BOURBAKI

COLEGIO DE MATEMÁTICAS

BOURBAKI
COLEGIO DE MATEMÁTICAS

Índice

01 Introducción

Bienvenidos a nuestro curso de Matemáticas Avanzadas para la Ciencia de Datos, nuestro curso tiene tres módulos dedicados a estudiar las ideas matemáticas más útiles para comprender los algoritmos y modelos matemáticos más comúnes en Ciencia de Datos. Los tres módulos son los siguientes

- Probabilidad y Estadística
- Álgebra Lineal
- Optimización y cálculo diferencial

Todos los módulos tienen una duración de 8 semanas. El curso está acompañado de ejercicios y tareas en Python para practicar y reforzar los conocimientos aprendidos así como las implementaciones en bases de datos de los algoritmos estudiados. Pueden consultar el repositorio de esta semana en este link.

La estructura de cada una de las semanas es la siguiente:

1. Treinta minutos dedicados a estudiar un artículo de referencia que motivará los conceptos matemáticos de esta semana.
2. Dos horas dedicadas a estudiar el tema de la semana y algunos ejercicios.

3. Una hora y media dedicada a practicar lo aprendido utilizando Python.

El primer módulo de probabilidad consta de los siguientes temas:

1. Kolmogorov, independencia y condicionamiento
2. Variables aleatorias discretas y sus momentos
3. Ley de los grandes números y máxima verosimilitud
4. Variables aleatorias continuas, el teorema límite central
5. Intervalos de confianza y tests estadísticos
6. Inferencia bayesiana
7. Cadenas de markov y muestreos de gibbs
8. Redes bayesianas

02 Lectura de referencia: las distribuciones de Laplace

La distribución gaussiana es ampliamente conocida y utilizada para modelar distintos fenómenos tanto en Ciencia de Datos como en muchas otras áreas del conocimiento.

En la semana 4 del curso hablaremos sobre un teorema que es llamado el Teorema Límite Central y que justifica su importancia. Por el momento solo haremos referencia a la idea intutiva que todos tenemos cuando escuchamos ”campana de gauss”.

Además del matemático alemán F. Gauss a quien le debe su nombre la distribución gaussiana, en su descubrimiento participó el matemático francés Laplace quien descubrió otra distribución de probabilidad hoy conocida como distribución de Laplace. Esta distribución tiene la importante diferencia con la gaussiana que la cantidad de outliers es mucho mayor.

Les compartimos este texto de referencia [1] sobre la presencia de las distribuciones de Laplace en numerosos fenómenos como la ingeniería, las finanzas, la teoría del control etc.

03 Las variables aleatorias y sus momentos

La teoría moderna de probabilidades se basa en el manejo de las variables aleatorias, trataremos de convencer al estudiante de sus ventajas con algunos ejemplos.

Antes de dar la definición formal de una variable aleatoria vamos a dar la definición de una aproximación entre variables aleatorias:

Definition 00.1. Una varibale aleatoria X es un fenómeno que es aproximado por un conjunto con N observaciones numéricas generadas por el mismo fenómeno $\{x_1, x_2, \ldots, x_N\}$.

Exercise 00.1. *Al elegir a N personas al azar y registrar sus estaturas* $\{x_1, x_2, \ldots, x_N\}$ *estamos aproximando una variable aleatoria que corresponde a la población total.*

Formalmente la definición de variable aleatoria es la siguiente:

Definition 00.2. Sea $(\Omega, \mathbb{P})$ una ley de probabilidad y $\mathbb{R}$ el conjunto de los números. Una variable aleatoria es una función $X : \Omega \to \mathbb{R}$.

Remark 00.2. *La enorme ventaja de las variables aleatorias respecto a los espacios de probabilidad es que podemos operar con ellas, es decir sumarlas, restarlas, etc...*

Exercise 00.3. *Si* $X : \Omega \to \mathbb{R}$ *es una variable aleatoria, demostrar que es posible* *definir una ley de probabilidad sobre* $\mathbb{R}$ *de la siguiente forma: para cada* $A \subseteq \mathbb{R}$ *definimos* $\mathbb{P}_X(A) = \mathbb{P}\left(X^{-1}(A)\right)$.

Exercise 00.4. *Si* $\Omega = \{1,\dots,6\}^2$ *es el espacio de probabilidad correspondiente a la experiencia aleatoria de lanzar dos dados de manera independiente, definimos la siguiente variable aleatoria:* $X_{sum}\left(i,j\right) = i + j$. *Graficar la función* $F:\mathbb{R}\to\mathbb{R}$ *definida por* $F(x) = \mathbb{P}_{X_{sum}}(\{x\})$.

Exercise 00.5. *Sea* $\left(\{0,1\}^n, \mathbb{P}_{Bernoulli,n,p}\right)$ *el producto de las leyes de probabilidad de Bernoulli. Para un* $i \leq n$ *y cada* $\omega = (\omega_1,\dots,\omega_n) \in \{0,1\}^n$ *definimos* $X_i(\omega) = \omega_i$. *Calcular* $\mathbb{P}_{X_k}$.

