Distribucions a priori: consideracions finals

Anabel Forte, Gonzalo García-Donato i Joaquín Martinez Minaya

El que hem vist fins ara.

Recordem que la distribució prèvia o a priori és la manera que tenim de transmetre a la inferència el nostre coneixement sobre un paràmetre d'interès.

Però això no és tan fàcil com sembla, i no ho és per dos motius. El primer és la dificultat de transformar un coneixement en una distribució de probabilitat. Encara que puguem més o menys intuir quina és la mitjana i / o la variància que hauria de tenir la distribució d'aquest paràmetre, determinar si aquesta ha de ser una normal, una uniforme o una altra distribució no és trivial.

D'altra banda, de l'elecció de la prèvia dependrà, en gran mesura, que puguem obtenir una distribució a posteriori amb una expressió coneguda o, com a mínim, tractable. Això és perquè el denominador del teorema de Bayes o, dit d'una altra manera, la constant d'integració d'aquesta distribució posterior, s'obté a partir d'una integral sobre la versemblança per la prèvia.

La primera solució que vam aplicar en aquest curs va ser utilitzar prèvies conjugades. Aquest tipus de prèvies es combinen amb la versemblança corresponent de manera que la distribució posterior pertany a la mateixa família de distribucions. És a dir, la prèvia i la posterior seran les dues gamma, beta, normal, etc. només que amb paràmetres diferents. Els paràmetres de la prèvia reflecteixen el coneixement a priori, mentre que els paràmetres obtinguts per a la posterior reflectiran la combinació del coneixement previ amb el que aporten les dades.

En el repte 3 vam estudiar l'ús dels mètodes MCMC per resoldre el problema de no tenir prèvies conjugades; això ens va començar a succeir quan ens vam allunyar dels models senzills estudiats en el repte 2.

Però a més, quan entrem en el món de les modelitzacions complexes, el problema no serà només que no hi hagi prèvies conjugades, el problema vindrà també de la mà de la manca de coneixement previ sobre la naturalesa de cada un dels paràmetres. Però què podem fer en aquest cas?

Prèvies en absència d'informació prèvia.

Més enllà de decidir quina família de prèvies triem en absència d'informació, el problema és com transmetem realment aquesta falta de coneixement.

La nostra intuïció ens pot fer pensar que la millor opció seria donar la mateixa probabilitat a tots els possibles valors del paràmetre. Això podria aconseguir-se amb una distribució uniforme sobre aquests valors, però quan tenim un paràmetre que es mou des de $-\infty$ a $+\infty$ o, fins i tot de 0 a $+\infty$, això ja no és tan senzill. En aquest cas, la distribució uniforme definida en aquests rangs integra infinit i és, per tant, impròpia.

Una possible solució és que, en lloc de fer que la uniforme arribi a tots els valors, agafem un interval “suficientment gran” però, quant és prou gran? com triem aquest valor límit?

De la mateixa manera, podem optar per qualsevol altra distribució no uniforme però que abasti un ampli rang de valors. Podria ser una normal amb una variància prou gran per paràmetres que es mouen entre $-\infty$ i $+\infty$ o una gamma amb molta variància per a paràmetres que van entre 0 a $+\infty$. No obstant això, el problema torna a ser el mateix, quan és bastant gran la variància?

Però a més, ens hem de fer una altra pregunta, com afecta el valor triat al resultat?

Doncs bé, la solució no és senzilla. A aquest tipus de distribucions se’ls anomena gandules (*vagas* en castellà) o mínim informatives i s’ha demostrat que, en alguns casos, el valor d’aquesta variància és determinant en els resultats finals arribant fins i tot al fet que la distribució posterior sigui impròpia (i això no és gens desitjable).

Però, ¿quina opció ens queda llavors?

Hi ha estudis que es centren en el desenvolupament matemàtic de prèvies anomenades *Objectives. Són distribucions que es construeixen per tal d’incorporar la mínima informació possible a l’estudi d’un paràmetre concret.

Per exemple, en el cas del model binomial, se sap que la prèvia que menys informació aporta a l’estudi de la probabilitat (paràmetre d’interès en aquest cas) és la Beta $(1/2,\ 1/2)$. Un cas curiós perquè, segurament, la nostra elecció de prèvia no informativa per a un paràmetre que sabem que es mou en l’interval 0,1, hagués estat una uniforme en aquest interval i, no obstant això, la teoria ens demostra que no és cert.

El problema de les prèvies objectives és que cal conèixer-les i pot ser que, per començar, no vulguis ficar-te en aquest món, ja que requereix un coneixement matemàtic més avançat i/o cercar en una literatura una mica més complicada.

En aquest cas, podem recórrer a prèvies mínim informatives que han demostrat tenir un bon comportament. Quines són?

Doncs bé, els dos casos més comuns i per als que més clara està la tria són:

- Els paràmetres d’un predictor lineal. Aquests admeten l’ús d’una distribució prèvia normal amb una variància àmplia i no es veuen afectats pel valor de la mateixa (sobretot si el nombre de dades és gran).
- Un paràmetre de variància. En aquest cas, el millor és utilitzar una uniforme $(0,\ a)$ sobre la desviació estàndard, on a és algun valor que considerem gran tenint en compte la variable a la qual representa aquesta variància. Per exemple, si parlem de la variabilitat de l’altura, evidentment una desviació de 4 metres sembla més que suficient i per això podem assignar una distribució uniforme (0,4).