Treball final de grau

GRAU DE MATEMÀTIQUES

Facultat de Matemàtiques i Informàtica
Universitat de Barcelona

INTRODUCCIÓ A L'ESTADÍSTICA BAYESIANA

Autor: Mónica Souto Abad

Director: **Dr. David Marquez**
Realitzat a: **Departament de Matemàtiques i Informàtica. UB**

Barcelona, **29 de juny de 2017**

Abstract

Bayesian statistics is that statistic that is based on the Bayes theorem, therefore, in the information we have before observe the data. As a result, we have the *a priori* distribution and *a posteriori* distribution, which we make inference.

In this work we will study the process of estimation from Bayesian inference and simulation methods of *a posteriori* distribution. In addition, we will see this knowledge applied to practical examples.

Resum

L'estadística Bayesiana és aquella estadística que es basa en el teorema de Bayes, per tant, en la informació que tenim abans d'observar les dades. Com a conseqüència, tenim la distribució *a priori* i la distribució *a posteriori*, sobre la qual fem inferència.

En aquest treball estudiarem el procés d'estimació a partir de la inferència Bayesiana i els mètodes de simulació de la distribució *a posteriori*. A més, veurem aquests coneixements aplicats a dos exemples pràctics.

Agraïments

Amb aquest treball donem fi a una gran etapa, amb les seves coses bones i dolentes, però sobretot plena de coneixement, admiració i sacrifici.

Durant tot el viatge m'han acompanyat persones molt importants que sempre m'han animat sempre a seguir quan ho veia tot negre. A no rendir-me mai.

Avui no podria presentar aquest treball sense haver tingut la presència d'algunes persones durant tota la seva realització, per això els hi vull agrair.

Al meu tutor per prestar-me ajuda just al moment en què ho necessitava, tot i tenir problemes d'incompatibilitat d'horaris.

A la meva família i amics per proporcionar-me la seva ajuda i temps encara no entenent res del tema a tractar, però sobretot per animar-me.

Índex

1 Introducció

Durant totes les assignatures de grau només hem après coneixements sobre l'estadística clàssica, i el més a prop que hem estat de l'estadística Bayesiana ha sigut a través del Teorema de Bayes. Va ser aquí on va sorgir la meva curiositat i motivació per aprendre i saber en què consistia aquesta estadística i així entendre perquè li donen aquesta poca importància, dins i fora de l'aula.

Per aquest motiu he tingut molt clar des del principi el fil conductor d'aquest treball i com a conseqüència la seva estructura. Per començar a entendre els conceptes claus del treball havia de tenir molt clar les idees principals de la probabilitat i l'estadística, de manera que m'ha semblat interessant posar-ho d'alguna manera en el treball perquè és la base de tot el que s'explica després, però com que no és l'objectiu del treball està plasmat en el primer Annex. Dins d'aquest Annex també podem trobar una introducció a la part històrica de l'estadística Bayesiana, que encara que no influeixi sobre l'objectiu del treball, em semblava interessant saber com va començar aquest canvi de mentalitat.

Amb totes aquestes idees clares al cap, ja es pot aprofundir en l'estadística Bayesiana. A partir de la probabilitat i les variables aleatòries es pot veure la part més bàsica però indispensable per entendre en què consisteix i en què es diferencia aquesta estadística, que corresponen als Capítols 2 i 3. En el següent Capítol parlem de la inferència Bayesiana, la qual és molt interessant i extensa. Aquí el que volia aconseguir era entendre el significat d'inferència i saber com afectava la inferència Bayesiana amb tot el que ja sabíem. Podria haver fet un treball només amb un subapartat d'aquest Capítol, però no volia enfocar-me en res en particular, volia arribar a veure el màxim de tipus, problemes i solucions, a partir de mètodes, algoritmes i contrastos, per no perdre mai l'objectiu principal del treball. Una de les solucions més importants i que soluciona el problema més greu de la inferència Bayesiana és la simulació, i per aquest motiu, hi ha un Capítol sencer dedicat als dos mètodes més famosos (MC i MCMC). De la mateixa manera que abans, aquest Capítol també podria ser molt més extens però des de l'inici no busco només parlar d'això. A més a més, ens acompanya durant tot el treball el mateix exemple aplicat als conceptes que estiguem treballant en els diferents apartats.

Com que el meu objectiu és entendre els problemes de l'estadística Bayesiana que implica la seva poca utilitat, he creat dos Capítols amb exemples per poder estudiar justament això. Una vegada hem vist i entès tots els passos que comporta fer inferència Bayesiana, volia veure un exemple real en el qual es vegi plasmat tot allò del que hem anat parlant. Tot i així, em semblava poc, ja que aquest estudi estava fet, tenia comentat els resultats, les conclusions, que tot i ser el que volia, no ho havia pogut experimentar per mi mateixa. Amb aquesta reflexió neix l'últim Capítol, en el qual faig a un nivell principiant inferència Bayesiana amb l'ajuda de R.

2 Probabilitat i Teorema de Bayes

En aquest Capítol explicarem com afecta el Teorema de Bayes a la probabilitat. El dividirem en dues parts: la primera on s'explica el Teorema de Bayes amb probabilitats i la segona és un exemple explicatiu per entendre de manera pràctica els conceptes explicats. Per poder entendre totes aquestes idees hem de tenir un mínim de coneixement en probabilitat adquirit durant les assignatures del grau. De manera que les propietats bàsiques no estan explicites en el treball però si recollides en l'Annex I.

Durant tot el Capítol denotem la probabilitat de que succeeixi un esdeveniment com P(·).

2.1 Teorema de Bayes

Teorema 2.1. ***(Teorema de Bayes)*** *Sigui* $\{A_1, A_2, \ldots, A_n\}$ *un sistema complet de successos on* $P(A_i) > 0\ \forall i$. *Aleshores, per a qualsevol succés* B *amb* $P(B) > 0$, *tenim que*

$$P(A_k \mid B) = \frac{P(B \mid A_k)P(A_k)}{\sum_{i=1}^{n} P(B \mid A_i)P(A_i)}.$$

Demostració. La demostració serà molt fàcil ja que només hem d'aplicar les proposicions i teoremes vistes en l'Annex I. A partir de la probabilitat condicionada tenim

$$P(A_k \mid B) = \frac{P(A_k \cap B)}{P(B)} = \frac{P(B \mid A_k)P(A_k)}{P(B \mid A_1)P(A_1) + ... + P(B \mid A_k)P(A_k)}.$$

A la segona igualtat en el numerador utilitzem la Regla del producte (Proposició 9.5 de l'Annex I) i en el denominador el Teorema de la probabilitat total (Teorema 9.8 de l'Annex I)

□

Per tant, si anomenem a cada part de la funció amb el nom propi del mètode Bayesià tenim que :

- Les probabilitats $P(A_i)$ s'anomenen *probabilitats a priori.*
- Les probabilitats $P(A_i \mid B)$ s'anomenen *probabilitats a posteriori.*
- Les probabilitats $P(B \mid A_i)$ s'anomenen *probabilitats versemblants.*

2.2 Exemple explicatiu

La manera més fàcil de poder entendre l'apartat anterior, acabar de comprendre en que es basa la estadística Bayesiana i veure les diferències amb la Freqüentista és a partir d'un exemple.
Les dades han sigut extretes de *Datos básicos del sistema universitario español del curso 2013-2014, Ministerio de Educación, Cultura y Deporte.*

Analitzarem quin tant per cent d'estudiants han estat matriculats en la Universitat diferenciats per branca d'ensenyament al curs 2012-2013. Segons el document son dades provisionals però nosaltres les treballarem com si fossin reals.

Les dades ens mostren que han hagut un total de 1.046.570 estudiants matriculats, dels quals 497.615 són estudiants de graus de Ciències Socials i Jurídiques, 191.583 d'Enginyeria i Arquitectura, 108.207 d'Arts i Humanitats, 187.765 de Ciències de la Salut i 61.400 estudiants de Ciències. És a dir, el 47'547% són estudiants de Ciències Socials i Jurídiques, el 18'306% d'Enginyeria i Arquitectura, el 10'339% d'Arts i Humanitats, el 17'941% de Ciències de la Salut i el 5'867% estudiants de Ciències.

A més a més, sabem que del total d'alumnes matriculats, ho han fet en una Universitat Pública el 87'2%, i la resta, 12'8%, en una Universitat Privada. Si dins de cada Universitat dividim els alumnes en les diferents brànques d'ensenyament tenim el següent resultat. En la Universitat Pública: Ciències Socials i Jurídiques un 84'7%, Enginyeria i Arquitectura un 91'8%, Arts i Humanitats un 95'3%, Ciències de la Salut un 81'0% i Ciències un 97'5%. En canvi, en la Universitat Privada: 15'3% en Ciències Socials i Jurídiques, 8'2% en Enginyeria i Arquitectura, 4'7% en Arts i Humanitats, 19'0% en Ciències de la Salut i 2'5% Ciències.

Si aproximen les dades anteriors en % a números enters, ja que corresponen al número d'alumnes, d'igual manera que a l'informe original. Anomenarem els diferents ensenyaments com: Ciències Socials i Jurídiques com *CSJ*, Enginyeria i Arquitectura com *EA*, Arts i Humanitats com *AH*, Ciències de la Salut com *CS* i Ciències com *C*. De manera que tenim la següent Taula:

Ensenyament	N^o Alumnes	N^o Alumnes U. Pública	N^o Alumnes U. Privada
CSJ	497.615	421.587	76.028
EA	191.583	175.874	15.709
AH	108.207	103.106	5.101
CS	187.765	152.150	35.615
C	61.400	59.837	1.563
Total	1.046.570	912.554	134.016

Taula 1:

En el següent quadre veurem aquestes probabilitats i d'altres calculades amb una explicació que ens ajudarà a entendre el concepte de probabilitat *a priori* i *a posteriori*. Anomenarem als successos d'escollir les diferents branques d'ensenyament com E, anar a una Universitat Pública com PL i anar a una Privada com PR.

Ensenyament E	P(E)	P(PL|E)	P(PR|E)	P(E|PL)	P(E|PR)
CSJ	0,475	0,847	0,153	0,462	0,567
EA	0,183	0,918	0,082	0,193	0,117
AH	0,103	0,953	0,047	0,113	0,038
CS	0,179	0,81	0,19	0,167	0,266
C	0,059	0,975	0,025	0,066	0,011

Taula 2:

A partir de la Taula 2, podem veure en la primera columna les diferents branques d'ensenyaments que permet la Universitat. En la segona columna la probabilitat *a priori* de que l'alumne s'hagi matriculat en l'àrea de les Ciències Socials i Jurídiques, Enginyeria i Arquitectura, etc. En les dues següents les probabilitas de matricular-se en una Universitat Pública o Privada condicionada per el tipus d'ensenyament.

Només explicarem els càlculs vincultas a la Universitat Pública, ja que els càlculs per a la Universitat Privada, són anàlogs.

La tercera columna de la Taula 2, recull les probabilitats de matricular-se en una Universitat Pública condicionada als diversos ensenyaments E_i. Aquesta columna que pren un valor de probabilitat diferent per a cadascuna de les branques és la funció de versemblança. És a dir, aquesta funció és l'aplicació que a cada succés E_i del sistema complet li assigna la probabilitat que ocorri el succés observat, en aquest cas estudiar en una Universitat Pública PL, condicionada per a cada E_i. A més, assigna a cada succés la probabilitat d'obtenir un resultat determinat, on la suma de tots ells no és 1, que implica, que aquesta funció no és una funció de probabilitat. Per tant, l'expressió del Teorema de Bayes és:

$$\text{Probabilitat } \textit{a posteriori} \propto \text{funció de versemblança} \cdot \text{probabilitat } \textit{a priori},$$

on $\propto$ vol dir que la probabilitat *a posteriori* és directament proporcional a la funció de versemblança per la probabilitat *a priori*.

Les dues últimes columnes senyalen la probabilitat *a posteriori* de cada branca d'ensenyament una vegada l'alumne decideix a quin del dos tipus de Universitat vol anar.

Per calcular la penúltima columna utilitzem el Teorema de Bayes de la següent manera:

$$P(CSJ \mid PL) = \frac{P(PL \mid CSJ)P(CSJ)}{P(PL)} = \frac{0,847 \cdot 0,475}{0,872} = \frac{0,403}{0,872} = 0,462,$$

$$P(EA \mid PL) = \frac{P(PL \mid EA)P(EA)}{P(PL)} = \frac{0,918 \cdot 0,183}{0,872} = \frac{0,168}{0,872} = 0,193,$$

$$P(AH \mid PL) = \frac{P(PL \mid AH)P(AH)}{P(PL)} = \frac{0,953 \cdot 0,103}{0,872} = \frac{0,098}{0,872} = 0,113,$$

$$P(CS \mid PL) = \frac{P(PL \mid CS)P(CS)}{P(PL)} = \frac{0,81 \cdot 0,179}{0,872} = \frac{0,145}{0,872} = 0,167,$$

$$P(C \mid PL) = \frac{P(PL \mid C)P(C)}{P(PL)} = \frac{0,975 \cdot 0,059}{0,872} = \frac{0,057}{0,872} = 0,066.$$

Per calcular el denominador, comú en tots els càlculs, utilitzem el Teorema de la probabilitat total de la següent manera:

$$\begin{aligned} P(PL) &= \sum P(PL \mid E_i)P(E_i) \\ &= P(PL \mid CSJ)P(CSJ) + P(PL \mid EA)P(EA) + P(PL \mid AH)P(AH) \\ &\quad + P(PL \mid CS)P(CS) + P(PL \mid C)P(C) = \\ &= 0,403 + 0,168 + 0,098 + 0,145 + 0,057 \\ &= 0.872. \end{aligned}$$

Tal i com hem dit abans, totes les probabilitats estan dividides per la mateixa probabilitat, P(PL), i això fa que la penúltima columna sumi 1. És la distribució de probabilitats *a posteriori* de l'elecció del tipus de Universitat que es matricula l'alumne.

En aquest exemple en concret tenim una probabilitat *a priori* de les diferents branques d'ensenyament, però a la realitat aquestes probabilitats canvien si preguntem als alumnes abans de matricular-se per como agruparien tots els graus que ofereix la Universitat, tant sigui Pública com Privada. Però aquestes probabilitats no canviarien un cop, ni dos, si no tantes vegades com alumnes participin, ja que cadascun tindria el seu criteri a l'hora d'ajuntar-los. És més, mantenint els mateixos cinc grups, no tots els alumnes posarien els mateixos graus en els mateixos grups perquè cadascun valoraria aspectes diferents de l'ensenyament. Quan qui decideix la distribució dels graus no pot tenir en compte (o no en té) tota la informació dels diferents aspectes de tots els graus obtinguda dels alumnes i professors, la probabilitat *a priori* té un paper molt subjectiu, basant en l'opinió concreta d'un petit grup de persones.

Per veure com actua la subjectivitat en les probabilitats tant *a priori* com *a posteriori*, a continuació repetim l'exemple per a diferents casos i diferents opinions o criteris a l'hora d'escollir estudiar un grau en una privada o en una pública. D'aquesta manera comprovem com canvia la probabilitat *a posteriori* partint de diferents probabilitats *a priori*. Per aquest segon estudi partim de diferents probabilitats *a priori* segons diverses situacions:

1. Probabilitat *a priori* sense dades prèvies. No disposem de informació sobre el nombre d'alumnes matriculats per grau, llavors les proporcions de les diferents branques estan repartides segons el nombre de graus que en té cadascuna. En la Taula 3 tenim aquestes probabilitats i les probabilitats *a posteriori*.

Ensenyament E	$P(E)^*$	P(PL|E)	P(PR|E)	P(E|PL)	P(E|PR)
CSJ	0,295	0,847	0,153	0,277	0,473
EA	0,368	0,918	0,082	0,373	0,315
AH	0,256	0,953	0,047	0,270	0,126
CS	0,038	0,81	0,19	0,034	0,075
C	0,042	0,975	0,025	0,046	0,011

(*)Les dades no estan basades en cap estudi real.

Taula 3:

2. Probabilitat *a priori* amb dades conegudes. És el primer cas corresponent a la Taula 2.

3. Probabilitat *a priori* segons l'opinió i l'experiència dels professors d'un batxillerat en concret tenint en compte que no es mouen els graus de les branques d'ensenyament assignades anteriorment. El consens de professors es basa en el nombre d'aprovats de les assignatures corresponents a les diferents branques i l'opinió dels seus alumnes a voler o no continuar amb els seus estudis.

   En un principi, podem veure que els graus es divideixen en les mateixen branques que les modalitats de batxillerat: Ciències Socials i Jurídiques amb el batxillerat Social, Enginyeria i Arquitectura juntament amb Ciències amb el batxillerat Tecnològic, Arts i Humanitats amb el batxillerat Artístic, Ciències de la Salut amb el batxillerat de Ciències de la Salut. Així mateix podem veure que la societat cada vegada és mes tecnològica i es creen nous lloc de treball i es perden uns altres.

   Amb tota aquesta informació, es pot intuir que cada vegada apareixeran més graus en les branques de ciència i arquitectura, que implicarà més alumnes, més demanda en aquest sector i menys en la resta. És a dir, la tendència de formar-se en l'àmbit artístic, humanístic i social disminueix, quan sempre han estat les branques predominants, i pugen el nom de matriculats en ciències, l'entorn minoritari. Aquesta informació la podem extreure de la comparativa de gràfics d'anys anteriors, nombre d'alumnes en batxillerat, opinió d'alumnes i professors, tant d'Universitat com d'estudis inferiors, criteri de professionals d'altres àmbits que ens arriben a través de la televisió, periòdics, Internet, etc.

   En la Taula 4 podem veure aquestes probabilitats *a priori* i les seves conseqüències amb les probabilitats *a posteriori*.

Si ens fixem en les tres diferents taules veiem que tenim tant per a les probabilitats *a priori* com *a posteriori* valors numèrics diferents. És normal, ja que les

Ensenyament E	$P(E)^*$	P(PL|E)	P(PR|E)	P(E|PL)	P(E|PR)
CSJ	0,413	0,847	0,153	0,400	0,508
EA	0,183	0,918	0,082	0,192	0,121
AH	0,085	0,953	0,047	0,093	0,032
CS	0,208	0,81	0,19	0,192	0,317
C	0,101	0,975	0,025	0,122	0,022

(*)Les dades no estan basades en cap estudi real.

Taula 4:

dades *a priori* són diferents en les tres. Amb la primera observació em refereixo al fet que, tot i així el nombre de vegades que el resultat de que una probabilitat *a posteriori* sigui major que una probabilitat *a priori* és el mateix en les tres, però amb la diferència que, cada vegada la probabilitat *a posteriori* es distancia més de la probabilitat *a priori* respecte la Taula 3, on es suposa que no hem utilitzat un raonament subjectiu, per tant Bayesià. Això ocorre perquè cada vegada que disposem de major informació sobre el succés a estudiar, la probabilitat *a posteriori* es veu dominada per aquest coneixement i no per la probabilitat *a priori* de la qual partim. Seguint aquesta lògica, on tenir més informació implica un valor major a posteriori que *a priori*, i vivint en un món ideal i perfecte, tindríem una progressió amb les tres taules, de manera que hi hauria més casos on els valors *a posteriori* serien més petits que *a priori* en la Taula 3, després aquest nombre disminuiria en la Taula 2 i seguiria disminuint en la Taula 4, ja que cada vegada formem una Taula amb més informació que l'anterior.

Podríem indagar més i fer estudis de perquè als estudiants els crida més l'atenció un grau que un altre, per tant una branca de l'ensenyament, perquè trien partint d'un grau en concret un ensenyament públic i no privat, perquè han augmentat el nombre de beques promovent l'ensenyament universitari i augmentat el nombre de matriculats tant en pública com en privada , etc. Amb tot això estem donant al nostre estudi, a través de probabilitats *a priori*, un caràcter subjectiu. Aquest el punt més criticat de l'estadística Bayesiana, però no l'únic pel qual el seu ús no és el més habitual. Hi ha un problema de dificultat en els càlculs de la probabilitat *a posteriori*, que gràcies als avanços dels programes informàtics s'estan solucionant, i això ajuda al fet que cada vegada s'utilitzi més aquest tipus d'estadística, com ja veurem en els següents apartats.

3 Variables aleatòries i Teorema de Bayes

En aquest Capítol explicarem com afecta el Teorema de Bayes a les variables aleatòries. Per poder entendre totes aquestes idees hem de tenir un mínim de coneixement sobre les definicions i propietats bàsiques de les variables aleatòries. Com que aquests coneixements han sigut temari impartit durant el grau els recollirem en l'Annex I.

A partir d'aquí fins al final del treball canviem la notació. Utilitzem la mateixa notació $p(\cdot)$ per a descriure la funció de densitat contínua i la funció de probabilitat discreta. A més, denotem $p(\cdot \mid \cdot)$ com la condicionada per arguments determinants pel context i $p(\cdot)$ per la marginal. Malgrat l'abús de notació, aconseguim una notació compacta i similar a la pràctica estàndar. Segons el context, per evitar confusions, s'utilitza la notació $Pr(\cdot)$ per a la probabilitat d'un esdeveniment. I per últim, quan treballem amb distribucions conegudes, utilitzem una notació basada en el nombre de la distribució.

A més, a partir d'ara, les probabilitats que es condicionen relacionades al conjunt són sempre positives.

3.1 Teorema de Bayes

Teorema 3.1. ***(Teorema de Bayes)*** *Siguin X i Y dues variables aleatòries discretes, aleshores*

$$p(Y = k \mid X = a) = \frac{p(X = a \mid Y = k)p(Y = k)}{p(X = a)} = \frac{p(X = a \mid Y = k)p(Y = k)}{\sum_{i \in X(\Omega)} p(X = a \mid Y = i)p(Y = i)}.$$

Teorema 3.2. ***(Teorema de Bayes)*** *Siguin X i Y dues variables aleatòries absolutament continues, aleshores*

$$p(y \mid x) = \frac{p(x \mid y)p(y)}{p(x)} = \frac{p(x \mid y)p(y)}{\int_{-\infty}^{+\infty} p(x \mid y)p(y)dy}.$$

Òbviament, per definició dels dos tipus de variables aleatòries, la diferència entre els Teoremes de Bayes és que, amb variables discretes estem treballant amb números, en canvi, per variables continues amb funcions. En aquest últim cas, el denominador sempre és un nombre constant per a cada x, de manera que podem escriure el Teorema de Bayes amb una funció proporcional tal que:

$$p(y \mid x) \propto p(x \mid y)p(y)$$

on $p(y \mid x)$ és la distribució *a posteriori* de la variable Y, $p(x \mid y)$ és la distribució de X condicionada a un valor de la variable Y i $p(y)$ és la distribució *a priori* de la variable Y.

Quan fem un estudi, volem saber la freqüència, la proporció, etc, de què un esdeveniment desconegut ocorri, és a dir, volem estimar el paràmetre desconegut θ.