Esperanza, varianza y otros momentos de las variables aleatorias

---

Definition 01.1. *Si* $X:\Omega\to\mathbb{R}$ *es una variable aleatoria con y* $\mathbb{P}$ *una ley de probabilidad sobre el conjunto* Ω *determinada por* $p_i = \mathbb{P}(\omega_i)$, *definimos la esperanza de* X *de la siguiente forma:*

$$\mathbb{E}[X] = \sum_{\omega_i\in\Omega} p_i \cdot X(\omega_i)$$

Remark 01.2. *Si* X *es una variable aleatoria denotaremos por* $Im(X)$ *al conjunto de valores* $x \in \mathbb{R}$ *tales que existe algún* $\omega \in \Omega$ *con* $X(\omega) = x$ *y por* $p_x = \mathbb{P}_X(\{x\})$.

Proposition 01.3. *Si X es una variable aleatoria, entonces*

$$\mathbb{E}(X) = \sum_{x \in Im(X)} x \cdot p_x.$$

Demostración.

$$\mathbb{E}(X) = \sum_{\omega_i \in \Omega} p_i \cdot X(\omega_i) = \sum_{x \in Im(X)} \left(\sum_{\omega_i : X(\omega_i) = x} p_i \cdot x \right) = \sum_{x \in Im(X)} \mathbb{P}(\{\omega : X(\omega_i) = x\}) \sum_{x \in Im(X)} x \cdot p_x$$

□

Example 01.4. *Sea $X : \{0,1\} \to \mathbb{R}$ la función identidad vista como una variable aleatoria sobre el espacio de probabilidad de Bernoulli $(\{0,1\}, \mathbb{P})$, entonces $\mathbb{E}[X] = (0 \cdot (1-p)) + (1 \cdot p) = p$.*

Exercise 01.5. *Definir dos variables aleatorias que corresponda a las 10 calificaciones durante un semestre de dos alumnos, todas ellas entre 5 y 10 y calcular sus esperanzas.*

Exercise 01.6. *Calcular la esperanza de X_{sum} del ejercicio 00.4.*

Una propiedad muy importante de la esperanza es que es una funcional lineal en el siguiente sentido:

Proposition 01.7. *Si X, Y son dos variables aleatorias y $a, b \in \mathbb{R}$, entonces $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$.*

Proposition 01.8. *Sea $(\{0,1\}^n, \mathbb{P}_{Bernoulli,n,p})$ y S_n la variable aleatoria definida en la proposición* **??**. *Entonces $\mathbb{E}[S_n] = np$.*

Example 01.9. *En el ejemplo* **??***, la esperanza es* $307 \times 0.005 = 1.535$ *que es lo que se esperaría en promedio de medallas ganadas por Argentina de acuerdo al tamaño de su población.*

Definition 01.1. Sean $X, Y : \Omega \to \mathbb{R}$ dos variables aleatorias

- Definimos la varianza de X como

$$Var(X) = \sum_{x \in Im(X)} (x - \mathbb{E}(X))^2 \cdot p_x$$

- Definimos la desviación estándar de una variable aleatoria como la raíz positiva de $Var(X)$, se denotará por σ_X.

- Definimos la covarianza entre X y Y como

$$Cov(X, Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y)$$

- Si X, Y son dos variables aleatorias, definimos su correlación como $Corr(X, Y) = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y}$

Remark 01.10. $Cov(X, X) = Var(X)$.

Example 01.11. *Sea* $X : \{0, 1\} \to \mathbb{R}$ *la variable aleatoria definida en el ejemplo 01.4 entonces* $Var(X) = (0 - p)^2 \cdot (1 - p) + (1 - p)^2 \cdot p = p \cdot (1 - p)$.

Exercise 01.12.

- *Calcular la varianza de la variable aleatoria definida en el ejercicio 01.5.*

- *Calcular la varianza de la variable* X_{sum} *definida en el ejercicio 00.4.*

Exercise 01.13. *Si X, Y son variables aleatorias independientes entonces $Cov(X, Y) = 0$.*

La importancia de la correlación es que nos permite medir qué tan cerca de estar en regresión están dos variables.

Proposition 01.14. *Entre más cercano a -1 o +1 esté la correlación entre dos variables $S = \{x_1, \ldots, x_N\}$, $S' = \{y_1, \ldots, y_N\}$ más cercana de estar en regresión lineal simple estará la base de datos $\{(x_1, y_1), \ldots, (x_N, y_N)\}$. De hecho*

$$\beta_0 = \frac{Cov(S, S')}{Var(S)}$$

$$\beta_1 = \mu_{S'} - \beta_0 \mu_S$$

La kurtosis y Laplace

Definition 02.1. Sea $S = \{x_1, x_2, \ldots, x_N\}$, definimos la kurtosis de S

$$\kappa_S = \frac{\frac{1}{N}\sum_{i \leq N}(x_i - \mu_S)^4}{\left(\frac{1}{N}\sum_{i \leq N}(x_i - \mu_S)^2\right)^2} - 3$$

Definition 02.2. (Distribución de Laplace) La variable aleatoria de Laplace con media μ y varianza $2b^2$ se define el fenómeno de obtener un número en la recta real con la siguiente probabilidad.

$$\mathbb{P}_{Laplace}(-\infty, x) = \frac{1}{2b}\int_{-\infty}^{x} e^{-\frac{|t-\mu|}{b}}\,dt$$

Example 02.1.

1. *La kurtosis de una distribución gaussiana centrada en cero y con varianza uno es igual a cero.*

2. *La kurtosis de una distribución de Laplace centrada en cero y con varianza uno es igual a tres.*

Bibliografía

[1] S. Kotz, T. J. Kozubowski y K. Podgórski, *The Laplace Distribution and Generalizations*, Birkhäuser Boston, **2001**.

BOURBAKI

COLEGIO DE MATEMÁTICAS

escuela-bourbaki.com