{v^1}{\|v\|_{\theta_0}} \frac{\partial \log f(\cdot, \theta)}{\partial \theta_1} \Big|_{\theta_0} + \dots + \frac{v^n}{\|v\|_{\theta_0}} \frac{\partial \log f(\cdot, \theta)}{\partial \theta_n} \Big|_{\theta_0}$$

■

Interpretació:

Si partim de la informació que ens dóna una variable aleatòria X , podem pensar que la capacitat que tenim per detectar com a diferents (o discriminar) dues poblacions properes, θ_0 i $\theta_0 + \Delta\theta$, és proporcional a la variació del valor mig d'aquesta variable, al passar d'una població a l'altra, una vegada fixada la variància en un punt θ_0 , per exemple $\text{var}_{\theta_0}(X) = 1$. Així una mesura global de la nostra capacitat de discriminació d'aquestes dues poblacions es pot definir prenent el suprem, d'aquesta quantitat, dins una classe convenient de variables aleatòries X . El resultat obtingut ens mostra que aquest suprem ve donat, localment, per la mètrica Riemanniana, induïda per la matriu d'informació de Fisher, o distància de Rao.

Si prenem $v^1 = \Delta\theta_1, \dots, v^n = \Delta\theta_n$ i usem el conveni de la suma d'índexs repetits, el teorema anterior permet escriure

$$\Delta\Phi_X(\theta_0) \approx \text{cov}_{\theta_0}(X, \frac{\partial \log f(\cdot, \theta)}{\partial \theta_i} \Big|_{\theta_0}) \Delta\theta_i$$

i això és màxim i igual a $\sqrt{g_{ij}(\theta_0) \Delta\theta_i \Delta\theta_j}$ quan

$$X = \Delta\theta_1 \frac{\partial \log f(\cdot, \theta)}{\partial \theta_1} \Big|_{\theta_0} + \dots + \Delta\theta_n \frac{\partial \log f(\cdot, \theta)}{\partial \theta_n} \Big|_{\theta_0}$$

Com a referències bibliogràfiques, on apareixen idees o conceptes semblants, podem citar Efron [6] i Fisher [16].

Una altra propietat interessant de la distància de Rao ve donada pel següent

Teorema 1.4.2 *Sigui $X \in S$ on S és el conjunt de variables aleatòries al començament de la secció i sigui α la corba de classe C^∞ que uneix $\nu, \xi \in \Theta$.*

Considerem $\Phi_X(\theta)$ també definida anteriorment i $l(\nu, \xi)$ com la longitud de α .

Aleshores si $\Delta\Phi_X = \Phi_X(\xi) - \Phi_X(\nu)$

$$|\Delta\Phi_X| \leq \left(\max_{\alpha} \sigma_X(\theta) \right) l(\nu, \xi)$$

on

$$\sigma_X(\theta) = \sqrt{\int_{\mathcal{X}} X^2(x) f(x, \theta) \mu(dx) - \Phi_X^2(\theta)}$$

En particular, si γ és una geodèsica minimal, tindrem

$$|\Delta\Phi_X| \leq \left(\max_{\gamma} \sigma_X(\theta) \right) \rho(\nu, \xi)$$

amb ρ la distància de Rao.

Demostració:

Siguin $\frac{d\theta_i}{ds}$ les components del vector tangent a la corba α parametritzada

per la longitud d'arc, usant el conveni de la suma d'índexs repetits tenim

$$\frac{d\Phi_X}{ds} = \frac{\partial\Phi_X}{\partial\theta_i} \frac{d\theta_i}{ds}$$

i usant també el resultat del lema (1.4.1)

$$\begin{aligned} |\Delta\Phi_X| &= \left| \int_{\alpha} \frac{d\Phi_X}{ds} ds \right| = \left| \int_{\alpha} \text{cov}_{\theta}(X, \frac{\partial \log f(\cdot, \theta)}{\partial \theta_i}) \frac{d\theta_i}{ds} ds \right| \\ &\leq \int_{\alpha} \left| \text{cov}_{\theta}(X, \frac{\partial \log f(\cdot, \theta)}{\partial \theta_i}) \frac{d\theta_i}{ds} \right| ds \\ &\leq \int_{\alpha} \sigma_X(\theta) \sqrt{\text{var}_{\theta}(\frac{\partial \log f(\cdot, \theta)}{\partial \theta_i})} \frac{d\theta_i}{ds} ds \end{aligned}$$

l'última desigualtat segons Cauchy-Schwarz.

Tenim però que

$$\begin{aligned} \text{var}_\theta \left(\frac{\partial \log f(\cdot, \theta)}{\partial \theta_i} \frac{d\theta_i}{ds} \right) &= E_\theta \left(\frac{\partial \log f(\cdot, \theta)}{\partial \theta_i} \frac{\partial \log f(\cdot, \theta)}{\partial \theta_j} \right) \frac{d\theta_i}{ds} \frac{d\theta_j}{ds} \\ &= g_{ij} \frac{d\theta_i}{ds} \frac{d\theta_j}{ds} = 1 \end{aligned}$$

Per tant s'obté

$$|\Delta \Phi_X| \leq \int_\alpha \sigma_X(\theta) ds \leq \left(\max_\alpha \sigma_X(\theta) \right) l(\nu, \xi)$$

En particular, si $\alpha = \gamma$ (geodèsica minimal)

$$|\Delta \Phi_X| \leq \left(\max_\gamma \sigma_X(\theta) \right) \rho(\nu, \xi)$$

■

Nota 1: En alguns casos pot ser convenient considerar

$$\Delta \Phi_X \leq \left(\sup_{\theta \in \Theta} \sigma_X(\theta) \right) \rho(\nu, \xi)$$

cal tenir en compte, però, que aquest suprem pot ser infinit.

Nota 2: Aquest resultat es pot generalitzar lleugerament considerant $X(x; \theta)$ una variable aleatòria tal que compleixi

a) podem derivar sota signe integral respecte θ .

b) $E \left[\frac{\partial X}{\partial \theta} \right] = 0$.

Nota 3: El teorema es pot estendre també a corbes trencades de classe C^∞ a trossos.

Hem vist doncs, una nova metodologia en la qual fem servir la distància de Rao, tot i que ja hem comentat que se'n podrien usar d'altres, donat que té unes bones

propietats dintre de las varietats paramètriques. Els motius que ens fan decidir, doncs, per la distància de Rao són, en primer lloc, que la topologia induïda per la distància Riemanniana és sempre la mateixa que la topologia induïda per l'atlas en la varietat Θ , la qual cosa no passa per exemple amb altres distàncies de tipus "global", com la de Hellinger.

Per altra banda, les varietats de paràmetres amb la distància de Rao són, generalment, un tipus particular d'espais mètrics, coneguts com Espais de Longitud (traducció literal del francès *Espace de Longueur*), per exemple quan Θ és una varietat Riemanniana completa, o bé una varietat Riemanniana amb frontera, mentre que en altres distàncies, com per exemple la de Hellinger, això no es dona. Un espai mètric (X, d) és un Espai de Longitud si la distància entre dos punts qualsevol és sempre igual a l'ínfim de la longitud de les corbes que els uneixen, i sabem que, en una varietat Riemanniana, la distància entre dos punts està definida com l'ínfim de la longitud de les corbes trencades C^∞ que els uneixen, per a més detalls veure Gromov [21].

Els espais mètrics "de longitud" tenen doncs la propietat que la "distància" pot mesurar-se amb "cinta mètrica", cosa que els fa més semblants a l'espai que percebem a través dels sentits, i això creiem que és una bona característica, doncs facilita la comprensió i intuïció dels objectes d'estudi.

Capítol 2

Alternatives al contrast d'hipòtesis clàssic

Una idea clau, potser una mica imprecisa al començament, per entendre el següent desenvolupament és que acceptar la hipòtesi nul·la (o submodel) significarà que sota certes condicions (mostres de mida suficientment petita), podrem fer *en algun sentit* millors estimacions de la vertadera mesura de probabilitat si restringim les estimacions al submodel que si usem el model més general.

Això està d'acord amb el sentit comú i l'experiència de l'estadístic aplicat; per exemple, cap estadístic proposarà un model de regressió polinòmica amb un elevat nombre de paràmetres si es tenen "poques" observacions respecte el total de paràmetres, preferirà fer, en canvi, una simple regressió lineal.

Sigui $\{(\mathcal{X}, a, \mu); \Theta; f\}$ un model estadístic n -paramètric regular i essent Θ una varietat Riemanniana n -dimensional, real i C^∞ . Suposarem, a més, que Θ és Hausdorff i connexa i sigui ρ^2 el quadrat de la distància de Rao sobre aquesta varietat. Considerem el problema de contrastar:

$$H_0 : \theta \in \Theta_0 \quad \text{contra} \quad H_1 : \theta \notin \Theta_0$$

tal com ja s'ha introduït en el capítol anterior, on Θ_0 és una varietat r -dimensional, real, C^∞ , inclosa en Θ , amb $0 \leq \dim \Theta_0 < \dim \Theta$, o bé el problema relacionat amb la selecció del model, model Θ_0 contra Θ , $\Theta_0 \subset \Theta$. Considerem també una mostra aleatòria simple $x = (x_1, \dots, x_k) \in \mathcal{X}^k$ de mida $k \in \mathbb{N}$. A continuació ens podem plantejar una qüestió diferent als tests habituals, en la qual, d'acord amb la informació obtinguda a partir de la mostra de grandària k , volem decidir entre hipòtesis o models en els quals es treballi amb la mateixa (o equivalent) població estadística però manegant informació de mostres de grandària m .

S'exposaran respostes parcials a aquesta qüestió, desenvolupant un mètode estadístic el qual *exigeix* l'acceptació d'un *estimador* com a estimador convenient pel model més general, per exemple l'estimador màxim-versemblant, encara que es poden triar altres estimadors. En altres paraules, les respostes proposades seran *dependents* del corresponent estimador. A partir d'ara considerarem:

$$\mathcal{U}_m : \chi^m \rightarrow \Theta$$

l'estimador proposat corresponent a una mostra de mida m , pel model més general.

Això no creiem en el fons que sigui un problema, potser al contrari, doncs segons el nostre parer, tots els aspectes de la inferència estan relacionats i això posa de relleu la relació entre estimació, per una banda i contrast d'hipòtesis, o elecció de model, per altra.

2.1 Els Mètodes

Imaginem, en un primer moment, que el vertader valor del paràmetre, diem-li θ , és conegut. Aleshores, una mesura *raonable* de l'error d'una estimació vindrà donada per:

$$\rho^2(\mathcal{U}_m(x), \theta)$$

mentre que l'error corresponent a la hipòtesi nul·la (o submodel) pot venir mesurat per:

$$\rho^2(\Theta_0, \theta) = \inf\{\rho^2(\nu, \theta) : \nu \in \Theta_0\}$$

Des d'aquesta perspectiva i tenint en compte aquestes consideracions, podria ser convenient introduir les següents regles:

2.1.1 Regles p i m

regla – p .

Rebutjarem la hipòtesi nul·la, o el que és el mateix, triarem el model total Θ si

$$p = P(\rho^2(\mathcal{U}_m(x), \theta) < \rho^2(\Theta_0, \theta)) - P(\rho^2(\mathcal{U}_m(x), \theta) > \rho^2(\Theta_0, \theta)) > 0 \quad (2.1)$$

altrament acceptarem el model més simple Θ_0 .

En un llenguatge més planer, si la probabilitat de fer més error en l'estimació del paràmetre agafant tot el model Θ , és més gran o igual que no pas prenent Θ_0 , llavors ens quedarem amb Θ_0 .

Per il·lustrar amb uns exemples senzills aquesta regla, considerem la següent taula

| | A | B | C | D | E |
|--|------------|------------|----------|------------|------------|
| $P(\rho^2(\mathcal{U}_m(x), \theta) < \rho^2(\Theta_0, \theta))$ | 40% | 40% | 50% | 50% | 40% |
| $P(\rho^2(\mathcal{U}_m(x), \theta) = \rho^2(\Theta_0, \theta))$ | 11% | 10% | 10% | 0% | 20% |
| $P(\rho^2(\mathcal{U}_m(x), \theta) > \rho^2(\Theta_0, \theta))$ | 49% | 50% | 40% | 50% | 40% |
| criteri: ens quedarem amb | Θ_0 | Θ_0 | Θ | Θ_0 | Θ_0 |

En aquesta taula es representen 5 situacions diferents: A, B, C, D i E. Per a cada situació tenim tres caselles en les quals ens donen el percentatge de casos que es troben dintre de cada una de les tres possibilitats: $<$, $=$ i $>$. Finalment, segons la regla $\text{—}p$ i en l'última fila, obtenim el criteri segons el qual ens quedarem amb el submodel Θ_0 , o bé amb el model general Θ .

Fixem-nos que, amb aquesta regla, només ens quedarem amb el model més general Θ quan, el percentatge de casos pels quals l'error al considerar el vertader valor del paràmetre dintre del submodel Θ_0 sigui estrictament més gran que si el considerem dintre del model Θ . En qualsevol altre cas es tendeix a quedar o preferir el model més senzill, la qual cosa és coherent amb el sentit comú.

Observem que ens quedem amb un model, no perquè sigui el vertader, sinó perquè fem millors estimacions del paràmetre suposant que és cert o, dit d'una altra manera, la probabilitat d'error que fem en estimar el paràmetre suposant que pertany a dit model és menor que si no ho féssim.

regla $\text{—} m$.

Rebutjarem la hipòtesi nul·la, o el que és el mateix, triarem el model total Θ si

$$m = \rho^2(\Theta_0, \theta) - E(\rho^2(\mathcal{U}_m(x), \theta)) > 0 \quad (2.2)$$

altrament acceptarem el model més simple Θ_0 .

És un criteri molt semblant a l'anterior, però fent servir moments en lloc de probabilitats: si per terme mig l'error en l'estimació del paràmetre agafant el model general Θ és superior o igual al que fem agafant Θ_0 , ens quedarem amb Θ_0 .

La primera regla, a la qual ens referirem a partir d'ara com regla- \mathfrak{p} , donat que està basada en probabilitats, és invariant sota transformacions monòtones de la distància de Rao. Per tant, és menys dependent de la mètrica considerada, o de la *funció de pèrdua* usada. La segona regla, a la qual ens referirem a partir d'ara com regla- \mathfrak{m} , basada en moments és, algunes vegades, més fàcil de calcular.

Notem que, si $\rho^2(\Theta_0, \theta) > 0$, per a qualsevol estimador consistent \mathcal{U}_m tenim

$$\lim_{m \rightarrow \infty} P(\rho^2(\mathcal{U}_m(x), \theta) < \rho^2(\Theta_0, \theta)) = 1$$

la qual cosa passa amb probabilitat 1 sota les nostres suposicions. A més, per a qualsevol estimador \mathcal{U}_m , L^2 -consistent (respecte la mètrica Riemanniana):

$$\lim_{m \rightarrow \infty} E(\rho^2(\mathcal{U}_m(x), \theta)) = 0$$

rebutjant sempre la hipòtesi nul·la per a mostres de mida gran.

Per altra banda, si $\rho^2(\Theta_0, \theta) = 0$, la qual cosa passa amb probabilitat zero sota les nostres suposicions, aleshores $\mathfrak{p} \leq 0$ i $\mathfrak{m} \leq 0$, acceptant sempre la hipòtesi nul·la (o submodel), per a tot $m \in \mathbb{N}$.

Observem que les decisions basades en ambdues regles són dependents de la grandària mostral m , essent això raonable. Un model serà o no acceptable segons el *nivell de resolució* amb el que es vulgui o es pugui treballar, nivell que estarà fixat per m .

2.1.2 Procediments \mathfrak{f} i \mathfrak{b}

En realitat però, no es coneix el vertader valor del paràmetre, per tant s'han de suggerir diferents camins per a poder donar respostes. Aquí es proposen dos procediments:

procediment – \mathfrak{f} .

Una solució freqüentista. Amb l'objectiu d'aplicar les regles definides anteriorment, donarem un $(1 - \alpha)$ -interval de confiança per a \mathfrak{p} o \mathfrak{m} , de la forma $[\mathfrak{p}_*, 1]$ or $[\mathfrak{m}_*, \infty)$ basat en una mostra $x =$

$(x_1, \dots, x_k) \in \chi^k$, on p_* i m_* es pot considerar que són estimacions o estimacions per defecte de p i m respectivament. A més, una condició que es requereix és que hem de triar un interval de confiança de manera que:

$$p_* \xrightarrow{\mathcal{P}} p \quad \text{o bé} \quad m_* \xrightarrow{\mathcal{P}} m$$

quan $k \rightarrow \infty$, i on \mathcal{P} indica la convergència en probabilitat. Aquesta condició garanteix un comportament adequat al procediment-f.

En general p i m són funcions que depenen de θ i de $\rho^2(\Theta_0, \theta)$, però en molts casos, són funcions monòtones creixents de $\rho^2(\Theta_0, \theta)$, de manera que depenen de θ exclusivament a través de $\rho^2(\Theta_0, \theta)$. En aquest cas, es pot obtenir l'interval de confiança per a p a partir d'un interval de confiança per a la distància $\rho^2(\Theta_0, \theta)$ de la forma $[\rho_*^2, \infty)$, essent ρ_*^2 una estimació o estimació per defecte de l'anterior distància.

Aquesta estratègia és convenient donat que, freqüentment, existeix un estimador de $\rho^2(\Theta_0, \theta)$ tal que la seva distribució depèn només de $\rho^2(\Theta_0, \theta)$, com únic paràmetre desconegut, almenys asimptòticament. Així doncs, es pot reemplaçar $\rho^2(\Theta_0, \theta)$ per ρ_*^2 a les regles p o m .

procediment- b.

És una solució *compatible* amb l'aproximació Bayesiana. Suposem una *distribució a priori* no-informativa sobre Θ , en particular una *a priori* proporcional al volum Riemannià, tot i que es poden usar altres distribucions *a priori*, si hi ha bones raons per fer-ho. Podem considerar el promig de les quantitats (2.1) o (2.2) respecte aquesta *distribució a priori* no-informativa, en general impròpia, o respecte a la seva corresponent *distribució a posteriori*, si adicionalment tenim una mostra $x = (x_1, \dots, x_k) \in \chi^k$. En qualsevol cas, una condició requerida és que la distribució de probabilitat *a posteriori* convergeixi feblement a la verdadera mesura de probabilitat, per assegurar un comportament correcte del procediment-b.

Encara que sota aquest procediment es pot usar qualsevol estimador convenient, com el màxim-versemblant, donat que el quadrat de la distància de Rao és una funció de pèrdua *intrínseca* natural, veure

Oller i Corcuera [35], en aquest treball es farà servir l'estimador de Bayes (generalitzat) prenent com *a priori* (generalment impròpia) el volum Riemannià, que s'obtindrà minimitzant la funció $\nu \mapsto \mathcal{H}(\nu|x)$, on

$$\mathcal{H}(\nu|x) = \int_{\Theta} \rho^2(\theta, \nu) f(\theta|x) V(d\theta),$$

essent V el volum Riemannià, i $f(\theta|x)$ la densitat *a posteriori*. Les solucions al problema de minimització anterior són *els centres de massa* de la mesura "a posteriori" sobre Θ . Suposant algunes condicions sobre Θ i la mesura "a posteriori", que es donaran en la geometria dels models estadístics més usuals, el centre de masses existeix i està definit de manera única, veure com bibliografia: Emery i Mokobodzki [14] i Karcher [26], la qual cosa permet una definició simple de l'estimador de Bayes generalitzat:

$$x = (x_1, \dots, x_k) \mapsto \mathcal{U}_k(x) = \mathcal{M}_k(x)$$

on $\mathcal{M}_k(x)$ és **el centre de massa** de la distribució *a posteriori*, a partir d'una mostra de grandària k .

Els anteriors apartats suggereixen la possibilitat de combinar les regles **p** o **m** amb els procediments **f** i **b** obtenint quatre mètodes bàsics similars, cada un dependent de l'estimador \mathcal{U}_m seleccionat. Aquests mètodes s'anomenaran a partir d'ara de la següent manera: mètodes– **pf**, **pb**, **mf**, **mb**, els quals seran desenvolupats en alguns casos particulars.

Notem també que en el procediment–**f** el nivell α no juga el paper de una *probabilitat d'error condicionada* (com el nivell de significació d'un test), donat que controla les estimacions p_* o m_* per a **p** o **m**. De manera raonable α hauria de ser més gran que 0 i més petita o igual que 0.5, però probablement no sigui convenient triar valors tant baixos com 0.05, perquè a valors més petits per a α , valors més baixos per a p_* i m_* , obtenint així estimacions per defecte massa petites per a **p** i **m**.

2.2 Relació entre hipòtesis simples i compostes

En aquesta secció s'intentarà establir una relació entre els tests amb hipòtesi nul·la simple i els tests amb hipòtesi nul·la composta usant les regles **p** i **m**. Per fer-ho usarem les següents proposicions:

Proposició 2.2.1 Sigui $F : \mathbb{R} \rightarrow \mathbb{R}$ una funció monòtona creixent, considerem un conjunt real $A \subset \mathbb{R}$ i sigui $\beta = \inf(A)$. Aleshores

$$F(\beta) \leq \inf F(A)$$

Demostració:

Com $A \neq \emptyset$, doncs altrament no tindria ínfim, $\inf F(A)$ existeix.

Per altra banda, com F és creixent

$$\forall \theta \in A \implies F(\beta) \leq F(\theta)$$

aleshores $F(A)$ acotat inferiorment i té ínfim. Com $F(\beta)$ és una cota inferior de $F(A)$ resulta:

$$F(\beta) \leq \inf F(A)$$

■

Proposició 2.2.2 Sigui $F : \mathbb{R} \rightarrow \mathbb{R}$ una funció monòtona creixent i contínua per la dreta, considerem un conjunt real $A \subset \mathbb{R}$ i sigui $\beta = \inf(A)$. Aleshores

$$F(\beta) = \inf F(A)$$

Demostració:

Com β és l'ímfim d' A es compleix que $\beta \in A$ o bé $\beta \in A'$, on A' és el conjunt de punts d'acumulació d' A , doncs altrament no seria un ímfim.

Si $\beta \in A \implies \inf F(A) \leq F(\beta)$, i per la proposició anterior tindrem la igualtat.

Si $\beta \notin A \implies \beta \in A'$, llavors existiran punts d' A tant propers a β com es vulgui, és a dir:

$$\forall \delta > 0 \quad (\beta, \beta + \delta) \cap A \neq \emptyset$$

usant la continuïtat a la dreta de F tenim

$$\forall \varepsilon > 0 \quad \exists \delta_\varepsilon > 0 \quad \text{tal que si } \theta \in (\beta, \beta + \delta_\varepsilon) \cap A \text{ llavors } |F(\theta) - F(\beta)| < \varepsilon$$

usant el creixement de F tenim que

$$\forall \theta \in A \quad F(\beta) \leq F(\theta) \tag{2.3}$$

i per tant, resulta

$$\forall \varepsilon > 0 \exists \theta \in A : F(\theta) < F(\beta) + \varepsilon \quad (2.4)$$

aleshores, tenint en compte l'arbitrarietat de ε , amb (2.3) i (2.4) podem arribar a la conclusió que

$$F(\beta) = \inf F(A)$$

per la pròpia definició d'ínfim. ■

Veiem ara com usar aquestes proposicions per establir el lligam entre les hipòtesis simples i les compostes, en el nostre cas. Definirem el conjunt A com

$$A = \{\rho^2(\nu, \theta) : \nu \in \Theta_0\}$$

i

$$\beta = \inf A = \rho^2(\Theta_0, \theta)$$

Cas regla – p

Aquesta regla es pot escriure en aquest cas com:

$$p = P(\rho^2(\mathcal{U}_m(x), \theta) < \beta) - P(\rho^2(\mathcal{U}_m(x), \theta) > \beta)$$

En el cas que $\rho^2(\mathcal{U}_m(x), \theta)$ tingui una distribució absolutament contínua o, en el cas que no ho sigui però $P(\rho^2(\mathcal{U}_m(x), \theta) = \beta) = 0$, tindrem la següent equivalència

$$p > 0 \iff P(\rho^2(\mathcal{U}_m(x), \theta) \leq \beta) > 1/2$$

Aleshores si anomenem $F_{\rho^2(\mathcal{U}_m(x), \theta)}$ a la funció de distribució de $\rho^2(\mathcal{U}_m(x), \theta)$, tindrem que, en ser contínua per la dreta, per la Proposició 2.2.2

$$p > 0 \iff F_{\rho^2(\mathcal{U}_m(x), \theta)}(\beta) = \inf_{\nu \in \Theta_0} \{F_{\rho^2(\mathcal{U}_m(x), \theta)}(\rho^2(\nu, \theta))\} > 1/2$$

i tindrem

$$\forall \nu \in \Theta_0 \quad F_{\rho^2(\mathcal{U}_m(x), \theta)}(\rho^2(\nu, \theta)) \geq F_{\rho^2(\mathcal{U}_m(x), \theta)}(\beta) > 1/2$$

Llavors si rebutgem la hipòtesi nul·la composta $H_0 : \theta \in \Theta_0$, rebutjarem totes les hipòtesis nul·les simples $H_0 : \theta = \theta_0 \in \Theta_0$, o bé, si acceptem alguna hipòtesi nul·la simple, acceptarem la composta.

En el cas que $\rho^2(\mathcal{U}_m(x), \theta)$ tingui una distribució discreta amb $P(\rho^2(\mathcal{U}_m(x), \theta) = \beta) \neq 0$, la regla – p quedarà

$$\rho > 0 \iff P(\rho^2(\mathcal{U}_m(x), \theta) < \beta) > P(\rho^2(\mathcal{U}_m(x), \theta) > \beta)$$

Considerem, $\forall \alpha > 0$, la funció $P(\rho^2(\mathcal{U}_m(x), \theta) < \alpha)$, que sabem és una funció creixent respecte α , aleshores la funció $P(\rho^2(\mathcal{U}_m(x), \theta) > \alpha)$ serà decreixent. Anomenem

$$F(\alpha) = P(\rho^2(\mathcal{U}_m(x), \theta) < \alpha) - P(\rho^2(\mathcal{U}_m(x), \theta) > \alpha)$$

que serà una funció creixent respecte α . Llavors la regla de decisió anterior, recordant que $\beta = \rho^2(\Theta_0, \theta)$, es pot escriure

$$\rho > 0 \iff F(\beta) > 0$$

i per la Proposició 1.2.1, com F és una funció creixent, tindrem

$$F(\beta) = F(\inf_{\nu \in \Theta_0} \rho^2(\nu, \theta)) \leq \inf_{\nu \in \Theta_0} F(\rho^2(\nu, \theta)) \leq F(\rho^2(\nu, \theta)) \quad \forall \nu \in \Theta_0$$

Aleshores si rebutgem la hipòtesi nul·la composta $H_0 : \theta \in \Theta_0$, donat que $F(\beta) > 0$, rebutjarem totes les nul·les simples $H_0 : \theta = \theta_0 \in \Theta_0$, o bé, si acceptem alguna hipòtesi nul·la simple, acceptarem la composta.

Observem que el fet de rebutjar totes les hipòtesis nul·les simples no serà suficient per assegurar el rebuigi de la hipòtesi composta.

Cas regla – m

En aquest cas la regla de decisió es transforma en

$$m > 0 \iff \rho^2(\Theta_0, \theta) = \inf_{\nu \in \Theta_0} \rho^2(\nu, \theta) > E(\rho^2(\mathcal{U}_m(x), \theta))$$

Però sabem que

$$\forall \nu \in \Theta_0, \quad \rho^2(\nu, \theta) \geq \inf_{\nu \in \Theta_0} \rho^2(\nu, \theta)$$

Llavors també arribem a la conclusió que, si rebutgem la hipòtesi nul·la composta rebutjarem totes les simples, o bé, si acceptem alguna simple acceptarem la composta.

Hem vist, doncs, que les regles –p i m tenen un comportament raonable respecte la relació entre hipòtesis nul·les simples i compostes.

2.3 Distribucions asimptòtiques

Moltes vegades, a l'hora d'aplicar els mètodes definits anteriorment, sorgeixen dificultats a l'hora de treballar amb determinades distribucions estadístiques, per exemple a l'hora de construir regions de confiança, i pot ser necessari utilitzar resultats asimptòtics com els que veurem a continuació, alguns d'ells ja coneguts dintre de la literatura estadística, però que aquí es demostren detalladament i en el nostre context.

Introduïrem prèviament la definició de distribució normal sobre una varietat. Per a més detalls veure Jupp i Mardia [25]. Definim primer el que s'entén per distribució Normal d'un vector aleatori sobre un espai vectorial general, que en particular s'aplicarà a l'espai tangent.

Definició 2.3.1 *Sigui Z un vector aleatori en un espai vectorial E de dimensió n . Direm que Z està normalment distribuït si existeix un vector μ i un tensor 2 vegades contravariant definit positiu Σ tal que per a qualsevol base d'aquest espai vectorial, $(Z^i)_{n \times 1} \sim N((\mu^i)_{n \times 1}, (\Sigma^{ij})_{n \times n})$. Escrivem $Z \sim N(\mu, \Sigma)$.*

En aquest treball considerarem només objectes aleatoris Y que prendran valors, quasi segurament, en entorns regulars normals de qualsevol punt de la varietat geodèsicament completa Θ , veure les definicions a l'apèndix. Per aquest tipus d'objectes aleatoris el camp vectorial $\exp_\theta^{-1}(Z)$, per a $\theta \in \Theta$, suposarem que estarà, quasi segurament, ben definit.

Definició 2.3.2 *Sigui Y un objecte aleatori que pren valors en una varietat completa Θ . Direm que Y està distribuït normalment respecte θ i amb paràmetres (μ, Σ) , si existeix un vector aleatori $Z \in T_\theta(\Theta)$, amb $Z \sim N(\mu, \Sigma)$, tal que $Y = \exp_\theta(Z)$ i es notará $Y \sim N(\mu, \Sigma)_\theta$.*

Un resultat asimptòtic útil, pels estimadors eficients de primer ordre $\mathcal{U}_k(x)$ (com l'estimador màxim-versemblant), és el següent:

Teorema 2.3.1 *Sigui Θ una varietat Riemanniana d'un model estadístic, C^∞ , real, n -dimensional, connexa i geodèsicament completa. Sigui $\Theta_0 = \{\theta_0\}$ i $\mathcal{U}_k(x)$ una successió d'estimadors màxim-versemblants de $\theta \in \Theta$, on θ és el vertader valor del paràmetre. Aleshores*

a) Si $\theta_0 \neq \theta$

$$\sqrt{k} (\rho(\mathcal{U}_k(x), \theta_0) - \rho(\theta, \theta_0)) \xrightarrow{\mathcal{L}} Z \sim N(0, 1)$$

b) Si $\theta = \theta_0$

$$k\rho^2(\mathcal{U}_k, \theta) \xrightarrow{\mathcal{L}} W \sim \chi_n^2$$

on \mathcal{L} indica la convergència en llei, $N(0, 1)$, la distribució Normal estàndard univariant i χ_n^2 la distribució Chi-quadrat amb n graus de llibertat.

Demostració:

Si \mathcal{U}_k és una successió d'estimadors màxim-versemblants de θ , com a conseqüència directa de les propietats d'aquests estimadors, sabem que

$$\sqrt{k} \exp_{\theta}^{-1}(\mathcal{U}_k) \xrightarrow{\mathcal{L}} Y \sim N(0, G^{-1})$$

i llavors obtenim b)

$$k\rho^2(\mathcal{U}_k, \theta) = k \sum_{i,j=1}^n g_{ij}(\theta) (\exp_{\theta}^{-1}(\mathcal{U}_k))^i (\exp_{\theta}^{-1}(\mathcal{U}_k))^j \xrightarrow{\mathcal{L}} W \sim \chi_n^2$$

Notem que $G = (g_{ij})$ és el tensor mètric donat per la matriu d'informació de Fisher definit en el capítol anterior.

