Distribuciones bivariadas y regla de Bayes

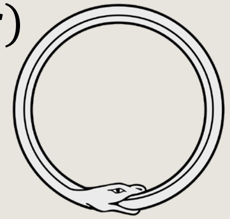
Dr. Héctor Nájera
Dr. Curtis Huffman

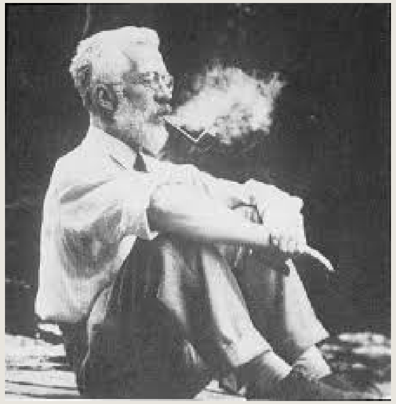
23/02/2023

De la sesión anterior

- ¿A qué nos referimos con "probabilidad" cuando lo aplicamos a la "vida real"/ investigación empírica?
- Para **calcular** medidas de probabilidad usando la teoría de la probabilidad (Kolmogorov) en la "vida real" (investigación empírica) es necesario identificar el espacio de eventos con suficiente precisión.

$$P(A) = \frac{M}{N} = \frac{(\boldsymbol{Numero}\ de\ casos\ favorables\ a\ A)}{(\boldsymbol{Número}\ total\ de\ casos\ \boldsymbol{igualmente\ possibles})}$$





- Notábamos el papel que juega la asignación aleatoria/justa/equiprobable/sin-información-asociada en la propuesta de Fisher.

¿Dentro o fuera de nuestra cabeza?

- No hay una sola aplicación de la teoría de la probabilidad en la que se pueda evadir ese importantísimo primer paso: **asignar algún valor inicial numérico de probabilidades para arrancar con los cálculos**.
- ¿Bajo cuál razonamiento se llega a ese argumento inicial?, y si es puesto en tela de juicio, ¿cómo podemos defenderlo?

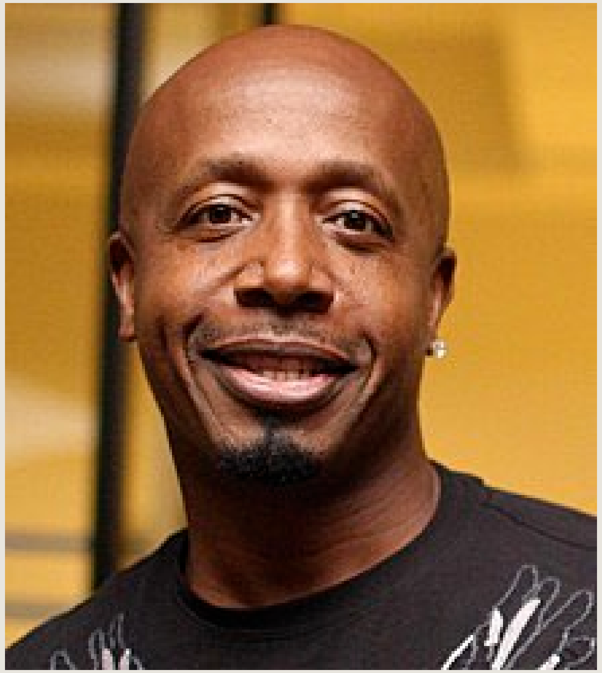
- Principio de razón insuficiente/indiferencia. (Bernoulli, J., 1713)
  - Reconocemos que una asignación probabilística es sólo un medio para describir un estado del conocimiento
  - Si la evidencia no nos da razones para considerar al enunciado A como más probable que el enunciado B, entonces la única manera honesta de describir tal estado de conocimiento es asignarles probabilidades iguales.

Jakob Bernoulli
1655-1705

¿Dentro o fuera de nuestra cabeza?

When you measure, include the measurer.

Burrell, K. S. [@MCHammer]. (2021, febrero 22). You bore us. If science is a "commitment to truth" shall we site all the historical non-truths perpetuated by scientists ? Of course not. It's not science vs Philosophy ... It's Science + Philosophy. Elevate your Thinking and Consciousness. When you measure include the measurer. [Tweet]. Twitter. https://twitter.com/MCHammer/status/1363908982289559553



Stanley Kirk Burrell
1962

Probabilidad condicional e independencia

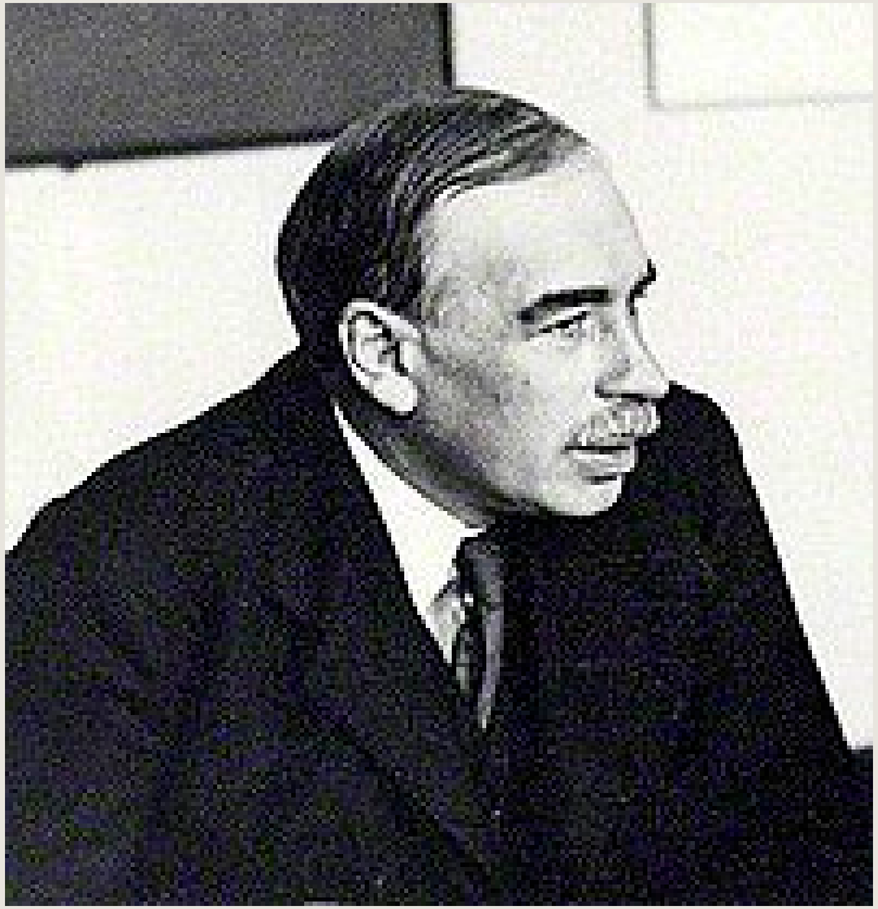
DISTRIBUCIONES BIVARIADAS

OK! Pero…

- La probabilidad de que un dado caiga en cierta cara es interesante
- La probabilidad es más interesante cuando un evento (hipótesis) es condicional a otro (datos)
  - La probabilidad de que llueva en un día en agosto condicional a que amaneció nublado!
  - La probabilidad de tener COVID-19 dado que el test fue positivo
  - La probabilidad de tener COVID-19 dado que el test fue positivo (conociendo un a priori del error del test)
  - La probabilidad de que el efecto β de que una vacuna reduzca el riesgo de fallecimiento de COVID-19 condicional en los datos de la prueba fase tres (cierto a priori sobre el efecto esperado)
  - La probabilidad de que el efecto β de CO2 sobre la temperature promedio anual sea positiva condicional en la serie de tiempo de NASA (y un prior informativo a partir de estudios de laboratorio)

