



UNAM
POSGRADO



Programa
Universitario
de Estudios
del Desarrollo
UNAM

¿Qué es esa cosa llamada probabilidad? (en la vida real / investigación empírica)

Dr. Héctor Nájera
Dr. Curtis Huffman

De la sesión anterior

- **Dos** ideas fundacionales del análisis Bayesiano
 - Reubicación de **credibilidad** entre **posibilidades**
 - Las **posibilidades** en clave parámetros de modelos probabilísticos
- Ambas dan forma a un proceso similar al de **Exoneración judicial**
 - «¿Cuántas veces le he dicho que una vez eliminado lo imposible, lo que queda debe ser la verdad, por improbable que parezca?»
Sherlock Holmes a Watson.

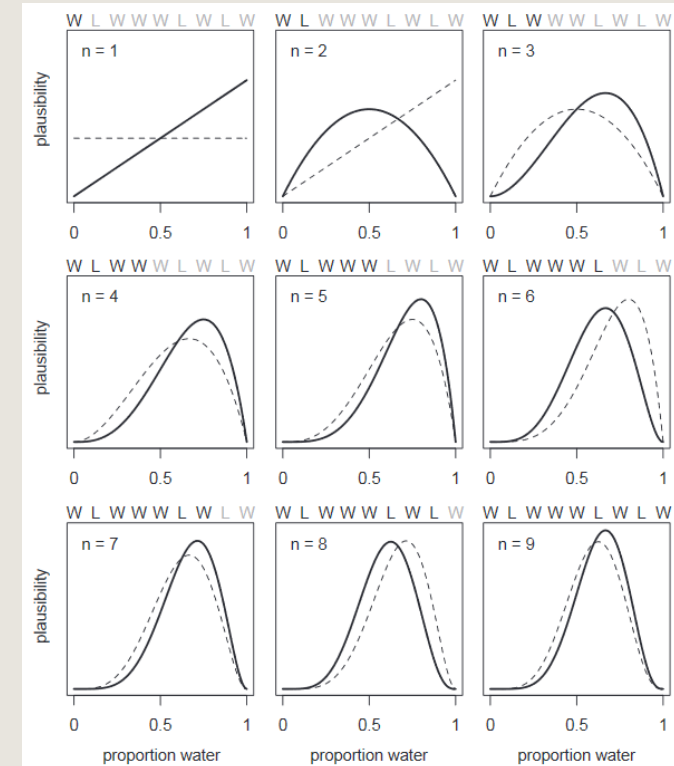


FIGURE 2.5. How a Bayesian model learns. Each toss of the globe produces an observation of water (W) or land (L). The model's estimate of the proportion of water on the globe is a plausibility for every possible value. The lines and curves in this figure are these collections of plausibilities. In each plot, previous plausibilities (dashed curve) are updated in light of the latest observation to produce a new set of plausibilities (solid curve).



Nos quedamos en...

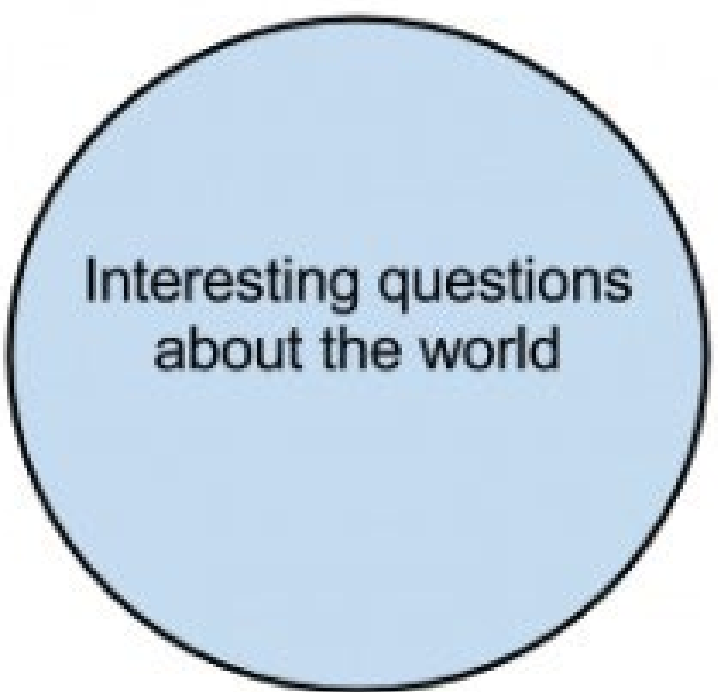
Intervalo creíble

VS

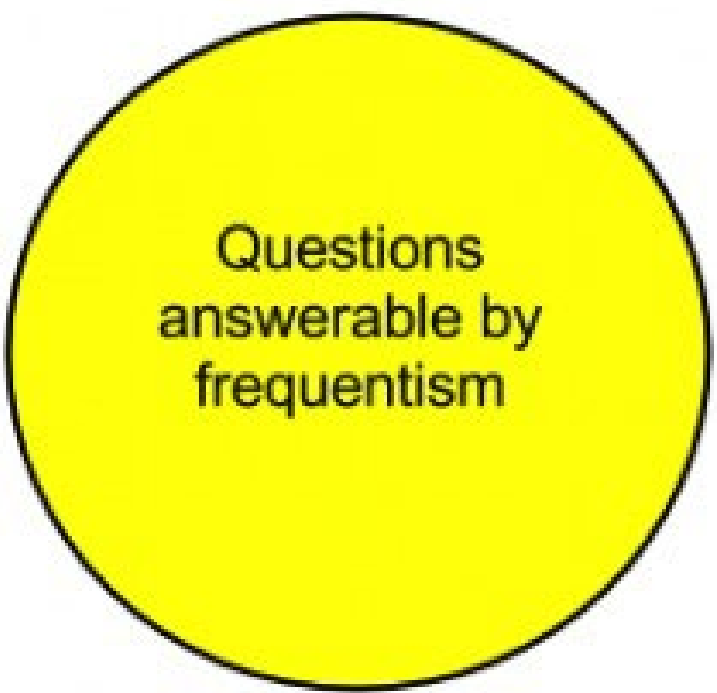
Intervalo de confianza

A series of thin, light-colored wavy lines that sweep across the bottom of the slide, creating a sense of motion or a decorative base.

Conditional on randomness?

A large light blue circle with a black outline, containing the text "Interesting questions about the world".

Interesting questions
about the world

A large yellow circle with a black outline, containing the text "Questions answerable by frequentism".

