Métodos Estadísticos Bayesianos con R

Stats 101

Gibrán Peniche

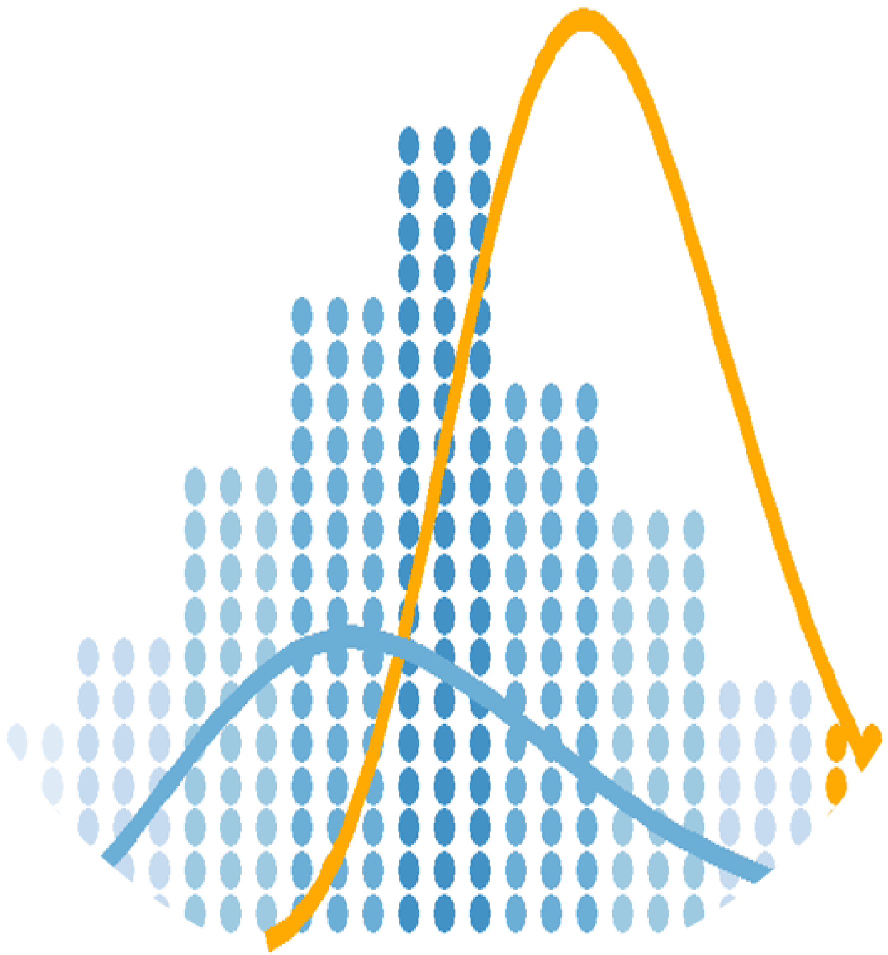
v. 0.0.1

2020-06-09

jgpeniche

PenicheGibran

jgpeniche@gmail.com



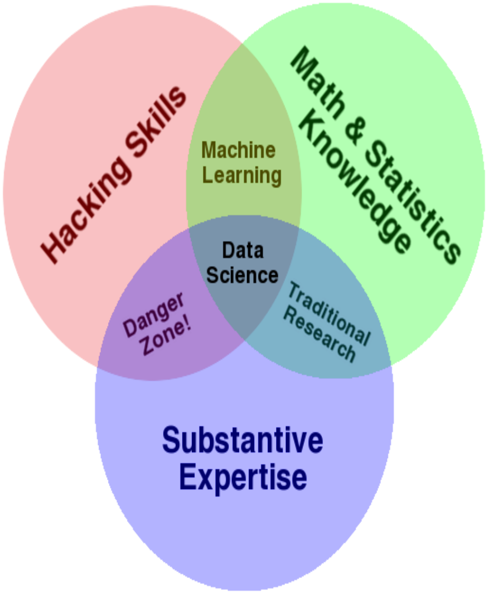
¿Porqué Estadística?

¿Porqué Estadística?

El problema fundamental en cualquier profesión es:

> ¿Cómo tomar decisiones **hoy** sujetas a la *incertudumbre* de **mañana**?

