



UNAM  
POSGRADO



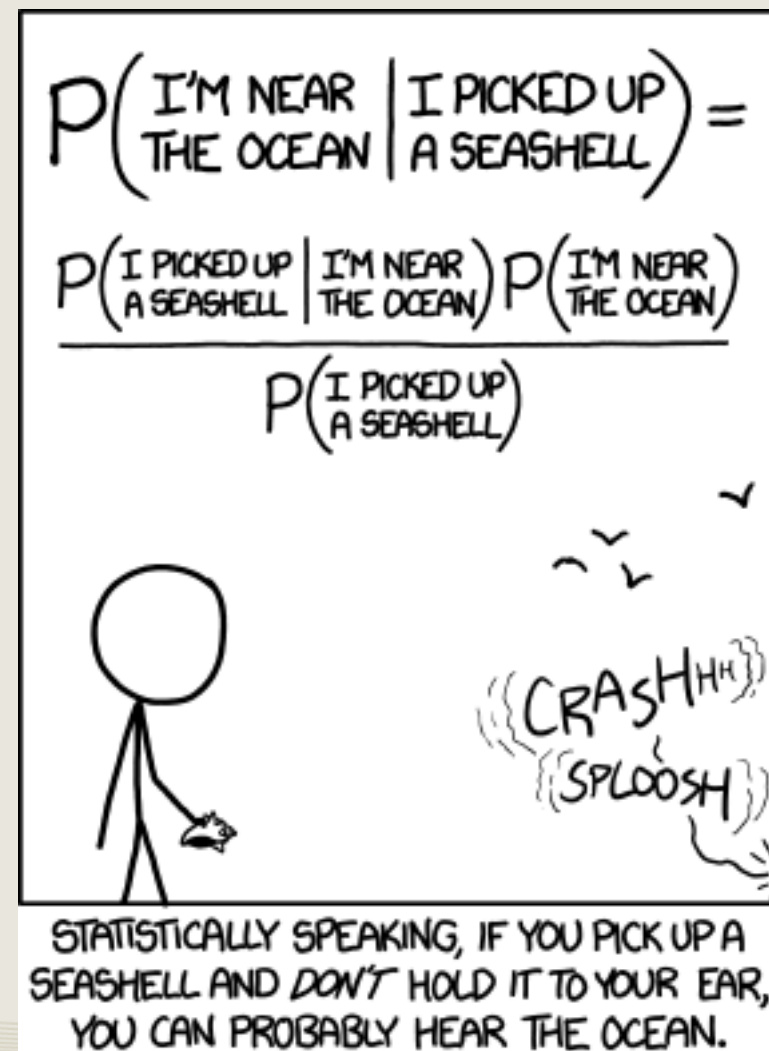
Programa  
Universitario  
de Estudios  
del Desarrollo  
UNAM

# Marco conceptual del análisis bayesiano: credibilidad, modelos y parámetros

Dr. Héctor Nájera  
Dr. Curtis Huffman

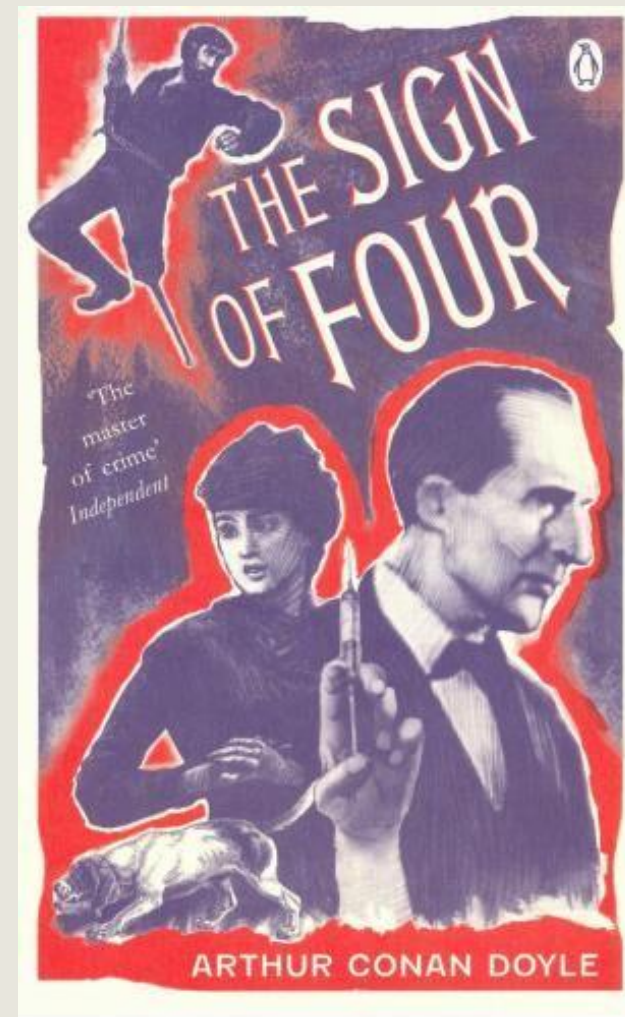
# Análisis Bayesiano de datos

*O sobre cómo uso a la estadística para ajustar mis conclusiones (respecto a una **hipótesis/parámetro**) cuando combino nueva y vieja información*



# ¿De qué va el capítulo?

- **Dos** ideas fundacionales del análisis Bayesiano
  - Reubicación de **credibilidad** entre **posibilidades**
  - Las **posibilidades** en clave de 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.





# Primera idea: Credibilidad y posibilidades, nueva y vieja información

# Reubicación de credibilidad entre posibilidades

- Las posibilidades son hipótesis/explicaciones
- Estas posibilidades tienen asociada cierta credibilidad *a priori* (prior)
- Esta credibilidad *a priori* viene de nuestro conocimiento previo
- Conforme tenemos más datos nuestra credibilidad se “reubica” (Tenemos una distribución posterior)

Datos

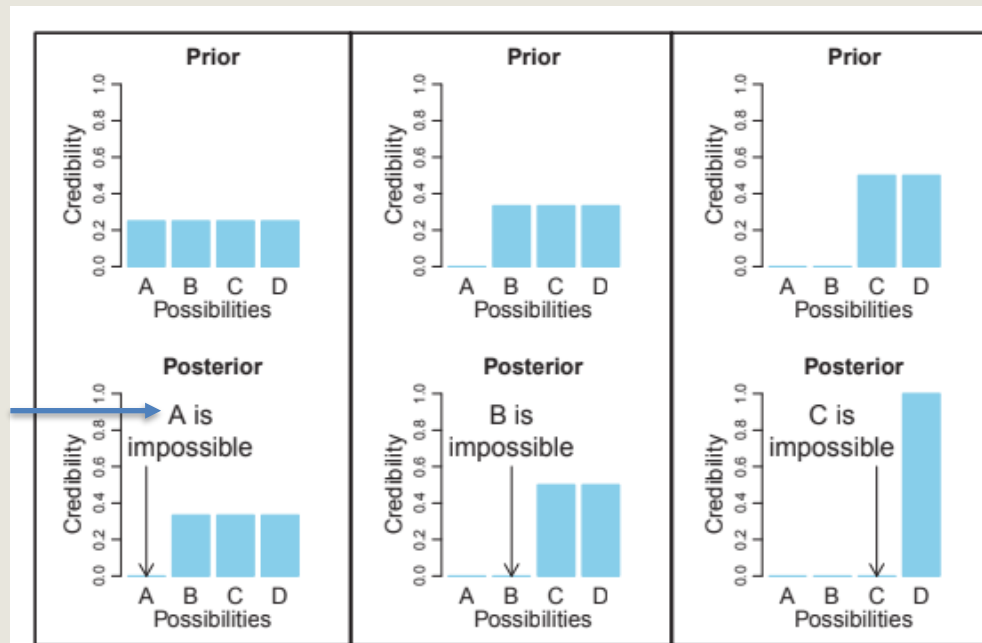


Figure 2.1: The upper-left graph shows the credibilities four possible causes for an outcome. The causes, labeled A, B, C and D, are mutually exclusive and exhaust all possibilities. The causes happen to be equally credible at the outset, hence all have prior credibility of 0.25. The lower-left graph shows the credibilities when one cause is learned to be impossible. The resulting posterior distribution is used as the prior distribution in the middle column, where another cause is learned to be impossible. The posterior distribution from the middle column is used as the prior distribution for the right column. The remaining possible cause is fully implicated by Bayesian re-allocation of credibility. Copyright © Kruschke, J. K. (2014).

