

Mètodes Bayesians: Sessions Pràctiques

Sessió 1: Distribució a Priori, Predictiva a Priori i Versemblança.

Objectius:

Calcular i representar:

- la distribució a priori
- la versemblança
- la distribució predictiva a priori

Exercici 1.1. Asma: Estimació d'una proporció (Binomial). És vol estimar el percentatge d'asmàtics de la ciutat de Sabadell, anomenem aquest paràmetre com θ . Amb aquest objectiu es pren una mostra aleatòria simple de 200 ciutadans. D'aquests 11 van resultar ser asmàtics. Els professionals de la salut a través de la seva experiència creuen que la prevalença estar al voltant del 5%, i que és molt improbable que prengui taxes superiors al 20%.

- Escull la distribució a priori conjugada
- Dibuixa en un mateix gràfic la distribució a priori i la versemblança
- Dibuixa la predictiva a priori

Exercici 1.2 Hospitals: Estimació d'una proporció (binomial). Considerem la taxa de mortalitat en 12 hospitals d'una determinada cirurgia cardíaca. Les dades són les següents:

Hospital	N operacions	N morts
A	47	0
B	148	18
C	119	8
D	810	46
E	211	8
F	196	13
G	148	9
H	215	31
I	207	14
J	97	8
K	256	29
L	360	24

El nombre de morts y_i de l'hospital i el podem modelar pensant que segueix una distribució binomial.

- Escull els paràmetres de la distribució a priori conjugada
- Dibuixa la funció de versemblança

Exercici 1.3. Alçada. Estimació de la mitjana (Normal). És vol estimar l'alçada mitjana dels catalans. Suposem $y|\mu \sim \text{Normal}(\mu, \sigma=8)$, $\sigma=8$ vol dir que en un rang de 50cm tenim el 99,7% de la població. Sabem, a priori, coses de μ ? Podem tenir alçades negatives? Podem tenir alçades de 80 m?. Tria el valor dels paràmetres de la distribució a priori conjugada ($\mu \sim \text{Normal}(m, s)$) i dibuixa-la. Dibuixa també la predictiva a priori.

Exercici 1.4 Bombetes: Temps de vida bombetes (Exponencial). Un fabricant de bombetes vol estimar el temps de vida d'unes noves bombetes que han dissenyat. Ell ja sap que el temps de vida en hores d'una bombeta segueix una distribució exponencial. Es mesuren el temps de vida de 10 d'aquestes bombetes mitjançant un mecanisme de desgast accelerat (26293, 10123, 3168, 23340, 5459, 13143, 10270, 1699, 15061, 29010). S'ha escollit com a distribució a priori una Gamma($a=0.01$, $b=0.01$). Dibuixa en un mateix gràfic la distribució a priori i la versemblança.

Exercici 1.5 Gols. Construïu un model Bayesià pel nombre de gols que es fan en un partit de futbol de primera divisió espanyola. Descriu i justifica l'elecció de la distribució a priori.

Exercici 1.6 Sèpia Verda: Estimació d'una freqüència (Poisson). El portal d'internet de l'associació cultural *La Sèpia Verda* vol aprendre sobre la freqüència de visitants setmanals al portal. Amb aquest objectiu s'ha recollit el nombre de visitants de les darreres 10 setmanes. Les dades es troben a l'arxiu *sepiaverda.txt*.

En el supòsit de que els responsables de l'associació visquessin en el convenciment de que rarament el nombre de visitants serà inferior a 5 i rarament superior a 40.

- Tria els paràmetres de la distribució a priori (Nota: Pot ser útil triar els paràmetres de la a priori utilitzant la predictiva a priori).
- Dibuixa la distribució a priori i la versemblança en un mateix gràfic
- Dibuixa la predictiva a priori

Sota el supòsit d'una ignorància total sobre el nombre de visitants:

- Tria els paràmetres de la distribució a priori.

Exercici 1.7 Pressió sanguínea: Estimació d'una mitjana (Normal): Tots els estudis d'arreu del món convergeixen en que la desviació estàndard d'aquesta mesura és de 13. No obstant la mitjana de la pressió sanguínia és lleugerament diferent, situant-se la mitjana global de tots els estudis en 125. Les autoritats sanitàries d'Andorra volen estimar μ , la mitjana de la pressió sanguínia dels habitants d'Andorra. Per a tal efecte prenen una mostra de 20 pressions sistòliques sanguínies, que notem per y_i , d'habitants d'Andorra. Assumim que la desviació estàndard de y és 13 (coneguda). Les dades de la mostra són: 98, 160, 136, 128, 130, 114, 123, 134, 128, 107, 123, 125, 129, 132, 154, 115, 126, 132, 136 i 130.

