



Instituto Politécnico Nacional
Escuela Superior de Física y Matemáticas
Introducción a los modelos bayesianos usando PyMC



Introduction:

A lo largo de este documento se realiza una abstracción y resumen de un panel presentado en la PyCon Chile 2021, PyCon es un evento que tiene como objetivo reunir a desarrolladores de Python, desarrolladores de software, entre otros tipos. Para que puedan hablar sobre las nuevas mejoras en el ecosistema de desarrollo de Python.

Este panel en específico llevaba por título el mismo título de este resumen “Introducción a los modelos bayesianos usando PyMC”. Y fue presentado por Fernando Iravarrazal, economista y estadístico egresado de la universidad de Harvard, donde el mismo nos aclara que se mostrará la utilidad de modelos bayesianos, donde definiremos algunos conceptos que son fundamentales en la estadística bayesiana, para luego llegar a un ejemplo donde se hará uso de la librería PyMC. Esta última es una librería de Python diseñada para la inferencia estadística bayesiana mediante técnicas de muestreo y bastante utilizada para construir modelos estadísticos complejos y realizar inferencias probabilísticas.

Abstract:

De manera casi inmediata se abordaremos la pregunta “¿Qué es la estadística bayesiana?”, se nos menciona que a diferencia de la estadística frecuentista que es aquella que se suele enseñar en la mayoría de los programas académicos, la estadística bayesiana asume que los parámetros de los modelos son aleatorios y las inferencias se actualizan con información previa y nuevos datos.

La estadística bayesiana de igual manera se encuentra basada en el teorema de Bayes

Teorema de Bayes

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

La forma en que la estadística bayesiana interpreta este teorema es de la siguiente forma:

$$P(\text{Parametro} | \text{Datos}) \propto P(\text{Datos} | \text{Parametro})P(\text{Parametro})$$

Del lado izquierdo de la expresión tenemos lo que se desea calcular ¿Cuál es nuestro conocimiento del parámetro dado los datos que hemos visto? A esta parte la denominaremos ‘Probabilidad posterior’.

Instituto Politécnico Nacional

Escuela Superior de Física y Matemáticas

Introducción a los modelos bayesianos usando PyMC

Del lado derecho de la expresión tenemos el producto de lo que sería el Likelihood que en palabras nos respondería ¿Cuál es la probabilidad de ver los datos que vimos en distintos casos hipotéticos? con algo que llamaremos Priori que será el conocimiento que nosotros tengamos del parámetro de manera intuitiva, sin haber tenido ningún tipo de dato, por lo general se suele suponer que los casos son igualmente probables a priori, esto cambia una vez que se han introducido datos y se ha actualizado la información de los parámetros.

A continuación vamos a abordar un ejemplo con el que se mostrara como se procedió en un caso práctico en Santiago, Chile en mayo del 2020, para fines pedagógicos y con el objetivo de honrar solo en el aspecto estadístico se ignoraran algunas propiedades de la epidemia que dio lugar el COVID.

Caso de estudio:

Es mayo del 2020, acaba de comenzar la pandemia, y el gobierno quiere saber cuánta gente se ha contagiado de COVID en Santiago, Chile. Dependiendo de este número el

gobierno seguirá (o no) una estrategia de inmunidad de rebaño. Para esto usaremos tests que detectan anticuerpos SARS-CoV-2.

Primero se asume que podemos aleatorizar a todo Santiago y realizamos una encuesta.

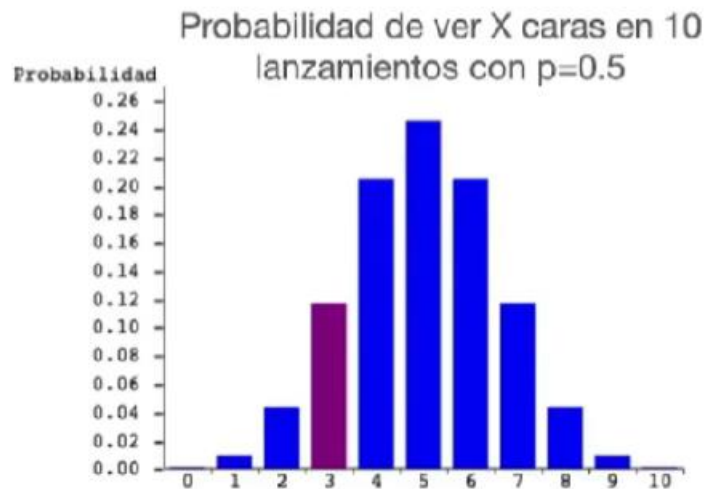
El panelista, Fernando abre un tipo de paréntesis en la presentación para hacer un pequeño recordatorio sobre como funciona la distribución binomial.

Distribución Binomial:

Hay N experimentos que resultan en un éxito (un 1) o un fracaso (un 0) con probabilidad p .

Por ejemplo si tuviéramos una moneda justa y tiramos la moneda 10 veces, el modelo binomial nos dice lo siguiente

Instituto Politécnico Nacional
Escuela Superior de Física y Matemáticas
Introducción a los modelos bayesianos usando PyMC



En palabras: “Describe cuales son las probabilidades de los escenarios posibles”

El anterior recordatorio de la distribución binomial fue ya que la usaremos para modelar los datos de los tests.

Tomaremos una muestra aleatoria de 50 personas en Santiago.

De los 50 tests:

- 40 son negativos (No se detectan anticuerpos)
- 10 son positivos (Se detectan anticuerpos)

En base a estos datos vamos a construir tres modelos.

Modelo 1: Asume que el test es perfecto

Modelo 2: Incluiremos que el test a veces da falsos positivos

Modelo 3: Incluiremos la incertidumbre sobre la tasa de falsos positivos

Para todos los modelos la definición de un “experimento” es la siguiente:

Éxito = Test positivo; Fracaso = Test negativo

Fracaso tiene probabilidad p de ocurrir.

Instituto Politécnico Nacional

Escuela Superior de Física y Matemáticas

Introducción a los modelos bayesianos usando PyMC

Ahora pasamos al código:

Primero importamos PyMC y Arviz

```
import pymc3 as pm
import arviz as az
```

Definimos nuestros datos

```
tests_totales = 50
tests_positivos = 10
```

Definimos el modelo en este caso donde se cree que el test es perfecto

```
with pm.Model() as modelo_test_perfecto:
    prob = pm.Uniform(name='prob',
                      lower=0,
                      upper=1)
    casos_positivos = pm.Binomial(name='casos_positivos',
                                  p=prob,
                                  n=tests_totales,
                                  observed=tests_positivos)
    trace_test_perfecto = pm.sample(3000)
```

Lo que nosotros hicimos con este modelo es darle una probabilidad Priori (conocimiento a priori sobre el parámetro) luego le introducimos la distribución de los datos para obtener la 'Probabilidad posterior'

Nosotros lo podemos graficar con un histograma la distribución de las muestras que nos arroja nuestro modelo.

Ahora supongamos que la ministra de salud nos pregunta

¿Cuál es la probabilidad de que menos del 15% de la población se haya infectado?

Sencillamente tomamos aquellas muestras donde se cumple esto y dividimos entre el número total de muestras para obtener la respuesta.