# Primera idea: Credibilidad, nueva y vieja información

- Cuantificación de la incertidumbre en el paradigma bayesiano
  - Información vieja, independiente de los datos, a priori (priors) y
  - Credibilidad a posteriori (posterior probability), después de recibir/procesar información nueva, los datos.

prior probability  data  posterior probability

¿Siempre existe?

¿Qué significa que exista?

¿Sólo datos?

¿Cómo se combinan y qué peso tienen?

- Parece reflejar la manera en la que piensa la mayoría de las personas (y el desarrollo científico)



Primera idea: Credibilidad y posibilidades,  
nueva y vieja información



# Credibilidad, nueva y vieja información

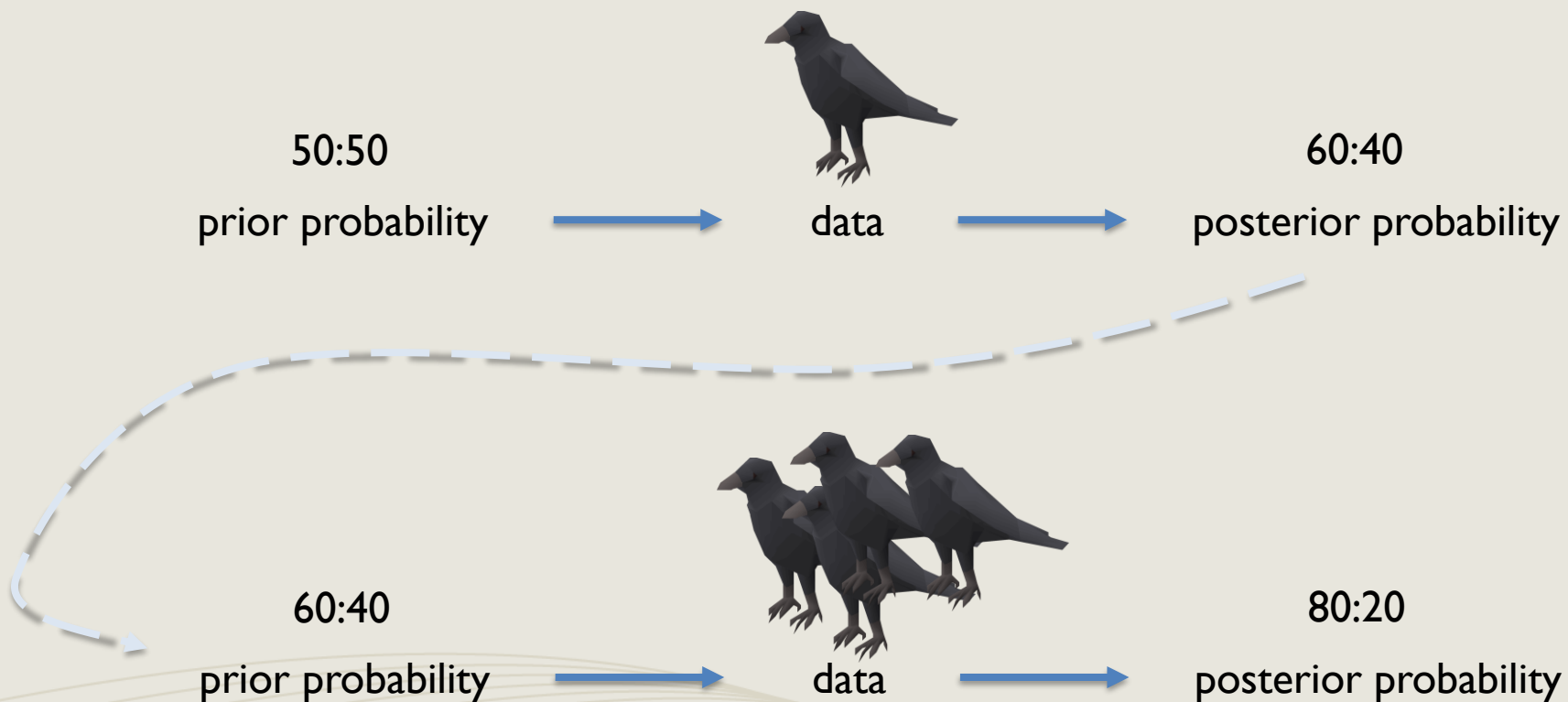
- Suponga que usted quiere determinar si todos los cuervos son negros.
  - Se arranca con una noción/conocimiento **a priori** de la verdad de dicha afirmación
    - Digamos que 50:50 (no sabemos nada de cuervos)
  - Se recolectan datos sobre avistamientos de cuervos.
    - Suponga que se avista un cuervo negro, luego su probabilidad posterior aumenta.





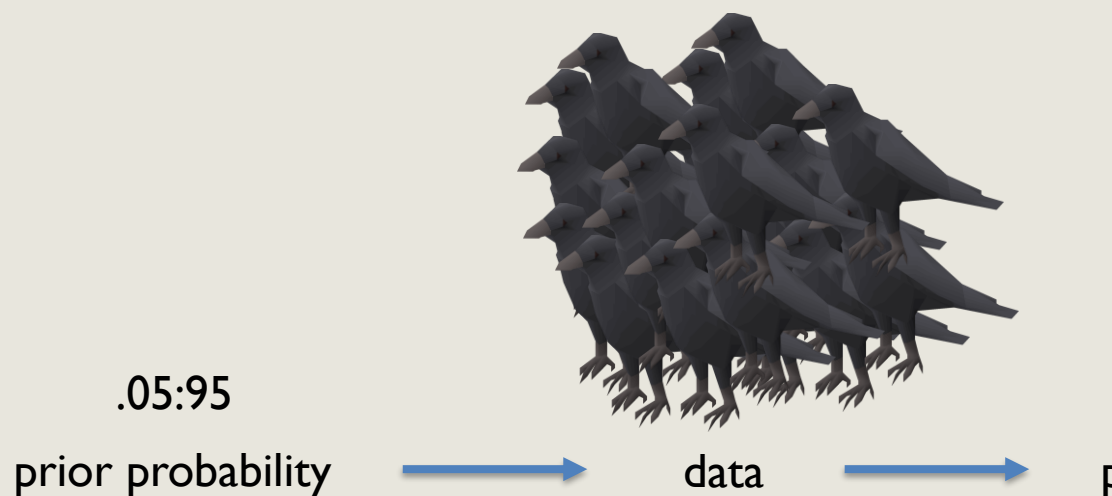
# Credibilidad, nueva y vieja información

- La próxima vez que se observen cuervos, su nuevo prior (la vieja posterior) es mayor que 50 % y continuará aumentando como aviste nuevos cuervos, los cuales son todos negros.



# Credibilidad, nueva y vieja información

- Por otro lado, una persona puede entrar al proceso prácticamente convencida de que los cuervos son blancos (strong prior), tanto que puede requerir una cantidad descomunal de datos para hacerla cambiar de opinión.



¡Pareciera que una persona puede calcular



# Credibilidad, nueva y vieja información

- ¿Qué sentido tiene seguir hablando de probabilidad en estos términos (subjetivos) en los que **todo vale**?



- ¿Cuál es la alternativa?, ¿es ésta mejor? (próxima clase: ¿Qué es esa cosa llamada probabilidad? Keynes, Kolmogorov, Savage-de Finetti)
- ¿Qué dice Kruschke sobre este punto?
  - “The prior must **pass muster** (be of satisfactory quality) with the audience of the analysis, such as skeptical scientists.”

# Credibilidad, nueva y vieja información

- ¿Qué sentido tiene seguir hablando de probabilidad en estos términos (subjetivos) en los que **todo vale**?