Podem considerar, ϕ , un sistema de coordenades geodèsiques polars entorn a θ i amb origen en θ_0 de manera que $\phi^1(\theta) = \rho(\theta, \theta_0)$, i $\phi^2(\theta), \dots, \phi^n(\theta)$ en un entorn de dimensió $n - 1$ de θ , és ben sabut que les components del tensor mètric satisfan

$$g_{11}(\theta) = \langle \partial_1, \partial_1 \rangle|_{\theta} = 1, \quad g_{1i} = g_{i1} = \langle \partial_1, \partial_i \rangle|_{\theta} = 0, \quad i = 2, \dots, n$$

Considerem un entorn normal U de θ , i donat que \exp_{θ} és un difeomorfisme local, sabem que existeix un entorn estrellat U^* del punt $0 \in T_{\theta}(\Theta)$ de dimensió n , tal que, fent un abús de notació i utilitzant un sistema de coordenades φ en $T_{\theta}(\Theta)$ que identifiqui els seus punts amb les seves coordenades a \mathbb{R}^n , és a dir, $\forall v \in U^*$, $\varphi(v) = (v^1, \dots, v^n)$, la següent composició està ben definida

$$\begin{array}{ccccccc} \varphi(U^*) \subset \mathbb{R}^n & \xrightarrow{\varphi^{-1}} & U^* \subset T_{\theta}(\Theta) & \xrightarrow{\exp_{\theta}} & U \subset \Theta & \xrightarrow{\rho_{\theta_0}} & \mathbb{R} \\ (v^1, \dots, v^n) & \longmapsto & v & \longmapsto & \xi = \exp_{\theta}(v) & \longmapsto & \rho_{\theta_0}(\xi) = \rho(\xi, \theta_0) \\ (0, \dots, 0) & \longmapsto & 0 & \longmapsto & \theta = \exp_{\theta}(0) & \longmapsto & \rho_{\theta_0}(\theta) = \rho(\theta, \theta_0) \end{array}$$

Notarem doncs $h = \rho_{\theta_0} \circ \exp_{\theta} \circ \varphi^{-1}$ a la funció real definida anteriorment i tal que

$$h(0, \dots, 0) = \rho_{\theta_0}(\exp_{\theta}(0)) = \rho(\theta, \theta_0)$$

$$h(\{(\exp_{\theta}^{-1}(\mathcal{U}_k))^i\}_{i=1}^n) = \rho_{\theta_0}(\exp_{\theta}(\exp_{\theta}^{-1}(\mathcal{U}_k))) = \rho(\mathcal{U}_k, \theta_0)$$

Sabem que $h'(0, \dots, 0) \neq (0, \dots, 0)$ doncs, amb el sistema de coordenades agafat

$$\frac{\partial h}{\partial u^1}(0, \dots, 0) = \frac{\partial}{\partial \phi^1}(\rho_{\theta_0}(\theta)) = 1$$

$$\frac{\partial h}{\partial u^i}(0, \dots, 0) = \frac{\partial}{\partial \phi^i}(\rho_{\theta_0}(\theta)) = 0 \quad i = 2, \dots, n$$

on u^i són les projeccions naturals a \mathbb{R}^n i llavors $(h'(0, \dots, 0))^t = (1, 0, \dots, 0)$. Aleshores segons les propietats asimptòtiques de les funcions de vectors normals, veure Serfling [45] per a més detalls,

$$\sqrt{k} [h(\{(\exp_{\theta}^{-1}(\mathcal{U}_k))^i\}_{i=1}^n) - h(0, \dots, 0)] \xrightarrow{\mathcal{L}} z \sim N(0, (h'(0))^t G^{-1}(h'(0)))$$

substituint la funció h pel seu valor, i donat que $(h'(0))^t G^{-1}(h'(0)) = 1$, ens queda

$$\sqrt{k} (\rho(\mathcal{U}_k(x), \theta_0) - \rho(\theta, \theta_0)) \xrightarrow{\mathcal{L}} Z \sim N(0, 1)$$

que és el que volíem veure en a). ■

Observem que aquest resultat asimptòtic ens servirà per trobar distribucions límit quan apliquem el mètode- \mathbf{pf} . Veurem, per exemple, la seva aplicació en capítol 4, on es farà un contrast en un model paramètric general amb hipòtesi nul·la simple del tipus $H_0 : \theta = \theta_0$, front l'alternativa $H_1 : \theta \neq \theta_0$.

Ens interessaria provar també el resultat asimptòtic anterior en un cas més general, $\dim(\Theta_0) = r < n$. És a dir, ens interessarà una generalització del Teorema 2.3.1 al cas d'hipòtesi nul·la simple r -dimensional. Però en aquest cas general la demostració es farà a partir d'un seguit de lemes usats per anar provant petites parts que finalment ens portaran al resultat desitjat. Els enunciats tenen un caràcter general que després s'aplicarà al nostre cas concret. Veiem doncs els lemes.

Lema 2.3.1 *Siguin f i h funcions contínues a valors reals definides en un entorn obert de $\theta \in \Theta$, on Θ és una varietat Riemanniana amb una mètrica associada (segons la definició que hem usat en el capítol 1 i a l'apèndix), de manera que aquestes funcions són iguals fins l'ordre 1 en θ , això és:*

$$\lim_{\xi \rightarrow \theta} \frac{f(\xi) - h(\xi)}{\rho(\xi, \theta)} = 0$$

Siguin $\{a_k\}_{k \in \mathbb{N}}$ una successió de nombres reals positius tals que compleixen $\lim_{k \rightarrow \infty} a_k = +\infty$. Sigui $\mathcal{U}_k : \chi^k \rightarrow \Theta$ una aplicació mesurable (respecte l'àlgebra de Borel en Θ) de manera que $a_k \rho(\mathcal{U}_k, \theta) \xrightarrow{\mathcal{L}} Y$ (on Y és una variable aleatòria de distribució coneguda arbitrària). Sigui $V_k = a_k f(\mathcal{U}_k)$ i $W_k = a_k h(\mathcal{U}_k)$, de manera que $V_k \xrightarrow{\mathcal{L}} Z$ (on Z és una altra variable aleatòria de distribució coneguda arbitrària). Aleshores

$$W_k \xrightarrow{\mathcal{L}} Z$$

Demostració:

Hem de veure doncs, que

$$|V_k - W_k| = |a_k f(\mathcal{U}_k) - a_k h(\mathcal{U}_k)| \xrightarrow{\mathcal{L}} 0$$

Multiplicant i dividint l'expressió anterior per $a_k \rho(\mathcal{U}_k, \theta)$ obtenim que hem de provar, de manera equivalent, que

$$\frac{|f(\mathcal{U}_k) - h(\mathcal{U}_k)|}{\rho(\mathcal{U}_k, \theta)} a_k \rho(\mathcal{U}_k, \theta) \xrightarrow{\mathcal{L}} 0$$

Sabem, per les hipòtesis del teorema, que

$$\lim_{\mathcal{U}_k \rightarrow \theta} \frac{|f(\mathcal{U}_k) - h(\mathcal{U}_k)|}{\rho(\mathcal{U}_k, \theta)} = 0$$

aleshores si definim la funció

$$F(\xi) = \begin{cases} \frac{|f(\xi) - h(\xi)|}{\rho(\xi, \theta)} & \text{si } \xi \neq \theta \\ 0 & \text{si } \xi = \theta \end{cases}$$

tenim que F és una funció contínua en \mathcal{U}_k i llavors com $\mathcal{U}_k \xrightarrow{\mathcal{P}} \theta$, podem assegurar que

$$F(\mathcal{U}_k) = \frac{|f(\mathcal{U}_k) - h(\mathcal{U}_k)|}{\rho(\mathcal{U}_k, \theta)} \xrightarrow{\mathcal{P}} F(\theta) = 0$$

Per altra banda sabem que

$$a_k \rho(\mathcal{U}_k, \theta) \xrightarrow{\mathcal{L}} Y$$

i, com aplicació directa també del Teorema de Slutsky, s'obté que

$$F(\mathcal{U}_k) a_k \rho(\mathcal{U}_k, \theta) = |V_k - W_k| \xrightarrow{\mathcal{L}} 0$$

■

Lema 2.3.2 *Sigui Θ modelada en \mathbb{R}^n amb dues estructures Riemannianes amb funcions distància ρ_1 i ρ_2 respectivament, tals que*

$$\lim_{(\nu, \xi) \rightarrow (\theta, \theta)} \frac{\rho_1(\nu, \xi) - \rho_2(\nu, \xi)}{\rho_1(\nu, \xi)} = 0$$

Sigui $\Theta_0 \subset \Theta$ una subvarietat tal que $\theta \in \Theta_0$. Notem que $\Theta \times \Theta$ té estructura de varietat i té sentit considerar el límit anterior. Definim

$$\rho_i(\nu, \Theta_0) = \inf\{\rho_i(\nu, \gamma), \gamma \in \Theta_0\} \quad i = 1, 2.$$

Aleshores

$$\lim_{\nu \rightarrow \theta} \frac{|\rho_1(\nu, \Theta_0) - \rho_2(\nu, \Theta_0)|}{\rho_1(\nu, \theta)} = 0$$

Demostració:

Siguin

$$B_i(\theta, \delta) \equiv \{\xi \in \Theta : \rho_i(\theta, \xi) < \delta\} \quad i = 1, 2 \quad \delta \geq 0$$

$$\overline{B}_i(\theta, \delta) \equiv \{\xi \in \Theta : \rho_i(\theta, \xi) \leq \delta\} \quad i = 1, 2 \quad \delta \geq 0$$

i

$$\Delta(x) \equiv \sup\{\rho_2(\theta, \gamma) : \gamma \in B_1(\theta, x)\} \quad x \geq 0$$

Cal observar que $\Delta(x) > 0$ si $x \neq 0$.

Tenint en compte que Θ és un espai mètric (tant amb ρ_1 com amb ρ_2 , mètriques Riemannianes equivalents des d'un punt de vista topològic), tindrem d'una forma natural una mètrica a $\Theta \times \Theta$ donada per

$$d((\nu, \xi), (\theta, \gamma)) \equiv \sqrt{\rho_1^2(\nu, \theta) + \rho_1^2(\xi, \gamma)}$$

Per tant, $\forall \varepsilon > 0 \quad \exists \delta' > 0$ tal que si $\rho_1(\nu, \theta) < \frac{\delta'}{\sqrt{2}}$, $\rho_1(\xi, \theta) < \frac{\delta'}{\sqrt{2}}$,
 $(\nu, \xi) \neq (\theta, \theta)$, $\nu \neq \xi$

es compleix $d((\nu, \xi), (\theta, \theta)) < \delta'$ i aleshores, per hipòtesi

$$\frac{|\rho_1(\nu, \xi) - \rho_2(\nu, \xi)|}{\rho_1(\nu, \xi)} < \frac{\varepsilon}{2} \tag{2.5}$$

Definim ara

$$W = B_1\left(\theta, \frac{\delta'}{4\sqrt{2}}\right) \cap B_2(\theta, \hat{\Delta}) \quad \text{amb} \quad \hat{\Delta} = \frac{\Delta\left(\frac{\delta'}{4\sqrt{2}}\right)}{2}$$

essent W un obert respecte les topologies associades ρ_1 ó ρ_2 (que, de fet, són la mateixa).

Per tant $\exists \delta > 0 : B_1(\theta, \delta) \subset W$.

Observem que si $\nu \in B_1(\theta, \delta)$, com $\theta \in \Theta_0$, tindrem

$$\rho_1(\nu, \Theta_0) < \delta \leq \frac{\delta'}{4\sqrt{2}}$$

i

$$\rho_1(\nu, \Theta_0) = \rho_1(\nu, \Theta_0 \cap B_1(\theta, 2\delta)) = \rho_1(\nu, \Theta_0 \cap B_1(\theta, \frac{\delta'}{2\sqrt{2}}))$$

per altra banda com $\nu \in B_1(\theta, \delta)$,

$$\nu \in B_2(\theta, \hat{\Delta}) \quad \text{i} \quad \rho_2(\nu, \Theta_0) \leq \frac{\Delta\left(\frac{\delta'}{4\sqrt{2}}\right)}{2}$$

i també

$$\rho_2(\nu, \Theta_0) = \rho_2(\nu, \Theta_0 \cap B_1(\theta, \frac{\delta'}{4\sqrt{2}})) = \rho_2(\nu, \Theta_0 \cap B_1(\theta, \frac{\delta'}{2\sqrt{2}}))$$

Aleshores si $\nu, \xi \in B_1(\theta, \frac{\delta'}{2\sqrt{2}})$, per (2.5), resulta que

$$\rho_1(\nu, \xi) < \rho_2(\nu, \xi) + \rho_1(\nu, \xi) \frac{\varepsilon}{2} \iff \rho_1(\nu, \xi) \left(1 - \frac{\varepsilon}{2}\right) < \rho_2(\nu, \xi)$$

també

$$\rho_1(\nu, \xi) > \rho_2(\nu, \xi) - \rho_1(\nu, \xi) \frac{\varepsilon}{2} \iff \rho_1(\nu, \xi) \left(1 + \frac{\varepsilon}{2}\right) > \rho_2(\nu, \xi)$$

així si $\nu \in B_1(\theta, \delta)$ llavors

$$\begin{aligned} \rho_2(\nu, \Theta_0) &= \rho_2(\nu, \Theta_0 \cap B_1(\theta, \frac{\delta'}{2\sqrt{2}})) \leq \rho_1(\nu, \Theta_0 \cap B_1(\theta, \frac{\delta'}{2\sqrt{2}})) \left(1 + \frac{\varepsilon}{2}\right) \\ &= \rho_1(\nu, \Theta_0) \left(1 + \frac{\varepsilon}{2}\right) \end{aligned}$$

$$\begin{aligned}\rho_2(\nu, \Theta_0) &= \rho_2(\nu, \Theta_0 \cap B_1(\theta, \frac{\delta'}{2\sqrt{2}})) \geq \rho_1(\nu, \Theta_0 \cap B_1(\theta, \frac{\delta'}{2\sqrt{2}}))(1 - \frac{\varepsilon}{2}) \\ &= \rho_1(\nu, \Theta_0)(1 - \frac{\varepsilon}{2})\end{aligned}$$

Cal observar que tant si $\varepsilon/2 < 1$ com si $\varepsilon/2 > 1$ la última desigualtat és certa.

Per tant:

$$|\rho_1(\nu, \Theta_0) - \rho_2(\nu, \Theta_0)| \leq \frac{\varepsilon}{2}\rho_1(\nu, \Theta_0) \leq \frac{\varepsilon}{2}\rho_1(\nu, \theta) < \varepsilon\rho_1(\nu, \theta)$$

Aleshores podem concloure que $\forall \varepsilon > 0 \exists \delta > 0$ tal que si $\rho_1(\nu, \theta) < \delta$ amb $\nu \neq \theta$, aleshores

$$\frac{|\rho_1(\nu, \Theta_0) - \rho_2(\nu, \Theta_0)|}{\rho_1(\nu, \theta)} < \varepsilon$$

això és:

$$\lim_{\nu \rightarrow \theta} \frac{|\rho_1(\nu, \Theta_0) - \rho_2(\nu, \Theta_0)|}{\rho_1(\nu, \theta)} = 0$$

■

Abans de continuar amb els lemes veurem dues observacions referents a la notació que s'utilitzarà a partir d'ara.

Observació 2.3.1 *En general, la notació*

$$f(t) = O(t^r) \quad \text{quan } t \rightarrow t_0$$

significarà que existeix una constant $C > 0$ tal que $|f(t)| \leq C|t^r|$ per a t suficientment proper a t_0 .

Observació 2.3.2 *Si ϕ és un sistema de coordenades en un entorn $U \subset \Theta$ del paràmetre θ , això és*

$$\begin{aligned}\phi: U &\longrightarrow \mathbb{R}^n \\ \theta &\longmapsto (\phi^1(\theta), \dots, \phi^n(\theta))\end{aligned}$$

Donats $\phi, \xi \in U$ definirem la següent norma

$$\|\phi(\theta) - \phi(\xi)\|^2 = \sum_{i,j=1}^n g_{ij}(\theta)(\phi^i(\theta) - \phi^i(\xi))(\phi^j(\theta) - \phi^j(\xi))$$

Continuem ara enunciant els lemes que necessitem per arribar al resultat general desitjat. De fet, els Lemes 2.3.3 fins al 2.3.7 ens donen les eines per a poder demostrar el resultat del Lema 2.3.8, que juntament amb el 2.3.1 i el 2.3.2, són els que s'usaran en el desenvolupament de la demostració del teorema final que ens dóna el resultat asimptòtic al qual volem arribar.

Lema 2.3.3 *Sigui Θ una varietat Riemanniana amb les hipòtesis del Lema 2.3.2. Sigui $\theta \in \Theta$, i anomenem ϕ a un sistema de coordenades al voltant de θ . Llavors si $\xi \rightarrow \theta$ tenim*

$$\rho^2(\theta, \xi) = \|\phi(\theta) - \phi(\xi)\|^2 + O(\rho^3(\theta, \xi))$$

Demostració:

Tenint en compte que sempre existeix una geodèsica minimal, a nivell local, per unir dos punts prou propers de manera que la distància Riemanniana entre els extrems coincideix amb la longitud d'arc, i si ϕ és un sistema de coordenades al voltant de θ , anomenem σ a la geodèsica minimal que uneix θ amb ξ parametritzada per la longitud d'arc, s , de manera que $\sigma(0) = \theta$, $\sigma(t) = \xi$, per a més detalls veure l'apèndix. Així tindrem

$$(\phi \circ \sigma)(s) = \phi(\theta) + v s + O(s^2)$$

on $v = (\phi \circ \sigma)'(0)$ és el vector tangent a l'origen, respecte del sistema de coordenades ϕ , per tant

$$\begin{aligned} \phi(\xi) - \phi(\theta) &= v t + O(t^2) \\ \|\phi(\xi) - \phi(\theta)\|^2 &= \|v\|^2 t^2 + O(t^3) \end{aligned}$$

però $\|v\| = 1$ en ser la geodèsica parametritzada per la longitud d'arc, $\|v\|$ és la norma en el espai tangent $T_\theta(\Theta)$ definida amb més detall a l'apèndix, i $t = \rho(\theta, \xi)$ si ξ és prou proper a θ . D'aquí obtenim

$$\rho^2(\theta, \xi) = \|\phi(\theta) - \phi(\xi)\|^2 + O(\rho^3(\theta, \xi))$$

■

Lema 2.3.4 *Sigui Θ una varietat Riemanniana amb les hipòtesis del lemes anteriors. Sigui σ una corba de classe C^2 , almenys, parametritzada per la longitud d'arc, de manera que $\sigma(0) = \theta \in \Theta$. Llavors*

$$\lim_{t \rightarrow 0} \frac{\rho(\theta, \sigma(t))}{t} = 1$$

Demostració:

Definim en primer lloc el camp tangent a la corba en cada punt t com $\sigma'(t) = d\alpha\left(\frac{d}{du}\Big|_t\right)$, on u és un sistema de coordenades consistent en la funció identitat sobre els reals, veure l'Apèndix. Observem que aquest camp és unitari en ser la corba parametritzada per la longitud d'arc. Per altra banda, si ϕ és un sistema de coordenades al voltant de θ (i.e.: $\phi : W \rightarrow E$, ϕ homeomorfisme, $E = \mathbb{R}^n$ ó $E = \text{Espai de Hilbert}$, $\theta \in W \subset \Theta$, W obert), tindrem pel Lema 2.3.3, que per a $t \rightarrow 0$

$$\rho^2(\theta, \sigma(t)) = \|\phi(\theta) - \phi(\sigma(t))\|^2 + O(\rho^3(\theta, \sigma(t)))$$

però

$$\phi(\sigma(t)) = \phi(\theta) + v t + O(t^2)$$

essent v el vector tangent en el sistema de coordenades ϕ . Per tant i com també hem vist en el lema anterior

$$\rho^2(\theta, \sigma(t)) = \|v\|^2 t^2 + O(t^3)$$

doncs $0 \leq \rho(\theta, \sigma(t)) \leq t$, i donat que $|v| = 1$, tenim finalment

$$\rho^2(\theta, \sigma(t)) = t^2 + O(t^3)$$

i

$$\lim_{t \rightarrow 0} \frac{\rho(\theta, \sigma(t))}{t} = 1$$

■

Lema 2.3.5 *Sigui Θ una varietat Riemanniana amb les hipòtesis dels lemes anteriors. Sigui Θ_0 i Θ_1 dues subvarietats de Θ amb $\theta \in \Theta_0 \cap \Theta_1 \subset \Theta$, amb immersió almenys de classe \mathcal{C}^2 . Sigui σ i $\bar{\sigma}$ dues corbes també almenys de classe \mathcal{C}^2 en Θ_0 i Θ_1 respectivament, amb $\sigma(0) = \bar{\sigma}(0) = \theta$ i $\sigma'(0) = \bar{\sigma}'(0)$, és a dir, amb els mateixos vectors tangents a l'origen. Aleshores*

$$\begin{aligned} i) \quad & \lim_{t \rightarrow 0} \frac{\rho(\sigma(t), \bar{\sigma}(t))}{t} = 0 \\ ii) \quad & \lim_{t \rightarrow 0} \frac{\rho(\sigma(t), \bar{\sigma}(t))}{\rho(\sigma(t), \theta)} = \lim_{t \rightarrow 0} \frac{\rho(\sigma(t), \bar{\sigma}(t))}{\rho(\bar{\sigma}(t), \theta)} = 0 \end{aligned}$$

Demostració:

Considerem un sistema de coordenades al voltant de θ i definim $u = \phi \circ \sigma$ i $\bar{u} = \phi \circ \bar{\sigma}$ en un entorn del punt $t = 0$. Si v és el vector tangent a l'origen en aquest sistema de coordenades tenim

$$\begin{aligned} u(t) &= \phi(\theta) + vt + O(t^2) \\ \bar{u}(t) &= \phi(\theta) + vt + O(t^2) \end{aligned}$$

i segons el Lema 2.3.3

$$\rho^2(\sigma(t), \bar{\sigma}(t)) = \|u(t) - \bar{u}(t)\|^2 + O(\rho^3(\sigma(t), \bar{\sigma}(t)))$$

però $u(t) - \bar{u}(t) = O(t^2)$ i per la desigualtat triangular $\rho(\sigma(t), \bar{\sigma}(t)) \leq 2t$, per tant

$$\rho^2(\sigma(t), \bar{\sigma}(t)) = O(t^3)$$

i

$$\rho(\sigma(t), \bar{\sigma}(t)) = O(t^{3/2})$$

dividint per t i fent límits obtenim l'apartat *i*).

Tenint ara en compte el lema anterior

$$\lim_{t \rightarrow 0} \frac{\rho(\theta, \sigma(t))}{t} = \lim_{t \rightarrow 0} \frac{\rho(\theta, \bar{\sigma}(t))}{t} = 1$$

i segons l'apartat *i*), obtenim finalment

$$\lim_{t \rightarrow 0} \frac{\rho(\sigma(t), \bar{\sigma}(t))}{\rho(\sigma(t), \theta)} = \lim_{t \rightarrow 0} \frac{\rho(\sigma(t), \bar{\sigma}(t))}{t} \frac{t}{\rho(\sigma(t), \theta)} = 0$$

que és el primer límit de l'apartat *ii*), el segon es fa de manera idèntica. ■

Lema 2.3.6 *Sigui Θ una varietat Riemanniana amb les hipòtesis dels lemes anteriors. Sigui Θ_0 una subvarietat de Θ amb $\theta \in \Theta_0 \subset \Theta$. Llavors $\forall \varepsilon > 0 \exists \delta > 0$ tals que*

$$B_{\Theta}(\theta, \delta) \cap \Theta_0 \subset B_{\Theta_0}(\theta, \varepsilon)$$

on B_{Θ} i B_{Θ_0} són boles Riemannianes obertes en Θ i Θ_0 respectivament.

Demostració:

Segons la definició de subvarietat adoptada, veure O'Neill [36], podem assegurar que la inclusió $j : \Theta_0 \longrightarrow j(\Theta_0)$ és un difeomorfisme.

Per tant, donat un obert en Θ_0 que contingui a θ , diguem-li W , existeix un obert de $j(\Theta_0)$, anomenem-lo U , tal que $j^{-1}(U) \subset W$. Però U és de la forma $U = V \cup \Theta_0$, amb V un obert de Θ .

Finalment és suficient considerar que la topologia en Θ i Θ_0 està generada per les mètriques Riemannianes d'aquestes dues varietats, on la mètrica en Θ_0 és induïda per la de Θ . ■

Lema 2.3.7 *Sigui Θ amb les mateixes hipòtesis dels lemes anteriors. Siguin Θ_0 i Θ_1 dues subvarietats de Θ tals que $\theta \in \Theta_0 \cap \Theta_1 \subset \Theta$, de la mateixa dimensió i tangents en el punt θ (i.e.: donada una corba en Θ_0 i una en Θ_1 que passin per θ , els seus vectors tangents en aquest punt estan en el mateix subespai vectorial de Θ_θ). Llavors existeix un obert en Θ , diem-li U , amb $\theta \in U$, i una aplicació $\tilde{f} : U \cap \Theta_0 \longrightarrow \Theta_1$ tal que*

$$\lim_{\xi \rightarrow \theta} \frac{\rho(\xi, \tilde{f}(\xi))}{\rho(\theta, \xi)} = 0$$

on ρ és, com fins ara, la mètrica Riemanniana en Θ .

Demostració:

Observem, en primer lloc, que Θ_0 i Θ_1 tenen estructura de varietat Riemanniana induïda per la de Θ . Per tant existirà un entorn obert en Θ_0 del punt θ , geodèsicament convex, que serà de la forma $U \cap \Theta_0$, (U obert de Θ), en ser la topologia de Θ_0 la induïda per la de Θ .

Sigui $\xi \in U \cap \Theta_0$, existeix una geodèsica de Θ_0 minimal parametrizada per la longitud d'arc de manera que $\sigma(0) = \theta$ i $\sigma(\tilde{t}) = \xi$.

Llavors afegim una corba $\bar{\sigma}$, també parametrizada per la longitud d'arc, en Θ_1 de manera que $\bar{\sigma}(0) = \theta$ i $\sigma'(0) = \bar{\sigma}'(0)$.

Sigui $\tilde{\xi} = \bar{\sigma}(\tilde{t})$. Aquest punt està en Θ_1 i està ben definit en fixar $\bar{\sigma}$ (que pot ser una geodèsica de Θ_1), sempre que agafem l'entorn $U \cap \Theta_0$ prou petit. Per tant podem construir una aplicació

$$\begin{aligned} \tilde{f} : U \cap \Theta_0 &\longrightarrow \Theta_1 \\ \xi &\longrightarrow \tilde{f}(\xi) = \tilde{\xi} \end{aligned}$$

Per altra banda, pel Lema 2.3.5

$$\lim_{\tilde{t} \rightarrow 0} \frac{\rho(\sigma(\tilde{t}), \bar{\sigma}(\tilde{t}))}{\rho(\sigma(\tilde{t}), \theta)} = 0$$

però aquest límit coincideix amb

$$\lim_{\xi \rightarrow \theta} \frac{\rho(\xi, \tilde{f}(\xi))}{\rho(\theta, \xi)} = 0$$

■

Lema 2.3.8 *Amb les mateixes hipòtesis del lema anterior*

$$\lim_{\xi \rightarrow \theta} \frac{\rho(\xi, \Theta_0) - \rho(\xi, \Theta_1)}{\rho(\xi, \theta)} = 0$$

on

$$\rho(\xi, \Theta_0) = \inf\{\rho(\xi, \gamma), \gamma \in \Theta_0\}$$

$$\rho(\xi, \Theta_1) = \inf\{\rho(\xi, \gamma), \gamma \in \Theta_1\}$$

Demostració:

Segons el Lema 2.3.7 existeix una bola en Θ , centrada en θ , de radi δ i inclosa en un obert U , de manera que si $\xi \in B(\theta, \delta/2)$, llavors al tenir

$$\rho(\xi, \Theta_0) = \rho(\xi, \Theta_0 \cap B(\theta, \delta))$$

$$\exists \gamma \in U \cap \Theta_0 : |\rho(\xi, \Theta_0) - \rho(\xi, \gamma)| < \varepsilon \rho(\theta, \xi)$$

amb ε arbitrari i $\rho(\theta, \xi)$ fixat.

Però, també pel 2.3.7, existirà un $\tilde{\gamma} \in \Theta_1$ de manera que $\rho(\gamma, \tilde{\gamma}) < \varepsilon \rho(\gamma, \theta)$, per tant

$$\rho(\xi, \Theta_1) \leq \rho(\xi, \gamma) + \rho(\gamma, \tilde{\gamma}) \leq \rho(\xi, \Theta_0) + \varepsilon \rho(\theta, \xi) + \varepsilon \rho(\gamma, \theta)$$

però $\rho(\gamma, \theta) \leq \rho(\gamma, \xi) + \rho(\xi, \theta)$, i sempre podem agafar γ de manera que $\rho(\gamma, \xi) \leq \rho(\xi, \theta)$, aleshores tindrem

$$\rho(\xi, \Theta_1) - \rho(\xi, \Theta_0) \leq 3\varepsilon\rho(\theta, \xi)$$

Anàlogament es prova

$$\rho(\xi, \Theta_0) - \rho(\xi, \Theta_1) \leq 3\varepsilon\rho(\theta, \xi)$$

i podem concloure amb

$$|\rho(\xi, \Theta_0) - \rho(\xi, \Theta_1)| \leq 3\varepsilon\rho(\theta, \xi)$$

sempre que $\xi \in B(\theta, \delta/2)$, provant així el lema. ■

Finalment, encadenarem tots aquests lemes per demostrar una generalització del Teorema 2.3.1 al cas r - dimensional.

Teorema 2.3.2 *Sigui Θ una varietat Riemanniana, C^∞ , real, n -dimensional, connexa i geodèsicament completa. Sigui Θ_0 una subvarietat de Θ r -dimensional, ($r < n$). Considerem la distància Riemanniana, ρ , associada a la varietat i definim*

$$\rho(\nu, \Theta_0) = \inf \{ \rho(\nu, \xi), \xi \in \Theta_0 \}$$

Sigui, a més, $\mathcal{U}_k(x)$ una successió de punts de Θ tal que:

$$\sqrt{k} \exp_\theta^{-1}(\mathcal{U}_k(x)) \xrightarrow{\mathcal{L}} Y \sim N(0, G^{-1})$$

aleshores, sota la hipòtesi nul·la $\theta \in \Theta_0$:

$$k\rho^2(\mathcal{U}_k(x), \Theta_0) \xrightarrow{\mathcal{L}} \mathcal{Z} \sim \chi_{n-r}^2$$

Demostració:

Per hipòtesi, $\theta \in \Theta_0$, i aleshores $\rho(\theta, \Theta_0) = 0$.

Considerem \mathcal{U}_k la successió d'estimadors màxim-versemblants que, sabem, convergeixen en probabilitat al vertader valor del paràmetre $\theta \in \Theta_0$. És a dir, sabem que

$$\mathcal{U}_k \xrightarrow{P} \theta$$

i per tant podem assegurar que $\forall \varepsilon > 0$, existirà un valor k_0 a partir del qual, si $k > k_0$, llavors $\rho(\mathcal{U}_k, \theta) < \varepsilon$.

Donats dos punts $v, w \in T_\theta(\Theta)$, podem considerar $d_\theta(v, w)$ la distància associada a la norma $\|\cdot\|_\theta$ en el espai tangent $T_\theta(\Theta)$, definida anteriorment en el capítol 1, llavors és conegut el següent resultat

$$\lim_{(v,w) \rightarrow (0,0)} \frac{d_\theta(v, w)}{\rho(\exp_\theta(v), \exp_\theta(w))} = \frac{\|v - w\|_\theta}{\rho(\exp_\theta(v), \exp_\theta(w))} = 1$$

es pot veure Helgasson [22], entre altres, per a més detalls.

Anomenarem $\xi = \exp_\theta(v)$, $\nu = \exp_\theta(w)$ i com $\theta = \exp_\theta(0)$, el límit anterior es pot escriure

$$\lim_{(\xi, \nu) \rightarrow (\theta, \theta)} \frac{\|\exp_\theta^{-1}(\xi) - \exp_\theta^{-1}(\nu)\|_\theta - \rho(\xi, \nu)}{\rho(\xi, \nu)} = 0$$

Per la Proposició 9.8.2 de l'apèndix, per a cada punt $\theta \in \Theta$ existeix un entorn obert U^* de $0 \in T_\theta(\Theta)$ tal que \exp_θ és un difeomorfisme sobre $U = \exp_\theta(U^*)$ entorn obert de $\theta \in \Theta$.

