TECNICATURA UNIVERSITARIA EN PROGRAMACIÓN

Estadística

Unidad 3:

Variables aleatorias discretas

Contenido

Variables aleatorias

Una variable aleatoria es una función que asigna a cada elemento del espacio muestral de un experimento aleatorio un número real. Estas variables suelen denotarse con letras mayúsculas (X, Y, Z..). Dependiendo de la naturaleza del fenómeno estudiado, las variables aleatorias pueden ser discretas o continuas.

Las variables aleatorias nos permiten describir y comprender mejor los fenómenos aleatorios a través de diferentes medidas estadísticas, como la media, la varianza y la función de distribución. Estas medidas proporcionan información sobre las características centrales y la dispersión de los valores que puede tomar una variable aleatoria, así como la probabilidad asociada a cada uno de ellos.

En esta unidad vamos a tratar las variables aleatorias discretas, dejando a las continuas para la siguiente unidad.

Variables aleatorias discretas (V.A.D.)

Una variable aleatoria discreta es aquella que toma valores aislados y contables, como por ejemplo el número de caras obtenidas al lanzar una moneda varias veces. Dependiendo de la cantidad de veces que arrojemos la moneda, la variable va a tener un rango determinado. Por ejemplo, si se arroja la moneda tres veces, nuestra V.A.D. puede tomar los valores 0, 1, 2 o 3 caras.

Ejemplo:

Siendo nuestro experimento aleatorio: “Se arroja una moneda ideal 2 veces” y nuestra variable aleatoria X: “número de caras obtenido”.

El espacio muestral de nuestro experimento es E = {(cara, cara), (cara, ceca), (ceca, cara), (ceca, ceca)}

Si contamos el número de caras obtenido en cada posible resultado del espacio muestral, podemos observar que nuestra variable X puede tomar los valores 0, 1 y 2. Las probabilidades de que tome cada uno de esos valores va a ser la suma de las probabilidades de cada caso favorable a dicho valor. En este caso, todos los resultados de lanzar una moneda dos veces son equiprobables (tienen la misma probabilidad de ocurrir), ya que al ser una moneda ideal hay una probabilidad de 0.5 de que salga cara y también de 0.5 de que salga ceca, lo que nos deja una probabilidad de 0.25 para cada elemento de nuestro espacio muestral. Para X=0 tenemos un solo caso favorable (ceca, ceca), para X=1 tenemos 2 casos favorables ((cara, ceca) y (ceca, cara)) y para X=2 nuevamente tenemos un solo caso favorable.

Organizando los resultados en una tabla nos queda que:

X	0	1	2
$P(X=x_i)$	0.25	0.5	0.25

Una vez que tenemos los datos organizados, si quisiéramos por ejemplo saber la probabilidad de que salga a lo sumo una cara (lo que sería lo mismo que decir $P(X \leq 1)$) podemos calcularlo en base a esos valores:

$P(X \leq 1) = P(X = 0) + P(X = 1) = 0.25 + 0.5 = 0.75$

Esperanza de una V.A.D.

La esperanza es el valor al que va a tender nuestra variable aleatoria si repitiéramos nuestro experimento aleatorio un número grande de veces. A la esperanza también se la conoce como valor esperado, valor medio o promedio y se denota como E(X) (siendo X la variable aleatoria).

Para calcular la esperanza lo que hacemos es sumar todos los valores que puede tomar nuestra variable, ponderados por su respectiva probabilidad. En otras palabras, es sumar el producto de cada valor x_i multiplicado por su respectiva probabilidad ($P(X=x_i)$). Expresado en una fórmula nos queda:

$$E(X) = \sum_{i=1}^{n} (x_i \, . \, P(X = x_i))$$

Propiedades de la esperanza

Siendo X e Y dos variables aleatorias independientes y a y b constantes:

- E(a . X ± bY) = a . E(X) ± b . E(Y) (conservando el signo de la operación)
- E(a) = a

Varianza de una V.A.D.

La varianza es una medida de dispersión que nos dice que tan cercanos al valor medio se encuentran los valores de nuestra variable y se la denota como V(X). Mientras mas grande sea la varianza, más dispersos estarán nuestros valores respecto de la esperanza.