Ademas según Drew Conway



La Ciencia de Datos está de moda...

¿Qué es la estadística?

En palabras de Manuel Mendoza...

"Un conjunto de **técnicas** para describir **fenómenos** que se manifiestan a través de **datos** que presentan **aleatoriedad**"

Un conjunto de técnicas...

Entendemos que describimos estos fenómenos a través de **resúmenes** que deben ser **suficientes** y **minimales**

Pero... ¿Para que "hacemos estadística"?

El fin último de la inferencia es **proveer** de información **confiable** para la **toma de decisiones**

En otras palabras las inferencias deben ser **ÓPTIMAS**

Pero... ¿Qué es la inferencia (estadística)?

Inferencia Estadística

Si bien el problema de la inferencia sigue siendo el mismo, puede ocurrir que no contemos con toda la información (datos) sobre la manifestación del fenómeno que nos interesa estudiar

Si este es el caso, decimos que estamos ante una **muestra** del fenómeno y nuestra información es **incompleta** por lo que solo podemos aspirar a realizar descripciones **aproximadas** del fenómeno

El problema de la inferencia es de natureleza *inductiva*, ya que a partir de la información que aporta la muestra buscamos realizar una *conjetura* o *juicio* de carácter **general** sobre el fenómeno de interés

Inferencia Estadística (Paramétrica)

Cuando, adicional a las consideraciones anteriores, decidimos representar la *incertidumbre* del fenómeno a través de la forma funcional de una **familia de distribuciones de probabilidad** caracterizadas por cierto conjunto de parámetros, esto es:

$$f \in \mathcal{F} = \{ f_\theta \, | \theta \in \Theta \}$$

El problema de inferencia se reduce al conjunto de técnicas que permiten encontrar el parámetro θ en el espacio parametral Θ que mejor describe la incertidumbre del fenómeno

A esto se le conoce como **Inferencia Estadística Paramétrica**

Inferencia Estadística Paramétrica

A lo largo de la historia, particularmente durante el siglo XX, se confrontaron dos escuelas de inferencia:

- La escuela de la Estadística Matemática (o escuela frecuentista)
- La escuela de la Teoría de Inferencia Bayesiana

Si bien la primera, cuyos principales exponentes son Ronald A. Fisher, Karl Pearson y Jerzy Neyman, sigue siendo la escuela dominante debido a lo sencillo del los procedimientos, exhibe varias patologías que exploraremos a continuación

La segunda, cuyos máximos exponentes son Leonard J. Savage, Dennis Lindley, Harold Jeffreys y Alan E. Gelfand, propone una teoría formal de inferencia que estudiaremos a lo largo del curso

Algunos problemas con la Estadística Matemática...

1

¿Qué es la probabilidad para la escuela de Fisher y Neyman?

$\rightarrow$ Es el **límite** de **frecuencias relativas** al repetir un experimento en las **mismas condiciones** una infinidad de veces

$$P(A) = \lim_{N\to\infty} \frac{a \in A}{N}$$

$\rightarrow$ Esta definición proviene del **positivismo** de Auguste Comte

$\rightarrow$ En esta corriente se entiende a la **experiencia** cómo la **única** fuente de conociento.

Pregunta:

¿Cuál es la probabilidad de que Benito Juárez haya nacido en Jueves?

¿Acaso Benito Juárez puede nacer a veces en Jueves, otras en Domingo y algunas otras en Lunes?

Estadística Matemática

El proceso de inferencia frecuentista, descansa en una serie de criterios **matemáticos** deseables a través de los cuales se caracteriza la **optimlaidad** de una inferencia (de ahí el nombre) entre los cuales se encuentran:

1. Insesgamiento
2. Consistencia
3. Eficiencia
4. Mínima Varianza

Sin embargo, el proceso de inferencia **NO** está unificado, si bien el método de *Máxima Verosimilitud* es la principal herramienta de estimación puntual, la mayoría de los casos complejos (y no tan complejos) requieren de soluciones *ad-hoc* y en todos los casos existe un *trade-off* de estas características deseables. Este tipo de problemas siguen apareciendo en otras ramas de la inferencia como lo es el *contraste de hipótesis*

¿No sería deseable tener un solo procedimiento de inferencia y que siempre cumpliera con las características deseables?

