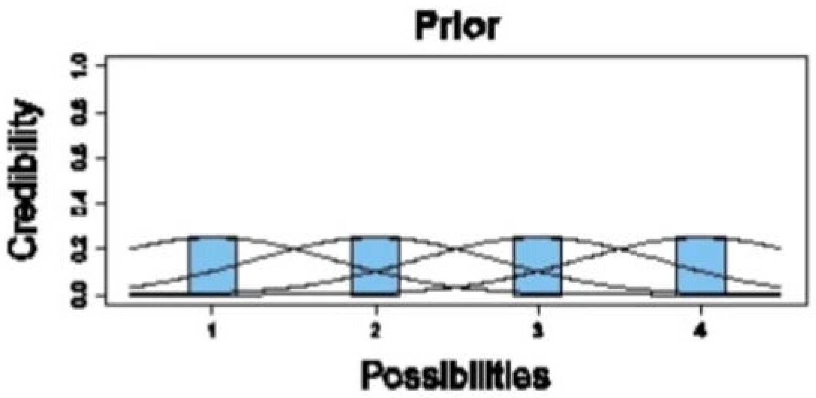
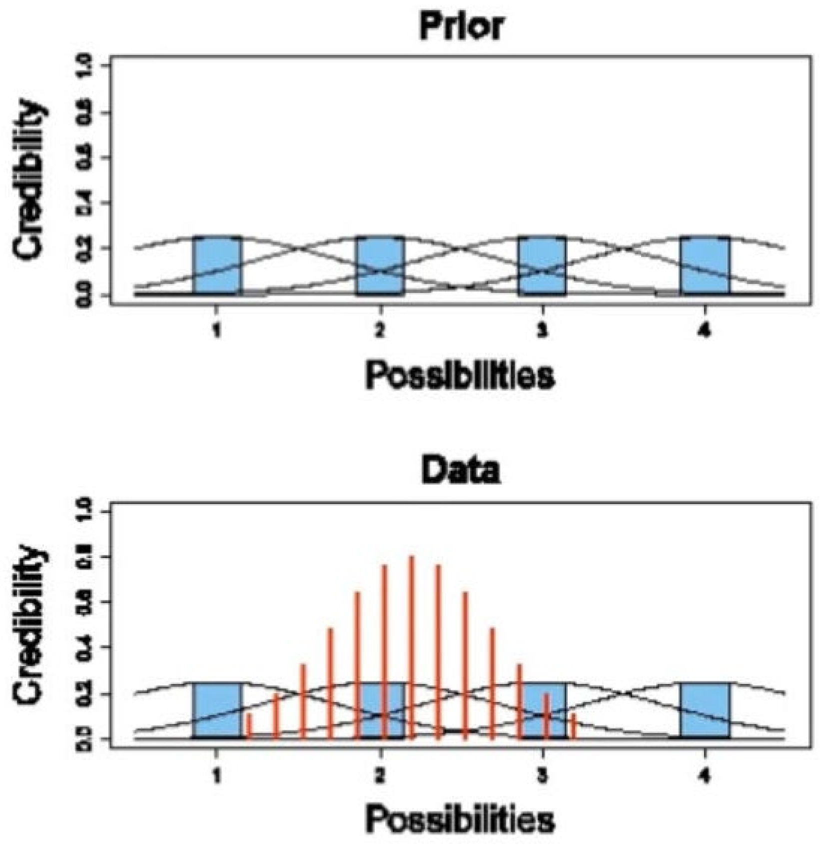
Teorema de Bayes: priors, verosimilitud y posterior

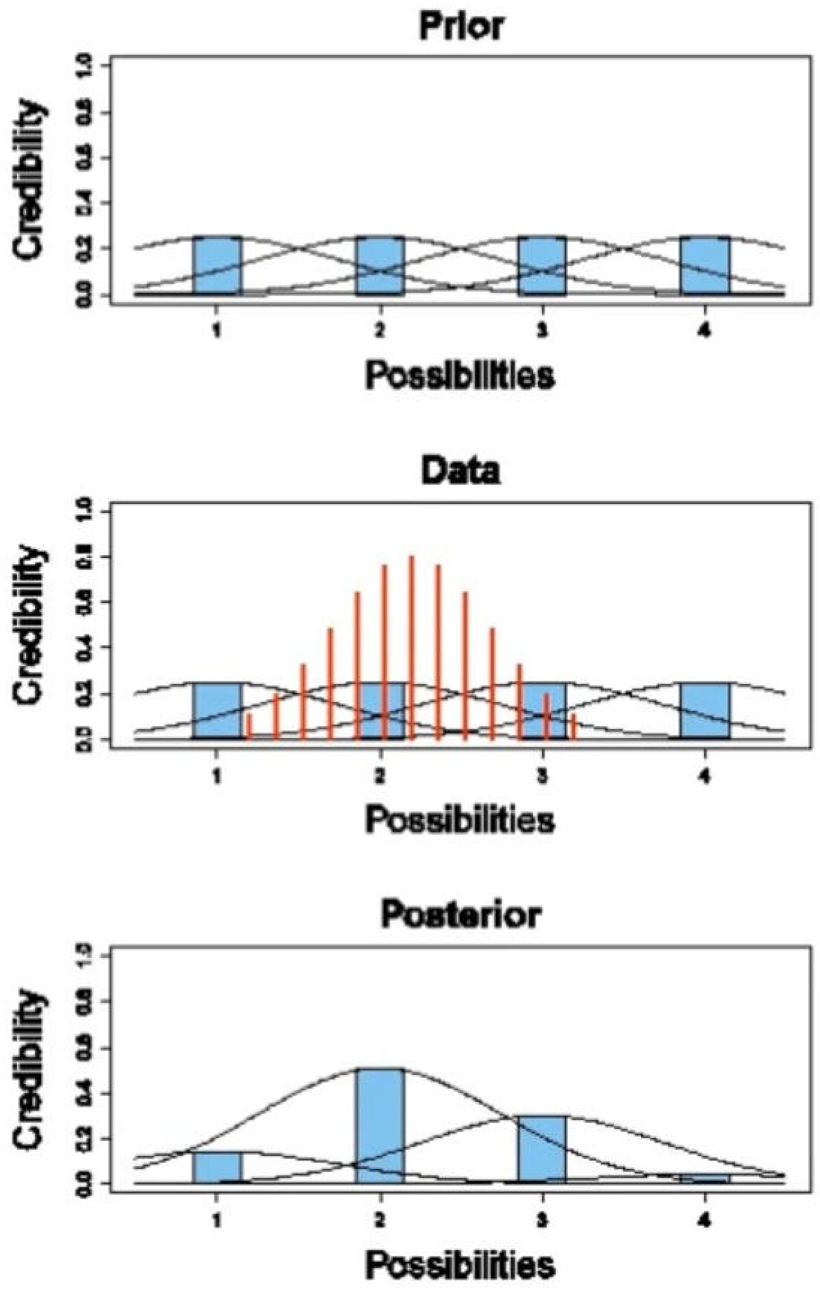
Dr. Héctor Nájera
Dr. Curtis Huffman

- Análisis bayesiano de datos
  - El papel de los datos es reubicar credibilidad entre posibilidades (i.e., aprender).
  - Las posibilidades son valores de parámetros en un modelo probabilístico (μ, σ)
  - Reubicamos credibilidad a valores de parámetros consistentes con los datos



- Análisis bayesiano de datos
  - El papel de los datos es reubicar credibilidad entre posibilidades (i.e., aprender).
  - Las posibilidades son valores de parámetros en un modelo probabilístico (μ, σ)
  - Reubicamos credibilidad a valores de parámetros consistentes con los datos




Prior
Credibility
Possibilities
Data
Credibility
Possibilities
Posterior
Credibility
Possibilities
1
2
3
4

- Análisis bayesiano de datos
  - El papel de los datos es reubicar credibilidad entre posibilidades (i.e., aprender).
  - Las posibilidades son valores de parámetros en un modelo probabilístico (μ, σ)
  - Reubicamos credibilidad a valores de parámetros consistentes con los datos

