Algunas Distribuciones Discretas de Probabilidad

UCR – ECCI
CI-0115 Probabilidad y Estadística
Prof. Kryscia Daviana Ramírez Benavides

Introducción

- El comportamiento de una variable aleatoria queda descrito por su distribución de probabilidad sin importar si esta se representa de forma gráfica, en forma tabular o con una fórmula.
    - A menudo, las observaciones de diferentes experimentos estadísticos tienen el mismo tipo general de comportamiento.
    - En consecuencia, las variables aleatorias discretas asociadas se pueden describir con la misma distribución de probabilidad y se pueden representar mediante una sola fórmula.
- De hecho, se necesita sólo un conjunto de distribuciones de probabilidad importantes para describir muchas de las variables aleatorias discretas que se encuentran en la práctica.

Distribución Uniforme Discreta

- Es la más simple de todas las distribuciones discretas de probabilidad.
- Si la variable aleatoria X toma los valores x_1, x_2, …, x_k, con probabilidades idénticas, entonces la **distribución uniforme discreta** está dada por

$$f(x;k) = \frac{1}{k} \qquad x = x_1, x_2, \ldots, x_k$$

- La media y la varianza de la distribución uniforme discreta $f(x;k)$ son

$$\mu = \frac{\sum_{i=1}^{k} x_i}{k} \qquad \sigma^2 = \frac{\sum_{i=1}^{k} (x_i - \mu)^2}{k}$$

Distribución Binomial y Multinomial

- Un experimento a menudo consiste en pruebas repetidas, cada una con dos posibles resultados que se pueden etiquetar como éxito o fracaso.
- Este tipo de proceso se denomina **proceso de Bernoulli**. Cada ensayo se llama **experimento de Bernoulli**.

Distribución Binomial y Multinomial (cont.)

- **Propiedades del Proceso de Bernoulli**:
  - El experimento consiste en n pruebas que se repiten.
  - Cada prueba produce un resultado que se puede clasificar como éxito o fracaso.
  - La probabilidad de un éxito, que se denota con p, permanece constante en cada prueba.
  - Las pruebas que se repiten son independientes.

Distribución Binomial y Multinomial (cont.)

- El número X de éxitos en n experimentos de Bernoulli se denomina **variable aleatoria binomial**.
- La distribución de probabilidad de esta variable aleatoria discreta se llama **distribución binomial**, y sus valores se denotarán como $b(x;n,p)$, pues dependen del número de pruebas y de la probabilidad de éxito en una prueba dada.

Distribución Binomial y Multinomial (cont.)

- Un experimento de Bernoulli puede tener como resultado un éxito con probabilidad p y un fracaso con probabilidad $q = 1 - p$. Entonces la distribución de probabilidad de la variable aleatoria binomial X, el número de éxitos en n pruebas independientes, es

$$b(x; n, p) = \binom{n}{x} p^x q^{n-x} \qquad x = 0,1,2,\ldots,n$$

- La media y la varianza de la distribución binomial $b(x;n,p)$ son

$$\mu = np \qquad \sigma^2 = npq$$

Distribución Binomial y Multinomial (cont.)

- La distribución de probabilidad acumulada de la variable aleatoria binomial *X* es

$$B(r; n, p) = \sum_{x=0}^{r} b(x; n, p)$$

- La tabla A.1 del libro de texto brinda las sumas binomiales para $n = 1, 2, \ldots, 20$, y para valores seleccionados de p entre 0.1 y 0.9.

Distribución Binomial y Multinomial (cont.)

- Aplicaciones:
  - Campos científicos.
  - Ingeniería industrial.
  - Mediciones de control de calidad y esquemas de muestreo para procesos.
  - Aplicaciones médicas y militares.

Distribución Binomial y Multinomial (cont.)

- El experimento binomial se convierte en un **experimento multinomial** si cada prueba tiene más de dos resultados posibles.
- En general, si una prueba dada puede tener como consecuencia cualquiera de los k resultados posibles E_1, E_2, …, E_k con probabilidades p_1, p_2, …, p_k, entonces **distribución multinomial** dará la probabilidad de que E_1 ocurra x_1 veces; E_2 ocurra x_2 veces; …; y E_k ocurra E_k veces en n pruebas independientes, donde $x_1 + x_2 + \ldots + x_k = n$

Distribución Binomial y Multinomial (cont.)

- Se denota esta distribución de probabilidad conjunta como

$$f\left(x_1, x_2, \ldots, x_k; p_1, p_2, \ldots, p_k, n\right)$$

- Claramente, $p_1 + p_2 + \ldots + p_k = 1$, pues el resultado de cada prueba debe ser uno de los *k* resultados posibles.

Distribución Binomial y Multinomial (cont.)

