TECNICATURA UNIVERSITARIA EN PROGRAMACIÓN

Estadística

Unidad 4:

Variables aleatorias continuas

UTN FRGP

Contenido

Variables aleatorias continuas (V.A.C.)

La unidad anterior definimos lo que era una variable aleatoria y estuvimos trabajando con distintos tipos de variables discretas. Las variables aleatorias continuas, a diferencia de las discretas, son aquellas que pueden tomar cualquier valor en un intervalo dado. Si bien en la teoría este intervalo puede ser infinito (tomando valores desde menos infinito a más infinito, como es el caso de las distribuciones normales de las que hablaremos mas adelante), en la práctica suele trabajarse con un intervalo acotado para facilitar los cálculos y el análisis matemático.

Algunos ejemplos de variables aleatorias continuas pueden ser el peso o la altura de las personas de un determinado grupo. Si bien en la teoría la altura o el peso puede variar desde cero hasta infinito, en la práctica sabemos que van a estar acotados entre un valor mínimo y un valor máximo. En el caso de la altura, podría acotarse entre 70 cm y 230 cm, que cubriría la mayoría de las alturas en una población especifica.

Función de densidad de probabilidad para V.A.C.

La función de densidad de probabilidad de una V.A.C. X se define como una función real continua f(x) no negativa e integrable, tal que la probabilidad de que un determinado valor x pertenezca al intervalo (a, b) es igual al área bajo la curva entre esos dos puntos. Es decir, el área bajo la función de densidad entre dos valores se interpreta como la probabilidad de que X caiga en ese intervalo, que denotamos $P(a < X < b)$. Nótese que al trabajar con áreas bajo la curva, la función de densidad no puede calcularse para valores puntuales (por ejemplo $P(X = a)$), ya que no existe área bajo un punto. Esto implica que $P(a < X < b)$ es igual a $P(a \leq X \leq b)$, $P(a < X \leq b)$ y $P(a \leq X < b)$ o expresado en términos de intervalos, la probabilidad de que X caiga en el intervalo (a, b) es igual a la de que caiga en el intervalo [a, b], (a, b] y [a, b).

Si bien las funciones de densidad de probabilidad (FDP) pueden tener diferentes formas, en esta materia solo vamos a trabajar con dos distribuciones específicas: la distribución uniforme y la distribución lineal. La distribución uniforme es útil para describir situaciones en las que todos los valores en un intervalo tienen la misma probabilidad, mientras que la distribución lineal se refiere a una relación lineal entre variables. Se seleccionaron estas distribuciones porque no requieren el uso de métodos de integración o funciones primitivas, lo que facilita su estudio y aplicación en el contexto de esta materia.

Propiedades de las FDP

Siendo f(x) una función cualquiera, para que sea una FDP en el intervalo (a, b), debe cumplir con las siguientes propiedades:

- No negatividad: La función siempre debe ser positiva en el intervalo (a, b), es decir que f(x) ≥ 0 para todo x entre a y b.
- El área bajo la función en el intervalo (a, b) tiene que ser igual a 1.

FDP uniforme

Un ejemplo de FDP uniforme podría ser:

$$f(x) = \begin{cases} 1/5 & si\ 0 < x < 5 \\ 0 & en\ otro\ caso \end{cases}$$

Donde la función f(x) va a valer 1/5 en el intervalo (0, 5) y 0 fuera de ese intervalo.

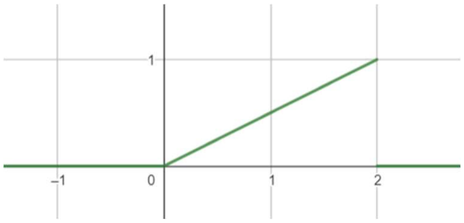
Para este tipo de funciones vamos a usar la fórmula del área de un rectángulo para calcular la probabilidad de que X caiga en un determinado intervalo, siendo la base la longitud de nuestro intervalo y la altura el valor que toma la función. Cabe destacar que si alguno de los límites del intervalo al que le queremos calcular la probabilidad cae fuera del intervalo sobre el que nuestra f(x) no es cero, vamos a acotarlo al límite de nuestra función para poder calcular el área.

FDP lineal

En el caso de que nuestra función de probabilidad esté dada por una función lineal, vamos a trabajar con la fórmula del área del triángulo (base x altura / 2), dado que nuestra recta va a formar un triángulo rectángulo con el eje de abscisas.