Questions
answerable by
frequentism

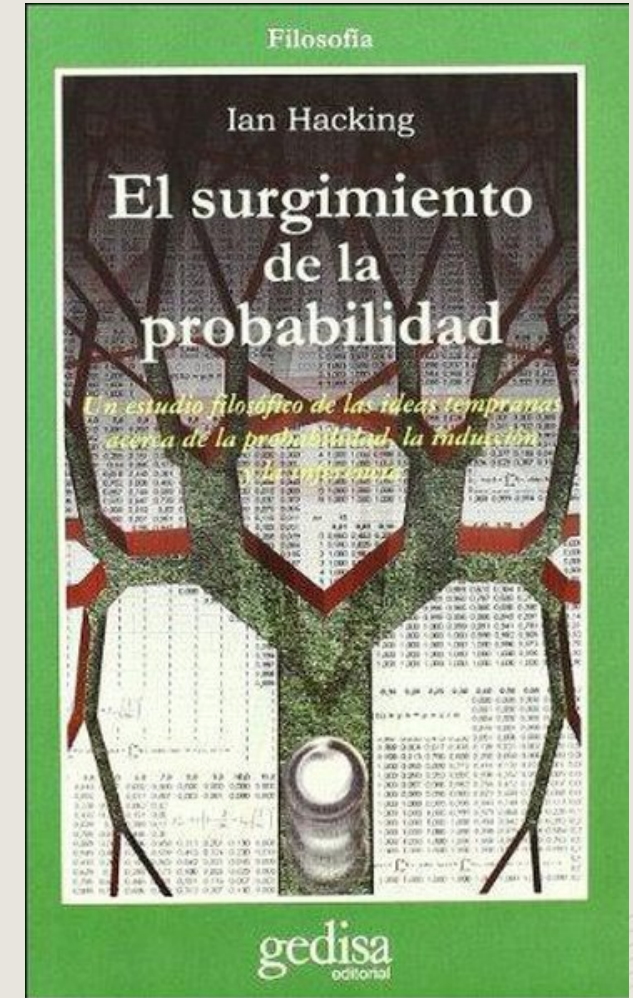


HOY: Capítulo 4

¿Qué es esa cosa llamada probabilidad?

La dualidad de la probabilidad

- La dualidad de la probabilidad no es noticia
- Desde mediados del siglo XVII (época de Pascal) la probabilidad tiene dos aspectos.
 - Por un lado es epistemológica, conectada con el grado de creencia garantizado por la evidencia (que de ningún modo son cuestiones de azar).
 - Por otro lado le conciernen las leyes estocásticas de proceso aleatorios, conectada con la tendencia, exhibida por algunos dispositivos de azar (diseños experimentales, aleatorizadores), a producir frecuencias relativas estables en largas sucesiones de ensayos repetidos.
 - La tendencia de una moneda a salir cara es una propiedad de la moneda tanto como de su masa y la frecuencia estable de una serie larga contabilizada de experimentos repetidos es un factor objetivo de la naturaleza, independientemente del conocimiento que cualquiera tenga de él o de su evidencia.



La incertidumbre se mide en términos de probabilidad

- Máxima de la estadística: Todos los modelos están mal, pero algunos son más útiles que otros (Box, 1976)
 - **2.3 Parsimony**

Since all models are wrong the scientist cannot obtain a "correct" one by excessive elaboration (...) Just as the ability to devise simple but evocative models is the signature of the great scientist so *overelaboration and overparameterization is often the mark of mediocrity.*
 - **2.4 Worrying Selectively**

Since all models are wrong the scientist must be alert to what is importantly wrong. It is inappropriate to be concerned about mice when there are tigers abroad.



George Box
1919 – 2013

- La estadística es el arte de modelar la incertidumbre
- “La incertidumbre se mide en términos de probabilidad” (Kruschke, 2015, p.71)
- Este capítulo nos habla de las ideas básicas de probabilidad.
- “Probability is relative, in part to this ignorance, in part to our knowledge” Laplace



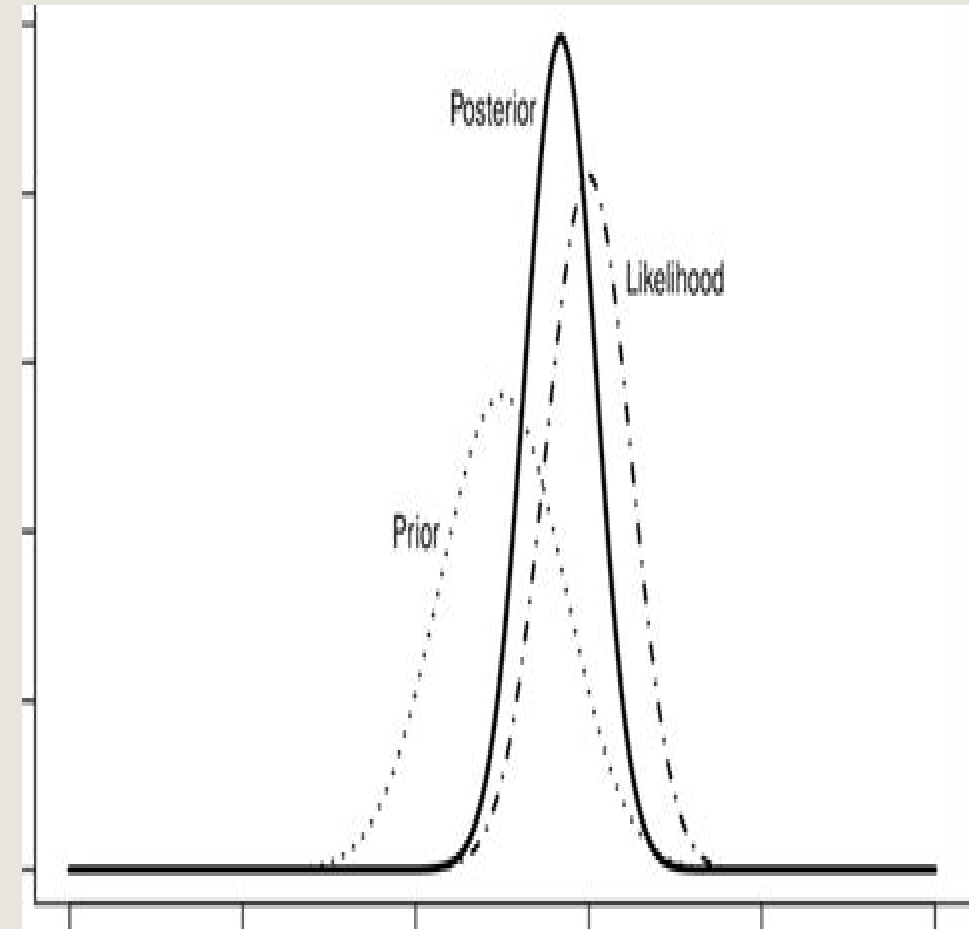
Cuatro aspectos:

I. El set de todas las posibilidades

II. La probabilidad dentro y fuera de tu cabeza

III. Distribuciones de probabilidad

?



El set de todas las posibilidades: Nociones

- La incertidumbre la medimos en términos de probabilidad
 - Qué tan probable es que mi $H(\theta)$ se sostenga?
- Qué tan probable es H considerando las alternativas?
- Siempre que nos preguntamos qué tan probable es un evento, lo que tenemos en mente son todos los posibles resultados
- Esos posibles resultados se ubican en el **espacio de muestreo**
- Ese espacio no es trivial de construir:
Medición



El set de todas las posibilidades: Nociones

En realidad hay dos espacios de muestreo

- El espacio de los resultados
- El espacio de nuestras creencias sobre el posible resultado

$P(\theta|D)$ El espacio de la observación de los resultados

$p(\theta = .5) = .99$ Nuestra creencia de que la moneda sea justa



El set de todas las posibilidades: Nociones

¿Funciona la vacuna?

¿Qué tan probable es que una persona en México que recibió la vacuna fallezca por COVID-19?

$p(\theta|D)$

- considerando $p(\theta) = .92$ —estudios en Europa, la posterior de los estudios europeos se usa como a priori—
- ¿Pero es .92 con $CrI \ +/-0$?

Los eventos posibles para la pregunta sobre México en este caso se determinan considerando los datos mexicanos

¿Cómo se generan los datos?