Formula:

Likelihood distribution

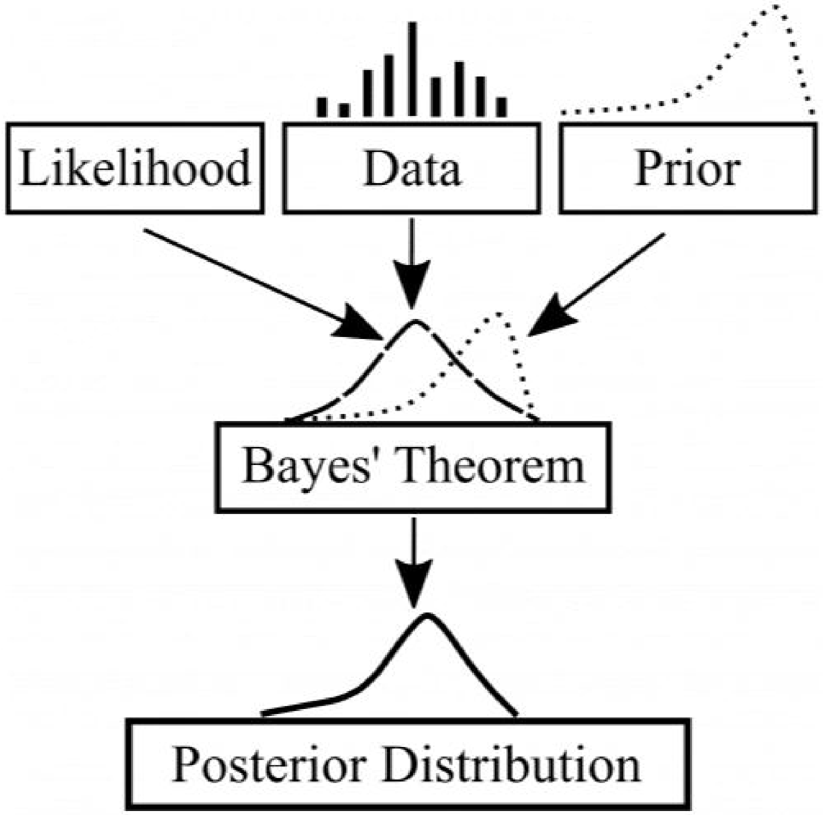
Prior distribution

$$f(\text{model} \mid \text{data}) = \frac{f(\text{data} \mid \text{model}) \times f(\text{model})}{f(\text{data})}$$

Bayesian Model

Data

Bayesian Model:



Formula:

Likelihood distribution

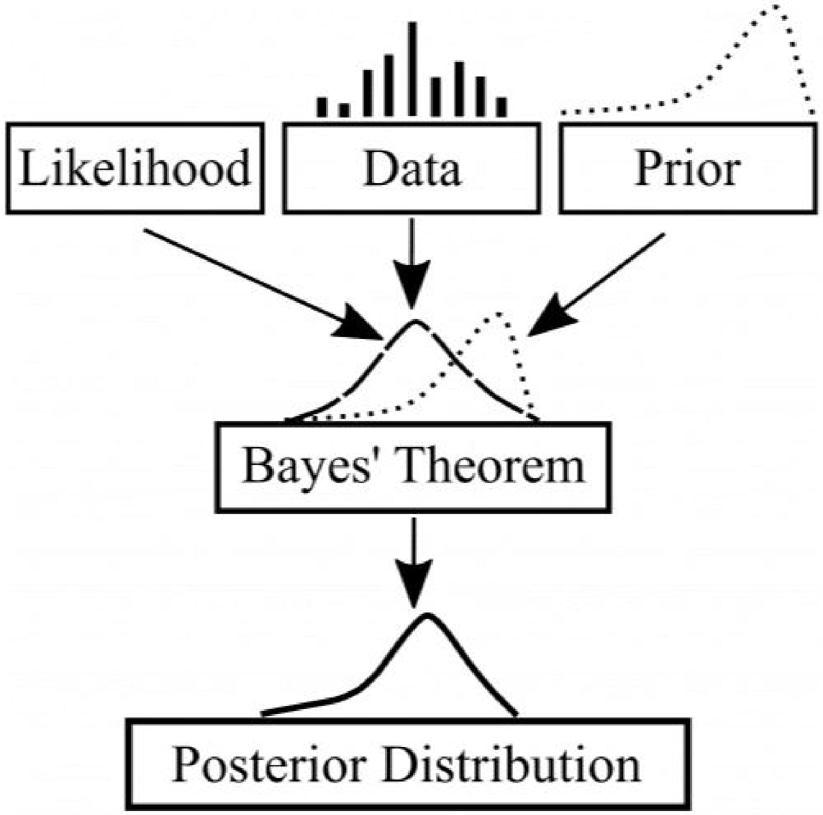
Prior distribution

$$f(\text{model} \mid \text{data}) = \frac{f(\text{data} \mid \text{model}) \times f(\text{model})}{f(\text{data})}$$

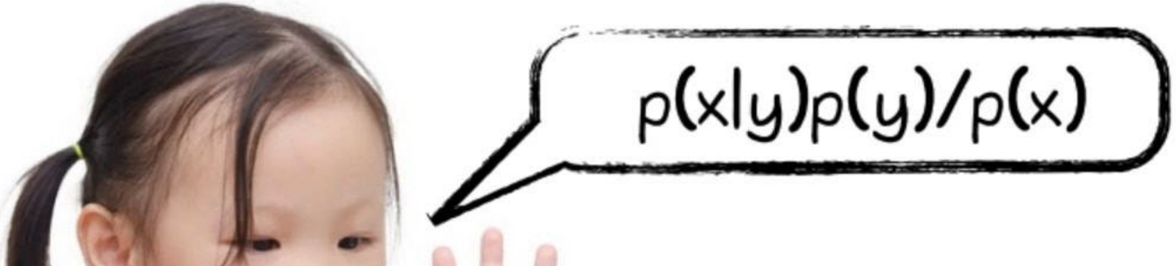
Bayesian Model

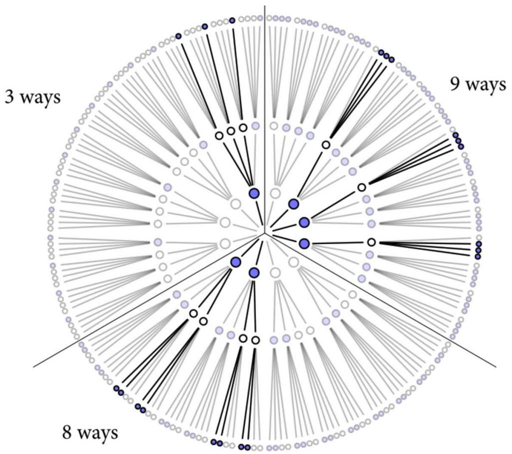
Data

Bayesian Model:



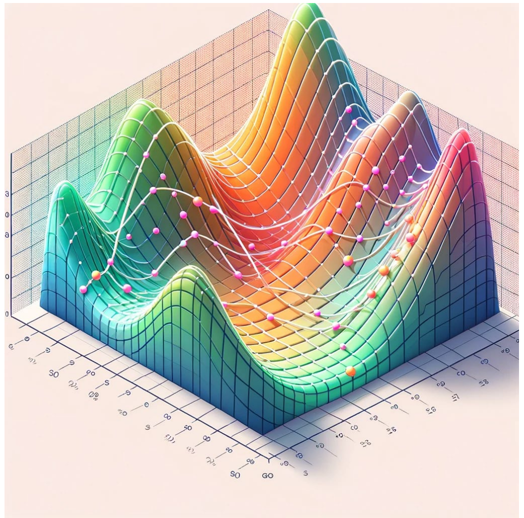
La inferencia bayesiana no es más que contar