- Si una prueba dada puede conducir a los k resultados posibles $E_1, E_2, \ldots, E_k$ con probabilidades $p_1, p_2, \ldots, p_k$, entonces la distribución de probabilidad de las variables aleatorias $X_1, X_2, \ldots, X_k$, que representan el número de ocurrencias para $E_1, E_2, \ldots, E_k$ en n pruebas independientes es

$$f(x_1, x_2, \ldots, x_k; p_1, p_2, \ldots, p_k, n) = \binom{n}{x_1, x_2, \ldots, x_k} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k}$$

con

$$\sum_{i=1}^{k} x_i = n \quad \sum_{i=1}^{k} p_i = 1$$

Distribución Hipergeométrica

- La manera más simple de ver la diferencia entre la distribución binomial y la distribución hipergeométrica está en la forma en que se realiza el muestreo.
- Los tipos de aplicaciones de la distribución hipergeométrica son muy similares a los de la binomial.
- Se interesa en el cálculo de probabilidades para el número de observaciones que caen en una categoría en particular. Pero a diferencia de la **binomial**, que son **pruebas independientes**, la **hipergeométrica** son **pruebas dependientes**.

Distribución Hipergeométrica (cont.)

- Como resultado, si se aplica la binomial a tomar muestras de un lote de artículos, el muestreo se debe efectuar **con reemplazo** de cada artículo después de que se observe. Por otro lado, la distribución hipergeométrica se basa en el muestreo que se realiza **sin reemplazo**.
- Aplicaciones:
  - Muestreo de aceptación.
  - Pruebas electrónicas.
  - Garantía de calidad.
- Para muchos de estos campos el muestreo se realiza a expensas del artículo que se prueba. Es decir, el artículo se destruye y por ello no se puede reemplazar en la muestra.

Distribución Hipergeométrica (cont.)

- En general, interesa la probabilidad de seleccionar x éxitos de los k artículos considerados como éxito y $n - x$ fracasos de los $N - k$ artículos que se consideran fracasos cuando se selecciona una muestra aleatoria de tamaño n de N artículos.
- Esto se conoce como **experimento hipergeométrico**; es decir, uno que posee las siguientes dos propiedades:
  - Se selecciona sin reemplazo una muestra aleatoria de tamaño n de N artículos.
  - Los k de los N artículos se pueden clasificar como éxitos y $N - k$ se clasifican como fracasos.

Distribución Hipergeométrica (cont.)

- El número *X* de éxitos de un experimento hipergeométrico se denomina **variable aleatoria hipergeométrica**.
- En consecuencia, la distribución de probabilidad de la variable hipergeométrica se llama **distribución hipergeométrica**, y sus valores se denotan como $h(x;N,n,k)$, debido a que dependen del número de éxitos k en el conjunto N del que seleccionamos n artículos.

Distribución Hipergeométrica (cont.)

- La distribución de probabilidad de la variable aleatoria hipergeométrica X, el número de éxitos en una muestra aleatoria de tamaño n que se selecciona de N artículos de los que k se denominan **éxito** y $N - k$ **fracaso**, es

$$h(x;N,n,k)=\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \qquad x=0,1,2,\ldots,n$$

- La media y la varianza de la distribución hipergeométrica $h(x;N,n,k)$ son $\mu=\frac{nk}{N}$ $\quad \sigma^2=\frac{N-n}{N-1}\cdot n\cdot\frac{k}{N}\left(1-\frac{k}{N}\right)$

Distribución Hipergeométrica (cont.)

- Hay una relación interesante entre la distribución hipergeométrica y la binomial.
- Como se podría esperar, si n es pequeño comparado con N, la naturaleza de los N artículos cambia muy poco en cada prueba.

$$\frac{n}{N} \leq 0.05$$

- Así la cantidad k/N juega el papel del parámetro binomial p.
- Como consecuencia, la distribución binomial se puede ver como una versión de población grande de las distribuciones hipergeométricas.

Distribución Hipergeométrica (cont.)

- La media y la varianza entonces se obtienen de las fórmulas

$$\mu = np = \frac{nk}{N} \qquad \sigma^2 = npq = n \cdot \frac{k}{N}\left(1 - \frac{k}{N}\right)$$

- Al comparar estas fórmulas, se puede observar que la media es la misma mientras que la varianza difiere por un factor de corrección $(N - n)/(N - 1)$, que es insignificante cuando n es pequeña en relación con N.