Notem que el límit anterior es pot escriure també

$$\lim_{(\xi, \nu) \rightarrow (\theta, \theta)} \frac{d_\theta(\exp_\theta^{-1}(\xi), \exp_\theta^{-1}(\nu)) - \rho(\xi, \nu)}{\rho(\xi, \nu)} = 0$$

Aleshores, si ens restringim al subconjunt obert $U \subset \Theta$, on $\exp^{-1}|_U$ és un difeomorfisme, la següent distància d sobre Θ estarà ben definida:

$$d(\xi, \nu) = d_\theta(\exp_\theta^{-1}(\xi), \exp_\theta^{-1}(\nu))$$

Llavors, com $\theta \in U \cap \Theta_0$, aplicant el Lema 2.3.2 a $\tilde{\Theta}_0 = U \cap \Theta_0$, subvarietat de dimensió r de Θ , tindrem que

$$\lim_{\xi \rightarrow \theta} \frac{|d(\xi, \tilde{\Theta}_0) - \rho(\xi, \tilde{\Theta}_0)|}{\rho(\xi, \theta)} = \lim_{\xi \rightarrow \theta} \frac{|d_\theta(\exp_\theta^{-1}(\xi), \exp_\theta^{-1}(\tilde{\Theta}_0)) - \rho(\xi, \tilde{\Theta}_0)|}{\rho(\xi, \theta)} = 0$$

Sigui doncs $\exp_\theta^{-1}(\tilde{\Theta}_0)$ una subvarietat r -dimensional de $T_\theta(\Theta)$ que conté el punt $0 = \exp_\theta^{-1}(\theta)$. Definim Θ_1 com la subvarietat Euclidiana de $T_\theta(\Theta)$ tangent a $\exp_\theta^{-1}(\tilde{\Theta}_0)$ en el punt 0 i difeomòrfica a $T_0(\exp_\theta^{-1}(\tilde{\Theta}_0))$. És a dir Θ_1 i $\exp_\theta^{-1}(\tilde{\Theta}_0)$ són subvarietats de la mateixa dimensió r i tangents en el punt 0 . Així aplicant el Lema 2.3.8 tenim

$$\lim_{\xi \rightarrow \theta} \frac{d_\theta(\exp_\theta^{-1}(\xi), \exp_\theta^{-1}(\tilde{\Theta}_0)) - d_\theta(\exp_\theta^{-1}(\xi), \Theta_1)}{d_\theta(\exp_\theta^{-1}(\xi), 0)} = 0$$

Com que $d_\theta(\exp_\theta^{-1}(\xi), 0) = \rho(\xi, \theta)$, i tenint en compte que $\rho(\xi, \tilde{\Theta}_0) = \rho(\xi, \Theta_0)$, per a punts ξ prou propers a θ , restant els dos últims límits s'obté

$$\lim_{\xi \rightarrow \theta} \frac{\rho(\xi, \Theta_0) - d_\theta(\exp_\theta^{-1}(\xi), \Theta_1)}{\rho(\xi, \theta)} = 0$$

Aleshores, com U és un conjunt obert, sabem que $\exists \varepsilon > 0$ tal que $B(\theta, \varepsilon) \subset U$, i donat que $\mathcal{U}_k \xrightarrow{\mathcal{P}} \theta$, tal i com hem vist a l'inici d'aquesta demostració, $\exists k_0$ a partir del qual $\rho(\mathcal{U}_k, \theta) < \varepsilon$, amb la qual cosa podem assegurar que almenys $\forall k > k_0$, $\mathcal{U}_k \in B(\theta, \varepsilon) \subset U$. Llavors, aplicant aquest darrer límit a \mathcal{U}_k , obtenim

$$\frac{\rho(\mathcal{U}_k, \Theta_0) - d_\theta(\exp_\theta^{-1}(\mathcal{U}_k), \Theta_1)}{\rho(\mathcal{U}_k, \theta)} \xrightarrow{\mathcal{P}} 0$$

Si anomenem $\pi_{\Theta_1}(\exp_\theta^{-1}(\mathcal{U}_k))$ a la projecció ortogonal dels punts $\exp_\theta^{-1}(\mathcal{U}_k)$ sobre Θ_1 en el tangent $T_\theta(\Theta)$, podem escriure

$$d_\theta^2(\exp_\theta^{-1}(\mathcal{U}_k), \Theta_1) = \|\exp_\theta^{-1}(\mathcal{U}_k) - \pi_{\Theta_1}(\exp_\theta^{-1}(\mathcal{U}_k))\|_\theta^2$$

Donades les propietats dels estimadors màxim-versemblants i per ortogonalitat, sabem que

$$k d_\theta^2(\exp_\theta^{-1}(\mathcal{U}_k), \Theta_1) = k \|\exp_\theta^{-1}(\mathcal{U}_k) - \pi_{\Theta_1}(\exp_\theta^{-1}(\mathcal{U}_k))\|_\theta^2 \xrightarrow{\mathcal{L}} \mathcal{Z} \sim \chi_{n-r}^2$$

Llavors, si anomenem

$$W_k = a_k h(\mathcal{U}_k) = \sqrt{k} \rho(\mathcal{U}_k, \Theta_0)$$

$$V_k = a_k f(\mathcal{U}_k) = \sqrt{k} d_\theta(\exp_\theta^{-1}(\mathcal{U}_k), \Theta_1)$$

com que la funció $\sqrt{\cdot}$ és una funció contínua, sabem que

$$V_k = \sqrt{k} d_\theta(\exp_\theta^{-1}(\mathcal{U}_k), \Theta_1) \xrightarrow{\mathcal{L}} \sqrt{\mathcal{Z}}$$

per les propietats dels estimadors màxim-versemblants, sabem també que:

$$k \rho^2(\mathcal{U}_k, \theta) \xrightarrow{\mathcal{L}} \mathcal{Y} \sim \chi_n^2$$

utilitzant, altra vegada, que la funció $\sqrt{\cdot}$ és contínua, obtenim

$$\sqrt{k} \rho(\mathcal{U}_k, \theta) \xrightarrow{\mathcal{L}} \sqrt{\mathcal{Y}}$$

Aleshores, com aplicació directa del Lema 2.3.1, es compleix que W_k té la mateixa convergència en llei que V_k , és a dir

$$W_k = a_k h(\mathcal{U}_k) = \sqrt{k} \rho(\mathcal{U}_k, \Theta_0) \xrightarrow{\mathcal{L}} \sqrt{\mathcal{Z}}$$

i elevant al quadrat, tenint en compte que $(\cdot)^2$ és una funció contínua, arribem a la conclusió

$$k \rho^2(\mathcal{U}_k, \Theta_0) \xrightarrow{\mathcal{L}} \mathcal{Z} \sim \chi_{n-r}^2$$

que és el que volíem veure. ■

Capítol 3

Model Normal

3.1 Model normal n -variant amb Σ coneguda

Considerem la distribució normal n -variant amb $\Sigma = \Sigma_0$ una matriu $n \times n$ definida positiva, simètrica i coneguda, $X \sim N_n(\mu, \Sigma_0)$, i plantegem el contrast, o selecció de model, que ve donat per

$$H_0 : H\mu + b = 0 \quad H_1 : H\mu + b \neq 0$$

on H és una matriu $r \times n$ amb $0 < \text{rank } H = r \leq n$ i b és un vector $r \times 1$.

Observem, en particular, el cas d'hipòtesi nul·la simple:

$$H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0$$

obtingut si agafem $H = I$, $b = -\mu_0$, i aleshores $r = n$.

L'espai total de paràmetres (model) serà $\Theta = \mathbb{R}^n$, i l'espai de paràmetres corresponent a la hipòtesi nul·la (submodel) $\Theta_0 = \{\mu \in \mathbb{R}^n : H\mu + b = 0\}$ que és una varietat lineal $(n - r)$ -dimensional.

La mètrica informativa ve donada, en notació matricial, per:

$$ds^2 = d\mu' \Sigma_0^{-1} d\mu,$$

la geometria és euclidiana, la distància de Rao és en aquest cas la *distància de Mahalanobis*, el quadrat de la qual és

$$\rho^2(\mu_1, \mu_2) = (\mu_1 - \mu_2)' \Sigma_0^{-1} (\mu_1 - \mu_2)$$

i el volum Riemannià és

$$dV = (\det \Sigma_0)^{-\frac{1}{2}} d\mu$$

Donada una mostra de grandària m , $x = (x_1, \dots, x_m) \in (\mathbb{R}^n)^m$, un estimador natural per a aquest model és l'estimador màxim-versemblant,

$$\mathcal{U}_m(x) = \bar{x}_m = \frac{1}{m}(x_1 + \dots + x_m)$$

Observem que aquest estimador és també el *centre de masses* de la mesura de probabilitat *a posteriori*, basada en una *a priori no informativa* proporcional al volum Riemannià, que és, donada una mostra de mida m , $x = (x_1, \dots, x_m) \in (\mathbb{R}^n)^m$,

$$\mu \sim N_n(\bar{x}_m, \frac{1}{m} \Sigma_0), \quad (3.1)$$

donat que en el cas euclidià, amb una parametrització convenient, el centre de masses és just l'esperança.

3.1.1 mètode – p f

Observem, primerament, que en el cas que $\rho^2(\mathcal{U}_m(x), \theta)$, o qualsevol transformació monòtona, tingui una distribució absolutament contínua podem escriure

$$P(\rho^2(\mathcal{U}_m(x), \theta) > \rho^2(\Theta_0, \theta)) = 1 - (P(\rho^2(\mathcal{U}_m(x), \theta) < \rho^2(\Theta_0, \theta)))$$

i per tant

$$p = 2P(\rho^2(\mathcal{U}_m(x), \theta) < \rho^2(\Theta_0, \theta)) - 1$$

Així doncs en el nostre cas particular, $\theta = \mu$, tenim que p és una funció monòtona creixent de $\rho^2(\Theta_0, \mu)$, i llavors podem obtenir un interval de confiança per a p a partir d'un interval de confiança per a $\rho^2(\Theta_0, \mu)$ de la forma $[\rho_*^2, \infty)$, essent ρ_*^2 una estimació per defecte de $\rho^2(\Theta_0, \mu)$.

Hem de determinar, doncs, donada una mostra de mida k , $x = (x_1, \dots, x_k) \in (\mathbb{R}^n)^k$ un $(1 - \alpha)$ -interval de confiança per a p , de la forma $[p_*, 1]$. Per aconseguir aquest propòsit, notem que, si prenem $\bar{X}_k \equiv \mathcal{U}_k(\cdot)$,

$$k \rho^2(\bar{X}_k, \mu) = k (\bar{X}_k - \mu)' \Sigma_0^{-1} (\bar{X}_k - \mu)$$

té una distribució χ^2 ordinària amb n -graus de llibertat.

Per obtenir aquest interval, notem que la geometria corresponent al model és euclidiana, i per tant el quadrat de la distància entre $\eta \in \Theta$ i Θ_0 ve donat per,

$$\rho^2(\Theta_0, \eta) = (\eta + H^{-1}b)' H' (H \Sigma_0 H')^{-1} H (\eta + H^{-1}b)$$

independentment de la matriu g -inversa, H^{-1} , escollida. Així doncs, el quadrat de la distància des del vertader valor de μ fins a Θ_0 , i des del valor estimat a partir de \bar{x}_k fins a Θ_0 vindrà donat, aplicant la darrera fórmula, per $\rho^2(\Theta_0, \mu)$ i $\rho^2(\Theta_0, \bar{x}_k)$ respectivament.

A més, observem que $W = \Sigma_0^{-1/2} H' (H \Sigma_0 H')^{-1} H \Sigma_0^{-1/2}$ és l'operador projecció, d'acord amb la mètrica informativa, sobre el subespai ortogonal de $\{\eta \in \mathbb{R}^n : H\eta = 0\}$, aquest és el subespai director corresponent a Θ_0 . Per tant W és una matriu simètrica idempotent de rang r , i tenint en compte que l'estadístic

$$\sqrt{k} \Sigma_0^{-1/2} (\bar{X}_k + H^{-1}b) \sim N_n(\sqrt{k} \Sigma_0^{-1/2} (\mu + H^{-1}b), I)$$

es dedueix que

$$\begin{aligned} k \rho^2(\Theta_0, \bar{X}_k) &= k (\bar{X}_k + H^{-1}b)' H' (H \Sigma_0 H')^{-1} H (\bar{X}_k + H^{-1}b) \\ &= k (\bar{X}_k + H^{-1}b)' \Sigma_0^{-1/2} W \Sigma_0^{-1/2} (\bar{X}_k + H^{-1}b) \end{aligned}$$

té una distribució χ^2 no-central, $\chi_r^2(\delta)$, amb r graus de llibertat paràmetre de no-centralitat $\delta = k \rho^2(\Theta_0, \mu)$, veure Muirhead [31] [pp 26–32].

Així doncs, si definim $\gamma_{r,\alpha}(\delta)$ a partir de

$$P(\chi_r^2(\delta) \leq \gamma_{r,\alpha}(\delta)) = 1 - \alpha$$

on $\gamma_{r,\alpha}(\delta)$ és, clarament, una funció monòtona estrictament creixent de δ , i donada $\rho^2(\Theta_0, \bar{x}_k)$, el $(1 - \alpha)$ -interval de confiança per a $\rho^2(\Theta_0, \mu)$ serà $[\rho_*^2, \infty)$ amb

$$\rho_*^2 = \begin{cases} \frac{\gamma_{r,\alpha}^{-1}(k \rho^2(\Theta_0, \bar{x}_k))}{k} & \text{si } k \rho^2(\Theta_0, \bar{x}_k) > \gamma_{r,\alpha}(0) \\ 0 & \text{altrament} \end{cases}$$

Aleshores hem d'avaluar

$$p_* = 2P(\rho^2(\bar{X}_m, \mu) < \rho_*^2) - 1 = 2P(m \rho^2(\bar{X}_m, \mu) < m \rho_*^2) - 1$$

a partir de les taules de la distribució χ_n^2 i, finalment, acceptarem H_0 si p_* , estimació per defecte de p , és menor o igual a zero, o el que és el mateix, si $P(m \rho^2(\bar{X}_m, \mu) < m \rho_*^2) \leq 1/2$. Observem que si considerem ξ_n la mediana de una distribució χ_n^2 , rebutjarem el submodel si

$$m\rho_*^2 = \frac{m}{k} \gamma_{r,\alpha}^{-1}(k\rho^2(\Theta_0, \bar{x}_k)) > \xi_n$$

acceptant-lo altrament.

Notem, també que, si comparem aquest resultat amb el clàssic i habitual test d'hipòtesis, el qual sabem que rebutjarà la hipòtesi nul·la, H_0 , per a valors grans de $k\rho^2(\Theta_0, \bar{X}_k)$, obtindrem doncs que, si em acceptéssim en el cas clàssic H_0 amb un nivell de significació $\varepsilon = \alpha$, aleshores acceptarem també la hipòtesi nul·la amb el mètode-**pf**, per diferents raons, ja que en aquest cas $k\rho^2(\Theta_0, \bar{x}_k) \leq \gamma_{r,\alpha}(0)$, per tant $\rho_*^2 = 0$ i llavors $P(\rho^2(\bar{X}_m, \mu) < \rho_*^2) = 0$, obtenint $p^* = -1 < 0$. Però, si amb el test clàssic rebutgem H_0 , aleshores amb el mètode-**pf** també, sempre que es doni

$$2P(\chi_n^2 < m\rho_*^2) - 1 > 0$$

i encara acceptarem H_0 si l'anterior quantitat és menor o igual que zero, la qual cosa depèn del valor de m : valors petits faran que s'accepti el submodel.

3.1.2 mètode – m_f

Si $k\rho^2(\Theta_0, \bar{x}_k) > \gamma_{r,\alpha}(0)$, l'estimació de la regla-**m** (2.2) donada en el capítol anterior esdevé, en aquest cas

$$m_* = \rho_*^2 - E((\rho^2(\bar{X}_m, \mu)))$$

Aleshores, rebutjarem H_0 quan $m_* > 0$, acceptant altrament.

Multiplicant l'anterior equació per m , i donat que $E(m\rho^2(\bar{X}_m, \mu)) = E(\chi_n^2) = n$, la regla es pot formular de manera equivalent dient que:

Rebutjarem la hipòtesi nul·la H_0 si

$$m\rho_*^2 - n = \frac{m}{k} \gamma_{r,\alpha}^{-1}(k\rho^2(\Theta_0, \bar{x}_k)) - n > 0 \quad (3.2)$$

Notem que si $k\rho^2(\Theta_0, \bar{x}_k) \leq \gamma_{r,\alpha}(0)$, acceptarem sempre la hipòtesi nul·la, doncs aleshores $\rho_*^2 = 0$ i per tant $m_* = -E(m\rho^2(\bar{X}_m, \mu)) < 0$.

Fem la comparació amb el cas clàssic, el qual sabem que rebutjarà la hipòtesi nul·la, amb un nivell de significació $\varepsilon = \alpha$, quan $k\rho^2(\Theta_0, \bar{X}_k)$ sigui gran. Llavors el mètode-**m_f** també rebutjarà la hipòtesi nul·la si $m\rho_*^2 > n$, però encara l'acceptarà quan $m\rho_*^2 \leq n$, és a dir dependent del valor de m .

3.1.3 mètode – ρ^2

Observem que ara hem de considerar la mitjana de la quantitat

$$P(\rho^2(\bar{X}_m, \mu) < \rho^2(\Theta_0, \mu))$$

respecte la *a posteriori* (3.1), però ara per una mostra de grandària k : $\mu \sim N_n(\bar{x}_k, \frac{1}{k} \Sigma_0)$.

Hem de tenir en compte que donada \bar{x}_k , amb la mateixa notació bàsica que s'ha utilitzat en els anteriors apartats,

$$\sqrt{k} \Sigma_0^{-1/2}(\mu + H^{-1}b) \sim N_n(\sqrt{k} \Sigma_0^{-1/2}(\bar{x}_k + H^{-1}b), I)$$

i aleshores,

$$k \rho^2(\Theta_0, \mu) = k (\mu + H^{-1}b)' \Sigma_0^{-1/2} W \Sigma_0^{-1/2} (\mu + H^{-1}b)$$

té una distribució χ^2 no-central, $\chi_r^2(\delta)$, amb r graus de llibertat i paràmetre de no-centralitat $\delta = k \rho^2(\Theta_0, \bar{x}_k)$. A més, $m \rho^2(\bar{X}_m, \mu) \sim \chi_n^2$, independentment de μ . Per tant

$$E(P(\rho^2(\bar{X}_m, \mu) < \rho^2(\Theta_0, \mu)) | \bar{x}_k) = P\left(\frac{k n}{m r} < Y\right)$$

on

$$Y = \frac{k \rho^2(\Theta_0, \mu) / r}{m \rho^2(\bar{X}_m, \mu) / n}$$

és una variable aleatòria amb distribució $F_{r,n}(\delta)$, això és: una distribució F no-central, amb r i n graus de llibertat i paràmetre de no-centralitat $\delta = k \rho^2(\Theta_0, \bar{x}_k)$.

Si aquesta probabilitat és més gran que $1/2$ aleshores rebutjarem la hipòtesi nul·la o acceptarem el model més general, i farem el contrari en l'altre cas. Observem que si prenem $\zeta_{r,n}(\delta)$ com la mediana de una distribució $F_{r,n}(\delta)$, llavors:

Rebutjarem la hipòtesi nul·la o submodel si

$$\zeta_{r,n}(\delta) > \frac{k n}{m r}$$

acceptant-lo si això no passa. Notem també que si fem tendir $m \rightarrow \infty$ podem certament rebutjar la hipòtesi nul·la donat que $\zeta_{r,n}(\delta) > 0$.

3.1.4 mètode – mb

La mitjana de la distància de Rao al quadrat per a \mathcal{U}_m ve donada, donat que en el nostre cas $\mathcal{U}_m = \bar{X}_m$, per

$$E_\mu (\rho^2(\mathcal{U}_m, \mu)) = E_\mu ((\bar{X}_m - \mu)' \Sigma_0^{-1} (\bar{X}_m - \mu)) = \frac{n}{m}$$

i sota la hipòtesi nul·la, l'error estarà mesurat com: $\rho^2(\Theta_0, \mu)$.

Ara s'haurà de calcular la mitjana d'ambdues quantitats respecte la *a posteriori*: $\mu \sim N_n(\bar{x}_k, \frac{1}{k} \Sigma_0)$, amb la qual cosa tindrem

$$E (E_\mu (\rho^2(\mathcal{U}_m, \mu)) \mid \bar{x}_k) = \frac{n}{m}$$

i com que donada \bar{x}_k , $k \rho^2(\Theta_0, \mu)$ té una distribució χ^2 no-central, $\chi_r^2(\delta)$ amb r graus de llibertat i paràmetre de no-centralitat $\delta = k \rho^2(\Theta_0, \bar{x}_k)$, amb $E(\chi_r^2(\delta)) = r + \delta$, tenim

$$E (\rho^2(\Theta_0, \mu) \mid \bar{x}_k) = \frac{r}{k} + \rho^2(\Theta_0, \bar{x}_k)$$

Per tant, rebutjarem H_0 si

$$\frac{r}{k} + \rho^2(\Theta_0, \bar{x}_k) - \frac{n}{m} > 0$$

és a dir, en aquest cas resultarà avantatjós, des del present punt de vista, rebutjar triar el model general $X \sim N_n(\mu, \Sigma_0)$, $\mu \in \mathbb{R}^n$. Altrament acceptarem la hipòtesi nul·la o submodel $X \sim N_n(\mu, \Sigma_0)$, $\mu \in \Theta_0$.

De manera equivalent, si hem de treballar amb mostres de mida m i

$$m \leq \frac{nk}{r + k \rho^2(\Theta_0, \bar{x}_k)}$$

és preferible treballar amb el submodel $X \sim N_n(\mu, \Sigma_0)$, $\mu \in \Theta_0$, és a dir, acceptar la hipòtesi nul·la, donat que en aquest cas no tindrem suficient *potència resolutiva* per a fer bones estimacions amb el model complet, ja que la pèrdua esperada serà més gran sota el model més general. Notem que, per a mostres de grandària k suficientment gran, aquesta cota superior per a m és aproximadament igual a $n/\rho^2(\Theta_0, \bar{x}_k)$.

A més, observem que si

$$\frac{nk}{r + k \rho^2(\Theta_0, \bar{x}_k)} < 1 \quad \text{o equivalentment} \quad \frac{nk - r}{k} < \rho^2(\Theta_0, \bar{x}_k)$$

aleshores la hipòtesi nul·la H_0 serà sempre rebutjada, independentment de la grandària mostral m , doncs aleshores, $m > 0$ independentment de m .

3.2 Comparació dels mètodes

Per resumir tot el que hem vist podem comparar les condicions a favor del rebuig de la hipòtesi nul·la (o submodel), entre els diferents mètodes considerats, amb la següent taula:

| mètode | condició |
|--------|--|
| pf | $m \rho_*^2 > \xi_n$ |
| mf | $m \rho_*^2 > n$ |
| pb | $m \frac{r}{k} \zeta_{r,n}(\delta) > n$ |
| mb | $m \left(\frac{r}{k} + \rho^2(\Theta_0, \bar{x}_k) \right) > n$ |

De la taula anterior, usant aproximacions a la normal per a les distribucions χ_n^2 , $\chi_r^2(\delta)$ i $F_{r,n}(\delta)$, donades en: Abramowitz [1] [pp 942–948], s'obté que

$$\xi_n \sim n \left(1 - \frac{2}{9n}\right)^3 \sim n - \frac{2}{3}$$

i

$$\zeta_{r,n}(\delta) \sim \left(1 + \frac{k}{r} \rho^2(\Theta_0, \bar{x}_k)\right) \left(\frac{9n}{9n-2} \left(1 - \frac{2(r + 2k\rho^2(\Theta_0, \bar{x}_k))}{9(r + k\rho^2(\Theta_0, \bar{x}_k))^2}\right)\right)^3$$

aleshores

$$\frac{r}{k} \zeta_{r,n}(\delta) \sim \beta_{rn} \left(\frac{r}{k} + \rho^2(\Theta_0, \bar{x}_k)\right)$$

on

$$\beta_{rn} \sim \left(\frac{n(9r-2)}{(9n-2)r}\right)^3 \quad \text{o} \quad \beta_{rn} \sim \left(\frac{9n}{9n-2}\right)^3$$

per a valors petits de $\rho^2(\Theta_0, \bar{x}_k)$ o per a valors grans, respectivament.

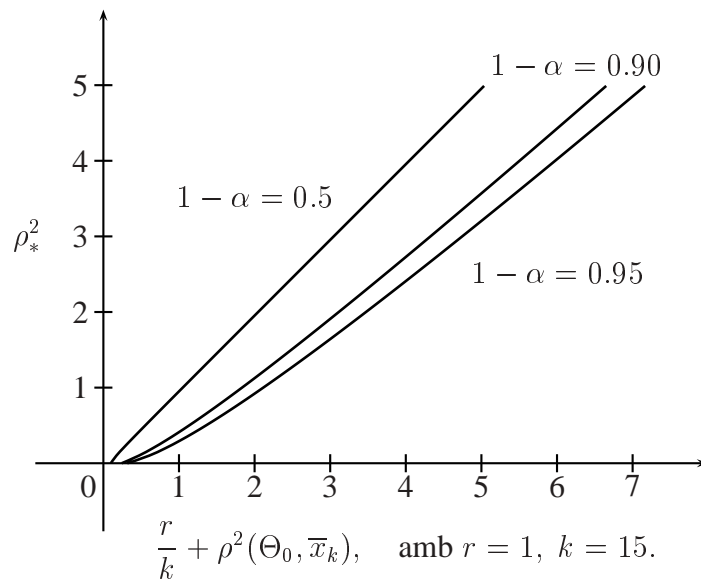
Per tant, per una banda **pf** i **mf**, proporcionaran conclusions similars donat que $\xi_n \sim n - 2/3$, per l'altra, **pb** i **mb** es comporten de manera similar, ja que, aproximadament $\frac{r}{k} \zeta_{r,n}(\delta)$ és proporcional a $\frac{r}{k} + \rho^2(\Theta_0, \bar{x}_k)$. És a dir les regles **p** i **m** funcionen de manera semblant tant si utilitzem el procediment **f** com si utilitzem el **b**.

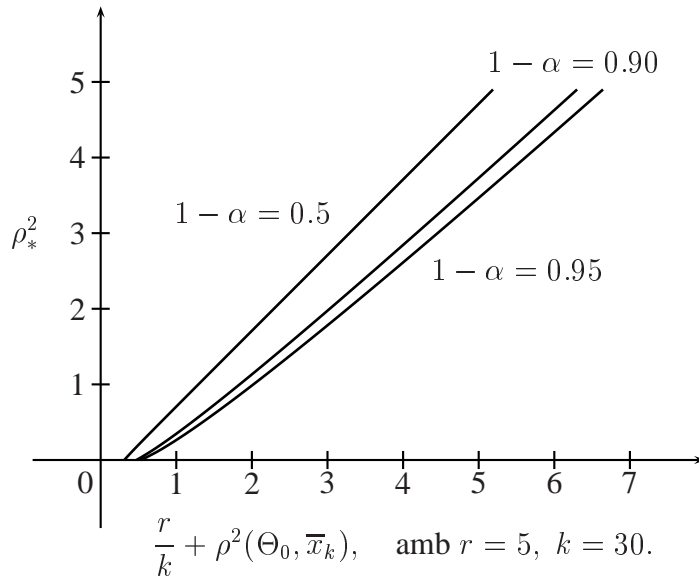
Amb l'objectiu de comparar els procediments **f** i **b**, podem fer la gràfica de ρ_*^2 respecte a $\frac{r}{k} + \rho^2(\Theta_0, \bar{x}_k)$ per a diferents valors de r i k , i considerant també diferents $(1 - \alpha)$ -intervalls de confiança per a ρ_*^2 . Per obtenir aquestes gràfiques s'ha usat l'aproximació a una normal per a la distribució χ^2 no central: si $U \sim \chi_r^2(\delta)$, aleshores la variable aleatòria

$$\sqrt{\frac{9(r+2\delta)^2}{2(r+2\delta)}} \left(\left(\frac{U}{r+\delta} \right)^{1/3} + \frac{2(r+2\delta)}{9(r+2\delta)^2} - 1 \right)$$

té, aproximadament, una distribució normal estàndard univariant.

Veurem doncs, a continuació, dues gràfiques comparatives usant en cada una d'elles diferents valors de r i k .





Observem, primerament, que la gràfica no depèn de manera sensible d'uns valors particulars de r i k . D'ambdós gràfics, és clar que els procediments **f** i **b** es comporten de manera similar sempre que $\frac{r}{k} + \rho^2(\Theta_0, \bar{x}_k)$ sigui més gran que una certa quantitat, c_α , la qual depèn del nivell de confiança ($1 - \alpha$). Per a nivells de confiança entre 0.50 i 0.95, c_α varia entre 0.098 i 0.535. Per a valors més petits de $\frac{r}{k} + \rho^2(\Theta_0, \bar{x}_k)$, els procediments **f** i **b** difereixen considerablement: el procediment **f** dóna prioritat al submodel (hipòtesi nul·la) sobre el model total, comparat amb el procediment **b**. Un fet que ens indica que el procediment **f** és més robust que el procediment **b** respecte al fet de partir de la suposició bàsica que la distribució *a priori* desconeguda està dominada pel volum Riemannià.

Capítol 4

Una aproximació asimptòtica del mètode- ρ

En aquest capítol donarem una solució asimptòtica al problema de test que s'ha plantejat en un principi. Usarem aquí els resultats asimptòtics obtinguts en la secció 2 del capítol 2. Treballarem amb un model regular paramètric de dimensió n i ho farem desenvolupant el mètode ρ .

4.1 Cas general amb hipòtesi nul·la simple

Considerem ara un model estadístic paramètric regular n -dimensional genèric, $\{(\mathcal{X}, \alpha, \mu); \Theta; f\}$, on Θ és l'espai de paràmetres, i el contrast:

$$H_0 : \theta = \theta_0, \quad H_1 : \theta \neq \theta_0$$

Sigui $\Theta_0 = \{\theta_0\}$ l'espai de paràmetres corresponent a la hipòtesi nul·la (submodel), $\rho(\theta_1, \theta_2)$ la distància de Rao, i sigui $\mathcal{U}_m(\cdot)$ un estimador eficient de primer ordre, com l'estimador màxim-versemblant.

Donada una mostra de grandària k , $x = (x_1, \dots, x_k) \in \mathcal{X}^k$, un $(1 - \alpha)$ -interval de confiança exacte per a ρ , de la forma $[\rho_*, 1]$, dependrà de cada model estadístic particular, i en general serà difícil obtenir-lo explícitament. Presentarem aquí, una aproximació basada en un resultat asimptòtic.

Primerament observem, segons hem vist al capítol 2, que

$$m \rho^2(\mathcal{U}_m(\cdot), \theta) \xrightarrow{\mathcal{L}} \chi_n^2$$

Com que $m \rho^2(\mathcal{U}_m(\cdot), \theta)$ té una distribució absolutament contínua, recordem que ρ es pot escriure

$$p = 2P(\rho^2(\mathcal{U}_m(\cdot), \theta) < \rho^2(\theta_0, \theta)) - 1$$

i aleshores, asimptòticament, p és una funció creixent de $\rho^2(\theta, \theta_0)$. Per tant podem obtenir un interval de confiança asimptòtic per a p de la forma $[p_*, 1]$, a partir d'un interval asimptòtic per a $\rho^2(\theta, \theta_0)$ de la forma $[\rho_*^2, \infty)$. Amb aquest propòsit, notem que

$$\sqrt{k} (\rho(\mathcal{U}_k(\cdot), \theta_0) - \rho(\theta, \theta_0)) \xrightarrow{L} Z \sim N(0, 1) \quad (4.1)$$

tal com s'ha obtingut en Teorema 2.3.1 del segon capítol.

Aleshores si definim ζ_α a partir de

$$P(Z \leq \zeta_\alpha) = 1 - \alpha$$

on Z té una distribució normal estàndard univariant, una vegada obtinguda la distància $\rho^2(\mathcal{U}_k(x), \theta_0)$, un $(1 - \alpha)$ -interval de confiança asimptòtic per a $\rho^2(\theta, \theta_0)$ serà $[\rho_*^2, \infty)$ amb

$$\rho_* = \begin{cases} \rho(\mathcal{U}_k(x), \theta_0) - \frac{\zeta_\alpha}{\sqrt{k}} & \text{si } \rho(\mathcal{U}_k(x), \theta_0) > \zeta_\alpha/\sqrt{k} \\ 0 & \text{altrament} \end{cases}$$

Després d'això, hem d'avaluar

$$P(\rho^2(\mathcal{U}_m(\cdot), \theta) < \rho_*^2) = P(m \rho^2(\mathcal{U}_m(\cdot), \theta) < m \rho_*^2)$$

usant, la ja mencionada abans, distribució asimptòtica: χ_n^2 . Llavors el criteri quedarà de la manera següent:

Acceptarem H_0 si

$$P(m \rho^2(\mathcal{U}_m(\cdot), \theta) < m \rho_*^2) \leq 1/2$$

i rebutjarem altrament.