Normalment, s'utilitzen tant el criteri del mètode dels moments com el de màxima versemblança per estimar el valor de θ a partir d'una mostra. Es suposa que la llei de probabilitat d'una variable aleatòria X, on es pugui aplicar la tècnica de mostreig, depèn del valor desconegut de θ, i per tant, s'utilitza aquest valor de la mostra per estimar un valor de θ. Cap dels dos mètodes esmentats necessita informació addicional als valors mostrals per estimar θ, però d'existir, tampoc s'hagués pogut usar-la. En moltes situacions existeix informació addicional en relació amb el valor de θ, però no sempre es pot utilitzar. En el cas de que si es pugui formar una distribució *a priori* per al paràmetre, llavors es treballa amb el mètode Bayesià.

Podem esquematitzar el mètode Bayesià com:

$\boxed{\{p(y \mid \theta), y \in \Omega, \theta \in \Theta\};\ \pi(\theta)}$ $\leftarrow$ Informació inicial

on $\pi(\theta)$ és la distribució *a priori* — Informació nova

$\downarrow$ $\downarrow$

$\boxed{\text{Teorema de Bayes}}$ $\longleftarrow$ $\boxed{\text{Les Dades (y) } L_{Y=y}(\theta)}$

$\downarrow$ on $L_{Y=y}(\theta)$ és la funció de versemblança

$\boxed{\{p(y \mid \theta), y \in \Omega, \theta \in \Theta\};\ \pi(\theta \mid y)}$

on $\pi(\theta \mid y)$ és la distribució *a posteriori*

En els següents apartats fem una explicació més exhaustiva de cada part. Tant en l'esquema anterior com a la resta del treball, només quan explicitem el model Bayesià, denotem distribució *a priori* i *a posteriori* del paràmetre com $\pi(\theta)$ i $\pi(\theta \mid \cdot)$, respectivament. D'aquesta manera queda identificada la distribució del paràmetre que ens interessa saber el resultat i diferenciat el model Bayesià del estadític $M = \{p(Y \mid \theta), \theta \in \Theta, Y \in \Omega\}$.

3.1.1 Distribució *a priori*

Definició 3.3. *La distribució a priori de un paràmetre θ és una funció de probabilitat o funció de densitat de probabilitat que expressa el grau de creença en relació amb el valor de θ, abans d'observar una mostra d'una variable aleatòria X.*

El model Bayesià parteix del model estadístic $M = \{p(Y \mid \theta), \theta \in \Theta, Y \in \Omega\}$, però a més a més, tracta el paràmetre com una variable aleatòria que tindrà la seva distribució de probabilitat $\pi(\theta)$ (distribució *a priori* del paràmetre) que ha de reflectir el nostre coneixement *a priori* sobre θ abans d'observar les dades.

Per tant, el model Bayesià $M = \{p(Y \mid \theta), \theta \in \Theta, Y \in \Omega\}$; $\pi(\theta)$ és una llista de distribucions de probabilitat ordenada per la distribució *a priori* $\pi(\theta)$. Com que aquesta estadística es subjectiva, si $\pi(\theta_1) > \pi(\theta_2)$ (distribució *a priori* de θ_1 influeix sobre la distribució *a priori* de θ_2) creiem que és més probable que $p(Y \mid \theta_1)$ hagi generat les dades que no pas $p(Y \mid \theta_2)$. Aquest model ordena la llista de distribucions de probabilitat a diferència del model estadístic.

Per crear un model d'aquest estil, la distribució *a priori* $\pi(\theta)$ s'ha de triar abans d'observar les dades i de reflectir el que se sap a partir del coneixement d'experts o d'estudis previs.

Tant el model Bayesià com el Freqüentista han de triar el model estadístic M. A més a més, el model Bayesià ha de triar, de manera subjectiva, la distribució *a priori*. Com el model Freqüentista parteix només del model estadístic M es suposa, indirectament la majoria de cops, que tots els models de la llista són equiprobables. En canvi, en el model Bayesià no té perquè, però una possible ordenació de la llista també pot ser la que tots els models són equiprobables. Com que a més hem de triar una distribució *a priori* que ens permetrà ordenar la llista, pot canviar les interpretacions.

El model estadístic M ha de plasmar el coneixement *a priori* de les dades que s'observaran i la distribució *a priori* ha de reflectir el coneixement *a priori* sobre el paràmetre. Per això, el model Bayesià ha de fer l'esforç d'estudiar que se sap del tema i traduir-ho en una distribució de probabilitat. Aquesta distribució es tria en base a estudis previs i al coneixement del experts.

Com hem dit abans, en el model Bayesià a part del model estadístic M també ha de triar la distribució *a priori* $\pi(\theta)$. La $\pi(\theta)$ és una distribució de probabilitat $\pi(\theta) = p(\theta) = p(a, b)$ que ha de reflectir el que sabem sobre el paràmetre θ. Denotem $p(a, b)$ com qualsevol distribució de probabilitat coneguda com una Normal, Beta, etc i (a, b) els corresponents paràmetres. Aleshores quan diem que hem d'escollir una distribució *a priori* també vol dir que hem de triar el valor dels seus paràmetres.

Pot passar que no hi hagi distribució *a priori* o que sigui molt difícil de trobar una distribució coneguda que expliqui el comportament *a priori* del paràmetre o simplement que els experts no arribin a un consens, llavors en aquests casos utilitzem les prioris no informatives.

Podem classificar les distribucions *a priori* com:

1. **Prioris informatives:** Analitzem el que se sap sobre l'espai mostral Ω amb una distribució *a priori* $p(\theta) = p(a, b)$. Cal triar els paràmetres de la distribució *a priori* de a i b. Es poden triar a partir del mètode per prova i error a base de $p(\theta)$ o bé a base de formular un sistema d'equacions fruit d'igualar moments, esperances, etc. Aquestes poden ser:

   (a) **Conjugades:** Una distribució *a priori* és la distribució conjugada d'un model estadístic si la distribució *a posteriori* és de la mateixa família que la distribució *a priori*, és a dir, els paràmetres de la distribució *a posteriori* són funció dels paràmetres de la distribució *a priori* i les dades.

   (b) **No conjugades:** Una distribució *a priori* és la distribució no conjugada d'un model estadístic si la distribució *a posteriori* no és de la mateixa família que la distribució *a priori*. En aquest cas, tenim més distribucions i això fa que sigui més difícil de triar-la. A més el càlcul de la distribució *a posteriori* és més complex.

2. **Prioris no informatives:** A aquest tipus també se'ls coneix com prioris objectives o de referència. Són aquelles distribucions que proporcionen molt poca informació en relació amb l'experiment que s'estigui estudiant. Això no vol dir que es parteixi del desconeixement sobre la variable, mai es parteix de la ingnorància absoluta. La solució no és pressuposar un comportament inicial específic per aquesta variable. Hi ha diferents mètodes per solucionar-ho que el que volen és que $p(\theta)$ interfereixi el mínim possible amb les dades. Els mètodes més utilitzats són:

   (a) **Prioris de Laplace o principis de la raó insuficient:** És un procediment molt intuïtiu, ja que si no hi ha informació per diferenciar entre diferents valors de θ, aleshores donem la mateixa probabilitat a tots els valors de $\theta \in \Theta$, és a dir, utilitzem una distribució *a priori* uniforme de θ. Una de les característiques és que la distribució uniforme no és invariant en cas de transformacions, per exemple, amb els canvis d'unitats de mesura. Si el suport de θ és infinit, la distribució *a priori* serà impròpia que implica que: $p(\theta) \propto 1$. Una distribució *a priori* és impròpia quan:

   $$\int p(\theta)d\theta = +\infty.$$

   Per tant, la teoria de la probabilitat no pot ser aplicada tret que la definició de la distribució *a posterior* proporcioni una distribució pròpia,

   $$\int p(\theta \mid x)d\theta = 1.$$

   Per això, Laplace motiva a donar la mateixa probabilitat als valors de θ només quan Ω és discret.

   (b) **Prioris de Jeffreys:** Un problema d'estimació, des de la visió de Jeffreys, és que donada la forma de la distribució, existeixin paràmetres desconeguts, sense preferències inicials per donar-los valor, amb l'objectiu de conèixer la distribució de probabilitats associada a aquests paràmetres basant-se en la informació que proporcionen les dades conegudes.

   Jeffreys va establir que la distribució *a priori* del paràmetre havia de ser:

   $$p(\theta) \propto \sqrt{I(\theta)},$$

   on $I(\theta)$ és la informació de Fisher, que mesura la informació del model en funció del paràmetre, expressada com:

   $$I(\theta) = E\left[\left(\frac{\partial lnL_{X=x}(\theta)}{\partial\theta}\right)^2\right],$$

   on $lnL_{X=x}(\theta)$ és el logaritme de la funció de versemblança.

La justificació d'aquest mètode es deu a que les funcions resultants compleixen la propietat d'invariància davant reparametritzacions del model. El problema és que no funciona adequadament per a paràmetres multivariants.

(c) **Límit de les conjugades:** Es tracta la distribució *a priori* conjugada triant els paràmetres que fan que la variància sigui infinit.

(d) **Priori de Zellner:** Proposa la recerca de la distribució *a priori* aplicant la teoría de la informació. Teoría creada per Claude Elwood Shannon, on estudía les lleis matemàtiques relacionades amb la transmissió i el processament de la informació. Per aplicar aquest procediment, parteix d'una funció de densitat $p(x, \theta)$ de probabilitats conjunta, per a una observació x i un paràmetre θ, i aplica el concepte d'entropia o mesura proporcionat per Shannon.

Un cop s'ha escollit el model estadístic M i la distribució *a priori*, que es creu correcte, es recomanable fer sempre un anàlisi de sensibilitat d'aquesta distribució. Consisteix a triar diferents possibles distribucions *a priori* i examinar com canvia la distribució *a posteriori* per poder treure conclusions, com per exemple, si tens una bona mida de la mostra al fer aquests canvis inicials es pot observar que la nostra referència no canvia.

3.1.2 Distribució *a posteriori*

De la mateixa manera que la distribució *a priori* explica el nostre coneixement abans d'observar les dades sobre el paràmetre θ, la distribució *a posteriori* reflexa el nostre coneixement sobre θ un cop estan observades les dades.

Partim d'un model Bayesià amb la seva distribució *a priori* $p(\theta)$, després observem les dades (Y=y), on tenim tota la informació sobre θ a la versemblança $L_{Y=y(\theta)}$ per poder fer inferència sobre θ, tenint en compte la informació que ens han aportat les dades. Per tant, el que volem és la distribució *a posteriori* del paràmetre $p(\theta \mid y)$.

La versemblança és una funció proporcional a la distribució de probabilitat de Y avaluada a les dades i expressada en funció del paràmetre θ. Per tant,

$$L_{Y=y(\theta)} \propto p(y \mid \theta).$$

La idea és que la versemblança t'ordena els possibles valors dels paràmetre de més a menys creïbles. Doncs, si donades unes dades tenim: $L_{Y=y(\theta_1)} > L_{Y=y(\theta_2)}$ és més creïble que el valor real del paràmetre θ sigui θ_1 en lloc de θ_2.
"El principi de versemblança" diu que tota la informació que proporcionen les dades sobre el paràmetre es troba en la funció de versemblança.

En conseqüència, el model Bayesià queda com:

$$\{p(Y \mid \theta), \theta \in \Theta\}\,;\ \pi(\theta \mid y)$$

on $\pi(\theta \mid y) = p(\theta \mid y)$ és la distribució *a posteriori* que combina la informació *a priori* $p(\theta)$, amb la informació de les dades $L_{Y=y}(\theta)$, utilitzant el Teorema de

Bayes. Tot el que sabem del paràmetre θ, un cop observades les dades, es troba a la distribució *a posteriori* $p(\theta \mid y)$, i per tant, serà l'objecte que utilitzarem per fer inferència (estimació puntual, interval i proves d'hipòtesis).

Per tant, la distribució *a posteriori* es calcula a partir del Teorema de Bayes:

$$p(\theta \mid y) = \frac{p(y, \theta)}{p(y)} = \frac{p(y \mid \theta)p(\theta)}{p(y)}.$$

Ara ens trobem com sempre amb el problema de la complexitat dels càlculs ja que les dues igualtats estan dividides per una integral. Com que $p(y)$ és una número, ja que és la distribució marginal de la variable Y, un cop observades les dades, podem escriure la distribució *a posteriori* com: $p(\theta \mid y) \propto p(y \mid \theta)p(\theta)$. L'inconvenient de treballar amb una funció proporcional és que no es una distribució i això s'ha de tenir en compte. Vol dir que, $p(y)$ és la constant que fa que:

$$\int_\Omega p(\theta \mid y)d\theta = \int_\Omega \frac{p(y, \theta)}{p(y)}d\theta = \int_\Omega \frac{p(y \mid \theta)p(\theta)}{\int_\Omega p(y \mid \theta)p(\theta)d\theta}d\theta = \frac{\int_\Omega p(y \mid \theta)p(\theta)d\theta}{\int_\Omega p(y \mid \theta)p(\theta)d\theta} = 1.$$

Observació 3.4. Tenim que $p(\theta \mid y) \propto p(y \mid \theta)p(\theta)$, on $p(y \mid \theta)$ és la distribució de probabilitat de la variable aleatòria Y avaluada en les dades $Y = y$ i expressada en funció del paràmetre θ, és a dir, la versemblança. Això implica que: $p(\theta \mid y) \propto L_{Y=y}(\theta)p(\theta)$. Podem expressar la distribució *a posteriori* com un compromís entre la versemblança i la distribució *a priori*, és a dir, és un compromís entre la informació que ens han aportat les dades i la informació que teníem abans de recollir les dades.

Com hem dit al principi de l'apartat, podem calcular la distribució *a posteriori* a partir de la distribució *a priori* i la funció de versemblança. Dit això, queda clar que primer hem d'obtenir una distribució *a priori*, però si tenim noves dades podem tractar la distribució *a posteriori* com una *a priori*, obtenint així una nova distribució *a posteriori*. El resultat és el mateix que si agaféssim totes les dades a la vegada, en parts o d'una en una, gràcies a la proposició 9.6 de l'Annex. Tal i com vam fer al exemple explicatiu del Capítol 3, si no tenim suficient informació sobre la variable i necessitem començar per la distribució *a priori*, una bona solució és que la funció de probabilitat sigui constant, és a dir, que la distribució *a priori* sigui uniforme.

Quan la distribució *a posteriori* resultant sigui de la mateixa família que la distribució *a priori*, es defineix com una família de distribucions conjugades.

En l'estadística Freqüentista existeix el concepte *interval de confiança*, que fa referència a la probabilitat que l'estimador calculat es trobi dins de dos nivells considerats de confiança, on sol ser del 95%, és a dir, en repetir una multitud de vegades l'experiment, l'estimador calculat es trobaría dins de l'interval el 95% de les vegades i la resta, el 5%, calcularíem un estimador incorrecte.

Per al mètode Bayesià tenim com a equivalència el que denotem com *interval de probabilitat*. En aquest cas, utilitzem per al càlcul de l'interval, la funció de densitat obtinguda *a posteriori*, de manera que, amb l'àrea de sota aquesta corba i uns certs valors X i Y amb una probabilitat del 95%, la majoria dels casos, es construeix aquest interval de probabilitat del 95% entre aquests punts X i Y.

3.1.3 Distribució predictiva *a priori* i *a posteriori*

Definició 3.5. *La distribució predictiva a priori és la distribució marginal de la variable aleatòria Y que expressa el nostre coneixement sobre l'espai mostral Ω, és a dir, el coneixement de nous valors de la variable que observem abans d'observar les dades.*

De manera que si denotem a la variable futura com $\tilde{y}$, la seva distribució *a priori* és $p(\tilde{y})$.

Per escollir un bon model Bayesià s'ha de triar primer el model estadístic M i la distribució *a priori* especificant el valor dels paràmetres. Per triar aquest valor els dos mètodes més emprats són:

1. Utilitzant la distribució *a priori* si la informació que tens a priori és sobre l'espai de paràmetres Θ.
2. Utilitzant la distribució predictiva *a priori* si la informació que tens a priori és sobre l'espai mostral Ω.

Aleshores, un cop definit el model Bayesià i recollides les dades obtenim un nou model Bayesià, és a dir, un model Bayesià actualitzat amb la informació que ens han aportat les noves dades per fer aquest nou model però amb la diferència de que ara la distribució *a priori* és la distribució *a posteriori.* Calculem la distribució predictiva *a posteriori* com:

$$\begin{aligned} p(\tilde{y} \mid y) &= \int p(\tilde{y}, \theta \mid y) d\theta \\ &= \int p(\tilde{y} \mid \theta, y) p(\theta, y) d\theta \\ &= \int p(\tilde{y} \mid \theta) p(\theta, y) d\theta, \end{aligned}$$

on $\tilde{y}$ és una observació futura, y és una dada observada, $p(\tilde{y}, \theta)$ és la funció de distribució conjunta i $p(\theta)$ expressa el nostre coneixement sobre l'espai de paràmetres Θ abans d'observar les dades.

Per tant, la distribució *a posteriori* $p(\theta \mid y)$ ens dóna informació sobre tot el que sabem del paràmetre després de recollir les dades i $p(\tilde{y} \mid y)$ ens aportarà el nostre coneixement sobre l'espai mostral Ω i la aprofitarem per fer prediccions.

4 Inferència Bayesiana

4.1 Definició

Definició 4.1. *La inferència estadística és el conjunt de mètodes i tècniques que permeten treure conclusions d'una població, a partir de tota la informació que se li pugui extreure a una mostra d'aquesta població que estem estudiant.*

Els principals procediments d'inferència són: els procediments d'estimació i els procediments de contrast d'hipòtesis. Per aquest motiu en aquest Capítol hi han dos apartats dedicats a explicar-los. En el punt 4.4, utilitzem variables binàries per ensenyar els procediments d'estimació de les diferents distribucions basant-nos en la resolució de problemes de decisions. L'altre procediment, ho argumentem a partir de l'explicació de dues proves d'hipòtesi.

La inferència estadística es pot dividir en dues branques segons el coneixement que tenim sobre la distribució de la població. En la inferència paramètrica es coneix la forma de la distribució però no els seus paràmetres i per tant, es fa inferència sobre aquests amb la distribució coneguda. En la inferència no paramètrica són desconeguts tant els paràmetres com la forma de la distribució, és a dir, no pot ser definida *a priori*, i per tant es fa inferència sobre característiques, per exemple la mitjana o l'ordre. Algunes de les proves no paramètriques més utilitzades són: processos neutrals per la dreta, prova de la $\mathcal{X}^2$ de Pearson, prova binomial, prova de Anderson-Darling, prova de Cochran, prova de Cohen Kappa, prova de Fisher i prova de Friedman.

Per poder entendre millor el que s'acaba d'argumentar i explicar conceptes relacionats amb la inferència per veure com canvien respecte a la inferència Freqüentista, hi ha un apartat específic dins d'aquest Capítol anomenat Inferència paramètrica Bayesiana. També hi ha un apartat dedicat a la inferència Bayesiana no paramètrica on es fa una breu pinzellada sobre el procés que conté les distribucions *a priori* més senzilles.

Hi ha una altre manera de classificar la inferència: deductiva i inductiva. La inferència és deductiva quan l'argument assegura que la veritat de les seves premisses garanteix la veritat de la seva conclusió, a partir de les lleis de la lògica i del conjunt d'hipòtesis que considerem veritables. Per l'altre part, es diu inferència inductiva quan un argument únicament assegura que la veritat de les seves premisses fa més probable que la conclusió sigui certa.

Definició 4.2. *La inferència Bayesiana és una inferència estadística basada en la distribució de probabilitat un cop donades les dades, anomenada distribució a posteriori.*

La diferència principal és que la inferència Bayesiana treballa amb el paràmetre com si fos una variable aleatòria, i es determina la distribució de la variable a partir de les dades que es van obtenint, per poder calcular les probabilitats que necessitem per aquest paràmetre. Per això, aquesta inferència esta basada en la distribució de probabilitat del paràmetre donades les dades, distribució *a posteriori*. L'únic requeriment que hem de tenir en compte per fer inferència Bayesiana és el coneixement sobre el paràmetre abans d'obtenir qualsevol informació respecte a les dades, distribució *a priori*.

En aquests últims anys la inferència ha donat una empenta a la metodologia Bayesiana en diferents àmbits. Els més destacats són l'àmbit financer i l'àmbit de la salut.

En general, els mètodes Bayesians s'han generalitzat perquè generen estudis que són més útils per a la solució de problemes en la presa de decisions i per a dur a terme estudis on la població és massa gran com per poder realitzar mesures de tots i cadascun dels individus.

Així, el problema general de la inferència consisteix a utilitzar la informació disponible per descriure, el comportament de la variable aleatòria.

4.2 Inferència paramètrica Bayesiana

Com hem vist a l'apartat anterior, l'objectiu principal de la Inferència Paramètrica Bayesiana és trobar solució als problemes de decisió amb incertesa, perquè encara que la funció de distribució de la variable aleatòria x és coneguda, el paràmetre θ és desconegut. No saber el valor d'aquest paràmetre crea una situació d'incertesa que el mètode Bayesià té en compte en el procés de presa de decisions.

Per resoldre aquest el problema, utilitzem el model predictiu de la variable aleatòria x:

$$p(x) = \int p(x \mid \theta) p(\theta) d\theta.$$

on $p(\theta)$ és el model de probabilitat que descriu la incertesa dels experts sobre θ respecte a la mostra. També pot ser a partir de la distribució *a posteriori*.

Quan a partir de la descripció de la variable fem una anàlisi d'algun atribut, com poden ser els moments, la moda o les probabilitats específiques, ens trobem en el mateix problema d'incertesa, ja que si denotem com a atribut a $\eta = \eta(\theta)$ també serà desconegut perquè el paràmetre ho és. La manera de resoldre-ho és la mateixa que abans perquè la distribució *a priori* $p(\eta)$ pot derivar-se en la distribució $p(\theta)$.

Com hem vist, és important la informació mostral en la creació d'aquests pronòstics, ja que es contempla l'efecte que té la mostra en la transformació de distribució *a priori* $p(\theta)$ a distribució *a posteriori* $p(\theta \mid x)$.

Per això mateix és tan important entendre els elements que participen en el Teorema de Bayes. Un dels primers resultats a tenir en compte és el principi de versemblança. No és sol típic del raonament Bayesià, però aquest li dóna una major importància. A grans trets ens ve a dir que: Si dues variables aleatòries X i Y tenen funcions de densitat conjuntes $p(x \mid \theta)$ i $p(y \mid \theta)$, on θ és el mateix paràmetre en tots dos models, i a més ocorre que $p(x \mid \theta) = k \cdot p(y \mid \theta)$ amb k una constant, aleshores, per a una distribució *a priori* comuna $p(\theta)$, les distribucions *a posteriori* $p(\theta \mid x)$ i $p(\theta \mid y)$ coincideixen, i per tant donen lloc a les mateixes inferències.

