

Tocarà la **loteria**?

Probabilitat i progrés científic

Albert Gras Martí

Físic · Membre de l'IEC

Incomprensió popular del significat de probabilitat

Suposem que els darrers 5 premis de la Grossa de Cap d'Any acaben en 9. Aleshores, si comptem participacions acabades del 0 al 8 augmentarem la probabilitat que, almenys, ens tornem els diners. Cert? FALS! La probabilitat que el proper premi acabe en un dígit qualsevol, 0 a 9, és sempre del 10 %.

Suposem que la probabilitat de pluja a Gandia és d'un 55 %. Hi plourà? Potser.



Grossa de Cap d'Any. (Font: Vilaweb) / Previsió del temps. (Font: cel·lular propi)

Dues definicions de probabilitat

La probabilitat (freqüentista) d'un esdeveniment aleatori A és la possibilitat d'ocurrència de l'esdeveniment, $P(A) = fA/N$, la freqüència fA d'A (les vegades que s'observa) dividida per N (les vegades que pot ocórrer).

Exemples: 1) La probabilitat d'obtenir cara llançant una moneda és del 50 %. Obtindrem aproximadament la meitat de cares si fem moltíssims llançaments. 2) La probabilitat de traure un 5 si llancem un dau és 1/6. De 6 milions de llançaments, eixirà vora un milió de vegades. Tanmateix, cada vegada que llancem la moneda o el dau, totes les opcions tenen la mateixa probabilitat, 1/2 o 1/6 respectivament.

I com hem de calcular probabilitats si disposem d'informació parcial? Fem servir la probabilitat condicionada, $P(A|B)$: la probabilitat de l'esdeveniment A, si sabem que l'esdeveniment B ha ocorregut.

Exemples: 1) Quina probabilitat hi ha que tinguem un virus, si el test dona positiu? 2) Només un de tres camins (I, II i III) està transitable. Triem l'I i aleshores ens diuen que el camí II està tallat. Convé canviar de tria al III o mantenir-nos-hi?

Veurem que el fet de tenir informació addicio-

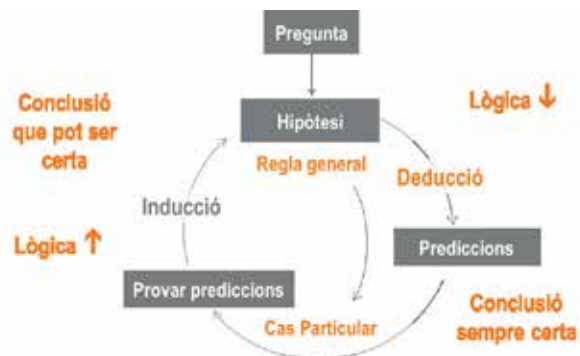
nal (un test positiu o el camí II tallat) modifica la probabilitat de tenir el virus o d'haver triat el camí correcte.

La ciència progressa per inducció (bayesiana), no per deducció

L'objectiu de la ciència és desentranyar els secrets de la naturalesa a força d'identificar i comprendre les estructures o els fets observables. Anomenem comprensió al resultat de desenvolupar models teòrics que expliquen les observacions i que siguin capaços de fer prediccions que puguem posar a prova.

La lògica deductiva es basa en coneixements axiomàtics (quan la informació disponible és completa) que condueixen a conseqüències que són certes o falses, sense opcions intermèdies: la probabilitat que una afirmació siga certa és de 0 o 1.

Per exemple, si sempre apareix un exemplar nou de DAUALDEU al març, de la premissa «Estem en el mes de març» només podem concloure que aviat llegirem un nou número de DAUALDEU.



Inducció i deducció. (Font: https://www.researchgate.net/figure/Processos-de-deduccio-i-dinduccio_fig43_372027576, sota llicència CC)

I com comparem les prediccions d'una teoria amb les dades experimentals, si aquestes tenen precisió limitada i són incompletes? No podem fer servir raonaments deductius per tal de veure si la teoria l'encerta. Les probabilitats condicionades ens ajuden a interpretar les dades disponibles en cada moment, en un procés de lògica inductiva. En ciència i en la vida real mai sabem res amb certesa absoluta, i les probabilitats estan entre 0 i 1.

Fórmula de Bayes per a analitzar noves evidències

La fórmula o teorema de Bayes relaciona probabilitats condicionades:

$$P(H | E) = \frac{P(E | H) P(H)}{P(E)}$$

probabilitat
a priori

a posteriori

La fórmula de Bayes es fa servir per a calcular la probabilitat que el correu electrònic identifique correctament l'spam (correu brossa) si el missatge conté determinades paraules.

Aplicació als tests mèdics i falsos positius
Vicent Soler aplica el teorema de Bayes als tests de la COVID.

