¿Qué es esa cosa llamada probabilidad?

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.

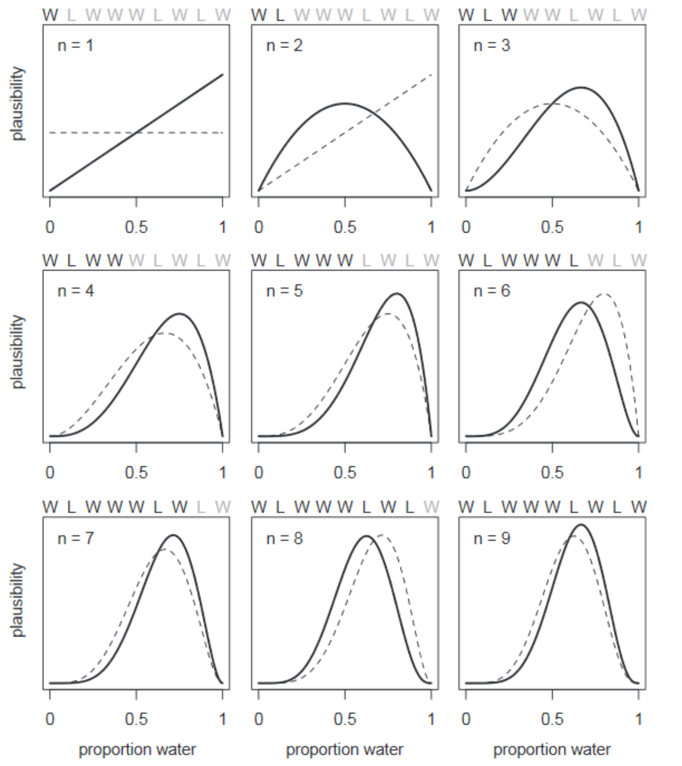


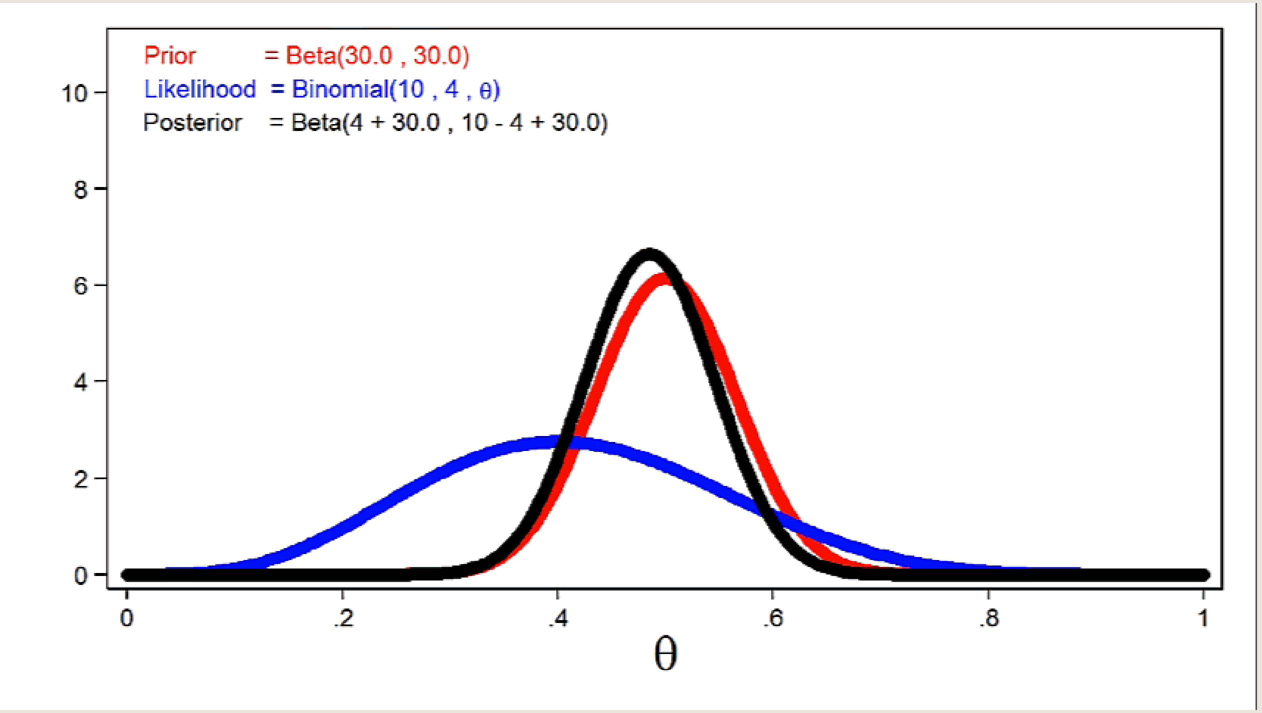
FIGURE 2.5. How a Bayesian model learns. Each toss of the globe produces an observation of water (W) or land (L). The model's estimate of the proportion of water on the globe is a plausibility for every possible value. The lines and curves in this figure are these collections of plausibilities. In each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

La inferencia bayesiana es sobre tres cosas

1. Hipótesis: Explicaciones (Posibilidades)

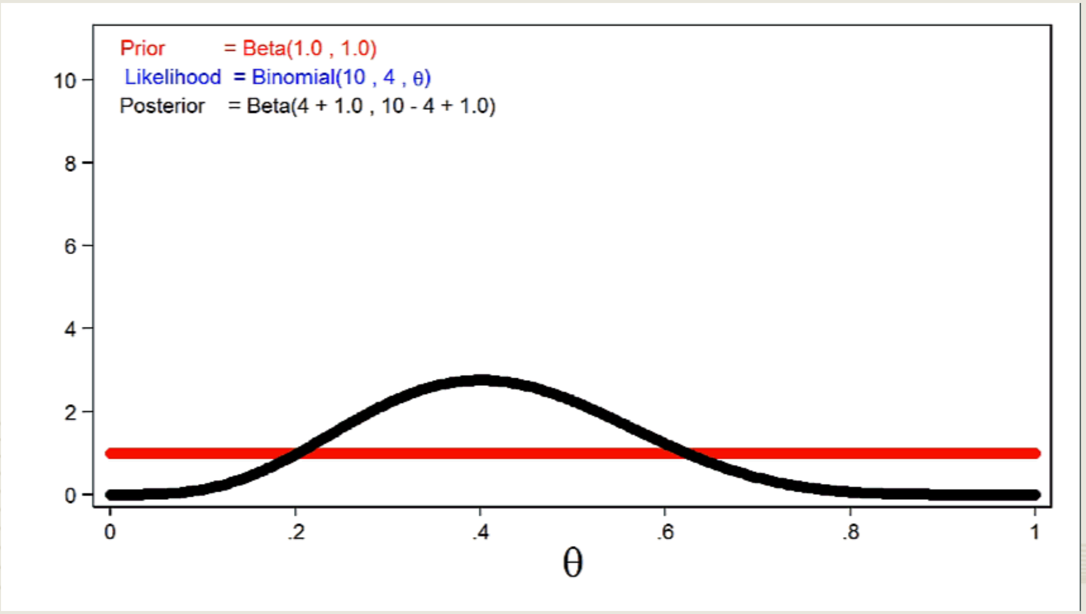
2. Credibilidad de H: Lo que pensamos respecto a H

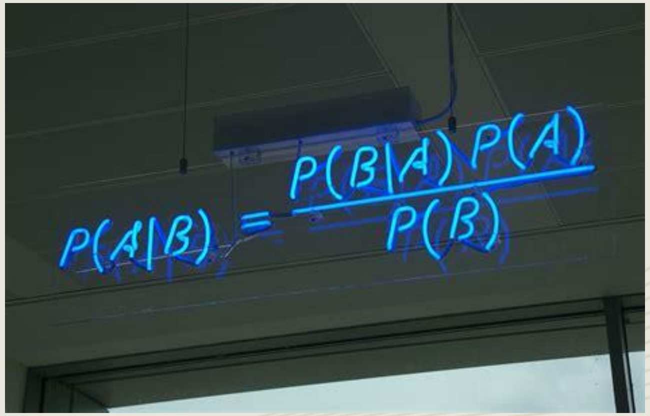
3. La probabilidad de H dados D: La probabilidad de la hipótesis dados los datos



Incetidumbre

- El objetivo de la inferencia es poder hacer afirmaciones de conocimiento con cierto grado de incertidumbre
  - Debemos poder cuantificar el tamaño de la incertidumbre asociada a nuestras afirmaciones
  - La probabilidad de H dados D: La probabilidad de la hipótesis dados los datos

