Variables Aleatorias Continuas

Clase 4

Chan-Stein-Howlin-Casparian-Arceo-Spano

Organización

1. Variables Aleatorias Continuas
2. Percentiles
3. Esperanza variables aleatorias continuas
4. Varianza de una variable aleatoria continua
5. Distribuciones Conjuntas y Marginales

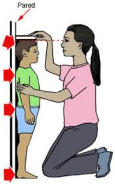
Variables Aleatorias Continuas

Cuándo una variable aleatoria es continua?

Una variable aleatoria es continua cuando su recorrido es un intervalo real. Este intervalo puede estar acotado superiormente, inferiormente, ambas cosas o no estar acotado.

Son ejemplos de variables aleatorias continuas:

- La altura de un adulto seleccionado al azar en una población.
- Un número real elegido al azar entre 0 y 1.
- Se pesa un paquete de arroz elegido al azar de la producción de una fábrica.



Variable aleatoria continua (definición)

Función de Densidad

Toda variable aleatoria continua X tiene asociada una función llamada función de densidad, f, tal que:

- $f(x) \geq 0$
- $\int_{\mathbb{R}} f(x)dx = 1$
- $P_X(B) = \int_B f(x)dx$ siendo B cualquier subconjunto de $\mathbb{R}$

Se pueden hacer las siguientes observaciones:

- Si X es una variable aleatoria continua entonces para todo $x \in \mathbb{R}$, $P_X(x) = 0$.
- La función de densidad puede no ser una función continua, aunque generalmente lo es.

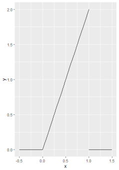
Ejemplo 1

Dada la función:

$$f(x) = \begin{cases} ax & \text{si } 0 \leq x \leq 1 \\ 0 & \text{si} x \notin [0,1] \end{cases}$$

Para qué valor de a resulta una función de densidad?

- Necesitamos que $a \geq 0$ para satisfacer la condicion 1)
- Necesitamos que se cumpla también $\int_0^1 axdx = 1$ para cumplir con la segunda condición. Es decir $a\frac{x^2}{2}|_0^1 = a\frac{1}{2} - a\frac{0}{2} = 1$, luego $a/2 = 1$ por lo cual $a = 2$.

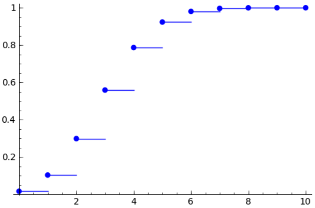


Función de distribución

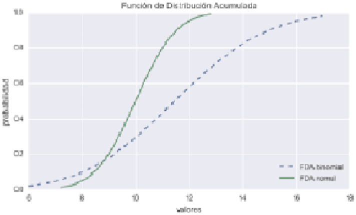
Función de distribución de Probabilidades Acumuladas

Siendo X una variable aleatoria continua con función de densidad f_X. Se llama **función de distribución acumulada** de X a la función $F : \mathbb{R} \to [0, 1]$ tal que:

$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt$



variable discreta



variable continua

Función de Distribución

Propiedades

Propiedades de la Función de Distribución (probabilidades acumuladas)

- $0 \leq F_X \leq 1$ para todo $x \in \mathbb{R}$.
- F_X es monótona no decreciente.
- F_X es continua.
- $F'_X(x) = f_X(x)$ por el Teorema Fundamental del Cálculo Integral.
- $P_X([a, b]) = F_X(b) - F_X(a)$ para todo $a < b$.

