Principios generales

Las bases para entender por qué el modelamiento bayesiano es útil en ciencias cognitivas

Probabilidad e Inferencia Probabilística

Lidiando con la Incertidumbre

Inferencia Probabilística

Implica hacer un cómputo sobre la probabilidad de que ocurra un evento X

Probabilidad

Se define como **un número real del 0 al 1 que define la certidumbre** que se tiene respecto de la ocurrencia o no ocurrencia de un evento específico.

0 "Imposible" — 1 "Seguro que sí"

- La probabilidad de que mañana salga el Sol
- La probabilidad de que los cerdos vuelen
- La probabilidad de que me salga águila en un volado

Estimación de Probabilidades

La probabilidad real con que ciertos eventos en el mundo ocurren, está oculta.

Sin información.

Se asume que todos los eventos posibles son igualmente probables.

$$\frac{\#Evento\ particular}{\#\ TOTAL\ Eventos\ posibles}$$

Con información (Observaciones)

La probabilidad se estima como una tasa

$$\frac{\#Veces\ que\ apareció\ el\ evento}{\#\ Total\ de\ observaciones}$$

Estimación de Probabilidades

Sacar un As	$\frac{4}{52} = .076$
Entrar a la UNAM	$8\% = \frac{8}{100}$
Encontrar a Jhonny Depp en Iztapalapa	0.00001

INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA

Continuemos..

¿Qué tan probable es que un individuo X sea zurdo?

1. Antes de haber interactuado con X

   Sabemos que 15% de la población es zurda; asumimos que hay una probabilidad de .15 de que X sea zurdo.

CONTINUEMOS..

¿Qué tan probable es que un individuo X sea zurdo?

2. Después de haber interactuado con X

Voy recopilando información:

- ¿En qué mano tiene su reloj?
- ¿En qué hombro carga su mochila?
- ¿Con qué mano manipula los objetos en su entorno?

Probabilidad Condicional

"La probabilidad de un evento A, dada cierta evidencia B"

$$P(A|B)$$

p(X es Zurdo|Escribe con la derecha) **<** p(X es zurdo)

p(X es Zurdo|Escribe con la izquierda) **>** p(X es zurdo)

Eventos Independientes

La ocurrencia de B NO me dice nada sobre la posible ocurrencia de A.

Evento A:

La probabilidad de que va a llover

Evento B:

Hoy es miércoles

$$p(A|B) = p(A)$$

Eventos NO independientes

La ocurrencia de B me permite hacer una mejor predicción sobre la posible ocurrencia de A.

Evento A:

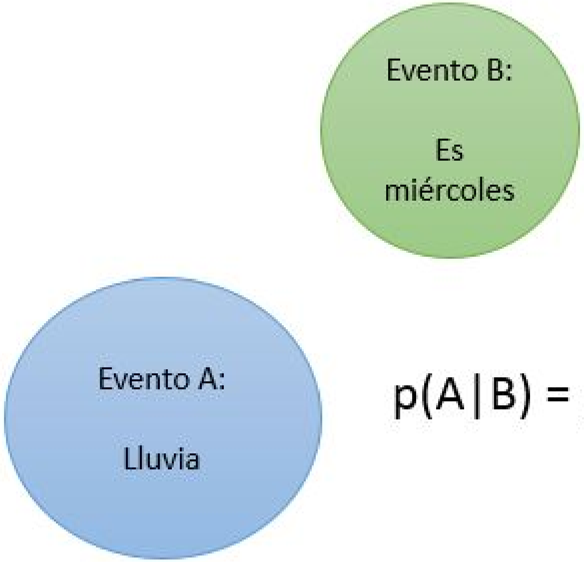
La probabilidad de que va a llover

Evento B:

Está nublado

$$p(A|B) > p(B)$$

Eventos Independientes



$$p(A|B) = p(A)$$

Eventos NO independientes

$$p(A|B) > p(A)$$

$p(A|B) > p(A)$

OJO!

$$p(A|B) \neq p(B|A)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A \cap B)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A \cap B)$$

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

Teorema de Bayes

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

Teorema de Bayes en acción:

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

Teorema de Bayes en acción:

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

- **La probabilidad de un evento A, dada la evidencia B.**
  - La probabilidad de que llueva dado que está nublado.

Teorema de Bayes en acción:

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

- **La probabilidad de un evento A, dada la evidencia B.**

- **La probabilidad del evento A**
  - La probabilidad de que llueva

Teorema de Bayes en acción:

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

- **La probabilidad de un evento A, dada la evidencia B.**
- **La probabilidad del evento A**
- **La probabilidad de observar la evidencia B, cuando el evento A ocurre.**
    - La probabilidad de que el cielo esté nublado, dado que está lloviendo

Teorema de Bayes en acción:

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

- La probabilidad de un evento A, dada la evidencia B.
- La probabilidad del evento A
- La probabilidad de observar la evidencia B, cuando el evento A ocurre.
- La probabilidad de observar la evidencia.
  - La probabilidad de que esté nublado

Acerca de la Verosimilitud Marginal

$$p(B) = \sum p(A_i \cap B)$$

$$p(B) = p(A \cap B) + p(A' \cap B)$$

¿Para qué sirve hacer una inferencia?

Para lidiar con la incertidumbre y guiar nuestro comportamiento de la mejor manera posible dada la estructura del entorno.

Por ejemplo:

Imaginemos el caso de un doctor que intenta determinar si el paciente tiene una enfermedad A, a partir de los resultados obtenidos en una prueba B.

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

Ejemplo 1:

¿La tarta está envenenada?

¿Cuál es la probabilidad de morir al comer la tarta?

1. **Definimos el problema**

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

¿Cuál es la probabilidad de morir al comer la tarta?

2. Repasamos los datos

Tartas de Fresa

50 envenenadas

30 NO envenenadas

Tartas de Vainilla

50 envenenadas

170 NO envenenadas

¿Cuál es la probabilidad de morir al comer la tarta?

3. Ubicamos los elementos del Teorema de Bayes

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

La prior

P(A|B) = ???

P(A) =
p(Veneno)

¿CUÁL ES LA PROBABILIDAD DE MORIR AL COMER LA TARTA?

3. Ubicamos los elementos del Teorema de Bayes

La verosimilitud

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

P(A|B) = ???

P(A) =

P(B|A) =

p(Tarta de Vainilla | Veneno)

¿Cuál es la probabilidad de morir al comer la tarta?

3. Ubicamos los elementos del Teorema de Bayes

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

La verosimilitud marginal

P(A|B) = ???

P(A) =

P(B|A) =

P(B) =
p(Tarta de Vainilla)

(((Notemos que para el Cálculo de la verosimilitud marginal...)))

$$p(B) = \sum p(A_i \cap B)$$

¿Cuál es la probabilidad de morir al comer la tarta?

4. Sustituimos y resolvemos la ecuación

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

P(A|B) = ???

P(A) =

P(B|A) =

P(B) =

(((Comprobando que los eventos sean complementarios…)))

$$p(A|B) + p(A'|B) = 1$$

Ejemplo 2:

¿Es mi maleta?

¿Es mi Maleta?

¿Qué implica decir "Bayesiano"?