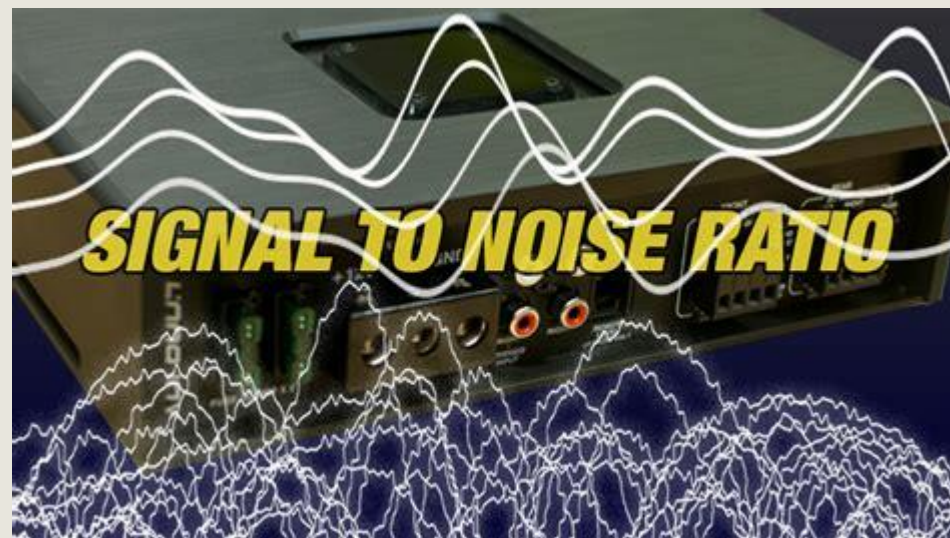


- Veremos que hay:
  - “a priori inadecuados” –Improper priors-
  - Malos priors llevan a posteriores poco confiables y/o difíciles de obtener.

## Antes de ir a la segunda idea fundamental...

Kruschke nos recuerda  
que:

Los datos son *ruidosos* y la  
inferencia es *problemática*!



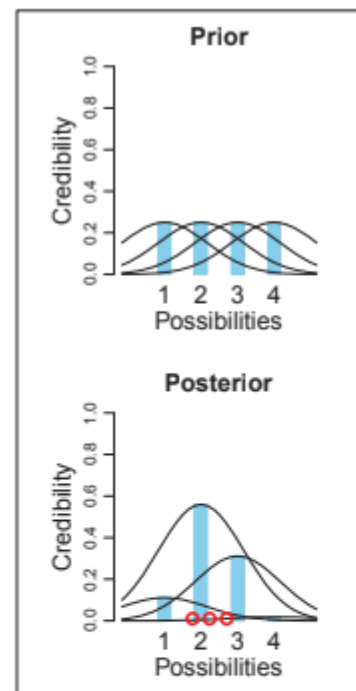


# Los datos son ruidosos y la inferencia es problemática

Uno de los usos más notables de la inferencia bayesiana es en contexto de error de medición.

Es útil porque es más intuitivo introducir el tipo de información a priori

- Recuerden que la posterior es una mezcla de los priors y los datos
- ¿Qué si  $D$  tiene defectos? (Error de medición)
- Las mediciones son imperfectas y la relación entre la causa y el efecto medido esta llena de ruido.
- Debemos reubicar nuestra credibilidad en un contexto de variabilidad indeseable



← Misma incertidumbre a priori respecto a las 4 posibilidades

Figure 2.3: The upper graph shows the prior credibilities of four candidate means in normal distributions, located at values of 1, 2, 3, and 4. Superimposed on the means are the corresponding normal distributions. The horizontal axis is playing double duty as a scale for the means (marked by the blue bars) and for the data (suggested by the normal distributions). The three observed data values are plotted as circles on the floor of the lower panel. Bayesian reallocation of credibility across the four candidate means indicates that the mean at 2 is most credible given the data, the mean at 3 is somewhat credible, and so on. Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. 2nd Edition. Academic Press / Elsevier.





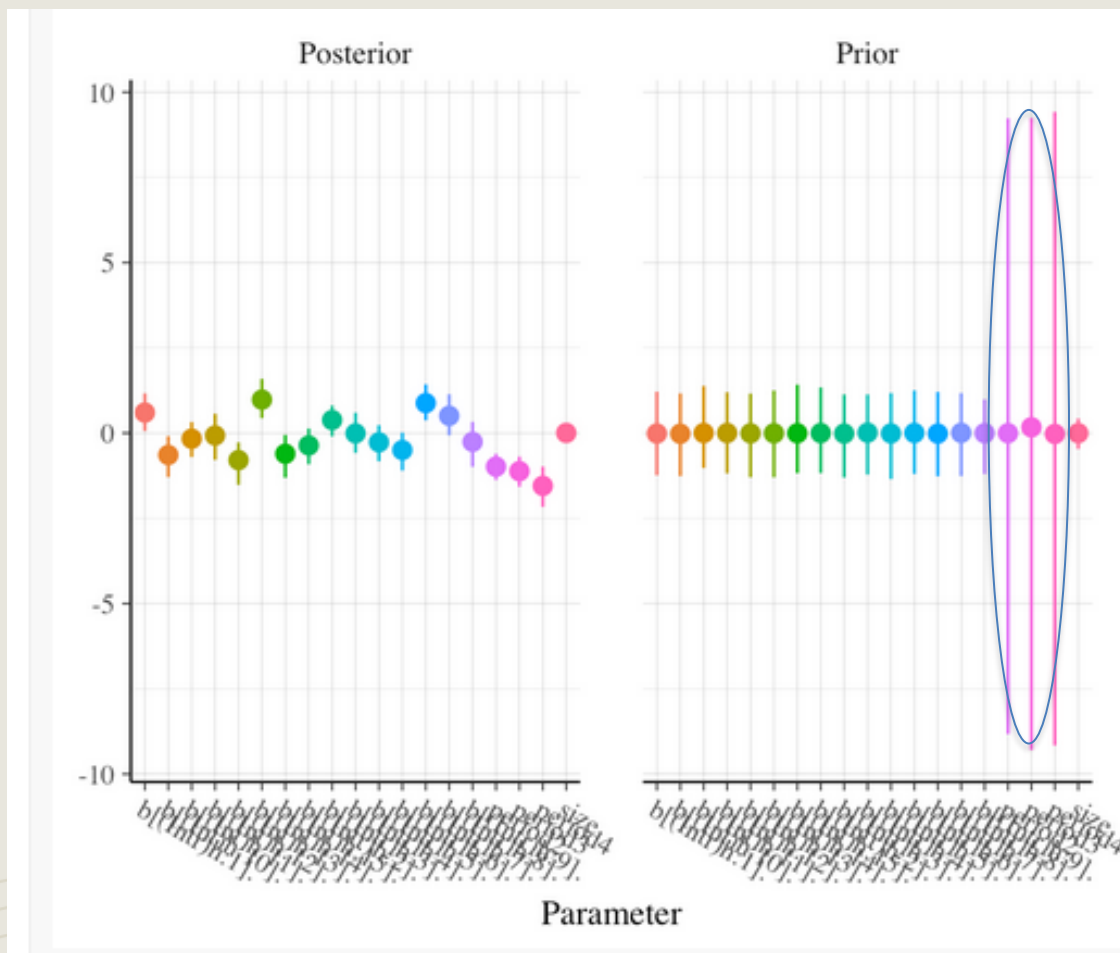
# Los datos son ruidosos y la inferencia es problemática

- Pero como vimos en el ejemplo de los cuervos, podemos tener a priori muy fuertes
- Ejemplo:
  - Google search: “radom”
  - Algoritmo de google: ¿Cómo debería leer dicha palabra? Random, radon or radom
  - ¿Cuál es la probabilidad de que sea “Random”?
  - Información a priori: Google asigna probabilidades a las diferentes palabras
  - Google además tiene datos sobre la búsqueda que:  $p(\text{radom}|\text{radom}^*)$
  - Puede calibrar conjuntamente **el a priori y los datos** para obtener la posterior de que la persona esté buscando “Random”

¿Qué pasaría con el buscador de Google si asignáramos probabilidades iguales?