Para calcular la varianza tenemos que hacer la diferencia entre la esperanza de nuestra variable aleatoria elevada al cuadrado y la esperanza al cuadrado de nuestra variable aleatoria. Esto es:

$$V(X) = E(X^2) – [E(X)]^2$$

El sustraendo ya vimos como calcularlo, solo hace falta elevar al cuadrado el valor esperado de X. Para el minuendo lo que hacemos es calcular la esperanza, pero elevando al cuadrado a los x_i:

$$E(X^2) = \sum_{i=1}^{n} ({x_i}^2 \,.\, P(X = x_i))$$

Nótese como el cuadrado solo afecta a x_i y no a su probabilidad y que su valor está expresado en unidades al cuadrado respecto de la unidad de nuestra variable aleatoria. Esto quiere decir que, si nuestra variable aleatoria está expresada en litros, la varianza representara litros al cuadrado. Para solventar este problema tenemos el desvío estándar.

Propiedades de la varianza

Siendo X e Y dos variables aleatorias independientes y a y b dos constantes:

- V(a) = 0
- V(a . X ± b Y) = a^2 . V(X) + b^2 . V(Y) (siempre se suman las varianzas).

Desvío estándar

El desvío estándar se calcula como la raíz cuadrada de la varianza y se expresa en las mismas unidades que nuestra variable aleatoria original, facilitando su interpretación y comparación con otros valores. Se lo denota con la letra griega sigma minúscula (σ).

Distribución binomial

La distribución binomial es uno de los modelos de probabilidad más comunes y ampliamente utilizados en estadística y teoría de las probabilidades. Este modelo es aplicable a experimentos que cumplan las siguientes condiciones:

- Experimentos independientes: Cada experimento no debe afectar el resultado de los siguientes.
- Resultados dicotómicos: Los experimentos deben tener dos resultados posibles, denominados favorable o éxito y desfavorable o fracaso.
- Probabilidad constante: La probabilidad de éxito del experimento (denotada como **p**), debe mantenerse constante para cada experimento.
- Número finito de experimentos: Se fija un número especifico de ensayo, denotado con la letra **n**.

Si se cumplen estas condiciones, se dice que una variable aleatoria X sigue una distribución binomial y se expresa como: X~Bi(n, p), siendo **n** el número de ensayos y **p** la probabilidad de éxito (o de obtener un resultado favorable a nuestra variable). El complemento de la probabilidad de éxito (1 – p) es la probabilidad de fracaso y se la denota con la letra **q**.

Para poder calcular la probabilidad de que nuestra variable tome un valor **k**, usamos la siguiente fórmula:

$$P(X = k) = C_{n;\,k} \,.\, p^k \,.\, q^{n-k}$$

La esperanza en las distribuciones binomiales puede calcularse como: E(X) = n . p

La varianza se calcula como V(X) = n . p . q y el desvío es la raíz cuadrada de ese valor.

Distribución de Poisson

Otro modelo probabilístico ampliamente utilizado es la distribución de Poisson, que se utiliza en para casos en los que los eventos ocurren en un lapso o en un espacio determinado. Para poder utilizar esta distribución deben cumplirse las siguientes características:

- Los eventos deben ser aleatorios e independientes.
- La tasa promedio (denotada con la letra griega lambda **λ**) debe ser constante.
- Los resultados deben estar acotados a un intervalo de tiempo o espacio (superficie, volumen..).

Si se cumplen estas características entonces se dice que la variable aleatoria sigue una distribución de Poisson de parámetro **λ** y se expresa como: $X \sim P(\lambda)$, siendo lambda el promedio en un **determinado lapso o espacio** y puede llegar a recalcularse en caso de que queramos extender o reducir dicho lapso o espacio.

Para calcular la probabilidad de que nuestra variable X tome un determinado valor k en este tipo de distribución es la siguiente:

$$P(X = k) = \frac{e^{-\lambda} . \lambda^k}{k!}$$

Siendo **e** la constante de Napier o número de Euler, que es aproximadamente igual a 2.7183.

En las distribuciones de Poisson, tanto la esperanza como la varianza son iguales a lambda.