Función de Distribución

Relación con la función de densidad

Dadas las siguientes tres funciones de densidad de probabilidad:

$$f_1(x) = \begin{cases} \frac{1}{2} & \text{si } 0 \leq x \leq 2 \\ 0 & \text{si } x \notin [0;2] \end{cases}$$

$$f_2(x) = \begin{cases} \frac{x}{2} & \text{si } 0 \leq x \leq 2 \\ 0 & \text{si } x \notin [0;2] \end{cases}$$

$$f_3(x) = \begin{cases} 1 - \frac{1}{2}x & \text{si } 0 \leq x \leq 2 \\ 0 & \text{si } x \notin [0;2] \end{cases}$$

Función de Distribución

Relación con la función de densidad (continuación)

Cuál es la correspondencia? Por qué?

$$F_a(x) = \begin{cases} 0 & \text{si } x < 0 \\ \dfrac{1}{2}x & \text{si } 0 \leq x \leq 2 \\ 1 & \text{si } x \geq 2 \end{cases}$$

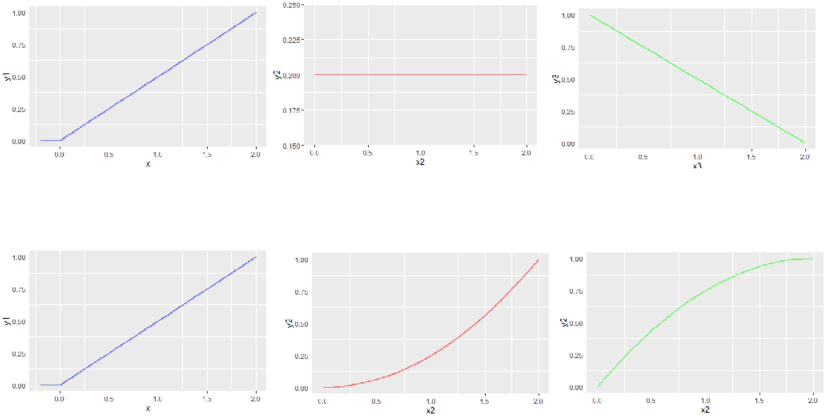
$$F_b(x) = \begin{cases} 0 & \text{si } x < 0 \\ x - \dfrac{x^2}{4} & \text{si } 0 \leq x \leq 2 \\ 1 & \text{si } x \geq 2 \end{cases}$$

$$F_c(x) = \begin{cases} 0 & \text{si } x < 0 \\ \dfrac{x^2}{2} & \text{si } 0 \leq x \leq 2 \\ 1 & \text{si } x \geq 2 \end{cases}$$

Función de Distribución

Relación con la función de densidad (continuación)

Cuál es la correspondencia? Por qué?



Ejemplo 3

Sea X una variable aleatoria con función de densidad:

$$f_X(x) = \begin{cases} 3e^{-3x} & \text{si } x > 0 \\ \\ 0 & \text{si } x \leq 0 \end{cases}$$

Notemos que el soporte (donde la densidad es estrictamente positiva) de esta función de densidad es no acotado.

Entonces la función de distribución acumulada es:

$$F_X(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ 1 - e^{-3x} & \text{si } x > 0 \end{cases}$$

Notemos que vale que
$P(1/3 < X < 1/2) = F(1/2) - F(1/3) = e^{-1} - e^{-3/2}$

Organización

Percentil de una variable continua

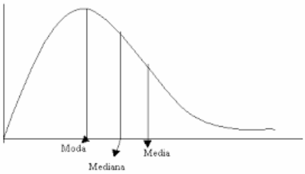
Percentil α

El **percentil 100*p** de la distribución F es el único valor x_p tal que:

$$F(x_p) = p = P(X \leq x_p)$$

La Mediana

Es un caso particular de percentil, corresponde al percentil 50 de la distribución. Es el valor del rango de la variable acumula el 50% de la probabilidad.



La media, la mediana y la moda pueden coincidir o no...depende de la forma de la distribución

Organización

Esperanza de variables aleatorias continuas

Esperanza Matemática

Sea X una variable aleatoria continua con función de densidad f_X.
Entonces se define la **Esperanza** de X y se nota $E(X)$ como:

$$E(X) = \int_{x \in \mathbb{R}} x f_X(x) dx$$

La esperanza queda definida cuando esta integral converge.

