Variables aleatorias

Curso 2022-23

13 de Febrero al 17 de Febrero de 2022

Variables aleatorias

- El concepto de variable aleatoria es un modo de unir técnicas que nos han aparecido en probabilidad y estadística descriptiva
- Entre otros conceptos:
  - De probabilidad
    - Sucesos y probabilidades asociados a ellos
    - Probabilidad, independencia, probabilidad condicionada y total, regla de Bayes
  - De estadística descriptiva
    - Medidas centrales y de dispersión como la media y la desviación típica e histogramas
    - Percentiles y datos atípicos
- Para poder usar todas estas técnicas, los sucesos a estudiar deben de ser numéricos

Variables aleatorias

- Partimos de un experimento aleatorio cualquiera
- Los sucesos no tienen por qué ser números
- Nos fijamos en una característica numérica de estos sucesos
- La variable aleatoria permite estudiar esa característica numérica
- Más formalmente
  - Una variable aleatoria es asignar a cada resultado del experimento un número
  - En vez de estudiar probabilidades de los sucesos del experimento estudiamos probabilidades de los números asociados
  - Matemáticamente, una variable aleatoria es una función $X : \Omega \to R$

Variables aleatorias discretas

Variables aleatorias discretas

- Una variable aleatoria discreta viene dada por
    - Una lista de los valores que puede tomar esa variable
    - Una lista de probabilidades de esos valores ya calculadas (con cualquier técnica)
- A partir de aquí
    - Al dibujar los valores junto a sus probabilidades se obtiene el equivalente al barplot o histograma de densidades
    - Faltaría
        - Definir media y varianza en función de las dos listas
        - Percentiles

Media y varianza de una variable aleatoria discreta

- Partimos de una variable aleatoria discreta X dada por:
  - Una lista de posibles valores $(x_1, \cdots, x_n)$
  - Una lista de probabilidades de esos valores ya calculadas $p_i = P(X = x_i)$
  - Suele denotarse $f(x_i) = P(X = x_i)$
- Media o esperanza de X

$$m = E[X] = x_1 \cdot p_1 + \cdots + x_n \cdot p_n$$

- En general, si $g(X)$ es cualquier función de la variable X

$$E[g(X)] = g(x_1) \cdot p_1 + \cdots + g(x_n) \cdot p_n$$

- Varianza

$$E[(X - m)^2] = (x_1 - m)^2 \cdot p_1 + \cdots + (x_n - m)^2 \cdot p_n$$

Función de distribución y percentiles

- Se llama función de distribución de X ,y se denota por F(x), a la función que asigna a cada valor x su percentil
- Una forma diferente de verlo es que F(x) asigna a x la probabilidad de ser menor o igual que ese valor

$$F(x) = P(X \leq x)$$

- Una forma sencilla de calcular F(x)
  - Calcular una lista con probabilidades acumuladas usando la función cumsum() en R
  - Para cualquier valor x_i de la lista de valores de X, el valor de $F(x_i)$ es el valor correspondiente en la lista calculada anteriormente
  - Para cualquier otro valor
    - Buscar en qué intervalo está de entre

$$(-\infty, x_1), [x_1, x_2), \cdots, [x_{n-1}, x_n), [x_n, \infty)$$

    - Si es menor que x_1 entonces $F(x) = 0$
    - En los demás casos, si está en el intervalo $[a, b)$, entonces $F(x) = F(a)$

Variables aleatorias continuas

Variables aleatorias continuas

- Una variable aleatoria se dice continua si puede tomar todos los valores de un intervalo.
- Una variable aleatoria continua viene dada por:
  - Un intervalo $[a, b]$ de posibles valores (puede ser abierto, etc)
  - Una función $f(x)$ (llamada de densidad) que nos indica qué regiones del intervalo son más probables que otras
- Esta función debe cumplir
  - $f(x) \geq 0, \forall x \in [a, b]$
  - $\int_a^b f(x)dx = 1$
- A partir de aquí
  - Al dibujar la función $f(x)$ se obtiene el equivalente al histograma de densidades

Media, varianza y esperanza para variables aleatorias continuas

- Son parecidas a las variables aleatorias discretas. Ahora:
  - Los valores vienen representados por la variable x
  - Las probabilidades por la función de densidad $f(x)$
  - Las sumas se convierten en integrales
- Con estos cambios
  - Media o esperanza de X: $m = E[X] = \int_a^b x \cdot f(x)dx$
  - Esperanza de una función $g(X)$: $E[g(X)] = \int_a^b g(x) \cdot f(x)dx$
  - Varianza: $v = E[(X-m)^2] = \int_a^b (x-m)^2 \cdot f(x)dx$

Función de distribución de una variable aleatoria continua

- Tiene el mismo papel que en variables aleatorias discretas
  - $F(x) = P(X \leq x)$
  - Por tanto $F(x)$ vale
    - 0 xi $x < a$
    - 1 si $x > b$
    - $\int_a^x f(t)dt$ si $a \leq x \leq b$
- Se utiliza para
  - Calcular probabilidades en subintervalos
  - Calcular percentiles

Cálculo de probabilidades

- En una variable aleatoria discreta
  - La probabilidad de estar en un intervalo es sumar las probabilidades de los valores del intervalo
  - Ejemplo
    - Una variable aleatoria toma valores $\{1, 2, \cdots, 10\}$ con probabilidades $P(X = i) = \frac{i^2}{385}$
    - En R tendríamos las listas $val = 1 : 10$, $prob = val^2/385$
    - Para calcular $P(3 < X <= 7)$ sumamos las probabilidades de 4, 5, 6, 7 que son los valores que lo cumplen
    - En R: $sum(prob[3 < val \,\&\, val <= 7])$

Calcular probabilidades

Cálculo de probabilidades

- En una variable aleatoria continua
  - La probabilidad de estar en un intervalo es sumar las probabilidades de los valores del intervalo
  - Ejemplo
    - Una variable aleatoria toma valores en el intervalo $[0,3]$ con función de densidad

$$f(x) = \begin{cases} 0, & \text{si } x < 0 \\ \frac{x^2}{9}, & \text{si } 1 \leq x \leq 3 \\ 0, & \text{si } x > 3 \end{cases}$$

    - $P(1 < X < 2) = \int_1^2 f(x)dx$
- Por tanto, en una variable aleatoria continua, $P(X = a) = \int_a^a f(x)dx = 0$
- No es cierto que $P(X = a) = f(a)$

Propiedades importantes

Variables independientes

- Sean X e Y dos variables aleatorias discretas. Se dice que X e Y son independientes si los sucesos $P(X = x)$ y $P(Y = y)$ son independientes para cualquier posible valor x de X y cualquier posible valor y de Y
- Si X e Y son dos variables aleatorias discretas, se dice que X e Y son independientes si los sucesos $P(X \leq x)$ y $P(Y \leq y)$ son independientes para cualquier posible par de números reales x e y.

Propiedades de la media y la varianza

- Sen X e Y dos variables aleatorias, a, b dos números
- Propiedades de la media
  - $E[a \cdot X + b] = a \cdot E[X] + b$
  - $E[X + Y] = E[X] + E[Y]$
  - Hay que tener cuidado con otras operaciones, por ejemplo $E[X \cdot Y] \neq E[X] \cdot E[Y]$
- Propiedades de la varianza
  - $Var(a \cdot X + Y) = a^2 \cdot Var(X)$
  - Si X e Y son independientes, $Var(X + Y) = Var(X) + Var(Y)$