Apunte 2:

Las probabilidades como una extensión de la lógica

1. Introducción - La estadística frecuentista

Lo que vimos hasta el momento se enmarca dentro de una perpectiva *frecuentista* de la estadística.

El relato es el siguiente:

- Existe una *variable aleatoria* X
  ej: $X \equiv$ resultado de tirar un dado.

- La variable tiene realizaciones concretas x, que son impredecibles.
  ej: x=3 (tiré el dado y salió 3)

- Defino la probabilidad como la *frecuencia* de que aparezca un resultado:

$$P(r) = \lim_{N \to \infty} \frac{\#\{x = r\}}{N}$$

En 1933, el matemático ruso Andrey Kolmogorov publica su libro: ***Fundamentos de la Teoría de Probabilidades*** en el que se presenta un formalismo axiomático. Una versión de los axiomas es:

- La probabilidad es positiva
- Regla de la suma:

$$P(A) + P(\overline{A}) = 1$$

- Regla del producto

$$\begin{aligned} P(AB) &= P(A|B)\ P(B) \\ &= P(B|A)\ P(A) \end{aligned}$$

Esta última relación algunos dicen que es un axioma y otros que es la definición de la probabilidad condicional, ya que $P(A|B)$ se lee "*la probabilidad de A dado que sé que ocurrió B*".

De la regla del producto, igualando las dos versiones se deduce el **Teorema de Bayes** que me dice cómo dar vuelta los condicionales:

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

que no parece tener nada tan raro como para armar mucho lío.

2. Medición de un péndulo: lo que calculo vs lo que interesa

Supongamos que tengo un péndulo. Armo un **modelo platónico** que dice que
en un péndulo

$$T = 2\pi\sqrt{\frac{l}{g}}$$

Según este modelo, si mido el período t me tiene que dar siempre lo mismo ($t = T$). Pero eso no sucede !

Tengo que mejorar mi modelo. Armo un **modelo estadístico** que incluye las molestias del mundo. Digo que en realidad estoy midiendo $t = T + \epsilon$ donde $\epsilon \sim \mathcal{N}(0, \sigma^2)$ es el mundo incidiendo sobre la medición con una distribución normal de promedio cero y de dispersión σ .

Entonces cuando mida, voy a estar midiendo una variable aleatoria t y puedo hacer su histograma (gaussiano él) y hablar de la probabilidad de que una medición tenfa cierto valor: $P(t)$.

Pero a mí, que me gustan tanto las mediciones t lo que me interesa no es la distribución de las mediciones. Mi objetivo es conocer algo del mundo, saber de el período del péndulo T.

Pero en el enfoque frecuentista no puedo ni siquiera hablar de la probabilidad de T, ya que no se trata de una variable aleatoria, sino de un valor platónico que existe antes de que los monos del mundo vengan a molestarme.

De lo que puedo hablar es de $P(t|T)$ y puedo calcular lo que quiera. Pero la pregunta que me intresa es cuál es la probabilidad de que T tenga un valor dadas las mediciones que hice ? Me estoy preguntando por $P(T|t)$, pero de eso no se puede hablar.

Lo que es peor: es una limitación conceptual, ya que el Teorema de Bayes nos explica bien sencillito cómo dar vuelta el condicional.

3. Plausibilidad: probabilidades bayesianas

No fue el primero, pero un lindo antecedente de esta discusión fue la que planteó el matemático húngaro George Pólya en 1954. Pólya como estudiante se maravillaba por la perfección de las demostraciones matemáticas, pero siempre se preguntó acerca de cómo se le pueden ocurrir esas cosas a álguien. Esto está descripto en su libro “Matemáticas y razonamiento plausible” donde desarrolla la idea del razonamiento plausible.

El ejemplo del ladrón es de él, aunque no de ese libro.

> *Supongamos que un policía escucha en la noche que se rompe un vidrio y suena una alarma.*
> *Al aproximarse al lugar ve salir por la vidriera rota de una joyería un hombre con antifaz y las joyas en el bolsillo.*
> *Inmediatamente lo detiene porque interpreta que se trata de un ladrón, pero esa conclusión no se dedice lógicamente de las premisas.*

Efectivamente, los vidrios pueden romperse sin que los rompan a propósito para robar, los antifases no necesariamente los usan los culpables de robo y llevar joyas en el bolsillo no es delito. No siquiera está claro a quién le pertenecen las joyas. Incluso todos esos hechos a la vez tienen una explicación inocente.

Sin embargo, muchos estaríamos dispuestos a pensar que se trata de un robo, porque todo lo que ocurre es tan infrecuente si no se trata de un robo, que si bien no puedo demostrarlo, es *plausible* que así sea.