Distribución Hipergeométrica Multivariada

- La distribución hipergeométrica se puede extender para tratar el caso donde los N artículos se pueden dividir en k celdas A_1, A_2, …, A_k con a_1 elementos en la primera celda, a_2 elementos en la segunda celda, …, y a_k elementos en la k-ésima celda. Se interesa ahora en la probabilidad de que una muestra aleatoria de tamaño n dé x_1 elementos de A_1, x_2 elementos de A_2,…, y x_k elementos de A_k. Esta probabilidad se representa por

$$f\left(x_1, x_2, \ldots, x_k; a_1, a_2, \ldots, a_k, N, n\right)$$

Distribución Hipergeométrica Multivariada (cont.)

- Si N artículos se pueden dividir en las k celdas $A_1, A_2, \ldots, A_k$ con $a_1, a_2, \ldots, a_k$ elementos, respectivamente, entonces la distribución de probabilidad de las variables aleatorias $X_1, X_2, \ldots, X_k$, que representan el número de elementos que se seleccionan de $A_1, A_2, \ldots, A_k$ en una muestra aleatoria de tamaño n, es

$$f(x_1, x_2, \ldots, x_k; a_1, a_2, \ldots, a_k, N, n) = \frac{\binom{a_1}{x_1}\binom{a_2}{x_2}\cdots\binom{a_k}{x_k}}{\binom{N}{n}}$$

con

$$\sum_{i=1}^{k} x_i = n \quad \sum_{i=1}^{k} a_i = N$$

Distribuciones Binomial Negativa y Geométrica

- Se considera un experimento donde las propiedades son las mismas que las que se indican para un experimento binomial, con la excepción de que las pruebas se repetirán hasta que ocurra un número **fijo** de **éxitos**.
- Por lo tanto, en lugar de encontrar la probabilidad de x éxitos en n pruebas, donde n es fija, interesa la probabilidad de que ocurra el k-ésimo éxito en la x-ésima prueba.
- Los experimentos de este tipo se llaman **experimentos binomiales negativos**.

Distribuciones Binomial Negativa y Geométrica (cont.)

- El número X de pruebas que produce k en un experimento binomial negativo se llama **variable aleatoria binomial negativa** y su distribución de probabilidad se llama **distribución binomial negativa**.
- Como sus probabilidades dependen del número de éxitos que se desean y la probabilidad de un éxito en una prueba dada, se denotará con $b^*(x;k,p)$.

Distribuciones Binomial Negativa y Geométrica (cont.)

- Si pruebas independientes repetidas pueden tener como resultado un éxito con probabilidad p y un fracaso con probabilidad $q = 1 - p$, entonces la distribución de probabilidad de la variable aleatoria X, el número de la prueba en la que ocurre el k-ésimo éxito, es

$$b^*(x;k,p) = \binom{x-1}{k-1} p^k q^{x-k} \qquad x = k, k+1, k+2, \ldots$$

- La media y la varianza de la distribución binomial negativa $b^*(x;k,p)$ son

$$\mu = \frac{k}{p} \qquad \sigma^2 = \frac{kq}{p^2}$$

Distribuciones Binomial Negativa y Geométrica (cont.)

- En teoría de probabilidad y estadística, la **distribución geométrica** es cualquiera de las dos distribuciones de probabilidad discretas siguientes:
  - La distribución de probabilidad del número X del experimento de Bernoulli necesaria para obtener un éxito, cuando $x = 1, 2, 3, \ldots$
  - La distribución de probabilidad del número $Y = X - 1$ de fallos antes del primer éxito, cuando $x = 0, 1, 2, 3, \ldots$

Distribuciones Binomial Negativa y Geométrica (cont.)

- Si se considera el caso especial de la distribución binomial negativa donde $k = 1$, se tiene una distribución de probabilidad para el número de pruebas que se requieren para un solo éxito. La distribución binomial negativa se reduce a la forma

$$b^*(x;1,p) = pq^{x-1} \qquad x = 1,2,3,\ldots$$

- Como los términos sucesivos constituyen una progresión geométrica, se acostumbra referirse a este caso especial como la **distribución geométrica** y denotar sus valores con $g(x;p)$.

Distribuciones Binomial Negativa y Geométrica (cont.)

- Si pruebas independientes repetidas pueden tener como resultado un éxito con probabilidad p y un fracaso con probabilidad $q = 1 - p$, entonces la distribución de probabilidad de la variable aleatoria X, el número de la prueba en el que ocurre el primer éxito, es

$$g(x; p) = pq^{x-1} \quad x = 1,2,3,\ldots$$

- La media y la varianza de una variable aleatoria que sigue distribución geométrica son

$$\mu = \frac{1}{p} \quad \sigma^2 = \frac{q}{p^2}$$

Distribuciones Binomial Negativa y Geométrica (cont.)

