Probabilidad bayesiana a partir de la estimación plausible de cantidades

Sergio Davis <sergio.davis@cchen.cl>

Centro de Investigación en la Intersección de Física de Plasma, Materia y Complejidad (P^{2}mc),
Comisión Chilena de Energía Nuclear

Departmento de Física, Facultad de Ciencias Exactas, Universidad Andrés Bello

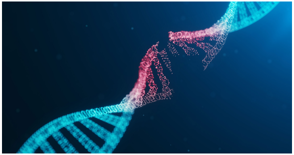
Comisión Chilena de Energía Nuclear
Ministerio de Energía

UNIVERSIDAD ANDRES BELLO

Bayes del Sur | Santiago del Estero (Argentina)

P^{2}mc

Plasma Physics, Matter and Complexity


Cristian Pavez, José Moreno, Leopoldo Soto, Biswajit Bora, Sergio Davis, Jalaj Jain
Marcelo Vásquez, Francisco Molina, Marcelo Zambra

Estudiantes y colaboradores

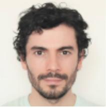
Estudiantes de Magíster y Doctorado


Constanza Farías
(U. Andrés Bello)


Vivianne Olguín
(U. de Chile)


Abiam Tamburrini
(U. de Chile)


Boris Maulén
(U. Andrés Bello)


Leonardo Herrera
(U. Andrés Bello)

Colaboradores


Gonzalo Gutiérrez
(U. de Chile)


Yasmín Navarrete
(IFICC)

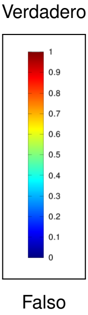

Joaquín Peralta
(U. Andrés Bello)


Claudia Loyola
(U. Andrés Bello)


Pablo Moya
(U. de Chile)


Gonzalo Avaria
(UTFSM)

Motivación: el origen de la probabilidad bayesiana



La probabilidad $P(A|I)$ de una *afirmación* A en un *contexto* I es un número entre 0 y 1 que representa el **grado de creencia plausible** en A según la información en I.

Para A, B, I afirmaciones cualesquiera debe cumplirse la **regla de la suma**,

$$P(\text{not } A|I) = 1 - P(A|I), \quad (1)$$

y la **regla del producto**,

$$P(A \text{ and } B|I) = P(A|I) \cdot P(B|A \text{ and } I). \quad (2)$$

Las reglas (1) y (2) pueden tomarse como axiomas, pero también pueden deducirse a partir de axiomas más fundamentales.

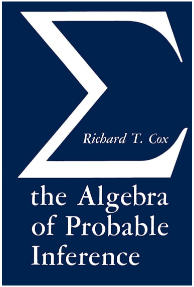
Richard T. Cox: Una axiomática bayesiana

(1) La probabilidad $P(\text{not } A|I)$ es una función de $P(A|I)$:

$$P(\text{not } A|I) = \mathcal{S}\Big(P(A|I)\Big)$$

(2) La probabilidad $P(A \text{ and } B|I)$ es una función de $P(A|I)$ y $P(B|A \text{ and } I)$:

$$P(A \text{ and } B|I) = \mathcal{F}\Big(P(A|I),\;\; P(B|A \text{ and } I)\Big)$$



Las únicas* funciones $\mathcal{S}$ y $\mathcal{F}$ consistentes con la *lógica proposicional* son

$$\boxed{\mathcal{S}(x) = 1 - x} \qquad \boxed{\mathcal{F}(x,y) = x \cdot y}$$

de forma que recuperamos la regla de la suma y del producto,

$$P(\text{not } A|I) = 1 - P(A|I),$$
$$P(A \text{ and } B|I) = P(A|I) \cdot P(B|A \text{ and } I).$$

*R. T. Cox. Am. J. Phys. **14**, 1 (1946) | R. T. Cox. *The algebra of probable inference*, John Hopkins Press (1961).

