Sociedad Andaluza de Enfermedades Infecciosas

INFERENCIA ESTADISTICA SEGUN EL MODELO BAYESIANO

Ya vimos con anterioridad cómo en el modelo frecuentista el parámetro poblacional que se desea inferir (θ) es considerado una constante, mientras que se consideran aleatorios los datos obtenidos de la muestra (y), y por tanto el estadístico que se empleará para determinar el valor del parámetro poblacional. De esta forma el *intervalo de confianza* era definido como un rango de valores tal que, si repitiéramos infinidad de veces el experimento, el 95% de las estimaciones del parámetro poblacional estarían contenidas en él.

En el modelo bayesiano el planteamiento es radicalmente diferente:

- Lo que se considera constante es la información que se posee de la población, es decir, los datos (y)
- El parámetro poblacional que se desea estimar (θ) es considerado variable. En este sentido, ya no es necesaria la búsqueda de buenos estimadores, como ocurría en la estadística frecuentista.
- Para la estimación del parámetro poblacional es necesario un conocimiento previo de la distribución que pueda seguir el parámetro poblacional. Es lo que se llama distribución *a priori*.
- No se hace referencia a intervalo de confianza, sino a *intervalo de credibilidad*, que sería el intervalo que contendría al parámetro poblacional con una probabilidad del 95% (concepto muy usado, pero erróneo, para referirse al intervalo de confianza dentro de la estadística frecuentista).

Las estimaciones en el modelo de bayesiano se realizan basándose en el teorema de Bayes:

$$P(B|A) = P(A|B) \cdot P(B) / P(A)$$

Una pieza fundamental de este modelo es el conocimiento *a priori* del parámetro que se quiere inferir. En este sentido, hay que decir que si bien θ es el parámetro que se desea estimar, siempre se tiene alguna información sobre él (ya sea una información subjetiva a través de la experiencia u objetiva a través de otro estudio previo) (Figura 1); y en caso contrario, siempre se prodrían realizar experimentos seriados que proporcionarían dicho conocimiento (Figura 2).

Componentes del modelo bayesiano

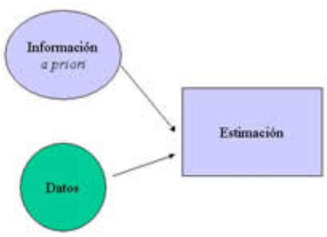

Figura 1

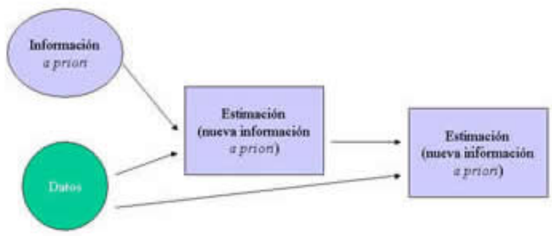

Figura 2

Para el modelo bayesiano esta experiencia previa es muy valiosa; por el contrario para el modelo frecuentista este conocimiento previo no tiene valor. Precisamente es este aspecto el más original pero también el más criticado del enfoque bayesiano: este conocimiento *a priori* podría ser "manipulado" para que los resultados del estudio sean convenientes. En este sentido, también el enfoque frecuentista tiene sus posibilidades de manipulación: es necesario prever un modelo con unos determinados errores tipo I y tipo II y predeterminar un tamaño muestral adecuado. El valor de p no es más que el resultado de un cálculo matemático directamente relacionado con el tamaño muestral: la significación estadística podría reducirse a la capacidad para aportar una muestra suficientemente grande. Por tanto el problema no se encontraría en el enfoque dado al estudio, sino en la poca honestidad de quien diseña el estudio.

Este conocimiento *a priori* se trata en resumen de una distribución de frecuencias, y como tal, con valores más frecuentes (dados por más probables por el investigador) y otros menos frecuentes (tenidos por menos probables por el investigador). La valoración de los distintos escenarios respecto a la información previa es lo que se conoce como *análisis de sensibilidad*. Dentro de esos posibles escenarios estaría aquel en el que no poseemos información previa (que se corresponde con una probabilidad a priori de 0.5 si se tratara de un suceso frente a su complementario).

Supongamos un estudio en el que se desea conocer la prevalencia de EPOC en una población, para lo que se entrevista a 170 personas. El investigador sostiene que lo más probable es que la prevalencia sea del 9%, en consonancia con los estudios poblacionales previos. Un prevalencia del 6% ó del 12% se consideran menos probables, y prevalencias del 2% ó 16% muy improbables. Se entrevistó a los 170 participantes y 10 confirmaron la presencia de EPOC (5.9%). Podría estudiarse la probabilidad de cada uno de los escenarios descritos dados los datos recogidos.