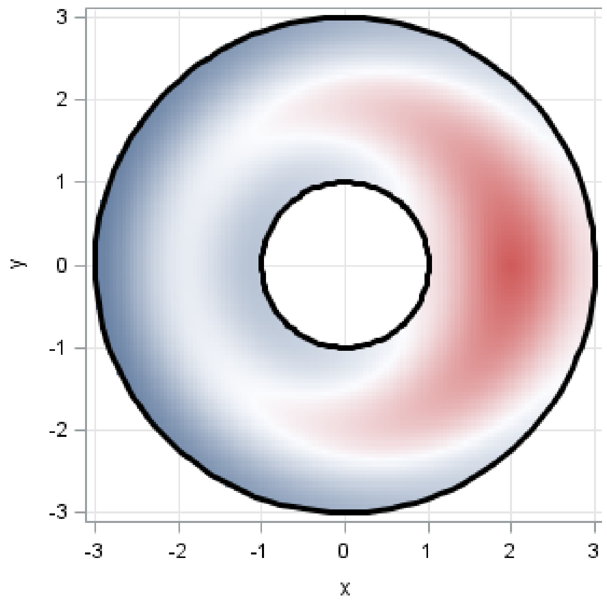




- Cuente todas las formas en que, **de acuerdo con su modelo**, los datos (variables observadas) pudieron haberse observado
- Parámetros con más formas en las que los datos pudieron haberse observado son más plausibles

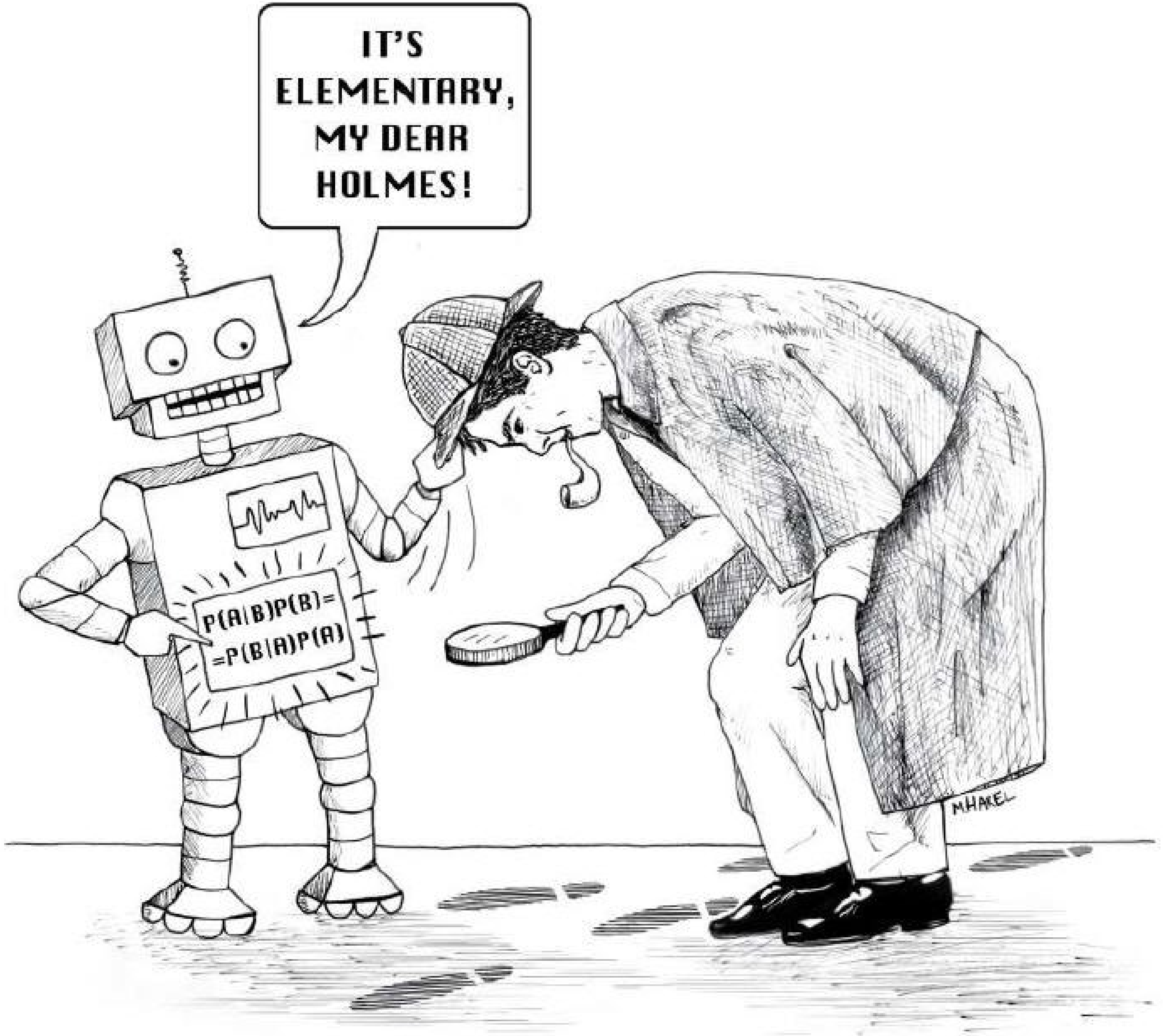
El jardín de los datos que se bifurcan





Restringir la región del espacio de parámetros donde viven los mejores predictores

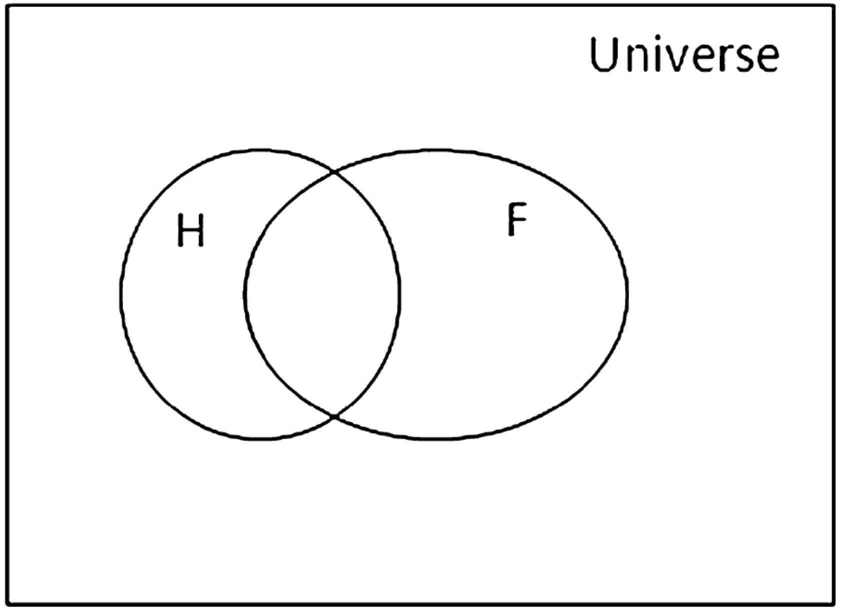
¿EL PESO DE LA EVIDENCIA?