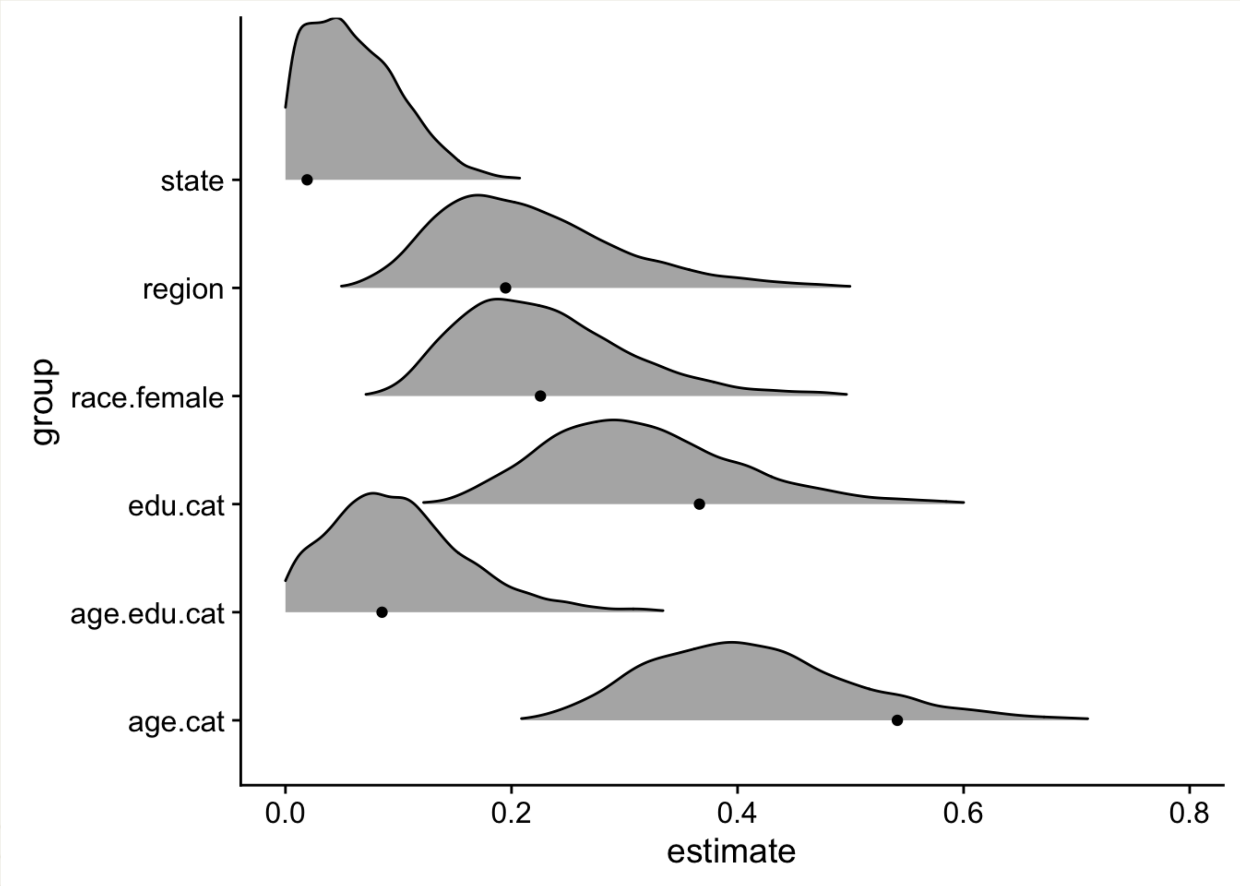




Incertidumbre

La probabilidad de H dados D: La probabilidad de la hipótesis dados los datos

?




```
. regress logdelay temp L.london_se

      Source |       SS           df       MS      Number of obs   =       259
-------------+----------------------------------   F(2, 256)       =     68.07
       Model |  11.8580863         2  5.92904313   Prob > F        =    0.0000
    Residual |  22.2991825       256  .087106182   R-squared       =    0.3472
-------------+----------------------------------   Adj R-squared   =    0.3421
       Total |  34.1572688       258  .132392515   Root MSE        =    .29514

------------------------------------------------------------------------------
    logdelay |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        temp |  -.0180465   .0044372    -4.07   0.000    -.0267845   -.0093086
             |
   london_se |
         L1. |   .1374573   .0130854    10.50   0.000     .1116885    .1632261
             |
       _cons |   .7930698   .0635816    12.47   0.000     .6678603    .9182793
------------------------------------------------------------------------------
```

¿Qué probabilidad tengo acá?

P(D|Theta)