Un altre concepte important a tenir en compte és l'estadístic suficient Bayesià. Per poder defini-lo, denotem una mostra aleatòria com $X = (x_1, ..., x_n)$.

Definició 4.3. *Es diu que* $\mathcal{T}_n : X \to \mathbb{R}^{k(n)}$ *és un estadístic si és una variable aleatòria que és funció de la mostra i no involucra en la seva expressió cap paràmetre desconegut. Es diu a més que* $\mathcal{T}_n(X))$ *és de dimensió fixa si* $k(n) = k \quad \forall n$.

Definició 4.4. *Sigui* X *una mostra aleatòria d'una variable aleatòria* Y *amb funció de densitat* $p(y \mid \theta)$, $\theta \in \Theta$ *i sigui* $\mathcal{T}_n(X)$ *un estadístic de les dades, es diu que* $\mathcal{T}_n(X)$ *és suficient per* $\theta \Leftrightarrow p(\theta \mid X)$ *depèn de* X *només a través de* $\mathcal{T}_n(X)$ $\forall n$ *i* $\forall p(\theta)$.

Aquest concepte és equivalent tant per a l'estadística Bayesiana com per l'estadística Freqüentista, tal i com podem veure en el Teorema.

Teorema 4.5. *Sigui* X *una mostra aleatòria d'una variable aleatòria* Y, *discreta o contínua, amb funció de densitat* $p(y \mid \theta)$, $\theta \in \Theta$, *llavors:*
$\mathcal{T}_n(X)$ *és suficiènt Bayesiana* $\Leftrightarrow$ $\mathcal{T}_n(X)$ *és suficiènt Freqüentista.*

4.3 Inferència no paramètrica Bayesiana

En aquest apartat veurem els conceptes bàsics de mètodes de resolució de la inferència no paramètrica Bayesiana. Com hem dit anteriorment, ens trobem davant d'aquest tipus d'inferència quan ni tan solament coneixem la forma de les distribucions, és a dir, quan és la pròpia distribució el paràmetre a estudiar i estimar, i l'espai de paràmetres és el conjunt de totes les distribucions de probabilitats definides en un espai mostral.

La teoria de la decisió no paramètrica Bayesiana ha trobat serioses dificultats, per tant, la seva aplicació no és molt àmplia. En 1973, Ferguson utilitza els processos de Dirichlet per induir les mesures de probabilitat sobre l'espai paramètric. A l'any següent, Doksum, indueix les mesures de probabilitat a partir dels processos neutrals per la dreta, marcant una pauta per a la decisió Bayesiana no paramètrica.

Quan fem referencia a un problema de desició no paramètric canviem la visió de l'espai, és a dir, l'espai paramètric Θ ara és l'espai de totes les distribucions de probabilitat definides en un espai mostral Ω. Al parlar de procediment Bayesià com a criteri de decisió, parlem a la vegada de distribució de probabilitat sobre l'espai de paràmetres, anomenada distribució de probabilitat *a priori*.

Tal i com hem explicat abans, un dels processos importants per resoldre aquest tipus d'inferència són els anomenats: processos neutrals per la dreta. Només farem una petita introducció per veure la dificultat i el treball que comporta determinar la distribució *a posteriori* i l'esperança d'aquests problemes concrets d'estimació.

Definició 4.6. *Direm que la funció de distribució aleatòria F(t) és un procés neutral per la dreta, si es pot escriure de la forma*

$$F(t) = 1 - e^{-Y_t}$$

on Y_t *és un procés amb increments independents, tal que:*

1. Y_t *és creixent,*

2. Y_t *és continua per la dreta,*

3. $\lim_{t\to-\infty} Y_t = 0$,

4. $\lim_{t\to\infty} Y_t = \infty$.

Amb aquesta petita idea en el cap, explicarem d'una manera més estesa en que consisteixen els processos de Dirichlet P. Aquests són un cas particular dels processos neutrals per la dreta, tot i utilitzant-se abans, ja que són les probabilitats *a priori* p més senzilles dins dels problemes no paramètrics.

Les característiques principals dels processos de Dirichlet P són:

1. p és no paramètrica en el sentit que té una classe de probabilitats 'gran' o 'no paramètrica' com suport seu en la topologia de la convergència feble.

2. Si P és considerat com un paràmetre amb distribució *a priori* p, llavors la distribució *a posteriori* de P, donada una mostra, també té una distribució de Dirichlet.

3. P és una probabilitat discreta amb probabilitat 1.

A continuació farem una introducció als conceptes bàsics d'aquest procés: distribució de Dirichlet i procesos de Dirichlet(Ferguson (1973)).

Siguin $X_1, X_2, ..., X_n$ variables aleatòries independents amb distribució Gamma, $X_j \sim G(\alpha_j, 1) \forall j = 1, ..., n$ i amb $\alpha_j \geq 0$, existint algun i, de manera que $\alpha_i > 0$.

Considerem les variables aleatòries

$$Y_j = \frac{X_j}{\sum_{k=1}^{n} X_k}, \qquad j = 1, 2, ..., n.$$

A la distribució de la variable aleatòria n-dimensional $(Y_1, Y_2, ..., Y_n)$ se li denomina distribució de Dirichlet de paràmetres $(\alpha_1, \alpha_2, ..., \alpha_n)$, la qual representem per $\mathcal{D}(\alpha_1, \alpha_2, ..., \alpha_n)$.

Si $\alpha_j > 0$ per a $j = 1, 2, ..., n$, la variable aleatòria (n-1)-dimensional $(Y_1, ..., Y_{n-1})$ és absolutament contínua pel que fa a la mesura de Lebesgue en $\mathbb{R}^{n-1}$ i té per funció de densitat

$$f(y_1, y_2, ...y_{n-1} \mid \alpha_1, \alpha_2, ..., \alpha_n) = \frac{\Gamma(\alpha_1 + \alpha_2 + ... + \alpha_n)}{\Gamma(\alpha_1)\Gamma(\alpha_2)...\Gamma(\alpha_n)} \left(\prod_{j=1}^{n-1} y_j^{\alpha_j - 1} \right)$$

$$\cdot \left(1 - \sum_{j=1}^{n-1} y_j \right)^{\alpha_n - 1} I_s(y_1, y_2, ..., y_{n-1})$$

on S és el simplex,

$$S = \left\{ (y_1, y_2, ..., y_{n-1}) : y_i > 0, i = 1, ..., n-1, \sum_{i=1}^{n-1} y_i \leq 1 \right\}.$$

Proposició 4.7. *Si* $(Y_1, ..., Y_k) \in \mathcal{D}(\alpha_1, ..., \alpha_k)$ *i si* $r_1, ..., r_e$ *són nombres enters tals que* $0 < r_1 < ... < r_e = k$, *llavors,*

$$\left(\sum_{i=1}^{r_1} Y_i, \sum_{i=r_1+1}^{r_2} Y_i, ..., \sum_{i=r_{e-1}+1}^{r_e} Y_i \right) \in \mathcal{D} \left(\sum_{i=1}^{r_1} \alpha_i, \sum_{i=r_1+1}^{r_2} \alpha_i, ..., \sum_{i=r_{e-1}+1}^{r_e} \alpha_i \right).$$

Demostració. La demostració és directa de la definició de distribució de Dirichlet i de la reproductividad respecte al primer paràmetre de la distribució Gamma. □

Sigui $(\mathcal{X}, \mathcal{A})$ un espai mesurable i sigui α una mesura finita no nul·la sobre $(\mathcal{X}, \mathcal{A})$. Definim un procés estocàstic particular $\{P(A) : A \in \mathcal{A}\}$ de la següent manera:

Definició 4.8. *P és un procés de Dirichlet sobre* $(\mathcal{X}, \mathcal{A})$ *amb paràmetre* α*, si per a cada* $k \in \mathbb{N}$ *i per a cada partició mesurable* $(B_1, ..., B_k)$ *de* $\mathcal{X}$*, la distribució de la variable aleatòria k-dimensional* $(p(B_1), ..., p(B_k))$ *és de Dirichlet amb paràmetres* $(\alpha(B_1), ..., \alpha(B_k))$*.*

Per poder relacionar aquests nous coneixements amb la decisió Bayesiana no paramètrica, necessitem tenir clars uns conceptes.

Definició 4.9. *(Ferguson (1973)): Sigui P una mesura de probabilitat aleatòria sobre* $(\mathcal{X}, \mathcal{A})$*. Direm que* $X_1, ..., X_n$ *és una mostra extreta mitjançant P si* $\forall m \in \mathbb{N}$ *i conjunts mesurables* $A_1, ..., A_m, C_1, ..., C_n$ *és*

$$Pr\{X_1 \in C_1, ..., X_n \in C_n \mid P(A_1), ..., P(A_m), P(C_1), ..., P(C_n)\} = \prod_{j=1}^{n} P(C_j).$$

Teorema 4.10. ***(Kolmogorov (1933)).*** *Cada sistema de funcions de distribució* $F_{\mu_1, \mu_2, ..., \mu_n}$ *satisfent les condicions (1) i (2) a baix expressades, defineix una funció de probabilitat p(A) sobre* F^M *(àlgebra dels conjunts cilíndrics de Borel). Aquesta funció de probabilitat p(A) pot ser estesa, mitjançant el Teorema d'extensió a* BF^M *(mínima* δ*-àlgebra que conté a* F^M*).*

1. $F_{\mu_{i_1}, \mu_{i_2}, ..., \mu_{i_n}}(a_{i_1}, a_{i_2}, ..., a_{i_n}) = F_{\mu_1, \mu_2, ..., \mu_n}(a_1, a_2, ..., a_n)$
2. $F_{\mu_1, \mu_2, ..., \mu_k}(a_1, a_2, ..., a_k) = F_{\mu_1, \mu_2, ..., \mu_k}(a_1, a_2, ..., a_k, +, ..., +\infty)$

on $k < n$ *i*

$$\begin{pmatrix} 1, & 2, & ..., & n \\ i_1, & i_2, & ..., & i_n \end{pmatrix}$$

és una permutació arbitrària.

Definició 4.11. *Direm que* $(B_1, ..., Bk)$ *és una partició mesurable* X *si* $B_i \in \mathcal{A}$ $\forall i$, $B_i \cap B_j =$ *per a* $i \neq j$ *i* $B_1 \cup B_2 \cup ... \cup B_k = \mathcal{X}$.

Condició C. Ferguson (1973): Si $B_1^{'}, ..., B_k^{'}$ i $(B_1, ..., B_k)$ són particions mesurables, i si $(B_1^{'}, ..., B_k^{'})$ és un refinament de $(B_1, ..., B_k)$ amb $B_1 = \bigcup_{i=1}^{r_1} B_i^{'}$, $B_2 = \bigcup_{i=r_1+1}^{r_2} B_i^{'}, ..., B_k = \bigcup_{i=r_{k-1}+1}^{k^{'}} B_i^{'}$ llavors la distribució de

$$\left(\sum_{i=1}^{r_1} P(B_i^{'}), \sum_{i=r_1+1}^{r_2} P(B_i^{'}), ..., \sum_{i=r_{k-1}+1}^{k^{'}} P(B_i^{'}) \right)$$

determinada a partir de la distribució conjunta de $(P(B_1^{'}), ..., P(B_{k^{'}}^{'}))$ és idèntica a la distribució de $(P(B_1), ..., P(B_k))$.

A partir de la proposicó anterior 4.7 es veu claralment que es compleix la condició C de consistència de Ferguson. A més, com a col·lorari del Teorema 4.10, sabem que existeix una mesura de probabilitat *a priori* sobre l'espai de paràmetres.

I per acabar aquesta apartat, donem la idea de com seria una mostra extreta mitjançant un procés de Dirichlet. La col·lecció de variables aleatòries $X_1, ..., X_n$ amb valors sobre $(\mathcal{X}, \mathcal{A})$ es diu que és una mostra de grandària n obtinguda a través d'un procés de Dirichlet P sobre $(\mathcal{X}, \mathcal{A})$ amb paràmetre α, si para tot $m \in N$ i conjunts $\mathcal{A}$-mesurables $A_1, ..., A_m, C_1, ..., C_n$ és

$$Pr \{X_1 \in C_1, ..., X_n \in C_n \mid P(A_1), ..., P(A_m), P(C_1), ..., P(C_n)\} = \prod_{j=1}^{n} P(C_j).$$

Com es veu, aquesta definició no és més que un cas particular de la definició 4.9, per ser un procés de Dirichlet un cas particular de probabilitat aleatòria.

Teorema 4.12. *Ferguson (1973): Sigui* P *un procés de Dirichlet sobre* $(\mathcal{X}, \mathcal{A})$ *amb paràmetre* α*, i sigui* $X_1, ..., X_n$ *una mostra de grandària* n *extreta mitjançant* P*. Llavors, la distribució condicional de* P *donada* $X_1, ..., X_n$ *és un procés de Dirichlet de paràmetre* $\alpha + \sum_{i=1}^{n} \delta_{X_i}$*, on per a cada* $x \in \mathcal{X}$*,* δ_X *denota la mesura sobre* $(\mathcal{X}, \mathcal{A})$ *que dóna massa un al punt* x*:*

$$\delta_X(A) = \begin{cases} 1 & si \quad x \in \mathcal{A} \\ 0 & si \quad x \notin \mathcal{A} \end{cases}$$

Com es veu, la distribució *a posteriori* d'un procés de Dirichlet és un procés de Dirichlet molt fàcilment manejable, la qual cosa permet estimar amb facilitat.

4.4 Inferència Bayesiana en variables binàries

4.4.1 Definició

La pràctica més habitual és la d'utilitzar proporcions o percentatges per poder classificar en grups una població. Com ja sabem, una població és un conjunt d'unitats de la que a partir de la inferència volem saber-ho tot. Podem suposar que aquesta població esta dividida per un nombre d'èxits, on θ és la seva probabilitat (paràmetre desconegut), i per un nombre de fracassos. Per poder treure una conclusió del paràmetre s'utilitzen dades, de manera que, a partir d'una mostra aleatòria de n individus hi ha a èxits i b fracassos.

La manera més ràpida i senzilla, però també la que comporta diferents problemes, de fer inferència sobre el percentatge d'èxits d'aquesta mostra, és calcular el quocient: a/n.

Per evitar aquest problemes s'utilitza la distribució *a priori* i la distribució *a posteriori* que porta implícit l'anàlisi Bayesià. La distribució *a posteriori*, com hem dit en altres apartats, conté tota la informació de la proporció una vegada observades les dades. Amb aquesta distribució podem treure diferents conclusions segons el que ens interessi depenent de l'experiment, sense oblidar-nos que el nostre objectiu sempre és aconseguir el valor del paràmetre desconegut. Un dels casos més comuns és buscar un valor en particular de θ que tingui una probabilitat alta/mitjana/baixa sobre la distribució *a posteriori.* També podem estimar θ a través d'un interval de probabilitat. A la pràctica el nostre interès va lligat amb els esdeveniments futurs, així que en aquest cas ens trobem amb un problema de predicció de l'anàlisi Bayesià.

Tot això es tradueix a considerar una variable aleatòria X que segueix una distribució Binomial de paràmetres n i θ, denotada com: $X \sim \mathcal{B}(n, \theta)$. L'expressió que defineix la seva versemblança és:

$$p(x \mid \theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x} = \frac{n!}{x!(n-x)!}\theta^x(1-\theta)^{n-x}, \qquad x = 0, 1, ..., n.$$

Tenim que el paràmetre d'interès és θ que determina la probabilitat de tenir un èxit i n el nombre de vegades que es realitza l'experiment amb resultat d'èxit o fracàs. Considerem independents les proves però amb la mateixa probabilitat d'èxit, θ.

Com sempre, atès que el nostre interès s'enfoca en θ, de l'expressió anterior solament agafem les dades concretes en els quals està implicada. D'aquesta manera, podem escriure la funció de versemblança com una proporció:

$$L_{X=x}(\theta) \propto \theta^x(1-\theta)^{n-x}.$$

Per desenvolupar els mètodes Bayesians sobre aquest tipus de problemes de proporcions, a part de la funció de versemblança, necessitem construir la distribució *a priori*, que reflecteix l'aprovació o justificació dels experts per a cada valor del paràmetre. Hi ha dos mètodes per construir-la segons el tipus de variable aleatòria amb la que treballem. Els mètodes discrets, on consideren que θ solament pot tenir valors dins d'un conjunt finit, i els mètodes continus, on θ és valorada de manera contínua en l'interval (0,1).

4.4.2 Distribució *a priori* en el cas discret

A partir del percentatge que se li assigna als individus de la mostra segons el tema que estiguem estudiant, denotem com a distribució *a priori* la probabilitat de cadascun dels casos segons l'opinió dels experts en aquell moment. Habitualment, per assignar aquestes probabilitats poden prendre dos valors de referència per després assignar cadascuna de les probabilitats a la resta dels valors mitjançant comparació amb aquests valors de referència.

A través del càlcul de l'expressió proporcional a la funció de versemblança, obtenim la justificació que tenen cadascun de les dades observades sota cadascun dels possibles models considerats. Amb tot això, ja tenim tots els ingredients per poder aconseguir la distribució *a posteriori*, ja que solament queda dividir el valor extret de la multiplicació entre el valor de la distribució *a priori* i el valor de la funció de versemblança, per la suma de tots ells.

Amb aquesta nova informació, els experts ja poden tenir una nova opinió de l'experiment o acabar de confirmar que les seves idees inicials eren les correctes.

Tots aquests passos per aconseguir nova informació ens poden semblar coneguts. Això es deu al fet que és el procediment que ja hem explicat i vist a l'apartat 2.2.

4.4.3 Distribució *a priori* en el cas continu

Aquest cas és el que ens trobem en realitzar estudis reals, ja que el paràmetre θ pot tenir qualsevol valor dins de l'interval (0,1) que implica, que tindríem un model diferent per a cada valor assignat a θ.

L'anàlisi Bayesià, en treballar amb variables aleatòria contínues, recomana utilitzar densitats *a priori* que permetin un fàcil maneig. En el cas del problema de proporcions, s'utilitza la distribució Beta. Llavors tenim que: $\theta \sim \mathcal{B}e(a,b)$, amb $a, b > 0$.

Recordem l'expressió d'aquesta distribució per poder continuar amb l'explicació. L'expressió

$$\beta(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx$$

representa l'àrea sota la corba $f(x) = x^{a-1}(1-x)^{b-1}$ en l'interval [0,1], de manera que tenim:

$$\beta(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

on $\Gamma(a)$ representa la funció Gamma com

$$\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}dx = (a-1)\Gamma(a-1).$$

Aleshores per formar la funció de densitat, diem que θ és la proporció d'èxits i per tant, estarà en l'interval (0,1). Així que, si dividim qualsevol funció f(x), declarada anteriorment, per $\beta(a,b)$ ens surt aquesta funció de densitat amb la següent expressió:

$$p(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}, \qquad 0 < \theta < 1 \quad a, b > 0.$$

La distribució Beta és suficientment flexible com per recollir la major part de les possibles creences de l'expert ja que, la densitat adopta formes absolutament diferents per a diferents valors del paràmetre. Existeix un catàleg on es recullen les formes i les seves relacions amb els seus paràmetres gràcies a Homberg.

4.5 Contrast d'hipòtesi

4.5.1 Prova d'hipòtesi per a una mostra amb dues hipòtesis

Suposem donades dues hipòtesis i volem triar una:

$$\begin{cases} H_1 : \theta \in \Theta_1 \rightarrow P(X \mid \theta) \in \{P(X \mid \theta), \theta \in \Theta_1\} = M_1 \\ H_2 : \theta \in \Theta_2 \rightarrow P(X \mid \theta) \in \{P(X \mid \theta), \theta \in \Theta_2\} = M_2 \end{cases}$$

Si partim d'un model estadístic $M = \{P(Y \mid \theta), \theta \in \Theta\}$, per poder comprovar les hipòtesis hem de dividir l'espai de paràmetres en dos, de la següent manera:

$$M = \{P(Y \mid \theta), \theta \in \Theta\} = \{P(Y \mid \theta), \theta \in \Theta_1\} \cup \{P(Y \mid \theta), \theta \in \Theta_2\} = M_1 \cup M_2,$$

on $\Theta_1 \cup \Theta_2 = \Theta$ i $\Theta_1 \cap \Theta_2 = \emptyset$.

L'idea principal d'aquest contrast d'hipòtesis és trobar el verdader paràmetre poblacional θ^*, per tant, saber si $\theta^* \in \Theta_1$ o $\theta^* \in \Theta_2$. Per fer-ho, partirem de l'aproximació Bayesiana, és a dir, necessitem una distribució *a priori*. Aquestes són:

$$p(H_1) = p(\theta \in \Theta_1) = \int_{\Theta_1} p(\theta)d\theta, \qquad p(H_2) = p(\theta \in \Theta_2) = \int_{\Theta_2} p(\theta)d\theta.$$

Un cop observats els resultats, ens basem en la distribució *a posteriori* per acabar de treure conclusions:

$$p(H_1 \mid y) = p(\theta \in \Theta_1 \mid y) = \int_{\Theta_1} p(\theta \mid y)d\theta,$$

$$p(H_2 \mid y) = p(\theta \in \Theta_2 \mid y) = \int_{\Theta_2} p(\theta \mid y)d\theta.$$

La primera observació que fem és fixar-nos en la hipòtesis amb més probabilitat *a posteriori*, és a dir:

Si $p(H_1 \mid y) > p(H_2 \mid y)$ triem H_1 i si $p(H_2 \mid y) > p(H_1 \mid y)$ triem H_2 on $p(H_1 \mid y) + p(H_2 \mid y) = 1$.

Aquest seria el cas més fàcil perquè coneixem les distribucions. En el cas de que no fos així, que no coneguéssim $p(\theta \mid y)$ però si que la sabéssim simular la distribució, podem calcular la seva probabilitat com el tant per cent de valors simulats que pertanyen a l'espai del paràmetre.

L'altre observació és fixar-nos en el odds. En l'estadística el odds és la raó entre la probabilitat que un esdeveniment ocorri i la probabilitat que no. En el nostre cas, el oods *a posteriori* de la hipòtesis 1 envers de la hipòtesis 2 és de la següent manera:

$$\lambda_B = \frac{p(H_1 \mid y)}{1 - p(H_1 \mid y)} = \frac{p(H_1 \mid y)}{p(H_2 \mid y)} = \frac{\frac{p(H_1)p(y|H_1)}{p(y)}}{\frac{p(H_2)P(y|H_2)}{p(y)}} = \frac{p(H_1)p(y \mid H_1)}{p(H_2)p(y \mid H_2)},$$

on si el resultat és més gran que 1 triaríem H_1 i si és més petit, H_2.