2

$X \sim Bernoulli(\theta)$

Problema:

Estimar por regiones, usando el método pivotal, a θ suponiendo que tiene una muestra de tamaño n donde **todas** las observaciones fueron exitos...

Sabemos que $\overline{X} = \hat{\theta} = 1$

$\rightarrow$ ¿Concluimos que **no** hay incertidumbre sobre nuestro parámetro?

$\rightarrow$ ¿Cómo incorporamos al procedimiento de inferencia que nosotros sabemos que el valor real del parámetro debe estar alrededor del 75%?

$\rightarrow$ ¿Disminuimos $\overline{X}$ a pesar de que en los datos no hay información sobre este hecho? ¿Y la incertidumbre? ¿Cómo la incorporamos?

$X \sim Bernoulli(\theta)$

Además, por el método pivotal...

$$\frac{X - \hat{\theta}}{\sqrt{\hat{\theta} \cdot (1 - \hat{\theta})}} \sim ?$$

$\rightarrow$ Sabemos que esto definitivamente **NO** es normal

$\rightarrow$ Si hacemos trampa, suponemos que es normal y resolviendo la ecuasión de segundo grado para θ vamos a obtener que $\theta \epsilon [1 \pm Z_{(1-\frac{\alpha}{2})} \sqrt{\frac{1}{n^4}}]$

$\rightarrow$ ¿Truncamos el intervalo en 1? (De hecho si $n \rightarrow \infty$ se colapsa en 1)

$\rightarrow$ Este resultado persiste para otras aproximaciones al intervalo

3

Teorema de Fieller para cociente de medias

Problema:

Estimar por regiones el cociente de medias para dos muestras con distribución normal...

Problema más grande

Se puede probar que

$$\exists(1-\alpha) < 1.,.\rho = \frac{\mu_1}{\mu_2} \epsilon \mathbb{R}$$

$\rightarrow$ Tenemos un intervalo de confianza **FIJO** al nivel $(1-\alpha)\cdot$ 100% < 1 que abarca todo el espacio parametral

Más problemas

· Existen 2 estimadores **NO** Máximo Verosímiles que funcionan mejor en la práctica estimar la mediana la distribución *lognormal* que el estimador de MV

· Se puede probar que hay estimadores **insesgados** que estiman con probabilidad 0 eventos que en la distribución teórica tienen probabilidad distinta de 0

· La estimación de un fenómeno **aleatorio** en regresión líneal se reduce. o es equivalente, a un problema **geométrico** en el método tradicional de mínimos cuadrados, que además por el *Teorema de Gauss* los estimadores que obtenemos son únicos y además son insesgados

· En *teoría de valores extremos*, la estimación de los parametros se hace generalmente con estádisticas que no necesariamente son insesgadas o Máximo Verosímiles

En conclusión

La escuela frecuentista de Fisher - Neyman - Pearson nos da una serie de **criterios de optimalidad** (insesgamiento, consistencia en ECM, etc)

Estos criterios los entendemos en términos del **problema matemático** (Por eso se le conoce como **Estadística Matemática**) pero no en el contexto del *¿para qué?* de la inferencia

En conclusión

Además, estos criterios **dependen del problema de inferencia** que estemos tratando

No tenemos un **procedimiento general** para hacer inferencia, ni un criterio general de optimalidad para nuestras inferencias

Los padres de la estadística clásica **no lograron** establecer un cuerpo axiomático para consolidar a la estadística cómo una teoría en el primer cuarto del siglo XX

De hecho la Estadística Matemática tal como la concoemos es un collage de las ideas de varios pensadores y siempre se vió minada por grandes discusiones en su ceno que evitaron llegar al concenso

El problema de toma de decisiones en ambiente de incertidumbre

Lo que generalmente nos interesa como profesionistas (economistas, actuarios, financieros, politologos, etc) es tener un **número** con el que podamos decidir nuestro siguiente paso y nos gustaría tener algún grado de certidumbre sobre ese número

¿Cuál es la manera óptima de tomar decisiones en ambiente de incertidumbre?

¿Existe algún criterio que nos asegure, sin importar las circunstancias, que nuestra inferencia es óptima?