# Los datos son ruidosos y la inferencia es problemática

- Los a priori reflejan incertidumbre (pueden reflejar **poca o mucha** incertidumbre)



Estos *a priori* reflejan mucha incertidumbre (respecto a los datos o a la relación con el fenómeno)

El resto refleja cierta certidumbre

Hacia delante: Esto significa que hay distintos **tipos de a priori** y que su **clasificación** pasa por su nivel asociado de incertidumbre



## Segunda idea: Posibilidades en forma de parámetro

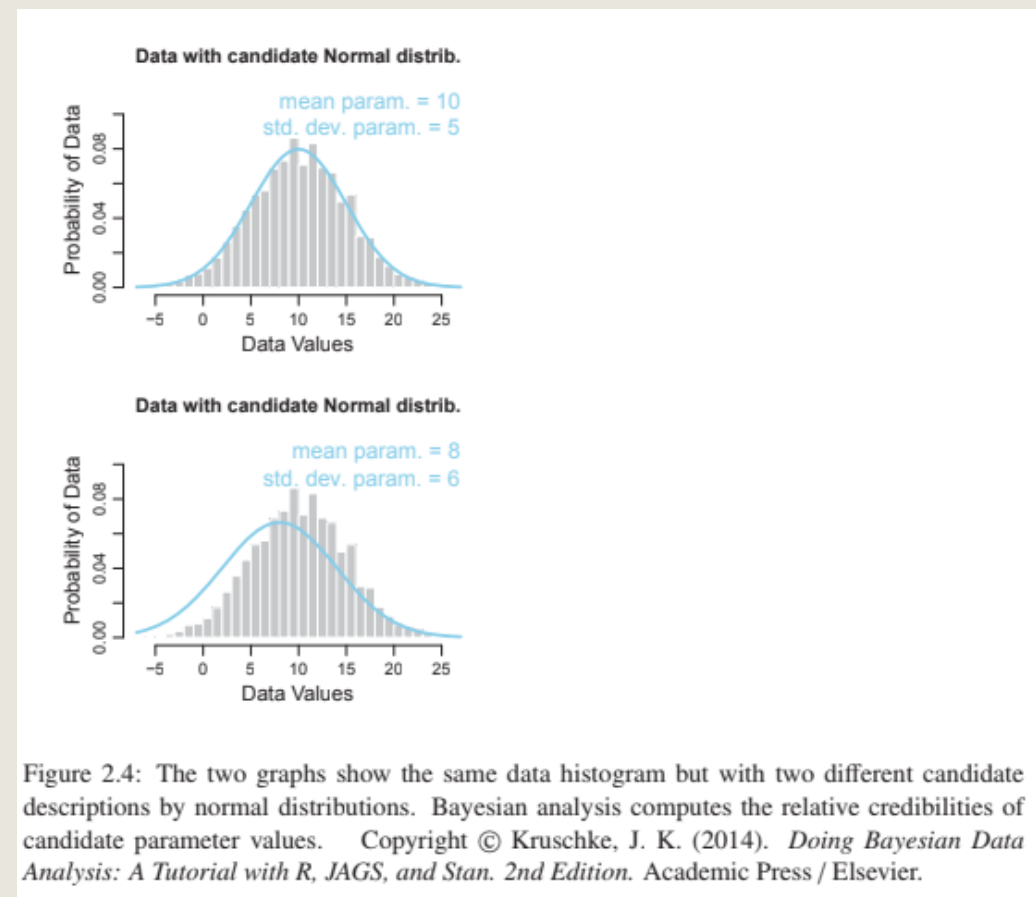
# Posibilidades

- Las posibilidades son las hipótesis/explicaciones de cierto fenómeno ( $\theta$ )
- Dados los datos y la credibilidad respecto a dichas hipótesis: ¿Cuál es la probabilidad posterior de la hipótesis/parámetro?  $P(\theta|D)$
- Nota: Los datos no son espontáneos, en realidad, son una combinación de lo observado y del mecanismo que los genera (**Modelo candidato**).
  - Esa descripción puede sintetizarse vía parámetros.  
(Control knobs)
  - Futuro: **Verosimilitud**
- Esto es lo que conocemos como  $P(D|\theta)$



# Posibilidades y parámetros: Verosimilitud

- Pensar a las posibilidades como parámetros es **muy útil y deseable** desde el punto de vista de inferencia estadística.
- En inferencia Bayesiana definimos el set de parámetros (**Modelo**) sobre el cual reubicamos la credibilidad para obtener la **posterior**
- “The role of Bayesian inference is to compute the exact relative credibilities of candidate parameter values, while also taking into account the prior probabilities” (p. 24)
- **¿Por qué esto es importante y fundamental para apreciar el alcance de la inferencia Bayesiana?**
  - Noisy data
  - Meaningful space of possibilities

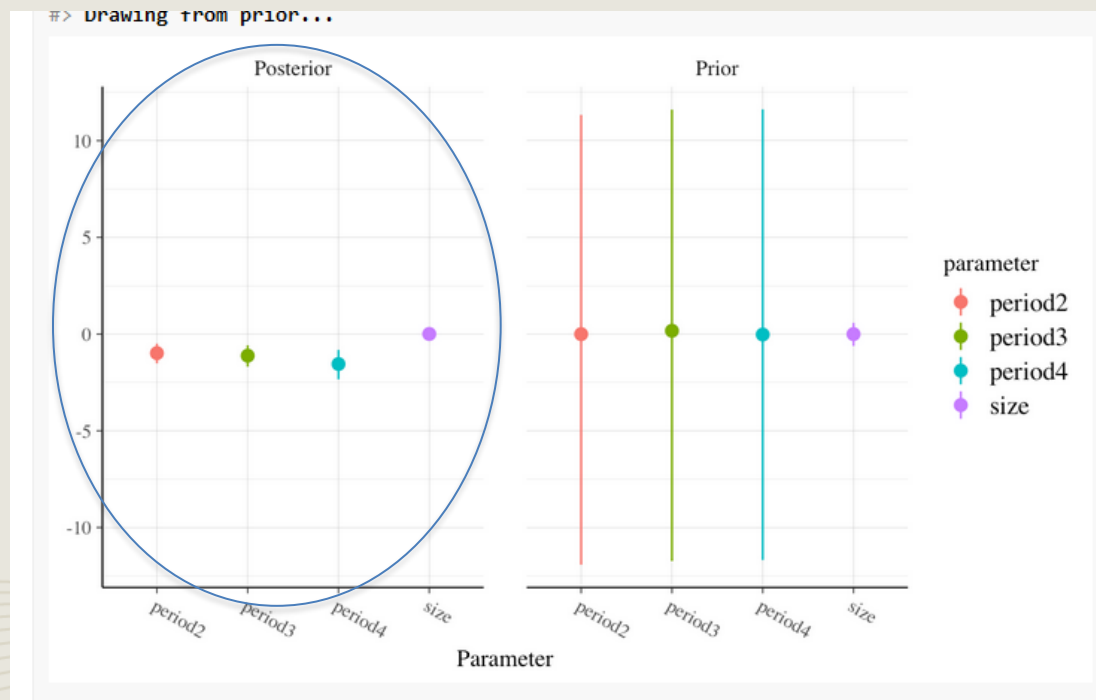


# Posibilidades y actualización:

- La distribución posterior: Los valores de nuestro parámetros una vez que incorporamos nueva información para recalibrar nuestras hipótesis.

$$P(\theta|D)$$

Noten que la posterior después se convierte en prior





# Pasos en el análisis Bayesiano de datos

1. Identificar los datos relevantes para abordar la pregunta de investigación.
2. Definir el modelo descriptivo de los datos: Modelo y parámetros (¿Qué modelos (Golems) hay?)
3. Especificar una distribución a priori de los parámetros (¿Tipos de priors?)
4. Usar inferencia Bayesiana para reubicar la credibilidad sobre el espacio de los valores de los parámetros: Posterior. (¿Cómputo Bayesiano?)
5. Interpretar las distribuciones posteriores (¿Pruebas de hipótesis?)
6. Hacer diagnósticos de las distribuciones posteriores y de la capacidad de los parámetros posteriores del modelo (para reproducir los datos). (¿Posibles diagnósticos y tests?)



---

# Ejemplo

A series of thin, light-colored wavy lines that sweep across the bottom of the slide, creating a sense of movement and design.