Ejemplo 4

Dada la variable aleatoria X (longitud de un corte de madera) de la cual conocemos la función de distribución

$$F_X(x) = \begin{cases} 0 & si \quad x < 2 \\ \frac{1}{2}x - 1 & si \quad 2 \leq x < 4 \\ 1 & si \quad x \geq 4 \end{cases} \tag{1}$$

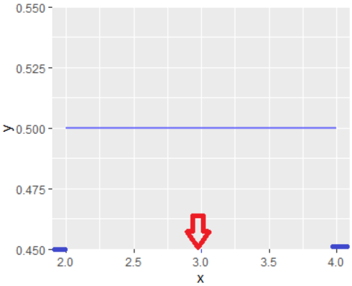
Derivando podemos obtener la expresión de la función de densidad:

$$f_X(x) = \begin{cases} \frac{1}{2} & si \quad 2 \leq x < 4 \\ 0 & \text{si no} \end{cases} \tag{2}$$

Con la expresión de la función de densidad podemos graficar la función y calcular el valor de su esperanza

Ejemplo 4 (continuación)

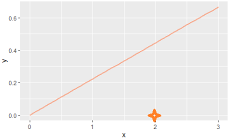
$E(X) = \int_2^4 \frac{1}{2} x dx = \frac{x^2}{4}\Big|_2^4 = 3$



Ejemplo 5

Dada la variable aleatoria X de la cual conocemos la función de densidad

$$f_X(x) = \begin{cases} \frac{2}{9}x & si \quad 0 \leq x < 3 \\ 0 & \text{si no} \end{cases} \tag{3}$$



Buscamos su valor esperado:

$E(X) = \int_0^3 x\frac{2}{9}xdx = \int_0^3 \frac{2}{9}x^2dx = \frac{2}{27}x^3|_0^3 = 2$ Buscamos su valor mediano:

$F_X(\widetilde{x}) = 0.5 \iff \frac{2}{27}x^3|_0^{\widetilde{x}} = 0.5 \iff \widetilde{x} = \sqrt[3]{\frac{9}{4}}$ NO COINCIDE CON LA ESPERANZA.

Esperanza de una función de la Variable Aleatoria

Para una función $h : \mathbb{R} \rightarrow \mathbb{R}$

Siendo X una variable aleatoria continua con función de densidad f_X vale que:

$$E(h(X)) = \int_{x \in \mathbb{R}} h(x) f_X(x) dx$$

Entonces como caso particular tenemos que: Es decir: $E(aX + b) = aE(X) + b$ para todo $a, b \in \mathbb{R}$.

Para dos variables aleatorias X e Y

Valen las mismas propiedades que las citadas para el caso discreto y vale también para ambos casos esta propiedad:

$$E(aX + bY) = aE(X) + bE(Y)$$

siendo $a, b \in \mathbb{R}$

Ejemplo 6

Sea X una variable aleatoria cuya función de densidad es:

$$f_X(x) = \begin{cases} 1 & si \quad 0 \leq x \leq 1 \\ \\ 0 & \text{si no} \end{cases} \tag{4}$$

Y sea h la función definida como:

$$h(x) = \begin{cases} 1 - x & si \quad 0 \leq x \leq \frac{1}{2} \\ \\ x & si \quad \frac{1}{2} < x \leq 1 \end{cases} \tag{5}$$

$E(h(X)) = \int_{x \in \mathbb{R}} h(x) f_X(x) dx = \int_0^1 h(x) dx = \int_0^{\frac{1}{2}} (1 - x) - dx + \int_{\frac{1}{2}}^1 x dx$
$E(h(X)) = \frac{3}{4}$.

Organización

Varianza de una variable aleatoria continua

Definición

Varianza de una variable aletoria continua

Sea X una variable aleatoria continua con función de densidad f_X y $E(X) = \mu$. Entonces se define la **Varianza** de X y se nota $V(X)$ como:

$$V(X) = E((X - \mu)^2) = \int_{x \in \mathbb{R}} (x - \mu)^2 f_X(x) dx$$

Desvío Standard de una variable aleatoria continua

Sea X una variable aleatoria continua. Se define el **desvío estándar** de X y se nota con σ_X a :

$$\sigma_X = \sqrt{V(X)}$$

Varianza de una variable aleatoria continua

Propiedades

Propiedades de la varianza

- $V(X) \geq 0$
- $V(X) = E(X^2) - (E(X))^2$
- $V(aX + b) = a^2 V(X)$
- $V(aX + bY) = a^2 V(X) + b^2 V(Y) + 2 \cdot a \cdot b \cdot Cov(X, Y)$ (ya veremos bien más adelante el significado de esta propiedad).