Tabla 1

Escenario	**P(datos|*θ*)† (V)**	**Prob. *a priori** * (P)**	**V x P**	**Prob. posterior (V x P / S)**	
P(5.9%|2%)	0.002	0.1	0.0002	P(2%|5.9%)	0.005
P(5.9%|6%)	0.128	0.2	0.0256	P(6%|5.9%)	0.598
P(5.9%|9%)	0.041	0.4	0.0164	P(9%|5.9%)	0.383
P(5.9%|12%)	0.003	0.2	0.0006	P(12%|5.9%)	0.014
P(5.9%|16%)	0.000	0.1	0.0000	P(16%|5.9%)	0
		Suma=1	0.0428 (S)		Suma=1

*: Se asume que la probabilidad de cada uno de estos escenarios es 0.1, 0.2, 0.4, 0.2, 0.1. La suma de estas probabilidades es 1, la probabilidad total de la distribución *a priori*. En este ejemplo, el estado no informativo sería aquel en el que se diera la misma importancia a los 5 escenarios: 0.2.

†: La verosimilitud, o distribución de los datos obtenidos dado el supuesto de un *θ* concreto, seguirá una distribución binomial (se aportan los resultados)

El investigador tendría que concluir que para su población el escenario más probable es el de una prevalencia del 6%, siendo menos probable un 9%.

Otro aspecto de interés en el enfoque bayesiano es que todo resultado de un estudio (toda información) puede ser de interés. Dentro del estricto modelo frecuentista, la ausencia de significación puede llevarnos a pensar en ausencia de interés; una p = 0.12 puede quedar sin valor alguno. El análisis de esos mismos datos desde un enfoque bayesiano y con información obtenida de estudios previos podría llevarnos a conclusiones muy diferentes.

Esquemáticamente podemos representar todo lo dicho anteriormente de la siguiente forma:

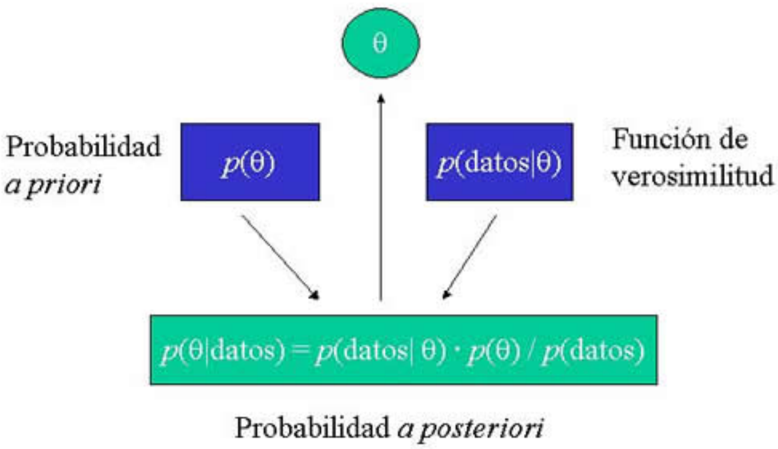

Figura 3

Deseamos realizar una estimación del parámetro poblacional θ, para lo cual se cuenta con una información *a priori* (p(θ)) y unos datos obtenidos de la población, los cuales se distribuyen de una forma determinada (p(datos|θ), *función de verosimilitud*). El Teorema de Bayes nos permite realizar dicha estimación del parámetro poblacional. Compare este esquema con la Tabla 1 del ejemplo previo.

El enfoque bayesiano presenta notables ventajas en determinados campos:

- En general, el enfoque bayesiano puede tener importantes beneficios en el campo de los ensayos clínicos:
  - Reducción del tamaño muestral, mediante el análisis contínuo de la información disponible. O en caso contrario, si la variabilidad observada es elevada, el incremento de la muestra para poder llegar a conclusiones fiables
  - Modificación del diseño del estudio durante la ejecución del mismo, mediante la adecuada planificación
  - Monitorización de ensayos clínicos: Los ensayos clínicos, desde una perspectiva clásica, son diseñados para obtener determinadas diferencias entre los grupos de comparación con un diseño predeterminado:
    - Un planteamiento de hipótesis nula y alternativa
    - Unos errores alfa y beta
    - Un tamaño muestral calculado previamente
    - En un tiempo dado