# ¿Cuál es la incertidumbre asociada a la relación peso y estatura?

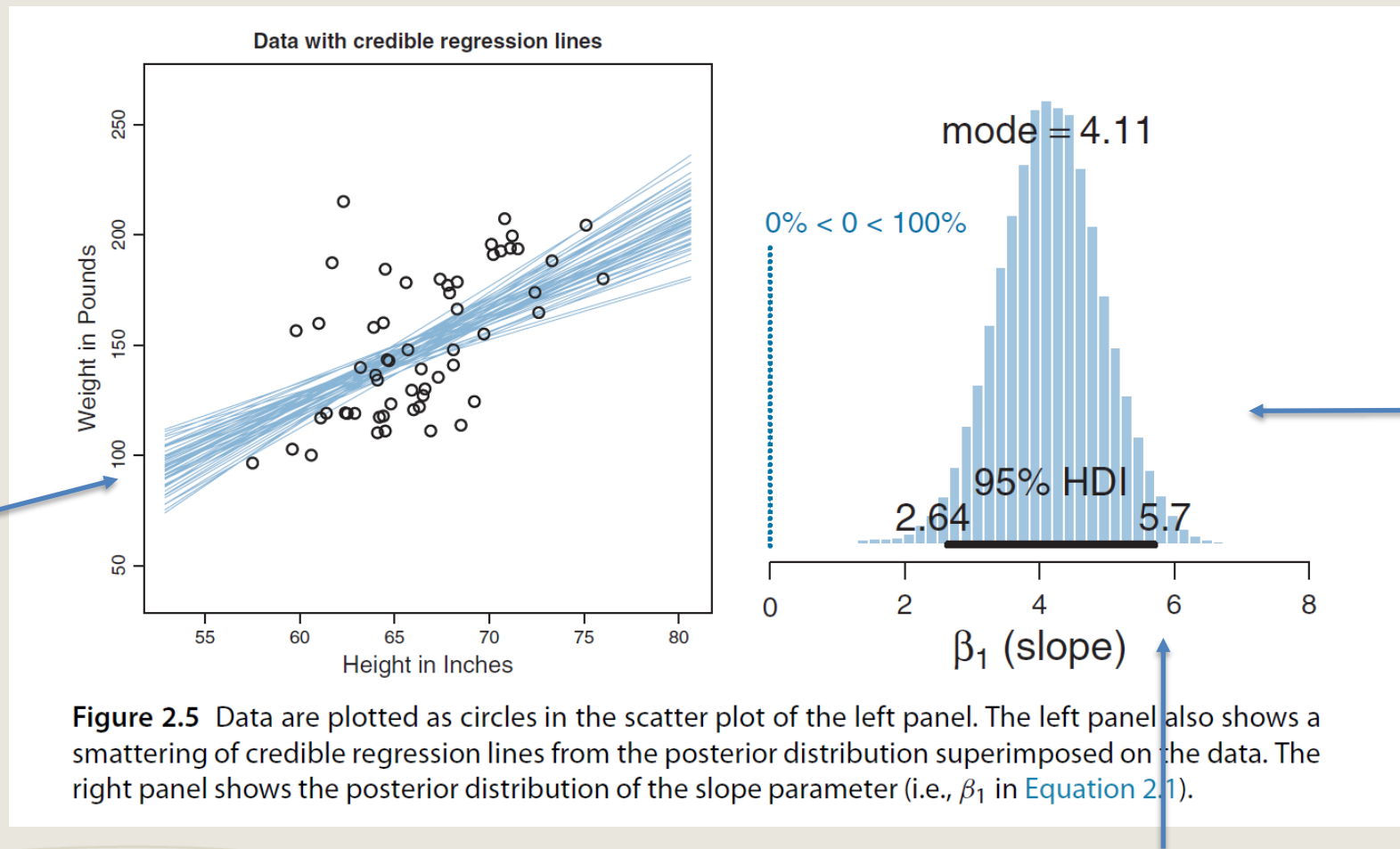
- Me interesa saber la incertidumbre asociada al parámetro que describe el cambio en el peso dada la estatura
- Modelo descriptivo de los datos:  $\hat{y} = \beta_1 X + \beta_0$
- Variación aleatoria:  $y \sim normal(\hat{y}, \sigma)$ 
  - Hay tres parámetros/hipótesis:  $\theta$
- Prior: Imaginemos que hay poca evidencia al respecto  $\theta \sim normal(0, 10^5)$
- Hacemos nuestros cálculos... -Ya veremos cómo-



# Ejemplo: ¿Cuál es la incertidumbre asociada a la relación peso y estatura?

Las posibles líneas que describen estos datos: Las Betas posteriores de la relación lineal

¿Qué diferencias notan respecto a su modelo de regresión lineal?

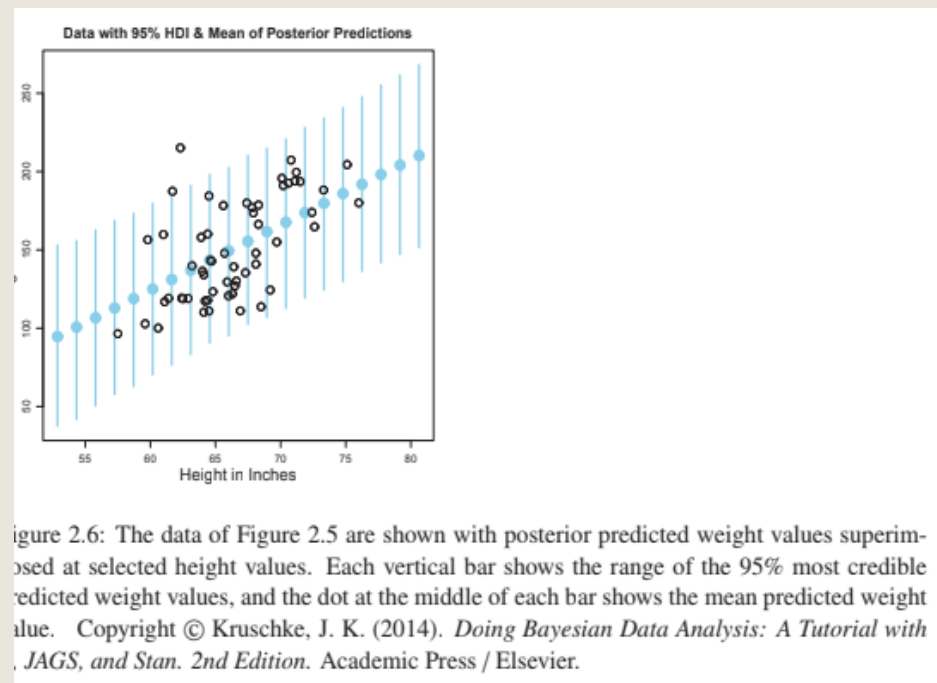


Estas betas resultan en una distribución de nuevas posibilidades (recuerden que supuse que se distribuía normalmente)

Intervalo creíble. Simplemente la desviación estándar de la distribución.

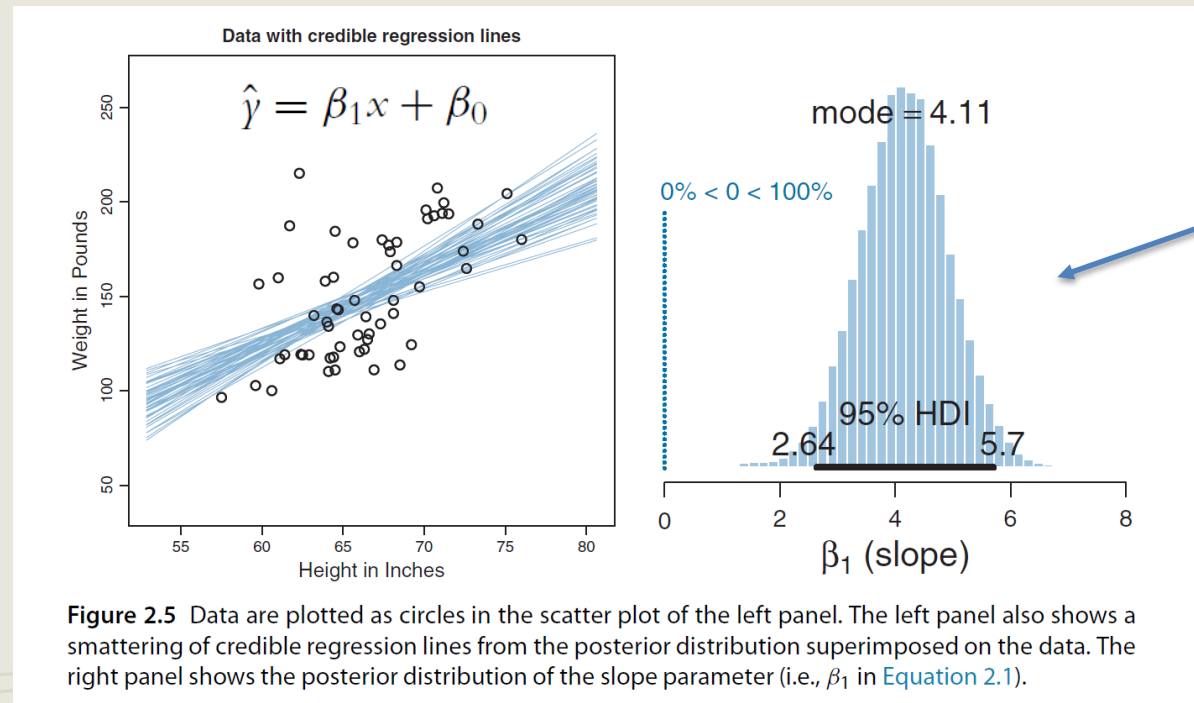
# ¿Cómo sé que mi beta es confiable?