La demostración de Cox es un *tour de force*

writing $F_{11}(p,q)$ for $\partial F_1(p,q)/\partial p$, and similarly representing the other second derivatives, we obtain

$$F_{11}(u,z)(\partial u/\partial x)(\partial u/\partial y)+F_1(u,z)\partial^2 u/\partial x\partial y = F_{12}(x,v)\partial v/\partial y, \quad (22)$$

$$F_{11}(u,z)(\partial u/\partial y)^2+F_1(u,z)\partial^2 u/\partial y^2 = F_{22}(x,v)(\partial v/\partial y)^2+F_2(x,v)\partial^2 v/\partial y^2, \quad (23)$$

$$F_{12}(u,z)\partial u/\partial y = F_{22}(x,v)(\partial v/\partial y)(\partial v/\partial z) + F_2(x,v)\partial^2 v/\partial y\partial z. \quad (24)$$

Differentiating Eq. (19) with respect to z, or Eq. (21) with respect to x, we obtain

$$F_{12}(u,z)\partial u/\partial x = F_{12}(x,v)\partial v/\partial z. \quad (25)$$

Among Eqs. (20), (22), $\cdots$, (25) we can now eliminate the functions of u and v other than their derivatives. Thus, eliminating $F_{12}(u,z)$ and $F_{12}(x,v)$ among Eqs. (22), (24) and (25), we find

$$[F_{11}(u,z)(\partial u/\partial y)^2 - F_{22}(x,v)(\partial v/\partial y)^2](\partial u/\partial x)(\partial v/\partial z) = F_2(x,v)(\partial^2 v/\partial y\partial z)(\partial v/\partial y)(\partial u/\partial x) - F_1(u,z)(\partial^2 u/\partial x\partial y)(\partial u/\partial y)(\partial v/\partial z)$$

Combining this with Eq. (23), we can eliminate $F_{11}(u,z)$ and $F_{22}(x,v)$ together, obtaining

$$F_1(u,z)(\partial v/\partial z)[(\partial^2 u/\partial y^2)(\partial u/\partial x) - (\partial^2 u/\partial x\partial y)(\partial u/\partial y)] = F_2(x,v)(\partial u/\partial x)[(\partial^2 v/\partial y^2)(\partial v/\partial z) - (\partial^2 v/\partial y\partial z)(\partial v/\partial y)].$$

Combining this with Eq. (20), we eliminate $F_1(u,z)$ and $F_2(x,v)$ together and obtain

$$\frac{\partial^2 u/\partial x\partial y}{\partial u/\partial x} - \frac{\partial^2 u/\partial y^2}{\partial u/\partial y} = \frac{\partial^2 v/\partial y\partial z}{\partial v/\partial z} - \frac{\partial^2 v/\partial y^2}{\partial v/\partial y}.$$

This may be written in the form,

$$\frac{\partial}{\partial y}\ln\left(\frac{\partial u/\partial x}{\partial u/\partial y}\right) = -\frac{\partial}{\partial y}\ln\left(\frac{\partial v/\partial y}{\partial v/\partial z}\right).$$

we obtain

$$\frac{\partial}{\partial x}\ln\left[\frac{F_1(x,y)}{F_2(x,y)}\right]dx + \frac{\partial}{\partial y}\ln\left[\frac{F_1(x,y)}{F_2(x,y)}\right]dy = -d\ln\Phi(x) + d\ln\Phi(y).$$

The left-hand member being now a complete differential, we may integrate and so find

$$F_1(x,y)/F_2(x,y) = h\Phi(y)/\Phi(x), \quad (29)$$

where h is a constant of integration.

To make use of this result, we divide Eq. (20) by Eq. (21), obtaining

$$\frac{F_1(u,z)}{F_2(u,z)}\frac{\partial u}{\partial y} = \frac{\partial v/\partial y}{\partial v/\partial z}.$$

The right-hand member is simply $F_1(y,z)/F_2(y,z)$, and, with the aid of Eq. (29), the equation may be written as

$$\frac{\Phi(z)}{\Phi(u)}\frac{\partial u}{\partial y} = \frac{\Phi(z)}{\Phi(y)}.$$

Replacing in this equation u by its value $F(x,y)$ we have

$$\partial F(x,y)/\partial y = \Phi[F(x,y)]/\Phi(y). \quad (30)$$

Similarly, from Eqs. (19) and (20) we obtain

$$\partial F(y,z)/\partial y = \Phi[F(y,z)]/\Phi(y),$$

which becomes, when x and y are written for y and z,

$$\partial F(x,y)/\partial x = \Phi[F(x,y)]/\Phi(x). \quad (31)$$

Combining Eqs. (30) and (31) to obtain the differential dF (the variables being understood as x and y) we find

$$dF/\Phi(F) = dx/\Phi(x) + dy/\Phi(y).$$

If we denote $\int[dp/\Phi(p)]$ by $\ln f(p)$, we obtain, by integrating and taking the exponentials of both members of this equation,

$$Cf(F) = f(x)f(y),$$

where C is a constant of integration. This then is the solu-

¿Es posible otra base para las ideas bayesianas?

