AULA ESTADÍSTICA

DISTRIBUCIÓN DE PROBABILIDADES

Material Didáctico para el estudio de Variables Aleatorias Discretas

Prof. Jorge Lorenzo

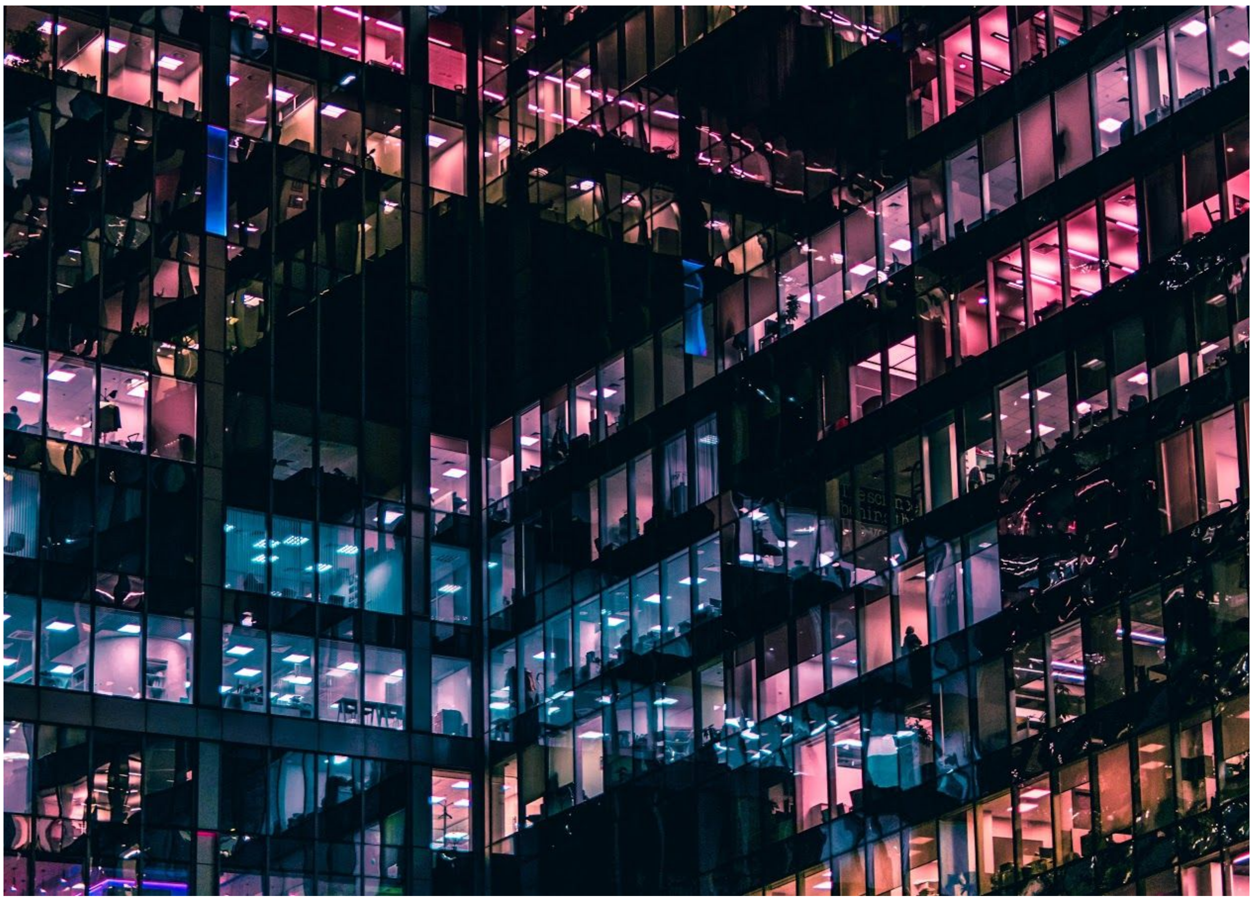

Foto cortesía de Mike Kononov (Unsplash)

Introducción

Comenzaremos por definir lo que se entiende por variable aleatoria: una variable aleatoria X es una función que asocia un número real a cada punto del espacio muestral. Dado un experimento aleatorio cualquiera cuyos sucesos elementales posibles pueden identificarse fácilmente mediante un número real, se denomina Variable Aleatoria X, al conjunto de estos números.