Probabilidad condicional

- All probabilities are conditional probabilities. Keynes, 1921



John Maynard Keynes
1883-1942

Probabilidad condicional

La probabilidad de cierto resultado dado que otro resultado es correcto

p(h|e)=Probabilidad de cierto color de cabello dado el color de ojos azules

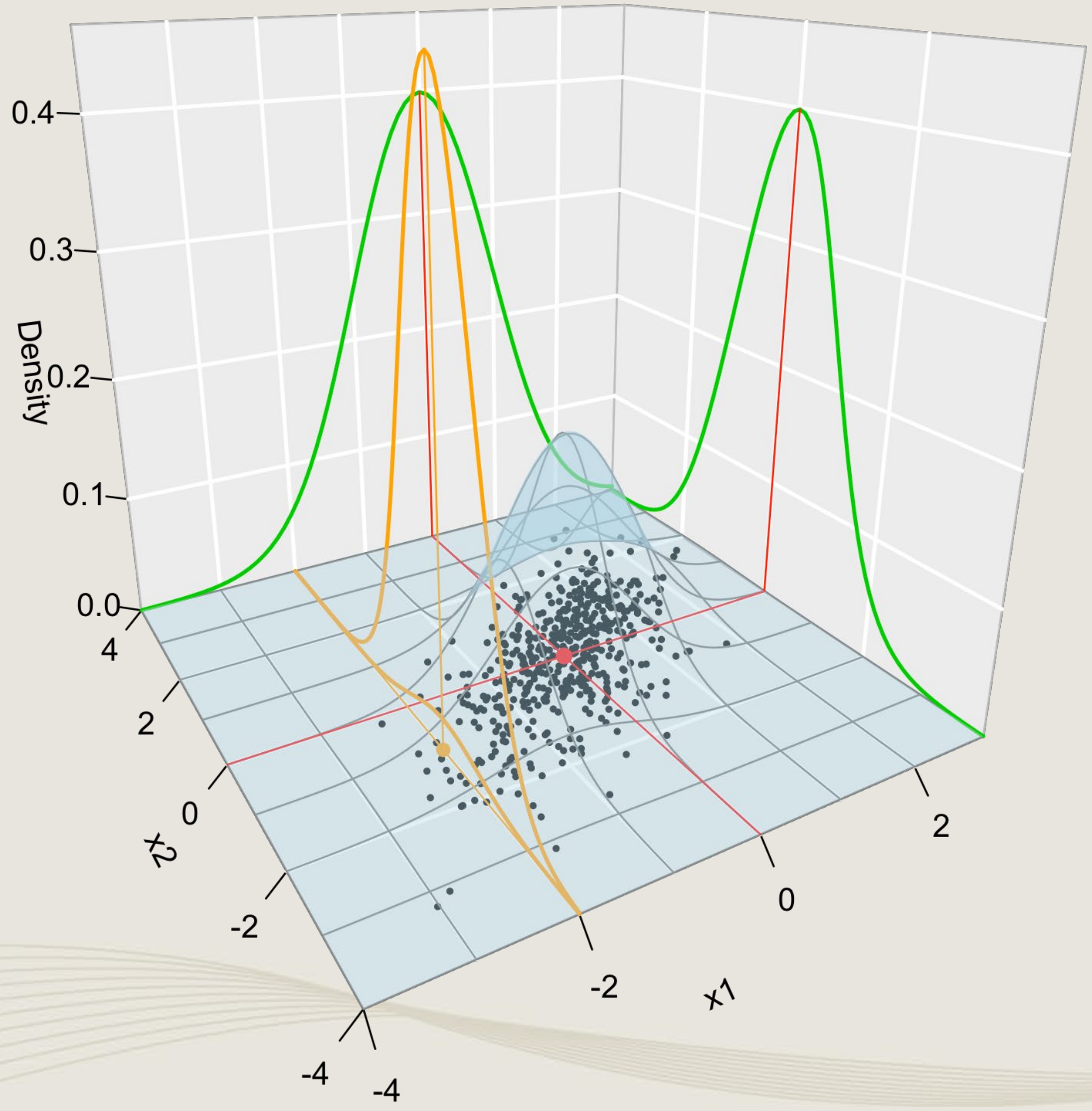
Table 4.2: Example of conditional probability. Of the blue-eyed people in Table 4.1, what proportion have hair color h? Each cell shows $p(h|\text{blue}) = p(\text{blue}, h)/p(\text{blue})$ rounded to two decimal points. Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier.

Eye Color	Hair Color: Black	Brunette	Red	Blond	Marginal (Eye Color)
Blue	.03/.36 = .08	.14/.36 = .39	.03/.36 = .08	.16/.36 = .45	.36/.36 = 1.0

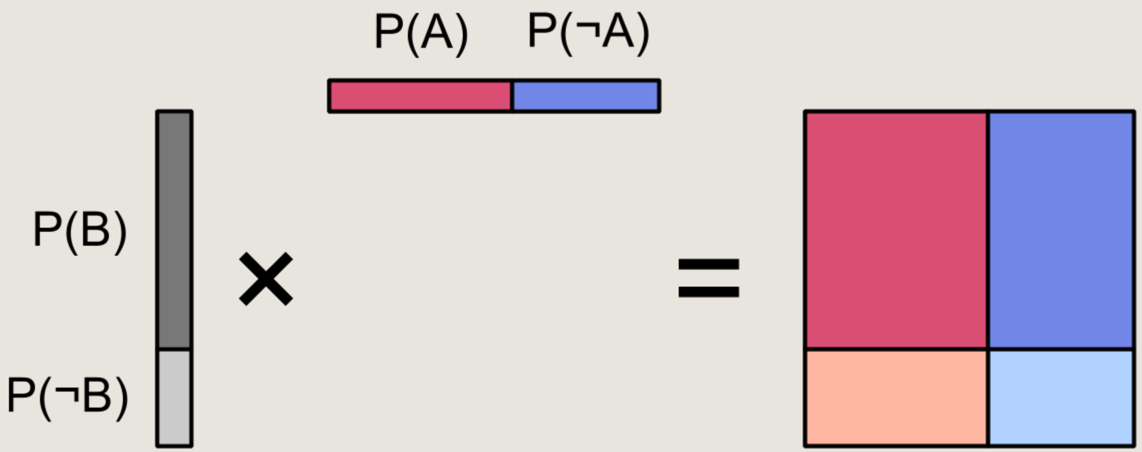
Veremos que es equivalente a la probabilidad de que:

P(β|D)