Organización

Probabilidad Conjunta

Sean X e Y variables aleatorias discretas con función de probabilidad conjunta dada por la siguiente tabla:

X \ Y	10	20	30	$p_X(x)$
100	0.1	0.15	0.25	0.5
200	0.3	0.15	0.05	0.5
$p_Y(y)$	0.4	0.3	0.3	1

- Se puede apreciar que que $R_X = \{100, 200\}$ y $R_Y = \{10, 20, 30\}$.
- Se visualizan las distribuciones marginales(margen) de X, de Y y la distribución conjunta de X e Y
- Así $p_Y(20) = 0.3$, $p_X(200) = 0.5$ y $p_{XY}(200, 30) = 0.05$.

Distribuciones Conjuntas y Marginales

Ejemplo 1

Se arrojan dos dados equilibrados de 4 caras (raros!!). Sean X = 'suma de los números e Y = 'número de ases' Podemos apreciar que los recorridos de las variables X e Y son:

- $R_X = \{2, 3, 4, 5, 6, 7, 8\}$
- $R_Y = \{0, 1, 2\}$

La distribución conjunta de ambas variables es:

Y \ X	**2**	**3**	**4**	**5**	**6**	**7**	**8**
0	**0**	**0**	**1/16**	**2/16**	**3/16**	**2/16**	**1/16**
1	**0**	**2/16**	**2/16**	**2/16**	**0**	**0**	**0**
2	**1/16**	**0**	**0**	**0**	**0**	**0**	**0**

Ambas v.a. sobre el mismo Espacio Muestral

X	1	2	3	4
1	2	3	4	5
2	3	4	5	6
3	4	5	6	7
4	5	6	7	8

Y	1	2	3	4
1	2	1	1	1
2	1	0	0	0
3	1	0	0	0
4	1	0	0	0

Algunos Cálculos

Hallemos las Distribuciones Marginales de X e Y

X	**2**	**3**	**4**	**5**	**6**	**7**	**8**
0	**1/16**	**2/16**	**3/16**	**4/16**	**3/16**	**2/16**	**1/16**

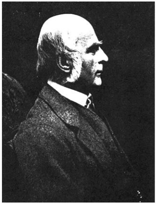
Y	$p_Y(y)$
0	**9/16**
1	**6/16**
2	**1/16**

Relación entre Variables Aleatorias

- Una pregunta elemental es si una de estas variables tiene relación o no con la otra? El interés es predecir o estimar una conociendo la otra?
- En nuestro problema si sé la cantidad de ases que salieron tengo información respecto del valor de la suma?
- y si conozco la suma tengo idea de la cantidad de ases que salieron??

Planteo de la Relación

El primero en indagar acerca de esta asociación fue un biólogo, Galton, preocupado por factores hereditarios, quien consultó al matemático Pearson.



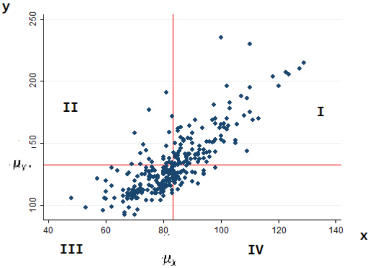
Sir Francis Galton

Karl Pearson

Idea de Pearson

En la Figura quedan determinados 4 cuadrantes

Por las líneas: $x = \mu_X$ (vertical roja) e $y = \mu_Y$ (horizontal roja).



Los puntos del primero y tercer cuadrante

Tienen el producto $(x - \mu_X)(y - \mu_Y) > 0$ Mientras que los puntos del segundo y cuarto cuadrantes tienen ese producto negativo.