Nos quedamos en…

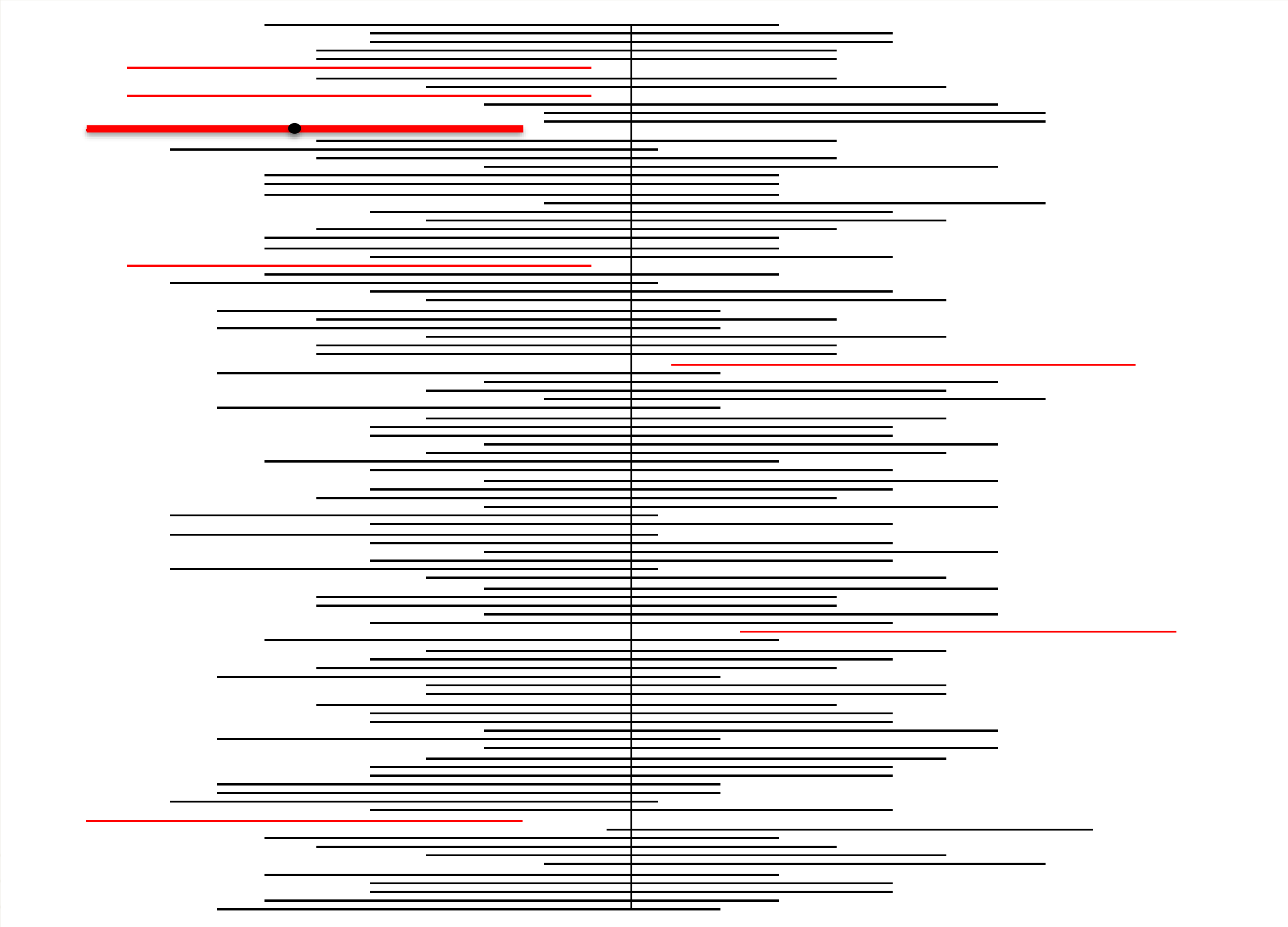
Intervalo creíble

vs

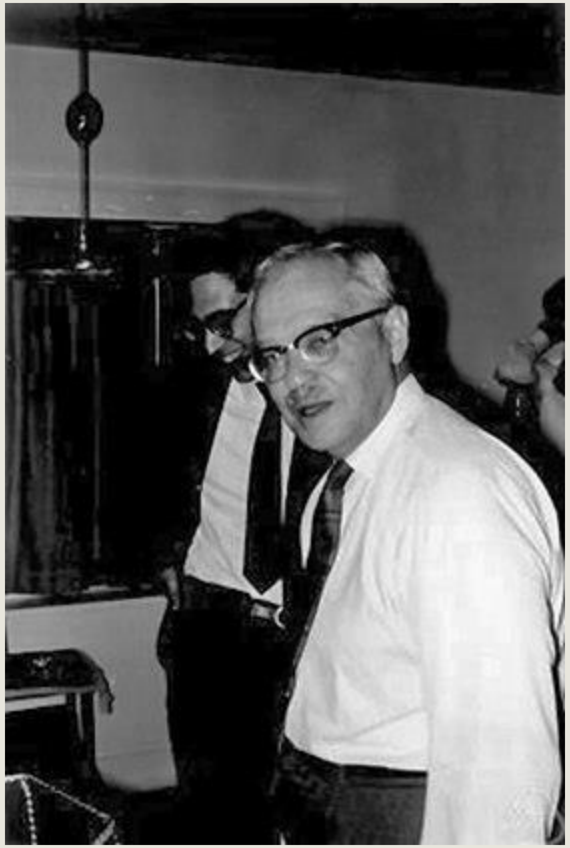
Intervalo de confianza

Incertidumbre en estadística clásica

El Intervalo de confianza es un estimador (de intervalo) **usualmente correcto**, que no tiene por qué ser creíble en cada ocasión que se use.



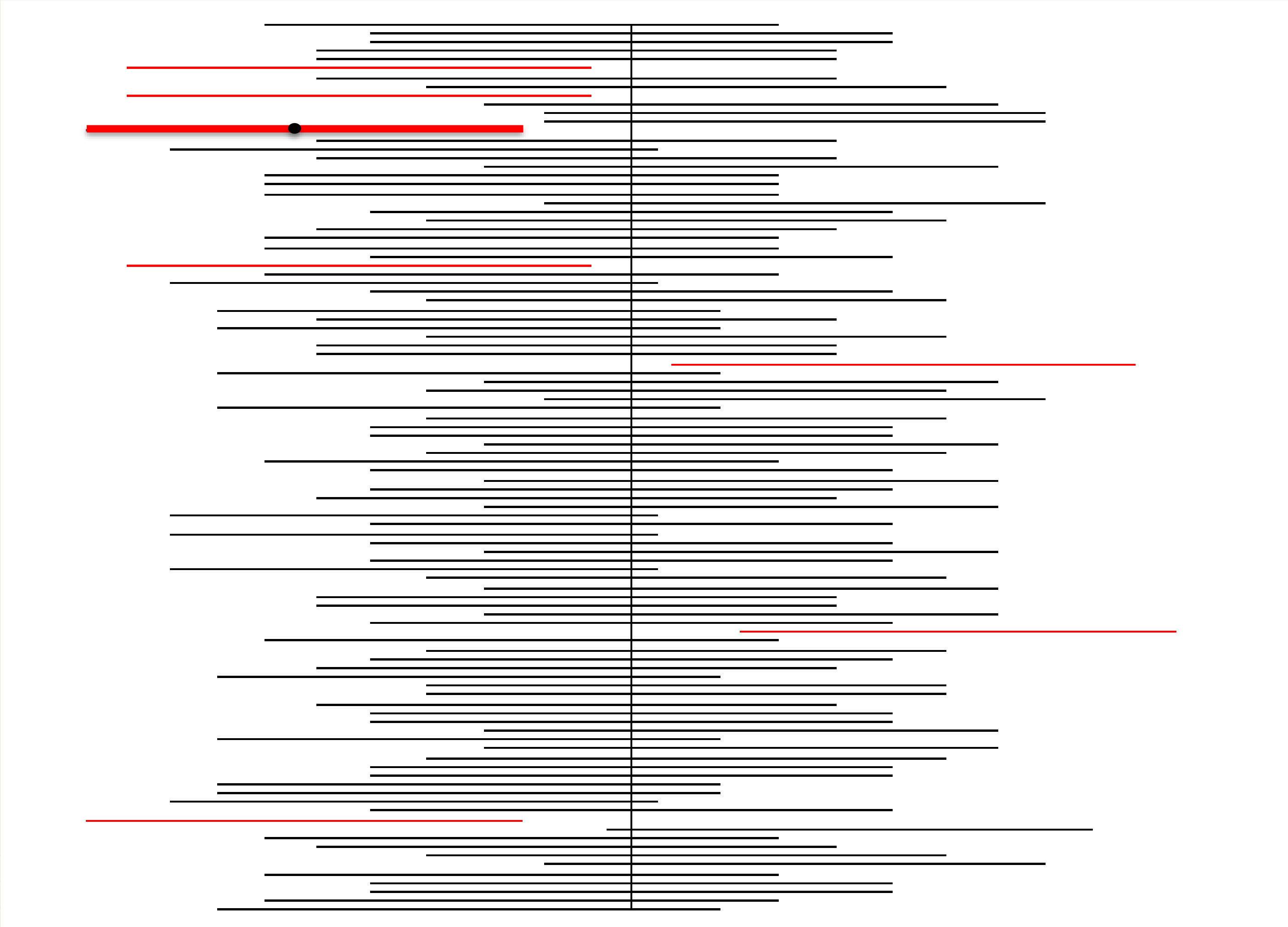
Hartnack y Roos (2021)



Jerzy Neyman 1894-1981
(foto de 1969)

Incertidumbre en estadística clásica

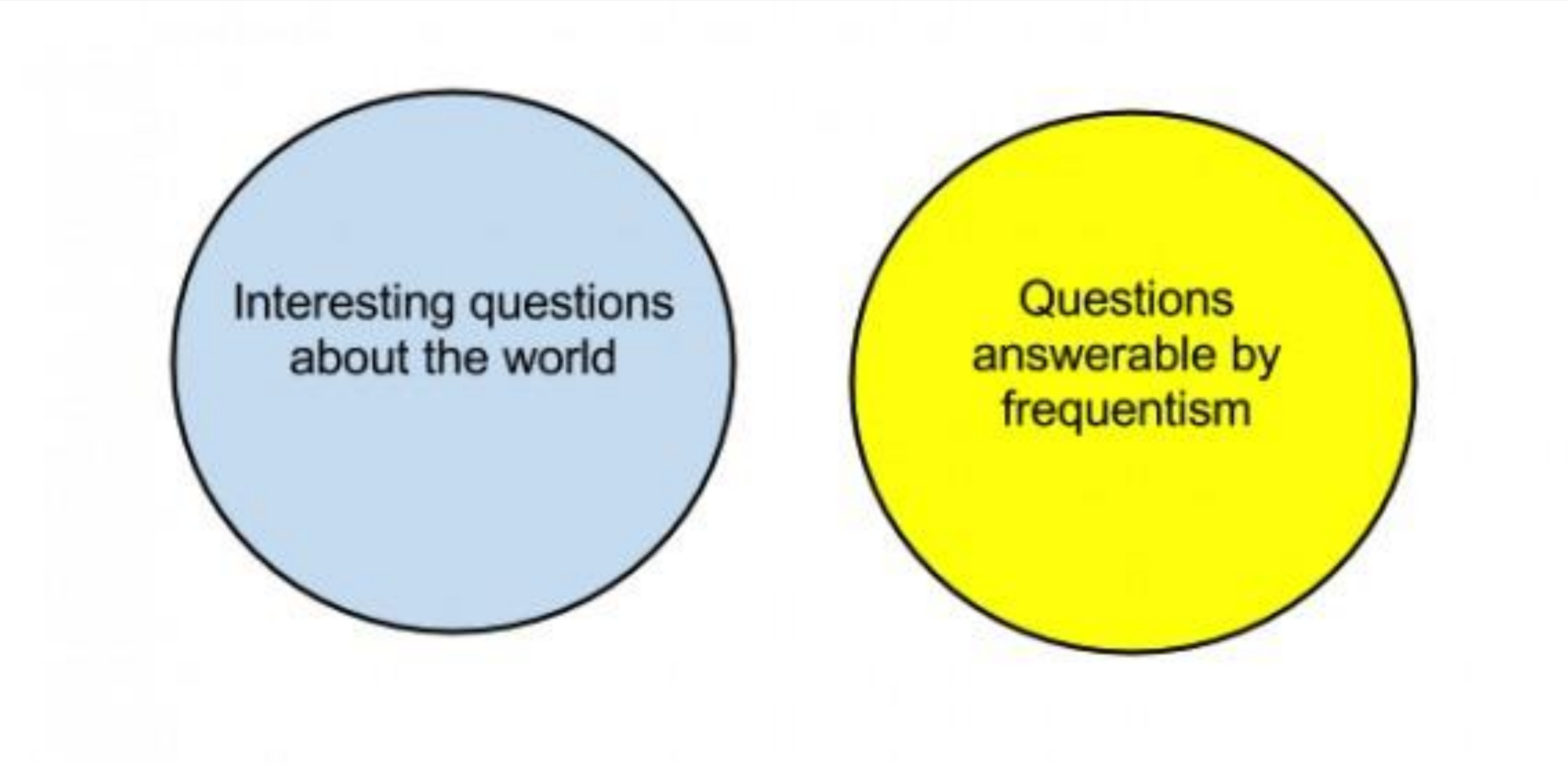
El Intervalo de confianza es un estimador (de intervalo) **usualmente correcto**, que no tiene por qué ser creíble en cada ocasión que se use.