Hay una actualización constante de la información que me permite reducir mi incertidumbre respecto a la probabilidad de ocurrencia de un evento X.

Ampliando Bayes

Continuo y más de una evidencia

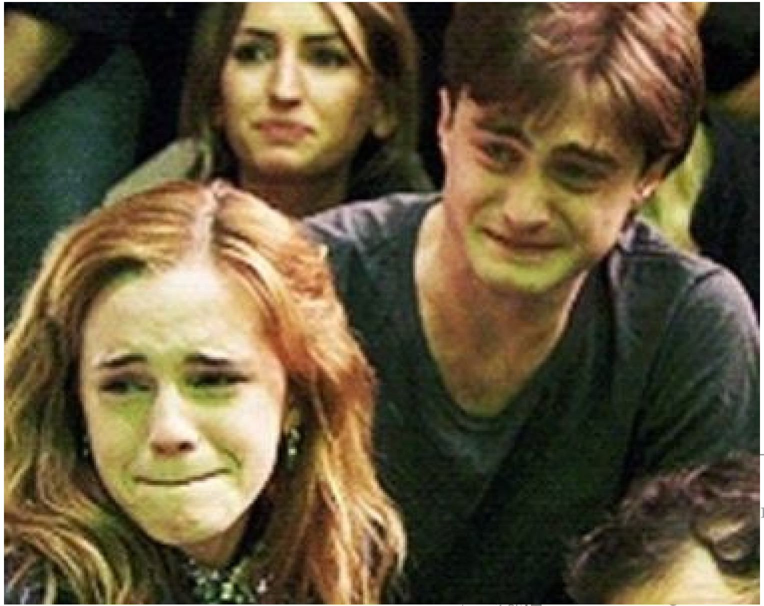
Bayes y más de una evidencia

El problema del sombrero seleccionador

Por años el sombrero había asignado a los estudiantes de Hogwards con una perfecta exactitud y un perfecto balance, es decir, a cada casa le tocaba 25% de los estudiantes que llegaban a Hogwards.

Pero todo cambió cuando Voldemort atacó!!!!!

Ahora el sombrero asigna al 40% de los estudiantes a Slytherin. Mientras que a las demás casas solo el 20% cada una.



Pero la profesora Sybill siempre sabe qué hacer!

Y crea una prueba psicomágica (KHÉ!!!) en la cual los estudiantes de las diferentes casas tienden a entrar en las siguientes categorías:

Slytherin (Ms)	Excelentes (Se)
Ravenclaw (Mr)	Excepcionales (So)
Gryffindor (Mg)	Aceptables (Sa)
Hufflepuff (Mh)	Pobres (Sp)

Y como siempre Hogwarts ya tenía datos de sus estudiantes antes de que sombrero fuera dañado que nos aproximan a las probabilidades de cada casa a un tipo particular de resultado

	Excelente (Se)	Excepcionales (So)	Aceptables (Sa)	Pobres (Sp)
Slytherin (Ms)	**0.80**	0.10	0.05	0.05
Gryffindor (Mg)	0.05	0.20	**0.70**	0.05
Ravenclaw (Mr)	0.05	**0.80**	0.15	0.00
Hufflepuff (Mh)	0.00	0.10	0.25	**0.65**

APLIQUEMOSLO A LOS GRIFFINDOREANOS!!!!

¿Cuál es la probabilidad de que alguien que saque un puntaje de Excelente en la prueba y que el sombrero haya seleccionado como Slytherin sea en realidad un Gryffindor?

$$P(\mathcal{M}_G|D_S, S_E) = \frac{P(\mathcal{M}_G)P(D_S, S_E|\mathcal{M}_G)}{P(D_S, S_E)}$$

$$P(\mathcal{M}_G|D_S,S_E)=\frac{P(\mathcal{M}_G)P(D_S,S_E|\mathcal{M}_G)}{P(D_S,S_E)}$$

$$P(\mathcal{M}_G|D_S,S_E)=\frac{P(\mathcal{M}_G)P(S_E|\mathcal{M}_G)P(D_S|S_E,\mathcal{M}_G)}{P(D_S,S_E)}.$$

$$\begin{aligned}P(\mathcal{M}_G|D_S,S_E) &= \frac{P(\mathcal{M}_G)P(S_E|\mathcal{M}_G)P(D_S|\mathcal{M}_G)}{P(D_S,S_E)}\\ &= \frac{0.25\times 0.05\times 0.20}{P(D_S,S_E)}.\end{aligned}$$

$$\begin{aligned}P(D_S,S_E) &= \sum_i P(\mathcal{M}_i)P(S_E|\mathcal{M}_i)P(D_S|\mathcal{M}_i)\\ &= P(\mathcal{M}_S)P(S_E|\mathcal{M}_S)P(D_S|\mathcal{M}_S)\\ &\quad + P(\mathcal{M}_G)P(S_E|\mathcal{M}_G)P(D_S|\mathcal{M}_G)\\ &\quad + P(\mathcal{M}_R)P(S_E|\mathcal{M}_R)P(D_S|\mathcal{M}_R)\\ &\quad + P(\mathcal{M}_H)P(S_E|\mathcal{M}_H)P(D_S|\mathcal{M}_H)\end{aligned}$$

$$\begin{aligned}&= 0.25\times 0.80\times 1.00\\ &\quad + 0.25\times 0.05\times 0.20\\ &\quad + 0.25\times 0.05\times 0.20\\ &\quad + 0.25\times 0.00\times 0.20\\ &= 0.2050.\end{aligned}$$

$$P(\mathcal{M}_G|D_S,S_E)=\frac{0.0025}{0.2050}=0.0122.$$

Ahora en continuo...

Bayes en términos continuos

COsas que deberíamos saber pa' iniciar

- Una variable aleatoria es una función que asigna un número real a cada evento del espacio muestra.

- Se acostumbra usar letras mayúsculas para las V.A.

- Por ejemplo si queremos expresar la probabilidad de que cierta variable aleatoria X asuma algún valor no menor a 2 ni mayor que 4, entonces:

$$P(2 \leq X \leq 4)$$

Clasificando a las V.A.

- Hay de dos sopas pa’ las variables aleatoria:

1) Discretas: Pueden tomar un número de valores que se puedan enumerar aunque sean infinitos.

2) Continuas: Están definidas en un conjunto no numerable (o continuo) de puntos que no se pueden contar, sino medir.