- Escull la distribució a priori d'acord amb l'enunciat i justifica la teva elecció
- Dibuixa en un mateix gràfic la distribució a priori i la versemblança. Explica com has calculat la versemblança i comenta el gràfic.
- Dibuixa la predictiva a priori. Comenta perquè i en què són diferents.

Exercici 1.8 Suïcidis: Estimació d'una proporció (Binomial). En un hospital danès volen estimar el percentatge de pacients ingressats per suïcidi, a través d'enverinament, que ho tornen a intentar suïcidar-se abans de 10 anys. Tot i que hi ha estudis americans al respecte, els investigadors danesos dubten de que el patró de comportament sigui intercanviable amb el dels seus pacients. Així es va fer un seguiment de 974 pacients durant 10 anys, dels quals 103 van intentar suïcidar-se de nou. Aquestes dades no tenen en compte 100 pacients que es van perdre durant el seguiment, els investigadors estan segurs que es tracte de dades faltants aleatòries, en el sentit que la proporció de nous intents de suïcidis no està relacionada amb la pèrdua en el seguiment.

- Escull els paràmetres de la distribució a priori conjugada
- Dibuixa la funció de versemblança

Exercici 1.9 Cremades: Estimació de la diferència de dues proporcions. Ens trobem davant d'un assaig clínic per valorar si els pacients afectats per cremades hipodèrmiques es recuperen més ràpidament quan el tractament combina certa crema antisèptica amb un apòsit hidrocoloide que quan utilitza solament la crema antisèptica. Es coneix amb força certesa que només aproximadament el 60% dels pacients es recupera amb aquest darrer recurs. L'equip investigador té, per altra banda, motius teòrics i indicis empírics sorgits del treball quotidià d'infermera que fan pensar amb força optimisme que el tractament combinat és més efectiu que el tractament simple. S'ha organitzat un experiment amb n pacients, $n/2$ dels quals s'escullen aleatòriament per ser atesos amb el tractament novetós (combinació de crema antisèptica i apòsit hidrocoloide), en tant que els $n/2$ restants se'ls aplicarà el tractament convencional (només crema). Les probabilitats a priori es defineixen sota la convicció anticipada de que P_c (tractament convencional) es torba quasi amb seguretat entre 0,4 i 0,8, amb alta probabilitat en un veïnatge de 0,6, i que es redueix ràpidament quan s'allunya d'aquest punt, i en quant a P_e (tractament experimental), es troba en un entorn de 0,8, amb minsa probabilitat d'estar fora de l'interval $[0,7; 0,9]$.

A la següent taula trobem la distribució d'una mostra de 80 pacients segons el tractament assignat i segons si es recuperen o no:

Tractament	Si	No	Total
Experimental	30	10	40
Convencional	24	16	40
Total	54	26	80

- a) Escull les distribucions a priori d'acord amb l'enunciat
- b) Dibuixa les distribucions a priori i les versemblances

Exercici 1.10 Rentadores: Estimació del temps de vida (Exponencial). Els tècnics d'una empresa de rentadores constaten que les rentadores que fabriquen gairebé sempre fallen abans dels 20 anys. L'empresa ha de decidir el temps de garantia que vol oferir als seus clients de manera que el nombre de reparacions mentre dura la garantia sigui inferior al 1%. Per a fer l'estimació de la garantia és imprescindible conèixer la distribució del temps de vida de les rentadores, per això es sotmeten 12 rentadores a un procés de desgast accelerat, i s'obtenen els següents temps, mesurats en dies, fins a la primera fallada: 2153, 839, 2810, 1314, 1423, 912, 5110, 2664, 3321, 3431, 2737 i 2372.

- a) Escull les distribució a priori d'acord amb l'enunciat
- b) Dibuixa en un mateix gràfic la distribució a priori i la versemblança
- c) Dibuixa la predictiva a priori

Exercici 1.11 Busca (o inventat) unes dades que trobis interessants modelar mitjançant una distribució exponencial. Formula el model Bayesià, fes una discussió de com has escollit la distribució a priori i dibuixa en un mateix gràfic la distribució a priori i la versemblança.