A partir d'aquest resultat podem observar dues idees importants del raonament Bayesià. Primer el valor de l'aproximació Bayesiana λ_B amb l'ajuda de presa de decisió, ja que si el seu valor és molt gran escollim l'hipòtesis 1. L'altre és que tenim diferents valors importants:

- odds *a priori* $= \frac{p(H_1)}{p(H_2)}$,
- odds *a posteriori* $= \frac{p(H_1|y)}{p(H_2|y)}$,
- Factor de Bayes $FB_{12} = \frac{p(y|H_1)}{p(y|H_2)}$,

on [odds *a posteriori*]=[odds *a priori*]· Factor de Bayes.

Els valors del Factor de Bayes oscil·len entre 0 i infinit, de manera que per poder valorar quina de les dues evidències és millor s'utilitza, normalment, una Taula de calibratge. Si observem com està definit el Factor de Bayes podem dir que si el seu valor és pròxim a 1, indica que és irrellevant distingir entre les dues hipòtesis, ja que els seus valors són molt semblants, i si el seu valor és superior a 1, aleshores donem més importància a la hipòtesi H1. Però per saber quin nivell de certesa l'hi hem de donar a la hipòtesi H1 davant de H2, la Taula de valors de referència més emprada és l'aportada per Jeffreys. Aquesta Taula ens diu que:

- Si $FB_{12} \in [1, 3.2] \Rightarrow$ Evidència a favor de H_1 fluixa.
- Si $FB_{12} \in [3.2, 10] \Rightarrow$ Evidència a favor de H_1 substancial.
- Si $FB_{12} \in [10, 32] \Rightarrow$ Evidència a favor de H_1 forta.
- Si $FB_{12} \in [32, 100] \Rightarrow$ Evidència a favor de H_1 molt forta.
- Si $FB_{12} \in [100, \infty) \Rightarrow$ Evidència a favor de H_1 decisiva.

4.5.2 Prova d'hipòtesi per a una mostra amb més de dues hipòtesis

Per resoldre aquest problema partim del mateix model estadístic que l'apartat anterior, $M = \{P(Y \mid \theta), \theta \in \Theta\}$, però en aquest cas no saben en quin dels k espais de paràmetre pertany θ^*.

El nostre contrast d'hipòtesis correspon a:

$$\begin{cases} H_1 : \theta \in \Theta_1 \to P(X \mid \theta) \in \{P(X \mid \theta), \theta \in \Theta_1\} = M_1 \\ H_2 : \theta \in \Theta_2 \to P(X \mid \theta) \in \{P(X \mid \theta), \theta \in \Theta_2\} = M_2 \\ \qquad\qquad\qquad\qquad \vdots \\ H_k : \theta \in \Theta_k \to P(X \mid \theta) \in \{P(X \mid \theta), \theta \in \Theta_k\} = M_k \end{cases}$$

on $\bigcup_{i=1}^{k} \Theta_i = \Theta$ i $\Theta_i \cap \Theta_j = \emptyset \ \forall i \neq j$.

Ara igual que abans, per triar la hipòtesis que més s'aproximi al model que volem, tenim dues maneres clau:

1. Calcular la distribució *a posteriori* associada a cada hipòtesis.
2. Calcular el Factor de Bayes de la H_i respecte H_j. D'aquesta manera aprofitant la proposta de Jeffreys, podem dir en quin interval es troba el valor de $log_{10}(FB_{ij})$ per tenir un criteri de quanta evidència tenen les dades a favor de H_i respecte de H_j.

5 Mètodes de simulació

5.1 Introducció

Com a conlusió del Capítol anterior, tenim que la infercència Bayesiana es basa en el comportament de la distribució *a posteriori*:

$$p(\theta \mid y) = \frac{p(y,\theta)}{p(y)} = \frac{p(y \mid \theta)}{\int p(y \mid \theta)p(\theta)d\theta} \propto \frac{L_{Y=y}(\theta)p(\theta)}{\int L_{Y=y}(\theta)p(\theta)d\theta}.$$

El càlcul del numerador és trivial ja que només és el producte de dos funcions. Si volem fer intervals de probabilitat o prediccions, que són els casos més comuns, necessitem el valor de tota l'expressió $p(\theta \mid y)$: hem de calcular la integral. Tenir la distribució *a posteriori* de manera explícita moltes vegades és impossible per culpa d'aquest càlcul. Per això es va crear el concepte de Computació Bayesiana, que fa referència a qualsevol procediment que ens permeti aproximar l'expressió $p(\theta \mid y)$.

Els mètodes més utilitzats són els explicats a continuació.

5.2 Mètode de Monte Carlo (MC)

Aquest mètode ens aporta una estructura general per estimar integrals definides finit-dimensionals. Les condicions amb les quals ens trobem per poder aplicar-ho són les següents.

Suposem que la integral que volem simular és $I = \int_\Omega f(x)dx$, on el seu volum sobre el domini que volem integrar és $V = \int_\Omega 1dx$. La visió més clàssica ens diu: suposem mostres suficientment grans $X_1, X_2, ..., X_n$ independents i idènticament distribuïdes d'una distribució uniforme definides en Ω. Podem fer una aproximació de la integral de la forma:

$$\tilde{I} \approx V \frac{1}{n}\sum_{i=1}^{n} f(X_i) = \left(\int_\Omega 1dx\right)\frac{1}{n}\sum_{i=1}^{n} f(X_i).$$

Podem donar per valgut el resultat anterior gràcies a la Llei dels grans nombres, que assegura que aquesta estimació convergeix en mitjana al valor de la integral original. A partir d'això, podríem calcular l'estimació de la variància de I per obtenir errors d'estimació.

Fins a aquí seria la idea bàsica per estimar una integral a partir del mètode de Monte Carlo. També ens pot interessar l'estimació de l'esperança d'una distibució en concret. Obtindríem:

$$E\left[f(x)\right] \approx \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Per tant, el mètode aplicat al raonament Bayesià ens diu: si suposem que els paràmetres $\theta^{(1)}, \theta^{(2)}, ...$ són una successió de variables independents amb distribució *a posteriori*, aleshores amb probabilitat un tenim

$$\lim_{m} \sum_{i=1}^{m} h(\theta^{(i)}) = E\left(h(\theta) \mid x\right) = \int h(\theta) p(\theta \mid x) d\theta,$$

on $h(\theta)$ és una funció mesurable del paràmetre θ, és a dir, $h(\theta)$ és un altre paràmetre i el podem denotar com $\upsilon = h(\theta)$. Per tant si tractem a θ com a una variable aleatòria, a υ també. Aleshores, la distribució *a posteriori* de υ és $\tilde{p}(\upsilon \mid x)$, sent

$$E\left(h(\theta) \mid x\right) = E\left(\upsilon \mid x\right) = \int \upsilon \tilde{p}(\upsilon \mid x) d\upsilon.$$

Una altra propietat d'aquesta integral és que té tantes dimensions com θ. Per contra, podem expressar qualsevol integral a l'espai de paràmetres com una simulació de la distribució *a posteriori* mitjançant la definició aproximada de $h(\theta)$.

Per tant per aproximar la integral $E(h(\theta) \mid x)$ n'hi ha prou amb simular la distribució *a posteriori* $\tilde{\pi}(\upsilon \mid x)$ un gran nombre de vegades. El mateix passa quan es vol simular la distribució *a posteriori* de $\upsilon = h(\theta)$. Es pot simular θ de $p(\theta \mid x)$ i calcular $\upsilon = h(\theta)$, és a dir, $\upsilon^{(1)} = h(\theta^{(1)}), \upsilon^{(2)} = h(\theta^{(2)}), ...$, són simulacions de la distribució *a posteriori* de $\upsilon = h(\theta)$.

Si d'altra banda, tenim el gràfic aproximat de la distribució *a posteriori* de θ^i, solament podem estimar la integral de la mostra promig com $\frac{1}{L}\sum_{i=1}^{L} h(\theta^i)$. Per a qualsevol nombre finit de simulacions de gràfics, l'exactitud de l'estimació pot avaluar-se aproximadament per la desviació estàndard dels valors de $h(\theta^i)$. Si no és fàcil extreure-ho de la distribució *a posteriori*, o si els valors de $h(\theta^i)$ són molt variables, de manera que la mitjana de la mostra és massa variable per ser un càlcul fàcil, llavors són necessaris mètodes més sofisticats per a la integració numèrica.

5.3 Mètode de Monte Carlo i cadenes de Markov (MCMC)

A continuació farem una iniciació al mètode de Monte Carlo i cadenes de Markov. En l'apartat anterior ja hem vist la simulació del mètode de Monte Carlo, però ara a més a més tenim les cadenes de Markov, que fa que sigui un mètode més complex. Aleshores, el primer que veurem serà una petita explicació sobre que són les cadenes de Markov, ja que és un terme que no ha sortit fins ara. Aquest tema és molt ampli, de fet, es podria fer un treball només d'això. Una breu explicació ens ajudarà a posar-nos en context.

5.3.1 Cadenes de Markov

Per simplificar l'explicació ens centrarem només en les cadenes amb espais d'estats finits. Tal i com ja hem dit, només farem un incís en definicions i propietats, per entendre la idea fonamental de les cadenes de Markov. Els dos conceptes claus són: la distribució de probabilitat *a priori* dels estats de la cadena i la matriu de probabilitats de transicions. Aquests dos conceptes regulen l'evolució d'una cadena de Markov. Un altre aspecte important és saber si després d'un número de transicions, la cadena de Markov convergeix a alguna distribució d'equilibri o estacionaria, independentment de la distribució de probabilitat *a priori* i si aquesta distribució d'equilibri és única.

Ho dividirem en dues parts: la primera on expliquem les cadenes de Markov a temps discret, i la segona part amb les cadenes de Markov a temps continu.

Sigui I un conjunt numerable. Cada $i \in I$ es diu estat i I és l'espai d'estat. Diem que $\lambda = (\lambda_i : i \in I)$ és una mesura en I si $0 \leq \lambda_i < \infty$ per a tots $i \in I$. Si a més $\sum_{i \in I} \lambda_i = 1$, llavors λ és una distribució.

La λ defineix una distribució, i si X és una variable aleatòria llavors, λ és la distribució de X. L'objectiu és modelar X que pren el valor i amb probabilitat λ_i. És a dir, $\lambda_i = Pr(X = i) = Pr(\{w : X(x) = i\})$.

Diem que una matriu $P = (p_{ij} : i, j \in I)$, la matriu que guarda les probabilitats de transició, és estocàstica si cada fila $(p_{ij} : j \in I)$ és una distribució. Amb aquesta informació ja podem donar una definició formal de les cadenes de Markov a partir de la matriu corresponent P.

Definició 5.1. *Diem que $(X_n)_{n \geq 0}$ és una cadena de Markov a temps discret amb una distribució a priori $\lambda = (\lambda_i, i \in I)$ i una matriu de transició $P = (p_{ij}, i, j \in I)$, per a $i_0, ..., i_{n+1} \in I$, si*

1. *X_0 té distribució λ: $Pr\{X_0 = i_0\} = \lambda_{i_0}$,*
2. *Per a $n \geq 0$, condicionada amb $X_n = i$, X_{n+1} té distribució $(p_{ij} : j \in I)$ i és independent de $X_0, ..., X_{n-1}$, és a dir,*

$$Pr(X_{n+1} = i_{n+1} \mid X_0 = i_0, X_1 = i_1, ..., X_n = i_n) = p_{i_n i_{n+1}}.$$

Diem que $(X_n)_{0 \leq n}$ és $Markov(\lambda, P)$ per abreujar.

Amb les condicions anteriors ja podríem entendre els exemples més habituals al món real que utilitzen cadenes de Markov, però matemàticament encara podríem plantejar i demostrar molts resultats. Per exemple, un d'ells seria la propietat forta de Markov. Per poder demostrar aquest teorema necessitem els coneixements previs següents.

Teorema 5.2. ***(Propietat de Markov).*** *Sigui $(X_n)_{n \geq 0}$ cadena de Markov(λ, P). Aleshores, està condicionat a que $X_m = i, (X_{m+n})_{n \geq 0}$ és Markov(δ_i, P) i és independent de les variables aleatòries $X_0, ..., X_m$.*

Anomenem temps de parada a una variable aleatòria $T : \Omega \to \{0, 1, 2, ...\} \cup \{\infty\}$ si l'esdeveniment $\{T = n\}$ només depèn de $X_0, X_1, ..., X_n$ por a $n = 0, 1, 2, ...$

Teorema 5.3. ***(Propietat forta de Markov).*** *Sigui $(X_n)_{n \geq 0}$ Markov (λ, P) i T el temps de parada de $(X_n)_{n \geq 0}$. Aleshores, condicionat amb $T < \infty$ i $X_T = i$, $(X_{T+n})_{n \geq 0}$ és Markov (δ_i, P) i independent de $X_0, X_1, ..., X_T$.*

Demostració. Si B és un esdeveniment determinat per $X_0, X_1, ..., X_T$, aleshores $B \cap \{T = m\}$ és determinat per $X_0, X_1, ..., X_m$. Per tant, per la propietat de Markov al temps m tenim que

$$\begin{aligned} &Pr(\{X_T = j_0, X_{T+1} = j_1, ..., X_{T+n} = j_n\} \cap B \cap \{T = m\} \cap \{X_T = i\}) \\ &\quad = Pr_i(X_0 = j_0, X_1 = j_1, ..., X_n = j_n) Pr(B \cap \{T = m\} \cap \{X_T = i\}) \end{aligned}$$

on hem utilitzat de $T = m$ per reemplaçar m per T. Ara, sumem per a $m = 0, 1, 2, ...$ i dividim per $Pr(T < \infty, X_T = i)$ per obtenir

$$\begin{aligned} &Pr(\{X_T = j_0, X_{T+1} = j_1, ..., X_{T+n} = j_n\} \cap B \mid T < \infty, X_T = i) \\ &\quad = Pr_i(X_0 = j_0, X_1 = j_1, ..., X_n = j_n) Pr(B \mid T < \infty, X_T = i). \end{aligned}$$

□

Quan parlem de les cadenes de Markov a temps continu hem de canviar la nostra perspectiva: ens trobem amb un procés estocàstic de temps continu. Si $\{X_t\}_{t \geq 0}$ és un procés estocàstic a temps continu, $t \in [0, T]$ amb $T \in \mathbb{R}$ és fix i pren valors en un conjunt numerable.

Definició 5.4. *$\{X_t\}_{t \geq 0}$ és una cadena de Markov a temps continu si per a tot $s, t \geq 0$ i possibles estats $i, j, x(u) \in I$, $0 \leq u < s$, tenim*

$$\begin{aligned} &Pr(X_{t+s} = j \mid X_s = i, X_u = x(u), 0 \leq u < s) = Pr(X_{t+s} = j \mid X_s = i) \\ &\quad = Pr(X_t = j \mid X_0 = i) := p_{ij}(t). \end{aligned}$$

Definició 5.5. *Si $\{X_t\}_{t \geq 0}$ és una cadena de Markov en temps continu i a més verifica que $Pr(X_{t+s} = j \mid X_s = i)$ és independent de s llavors es diu que la cadena de Markov en temps continu té probabilitats de transició estacionáries o homogènies.*

Aleshores, una cadena de Markov a temps continu és un procés estocàstic que verifica la propietat de Markov: la probabilitat condicional d'un futur estat en el temps $t + s$, donat l'estat present en el temps s i tots els estats passats, solament depèn del present estat i és independent del passat.

Si denotem per T_i al temps que el procés es troba dins de l'estat i abans d'anar a un altre estat diferent, llavors per la propietat de Markov i l'homogeneïtat en temps

$$Pr(T_i > s + t \mid T_i > s) = Pr(T_i > t)$$

i es pot demostrar llavors que T_i ha de tenir una distribució exponencial. Aquest temps de permanència dins d'un estat amb distribució exponencial és independent per a cada estat.

A partir d'aquesta propietat i amb la primera definició, podem donar una altra definició molt més intuïtiva sobre les cadenes de Markov a temps continu:

- La quantitat de temps que el procés es troba dins de l'estat i, una vegada entra en ell, té una distribució exponencial de paràmetre q_i.

- Quan el procés deixa l'estat i i entra a l'estat j, ho fa amb probabilitat de transició π_{ij}. Aquesta probabilitat ha de satisfer

$$\pi_{ii} = \begin{cases} 0 & si \;\; q_i \neq 0 \\ 1 & si \;\; q_i = 0 \end{cases} \qquad , \sum_{j \in I} \pi_{ij} = 1.$$

Igual que en el cas de les cadenes de Markov a temps discret, per a temps continu tenim que

$$Pr(X_{t+s} = j \mid X_t = i) = p_{ij}(t, s)$$

és la probabilitat de passar en temps t de l'estat i al j en s unitats de temps, i

$$P(t, s) = (p_{ij}(t, s))_{i,j \in I},$$

on P denota la matriu de transicions de igual manera que en el cas discret, però les entrades de la matriu p_{ij} són diferents encara que tinguin la mateixa notació.

En tots dos casos volem resoldre una qüestió important: si la cadena de Markov convergeix a una distribució límit, independent de qualsevol distribució *a priori*, és a dir, necessitem que P^n convergeixi a alguna matriu invariant on en el límit té files iguals. Per poder explicar això afegim nova notació. Denotem a $\lambda^{'(n)}$ com el vector fila N-dimensional que denota la distribució de probabilitat de X_n. De manera que la component i-ésima de λ^n és $\lambda^n(i) = Pr(X_n = i), i \in I$.

Suposem que λ, des de $\lambda^{'(n)} = \lambda^{'(0)} P^n$, és un vector de probabilitats límit com $\lambda^{'} = \lim_{n \to \infty} \lambda^{'(0)} P^n$. Aleshores si la distribució λ compleix que

$$\begin{aligned} \lambda^{'} &= \lim_{n \to \infty} \lambda^{'(0)} P^{n+1} \\ &= \left(\lim_{n \to \infty} \lambda^{'(0)} P^n \right) P \\ &= \lambda^{'} P \end{aligned}$$

diem que és una distribució estacionaria o d'equilibri.

5.3.2 Mètode MCMC

Coneixent les principals propietats i definicions de les cadenes de Markov, podem especificar que el métode de Monte Carlo, explicat anteriorment, és un cas particular del de Monte Carlo amb cadenes de Markov. Quan $\{X_t\}_{t \geq 0}$ són independets i idènticament distribuïdes utilitzem el métode Monte Carlo, ja que les cadenes de Markov són, en aquest cas, estacionàries i reversibles. Per entendre millor la diferència entre aquests dos mètodes expliquem aquests dos últims conceptes.

Com ja hem dit, les cadenes de Markov són un cas particular de un procés estocàstic. Un procés estocàstic és estacionari si per a cada $k \geq 0$ la k-èssima distribució $(X_{n+1}, ..., X_{n+k})$ és independent de n, és a dir, si es invariant davant una translació de temps. Per tant, una cadena de Markov és estacionària si es tracta de un procés estacionari. Per l'altra banda, una cadena de Markov és reversible respecte a una distribució *a priori* λ_i si, $\forall i, j \in I$ tenim que $\lambda_i p_{ij} = \lambda_i p_{ji}$. Hi ha una relació unidireccional entre aquestes dues propietats: si una cadena és reversible implica que és estacionària.

El mètode de Monte Carlo amb cadenes de Markov ha esdevingut una eina computacional molt important en la estadística Bayesiana, des de que permet dibuixar inferècies donades a partir d'una distribució *a posteiori* complexe, que les altres tècniques analítiques o numèriques d'integració no són capaces de resoldre. La idea principal d'aquest mètode és generar una cadena de Markov iterant la simulació de Monte Carlo, per tenir la distribució *a posteriori* desitjada com distribució estacionària.

Per poder explicar en que consisteix el mètode MCMC treballem amb cadenes de Markov en temps continu. A més utilitzem una notació diferent, ja que en el cas continu no té sentit escriure $p(i,j) = Pr(X_n = j \mid X_{n-1} = i)$ perquè sempre és zero. Per tant, anomenrarem les matrius de transició conegudes fins ara en el cas discret com nucli de transició P(.,.). Suposem que X és un vector p-dimensional estocàstic amb densitat f en $\mathbb{R}^p$ i que A és un subconjunt de $\mathbb{R}^p$. Aleshores P(.,.) especifica la probabilitat condicionada que X_{n+1} té en A donat que $X_n = x$ i s'escriu com

$$P(x, A) = Pr(X_{n+1} \in A \mid X_n = x). \tag{5.1}$$

Per tant, en les cadenes de Markov la matriu de probabilitat de la transició P és coneguda ja que es pot calcular, però el cas dels algoritmes de MCMC és just al contrari: la distribució d'equilibri és coneguda però el nucli de transició és desconegut.

Denotem X com un vector estocàstic real de paràmetres desconeguts amb una densitat *a posteriori* p en $\mathbb{R}^d$. Normalment, aquesta densitat sol tenir una forma bastant complexe tal que els resultats que es necessiten per fer l'estudi no es poden calcular analíticament o utilitzant tècniques estàndard de integració numèrica. En particular, la densitat p només pot ser coneguda fins a una desconeguda constant normalitzada.

La simulació directe de p pot ser molt difícil de calcular, per això es construeixen cadenes de Markov $(X_n)_{n\geq 0}$, on la seva construcció és bastant més fàcil, i la distribució invariant té la densitat donada per p. Les cadenes de Markov estàn especificadas en termes de la distribució de l'estat inicial X_1 i del nucli de transició P, que especifica la distribució condicional de X_{i+1} donat l'estat anterior X_i. Si el valor d'estat actual és $X_i = x$, llavors la probabilitat que X_{i+1} estigui al conjunt $A \subseteq \mathbb{R}^d$ és donada per la probabilitat del nucli de transició (5.1).

Si la cadena de Markov generada és irreductible amb la distribució invanriant p, passem a utilitzar la estimació de Monte Carlo de diverses esperances $E(h(X))$ pel que fa a p, és a dir, per a qualsevol funció h en $\mathbb{R}^d$ amb esperança finita $E(h(X))$,

$$E(h(X)) = \int h(x)p(x)dx = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} h(X_i).$$

Aleshores podem aproximar $E(h(X))$ amb la mitjana de la mostra de la forma

$$E(h(X)) \approx \frac{1}{N}\sum_{i=1}^{N} h(X_i),$$

per a N suficientment grans.

La densitat p és invariant en les cadenes de Markov si el nucli de transició P de la cadena de Markov conserva p, és a dir, si

$$\int_{\mathbb{R}^d} P(x,B)p(x)dx = \int_B p(x)dx,$$

on el dos termes de la expressió estàn igualats a la probabilitat marginal que $X \in B$, sempre que p sigui la densitat invariant. Per a verificar que p és la densitat invariant utilitzant la darrere igualtat és una tasca difícial, ja que tenim una integració respecte a p. Per aquesta raó s'utilitza abans el mètode de Monte Carlo amb cadenes de Markov. No obstant, escollir un nucli de transició que imposi una forta condició de reversibilitat pel que fa a p és suficient per garantir que p és invariant per la cadena de Markov.