Probabilidad ¿Dentro o afuera de tu cabeza?

Dentro: Cuando hablamos en términos de nuestra creencia sobre una hipótesis

Hablamos del grado de creencia: No pensamos

$p(\theta) = .92$; Sino $p(\theta = .92) = .6$

Fuera: Cuando vía experiento estimamos la incertidumbre –reubicamos la creencia sobre $p(\theta|D)$. Pensamos en cómo se generan los datos observados.

TODO BAJO CIERTAS REGLAS QUE VEREMOS A CONTINUACIÓN



Probabilidad: ¿en el mundo allá afuera?

- ¿A qué nos referimos con probabilidad cuando lo aplicamos a la “vida real”/ investigación empírica?
- Andrei Kolmogorov (1933): Probabilidad es una medida de los conjuntos en un **espacio abstracto de eventos** (The set of all posible events)

DEFINICIÓN. (MEDIDA DE PROBABILIDAD). Sea (Ω, \mathcal{F}) un espacio medible. Una medida de probabilidad es una función $P : \mathcal{F} \rightarrow [0, 1]$ que satisface

1. $P(\Omega) = 1$.
2. $P(A) \geq 0$, para cualquier $A \in \mathcal{F}$.
3. Si $A_1, A_2, \dots \in \mathcal{F}$ son ajenos dos a dos, esto es, $A_n \cap A_m = \emptyset$ para $n \neq m$, entonces $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$.

<http://lya.fciencias.unam.mx/lars/libros/cip.pdf>



Андрей Николаевич Колмогоров
1903-1987

¿El conjunto de todos los eventos posibles?

- En la teoría axiomática de la probabilidad de Kolmogorov se supone que hay un espacio abstracto de elementos llamados “eventos”.
 - Conjuntos de eventos pueden medirse de la misma forma que se mide un área/volumen.
 - Si esta medida sobre el espacio abstracto cumple con los axiomas anteriores (1-3), es una medida de probabilidad.



Андрей Николаевич Колмогоров
1903-1987

Probabilities assign numbers to possibilities



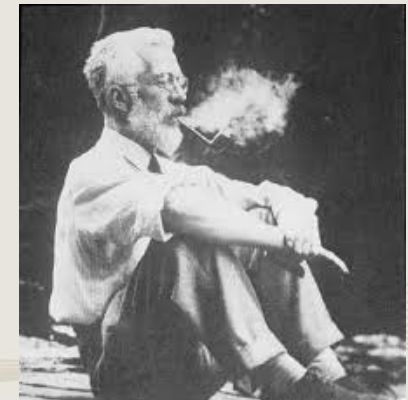
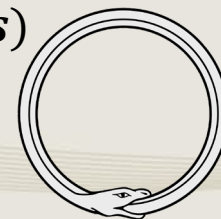
Probabilidad: ¿en el mundo allá afuera?

- Para **calcular** probabilidades usando la teoría de la probabilidad (probabilidad matemática) en la “vida real” (investigación empírica) es necesario identificar/describir el espacio de eventos con suficiente precisión.
- A menos que podamos identificar/describir el espacio abstracto de Kolmogorov, los enunciados probabilísticos que emerjan del análisis estadístico pueden tener muchos significados diferentes, algunos contradictorios.

Probabilidad: ¿en el mundo allá afuera?

- Karl Pearson, antes de Kolmogorov, creía que las distribuciones de probabilidad eran observables meramente recolectando datos.
- William Gosset intentó describir el espacio de eventos para un experimento (el conjunto de todos los posibles resultados de ese experimento. ¡Ingenioso!, pero inútil pues hay que describir a detalle la distribución de todos los resultados posibles del experimento).
- R.A. Fisher desarrolló la intuición de Gosset en sus diseños experimentales donde un “tratamiento” se asigna **aleatoriamente** a las unidades de experimentación/análisis. El espacio de eventos es el conjunto (**finito**) de todas las posibles asignaciones aleatorias (con igual probabilidad) que pudieron haber tenido lugar: la probabilidad como frecuencia en un experimento aleatorio.
- Lamentablemente no se puede aplicar a todos los problemas: ¿Qué hay entonces de los estudios observacionales?, ¿sin manipulación no hay probabilidad?

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



Strong repeated sampling principle

- Repeticiones hipotéticas del “experimento” que generó los datos.
- Aproximar la probabilidad por la frecuencia relativa de ocurrencia en el largo plazo

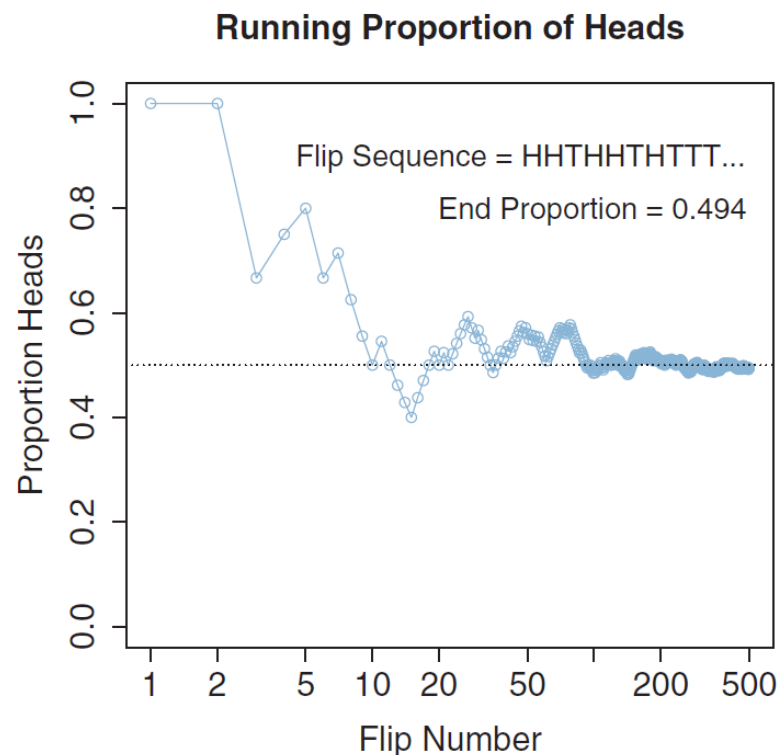


Figure 4.1 Running proportion of heads when flipping a coin. The x-axis is plotted on a logarithmic scale so that you can see the details of the first few flips but also the long-run trend after many flips. R code for producing this figure is discussed in [Section 4.5](#).

Pro tip!

La computación
Bayesiana es el
arte de explorar
el espacio de
muestreo.

¿Espacio de eventos en la vida real?



- ¿Cuál es el espacio de eventos de un experimento?
 - El conjunto de todos los posibles resultados (Gosset)
 - Bajo asignación aleatoria, el conjunto de todas las posibles asignaciones (Fisher)
- El espacio de eventos también está bien definido en la teoría de encuestas por muestreo
- ¿Cuál es el espacio de eventos bajo medición cuando se dice que la probabilidad de lluvia mañana es del 95 %?
 - El conjunto de todas las personas que saldrán mañana, ¿95 % de ellos habrán de mojarse?
 - El conjunto de todos los momentos en el tiempo, ¿95 % del tiempo caerá agua?
 - El conjunto de todos los metros cuadrados en la región, ¿95 % de éstos habrán de mojarse?