- Si I es un intervalo de confianza particular al 95 % obtenido de algunas observaciones, **NO** se puede decir: "es 95 % probable que I incluya el valor verdadero del parámetro".
- No se puede determinar la credibilidad de cualquier estimación particular porque sólo se puede determinar la confiabilidad de un sistema de estimación de largo plazo.

Hartnack y Roos (2021)

Conditional on randomness?



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

17/02/2022

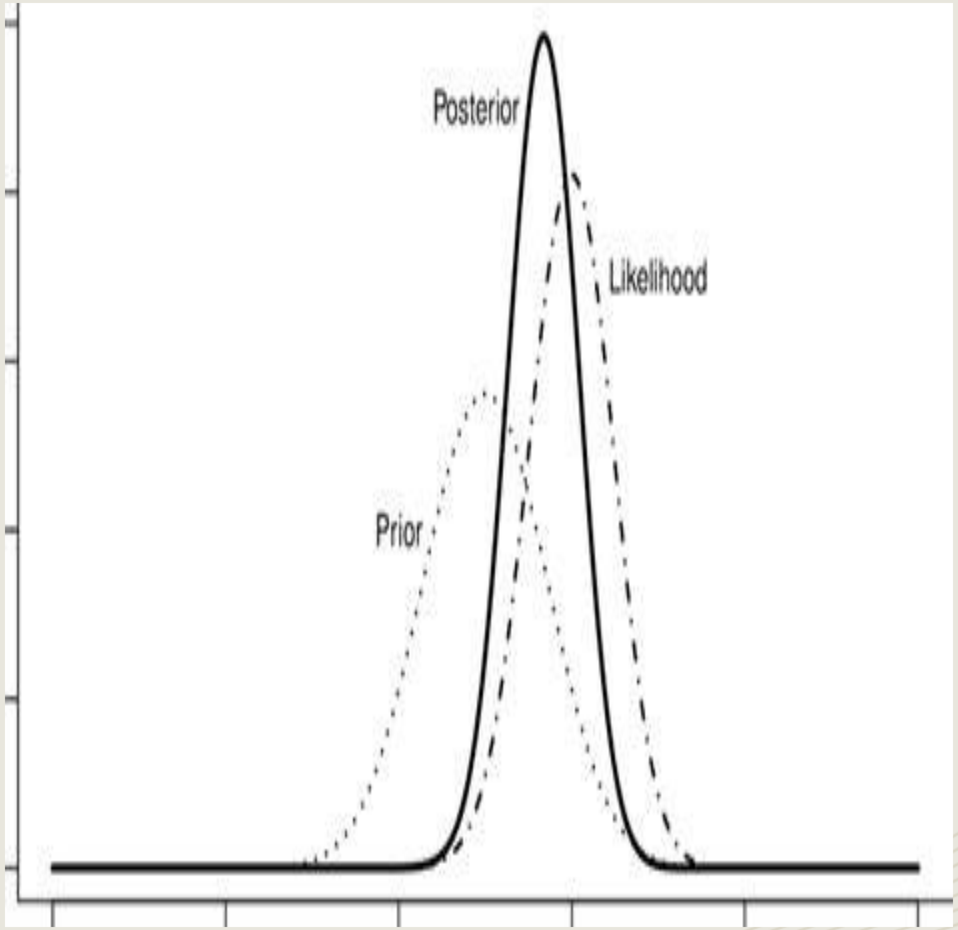
Cuatro aspectos:

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 (θ) sea cierta?
  - 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 (Verosimilitud)
- El espacio de nuestras creencias sobre el posible resultado (Priors)

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

$p(\theta=.5) = .99$ 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 NO 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 obervados.

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 $(\Omega, \mathscr{F})$ un espacio medible. Una medida de probabilidad es una función $P : \mathscr{F} \to [0,1]$ que satisface

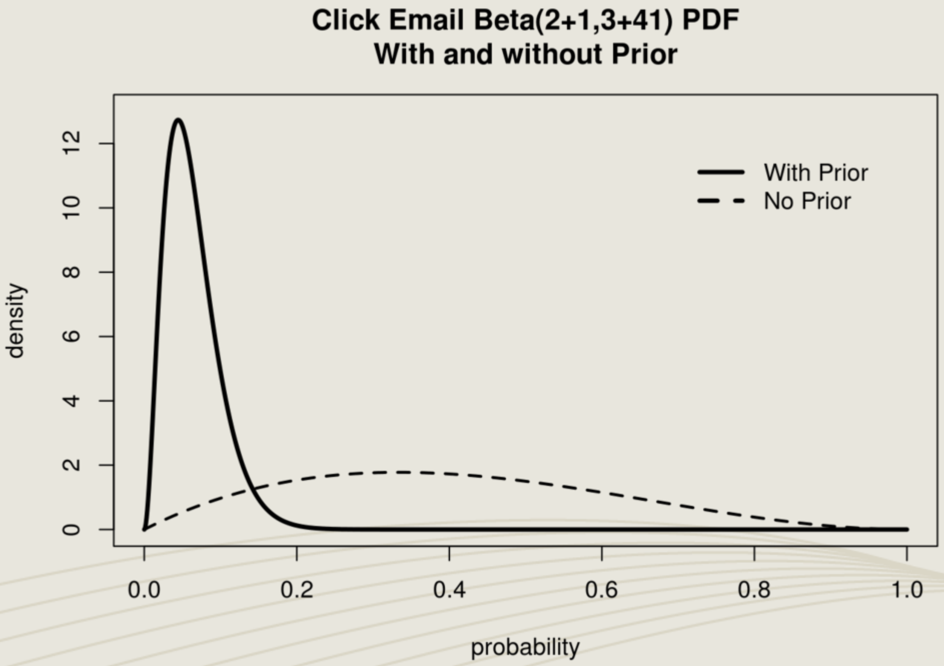
1. $P(\Omega) = 1$.
2. $P(A) \geq 0$, para cualquier $A \in \mathscr{F}$.
3. Si $A_1, A_2, \ldots \in \mathscr{F}$ son ajenos dos a dos, esto es, $A_n \cap A_m = \emptyset$ para $n \neq m$, entonces $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$.

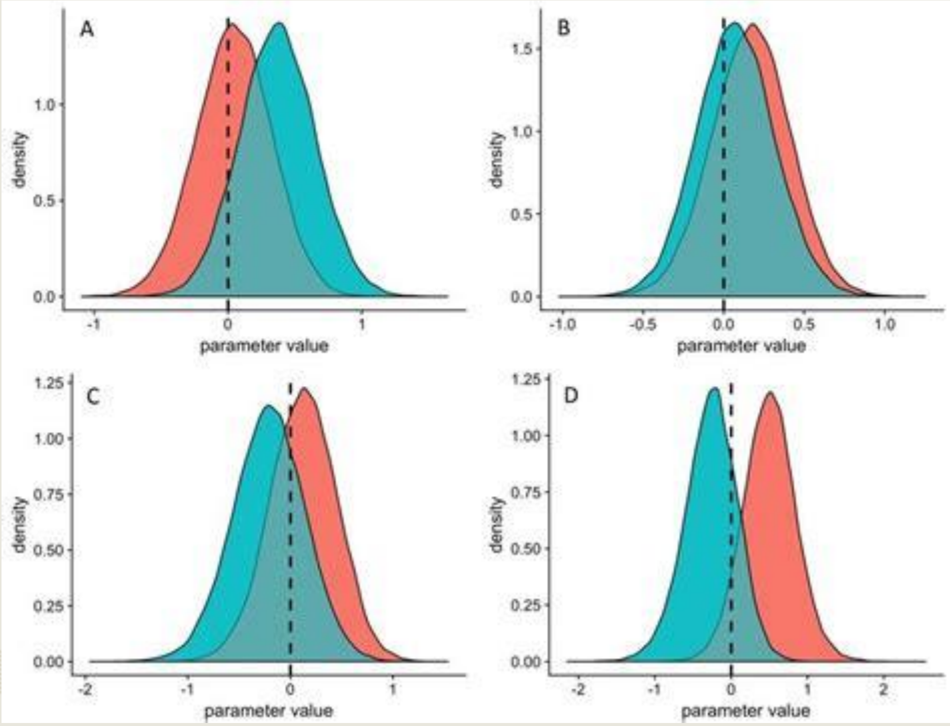
http://lya.fciencias.unam.mx/lars/libros/cip.pdf