L'algorisme de Metropolis–Hastings ofereix una recepta pràctica per la construcció d'una cadena de Markov reversible amb la desitjada distribució invariant per una adequada selecció d'un nucli de transició. A més a més d'aquest algoritme hi ha d'altres també molt coneguts: Gibbs Sampler, cas particular de l'algoritme Metropolis–Hastings i molt popular per la seva simplicitat computacional, l'algoritme Langevin–Hastings i Reversible Jump MCMC.

6 Aplicació al món real

6.1 Introducció

En aquest Capítol podem observar com és en realitat un estudi d'un cas real utilitzant la inferència Bayesiana. D'alguna manera, podem observar tots i cadascun dels conceptes Bayesians descrits al llarg d'aquest treball utilitzats per poder treure unes bones conclusions sobre un estudi en concret. Per tant, aquest apartat esta destinat a la visió real i pràctica de la inferència Bayesiana.

Dividim el Capítol en diferents parts on expliquem per separat les conclusions de manera gràfica dels paràmetres de cada model que conjuntament expliquen l'estudi sobre 'L'economia oberta amb dolarització parcial'. Al ser un estudi econòmic hi ha molt vocabulari especifíc que estarà explicat a peu de pàgina per fer-ho més entenedor.

L'objectiu no és fer una réplica exacta d'aquest document, si no, a partir de la informació donada, poder descriure aquest informe d'una manera clara i entenedora fent èmfasis en la part Bayesiana vista en el treball. Totes les taules necessàries estan a l'Annex II.

Aquest informe[1] ha sigut cedit pel departament de Econometría de la Facultat d'Economía de la UB.

6.2 Motivació del problema

Aquest estudi explica un model macroeconòmic simple[2] per a una petita economia oberta amb dolarizació parcial al Perú en la tradició de la síntesi neokeynesiana[3]. Aquest informe té molts punts en comú amb el Model de Projecció Trimestral (MPT) del Banc Central de Reserva del Perú (BCRP) fet anteriorment, ja que és una versió adaptada d'aquest, per la qual cosa al llarg del treball es fan diverses referències comparatives. Per aquest motiu, la font principal de les dades és el BCRP.

Un dels problemes dels països emergents econòmicament és la falta de dades econòmiques fiables. Aquest és el primer motiu pel qual s'ha decidit utilitzar els mètodes Bayesians, ja que permeten extreure conclusions a partir d'analitzar sèries a curt termini. Un altre dels motius de la utilització d'aquests mètodes, és que no depèn de la grandària de la mostra per poder utilitzar de forma eficient tota la informació *a priori*, que com s'ha repetit al llarg de tot el treball, és la informació sobre els paràmetres que s'obté abans d'analitzar les dades.

[1] Disponible en www.bcrp.gob.pe/publicaciones/revista-estudios-economicos/estudios-economicos-no-22.html

[2] Models macroeconòmics d'equilibri general amb expectatives racionals que consisteixen en un conjunt d'equacions de comportament que no tenen fundaments explícits a nivell microeconòmic, però que es sustenten amb solidesa en termes d'interpretació econòmica.

[3] La macroeconomia del neokeynesianisme està orientada en la discussió de la política monetària amb metes d'inflació. El model bàsic consta de tres equacions estudiades en aquest treball.

Al principi de l'informe hi ha una breu explicació del context econòmic del Perú d'aquests anys, necessària per poder entendre les conclusions extretes dels gràfics que es veuen a continuació. Per aquest motiu, totes les conclusions que es veuen en aquest Capítol estan exposades en l'informe que estem analitzant.

6.3 Estimació del model

Per poder realitzar l'estudi s'utilitzen dades trimestrals de la economía peruana desde el primer trimestre de l'any 2000 fins al tercer trimestre del 2008.

És un model a curt termini on les variables estan expressades en termes de bretxes, és a dir, com a desviacions dels seus valors d'equilibri o a llarg termini. Les variables d'equilibri són exògenes[4], independents i segueixen un procés autoregressiu de primer ordre[5].

Dividim el model en quatre equacions de comportament explicades a continuació cadascuna per separat amb les seves respectives conclusions. Els següents subapartats corresponen a cadascuna d'aquestes equacions anomenades: Demanda ag, Oferta agregada o corba de Philips, Regla de política monetària i Paritat descoberta de les taxes d'interès. Com veurem a continuació, el model està estimat per 14 variables observables[6] i conté 28 paràmetres per estimar.

A partir del criteri dels qui van modelar inicialment el MPT es va delimitar els rangs de la majoria de coeficients *a priori*, tal que es redueix la càrrega computacional del problema d'estimació. Per això, s'utilitzen majoritàriament distribucions Beta per a coeficients fitats, distribucions Gamma per a coeficients positius i distribucions *a priori* Gamma-Inverses per a desviacions estàndards de les pertorbacions[7] per garantir que siguin estrictament positives.

Per observar les estimacions *a posteriori* s'utiltiza per al seu càlcul l'algoritme de Metropolis-Hastings, on es va ajustar la variància per tenir una taxa d'acceptació del 20% - 30% aproximadament. A parir d'unes 50.000 repeticions es van extreure els resultats de la distribució *a posteriori*. Normalment seria un número relativament baix, peró s'ha de tenir en compte que és una versió simplificada, de manera que permet realitzar una recerca exhaustiva de bons valors per als paràmetres *a prioir*.

Els resultats d'aquesta estimació són la distribució, la mitjana i la desviació estàndard *a priori* i la moda i els valors dels percentils 5 i 95 de la distribució *a posteriori*, mostrats en quadres i gràfics, diferenciat per les diferents equacions d'explicació del model. L'estudi no presenta els resultats d'alguns paràmetres de menor rellevància.

[4]Una variable exògena o explicativa és una variable independent que afecta al model des de l'exterior.

[5]Un procés autoregressiu és un model de regressió on les variables explicatives/exógenes són la mateixa variable dependent/endógena retardada. Si és de primer ordre aquesta només està retardada una vegada.

[6]Una dada que cal puntualitzar alhora d'analitzar i treure conclusions és que es van transformar les taxes trimestrals d'inflació en taxes anuals.

[7]La pertorbació és la desviació pel que fa a la mitjana. Substitueix o representa a les variables exògenes omeses que poden afectar a la variable endògena per que no estan incloses en el model.

6.3.1 Demanda agregada

Descriu la dinàmica de la bretxa del producte y_t. El model és de la forma:

$$\begin{aligned} y_t = a_y y_{t-1} + a_{re} y_{t+1} - a_{rmc}(\beta_r r_{t-1} + \beta_{rs} r^{\$}_{t-1}) + a_{tot}[\gamma tot_t + (1-\gamma) tot_{t-1}] + a_q q_t \\ + a_{fis} fis_t + a_{y^*} y^*_{t-1} + \epsilon^y_t \end{aligned}$$

on:

- y_{t-1} i y_{t+1} són esdeveniments passats i futurs, respectivament.
- r_t i $r^{\$}_t$ són les taxes d'interès reals de llarg termini en moneda nacional i moneda estrangera, respectivament. Estan retardades i afecten a y_t a través d'un coeficient comú, a_{rmc}. Tot i així, cadascuna afecta per separat amb un paràmetre que no té perque tenir el mateix valor.
- tot_t és el preu de les exportacions en relació al preu de les importacions (terme d'intercanvi).
- q_t és el tipus de canvi real.
- y_{t-1} representa un mesurament explícit de la demanda externa en la forma d'un retard de la bretxa de la mitjana ponderada del producte dels socis comercials.
- fis_t representa el paper de la política fiscal mitjançant la primera diferència del balanç estructural.
- ϵ^y_t és el terme de pertorbació.

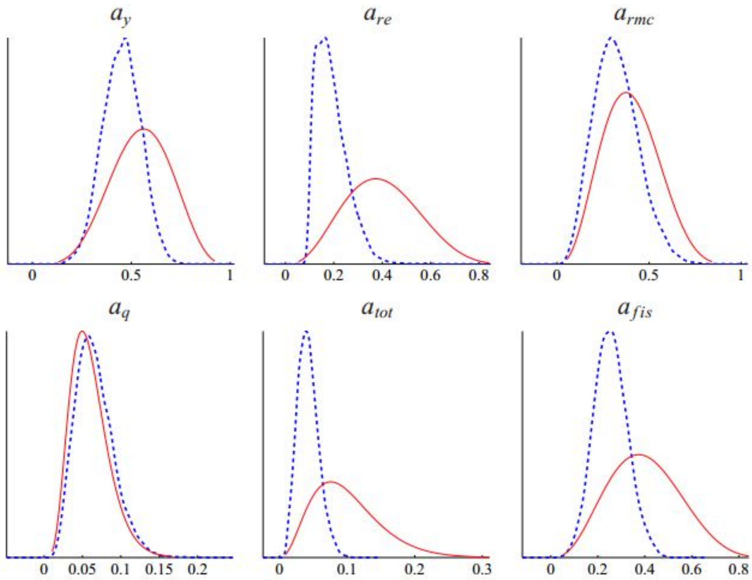
En la Taula 6 tenim els resultats de l'inferència Bayesiana per a cadascun dels paràmetres estudiats, que acompanyen a les variables de la primera equació explicades abans.

A partir d'aquests resultats construïm un gràfic per a cada paràmetre, corresponent a la Figura 1. D'aquesta manera és mes senzill analitzar el comportament de les variables a partir del paràmetres.

A partir d'analitzar tant la Taula 6 com la Figura 1, una de les primeres conclusions és que la moda *a posteriori* del coeficient d'inèrcia (a_y) és més gran que la del component prospectiu (a_{re}). Observem que el pes en la bretxa del producte anticipat és més gran que zero, a diferència del que ocorre en MPT en la qual aquest terme no ha sigut considerat. A més, en vista dels valors que s'assumeixen per a β_r i β_{rs}, així amb una moda *a posteriori* del coeficient a_{rmc}, el pes en la bretxa de la taxa d'interès real en moneda nacional (r_t) és aproximadament del 8%, mentre que el pes del seu parell en moneda estrangera ($r^{\$}_t$) és del 4%.

Només la variància *a posteriori* del coeficient de la bretxa dels termes d'intercanvi és considerablement més petita que en la seva variància *a priori*, que implica que les dades donen informació significativa sobre aquest paràmetre, però no sobre els coeficients relacionats amb la bretxa del producte extern (tot_t) i amb la bretxa el tipus de canvi real (q_t). Per tant, els valores de les modes *a posteriori* d'aquest tres paràmetres (a_{tot}, γ, a_q) es troben en un rang d'entre 0.04 i 0.08. Finalment, el valor estimat de la moda de la variable fiscal (a_{fis}) és bastant elevat.

Figura 1: Distribucions *a priori* i *a posteriori*



Les línies contínues/discontínues corresponen a les distribucions a priori/posteriori.

6.3.2 Oferta agregada o corba de Philips

Determina la inflació subjacent[8] π_t^c. El model és de la forma:

$$\pi_t^c = b_{p^*}\pi_t^m + (1 - b_{p^*})[b_{p^*}\pi_{t-1}^c + (1 - b_p)\pi_{t+1}] + b_y y_{t-1} + \epsilon_t^\pi$$

on:

- π_t^m és la inlfació importada. Aquesta segueix el model

  $$\pi_t^m = c_p\pi_{t-1}^m + c_{pf}(4s_t + \pi_t^*) + (1 - c_p - c_{pf})(4s_{t-1} + \pi_{t-1}^{rm}) + \epsilon_t^m$$

  que depèn de la seva evolució passada (π_{t-1}^m), de la inflació externa (π_t^*) i el retard de la inflació de les matèries primes i béns intermedis importats (π_{t-1}^{rm}). Aquestes dues últimes estan expressades en moneda nacional.

- π_{t-1}^c i π_{t+1} són un component inercial i un component d'expectatives, respectivament.

- y_{t-1} és la bretxa del producte, de l'equació anterior, amb un retard.

[8] La inflació subjacent és un indicador que mostra la variabilitat dels preus de consum a curt termini.

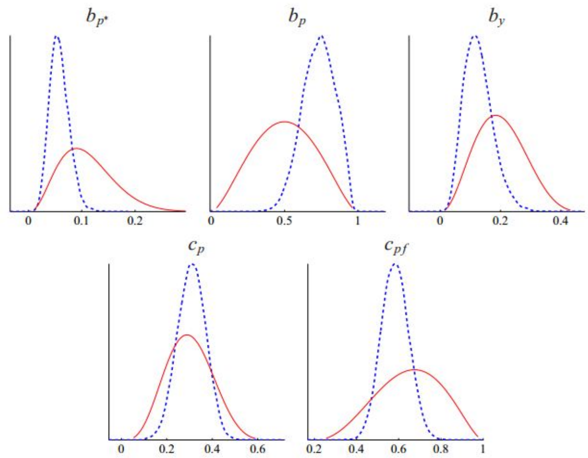
- ϵ_t^m és el terme de pertorbació.

En la Taula 7 tenim els resultats de l'inferència Bayesiana per a cadascun dels paràmetres estudiats, que acompanyen a les variables de la segona equació del model conjunt de l'estudi explicades abans.

Per un altre part, tenim els resultats de l'equació que explica la inflació importada (π_t^m) (Taula 8).

A partir d'aquests resultats construïm un gràfic per a cada paràmetre, corresponent a la Figura 2. D'aquesta manera és mes senzill analitzar el comportament de les variables a partir del paràmetres.

Figura 2: Distribucions *a priori* i *a posteriori*



Les línies contínues/discontínues corresponen a les distribucions a priori/posteriori.

A partir d'analitzar tant les taules 7 i 8 com la Figura 2, una de les primeres conclusions és que igual que passava en el cas de l'equació de la demanda agregada, la moda *a posteriori* del coeficient retrospectiu és també més gran que el del component d'expectatives. A més, el coeficient *a posteriori* de la bretxa del producte (b_y) és de 0.10, inferior al calibratge original. La inflació importada (π_t^m) també és rellevant per determinar la dinàmica de la inflació subjacent (π_t^c), amb una moda *a posteriori* de b_{p^*} propera al 5%.

6.3.3 Regla de política monetària

L'equació descriu una regla de Taylor que defineix la taxa d'interès a curt termini (i_t). El model és de la forma:

$$i_t = f_i i_{t-1} + (1 - f_i)[\bar{l}_t + f_p(\pi_{4,t+4}^c - \bar{\pi}) + f_y y_t] + \epsilon_t^i$$

on:

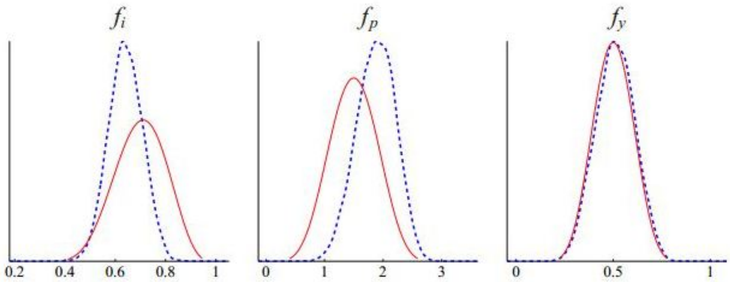
- i_t és l'instrument de política monetària en l'equació i i_{t-1} és el seu primer retard.

- l'equació és una funció de la desviació de la inlfació anual esperada (en els següents 4 trimestres) respecte a la meta d'inlfació $\pi^c_{4,t+4} - \bar{\pi}$ i de la bretxa del producte corrent y_t.
- $\bar{l_t}$ és el nivell d'equilibri de la taxa d'interès, que es crea a llarg termini, quan la bretxa del producte i la desviació de la inflació respecte de la meta són igual a 0.
- ϵ^i_t és el terme de pertorbació.

En la Taula 9 tenim els resultats de l'inferència Bayesiana per a cadascun dels paràmetres estudiats, que acompanyen a les variables de la tercera equació del model conjunt de l'estudi explicades abans.

A partir d'aquests resultats construïm un gràfic per a cada paràmetre, corresponent a la Figura 3. D'aquesta manera és mes senzill analitzar el comportament de les variables a partir dels paràmetres.

Figura 3: Distribucions *a priori* i *a posteriori*



Les línies contínues/discontínues corresponen a les distribucions a priori/posteriori.

A partir d'analitzar tant la Taula 9 com la Figura 3, una de les primeres conclusions és que els coeficients *a posteriori* són consistents amb l'evidència general d'altres països. D'altra banda, com que la variabilitat *a priori* i *a posteriori* del coeficient de la bretxa del producte (f_y) són gairebé les mateixes, implica que les dades són insuficients per identificar aquest paràmetre.

6.3.4 Paritat descoberta de les taxes d'interès

EL tipus de canvi nominal està definit per la condició de paritat de la taxa d'interès $(4(s^e_{t+1} - s_t) = i_t - i^*_t - rp_t + \epsilon^s_t)$. Per tant, podem explicar la quarta equació a partir de les expectatives de tipus de canvi (s^i_{t+1}). De manera que l'última equació queda de la forma

$$s^e_{t+1} = \rho s_{t-1} + (1-\rho)s_{t+1} + \epsilon^e_t$$

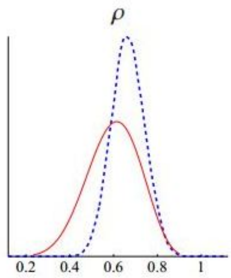
on

- l'equació està definida com la mitjana ponderada d'un component retrospectiu (s_{t-1}) i un de prospectiu (s_{t+1}).
- ϵ_t^e és el terme de pertorbació.

En Taula 10 tenim els resultats de l'inferència Bayesiana per a cadascun dels paràmetres estudiats, que acompanyen a les variables de la quarta equació del model conjunt de l'estudi explicades abans.

A partir d'aquests resultats construïm un gràfic per al paràmetres, corresponent a la Figura 4. D'aquesta manera és mes senzill analitzar el comportament de les variables a partir del paràmetres.

Figura 4: Distribucions *a priori* i *a posteriori*



Les línies contínues/discontínues corresponen a les distribucions a priori/posteriori.

A partir d'analitzar tant la Taula 10 com la Figura 4, una de les primeres conclusions és que els termes adaptatius i anticipats són rellevants per explicar el tipus de canvi esperat. A més, la moda *a posteriori* és més gran que el component retrospectiu ($\rho > 0.5$). Això podria reflectir el paper que exerceixen les intervencions canviaries en atenuar la volatilitat[9] del tipus de canvi.

6.3.5 Conclusió conjunta

Una vegada hem analitzat i hem extret conclusions de totes les taules corresponents a cadascuna de les equacions que conjuntament expliquen el model, podem dir que, totes les taules presenten resultats en relació als coeficients autoregressius (ρ^j) i volatilitat de les principals pertorbacions o xocs estructurals vinculats amb els blocs de les equacions centrals ($SD_{\epsilon_t^j}$). A partir dels coeficients *a posteriori* s'observa, sobretot en l'equació de la demanda agregada, una mica d'inèrcia en els xocs. A més, la desviació estàndar dels xocs, en la primera i segona equació, és pràcticament del mateix tamany, en l'úlima és significativament major, i per últim en la tercera equació aquest valor també es molt gran, però podria ser per l'alta volatilitat de la taxa d'interés a curt termini en els primers dos anys del període mostral.

[9] La volatilitat és la variabilitat de la rendibilitat d'una acció respecte a la seva mitjana en un període de temps determinat.

Amb aquests resultats també es pot comprovar si els criteris triats pels tècnics i responsables d'aquest estudi amb metodologia Bayesiana es corresponen a la formulació de polítiques del BCRP en funció dels paràmetres definits en el MPT original. Amb això, podem dir que hi ha una evidència empírica massa escassa per implementar aquests contrastos. De manera, que les conclusions obtingudes són mixtes i observables en la Taula 11.

Podem veure que els valors *a posteriori* d'alguna paràmetres s'aproximen bastant als originals, de manera que mostra el bon criteri dels experts del banc central. També podem apreciar que els mètodes Bayesians suggereixen que alguns paràmetres del MPT s'haurien de canviar perquè fossin més consistents amb les dades, com per exemple, a_{re}, a_{tot} i a_{fis} de l'equació de demanda agregada. En general, podem dir que la majoria d'aquests ajustos són bastant moderats.

6.4 Conclusions

En aquest apartat veurem un conjunt de comentaris relacionats amb la inferència Bayesiana, extrets d'aquest informe, on s'explica per quins motius s'ha escollit aquesta manera de treballar i els avantatges i inconvenients que comporta. Per tant, podem veure l'opinió real d'experts.

Els mètodes Bayesians ens permeten poder estimar un gran nombre de paràmetres. Conseqüentment, ens permet utilitzar de forma eficient la informació existent en les dades i la informació *a priori*, per poder treballar directament d'aquesta informació. Però, tenir que treballar amb informació *a priori* implica una major investigació per a formar evidències empíriques significatives. En aquest cas, al ser una verisió del MPT, algunes distribucions *a priori* són similars als valors del treball original i no han tingut aquest treball darrere d'investigació. Si s'assignen distribucions *a priori* relativament difuses[10], les dades tenen un paper important en la determinació de les distribucions *a posteriori*.

Com ja hem dit al principi, un dels motius fonamentals per als quals s'han utilitzat aquests mètodes,és la falta de dades. Això implica que ens molts casos les distribucions *a priori* són iguals a les distribucions *a posteriori*. És a dir, el fet de la limitació de dades comporta que la inferèrencia sobre aquests coeficients depengui en gran mesura de l'informació *a priori*.

Fins ara només ens hem centrat en la utilització de la simulació Bayesiana per l'estimació de paràmetres, però té moltes més aplicacions. Una altra utilitat que se li dóna en aquest treball és l'extracció de variables latents[11], com són la bretxa del producte. A partir d'això podem saber si es possible estendre el model per estimar altres variables no observades.

Una altra conclusió important és la necessitat de la realització d'un estudi més exhaustiu per aquells paràmetres on la distribució *a posteriori* té els mateixos valors que la distribució *a priori*. En aquests casos, la solució més emprada és la

[10]Distribucions amb variància relativament alta.

[11]Una variable latent és una variable que no s'inclou entre les variables estudiades i que, no obstant això, té un important efecte sobre la relació que existeix entre elles.

introducció de nous elements al model i evaluar la seva millora. Per exemple, els coeficients del tipus de canvi real i la bretxa del producte extern en l'equació de la demanda agregada compleixen aquesta condició, de manera que podríem augmentar la grandària de l'equació amb els vincles real-financer, com a nova variable.

7 Simulació d'inferènica Bayesiana amb R

En aquest Capítol simularem una inferència Bayesiana a partir de les distribucions més conegudes i fàcils d'utilitzar. Explicarem un exemple relacionat amb l'exemple explicatiu que ens ha acompanyat durant tot el treball. Abans de tot, comentar que són dades falses amb l'únic objectiu de poder treballar el màxim de conceptes treballats anteriorment i podem experimentar els avantatges i inconvenients de l'estadística Bayesiana.