IT'S ELEMENTARY, MY DEAR HOLMES!
P(A|B)P(B)=
=P(B|A)P(A)

Teorema de Bayes

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

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

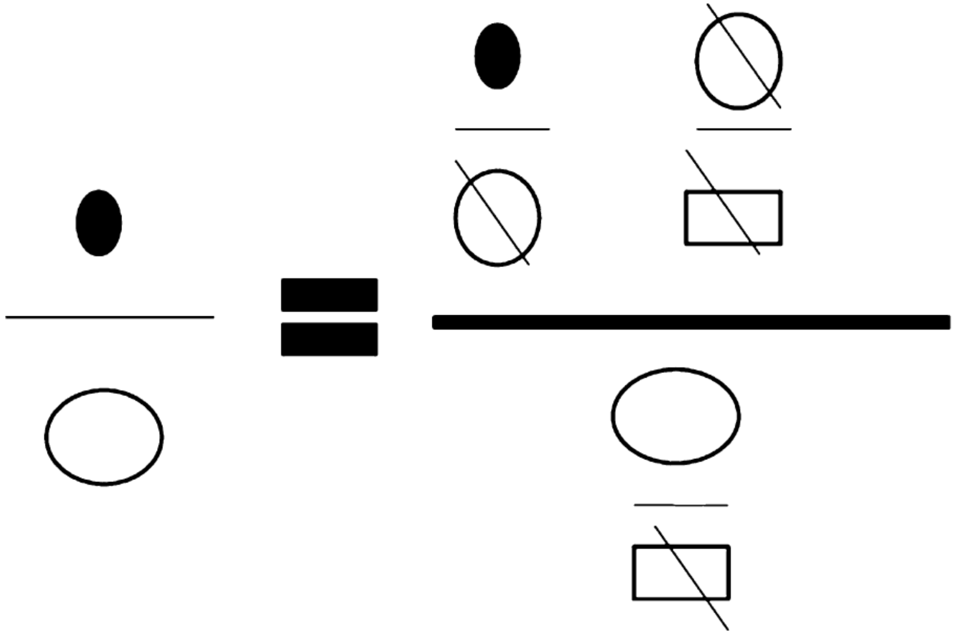


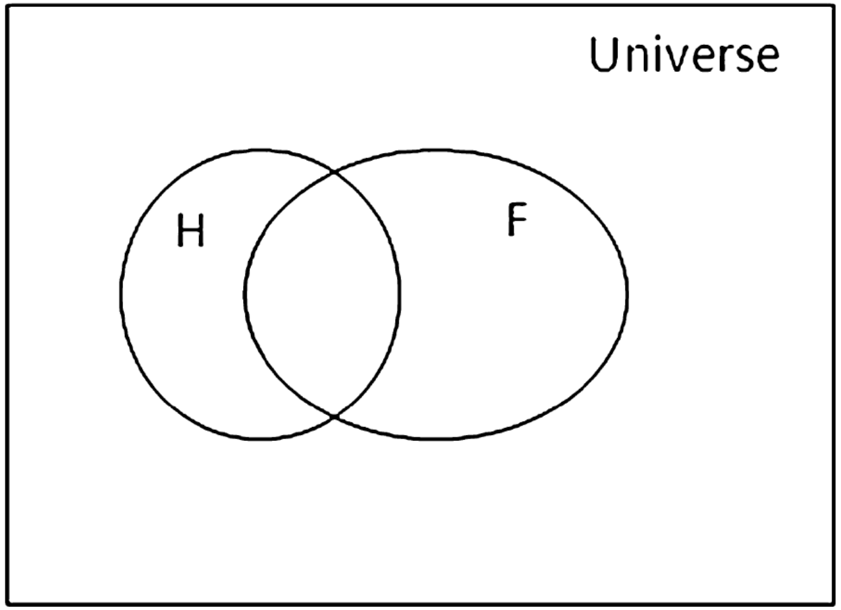
Trafimow, D. (2011).

Teorema de Bayes

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

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



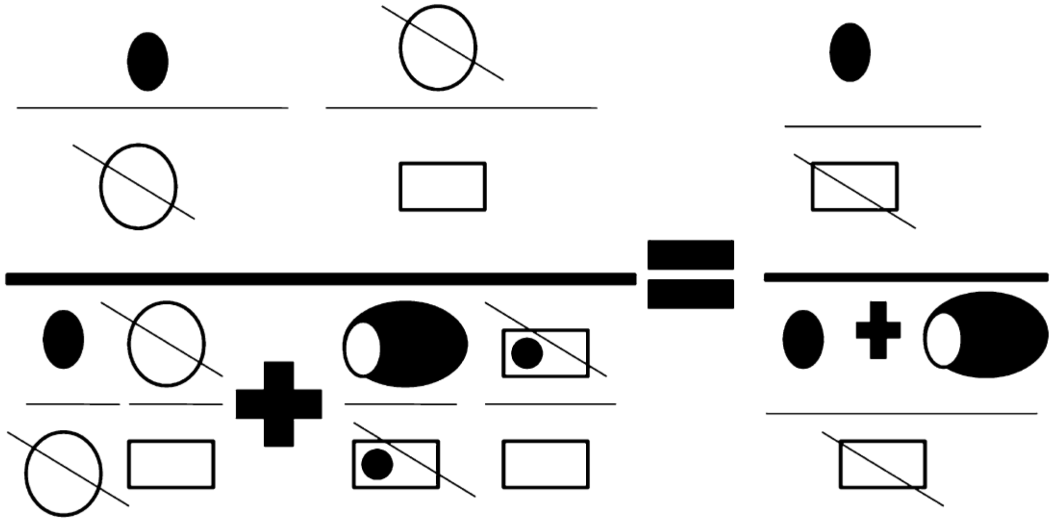


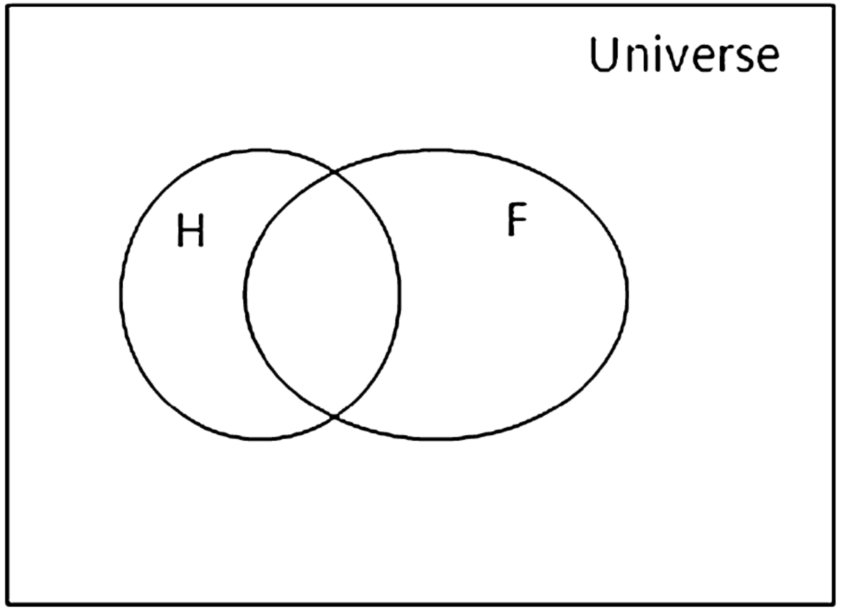
Trafimow, D. (2011).

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:

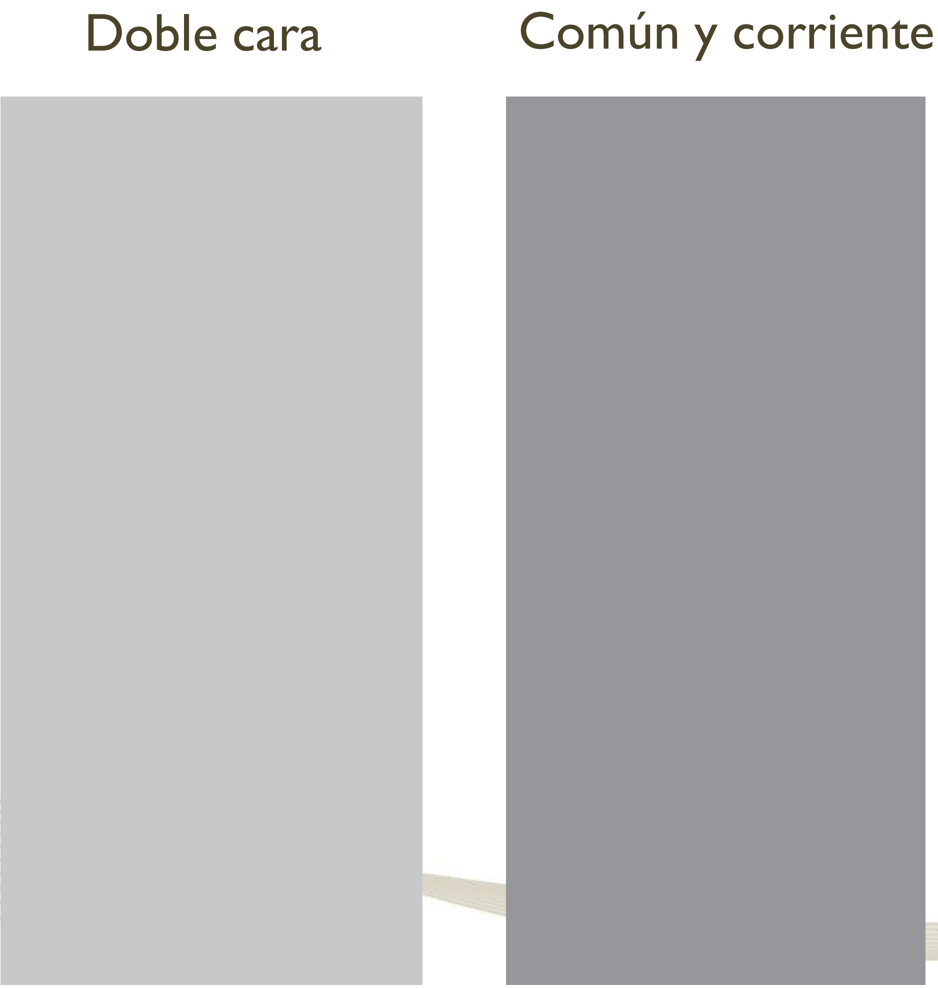
Común y corriente

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

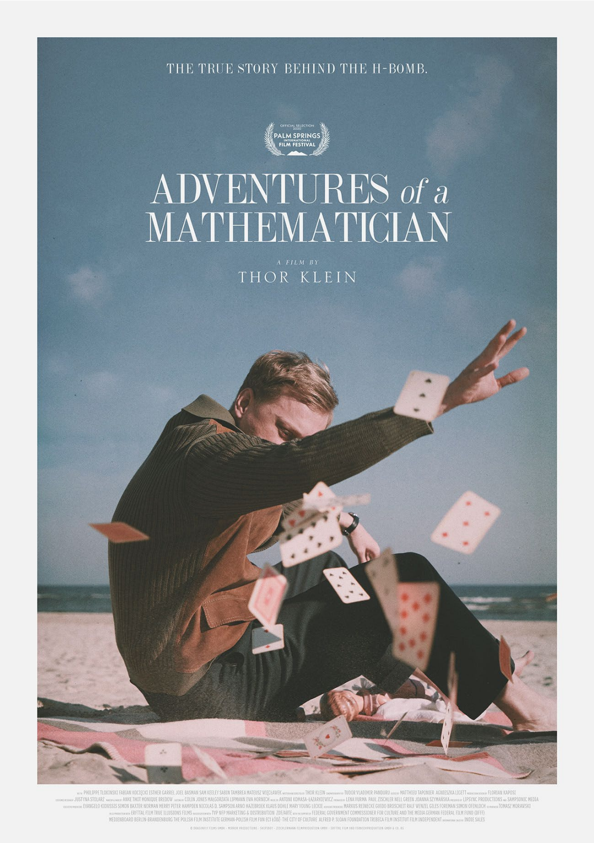
$$P(F1|H1) = \frac{P(F1).P(H1|F1)}{P(F1).P(H1|F1) + P(U1).P(H1|U1)}$$
$$= \frac{1/2.1/2}{1/2.1/2 + 1/2.1} = \frac{1}{3}$$

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

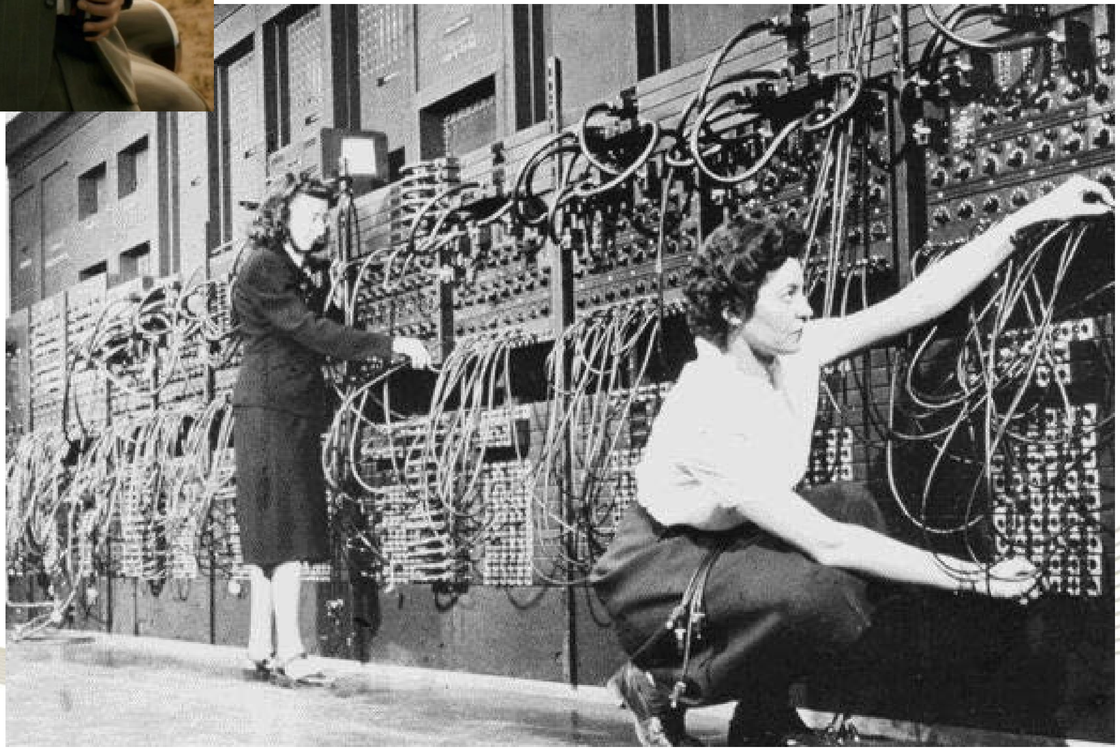
Erickson, T. (2017)

Método de Monte Carlo

Método de Monte Carlo




Stanisław Ulam and his daughter Claire with MANIAC



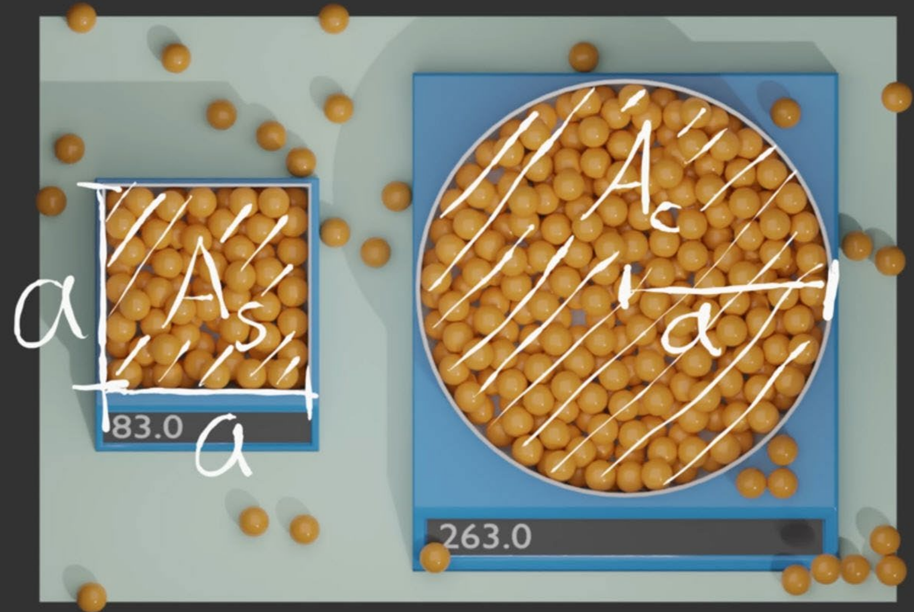
Método de Monte Carlo

- Un método para estimar el valor de una cantidad desconocida usando principios de la inferencia estadística
- Inferencia estadística
  - Población (conjunto de ejemplos)
  - Muestra (subconjunto propio de la población)
  - Una ***muestra aleatoria*** tiende a exhibir las mismas propiedades que la población de la que fue extraída

Método de Monte Carlo

¿Qué hay de aleatorio en π?