Una variables aleatoria discreta produce como resultado un número finito de valores, por lo que su recorrido es finito. Así, la variable aleatoria discreta, tiene un conjunto definido de valores posibles $x_1,x_2,x_3,.....x_n$, con probabilidades respectivas $p_1,p_2,p_3,.....p_n$. Es decir, sólo puede tomar ciertos valores dentro de un campo de variación dado. Como X ha de tomar uno de los valores de este conjunto, entonces $p_1+p_2+p_3,.....+p_n=1$.

En general, una variable aleatoria discreta X representa los resultados de un espacio muestral en forma tal que por $P(X = x)$ se entenderá la probabilidad de que X tome el valor de x. De esta forma, al considerar los valores de una variable aleatoria es posible desarrollar una función matemática que asigne una probabilidad a cada realización x de la variable aleatoria X. Esta función recibe el nombre de *función de la probabilidad*.

En un experimento aleatorio, la esperanza matemática se define como la suma del producto de cada valor de la variable aleatoria considerada por su probabilidad. Para cada distribución de probabilidad es posible calcular este valor. En estadística la esperanza matemática de una variable aleatoria X, es el número E(X) que formaliza la idea de valor medio de un fenómeno aleatorio. Cuando la variable aleatoria es discreta, la esperanza es igual a la suma de la probabilidad de cada posible suceso aleatorio multiplicado por el valor de dicho suceso. Por lo tanto, representa la cantidad media que se "espera" como resultado de un experimento aleatorio cuando la probabilidad de cada suceso se mantiene constante y el experimento se repite un elevado número de veces. Por ser valores calculados a partir de un conjunto de datos, es posible que la esperanza matemática no esté contenido dentro de los valores reales observados.

Cuando se analiza un experimento aleatorio, es posible describir el comportamiento de la variables mediante modelos particulares. Por ello, es frecuente asociar a estos experimentos una función de probabilidad, que puede adoptar diversas formas y regirse por principios diferentes, y cuyo estudio describe la naturaleza y las características del fenómeno ligado al experimento.

En teoría de la probabilidad y estadística, la distribución de probabilidad de una variable aleatoria es una función que asigna a cada suceso definido sobre la variable aleatoria, la probabilidad de que dicho suceso ocurra. La distribución de probabilidad está definida sobre el conjunto de todos el rango de valores de la variable aleatoria.

Cuando la variable aleatoria toma valores en el conjunto de los números reales, la distribución de probabilidad está completamente especificada por la función de distribución, cuyo valor en cada valor de *x* es la probabilidad de que la variable aleatoria sea menor o igual que *x*. Dada una variable aleatoria X, se llama función de distribución a aquella que proporciona la probabilidad de que la variable aleatoria tome un valor menor o igual que x_i. Es decir:

$$F\left(x_i\right) = P\left(X \leq x_i\right)$$

Si se conoce la función de distribución F(x) de una variable aleatoria X, siempre se cumple que la probabilidad de que la variable aleatoria tome valores en el intervalo [a, b] es:

$$P(a < X \leq b) = F(b) - F(a)$$

Así como puede calcularse la esperanza matemática o promedio, es posible calcular la varianza. Ésta se define como el promedio de las diferencias con de cada valor respecto a la media, elevadas al cuadrado. En otras palabras, es la desviación cuadrática de cada valor de x respecto a su promedio. Para conocer el grado de dispersión de los valores de una variable es necesario considerar su varianza. En variables aleatorias discretas, la varianza se define como:

$$V(X) = \sum_{i=1}^{n} P_i x_i^2 - E(X)^2$$

Tratándose la varianza de desviaciones cuadráticas respecto de la media, sus unidades también estarán expresadas en ese sentido, por lo tanto, para interpretar la dispersión en la mismas unidades en que se reporta la media aritmética, es necesario utilizar la raíz cuadrada de la varianza, conocida como desviación estándar.

