Clase 4: Variables Aleatorias y Distribución de Probabilidad.

Harvard Statistics 110: Probability

Resumen

En esta clase se realiza una introducción a los conceptos de variable aleatoria y de distribución de probabilidad, con mayor detalle en el primero. Ambos son fundamentales para comprender los próximos temas que se estudiarán en este curso.

1. Variables Aleatorias.

Consideremos el experimento de lanzar cuatro veces una moneda no sesgada de dos lados. Su espacio muestral S consiste de $2^4 = 16$ elementos o resultados muestrales. Un evento $E \subseteq S$ sería obtener tres caras, donde:

$$E = \{CSCC,\ SCCC,\ CCSC,\ CCCS\}$$

En este caso, $|E| = 4$, pero si decidimos lanzar ocho veces la misma moneda, entonces[1]:

$$|E| = \frac{8!}{3! \cdot 5!} = 56$$

Es decir, el conjunto E consistirá de 56 secuencias con tres caras y cinco sellos.

Es claro que si realizamos más lanzamientos de la moneda, el evento E se hará cada vez más complejo de expresar. Sin embargo, si solo nos interesan las tres caras, las secuencias pasan

[1] Uso la siguiente fórmula para calcular la cantidad de permutaciones de un multiconjunto, conocido también como coeficiente multinomial:

$$\binom{n}{n_1\ n_2\ \dots\ n_r} = \frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_r!}; \quad \text{donde} \sum_{i=1}^{r} n_i = n$$

Ver más en Larsen y Marx (2018). *An Introduction to Mathematical Statistics and Its Applications.* Pp 77.

a ser irrelevantes porque en todas ellas se cumple aquello. Para reducir toda esa tarea, es más conveniente definir una función X que **resuma** numéricamente esa cantidad.

Volvamos al experimento de lanzar cuatro veces una moneda. Para mayor legibilidad, establezcamos que $E = \{CSCC,\ SCCC,\ CCSC,\ CCCS\} = \{e_1,\ e_2,\ e_3,\ e_4\}$. Si X es una función de la cantidad de caras de los resultados del espacio muestral S, entonces:

$$X(e_1) = X(e_2) = X(e_3) = X(e_4) = 3$$

En probabilidades, X recibe el nombre de **Variable Aleatoria** y es definida como una **función** $X\colon S \mapsto \mathbb{R}$, con S siendo el espacio muestral de un experimento aleatorio. En ese sentido, cada imagen de ella se expresa como $x = X(s)$. Por lo tanto, su **rango** será el conjunto $\{x\colon X(s) = x;\ \forall s \in S\}$.

Como una variable aleatoria X es una regla que transforma a cada resultado muestral $s \in S$ en un número $x \in \mathbb{R}$, su **rango** pasa a ser un **espacio muestral resumido** del mismo experimento.

Por ejemplo, en el experimento de lanzar ocho veces una moneda, su espacio muestral consiste de $2^8 = 256$ resultados muestrales. Si X es una variable aleatoria de la cantidad de caras obtenidas, su rango será el conjunto $\{0,\ 1,\ 2,\ 3,\ \ldots,\ 8\}$. También es un espacio muestral (cada número es un resultado del experimento), pero asociado solo a X.

Como a veces es difícil expresar un resultado muestral[2] $s \in S$, se denotará al resultado de una variable aleatoria X como $X = x$ en vez de $X(s) = x$. De igual modo se hará para otros operadores de relación, tales como $>$, $<$, etc.

2. Distribución de probabilidad.

Puesto que las variables aleatorias están definidas en el espacio muestral de un experimento aleatorio, un paso lógico a seguir es calcular la probabilidad de que suceda ese valor o qué tan posible es que ocurra dentro de un intervalo (cerrado o abierto) de números. En otras palabras, a toda variable aleatoria se le puede asociar una función (o medida) de probabilidad.

Un concepto que surge del vínculo entre las variables aleatorias y sus funciones de probabilidad es el de **distribución de probabilidad**. Corresponde a una **colección** de **todas las probabilidades asociadas a cada valor que puede tomar una variable aleatoria**. En ese sentido, dicho listado describe el **comportamiento** probabilístico de esta última

[2]Sea en una cadena de caracteres, en notación de conjuntos, etc.

función.

Producto del desarrollo teórico de las probabilidades y de su aplicación en muchas áreas del conocimiento, se han podido generalizar distintas distribuciones de probabilidad las que, debido a la relevancia en cuanto a sus características, han llegado a ser nombradas. Algunos ejemplos son las distribuciones binomial o la de Gauss.

Dependiendo del hecho o fenómeno que se quiera representar mediante una variable aleatoria, se va a querer asumir (al menos hipotéticamente) que sigue una distribución de probabilidad determinada. Por ejemplo, si señalamos que X es una variable aleatoria que es descrita por una distribución binomial, entonces lo denotamos como:

$$X \sim \text{Binom}(n,\ p)$$

En la expresión $\text{Binom}(n,\ p)$, tanto n como p son los parámetros que definen a la distribución binomial y que debemos conocer de algún modo[3] para afirmar que, efectivamente, la variable aleatoria X sigue dicha distribución.

En el estudio de las probabilidades, las variables aleatorias han sido divididas en dos grandes grupos[4]:

1. Variables aleatorias discretas.
2. Variables aleatorias continuas.

En las siguientes clases estudiaremos ambos grupos de variables aleatorias, comenzando con las discretas. Al hacerlo, también profundizaremos en sus funciones y distribuciones de probabilidad más relevantes.

[3] O al menos estimarlas, que es algo que se estudia más profundamente en Estadística.

[4] También se identifica un tercer grupo de variables aleatorias conocidas como **mixtas**, que se caracterizan por ser tanto discretas como continuas.