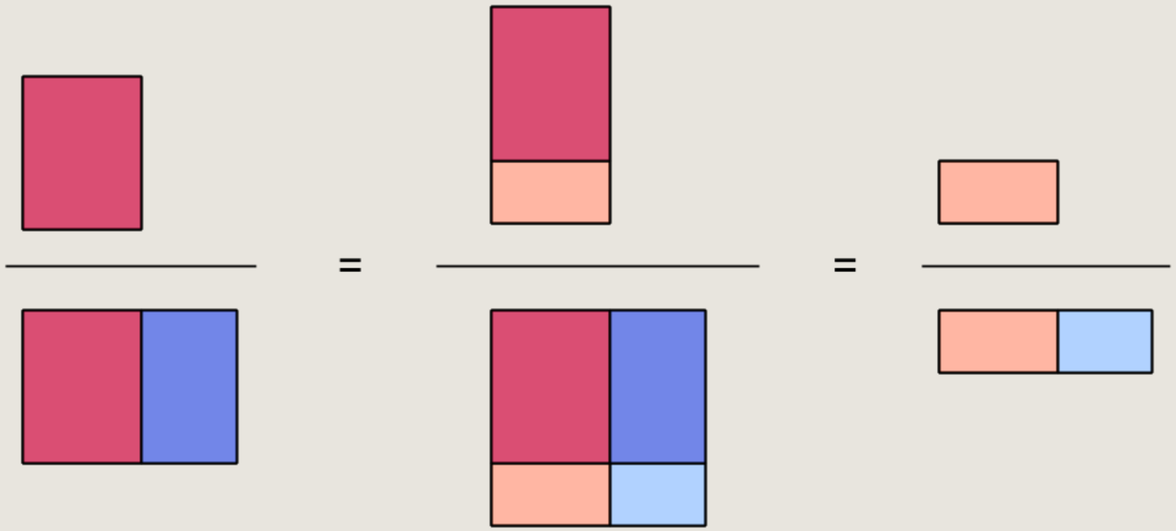
Probabilidad condicional



Independencia

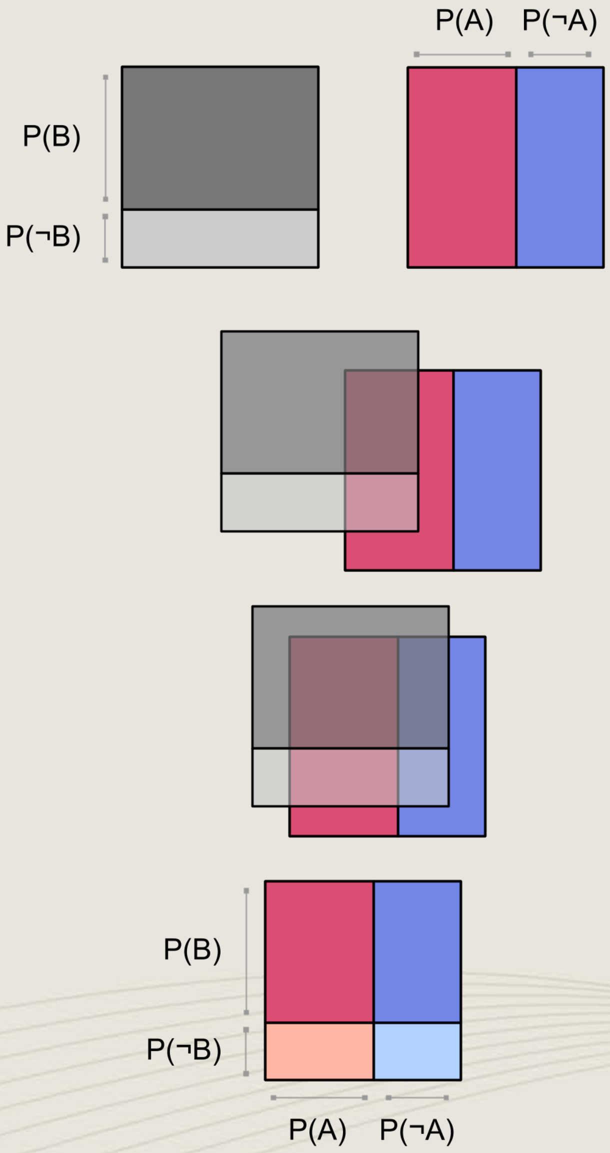


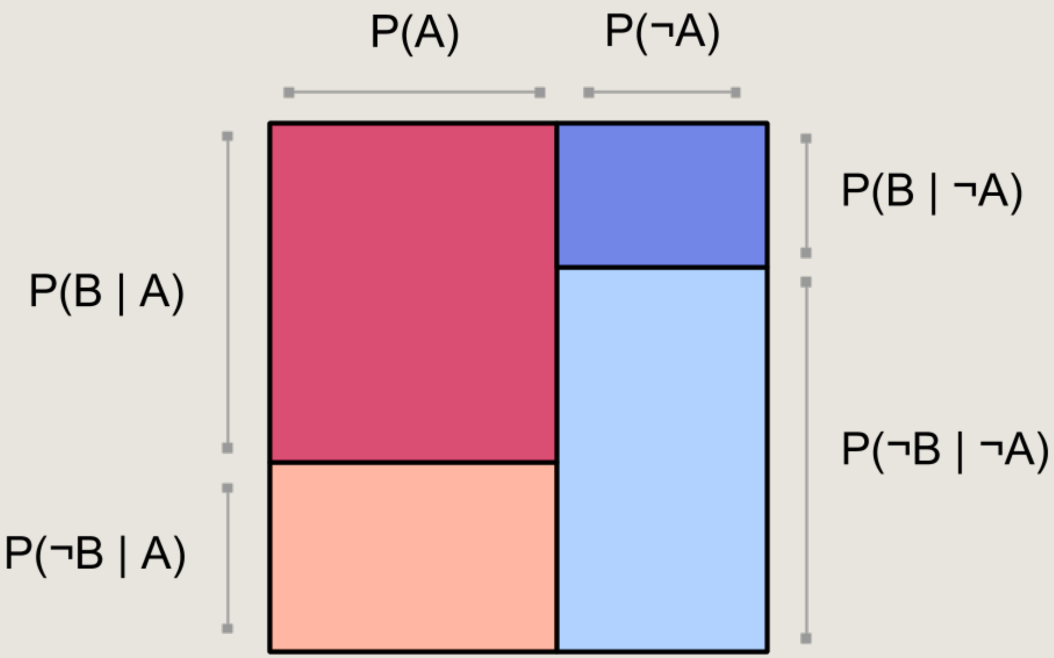
$$P(A \mid B) = P(A) = P(A \mid \neg B)$$



$$\frac{P(A, B)}{P(B)} = \frac{P(A, B) + P(A, \neg B)}{P(B) + P(\neg B)} = \frac{P(A, \neg B)}{P(\neg B)}$$

No-independencia





Capítulo 5. Teorema de Bayes

- El capítulo 5 se enfoca en como las probabilidades marginales se relacionan con las probabilidades condicionales

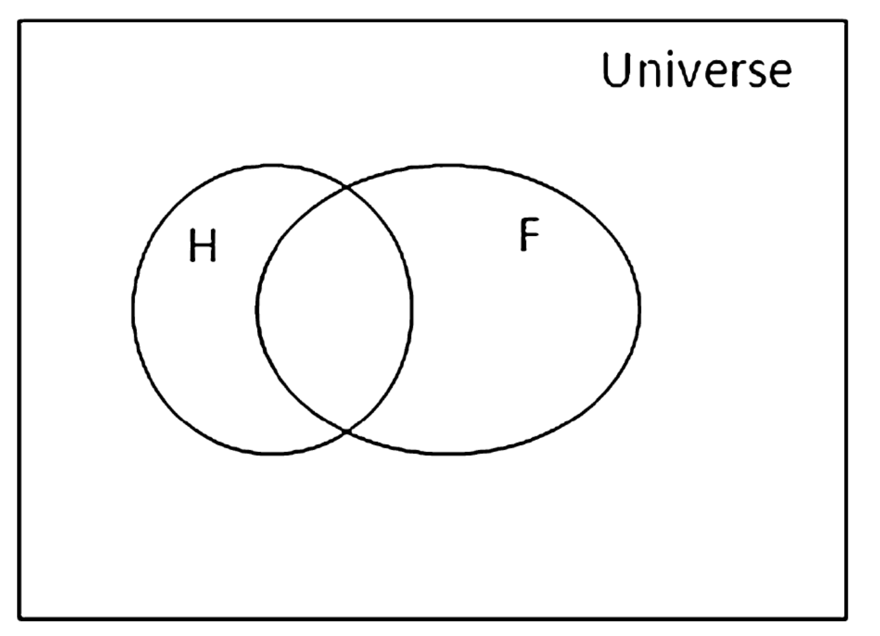
Ideas centrales:

  - Definición formal de probabilidad condicional
  - Aplicación del teorema de Bayes a parámetros y datos
  - Por qué la inferencia bayesiana es difícil

Teorema de Bayes

El objetivo típico consiste en determinar la probabilidad de una hipótesis (H) dados los datos (F; findings)