La Función de Distribución Acumulada corresponde a la probabilidad de que la variable aleatoria X tome un valor numérico menor o igual a x, en este sentido representa las probabilidades acumuladas hasta alcanzar ese valor. Esto se expresa como:

$$F(x) = P\left[X \leq x\right]$$

En este trabajo revisaremos las principales funciones de probabilidades. En el caso en que los docentes lo deseen, pueden ejercitar estos contenidos es a través de una aplicación gratuita que puede descargarse a cualquier tableta o celular. Al presentar los conceptos esenciales y luego la aplicación, los alumnos pueden verificar resultados y realizar cambios en los parámetros de las distribuciones en sus dispositivos. De esta manera, se puede conectar lo artefactual con contenidos de estadística de manera que puedan usarse en contextos áulicos con la apoyatura de TICs. Una aplicación de referencia se es f(x): distribución de probabilidades, versión 5.6.2, desarrollada por Mathew Bognar del Departamento de Estadística y Ciencias Actuariales de la Universidad de Iowa. La aplicación está disponible de manera gratuita a través de la tienda Google Play.

En caso de que no se utilice una aplicación como la mencionada, en los siguientes apartados expondremos los fundamentos de distintas distribuciones de probabilidades acompañadas de un ejemplo. Modificando la consigna y los valores presentados, el docente podrá ampliar el repertorio de posibilidades para la ejercitación. Cabe destacar que en la misma, se encuentra detallado las fórmulas de cálculo de los parámetros correspondientes a media, varianza y desviación estándar.

Espero que este aporte pueda ser de utilidad.

OER

OPEN EDUCATIONAL
RESOURCES

Distribuciones de Variables Aleatorias Discretas

Distribución Uniforme

La variable aleatoria discreta X se dice que tiene una distribución uniforme si puede tomar cualquiera de *n* valores x_1, x_2, x_3 ..., x_n, con probabilidad

$$P\left(X = x_i \right) = \frac{1}{n} \ \forall i$$

Los parámetros de esta distribución de media, varianza y desviación estándar serán:

$$\mu_x = \frac{n+1}{2}$$

$$\sigma_x^2 = \frac{n^2 - 1}{12}$$

$$\sigma_x = \sqrt{\frac{n^2 - 1}{12}}$$

Ejemplo: estamos interesados en un suceso específico en el lanzamiento de un dado, obtener un valor menor a cuatro en un lanzamiento. Sabemos que las probabilidades para cada uno de los valores en las caras es igual a ⅙, ¿qué probabilidad le corresponde al conjunto de valores {1,2,3}? Como se trata de un lanzamiento, dicha probabilidad es efectivamente p=3/6= 0,5. La media de esta distribución será igual a 2, con varianza igual a 0,66 y desviación estándar igual a 0, 81.

Distribución de Bernouilli

La distribución de Bernoulli, describe un experimento aleatorio que sólo admite dos resultados excluyentes (éxito y fracaso). La variable aleatoria discreta X asociada a este experimento toma el valor 1 cuando ocurre el suceso éxito con probabilidad P(A)=p y el valor 0 cuando ocurre el suceso fracaso con probabilidad P(A)=q. Entonces:

X=0; $P(X=x_i)=q$

X=1; $P(X=x_i)=p$

Los parámetros de esta distribución de media, varianza y desviación estándar serán:

$\mu_x = p$

$\sigma_x^2 = p \cdot q$

$\sigma_x = \sqrt{p \cdot q}$

Ejemplo: El gerente de un restaurante que sólo da servicio mediante reservas sabe, por experiencia, que el 20% de las personas que reservan una mesa no asistirán. Si el restaurante acepta 25 reservas pero sólo dispone de 20 mesas, ¿cuál es la probabilidad de que a todas las personas que asistan al restaurante se les asigne una mesa?

Representemos por la variable aleatoria X decisión de asistir (p=0), o no asistir (p=1), al restaurante por parte de una persona que ha hecho una reserva. Esta variable sigue una distribución de Bernoulli de parámetro p=0,2, de acuerdo con el enunciado. Suponiendo que las distintas reservas son independientes entre sí, se tiene que, de un total de n reservas (X_1, X_2....,X_n), el número de ellas que acuden finalmente al restaurante es una variable aleatoria

$$y_n = \sum_{i=1}^{n} = X_i$$

con distribución binomial de parámetros *n* y p=0,2. En el caso particular del problema, n=25. Entonces, para aquellas personas que asistan al restaurante de las 25 que han hecho la reserva puedan disponer de una mesa, debe ocurrir que acudan 20 o menos. Así se tiene que:

$$P(y \leq 20) \sum_{i=0}^{20} \binom{25}{0} \cdot 0.2^i \cdot (1 - 0.2)^{25-i} = 0,5799$$

Distribución Binomial

Cuando se realizan *n* pruebas de Bernouilli sucesivas e independientes, la variable aleatoria discreta X se denomina variable binomial cuando: X="número de veces que ocurre el suceso éxito en *n* pruebas" ~X B(n, p).