La Estadística Bayesiana nos provee de un procedimiento de inferencia ÓPTIMO partiendo del problema de la toma de decisiones ante un ambiente de incertidumbre

(Que si lo piensan bien, en realidad es nuestro día a día)

PREGUNTA: ¿Qué es la Estadística Bayesiana?

R: Una TEORÍA DE INFERENCIA

Esatídistica Bayesiana

Proviene de la corriente filosófica del **idealismo trascendetal** de Immanuel Kant en el siglo XIX

Esta corriente (a grandes rasgos) propone conciliar a la razón y a la experiencia como fuentes de conocimiento

Partiendo de lo anterior la Teoría Bayesiana entiende la probabilidad como una **medida de la ignorancia de uno mismo** con respecto a algún fenómeno, razón por la cual históricamente se la ha conocido cómo *probabilidad subjetiva*

La necesidad de una teoría estadística

Un indicador unívoco de la madurez de una **técnica** es cuando esta logra consolidarse cómo una teoría a través de la construcción un cuerpo axiomático

En mi opinión la estadística es la herramienta por excelencia de la ciencia, ya que a través de ella se enfrenta la abstracción con el mundo real

Al final, muchas decisiones (políticas públicas, medidas de control de medicamentos, regulación bancaria/financiera) descansan en tener cierto grado de certidumbre y de ellas dependen a veces millones de dolares/miles de empleos/ la salud de cierta población, etc...

La estadística es la rama verdaderamente práctica de las matemáticas

¿Qué NO es la Estadística Bayesiana?

- **NO** es tratar a los parámetros como variables aleatorias (El parámetro es **SIEMPRE FIJO** y **DESCONOCIDO**)
- **NO** es `model.bayesian.fit`, hay que conocer la teoría detrás del procedimiento de inferencia (esto también es por ética profesional)
- **NO** es *MCMC go prrrrrr*, hay que conocer sobre la teoría que está detrás de los métodos númericos (También aplica para la ciencia en general)

Breviario Histórico de los Bayesianos

Un poco de historia...

"Bayesianos" era un término que Fisher utilizaba para referirse despectivamente a los partidiarios de esta escuela

Aunque hoy en día el teorema a partir del cual se contruye el procedimiento de inferencia que vamos a estudiar lleva por nombre el apellido del reverendo **Thomas Bayes**, fue en realidad **Pierre-Simon Laplace** quien desarrolló la expresión que hoy utlizamos en los cursos de cálculo de probabilidades y quien fué el verdadero precursor de esta corriente

Aunque hoy en día la estadística bayesiana se ha vuelto popular debido a los avances en materia de poder de computo y consecuentemente a la popularización de los modelos de Machine Learning e Intelgencia artificial, lo cierto es que siempre ha estado allí...

Un poco de historia...

- Fue con métodos bayesianos que se descifraron los códigos ENIGMA alemanes (WWII)
- Se uso ampliamente en la milicia en los manuales de artillería Ingleses y Americanos para apuntar cañones
- Se utilizó para encontrar bombas nucleares perdidas en la guerra fría, para trazar trayectorias de balística contra submarinos rusos y para diseñar rutas de búsqueda de naufragos por la marina estadounidense
- Ha sido el común denominador entre la ciencias actuariales desde los tiempos de Edmun Halley
- La BBC usó durante 20 años métodos bayesianos para estudiar la probabilidad de éxito de los condidatos a la presidencia de Estados Unidos
- Personajes como Robert Schailfer desarrollaron manuales de negocios en Harvard con principios Bayesianos de Teoría de Decisión

Obviamente, toda esta información era secreta...

...mientras tanto Fisher y anexos se encargaron de desacreditar sistemáticamente esta corriente por muchos, muchos años además de las (obvias) limitaciones computacionales

¿Qué sigue?

En las siguientes sesiones vamos a aprender:

1. La definición de problema de decisión
2. Los axiomas de coherencia
3. Costruiremos a partir de los axiomas el criterio óptimo para la toma de decisiones
4. La inferencia cómo problema de decisión
5. Estimadores de Pérdida Mínima, estimación puntual y por regiones (Modelos Conjugados)
6. Pruebas de hipótesis
7. Pronóstico
8. Estadística Bayesiana Computacional