- En inferencia bayesiana el análisis comienza (no termina) con el cálculo de las distribuciones posteriori.
- Si la beta que obtuvimos hace un buen trabajo...  $P(D|\hat{\beta})$
- La beta debería hacer un buen trabajo en reproducir los datos.
- Esto se conoce como chequeos posteriores predictivos (posterior predictive checks). Verémos de qué se trata.



# Posibilidades e incertidumbre

- Reubicación de **credibilidad** (distribuciones posteriores) entre **posibilidades** (en clave parámetros de modelos probabilísticos)
- Incertidumbre en torno a parámetros (hyper-parámetros: parámetros que describen la distribución de parámetros)



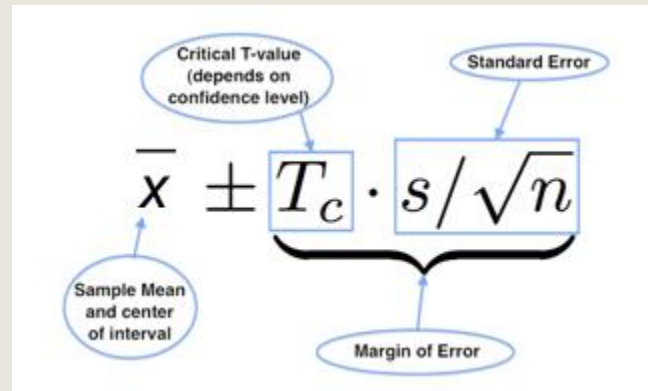
Cuantificación de la incertidumbre en el paradigma bayesiano (¿qué tan buenas son nuestras estimaciones?)

En el enfoque clásico/frecuentista se estima valor puntual del parámetro, no su distribución



# Incertidumbre en estadística clásica

- ¿Pero qué no el intervalo de confianza hace lo mismo?, ¿qué no calcula una probabilidad?
  - Sí, ¿pero la probabilidad de qué?



The diagram illustrates the formula for a confidence interval:  $\bar{x} \pm T_c \cdot s/\sqrt{n}$ . It includes the following labels and arrows:

- Sample Mean and center of interval**: Points to  $\bar{x}$ .
- Critical T-value (depends on confidence level)**: Points to  $T_c$ .
- Standard Error**: Points to  $s/\sqrt{n}$ .
- Margin of Error**: A bracket under the product  $T_c \cdot s/\sqrt{n}$  points to this label.

# Incertidumbre en estadística clásica

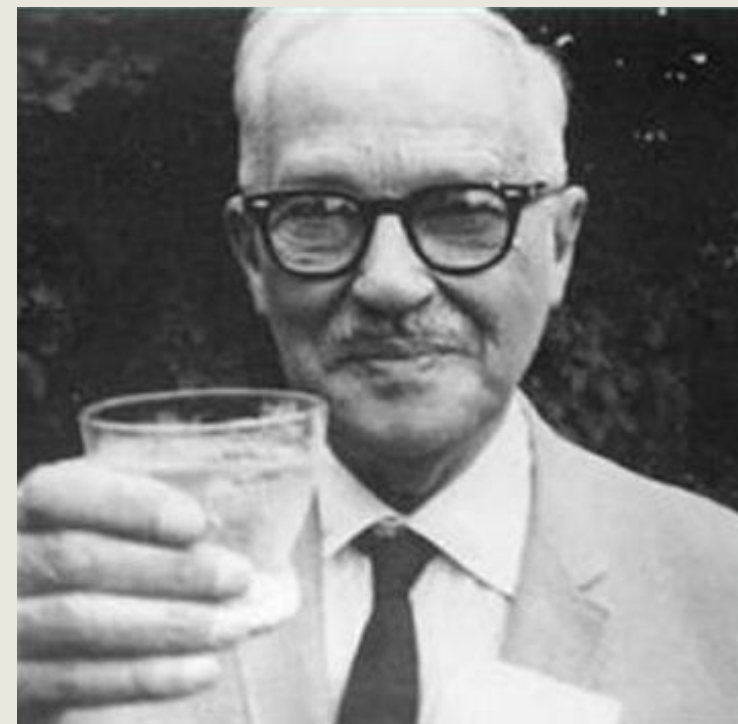
- ¿Cómo calcula una un intervalo de confianza?
- Neyman, J. 1934. On the two different aspects of the representative method. Plática ante la Royal Statistical Society.
  - "I am not at all sure that the 'confidence' [in confidence interval] is not a 'confidence trick.'"
- Bowley, A. L.
- Las cuatro líneas de Neyman en el apéndice conducen al cálculo elegante de una probabilidad, pero no está claro a qué corresponde esa probabilidad en “la vida real”.



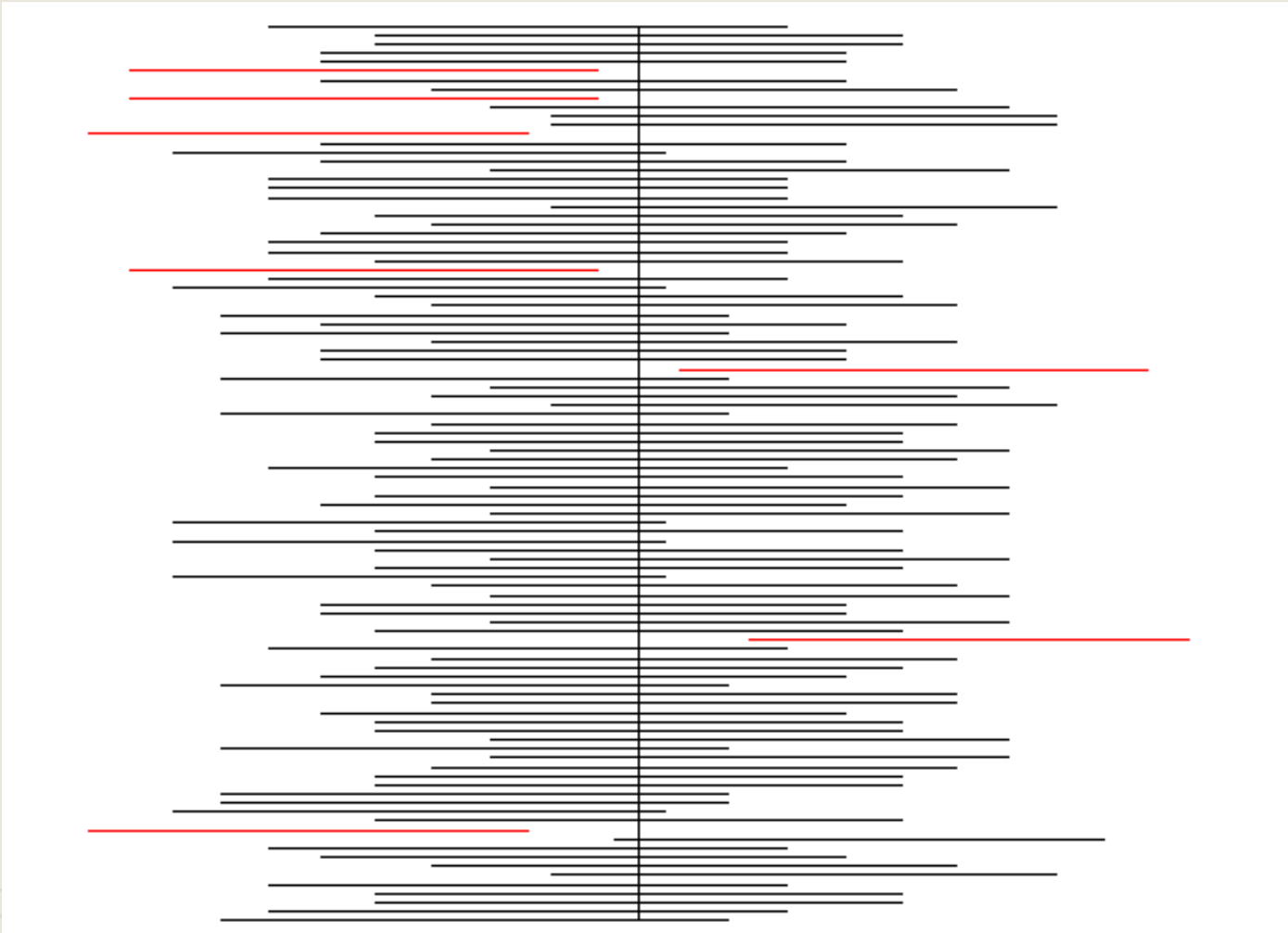
Jerzy Neyman 1894-1981  
(foto de 1969)

# Incertidumbre en estadística clásica

- De acuerdo con Neyman, el intervalo de confianza no debe verse en términos de cada conclusión, sino como un proceso (aleatorio). Si una investigadora calcula intervalos de confianza **de esta manera** a partir de muchas muestras **aleatorias** (repeticiones idénticas e independientes del **experimento** de muestreo estadístico subyacente), 95 % de los intervalos incluirán, en el largo plazo, el valor verdadero del parámetro de interés.
  - Fíjense que no dice nada acerca de la probabilidad de que una estimación en particular esté en lo correcto, ni cuán “acertada” es nuestra estimación actual.
  - Es la frecuencia de afirmaciones correctas que la investigadora que use este método (en una secuencia larga de ensayos aleatorios) hará en el largo plazo.