Suposem que ens trobem en el cas de la Universitat Pública. Ara ens preocupa la baixa assistència a les aules durant tot el curs. Començarem l'estudi fent inferència a una classe estàndard de 90 alumnes, i voldrem estendre aquest resultat a tots els alumnes d'una facultat, on la mostra d'alumnes augmentaria fins a 1560 alumnes.

Comencem primer de tot amb la informació *a priori* que ens poden aportar els experts, que en aquest cas són els professors, dient que la millor manera d'aprovar qualsevol assignatura és assistir a totes les classes. Com hem dit, comencem amb una mostra petita de 90 estudiants on θ denota la proporció d'alumnes que assisteixen a totes les classes d'una assignatura en concret, i sobre la qual volem fer inferència ja que és un paràmetre desconegut. Els mètodes Bayesians representen les creences i les opinions inicials dels experts sobre la proporció a partir d'una distribució de probabilitat sobre aquest paràmetre. Per poder construir aquesta distribució *a priori* els professors recullen les notes dels exàmens de tots els alumnes dels 7 anys anteriors, que va ser quan va començar el grau, i treuen les primeres conclusions.

Denotem a la densitat *a priori* com $p(\theta)$. Si un èxit es considera anar durant tot el curs cada dia a classe, denotem e els èxits i f els fracassos. És a dir, la nostra probabilitat és una beta. Amb això podem construir la funció de versemblança tal que: $L_{X=x}(\theta) = \theta^e(1-\theta)^f$, on $0 < \theta < 1$. Tal i com hem vist en altres Capítols, podem obtenir la densitat a posteriori com $p(\theta \mid x) \propto L_{X=x}(\theta)p(\theta)$.

Ara calcularem la distribució *a priori* i *a posteriori* a partir de tres mètodes bàsics utilitzant l'eina R.

Una primera manera molt simple de calcular la probabilitat *a priori* és a partir d'una llista de possibles valors de la proporció θ amb els seus corresponents pesos a partir de les conclusions dels experts. A l'Annex III tenim el script de la Figura 12 de l'Annex III on s'expliquen els càlculs. Ens donen un resultat plasmat en la Figura 5. Els experts han extret com a conclusió a partir de la mitjana que en una classe de 90, 63 van a classe diàriament i 27 no. De manera que la nostra funció de versemblança queda com: $L_{X=x}(\theta) = \theta^{63}(1-\theta)^{27}$. A partir del script de la Figura 13 de l'Annex III podem treure la Taula amb les probabilitats *a priori* i *a posteriori.* Amb aquest valor podem obtenir l'histograma de les probabilitats *a posteriori* que correspon a la Figura 6. En aquesta podem observar que els valors de $\theta = \{0.65, 0.70, 0.75\}$ tenen el valor de la probabilitat *a posteriori* més alta.

La segona manera és a partir de la densitat *a priori*, que com hem vist, és una densitat beta. Donat que la proporció θ és un paràmetre continu, construïm una densitat *a priori* $p(\theta)$ en l'interval (0,1) que representa les idees prèvies. Suposem

Figura 5: Probabilitats *a priori*

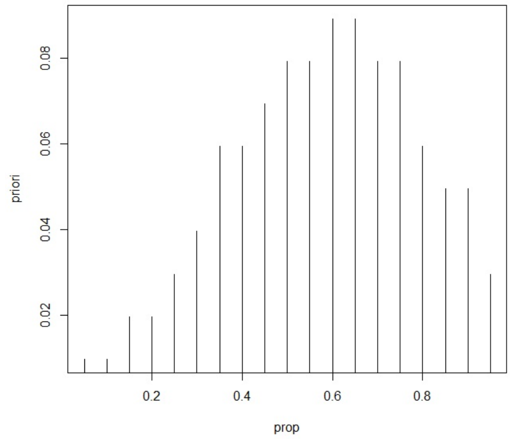
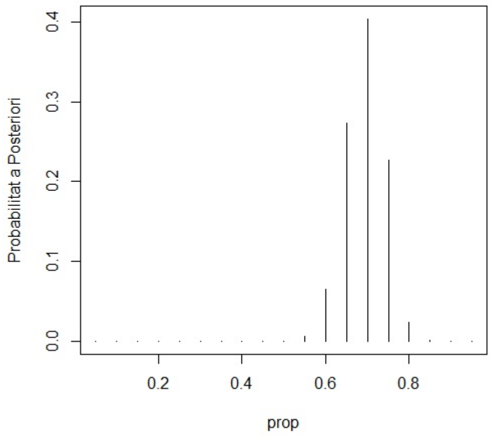


Figura 6: Probabilitats *a posteriori*



que els professors creuen que la proporció θ pot ser aproximadament més gran o més petita que $\theta = 0.70$, i a més, estan confiats en el fet que θ és menor que 0.90. Per tant, una família de densitats *a priori* per a la proporció és una distribució beta proporcional:

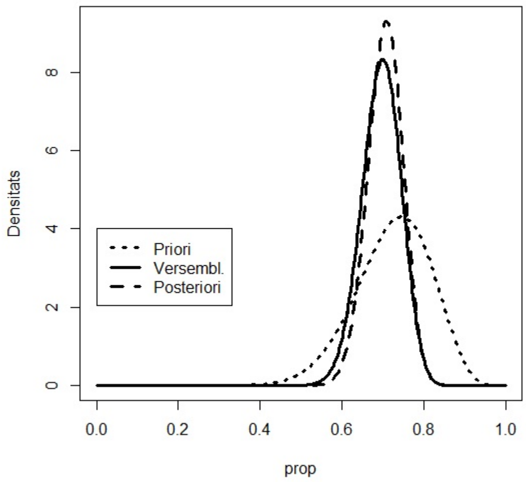
$$p(\theta) \propto \theta^{a-1}(1-\theta)^{b-1} \quad on \ 0 < \theta < 1.$$

Els paràmetres a i b representen les creences prèvies dels experts sobre θ, que es troben a partir del mètode de prova i error. Els resultats i proves fetes per trobar aquests valors no les contemplen en aquest treball, ja que consisteix en anar provant valors fins que tenim que la densitat beta corresponent a la densitat *a priori* té valors més alts per als valors més alts de θ decidits pels experts prèviament. Si combinem aquesta densitat amb la funció de versemblança, tenim que la densitat *a posteriori* també és una densitat beta amb paràmetres $a + e$ i $b + f$, tal que:

$$p(\theta \mid x) \propto \theta^{a+e-1}(1-\theta)^{b+f-1} \quad on \ 0 < \theta < 1.$$

A partir del script de la Figura 14 de l'Annex III calculem les densitats i la funció de versemblança, tenint en compte que les tres són densitats beta, i ho plasmem de manera gràfica en la Figura 7. Aquest gràfic ens ajuda a veure que la densitat *a posteriori* en aquest cas és una combinació entre la densitat *a priori* i la funció de versemblança, encara que s'assembla molt més a aquesta última. Un motiu pot ser perquè la informació *a priori* ha sigut seleccionada sense molt de criteri.

Figura 7:



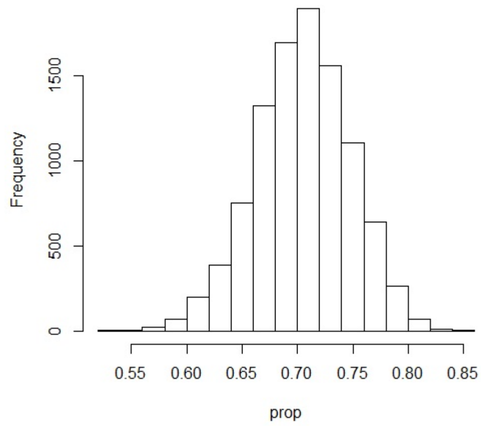
A partir de la distribució *a posteriori* podem treure moltes més conclusions. Amb diferents comandes de R, podem resoldre dubtes com: podria ser que la proporció d'estudiants que assisteixen a classe diàriament fos inferior a 0,50? És a dir, calculem a partir del script de la Figura 15 de l'Annex III la probabilitat *a posteriori*: $p(\theta \leq 0.50 \mid x) = 4.661643e - 06$. En el mateix script trobem la solució d'una altra pregunta: $p(\theta \geq 0.90 \mid x) = 5.246898e - 09$. Per tant, les probabilitats tenen valors tan baixos que és poc probable que θ sigui inferior a 0.50 o superior a 0.90. A més a més, volem una bona estimació d'un interval de valors de θ. A partir del script de la Figura 15 de l'Annex III podem dir que la proporció θ tindrà valor dins d'aquest interval [0.6323986, 0.7730294].

Podem dir que fins aquí les dades són exactes, ja que utilitzem funcions de R per a densitats beta. Hi ha una manera alternativa de trobar valors de la densitat *a posteriori* a partir de la simulació. En aquest cas, simulem un gran nombre de valors per a la distribució *a posteriori* beta i sintetitzem la sortida simulada. Aquest resultat és pot veure a la Figura 8. Ara igual que abans, a partir de la distribució *a posteriori* volem resoldre les mateixes preguntes per poder comparar resultats. En aquest cas tenim a partir del script de la Figura 16 de l'Annex III que

$$p(\theta \leq 0.50 \mid x) = 1e - 04 \quad i \quad p(\theta \geq 0.90 \mid x) = 0,$$

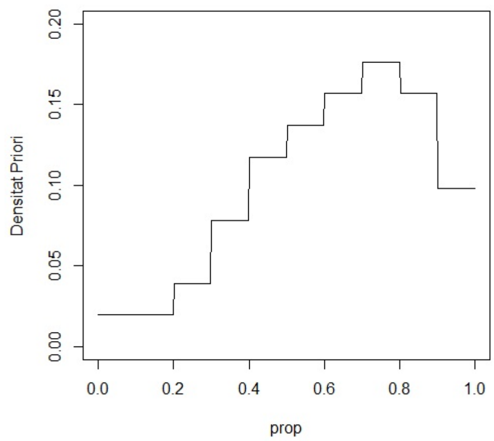
i l'interval correspon a [0.6327473, 0.7731103]. Podem treure com a conclusió que els valors de les densitats *a posteriori*, amb les dues maneres diferents de treballar, donen valors molts similars.

Figura 8:



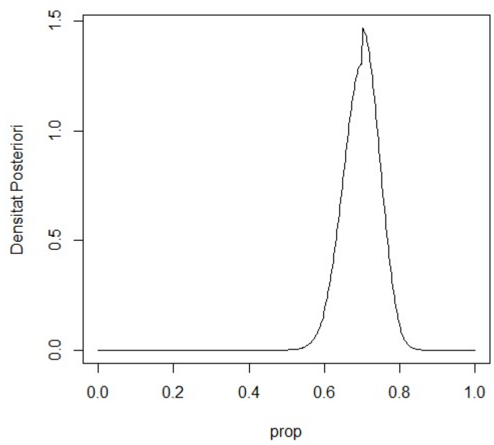
I per l'últim utilitzarem un mètode de 'força bruta' que ens permetrà fer els càlculs posteriors per a una densitat *a priori* arbitrària $p(\theta)$, ja que sabem que hi ha molts avantatges computacionals amb l'ús d'una densitat *a priori*. Explicarem aquest algoritme a partir del càlcul d'un histograma de la densitat *a priori*, que pugui reflectir millor l'opinió prèvia dels experts sobre la proporció θ. Comencem dividint el rang de θ en 10 subintervals (0,0.10), (0.10 , 0.20), ..., (0.90, 1) i assignant una probabilitat a cadascun. En aquest cas, els experts donen un pes similar al del primer mètode que hem utilitzat. L'histograma resultant és el de la Figura 9. Multiplicant aquest histograma per la funció de versemblança obtenim la densitat *a posteriori*, que gràficament correspon a la Figura 10. Aquests càlculs es troben a la Figura 17 de l'Annex III.

Figura 9: Densitat *a priori*



Fins aquí l'únic que hem fet ha sigut intentar trobar el valor de la proporció d'alumnes que van a classe diàriament només centrant-nos en una classe. L'altre

Figura 10: Densitat *a posteriori*

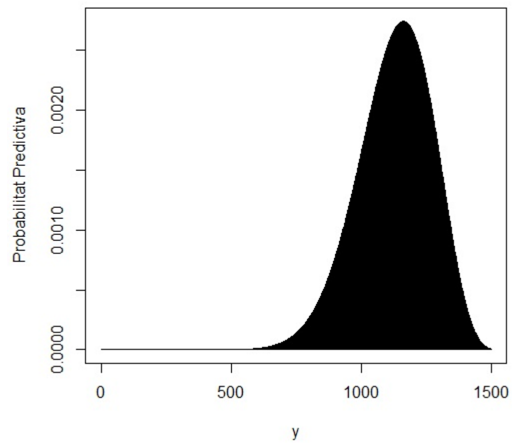


objectiu d'aquest estudi era predir quants alumnes de tota la facultat no anirien a classe. Tal i com hem fet en Capítols anteriors, denotem aquest número que volem predir com $\tilde{x}$. Ara la nostra mostra ha augmentat fins a tenir 1560 alumnes, dels quals voldríem tenir almenys 1500 alumnes que assistissin diàriament a classe. Amb les probabilitats predictives que s'observen al script de la Figura 18 de l'Annex III podem concloure que el número més probable d'èxists en aquesta mostra futura és qualsevol valor de $\tilde{x}$ que pertanyi a l'interval [1469, 1496]. Suposem que els experts li donen forma a les seves creences sobre θ utilitzant una densitat *a priori* beta (a,b), on els valors dels paràmetres són els mateixos que abans. Aleshores, calcularem les probabilitats predictives utilitzant la densitat beta a partir de la funció 'pbetap' de R. A partir del script de la Figura 19 de l'Annex III podem veure que el número més probable d'èxists en aquesta mostra futura és qualsevol valor de $\tilde{x}$ que pertanyi a l'interval [1139, 1188]. De manera gràfica podem veure aquests resultats en la Figura 11.

Hem calculat les probabilitats predictives a partir de dos mètodes d'obtenció de la densitat *a priori*. De la mateixa manera que abans, una altra manera de calcular una densitat predictiva és a partir de la simulació per a qualsevol densitat *a priori*. Els càlculs serien molt semblants als fets anteriorment, i com ja hem vist, la solució és molt similar, per tant, en aquest cas no els farem i ens centrarem en els mètodes posteriors.

A continuació utilitzem el mètode de Monte Carlo per calcular la distribució *a posteriori*. Continuem amb el mateix exemple, on θ representa la proporció d'alumnes que van a classe cada dia, on la densitat *a posteriori* beta té paràmetres, $a + e = 16.5 + 63 = 79.5$ i $b + f = 6.3 + 27 = 33.3$. Per utilitzar aquest mètode suposem que estem interessats en saber quantes noies compleixen aquest requeriment. Suposem que en la classe tenim que el 63% d'alumnes són noies, és a dir, estem interessats en la mitjana *a posteriori* de $\theta^{56.7}$. Simulem la distribució *a posteriori* beta tal com es veu al script de la Figura 20 de l'Annex III. Obtenim com a conclusió que l'estimació de Monte Carlo és $E(\theta^{56.7} \mid x) = 1.486574e - 07$ amb un

Figura 11: Probabilitat Predictiva



error estàndard de simulació associat de 1.163329e-08.

Un altre mètode explicat al treball que serveix per simular la distribució *a posteriori* és el mètode de Monte Carlo amb Cadenes de Markov (MCMC). Com ja sabem, hi ha molts algoritmes que ens permeten fer aquest càlcul. Nosaltres a partir de l'eina R només utilitzarem el més conegut: Metropolis–Hastings.

Explicarem dues versions d'aquest l'algoritme, ja que són aplicables a una àmplia varietat de problemes d'inferència Bayesiana. Aquestes dues versions són: la cadena d'independència i la cadena del camí aleatori. L'algoritme de Metropolis–Hastings comença amb un valor inicial θ^0 i especifica una pauta per simular el t-èsim valor en la seqüència θ^t donat el valor de (t-1) en la seqüència θ^{t-1}. Aquesta pauta consisteix en una densitat proposada p que simula el valor del candidat θ^*, i en el càlcul d'una probabilitat d'acceptació P, que indica la probabilitat de què el següent valor candidat sigui acceptat com el següent valor de la seqüència. Diferenciem les dues versions en el moment d'escollir aquesta densitat p. L'algoritme de les cadenes independents, s'utilitza quan la densitat és independent del valor actual en la seqüència. És a dir,

$$p(\theta^* \mid \theta^{t-1}) = p(\theta^*).$$

En canvi, l'algoritme del camí aleatori defineix la densitat a partir d'una densitat h que és simètrica al voltant de l'origen. És a dir,

$$p(\theta^* \mid \theta^{t-1}) = h(\theta^* - \theta^{t-1}).$$

En R tenim dues funcions que ens permeten fer aquests càlculs juntament amb la llibreria LearnBayes: 'rwmetrop' i 'indepmetrop'.

Per a aquest algoritme, utilitzem l'exemple del fet que volem saber la mitjana i la desviació estàndard de les hores d'estudi dels alumnes d'una Universitat dues setmanes abans dels exàmens finals. Agafarem una mostra aleatòria de 211 alumnes i observarem les dades de forma agrupada juntament amb la seva freqüència. Escrivim aquestes dades en la Taula 5. Amb aquest exemple estem estudiant dades

Taula 5:

Intervals en hores	Freqüència
menys de 66	14
entre 66 i 68	30
entre 68 i 70	49
entre 70 i 72	70
entre 72 i 74	33
més de 74	15

multinomials (generalització d'una distribució binomial) amb probabilitats desconegudes $p_1, ..., p6$ de la forma, $p_2 = \phi(68, \mu, \sigma) - \phi(66, \mu, \sigma)$, on μ és la mitjana, σ la desviació estàndard i $\phi(\cdot, \mu, \sigma)$ la funció de distribució acumulativa d'una variable aleatòria normal (μ, σ). Ens trobem davant d'un cas de distribució *a priori* no informativa, de manera que suposem que és proporcional a $1/\sigma$. Aleshores la densitat *a posteriori* dels paràmetres és

$$p(\mu, \sigma \mid x) \propto \frac{1}{\sigma} L_{X=x}(\mu, \sigma),$$

on $L_{X=x}(\mu, \sigma) = \phi(66, \mu, \sigma)^{14}(\phi(68, \mu, \sigma) - \phi(66, \mu, \sigma))^{30}...(1 - \phi(74, \mu, \sigma))^{15}$ és la funció de versemblança. De forma equivalent,

$$p(\mu, \sigma \mid x) \propto L_{X=x}(\mu, e^{\lambda})$$

si transformem la desviació estàndard positiva per $\lambda = log(\sigma)$.

Amb aquestes premisses podríem utilitzar la versió del camí aleatori que correspon a la funció $rwmetrop(logctablepost, proposal, start, m, data)$, on *logctablepost* és la funció que defineix la densitat *a posterior* logarítmica, *proposal* consisteix en una matriu de variàncies i covariàncies estimades i una escala (el factor d'escala de Metropolis–Hastings), *start* és el vector que conté el valor inicial del paràmetre, m és el número d'iteracions de la cadena i *data* són les dades que s'utilitzen en la funció *logctablepost*.

Aquests càlculs ja tenen un nivell de complexitat alt ja que requereixen uns coneixements avançats de R i gran maneig de les distribucions estàndards, traient així una millor conclusió.

8 Conclusions

Com hem pogut comprovar al llarg de tot el treball l'estadística Bayesiana és molt versàtil i ens permet fer un estudi estadístic sobre un tema en concret a partir d'un mostreig i possibles coneixements previs de manera subjectiva, aprofitant al màxim tota la informació existent.

Com ja hem vist, que els experts donin la seva opinió, creença, coneixements sobre l'esdeveniment que ens interessa, no sempre té avantatges, i això és un problema, ja que aquestes dades tenen molt de pes en resultats posteriors. Un dels inconvenients més comuns, que he pogut viure jo mateixa al proposar els exemples, és la falta d'informació *a priori* i la poca cura científica en el moment de donar-los a conèixer. Per evitar aquest problema, quan la mostra és suficientment gran, utilitzem l'estadística clàssica que dóna els mateixos resultats.

En la part d'inferència Bayesiana, hi ha una gran diferenciació amb l'estadística clàssica, ja que aquesta última es basa en els contrastos d'hipòtesis majoritàriament, i l'estadística Bayesiana amb la distribució *a posteriori.* Com que aquesta distribució ha sigut extreta a partir de coneixements previs (distribució *a priori*) i de tota la informació i coneixements nous (funció de versemblança) sobre el tema que estem estudiant, és millor la precisió de l'estimació. El problema que comporta, que ja hem vist anteriorment, és el seu càlcul complex. Aquest és un dels motius principals pels quals crec que fins ara no era molt coneguda, ja que fins fa poc no hi havia programes o simulacions computacionals que ens permetessin fer aquests càlculs per tenir una bona aproximació.

D'altra banda, hi ha molts aspectes interessants que només hem treballat de manera superficial, com poden ser, els mètodes de la inferència no paramètrica o tots els algoritmes i versions del mètode de Monte Carlo amb Cadenes de Markov. De tota manera, conèixer'ls ens ha aportat coneixements nous i hem pogut veure solucions i problemes que aporten els mètodes de simulació, tal com volíem des del principi.

Per tot aixó, podem concloure que l'existència i l'ús de l'estadística Bayesiana pot ser de gran utilitat, en diferents situacions, com per exemple, en aquells estudis on tenim una mostra estadística molt petita o quan els experts tenen informació rellevant sobre l'estudi. De tota manera, poden coexistir les dues estadístiques comparades en aquest treball, on cada una pot ser utilitzada en les parts de l'estudi en les quals aporti més utilitat.

9 Annex I: Inicis de l'estadísitca i propietats bàsiques

9.1 Definició

No existeix acord entre els autors sobre l'origen de la paraula estadística; segons uns prové del llatí status, estat o situació actual d'una cosa; segons uns altres de l'alemany *staat*, estat; alguns fixen la seva providència del llatí *statera*, balança, perquè mesura, pesant, els efectes, les causes i les forces socials. No obstant això si sembla cert que la paraula estadístic va ser introduïda per Godofredo Achenwall cap a 1748.

Bàsicament, en molts llibres surt una definició i un origen dâaquesta paraula comú basat en la seva utilitat al llarg de la història. La paraula Estadística procedeix del llatí *statisticum collegium* que vol dir 'consell d'Estat' i del seu derivat italià *statistica* amb significat 'home d'Estat' o 'polític'. Pot ser degut a que, una de les funciones principals dels Governs dels diferents Estats era establir registres de població, naixements, defuncions, impostos, collites, etc. Des dels inicis, les societats han sentit la necessitat de posseir dades xifrades sobre la població i les seves condicions materials d'existència. Per tant, podem considerar-la com la ciència, el mètode o la tècnica que permet estudiar numèricament, amb la màxima precisió, els fenòmens col·lectius incompletament coneguts.