¿Dentro o fuera de nuestra cabeza?

- “But here, finally, we seem to have met our problem, since this may be done only in a very few cases and almost nowhere other than in games of chance the inventors of which in order to provide equal chances for the players, took pains to set up so that the numbers of cases would be known and so that all the cases could happen with equal ease.... But what mortal will ever determine, for example, the number of diseases ---these and other such things depend upon causes completely hidden from us ---.”

¿Es difícil suponer que el número de enfermedades se mantiene constante en el curso del tiempo!



Jakob Bernoulli
1655-1705

¿Dentro o fuera de nuestra cabeza?

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

¿Probabilidad en la vida real?

- Ya sea que la probabilidad sea concebida como una frecuencia relativa de largo plazo de resultados posibles en el mundo-allá-afuera, o como un medio para describir un estado del conocimiento, se comporta de la misma forma matemáticamente.



A. N. Kolmogorov with students

¿Probabilidad en la vida real?

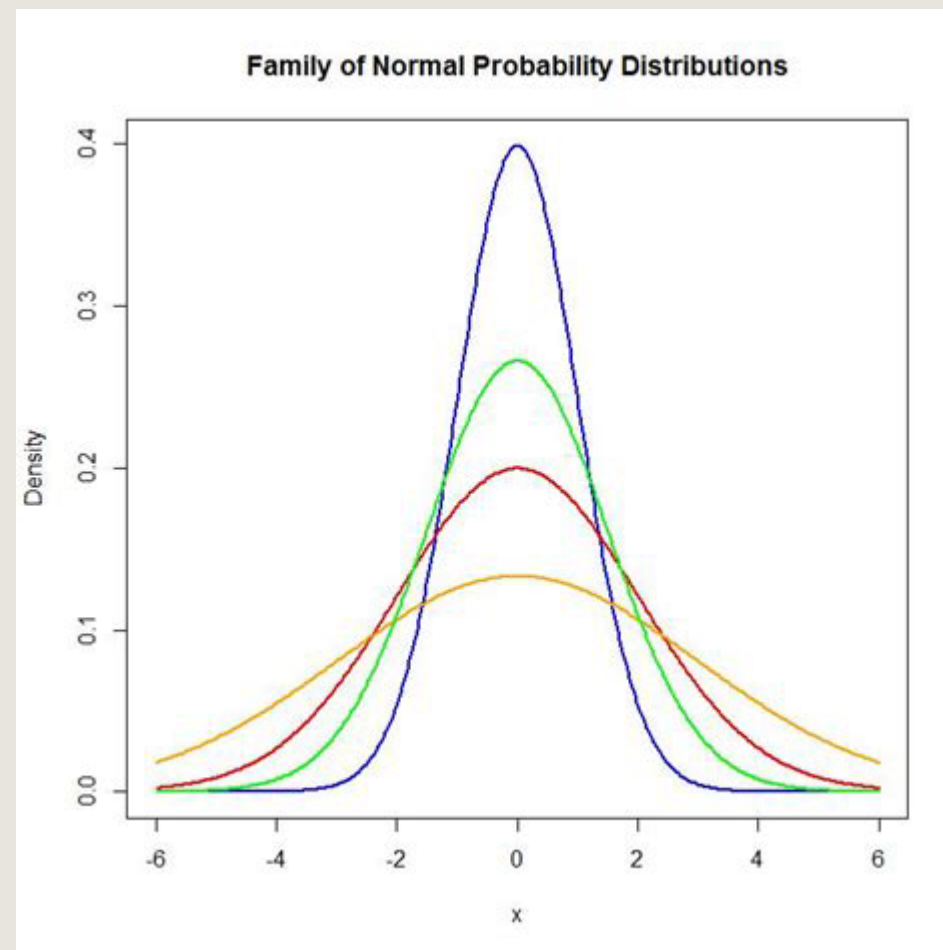
- Kolmogorov dedicó sus últimos años (1965) a describir la probabilidad en términos de las propiedades de secuencias finitas de símbolos, donde la teoría de la información no es el resultado de cálculos probabilísticos sino la progenitora de la probabilidad misma (¿MaxEnt?)



A. N. Kolmogorov with students

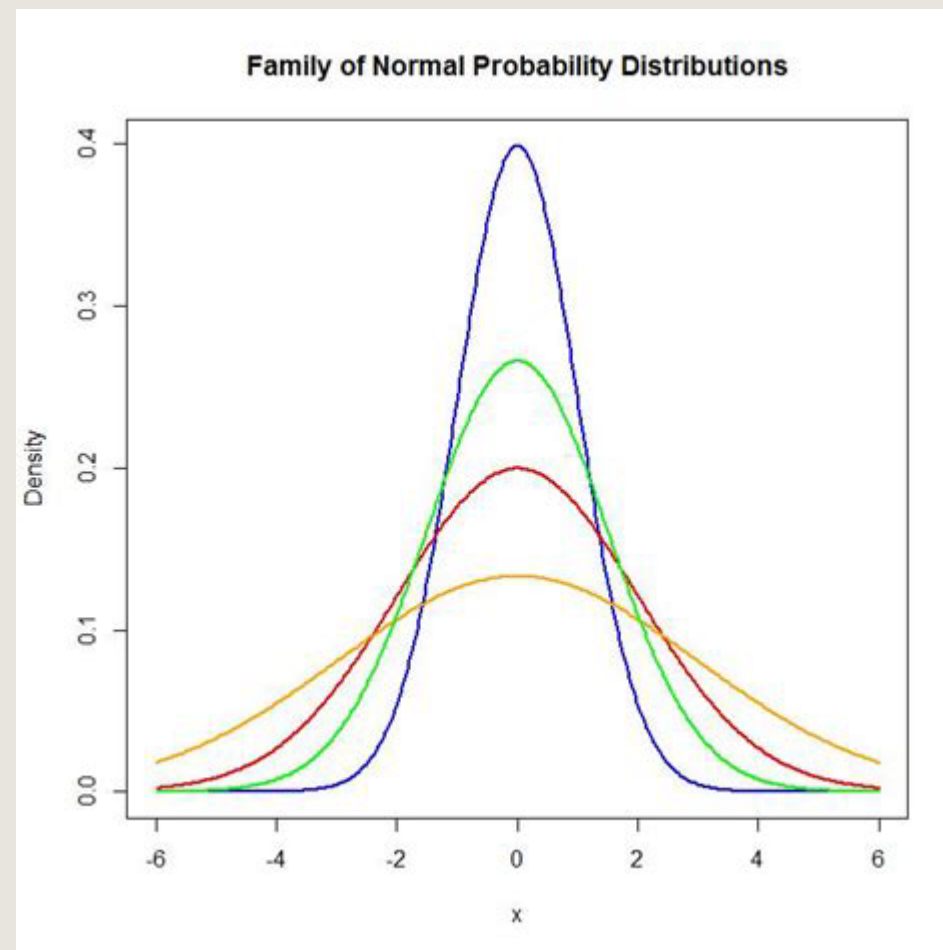
Describiendo nuestras creencias

- La estatura promedio de las mujeres en México es 1.55 pero estás abierto a posibilidades (incertidumbre)
- Puedes asignar probabilidades a distintas medidas 1.40 o 1.70
- Hiperparámetros $\theta \sim N(0, SD)$: El parámetro θ tiene sus propios parámetros
- Distribuciones de probabilidad