Notem que, si anomenem ξ_n a la mediana de la distribució χ_n^2 , aquest criteri es pot expressar de manera equivalent dient que:

Acceptarem H_0 si

$$m \rho_*^2 = m \left(\rho(\mathcal{U}_k(x), \theta_0) - \frac{\zeta_\alpha}{\sqrt{k}} \right)^2 \leq \xi_n \quad (4.2)$$

i rebutjarem altrament.

Cal observar que, si $\rho_* = 0$, llavors sempre acceptarem la hipòtesi nul·la doncs $p_* = -1 < 0$.

Podem comparar aquest mètode amb el test d'hipòtesis clàssic, el qual rebutja la hipòtesi nul·la per a valors grans de $k\rho^2(\mathcal{U}_k(x), \theta_0)$ i tenint en compte la seva distribució asimptòtica, sota la hipòtesi nul·la, ve donada per

$$k\rho^2(\mathcal{U}_k(\cdot), \theta_0) \xrightarrow{\mathcal{L}} \chi_n^2$$

Així, fixat el p -valor obtingut en el test clàssic a partir d'una mostra $x = (x_1, \dots, x_k)$, de la forma

$$P(\chi_n^2 \geq k\rho^2(\mathcal{U}_k(x), \theta_0)) = p\text{-valor}$$

i la confiança $(1 - \alpha)$ de l'interval $[p_*, 1]$, que s'obindrà de $[\rho_*^2, \infty)$, donada la mateixa mostra anterior de mida k i a partir de (4.1)

$$P\left(\rho^2(\theta, \theta_0) \geq \left(\rho(\mathcal{U}_k(x), \theta_0) - \frac{\zeta_\alpha}{\sqrt{k}}\right)^2\right) = 1 - \alpha$$

podem calcular el valor del quocient m/k (la raó entre la segona i la primera mostres) començant des del valor a partir del qual, fins a l'infinit, rebutgem la hipòtesi nul·la (submodel), que s'obté del criteri (4.2), de la següent manera:

Rebutjarem H_0 si

$$\frac{m}{k} > \frac{\xi_n}{(\sqrt{k}\rho(\mathcal{U}_k(x), \theta_0) - \zeta_\alpha)^2}$$

Aleshores, per a diferents valors del nivell de confiança $(1 - \alpha)$ i del p -valor, donada una dimensió $n = \dim\Theta$, obtenim les següents taules.

| | | confiança per a $[p_*, 1]$ | | | | |
|-------------|--|----------------------------|-------|-------|----------|----------|
| p -valors | | 0.5 | 0.75 | 0.90 | 0.95 | 0.99 |
| 0.1 | | 0.168 | 0.483 | 3.453 | ∞ | ∞ |
| 0.05 | | 0.118 | 0.275 | 0.990 | 4.590 | ∞ |
| 0.01 | | 0.060 | 0.126 | 0.272 | 0.525 | 7.289 |
| 0.001 | | 0.042 | 0.066 | 0.113 | 0.168 | 0.489 |

$\dim \Theta = 1$. Valors crítics m/k .

| | | confiança per a $[p_*, 1]$ | | | | |
|-----------------|------------|--|-------------|-------------|-------------|--|
| <i>p-valors</i> | 0.5 | 0.75 | 0.90 | 0.95 | 0.99 | |
| 0.1 | 0.471 | 0.778 | 1.409 | 2.239 | 8.557 | |
| 0.05 | 0.393 | 0.618 | 1.040 | 1.538 | 4.341 | |
| 0.01 | 0.288 | 0.422 | 0.643 | 0.868 | 1.792 | |
| 0.001 | 0.212 | 0.293 | 0.413 | 0.523 | 0.896 | |

$\dim \Theta = 5$. Valors crítics m/k .

| | | confiança per a $[p_*, 1]$ | | | | |
|-----------------|------------|--|-------------|-------------|-------------|--|
| <i>p-valors</i> | 0.5 | 0.75 | 0.90 | 0.95 | 0.99 | |
| 0.1 | 0.584 | 0.845 | 1.266 | 1.687 | 3.340 | |
| 0.05 | 0.510 | 0.719 | 1.040 | 1.347 | 2.450 | |
| 0.01 | 0.403 | 0.544 | 0.747 | 0.928 | 1.505 | |
| 0.001 | 0.316 | 0.411 | 0.540 | 0.649 | 0.964 | |

$\dim \Theta = 10$. Valors crítics m/k .

| | | confiança per a $[p_*, 1]$ | | | | |
|-----------------|------------|--|-------------|-------------|-------------|--|
| <i>p-valors</i> | 0.5 | 0.75 | 0.90 | 0.95 | 0.99 | |
| 0.1 | 0.729 | 0.912 | 1.145 | 1.328 | 1.816 | |
| 0.05 | 0.670 | 0.831 | 1.031 | 1.187 | 1.594 | |
| 0.01 | 0.576 | 0.703 | 0.857 | 0.974 | 1.269 | |
| 0.001 | 0.491 | 0.590 | 0.706 | 0.793 | 1.006 | |

$\dim \Theta = 30$. Valors crítics m/k .

A la vista dels resultats obtinguts a les taules podem dir que, si usant els mètodes clàssics i donada una determinada mostra de grandària k , s'obté un $p - valor \leq 0.01$ amb la qual cosa es rebutjaria la hipòtesi nul·la amb un nivell de significació de l'ordre $\varepsilon > 0.01$, amb el mètode $-pf$ en canvi, usant nivells de confiança entre el 50% i el 95%, encara s'acceptaria la hipòtesi per a mostres de grandàries $m < k$, i en alguns casos $m \simeq k$. Per altra banda, si usem confiances a partir del 95%, podrem acceptar la hipòtesi nul·la fins i tot per algunes mostres de grandàries més grans que k .

També podem dir, per exemple, que si amb els mètodes clàssics obtenim un $p - valor \geq 0.05$ amb la qual cosa s'acceptaria la hipòtesi nul·la amb nivells de significació $\varepsilon \leq 0.05$, amb el mètode $-pf$, també s'acceptarà per a mostres de grandàries $m \leq k$, i fins i tot per a confiances $1 - \alpha \geq 90\%$, encara s'acceptarà per a algunes mostres de grandària $m > k$.

Cal observar, finalment, que els mètode es comporta aproximadament de la mateixa manera encara que la dimensió de l'espai de paràmetres augmenti.

Capítol 5

Alguns models uniparamètrics

5.1 Model Poisson

Considerem la distribució Poisson, $X \sim P(\lambda)$, amb $\lambda > 0$, i el següent contrast:

$$H_0 : \lambda = \lambda_0, \quad H_1 : \lambda \neq \lambda_0$$

Desenvoluparem el mètode --mb per aquest exemple particular. L'espai general de paràmetres (model) és $\Theta = \mathbb{R}^+$, i l'espai de paràmetres corresponent a la hipòtesi nul·la (submodel), $\Theta_0 = \{\lambda_0\}$.

La mètrica informativa ve donada per:

$$ds^2 = \frac{1}{\lambda} d\lambda^2$$

la geometria és, com en tots els casos unidimensionals, euclidiana i la distància de Rao és en aquest cas:

$$\rho(\lambda_1, \lambda_2) = 2 |\sqrt{\lambda_1} - \sqrt{\lambda_2}|$$

Donada una mostra de mida m , $x = (x_1, \dots, x_m) \in \mathbb{N}^m$, un estimador natural per aquest model és el centre de masses de la mesura de probabilitat *a posteriori*, basada en una *a priori no-informativa* proporcional al volum Riemannià. Aquesta distribució *a posteriori*, sota la parametrització λ , ve donada per

$$h(\lambda|s) = \frac{m^{s+\frac{1}{2}}}{\Gamma(s+\frac{1}{2})} e^{-m\lambda} \lambda^{s-\frac{1}{2}} \quad \lambda > 0 \quad s \in \mathbb{N} \quad (5.1)$$

on $s = \sum_i^m x_i$.

A més, si fem un canvi de parametrització $\lambda \mapsto \theta(\lambda) = 2\sqrt{\lambda}$, la mètrica informativa esdevé $ds^2 = d\theta^2$ i sota aquesta parametrització, el centre de masses és just l'esperança. Per tant, tenint en compte que

$$E(2\sqrt{\lambda}|s) = \frac{2 m^{s+\frac{1}{2}}}{\Gamma(s+\frac{1}{2})} \int_0^\infty e^{-m\lambda} \lambda^s d\lambda = \frac{2 \Gamma(s+1)}{\sqrt{m} \Gamma(s+\frac{1}{2})} \quad s \in \mathbb{N}$$

l'estimador de Bayes generalitzat, donat per una mostra de mida m , sota la parametrització λ , és

$$S_m(x_1, \dots, x_m) = \left(\frac{\Gamma(s+1)}{\sqrt{m} \Gamma(s+\frac{1}{2})} \right)^2$$

donat que $\lambda = (\theta/2)^2$. La mitjana de la distància de Rao al quadrat, per a S_m , ve donada per

$$\begin{aligned} E_\lambda(\rho^2(S_m, \lambda)) &= 4 \sum_{s=0}^{\infty} \left(\frac{s!}{\sqrt{m} \Gamma(s+\frac{1}{2})} - \sqrt{\lambda} \right)^2 e^{-m\lambda} \frac{(m\lambda)^s}{s!} \\ &= \frac{4e^{-m\lambda}}{m} \sum_{s=0}^{\infty} \frac{s! s!}{\Gamma(s+\frac{1}{2}) \Gamma(s+\frac{1}{2})} \frac{(m\lambda)^s}{s!} + 4\lambda e^{-m\lambda} \sum_{s=0}^{\infty} \frac{(m\lambda)^s}{s!} \\ &\quad - \frac{8\sqrt{\lambda}}{\sqrt{m}} e^{-m\lambda} \sum_{s=0}^{\infty} \frac{s! (m\lambda)^s}{\Gamma(s+\frac{1}{2}) s!} \\ &= 4\lambda + \frac{4e^{-m\lambda}}{\pi m} \sum_{s=0}^{\infty} \frac{(1)_s (1)_s}{(\frac{1}{2})_s (\frac{1}{2})_s} \frac{(m\lambda)^s}{s!} - \frac{8\sqrt{\lambda}}{\sqrt{\pi m}} e^{-m\lambda} \sum_{s=0}^{\infty} \frac{(1)_s}{(\frac{1}{2})_s} \frac{(m\lambda)^s}{s!} \end{aligned}$$

tenint en compte que:

$$\frac{\Gamma(s+\frac{1}{2})}{\Gamma(\frac{1}{2})} = \left(\frac{1}{2}\right)_s \quad \text{on} \quad (a)_s = a(a+1)\cdots(a+s-1)$$

$$s! = 1(1+1)\cdots(1+s-1) = (1)_s \quad \text{i} \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

A més, ${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; z)$ són les funcions hipergeomètriques generalitzades, les quals convergeixen si $p \leq q$, $\forall z$ finit, veure Muirhead [31][pp 20–22].

Així finalment ens queda:

$$E_\lambda(\rho^2(S_m, \lambda)) = 4\lambda + \frac{4e^{-m\lambda}}{\pi m} \left({}_2F_2\left(1, 1; \frac{1}{2}, \frac{1}{2}; m\lambda\right) - 2\sqrt{\lambda\pi m} {}_1F_1\left(1; \frac{1}{2}; m\lambda\right) \right)$$

Sota la hipòtesi nul·la, l'error ve mesurat per $\rho^2(\lambda, \lambda_0) = 4(\sqrt{\lambda} - \sqrt{\lambda_0})^2$.

Calculant ara la mitjana d'ambdues quantitats respecte la *a posteriori* (5.1), però ara per una mostra de mida k : $s = \sum_{i=1}^k x_i$, respecte la parametrització λ , i utilitzant que:

$$\begin{aligned} & \int_0^\infty e^{-zt} t^{a-1} {}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; kt) dt \\ &= \Gamma(a) z^{-a} {}_{p+1}F_q(a_1, \dots, a_p, a; b_1, \dots, b_q; kz^{-1}) \end{aligned}$$

per a $p < q$, $Re(a) > 0$, $Re(z) > 0$ o bé $p = q$, $Re(a) > 0$, $Re(z) > Re(k)$

tindrem aleshores

$$\begin{aligned} E(E_\lambda(\rho^2(S_m, \lambda)) | s) &= \frac{4k^{s+\frac{1}{2}}}{\Gamma(s+\frac{1}{2})} \int_0^\infty \left(\lambda + e^{-m\lambda} \frac{1}{\pi m} \right. \\ & \left. \left({}_2F_2\left(1, 1; \frac{1}{2}, \frac{1}{2}; m\lambda\right) - 2\sqrt{\pi m \lambda} {}_1F_1\left(1; \frac{1}{2}; m\lambda\right) \right) \right) e^{-k\lambda} \lambda^{s-\frac{1}{2}} d\lambda \\ &= \frac{4s+2}{k} + \frac{4(1+\frac{m}{k})^{-(s+\frac{1}{2})}}{\pi m} \left({}_3F_2\left(1, 1, s+\frac{1}{2}; \frac{1}{2}, \frac{1}{2}; \frac{m}{m+k}\right) \right. \\ & \quad \left. - \frac{2s! \sqrt{\pi m}}{\sqrt{m+k} \Gamma(s+\frac{1}{2})} {}_2F_1\left(1, s+1; \frac{1}{2}; \frac{m}{m+k}\right) \right) \end{aligned} \quad (5.2)$$

i

$$\begin{aligned} E(\rho^2(\lambda, \lambda_0) | s) &= \frac{4k^{s+\frac{1}{2}}}{\Gamma(s+\frac{1}{2})} \int_0^\infty (\sqrt{\lambda_0} - \sqrt{\lambda})^2 \lambda^{s-\frac{1}{2}} e^{-k\lambda} d\lambda \\ &= 4 \left(\lambda_0 - \frac{2\sqrt{\lambda_0} s!}{\sqrt{k} \Gamma(s+\frac{1}{2})} \right) + \frac{4s+2}{k} \end{aligned} \quad (5.3)$$

Tenint en compte que si ${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; z)$ són tals que $p > q$, aquestes funcions seran convergents quan $z < 1$, la qual cosa passa en el nostre cas donat que $\frac{m}{m+k} < 1$.

Per tant, hem de comparar ambdues quantitats. Si (5.3) és estrictament més gran que (5.2) aleshores triarem el model general, si això no passa triarem el submodel (hipòtesi nul·la).

D'una manera més explícita, si

$$\frac{2s!}{\Gamma(s + \frac{1}{2})} \sqrt{\frac{\pi m}{m+k}} {}_2F_1(1, s+1; \frac{1}{2}; \frac{m}{m+k}) - {}_3F_2(1, 1, s + \frac{1}{2}; \frac{1}{2}, \frac{1}{2}; \frac{m}{m+k}) + \pi m (1 + \frac{m}{k})^{(s+\frac{1}{2})} \left(\lambda_0 - \frac{2\sqrt{\lambda_0} s!}{\sqrt{k} \Gamma(s + \frac{1}{2})} \right) > 0$$

aleshores *rebutjarem* H_0 .

La següent taula mostra el comportament del mètode – **m6**. Per a $\lambda_0 = 1$ i per a diferents valors de k i s , es calcularà la màxima grandària mostral, m^* , que encara permet l'ús del submodel: per a m menor o igual que m^* acceptarem H_0 . Observem que $m^* \leq k$, independentment de k i s , per a m suficientment gran, sempre serà preferible el model més general.

| | | s | | | | | | | | | |
|----|--|---|---|---|---|----|----|----|----|----|----|
| k | | 1 | 3 | 5 | 7 | 10 | 15 | 20 | 25 | 30 | 35 |
| 5 | | – | 2 | 5 | 5 | – | – | – | – | – | – |
| 10 | | – | – | 1 | 2 | 10 | 3 | 1 | – | – | – |
| 15 | | – | – | – | 1 | 5 | 15 | 5 | 2 | 1 | – |
| 20 | | – | – | – | – | 2 | 8 | 20 | 9 | 3 | 2 |
| 25 | | – | – | – | – | – | 4 | 12 | 25 | 12 | 5 |
| 30 | | – | – | – | – | – | 2 | 6 | 16 | 30 | 16 |

valors de m^* : per a $m \leq m^*$ acceptarem H_0

5.2 Model Gamma

Considerem la distribució Gamma, $X \sim G(1/\lambda, \alpha)$, on $\alpha > 0$ és conegut, i $\lambda > 0$, on el paràmetre λ és tal que $E(X) = \alpha\lambda$, amb el contrast:

$$H_0 : \lambda = \lambda_0, \quad H_1 : \lambda \neq \lambda_0$$

En aquest exemple desenvoluparem el mètode $\mathbf{m\delta}$. L'espai general de paràmetres (model) és $\Theta = \mathbb{R}^+$, i l'espai de paràmetres corresponent a la hipòtesi nul·la (submodel), $\Theta_0 = \{\lambda_0\}$.

La mètrica informativa ve donada per:

$$ds^2 = \frac{\alpha}{\lambda^2} d\lambda^2,$$

la geometria és euclidiana, i la distància de Rao és en aquest cas

$$\rho(\lambda_1, \lambda_2) = \sqrt{\alpha} \left| \log \frac{\lambda_1}{\lambda_2} \right|$$

Donada una mostra de grandària m , $x = (x_1, \dots, x_m) \in (\mathbb{R}^+)^m$, un estimador natural per a aquest model és el centre de masses de la probabilitat *a posteriori*, basat en una *a priori no-informativa* proporcional al volum Riemannià. Aquesta *a posteriori* és, sota la parametrització λ ,

$$h(\lambda|s) = \frac{s^{m\alpha}}{\Gamma(m\alpha) \lambda^{m\alpha+1}} e^{-s/\lambda} \quad \alpha, s \in \mathbb{R}^+ \quad (5.4)$$

on $s = \sum_{j=1}^m x_j$ i havent tingut en compte que $\sum_{j=1}^m X_j \sim G(1/\lambda, m\alpha)$.

Per tant, si fem un canvi de parametrització: $\lambda \mapsto \theta(\lambda) = \sqrt{\alpha} \log \lambda$, la mètrica informativa esdevé $ds^2 = d\theta^2$ i sota aquesta parametrització, el centre de masses és just l'esperança. Per tant, tenint en compte que

$$\begin{aligned} E(\sqrt{\alpha} \log \lambda | s) &= \frac{\sqrt{\alpha} s^{m\alpha}}{\Gamma(m\alpha)} \int_0^\infty \frac{\log \lambda}{\lambda^{m\alpha+1}} e^{-s/\lambda} d\lambda \\ &= \sqrt{\alpha} (\log s - \Psi(m\alpha)) \end{aligned}$$

on $\Psi = \Gamma'/\Gamma$, i l'estimador de Bayes generalitzat S_m , donat per una mostra de mida m , respecte la parametrització λ , és

$$S_m(x_1, \dots, x_m) = s e^{-\Psi(m\alpha)} = e^{-\Psi(m\alpha)} \sum_{j=1}^m x_j$$

donat que $\lambda = e^{\theta/\sqrt{\alpha}}$. Aleshores la mitjana de la distància de Rao al quadrat, per a S_m , ve donada per

$$\begin{aligned} E_\lambda (\rho^2(S_m, \lambda)) &= \frac{\alpha}{\Gamma(m\alpha)} \int_0^\infty \log^2 \left(\frac{se^{-\Psi(m\alpha)}}{\lambda} \right) s^{m\alpha-1} e^{-s/\lambda} ds \\ &= \alpha \Psi'(m\alpha) \end{aligned}$$

Sota la hipòtesi nul·la, l'error ve mesurat per $\rho^2(\lambda, \lambda_0) = \alpha \log^2 \left(\frac{\lambda}{\lambda_0} \right)$.

Calculant ara la mitjana d'ambdues quantitats respecte la *a posteriori* (5.4), però ara per a una mostra de mida k : $s = \sum_{i=1}^k x_i$, tindrem

$$E(E_\lambda (\rho^2(S_m, \lambda)) \mid s) = \alpha \Psi'(m\alpha) \quad (5.5)$$

i

$$\begin{aligned} E(\rho^2(\lambda, \lambda_0) \mid s) &= \frac{\alpha s^{k\alpha}}{\Gamma(k\alpha)} \int_0^\infty \log^2 \left(\frac{\lambda_0}{\lambda} \right) \frac{1}{\lambda^{k\alpha+1}} e^{-s/\lambda} d\lambda \\ &= \alpha \left(\log^2 \left(\frac{se^{-\Psi(k\alpha)}}{\lambda_0} \right) + \Psi'(k\alpha) \right) \end{aligned} \quad (5.6)$$

Per tant, hem de comparar ambdues quantitats. Si (5.6) és estrictament més gran que (5.5) aleshores triarem el model general, si això no passa, triarem el submodel (hipòtesi nul·la). Explícitament, si

$$\alpha (\Psi'(m\alpha) - \Psi'(k\alpha)) < \alpha \log^2 \left(\frac{se^{-\Psi(k\alpha)}}{\lambda_0} \right) = \rho^2(S_k, \lambda_0)$$

llavors rebutgem la hipòtesi nul·la, H_0 , acceptant-la en qualsevol altre cas. El comportament del mètode és, en aquest cas, similar al dels exemples previs, tal com es pot observar en la següent taula, on s'han obtingut valors crítics per a $\rho^2(S_k, \lambda_0)$, anomenats $\hat{\rho}^2$, en el cas $\alpha = 1$ i per a diferents valors de k i m .

| k | m | | | | | | | |
|----|--------|-------|-------|-------|-------|-------|-------|-------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 5 | 1.4236 | .4236 | .1736 | .0625 | .0000 | .0000 | .0000 | .0000 |
| 10 | 1.5398 | .5398 | .2898 | .1787 | .1162 | .0762 | .0484 | .0280 |
| 15 | 1.5760 | .5760 | .3260 | .2149 | .1524 | .1124 | .0846 | .0642 |
| 20 | 1.5937 | .5937 | .3437 | .2326 | .1701 | .1301 | .1023 | .0819 |
| 25 | 1.6041 | .6041 | .3541 | .2430 | .1805 | .1405 | .1127 | .0923 |
| 30 | 1.6110 | .6110 | .3610 | .2499 | .1874 | .1474 | .1197 | .0992 |
| 35 | 1.6160 | .6160 | .3660 | .2548 | .1923 | .1523 | .1246 | .1042 |

Valors crítics $\hat{\rho}^2$: *acceptarem* H_0 si $\rho^2(S_k, \lambda_0) \leq \hat{\rho}^2$

Observem que acceptarem el submodel, hipòtesi nul·la, per a valors petits de m , si $m \geq k$, aleshores rebutjarem el submodel.

Capítol 6

Model lineal normal

6.1 Cas general

Considerem un model lineal normal $Y \sim N_n(X\beta, \sigma^2 I)$, on $\beta \in \mathbb{R}^s$ i $\sigma > 0$, són paràmetres desconeguts i X és una matriu $n \times s$ els coeficients de la qual són coneguts i tal que $0 < \text{rang } X = s < n$, amb el contrast:

$$H_0 : H\beta = 0, \quad H_1 : H\beta \neq 0$$

on H és una matriu constant $r \times s$ amb $0 < \text{rang}(H) = r \leq s$.

Nota: si considerem el cas general:

$$H_0 : H\beta = C, \quad H_1 : H\beta \neq C$$

aquest es pot reduir al primer. Per exemple: suposem que existeix β_0 amb $H\beta_0 = C$, aleshores:

$$H\beta = C \iff H\beta = H\beta_0 \iff H(\beta - \beta_0) = 0$$

Si definim $\gamma = \beta - \beta_0$ el model lineal esdevé:

$$Y = X\beta + e = X(\beta - \beta_0) + X\beta_0 + e = X\gamma + X\beta_0 + e$$

i podem considerar $Z = Y - X\beta_0 = X\gamma + e$ amb $Z \sim N_n(X\gamma, \sigma^2 I)$, aleshores el contrast d'hipòtesis és:

$$H_0 : H\gamma = 0, \quad H_1 : H\gamma \neq 0$$

a partir d'ara, en aquest exemple, els elements de \mathbb{R}^k seran considerats en notació matricial, com a vectors columna. Sota aquesta parametrització l'espai de

paràmetres corresponent al model total és $\Theta = \mathbb{R}^s \times \mathbb{R}^+$ i l'espai de paràmetres corresponent a la hipòtesi nul·la $\Theta_0 = \{(\beta, \sigma) \in \mathbb{R}^s \times \mathbb{R}^+ : H\beta = 0\}$.

La mètrica informativa ve donada, en notació matricial, per

$$ds^2 = \frac{1}{\sigma^2} (d\beta' X' X d\beta + 2n d\sigma^2)$$

que és la *mètrica hiperbòlica de Poincaré* del semi-espai superior $\mathbb{R}^s \times \mathbb{R}^+$, veure Burbea i Oller [9], i la distància Riemanniana és

$$\rho(\beta_1, \sigma_1; \beta_2, \sigma_2) = \sqrt{2n} \log \frac{1 + \delta_{12}}{1 - \delta_{12}} = 2\sqrt{2n} \tanh^{-1}(\delta_{12}) \quad (6.1)$$

on

$$\delta_{12} = \sqrt{\frac{\|\beta_2 - \beta_1\|^2 + 2n(\sigma_1 - \sigma_2)^2}{\|\beta_2 - \beta_1\|^2 + 2n(\sigma_1 + \sigma_2)^2}}$$

i

$$\|\beta_2 - \beta_1\|^2 = (\beta_1 - \beta_2)' X' X (\beta_1 - \beta_2)$$

Per aquest exemple anem a desenvolupar el mètode **pl**, prenent l'estimador màxim-versemblant. Considerarem una mostra de mida m , $\mathbf{Y} = (Y_1, \dots, Y_m)$, del model lineal definit anteriorment, i definim una nova matriu $\tilde{X}_{(m)}$, que serà una matriu $nm \times s$ obtinguda copiant convenientment les files de la matriu X de la manera següent

$$\tilde{X}_{(m)} = I_m \otimes X$$

on $I_m' = (1, \dots, 1)'$, essent \otimes el producte de Kronecker. Per altra banda la matriu $n \times m$ aleatòria $\mathbf{Y} = (Y_1, \dots, Y_m)$, les columnes de la qual són vectors aleatoris corresponents a una mostra de mida m d'un model lineal, pot ser considerada com una mostra de mida 1 del model lineal $\text{vec}(\mathbf{Y}) \sim N_{nm}(\tilde{X}_{(m)}\beta, \sigma^2 I)$, a més, tenint en compte que:

$$(I_m \otimes X)' I_m \otimes X = m X' X$$

$$(I_m \otimes X)' \text{vec}(\mathbf{Y}) = m X' \bar{Y}_m$$

i

$$\text{vec}(\mathbf{Y} H_m)' \text{vec}(\mathbf{Y} H_m) = \sum_{i=1}^m (Y_i - \bar{Y}_m)' (Y_i - \bar{Y}_m) = \text{tr}(\mathbf{Y} H_m \mathbf{Y}')$$

on $\bar{Y}_m = \frac{1}{m}(Y_1 + \dots + Y_m)$ i H_m és la matriu $m \times m$ simètrica idempotent $H_m = I - \frac{1}{m} I_m I_m'$, aleshores l'estimador màxim-versemblant vindrà donat per:

$$\begin{aligned}\beta_{(m)}^* &= (\tilde{X}'_{(m)} \tilde{X}_{(m)})^{-1} \tilde{X}'_{(m)} \text{vec}(\mathbf{Y}) = (X'X)^{-1} X' \bar{Y}_m \\ \sigma_{(m)}^* &= \sqrt{\frac{1}{nm} \text{vec}(\mathbf{Y})' \left(I - \tilde{X}_{(m)} (\tilde{X}'_{(m)} \tilde{X}_{(m)})^{-1} \tilde{X}'_{(m)} \right) \text{vec}(\mathbf{Y})} \\ &= \sqrt{\frac{1}{nm} \text{tr}(\mathbf{Y} H_m \mathbf{Y}') + \frac{1}{n} \bar{Y}'_m (I - X(X'X)^{-1} X') \bar{Y}_m}\end{aligned}$$

Sabem que $\bar{Y}_m \sim N_n(X\beta, \frac{\sigma^2}{m}I)$, aleshores:

$$\frac{m}{\sigma^2} \bar{Y}'_m (I - X(X'X)^{-1} X') \bar{Y}_m \sim \chi_{n-s}^2$$

i

$$\frac{1}{\sigma^2} \text{tr}(\mathbf{Y} H_m \mathbf{Y}') = \frac{1}{\sigma^2} \sum_{j=1}^n \sum_{i=1}^m (Y_{ij} - \bar{Y}_{.j})^2 \sim \chi_{nm-n}^2$$

Si definim:

$$U_{(m)} = m \|\beta_{(m)}^* - \beta\|^2 / \sigma^2 \sim \chi_s^2 \quad V_{(m)} = nm(\sigma_{(m)}^*)^2 / \sigma^2 \sim \chi_{nm-s}^2$$

amb $U_{(m)}$ i $V_{(m)}$ independents, tenim

$$\begin{aligned}P\left(\rho^2(\beta_{(m)}^*, \sigma_{(m)}^*; \beta, \sigma) < c\right) &= P\left(m\rho^2(\beta_{(m)}^*, \sigma_{(m)}^*; \beta, \sigma) < mc\right) \stackrel{(*)}{=} \\ P\left(U_{(m)} + 2nm\left(\sqrt{\frac{V_{(m)}}{nm}} - 1\right)^2 < 8\sqrt{nmV_{(m)}} \sinh^2\left(\sqrt{\frac{c}{8n}}\right)\right) &\stackrel{(**)}{=} \\ \alpha_{s, nm} \int_{nm e^{-\sqrt{2c/n}}}^{nm e^{\sqrt{2c/n}}} \int_0^{4\sqrt{nmv} \cosh(\sqrt{c/2n}) - 2v - 2nm} u^{\frac{s}{2}-1} v^{\frac{nm-s}{2}-1} e^{-\frac{1}{2}(u+v)} du dv &\end{aligned} \quad (6.2)$$

on $c > 0$ i

$$\alpha_{s, nm} = \frac{1}{2^{\frac{nm}{2}} \Gamma(\frac{s}{2}) \Gamma(\frac{nm-s}{2})}$$

Provem ara com s'ha obtingut la igualtat (*). Hem vist que la distància Riemanniana complex, segons (6.1), la següent igualtat

$$\rho^2(\beta_{(m)}^*, \sigma_{(m)}^*; \beta, \sigma) = 8n \left(\tanh^{-1}(\delta_{(m)})\right)^2$$

on $\delta_{(m)}$ és en el nostre cas

$$\delta_{(m)} = \sqrt{\frac{\|\beta - \beta_{(m)}^*\|^2 + 2n(\sigma_{(m)}^* - \sigma)^2}{\|\beta - \beta_{(m)}^*\|^2 + 2n(\sigma_{(m)}^* + \sigma)^2}}$$

Considerarem, llavors, la següent cadena d'equivalències

$$\begin{aligned} \rho^2(\beta_{(m)}^*, \sigma_{(m)}^*; \beta, \sigma) < c &\iff 8n (\tanh^{-1}(\delta_{(m)}))^2 < c \\ &\iff \tanh^{-1}(\delta_{(m)}) < \sqrt{\frac{c}{8n}} \iff \delta_{(m)} < \tanh\left(\sqrt{\frac{c}{8n}}\right) \\ &\iff 1 - \delta_{(m)}^2 > 1 - \tanh^2\left(\sqrt{\frac{c}{8n}}\right) = \frac{1}{\cosh^2\left(\sqrt{\frac{c}{8n}}\right)} \\ &\iff \frac{8n\sigma_{(m)}^*\sigma}{\|\beta - \beta_{(m)}^*\|^2 + 2n(\sigma_{(m)}^* + \sigma)^2} > \frac{1}{\cosh^2\left(\sqrt{\frac{c}{8n}}\right)} \\ &\iff 8n\sigma_{(m)}^*\sigma(\sinh^2\left(\sqrt{\frac{c}{8n}}\right) + 1) > \|\beta - \beta_{(m)}^*\|^2 + 2n(\sigma_{(m)}^* + \sigma)^2 \\ &\iff 8n\sigma_{(m)}^*\sigma \sinh^2\left(\sqrt{\frac{c}{8n}}\right) > \|\beta - \beta_{(m)}^*\|^2 + 2n(\sigma_{(m)}^* - \sigma)^2 \end{aligned}$$

Llavors si multipliquem aquesta última desigualtat per m i la dividim entre σ^2 , obtenim l'expressió

$$m \frac{\|\beta - \beta_{(m)}^*\|^2}{\sigma^2} + 2nm \left(\frac{\sigma_{(m)}^*}{\sigma} - \sigma\right)^2 < 8nm \frac{\sigma_{(m)}^*}{\sigma} \sinh^2\left(\sqrt{\frac{c}{8n}}\right)$$

que de manera equivalent, amb les definicions utilitzades anteriorment és

$$U_{(m)} + 2nm \left(\sqrt{\frac{V_{(m)}}{nm}} - 1\right)^2 < 8\sqrt{nmV_{(m)}} \sinh^2\left(\sqrt{\frac{c}{8n}}\right)$$

A continuació, indicarem com s'ha obtingut la desigualtat (**). Sabem que $U_{(m)}$ i $V_{(m)}$ són variables aleatòries independents i aleshores considerem el vector aleatori $(U_{(m)}, V_{(m)})$ que pren valors $(u, v) \in (\mathbb{R}^+ \times \mathbb{R}^+)$, hem de determinar la regió del pla que compleix la desigualtat

$$u + 2nm \left(\sqrt{\frac{v}{nm}} - 1\right)^2 < 8\sqrt{nmv} \sinh^2\left(\sqrt{\frac{c}{8n}}\right)$$

Treballant amb aquesta desigualtat ens queda una inequació de segon grau de la forma

$$(\sqrt{v})^2 - 2nm(1 + 2 \sinh^2(\sqrt{\frac{c}{8n}}))\sqrt{v} + nm + \frac{u^2}{2} < 0 \quad (6.3)$$

les solucions de la qual en funció de la variable u són

$$\sqrt{nm}(1 + 2 \sinh^2(\sqrt{\frac{c}{8n}})) \pm \sqrt{nm(1 + 2 \sinh^2(\sqrt{\frac{c}{8n}}))^2 - (nm + \frac{u^2}{2})}$$

Notem que els valors màxim i mínim de \sqrt{v} s'assoliran quan $u = 0$, llavors les solucions quedaran

$$\sqrt{nm} \left(1 + 2 \sinh^2(\sqrt{\frac{c}{8n}}) \pm 2 \sinh(\sqrt{\frac{c}{8n}}) \cosh(\sqrt{\frac{c}{8n}}) \right)$$

que, utilitzant les propietats de les funcions trigonomètriques hiperbòliques i la seva expressió exponencial, ens determinen un rang de valors per a \sqrt{v} de la forma

$$-\sqrt{nm} e^{-2\sqrt{\frac{c}{8n}}} < \sqrt{v} < \sqrt{nm} e^{2\sqrt{\frac{c}{8n}}}$$

i en elevar aquesta expressió al quadrat s'obté el següent rang de valors per a v

$$nm e^{-\sqrt{\frac{2c}{n}}} < v < nm e^{\sqrt{\frac{2c}{n}}}$$

Per obtenir el camp de variació de u cal tenir en compte que ha de ser positiva i que fixada v el valor màxim s'obté aïllant u de la inequació (6.3). Llavors, considerant que $U_{(m)}$ i $V_{(m)}$ són variables aleatòries amb distribució χ^2 obtenim finalment la integral de la igualtat (**).