La función de probabilidad de esta distribución viene dada por:

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot q^{n-k}$$

Los parámetros de esta distribución de media, varianza y desviación estándar serán:

$$\mu_x = n \cdot p$$

$$\sigma_x^2 = n \cdot p \cdot q$$

$$\sigma_x = \sqrt{n \cdot p \cdot q}$$

Si el experimento consiste en extracciones sucesivas de una población finita, el proceso ha de ser con reemplazo para mantener la probabilidad de éxito a lo largo de todas los ensayos. Si X~B(n, p), cuando *n* es grande, y, p y q son distantes de cero, se puede considerar la aproximación a la distribución normal X~N (n.p; n.p.q)

$$P(X=k) = \binom{n}{k} \cdot p^k \cdot q^{n-k} \xrightarrow{n \cdot p \geq 5} \left(n.p;\ \sqrt{n.p.q}\right)$$

de esto se deduce que la variable:

$$z = \frac{X - n.p}{\sqrt{n.p.q}} \sim N(0, 1)$$

Cuando el tamaño *n* es grande y la probabilidad *p* es pequeña, la distribución de Poisson es una buena aproximación de la distribución binomial. Esto se cumple cuando $n \geq 30$ y $p \leq 0.1$

$$B(n, k) = \binom{n}{k} p^k\, q^{n-k} \xrightarrow{n.p<5} P(\lambda) = \frac{\lambda^k}{k!}\, e^{-\lambda}\ cuando\ \lambda = n.p$$

Ejemplo: Una empresa dedicada a la venta de artículos del hogar, ofrece a sus clientes dos formas de pago: a) pago al contado; b) pago en cuotas. Se sabe que el 30% de los artículos comprados fueron en forma de pago al contado. Si en un período de tiempo se adquirieron 6 unidades, determinar la probabilidad de que dos unidades o más hayan sido pagadas al contado.

$X = n^{o}$ *de unidades pagadas al contado*

$N = 6$ *unidades independientes*

$p = P(contado) = 0.3$

$q = P(cuotas) = 0.7$

$$P(X \geq 2) = 1 - P(X < 2) = 1 - P(X = 0) - P(x = 1)$$

$$P(X = 1) = \binom{6}{1} 0.3^{1} \cdot 0.7^{5} = 6 \cdot 0.3 \cdot 0.7^{5} = 0.3025$$

$$P(X = 0) = \binom{6}{0} 0.3^{0} \cdot 0.7^{6} = 1 \cdot 1 \cdot 0.7^{6} = 0.1176$$

$$P(X \geq 2) = 1 - 0.3025 - 0.1176 = 0.5798$$

Distribución de Poisson

Una variable X se dice que sigue una distribución de probabilidad de Poisson si puede tomar todos los valores enteros (0,1, 2, … *n*), con probabilidades

$$P(X = k) = \frac{\lambda^{k}}{k!} e^{-\lambda} \; cuando \; \lambda > 0$$

Los parámetros de esta distribución de media, varianza y desviación estándar serán:

$\mu_x = \lambda$

$\sigma_x^2 = \lambda$

$\sigma_x = \sqrt{\lambda}$

En este caso, X es el número de ocurrencias de un suceso durante un gran número de pruebas. λ es un parámetro positivo que representa el número de veces que se espera que ocurra el fenómeno durante un intervalo dado. Por ejemplo, si el suceso estudiado tiene lugar en promedio 4 veces por minuto y estamos interesados en la probabilidad de que ocurra k veces dentro de un intervalo de 10 minutos, usaremos un modelo de distribución de Poisson con λ = 10×4 = 40. La constante *e*, es la base de los logaritmos naturales (*e*=2,71828...).

La distribución de Poisson se ajusta a varios modelos experimentales, pero donde es más utilizada es cuando se espera que un evento se produzca en un determinado período de tiempo.