Actualment el Diccionari de la Real Acadèmia Espanyola (RAE) ens indica etimològicament que la paraula estadística prové del italià *Statistica* amb significat 'home d'Estat'. La RAE ho defineix com:

Estudi de les dades quantitatives de la població, dels recursos naturals i industrials, del tràfic o de qualsevol altra manifestació de les societats humanes.

Branca de la matemàtica que utilitza grans conjunts de dades numèriques per obtenir inferències basades en el càlcul de probabilitats.

9.2 Historia

No es coneix exactament una data com a inici de l'Estadìstica. Des dels començaments de la civilització hi han hagut petites representacions gràfiques i altres símbols en pells, roques i parets de coves per comptabilitzar. Aproximadament a l'any 3000 a.C. els babilonis utilitzaven petits objectes d'argila per recopilar les dades sobre la producció agrícola i els diferents gèneres venuts i comprats. En els següents anys, tant els xinesos com els egipcis portaven una bona organització i administració dels moviments poblacionals i contínuament feien censos. A Grècia tenim els primers noms coneguts com Heredot i Aristòtil, els quals van incentivar la importància de les observacions estadístiques en el que es refereix a distribucions de terrenys o servei militar. A més a més, també s'efectuaven censos amb finalitats tributàries, socials i militars. Cap a l'any 500 a.C. a Roma es va crear la figura del censor, responsable de la producció periòdica dels censos ja que controlava el nombre d'habitants i les seves distribucions pels diversos territoris.

9.2.1 Estadística Subjectiva

L'església també va tenir una gran importància en el desenvolupant de l'Estadística a partir de l'any 762, amb la voluntat de Carlomango de voler tenir un registre de totes les seves propietats inclosos els béns d'aquesta institució religiosa. Tant va ser així que al voltant de l'any 1563, va establir l'obligació de la inscripció dels naixements, matrimonis i defuncions. Tot i així, la primera persona qui porta un recompte e intenta analitzar les causes de les defuncions va ser el botiguer anglès John Graunt (1620-1674). Al 1662 va publicar un tractat titulat *Observations on the London Bills of Mortality*, on va posa de manifest les xifres de naixements i defuncions ocorregudes a Londres durant els anys 1604 al 1661 i les influències que exercien les causes naturals, socials i polítiques sobre aquests esdeveniments. És considerat el primer treball estadístic seriós sobre la població. Per aquest motiu i per ser el primer home en extreure conclusions per a un col·lectiu a partir de l'estudi d'una mostra, John Graunt és conegut com fundador de l'estadística Moderna i de la Demografia.

El creador de l'Estadística Bayesiana, Thomas Bayes, canvia la visió de la probabilitat introduint el concepte de subjectivitat, és a dir, com canvien els judicis d'una persona sobre l'ocurrència d'un esdeveniment a la llum de nova informació o evidències. Bayes va estudiar teologia a la Universitat d'Edimbug i per això, el seu primer tractat publicat a l'any 1731 es titula *Divina Benevolència*, un intent de provar que la fi principal de la Divina Providència és la felicitat de les seves criatures. En 1734 el bisbe George Berkeley atacava al càlcul diferencial de Newton en un tractat analista. Argumentava que els infinitèsims no presenten una distinció clara entre zero i quantitats infinitament petites. Entre les rèpliques a Berkeley, apareix una signada per Bayes amb el pseudònim de John Noon. Aquestes son les dues úniques obres que va publicar Bayes. A la seva mort, la seva família envia a Price els documents sobre matemàtiques redactats per Bayes en vida. Determinava per primera vegada un interval de confiança Bayesià per al paràmetre θ d'una distribució de Bernoulli a partir de n repeticions de l'experiment de Bernoulli. També hi havia una aproximació a la integral de la funció beta però que donava lloc a un nombre massa alt de repeticions de l'experiment de Bernoulli per obtenir amb precisió l'interval sobre la probabilitat d'èxit. Tot i així, la formulació del problema era precisa i les matemàtiques emprades eren correctes.

En 1763 Bayes, tot i estar mort, va ser comunicat a la Royal Society per el seu treball titulat *un assaig cap a la solució d'un problema de la doctrina de probabilitats.*Prince va enviar el treball que començava enunciant el problema que diu: *Donat el nombre de vegades que un succés ha ocorregut i no s'ha presentat, calcular la probabilitat que la probabilitat que ser present en una sola repetició estigui compresa entre dos valors de probabilitat coneguts.* Actualment, traduiríem aquest enunciat com: tenim una mostra $(x_1, x_2, \ldots, x_n)$ de una població de Bernoulli de paràmetre θ desconegut i es pretén calcular la probabilitat condicionada

$$P\left\{a < \theta < b \mid \sum_{i=1}^{n} x_i = r\right\}.$$

Fins aquell moment, es sabia calcular la probabilitat de tenir r èxits en n repeticions si era coneguda la probabilitat d'éxit θ de la distribució de Bernoulli. Dit per Price, es tractava del *problema invers* al plantejat per De Moivre. Per resoldre-ho es requeria saber calcular la distribució a posteriori, mitjançant la versió contínua del Teorema de Bayes, i introduir una distribució a priori adequada per al paràmetre θ de la distribució de Bernoulli. Bayes va utilitza com a distribució a priori la uniforme en l'interval (0,1). Mitjançant el mètode geomètric, va anar desglossant les regles del càlcul de probabilitats a posteriori.

9.2.2 Estadística Objectiva

Al inici del segle XVII, un noble francès estava interessat en alguns jocs d'atzar que en aquells dies es jugaven en Monte Carlo. Va intentar descriure de forma matemàtica la proporció relativa de temps en què es podrien guanyar certes apostes i els hi va ensenyar a dos grans matemàtics de l'època que coneixia, Pascal i Fermat. Aquí va començar l'inici de la teoria de la probabilitat, amb el famós intercanvi de correspondència entre els dos matemàtics refent a l'aplicació correcta de les matemàtiques per poder calcular les freqüències relatives d'ocurrències en jocs senzills d'apostes.

Van continuar els avanços en la teoria de la probabilitat a partir del matemàtic rus A.N.Kolmogorov que va anunciar tres axiomes.

9.3 L'estadística avui dia

A partir del segle XIX, cada vegada la ciència matemàtica evolucionava de manera més empírica basada en la instrumentació dels mesuraments on la intuïció i la subjectivitat no tenien cabuda, desembocant en la interpretació Freqüentista de la probabilitat. Galton (1822-1911) i Pearson (1857-1936) es poden considerar els pares de l'estadística moderna, ja que gràcies a ells es va passar d'emprar l'estadística deductiva a la estadística inductiva. Més tard, Ronald Fisher (1890-1962) entre d'altres, assenten les bases d'un enfocament objectiu en l'Estadística, és a dir, per extreure conclusions d'un experiment es depèn exclusivament dels resultats i no requereix cap referència externa. Aquest nou enfocament, basat en la noció de versemblança i en la interpretació Freqüentista de la probabilitat, va resultar ser relativament fàcil d'aplicar en la pràctica, la qual cosa va motivar el seu desenvolupament durant tot el segle següent.

A partir del segle XX es va començar a denominar com l'Estadística moderna. La metodologia estadística s'inicia cap a la computació intensiva aplicades a grans masses de dades i es comença a considerar el mètode estadístic com un procés iteratiu de cercar el model ideal. L'aparició de l'ordinador digital, ens ajuda a poder treballar amb aquestes dades d'uns models més complexos de manera que ens permet tractar adequadament problemes dinàmics i multivariants.

Actualment sorgeixen diferents problemes alhora d'escollir un bon model estadístic, perquè ens ajudi a fer prediccions de les nostres dades i poder aproximarnos el màxim possible a la realitat. Hem de saber escollir bé des de un principi aquest model i saber amb claredat el que estem fent. Alguns d'aquests problemes més freqüents són:

- **Disseny d'experiments:** Dissenyar com recollir les dades i per tant, escollir un bon model estadístic M. Normalment, el model estadístic s'escull abans de recollir les dades. Definim un model estadístic com: $M = \{P(Y \mid \theta), \theta \in \Theta, Y \in \Omega\}$, on $P(Y \mid \theta)$ és el model de probabilitat per cada possible valor de $\theta \in \Theta$, Θ és l'espai de paràmetres i Ω és l'espai mostral, conjunt de tots els resultats possibles d'un experiment.

- **Identificar/Validar el model estadístic:** Una vegada hem recollit les dades Y , hem de decidir si el model de probabilitat que ha generat les dades $P(Y \mid \theta)$ es dins del model estadístic, és a dir, $P(Y \mid \theta) \in M$. Si és així, hem fet una bona elecció des del principi.

- **Inferència estadística:** Un cop arribat aquí, ja tenim les dades recollides i valida't el model. Ara es tracta d'endevinar el model de probabilitat $P(Y \mid \theta)$ que ha generat les dades, θ. En aquest punt, ja trobem diferencies entre les dues estadístiques més emprades alhora de calcular la inferència estadística. Aquest concepte és el conjunt de mètodes i tècniques que permeten induí, a partir de la informació empírica proporcionada per una mostra, com és el comportament d'una determinada població amb un risc d'error mesurable en termes de probabilitat. L'estadística Freqüentista només utilitza el model estadístic i les dades, i en canvi, l'estadística Bayesiana a part d'utilitzar aquesta informació també es basa en la informació a priori que es pugui tenir. La inferència principalment aborda els problemes de :

  - **Estimació puntual:**: a partir de les dades donar un paràmetre estimat que millor s'aproximi a θ,per tant, que el paràmetre estimat pertanyi a Ω.
  - **Estimació per intervals:** a partir de les dades donar una regió (interval) amb alta probabilitat o confiança en el cas freqüentísta de que θ hi pertanyi.
  - **Proves d'hipòtesis:** reduir l'espai de paràmetres, que és equivalent a triar un subconjunt de M $\Rightarrow M = \{P(Y \mid \theta) \in \Omega\} = \{P(Y \mid \theta) \in \Omega_0\} \cup \{P(Y \mid \theta) \in \Omega_1\} = M_o \cup M_1$.
  - **Predicció:** serveix per predir com seràn les dades futures.

- **Presentació dels resultats:** un cop tenim totes les noves dades calculades i les conclusions pertinents, el problema es presenta quan hem de compartir aquests resultats. La pràctica més habitual és a partir de gràfiques comparatives o de taules amb els valors directament. L'inconvenient moltes vegades és saber interpretar aquests resultats només amb la presentació.

9.4 Probabilitat

9.4.1 Definició

La probabilitat és un concepte que sorgeix amb la necessitat de mesurar quantitativament la certesa de que un succés donat ocorri. Com hem vist al Capítol anterior, originàriament es va començar mesurant l'ocurrència de successos sobre jocs d'atzar, però amb el pas del temps s'ha anat utilitzant en altres camps com la medicina, l'educació, la biologia, l'economia, entre d'altres. Per poder calcular la probabilitat es necessari abans de tot tenir molt clar quin és l'espai mostral de l'experiment denotat com Ω. L'espai mostral és el conjunt de tots els resultats possibles de l'experiment i, sobre aquest, es construeixen el conjunt de tots els successos, on es defineixen tres operacions: la unió, la intersecció i la complementarietat. El conjunt de successos, també anomenat parts de Ω i denotat com $\mathcal{F}$, amb aquestes tres operacions pot tenir una estructura diferent segons la mida de l'espai mostral. Si aquest és finit té estructura d'àlgebra de Booble, i si es infinit té estructura de sigma àlgebra, on les propietats que compleixen les operacions són les mateixes però esteses a la unió i la intersecció d'un nombre infinit numerable de successos. A més a més, li assignem a cada un d'aquests successos o esdeveniments una lletra majúscula. L'àlgebra de Booble i la sigma àlgebra verifiquen les següents propietats:

1. $\Omega \in \mathcal{F}$.
2. Si $A \in \mathcal{F}$ aleshores el seu complementari també : $A^c \in \mathcal{F}$.
3. Si un conjunt infinit i numerable d'esdeveniments, $\{A_m, m \geq 1\}$, esta contingut en el conjunt de successos, aleshores la seva unió també.

Amb el pas del temps s'ha anat modificant fins trobar la millor definició de probabilitat. Al principi, Laplace va donar una aproximació: Si A és un succés, la probabilitat es defineix com el quocient de casos favorables per nombre de casos possibles. És a dir,

$$P(A) = \frac{Número\ de\ casos\ favorables\ de\ A}{Número\ de\ casos\ possibles}.$$

Ràpidament es va veure que no servia com definició ja que només és vàlida per als casos equiprobables, cada experiment té la mateixa probabilitar d'èxit. A través d'un exemple és molt més fàcil veure que no serveix i ens donarà pas a poder introduir una definició més axiomàtica.

Exemple 9.1. Si considerem treure a l'atzar una carta de la baralla espanyola podem observar que els resultats possibles són per a cadascuna de les quatre famílies: 1,2,3,4,5,6,7,8,9,10,11 i 12. La probabilitat que surti una Figura, corresponent a les cartes numerades per 10, 11 i 12, és 12/48; la probabilitat que surti l'1 només d'una família és 1/48. Aquest és un cas bastant comú i tractat, però canviaria si es modifiquessin les hipòtesis inicials.

Ara la baralla espanyola esta formada per les quatres famílies on cada una té 12 cartes numerades de la següent forma: 1,2,3,4,5,6,7,7,7,10,11,12. Amb les cartes d'aquesta forma només hi ha 10 casos possibles i per tant, les probabilitats canvien. La probabilitat que surti una Figura, corresponent a les cartes numerades per 10, 11 i 12, és 12/40; la probabilitat que surti l'1 només d'una família és 1/40. Es pot observar que això no és correcte, ja que la baralla segueix tenint 48 cartes. En aquest cas la igualtat donada no funciona, encara que, sense allunyar-nos d'ella, podem assignar probabilitats als resultats d'aquesta nova baralla.

Un altre cas és tenir la primera baralla però trucada per algunes cartes, de manera que, hi ha 48 resultats possibles però no tots tenen la mateixa probabilitat d'ocórrer. Si no es disposa de més informació no es pot assignar una probabilitat a cada resultat, amb la única solució de treure repetidament cartes de la baralla i anar observant la freqüència de cada esdeveniment. En aquest cas, la igualtat tampoc es pot aplicar.

Per solucionar aquests problemes és va donar una alternativa. Es va dir que: la probabilitat d'un succés és el límit de la freqüència relativa del succés. És a dir,

$$\lim_{n\to\infty} \frac{n_A}{n}$$

Aquesta igualtat tampoc és una definició de probabilitat, perquè a la pràctica no es poden fer un nombre infinit de repeticions per trobar la freqüència de l'esdeveniment, sinó que es tenen aproximacions lo més exactes possibles de la veritable probabilitat del succés.

Per solucionar aquest problema de no tenir una definició de probabilitat correcte i tenir dues igualtats que només serveixen en casos puntuals, i per tant, no valen com una definició general, s'acostuma a treballar amb una definició axiomàtica.

Definició 9.2. *Un espai de probabilitat és un tern* $(\Omega, \mathcal{F}, P)$, *on:*

- Ω *és un conjunt, anomenat espai mostral, que correspon al dels resultats possibles de l'experiència aleatòria.*
- $\mathcal{F}$ *és una família de parts d'* Ω *que té estructura de* $\sigma-$*àlgebra.*
- P *denota la probabilitat i determina l'assignació de versemblança als esdeveniments. És una aplicació* $P : \mathcal{F} \longrightarrow [0,1]$ *que té les propietats següents:*
  1. $P(\Omega) = 1$.
  2. $(\sigma-$ *additivitat)* *Si* $\{A_n, n \geq 1\}$ *és una successió de conjunts de* $\mathcal{F}$ *disjunts dos a dos, aleshores*
  $$P\left(\bigcup_{n=1}^{\infty}\right) = \sum_{n=1}^{\infty} P(A_n).$$

A partir d'aquests axiomes es deriven moltes propietats que tota funció de probabilitat compleix. Les següents propietats només estan anunciades ja que és material de classe però necessàries per entedre conceptes posteriors.

Propietats:

- Sigui $\emptyset \in \mathcal{F}$ tenim que $P(\emptyset) = 0$ per qualsevol conjunt Ω.
- Sigui $A \in \mathcal{F}$ tenim que P(A)=1-$P(A)^c$, on A^c és el complementari de A respecte a Ω. Anàlogament, tenim que $P(A)^c =$ 1-P(A).
- Sigui $A \in \mathcal{F}$ tal que P(A)$\leq$ 1.
- Sigui $A \in \mathcal{F}$ i B $\in \mathcal{F}$ tenim que $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Si $A \cap B = \emptyset$ aleshores $P(A \cup B) = P(A) + P(B)$.

9.4.2 Probabilitat condicionada

Definició 9.3. *Sigui* $(\Omega, \mathcal{F}, P)$ *un espai de probabilitat i sigui* $B \in \mathcal{F}$ *tal que* $P(B) > 0$*. Definim una nova funció* $P(\cdot \mid B)$*, anomenada la probabilitat condicionada donat* B*, de* $\mathcal{F}$ *en els reals mitjançant* $P(A \mid B) = P(A \cap B)/P(B)$ *per a tot succés* $A \in \mathcal{F}$*. És a dir,*

$$P(\cdot \mid B) : \mathcal{F} \longrightarrow \mathbb{R}$$
$$A \longrightarrow \frac{P(A \cap B)}{P(B)}$$

En altres paraules, el que ens diu la definició és que sabent que ha ocorregut un esdeveniment B es vol saber quina és la probabilitat que també hi hagi ocorregut un altre esdeveniment A, òbviament tots dos pertanyents a l'espai mostral. És a dir, a partir d'aquesta nova funció es mesura la probabilitat que A ocorri també utilitzant la proporció relativa del temps en que B i A ocorren juntes (relatiu a la probabilitat total que ocorri B).

A continuació donarem algunes propietats útils per més endavant, però vistes i demostrades en classe, de manera que, igual que a l'apartat anterior solament les enunciarem. Demostrarem aquells resultats no coneguts per l'assignatura correspo- nent.

Proposició 9.4. $P(\cdot \mid B)$ *és una probabilitat.*

Proposició 9.5. ***(Regla del producte)*** *Sigui* $A_1, A_2, \ldots, A_k$ *una col·lcció finita d'esdeveniments pertanyents a* $\mathcal{F}$*. Si* $P(A_1 \cap A_2 \cap \ldots \cap A_{k-1}) > 0$ *aleshores* $P(A_1 \cap A_2 \cap \ldots \cap A_k) = P(A_1)P(A_2 \mid A_1) \ldots P(A_k \mid P(A_1 \cap A_2 \cap \ldots \cap A_{k-1})$*.*

Proposició 9.6. *Siguin* A_1, A_2 *i* B *tres successos pertanyents a* $\mathcal{F}$ *amb* $P(B) > 0$ *aleshores* $P(A_1 \cap A_2 \mid B) = P(A_1 \mid A_2 \cap B)P(A_2 \mid B) = P(A_2 \mid A_1 \cap B)P(A_1 \mid B)$*.*

A més a més, entre moltes d'altres, existeixen dues propietats importants:

- Si A i B són dos esdeveniments pertanyents a $\mathcal{F}$ disjunts aleshores $P(A \mid B) = 0$.
- Si $A \subset B$ són dos esdeveniments pertanyents a $\mathcal{F}$ aleshores $P(A \mid B) = 1$.

Definició 9.7. *Un sistema complet de successos és un conjunt se successos* $\{A_1, \ldots, A_n\}$*, tals que els successos* A_i *són disjunts dos a dos i la seva unió és* Ω*, és a dir:*

1. $A_i \cap A_j = \emptyset \ \forall i \neq j$.

2. $A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$.

Utilitzem en moltes hipòtesis aquest tipus de conjunt perquè ens són molt útils les seves propietats, com poden ser:

- $P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n) = 1$.
- Un succés A i el seu complementari $\overline{A}$ formen un sistema complet.
- Si un conjunt de successos disjunts $\{A_1, A_2, \ldots, A_{n-1}\}$ no formen un sistema complet sempre podem complementar-lo amb $A_n = \overline{A_1 \cup A_2 \cup \ldots \cup A_n} = \overline{A_1} \cap \overline{A_2} \cap \ldots \cap \overline{A_{n-1}}$ de manera que ja ho és.

Teorema 9.8. ***(Teorema de la probabilitat total)*** *Sigui $\{A_1, A_2, \ldots, A_k\}$ un sistema complet de successos, on la probabilitat de cadascun d'ells és diferent de zero, $P(A_i) > 0 \ \forall i \in \mathbb{N}$. Sigui B també un esdeveniment de $\mathcal{F}$ aleshores*

$$P(B) = P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2) + \ldots + P(B \mid A_k)P(A_k).$$

Demostració. Podem escriure B com la intersecció de B amb Ω que és la unió de tots els esdevenimets A_i, de manera que:

$$B = B \cap \Omega = B \cap (A_1 \cup A_2 \cup \ldots \cup A_k) = (B \cap A_1) \cup (B \cap A_2) \cup \ldots \cup (B \cap A_k).$$

Com que $(B \cap A_i) \cup (B \cap A_j) \neq \emptyset \ \forall i \neq j$ tenim que:

$$\begin{aligned} P(B) &= P((B \cap A_1) \cup (B \cap A_2) \cup \ldots \cup (B \cap A_k)) \\ &= P(B \cap A_1) + (B \cap A_2) + \ldots + (B \cap A_k) \\ &= P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2) + \ldots + P(B \mid A_k)P(A_k). \end{aligned}$$

□

Definició 9.9. *Siguin A i B dos successos pertanyents a $\mathcal{F}$ són independents si $P(A \cap B) = P(A)P(B)$.*

Proposició 9.10. *Si A i B dos successos pertanyents a $\mathcal{F}$ independents aleshores $P(A \mid B) = P(A)$.*

9.5 Variables aleatòries

9.5.1 Definició

Tant en aquest apartat com en el següent, Distribució d'una variable aleatòria, veurem conceptes bàsics estudiats a classe però que ens ajudaran acabar d'afiançar coneixements per poder entendre bé com influeix l'estadística Bayesiana en aquest tipus de variables.

A la pràctica una variable aletòria X és una funció que pren valors reals associats a cada element del espai mostral Ω. Si volem una definició més correcta hauria de ser la següent.

Definició 9.11. *Una variable aleatòria és una aplicació* $X : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}), Q)$ *on* $Q = P \circ X^{-1}$ *que compleix la propietat següent:*

$$\forall B \in \mathcal{B}(\mathbb{R}), X^{-1}(B) = \{w \in \Omega, X(w) \in B\} \in \mathcal{F}.$$

Exemple 9.12. Considerem l'experiment de comptar el nombre d'assignatures associades a les llengües d'un alumne de primària. Tenim un espai mostral com:

$$\Omega = \big\{\text{Llengua Catalana, Llengua Castellana, Anglès}\big\}.$$

Aleshores podem definir una variable aletòria X tal que: X(Llengua Catalana)=0 $\in \mathbb{R}$, X(Llengua Castellana)=1 $\in \mathbb{R}$ i X(Anglès)=2 $\in \mathbb{R}$.