- Aplicaciones:
  - Binomial negativa:
    - Similares a la naturaleza.
    - Los intentos son costosos en algún sentido y ocurren en sucesión; para ver si hay un alta probabilidad de requerir un número grande de intentos para experimentar un número fijo de éxitos, ya que no es benéfico para el científico o ingeniero.
  - Geométrica:
    - Las pruebas ocurren antes de que un éxito represente un costo.
    - Si hay una alta probabilidad de hacer varios intentos antes del enlace, entonces se deben hacer planes para rediseñar el sistema o proceso.

Distribución de Poisson y Proceso de Poisson

- Los experimentos que dan valores numéricos de una variable aleatoria X, el número de resultados que ocurren durante un intervalo dado o en una región específica, se llaman **experimentos de Poisson**.
- El intervalo dado puede ser de cualquier longitud, como un minuto, un día, una semana, un mes o incluso un año.
- La región específica podría ser un segmento de línea, un área o quizá una pieza de material.

Distribución de Poisson y Proceso de Poisson (cont.)

- Un experimento de Poisson se deriva del proceso de Poisson y posee las siguientes propiedades:
  - El número de resultados que ocurren en un intervalo o región específica es independiente del número que ocurre en cualquier otro intervalo o región del espacio disjunto. No tiene memoria.
  - La probabilidad de que ocurra un solo resultado durante un intervalo muy corto o en una región pequeña es proporcional a la longitud del intervalo o al tamaño de la región, y no depende del número de resultados que ocurren fuera de este intervalo o región.
  - La probabilidad de que ocurra más de un resultado en tal intervalo corto o que caiga en tal región pequeña es insignificante.

Distribución de Poisson y Proceso de Poisson (cont.)

- El número X de resultados que ocurren durante un experimento de Poisson se llama **variable aleatoria de Poisson** y su distribución de probabilidad se llama **distribución de Poisson**.
- El número medio de resultados se calcula de $\mu = \lambda t$, donde t es el tiempo o región específico de interés.
- Como sus probabilidades dependen de λ, la tasa de ocurrencia de los resultados, se denota con $p(x;\lambda t)$.

Distribución de Poisson y Proceso de Poisson (cont.)

- La distribución de probabilidad de la variable aleatoria de Poisson X, que representa el número de resultados que ocurren en un intervalo dado o región específica que se denota con t, es

$$p(x;\lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!} \quad x = 0,1,2,\ldots$$

- Donde λ es el número promedio de resultados por unidad de tiempo o región y $e = 2.71828\ldots$
- La media y la varianza de la distribución de Poisson $p(x;\lambda t)$ tienen el valor λt.

Distribución de Poisson y Proceso de Poisson (cont.)

- La distribución de probabilidad acumulada de la variable aleatoria Poisson X es

$$P(r;\lambda t)=\sum_{x=0}^{r} p(x;\lambda t)$$

- La tabla A.2 del libro de texto brinda la suma de probabilidad de Poisson para algunos valores selectos de λt que van de 0.1 a 18.

Distribución de Poisson y Proceso de Poisson (cont.)

- Aplicaciones:
  - Control de calidad.
  - Seguro de calidad.
  - Muestreo de aceptación.
  - Ciertas distribuciones continuas importantes que se usan en la teoría de confiabilidad y teoría de colas depende del proceso de Poisson.

Distribución de Poisson y Proceso de Poisson (cont.)

- Aunque el proceso de Poisson por lo general encuentra aplicaciones en problemas de espacio y tiempo, se puede ver como una forma limitante de la distribución binomial.
- En el caso de la binomial, si n es bastante grande y p es pequeña, las condiciones comienzan a simular las implicaciones de **espacio continuo** o **región temporal** del proceso de Poisson.

Distribución de Poisson y Proceso de Poisson (cont.)

- La independencia entre las pruebas de Bernoulli en el caso binomial es consistente con la propiedad 2 del proceso de Poisson.
- Si se hace al parámetro p cercano a cero se relaciona con la propiedad 3 del proceso de Poisson.
  - Si p es cercana a 1, aún se puede utilizar la distribución de Poisson para aproximar probabilidades binomiales mediante el intercambio de lo que definimos como éxito o fracaso, se cambian los valores de p y q.

Distribución de Poisson y Proceso de Poisson (cont.)

- Distribución de Poisson como forma limitante de la binomial, teorema:
  - Sea X una variable aleatoria binomial con distribución de probabilidad $b(x;n,p)$. Cuando $n \to \infty$, $p \to 0$, y $\mu = np$ permanece constante, $b(x;n,p) \to p(x;\mu)$.

$$b(x;n,p) \to \frac{e^{-\mu}\mu^{x}}{x!} \qquad x = 0,1,2,\ldots$$

Referencias Bibliográficas

- Walpole, R.E.; Myers, R.H.; Myers, S.L. & Ye, K. "Probabilidad y estadística para ingeniería y ciencias". Octava Edición. Pearson Prentice-Hall. México, 2007.