- Para deducir las funciones $\mathcal{S}(x) = 1 - x$ y $\mathcal{F}(x, y) = x \cdot y$ de Cox se necesita resolver ecuaciones funcionales.
- No es claro por qué $P(A \text{ and } B|I)$ debe depender únicamente de $P(A|I)$ y $P(B|A \text{ and } I)$. ¿Por qué no de $P(B|I)$ o de $P(A \text{ or } B|I)$?

> En este trabajo mostraremos que es posible comenzar a partir de otro concepto, el de **estimación**, y llegar a la teoría bayesiana usual, donde las probabilidades son objetos que surgen a partir de las estimaciones.

Estimación plausible de cantidades desconocidas

Estimación: Idea aproximada del valor de una cantidad dada cierta información *incompleta*.




Escribiremos la estimación de una cantidad X usando la información I por medio de la notación $\langle X \rangle_I$

Postulados para la estimación (1)

Vamos a imponer **cuatro requerimientos** o postulados que definirán la operación estimación de manera única.

Estos requerimientos **extienden** ciertas propiedades de la **evaluación de funciones**, por ejemplo:

$$\Big(X^2+3Y\Big)_{X=3,Y=2}=3^2+3\cdot 2=15.$$

Postulado 1: Aditividad

$$\big\langle X+Y\big\rangle_I=\big\langle X\big\rangle_I+\big\langle Y\big\rangle_I \quad \text{para todo } I$$

Es decir, la estimación de una suma de cantidades es la suma de las estimaciones por separado.

Postulados para la estimación (2)

Si transformamos linealmente una cantidad X usando parámetros a y b constantes, tenemos el siguiente postulado.

Postulado 2: Transformación lineal constante

$$\big\langle aX+b\big\rangle_I = a\big\langle X\big\rangle_I + b \quad \text{para } a \text{ y } b \text{ constantes bajo } I$$

En otras palabras,

"Estimar la cantidad transformada es equivalente a estimar la cantidad original y transformar el resultado".

Consideremos dos casos:

$a=0$: En este caso se tiene que $\big\langle b\big\rangle_I = b$, es decir, la estimación de una constante es la propia constante.

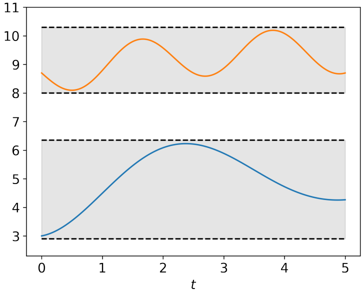
$b=0$: Ahora en cambio se tiene $\big\langle aX\big\rangle_I = a\big\langle X\big\rangle_I$, esto es, un factor constante puede extraerse de la estimación.

Postulados para la estimación (3)

Si dos cantidades están ordenadas (una siempre es mayor que la otra), entonces las estimaciones de dichas cantidades deben reflejar ese orden.

Postulado 3: Conservación del orden

$$\big\langle X\big\rangle_I > \big\langle Y\big\rangle_I \quad \text{si } X > Y \text{ bajo } I$$



Consecuencia de esto es que si $a \leq X \leq b$, debe cumplirse que

$$a \leq \big\langle X\big\rangle_I \leq b.$$

En particular, la estimación de una cantidad positiva siempre será positiva.

Postulados para la estimación (4)

Consideremos que pasa con la **evaluación parcial** de una función:

$$\Big(X^2+3Y\Big)_{X=3} = 9+3Y,$$
$$\Big(9+3Y\Big)_{Y=2} = 9+3\cdot 2 = 15,$$

que es lo mismo que evaluar simultáneamente en X y en Y,

$$\Big(X^2+3Y\Big)_{X=3,Y=2} = 3^2+3\cdot 2 = 15.$$

Postulado 4: Doble estimación

Si $Y_p(x) := \big\langle Y \big\rangle_{X=x,I}$ entonces $\big\langle Y_p \big\rangle_I = \big\langle Y \big\rangle_I$ para todo I

En otras palabras,

"La estimación de una estimación parcial es la estimación total".