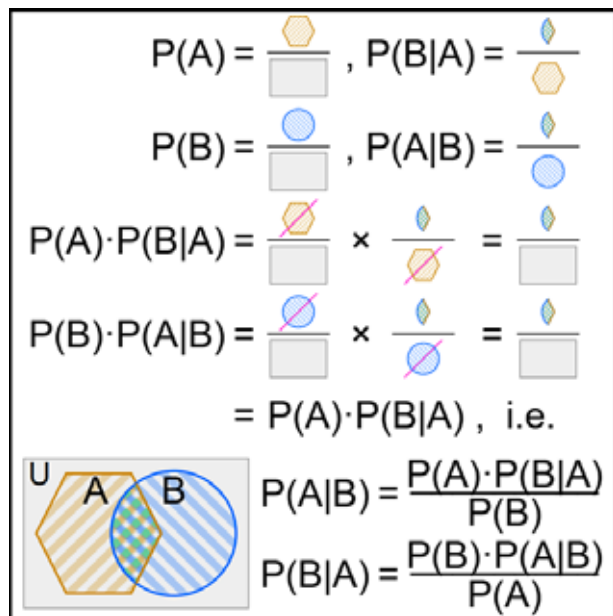
en què $P(H)$ és la probabilitat (coneguda o suposada) de la hipòtesi H , $P(E)$ és la probabilitat d'una nova evidència E , $P(H|E)$ és la probabilitat condicionada que la hipòtesi siga certa si sabem que l'evidència és certa, i anàlogament $P(E|H)$.

Per exemple, la nova evidència pot ser que un test haja resultat positiu, i aleshores la hipòtesi que volem estimar és si estem realment infectats.

La figura següent mostra gràficament com s'obtenen les probabilitats esmentades (en trobareu més detalls en la Viquipèdia).



Test de COVID. (Font: Pixabay, amb llicència)



Probabilitats com a quocient d'àrees, explicació visual. (Font: By Cmglee - Own work, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=143581610>)

Algunes aplicacions del teorema

Suposem que una persona diu la veritat 2 de cada 3 vegades. Si ara llança un dau i diu que ha tret un 4, quina és la probabilitat que el nombre obtingut siga efectivament el 4? Segons la fórmula de Bayes, aquesta probabilitat és $2/7$ (i no $2/3$, com podríem pensar). Per tant, el coneixement que la persona afirma que ha tret un 4 modifica la probabilitat que estiga dient la veritat.

El problema esmentat abans, de triar l'únic camí transitable entre tres opcions, s'anomena problema de Monty Hall i podeu trobar-ne la solució en Internet: si canviem de tria de camí, tenim el doble de possibilitats d'encertar.

La previsió del temps, així com el tractament de fenòmens atmosfèrics complexos, fan servir tècniques que barregen mètodes freqüentistes i bayesians de probabilitat.

Ací analitzarem un exemple simplificat. Suposem que un test identifica correctament el 99 % de les persones que tenen un virus. El test identificarà incorrectament com a infectats l'1 % de les persones que no tenen el virus (falsos positius). Suposem també que una persona de cada mil està infectada. La qualitat del test (99 %) i el percentatge d'infectats (0,1 %) són dades certes. La probabilitat d'estar infectats canviarà si ens fem el test una vegada, o dues.

Probabilitat d'infecció i precisió del test

Amb les dades anteriors, segons la fórmula de Bayes, la probabilitat que estiguem infectats (si el test dona positiu) és només del 9 %, no del 99 %!

Per a entendre millor aquesta conclusió tan sorprenent, imaginem que apliquem el test a 1000 persones. Només una persona tindrà el virus, i el test l'identificarà com a infectada. Tanmateix, de les altres 999 persones, l'1 % seran falsos positius: unes 10 persones seran falsament identificades com a infectades. Així, com que som una de les 11 persones que hem donat positiu i només una de nosaltres està infectada, la probabilitat que efectivament ho estiguem és $1/11 = 9 \%$.

La fórmula de Bayes s'ha d'emprar cada vegada que disposem de dades noves. Si repetim el càlcul anterior suposant que en un segon test hem tornat a donar positiu, ara la probabilitat a priori d'estar infectats és del 9 % (no 0,1 % com abans). Conclourem que la probabilitat d'estar efectivament infectats després de dos tests positius és del 91 %. La probabilitat ha augmentat bastant, però està lluny del grau de fiabilitat del test, 99 %.