Exemple 9.13. Considerem l'experiment de comptar quant de temps tarda en fondre's una bombeta. Si suposem que una bombeta convencional té una vida màxima de 5 anys, aleshores l'espai mostral és $\Omega = \big[0, 5\big]$.

Tal i com hem vist en els dos exemples anteriors hi ha dos tipus de variables aletòries: discretes i continues. Les primeres són variables aleatòries amb un càcul i raonament més senzill, com observem a partir de l'exemple 4.2, que les segones. També es podria definir un altre tipus com la unió d'aquestes dos, però la seva utilització és molt més complexa i la seva explicació no ens aporta nova informació rellevant per al treball. Per això, no parlarem d'elles.

Definició 9.14. *Una variable aleatòria X direm que és discreta si la seva llei està concentrada en un conjunt* $B_0 \in \mathcal{B}$.

Per tant, una variable aleatòria és discreta si el conjunt de possibles valors és numerable. Les variables aleatòries discretes poden ser finites o infinites. Les variables aleatòries discretes finites són aquelles que prenen un número finit determinat de valors diferents, en canvi, en les variables aleatòries discretes infinites aquest nombre pot ser infinit, com passa en el cas de llançar un dau fins a obtenir un número en concret.

Per l'altra banda tindríem les variables aleatòries continues, que són aquelles on el conjunt de possibles valors és infinit i no numerable dins d'un interval o unió d'intervals sobre la recta dels reals. A més compleixen que $P\{X = x\} = 0, \forall x \in \mathbb{R}$. Habitualment s'utilitzen més les variables aleatòries contínues ja que són més fàcils de tractar. La seva definició formal està en el següent apartat. Amb les variables aleatòries continues es poden definir operacions com la suma, la resta o la multiplicació i el resultat també és una variable aleatòria.

9.5.2 Distribució d'una variable aleatòria

Definició 9.15. *La funció de distribució associada a una variable aleatòria X és la funció* $F : \mathbb{R} \longrightarrow [0, 1]$ *definida per*

$$F(x) = P \circ X^{-1}\left((-\infty, x]\right).$$

Proposició 9.16. *Sigui F la funció de distribució d'una variable aleatòria X. Es compleix que*

1. F és creixent,

2. F és contínua per la dreta,

3. $\lim_{x\to\infty} F(x) = 1$ i $\lim_{x\to-\infty} F(x) = 0$.

Definició 9.17. *Una funció $f : \mathbb{R} \longrightarrow \mathbb{R}$ s'anomena una densitat si compleix les condicions següents:*

1. $f \geq 0$,

2. f és integrable (en el sentit de Riemann) en $\mathbb{R}$,

3. es té que

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

Definició 9.18. *Es diu que una variable aleatòria X és absolutament contínua (o té llei absolutament contínua) amb densitat f si la seva funció de distribució F es pot escriure com*

$$F(x) = \int_{-\infty}^{\infty} f(y)dy,$$

per a tot $x \in \mathbb{R}$, on la funció f satisfà les condicions de la definició anterior.

Teorema 9.19. ***(Teorema de Probabilitats Totals)*** *Siguin X i Y dues variables aleatòries discretes, aleshores*

$$P(X = a) = \sum_{i=0}^{n} P(X = a \mid Y = i)P(Y = i)$$

on $P(X = a \mid Y = i)$ és la probabilitat de que succeeixi $\{X = a\}$ sabent que ha succeït $\{Y = i\}$ i P(Y=i) la probabilitat que succeeixi $\{Y = i\}$.

Teorema 9.20. ***(Teorema de Probabilitats Totals)*** *Siguin X i Y dues variables aleatòries absolutament continues, aleshores*

$$p(x) = \int_{-\infty}^{+\infty} p(x \mid y)p(y)dy$$

on podem obtenir les distribucions marginals de les seves variables i observar que podem calcular la probabilitat de que la variable es trobi entre dos valors però no la que sigui igual a un.

9.6 Vectors aleatoris

D'igual manera que hem fet amb una variable aleatòria, ara ho fem amb un vector d'elles. Com moltes de les propietats són semblants, només explicarem les definicions bàsiques i si cal alguna proposició.

Definició 9.21. *Un vector aleatori m-dimensional és una aplicació* $X : \Omega \longrightarrow \mathbb{R}^m$, $X = (X_1, ..., X_m)$, *tal que cadascun dels components* $X_i : \Omega \longrightarrow \mathbb{R}$, $i = 1, ..., m$, *és una variale aleatòria.*

Definició 9.22. *La funció de distribució d'un vector aleatori X és una aplicació* $F : \mathbb{R}^m \longrightarrow [0, 1]$ *definida per*

$$F(x) = P\{X \leq x\} = P\{X_1 \leq x_1, ..., X_m \leq x_m\},$$

per a tot $(x_1, ..., x_m)$.

La funció de distribució d'un vector aleatori té propietats similars a les presentades en la proposició 9.16 per a m=1.

Definició 9.23. *Una funció* $f : \mathbb{R}^m \longleftarrow \mathbb{R}$ *és una densitat en* $\mathbb{R}^m$ *si es compleixen les condicions següents:*

1. $f \geq 0$,
2. *existeix la integral*
$$\int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f(x)dx,$$
en el sentit de Riemann, i
$$\int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f(x)dx = 1.$$

Definició 9.24. *Un vector aleatori X m-dimensional és absolutament continu (o té llei absolutament contínua) amb densitat f, si la seva funció de distribució F es por escriure com*

$$F(x) = \int_{-\infty}^{x_1} ... \int_{-\infty}^{x_m} f(x)dx,$$

on f és una funció de densitat en $\mathbb{R}^m$, *és a dir, compleix les condicions de la definició anterior, i* $x = (x_1, ..., x_m)$.

Proposició 9.25. *Sigui X un vector aleatori amb llei absolutament contínua. Aleshores, cadascuna de les variables aleatòries components,* X_i, $i = 1, ..., m$, *són també absolutament contínues i les seves densitats respectives, que denotarem per* f_i, *s'expressen com*

$$f_i(y) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f(x_1 ... x_{i-1}, y, x_{i+1}, ..., x_m)dx_1 ... dx_{i-1}dx_{i+1} ... dx_m.$$

Les densitats f_i s'anomenen densitats marginals.

10 Annex II: Taules del Capítol 6

Taula 6: Demanda agregada

	A priori			*A posteriori*	
Variable	Distribució	Mitjana	Desviació estàndar	Moda	Interval del 90% de probabilitat
a_y	Beta	0.55	0.15	0.49	0.28 / 0.61
a_{re}	Beta	0.40	0.15	0.16	0.10 / 0.28
a_{rmc}	Beta	0.40	0.15	0.28	0.12 / 0.52
a_q	Gamma	0.06	0.025	0.06	0.03 / 0.10
a_{tot}	Beta	0.10	0.05	0.04	0.02 / 0.07
a_{y^*}	Gamma	0.10	0.05	0.08	0.02 / 0.18
a_{fis}	Beta	0.40	0.15	0.25	0.13 / 0.37
ρ_y	Beta	0.50	0.10	0.47	0.33 / 0.65
$SD_{\epsilon_t^y}$	GammaInv	0.65	0.15	0.49	0.41 / 0.64

Taula 7: Oferta agregada o corba de Philips

	A priori			*A posteriori*	
Variable	Distribució	Mitjana	Desviació estàndar	Moda	Interval del 90% de probabilitat
b_{p^*}	Beta	0.11	0.05	0.05	0.03 / 0.09
b_p	Beta	0.50	0.20	0.68	0.56 / 0.91
b_y	Beta	0.20	0.08	0.10	0.05 / 0.20
ρ_π	Beta	0.15	0.05	0.12	0.06 / 0.21
$SD_{\epsilon_t^\pi}$	GammaInv	0.75	0.25	0.52	0.44 / 0.73

Taula 8: Corba de Philips (inlfació importada)

	A priori			*A posteriori*	
Variable	Distribució	Mitjana	Desviació estàndar	Moda	Interval del 90% de probabilitat
c_p	Beta	0.30	0.10	0.31	0.21 / 0.44
c_{pf}	Beta	0.65	0.15	0.58	0.47 / 0.68

Taula 9: Regla de política monetària

	A priori			*A posteriori*	
Variable	Distribució	Mitjana	Desviació estàndar	Moda	Interval del 90% de probabilitat
f_i	Beta	0.70	0.10	0.66	0.53 / 0.75
f_p	Beta	1.50	0.40	1.93	1.34 / 2.43
f_y	Beta	0.50	0.10	0.51	0.35 / 0.68
ρ_i	Beta	0.15	0.05	0.12	0.07 / 0.21
$SD_{\epsilon_t^i}$	GammaInv	4.20	0.60	4.23	3.40 / 5.38

Taula 10: Paritat descoberta de les taxes d'interès

	A priori			*A posteriori*	
Variable	Distribució	Mitjana	Desviació estàndar	Moda	Interval del 90% de probabilitat
ρ	Beta	0.60	0.12	0.66	0.54 / 0.79
ρ_s	Beta	0.30	0.10	0.39	0.23 / 0.55
$SD_{\epsilon_t^s}$	GammaInv	1.60	0.30	1.60	1.38 / 1.94

Taula 11: Comparació de paràmetres

Paràmetres	Valors Originals	Moda *a posteriori*
a_y	0.50	0.49
a_{re}	0.00	0.016
a_{rmc}	0.26	0.28
a_q	0.02	0.06
a_{tot}	0.09	0.04
a_{y^*}	0.01	0.08
a_{fis}	0.15	0.25
b_{p^*}	0.08	0.05
b_p	0.92	0.68
b_y	0.20	0.10
f_i	0.70	0.66
f_p	1.50	1.93
f_y	0.50	0.51
ρ	0.50	0.66

11 Annex III: Script del Capítol 7

Explicació **Figura 12**: Els experts creuen que els possibles valors per a θ són els que posem en el vector prop (proporció). A partir de les conclusions extretes dels anys anteriors a cada valor li donen un pes, que correspon al vector anomenat priori. Podem convertir aquests pesos en probabilitats *a priori* si dividim cada un per la suma de tots. Dibuixem la probabilitat *a priori* a partir de la funció de R anomenada 'plot'.

Figura 12: Correspon a la Figura 5

```
>
> prop = seq(0.05, 0.95, by = 0.05)
> priori = c(1,1,2,2,3,4,6,6,7,8,8,9,9,8,8,6,5,5,3)
> priori = priori / sum(priori)
> plot(prop, priori, type = "h", ylab = "Probabilitat a Priori")
>
```

Explicació **Figura 13**: Per poder calcular les probabilitats *a posteriori* utilitzem la funció 'pdisc' que es troba en la llibreria 'LearnBayes' de R.

Figura 13: Correspon a la Figura 5

```
>
> library(LearnBayes)
> x = c(63, 27)
> posteriori = pdisc ( prop, priori, x)
> cbind(prop, priori, posteriori)
      prop     priori   posteriori
 [1,] 0.05 0.00990099 1.031155e-60
 [2,] 0.10 0.00990099 2.209144e-42
 [3,] 0.15 0.01980198 1.171565e-31
 [4,] 0.20 0.01980198 1.694438e-24
 [5,] 0.25 0.02970297 5.671163e-19
 [6,] 0.30 0.03960396 1.142937e-14
 [7,] 0.35 0.05940594 3.826264e-11
 [8,] 0.40 0.05940594 1.984681e-08
 [9,] 0.45 0.06930693 3.689396e-06
[10,] 0.50 0.07920792 2.455085e-04
[11,] 0.55 0.07920792 5.785668e-03
[12,] 0.60 0.08910891 6.502304e-02
[13,] 0.65 0.08910891 2.737142e-01
[14,] 0.70 0.07920792 4.038468e-01
[15,] 0.75 0.07920792 2.269896e-01
[16,] 0.80 0.05940594 2.400527e-02
[17,] 0.85 0.04950495 3.859247e-04
[18,] 0.90 0.04950495 2.488424e-07
[19,] 0.95 0.02970297 3.354090e-14
> plot (prop, posteriori, type = "h", ylab = "Probabilitat a Posteriori")
>
```

Explicació **Figura 14**: Com les tres densitats són una densitat beta, utilitzem la funció de R 'debta' per calcular els valors de la probabilitat *a priori*, la versemblança i la *posteriori*, cadascuna amb els corresponents valors dels paràmetres. Amb la funció 'plot' les pintem en el mateix gràfic.

Explicació **Figura 15**: A partir de les funcions 'pbeta' trobem les probabilitats que θ tingui valors per sota del 0.50 o per sobre del 0.90. A més, amb la funció 'qbeta' trobem l'estimació de l'interval on trobem el valor de θ.

Explicació **Figura 16**: Corresponen als càlculs fets amb el mètode de la simulació per treure resultats a partir de la distribució *a posteriori*. A partir de la funció

Figura 14:

```
>
> prop = seq (0, 1, length = 500)
> a = 16.5
> b = 6.3
> e = 63
> f = 27
> priori = dbeta (prop, a, b)
> ver = dbeta (prop, e+1, f+1)
> posteriori = dbeta (prop, a+s, b+f)
> plot (prop, posteriori, type = "l", ylab = "Densitats", lty = 2, lwd = 3)
> lines (prop, ver, lty =1, lwd=3)
> lines (prop, priori, lty = 3, lwd = 3)
> legend (.0,4,c("Priori","Versembl.","Posteriori"),lty=c(3,1,2),lwd=c(3,3,3))
>
```

Figura 15:

```
>
> pbeta(0.50, a+e, b+f)
[1] 4.661643e-06
> 1 - pbeta(0.90, a+e, b+f)
[1] 5.246898e-09
> qbeta ( c(0.05, 0.95), a+e, b+f)
[1] 0.6323986 0.7730294
>
```

'rbeta' simulem 10.000 valors de la proporció *a posteriori* beta amb paràmetres igual que abans. A més responem, amb aquest mètode les mateixes preguntes que a la Figura 15.

Figura 16:

```
>
> pv = rbeta (10000, a+e, b+f)
> hist (pv, xlab = "prop", main = " ")
>
> sum(pv <= 0.50)/10000
[1] 1e-04
> sum(pv >= 0.90)/10000
[1] 0
> quantile (pv, c(0.05, 0.95))
       5%       95%
0.6327473 0.7731103
>
```

Explicació **Figura 17**: Amb la llibreria LearnBayes, podem crear l'interval i un vector que conté els pesos *a priori* que convertim amb probabilitat dividint per la suma de tots. Per poder crear l'histograma fem servir la funció 'histprior'.

Explicació **Figura 18**: Amb la llibreria LearnBayes, creem com abans l'interval i a cadascun li donem un pes que després convertirem en probabilitat. A més, el valor de m correspon al valor de la mostra que volem predir i ys és un vector que conté el número de èxits d'interès. Amb la funció 'cbind' podem veure la probabilitat predictiva. Òbviament al tenir una mostra tan gran no ensenyarem tots els valors.

Explicació **Figura 19**: Calculem les probabilitats predictives a partir de la funció 'pbetap' corresponents a cadascun dels valori del vector ys. ab és un vector dels paràmetres beta a i b, i m, ys, 'cbind' corresponen igual que abans. Amb la funció 'plot' podem fer la gràfica d'aquesta valors que correspon a la Figura 11 del Capítol 7.

Explicació **Figura 20**: Simulem 10.000 densitats *a posteriori* beta. També

Figura 17:

```
>
> interval = seq (0.05, 0.95, by = 0.10)
> pes = c(1, 1, 2, 4, 6, 7, 8, 9, 8, 5)
> priori = pes / sum(pes)
> prop = seq(0, 1, length = 500)
> plot (prop, histprior(prop, interval, priori), type = "l", ylab = "Densitat Priori", ylim = c(0, 0.20))
>
>
> e = 63
> f = 27
> vers = dbeta (prop, e+1, f+1)
> posteriori = vers * histprior(prop, interval, priori)
> plot(prop, posteriori, type = "l", ylab="Densitat Posteriori")
>
```

Figura 18:

```
>
> prop = seq(0.05, 0.95, by= 0.10)
> pes = c(1, 1, 2, 4, 6, 7, 8, 9, 8, 5)
> priori = pes / sum(pes)
> m = 1560 ; ys = 0: 1500
> pred=pdiscp(prop, priori, m, ys)
> cbind(0:1500,pred)
               pred
 [1,]    0 3.477395e-37
 [2,]    1 2.855125e-35
 [3,]    2 1.171352e-33
 [4,]    3 3.201697e-32
 [5,]    4 6.559265e-31
 [6,]    5 1.074339e-29
 [7,]    6 1.465436e-28
 [8,]    7 1.712246e-27
 [9,]    8 1.749420e-26
[10,]    9 1.587777e-25
[11,]   10 1.296128e-24
[12,]   11 9.612430e-24
[13,]   12 6.530550e-23
[14,]   13 4.092831e-22
[15,]   14 2.380304e-21
[16,]   15 1.291211e-20
[17,]   16 6.562238e-20
[18,]   17 3.136872e-19
[19,]   18 1.415261e-18
[20,]   19 6.045243e-18
```

calculem l'error estàndard simulat a partir de la desviació estàndar dividida per l'arrel quadrada de la mida de la mostra de la simulació.

Figura 19:

```
>
> ab = c(16.5, 6.3)
> m = 1560; ys = 0:1500
> pred = pbetap (ab, m, ys)
> cbind (0:1500, pred)
                 pred
 [1,]     0 5.308690e-35
 [2,]     1 8.729680e-34
 [3,]     2 7.612590e-33
 [4,]     3 4.678515e-32
 [5,]     4 2.273039e-31
 [6,]     5 9.287823e-31
 [7,]     6 3.316832e-30
 [8,]     7 1.062501e-29
 [9,]     8 3.110481e-29
[10,]     9 8.438602e-29
[11,]    10 2.144516e-28
[12,]    11 5.148727e-28
[13,]    12 1.175893e-27
[14,]    13 2.569124e-27
[15,]    14 5.395028e-27
[16,]    15 1.093241e-26
[17,]    16 2.144960e-26
[18,]    17 4.086632e-26
[19,]    18 7.579640e-26
[20,]    19 1.371589e-25
[21,]    20 2.426226e-25
[22,]    21 4.202548e-25
[23,]    22 7.138849e-25
[24,]    23 1.190877e-24
[25,]    24 1.953250e-24
[26,]    25 3.153385e-24
[27,]    26 5.015968e-24
[28,]    27 7.868321e-24
[29,]    28 1.218188e-23
```

Figura 20:

```
>
> prop = rbeta(10000, 79.5, 33.3)
> mit = mean (prop^56.7)
> errstd = sd (prop^56.7)/sqrt(10000)
> c (mit, errstd)
[1] 1.486574e-07 1.163329e-08
>
```

Referències

[1] Andrew Gelman, John B. Carlin, Hal S. Stern and Donald B. Rubin; *Bayesian Data Analysis,* Second edition; Chapman & Hall/CRC Texts in Statistical Science, 2003.

[2] William M. Bolstad; *Introduction to Bayesian Statistics,* Second Edition; Wiley, 2007.

[3] José Serrano Angulo; *Iniciación a la estadísitica Bayesiana*; Cuadernos de estadísitca, LA MURALLA, S.A. / HESPÉRIDES, 2003.

[4] Harold J. Larson; *Introducción a la teoría de probabilidades e inferencia estadística*; LIMUSA, México, 1993.

[5] J.R. Norris *Markov Chains;* Cambridge Series in Statistical and Probabilistic Mathematics: Cambridge University Press, 1997.

[6] Patrick Gordon; *Cadenas Finitas de Markov y sus Aplicaciones*; Barcelona: Hispano Europea, 1967 (Colección ESADE (Hispano Europea)).

[7] Alfonso García Pérez; *Problemas de decisión Bayesiana no paramétrica: aproximación lineal*; Universidad Complutense de Madrid. Servicio de Reprografía, 1983; Col·lecció Tesis doctorales (Universidad Complutense de Madrid).

[8] Marta Sanz; *Probabilitats*; Edicions Universitat de Barcelona, 1999.

[9] Jim Albert; *Bayesian computation with R*; New York: Springer, cop. 2008.

[10] *Pàgina web:* Datos básicos del sistema universitario español; Ministerio de Educación, Cultura y Deporte.Curso 2013-2014.

`www.mecd.gob.es/dms/mecd/prensa-mecd/actualidad/2014/02/20140213-datos-univer/datos-cifras-13-14.pdf`

[11] *Pàgina web:* Manuel Mendoza R., Pedro Regueiro M.; *Estadística Bayesiana*; Departamento de Estadística Instituto Tecnol'ogico Autónomo de México, 2011.

`allman.rhon.itam.mx/~lnieto/index_archivos/NotasBayesMR.pdf`

[12] *Pàgina web: Capítulo 1. Preliminares: elementos de estadística Bayesiana*

`www.iuma.ulpgc.es/~nunez/mastertecnologiastelecomunicacion/Tema2InferenciaEstadisticaInferenciaEstadistica/inferencia-estadistica-Bayesiana/Bayes-polo-asignatura-preliminares_1.pdf`

[13] *Pàgina web:* Diego Salmerón Martínez; *Bioestadística. Análisis Bayesiano de datos experimentales.*

[14] *Pàgina web: Tema 1: Introducción a la Estadística Bayesiana.*

`halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/Bayes/tema1Bayes.pdf`

[15] *Pàgina web:* Sergio Hernández González; *Historia de la estadística.*

www.uv.mx/cienciahombre/revistae/vol18num2/articulos/historia/

[16] *Pàgina web:* David Ruiz Muñoz; *Manual de Estadística.*

www.eumed.net/cursecon/libreria/drm/index.htm

[17] *Pàgina web:* Miguel A. Gómez Villegas; *Thomas Bayes en su tricentenario.*

www.mat.ucm.es/∼villegas/AnBayesSeio.pdf

[18] *Pàgina web:* Miguel A. Gómez Villegas;*Estadísticos significativos.*

www.mat.ucm.es/∼villegas/EstadSig.pdf

[19] *Pàgina web:* Ms Carlos López de Castilla Vásquez; *Estadística Bayesiana: Modelos uniparamétricos*

[20] *Pàgina web:* Manuel Mendoza R. y Pedro Regueiro M.; *Estadística Bayesiana.*

allman.rhon.itam.mx/∼lnieto/index_archivos/NotasBayesMR.pdf

[21] *Pàgina web:* Conchi Ausín; *Tema 6: Introducción a la Inferencia Bayesiana.*

halweb.uc3m.es/esp/Personal/personas/causin/esp/2012-2013/SMB /Tema6.pdf

[22] *Pàgina web:* Conchi Ausín; *Tema 8:: Métodos de cadenas de Markov Monte Carlo.*

halweb.uc3m.es/esp/Personal/personas/causin/esp/2012-2013/SMB /Tema8.pdf