Inspirado en estas ideas, el físico Richard Cox, en 1961 publicó un teorema que dice que si uno quiere definir una medida de la plausibilidad de una afirmación, con la condición de que

- sea un número entre cero y uno.
- que valga cero si algo es falso
- que valga uno si es verdadero
- quiero que satisfaga algunas hipótesis sencillas y razonables

Entonces

Las reglas de la Plausibilidad de una afirmación satisfacen los axiomas de la probabilidad de Kolmogorov.

> *Eso quiere decir que la probailidad y la plausibilidad son términos sometidos a la mismas reglas matemáticas.*

4. El péndulo revisitado

los datos medidos corresponden a un conjunto de mediciones independientes:

$$D = \{t_i\}_N$$

Entonces puedo calcular (Teorema de Bayes):

$$P(T|D) = \frac{P(D|T)}{P(D)} \ P(T)$$

Los personajes de esta historia son:

- $P(T|D)$: es lo que quiero calcular. Es la distribución plausible de T.
- $P(D|T)$: es lo que sé calcular. Como función de T no es una probabilidad. Por ejemplo no está normalizada. Se la denomina la *likelihood* (me gusta traducirla como la función *verosimilitud*)
- $P(D)$: es la probabilidad de los datos, pero independiéntemente del valor que tome T. Claramente es un valor que no depende de T y por lo tanto es un factor multiplicativo en la cuenta que me interesa. Lo pensamos como un factor de normalización.
- $P(T)$: es la probabilidad de T *sin tener en cuenta los datos medidos*. A esto se lo llama distribución *prior*.

Hay ríos de tinta discutiendo acerca del criterio con el que elegir la función prior. Básicamente hay tres criterios:

- expresar que no sé nada sobre esta variable y proponer una prior uniforme (constante). Es muy cómodo numéricamente y lo vamos a hacer muchas veces. Pero es un poco raro porque si hacemos un cambio de variables, por ejemplo si expresamos todo en términos de la frecuencia en lugar del período, la prior cambia de forma.
- expresar que no se nada de nada de nada sobre la variable. Y para hacerlo objetivamente, defino una entropía y la maximizo.
- lo mejor es reconocer que algo difuso se conoce de la variable en cuestión y lo mejor es poner acá lo que efectivamente se conoce. Aquí se pueden incluir mediciones anteriores.

Lo cierto es que mi medición va a tener mucha más información que la prior (al menos eso es lo que me gustaría sacar de una medición: información). Eso implica que el término $P(D|T)$ está muy localizado. Los datos que medí no dejan que T puedan valer cualquier cosa. Pero la prior es un flan.

De manera que al final del día, si mi medición contiene buena información, la elección de la prior no afecta el resultado y estamos pensando demasiado cosas poco interesantes. A lo sumo, cuando lleguemos a un resultado, siempre podemos chequear la influencia de la prior en el mismo.

Hacemos ahora las cuentas del péndulo:

$$\boxed{P(T|D) \propto P(D|T)\; P(T)}$$

donde proponemos una prior uniforme $P(T) \propto 1$. Resulta:

$$\begin{aligned} P(D|T) &= P(\{t_i\}_N \mid T) \\ &= \prod_{i=1}^{N} P(t_i \mid T) \propto \prod_{i=1}^{N} \exp - \frac{(t_i - T)^2}{2\sigma^2} \\ &\propto \exp - \frac{\sum_{i=1}^{N}(t_i - T)^2}{2\sigma^2} \end{aligned}$$

Pero si desarrollamos la sumatoria del exponente resulta que

$$\sum_{i=1}^{N}(t_i - T)^2 = \sum_{i=1}^{N}(t_i^2 + 2t_iT + T^2) = N(\overline{t_i^2} + 2T\bar{t} + T^2)$$

Aquí vale la pena recordar que todos los términos que no dependan de T son constantes multiplicativas para lo que me interesa, porque están en el exponente.

Por lo tanto puedo tirar el primer término y reemplazarlo por $\bar{t}^2$ sin problemas y entonces resulta que