Formalización de la Idea de Pearson

- Dos variables tienen **asociacion positiva** cuando al aumentar una, aumenta la otra y al disminuir una disminuye la otra, la covarianza es positiva y los puntos aparecen en su mayoria en los cuadrantes I y III.
- Dos variables tienen **asociacion negativa** cuando al aumentar una, la otra disminuye, la covarianza es negativa y los puntos aparecen en su mayoria en los cuadrantes II y IV.
- Cuando no haya ninguno de estos dos comportamientos diremos que no están asociadas o tienen covarianza 0.

Definición de Covarianza

Siendo X e Y dos v.a., tales que sus esperanzas existen, entonces:

$$Cov(X, Y) = E((X - E(X))(Y - E(Y))$$

Propiedades de la Covarianza

- $Cov(X,X) = Var(X)$
- $Cov(X,Y) = E(XY) - E(X)E(Y)$

$$\begin{aligned}
Cov(X,Y) &= E((X-E(X))(Y-E(Y)) \\
&= E(XY - E(X)Y - XE(Y) + E(X)E(Y)) \\
&= E(XY) - E(X)E(Y) - E(X)E(Y) + \\
&\quad +E(X)E(Y)) \\
Cov(X,Y) &= E(XY) - E(X)E(Y)
\end{aligned}$$

- $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$
  $Var(X-Y) = Var(X) + Var(Y) - 2Cov(X,Y)$

Calculemos la Covarianza de Nuestro Ejemplo

- $E(X) = 2 * \frac{1}{16} + 3 * \frac{2}{16} + 4 * \frac{3}{16} + 5 * \frac{4}{16} + 6 * \frac{3}{16} + 7 * \frac{2}{16} + 8 * \frac{1}{16} = 5$
- $E(Y) = 0 * \frac{9}{16} + 1 * \frac{6}{16} + 2 * \frac{1}{16} = \frac{1}{2}$
- $E(XY) = 0 * \frac{9}{16} + 3 * \frac{2}{16} + 4 * \frac{3}{16} + 5 * \frac{2}{16} = \frac{7}{4}$
- $Cov(X, Y) = E(XY) - E(X)E(Y) = \frac{7}{4} - 5 * \frac{1}{2} = -\frac{3}{4}$

Observemos que:

Nos dio negativa la covarianza!! Es correcto?

El problema de la covarianza es que es sensible a las unidades y que no nos da una idea de la fuerza de esa asociación. Por este motivo Pearson definió la correlación entre dos variables.

Correlación

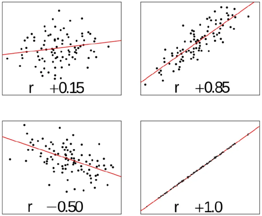
Coeficiente de Correlación Lineal de Pearson

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y}$$

Rango de Variación e Interpretación

La correlación toma valores entre -1 y 1.

$$-1 \leq \rho_{XY} \leq 1$$



Coeficiente de correlación lineal para nuestro Ejemplo

Necesitamos las dos Desviaciones Standard

$$E(X^2) = 2^2 * \frac{1}{16} + 3^2 * \frac{2}{16} + 4^2 * \frac{3}{16} + 5^2 * \frac{4}{16} + 6^2 * \frac{3}{16} + 7^2 * \frac{2}{16} + 8^2 * \frac{1}{16}$$

$$E(X^2) = 27.5$$

$$E(Y^2) = 0^2 * \frac{9}{16} + 1^2 * \frac{6}{16} + 2^2 * \frac{1}{16} = \frac{5}{8}$$

$$V(X) = 27.5 - 5^2 = 2.5$$

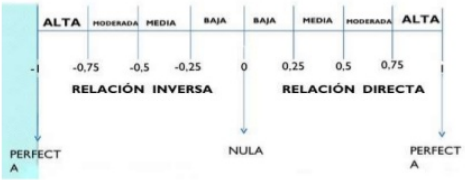
$$V(Y) = \frac{5}{8} - 0.5^2 = 0.375$$

$$\sigma_X = 1.58$$

$$\sigma_Y = 0.612$$

Coeficiente de correlación lineal para nuestro Ejemplo

$$\rho_{XY} = \frac{cov(X, Y)}{\sigma_X \cdot \sigma_Y} = \frac{-0.75}{1.58 * 0.612} = -0.776$$



MUCHAS GRACIAS!