Distribuciones de probabilidad

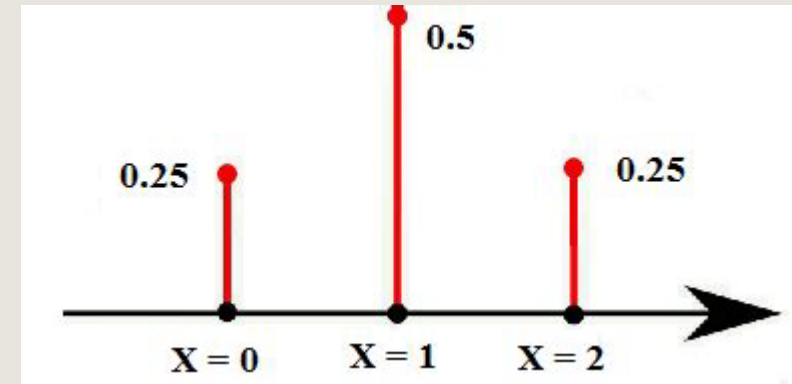
- Las distribuciones de probabilidad se conforman de la lista de todas las posibilidades con sus probabilidades correspondientes
- Resultados/observaciones discretos o continuos?



Distribuciones discretas: Masas

- En el caso de eventos discretos:
Cara/cruz; lluvia/no lluvia;
cáncer/no cáncer;
- Es sencillo atribuirle al evento puntual cierta probabilidad

- Por ejemplo





PUED Distribuciones discretas: Masas y densidades

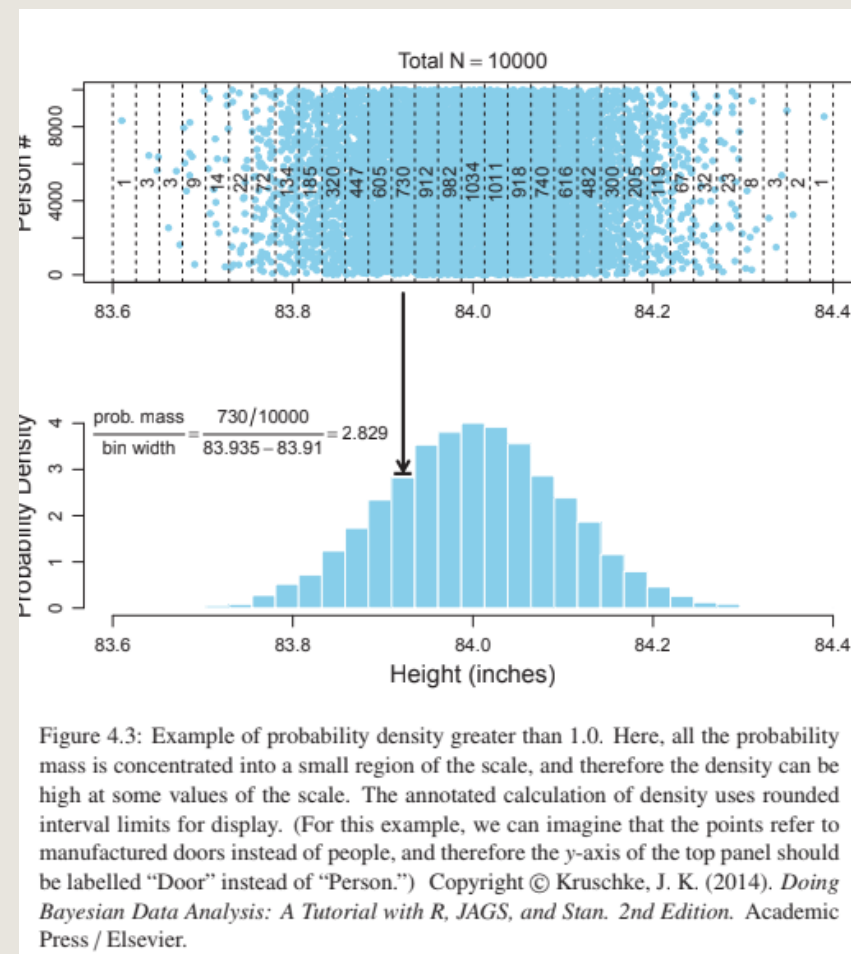
Distribuciones continuas (espacios pequeños)

La probabilidad de 23.534%

La probabilidad de 23?

¿Cuál es el intervalo adecuado?

Hablar en clave de densidades de probabilidad (masa / espacio)

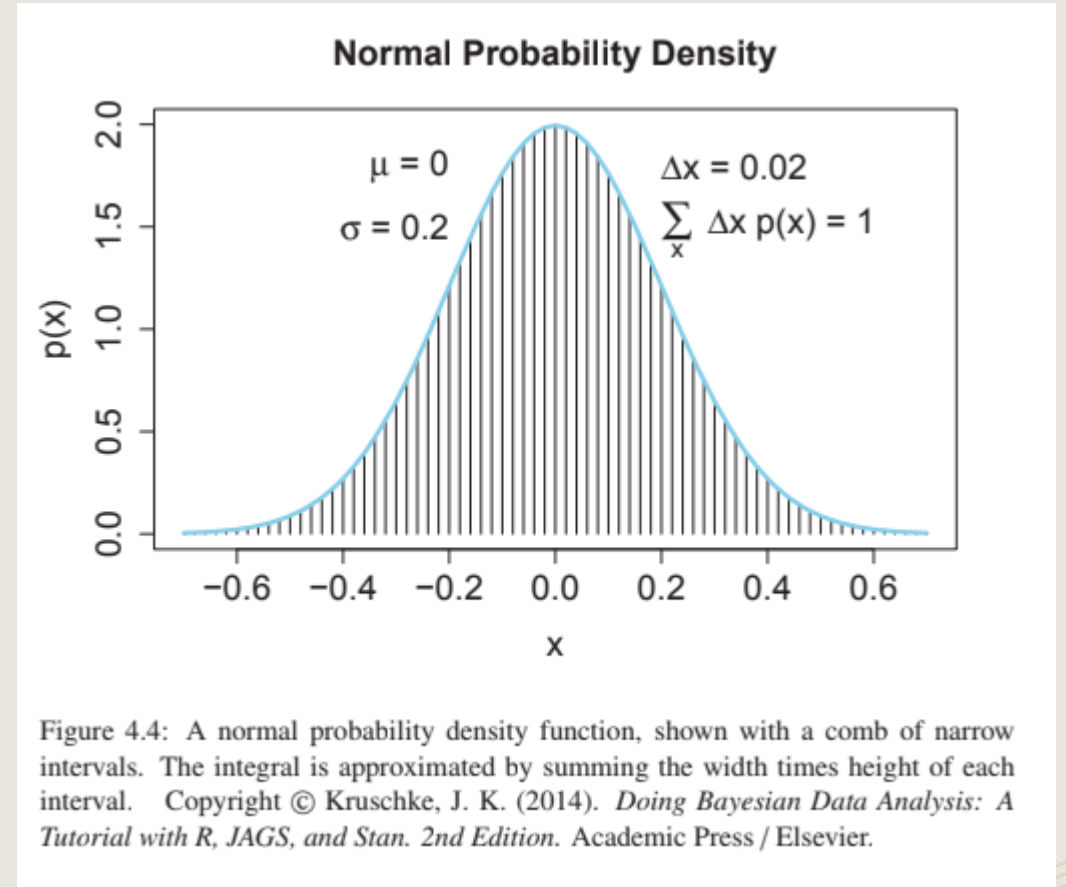


Propiedades

- ¡OK! Si me interesa entonces encontrar la masa y la densidad de una distribución (posterior), eso significa que tengo que calcular el área bajo la curva
- La integral