Андрей Николаевич Колмогоров
1903-1987

Probabilidad: Dentro y fuera

- La probabilidad no puede ser negativa
- La suma de las probabilidades de todos los eventos posibles es 1
- Para cada dos eventos mutuamente exluyentes, la probabilidad de que los dos ocurran es la suma de sus probabilidades





Fuera de la cabeza

Probabilidad y frecuencias

- Mecanismo generador de datos
- Generación de datos conforme al mecanismo
- Por eso podemos simular volados

- Derivar la frecuencia
- Experimentación y conteo

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

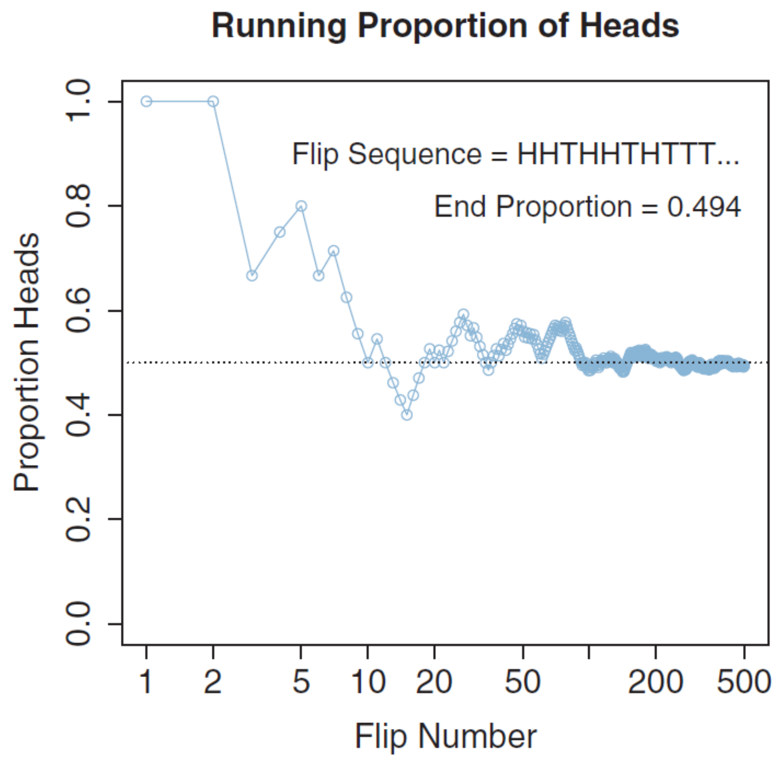
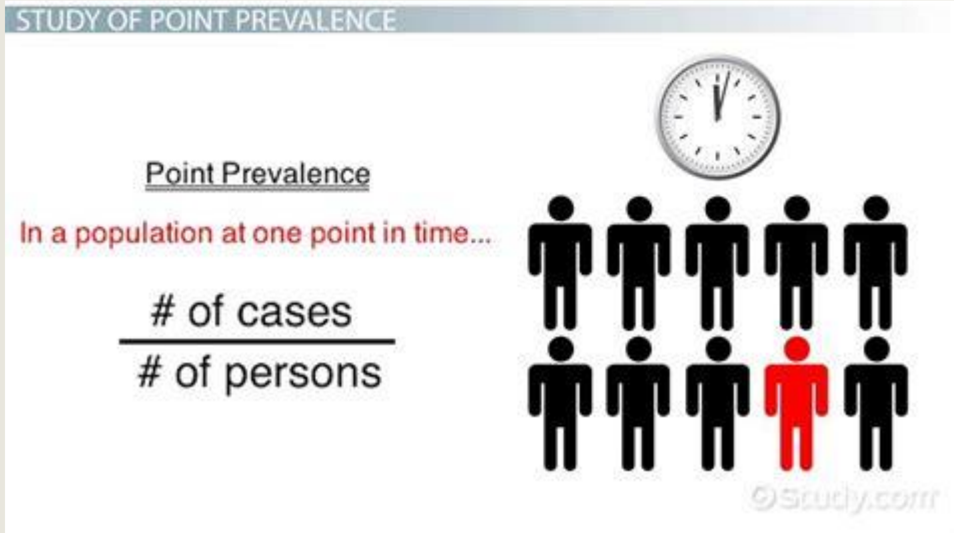


Figure 4.1 Running proportion of heads when flipping a coin. The x-axis is plotted on a logarithmic scale so that you can see the details of the first few flips but also the long-run trend after many flips. R code for producing this figure is discussed in Section 4.5.

¿Parámetros?



- Me interesa p(Datos | Theta)
- Me interesa p(Theta | Datos)
- Y si los datos tienen error?

Los estimadores de encuestas tienen sentido y ¿el IC?