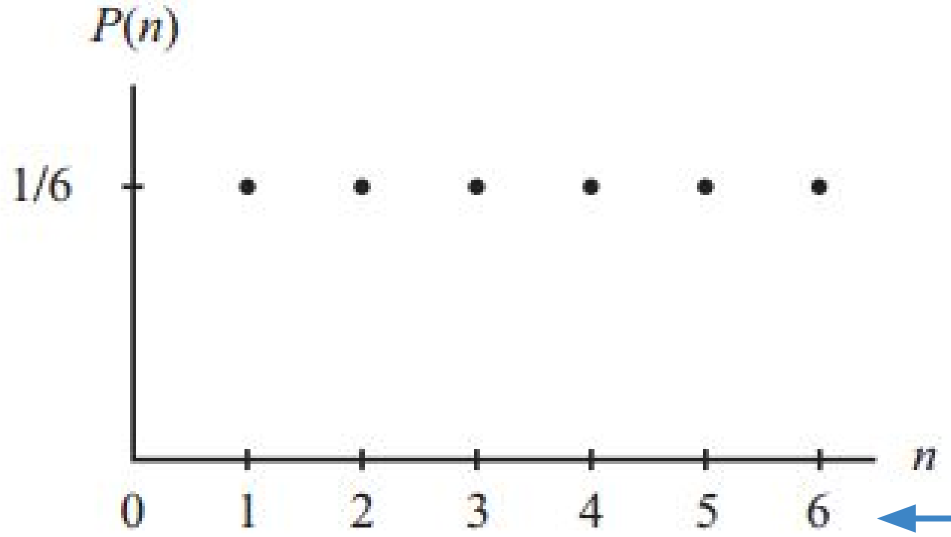
Primero aclaremos algo... Distribuciones de probabilidaD

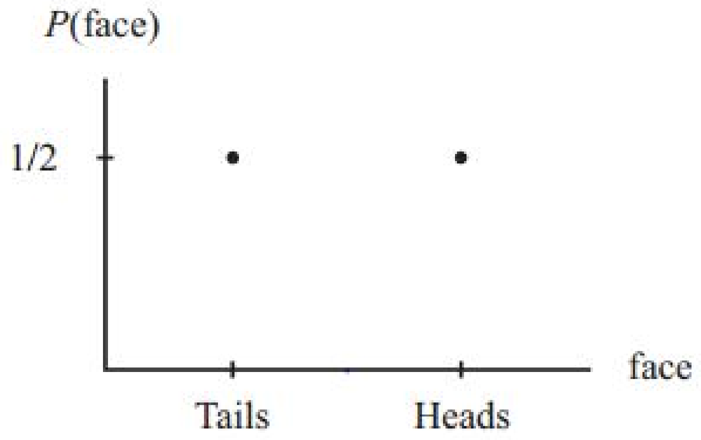
- Una distribución de probabilidad como una simple lista de posibles resultados y sus correspondientes probabilidades.

- Simplemente nos dice cuál es la probabilidad que puede tomar cierta variable aleatoria…

- O cómo distribuimos nuestras creencias acerca del valor que puede tomar una variable aleatoria

La distribución de proba para algunos resultados





La v.a. tomaría el valor de 0 y 1

Los posibles valores que la v.a. puede tomar

Tons, si ya estamos un poco más claros...

- A cada V.A. (sin importar que sea discreta o continua) tiene asociada una distribución de probabilidad, la cual se expresa generalmente por medio de una fórmula.

- Y la distribución de probabilidad es una especie de ley matemática que rige el comportamiento estocástico (o aleatorio) de la variable en cuestión.

SEAMOS MÁS EXQUISITOS...

Cuando el espacio muestral está compuesto de eventos discretos vamos a hablar de **MASA DE PROBABILIDAD**, pero cuando hablamos de eventos continuos hablamos de **DENSIDAD DE PROBABILIDAD**. Peeeroooo… ¿Cuál es la diferencia?

Distingamos un poco...

Densidad de probabilidad NO ES IGUAL a masa de probabilidad.

$$p(A) \neq P(A)$$

¿¿¿¿Cuál es la probabilidad de un evento Continuo????

Entonces ¿¿¿cómo calculamos probabilidades para casos continuos???

Veamos algunas soluciones...

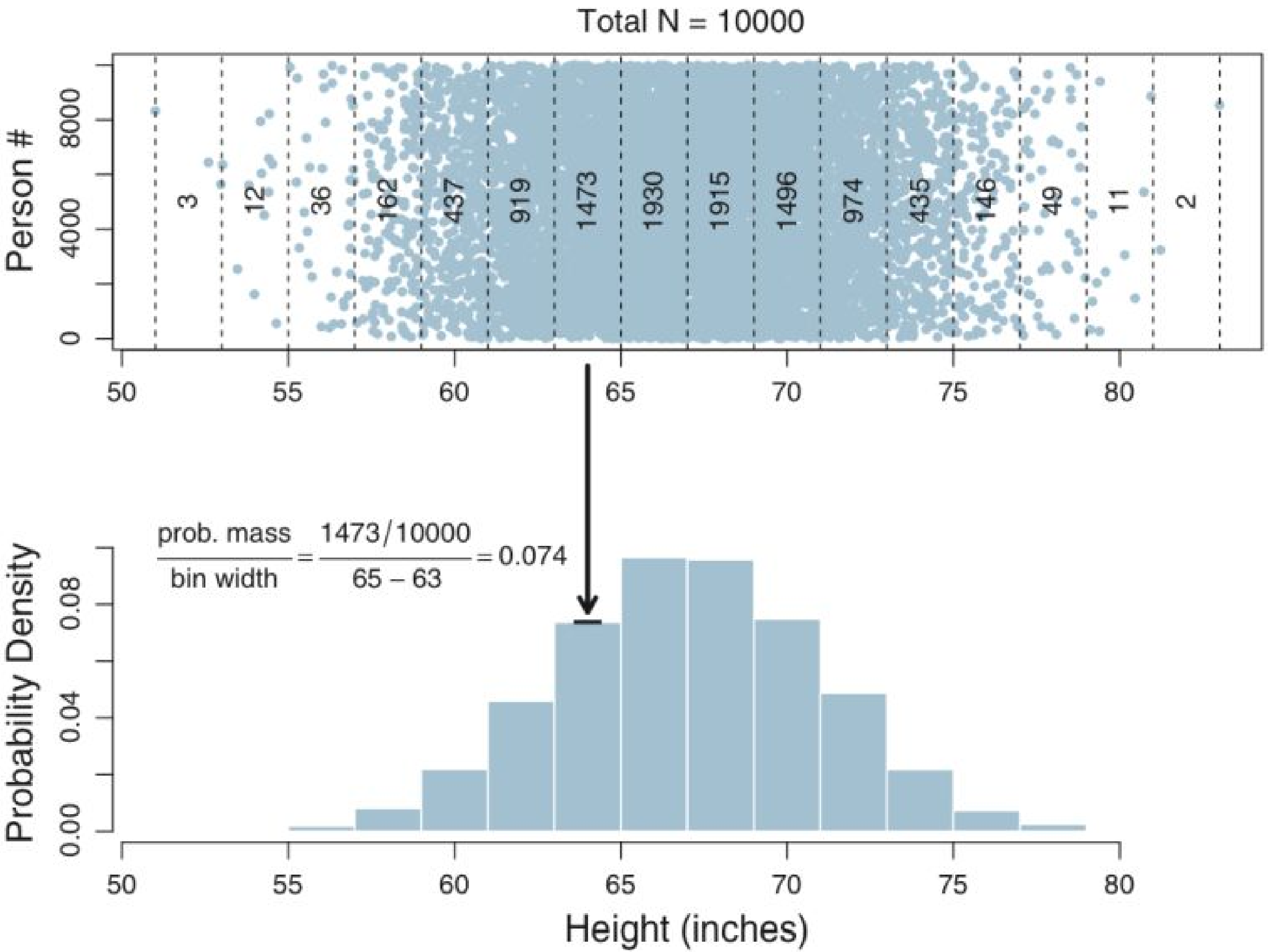
- Primera Solución: Tal vez no podemos hablar de eventos puntuales pero podemos hablar de la masa de probabilidad de intervalos, peeeeroooo… el ancho y borde podrían ser arbitrarios. Por lo que la segunda solución es:

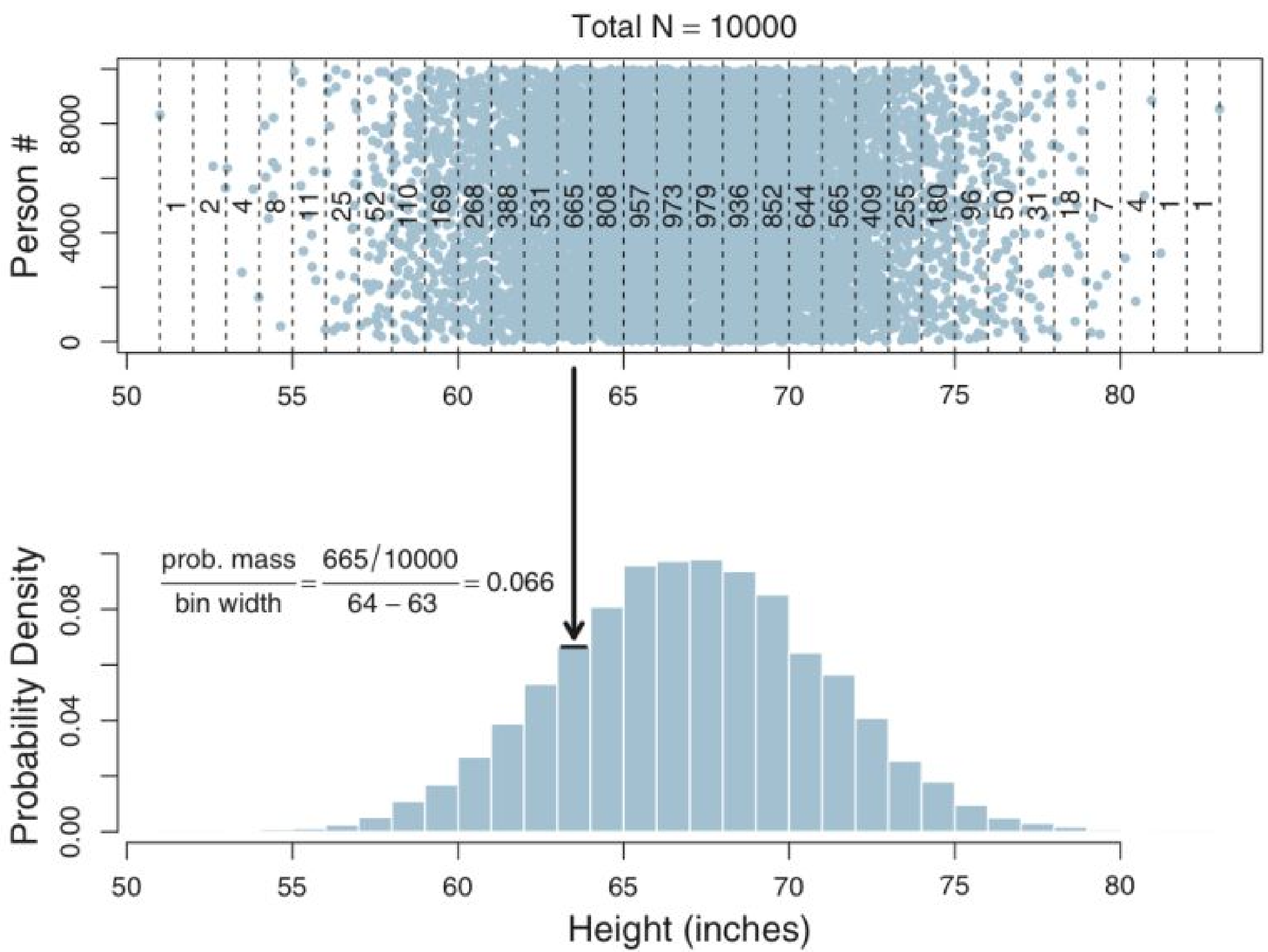
- Hacer los intervalos de un ancho infinitesimal.

Ahondemos un poco más...

- La densidad de probabilidad, más intuitivamente es la razón entre la masa de probabilidad y el ancho del intervalo.

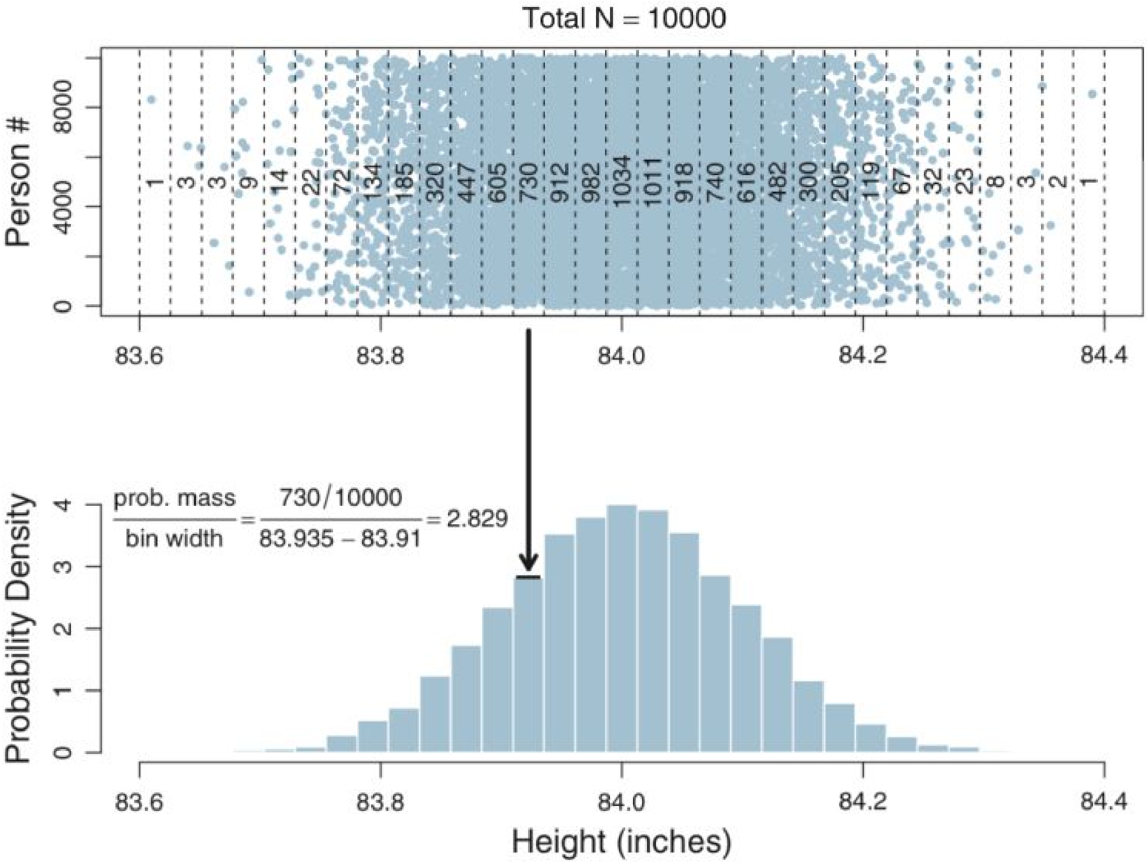
- Digamos que es la cantidad de algo por unidad de espacio que ocupa.