$$p(H|F) = \frac{p(F|H)\,p(H)}{p(F)}$$

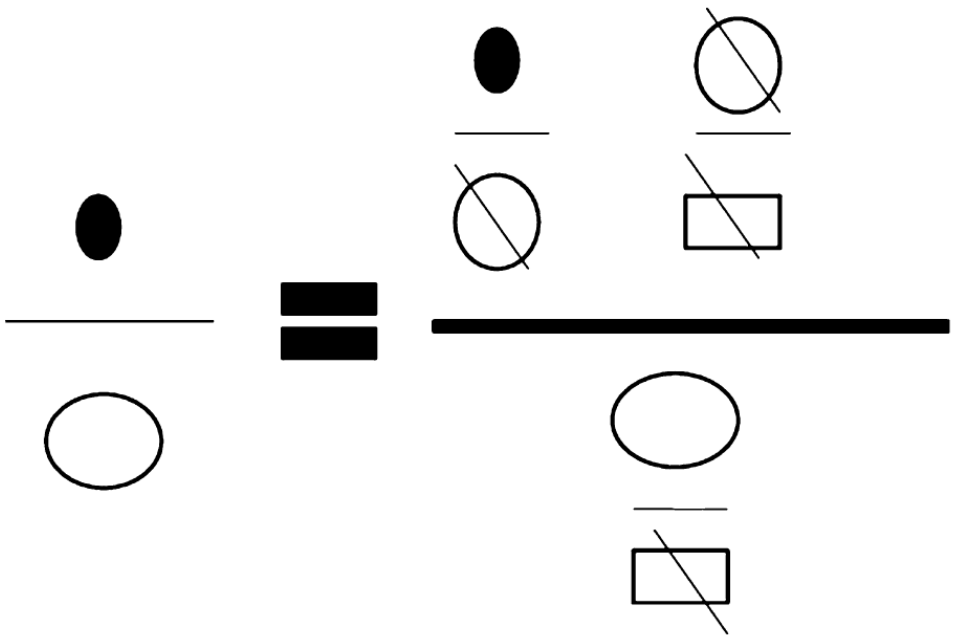


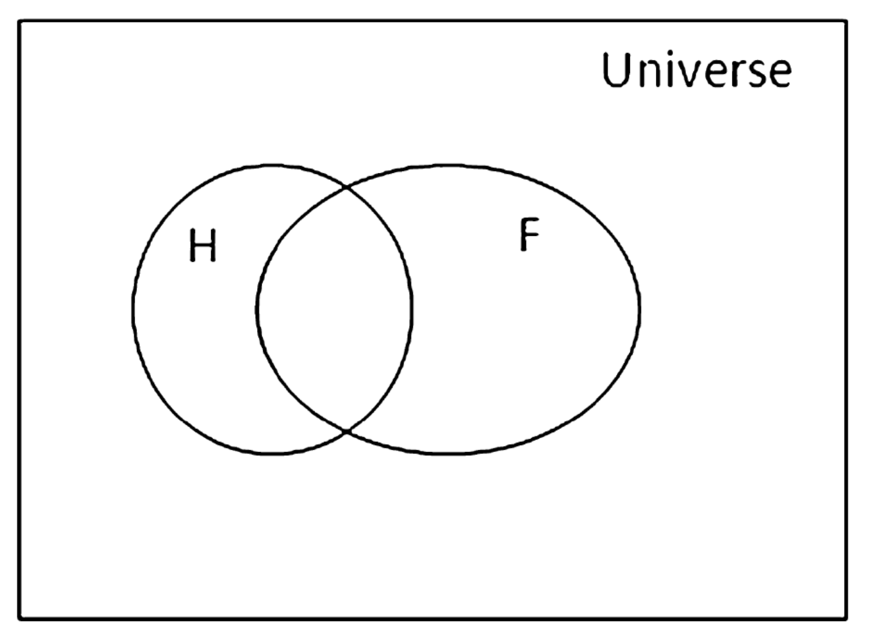
Trafimow, D. (2011).

Teorema de Bayes

El objetivo típico consiste en determinar la probabilidad de una hipótesis (H) dados los datos (F; findings)

$$p(H|F) = \frac{p(F|H)p(H)}{p(F)}$$



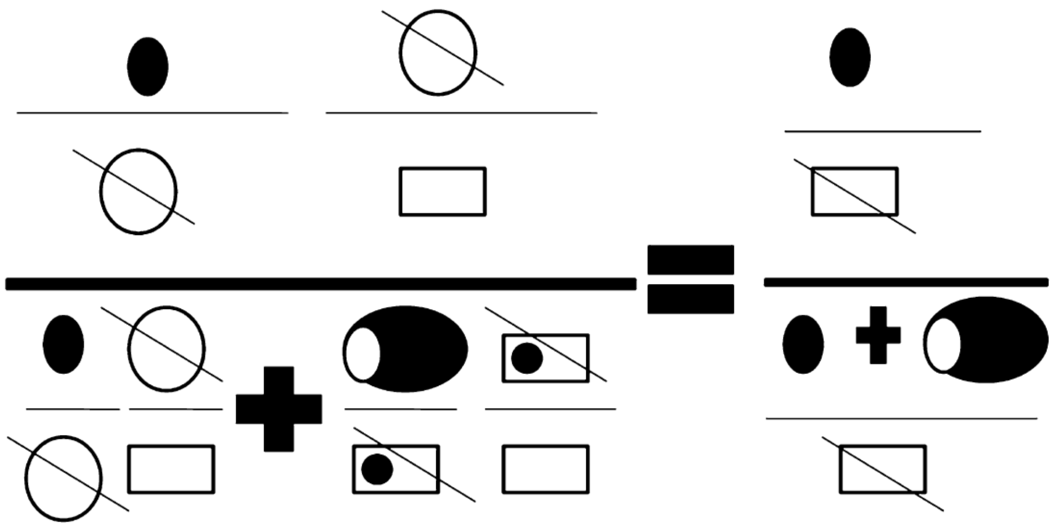


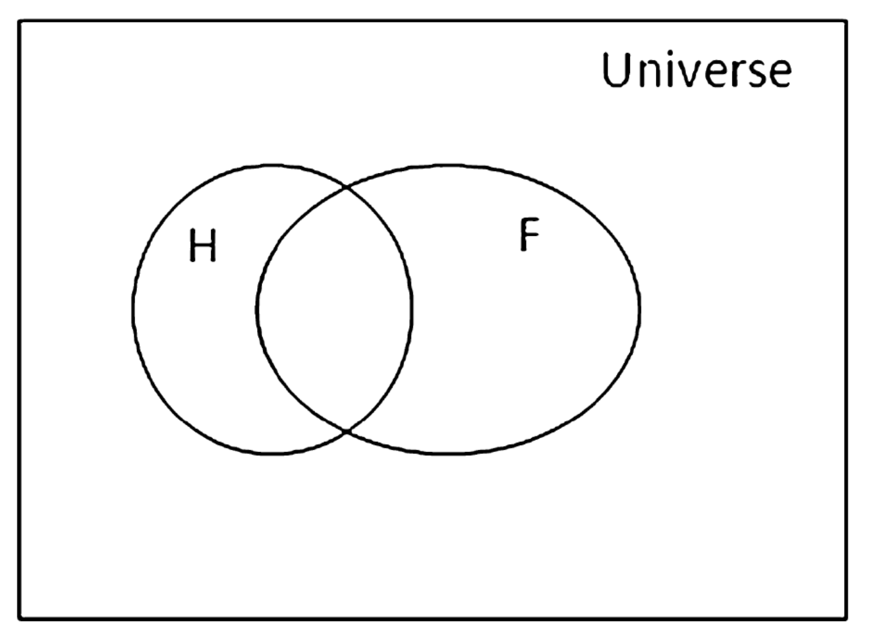
Trafimow, D. (2011).

Teorema de Bayes

El objetivo típico consiste en determinar la probabilidad de una hipótesis (H) dados los datos (F; findings)

$$\frac{p(F|H)p(H)}{p(F|H)p(H) + p(F|{\sim}H)p({\sim}H)} = p(H|F)$$





Trafimow, D. (2011).

El caso de las dos monedas

Suponga que tenemos 2 monedas:

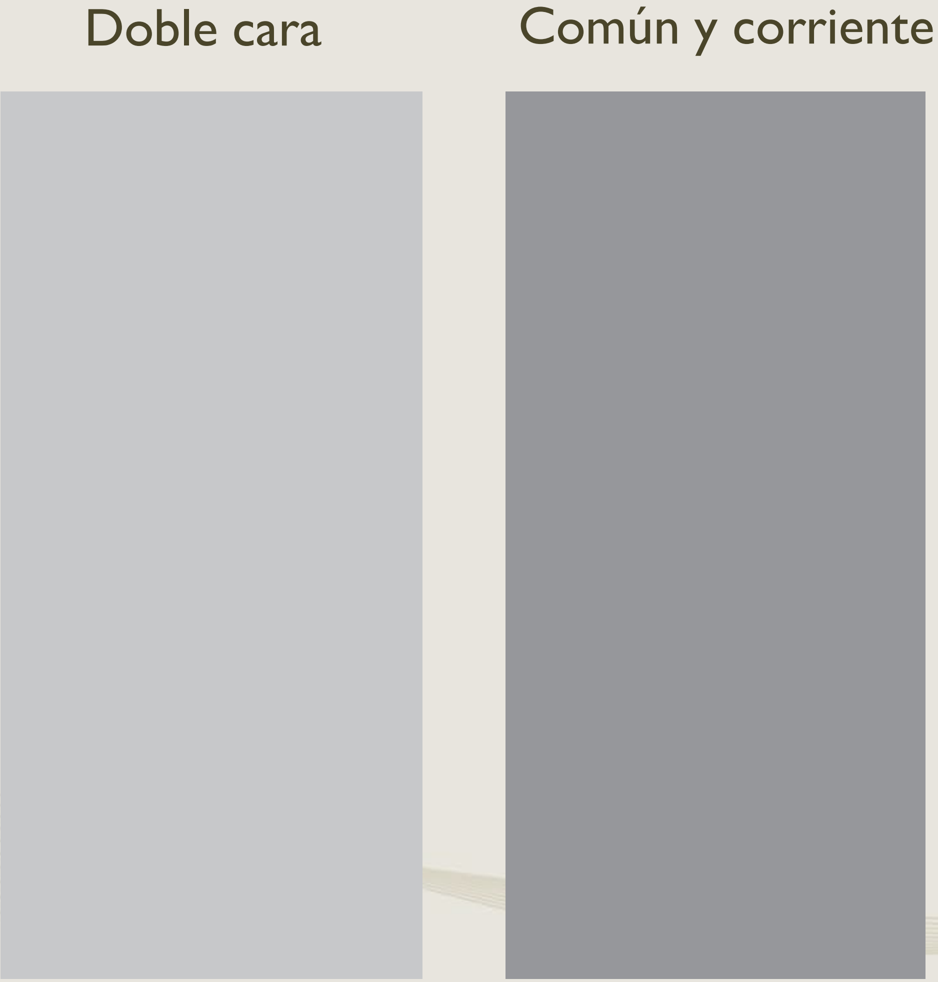
Común y corriente

Doble cara

¿cuál es la probabilidad de que hayamos escogido (***aleatoriamente***) la común y corriente?