Dentro de la cabeza

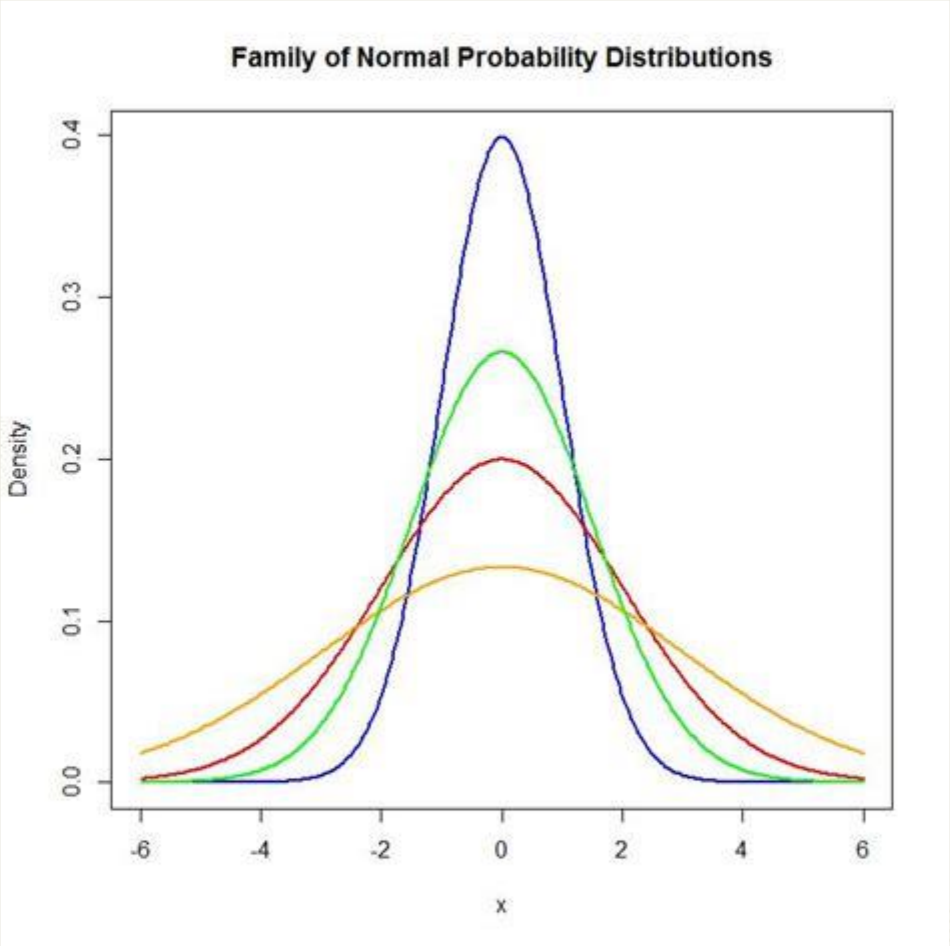
Describiendo nuestras creencias

- La estura 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

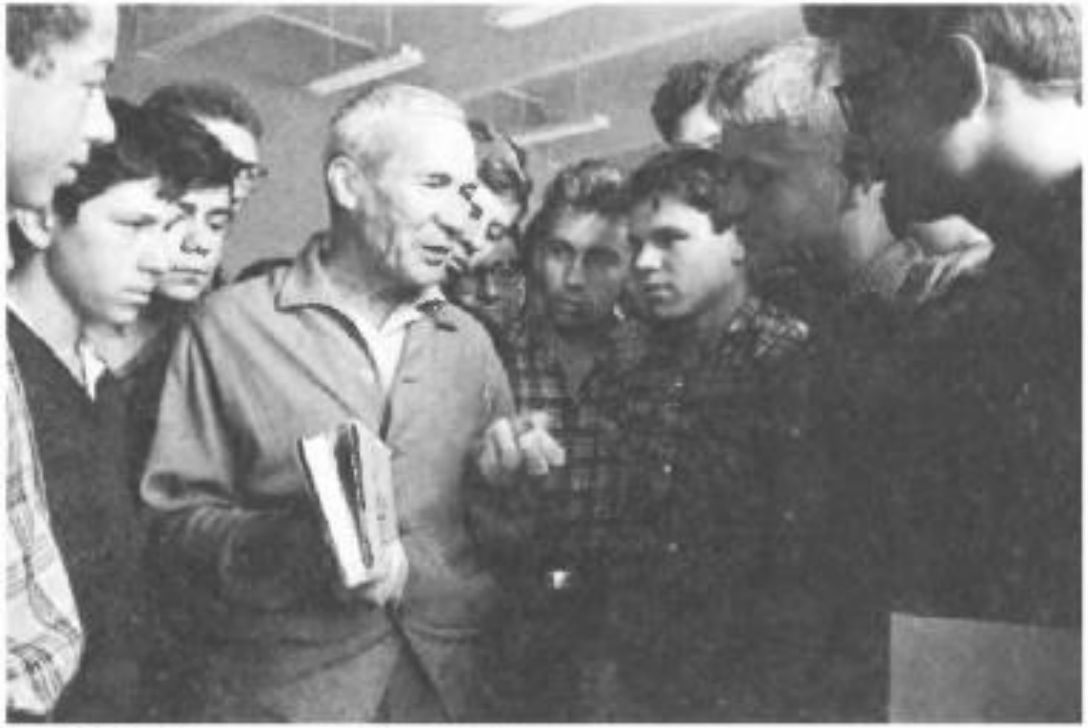
- Hiperparámetros $\theta \sim N(0,SD)$: El parámetro θ tiene sus propios parámetros

- Distribuciones de probabilidad



¿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

¿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 form us ---.”

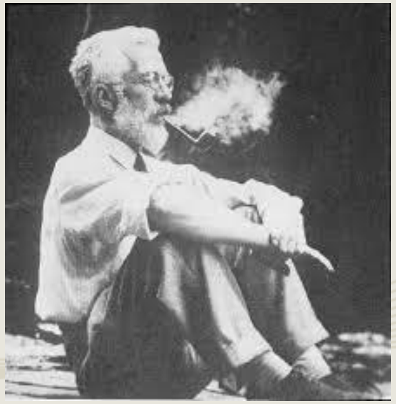
Jakob Bernoulli
1655-1705

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

Probabilidad: ¿en el mundo allá afuera?

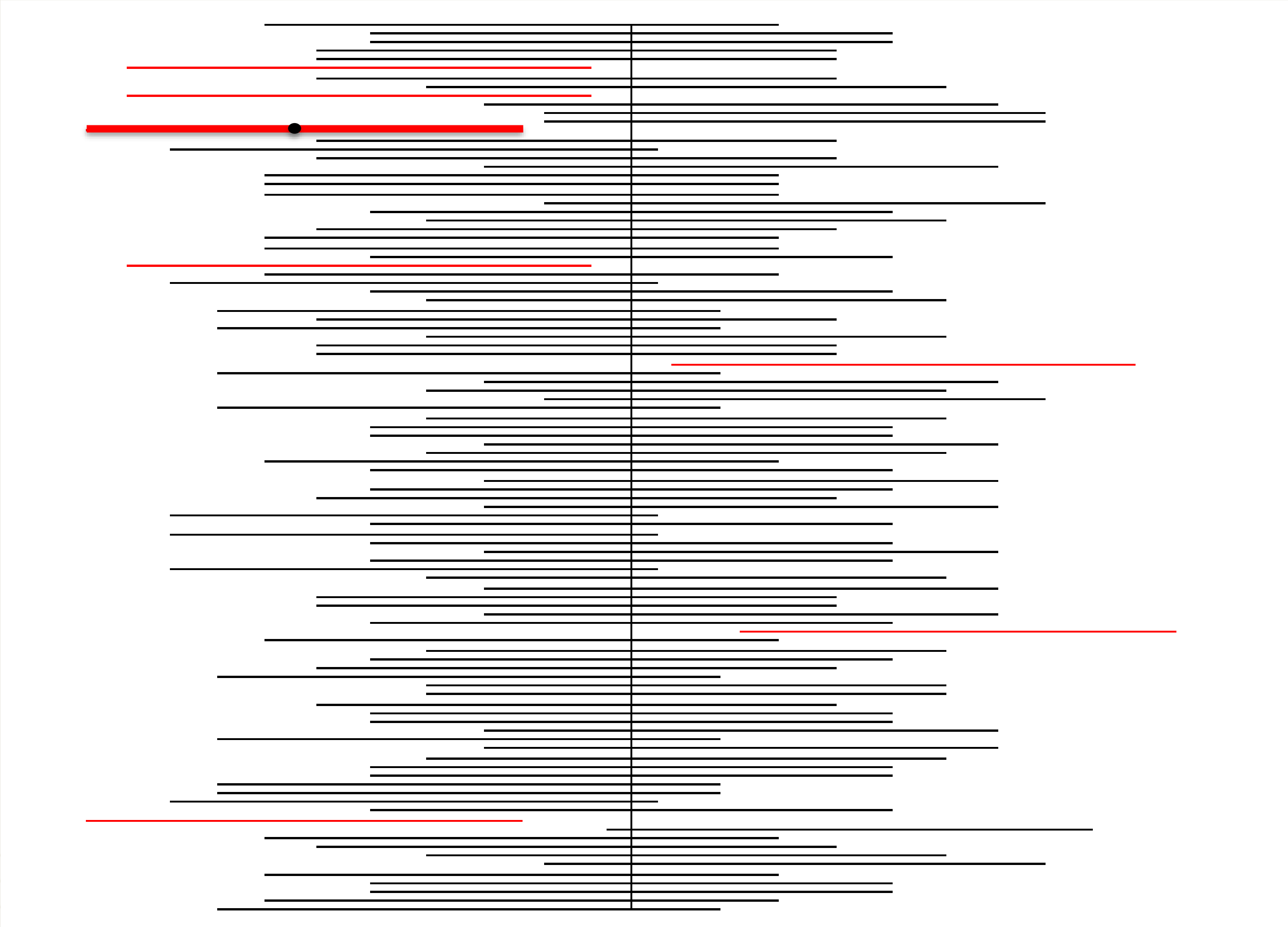
- Karl Pearson, antes de Kolmogorov, creía que las distribuciones de probabilidad eran observables meramente recolectando datos.
- William Gosset intentó describir el espacio de eventos para un experimento (el conjunto de todos los posibles resultados de ese experimento. ¡Ingenioso!, pero inútil pues 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 todos 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?

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



Incertidumbre en estadística clásica

El Intervalo de confianza es un estimador (de intervalo) **usualmente correcto**, que no tiene por qué ser creíble en cada ocasión que se use.



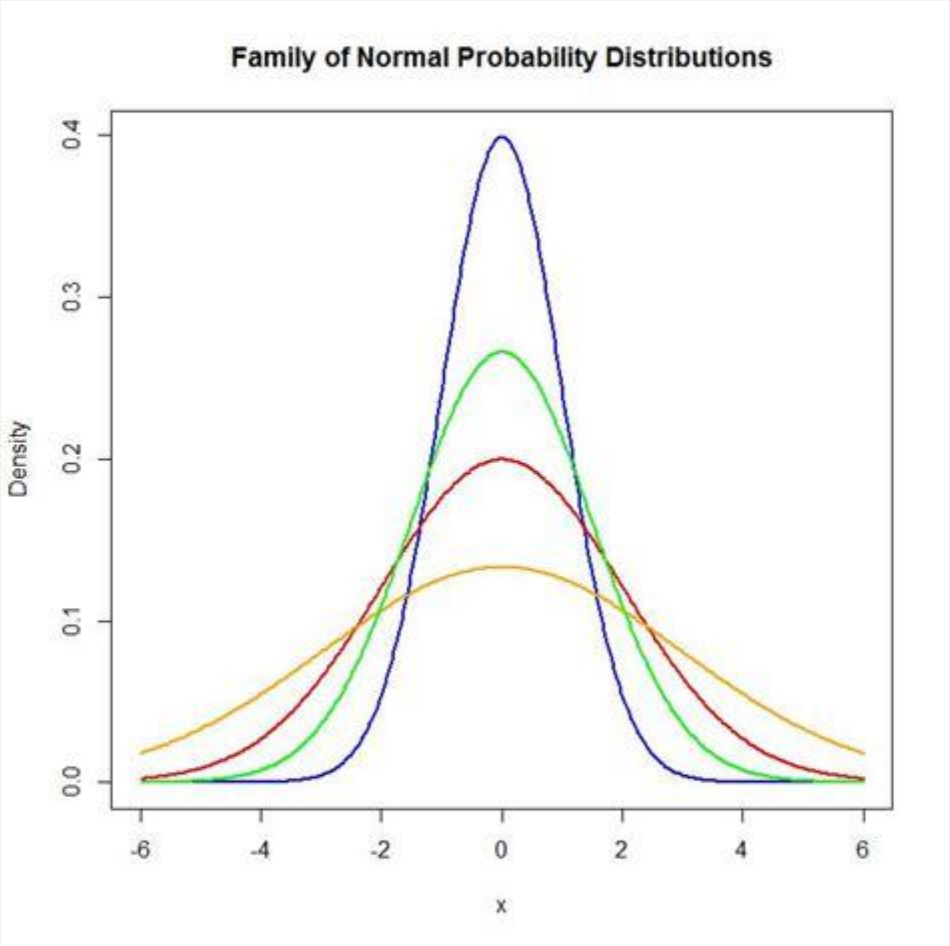
- Si I es un intervalo de confianza particular al 95 % obtenido de algunas observaciones, **NO** se puede decir: "es 95 % probable que I incluya el valor verdadero del parámetro".
- No se puede determinar la credibilidad de cualquier estimación particular porque sólo se puede determinar la confiabilidad de un sistema de estimación de largo plazo.