Buffon-Laplace



$$A_s = a^2 \qquad A_c = \pi a^2$$

$$\frac{A_c}{A_s} = \frac{\pi \cancel{a^2}}{\cancel{a^2}} = \pi$$

El caso de las dos monedas

Simulación de la regla de Bayes por el método de Monte Carlo (métodos para estimar el valor desconocido de una cantidad usando principios de la estadística inferencial)

D10004 | =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 (métodos para estimar el valor desconocido de una cantidad usando principios de la estadística inferencial)

E10004 =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,"")

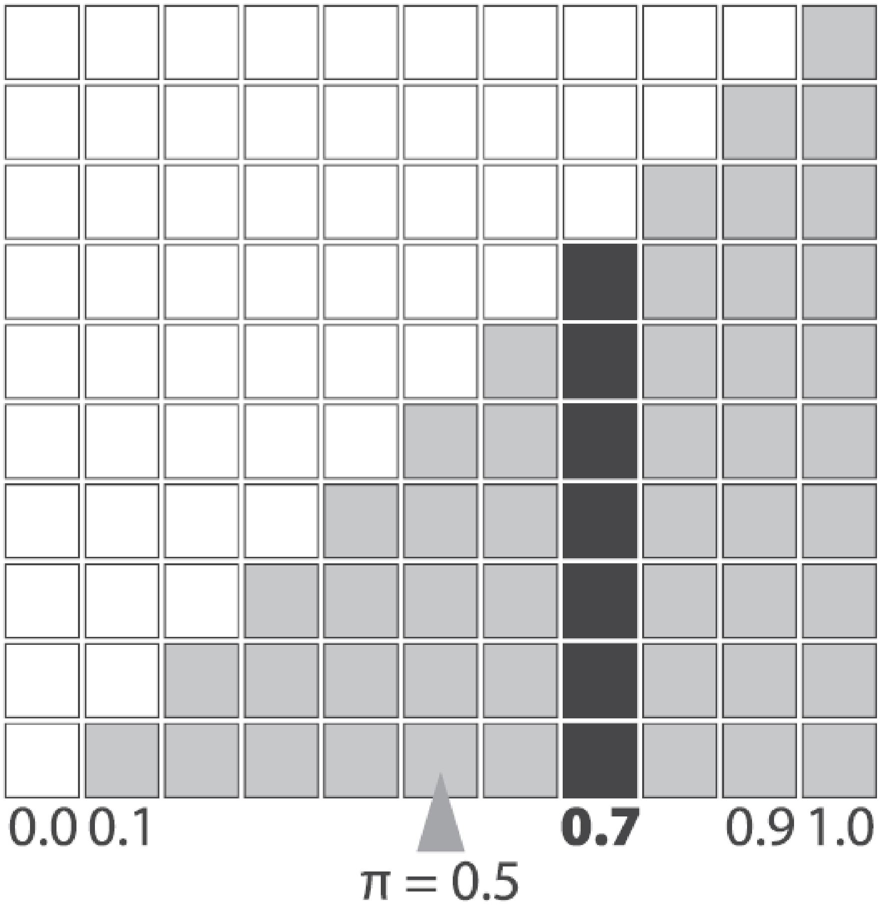
=E10002/D10002

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)?



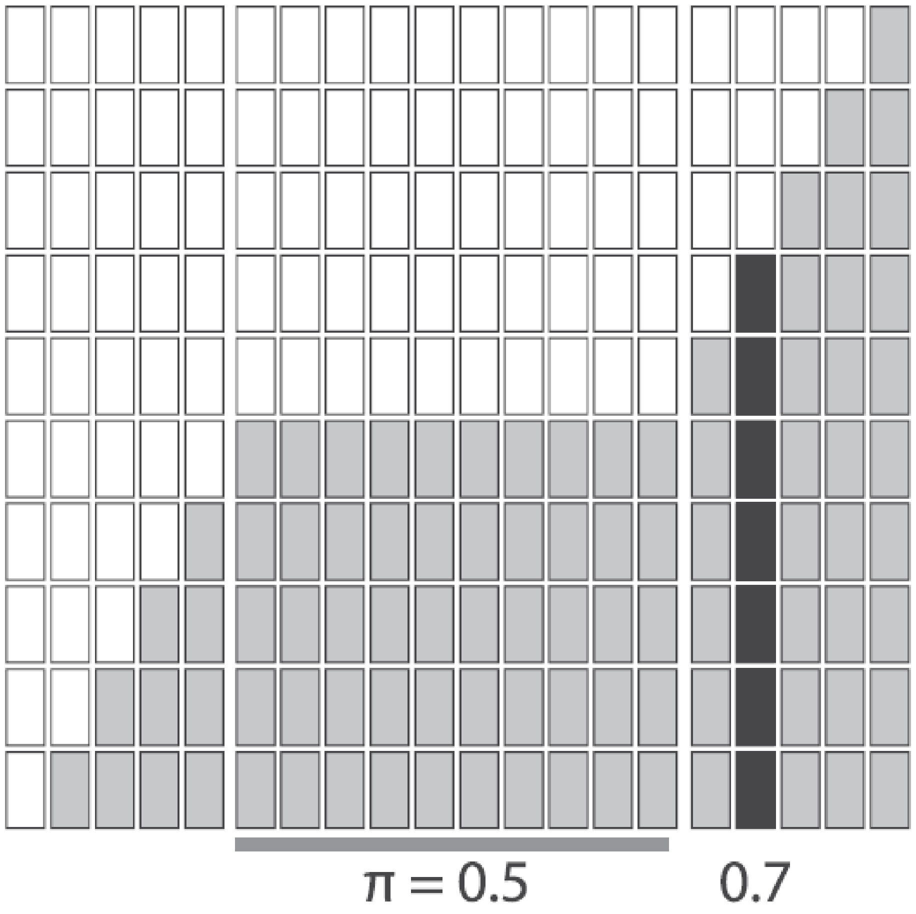
$$P(\pi = 0.7|H1) = 7/55$$

(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 “arreglo” entre los a prioris y la función de verosimilitud
- Definimos dos $p(\theta)$
- 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