Aleshores és clar que la probabilitat (6.2) és una funció monòtona creixent de c , i per tant p és una funció monòtona creixent de $\rho^2(\Theta_0, (\beta, \sigma))$. Així doncs, donada una matriu $n \times k$ aleatòria $\mathbf{y} = (y_1, \dots, y_k)$, les columnes de la qual corresponen a una mostra de mida k del model lineal, podem obtenir un $(1 - \alpha)$ interval de confiança per a p a partir d'un interval de confiança per a $\rho^2(\Theta_0, (\beta, \sigma))$, de la forma $[\rho_*^2, \infty)$, essent ρ_*^2 una estimació per defecte de $\rho^2(\Theta_0, (\beta, \sigma))$ donada aquesta mostra de mida k .

Per obtenir aquest interval observem que, després d'alguns càlculs obtinguts també de l'equació (6.1) sota la hipòtesi nul·la, podem expressar

$$\rho(\Theta_0, (\beta, \sigma)) = 2\sqrt{2n} \tanh^{-1}\left(\frac{\sqrt{D^2 + 1} - 1}{D}\right)$$

on

$$D^2 = D^2(\beta, \sigma) = \frac{\beta' H' (H (X' X)^{-1} H')^{-1} H \beta}{2n\sigma^2}$$

essent D^2 independent de la g -inversa escollida. Així doncs, $\rho(\Theta_0, (\beta, \sigma))$ és una funció monòtona creixent de D^2 , i es pot obtenir un interval de confiança per a $\rho^2(\Theta_0, (\beta, \sigma))$ de la forma $[\rho_*^2, \infty)$ a partir d'un interval de confiança per a D^2 de la forma $[D_*^2, \infty)$.

Notem ara que

$$\mathcal{Y} = \frac{2(nk - s)}{r} D^2(\beta_{(k)}^*, \sigma_{(k)}^*)$$

és una variable aleatòria amb distribució $F_{r, nk-s}(\Delta)$, amb paràmetre de no-centralitat $\Delta = 2nkD^2(\beta, \sigma)$. Definim $\phi_{r, nk-s, \alpha}(\Delta)$ a partir de

$$P(\mathcal{Y} \leq \phi_{r, nk-s, \alpha}(\Delta)) = 1 - \alpha$$

on $\phi_{r, nk-s, \alpha}(\Delta)$ és una funció monòtona estrictament creixent de Δ , per tant donada $\hat{D}_{(k)}^2 = D^2(\hat{\beta}_{(k)}^*, \hat{\sigma}_{(k)}^*)$, on $\hat{\beta}_{(k)}^*$ i $\hat{\sigma}_{(k)}^*$ són les estimacions particulars corresponents a la mostra de mida k \mathbf{y} , i si considerem $\delta \equiv \phi_{r, nk-s, \alpha}(0)$, un $(1 - \alpha)$ -interval de confiança per a D^2 serà $[D_*^2, \infty)$ amb

$$D_*^2 = \begin{cases} \frac{1}{2nk} \phi_{r, nk-s, \alpha}^{-1}\left(\frac{2(nk - s)}{r} \hat{D}_{(k)}^2\right) & \text{si } \frac{2(nk - s)}{r} \hat{D}_{(k)}^2 > \delta \\ 0 & \text{altrament} \end{cases}$$

i llavors

$$\rho_* = 2\sqrt{2n} \tanh^{-1}\left(\frac{\sqrt{D_*^2 + 1} - 1}{D_*}\right)$$

Per altra banda hem d'avaluar $\mathfrak{p}_* = 2P(m\rho^2(\beta_{(m)}^*, \sigma_{(m)}^*; \beta, \sigma) < m\rho_*^2) - 1$, a partir de la distribució donada en (6.2). De fet, només necessitem conèixer la mediana de la distribució $m\rho^2(\beta_{(m)}^*, \sigma_{(m)}^*; \beta, \sigma)$, doncs ens interessarà determinar quan $P(m\rho^2(\beta_{(m)}^*, \sigma_{(m)}^*; \beta, \sigma) < c) > 1/2$.

Primerament, observem que

$$\begin{aligned} 8nm\sqrt{\frac{V_{(m)}}{nm}} \sinh^2\left(\sqrt{\frac{c}{8nm}}\right) &= 8nm\sqrt{\frac{V_{(m)}}{nm}} \frac{1}{2} \left(\cosh\left(\sqrt{\frac{c}{2nm}}\right) - 1\right) \\ &= \sqrt{\frac{V_{(m)}}{nm}} \left(c + O\left(\frac{1}{nm}\right)\right) \end{aligned}$$

que es troba a partir del desenvolupament de Taylor de segon ordre de la funció \cosh . A més, quan $nm \rightarrow \infty$, es pot provar fàcilment, a partir del Teorema Central del Límit, que

$$\frac{V_{(m)} - nm}{\sqrt{2nm}} \xrightarrow{\mathcal{L}} Z \sim N(0, 1)$$

i a partir del Teorema de Slutsky tenim que, si $\frac{V_{(m)}}{nm}$ és $AN(\mu = 1, \sigma_{nm}^2 = 2/nm)$ amb $\sigma_{nm}^2 \rightarrow 0$ quan $nm \rightarrow \infty$, i a més considerem una funció real $g(x) = \sqrt{x}$ derivable a $x = \mu$ amb $g'(\mu) \neq 0$, podem assegurar que $g(\frac{V_{(m)}}{nm})$ és $AN(g(\mu), g'(\mu)^2 \sigma_{nm}^2)$, és a dir

$$\sqrt{2nm} \left(\sqrt{\frac{V_{(m)}}{nm}} - 1 \right) \xrightarrow{\mathcal{L}} Z \sim N(0, 1),$$

A més, com $\sigma_{(m)}^*$ és un estimador màxim-versemblant de σ , és consistent i es compleix

$$\frac{(\sigma_{(m)}^*)^2}{\sigma^2} = \frac{V_{(m)}}{nm} \xrightarrow{\mathcal{P}} 1$$

Per tant,

$$8nm \sqrt{\frac{V_{(m)}}{nm}} \sinh^2 \left(\sqrt{\frac{c}{8nm}} \right) = c + O_P\left(\frac{1}{nm}\right)$$

Finalment, fixada s , podem concloure que

$$m\rho^2(\beta_{(m)}^*, \sigma_{(m)}^*; \beta, \sigma) \xrightarrow{\mathcal{L}} \chi_{s+1}^2$$

això és, la distribució límit de $m\rho^2(\beta_{(m)}^*, \sigma_{(m)}^*; \beta, \sigma)$ és una distribució χ^2 amb $s + 1$ graus de llibertat.

Així doncs, per a valors grans de nm si considerem ξ_{s+1} la mediana d'una distribució χ_{s+1}^2 , rebutjarem el submodel si $m\rho_*^2 > \xi_{s+1} \sim s + \frac{1}{3}$, acceptant-lo en qualsevol altre cas. Per a valors petits s'ha calculat la mediana numèricament, obtenint la següent Taula

| nm | s | | | | |
|----------------|-------|-------|-------|-------|--------|
| | 1 | 2 | 3 | 4 | 5 |
| 10 | 1.586 | 3.022 | 4.871 | 7.311 | 10.638 |
| 15 | 1.513 | 2.763 | 4.224 | 5.942 | 7.982 |
| 20 | 1.479 | 2.650 | 3.964 | 5.438 | 7.098 |
| 25 | 1.459 | 2.588 | 3.824 | 5.177 | 6.658 |
| 30 | 1.447 | 2.548 | 3.737 | 5.017 | 6.395 |
| 50 | 1.422 | 2.471 | 3.574 | 4.726 | 5.928 |
| χ_{s+1}^2 | 1.386 | 2.366 | 3.357 | 4.351 | 5.348 |

Taula 6.1.1: Mediana de $m\rho^2(\beta_{(m)}^*, \sigma_{(m)}^*; \beta, \sigma)$.

Notem que, quan més gran és la dimensió de l'espai de paràmetres, més difícil és rebutjar la hipòtesi nul·la, és a dir, necessitem una grandària m més gran. En canvi, quan més gran és la dimensió n del model lineal donat, més fàcil és que la grandària final nm sigui gran, i lògicament, més fàcil és rebutjar la hipòtesi nul·la.

6.2 Exemple

6.2.1 Cas I

Considerarem ara l'estudi d'un exemple particular, basant-nos en els resultats obtinguts per al mètode- pf en el cas general, de la secció anterior.

S'han considerat els salaris anuals, en dòlars, de 15 homes i 30 dones, nascuts entre els anys 1966 i 1970, de la mateixa categoria laboral (administratius), segons les dades de la taula que apareix a la pàgina següent.

TAULA DE SALARIS D'ADMINISTRATIUS

| individu | sexe | data naixement | salari (en \$) |
|-----------------|-------------|-----------------------|-----------------------|
| 1 | h | 18-09-69 | 25500 |
| 2 | h | 27-02-69 | 27750 |
| 3 | h | 09-03-69 | 21300 |
| 4 | h | 08-11-66 | 31050 |
| 5 | h | 10-10-68 | 26700 |
| 6 | h | 04-08-68 | 31500 |
| 7 | h | 27-05-68 | 31650 |
| 8 | h | 19-07-69 | 28500 |
| 9 | h | 08-10-66 | 27450 |
| 10 | h | 22-07-67 | 37800 |
| 11 | h | 14-11-67 | 28050 |
| 12 | h | 03-08-66 | 26400 |
| 13 | h | 15-11-66 | 33540 |
| 14 | h | 26-07-67 | 29400 |
| 15 | h | 05-04-66 | 30750 |
| 16 | d | 02-05-69 | 27450 |
| 17 | d | 09-05-69 | 25500 |
| 18 | d | 12-05-70 | 21600 |
| 19 | d | 19-09-69 | 34500 |
| 20 | d | 17-01-70 | 22950 |
| 21 | d | 12-01-69 | 24450 |
| 22 | d | 24-06-69 | 27900 |
| 23 | d | 04-02-70 | 29100 |
| 24 | d | 21-11-70 | 30600 |
| 25 | d | 10-10-66 | 28200 |
| 26 | d | 13-10-67 | 21900 |
| 27 | d | 06-08-69 | 23850 |
| 28 | d | 12-08-69 | 22500 |
| 29 | d | 13-05-67 | 26700 |
| 30 | d | 02-07-67 | 26550 |
| 31 | d | 14-06-66 | 35550 |
| 32 | d | 27-01-66 | 27900 |
| 33 | d | 29-04-66 | 33300 |
| 34 | d | 24-08-66 | 24600 |
| 35 | d | 24-11-66 | 29850 |
| 36 | d | 18-09-69 | 29700 |
| 37 | d | 09-11-68 | 24150 |
| 38 | d | 09-06-66 | 28800 |
| 39 | d | 13-11-66 | 28500 |
| 40 | d | 16-08-68 | 22950 |
| 41 | d | 05-11-68 | 29400 |
| 42 | d | 14-10-66 | 29150 |
| 43 | d | 11-08-68 | 29400 |
| 44 | d | 05-05-69 | 22200 |
| 45 | d | 18-08-67 | 32850 |

Suposarem que la variable "salari homes" segueix una distribució $N(\beta_1, \sigma^2)$ i que la variable "salari dones" segueix una distribució $N(\beta_2, \sigma^2)$.

Tot i que podem considerar aquest exemple com un cas clar d'aplicació del test *t de Student* de comparació de mitjanes, el plantejarem, amb l'objectiu d'aplicar els resultats obtinguts en la secció anterior, suposant que la mostra obtinguda prové d'un model lineal general amb la següent estructura

$$Y = X\beta + e \quad Y \sim N_n(X\beta, \sigma^2 I)$$

on $n = n_1 + n_2$, en el nostre cas $n_1 = 1$, $n_2 = 2$, el vector e amb components $e_j \sim N(0, \sigma^2)$, per a $j = 1, \dots, n$, el vector de paràmetres $\beta' = (\beta_1, \beta_2) \in \mathbb{R}^s$, $s = 2$, i X la matriu del disseny $n \times s = 3 \times 2$, tal que $\text{rang}(X) = s = 2$, i amb la següent estructura, donat que hi doble número d'dones que d' homes,

$$X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}$$

Volem contrastar si hi ha diferències entre les mitjanes dels salaris dels homes i de les dones. És a dir:

$$H_0 : \beta_1 = \beta_2, \quad H_1 : \beta_1 \neq \beta_2$$

que matricialment s'expressa

$$H_0 : H\beta = 0, \quad H_1 : H\beta \neq 0$$

amb $H = (1, -1)$, amb $\text{rang}(H) = r = 1$.

L'espai general de paràmetres és

$$\Theta = \{(\beta, \sigma) \in \mathbb{R}^2 \times \mathbb{R}^+\}, \quad \dim(\Theta) = 3$$

i el subespai sota l'hipòtesi nul·la

$$\Theta_0 = \{(\beta, \sigma) \in \mathbb{R}^2 \times \mathbb{R}^+ : \beta_1 = \beta_2\}, \quad \dim(\Theta_0) = 2$$

Aleshores podem suposar que disposem d'una mostra de grandària $k = 15$ del model lineal anterior, i podem escriure

$$Y_{ij} \sim N(\beta_1, \sigma^2) \quad i = 1, \dots, 15 \quad j = 1$$

$$Y_{ij} \sim N(\beta_2, \sigma^2) \quad i = 1, \dots, 15 \quad j = 2, 3$$

amb

$$\bar{Y}_{.j} = \frac{1}{15} \sum_{i=1}^{15} Y_{ij}$$

llavors, en aquest cas particular

$$\bar{Y}_{15} = \begin{pmatrix} \bar{Y}_{.1} \\ \bar{Y}_{.2} \\ \bar{Y}_{.3} \end{pmatrix} = \begin{pmatrix} 29156 \\ 26250 \\ 28554 \end{pmatrix}, \quad \sum_{j=1}^3 \sum_{i=1}^{15} (Y_{ij} - \bar{Y}_{.j})^2 = 587530920$$

A partir de la mostra anterior hem d'obtenir un interval de confiança per a la distància $\rho^2(\Theta_0, (\beta, \sigma))$ de la forma $[\rho_*^2, \infty)$, a partir d'un interval de confiança per a $D^2(\beta, \sigma)$ de la forma $[D_*^2, \infty)$ on, tal i com hem vist en la secció anterior

$$\rho(\Theta_0, (\beta, \sigma)) = 2\sqrt{2n} \tanh^{-1}\left(\frac{\sqrt{D^2 + 1} - 1}{D}\right)$$

amb

$$D^2 = D^2(\beta, \sigma) = \frac{\beta' H' (H(X'X)^{-1} H')^{-1} H \beta}{2n\sigma^2}$$

i sabem que la variable

$$\mathcal{Y} = \frac{2(nk - s)}{r} D^2(\beta_{(k)}^*, \sigma_{(k)}^*)$$

és, en aquest cas particular, una variable aleatòria amb distribució $F_{1,43}(\Delta)$, amb paràmetre de no-centralitat $\Delta = 2nkD^2(\beta, \sigma) = 90D^2(\beta, \sigma)$. Definim $\phi_{1,43,\alpha}(\Delta)$ a partir de

$$P(\mathcal{Y} \leq \phi_{1,43,\alpha}(\Delta)) = 1 - \alpha \quad (6.4)$$

Necessitem ara calcular $\hat{D}_{(15)}^2 = D^2(\hat{\beta}_{(15)}^*, \hat{\sigma}_{(15)}^*)$, a partir de la nostra mostra particular. Primerament, a partir d'uns senzills càlculs, obtenim les estimacions màxim-versemblants de β i σ per a la mostra donada.

$$\hat{\beta}_{(15)}^* = (X'X)^{-1} X' \bar{Y}_{15} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \end{pmatrix} \begin{pmatrix} 29156 \\ 26250 \\ 28554 \end{pmatrix} = \begin{pmatrix} 29156 \\ 27402 \end{pmatrix}$$

$$\begin{aligned} \frac{1}{3} \bar{Y}'_{15} (I - X(X'X)^{-1}X') \bar{Y}_{15} &= \\ \frac{1}{3} (29156, 26250, 28554) \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1/2 & -1/2 \\ 0 & -1/2 & 1/2 \end{pmatrix} \begin{pmatrix} 29156 \\ 26250 \\ 28554 \end{pmatrix} &= \\ \frac{1}{3} (-30240000 + 32894208) &= 884736 \end{aligned}$$

i

$$\frac{1}{3} \text{tr}(\mathbf{Y}H_{15}\mathbf{Y}') = \frac{1}{3} \sum_{j=1}^3 \sum_{i=1}^{15} (Y_{ij} - \bar{Y}_{.j})^2 = \frac{1}{3} 587530920$$

així obtenim

$$\begin{aligned} \hat{\sigma}_{15}^* &= \sqrt{\frac{1}{45} \text{tr}(\mathbf{Y}H_{15}\mathbf{Y}') + \frac{1}{3} \bar{Y}'_{15} (I - X(X'X)^{-1}X') \bar{Y}_{15}} \\ &= \sqrt{\frac{1}{3} 587530920 + 884736} = 3733.762 \end{aligned}$$

i tenint en compte que

$$(H(X'X)^{-1}H')^{-1} = \left((1, -1) \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right)^{-1} = \frac{2}{3}$$

llavors podem calcular

$$\begin{aligned} \hat{D}_{(15)}^2 &= \frac{(29156, 27402) \begin{pmatrix} 1 \\ -1 \end{pmatrix} \frac{2}{3} (1, -1) \begin{pmatrix} 29156 \\ 27402 \end{pmatrix}}{6 (3733.762)^2} \\ &= \frac{\frac{2}{3} (29156 - 27402)^2}{6 (3733.762)^2} = 0.0245201 \end{aligned}$$

i

$$\hat{\mathcal{Y}}_{(15)} = \frac{2(nk - s)}{r} \hat{D}_{(15)}^2 = \frac{2(45 - 2)}{1} 0.0245201 = 2.109$$

Buscarem ara el valor de $\delta_\alpha = \phi_{1,43,\alpha}(0)$ per a diferents valors de α i el comparem amb el valor anterior.

$$\delta_\alpha = \phi_{1,43,\alpha}(0) = \begin{cases} 4.07 & \text{si } \alpha = 0.05 \\ 2.83 & \text{si } \alpha = 0.10 \\ 1.36 & \text{si } \alpha = 0.25 \\ 0.46 & \text{si } \alpha = 0.50 \end{cases}$$

aleshores

$$\frac{2(45-2)}{1} \hat{D}_{(15)}^2 = 2.109 < \delta_\alpha \quad \text{per a } \alpha = 0.05 \quad \text{i} \quad \alpha = 0.10$$

mentre que

$$\frac{2(45-2)}{1} \hat{D}_{(15)}^2 = 2.109 > \delta_\alpha \quad \text{per a } \alpha = 0.25 \quad \text{i} \quad \alpha = 0.50$$

i podem obtenir

$$\hat{D}_*^2 = \begin{cases} 0 & \text{si } \alpha = 0.05 \\ 0 & \text{si } \alpha = 0.10 \\ \frac{1}{90} \phi_{1,43,0.25}^{-1}(2.109) & \text{si } \alpha = 0.25 \\ \frac{1}{90} \phi_{1,43,0.50}^{-1}(2.109) & \text{si } \alpha = 0.50 \end{cases}$$

Llavors podem concloure que

$$\text{per a } \alpha = 0.05 \quad \text{i} \quad \alpha = 0.10 \quad \implies \quad \text{acceptarem } H_0$$

doncs, en aquest cas, $\rho_* = 0$, amb la qual cosa $p_* = -1 < 0$.

Calculem ara, per a $\alpha = 0.25$ i per a $\alpha = 0.50$, D_*^2 a partir del paràmetre de no centralitat Δ que s'obté de l'equació (6.4). Notem que donat el valor mostrat $\hat{Y}_{(15)} = 2.109$ obtindrem, a partir de la funció de distribució $F_{1,43}(\Delta)$, que

$$\Delta = \phi_{1,43,\alpha}^{-1}(2.109) = \begin{cases} 0.499327 & \text{si } \alpha = 0.25 \\ 2.068338 & \text{si } \alpha = 0.50 \end{cases}$$

Aleshores, com que

$$D_*^2 = \frac{1}{2nk} \Delta$$

obtenim que

$$D_*^2 = \begin{cases} \frac{1}{90} 0.499327 = 0.00555 & \text{si } \alpha = 0.25 \\ \frac{1}{90} 2.068338 = 0.023 & \text{si } \alpha = 0.50 \end{cases}$$

Llavors l'interval de confiança al 75% per a $D^2(\beta, \sigma)$ serà $[0.00555, \infty)$, i al 50% serà $[0.023, \infty)$.

Notem que, per obtenir $[\rho_*^2, \infty)$, necessitem calcular

$$\rho_* = 2\sqrt{6} \tanh^{-1} \left(\frac{\sqrt{D_*^2 + 1} - 1}{D_*} \right)$$

i amb els resultats obtinguts anteriorment sabem que

$$D_* = \begin{cases} \sqrt{0.005555} = 0.0745 & \text{si } \alpha = 0.25 \\ \sqrt{0.023} = 0.152 & \text{si } \alpha = 0.50 \end{cases}$$

aleshores

$$\rho_* = \begin{cases} 2\sqrt{6} \tanh^{-1} \left(\frac{\sqrt{0.0745^2 + 1} - 1}{0.0745} \right) = 0.1823 & \text{si } \alpha = 0.25 \\ 2\sqrt{6} \tanh^{-1} \left(\frac{\sqrt{0.152^2 + 1} - 1}{0.152} \right) = 0.3723 & \text{si } \alpha = 0.50 \end{cases}$$

Així, els intervals de confiança de la forma $[\rho_*^2, \infty)$ obtinguts per a $\rho^2(\Theta_0, (\beta, \sigma))$ al $(1 - \alpha)\%$ seran

$$[\rho_*^2, \infty) = \begin{cases} [0.1823^2, \infty) = [0.0332, \infty) & \text{al } 75\% \\ [0.3723^2, \infty) = [0.1386, \infty) & \text{al } 50\% \end{cases}$$

Sabem doncs, que acceptarem H_0 si

$$m\rho_*^2 \leq \text{mediana}(m\rho^2(\beta_{(m)}^*, \sigma_{(m)}^*; \beta, \sigma))$$

i, en altre cas, la rebutjarem.

Llavors a partir de la Taula 6.1.1, de la secció anterior, per a $n = 3$, $s = 2$ i sabent que

$$m\rho_*^2 = \begin{cases} m 0.0332 & \text{si } \alpha = 0.25 \\ m 0.1386 & \text{si } \alpha = 0.50 \end{cases}$$

el criteri de decisió quedarà

Per a $\alpha = 0.25$, si $m \leq 71 \implies$ acceptarem H_0

Per a $\alpha = 0.50$, si $m \leq 17 \implies$ acceptarem H_0

Fem ara una comparativa dels resultats que s'obtenen amb aquest mètode-pf i els que s'obtidrien amb els tests clàssics que, en aquest cas, serien el test *t de Student* de comparació de mitjanes, equivalent també al d'anàlisi de la variància univariant d'un factor.

Suposem primer que realitzem el test d'hipòtesis de l'exemple amb algun dels dos tests clàssics anomenats anteriorment. Si prenem un nivell de significació $\varepsilon = 0.05$, $\varepsilon = 0.10$ o bé, en general, valors de ε menors o iguals que el p -valor, és a dir, si

$$\varepsilon \leq P(|t_{43}| \geq 1.452) = P(F_{1,43} \geq 2.109) = 0.154$$

llavors acceptarem la hipòtesi nul·la. Mentre que si prenem nivells de significació superiors al p -valor, és a dir, $\varepsilon > 0.154$, aleshores rebutjarem H_0 . Cal tenir en compte que, en aquests tests clàssics, es treballa amb una única mostra donada i es prenen decisions, segons els resultats obtinguts, per a aquesta mostra.

Considerem ara els resultats obtinguts amb el mètode-**pf**. Notem que si prenem valors de α del mateix ordre que s'acostumen a prendre els nivells de significació, ε , dels tests clàssics, és a dir, $\alpha = 0.05$, $\alpha = 0.10$ i, en general, $\alpha \leq 0.154$, llavors acceptarem la hipòtesi nul·la, tal i com passava en els tests clàssics, doncs els intervals de confiança al nivell $(1 - \alpha)\%$ obtinguts per a $\rho^2(\Theta_0, (\beta, \sigma))$, de la forma $[\rho_*^2, \infty)$, són en aquest cas $[0, \infty)$, intervals massa amplis.

Notem però, que el valor α del mètode-**pf** no juga el mateix paper que el nivell de significació en els tests clàssics, no és una probabilitat d'error. En aquest nou mètode, el valor α ens serveix per obtenir una estimació de la distància $\rho^2(\Theta_0, (\beta, \sigma))$ amb nivells de confiança de l'ordre del $(1 - \alpha)\%$. Sembla, doncs, més convenient trobar estimacions ρ_*^2 més realistes o, en certa manera, més centrades, és a dir, intervals de confiança de la forma $[\rho_*^2, \infty)$ fins i tot del 50%, de manera que el 50% de les vegades obtinguéssim estimacions de la distància per sobre de ρ_*^2 , i l'altre 50% per sota.

Així, en aquest cas particular, s'obté que, si amb el mètode-**pf** prenem $\alpha = 0.25$ i $\alpha = 0.50$ es rebutjarà la hipòtesi nul·la per a mostres de grandàries $m > 71$ i $m > 17$ respectivament, que és el mateix que passaria prenent nivells de significació $\varepsilon = 0.25$, $\varepsilon = 0.50$ i usant qualsevol dels dos tests clàssics anteriors, lògicament. Però encara s'acceptarà per a mostres de grandàries $m \leq 71$ i $m \leq 17$, també respectivament a $\alpha = 0.25$ i $\alpha = 0.50$. És a dir, donada una primera mostra de grandària $k = 15$, que usem per obtenir l'estimació ρ_*^2 obtenim un criteri de decisió que dependrà de la grandària, m , d'una segona mostra que poguéssim obtenir amb les mateixes característiques que la primera.

Veiem doncs, que aquest nou mètode ens dóna més informació que la que obtenim en els tests clàssics, en el sentit que ens dóna més possibilitats de decisió, actuant, a més, de manera equivalent si treballem amb valors de α iguals al nivell de significació ε .

6.2.2 Cas II

En aquest segon cas, farem una modificació en les dades del nostre exemple de la següent manera: augmentarem en 2000 dòlars el salari anual dels 15 homes, deixant invariant el de les 30 dones. Aquesta modificació *artificial* es fa a fi d'obtenir, d'una manera molt senzilla, unes dades tals que la diferència entre les mitjanes dels salaris sigui més gran, és a dir, més significativa que en el Cas I, però mantenint la mateixa variància.

De fet, es pretén que amb aquestes dades es faci més entenedor i es vegin totes les possibilitats del paper que juga la grandària de la segona mostra m , en el fet de decidir acceptar o rebutjar la hipòtesi nul·la. És a dir, que depenent de la mostra amb la qual poguéssim treballar posteriorment, serà preferible o avantatjós acceptar el submodel.

Les estimacions de β i σ , en aquest nou cas, quedaran

$$\hat{\beta}_{(15)}^* = \begin{pmatrix} 31156 \\ 27402 \end{pmatrix}$$

i com la variància l'hem mantingut igual, el seu estimador coincidirà amb el del cas anterior, és a dir

$$\hat{\sigma}_{(15)}^* = 3733.762$$

i donats aquests dos valors, l'estimació de $D^2(\beta, \sigma)$ dona

$$\hat{D}_{(15)}^2 = \frac{\frac{2}{3}(31156 - 27402)^2}{6(3733.762)^2} = 0.1123188$$

i

$$\hat{Y}_{(15)} = \frac{2(nk - s)}{r} \hat{D}_{(15)}^2 = \frac{2(45 - 2)}{1} 0.1123188 = 9.659$$

Els valors de $\delta_\alpha = \phi_{1,43,\alpha}(0)$ per a diferents valors de α seran, lògicament, els mateixos que hem obtingut en el cas anterior i si els compararem amb el valor $\frac{2(nk-s)}{r} \hat{D}_{(15)}^2$, que acabem de calcular, s'obté que

$$\frac{2(45 - 2)}{1} \hat{D}_{(15)}^2 = 9.659 > \delta_\alpha \quad \forall \alpha$$

tenint en compte que hem considerat els quatre mateixos valors de α que en el Cas I.