Total N = 10000
Person #
0
4000
8000
3
12
36
162
437
919
1473
1930
1915
1496
974
435
146
49
11
2
50
55
60
65
70
75
80
prob. mass / bin width = 1473/10000 / 65 − 63 = 0.074
Probability Density
0.00
0.04
0.08
Height (inches)



Total N = 10000
Person #
0
4000
8000
1
2
4
8
11
25
52
110
169
268
388
531
665
808
957
973
979
936
852
644
565
409
255
180
96
50
31
18
7
4
1
1
50
55
60
65
70
75
80
prob. mass / bin width = 665/10000 / 64 − 63 = 0.066
Probability Density
0.00
0.04
0.08
Height (inches)

Aclarando, la densidad de proba puede ser mayor a 1



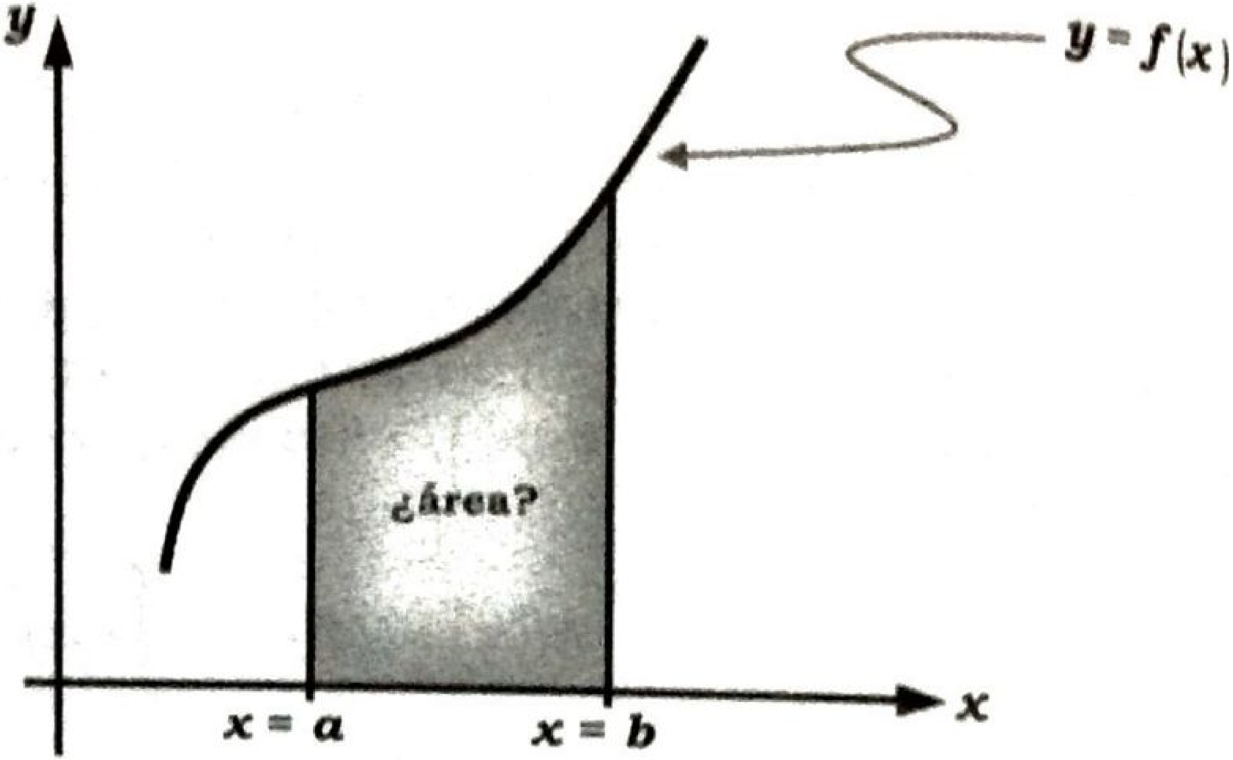
$$\sum_i p\left([x_i, x_i + \Delta x]\right) = 1$$

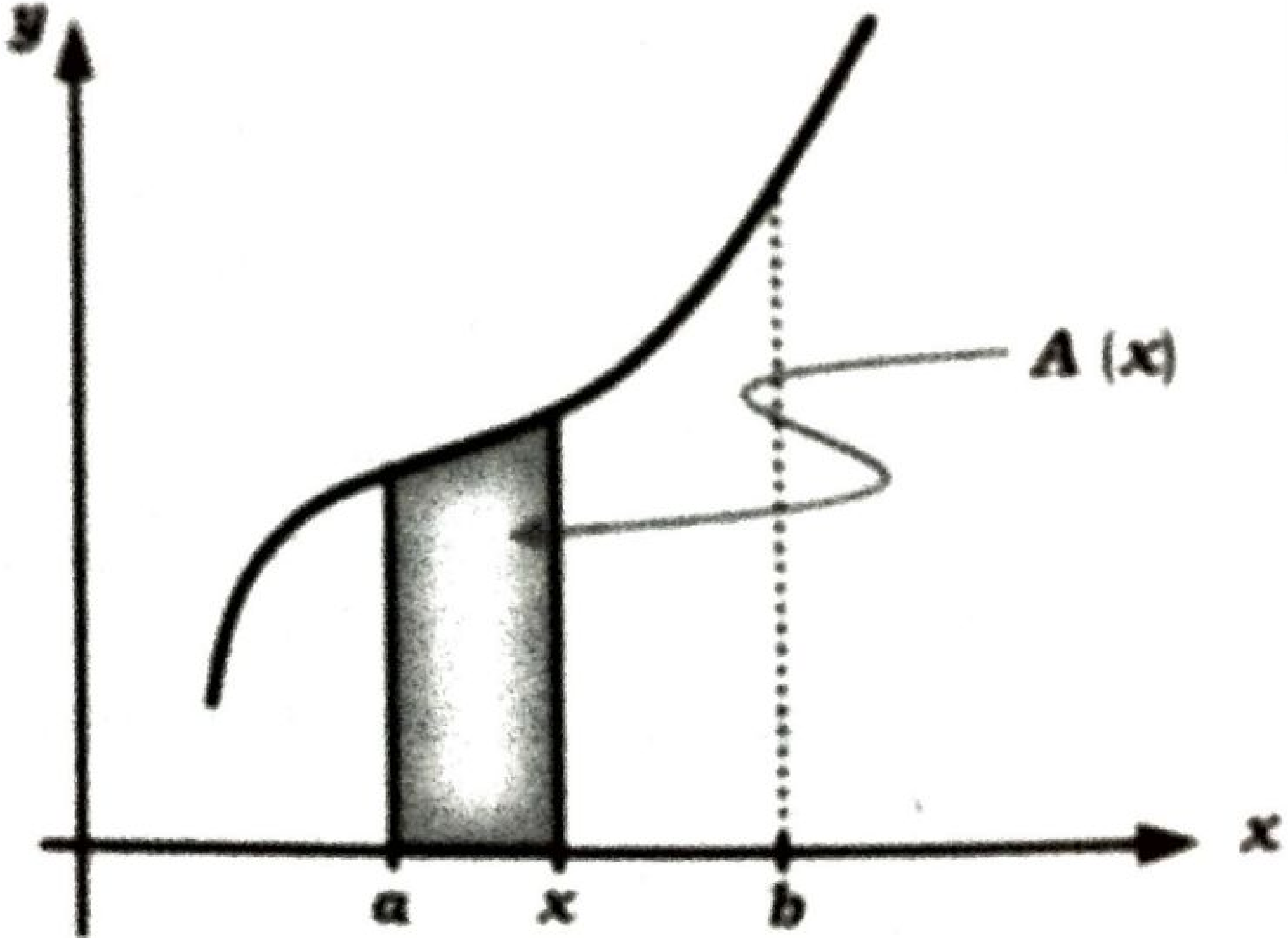
$$\sum_i \Delta x \, \frac{p\left([x_i, x_i + \Delta x]\right)}{\Delta x} = 1$$