El caso de las dos monedas

¿Cuál es la probabilidad (a priori) de que hayamos escogido (***aleatoriamente***) la común y corriente?



Erickson, T. (2017)

El caso de las dos monedas

Tiro la moneda al aire sin saber cuál es y al revisar descubrimos que ha caído cara, ¿sabemos cuál moneda escogimos?,
¿cuál es la probabilidad (posterior) de que la moneda escogida sea la común y corriente?

Doble cara | Común y corriente

$$P(F1|H1) = \frac{P(F1).P(H1|F1)}{P(F1).P(H1|F1) + P(U1).P(H1|U1)}$$
$$= \frac{1/2.1/2}{1/2.1/2 + 1/2.1} = \frac{1}{3}$$

$$P(F1|H1) = 1/3$$

Erickson, T. (2017)

El caso de las dos monedas

Simulación de la regla de Bayes por el método de Monte Carlo

D10004 | f_x =D10002/C10002

	A	B	C	D
1	Coin	Flip	Head	Head|Fair
2	U	H	1	
3	U	H	1	
4	F	H	1	1
5	U	H	1	
6	F	H	1	1
7	U	H	1	
8	F	T		
9995	F	H	1	1
9996	F	T		
9997	F	T		
9998	U	H	1	
9999	U	H	1	
10000	F	T		
10001	F	H	1	1
10002	5019		7471	2490
10003				
10004	**P(F)=**	**0,5019**	**P(F1|H1)=**	**0,33329**

=D10002/C10002

A2=SI(ALEATORIO()<0.5,"F","U")

B2=SI(A2="U","H",SI(ALEATORIO()<0.5,"H","T"))

C2=SI(B2="H",1,"")

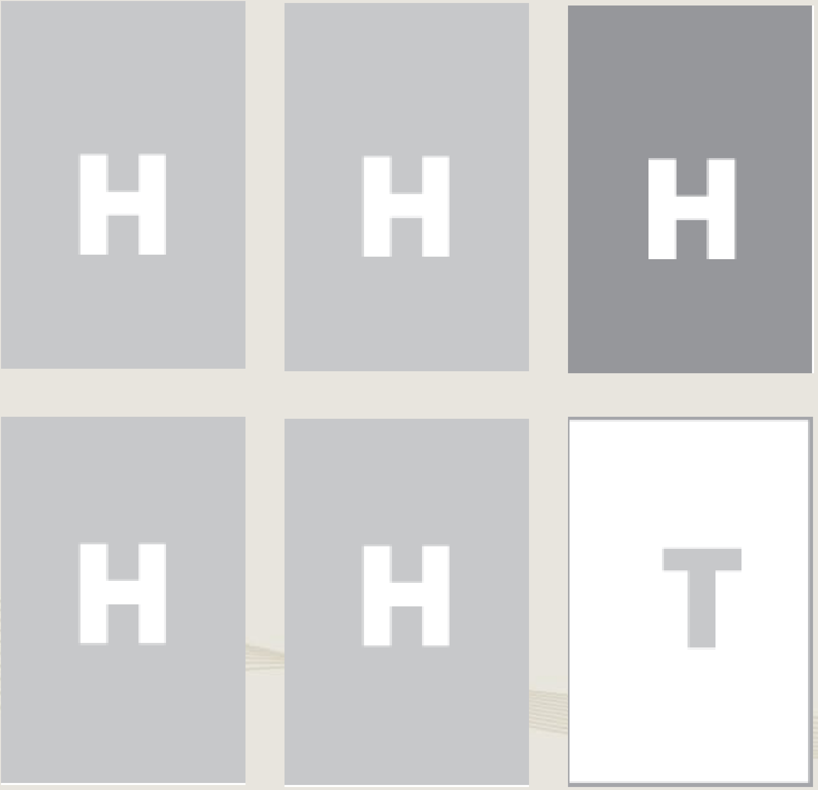
D2=SI(Y(A2="F",B2="H"),1,"")

Gangur, M., & Svoboda, M. (2018).
Teaching Statistics, Volume: 40, Issue: 3, Pages: 83-87, First published: 06 April 2018, DOI: (10.1111/test.12158)

El caso de las dos monedas

Tiro la moneda al aire por segunda vez (sin revisar de qué tipo es) y al revisar descubrimos que ha caído cara nuevamente,
¿ahora sí ya sabemos cuál moneda escogimos?,
¿cuál es la probabilidad (posterior) de que la moneda escogida sea la común y corriente?

Doble cara   Común y corriente



$$P(F2|H2) = \frac{P(F1).P(H2|F1)}{P(F1).P(H2|F1) + P(U1).P(H2|U1)} = \frac{1/3.1/2}{1/3.1/2 + 2/3.1} = \frac{1}{5}$$

$$P(F2|H2) = 1/5$$

Erickson, T. (2017)

El caso de las dos monedas

Simulación de la regla de Bayes por el método de Monte Carlo

E10004 fx =E10002/D10002

	A	B	C	D	E
1	Coin	First flip	Second flip	Head and Head	(Head and Head)|Fair
2	F	T			
3	U	H	H	1	
4	F	H	H	1	1
5	U	H	H	1	
6	F	H	T		
7	U	H	H	1	
8	U	H	H	1	
9995	F	T			
9996	F	T			
9997	U	H	H	1	
9998	U	H	H	1	
9999	F	H	H	1	1
10000	F	T			
10001	F	H	H	1	1
10002	5017			6239	1256
10003					
10004	**P(F)=**	**0,5017**		**P(F2|H2) =**	**0,20131**

A2=SI(ALEATORIO()<0.5,"F","U")

B2=SI(A2="U","H",SI(ALEATORIO()<0.5,"H","T"))

C2=SI(B2="H",SI(A2="U","H",SI(ALEATORIO()>0.5,"H","T")),"")

D2=SI(Y(B2="H",C2="H"),1,"")

E2=SI(Y(A2="F",D2=1),1,"")

=E10002/D10002

Gangur, M., & Svoboda, M. (2018).

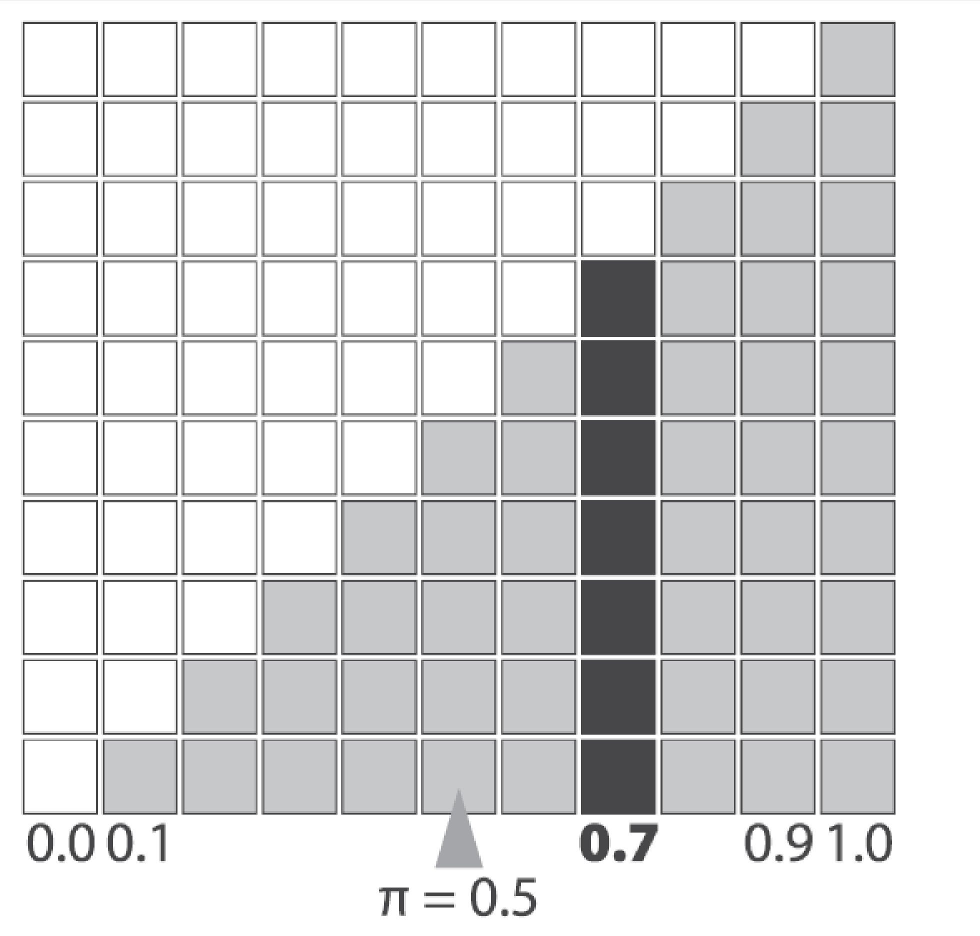
Teaching Statistics, Volume: 40, Issue: 3, Pages: 83-87, First published: 06 April 2018, DOI: (10.1111/test.12158)