Funciones indicador

Una función indicador lleva los *valores de verdad* a valores enteros $\in \{0, 1\}$.

$$\mathrm{Q}(A) = \begin{cases} 1 \text{ si } A \text{ es verdadero,} \\ 0 \text{ si } A \text{ es falso.} \end{cases}$$

Las tablas de verdad para not , and y or pueden escribirse como

$\mathrm{Q}(A)$	$\mathrm{Q}(\text{not } A)$
1	0
0	1

$\mathrm{Q}(A)$	$\mathrm{Q}(B)$	$\mathrm{Q}(A \text{ and } B)$	$\mathrm{Q}(A \text{ or } B)$
0	0	0	0
0	1	0	1
1	0	0	1
1	1	1	1

Estas tablas pueden resumirse en las siguientes identidades aritméticas,

$$\mathrm{Q}(\text{not } A) = 1 - \mathrm{Q}(A), \tag{3a}$$

$$\mathrm{Q}(A \text{ and } B) = \mathrm{Q}(A) \cdot \mathrm{Q}(B), \tag{3b}$$

$$\mathrm{Q}(A \text{ or } B) = \mathrm{Q}(A) + \mathrm{Q}(B) - \mathrm{Q}(A) \cdot \mathrm{Q}(B). \tag{3c}$$

¿Conexión con las probabilidades?

$$\mathrm{Q}(\text{not } A) = 1 - \mathrm{Q}(A)$$

versus

$$P(\text{not } A|I) = 1 - P(A|I)$$

$$\mathrm{Q}(A \text{ and } B) = \mathrm{Q}(A) \cdot \mathrm{Q}(B)$$

versus

$$P(A \text{ and } B|I) = P(A|I) \cdot P(B|A \text{ and } I)$$

$$\mathrm{Q}(A \text{ or } B) = \mathrm{Q}(A) + \mathrm{Q}(B) - \mathrm{Q}(A) \cdot \mathrm{Q}(B)$$

versus

$$P(A \text{ or } B|I) = P(A|I) + P(B|I) - P(A \text{ and } B|I)$$

El significado de la estimación

Para una variable X que toma valores en $\{x_1, \ldots, x_n\}$, se cumple que

$$\sum_{i=1}^{n} \mathrm{Q}(X = x_i) = 1, \tag{4}$$

y de la misma forma, para una función cualquiera f de X, se tiene

$$f(X) = \sum_{i=1}^{n} f(x_i)\mathrm{Q}(X = x_i). \tag{5}$$

Con esto podemos escribir la estimación de la función f como

$$\langle f \rangle_I = \left\langle \sum_{i=1}^{n} f(x_i)\mathrm{Q}(X = x_i) \right\rangle_I = \sum_{i=1}^{n} f(x_i)\langle \mathrm{Q}(X = x_i)\rangle_I = \sum_{i=1}^{n} f(x_i)p_i. \tag{6}$$

Vemos que $\langle f \rangle_I$ tiene la forma de un **valor de expectación**, donde la cantidad $p_i := \langle \mathrm{Q}(X = x_i)\rangle_I$ juega el rol de la probabilidad $P(X = x_i|I)$.

Probabilidad a partir de la estimación

Esto sugiere definir la probabilidad de una afirmación A en el estado I como

$$\boxed{P(A|I) := \big\langle \mathrm{Q}(A) \big\rangle_I}$$

Ya que $\mathrm{Q}(A) \in \{0, 1\}$, gracias al postulado de *conservación del orden* se garantiza que

$$0 \leq P(A|I) \leq 1. \tag{7}$$

Además, como $\sum_{i=1}^{n} \mathrm{Q}(X = x_i) = 1$, automáticamente se tiene que

$$\sum_{i=1}^{n} P(X = x_i|I) = \sum_{i=1}^{n} \big\langle \mathrm{Q}(X = x_i) \big\rangle_I = \left\langle \sum_{i=1}^{n} \mathrm{Q}(X = x_i) \right\rangle_I = \big\langle 1 \big\rangle_I = 1. \tag{8}$$

Sólo resta mostrar que la estimación de una función indicador sigue las reglas de la suma y del producto.

Derivación de la regla de la suma

Comenzamos con la identidad de la negación en términos de funciones indicador,

$$\mathrm{Q}(\text{not } A) = 1 - \mathrm{Q}(A), \tag{9}$$

y aplicamos estimación a ambos lados en el estado de conocimiento I,

$$\Big\langle \mathrm{Q}(\text{not } A) \Big\rangle_I = \Big\langle 1 - \mathrm{Q}(A) \Big\rangle_I = \big\langle 1 \big\rangle_I - \big\langle \mathrm{Q}(A) \big\rangle_I = 1 - \big\langle \mathrm{Q}(A) \big\rangle_I \tag{10}$$

Con esto, y usando $P(A|I) := \big\langle \mathrm{Q}(A) \big\rangle_I$, hemos llegado a la regla de la suma,

$$\boxed{P(\text{not } A|I) = 1 - P(A|I)} \tag{11}$$

Derivación de la regla del producto (1)