# Incertidumbre en estadística clásica





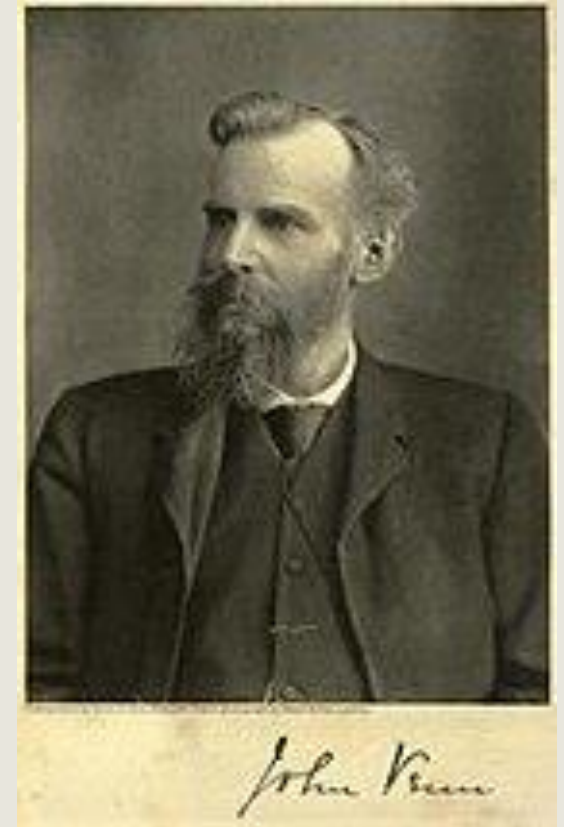
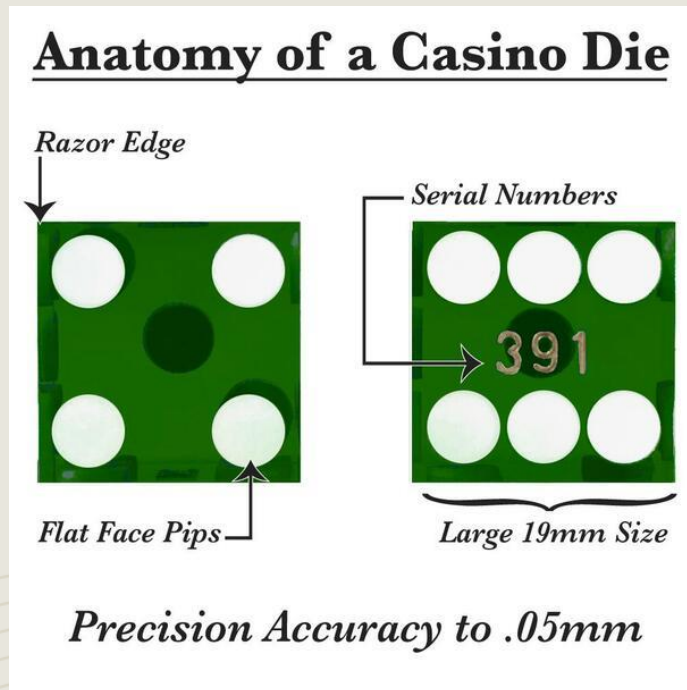
# ¿Rechazar la hipótesis nula es la felicidad?

- ¿Es el objetivo de la investigación científica?
  - Se pueden rechazar hipótesis nulas con
    - efectos despreciables
    - precisión extremadamente baja
- Paradoja de Meehl (Presidente de la American Psychological Association, 1962)
  - Ganar precisión en nuestras medidas debería hacer más difícil confirmar las hipótesis teóricas que les involucran. Exige más de nuestras teorías.
  - Predicciones teóricas en clave de hipótesis alternativa (i.e., rechazar la hipótesis nula) se verifican más fácilmente como la precisión en la medición aumenta. Después de todo, son raras las ocasiones en las que la hipótesis nula puede verificarse **exactamente**.
- ¿Qué hay de la estimación y el modelaje, representar/explicar e intervenir?



# Incertidumbre en estadística clásica

- Si un evento tiene una **probabilidad dada**, y corremos **ensayos aleatorios** una y otra vez, la proporción (**frecuencia**) con la que el evento ocurre se irá acercando más y más a esa probabilidad.