Más de 2 hipótesis

- Suponga 11 diferentes hipótesis (igualmente probables) sobre la moneda misteriosa↑ $\pi = \{0, 0.1, 0.2, \dots, 0.9, 1.0\}$
- Tiramos la moneda al aire y cae cara, ¿cuáles son las probabilidades posteriores?
- ¿Cuál es la probabilidad de $\pi = 0.7$?
- ¿Sigue siendo $1/11$ (algo así como 0.09)?

$$P(\pi = 0.7 | H1) = 7/55$$

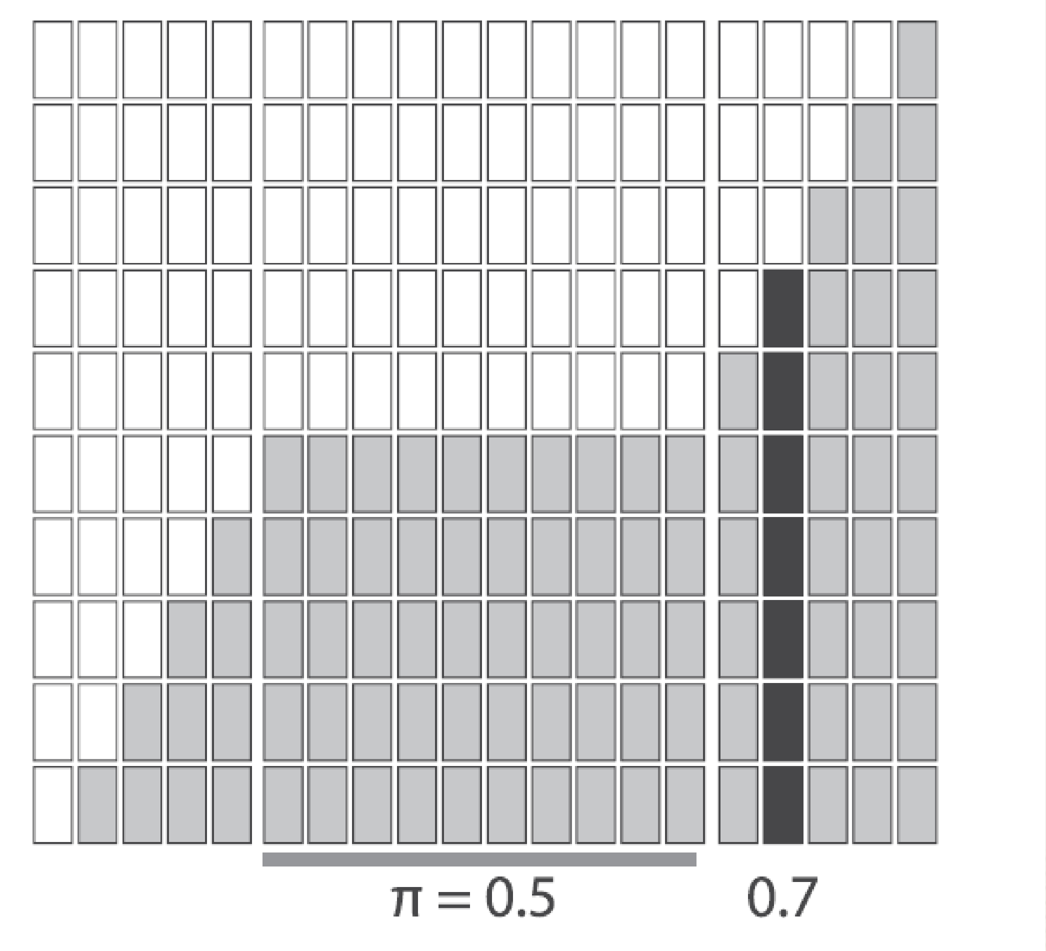
(algo así como 0.17)



Erickson, T. (2017)

Más de 2 hipótesis

- ¿Cómo se vería un prior no-uniforme sobre las diferentes hipótesis usando el mismo diagrama de área donde podemos contar cajas?



Erickson, T. (2017)

Conclusiones. Influencia del tamaño de los datos sobre la posterior

- Hemos visto que la distribución posterior es un compromiso entre los a prioris y la función de verosimilitud
- Definimos dos p(θ)
- La posterior se aproxima a los datos a medida que la muestra se incrementa (con priors vagos es más claro)
- Entre más datos tengamos, mejor es la precisión

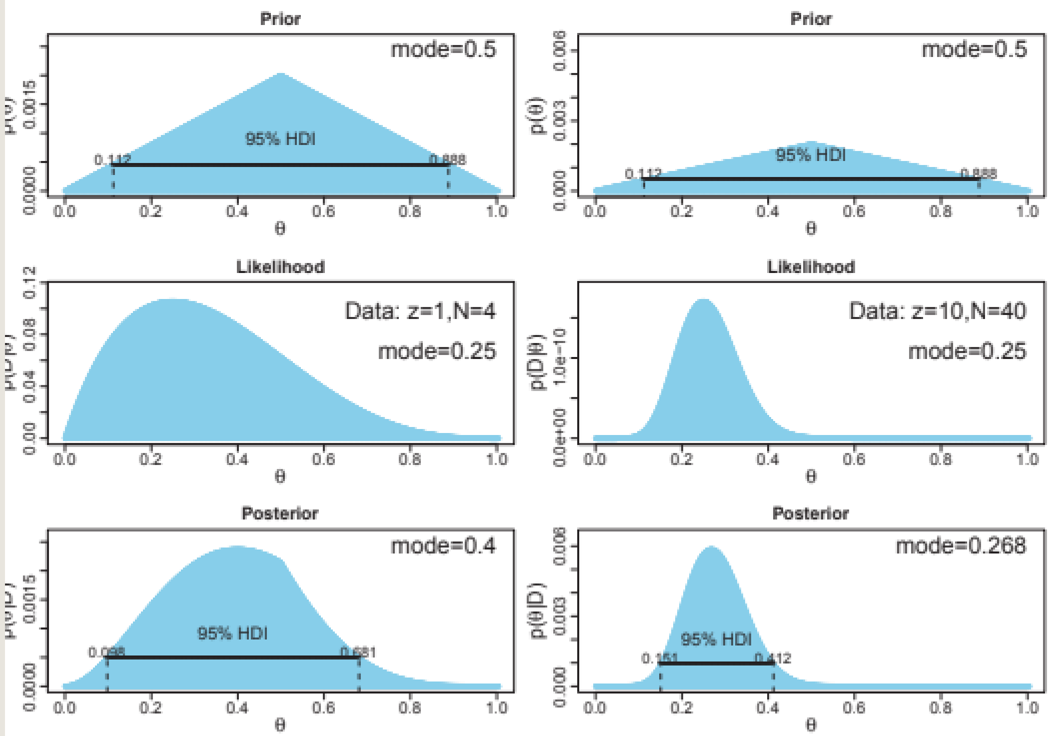


Figure 5.2: The two columns show different sample sizes with the same proportion of heads. The prior is the same in both columns but plotted on a different vertical scale. The influence of the prior is overwhelmed by larger samples, in that the peak of the posterior is closer to the peak of the likelihood function. Notice also that the posterior HDI is narrower for the larger sample. Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier.

Conclusiones. Influencia del a priori sobre la posterior

- En general, cuando la distribución de $p(\theta)$ es amplia comparada con la función de verosimilitud, $p(\theta)$ tiene poca influencia

- Con priors muy informativos (que están basados en muchos datos), se necesitarán muchos datos que lo contradigan para ver diferencias entre el prior y la distribución posterior.

- Con priors vagos, los datos fácilmente desplazan a la distribución a posteriori.