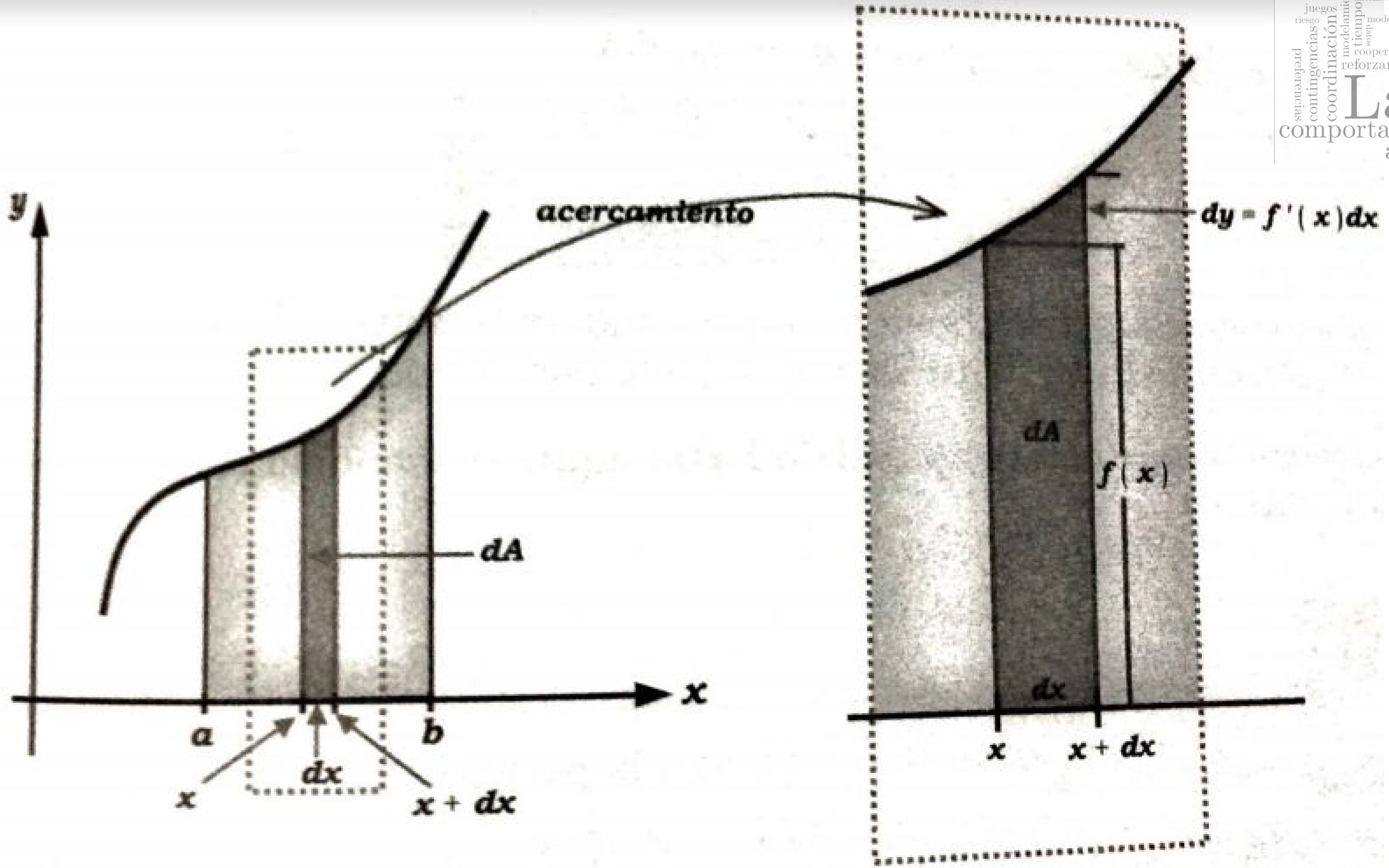
$$\underbrace{\sum_i}_{\int} \underbrace{\Delta x}_{\mathrm{d}x} \underbrace{\frac{p([x_i, x_i + \Delta x])}{\Delta x}}_{p(x)} = 1 \quad \text{that is,} \quad \int \mathrm{d}x \, p(x) = 1$$

Segunda solución

Recordemos nuestras clases de cálculo...




y
A (x)
x
a
x
b


y
acercamiento
dA
x
a
b
x
dx
x + dx
dy = f ′ (x)dx
dA
f (x)
dx
x
x + dx

Entonces...

Por definición el área bajo la curva de cierto rango puede expresarse como la densidad de probabilidad.

O algo así como…

"p(A) es una función de A, en la cual el área bajo la curva en p(A) entre A y A+ΔA da la probabilidad de que A esté entre alguno de estos valores."

$$1 = \int_A p(a)da,$$

Caso discreto	*Caso continuo*
$P(X = x) = p(x) = f(x)$	$f(x)$ = expresión matemática de la fdp
$0 \leq p(x) \leq 1$	$0 \leq f(x)$
$\sum_{\text{toda } x} p(x) = \sum_{i=1}^{n} p_i = 1$	$\int_{-\infty}^{\infty} f(x)dx = 1$
$F(x) = P(X \leq x) = \sum_{t \leq x} p(t)$	$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$
$P(a < X \leq b) = F(b) - F(a)$	$P(a < X < b) = P(a \leq X \leq b) =$ $= F(b) - F(a)$

Recordando...