Si $X_i \sim P(\lambda_i)$ donde i= 1,2,...n variables aleatorias independientes de Poisson,

$$Y = \sum_{i=1}^{n} x_i \sim P\left(\sum_{i=1}^{n} \lambda_i\right)$$

Para cada valor t>0, que representa el tiempo, el número de sucesos de un fenómeno aleatorio sigue una distribución de Poisson de parámetro λt , los tiempos transcurridos entre dos sucesos sucesivos sigue una distribución exponencial. Cuando λ≥10 la distribución de Poisson se aproxima a una distribución normal $N = \left(\lambda,\ \sqrt{\lambda}\right)$

Ejemplo: en una carretera se producen un promedio de 2 accidentes anuales. Calcula la probabilidad de que este año se produzcan más de 3 accidentes.

$$parámetro\ \lambda = 2$$

$$p(X>3) = 1 - p(X \leq 3)$$

$$1 - \left[p(X=0) + p(X=1) + p(X=2) + p(X=3)\right] =$$

$$= 1 - \frac{e^{-2} \cdot 2^0}{0!} - \frac{e^{-2} \cdot 2^1}{1!} - \frac{e^{-2} \cdot 2^2}{2!} - \frac{e^{-2} \cdot 2^3}{3!} = 0.143$$

Distribución Geométrica (Pascal)

La distribución geométrica o de Pascal consiste en la realización sucesiva de pruebas de Bernouilli, donde la variable aleatoria discreta: X = número de la prueba en que aparece por primera vez el suceso A, donde X~G(p). Para hallar la función de probabilidad o cuantía P(X=k) hay que notar que la probabilidad del suceso es: k-1 (...A)

Por lo tanto, P(X=k) q^{k-1}. p

Los parámetros de esta distribución de media, varianza y desviación estándar serán:

$$\mu = \frac{1}{p}$$

$$\sigma^2 = \frac{q}{p^2}$$

$$\sigma = \frac{\sqrt{q}}{p}$$

La distribución geométrica es un modelo adecuado para aquellos procesos en los que se repiten pruebas hasta la consecución del resultado deseado.

Ejemplo: un matrimonio desea tener una hija, y por ello deciden tener planificar los embarazos hasta el nacimiento de la hija deseada. Calcular el número esperado de hijos (entre varones y mujeres, suponiendo que la posibilidad de que el nacimiento de sea la misma para cada género), que tendrá el matrimonio. Calcular la probabilidad de que la pareja acabe teniendo tres hijos o más.

$$P(X \geq 3) = 1 - p(X<3) = 1 - \left[p(X=1) + p(X=2)\right] = 1 - \left[0.5 + 0.5^2\right] = 1 - 0.75 = 0.25$$

Distribución Binomial Negativa

La distribución binomial negativa Bn(n, p) es un modelo adecuado para tratar procesos en los que se repite *n* veces una prueba determinada o ensayo hasta conseguir un número determinado k de resultados favorables (por primera vez). Si el número de resultados favorables buscados fuera 1 sería el caso de una distribución geométrica, esto es, la distribución binomial negativa puede considerarse una extensión o ampliación de la distribución geométrica. La variable aleatoria discreta X= número de pruebas necesarias para lograr k-éxitos, donde X~Bn(n, p).

Los parámetros de esta distribución de media, varianza y desviación estándar serán:

$$P(X=n) = \binom{n-1}{k-1} p^k \cdot q^{n-k}$$

$$\mu = \frac{k \cdot q}{p}$$

$$\sigma^2 = \frac{k \cdot q}{p^2}$$

$$\sigma = \sqrt{\frac{k \cdot q}{p^2}}$$

Ejemplo: Se sabe que la probabilidad de que un niño expuesto a una enfermedad contagiosa la contraiga es de 0,4. Calcula la probabilidad de que el décimo niño estudiado sea el tercero en contraer la enfermedad. Podemos enfocar el problema como una binomial negativa de parámetros X=10, k=3 y p=0,4.

$$P(X=10) = \binom{9}{2} \cdot 0.4^3 \cdot 0.6^7 = 0.0645$$

Distribución Polinomial o Multinomial

Es una generalización de la distribución binomial cuando en cada prueba se consideran k sucesos excluyentes (A_1, A_2,..., A_k), con probabilidades respectivas (p_1, p_2..., p_k), siendo p_1+p_2+...+ p_k= 1.