Per obtenir l'interval de confiança $[D_*^2, \infty)$ al nivell $(1 - \alpha)\%$, cal que calculem els corresponents valors de D_*^2 , que s'obtenen a partir del paràmetre de no centralitat

Δ . Així doncs, donat el valor mostrat $\hat{\mathcal{Y}}_{(15)} = 9.659$ i a partir de la funció de distribució $F_{1,43}(\Delta)$, tenim

$$\Delta = \phi_{1,43,\alpha}^{-1}(6.659) = \begin{cases} 1.840215 & \text{si } \alpha = 0.05 \\ 3.024278 & \text{si } \alpha = 0.10 \\ 5.656730 & \text{si } \alpha = 0.25 \\ 9.543512 & \text{si } \alpha = 0.50 \end{cases}$$

Aleshores, com que

$$D_*^2 = \frac{1}{2nk} \Delta$$

obtenim que

$$D_*^2 = \begin{cases} \frac{1}{90} \phi_{1,43,0.05}^{-1}(9.659) = 0.0204468 & \text{si } \alpha = 0.05 \\ \frac{1}{90} \phi_{1,43,0.10}^{-1}(9.659) = 0.0336030 & \text{si } \alpha = 0.10 \\ \frac{1}{90} \phi_{1,43,0.25}^{-1}(9.659) = 0.0628525 & \text{si } \alpha = 0.25 \\ \frac{1}{90} \phi_{1,43,0.50}^{-1}(9.659) = 0.1060390 & \text{si } \alpha = 0.50 \end{cases}$$

Per obtenir $[\rho_*^2, \infty)$, necessitem calcular

$$\rho_* = 2\sqrt{6} \tanh^{-1} \left(\frac{\sqrt{D_*^2 + 1} - 1}{D_*} \right)$$

i amb els resultats obtinguts anteriorment sabem que

$$D_* = \begin{cases} \sqrt{0.0204468} = 0.143 & \text{si } \alpha = 0.05 \\ \sqrt{0.0336030} = 0.183 & \text{si } \alpha = 0.10 \\ \sqrt{0.0628525} = 0.251 & \text{si } \alpha = 0.25 \\ \sqrt{0.1060390} = 0.326 & \text{si } \alpha = 0.50 \end{cases}$$

aleshores

$$\rho_* = \begin{cases} 2\sqrt{6} \tanh^{-1} \left(\frac{\sqrt{0.143^2 + 1} - 1}{0.143} \right) = 0.3491 & \text{si } \alpha = 0.05 \\ 2\sqrt{6} \tanh^{-1} \left(\frac{\sqrt{0.183^2 + 1} - 1}{0.183} \right) = 0.4458 & \text{si } \alpha = 0.10 \\ 2\sqrt{6} \tanh^{-1} \left(\frac{\sqrt{0.251^2 + 1} - 1}{0.251} \right) = 0.6085 & \text{si } \alpha = 0.25 \\ 2\sqrt{6} \tanh^{-1} \left(\frac{\sqrt{0.326^2 + 1} - 1}{0.326} \right) = 0.7850 & \text{si } \alpha = 0.50 \end{cases}$$

Així, els intervals de confiança de la forma $[\rho_*^2, \infty)$ obtinguts per a $\rho^2(\Theta_0, (\beta, \sigma))$ al $(1 - \alpha)\%$ seran

$$[\rho_*^2, \infty) = \begin{cases} [0.3491^2, \infty) = [0.1219, \infty) & \text{al } 95\% \\ [0.4458^2, \infty) = [0.1987, \infty) & \text{al } 90\% \\ [0.6085^2, \infty) = [0.3703, \infty) & \text{al } 75\% \\ [0.7850^2, \infty) = [0.6162, \infty) & \text{al } 50\% \end{cases}$$

Sabem doncs, que acceptarem H_0 si

$$m\rho_*^2 \leq \text{mediana}(m\rho^2(\beta_{(m)}^*, \sigma_{(m)}^*; \beta, \sigma))$$

i, en altre cas, la rebutjarem.

Llavors a partir de la Taula 6.1.1, de la secció anterior, per a $n = 3$, $s = 2$ i sabent que

$$m\rho_*^2 = \begin{cases} m0.1219 & \text{si } \alpha = 0.05 \\ m0.1987 & \text{si } \alpha = 0.10 \\ m0.3703 & \text{si } \alpha = 0.25 \\ m0.6162 & \text{si } \alpha = 0.50 \end{cases}$$

el criteri de decisió quedarà

- Per a $\alpha = 0.05$, si $m \leq 19 \implies$ acceptarem H_0
- Per a $\alpha = 0.10$, si $m \leq 11 \implies$ acceptarem H_0
- Per a $\alpha = 0.25$, si $m \leq 6 \implies$ acceptarem H_0
- Per a $\alpha = 0.50$, si $m \leq 3 \implies$ acceptarem H_0

En conclusió, usant els tests clàssics veiem que rebutjaríem la hipòtesi nul·la per als nivells de significació més habituals $\varepsilon = 0.10, 0.05, 0.01$, de fet només l'acceptaríem per a nivells $\varepsilon < 0.003$, mentre que usant el mètode $-pf$ amb els valors de α anteriorment especificats, encara s'acceptarà la hipòtesi nul·la quan tinguem mostres de grandàries donades en el criteri de decisió anterior. Altrament, també la rebutjaríem.

6.2.3 Cas III

En aquest tercer cas, farem una modificació més exagerada en les dades del nostre exemple inicial. Augmentarem en 10000 dòlars el salari anual dels 15 homes, deixant invariant, també, el de les 30 dones. Aquesta modificació es fa a fi d'obtenir, ara, unes dades tals que la diferència entre les mitjanes dels salaris sigui molt més gran, és a dir, molt més significativa que en els casos anteriors, però mantenint també la mateixa variància.

Serà important veure el paper que jugarà la grandària de la segona mostra m , en el fet de decidir acceptar o rebutjar la hipòtesi nul·la, en una situació tant clarament significativa de rebuig de H_0 . Seguirem, lògicament, els mateixos passos que en els dos casos anteriors.

Les estimacions de β i σ quedaran ara

$$\hat{\beta}_{(15)}^* = \begin{pmatrix} 39156 \\ 27402 \end{pmatrix}$$

i com la variància també l'hem mantingut igual, el seu estimador coincidirà amb el dels casos anteriors, és a dir

$$\hat{\sigma}_{(15)}^* = 3733.762$$

i donats aquests dos valors, l'estimació de $D^2(\beta, \sigma)$ dóna

$$\hat{D}_{(15)}^2 = \frac{\frac{2}{3} (39156 - 27402)^2}{6 (3733.762)^2} = 1.1011224$$

i

$$\hat{Y}_{(15)} = \frac{2(nk - s)}{r} \hat{D}_{(15)}^2 = \frac{2(45 - 2)}{1} 1.1011224 = 94.697$$

Els valors de $\delta_\alpha = \phi_{1,43,\alpha}(0)$ per a diferents valors de α seran, lògicament també, els mateixos que hem obtingut en els casos anteriors i si els comparem amb el valor $\frac{2(nk-s)}{r} \hat{D}_{(15)}^2$, que acabem de calcular, s'obté que

$$\frac{2(45 - 2)}{1} \hat{D}_{(15)}^2 = 94.697 > \delta_\alpha \quad \forall \alpha$$

tenint en compte que hem considerat els quatre mateixos valors de α que en el Cas I i en el Cas II.

Per obtenir l'interval de confiança $[D_*^2, \infty)$ al nivell $(1 - \alpha)\%$, cal que calculem els corresponents valors de D_*^2 , que s'obtenen a partir del paràmetre de no centralitat Δ . Així doncs, donat el valor mostrat $\hat{Y}_{(15)} = 94.697$ i a partir de la funció de distribució $F_{1,43}(\Delta)$, tenim

$$\Delta = \phi_{1,43,\alpha}^{-1}(94.697) = \begin{cases} 53.440269 & \text{si } \alpha = 0.05 \\ 61.239675 & \text{si } \alpha = 0.10 \\ 75.561734 & \text{si } \alpha = 0.25 \\ 93.403857 & \text{si } \alpha = 0.50 \end{cases}$$

Aleshores, com que

$$D_*^2 = \frac{1}{2nk} \Delta$$

obtenim que

$$D_*^2 = \begin{cases} \frac{1}{90} \phi_{1,43,0.05}^{-1}(94.697) = 0.5937807 & \text{si } \alpha = 0.05 \\ \frac{1}{90} \phi_{1,43,0.10}^{-1}(94.697) = 0.6804408 & \text{si } \alpha = 0.10 \\ \frac{1}{90} \phi_{1,43,0.25}^{-1}(94.697) = 0.8395748 & \text{si } \alpha = 0.25 \\ \frac{1}{90} \phi_{1,43,0.50}^{-1}(94.697) = 1.0378206 & \text{si } \alpha = 0.50 \end{cases}$$

Per obtenir $[\rho_*^2, \infty)$, necessitem calcular

$$\rho_* = 2\sqrt{6} \tanh^{-1} \left(\frac{\sqrt{D_*^2 + 1} - 1}{D_*} \right)$$

i amb els resultats obtinguts anteriorment sabem que

$$D_* = \begin{cases} \sqrt{0.5937807} = 0.771 & \text{si } \alpha = 0.05 \\ \sqrt{0.6804408} = 0.825 & \text{si } \alpha = 0.10 \\ \sqrt{0.8395748} = 0.916 & \text{si } \alpha = 0.25 \\ \sqrt{1.0378206} = 1.019 & \text{si } \alpha = 0.50 \end{cases}$$

aleshores

$$\rho_* = \begin{cases} 2\sqrt{6} \tanh^{-1} \left(\frac{\sqrt{0.771^2 + 1} - 1}{0.771} \right) = 1.7338 & \text{si } \alpha = 0.05 \\ 2\sqrt{6} \tanh^{-1} \left(\frac{\sqrt{0.825^2 + 1} - 1}{0.825} \right) = 1.8422 & \text{si } \alpha = 0.10 \\ 2\sqrt{6} \tanh^{-1} \left(\frac{\sqrt{0.916^2 + 1} - 1}{0.916} \right) = 2.0103 & \text{si } \alpha = 0.25 \\ 2\sqrt{6} \tanh^{-1} \left(\frac{\sqrt{1.019^2 + 1} - 1}{1.019} \right) = 2.1917 & \text{si } \alpha = 0.50 \end{cases}$$

Així, els intervals de confiança de la forma $[\rho_*^2, \infty)$ obtinguts per a $\rho^2(\Theta_0, (\beta, \sigma))$ al $(1 - \alpha)\%$ seran

$$[\rho_*^2, \infty) = \begin{cases} [1.7738^2, \infty) = [3.0234, \infty) & \text{al } 95\% \\ [1.8422^2, \infty) = [3.3937, \infty) & \text{al } 90\% \\ [2.0103^2, \infty) = [4.0413, \infty) & \text{al } 75\% \\ [2.1917^2, \infty) = [4.8035, \infty) & \text{al } 50\% \end{cases}$$

Sabem doncs, que acceptarem H_0 si

$$m\rho_*^2 \leq \text{mediana}(m\rho^2(\beta_{(m)}^*, \sigma_{(m)}^*; \beta, \sigma))$$

i, en altre cas, la rebutjarem.

Llavors a partir de la Taula 6.1.1, de la secció anterior, per a $n = 3$, $s = 2$ i sabent que

$$m\rho_*^2 = \begin{cases} m 3.0234 & \text{si } \alpha = 0.05 \\ m 3.3937 & \text{si } \alpha = 0.10 \\ m 4.0413 & \text{si } \alpha = 0.25 \\ m 4.8035 & \text{si } \alpha = 0.50 \end{cases}$$

el criteri de decisió quedarà

$$\text{Només si } m \leq 1 \implies \text{acceptarem } H_0, \quad \forall \alpha$$

En conclusió, en un cas tant extrem, en el qual les diferències entre els sous són tant evidents, es rebutjarà la hipòtesi nul·la, per qualsevol valor de α , sempre que

tinguem mostres de grandària superior a 1. El mateix que passaria si utilitzéssim qualsevol dels dos tests clàssics, amb qualsevol nivell de significació ε .

Així doncs, veiem que en situacions no extremes, el mètode $-pf$ ens dóna informació adicional per decidir si seria millor, en cert sentit, treballar amb el submodel o amb el model general, depenent de la grandària mostral de la qual poguéssim disposar. Notem també que, en qualsevol situació, aquest nou mètode no dóna mai criteris de decisió contradictoris amb els que donarien els tests clàssics, sinó criteris addicionals.

Capítol 7

Altres situacions

En aquest darrer exemple estudiarem el comportament dels mètodes proposats quan no es compleix una de les suposicions bàsiques: quan no resulta adequat suposar que la probabilitat *a priori* corresponent a la hipòtesi nul·la, o submodel, és zero.

7.1 Quan $\dim \Theta_0 = \dim \Theta$

Per exemple podríem considerar el cas clàssic de Neyman-Pearson, amb dues hipòtesis simples, que en un model identificable podríem formular com

$$H_0 : \theta = \theta_0, \quad H_1 : \theta = \theta_1$$

tot i que nosaltres l'estudiarem en el sentit que ho hem fet fins ara, és a dir, decidir entre el model general que és, en aquest cas, $\Theta = \{\theta_0, \theta_1\}$ i el submodel, corresponent a la hipòtesi nul·la que vindrà donat per $\Theta_0 = \{\theta_0\}$, i per tant, es compleix que $\dim \Theta_0 = \dim \Theta = 0$.

Suposarem, a més, una *a priori* desconeguda tal que $\Pi(H_0) = \lambda$, $\Pi(H_1) = 1 - \lambda$, amb $\lambda > 0$, i considerarem una distància arbitrària $\delta^2 = d^2(\theta_0, \theta_1)$ (per exemple la distància de Hellinger), amb $\delta > 0$ donat que $\theta_0 \neq \theta_1$.

Donada una mostra de mida m , $x = (x_1, \dots, x_m) \in \chi^m$, on χ és l'espai mostral, considerem un estimador per a aquest model donat per

$$\mathcal{U}_m(x) = \begin{cases} \theta_0 & \text{si } x \in S_m^c \\ \theta_1 & \text{si } x \in S_m \end{cases}$$

on $S_m \subset \chi^m$, és una regió que compleix:

$$P(x \in S_m | H_0) = \nu_m \quad \text{i} \quad P(x \in S_m | H_1) = \eta_m$$

amb, de manera raonable, $1 > \nu_m > 0$, $1 > \eta_m > 0$.

7.1.1 mètode-pf

Calculem la distribució de $d^2(\mathcal{U}_m(x), \theta)$, $\theta \in \Theta$, on θ és el vertader valor del paràmetre.

$$d^2(\mathcal{U}_m(x), \theta = \theta_0) = \begin{cases} 0 & \text{si } x \in S_m^c \quad \text{amb prob. : } 1 - \nu_m \\ \delta^2 & \text{si } x \in S_m \quad \text{amb prob. : } \nu_m \end{cases}$$

$$d^2(\mathcal{U}_m(x), \theta = \theta_1) = \begin{cases} \delta^2 & \text{si } x \in S_m^c \quad \text{amb prob. : } 1 - \eta_m \\ 0 & \text{si } x \in S_m \quad \text{amb prob. : } \eta_m \end{cases}$$

Considerem ara l'estadístic:

$$d^2(\Theta_0, \mathcal{U}_m(x)) = d^2(\theta_0, \mathcal{U}_m(x)) = \begin{cases} 0 & \text{si } x \in S_m^c \\ \delta^2 & \text{si } x \in S_m \end{cases}$$

que és un estimador de:

$$d^2(\Theta_0, \theta) = d^2(\theta_0, \theta) = \begin{cases} 0 & \text{si } \theta = \theta_0 \\ \delta^2 & \text{si } \theta = \theta_1 \end{cases}$$

Volem calcular:

$$p = P(d^2(\mathcal{U}_m(x), \theta) < d^2(\Theta_0, \theta)) - P(d^2(\mathcal{U}_m(x), \theta) > d^2(\Theta_0, \theta))$$

Però observem que hi ha dos possibles valors de p dependent de la vertadera mesura de probabilitat, això és

$$\begin{aligned} p(\theta_0) &= P(d^2(\mathcal{U}_m(x), \theta_0) < 0 | H_0) - P(d^2(\mathcal{U}_m(x), \theta_0) > 0 | H_0) \\ &= -P(x \in S_m | H_0) = -\nu_m \end{aligned}$$

$$\begin{aligned} p(\theta_1) &= P(d^2(\mathcal{U}_m(x), \theta_1) < \delta^2 | H_1) - P(d^2(\mathcal{U}_m(x), \theta_1) > \delta^2 | H_1) \\ &= P(x \in S_m | H_1) = \eta_m \end{aligned}$$

aleshores si $p > 0$, rebutjarem la hipòtesi nul·la i acceptarem el model general Θ , i si $p \leq 0$ llavors podrem acceptar la hipòtesi nul·la i, per tant, el submodel Θ_0 .

Desenvoluparem el mètode- \mathbf{pf} donant un interval de confiança de la forma $[\mathbf{p}_*, 1]$, on \mathbf{p}_* és una estimació de \mathbf{p} que s'obté a partir d'una mostra aleatòria de mida k , $x \in \chi^k$, i d'una regió crítica $W_k \subset \chi^k$, de la següent manera

$$\mathbf{p}_*(x) = \begin{cases} \mathbf{p}(\theta_0) = -\nu_m & \text{si } x \in W_k^c \\ \mathbf{p}(\theta_1) = \eta_m & \text{si } x \in W_k \end{cases}$$

de manera que la seva confiança, en un sentit generalitzat, s'obté a partir de

$$P(\mathbf{p}(\theta_0) \in [\mathbf{p}_*, 1] | H_0) = P(-\nu_m \in [\mathbf{p}_*, 1] | H_0) = P(W_k^c | H_0) = 1 - \alpha_k$$

$$P(\mathbf{p}(\theta_1) \in [\mathbf{p}_*, 1] | H_1) = P(\eta_m \in [\mathbf{p}_*, 1] | H_1) = 1$$

és a dir, tindrem un nivell de confiança, com a mínim, de $1 - \alpha_k$.

Així doncs, resumint, el criteri de decisió quedarà:

- Si $x \in W_k$ i $\eta_m > 0$, rebutjarem la hipòtesi nul·la.
- Altrament, acceptarem la hipòtesi nul·la.

Per altra banda, si $\eta_m > 0$ i definim la regió crítica W_k a partir del lema de Neymann-Pearson, això és, per a qualsevol $k \in \mathbb{N}$

$$W_k = \{x \in \chi^k : f(x, \theta_1) \geq C_k f(x, \theta_0)\}$$

per a constants reals $C_k > 0$ convenients i amb un nivell de significació igual a α_k , (i utilitzant un test aleatoritzat en cas que sigui necessari), aleshores el mètode- \mathbf{pf} és, essencialment, la regla de Neymann-Pearson.

Notem que la distància entre θ_0 i θ_1 , i gairebé l'estimador \mathcal{U}_m (doncs només cal que $\eta_m > 0$), no juguen cap paper, essent per tant totes les distàncies equivalents.

7.1.2 mètode- \mathbf{mf}

En aquest cas tenim

$$\mathbf{m}(\theta_0) = -E(d^2(\mathcal{U}_m(x), \theta_0) | H_0) = -P(S_m | H_0)\delta^2 = -\nu_m\delta^2$$

i

$$\mathbf{m}(\theta_1) = \delta^2 - E(d^2(\mathcal{U}_m(x), \theta_1) | H_1) = \delta^2 - P(S_m^c | H_1)\delta^2 = \eta_m\delta^2$$

això és, hi ha dos possibles valors per a m , depenent de la vertadera mesura de probabilitat. Per altra banda, donada una mostra aleatòria de mida k , $x \in \mathcal{X}^k$, l'interval de confiança aleatori $[m_*, \infty)$, on

$$m_*(x) = \begin{cases} -\delta^2 \nu_m & \text{si } x \in W_k^c \\ \delta^2 \eta_m & \text{si } x \in W_k \end{cases}$$

té una confiança de, com a mínim, $1 - \alpha_k$. Aleshores, rebutjarem la hipòtesi nul·la H_0 si $x \in W_k$ i $\eta_m > 0$, acceptant-la altrament, això és si $x \in W_k^c$ o bé si $\eta_m = 0$.

Observem que el comportament del mètode m_f és idèntic al del mètode anterior i , per tant, es recupera el mètode de Neyman-Pearson si $\eta_m > 0$ i prenem la regió crítica W_k com abans.

7.1.3 mètode p_b

Donada una mostra de mida k , $x \in \mathcal{X}^k$, hem de calcular la probabilitat *a posteriori* sobre Θ

$$\gamma_k(x) = P(H_0 | x) \quad \text{i} \quad 1 - \gamma_k(x) = P(H_1 | x)$$

Aleshores hem d'agafar la mitjana de p , respecte a aquesta mesura de probabilitat, obtenint

$$E(p | x) = \eta_m(1 - \gamma_k(x)) - \nu_m \gamma_k(x) = \eta_m - (\nu_m + \eta_m) \gamma_k(x)$$

Aleshores rebutjarem el submodel (hipòtesi nul·la) si $E(p | x) > 0$, acceptant-la altrament. Observem que aquest mètode depèn de la distribució *a priori* considerada.

Comparem ara aquest mètode amb la versió Bayesiana del test de Neyman-Pearson clàssic. Donada la distribució *a priori*: $\Pi(H_0) = \lambda$, $\Pi(H_1) = 1 - \lambda$, és ben conegut que la regió crítica a partir d'una mostra de mida k , $x = (x_1, \dots, x_k)$, donada pel test Bayesià clàssic vindria donada per

$$\begin{aligned} W_k &= \left\{ x : \frac{f(x|\theta_0)}{f(x|\theta_1)} < \frac{1-\lambda}{\lambda} \right\} = \left\{ x : \frac{\Pi(\theta_0) f(x|\theta_0)}{\Pi(\theta_1) f(x|\theta_1)} < 1 \right\} \\ &= \left\{ x : \frac{P(H_0|x)}{P(H_1|x)} < 1 \right\} = \left\{ x : \frac{\gamma_k(x)}{1-\gamma_k(x)} < 1 \right\} \\ &= \left\{ x : \gamma_k(x) < 1/2 \right\} \end{aligned}$$

Notem que, el criteri de decisió donat pel mètode- \mathbf{pb} , el qual rebutja la hipòtesi nul·la quan $E(\mathbf{p} | x) > 0$, es pot reescriure també dient que rebutjarem la hipòtesi nul·la o submodel si

$$\gamma_k(x) < \frac{\eta_m}{\nu_m + \eta_m}$$

Hem vist, doncs, que el comportament del mètode- \mathbf{pb} és similar al test Neyman-Pearson clàssic Bayesià, i en el cas que es pugui obtenir un estadístic amb el qual $\nu_m \simeq \eta_m$, els criteris són pràcticament equivalents.

7.1.4 mètode- \mathbf{mb}

És similar al cas previ, però ara hem de considerar la mitjana de m , respecte de la mesura de probabilitat *a posteriori*, obtenint

$$E(m | x) = \delta^2 \eta_m (1 - \gamma_k(x)) - \delta^2 \nu_m \gamma_k(x) = \delta^2 (\eta_m - (\nu_m + \eta_m) \gamma_k(x))$$

Aleshores rebutjarem la hipòtesi nul·la si $E(m | x) > 0$, acceptant-la altrament.

Observem que el criteri de decisió del mètode- \mathbf{mb} , donat que $\delta^2 > 0$, és equivalent al del mètode anterior, i conseqüentment similar també al test Bayesià de Neyman-Pearson. En aquest cas, el mètode també depèn de la *a priori* escollida i, en canvi, no depèn de la distància considerada.

Notem, finalment, que fins i tot sense complir-se les suposicions bàsiques per a l'aplicació dels mètodes considerats en aquest treball, aquests encara proporcionen regles raonables i similars als tests clàssics, la qual cosa ens suggereix la seva robustesa.

7.1.5 Un cas particular quan la hipòtesi nul·la no és simple

Considerarem la distribució Normal univariant, $X \sim N(\mu, \sigma_0^2)$, amb $\sigma_0 > 0$ coneguda. Volem contrastar:

$$H_0 : \mu \leq \mu_0 \quad H_1 : \mu > \mu_0$$

Utilitzarem el mètode- \mathbf{pf} per desenvolupar aquest exemple en el qual també es compleix que $\dim \Theta = \dim \Theta_0$, donat que:

$$\Theta = (-\infty, \infty) = \mathbb{R} \quad \Theta_0 = (-\infty, \mu_0]$$

La distància de Rao, en aquest model estadístic paramètric, ve donada per

$$\rho(\mu_1, \mu_2) = \frac{|\mu_1 - \mu_2|}{\sigma_0}$$

i

$$\rho(\Theta_0, \mu) = \inf_{\theta_0 \in \Theta_0} \rho(\theta_0, \mu) = \begin{cases} 0 & \text{si } \mu \leq \mu_0 \\ \rho(\mu_0, \mu) = \frac{\mu - \mu_0}{\sigma_0} & \text{si } \mu > \mu_0 \end{cases}$$

En aquest cas, donada un m.a.s. de mida m , $x = (x_1, \dots, x_m) \in \chi^m$ i considerant l'estimador màxim-versemblant $\mathcal{U}_m = \bar{X}_m$ el criteri de decisió quedarà tal que si

$$p = 2P(\rho(\bar{X}_m, \mu) < \rho(\Theta_0, \mu)) - 1 > 0$$

aleshores rebutjarem H_0 .

Observem que

$$p = \begin{cases} -1 & \text{si } \mu \leq \mu_0 \\ 2P\left(\frac{|\bar{X}_m - \mu|}{\sigma_0} < \frac{\mu - \mu_0}{\sigma_0}\right) - 1 & \text{si } \mu > \mu_0 \end{cases}$$

En el cas $\mu > \mu_0$, podem escriure

$$\begin{aligned} P\left(\frac{|\bar{X}_m - \mu|}{\sigma_0} < \frac{\mu - \mu_0}{\sigma_0}\right) &= P\left(\sqrt{m} \frac{|\bar{X}_m - \mu|}{\sigma_0} < \sqrt{m} \frac{(\mu - \mu_0)}{\sigma_0}\right) \\ &= P\left(|Y| < \sqrt{m} \frac{(\mu - \mu_0)}{\sigma_0}\right) \end{aligned}$$

que és una funció creixent de μ , doncs sabem que

$$Y = \sqrt{m} \frac{(\bar{X}_m - \mu)}{\sigma_0} \sim N(0, 1)$$

Aleshores un interval de la forma $[p_*, 1]$ amb confiança $(1 - \alpha)$ s'obté considerant $[p(\mu_*), \infty)$, on μ_* serà una estimació per defecte de μ obtinguda a partir d'una m.a.s. de grandària k , i de manera que $P(\mu \geq \mu_*) = 1 - \alpha$. Amb aquest objectiu, notem també com abans, que

$$\sqrt{k} \frac{(\bar{X}_k - \mu)}{\sigma_0} \sim N(0, 1)$$

Aleshores definim ζ_α a partir de

$$P(Z \leq \zeta_\alpha) = 1 - \alpha$$

on $Z \sim N(0, 1)$, de manera que

$$P\left(\sqrt{k} \frac{(\bar{X}_k - \mu)}{\sigma_0} \leq \zeta_\alpha\right) = P\left(\mu \geq \bar{X}_k - \frac{\sigma_0}{\sqrt{k}} \zeta_\alpha\right) = 1 - \alpha$$

Llavors considerarem

$$\mu_* = \bar{X}_k - \frac{\sigma_0}{\sqrt{k}} \zeta_\alpha$$

i, per tant, podrem escriure

$$p(\mu_*) = \begin{cases} -1 & \text{si } \bar{X}_k - \frac{\sigma_0}{\sqrt{k}} \zeta_\alpha \leq \mu_0 \\ 2P\left(|Y| < \sqrt{m} \frac{(\bar{X}_k - \frac{\sigma_0}{\sqrt{k}} \zeta_\alpha) - \mu_0}{\sigma_0}\right) - 1 & \text{si } \bar{X}_k - \frac{\sigma_0}{\sqrt{k}} \zeta_\alpha > \mu_0 \end{cases}$$

Aleshores el criteri de decisió quedarà de manera que rebutjarem H_0 si $p(\mu_*) > 0$, és a dir, si

$$P\left(|Y| < \sqrt{m} \frac{(\bar{X}_k - \frac{\sigma_0}{\sqrt{k}} \zeta_\alpha) - \mu_0}{\sigma_0}\right) > 1/2$$

Com sabem que $Y \sim N(0, 1)$, la desigualtat anterior es complirà si

$$\sqrt{m} \frac{(\bar{X}_k - \frac{\sigma_0}{\sqrt{k}} \zeta_\alpha) - \mu_0}{\sigma_0} > \zeta_{0.25} = 0.67449$$

Així doncs, rebutjarem H_0 si

$$\sqrt{\frac{m}{k}} \left(\sqrt{k} \frac{(\bar{X}_k - \mu_0)}{\sigma_0} - \zeta_\alpha \right) > 0.67449$$

o, de manera equivalent, si

$$\sqrt{k} \frac{(\bar{X}_k - \mu_0)}{\sigma_0} > \sqrt{\frac{k}{m}} 0.67449 + \zeta_\alpha$$

Aleshores observem que si m és gran, el mètode-pf pràcticament equivalent al test clàssic (extensió per a hipòtesis compostes del test de Neyman-Pearson), quan considerem un nivell de significació igual al valor α que hem utilitzat per a la confiança. Mentre que si tenim mostres de grandària m petita, quan rebutgem

amb el mètode \mathbf{pf} rebutjaríem també amb el test clàssic, i si rebutgèssim amb el test clàssic, amb el mètode \mathbf{pf} encara podríem acceptar quan

$$\sqrt{\frac{m}{k}} > \frac{0.67449}{\left(\sqrt{k} \frac{(\bar{X}_k - \mu_0)}{\sigma_0} - \zeta_\alpha\right)}$$

Això ens mostra doncs, una vegada més, la robustesa d'aquest mètode quan partim d'unes suposicions diferents a les suposicions bàsiques plantejades a l'inici d'aquest treball per desenvolupar la nova metodologia.

Capítol 8

Resum de resultats i conclusions

Aquest capítol es diferenciarà en dues seccions. En la primera resumirem els principals resultats obtinguts al llarg de tot el treball, i en la segona es donarà un perspectiva de com es podria ampliar o generalitzar la nova metodologia en l'estudi dels tests d'hipòtesis.

8.1 Resum de resultats

Cal observar que la característica diferencial d'aquest treball no és tant donar una sèrie de resultats particulars sobre els tests d'hipòtesis, sinó donar un nou plantejament conceptual dels mateixos, des de la perspectiva de la selecció de models i fent servir eines d'estimació puntual. Resumirem el contingut d'aquesta memòria en els següents punts:

- Es posarà de relleu alguns problemes i paradoxes relacionades amb la teoria clàssica dels tests d'hipòtesis, i es donarà una metodologia alternativa, basada en distàncies sobre els models estadístics, que evitarà o, si més no, obrirà un nou camí per resoldre aquestes qüestions.
- Recordarem que, en els models estadístics paramètrics regulars, les mesures probabilístiques indueixen una mètrica Riemanniana en la varietat de paràmetres Θ , també coneguda com distància de Rao, passant a tenir doncs una varietat Riemanniana. Veurem, a més, algunes noves propietats de la distància de Rao.
- Es tracta el problema de contrastar:

$$H_0 : \theta \in \Theta_0 \quad H_1 : \theta \notin \Theta_0$$

on Θ_0 una subvarietat de Θ tal que $0 \leq \dim \Theta_0 < \dim \Theta$, des de la perspectiva de la selecció del model més senzill Θ_0 , front el model més general Θ , acceptant l'existència d'un *estimador*, com per exemple l'estimador màxim-versemblant, com a estimador convenient del model general.

- La mesura d'error estadístic es dona, com ja hem comentat, en termes de la distància Riemanniana. Així si anomenem θ al vertader valor del paràmetre, l'error d'una estimació, donada una mostra de grandària m , ve donat per $\rho^2(\mathcal{U}_m(x), \theta)$, mentre que l'error corresponent a la hipòtesi nul·la (o submodel) ve donat per $\rho^2(\Theta_0, \theta) = \inf\{\rho^2(\nu, \theta) : \nu \in \Theta_0\}$
- Es proposen quatre nous possibles mètodes, que comparteixen una mateixa filosofia de fons, per resoldre els tests d'hipòtesis plantejats anteriorment. Aquests s'anomenen mètodes-**pf**, **pb**, **mf**, **mb**, que resulten de la combinació de dos procediments, un freqüentista **f** i un Bayesià **b**, i dues regles, una basada en probabilitats **p** i una basada en moments **m**. Més endavant, i a partir d'uns exemples, es veu com, lògicament, els quatre mètodes donen regles de decisió semblants, podent-se aplicar indistintament un o altre.
- S'exposen els següents resultats asimptòtics, fent amb detall la seva demostració: Donada Θ una varietat Riemanniana, C^∞ , real, n -dimensional, connexa i geodèsicament completa i $\mathcal{U}_k(x)$ una successió d'estimadors màxim-versemblants de $\theta \in \Theta$, aleshores

a) Si $\Theta_0 = \{\theta_0\}$, amb $\theta_0 \neq \theta$

$$\sqrt{k} (\rho(\mathcal{U}_k(x), \theta_0) - \rho(\theta, \theta_0)) \xrightarrow{\mathcal{L}} Z \sim N(0, 1)$$

b) Si Θ_0 és una subvarietat de Θ r -dimensional, amb $r < n$ i sota la hipòtesi nul·la $\theta \in \Theta_0$

$$k \rho^2(\mathcal{U}_k(x), \Theta_0) \xrightarrow{\mathcal{L}} \mathcal{Z} \sim \chi_{n-r}^2$$

- Primerament s'aplica la nova metodologia al model Normal n -variant amb Σ coneguda, en aquest cas es desenvolupen els quatre mètodes, es comparen entre ells, donant resultats semblants, i amb els mètodes clàssics.
- A continuació es dona una aproximació asimptòtica del mètode-**pf** pel cas general amb hipòtesi nul·la simple.
- Després es desenvolupa el mètode-**mb** en els models uniparamètrics Poisson i Gamma.