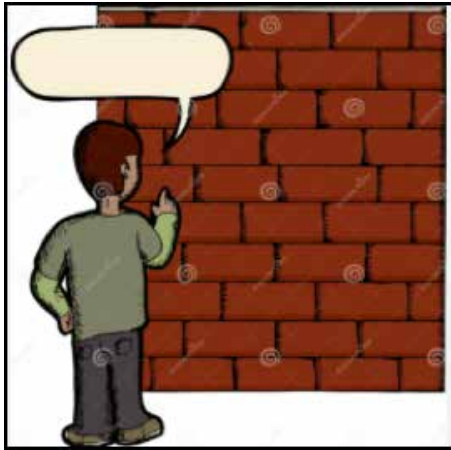
Les evidències milloren el grau de coneixement

Com acabem de veure, quantifiquen la cre-

dibilitat que assignem a determinades hipòtesis amb probabilitats, i actualitzem el grau de credibilitat quan disposem d'informació nova. Sovint es fa l'analogia amb un home que ix d'una cova on ha viscut sempre i, quan veu eixir el sol, es pregunta si és un fet extraordinari o regular. A mesura que passen els dies i el fet es repeteix, augmenta la seua confiança que així funciona el món. Anàlogament, amb cada nova evidència la fórmula de Bayes actualitza la probabilitat que la hipòtesi siga certa.

Creences i arguments

El teorema de Bayes dona resultats clars en dues situacions extremes: tant si estem totalment convençuts d'alguna cosa o de la contrària. Si la probabilitat a priori que una cosa siga certa és del 100 % o 0 %, cosa freqüent en els debats o en xarxes socials, la fórmula de Bayes indica que cap evidència o argument ens farà canviar d'opinió. No perdem el temps, doncs, tractant de convèncer els convençuts: és com parlar a un mur.



Parlar a un mur. (Font: <https://www.dreamstime.com/royalty-free-stock-photos-talking-to-brick-wall-image20693678>)

Un bayesià mai dona credibilitat zero a cap hipòtesi, per improbable que sembla, ni tampoc el 100 %, perquè sempre pot haver-hi noves evidències que modifiquen les nostres certeses.

Inferència bayesiana i progrés científic

El mètode d'anàlisi bayesià s'ha desenvolupat en els darrers cinquanta anys, però encara no ha impregnat la comunitat científica. Es fa servir per a acotar les conclusions que s'extrauen dels experiments i esbrinar fins a quin punt estan d'acord amb les prediccions teòriques. S'assignen probabilitats, a priori, a afirmacions sobre determinats esdeveniments, com ara que s'ha descobert el bosó de Higgs, o sobre paràmetres cosmològics que es volen ajustar i, a la vista de les noves evidències, s'actualitzen els nivells de credibilitat en les hipòtesis.

El mètode bayesià és consistent amb l'anomenada navalla d'Occam, un principi general que diu que les explicacions senzilles són sempre preferibles, perquè penalitza els models més complicats, i únicament els assigna probabilitats elevades quan la complexitat de les dades ho justifica.

Sobre la formació dels ciutadans

Sovint sentim que personatges públics i persones corrents es declaren incapaços d'entendre res que sone a números i fórmules. Per tal de combatre aquesta situació, els alumnes han d'eixir del batxillerat amb nocions bàsiques sobre probabilitat, incloent-hi el teorema de Bayes, i aplicacions pràctiques significatives. D'aquesta manera estaran millor equipats per a fer front a la desinformació a què estaran exposats en la vida.

1 Demostració: en la fórmula de Bayes, $P(E) = P(E|H) \cdot P(H) + P(E|-H) \cdot P(-H)$, en què -H indica hipòtesi falsa. Fem H: diu la veritat, E: trau un 4. Aleshores, $P(H) = 2/3$, $P(-H) = 1/3$, $P(E|H) = 1/6$, $P(E|-H) = 5/6$, $P(H|E) = (2/3) \cdot (1/6) / [(2/3) \cdot (1/6) + (1/3) \cdot (5/6)] = 2/7$.

2 Vicent Soler <https://www.facebook.com/vicent.soler.54/> 'BAYES', 27-8-2020.

3 Trampa bayesiana <https://www.youtube.com/watch?v=R13BD8qKeTg>

4 H: estem infectats, E: el test dona positiu. En $P(H|E)$, numerador de la fórmula de Bayes: probabilitat de donar positiu si estem infectats per la probabilitat d'estar-ho. Denominador: probabilitat total de donar positiu, és a dir, test positiu i tenir el virus sumat a test fals positiu i no tenir el virus. $P(H|E) = (0,99 \cdot 0,001) / (0,99 \cdot 0,001 + 0,01 \cdot 0,999) = 0,09 = 9\%$.



ASSOCIACIÓ
PER LA DIVULGACIÓ
DE LA CIÈNCIA
I LA TECNOLOGIA

<https://associaciomeridiazero.com>

FES-TE'N SOCI