Suponiendo que se realizan *n* pruebas independientes de este tipo y considerando las variables X_i = número de veces que ocurre el suceso Ai en las pruebas, entonces:

$$P\left(X_1=n_1;\, X_2=n_2;\, \ldots;\, X_k=n_k\right) = \frac{n!}{n_1!\; n_2! \ldots n_k!}\, p_1^{n_1}\; p_2^{n_2} \ldots p_k^{n_k}$$

Ejemplo: según establece la nueva ley, una persona es donante de órganos salvo que exprese fehacientemente su decisión de no hacerlo. Se sabe que un 15% de las personas está en contra de la ley, un 40% es indiferente y un 45% está a favor. Si se obtiene una muestra aleatoria de 20 personas ¿cuál es la probabilidad de que 5 esten en contra, 10 sean indiferentes y 5 estén a favor?

$$X_1 = en\ contra = 0.15$$

$$X_2 = indif erente = 0.4$$

$$X_3 = a\ f avor\ 0.45$$

$$f\left(x_1=5;\, x_2=10;\, x_3=5\right) = \frac{20!}{5!\; 10!\; 5!} \cdot 0.15^5 \cdot 0.4^{10} \cdot 0.45^5 = 0.41$$

Distribución Hipergeométrica

Es una variante de la distribución binomial. La distribución hipergeométrica corresponde a extracciones sin reemplazamiento. En las demás cuestiones presenta el mismo marco de consideraciones que la distribución binomial, es decir, dos situaciones excluyentes (éxito y fracaso) que se realizan en n pruebas.

Sean N elementos, con la probabilidad de éxito p en la primera extracción. Los N elementos se distribuyen en N.p éxitos y N.q fracasos.

La variable aleatoria X= número de éxitos k en n extracciones, donde $X \sim H[n,N,N_A]$

Los parámetros de esta distribución de media, varianza y desviación estándar serán:

$$P(X=k)=\frac{\begin{bmatrix} N\cdot p \\ k \end{bmatrix}\begin{bmatrix} N\cdot q \\ n-k \end{bmatrix}}{\begin{bmatrix} N \\ n \end{bmatrix}}$$

$$\mu=n\cdot p$$

$$\sigma_x^2=n\cdot p\cdot q\ \frac{N-n}{N-1}$$

$$\sigma=\sqrt{n\cdot p\cdot q\ \frac{N-n}{N-1}}$$

Cuando N es grande respecto a n, es decir, n/N<0.1, la variable hipergeométrica se aproxima a una distribución binomial.

$$P(X=k)\xrightarrow{\frac{n}{N}<0,1}\binom{n}{k}p^k\cdot q^{n-k}$$

En general, de forma análoga a la distribución polinomial, en una población con N elementos repartidos en k clases excluyentes (A_1,A_2,...,A_k), con elementos respectivos de cada clase (N_1, N_2,...,N_k), N_1+ N_2+...+N_k=N, al tomar consecutivamente n elementos sin reemplazo y cuando X= número de elementos que hay de la clase Ai en la muestra de tamaño n, tenemos que:

$$P\left(X_1=n_1;X_2=n_2;\ldots;X_k=n_k\right)=\frac{\binom{N_1}{n_1}\cdot\binom{N_2}{n_2}\cdot\ldots\binom{N_k}{n_k}}{\binom{N}{n}}\ \textit{cuando}\ \begin{cases} N_1+N_2+\ldots+N_k=N \\ n_1+n_2+\ldots+n_k=n \end{cases}$$

Ejemplo: De cada 20 piezas fabricadas por una máquina, hay 2 que son defectuosas. Para realizar un control de calidad, se observan 15 elementos y se rechaza el lote si hay alguna que que presente alguna falla. Calcular la probabilidad de que un lote sea rechazado.

$N = 20$

$n = 15$

$X =$ *piezas defectuosas de las 15 escogidas*

$P(X \geq 1) = 1 - p(x < 1) = 1 - p(x = 0)$

$$1 - \frac{\binom{2}{0} . \binom{20-2}{15}}{\binom{20}{15}} = 1 - \frac{816}{15504} = \frac{18}{19} = 0.947$$