- "It can be a serious blunder not to use strong prior information when it is available" p. 114.

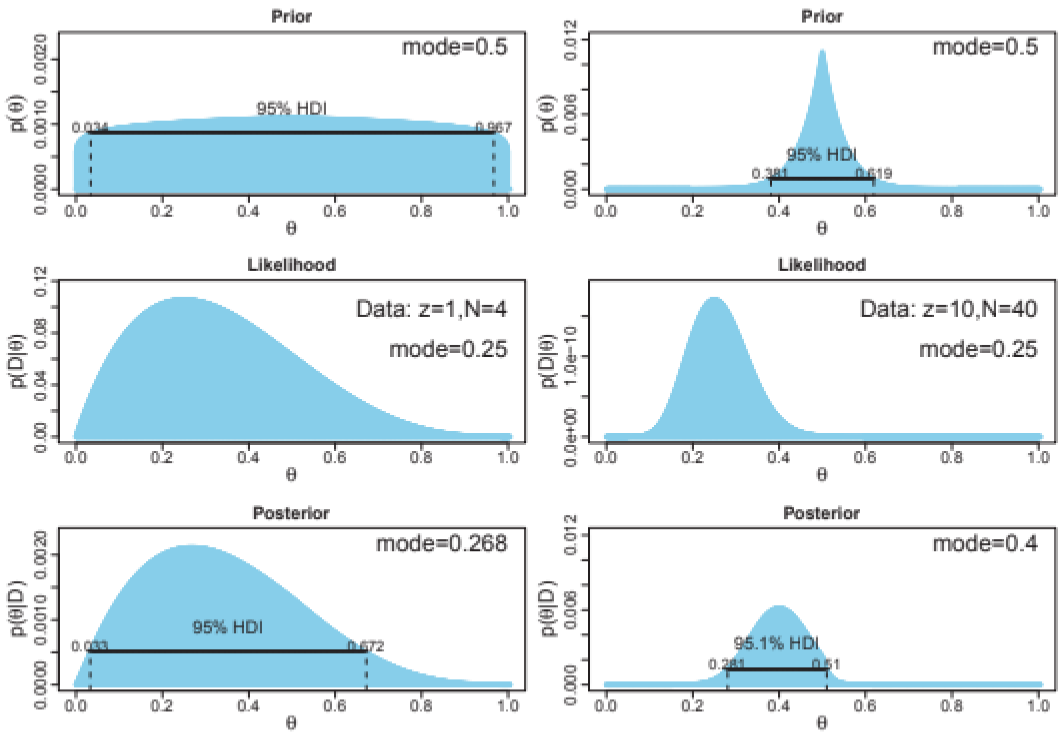


Figure 5.3: The left side is the same small sample as the left side of Figure 5.2 but with a flatter prior. The right side is the same larger sample as the right side of Figure 5.2 but with a sharper prior. Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier.

La inferencia Bayesiana puede ser difícil

$p(\theta|D) = p(D|\theta)\, p(\theta) / p(D)$

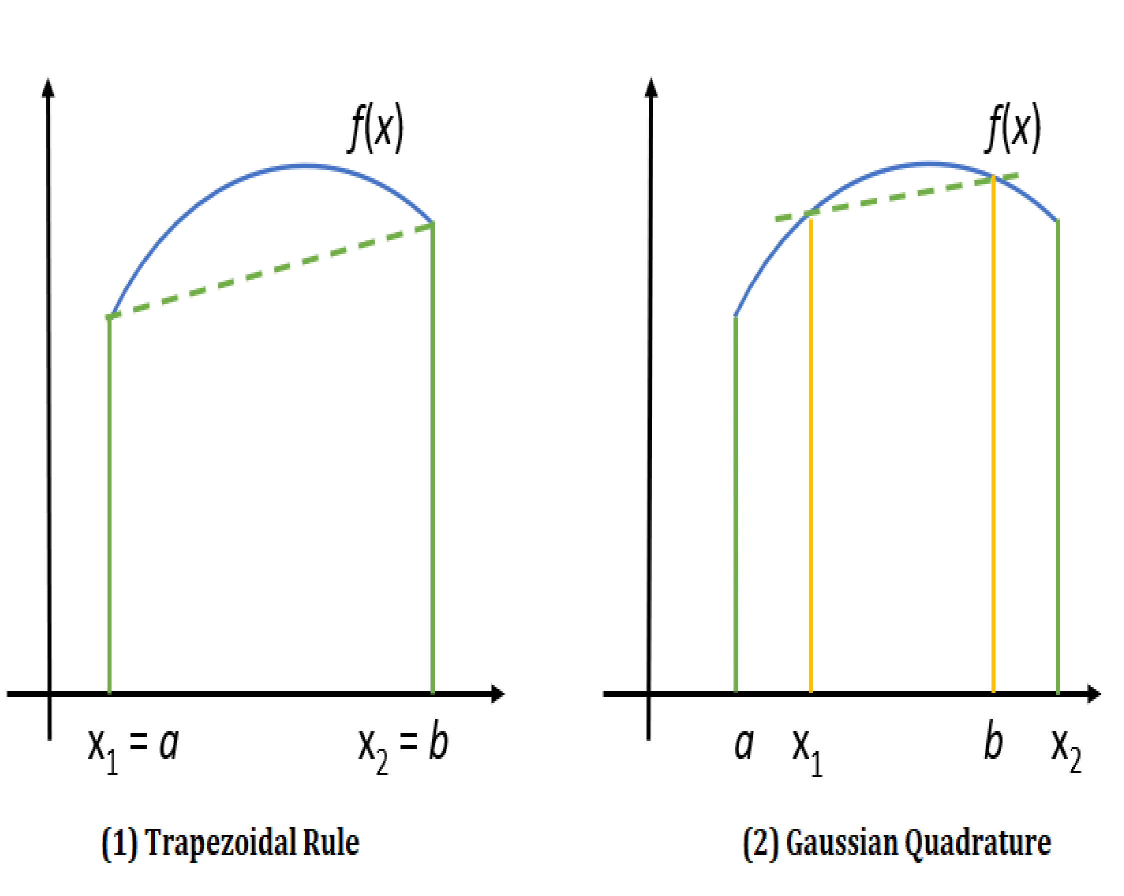
Posterior = verosimilitud a priori / evidencia

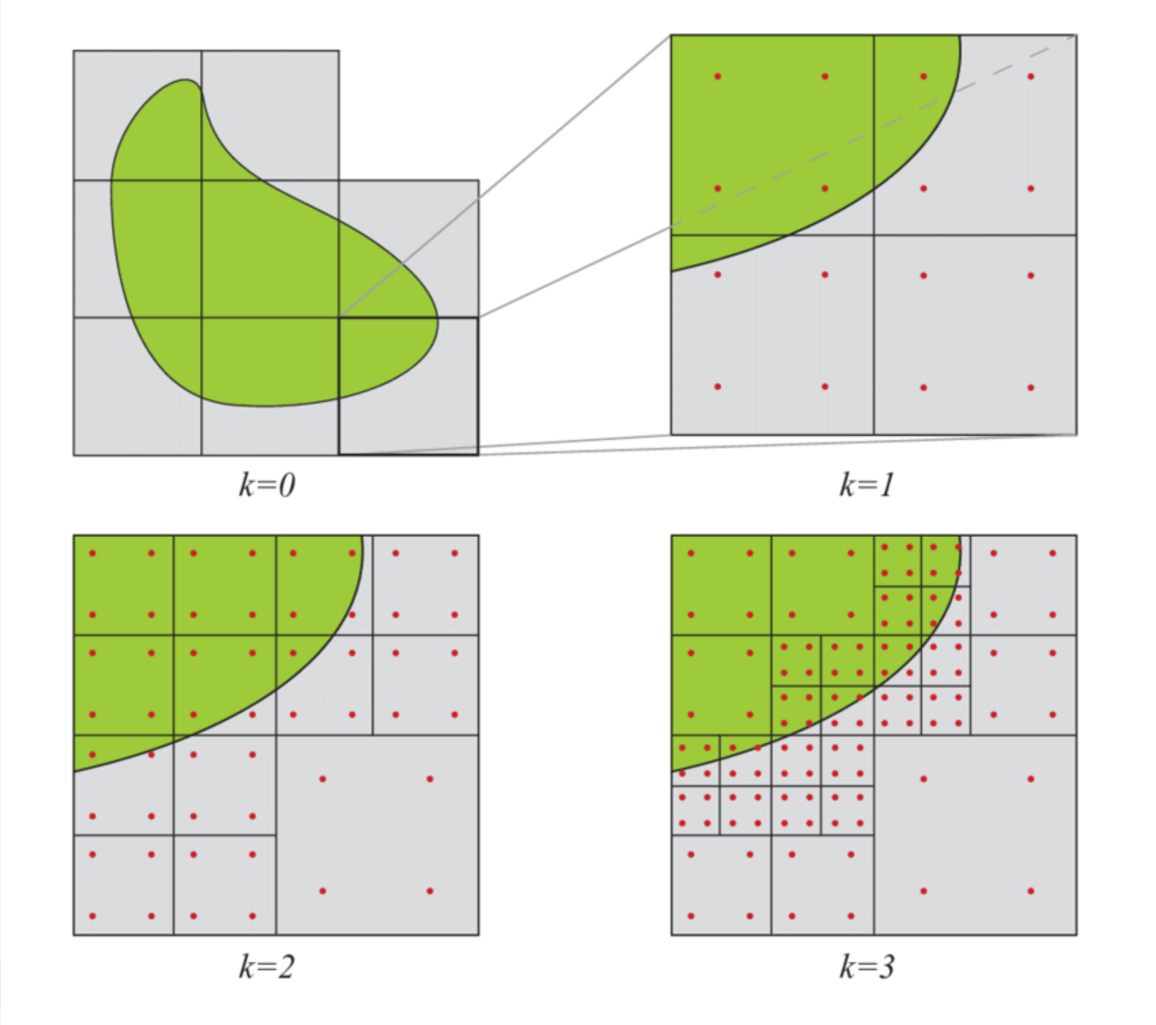
En el caso de parámetros continuos la ecuación puede ser imposible de resolver analíticamente