Probability Density Function

$$F(x) = P(a \leq x \leq b) = \int_a^b f(x)dx \geq 0$$



Intervalo/región de más alta de densidad

- Raramente buscamos conocer la probabilidad de un valor tan puntual. Generalmente nos interesa saber el rango de valores probables de un parámetro

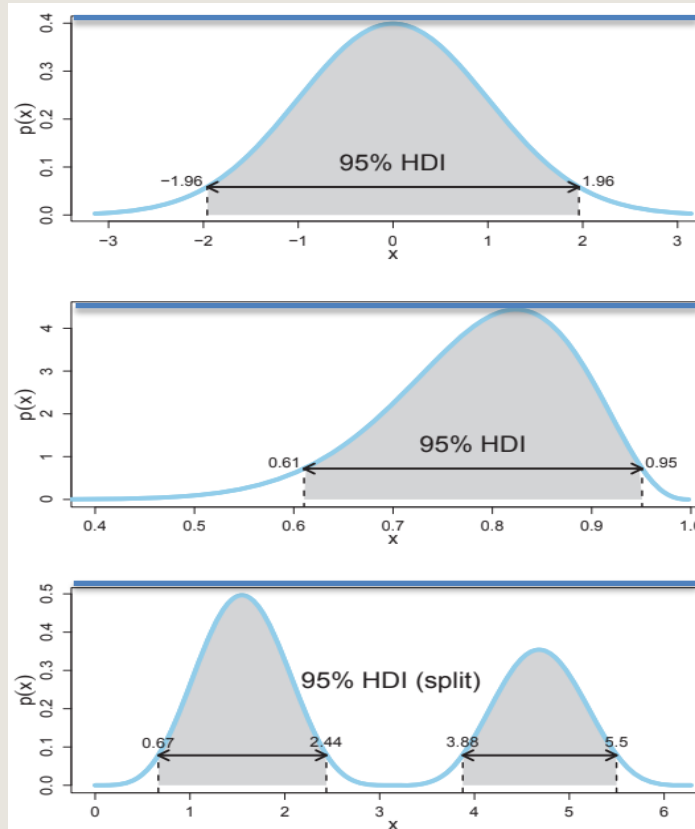


Figure 4.5: Examples of 95% highest density intervals (HDIs). For each example, all the x values inside the interval have higher density than any x value outside the interval, and the total mass of the points inside the interval is 95%. The 95% area is shaded, and it indicates the zone below the horizontal arrow. The horizontal arrow indicates the width of the 95% HDI, with its ends annotated by (rounded) x values.

- Dónde encaja $P(\beta|D)$:

$$Y = a + bX$$

- Y definimos nuestra credibilidad apriori (priors)

$$b \sim N(0,10)$$

- O qué pasa cuando tengo más variables

$$Y = a + \beta_1 X + \beta_2 Z$$

Probabilidad condicional

Table 4.1: Proportions of combinations of hair color and eye color. Some rows or columns may not sum exactly to their displayed marginals because of rounding error from the original data. Data adapted from Snee (1974). 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				Marginal (Eye Color)
	Black	Brunette	Red	Blond	
Brown	.11	.20	.04	.01	.37
Blue	.03	.14	.03	.16	.36
Hazel	.03	.09	.02	.02	.16
Green	.01	.05	.02	.03	.11
Marginal (Hair Color)	.18	.48	.12	.21	1.0

Reubicación de acuerdo al color de cabello

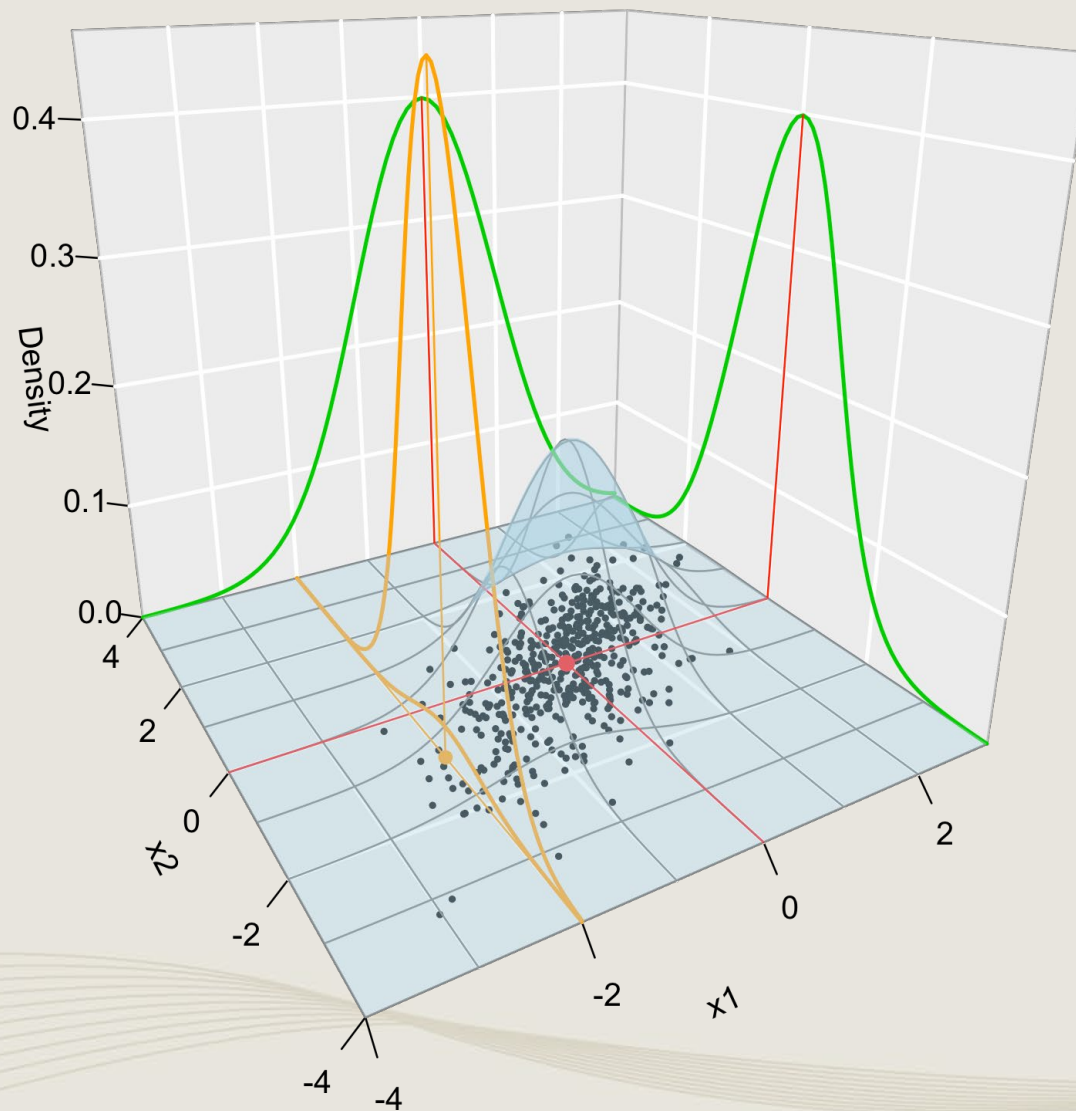
Table 4.2: Example of conditional probability. Of the blue-eyed people in Table 4.1, what proportion have hair color h ? Each cell shows $p(h|\text{blue}) = p(\text{blue}, h)/p(\text{blue})$ rounded to two decimal points. Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with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