- Finalment es desenvolupa el mètode \mathcal{P}_f en un model lineal normal general i s'aplica a un cas particular.
- En tots els exemples es comprova com els resultats obtinguts no contradueixen els obtinguts amb la metodologia clàssica, sinó que s'obtenen criteris equivalents, en alguns casos, i els complementen o amplien en altres. Cal remarcar que aquests criteris de decisió es donen en funció d'una grandària mostral m que poguéssim tenir, d'acord amb la informació obtinguda a partir d'una mostra de grandària k . En certa manera doncs, aquesta metodologia ens dóna més informació.
- Per últim, es comprova com actuen aquests mètodes quan no es compleix una de les suposicions bàsiques plantejades en el tercer punt: quan no podem suposar que la distribució a priori del submodel és zero. Concretament es desenvolupen alguns exemples en els quals $\dim \Theta_0 = \dim \Theta$, obtenint-se resultats o bé idèntics, o bé semblants als que donarien els test clàssics.

8.2 Perspectives de la metodologia proposada

Ja s'ha comentat al llarg d'aquest treball que, si bé aquí s'ha utilitzat la distància Riemanniana, es poden estendre els resultats obtinguts a qualsevol altra distància sobre models estadístics paramètrics. Així doncs, es pot mantenir la mateixa filosofia metodològica, però usant una altra mesura d'error estadístic.

Hem vist com aquest nou plantejament metodològic dels tests d'hipòtesis dóna bons resultats quan considerem una classe concreta d'estimadors, en aquest cas, els estimadors màxim-versemblants. És a dir, els criteris de decisió obtinguts depenen, o són funció, de l'estimador usat. Però semblaria raonable triar entre tot el conjunt d'estimadors possibles el que minimitzés l'error estadístic en el sentit anteriorment descrit, és a dir formular aquests nous mètodes prenent l'ínfim, dintre de la classe de tots els possibles estimadors, de les regles de decisió proposades. Queda, doncs, pendent per a treballs posteriors, el ampliar totes aquestes regles de decisió de manera que siguin independents de l'estimador usat, o fins i tot, triar l'estimador que faci òptima la regla de decisió.

Semblaria també raonable estendre aquesta nova filosofia metodològica a tot l'àmbit de la inferència estadística. És a dir, tractar d'una forma unificada els problemes d'estimació puntual, predicció, test d'hipòtesis i selecció de models, utilitzant en tots els casos unes mateixes eines metodològiques per mesurar l'error estadístic, o bé, optimitzar un criteri de decisió. De fet, en aquesta memòria s'intenta treballar en aquesta línia i s'obre un camí alternatiu per resoldre alguns problemes estadístics.

Capítol 9

Apèndix

9.1 Varietats i funcions C^∞

En aquest apèndix es presenten una sèrie de definicions i resultats propis de la geometria diferencial que són necessaris per demostrar moltes de les definicions que es fan en aquest treball. Es pot trobar informació addicional a O'Neill [36], Amari [3], Spivak [46], Hicks [23], Chavel [12] i Karcher [26] entre altres.

En un sentit ampli i general, una varietat és un espai topològic que, localment, s'assembla o es podria identificar amb un espai Euclidià. Una varietat C^∞ és una varietat Θ tal que aquesta semblança és suficientment forta per poder establir en ella la derivació parcial i, de fet, totes les eines essencials del càlcul. Donarem, però, unes definicions formals dels conceptes que utilitzem.

Definició 9.1.1 *Un sistema de coordenades en un espai topològic \mathcal{S} és un homeomorfisme ϕ d'un conjunt obert U de \mathcal{S} a un altre conjunt obert $\phi(U)$ de \mathbb{R}^n . Si escrivim*

$$\phi(p) = (\phi^1(p), \dots, \phi^n(p)) \quad \forall p \in U$$

on les funcions ϕ^1, \dots, ϕ^n s'anomenen les funcions coordenades de ϕ . Llavors

$$\phi = (\phi^1, \dots, \phi^n) : U \longrightarrow \mathbb{R}^n$$

Notem que $\phi^i = u^i \circ \phi$, on u^i són les projeccions naturals sobre \mathbb{R}^n

$$\begin{aligned} u^i : \quad \mathbb{R}^n &\longrightarrow \mathbb{R} \\ (y_1, \dots, y_n) &\longrightarrow y_i \end{aligned}$$

Respecte al canvi de coordenades podem dir que, dos sistemes de coordenades n -dimensionals sobre un espai topològic \mathcal{S} , $\phi : U \rightarrow \mathbb{R}^n$ i $\varphi : V \rightarrow \mathbb{R}^n$, tenen

una intersecció de classe C^∞ un respecte l'altre si les funcions $\phi \circ \varphi^{-1}$ i $\varphi \circ \phi^{-1}$ són ambdues de classe C^∞ , en els respectius dominis de definició, $\varphi(U \cap V)$ i $\phi(U \cap V)$.

Definició 9.1.2 *Un atlas \mathcal{A} de dimensió n sobre un espai topològic \mathcal{S} és una col·lecció de sistemes de coordenades n -dimensionals en \mathcal{S} tal que*

- a) *cada punt de \mathcal{S} està contingut en el domini d'algun sistema de coordenades*
- b) *dos sistemes de coordenades qualsevol en \mathcal{A} tenen una intersecció de classe C^∞ .*

Direm que un atlas \mathcal{A} sobre \mathcal{S} és *complet* si conté tots els sistemes de coordenades en \mathcal{S} amb intersecció de classe C^∞ .

Definició 9.1.3 *Una varietat Θ de classe C^∞ és un espai Hausdorff dotat d'un atlas complet.*

Hi ha altres variants de la definició de varietat, o varietats de classe C^∞ , però nosaltres usarem les definides anteriorment.

La dimensió d'una varietat Θ és la dimensió dels seus atlas. Un sistema de coordenades ϕ en una varietat Θ és un sistema de coordenades que pertany al seu atlas complet. Si el domini U de ϕ conté el punt $\theta \in \Theta$, aleshores ϕ s'anomena sistema de coordenades centrat a θ i U és un entorn de θ amb aquest sistema de coordenades. Si ϕ és un sistema de coordenades a Θ i V és un conjunt obert contingut en el domini de ϕ , aleshores per completitud, $\phi|_V$ és també un sistema de coordenades a Θ .

Considerem ara les funcions de classe C^∞ començant primer pel cas especial de les funcions reals sobre una varietat, això és, $f : \Theta \rightarrow \mathbb{R}$. Si $\phi : U \rightarrow \mathbb{R}^n$ és un sistema de coordenades a Θ , aleshores la funció composició $f \circ \phi^{-1} : \phi(U) \rightarrow \mathbb{R}$ s'anomenarà *expressió en coordenades per a f en termes de ϕ* . Aleshores podem donar la següent definició

Definició 9.1.4 *Direm que una funció $f : \Theta \rightarrow \mathbb{R}$ és de classe C^∞ si per a cada sistema de coordenades ϕ a Θ , la composició $f \circ \phi^{-1}$ és de classe C^∞ en el sentit Euclidià usual.*

Notarem

$$\mathcal{F}(\Theta) = \{f : \Theta \rightarrow \mathbb{R} \mid f \in C^\infty\}$$

Aleshores si $f, g \in \mathcal{F}(\Theta)$ llavors la suma $f + g$ i el producte fg també. Les regles algebraiques usuals funcionaran per a les dues operacions i això farà de $\mathcal{F}(\Theta)$ un anell commutatiu.

Aquesta noció de funció C^∞ per a funcions a valors reals sobre una varietat es pot estendre a funcions entre varietats usant la mateixa idea.

Definició 9.1.5 *Siguin Θ i Σ dues varietats de dimensions m i n respectivament. Una aplicació $f : \Theta \rightarrow \Sigma$ és de classe C^∞ si per a cada sistema de coordenades ϕ a Θ i φ a Σ , la composició $\varphi \circ f \circ \phi^{-1}$ és de classe C^∞ en el sentit Euclidià i definida en conjunt obert de \mathbb{R}^m .*

Definició 9.1.6 *Un difeomorfisme entre dues varietats $f : \Theta \rightarrow \Sigma$ és una aplicació de classe C^∞ tal que té una inversa f^{-1} que és també de classe C^∞ .*

Notem que cada sistema de coordenades ϕ és un difeomorfisme des del seu domini U a $\phi(U) \subset \mathbb{R}^n$. De igual manera, cada difeomorfisme f des d'un conjunt obert $V \subset \Theta$ a $f(V) \subset \mathbb{R}^n$ és un sistema de coordenades de Θ . De fet, per a qualsevol sistema de coordenades ϕ les aplicacions $f \circ \phi^{-1}$ i $\phi \circ f^{-1}$ són de classe C^∞ , i donat que l'atlas de Θ és complet, f hi pertany. Aleshores composant un sistema de coordenades amb un difeomorfisme de \mathbb{R}^n ens dóna novament un sistema de coordenades.

9.2 Vectors i Espais Tangents

Definició 9.2.1 *Sigui θ un punt de la varietat Θ . Un vector tangent a Θ en θ és una funció real $v : \mathcal{F}(\Theta) \rightarrow \mathbb{R}$ tal que*

$$(1) \quad v(af + bg) = av(f) + bv(g)$$

$$(2) \quad v(fg) = v(f)g + fv(g)$$

Per a tot $a, b \in \mathbb{R}$ i $f, g \in \mathcal{F}(\Theta)$, on $\mathcal{F}(\Theta)$ és el conjunt de totes les funcions reals sobre Θ de classe C^∞ .

Així doncs, anomenarem $T_\theta(\Theta)$ a l'espai tangent a Θ en θ , és a dir, serà el conjunt de tots els vectors tangents a Θ en el punt θ .

Definició 9.2.2 *Sigui $\phi = (\phi^1, \dots, \phi^n)$ un sistema de coordenades a Θ en el punt (o centrat en el punt) θ . Si $f \in \mathcal{F}(\Theta)$, llavors*

$$\frac{\partial f}{\partial \phi^i}(\theta) = \frac{\partial (f \circ \phi^{-1})}{\partial u^i}(\phi(\theta)) \quad (1 \leq i \leq n)$$

on u^1, \dots, u^n són les (funcions) coordenades naturals de \mathbb{R}^n .

Un senzill càlcul ens demostra que la funció

$$\partial_i|_\theta = \frac{\partial}{\partial \phi^i} \Big|_\theta : \mathcal{F}(\Theta) \longrightarrow \mathbb{R}$$

que fa correspondre cada $f \in \mathcal{F}(\Theta)$ a $(\partial f / \partial \phi^i)(\theta)$ és un vector tangent a Θ en el punt θ . Gràficament podem dibuixar $\partial_i|_\theta$ com un vector en el punt θ tangent a la corba de coordenades ϕ^i a través de θ .

El següent resultat és fonamental per establir el lligam entre coordenades i vectors tangents.

Teorema 9.2.1 *Si $\phi = (\phi^1, \dots, \phi^n)$ és un sistema de coordenades a Θ en un entorn del punt θ , aleshores els seus vectors tangents $\partial_1|_\theta, \dots, \partial_n|_\theta$ formen una base de l'espai tangent $T_\theta(\Theta)$, i*

$$v = \sum_{i=1}^n v(\phi^i) \partial_i|_\theta \quad \forall v \in T_\theta(\Theta)$$

Demostració:

Considerem un entorn U del punt θ , sense pèrdua de generalitat podem suposar que $\phi(\theta) = 0$. Per a qualsevol $f \in \mathcal{F}(\Theta)$ podem considerar $F = f \circ \phi^{-1}$ que serà una funció de classe C^∞ , almenys en un entorn $W \subset \phi(U)$ del punt $0 \in \mathbb{R}^n$. Aleshores, per un punt $(a_1, \dots, a_n) \in W$ podem escriure

$$\begin{aligned} F(a_1, \dots, a_n) &= F(a_1, \dots, a_n) - F(a_1, \dots, a_{n-1}, 0) + \\ &\quad F(a_1, \dots, a_{n-1}, 0) - F(a_1, \dots, a_{n-2}, 0, 0) + \\ &\quad \dots + F(a_1, 0, \dots, 0) - F(0, \dots, 0) + F(0, \dots, 0) \\ &= \sum_{i=1}^n \left\{ \int_0^1 \frac{\partial F(a_1, \dots, a_{i-1}, ta_i, 0, \dots, 0)}{\partial u^i} a_i dt \right\} + \\ &\quad F(0, \dots, 0) \end{aligned}$$

Si anomenem

$$F_i(a_1, \dots, a_n) = \int_0^1 \frac{\partial F(a_1, \dots, a_{i-1}, ta_i, 0, \dots, 0)}{\partial u^i} a_i dt$$

Podrem expressar

$$F(a_1, \dots, a_n) = \sum_{i=1}^n a_i F_i(a_1, \dots, a_n) + F(0, \dots, 0)$$

Com $a_i = u^i(a_1, \dots, a_n) = (\phi^i \circ \phi^{-1})(a_1, \dots, a_n)$ i recordant que $F = f \circ \phi^{-1}$, podem notar $f_i = F_i \circ \phi$ tenint aleshores la següent expressió per a f i a valors en el obert U

$$f = f(\theta) + \sum_{i=1}^n \phi^i f_i$$

Notem ara que, usant la Definició (9.2.2), podem escriure

$$\begin{aligned} f_i(\theta) &= F_i(\phi(\theta)) = F_i(0, \dots, 0) = \int_0^1 \frac{\partial F(0, \dots, 0)}{\partial u^i} dt \\ &= \frac{\partial F}{\partial u^i}(0, \dots, 0) = \frac{\partial (f \circ \phi^{-1})}{\partial u^i}(\phi(\theta)) = \frac{\partial f}{\partial \phi^i}(\theta) \end{aligned}$$

Aplicant ara el vector tangent v a l'equació que descriu f i usant la igualtat anterior obtenim

$$\begin{aligned} v(f) &= 0 + \sum_{i=1}^n v(f_i) \phi^i(\theta) + \sum_{i=1}^n f_i(\theta) v(\phi^i) \\ &= \sum_{i=1}^n \frac{\partial f}{\partial \phi^i}(\theta) v(\phi^i) = \sum_{i=1}^n v(\phi^i) \partial_i|_{\theta}(f) \end{aligned}$$

Hem vist, doncs, que $\partial_1|_{\theta}, \dots, \partial_n|_{\theta}$ generen $T_{\theta}(\Theta)$. Veiem ara que són linealment independents suposant primer que $\sum_{i=1}^n a_i \partial_i|_{\theta} = 0$, llavors si ho apliquem a la funció ϕ^j tenim

$$0 = \sum_{i=1}^n a_i \frac{\partial \phi^j}{\partial \phi^i}(\theta) = \sum_{i=1}^n a_i \delta_i^j = a_j$$

que és el que volíem veure. ■

Notem que aquest resultat ens mostra també que la dimensió de $T_{\theta}(\Theta)$ és la mateixa que la dimensió de Θ .

Intentarem ara aproximar una aplicació C^{∞} entre dues varietats $f : \Theta \rightarrow \Sigma$ en un entorn de cada punt $\theta \in \Theta$ per una transformació lineal dels seus espais tangents. Notem que si $v \in T_{\theta}(\Theta)$ aleshores la funció $v_f : \mathcal{F}(\Theta) \rightarrow \mathbb{R}$ que envia cada g a $v(g \circ f)$ és un vector tangent a Σ en el punt $f(\theta)$.

Definició 9.2.3 Sigui $f : \Theta \rightarrow \Sigma$ una aplicació de classe C^{∞} . Per a cada $\theta \in \Theta$, la funció

$$df_{\theta} : T_{\theta}(\Theta) \longrightarrow T_{f(\theta)}(\Sigma)$$

que envia v a v_f s'anomena diferencial de f en el punt θ .

Notem que df_θ ve caracteritzada per l'equació

$$df_\theta(v)(g) = v_f(g) = v(g \circ f)$$

per a tot $v \in T_\theta(\Theta)$ i $g \in \mathcal{F}(\Theta)$. De la pròpia definició es dedueix també que les diferencials són funcions lineals.

9.3 Subvarietats

Definirem ara el que s'entén per subvarietat. Simplificant-ho, una subvarietat Θ_0 de Θ és un subconjunt de Θ que adquireix la seva estructura de varietat de la varietat Θ on està continguda. En particular exigirem que Θ_0 sigui un espai topològic de Θ , és a dir, que tingui la topologia induïda, amb la qual cosa un subconjunt $V \subset \Theta_0$ és un obert si i només si existeix un obert $U \subset \Theta$ tal que $U \cap \Theta_0 = V$.

Definició 9.3.1 *Una varietat Θ_0 és una subvarietat d'una varietat Θ quan*

- a) Θ_0 és un espai topològic de Θ .
- b) *La funció inclusió $j : \Theta_0 \hookrightarrow \Theta$ és de classe C^∞ a cada punt $\theta \in \Theta_0$ i la seva aplicació diferencial $dj : T_\theta(\Theta_0) \rightarrow T_\theta(\Theta)$ és bijectiva.*

De fet es considerarà l'espai tangent $T_\theta(\Theta_0)$ com un subespai vectorial de $T_\theta(\Theta)$.

Proposició 9.3.1 *Si Θ_0 és una subvarietat de dimensió k d'una varietat Θ amb dimensió n , ($k < n$), aleshores per a cada punt de la subvarietat existeix una sistema de coordenades de Θ , $\phi : U \rightarrow \mathbb{R}^n$, amb U obert de Θ , de manera que prenent $n - k$ de les funcions coordenades constants, aquest resulti un sistema de coordenades de Θ_0 definit sobre el obert $U \cap \Theta_0$.*

Demostració:

Permutant les coordenades sempre podem suposar que són les $n - k$ últimes coordenades les que prendran un valor constant. Sigui $\phi_{\Theta_0} : U \cap \Theta_0 \rightarrow \mathbb{R}^k$ la restricció de (ϕ^1, \dots, ϕ^k) a $U \cap \Theta_0$.

Com Θ_0 és una subvarietat es pot provar, usant el teorema de la funció inversa, que ϕ_{Θ_0} és un difeomorfisme sobre el seu conjunt (obert) imatge a \mathbb{R}^k , essent doncs un sistema de coordenades. Per a més detalls veure O'Neill [36] o qualsevol altra bibliografia citada al començament del capítol.

■

9.4 Corbes

Una corba en una varietat Θ és una aplicació C^∞ , $\alpha : I \rightarrow \Theta$, on I és un interval obert de la recta real. Com a subvarietat oberta de \mathbb{R} , I té un sistema de coordenades consistent en la funció identitat u de I . A cada $t \in \mathbb{R}$ podem traçar el vector $(d/du)(t) \in T_t(\mathbb{R})$ com el vector unitat en el punt t en la direcció positiva u .

Definició 9.4.1 Sigui $\alpha : I \rightarrow \Theta$ una corba. Es defineix el vector velocitat de α en el punt t com

$$\alpha'(t) = d\alpha \left(\frac{d}{du} \Big|_t \right) \in T_{\alpha(t)}(\Theta)$$

Enunciem a continuació una sèrie de propietats bàsiques que cal tenir en compte a l'hora de manegar corbes

- Per la definició de $d\alpha$, el vector tangent $\alpha'(t)$ aplicat a la funció $f \in \mathcal{F}(\Theta)$ ens dóna

$$\alpha'(t) f = \frac{d(f \circ \alpha)}{du}(t)$$

Aleshores si α és una corba amb $\alpha'(0) = v$, aleshores $v(f) = (d(f \circ \alpha)/dt)(0)$.

- Sigui ϕ^1, \dots, ϕ^n un sistema de coordenades a Θ en el punt $\alpha(t)$ de α . Llavors es pot expressar en coordenades

$$\alpha'(t) = \sum_{i=1}^n \frac{d(\phi^i \circ \alpha)}{du}(t) \partial_i|_{\alpha(t)}$$

- Si $\alpha : I \rightarrow \Theta$ és una corba i $h : J \rightarrow I$ és una funció de classe C^∞ sobre un interval J , aleshores $\beta = \alpha \circ h : J \rightarrow \Theta$ és una corba anomenada reparametrització de α . Aleshores

$$\beta'(s) = \frac{dh}{du}(s) \alpha'(h(s)) \quad \forall s \in J$$

- Si $\alpha : I \rightarrow \Theta$ és una corba a Θ , aleshores una aplicació $f : \Theta \rightarrow \Sigma$ de classe C^∞ transporta α a $f \circ \alpha : I \rightarrow \Sigma$ corba de Σ . A més la funció diferencial de f preserva velocitats, això és

$$df(\alpha'(t)) = (f \circ \alpha)'(t) \quad \forall t \in I$$

Direm que una corba α és *regular* si $\alpha'(t) \neq 0$ per a tot t . Si $[a, b]$ és un interval tancat a \mathbb{R} llavors una *corba segment* és una funció $\alpha : [a, b] \rightarrow \Theta$ tal que té una extensió C^∞ a un interval obert que conté $[a, b]$. Llavors α' està ben definida fins i tot als punts frontera a i b .

Una funció $\beta : [a, b] \rightarrow \Theta$ és una corba segment de classe C^∞ a trossos si existeix una partició de $[a, b]$, $a = t_0 < t_1 < \dots < t_{k+1} = b$, tal que cada $\beta|_{[t_i, t_{i+1}]}$ és una corba segment. Per a un interval obert I , $\beta : I \rightarrow \Theta$ és de classe C^∞ a trossos si per a tots $a < b$ de I la restricció $\beta|_{[a, b]}$ és C^∞ a trossos.

Definició 9.4.2 *Direm que una varietat és connexa si i només si dos punts qualssevol de la varietat es poden unir per una corba segment de classe C^∞ a trossos.*

9.5 Camps vectorials

Un *camp vectorial* V sobre una varietat Θ és una funció que assigna a cada punt $\theta \in \Theta$ un vector tangent V_θ a Θ en cada punt θ . Intuitivament, V és una col·lecció de vectors, un cada punt de Θ . Si V és una camp vectorial sobre Θ i $f \in \mathcal{F}(\Theta)$, aleshores Vf denota la funció a valors reals sobre Θ donada per

$$(Vf)(\theta) = V_\theta(f) \quad \forall \theta \in \Theta$$

Aleshores V és C^∞ si Vf és C^∞ per a tot $f \in \mathcal{F}(\Theta)$.

Els camps vectorials sobre Θ es poden sumar, o multiplicar per una funció $f \in \mathcal{F}(\Theta)$, de la següent manera

$$(fV)_\theta = f(\theta) V_\theta$$

$$(V + W)_\theta = V_\theta + W_\theta \quad \forall \theta \in \Theta$$

Si V i W són camps vectorials C^∞ , llavors els camps vectorials $V + W$ i fV també ho són.

Si $\phi = (\phi^1, \dots, \phi^n)$ és un sistema de coordenades sobre $U \subset \Theta$, aleshores per a cada $1 \leq i \leq n$ el camp vectorial ∂_i sobre U envia cada θ a $\partial_i|_\theta$. Aquests camps vectorials són C^∞ , donat que $\partial_i(f) = \partial f / \partial \phi^i$. A més per a qualsevol camp vectorial V sobre U podem escriure

$$V = \sum_{i=1}^n V \phi^i \partial_i$$

Definició 9.5.1 Una derivada sobre $\mathcal{F}(\Theta)$ és una funció $D : \mathcal{F}(\Theta) \rightarrow \mathcal{F}(\Theta)$ tal que

- (1) $D(a f + b g) = a D(f) + b D(g)$, per $a, b \in \mathbb{R}$
- (2) $D(f g) = D(f) g + f D(g)$

La definició de vector tangent ens mostra que per a un camp vectorial V de classe C^∞ la funció $f \rightarrow Vf$ és una derivada sobre $\mathcal{F}(\Theta)$. De la mateixa manera, cada derivada D sobre $\mathcal{F}(\Theta)$ prové d'un camp vectorial. De fet, per a cada $\theta \in \Theta$ es defineix $V_\theta : \mathcal{F}(\Theta) \rightarrow \mathbb{R}$ per $V_\theta(f) = D(f)(\theta)$. Les anteriors propietats (1) i (2) de la derivada impliquen que V_θ és un vector tangent a Θ en el punt θ ; llavors V és un camp vectorial ben definit sobre Θ . Però $Vf = D(f) \in \mathcal{F}(\Theta)$ per a tot $f \in \mathcal{F}(\Theta)$, per tant V és C^∞ i determina la derivada D . Considerarem doncs, que els camps vectorials són derivades sobre $\mathcal{F}(\Theta)$.

9.6 Varietats Riemannianes

Definició 9.6.1 Un tensor mètric g sobre una varietat Θ de classe C^∞ és un camp tensorial sobre Θ simètric, no degenerat i dues vegades covariant.

En altres paraules, g assigna a cada punt $\theta \in \Theta$ un producte escalar sobre l'espai tangent $T_\theta(\Theta)$.

Definició 9.6.2 Una varietat Riemanniana és una varietat Θ de classe C^∞ dotada d'un tensor mètric g .

Usarem alternativament la notació $\langle \cdot, \cdot \rangle$ per a g , escrivint doncs $g(v, w) = \langle v, w \rangle \in \mathbb{R}$ per a vectors tangents, i $g(V, W) = \langle V, W \rangle \in \mathcal{F}(\Theta)$ per a camps vectorials.

Si ϕ^1, \dots, ϕ^n és un sistema de coordenades sobre $U \in \Theta$, les components del tensor mètric sobre U són

$$g_{ij} = \langle \partial_i, \partial_j \rangle \quad (1 \leq i, j \leq n)$$

Aleshores pels camps vectorials $V = \sum V^i \partial_i$ i $W = \sum W^j \partial_j$

$$g(V, W) = \langle V, W \rangle = \sum g_{ij} V^i W^j$$

Donat que g és no degenerat, a cada punt $\theta \in U$ la matriu $(g_{ij}(\theta))$ és invertible, i la seva matriu inversa es nota $(g^{ij}(\theta))$. La fórmula usual per a la inversa d'una matriu ens demostra que les funcions g^{ij} són C^∞ sobre U .

Donat que g és simètrica, $g_{ij} = g_{ji}$ i també $g^{ij} = g^{ji}$, per a $1 \leq i, j \leq n$. Finalment el tensor mètric sobre U es pot escriure com

$$g = \sum g_{ij} d\phi^i \otimes d\phi^j$$

i l'element mètric diferencial serà

$$ds^2 = \sum g_{ij} d\phi^i d\phi^j$$

9.7 Varietats amb connexió

Siguin V i W camps vectorials sobre una varietat Riemanniana Θ . L'objectiu d'aquesta secció serà definir un nou camp vectorial $\nabla_V W$ sobre Θ tal que el seu valor a cada punt θ sigui la taxa de canvi de W en la direcció de V_θ .

Definició 9.7.1 *Siguin u^1, \dots, u^n les coordenades naturals sobre \mathbb{R}^n . Si V i $W = \sum W^i \partial_i$ són camps vectorials sobre \mathbb{R}^n , el camp vectorial*

$$\nabla_V W = \sum_{i=1}^n V(W^i) \partial_i$$

s'anomena derivada covariant (natural) de W respecte V .

Definició 9.7.2 *Una connexió ∇ sobre una varietat Θ de classe C^∞ és una funció $\nabla : \Upsilon(\Theta) \times \Upsilon(\Theta) \rightarrow \Upsilon(\Theta)$ tal que, si $\Upsilon(\Theta)$ és el conjunt de tots els camps vectorials de classe C^∞ i $V_1, V_2, W_1, W_2 \in \Upsilon(\Theta)$, compleix*

- (1) $\nabla_{aV_1+bV_2}(W) = a\nabla_{V_1}(W) + b\nabla_{V_2}(W)$, amb $a, b \in \mathbb{R}$
- (2) $\nabla_V(W_1 + W_2) = \nabla_V(W_1) + \nabla_V(W_2)$
- (3) $\nabla_V(fW) = (Vf)W + f\nabla_V(W)$, amb $f \in \mathcal{F}(\Theta)$

$\nabla_V W$ s'anomena derivada covariant de W respecte V per a la connexió ∇ .

Sigui ϕ^1, \dots, ϕ^n un sistema de coordenades en un entorn U d'una varietat Riemanniana Θ . Els símbols de Christoffel de la connexió ∇ respecte aquest sistema de coordenades són les funcions reals Γ_{ij}^k de U tals que

$$\nabla_{\partial_i}(\partial_j) = \sum_{k=1}^n \Gamma_{ij}^k \partial_k \quad (1 \leq i, j \leq n)$$

i que es poden expressar per

$$\Gamma_{ij}^k = \frac{1}{2} \sum_{m=1}^n \left\{ \frac{\partial g_{jm}}{\partial \phi^i} + \frac{\partial g_{im}}{\partial \phi^j} - \frac{\partial g_{ij}}{\partial \phi^m} \right\}$$

9.8 Geodèsiques i aplicació exponencial

Generalitzarem ara la noció Euclidiana de línia recta. Una geodèsica en una varietat Riemanniana Θ és una corba $\sigma : I \rightarrow \Theta$ tal que el seu camp vectorial σ' és paral·lel, és a dir amb derivada nul·la. Equivalentment, les geodèsiques són corbes amb acceleració zero: $\sigma'' = 0$. Donem ara una definició més formal de geodèsica

Definició 9.8.1 *Sigui Θ una varietat diferenciable amb connexió ∇ . Una corba $\sigma : (a, b) \rightarrow \Theta$ de classe C^r , $r \geq 2$, és una geodèsica corresponent a ∇ si satisfà l'equació*

$$\frac{\nabla}{dt} \sigma' = 0$$

per a tot $t \in (a, b)$.

Si volem escriure aquesta definició usant coordenades tenim el següent

Corol·lari 9.8.1 *Sigui ϕ^1, \dots, ϕ^n un sistema de coordenades sobre $U \subset \Theta$. Una corba $\sigma : I \rightarrow \Theta$ és una geodèsica de Θ si i només si les seves funcions coordenades $x^k \circ \sigma$ satisfan*

$$\frac{d^2(\phi^k \circ \sigma)}{dt^2} + \sum_{i,j=1}^n \Gamma_{ij}^k \frac{d(\phi^i \circ \sigma)}{dt} \frac{d(\phi^j \circ \sigma)}{dt} = 0 \quad (1 \leq k \leq n)$$

Lema 9.8.1 *Si $v \in T_\theta(\Theta)$ existeix un interval I entorn al 0 i una única geodèsica $\sigma : I \rightarrow \Theta$ tal que $\sigma'(0) = v$.*

Proposició 9.8.1 *Donat qualsevol vector tangent $v \in T_\theta(\Theta)$ existeix una única geodèsica σ_v a Θ tal que*

- (1) *La velocitat inicial de σ_v és v ; això és, $\sigma_v'(0) = v$.*
- (2) *El domini I_v de σ_v és el major possible. És a dir, si $\alpha : J \rightarrow \Theta$ és una geodèsica amb $\alpha'(0) = v$, aleshores $J \subset I$ i $\alpha = \sigma_v|_J$.*

La geodèsica σ_v s'anomena freqüentment *geodèsica maximal*. Una varietat Riemanniana Θ per a la qual cada geodèsica maximal està definida en tota la recta real es diu que és una *varietat geodèsicament completa*.

Definirem ara l'aplicació exponencial en cada punt $\theta \in \Theta$, amb Θ varietat Riemanniana, de manera que amb aquesta aplicació es puguin recollir totes les geodèsiques que comencen, o tenen el seu punt inicial, a θ .

Definició 9.8.2 Sigui Θ una varietat Riemanniana amb $\theta \in \Theta$. Sigui \mathcal{B}_θ el conjunt de vectors $v \in T_\theta(\Theta)$ tals que la seva geodèsica maximal σ_v està definida almenys sobre $[0, 1]$. La funció exponencial de Θ en el punt θ és la funció

$$\exp_\theta : \mathcal{B}_\theta \longrightarrow \Theta$$

tal que $\exp_\theta(v) = \sigma_v(1)$ per a tot $v \in \mathcal{B}_\theta$.

Òbviament \mathcal{B}_θ és el major subconjunt de $T_\theta(\Theta)$ sobre el qual la funció \exp_θ pot ser definida. Si Θ és geodèsicament completa, llavors $\mathcal{B}_\theta = T_\theta(\Theta)$ per a cada punt $\theta \in \Theta$.

