**AXIOMAS DE PROBABILIDAD** :

La probabilidad está dada por el número de casos de éxito sobre la cantidad total(teórica) de casos.

* P = #-Casos de éxito/ # Casos-totales.
* **Suceso elemental**: Es una única ocurrencia, “Solo tienes una cara de la moneda como resultado”
* **Sucesos**: Son las posibilidades que tenemos en el sistema. Está compuesto de sucesos elementales,  
  por ejemplo, “El resultado de lanzar un dado es par”, hay tres sucesos (2,4,6) que componen este enunciado.

Espacio muestral: todas las posibilidades

* De la interpretación del axioma anterior divergen dos escuelas de pensamiento. **Frecuentista y Bayesiana**
* Ejemplo: “Solo tengo dos posibles resultados al lanzar una moneda, 50% de probabilidad para cada cara  
  , (1/2 y 1/2), si lanzo la moneda n veces, la moneda no cae la mitad de las veces en una cara, y luego la otra”
* Esta equiprobabilidad de ocurrencia en un espacio muestral ocurre bajo el supuesto de que  
  la proporción de exitos/totales tiende a un valor p. En otras palabras, solo lanzando la moneda  
  infinitas veces podemos advertir que el valor de la probabilidad es cercano a (1/2 o 50%).

**Escuela frecuentista**

“Toda variable aleatoria viene descrita por el espacio muestral que contiene todos los posibles sucesos  
de ese problema aleatorio.”

La probabilidad que se asigna como un valor a cada posible suceso tiene varias propiedades por cumplirse

 Veritasium dan una perspectiva sobre el Teorema de Bayes, también explica la diferencia entre los Frequentistas y los Bayesianos: [Cómo Escapar de la Trampa Bayesiana](https://www.youtube.com/watch?v=D7KKlC0LOyw" \t "_blank).

También vi otro video corto de Cassie Kozyrkov sobre estas escuelas: [Are you Bayesian or Frequentist?](https://www.youtube.com/watch?v=GEFxFVESQXc)

Ambos videos me ayudaron mucho a comprender la diferencia y a cambiar un poco la forma en que hago algunas cosas en mi rutina diaria.

**Probabilidad en Machine Learning**

**¿Cuáles son las fuentes de incertidumbre?**

* Datos: Debido a que nuestros instrumentos de medición tienen un margen de error, se presentan datos imperfectos e incompletos,
* Atributos del modelo: Son variables que representan un subconjunto reducido de toda la realidad del problema, estas variables provienen de los datos y por lo tanto presentan cierto grado de incertidumbre.
* Arquitectura del modelo: Un modelo en mates es una representación simplificada de la realidad, al ser una representación simplificada se considera mucho menos información.

**Y claro, todo esta incertidumbre se puede cuantificar con probabilidad:**

**Ejemplo, un clasificador de documento de texto:**

Diagram

Description automatically generated

Entonces, el modelo asignara cierta probabilidad a cada documento y así de determinara la clasificación de los documentos.

Pero, ¿cómo funciona por dentro nuestro modelo de clasificación?

Diagram

Description automatically generated

**So, ¿En dónde se aplica la probabilidad?**

No todos los modelos probabilístico, a la hora de diseñarlo nosotros elegimos si queremos que sea un modelo probabilístico o no.

Por ejemplo si escogemos el modelo de Naive Vayes, luego de que escogemos el diseño ahora definimos el entrenamiento y este es básicamente que el modelo aprenda el concepto de distribución de probabilidad y es una manera que yo uso para saber que probabilidades le asigno una de las posibles ocurrencias de mis datos, de ahí sirgue el esquema MLE que es el estimador de máxima verosimilitud y luego de esto esta la calibración se configuran los hiper-parámetros, esto se entiende mas en redes neuronales artificiales en donde el numero de neuronas de una capa tiene 10 neuronas y cada una tiene sus propios pesos que conectan a las neuronas, entonces esos pesos los podemos ir calibrando para que el modelo sea cada vez mas pequeño. Sin embargo, hay parámetros están fuera del modelo y no se pueden calibrar y a esos parámetros les llamamos hiper-parámetros, porque están fuera de todo ese esquema de optimización. Al final se hace la optimización de los hiper parámetros. Y al final tenemos la interpretación, para interpretar hay veces que se tiene que saber el funcionamiento del modelo y aplicar conceptos de estadística para poder interpretarlo.

Responder

**Los frecuentistas piensan en los parámetros como unos valores fijos pero desconocidos.**

La estimación se basa en la elección de aquellos valores de los parámetros que maximizan la probabilidad de observar los datos.

Por su parte, **los bayesianos interpretan los parámetros como variables aleatorias cuya distribución de probabilidad es estudiada en base al Teorema de Bayes.**La idea es simple, un bayesiano ha de tener una distribución subjetiva de los parámetros antes de ver los datos (a priori) que modificará en función de los datos que haya observado para obtener una distribución a posteriori que resumirá todo el conocimiento del investigador sobre los parámetros de interés dados los datos y sus creencias a priori.