Hartnack y Roos (2021)

Distribuciones de probabilidad

- Las distribuciones de probabilidad se conforman de la lista de todas las posibilidades con sus probabilidades correspondientes

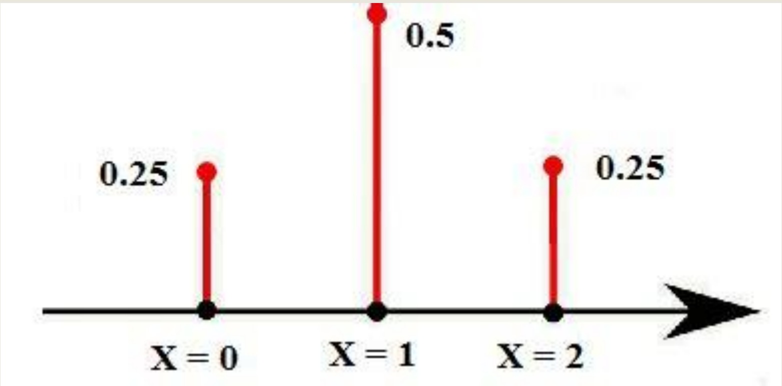
- Resultados/observaciones discretos o continuos?



Distribuciones de probabilidad

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

Distribuciones discretas: Masas y densidades

- Para eventos continuos: estatura; valor de β
- Ir de discreto a continuo no es muy difícil pero…
- Hay que tener claro el intervalo que se usa
- La clave es cómo la masa (monto de casos) da forma a la distribución

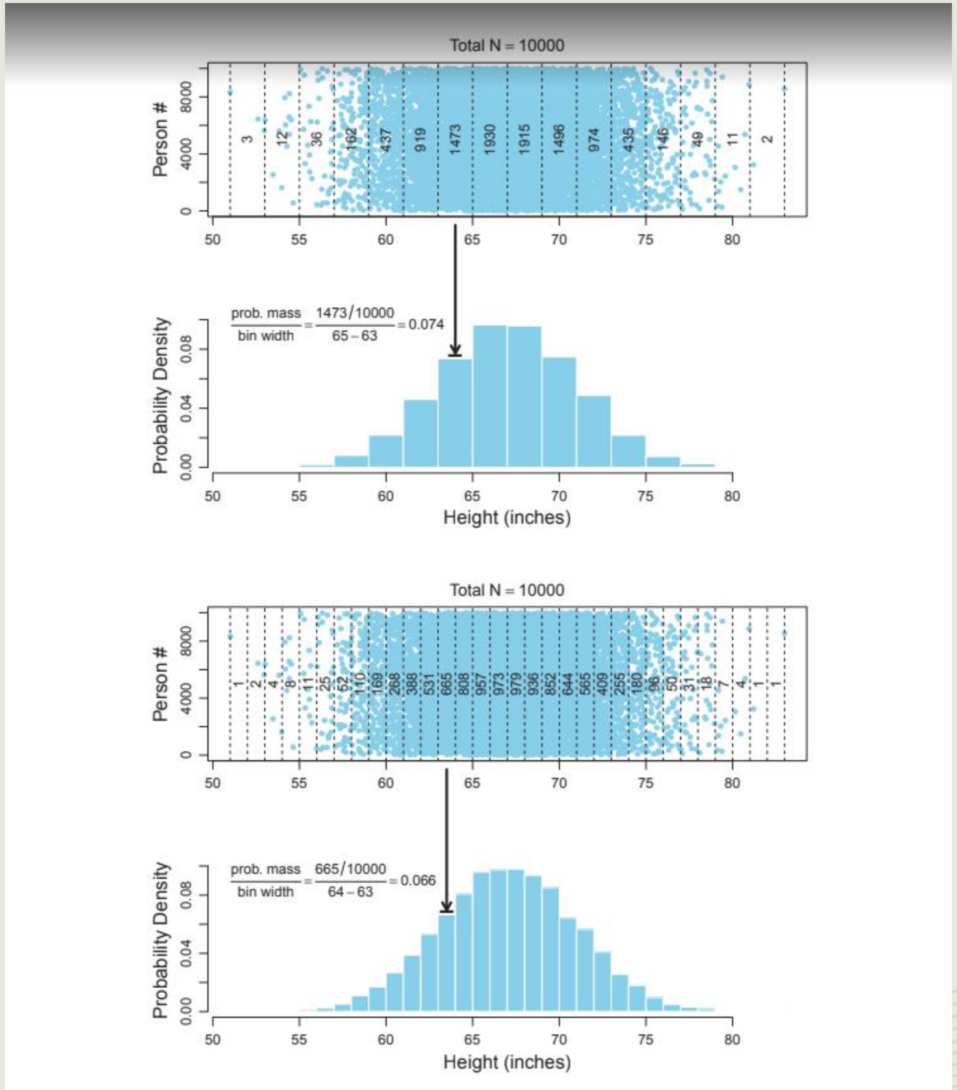


Figure 4.2: Examples of computing probability density. Within each main panel, the upper plot shows a scatter of 10,000 heights of randomly selected people, and the lower plot converts into probability density for the particular selection of bins depicted. Copyright Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier.

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)

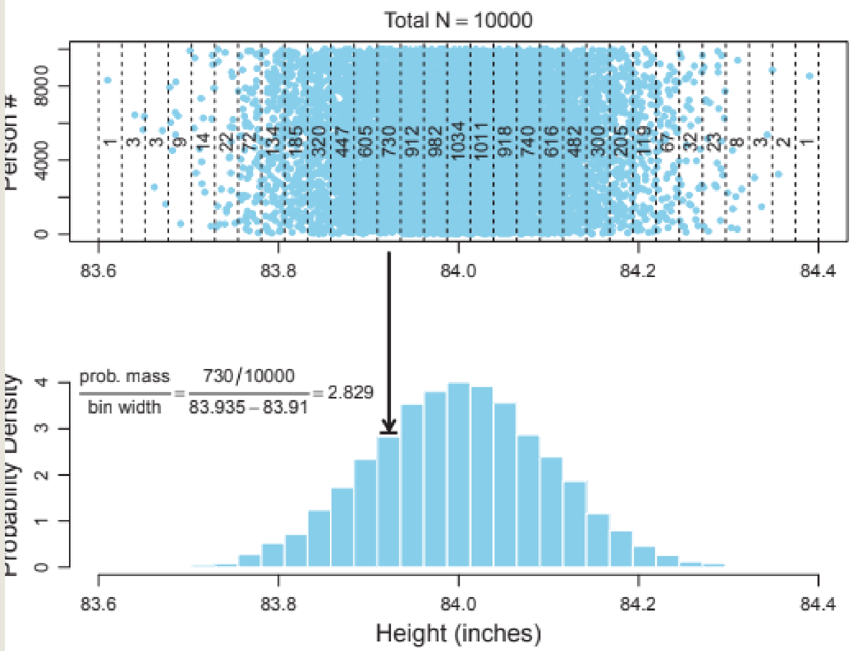


Figure 4.3: Example of probability density greater than 1.0. Here, all the probability mass is concentrated into a small region of the scale, and therefore the density can be high at some values of the scale. The annotated calculation of density uses rounded interval limits for display. (For this example, we can imagine that the points refer to manufactured doors instead of people, and therefore the *y*-axis of the top panel should be labelled “Door” instead of “Person.”) Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier.