Fixat $v \in T_\theta(\Theta)$ i $t \in \mathbb{R}$; aleshores la geodèsica $s \rightarrow \sigma_v(ts)$ té velocitat inicial $t\sigma'_v(0) = tv$. Per tant $\sigma_{tv}(s) = \sigma_v(ts)$ per a tot s i t tals que ambdues geodèsiques estiguin ben definides. En particular, si $v \in \mathcal{B}_\theta$ aleshores

$$\exp_\theta(tv) = \sigma_{tv}(1) = \sigma_v(t)$$

Aleshores l'aplicació exponencial \exp_θ transporta rectes que surten de l'origen de $T_\theta(\Theta)$ a geodèsiques de Θ que comencen en θ . Geomètricament, $\exp_\theta(v)$ és el punt de Θ que s'obté en recórrer la distància unitat, començant des de θ , al llarg de la geodèsica amb vector tangent v en el punt θ .

Proposició 9.8.2 Per a cada punt $\theta \in \Theta$ existeix un entorn U^* de $0 \in T_\theta(\Theta)$ sobre el qual la funció exponencial \exp_θ és un difeomorfisme sobre un entorn U de $\theta \in \Theta$.

Un subconjunt S d'un espai vectorial s'anomena *estrellat* entorn al 0 si quan $v \in S$ això implica que $tv \in S$ per a tot $t \in [0, 1]$. Aleshores S és la unió de segments de línies radials. Si U i U^* són com en la proposició anterior i U^* és estrellat entorn al $0 \in T_\theta(\Theta)$, aleshores U s'anomenarà *entorn normal* de θ . Donem-ne una definició més formal

Definició 9.8.3 Direm que un entorn U de $\theta \in \Theta$ és normal si U és la imatge difeomòrfica, per l'aplicació exponencial, d'un entorn obert estrellat U^* de $0 \in T_\theta(\Theta)$

Amb la següent proposició es prova que un entorn normal U de θ determina de manera única l'entorn estrellat U^* a $T_\theta(\Theta)$.

Proposició 9.8.3 Si U és un entorn normal de $\theta \in \Theta$, aleshores per a cada punt $\nu \in U$ existeix una única geodèsica $\sigma : [0, 1] \rightarrow U$ de θ a ν en U . A més $\sigma'(0) = \exp_\theta^{-1}(\nu) \in U^*$.

Definició 9.8.4 Direm que U és un entorn regular normal de $\theta \in \Theta$ si i només si la seva intersecció amb qualsevol entorn normal de θ segueix essent normal.

En el nostre cas, com Θ és una varietat Riemanniana i usant la Proposició (9.8.2), sabem que per a qualsevol punt de la varietat existeixen aquests tipus d'entorns donat que es pot restringir el difeomorfisme \exp_θ a la intersecció d'un entorn de $0 \in T_\theta(\Theta)$ amb un entorn estrellat de dit punt, i així obtindrem un entorn regular normal de $\theta \in \Theta$.

Sobre qualsevol entorn normal U de $\theta \in \Theta$ existeix un tipus especial de sistema de coordenades que és particularment simple. Sigui e_1, \dots, e_n una base ortonormal per a $T_\theta(\Theta)$, això és $\langle e_i, e_j \rangle = \delta_i^j$.

Definició 9.8.5 $\phi = (\phi^1, \dots, \phi^n)$ és un sistema normal de coordenades determinat per e_1, \dots, e_n si assigna a cada punt $\nu \in U$ el vector que té per coordenades, respecte la base e_1, \dots, e_n , les coordenades corresponents al punt, això és

$$\exp_\theta^{-1}(\nu) \in U^* \subset T_\theta(\Theta)$$

Escrit de forma abreujada

$$\exp_\theta^{-1}(\nu) = \sum_{i=1}^n \phi^i(\nu) e_i \quad (\nu \in U)$$

9.9 Longitud d'arc i distància Riemanniana

La noció familiar de longitud d'arc d'una corba segment en un espai Euclidià és generalitza d'una manera molt senzilla a varietats Riemannianes.

Definició 9.9.1 Sigui $\alpha : [a, b] \rightarrow \Theta$ una corba segment de classe C^∞ a trossos sobre una varietat Riemanniana Θ . La longitud d'arc de α és

$$L(\alpha) = \int_a^b \|\alpha'(t)\| dt$$

Recordem que el tensor mètric g sobre Θ assigna a cada punt de la varietat un producte escalar sobre l'espai tangent en aquest punt, i donat que $\alpha'(t) \in T_{\alpha(t)}(\Theta)$, tenim per definició que $\|\alpha'\| = \langle \alpha', \alpha' \rangle^{1/2}$. Si escrivim això en coordenades $\phi = (\phi^1, \dots, \phi^n)$ tenim

$$\|\alpha'\|^2 = \sum_{i,j=1}^n g_{ij} \frac{d(\phi^j \circ \alpha)}{dt} \frac{d(\phi^i \circ \alpha)}{dt}$$

Lema 9.9.1 *Sigui U un entorn normal de $\theta \in \Theta$ i sigui $\|\cdot\|_\theta$ la norma definida sobre l'espai tangent $T_\theta(\Theta)$. Si σ és la geodèsica radial des de θ fins a $\nu \in U$, aleshores*

$$L(\sigma) = \|\exp_\theta^{-1}(\nu)\|_\theta$$

Demostració:

Si anomenem v a la velocitat inicial de σ , és a dir $\sigma'(0) = v$, sabem per la Proposició (9.8.3) que $v = \exp_\theta^{-1}(\nu)$ i donat que $\|\sigma'\|$ és constant,

$$L(\sigma) = \int_0^1 \|\sigma'\|_\theta dt = \int_0^1 \|v\|_\theta dt = \|v\|_\theta = \|\exp_\theta^{-1}(\nu)\|_\theta$$

■

Lema 9.9.2 *Sigui U un entorn normal del punt $\theta \in \Theta$, amb Θ varietat Riemanniana. Si $\theta \in U$ aleshores el segment geodèsic radial*

$$\sigma : [0, 1] \longrightarrow U$$

des de θ fins a ν és la corba més curta en U que uneix aquests punts i, a més, és única.

Definició 9.9.2 *Per a dos punts qualsevol ν i ξ d'una varietat Riemanniana conexa Θ , la distància Riemanniana $\rho(\nu, \xi)$ des de ν fins a ξ és l'ímfim del conjunt $\{L(\alpha) : \alpha \in \Omega(\nu, \xi)\}$, on $\Omega(\nu, \xi)$ és el conjunt de totes les corbes segment de classe C^∞ a trossos en Θ des de ν fins a ξ .*

El següent resultat relaciona la distància Riemanniana amb la longitud de les geodèsiques que uneixen dos punts de la varietat.

Proposició 9.9.1 *Sigui U un entorn normal del punt $\theta \in \Theta$, amb Θ varietat Riemanniana. Aleshores la geodèsica radial σ des de θ fins a $\nu \in U$ és la (única) corba més curta a Θ des de θ fins a ν . En particular*

$$L(\sigma) = \|\exp_\theta^{-1}(\nu)\|_\theta = \rho(\theta, \nu)$$

Amb aquesta última proposició donaríem per finalitzat aquest apèndix de conceptes generals en geometria diferencial. Tal com hem indicat al començament d'aquest capítol, es poden obtenir més detalls a O'Neill [36], Amari [3], Spivak [46], Hicks [23], Chavel [12] i Karcher [26] entre altres.

Capítol 10

Llistat de programes

En aquest darrer capítol s'inclouen els llistats del programes que s'han desenvolupat per fer les gràfiques i taules que apareixen en els diferents capítols del treball.

Programa 1

Càlcul de la gràfica de ρ_*^2 respecte a $\frac{r}{k} + \rho^2(\Theta_0, \bar{x}_k)$, que apareixen en la secció 3.2, per a diferents valors de r i k i diferents $(1 - \alpha)$ -intervalls de confiança per a ρ_*^2 .

```

C   Fitxer de sortida amb les coordenades
C   dels punts de la gràfica
C
C       open(8,err=500,file='c:\f32\r5k30p95.dat')
C
C   Valors de "r", "k" i del valor crític
C   d'una normal amb cua a la dreta de 0.05
C
C       r=5
C       k=30
C       zp=1.64485
C
C       del=0.1
C       ros2=0.
C       do 100 i=1,50
C       aux=r+k*ros2
C       auxdk=aux/k
C       c=2.*(1.+ros2/auxdk)/(9.*aux)
C       rot2=auxdk*(zp*sqrt(c)+1.-c)**3
C       bb=(r + k*rot2)/k
C
C   Escriurà dels resultats
C
C       write(8,200) bb, ros2
C
C       ros2=ros2+del
100  continue
200  format('(',f6.4,',',f6.4,')')
500  stop
      end

```

Programa 2

Càlcul de les taules per als valors crítics m/k , de la secció 4.1, donada la dim Θ , per a diferents valors del nivell de confiança $(1 - \alpha)$ del mètode-pf i del p -valor del test clàssic.


```

C   Entrades:
C           c = valors crítics de la Chi-quadrat
C           f = valor crític de la Normal tipificada
C           x = mediana de la Chi-quadrat
C           ng = dimensió de la varietat
C           y = p-valor del test clàssic
C   Sortides:
C           w = valor crític del nou mètode
C
C           dimension c(4,4),f(5),x(4),ng(4),w(5),y(4)
C
C   Fitxer de sortida dels valors crítics
C
C           OPEN(7,ERR=600,file='c:\f32\test.dat')
C
C   Entrada de les dades
C
C           c(1,1)=2.706
C           c(1,2)=3.841
C           c(1,3)=6.635
C           c(1,4)=10.825
C           c(2,1)=9.236
C           c(2,2)=11.070
C           c(2,3)=15.086
C           c(2,4)=20.517
C           c(3,1)=15.987
C           c(3,2)=18.307
C           c(3,3)=23.209
C           c(3,4)=29.588
C           c(4,1)=40.256
C           c(4,2)=43.773
C           c(4,3)=50.892
C           c(4,4)=59.703
C
C           f(1)=0.0
C           f(2)=0.674
C           f(3)=1.282
C           f(4)=1.645
C           f(5)=2.326
C
C           x(1)=0.455
C           x(2)=4.351

```

```

        x(3)=9.342
        x(4)=29.336
C
        ng(1)=1
        ng(2)=5
        ng(3)=10
        ng(4)=30
C
        y(1)=0.1
        y(2)=0.05
        y(3)=0.01
        y(4)=0.001
C
C   Escriptura de la dimensió de la varietat
C
        do 100 i=1,4
        write(7,500) ng(i)
500 format(' dimensio varietat = ',I4)
C
C   Escriptura del p-valor del test clàssic
C
        do 110 j=1,4
        write(7,505) y(j)
505 format(' p valor del test classic ', f10.3)
C
C   Càlcul del valor crític del nou mètode
C
        do 120 l=1,5
        res=sqrt(c(i,j))
        rstar=res - f(l)
        if (rstar.le.0. ) then
            rstar=0.
            w(l)=99999999.
        else
            w(l)=x(i)/(rstar*rstar)
        endif
120 continue
C
C   Escriptura del nou valor crític
C
        write(7,510)(w(l), l=1,5)
510 format(1x,5f10.3)

```

```

C
  110 continue
  100 continue
  600 stop
      end

```

Programa 3

El següent programa dóna els valors que apareixen a la taula de la secció 5.1, on es mostra el comportament del mètode `mb` en el model Poisson. Es calcularà la màxima grandària mostral m^* que encara permet l'ús del submodel, per a diferents valors de k i s i $\lambda_0 = 1$.

```

C
C   Definició de la funció hipergeomètrica generalitzada
C   amb  $p=2$ ,  $q=1$ , per un factor multiplicatiu
C
      function dfu(s,x,tol)
      common bux
      dimension w(500)
C
C   Càlcul dels primers termes de la sèrie amb valor
C   absolut més gran que la tolerància
      sum=1./bux
      xxx=1.
      do 100 i=1,500
      xj=i
      xjh=xj-0.5
      xxx=xxx*(s+xj)*x/xjh
      w(i)=xxx/bux
      if(abs(xxx).lt.tol.and.i.gt.50) then
                                max=i
                                go to 200
                                end if
      100 continue
      write(6,*) ' error   dfu '
C
C   Suma dels termes calculats anteriorment
C   de major a menor
C
      200 do 300 i=max,1,-1
      300 sum=sum+w(i)

```

```

        dfu=sum
        return
    end

C
C Definició de la funció hipergeomètrica generalitzada
C amb p=2, q=3, per un factor multiplicatiu
C
    function tfd(s,x,tol)
        common bux
        dimension w(500)

C
C Càlcul dels primers termes de la sèrie amb valor
C absolut més gran que la tolerància
C
        sum=1./bux
        xxx=1.
        do 100 i=1,500
            xj=i
            xjh=xj-0.5
            xxx=xxx*((s+xjh)*x/xjh)*(xj/xjh)
            w(i)=xxx/bux
            if(abs(w(i)).lt.tol.and.i.gt.50) then
                max=i
                go to 200
            end if

100 continue
        write(6,*)' error en tfd '

C
C Suma dels termes calculats anteriorment
C de major a menor
C
200 do 300 i=max,1,-1
300 sum=sum+w(i)
        tfd=sum
        return
    end

C
C Entrades:
C         xxk = grandària "k" de la primera mostra
C         xs = valor "s" de la suma de la mostra
C Sortides:
C         xmat = màxima grandària mostral "m"
C

```

```

common bux
dimension xxk(6),xs(10),xmat(10,6)
xxk(1)=5.
xxk(2)=10.
xxk(3)=15.
xxk(4)=20.
xxk(5)=25.
xxk(6)=30.
xs(1)=1.
xs(2)=3.
xs(3)=5.
xs(4)=7.
xs(5)=10.
xs(6)=15.
xs(7)=20.
xs(8)=25.
xs(9)=30.
xs(10)=35.
C
C Fitxers de sortida de resultats
C
    tol=1.e-8
    open(6,err=500,file='c:\f32\poisson.dat')
    open(7,err=500,file='c:\f32\pois2.dat')
    open(8,err=500,file='c:\f32\pois3.dat')
C
    xlam0=1.
C
    do 77 l=1,3
    write(6,*) ' lambda_0 = ',xlam0
    write(8,*) ' lambda_0 = ',xlam0
C
    do 78 k=1,6
    xk=xxk(k)
    write(6,*) ' k = ',xk
C
    do 79 n=1,10
    s=xs(n)
    write(6,*) ' s = ', s
C
    as=s
    uuu=1./1.7724538
    do 33 j=1,10000

```

```

        if(as.lt.0.5) go to 34
        uuu=uuu*as/(as-0.5)
33  as=as-1.
C
34  zzz=4.*(uuu/sqrt(xk) - sqrt(xlam0))**2
C
        write(6,*) ' distancia^2 a H_0: ',zzz
C
C   Càlcul del factor multiplicatiu
C
        do 100 m=1,35
        xm=m
        bux=(1.+xm/xk)**(s+0.5)
        x=xm/(xm+xk)
        aux21=dfu(s,x,tol)
        aux32=tfd(s,x,tol)
C
C   Càlcul final del criteri de decisió
C
        daux=2.*uuu*aux21/sqrt(3.1415927*xm*(xm+xk))
        eaux=aux32/(3.1415927*xm)
        total=4.*(eaux-daux-xlam0+2.*sqrt(xlam0/xk)*uuu)
        if(total.ge.0) then
                write(7,*) xm, total
                else
                xm=xm-1.
                xmat(n,k)=xm
                write(6,*) ' m*= ',xm
                go to 79
                end if
C
100  continue
C
        79  continue
        78  continue
        do 22 n=1,10
C
C   Escriptura dels resultats finals
C
22  write(8,333)(xmat(n,k),k=1,6)
333 format(6f8.3)
        xlam0=xlam0+1.

```

```

    77 continue
C
    500 stop
    end

```

Programa 4

Aquest programa ens calcula els valors de la taula que apareix en la secció 5.2, on es veu el comportament del mètode `mb` pel model Gamma. Ens troba els valors crítics per a $\rho^2(S_k, \lambda_0)$, anomenats $\hat{\rho}^2$, en el cas $\alpha = 1$ per a diferents valors de k i m , pels quals encara acceptarem la hipòtesi nul·la.

```

    function psi(x)
C
C   Derivada logarítmica funció Gamma
C
    real*8 y,aux,aux2,z,tol,res
    y=dbl(x)
    aux=0.d0
    aux2=0.d0
10  if (y.gt.1.5) then
        y=y-1.d0
        aux=aux+1.d0/y
        go to 10
    endif
    aux=aux-.5772156649015329
    tol=1.d-20
    y=y-1.
    if (y.eq.0.) go to 500
    z=-1.
    do 100 i=2,1000
    if (dabs(z).lt.tol) go to 500
    z=-z*y
    call zeta(i,res)
100  aux2=aux2+res*z
500  psi=sngl(aux+aux2)
    return
    end
C
    function dpsi(x)
C
C   Derivada segona logaritme funció Gamma

```

C

```

real*8 y,aux,aux2,z,tol,res
y=dbl(x)
aux=0.d0
aux2=0.d0
10 if (y.gt.1.5) then
    y=y-1.d0
    aux=aux-1.d0/y/y
    go to 10
endif

tol=1.d-20
y=y-1.
z=1.
do 100 i=2,1000
if (dabs(z).lt.tol) go to 500
call zeta(i,res)
aux2=aux2+res*z*dfloat(i-1)
100 z=-z*y
500 dpsi=sngl(aux+aux2)
return
end

```

C

```

subroutine zeta(m,res)

```

C

C Funció zeta de Riemann

C

```

real*8 x,res
dimension x(49)
data x /1.644934066848226,1.202056903159594,
1 1.082323233711138,1.036927755143370,1.017343061984449,
2 1.008349277381923,1.004077356197944,
3 1.002008392826082,1.000994575127818,1.000494188604119,
4 1.000246086553308,1.000122713347578,
5 1.000061248135059,1.000030588236307,1.000015282259409,
6 1.000007637197638,1.000003817293265,
7 1.000001908212717,1.000000953962034,1.000000476932987,
8 1.000000238450503,1.000000119219926,
9 1.000000059608189,1.000000029803504,1.000000014901555,
1 1.000000007450712,1.000000003725334,1.000000001862660,
2 1.000000000931327,1.000000000465663,1.000000000232831,
3 1.000000000116416,1.000000000058208,1.000000000029104,
4 1.000000000014552,1.000000000007276,1.000000000003638,
5 1.000000000001819,1.000000000000909,1.000000000000455,

```



```

6 1.0000000000000227,1.000000000000114,1.000000000000057,
7 1.000000000000029,1.00000000000015,1.00000000000008,
8 1.000000000000004,1.00000000000002,1.00000000000001/
  if (m.le.1) then
    write(6,100) m
100 format(' funció zeta de Riemann no definida per x = ',i5)
    res=1.d75
    return
  endif
  if (m.le.50) then
    res=x(m-1)
  else
    res=1.d0
  endif

  return
end

C
C  Obtenció del criteri de decisió
C
  dimension crit(10)
  open(6,err=500,file='c:\f32\gammas.dat')
  alpha=1.
  do 90 l=1,3
  write(6,*)' alpha = ', alpha

C
  do 100 k=5,35,5
  do 110 m=1,10
  xk=k
  xm=m
110 crit(m)=amax1(alpha*(dpsi(xm*alpha)-dpsi(xk*alpha)),0.)

C
  write(6,130)(crit(m),m=1,10)
130 format(10f7.4)
100 continue
  alpha=alpha+1.
  90 continue
500 stop
end

```

Programa 5

Aquest programa crea els valors de la taula de la secció 6.1, buscant numèricament els valors de la mediana de $m \rho^2(\beta_{(m)}^*, \sigma_{(m)}^*; \beta, \sigma)$.

```

C
C   Càlcul de la integral pel mètode de Simpson
C
      function simp(a,b,xtol,ix)
      ic=0
      if(xtol.le.0.) xtol=1. e-8
      h=b-a
      suma=fun(a)+fun(b)
      c=0.
      xaux=0.
      sumc=0.
      max=1
      do 100 j=1,18
      c=xaux
      h=h*0.5
      max=max*2
      suma=suma+2.*sumc
      sumc=0.
      do 101 i=1,max,2
      aux=i
101  sumc=sumc+fun(a+aux*h)
      xaux=(suma+4.*sumc)/3.*h
      if(abs(c-xaux).lt.xtol) then
                                ic=ic+1
                                if(ic.ge.ix) go to 102
                                end if

100  continue
      xtol=-1.
102  simp=(16.*xaux-c)/15.
      return
      end

C
C   Funció final que integrarem per Simpson
C
      function fun(t)
      common xn,xm,tol,c,xxx,ix
      if(c.eq.0.) then
          fun=0.
          return
          end if

      t2=t*t
      aux=(f(t2)/xxx)*(t**(xn-xm-1.))*exp(-t2/2.)
      fun=aux

```

```

        return
    end
C
C Definició de la funció "f"
C
    function f(v)
    dimension auxi(500)
    common xn,xm,tol,c,xxx,ix
    tolmes=tol/100.
    z=4.*sqrt(xn*v)*cosh(sqrt(c/(2.*xn)))-2.*v-2.*xn
    if(z.lt.0.) then
        f=0.
        return
    end if

    ww=z**(xm/2.)
    xj=0.
    xcon=1.
    w=1.
    is=0
C
C Desenvolupament de Taylor de la funció que
C depèn de la variable "u"
C
    do 100 j=1,500
    if (abs(xcon).lt.tolmes) then
        is=is+1
        if(is.gt.10) then
            max=j-1
            go to 250
        end if
    end if

    xcon=w/(xm/2.+xj)
    auxi(j)=xcon
    xj=xj+1.
    w=w*z*(-0.5)/xj
100 continue
    write(8,*)' dificultats en avaluar la sèrie ',xn,xm,c
250 sum=0.
    do 110 j=max,1,-1
110 sum=sum+auxi(j)
    sum=sum*ww
    if(sum.lt.0) then

```

```

        f=0.
                else
        f=sum
        end if
    return
end
C
C Definició de la funció Gamma
C
    function gamma(x)
    if (x.lt.0.9) then
        gamma=1.7724538
        return
    else
        aux=1.
        end if

    z=x
    do 100 k=1,1000
    if(z.le.1.1) go to 500
    z=z-1.
100 aux=aux*z
500 if (z.lt.0.9) then
        aux=aux*1.7724538
        end if

    gamma=aux
    return
end
C
C Resultat de la integració
C
    subroutine dis(resul)
    common xn,xm,tol,c,xxx,ix
    ww=sqrt(xn)
    xinc=ww*exp(-sqrt(c/(2.*xn)))
    xsupc=ww*exp(sqrt(c/(2.*xn)))
    xtol=tol
    ixx=ix
    res=simp(xinc,xsupc,xtol,ix)
    if(xtol.lt.0.) write(6,*) ' error a la integració',
    xtol,c,xn,xm
C
    resul=2.*res
    return

```

```

        end
C
C   Ajust de la mediana
C
      subroutine adjust(c1,c2,xmedian)
      common xn,xm,tol,c,xxx,ix
      topinf=c1
      c=c1
      call dis(p)
      if(abs(p-0.5).lt.tol) then
                                xmedian=c1
                                return
                                end if

      yinf=p
C
      topsup=c2
      c=c2
      call dis(q)
      if(abs(q-0.5).lt.tol) then
                                xmedian=c2
                                return
                                end if

      ysup=q
C
      do 1000 k=1,25
      call motor(topinf,yinf,topsup,ysup,xmedian,*500)
1000 continue
      write(6,*) ' no puedo ',topinf,topsup,yinf,ysup
      500 return
      end
C
      subroutine motor(xinf,yinf,xsup,ysup,xmedian,*)
      common xn,xm,tol,c,xxx,ix
      x=(xinf+xsup)/2.
      c=x
      call dis(y)
      if(abs(y-0.5).lt.tol) then
                                xmedian=c
                                return 1
                                end if

      if(y.lt.0.5) then
                                xinf=x
                                yinf=y

```

```

        else
        xsup=x
        ysup=y
        end if

    return
end

C
C Definició dels factors que apareixen a la integració
C
    subroutine init
    common xn,xm,tol,c,xxx,ix
    xmh=xm/2.
    xnmh=(xn-xm)/2.
    xxx=((2.**((xn/2.)))*gamma(xmh)*gamma(xnmh))
    return
end

C
C Inici del programa principal i entrada de dades
C
    dimension xmeds(5,6),xmedi(5,6),xnum(6)
    common xn,xm,tol,c,xxx,ix
    xnum(1)=10.
    xnum(2)=15.
    xnum(3)=20.
    xnum(4)=25.
    xnum(5)=30.
    xnum(6)=50.
    ix=3

C
C Fitxers d'escriptura dels resultats
C
    open(8,err=500,file='c:\f32\mediana.dat')
    open(9,err=500,file='c:\f32\distrib.dat')

C
    write(6,*)' tolerància ?'
    read(5,*) tol
    write(6,*) ' tope màxim funció distribució ?'

C
    read(5,*) tmax
    do 150 m=1,5
    delta=0.3*(1.+aint((xm+1.)/3.))
    do 150 n=1,6
    xn=xnum(n)

```

```

xm=m
call init
c=0.
xmejors=-1.
xmejori=0.
do 155 k=1,40
call dis(p)
if((p.gt.0.5).and.(xmejors.lt.0.)) xmejors=c
if(p.lt.0.5) xmejori=c
write(9,*)'xm,xn,c,p',xm,xn,c,p
if(p.gt.tmax) go to 175
155 c=c+delta
175 xmeds(m,n)=xmejors
xmedi(m,n)=xmejori
150 continue
C
do 20 m=1,5
do 30 n=1,6
xm=m
xn=xnum(n)
call init
c1=xmedi(m,n)
c2=xmeds(m,n)
call adjust(c1,c2,xmedian)
write(8,*) xm,xn,' mediana = ',xmedian
30 continue
20 continue
500 stop
end
```


Bibliografia

- [1] M. Abramowitz. *Handbook of Mathematical Functions*. Dover Publications, Inc., New York, 1970.
- [2] H. Akaike. *2nd Int. Symp. Inf. Theory*, chapter Information Theory and An Extension of The Maximum Likelihood Principle, pages 267–281. B.N. Petrov and F. Czaki eds., Akademiai Kiado, Budapest, 1973.
- [3] S. Amari. *Differential-Geometrical Methods in Statistics*, volume 28 of *Lecture notes in statistics*. Springer Verlag, New York, 1985.
- [4] C. Atkinson and A.F.S. Mitchell. Rao's distance measure. *Sankhyà*, 43 A:345–365, 1981.
- [5] O. E. Barndorff-Nielsen. *Differential geometry in statistical inference*, volume 10 of *Lecture Notes - Monographs Series*, chapter Differential and integral geometry in statistical inference, pages 95–161. Institute of Mathematical Statistics, Hayward, California, 1987.
- [6] B.Efron. Defining the curvature of a statistical problem. *Ann. Statist.*, 3:1189–1242, 1975.
- [7] J.O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer Verlag, New York, 1985.
- [8] J. Burbea. Informative geometry of probability spaces. *Expositiones Mathematicae*, 4:347–378, 1986.
- [9] J. Burbea and J.M. Oller. The information metric for univariate linear elliptic models. *Statistics & Decisions*, 6:209–221, 1988.
- [10] J. Burbea and C.R. Rao. Entropy differential metric, distance and divergence measures in probability spaces: a unified approach. *J. Multivariate Anal.*, 12:575–596, 1982.

- [11] N.N. Cencov. *Statistical Decision Rules and Optimal Inference*. Trans. Math. Monographs, 53, Amer. Math. Soc. (English translation of the Russian book published in 1972, Nauka, Moscow), Providence, 1982.
- [12] I. Chavel. *Riemannian Geometry. A Modern Introduction*. Cambridge Tracts in Mathematics. Cambridge University Press, Cambridge, 1993.
- [13] J. Dieudonné. *Fundamentos de Análisis Moderno*. Ed. Reverté, Barcelona, 1966.
- [14] M. Emery and G. Mokobodzki. *Séminaire de Probabilités XXV*, volume 1485, chapter Sur le barycentre d'une probabilité dans une variété, pages 220–233. Springer Lecture Notes in Mathematics, Springer Verlag, Berlin, 1991.
- [15] R.A. Fisher. *Statistical Methods for Research Workers*. Oliver & Boyd, Edinburgh, 1925.
- [16] R.A. Fisher. The use of multiple measurements in taxonomic problems. *Ann. Eugen.*, 7:179–188, 1936.
- [17] R.A. Fisher. *Statistical Methods and Scientific Inference*. Hafner, New York, 1959.
- [18] S. Geisser and W.F. Eddy. A predictive approach to model selection. *Journal of the American Statistical Association*, 74:153–160, 1972.
- [19] I.J. Good. *Encyclopedia of Statistical Sciences*, chapter Scientific Method and Statistics, pages 291–304, Vol. 8. S. Kotz and N.L. Johnson Editors, John Wiley & Sons, New York, 1988.
- [20] I.J. Good and R.A. Gaskings. An exposition of the maximum penalized likelihood method of probability density estimation. *Virginia J. Sci.*, 23:171–193, 1972.
- [21] M. Gromov. *Structures Métriques pour les Variétés Riemanniennes*. Editions CEDIC, Paris, 1981.
- [22] S. Helgason. *Differential Geometry, Lie Groups and Symetric Spaces*. Academic Press, 1978.
- [23] N.J. Hicks. *Notes on Differential Geometry*. Van Nostrand Reinhold Comp., London, 1965.

- [24] H. Jeffreys. An Invariant Form for the Prior Probability in Estimation Problems. *Proc. Royal Soc.*, 186 A:453–461, 1946.
- [25] P.E. Jupp and K.V. Mardia. A unified view of the theory of directional statistics, 1975-1988. *International Statistical Review*, 57(3):261–294, 1989.
- [26] H. Karcher. Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.*, 30:509–541, 1977.
- [27] M.G. Kendall and R.L. Plackett. *Studies in the History of Statistics and Probability*. Griffin, London, 1977.
- [28] P.S. Laplace. *Théorie Analytique des Probabilités*. Paris, 1812. The 3rd edition of 1820 is reprinted as Vol. 7 of Laplace’s collected works.
- [29] E. Lehmann. *Testing Statistical Hypotheses*. Wiley, New York, 1959.
- [30] S.A. Mulaik L.L. Harlow and J.H. Steiger. *What If There Were No Significance Tests*. Lawrence Erlbaum Associates, Mahwah NJ, 1997.
- [31] R.J. Muirhead. *Aspects of Multivariate Statistical Theory*. John Wiley, New York, 1982.
- [32] J. Neyman. *A Selection of the Early Statistical Papers of J. Neyman*. University of California Press., Berkeley, 1967.
- [33] J. Neyman and E.S. Pearson. *Joint Statistical Papers*. Cambridge University Press, Cambridge, 1967. Reprints of the original sources for Neyman and Pearson’s alternative testing methodology.
- [34] J.M. Oller. *Statistical data analysis and inference*, chapter Some geometrical aspects of data analysis and statistics, pages 41–58. Elsevier science publishers B. V., North Holland, Amsterdam, 1989.
- [35] J.M. Oller and J.M. Corcuera. Intrinsic analysis of statistical estimation. *Ann. Statist.*, 23(5):1562–1581, 1995.
- [36] B. O’Neill. *Semi-Riemannian Geometry: with Applications to Relativity*. Academic Press, New York, 1983.
- [37] E.S. Pearson. *In Research Papers in Statistics*. Wiley, New York, 1966.
- [38] E.S. Pearson. *The Selected Papers of E.S. Pearson*. Cambridge University Press, Cambridge, 1966.

- [39] K. Pearson. On the Criterion that a Given System of Deviations From the Probable in the Case of a Correlated System Can Be Reasonably Supposed to Have Arisen From Random Sampling. *Philosophical Magazine Fifth Series*, 50:157–175, 1900.
- [40] K.R. Popper. *The Logic of Scientific Discovery*. Hutchinson, London, 1959. based on *Logik der Forschung* (1935).
- [41] M. Queisanne. *Álgebra Básica*. Ed. Vicens-Vives, Barcelona, 1971.
- [42] C.R. Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcuta Math. Soc.*, 37:81–95, 1945.
- [43] C.R. Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.*, 37:81–91, 1945.
- [44] R. Rollal. *Statistical Evidence, a Likelihood Paradigm*. Chapman & Hall, Boca Raton, 1997.
- [45] R.J. Serfling. *Aproximation Theorems in Mathematical Statistics*. John Wiley, New York, 1980.
- [46] M. Spivak. *A Comprehensive Introduction to Differential Geometry*. Publish or Perish, Berkeley, 1979.
- [47] "Student". *Student's Collected Papers*. Cambridge University Press, Cambridge, 1958.
- [48] L. Wittgenstein. *Tractatus Logico-Philosophicus*. Alianza Ed., Madrid, 1973.