**El Teorema de Bayes es la clave**

El siguiente ejemplo trata de ilustrar el funcionamiento del Teorema de Bayes en la combinación de las creencias subjetivas a priori con los datos para obtener la estimación bayesiana.

Imaginemos un juego de ruleta en el que hay 38 posibles huecos en los que la bola puede acabar con la misma probabilidad (desde el 1 hasta el 36 más el 0 y el 00). Si un jugador apuesta al número 10, yo pienso que la probabilidad de que gane es 1/38 (aproximadamente 2,6%). Esta es **mi creencia a priori, antes de ver ningún dato.** Ahora imaginemos que consigo ver que la bola ha parado y aunque no puedo ver el número en el que ha caído, puedo ver que el hueco es de color rojo. Esta visión puede interpretarse como los datos que observo, la muestra que tengo disponible. Pues bien, dado que hay 18 huecos rojos en la ruleta y el número 10 es uno de ellos, actualizo mis creencias y ahora pienso que la probabilidad de que gane es 1/18 (5.6%). **Habiendo observado los datos,** el hueco rojo en que la bola ha caído, **modifico mis creencias sobre la probabilidad de ganar (o parámetro de interés).**

La idea fundamental del Teorema de Bayes es, por tanto, la modificación de mis creencias una vez que observo los datos. De aquí surge **la principal crítica de los frecuentistas al enfoque bayesiano. Basar el análisis en unas creencias subjetivas del investigador de las que el resultado depende de forma crucial parece poco riguroso.** Sin embargo, cuando realmente hay creencias fuertes y consensuadas sobre determinados parámetros, como por ejemplo que el coeficiente del precio en una función de demanda es negativo, ¿por qué no hacerlas explícitas y transparentes a través del análisis bayesiano? Por otro lado, sabemos que si tenemos una muestra suficientemente grande (aquí, como siempre, no está claro qué es una muestra suficientemente grande porque dependerá de la complejidad del problema analizado) la creencia a priori del investigador se ve dominada por los datos y su influencia en el resultado final disminuye hasta hacerse inocua con una muestra de infinitas observaciones. Finalmente, sabemos que **hay muchos resultados frecuentistas que pueden obtenerse desde una perspectiva bayesiana aunque la filosofía subyacente sea diferente.** Mínimos cuadrados ordinarios (MCO) es el ejemplo más sencillo. Dicho estimador frecuentista coincide exactamente con la media de la distribución bayesiana bajo unas creencias concretas en el marco del modelo lineal normal. Existen en la literatura muchos otros ejemplos.

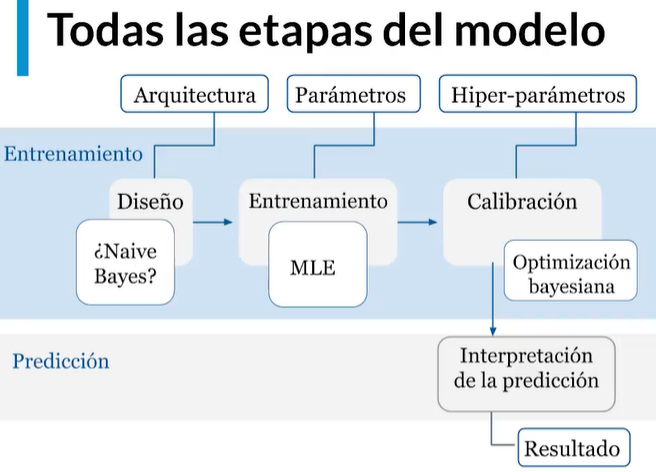
**Contrastando hipótesis**

Teniendo en cuenta la conexión entre ambos enfoques en la práctica y la posibilidad de obtener los mismos resultados bajo ambos enfoques, **los bayesianos argumentan que su interpretación del problema es siempre más intuitiva y natural.** Recordemos que

Un bayesiano: “hay un 95% de probabilidad de que el parámetro esté entre 0,3 y 0,8”,

Un frecuentista “si generáramos 100 muestras aleatorias del mismo tamaño y repitiéramos la estimación 100 veces, en 95 de ellas el parámetro estimado estaría entre 0,3 y 0,8”.

Los contrastes de hipótesis parecen por tanto más naturales en el marco bayesiano. Para un bayesiano convencido, un p-valor o un t-ratio es una herejía, ya que lo único que necesitamos para contrastar hipótesis es tener la distribución a posteriori de los parámetros. Salvo en contadas excepciones, **un frecuentista basa su contraste de hipótesis en el análisis asintótico, es decir, en calcular p-valores de la distribución asintótica del estimador (no del parámetro verdadero, que es un valor fijo). Esta distribución, mayoritariamente normal gracias a numerosos teoremas centrales del límite, es la que el estimador tendría si el investigador tuviera muchos más datos de los que realmente tiene… ¿es ésta una buena forma de hacer inferencia en general? Por el contrario, el bayesiano se basa en la distribución de los parámetros dados sus datos, que puede en principio tener cualquier forma no g**



MLE Maximum Likelihood