    De forma que dentro del modelo frecuentista, sólo cumpliéndose dicho diseño podríamos llegar a conclusiones con un determinado grado de incertidumbre.
    Sin embargo, en momentos intermedios entre el inicio y el final programado para el ensayo tenemos de hecho información suficiente que, dentro de un enfoque bayesiano, puede ayudarnos a tomar decisiones sobre la necesidad o no de la continuación del estudio, bien porque se evidencia un claro beneficio en uno de los grupos o por el contrario, porque se evidencia un claro perjuicio. Son los análisis intermedios.
- Meta-análisis: Es un campo ideal para el uso del enfoque bayesiano, pues se trata de acumular información diversa para llegar a una valoración de conjunto sobre un determinado problema.
- Valoración de estudios locales con la información de otros estudios. El desarrollo de grandes estudios no siempre es posible, siendo más factible el desarrollo de estudios de menor tamaño. Sin embargo, estudios con tamaño muestral suficiente aunque reducido podrían ser tenidos bajo sospecha desde un enfoque clásico. El enfoque bayesiano nos permite utilizar la información facilitada por otros estudios para ratificar o refutar con los datos de nuestra población el conocimiento que se tenga sobre el problema estudiado. Por otra parte, el enfoque bayesiano también nos permite la adaptación de modelos epidemiológicos complejos, y fundamentados en los datos de amplias poblaciones, a nuestra población mediante el uso de los datos derivados de ella.
- Análisis de decisión.

Veamos ahora algunos ejemplos:

Ejemplo 1

Estudios previos estiman que la prevalencia de la enfermedad A en una población es de 16/100.000 hab, y la prevalencia del síntoma guía X de 120/100.000 hab. Se sabe además que un 30% de los enfermos que padecen la enfermadad A presentan el síntoma guía X. ¿Cuál sería la probabilidad de que un paciente que llegara a Urgencias con el síntoma guía X padeciera la enfermedad A?

$$P(A|X) = P(X|A) \cdot P(A) / P(X) = 0.30 \cdot 0.00016 / 0.0012 = 0.04$$

... es decir, un 4% de los pacientes con el síntoma guía X padecerán realmente la enfermedad A.

Como puede verse, este enfoque (ir del síntoma a la enfermedad) es clínicamente más relevante que el que podemos encontrar en cualquier tratado médico (en el que se va de la enfermedad al síntoma).

Ejemplo 2

Se plantea a una persona un juego en el que otra persona podrá lanzar varias veces uno o dos dados (el primero desconoce cuántos dados se lanzarán, pero siempre serán uno o los dos). Se le pregunta por el número de dados que se lanzan según los datos que obtienen con cada lanzamiento.

En este caso como información *a priori* se tiene el conocimiento del dado (6 caras). Se sabe que en el caso de que se lancen dos dados, ambos dados son independientes; por tanto la probabilidad de que salga un determinado número será la siguiente según se trate de uno o dos dados:

Distribuciones *a priori*
Distribución a priori para 1 dado
p(1|1 dado) = p(2|1 dado) = ... = p(6|1 dado) = 1/6
Distribución a priori para 2 dados
p(1|2 dados) = 0 p(2|2 dados) = p(12|2 dados) = 1/6 · 1/6 = 1/36 p(3|2 dados) = p(11|2 dados) = 2 · 1/6 · 1/6 = 2/36 p(4|2 dados) = p(10|2 dados) = 3 · 1/6 · 1/6 = 3/36 p(5|2 dados) = p(9|2 dados) = 4 · 1/6 · 1/6 = 4/36 p(6|2 dados) = p(8|2 dados) = 5 · 1/6 · 1/6 = 5/36 p(7|2 dados) = 6 · 1/6 · 1/6 = 6/36

Supongamos que en la primera tirada le informan que el resultado ha sido 5, y que el jugador se inclina por que es más probable que su amigo haya tirado 2 dados, con un 60% de probabilidad.

p(5|1 dado) = 1/6; p(5|2 dados) = 4/36
p(5) = p(5|1 dado) + p(5|2 dados) = 1/6 · 0.4 + 4/36 · 0.6 = 0.13
p(1 dado|5) = (1/6 · 0.4) / p(5) = (1/6 · 0.4) / 0.13 = **0.5**
p(2 dado|5) = (4/36 · 0.6) / p(5) = (4/36 · 0.6) / 0.13 = **0.5**

... lo que indica que con el conocimiento previo de la prueba y con los datos actuales es igualmente probable que se trate de uno o dos dados. Por tanto le pide al amigo que tire nuevamente los dados. Esta vez sale un 3. Ahora empleará como probabilidad *a priori* estos 50% obtenidos como probabilidad *a posteriori* del paso previo.