Para la regla del producto, comenzamos usando la identidad para la conjunción,

$$\mathrm{Q}(A \text{ and } B) = \mathrm{Q}(A) \cdot \mathrm{Q}(B) \tag{12}$$

a la cual aplicamos estimación a ambos lados bajo I,

$$\Big\langle \mathrm{Q}(A \text{ and } B) \Big\rangle_I = \Big\langle \mathrm{Q}(A) \cdot \mathrm{Q}(B) \Big\rangle_I. \tag{13}$$

Definiendo las variables

$$a := \mathrm{Q}(A), \;\; b := \mathrm{Q}(B), \tag{14}$$

tal que $a, b \in \{0, 1\}$, tenemos

$$\big\langle \mathrm{Q}(A \text{ and } B) \big\rangle_I = \big\langle a \cdot b \big\rangle_I = \Big\langle \big\langle a \cdot b \big\rangle_{a,I} \Big\rangle_I = \Big\langle a \cdot \big\langle b \big\rangle_{a,I} \Big\rangle_I \tag{15}$$

Derivación de la regla del producto (2)

Usando el lema: $a \cdot f(a) = a \cdot f(1)$ para una variable $a \in \{0, 1\}$, vemos que

a	$a \cdot f(a)$	$a \cdot f(1)$
0	0	0
1	$f(1)$	$f(1)$

$$a \cdot \big\langle b \big\rangle_{a,I} = a \cdot \big\langle b \big\rangle_{a=1,I} = a \cdot \big\langle b \big\rangle_{A \text{ and } I} \tag{16}$$

Por lo tanto,

$$\big\langle \mathrm{Q}(A \text{ and } B) \big\rangle_I = \Big\langle a \cdot \underbrace{\big\langle b \big\rangle_{A \text{ and } I}}_{=\text{constante}} \Big\rangle_I = \big\langle a \big\rangle_I \cdot \big\langle b \big\rangle_{A \text{ and } I} \tag{17}$$

es decir, recordando que $a = \mathrm{Q}(A)$; $b = \mathrm{Q}(B)$, podemos escribir

$$\big\langle \mathrm{Q}(A \text{ and } B) \big\rangle_I = \big\langle \mathrm{Q}(A) \big\rangle_I \cdot \big\langle \mathrm{Q}(B) \big\rangle_{A \text{ and } I} \tag{18}$$

que es la regla del producto,

$$\boxed{P(A \text{ and } B|I) = P(A|I) \cdot P(B|A \text{ and } I)} \tag{19}$$

Bonus: proyección y el teorema de Bayes

Identidad de proyección:

$$\boxed{\Big\langle \omega \cdot \mathrm{Q}(E) \Big\rangle_I = \langle \omega \rangle_{E \text{ and } I} \cdot \langle \mathrm{Q}(E) \rangle_I}$$

para cualquier cantidad ω y afirmación E.

Con esta identidad podemos estimar cantidades en $(E \text{ and } I)$ si sabemos estimar cantidades en I.

Demostración:

Usamos $\langle \omega \cdot e \rangle_{e,I} = \langle \omega \rangle_{e,I} \cdot e = \langle \omega \rangle_{E \text{ and } I} \cdot e \quad // \; \langle \bullet \rangle_I$ donde $e := \mathrm{Q}(E)$.

Usando $\omega = \mathrm{Q}(T)$ y reconociendo que

$$\Big\langle \mathrm{Q}(T) \Big\rangle_{E \text{ and } I} = P(T|E \text{ and } I), \qquad \Big\langle \mathrm{Q}(T)\mathrm{Q}(E) \Big\rangle_I = P(T \text{ and } E|I),$$

vemos que la identidad de proyección contiene al teorema de Bayes,

$$P(T|E \text{ and } I) = \frac{P(T \text{ and } E|I)}{P(E|I)} = \frac{P(T|I) \cdot P(E|T \text{ and } I)}{P(E|I)}. \tag{20}$$

Conclusiones

- Es posible usar la idea de estimación de cantidades como la base de una teoría bayesiana.
- La idea de estimación tiene cierta "intuición física".
- Se necesitan cuatro postulados para la estimación, la cual resulta equivalente al valor esperado en la teoría bayesiana.
- ¡No se requiere resolver ecuaciones funcionales!

Agradecemos financiamiento de ANID
(Agencia Nacional de Investigación y Desarrollo, Chile)
Proyecto FONDECYT Regular 1220651

Para más información, ver el libro gratuito *"Estimación, probabilidad e inferencia"*
`https://sdavis.cl/epi/epi.pdf`