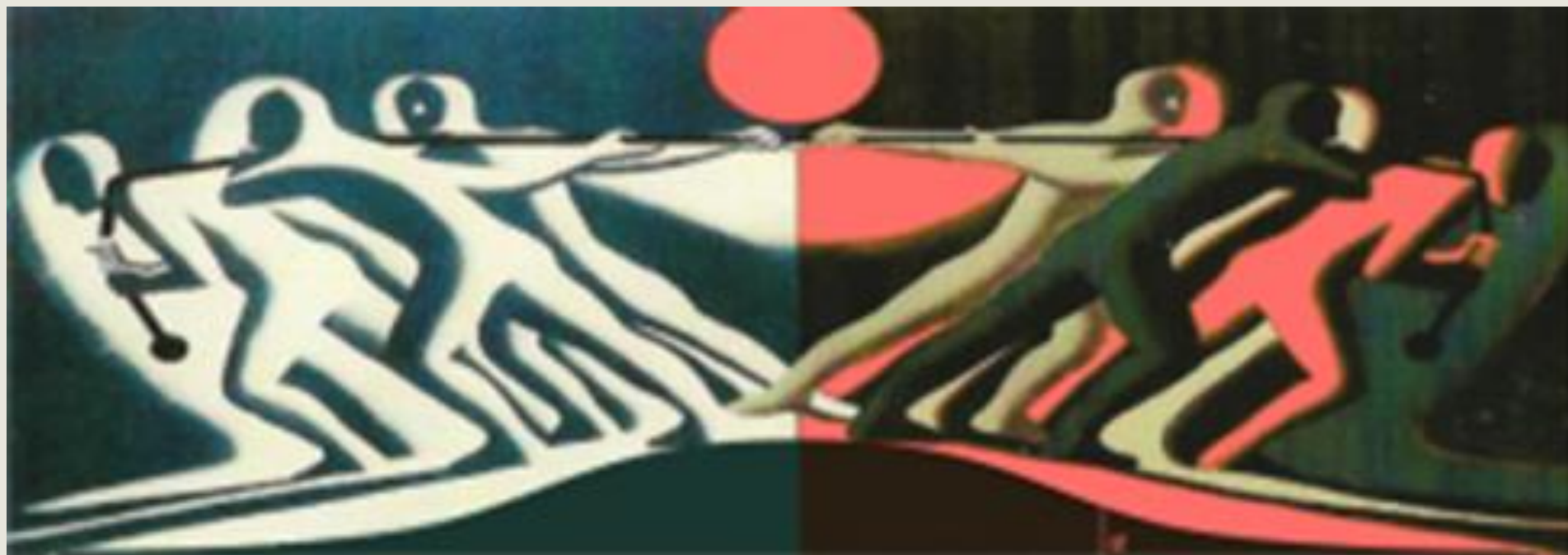
John Venn 1834-1923

**La ley de grandes número implica la probabilidad y no al revés**



# La Guerra de las Estadísticas

---



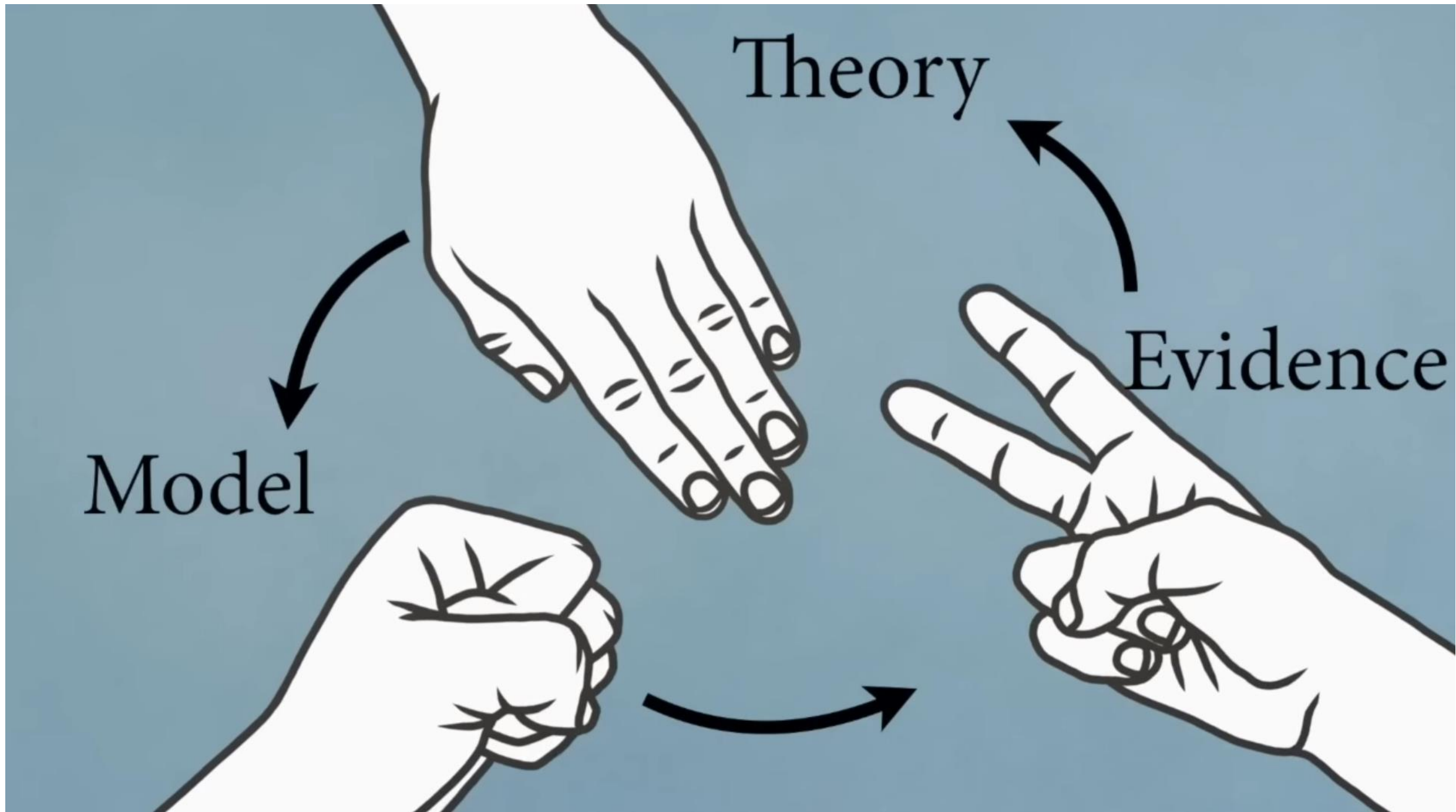
# La Guerra de las Estadísticas



# La Guerra de las Estadísticas



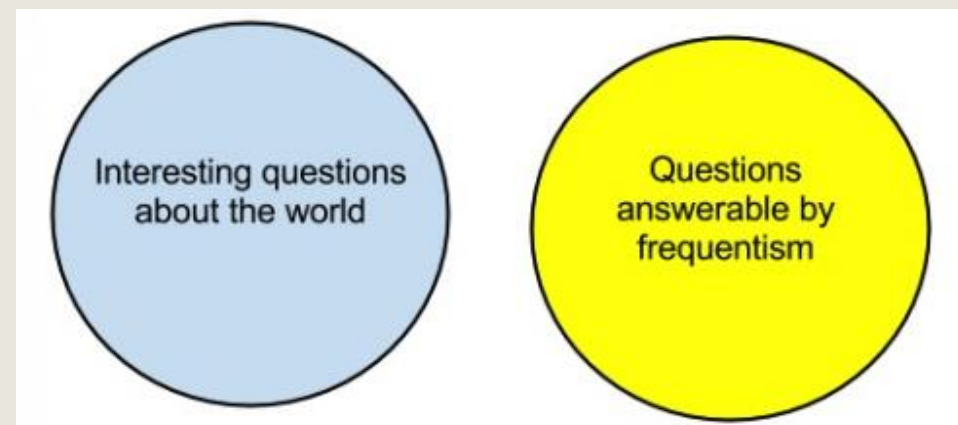




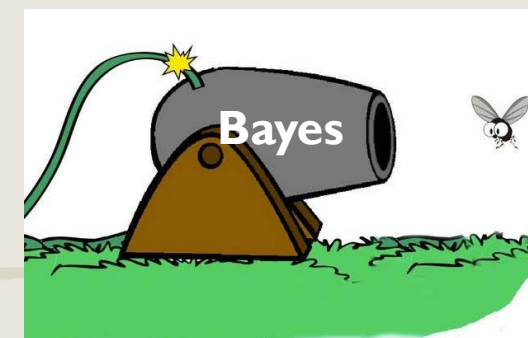
# Investigación antes que estadística

- Justificar el análisis de los datos (bayesiano o frecuentista) en la derivación de nuestras estimaciones y su incertidumbre asociada
- Poner la estadística al servicio de la investigación científica (conocimiento causal)
  - Explicar
  - Intervenir
- Modelos estadísticos justificados por (hechos a la imagen y semejanza de) modelos científicos/causales generativos transparentes (capaces de producir/sintetizar datos)

vs simple asociación y predicción  
(NHT)

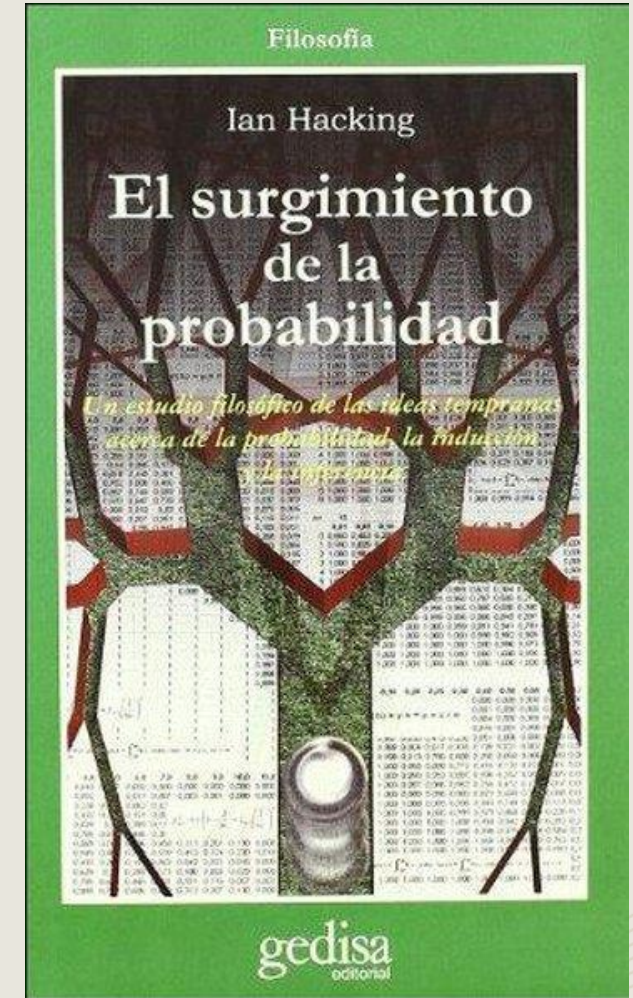


- control industrial de calidad
- experimentos de control aleatorio



# 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 zar).
  - 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.





# Investigación antes que estadística





# Próxima clase

---

## Capítulo 4

What is this stuff called probability?



- Andrew Gelman (2011), “Induction and Deduction in Bayesian Data Analysis“, Special Topic: Statistical Science and Philosophy of Science RMM Vol. 2, 2011, 67–78
- 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.
- Kruschke, J. (2014). Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan.
- Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science*, 34(2), 103-115.
- 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.
- Salsburg, D. (2001). The lady tasting tea: How statistics revolutionized science in the twentieth century. Macmillan.



# 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](mailto:hecatalan@hotmail.com), [chuffman@unam.mx](mailto:chuffman@unam.mx)