$$P(\mathcal{M}|X) = \frac{P(\mathcal{M})P(X|\mathcal{M})}{P(X)}.$$

$$P(\neg\mathcal{M}|X) = \frac{P(\neg\mathcal{M})P(X|\neg\mathcal{M})}{P(X)}$$

$$\begin{aligned} P(X) &= P(X,\mathcal{M}) + P(X,\neg\mathcal{M}) \\ &= P(\mathcal{M})P(X|\mathcal{M}) + P(\neg\mathcal{M})P(X|\neg\mathcal{M}) \end{aligned}$$

$$P(\mathcal{M}|X) = \frac{P(\mathcal{M})P(X|\mathcal{M})}{P(\mathcal{M})P(X|\mathcal{M}) + P(\neg\mathcal{M})P(X|\neg\mathcal{M})}$$

$$P(\mathcal{M}_i|X) = \frac{P(\mathcal{M}_i)P(X|\mathcal{M}_i)}{\sum_{k=1}^{K} P(\mathcal{M}_k)P(X|\mathcal{M}_k)}$$

Apliquemoslo a Bayes...

$$p(a,b) = p(a)\,p(b|a)$$

La densidad del parámetro continuo *a*

La densidad condicional de *b*

$$p(a,b) = p(a)\,p(b|a)$$

$$p(a,b) = p(b)\,p(a|b)$$

$$p(a) = \int_B p(a,b)\,db$$

$$p(a|b) = \frac{p(a,b)}{p(b)} = \frac{p(a)p(b|a)}{p(b)}$$

$$= \frac{p(a)p(b|a)}{\int_A p(a)p(b|a)da}.$$

Uso de Bayes

Estadística bayesiana

Pero primero hay que saber que...

- La estadística bayesiana es de lo que los chicos cool están hablando.

- Estadística Bayesiana guía una importante pregunta en ciencia:

  En cuál de mis hipótesis debo de creer, y qué tan fuerte debe ser mi creencia dado los datos recolectados.

Classical Statistics

¿Qué es probabilidad?

1) Probabilidad Epistémica: Es un número entre 0 y 1 que cuantifica qué tan fuerte debemos de creer que algo es verdadero dada nuestra información relevante que tenemos.

2) Probabilidad Aleatoria (o física): Probabilidad como una sentencia acerca de la frecuencia esperada sobre muchas repeticiones de un procedimiento. (Proba como la frecuencia relativa a largo plazo).

Y ESO QUÉ?

La estadística clásica conceptualiza la probabilidad como la frecuencia relativa a largo plazo de un evento.

→ **Estadística Frecuentista.**

Frecuentistas → Sus supuestos expresan propiedades de conducta a largo plazo de procesos bien definidos.

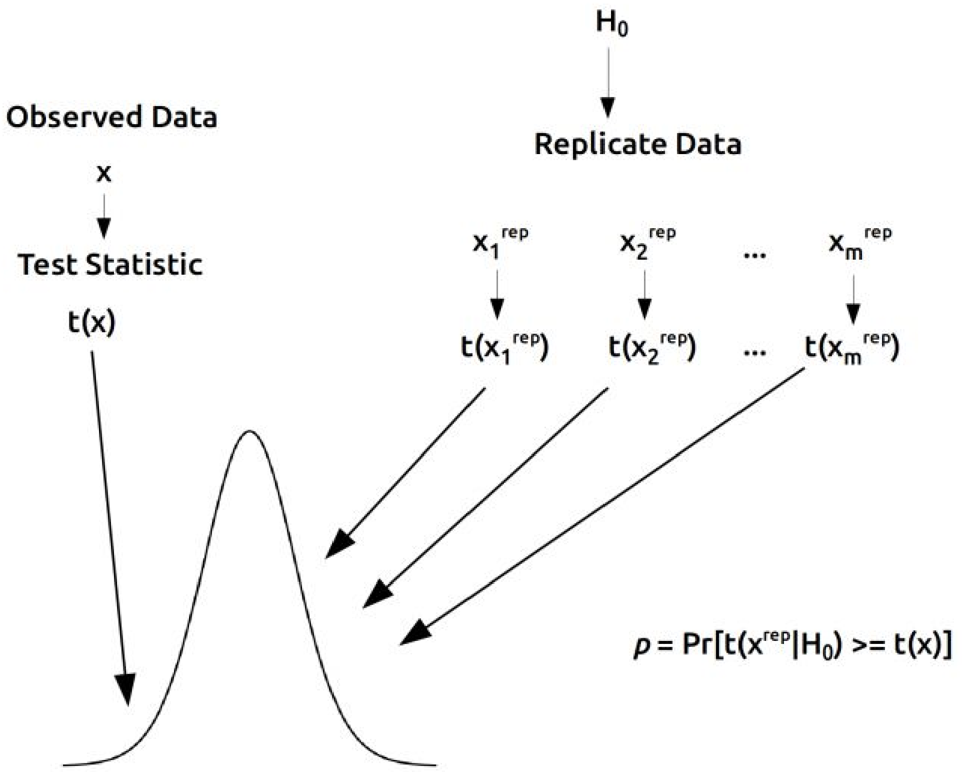
Se requiere de supuestos acerca de la repetibilidad física e independencia entre las repeticiones

Mucho menos asignar probabilidades a teorías ni a hipótesis!

Por la misma definición hay problemas…

1) La inferencia en la estadística clásica resulta contraintuitiva (para los estudiantes como algunos investigadores de pregrado).
2) Una notable dificultad resulta para el concepto de *p-value*.

p-value: probabilidad de obtener un resultado tan extremo como el que observamos o probabilidad de obtener un resultado tan


H_0
Observed Data
Replicate Data
x
Test Statistic
t(x)
x_1^{rep} x_2^{rep} ... x_m^{rep}
$t(x_1^{rep})$ $t(x_2^{rep})$ ... $t(x_m^{rep})$
$p = Pr[t(x^{rep}|H_0) >= t(x)]$

¿Qué son los parámetros?

La E.C. considera los parámetros poblacionales como fijos, mientras que los datos son los que varían (sampleo repetido).

Diferencia entre estaturas H-M = 15cm.

Asumamos que recolectamos muchos datos de H y M y computamos nuestro Intervalo de confianza de 95% alrededor de nuestra diferencia estimada…