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

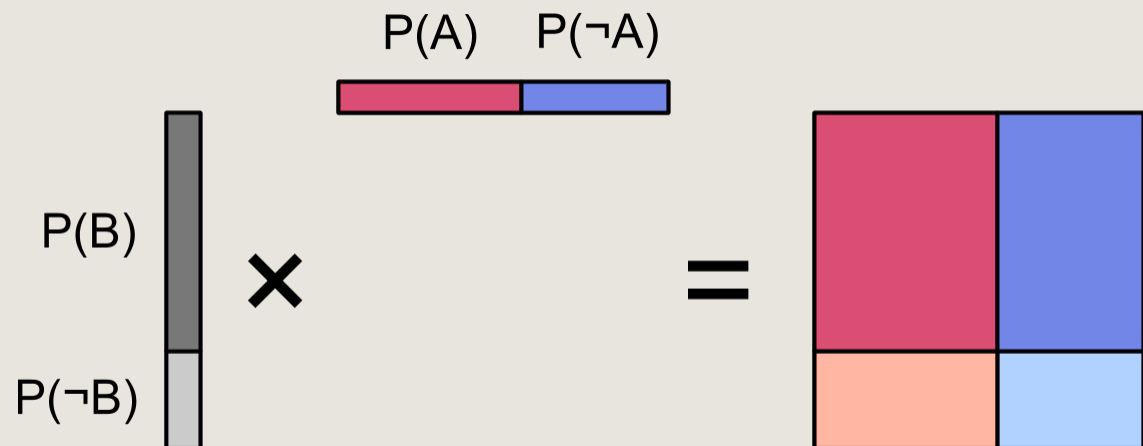
Esto es inferencia bayesiana:

$$P(\beta|D)$$

Probabilidad condicional



Independencia



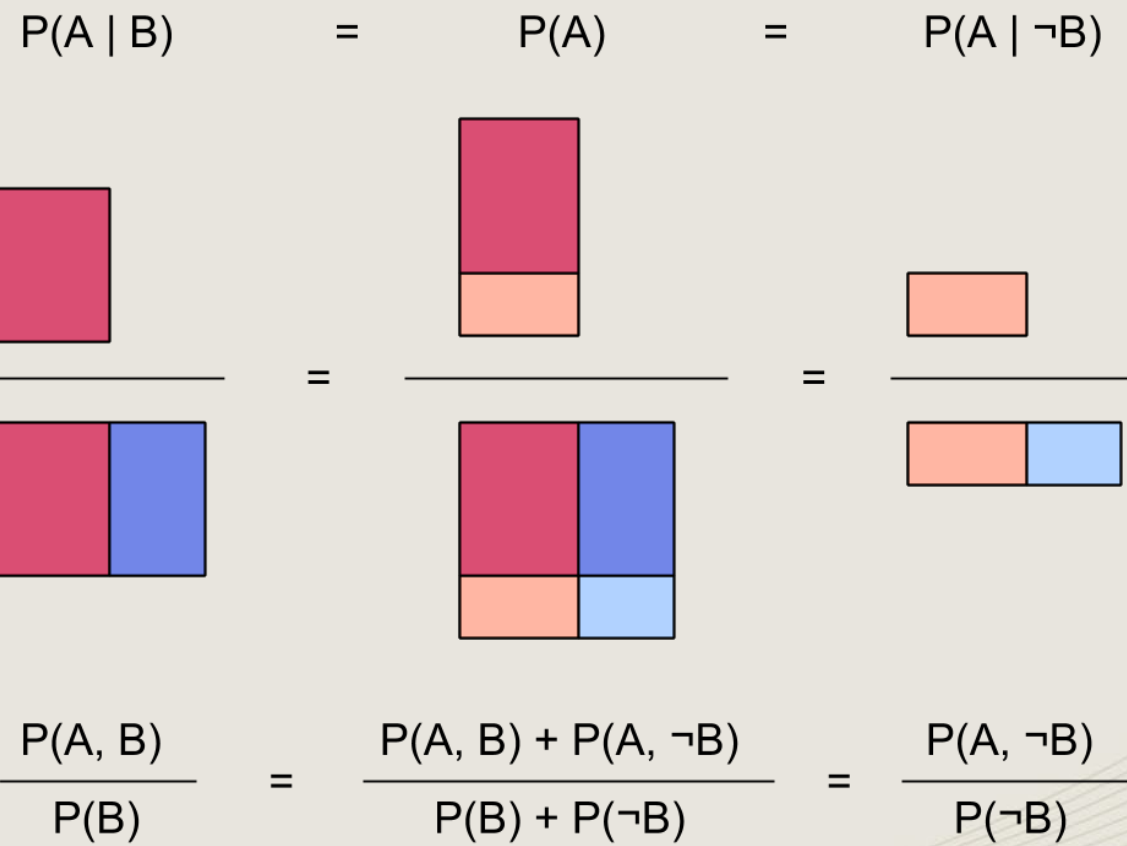
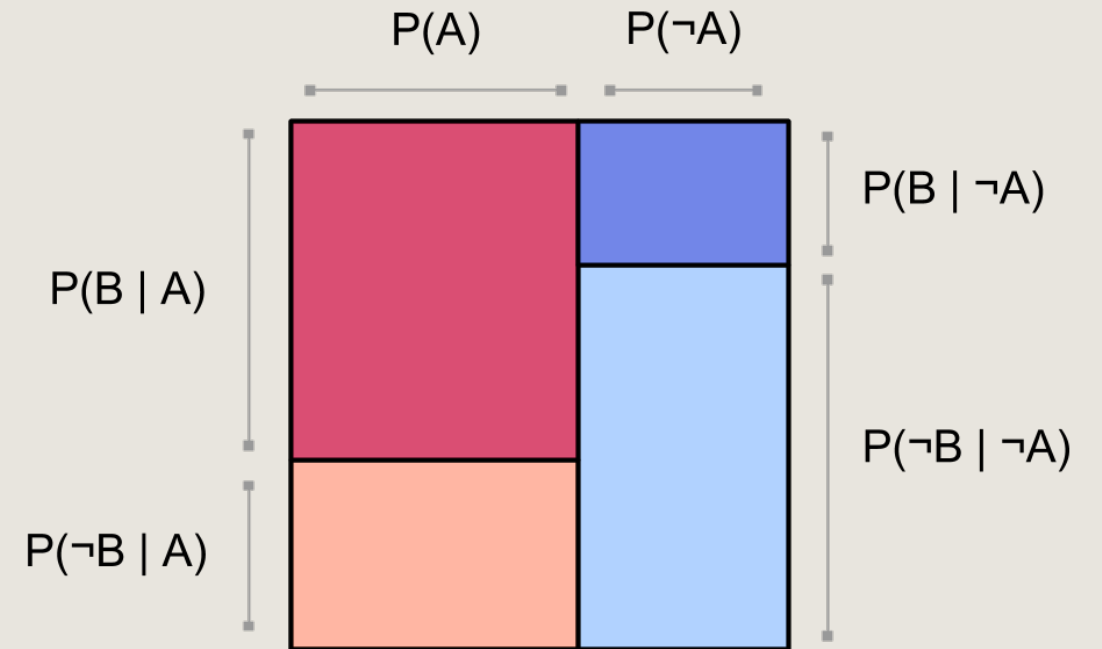
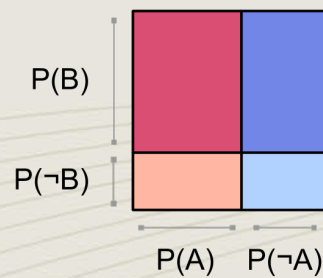
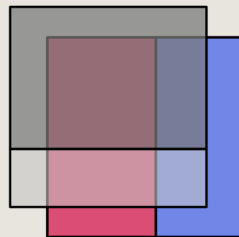
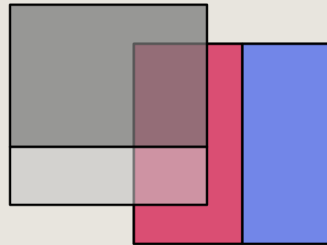
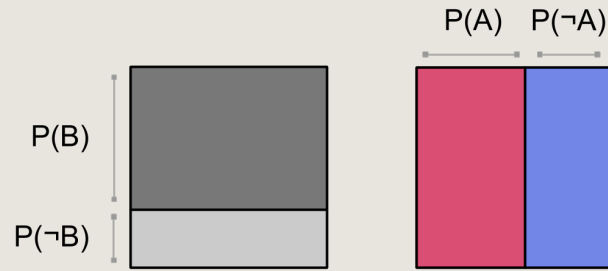
$$P(A | B) = \frac{P(A, B)}{P(B)} = \frac{P(A, B) + P(A, \neg B)}{P(B) + P(\neg B)} = \frac{P(A)}{P(A) + P(\neg A)} = P(A | \neg B)$$