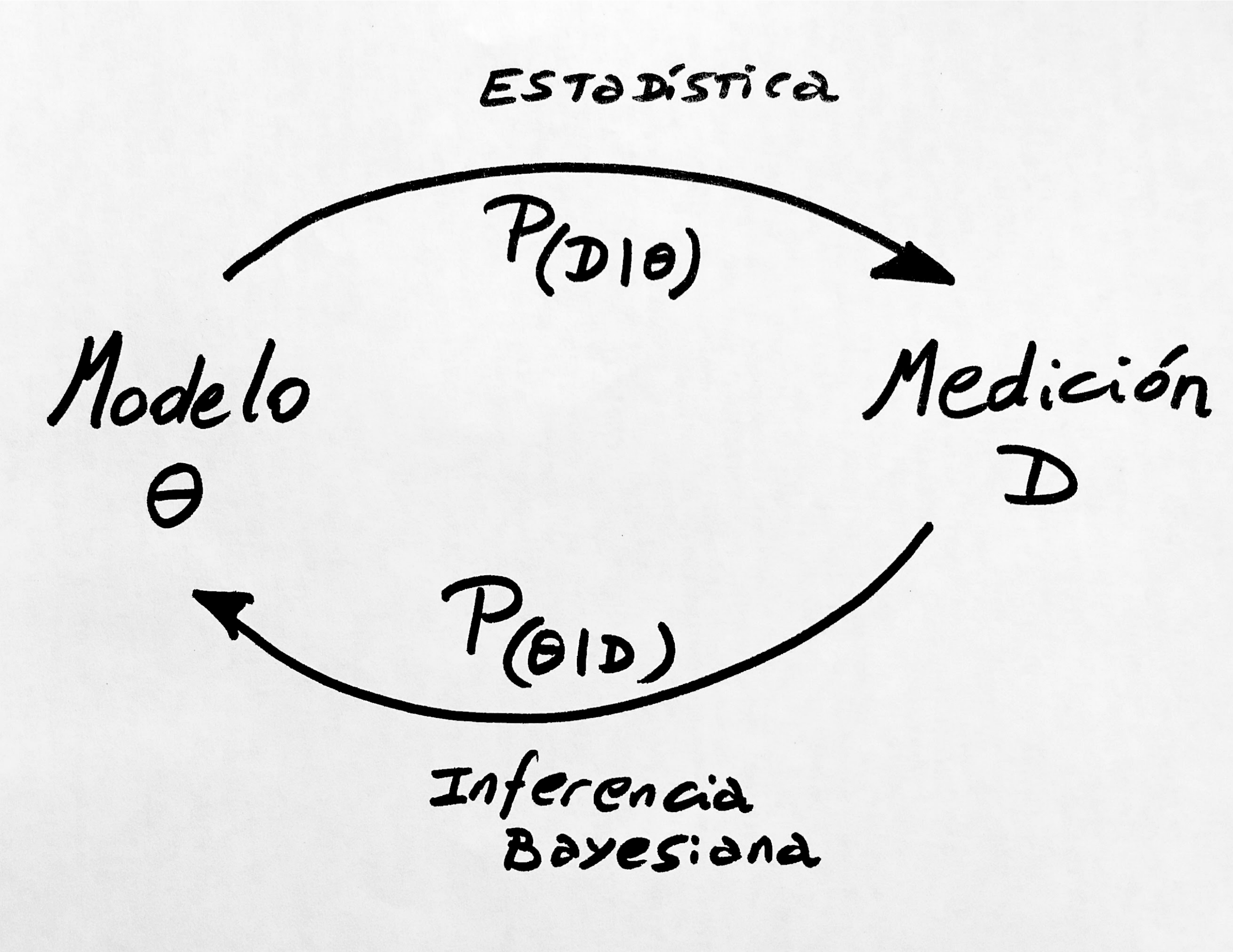
$$P(T|D) \propto \exp - \frac{(\bar{t} - T)^2}{2\sigma^2/N}$$

> *el resultado de medir N veces t es que el período T tiene una distribución gaussiana alrededor del valor medio de las mediciones, con una dispersión que es*
>
> $$\frac{\sigma}{\sqrt{N}}$$

y este es el mismo resultado que habíamos encontrado antes, pero aclarando las hipótesis y que significa lo que estamos diciendo. Antes hablaba de la dispersión del promedio, en el caso de eventuales repeticiones del experimento, pero ahora hablamos de lo que es razonable pensar acerca del valor de T dados los datos medidos.

5. Esquema de razonamiento

La relación entre teoría y experimento queda descripta por el siguiente círculo:



Por un lado tenemos un modelo de la realidad que tiene algunos parámetros θ sobre los que quiero obtener información. Aquí condensamos en θ todos los parámetros que necesita el modelo. Pueden ser muchos parámetros numéricos continuos o discretos. Incluso puede indicar diferentes modelos alternativos.

Con este enfoque es completamente equivalente hacer un test de hipótesis que hacer un ajuste de parámetros: siempre quiero saber qué me dicen los datos acerca de θ que refleja la flexibilidad de mi modelado.

Con esto tenemos toda la matemática que necesitamos para calcular lo que se nos ocurra. No quiere decir que sea fácil, pero que podemos dejar planteados nuestros problemas.

Lo que me interesa que quede como moraleja es que los experimentos nos dan información sobre los parámetros del modelo, pero no de manera contundente:

> *Todo lo que sé acerca de la realidad está expresado en distribuciones*

Salú.

(Fin de la segunda parte de las clases de errores. Willy 2017.)