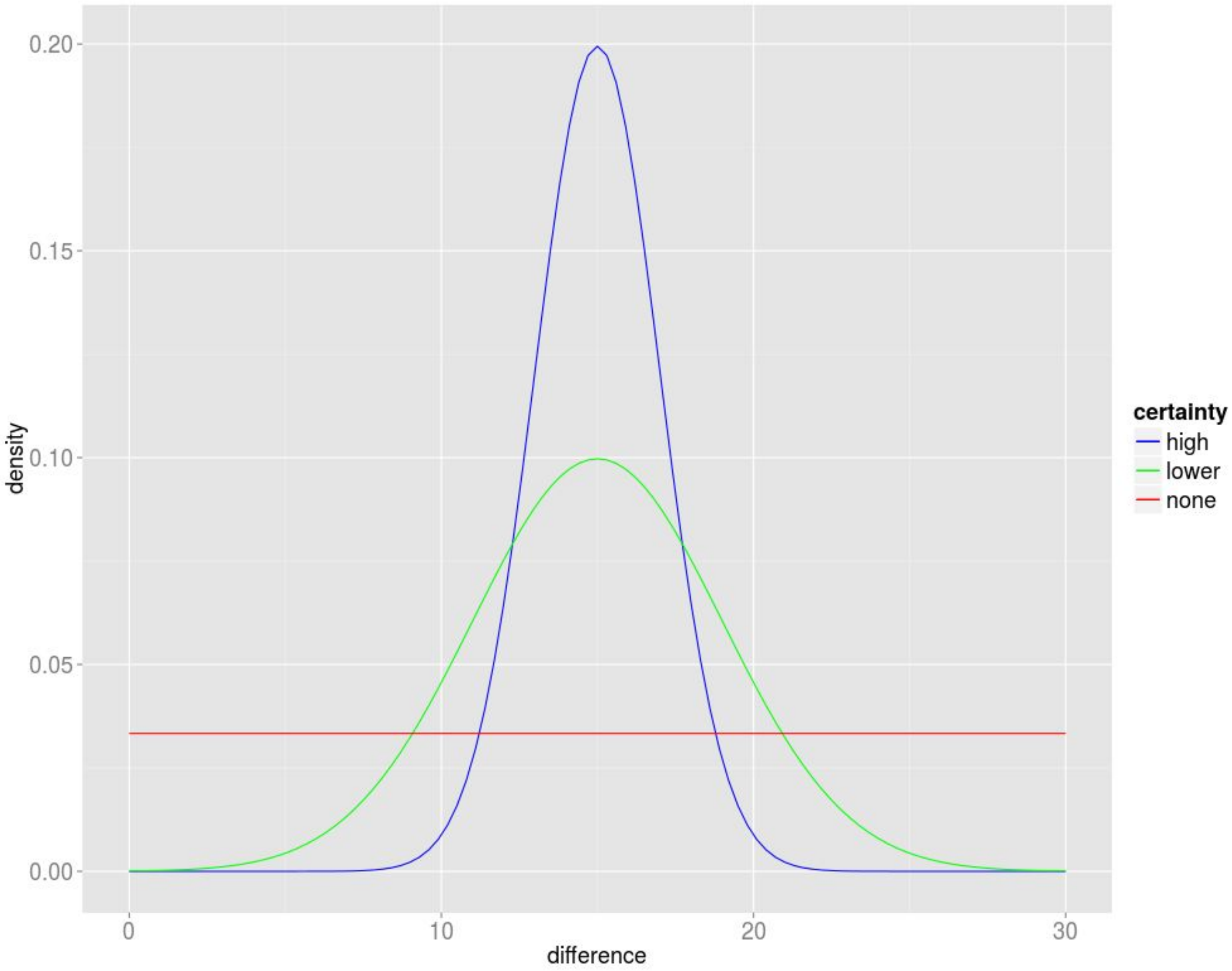
¿Qué significa este intervalo?

Por su parte la estadística bayesiana...

Asume la postura de probabilidad epistémica.

$$\text{difference} \sim \mathcal{N}(15, 16)$$

$$p(\theta|\mathbf{y}) = \frac{p(\theta) \times p(\mathbf{y}|\theta)}{p(\mathbf{y})} \qquad \text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{marginal likelihood}} \qquad p(\theta|\mathbf{y}) \sim p(\theta) \times p(\mathbf{y}|\theta)$$


density
0.20
0.15
0.10
0.05
0.00
0
10
20
30
difference
certainty
high
lower
none

Bayes in the Mind

Postulados principales

- La mente asigna probabilidades a distintas hipótesis y actualiza estas probabilidades de acuerdo a reglas de inferencia probabilística y a la evidencia.

- La mente es capaz de formar modelos ricos, abstractos y verídicos del mundo con solo un número escaso de observaciones poco claras.

Se ha estudiado en:

- Percepción
- Aprendizaje
- Memoria
- Lenguaje
- Toma de decisiones

Ejemplo:

Investigadores interesados en el lenguaje se preguntan

¿Cómo un niño es capaz de inferir las reglas gramaticales -un sistema generativo infinito- a partir de conjuntos de observaciones limitadas y poco claras?

Algunas respuestas a esto son:

Priors contextuales

Àrboles sintácticos para una secuencia de palabras.

Detractores

Bayesian Just-So Stories in Psychology and Neuroscience

Jeffrey S. Bowers
University of Bristol

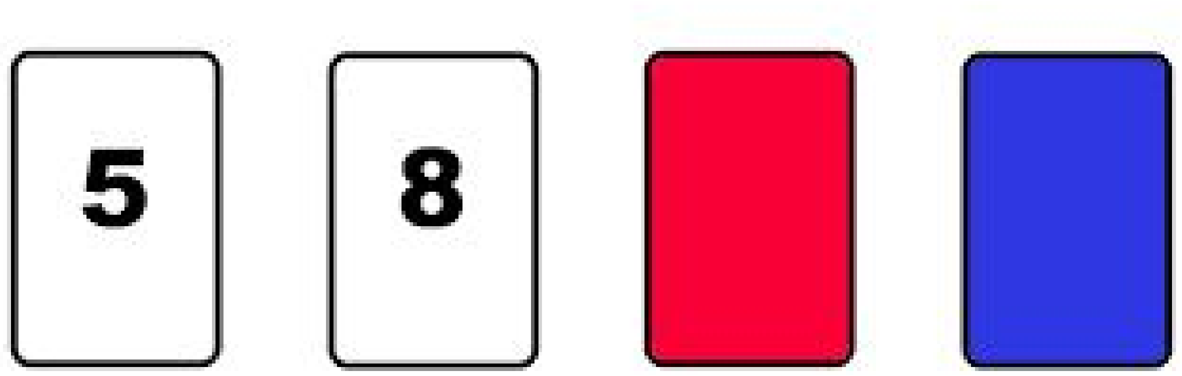
Colin J. Davis
Royal Holloway University of London

Principales críticas

Modelos bayesianos no son susceptibles a la falsación.

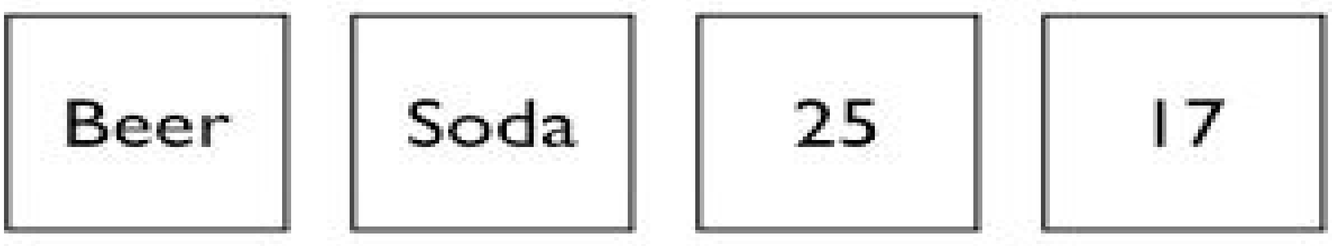
Tratan al humano como un agente racional óptimo.

Si la carta muestra un número par entonces la cara opuesta muestra rojo. ¿Qué cartas debo voltear para comprobar la veracidad de la proposición?



Soy un inspector que va a un restaurante y quiero saber si se cumple la regla de no vender alcohol a menores de edad

¿Qué debo de checar para saber si la regla se cumple?



Optimal Predictions in Everyday Cognition

Thomas L. Griffiths[1] and Joshua B. Tenenbaum[2]

[1]Department of Cognitive and Linguistic Sciences, Brown University, and [2]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology