2.6 Detecció de malalties. Fórmula de Bayes

Crec que el problema que estudiarem ens fa veure que per a prendre certes decisions no n'hi ha prou de tenir bones intencions, sinó que s'ha d'entendre una mica de números.

Suposem que hi ha una certa malaltia greu que afecta al 2 per mil de la població i que si és detectada prou a temps és curable. Suposem, a més, que es disposa d'una certa prova per a detectar la malaltia amb les característiques següents:

(i) Si una persona té la malaltia, en el 99% dels casos li detecta.

(ii) Si una persona no té la malaltia, en un 99% dels casos la prova diu que no la té.

Cal comentar que la situació descrita abans és corrent, ja que normalment les proves sempre tenen un cert error.

Un cop plantejat el problema, la qüestió és: és útil sotmetre tota la població a aquesta prova per tal de detectar els casos de malaltia? O en altres paraules: Si a una certa persona la prova li surt positiva (és a dir, la prova diu que pot tenir la malaltia), en quin tant per cent dels casos és cert que la persona en qüestió té efectivament la malaltia? La solució correcta del problema passa per la traducció d'aquest en termes de probabilitats i en l'aplicació de l'anomenada fórmula de Bayes. Aquesta solució es donarà en segon lloc. Primer donarem una explicació simplificada i potser més intuïtiva:

Suposem que la nostra població té moltes persones (per exemple 1 000 000), aleshores podem construir la taula següent, que reflecteix la situació esperada de la població. La simplificació del problema rau en la construcció d'aquesta taula, que no té per què ser correcta. Quan es diu que la malaltia afecta a un 2 per mil de la població, no vol dir que sigui quin sigui el nombre de persones que agafem, exactament 2 de cada mil seran malaltes. I el mateix pel que fa a la fiabilitat de la prova.

	Població malalta	Població no malalta	Total
Prova surt positiva	1 980	9 980	11 960
Prova surt negativa	20	988 020	988 040
Total	2 000	998 000	1 000 000

Taula 1. Distribució esperada de la població

Usant les dades de la taula tenim que, del total de persones a les quals la prova surt positiva (11 960), només 1 980 estaven malaltes, mentre que 9 980 no ho estaven.

Així tenim que només en un $\frac{1\,980}{11\,960}100 \simeq 16.56\%$ dels casos en què la prova surt positiva, la persona en qüestió té la malaltia. Aquest fet desaconsella fortament l'ús massiu de la prova de detecció. Només es aconsellable fer-la a persones que tinguin algun indici de tenir la malaltia.

En cas que la malaltia sigui molt comuna (per exemple que afecti un 10% dels individus), aleshores una prova amb les mateixes característiques sí que seria fiable (en un 91.7% dels casos en què la prova surt positiva la persona en qüestió té la malaltia).

Fem, per acabar, el mateix càlcul d'una manera rigorosa. La fórmula de Bayes serveix per a calcular la probabilitat d'un esdeveniment A sabent que se n'ha produït un altre B. Això es fa a partir de la probabilitat que es produeixi aquest últim esdeveniment B sabent que s'ha produït A. És, doncs, una fórmula d'inversió dels condicionaments. La versió més simple d'aquesta fórmula és

$$P(A/B) = \frac{P(A)P(B/A)}{P(A)P(B/A) + P(A^c)P(B/A^c)},$$

on la notació $P(A/B)$ vol dir la probabilitat que es produeixi l'esdeveniment A, sabent que s'ha produït l'esdeveniment B. Naturalment $P(A)$ vol dir la probabilitat que es produeixi l'esdeveniment A i també estem denotant per A^c el complementari o contrari de l'esdeveniment A, que és aquell esdeveniment que ocorre sempre que no ocorre A.

Si apliquem la fórmula al nostre exemple, tenim que, posant

$A =$ tenir la malaltia,
$A^c =$ no tenir la malaltia,
$B =$ donar la prova positiva,
$B^c =$ donar la prova negativa,
ens estan dient que:
$P(A) = 0.002$, d'on es dedueix que $P(A^c) = 0.998$,
$P(B/A) = 0.99$ (probabilitat que la prova doni positiva si l'individu té la malaltia),
$P(B^c/A^c) = 0.99$ d'on es dedueix que $P(B/A^c) = 0.01$, i ens demanen $P(A/B)$.

Thomas Bayes (1702-1761)

Aplicant la fórmula tenim que

$$P(A/B) = \frac{P(A)P(B/A)}{P(A)P(B/A) + P(A^c)P(B/A^c)} = \frac{0.002 \times 0.99}{0.002 \times 0.99 + 0.998 \times 0.01} \simeq 0.1656,$$

resultat que coincideix amb la forma més intuïtiva de fer-ho.