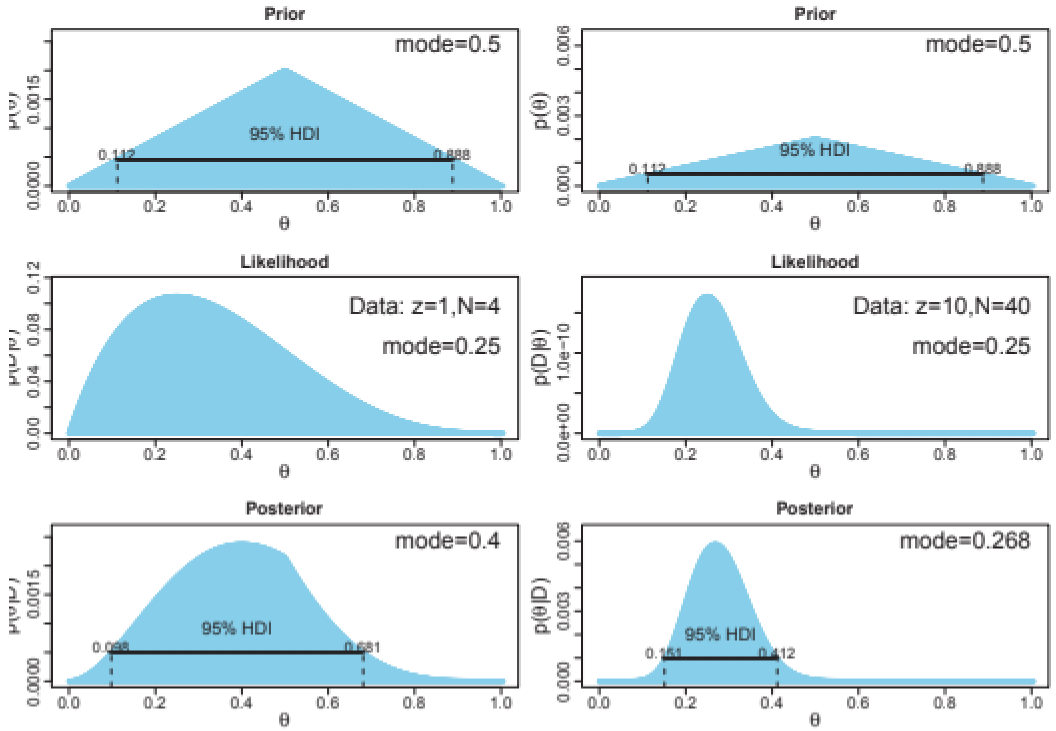


Figure 5.2: The two columns show different sample sizes with the same proportion of heads. The prior is the same in both columns but plotted on a different vertical scale. The influence of the prior is overwhelmed by larger samples, in that the peak of the posterior is closer to the peak of the likelihood 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.

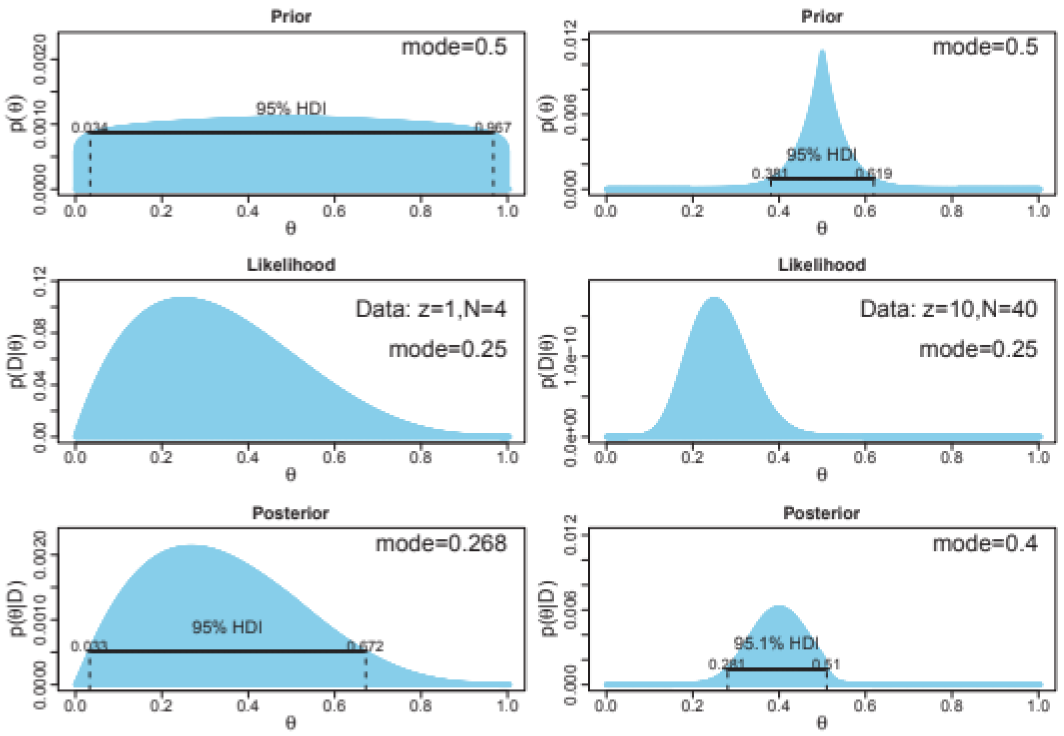


Figure 5.3: The left side is the same small sample as the left side of Figure 5.2 but with a flatter prior. The right side is the same larger sample as the right side of Figure 5.2 but with a sharper prior. Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and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

La inferencia Bayesiana puede ser difícil

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

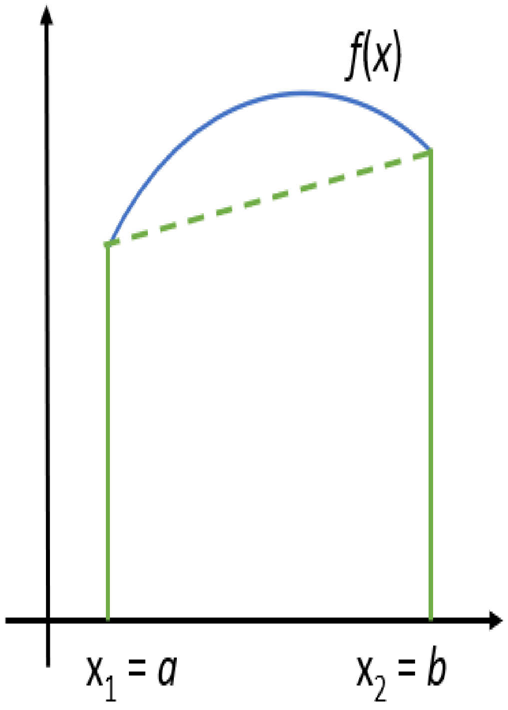
Posterior = verosimilitud * a priori / evidencia

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

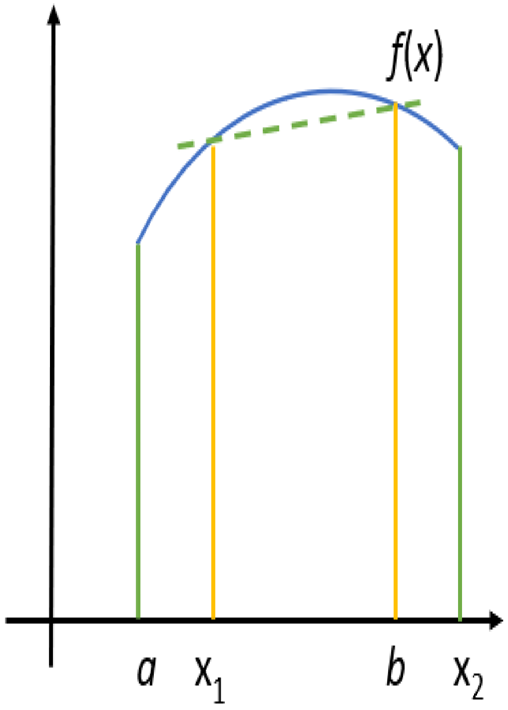
Si tenemos DOS parámetros con 1000 valores cada uno tenemos:

1,000^2 combinaciones de parámetros

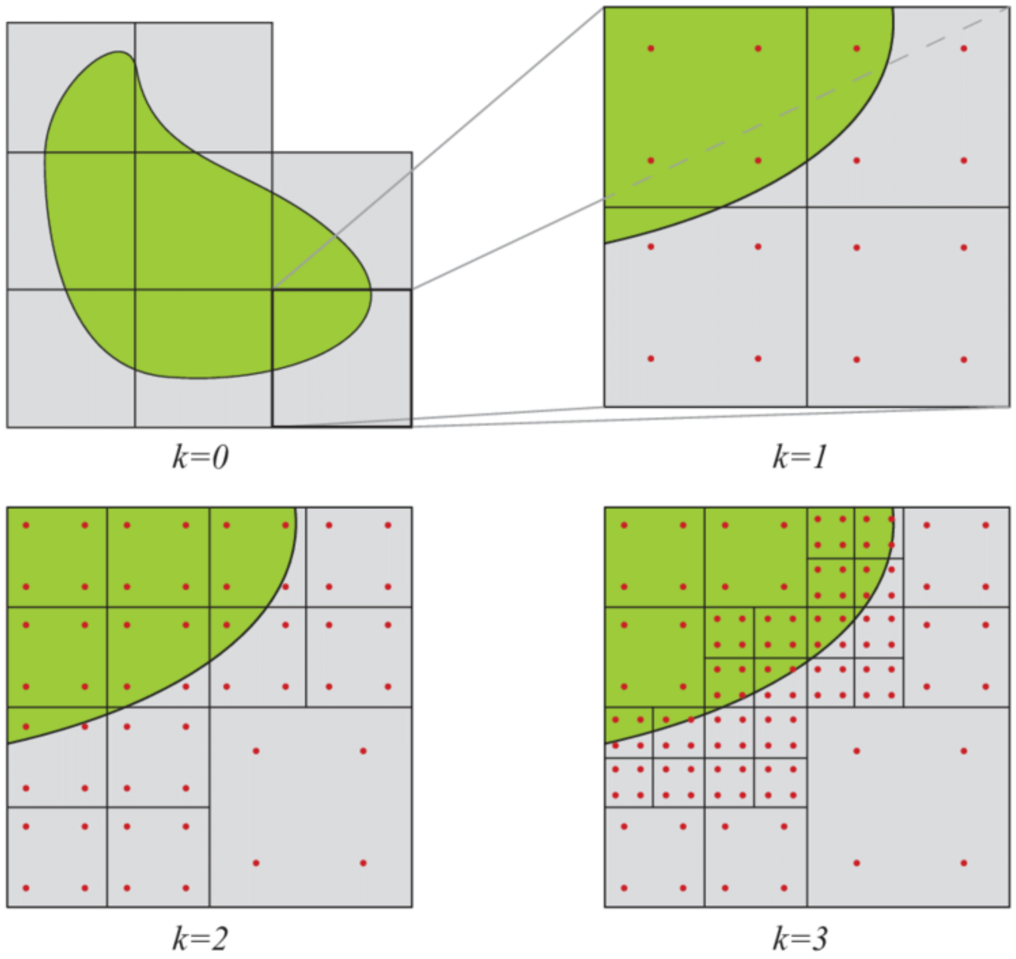
Soluciones (¿?)



(1) Trapezoidal Rule



(2) Gaussian Quadrature

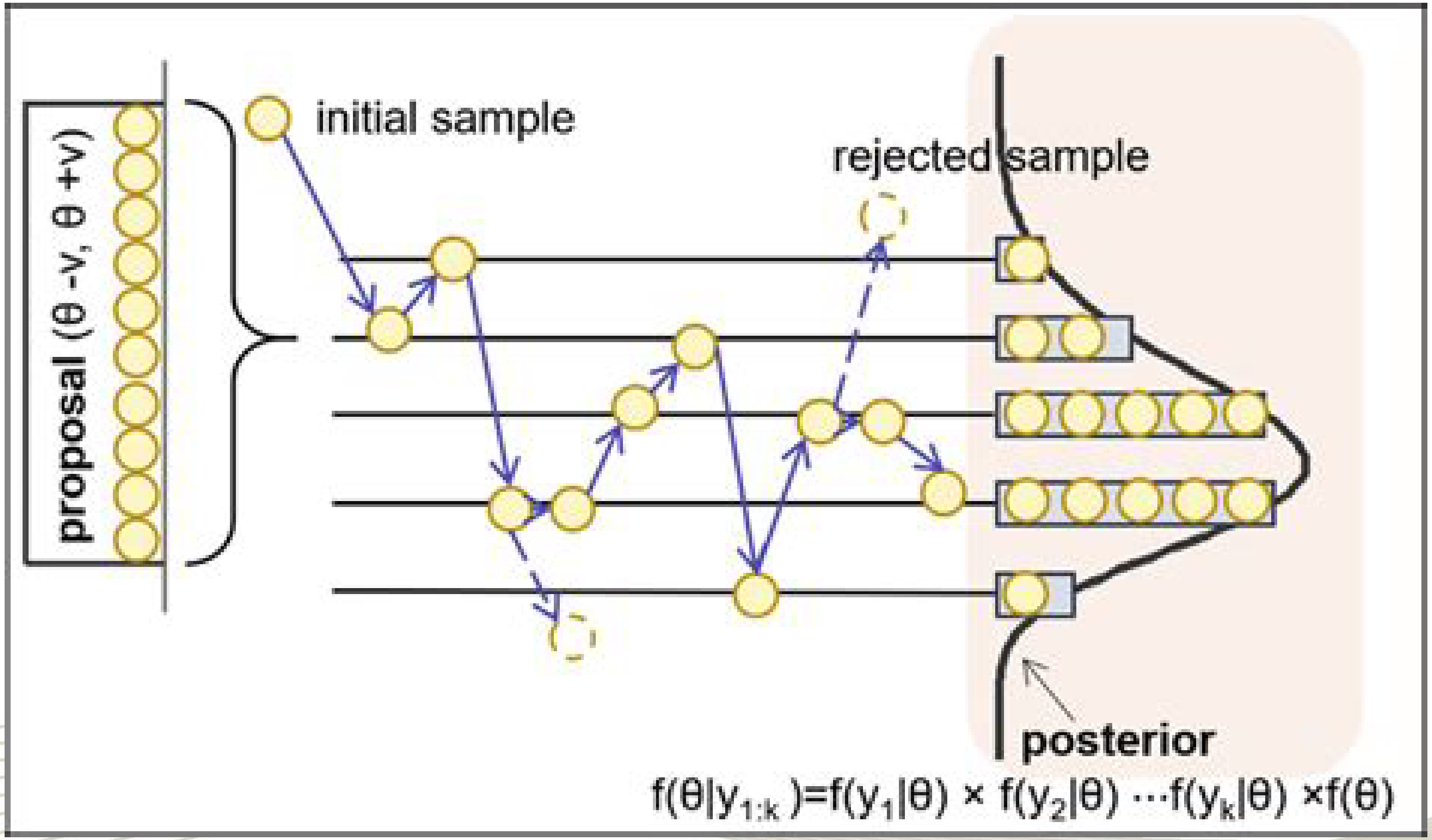


Calculando la posterior

- Solución analítica (a menudo imposible)
- Aproximación de rejilla (muy intensiva)
- Aproximación por cuadratura (limitada)
- Métodos de Monte Carlo basados en cadenas de Markov (intensiva)

Métodos estocásticos (de Monte Carlo)

No se trata más que de una manera de contar



Modestia bayesiana

- Sin garantías excepto que es lógico (matemáticamente correcto)
- La teoría de la probabilidad es un método para deducir implicaciones lógicas de los **datos bajo supuestos que usted debe escoger** (pasar el escrutinio de la audiencia más escéptica)
- Cualquiera que les venda algo diferente está ocultando sus supuestos (**sin modelo no se puede calibrar la incertidumbre**)

Sugerencias

- Sanderson, G. [3blue1brown]. (2019, December 22)The quick proof of Bayes' theorem [Video]. YouTube. https://youtu.be/U_85TaXbeIo?si=qOCChVC7RXrZOWg7
- Sanderson, G. [3blue1brown]. (2019, December 22) Bayes theorem, the geometry of changing beliefs [Video]. YouTube. https://youtu.be/HZGCoVF3YvM?si=QPzRU3LX-Qa0_X7J

Siguiente sesión

- Capítulo 3: Muestreando el imaginario. *Statistical rethinking: A Bayesian course with examples in R and Stan,* Chapman and Hall/CRC, 2020
- Capítulo 7: Markov Chain Monte Carlo. *Doing Bayesian analysis*. John K. Kruschke, 2018.

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 (SNI II)

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

¡Bienvenidos estudiantes!