Si temenos un parámetro con 1000 valores y tenemos DOS parámetros tenemos:

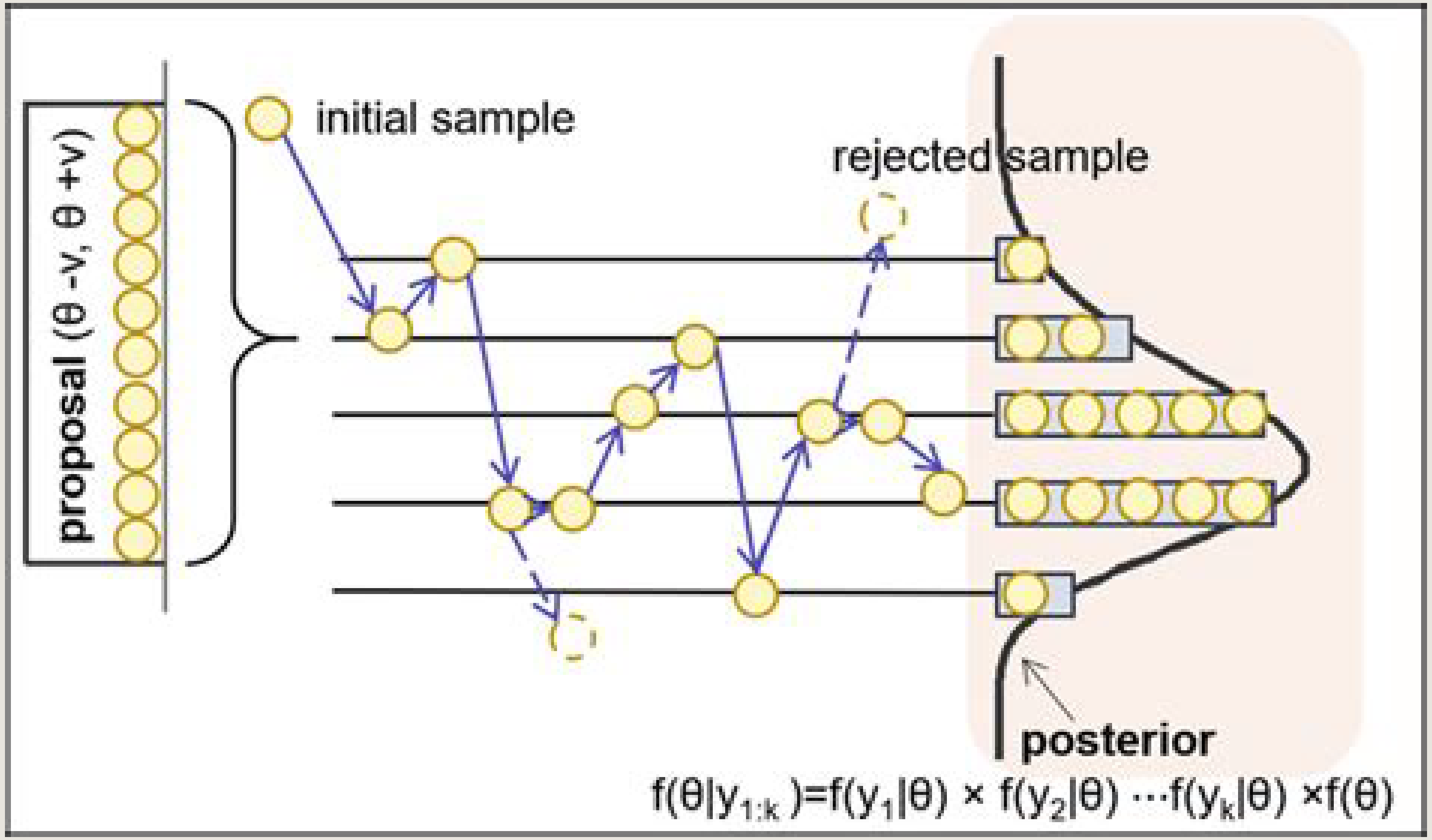
1,000^2 combinations de parámetros

Soluciones (?)


f(x)
f(x)
x1 = a
x2 = b
a
x1
b
x2
(1) Trapezoidal Rule
(2) Gaussian Quadrature


k=0
k=1
k=2
k=3

Métodos estocásticos



Próxima clase

Los peligros de los p-values en NHST

Capítulo 11: Null Hypothesis Significance Testing.
Doing Bayesian analysis. John K. Kruschke

Referencias

Andrew Gelman (2011), "Induction and Deduction in Bayesian Data Analysis", Special Topic: Statistical Science and Philosophy of Science RMM Vol. 2, 2011, 67–78

Bernoulli, J. (1713). *Ars conjectandi, opus posthumum: accedit tractatus de seriebus infinitis, et epistola Gallice scripta de ludo pilæ reticularis*. Impensis Thurnisiorum Fratrum.

Brooks, S. P. (2003). Bayesian computation: a statistical revolution. Philosophical Transactions of the Royal Society of London. Series A: Mathematical, *Physical and Engineering Sciences*, 361(1813), 2681-2697.

Erickson, T. (2017). Beginning Bayes. *Teaching Statistics*, 39(1), 30-35.

Gangur, M., & Svoboda, M. (2018). Simulation of Bayes' rule by means of Monte Carlo method. *Teaching Statistics*, 40(3), 83-87.

Keynes, J. M. A. (1921). *Treatise on Probability*, London: Macmillan.

Kruschke, J. (2014). Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan.

Neyman, J. (1992 [1934]). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. En *Breakthroughs in Statistics* (pp. 123-150). Springer, New York, NY.

Rincón, L. (2007). *Curso intermedio de probabilidad*. UNAM, Facultad de Ciencias.

Rosenkrantz, R. D. (1989). Where do we stand on maximum entropy?(1978). In *ET Jaynes: Papers on probability, statistics and statistical physics* (pp. 210-314). Springer, Dordrecht.

Salsburg, D. (2001). *The lady tasting tea: How statistics revolutionized science in the twentieth century*. Macmillan.

Shafer, G. (1996). The significance of Jacob Bernoulli's Ars Conjectandi for the philosophy of probability today. *Journal of Econometrics*, 75(1), 15-32.

Trafimow, D. (2011). Using pictures to enhance students' understanding of Bayes' theorem. *Teaching Statistics*, 33(3), 83-84.

CONTACTO

Dr. Héctor Nájera y Dr. Curtis Huffman
Investigadores

Programa Universitario de Estudios del Desarrollo (PUED)
Universidad Nacional Autónoma de México (UNAM)
Antigua Unidad de Posgrado (costado sur de la Torre II de Humanidades), planta baja.
Campus Central, Ciudad Universitaria, Ciudad de México, México.
Tel. (+52) 55 5623 0222, Ext. 82613 y 82616

Tel. (+52) 55 5622 0889
Email: hecatalan@hotmail.com, chuffman@unam.mx

PONTE PUMA
PONTE EL CUBREBOCAS

UNAM La Universidad de la Nación