Diagram illustrating the formula for conditional probability:

Top row: $P(A | B)$ = $\frac{P(A, B)}{P(B)}$ = $\frac{P(A, B) + P(A, \neg B)}{P(B) + P(\neg B)}$ = $\frac{P(A)}{P(A) + P(\neg A)}$ = $P(A | \neg B)$

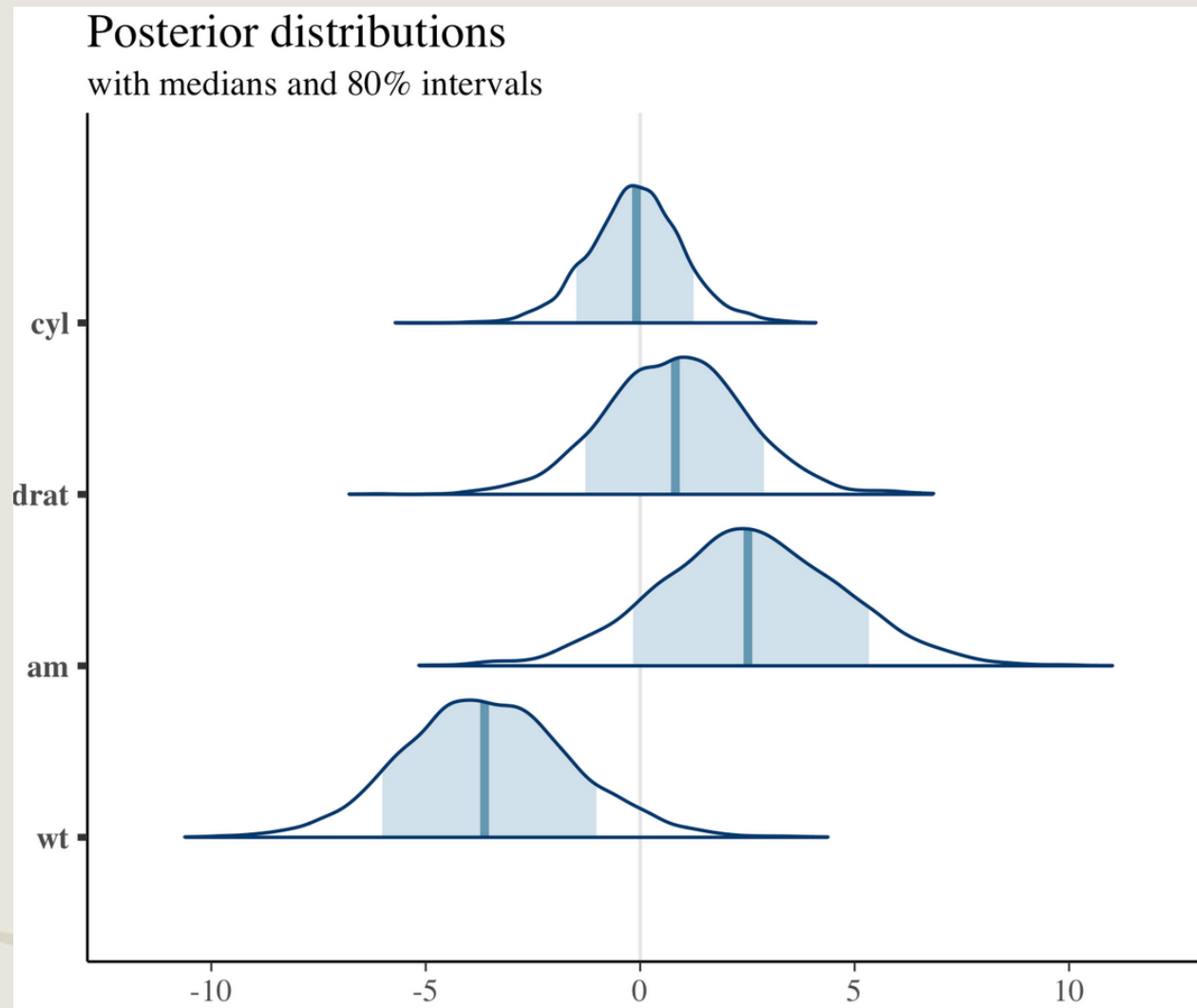
Bottom row: $\frac{P(A, B)}{P(B)}$ = $\frac{P(A, B) + P(A, \neg B)}{P(B) + P(\neg B)}$ = $\frac{P(A, \neg B)}{P(\neg B)}$

Independencia



El reto es...

- Dado un modelo probabilístico condicional (prior, likelihood y datos):
- **Cómo estimo todas las áreas bajo las curvas para múltiples parámetros**



- Capítulo 5.
Bayes' rule
(or Bayes rules)





- 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.
- Hartnack, S., & Roos, M. (2021). Teaching: confidence, prediction and tolerance intervals in scientific practice: a tutorial on binary variables. *Emerging Themes in Epidemiology*, 18(1), 1-14.
- Kerrich, J. E. (1950). *An experimental introduction to the theory of probability*. Belgisk Import Company.
- 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.
- Salsburg, D. (2001). The lady tasting tea: How statistics revolutionized science in the twentieth century. Macmillan.
- 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.
- Shafer, G. (1996). The significance of Jacob Bernoulli's Ars Conjectandi for the philosophy of probability today. *Journal of Econometrics*, 75(1), 15-32.



CONTACTO

Dr. Héctor Nájera y Dr. Curtis Huffman
Investigadores (SIN 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