Mediante la resta de triángulos podemos llegar a calcular el área de cualquier intervalo, calculando el área del triángulo que se forme desde donde la función corta al eje de abscisas hasta límite del intervalo que nos deje el triángulo más grande y restándole el área del triángulo que se forma hasta el otro límite. Por ejemplo, supongamos que tenemos la siguiente FDP:

$$f(x) = \begin{cases} \frac{1}{2}x & si\ 0 < x < 2 \\ 0 & en\ otro\ caso \end{cases}$$



Si queremos calcular P(1 < X < 2) podemos hacerlo restándole al triángulo que se forma para el intervalo (0, 2) el triángulo que se forma para el intervalo (0, 1). Lo que sería lo mismo que decir P(1 < X < 2) = P(X < 2) – P(X < 1). También se podría utilizar la fórmula del trapecio, que sería **(base mayor + base menor) x altura / 2**, pero hay que tener en cuenta que tanto la base mayor como la base menor en realidad son las alturas de nuestra función y la altura es la base.

Distribución normal

La distribución normal (también conocida como distribución gaussiana) es una de las más utilizadas para modelar el comportamiento de variables aleatorias continuas, ya que describe con precisión a muchos fenómenos que a primera vista pueden parecer muy distintos entre sí.

Suele emplearse cuando la variable aleatoria continua observada tiene las siguientes características:

- Los valores que toma tienden a agruparse alrededor de un valor central, decreciendo rápidamente al alejarse del mismo.
- Los valores son simétricos en cuanto a la frecuencia o probabilidad de ocurrencia, con respecto del valor central.

Como se trata de una variable continua, tiene una función de densidad asociada. Si una variable aleatoria tiene distribución normal de parámetros μ y σ, entonces vale que:

$$X \sim N(\mu, \sigma) \rightarrow f(X) = \frac{1}{\sigma\sqrt{2\pi}} . e^{-\frac{(X-\mu)^2}{2\sigma^2}}$$

Donde μ representa al promedio de la variable y σ a su desvío estándar.

La gráfica de la función es curva que tiene una forma de campana, centrada en la media. En el intervalo [μ – σ, μ + σ] se encuentra comprendida, aproximadamente, el 68,26% de la distribución; en el intervalo [μ – 2σ, μ + 2σ] se encuentra, aproximadamente, el 95,44% y en el intervalo [μ – 3σ, μ + 3σ] se encuentra, aproximadamente, el 99,74% de la distribución.

Distribución normal estándar

La distribución normal estándar es una forma especifica de la distribución normal que tiene una media igual a cero y un desvío estándar igual a uno. Se la denota con la letra Z y se expresa como Z~N(0, 1).

Las probabilidades acumuladas para los distintos valores que puede tomar la variable Z están tabulados para facilitar su búsqueda sin necesidad de realizar ningún cálculo. Hay distintos tipos de tablas:

- Acumulada a izquierda: nos da la probabilidad acumulada desde menos infinito hasta un valor determinado. Esta es la tabla que usaremos en la materia.
- Acumulada a derecha: nos da la probabilidad acumulada desde un valor determinado hasta infinito. Esta tabla es el complemento de la anterior.
- Acumulada desde la media: nos da la probabilidad acumulada entre 0 (la media) y un valor determinado.

En los márgenes de la tabla vamos a encontrar valores que corresponden a Z y en el cuerpo de esta, las probabilidades acumuladas para esos valores. En la primer columna se encuentran los valores de la parte entera y el primer decimal de Z y en la primera fila el segundo decimal de Z. Para buscar la probabilidad acumulada de un determinado valor de Z, tenemos que hallar el valor que está en la intersección de la fila que tiene la parte entera y el primer decimal y la columna que tiene el valor del segundo decimal.

Estandarización de una variable X~N(μ, σ)

Cualquier variable que siga una distribución normal puede ser estandarizada usando la siguiente fórmula:

$$z = \frac{x - \mu}{\sigma}$$

De este modo, para cualquier valor x de una distribución X~N(μ, σ) podemos hallar un valor equivalente Z para la distribución normal estándar y utilizar los valores tabulados para la misma.

Uso inverso de la tabla

Normalmente cuando queremos hallar la probabilidad acumulada para un determinado valor de Z, buscamos este valor en los márgenes y obtenemos el valor de probabilidad que corresponde a la intersección de la fila y la columna, pero supongamos que lo que necesitamos hallar es el valor de Z a partir de una probabilidad acumulada. Esto podemos lograrlo haciendo el camino inverso, en lugar de buscar en los márgenes, buscamos en el cuerpo de la tabla el valor mas aproximado a esa probabilidad. A partir de dicho valor, nos fijamos los valores de los márgenes en los que se encuentra, lo que nos da el valor de Z.

Cabe destacar que como nuestra tabla nos da las probabilidades a izquierda, esto solo es válido para $P(Z < z) = k$ (donde k es el valor dato de la probabilidad que buscamos en la tabla y z es el valor buscado). En caso de tener como dato $P(Z > z) = k$, podemos usar el complemento y buscar $1 - k$ para hallar el valor de z que nos sirve.

Aplicando la fórmula de la estandarización y despejando el valor de X, podemos utilizar esta técnica para cualquier distribución normal $X \sim N(\mu, \sigma)$.