p(3|1 dado) = 1/6; p(3|2 dados) = 2/36
p(3) = p(3|1 dado) + p(3|2 dados) = 1/6 · 0.5 + 2/36 · 0.5 = 0.11
p(1 dado|3) = (1/6 · 0.5) / p(3) = (1/6 · 0.5) / 0.11 = **0.75**
p(2 dado|3) = (2/36 · 0.5) / p(3) = (2/36 · 0.5) / 0.11 = **0.25**

... lo que implica tras dos tiradas que es muy probable que en vez de dos dados, como creía en un principio, su amigo esté tirando dos dados. Pero le pide que tire una tercera vez, y ahora sale un 2.

p(2|1 dado) = 1/6; p(2|2 dados) = 1/36
p(2) = p(2|1 dado) + p(2|2 dados) = 1/6 · 0.75 + 1/36 · 0.25 = 0.13
p(1 dado|2) = (1/6 · 0.75) / p(2) = (1/6 · 0.75) / 0.13 = **0.95**
p(2 dado|2) = (1/36 · 0.25) / p(2) = (1/36 · 0.25) / 0.13 = **0.05**

... lo que le lleva a asegurar que su amigo ha estado tirando un único dado.

Ejemplo 3

Vamos a ver ahora cómo es útil este enfoque en el análisis de decisión.

El Ministerio de Sanidad y Consumo tiene dudas sobre el beneficio de vacunar a la población de una infección viral banal pero con una alta tasa de absentismo laboral. Se cree, por estudios previos, que el 60% de la población está inmunizada frente a la enfermedad. Se sabe además que la prueba diagnóstica tiene el siguiente rendimiento:

Rendimiento de la prueba				
prob. de reacción [p(datos|θ)]	***Reacción***			
	-	+	++	+++
Inmune	*0.35*	*0.30*	*0.21*	*0.14*

Vulnerable	*0.09*	*0.17*	*0.25*	*0.49*

Por último se sabe que el coste de no vacunar a un sujeto y que enferme es de 20€ (el valor de la vacuna y el coste del absentismo). Las restantes situaciones se consideran sin coste.

Resultados I
Coste de cada situación

p(θ|datos)·p(θ)	***Reacción***			
	-	+	++	+++
Inmune	*0.35·0.6=0.21*	*0.30·0.6=0.18*	*0.21·0.6=0.126*	*0.14·0.6=0.084*
Vulnerable	*0.09·0.4=0.036*	*0.17·0.4=0.068*	*0.25·0.4=0.1*	*0.49·0.4=0.196*
p(datos) [suma]	*0.246*	*0.248*	*0.226*	*0.28*

Resultados II

p(θ|datos)·p(θ)/p(datos)	***Reacción***			
	-	+	++	+++
Inmune	*0.854*	*0.726*	*0.558*	*0.3*
Vulnerable	*0.146*	*0.274*	*0.442*	*0.7*

Y si aplicamos el coste de cada circunstancia podremos ver el posible coste según la probabilidad de enfermear en función del tipo de reacción de la prueba diagnóstica.

Resultados III

	Reacción			
	-	+	++	+++
Inmune y vacunado (x 8€)	*6.83€*	*5.81€*	**4.46€**	**2.40€**
Vulnerable y no vacunado (x 20€)	**2.93€**	**5.48€**	*8.85€*	*14€*
Decisión	*No vacunar*	*No vacunar*	*Vacunar*	*Vacunar*

Bibliografía:

1. Silva Ayçaguer LC, Benavides A. El enfoque bayesiano: otra manera de inferir. Gaceta sanitaria 2001;15:341-346.
2. Domenech JM. Fundamentos de diseño y estadística. Barcelona: Signo; 2001.
3. Silva Ayçaguer LC, Muñoz A. Debate sobre métodos frecuentistas vs bayesianos. Gaceta Sanitaria 2000;14:482-494.
4. Silva Ayçaguer LC, Benavides A. Apuntes sobre subjetividad y estadística en la investigación en salud. Rev Cubana Salud Pública 2003;29:170-173.

Ficha bibliográfica

Título:
Inferencia estad stica segn el modelo bayesiano
Direcci n:
http://saei.org/hemero/epidemiol/nota3.html
Descripcin:
Se describen los conceptos b sicos para poder entender el funcionamiento de la inferencia bayesiana y sus ventajas e inconvenientes frente al modelo frecuentista o clsico. Se aportan ejemplos.
Palabras clave:
estad stica bayesiana; inferencia; anlisis de decisi n, probabilidad a priori
Cdigo de idioma:
es
Autor:
Fco. Javier Caballero Granado
Editor:
Fco. Javier Caballero Granado

Google Analytics