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

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

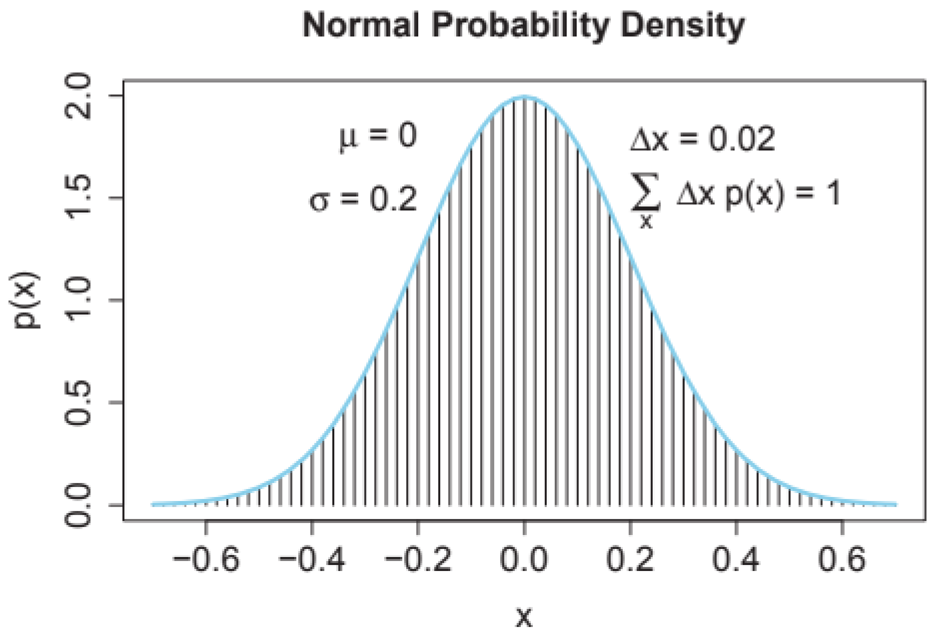


Figure 4.4: A normal probability density function, shown with a comb of narrow intervals. The integral is approximated by summing the width times height of each interval. Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier.

Intervalo más alto de densidad

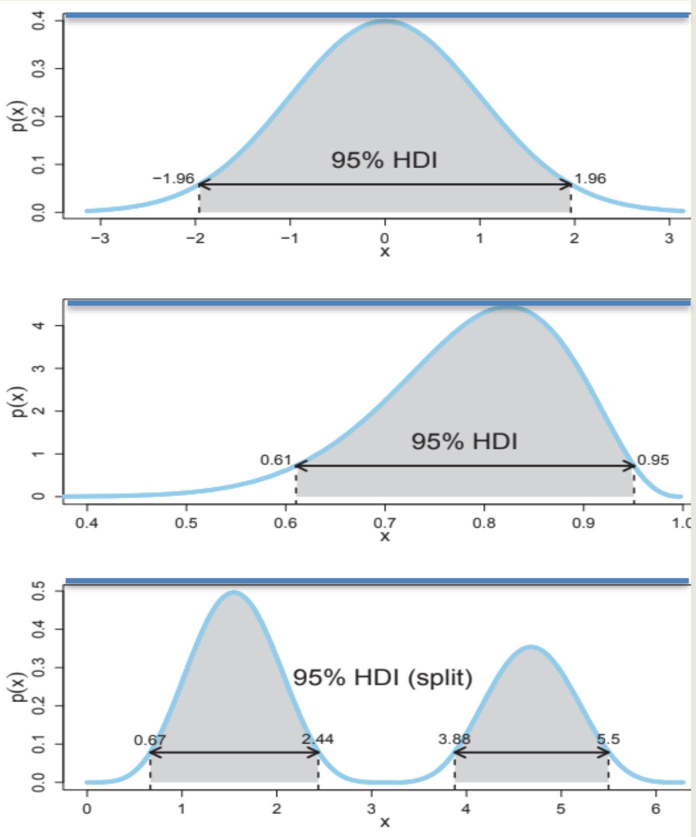


Figure 4.5: Examples of 95% highest density intervals (HDIs). For each example, all the x values inside the interval have higher density than any x value outside the interval, and the total mass of the points inside the interval is 95%. The 95% area is shaded, and it includes the zone below the horizontal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

Página 14 / 218

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

OK! Pero…

- Dónde encaja P(β|D) :
  - $\mathrm{Y} = a + bX$

- Y definimos nuestra credibildiad apriori (priors)

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

- O qué pasa cuando tengo más variables

$$\mathrm{Y} = a + \beta_1 X + \beta_2 Z$$

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

Si la persona tiene ojos azules ¿Cuál es la probabilidad de que sea rubia?

Probabilidad condicional

Reubicación de acuerdo al color de cabello

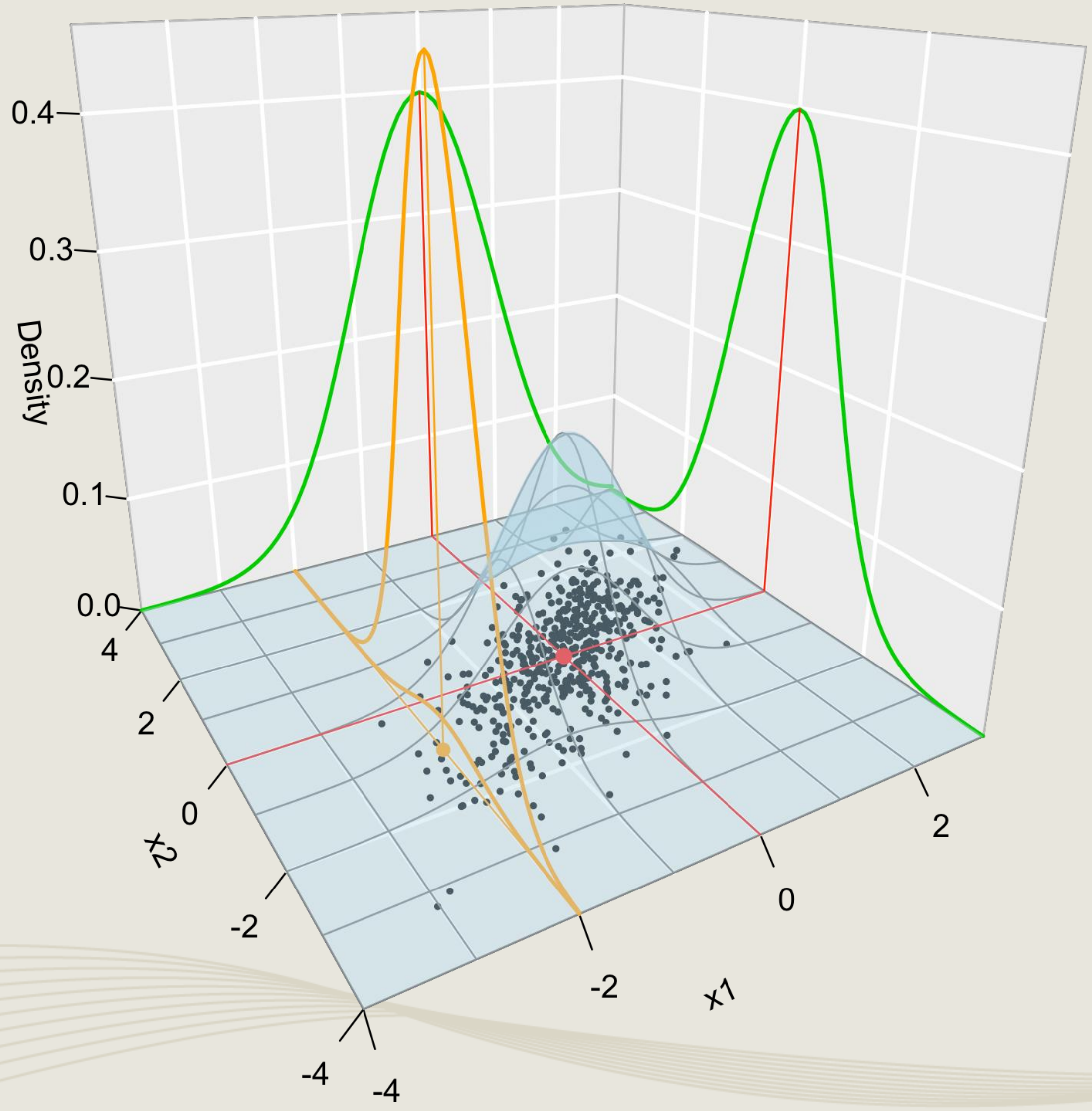
Table 4.2: Example of conditional probability. Of the blue-eyed people in Table 4.1, what proportion have hair color h? Each cell shows $p(h|\text{blue}) = p(\text{blue}, h)/p(\text{blue})$ rounded to two decimal points. Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with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

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

Esto es inferencia bayesiana:

P(β|D)

Probabilidad condicional



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.

Hartnack, S., & Roos, M. (2021). Teaching: confidence, prediction and tolerance intervals in scientific practice: a tutorial on binary variables. *Emerging Themes in Tpidemiology,* 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

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 EL CUBREBOCAS